Genetically encoded multivalent sensors to detect bivalent epigenetic modifications in living stem cells

THÈSE N^{O} 8041 (2017)

PRÉSENTÉE LE 24 NOVEMBRE 2017
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE CHIMIE BIOPHYSIQUE DES MACROMOLÉCULES
PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Aurore Marie-France DELACHAT

acceptée sur proposition du jury:

Dr A.-S. Chauvin, présidente du jury
Prof. B. Fierz, directeur de thèse
Prof. D. Schwarzer, rapporteur
Prof. W. Herr, rapporteur
Prof. C. Heinis, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

Abstract

Eukaryotic DNA is organized in the form of chromatin whose basic unit is the nucleosome. The four core histones forming the nucleosome, H2A, H2B, H3 and H4 can be highly post-translationally modified, especially on their N-terminal tail protruding from the nucleosome particle. Histone post-translational modifications (PTMs) work combinatorially to establish chromatin states defined by specific gene expression status. Found in embryonic stem cells (ESCs) at promoters of key developmental genes, bivalent chromatin is the combination of the active chromatin mark, trimethylation of lysine 4 on histone H3 (H3K4me3) and the repressive mark, H3K27me3. Established and maintained by Polycomb (Pc) and Trithorax (Trx) proteins, bivalency is proposed to keep gene transcription repressed but poised for activation. How bivalent domains are organized within the nucleus and how they are installed by Pc and Trx are still unknown. In this work, we aim to answer these questions by designing probes that enable live cell imaging of bivalent domains and by studying the installation and removal of H3K4 methylation on nucleosomes.

The current lack of live cell imaging methods for PTM patterns prompted us to engineer genetically encoded sensors that bind to bivalent marks in a multivalent fashion. These sensors contain a fluorescent protein and two reader domains, joined by flexible linkers. Their selectivity for bivalent nucleosomes was tested in a pulldown assay with a library of differently modified reconstituted nucleosomes. To this end, we obtained site-specifically modified histones via Expressed Protein Ligation (EPL). The best probe was then applied in live ESCs to visualize bivalent domains. Subsequent imaging by confocal microscopy revealed the organization of bivalent chromatin into discrete and local clusters. Furthermore, this probe was employed to monitor loss of bivalency upon treatment with a small molecule epigenetic inhibitor.

Then we studied the histone-lysine N-methyltransferase SET1B and lysine-specific histone demethylase 1A (LSD1), enzymes that install and remove methyl groups on H3K4, respectively. We measured a mid-micromolar affinity for LSD1 to reconstituted nucleosomes using microscale thermophoresis. This interaction might have an impact on the recruitment of LSD1 to its target genes. We then measured the

activity of SET1B complex on symmetrically modified H3K27me3 nucleosomes and on asymmetrically modified H3K4me3 nucleosomes. We showed that H3K27me3 does not influence the activity of SET1B complex, whereas H3K4me3 activates SET1B-mediated installation of methyl groups at K4 on the opposite H3 tail. This findings might have important implications concerning the generation of bivalent chromatin.

Together these results gave insights about multivalent binding of tandem reader domains, subnuclear organization of bivalent chromatin and installation of bivalent marks. In the future, we envisage to develop multivalent sensors for other PTM patterns of biological interest.

Key words: histone post-translational modifications, expressed protein ligation, bivalent chromatin, embryonic stem cells, reader domains, genetically encoded sensors, multivalent binding, site-specifically modified nucleosomes, LSD1, SET1B

Résumé

L'ADN eucaryote est organisé sous la forme de chromatine dont l'unité de base est le nucléosome. Les quatre histones formant le nucléosome peuvent porter des modifications post-traductionnelles, principalement sur leur queue N-terminale. Ces modifications fonctionnent de manière combinatoire pour établir des états distincts de chromatine définis par leur statut d'expression des gènes. La chromatine bivalente, trouvée dans les cellules souches embryonnaires aux promoteurs de gènes importants pour le développement, est la combinaison de la marque active tri-méthylation de la lysine 4 sur l'histone H3 (H3K4me3) et de la marque répressive H3K27me3. La bivalence est établie et maintenue par les protéines Polycomb (Pc) et Trithorax (Trx). Il a été proposé que la chromatine bivalente garde les gènes réprimés mais prêts à être activés. A ce jour, la manière dont les domaines bivalents sont organisés dans le noyau et la façon dont ils sont installés par Pc et Trx demeurent inconnues. Dans cette thèse, nous visons à répondre à ces questions en développant des senseurs qui permettent d'observer les domaines bivalents dans les cellules vivantes et en étudiant l'installation et le retrait des groupes méthyle sur H3K4.

Le manque actuel de méthodes pour imager les combinaisons de modifications dans les cellules vivantes nous a poussés à concevoir des senseurs génétiquement encodés qui reconnaissent les marques bivalentes de manière multivalente. Ces senseurs contiennent une protéine fluorescente et deux domaines « lecteur » joints par des connecteurs flexibles. Leur sélectivité pour des nucléosomes bivalents a été testée grâce à une bibliothèque de nucléosomes portant différentes modifications. Pour cela, nous avons obtenu des histones avec un (plusieurs) site(s) spécifiquement modifié(s) à l'aide de la réaction de ligature de protéines exprimées (EPL). Le meilleur senseur a ensuite été utilisé dans des cellules souches embryonnaires vivantes pour visualiser les domaines bivalents. Nous avons ensuite observé les cellules à l'aide d'un microscope confocal et révélé que la chromatine bivalente est organisée sous forme de grappes.

Par la suite, nous avons étudié les enzymes histone-lysine N-méthyl transférase SET1B et lysine-spécifique histone démethylase 1A (LSD1) qui respectivement dépose et enlève les groupes méthyle sur H3K4. Nous avons utilisé la thermophorèse à micro-

échelle pour mesurer une affinité micro-molaire entre LSD1 et des nucléosomes. Ensuite, nous avons évalué l'activité du complexe SET1B sur des nucléosomes portant H3K27me3 de manière symétrique et sur des nucléosomes portant H3K4me3 de manière asymétrique. Nous avons montré que H3K27me3 n'a pas d'influence sur l'activité de SET1B alors que H3K4me3 active la mise en place de groupes méthyle par SET1B sur la queue opposée.

En tout, ces résultats donnent des informations cruciales sur l'organisation sous-nucléaire de la chromatine bivalente et sur l'installation des marques bivalentes. A l'avenir, nous envisageons de développer des senseurs multivalents ciblant d'autres combinaisons de modifications.

Mots clés: modifications post-traductionnelles des histones, ligation de protéines exprimées, chromatine bivalente, cellules souches embryonnaires, domaines « lecteur », senseurs génétiquement encodés, liaison multivalente, nucléosomes spécifiquement modifiés, LSD1, SET1B

Acknowledgments

I greatly thank my thesis director Prof. Beat Fierz for the opportunity of joining LCBM. His persistence and enthusiasm have been helpful and inspiring for me. I acquired countless skills and have become an accomplished scientist thanks to him.

I thank the jury members Prof. Christian Heinis, Prof. Dirk Schwarzer and Prof. Winship Herr for taking time to read and review this thesis. I thank Dr. MER Anne-Sophie Chauvin for being the president of the jury.

Several other persons were involved, scientifically or not, in this thesis.

I would like to thank all current and former members of LCBM for making this lab a great professional environment. I thank Horst Pick and Carolin Lechner with whom the collaboration on the multivalent sensor project was fruitful. I thank warmly Sinan Kilic who was always available for my questions. Thanks a lot to Iuliia Boichenko for her kind advices about my thesis. I particularly appreciated the scientific and non-scientific support of Andreas Bachmann. I enjoyed the company of Ninad Agashe, Maxime Mivelaz and Eduard Ebberink for passionate discussions about random things (including science!). Many thanks to Louise Bryan for her English expertise on the thesis and for her support during hard times. I thank a lot Nora Guidotti for her help in the cMAP project and for her continual positivity. I wish to Anne, Karthik and Harsh to enjoy their time at LCBM as much as I did.

A special thank for Ruud Hovius for his support as my mentor.

I thank Cédric Deluz for his kind advices and for lending me some chemicals for the ESC culture. I want to thank Antonio C. A. Meireles-Filho for his willpower and efforts for the ChIP experiments. I thank Marie Munoz, Yoann Dind, Marie Jirousek and Jacques Gremaud for their help and positive attitude at all times.

During my time as a PhD student, I was given the opportunity of being assistant for Nora Guidotti, Oscar Pundel and Yang Liu. Additionally, I did different teaching tasks such as TPs and exercises. Throughout these experiences, I gained self-confidence and I am especially thankful to Beat Fierz, Anne-Sophie Chauvin and Nora Guidotti for that.

Many thanks to Lucas, Coraly, Ismaël, Céline and Sabine for their longstanding friendship.

I am thankful to my family for their support during my doctoral studies. Last but not least, I greatly thank Loris for encouraging me during hard times and for sharing my enthusiasm about science.

Abbreviations

(G)MEM	(Glasgow) Minimal essential medium
(m)ESC	(Mouse) embryonic stem cells
(q)PCR	(Quantitative) Polymerase chain reaction
(t)/(m)RNA	(Transfer)/(messenger) ribonucleic acid
3C	Chromatin conformation capture
(^{3}H-)SAM	S-adenosylmethionine (labelled with tritium)
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ac	Acetyl
Acetyl-CoA	Acetyl coenzyme A
AEBP2	Adipocyte enhancer-binding protein 2
AIEX	Anion exchange chromatography
AKT	RAC-alpha serine/threonine-protein kinase
AOD	Amine oxidase domain
ASC-2	Apoptosis-associated speck-like protein 2
ASCOM	ASC-2/NCOA6 complex
ASH2(L)	Absent small and homeotic disks protein (like)
ATP	Adenosine triphosphate
BET	Bromo- and extra-terminal domain
BHC80	BRAF35-HDAC complex protein 80
BiFC	Bimolecular fluorescence complementation
Boc	*Tert*-butoxycarbonyl
bp	Base pair
BPTF	Bromodomain PHD finger transcription factor
BRD3	Bromodomain-containing protein 3
BSA	Bovine serum albumin
C. Elegans	*Caenorhabditis Elegans*
CBX	Chromobox protein homolog
CD	Chromodomain
CFP1	Cyclin fold protein 1
CGI	CpG island
(Ch)IP(-seq)	(Chromatin) immunoprecipitation (sequencing)
cMAP	Chromatin sensing multivalent probe
CMV	Cytomegalovirus

COMPASS	Complex proteins associated with SET1 protein
CoREST	REST (RE1-silencing transcription factor) Corepressor
CpG	Cytosine-phosphate-guanine
CSC	Cancer stem cells
Cy5	Indodicarbocyanine/ Cyanine5
DAPI	4',6-diamidino-2-phenylindole
Dbz	Diamino benzoic acid
DCC	*N-N'*-dicyclohexylcarbodiimide
DIPEA	*N*-ethyldiisopropylamine
DMF	Dimethylformamide
DNA	Desoxyribonucleic acid
DNMT	DNA methyltransferase
Drosophila	*Drosophila Melanogaster*
DTNB	5,5′-Dithiobis(2-nitrobenzoic acid)
DTT	DL-Dithiothreitol
DYP	Protein dumpy
E. Coli	*Escherischia Coli*
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic ectoderm development protein
EGTA	Glycol ether diamine tetraacetic acid
EMSA	Electromobility shift assay
EPL	Expressed Protein Ligation
ESI-MS	Electrospray ionization mass spectrometry
EZH	Enhancer of zeste homolog
FAD	Flavine adenine dinucleotide
FISH	Fluorescent *in situ* hybridization
Fmoc	Fluorenylmethoxycarbonyl
FP	Fluorescent protein
FRET	Fluorescence resonance energy transfer
FZ	Frizzled
GdmCl	Guanidine hydrochloride
GSK3	Glycogen synthase kinase-3
H3, H4, H2A, H2B, H1, H5	Histone proteins
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate
HBTU	*N,N,N′,N′*-Tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate

HDAC	Histone deacetylase
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HMTase	Histone methyltransferase
HP1	Heterochromatin protein 1
ICM	Inner cell mass
IgG	Immunoglobulin G
ING2	Inhibitor of growth protein 2
iPSC	Induced pluripotent stem cells
IPTG	Isopropyl-β-D-thiogalactoside
JAK	Janus kinase
JARID2	Jumonji/ARID domain-containing protein 2
Kac	Acetyl lysine
KAHA	α-Ketoacid-Hydroxylamine
KDM	Lysine-specific demethylase
Klf4	*Krueppel-like factor 4*
Kme	Methyl lysine
KMT	Lysine methyltransferase
KUb	Ubiquitin lysine
LB	Luria-Bertani
LIF	Leukaemia inhibitory factor
LRP	Low-density lipoprotein receptor-related protein
LSD1	Lysine-specific histone demethylase 1A
MAPK/Erk	Mitogen-activated protein kinase
me	Methyl
MEF	Murine embryonic fibroblast
MEK	Dual specificity mitogen-activated protein kinase kinase
MESNa	2-Mercaptoethanesulfonic acid sodium
MLL	Mixed-lineage leukaemia
MPAA	4-mercaptophenylacetic acid
MS	Mass spectrometry
MST	Microscale Thermophoresis
MTG	Methyl thioglycolate
MWCO	Molecular weight cutoff
MYPT1	Myosin phosphatase-targeting subunit 1
NaF	Sodium fluoride
$NaNO_2$	Sodium nitrite
NCL	Native chemical ligation
Ni-NTA	Nickel nitrilotriacetic acid

NLS	Nuclear localisation signal
NURF	Nucleosome-remodelling factor
OCT4	Octamer-binding transcription factor 4
OSN	OCT4, SOX5, NANOG
PAGE	Polyacrylamide gel electrophoresis
PB	Polar bodies
Pbf	Pentamethyldihydrobenzofuran-5-sulfonyl
PBS	Phosphate buffer saline
Pc(G)	Polycomb (group)
PCGF	Polycomb group RING finger protein
PCL	Polycomblike protein
PHC	Polyhomeotic-like protein
PHD	Plant homeodomain
PI3K	Phosphoinositide 3-kinase
PMSF	Phenylmethylsulfonyl fluoride
PN	Pronuclei
Pol	Polymerase
PRC	Polycomb repressive complex
PTM	Post-translational modification
PVDF	Polyvinylidene difluoride
RBAP(RBBP)	Retinoblastoma-binding protein
Re-ChIP	Sequential chromatin immunoprecipitation
RING	Really Interesting New Gene
Rme	Methyl arginine
RP-HPLC	Reverse phase high pressure liquid chromatography
RT	Room temperature
RYBP	RING1 and YY1-binding protein
SAHA	Suberoylanilide hydroxamic acid
SAM	S-adenosylmethionine
SDS	Sodium dodecyl sulphate
SEA	Bis(2-sulfanylethyl) amido
SEC	Size exclusion chromatography
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SOX2	Sry box-containing gene 2
Sphos	Phosphorylated serine
SPPS	Solid phase peptide synthesis
STAT3	Signal transducer and activator of transcription 3
Su(var)	Suppressor of position-effect variegation

SUMO	Small ubiquitin-related modifier
SUZ12	Suppressor of zeste 12 protein
SWI/SNF	SWItch/Sucrose Non-Fermentable
SWIRM	Swi3p, Rsc8 and Moira
TAD	Topologically associated domain
TAF1/3	Transcription initiation factor TFIID subunit 1/3
TBS	Tris buffered saline
Tbx3	T-box transcription factor
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TCP	Tranylcypromine
TE	Trophectoderm
TEV	Tobacco etch virus
TF	Transcription factor
TFA	Trifluoroacetic acid
TFET	2,2,2-trifluoroethanethiol
TIS	Triisopropylsilane
Trr	Trithorax-related
Trx(G)	Trithorax (group)
TSS	Transcription start site
Uaa	Unnatural amino acid
Ub	Ubiquitin
ULP1	Ubiquitin-like-specific protease 1
UTX	Ubiquitously transcribed X chromosome tetratricopeptide repeat protein
WB	Western Blot
WD40	Tryptophan Aspartate 40 repeat
WDR	WD repeat-containing protein
WRAD	WRD5, RBBP5, ASH2L, DPY30

Table of contents

Chapter 1. Introduction

1.1 Gene expression regulation and chromatin in eukaryotes

1.1.1 Nuclear DNA under the form of chromatin

The structure of chromatin

The human genome with its ~ 6 billion DNA base pairs (bp) is about four meters in length and encodes for all genetic information required for development and maintenance of body functions throughout its lifetime. The nucleus of a human cell is in the range of 5 - 10 μm in diameter and yet contains all chromosomes. Nuclear DNA is organized in the form of chromatin, which compacts it and tightly controls access of the transcription machinery to genes. Chromatin, the nucleoprotein complex composed of DNA and histone proteins, adopts different conformations during the cell cycle. Chromatin takes the form of highly compact mitotic chromosomes during mitosis, whereas during interphase chromatin appears in a more relaxed form, organized in subnuclear chromosome territories. Chromatin has several layers of organization (Figure 1). The nucleosome, the basic unit of chromatin, is a pseudo-symmetric structure composed of 147 bp of DNA wrapped around a histone octamer in 1.65 turns of a left-handed helix. The first high resolution crystal structure of the nucleosome core particle was determined by the Richmond group in 1997.[1] The histone octamer is composed of two copies of the four core canonical histones H2A, H2B, H3, H4 or their variants (H2A.X for instance). Nucleosomes cover the whole genome with a repeat length of ~ 160 - 200 bp even though some portions of the genome are richer or poorer in nucleosomes.[2] A nucleosome array represents the structure known as "beads-on-a-string" first described by Don and Ada Olins in 1974.[3] Nucleosome arrays can further fold into 30 nm chromatin fibres;[4] the structure of these fibres has been of longstanding debate among the scientific community. Two models are proposed: the two-start helix model with zigzag conformation was initially proposed by Woodcock et al.[5] and the solenoid one-start helical conformation was first proposed by Finch and Klug in 1976.[4] Tetranucleosome

structures and other studies support the two-start helical model,[6–9] whereas studies by the Rhodes group support the solenoid model.[10,11] The 30 nm fibre is however a short range organizational motif, as over long length scales chromatin appears highly heterogeneous in interphase and mitotic cells.[12,13]

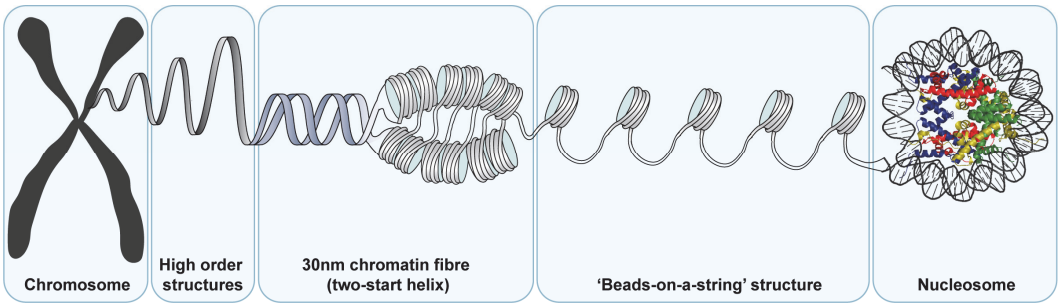

Figure 1. Chromatin organization. From the level of chromosomes to nucleosomes, chromatin is organized into different structures. From left to right: schematic view of a mitotic chromosome, high order structures, two-start helical 30 nm chromatin fibre, nucleosome array with "beads-on-a-string" conformation, ribbon representation of a human nucleosome core particle (PDB: 3AFA) with 146 bp double stranded DNA shown in grey, histone H3.1, H4, H2A, H2B represented in blue, red, yellow and green, respectively. Unstructured N-terminal histone tails are not displayed here.

Epigenetics: discovery and definition

Within eukaryotes, all cells composing an organism carry the same genetic information, although they differ in phenotype. Each cell type is characterized by a particular gene expression profile that defines cell function, morphology and location within the body. For proper functioning of the organism, a tight control of gene expression is required. The study of the biological mechanisms that control gene expression is known as the field of epigenetics, literally meaning "*over the genetics*". The term of "epigenetics" was first used by Waddington in 1939 for whom it referred to the link between genes and development.[14,15] This term has evolved and nowadays an epigenetic trait is considered as "a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence".[16]

The three following phenomena are generally considered as epigenetic mechanisms: DNA methylation (and other DNA modifications) (Figure 2. A), histone post-translational modifications (PTMs) (detailed in 1.1.2) (Figure 2. B) and RNA-mediated silencing. Historically, the first phenomenon proposed as epigenetic

was DNA methylation,[17] which is known to repress gene transcription and is involved in gene imprinting.[18,19] Histone PTMs (or marks) encompass a large diversity of chemical modifications and can be found on all four core histones. In 2015, more than 540 different histone PTMs were reported.[20] Histone marks are involved in a variety of chromatin-associated events such as transcription, genome stability, chromatin compaction, DNA replication and repair.[21] Finally, RNA-mediated silencing is based on sequence complementarity with the messenger RNA and is involved in X-chromosome inactivation.[22,23] Epigenetic phenomena are involved in many crucial cellular processes such as differentiation, mitosis, apoptosis and are involved in cancer as well as in mental disorders.[24,25]

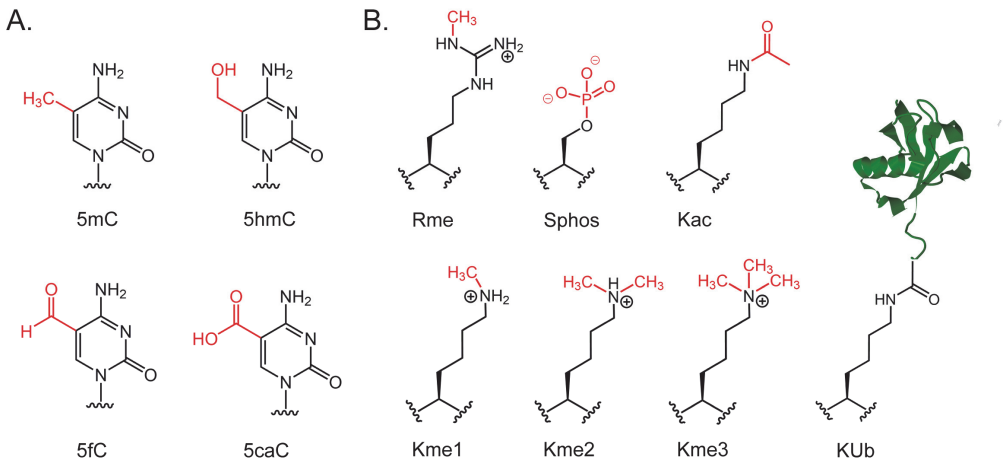

Figure 2. DNA and histone modifications are epigenetic marks. A. Chemical structures of DNA methylation (5mC: 5-methylcytosine) and other DNA modifications (5hmC: 5-hydroxymethylcytosine; 5fC: 5-formylcytosine; 5caC: 5-carboxylcytosine). B. Chemical structures of some histone post-translational modifications (Rme: Arginine monomethylation; Sphos: Serine phosphorylation; Kac: Lysine acetylation; Kme1/2/3: Lysine mono-, di- and trimethylation; KUb: Lysine ubiquitination).

Epigenetic inheritance

The definition of epigenetics contains the notion of inheritance through mitosis and meiosis. However, unlike DNA sequence, histone and DNA modifications are labile, i.e. they are reversibly deposited and removed from their substrate by specific enzymes (this topic is detailed in 1.1.2). This raises the question if histone and DNA modifications are indeed stably inherited during cell division and across multiple generations. Evidence that the DNA methyltransferase DNMT1 is

preferentially active on hemi-methylated DNA (DNA methylated on one strand only) supports the hypothesis that DNMT1 installs methyl groups on the newly synthesized DNA strand at methylated positions of the parental strand.[26,27] This provides a mechanism for mitotic DNA methylation inheritance. However, *de novo* DNA methylation by DNMT3A/B occurs as well and is particularly important during development.[28,29] Regarding histone modifications, several studies showed that methylation at Lysine 9 and 27 of histone H3 (H3K9me and H3K27me) are inherited during cell division or even from parents to offspring.[30–32] It is however unclear which and to what extent histone modifications are stably inherited.

Inheritance through multiple generations is assumed to be mainly driven by non-coding RNA.[33–35] DNA methylation at imprinted genes is also an example of transgenerationally inherited epigenetic marks.[18] However the transgenerational inheritance of histone marks is under debate. During spermatogenesis, most of the histones are replaced by protamines,[36] limiting but not abolishing paternal transmission of histone marks to offsprings.[37] Although oocytes and zygotes undergo major epigenetic reprogramming,[38] maternal transgenerational inheritance of H3K9me3 exists at pericentric heterochromatin.[39] Transgenerational inheritance of epigenetic modifications was first proved in 1999 by Morgan et al. at the agouti locus in mouse.[40] More recently, Greer et al. and Gu et al. showed transgenerationally inherited epigenetic footprints in *C. Elegans* for H3K4me3 and H3K9me3, respectively.[41,42] In *Drosophila* and mouse, paternal transgenerational inheritance of gene expression reprogramming or of chromatin state reprogramming was highlighted by the use of a specific paternal diet.[43,44]

1.1.2 Histone post-translational modifications regulate gene expression

Histone modifications and their functions

Histones are highly conserved chromatin-associated proteins that can be categorized as core histones (H3, H4, H2A and H2B) or linker histones (H1 and H5). Core and linker histones can be further classified as canonical histones (expressed

during S phase and cover the majority of the genome) or histone variants (can substitute for the canonical histones). All four core histones can be post-translationally modified, mainly on their unstructured N-terminal tail protruding from the nucleosome, but also on their globular domain. Examples of histone PTMs include lysine acetylation, ubiquitination, mono-, di- and trimethylation, arginine mono- and dimethylation, serine phosphorylation (Figure 2. B). Histone marks control gene expression via a number of mechanisms such as chromatin compaction or decompaction, recruitment of protein effectors, inhibition or activation of chromatin modifiers and remodelers.[21] Histone modifications are differently distributed along chromosomes and their level vary during cell cycle.[45] Some histone marks are known to be associated with euchromatin, a chromatin type characterized by a decondensed state and active gene transcription. Active chromatin marks include H3K4 di- and trimethylation (present at active promoters) and acetylation of H3 and H4.[46] Conversely, other histone PTMs are associated with heterochromatin, a compact chromatin type where genes are silenced. The repressive chromatin mark H3K9me3 is characteristic of constitutive heterochromatin, a highly condensed HP1-associated chromatin state found at gene-poor regions. H3K27me3 is the hallmark of facultative heterochromatin, a Polycomb-associated chromatin state. Histone PTMs can act via *trans* or *cis* effects. In *cis*, the covalent modification changes *per se* the compaction state and/or the dynamics of chromatin. Examples of marks having a *cis* effect include H4K20me3, which stabilizes a compact nucleosome fibre,[47] H2BK120Ub[48] and H4K16ac,[49] which open chromatin arrays. On the contrary, a mark acting in *trans* recruits (or prevent the recruitment of) chromatin effector proteins (or complexes), which consequently triggers (or blocks) downstream signalling. For example, H3K4me3 recruits transcription machinery, whereas H3K27me3 recruits Polycomb group proteins (detailed in 1.2.2).

Readers, writers and erasers of histone marks

Histone PTMs represent a very dynamic layer of epigenetic information. Modifications are deposited and removed on histones by enzymes called writers and erasers, respectively. Writers of acetyl marks include the acetyltransferase GCN5

(H3K9/K14/K18-specific), which uses acetyl-CoA.[50] Methyl groups are deposited by methyltransferases, which uses S-adenosylmethionine (SAM). Examples of methyltransferases include the H3K4-specific mixed-lineage leukaemia (MLL) complexes, the H3K27-specific Polycomb repressive complex 2 (PRC2) and the H3K9-specific Suv(ar) family.[45] Kinases such as Aurora kinase B use ATP to catalyse the installation of a phospho group on H3S10 and H3S28.[51] Examples of erasers include the lysine demethylases of the Jumonji family and LSD1.[45]

The cell uses small protein binding modules as a general way to decode protein PTMs and therefore trigger downstream signalling cascades.[52] Such interaction domains also exist for histone marks and are called reader domains. They are small (between 50 and 110 residues) folded protein domains that can recognize histone PTMs in a specific manner (Figure 3).[53] Reader domains typically harbour a cavity or a surface groove that accommodates the modified residue. Amino acids forming this cavity provide specificity for the PTM type (methylation versus acetylation for instance) and PTM state (mono- versus di- and trimethylation). Reader domains typically make interactions with flanking residues thereby recognizing sequence context.

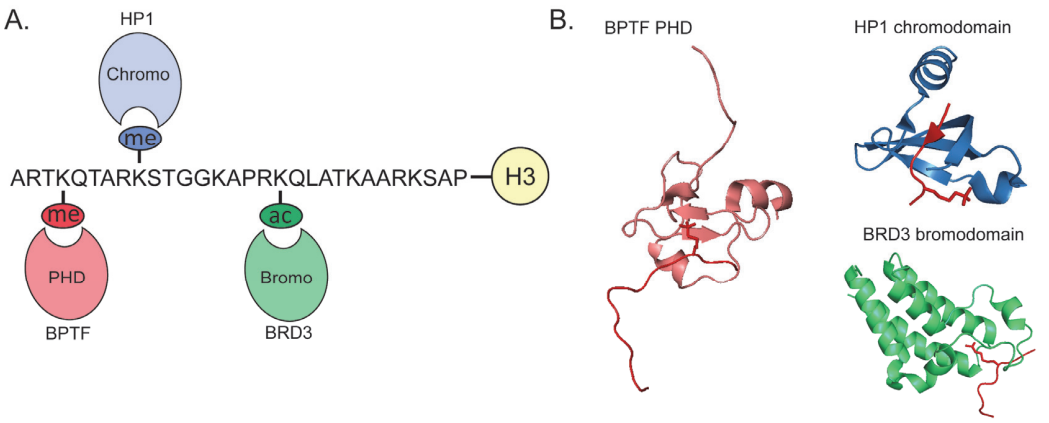

Figure 3. Reader domains as histone PTM binding modules. A. Amino acid sequence of histone H3.1 from residue 1 to 30 with examples of PTMs and associated reader domains. The protein in which reader domains are found are indicated. "me" corresponds to trimethylation. B. Ribbon representation of BPTF PHD bound to H3K4me3 peptide (PDB: 2FUU), HP1 chromodomain bound to H3K9me3 peptide (PDB: 1KNE) and BRD3 bromodomain bound to H3K18ac peptide (PDB: 5HJC).[54–56] Histone peptides are displayed in red with the modified residue in sticks.

The first reported reader-PTM interaction is the recognition of acetyl lysine by a bromodomain (Figure 4. A).[57] Bromodomains are found for example in the transcription regulators of the BET family.[58,59] Lysine methylation is recognized by chromodomains (CD), Tudor domains and Plant Homeodomains (PHD) among others (Figure 4. A). Examples include the recognition of H3K9me3 by the chromodomain of HP1,[60] the binding of methylated H3K4 by the PHD of BPTF[56] or the recognition of methylated H3K27 by the chromodomain of the Polycomb protein Pc.[61] Generally, reader domains specific for methyllysine harbour an aromatic cage to accommodate the di- or trimethyl ammonium group via cation-π interactions. Methyllysine binders also interact with residues flanking the modified amino acid to reach high sequence specificity. A striking example for this is provided by the chromodomains of HP1 and Polycomb (Pc). Although K9 and K27 residues are located in a similar sequence motif, chromodomain of HP1 is specific for H3K9me3, whereas chromodomain of Pc is specific for H3K27me3.[62] For PHD and chromodomains, specificity for the methyl state is generally inferior compared to sequence specificity, with increasing K_{DS} from trimethyl to unmethylated lysine.[56,62,63] Discrimination between unmethylated/monomethylated and di/trimethylated states is usually pronounced (1 or more orders of magnitude differences in K_{DS}), whereas preference for trimethyl versus dimethyl is less marked (1.7- and 10-fold difference in K_{DS} for HP1 CD and TAF3 PHD, respectively).

PHD zinc fingers are considered as a special case compared to other reader domains.[64] The majority of them are specific for methyl H3K4 but some recognize H3K14ac[65] or unmethylated H3K4.[66] For some PHD, the binding to methyl H3K4 is inhibited by methylation of H3R2, which is accommodated in an adjacent binding pocket.[63,67] This crosstalk mechanism is also observed for other reader domains such as the HP1 chromodomain, which is displaced from H3K9me3 by H3S10 phosphorylation.[68]

The histone code hypothesis and chromatin states

Reader domains are found in chromatin-associated proteins or complexes such as remodelers and modifiers. Numerous chromatin effectors contain several

reader domains and bind to histone marks in a multivalent fashion.[69–71] Several binding modes exist: *cis*- or *trans*-intranucleosomal binding and internucleosomal binding (Figure 4. B and C). TAF1 binds to diacetylated H4 tail with a tandem bromodomain in a *cis*-intranucleosomal fashion,[72] whereas BPTF recognizes PTMs on two different histones (H3K4me3 and H4K16ac) on the same nucleosome in a *trans*-nucleosomal interaction.[73] HP1 forms dimers which bind H3K9me3 in an internucleosomal fashion.[74,75] Individually, reader domains are modest affinity binders with typical K_Ds in the micromolar range. In the case of multivalent binding of tandem reader domains, the apparent affinity is higher. Indeed, once the first reader-PTM binding event occurred, the second binding is facilitated thanks to a significantly decreased cost in entropy (Figure 4. D and E).[69]

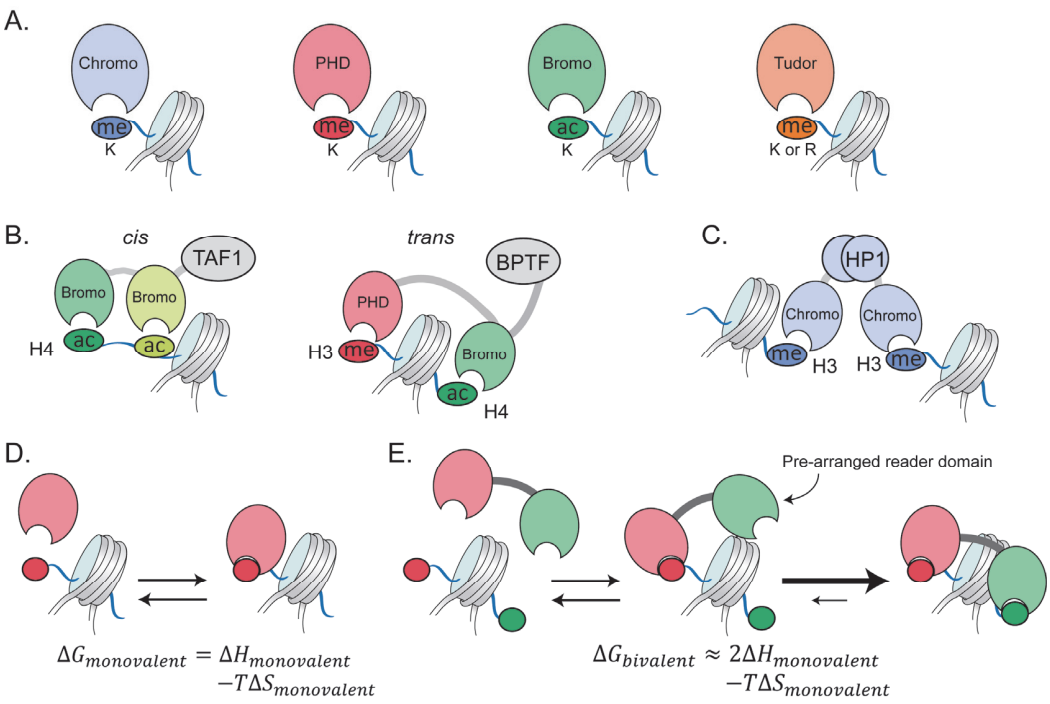

Figure 4. Reader domains recognize histone PTMs individually or combinatorially. A. Chromodomains and PHD are generally specific for methylated lysines, bromodomains recognize acetylated lysine residues and Tudor domains can bind methylated lysines or arginines. Here, "me" can correspond to mono-, di-, or trimethylation. B. *Cis*-intranucleosomal multivalent binding: tandem bromodomains of TAF1 recognize two modifications on the same tail (diacetylated H4). *Trans*-intranucleosomal multivalent binding: tandem PHD-bromodomain of BPTF bind to two marks on different histones (H3K4me3 and H4K16ac). C. Internucleosomal multivalent binding: HP1 binds H3K9me3 on different nucleosomes. D. and E. For similar individual reader domain affinities, bivalent binding is stronger than monovalent binding. The entropy cost is negligible for the second binding event thanks to a favourable pre-arrangement of the second domain.

Combinatorial readout of epigenetic marks by multivalent effectors forms the basis of the histone code hypothesis.[76,77] This hypothesis proposes that patterns of histone marks rather than individual marks serve as recruitment platforms and are combinatorially decoded by chromatin effectors containing multiple reader domains. It implies that combinations of histone marks produce distinct biological outputs via a multivalent readout. Some PTMs are often found together and others rarely coexist. Additionally, combinations of histone marks are often associated with a particular gene expression state (silenced or transcribed), and thus form a chromatin state. These observations support the histone code hypothesis and suggest that histone PTM patterns define specific chromatin states associated with a certain gene expression status.[78–81] The above-mentioned euchromatin and heterochromatin are two examples of chromatin states. Methods to detect and analyse combinations of histone marks involve sequential chromatin immunoprecipitation (ChIP) or mass spectrometry (MS) (discussed in 1.2.3).[82,83]

1.2 Bivalent chromatin in embryonic stem cells

1.2.1 Pluripotent stem cells

Cell potency and embryonic development

Stem cells are characterized by three features: the capacity to differentiate into any cell type forming an organism (i.e. pluripotency), the ability to proliferate indefinitely and the capacity to maintain an undifferentiated state (i.e. self-renewal).[84] Along the different stages of embryonic development, cell potency progressively decreases (Figure 5).[85] After fertilization, the zygote (or fertilized egg) undergoes several cellular divisions to form the morula embryo (16-cell state).[86] At this stage, all cells are totipotent, i.e. they can give rise to embryonic or extra-embryonic tissues.[87] After a few cell divisions, the morula becomes the hollow sphere known as blastocyst.[86] The outer layer of the blastocyst forms the trophectoderm (TE), which will form extra-embryonic tissues (such as placenta), whilst the inner cell mass (ICM) is composed of pluripotent cells, i.e. they can give rise to any of the

somatic cells forming an organism.[88] The ICM will then produce the three types of germs layers, endoderm, mesoderm and ectoderm, all composed of multipotent stem cells (i.e. having the ability to differentiate into multiple but limited cell types). Later during embryonic development, cell potency further decreases: oligopotent, unipotent and fully differentiated (or somatic) cells.[89] The adult organism is composed of somatic cells but also contains stem cells that represent a pool of undifferentiated cells dedicated to the production of new somatic cells.[90,91]

Keeping pluripotent cells in culture is of high interest not only for fundamental research but also for applications in regenerative medicine.[92] Induced pluripotent stem cells (iPSCs) are very promising for personalized regenerative medicine, whereas embryonic stem cells (ESCs) and cancer stem cells (CSCs) are rather used in fundamental research. iPSCs are obtained by reprogramming of somatic cells therefore representing an unlimited source of stem cells. Reprogramming can be achieved by several ways such as retroviral transduction of pluripotency-associated transcription factors,[93] by delivery of recombinant transcription factors[94,95] or by treatment with small molecule compounds.[96] CSCs arise from adult stem or progenitor cells that undergo mutation(s).[97] Found in several human cancers, they represent a minority of the cell population of tumours. Despite their low number, they are thought to be the source of all tumour cells, to be responsible for relapse after remission and for the appearance of metastasis.[97] Therefore they represent an important target in the development of anti-cancer drugs for a long-lasting cancer remission.

In this work, mouse ESCs (mESCs) were used. They were first isolated by Evans & Kaufman in 1981 from a mouse blastocyst.[98] ESCs are derived from the pluripotent cells forming the ICM. Upon culturing in a suitable medium, pluripotency and self-renewal capacities are maintained. Due to a short G1 phase, mouse ESCs have a cell cycle of ~ 4 - 5 hours,[99] much shorter compared to more differentiated cells. For a non-synchronized culture, 70 % of mouse ESCs are in S phase (and 20 % in G1), whereas this phase represents only 25 % of an embryonic fibroblast culture (and 70 % for G1).[100]

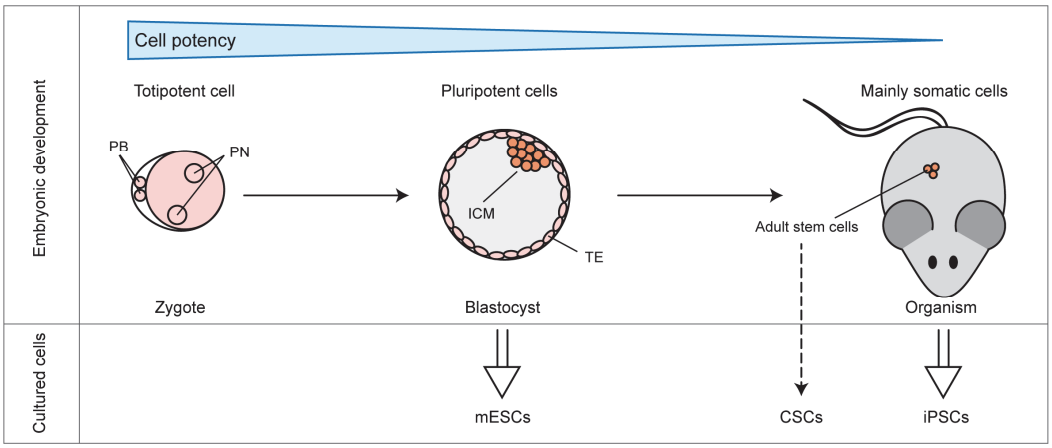

Figure 5. Schematic view of embryonic development and origin of cultured pluripotent stem cells. From the totipotent zygote to the somatic cells of an adult organism, cell potency decreases (blue triangle). Mouse embryonic stem cells (mESCs) are derived from the ICM of the blastocyst and induced pluripotent stem cells (iPSCs) can be obtained from somatic cells (plain double arrows). Cancer stem cells (CSCs) arise from adult stem cells (dotted arrow) and drive tumourgenesis. PB: Polar bodies; PN: Pronuclei; ICM: Inner cell mass; TE: Trophectoderm.

Pluripotent stem cells have particular epigenetic and gene expression landscapes that reflect their unique differentiation and self-renewal potentials.[101] Tight control of transcription factors (TFs) expression levels and chromatin machineries is required to keep the balance of pluripotent state and a disruption in regulatory mechanisms leads to differentiation. A variety of factors and pathways participates in pluripotency maintenance and some are discussed here (Figure 6). At the centre of pluripotency maintenance networks is the core regulatory circuit composed of three transcription factors, OCT4 (Octamer-binding transcription factor 4 or Pou5f1), SOX2 (Sry box-containing gene 2) and NANOG.[102,103] OCT4, SOX2 and NANOG (abbreviated OSN) are highly expressed in the ICM and knockouts of any of them in mouse embryos lead to development failure.[104–106] In mouse ESCs, OSN levels are tightly controlled to maintain pluripotency. A variation in the expression level of one or a combination of these three master TFs leads to differentiation to a specific lineage.[107–110] For example, a moderate level of OCT4 allows to maintain mouse ESCs in an undifferentiated state, whereas high OCT4 expression promotes lineage commitment to mesoderm or ectoderm. Finally, low OCT4 expression leads to trophectoderm formation.[107] OSN positively regulate the expression of many pluripotency-associated genes and negatively regulate lineage-

specific genes. They also positively regulate their own expressions via binding to their promoters, therefore forming an auto-regulatory feedback loop.[111,112]

LIF (Leukaemia Inhibitory Factor) is indispensable to maintain proliferation and pluripotency of mouse ESCs.[103,113] LIF is a cytokine secreted by the murine embryonic fibroblasts (MEFs) (so called feeder cells) and is routinely added as a supplement in the culture medium in the case of a feeder-free culture of mouse ESCs. LIF removal leads to spontaneous differentiation.[114] Upon binding to its membrane heterodimeric receptor (GP130/LIFR), LIF activates three major signalling pathways.[102,113] First, the JAK/STAT3 signalling pathway involves the phosphorylation of JAK1 & 2 (Janus kinases), which then phosphorylates and activates STAT3 (Signal transducer and activator of transcription 3). Phosphorylated STAT3 forms a homodimer that is translocated into the nucleus where it regulates transcription of pluripotency-associated genes such as *Klf4* and *Nanog*. Secondly, the PI3K/AKT pathway is activated upon LIF binding to it receptor. JAK proteins activate PI3K, which then phosphorylates and activates AKT kinases. AKTs inactivate GSK3 by phosphorylation of the latter at serine 9 and by favouring its nuclear export. GSK3 inactivation leads to upregulation of pluripotency-associated factors such as *Nanog* and *c-Myc*. The third signalling cascade activated by LIF is the MAPK/Erk pathway. JAK kinases phosphorylate SHP2, which then triggers a signalling cascade involving MEK and MAPK kinases. This cascade induces differentiation by down regulating pluripotency genes like *Nanog* and *Tbx3*. Therefore, LIF triggers not only pro-pluripotency pathways but also a pro-differentiation pathway.

Another crucial pathway for pluripotency maintenance is the canonical WNT signalling. In the absence of the glycoprotein WNT, GSK3 forms a complex with other proteins that binds, phosphorylates and triggers the proteosomal degradation of ß-catenin. Upon binding of WNT to the heterodimeric receptor Frizzled (FZ)/ LRP, GSK3 is recruited at the receptor and displaced from ß-catenin.[115] This leads to an accumulation of unphosporylated ß-catenin, which enters the nucleus and upregulates pluripotency-associated genes such as *Oct4*.[116] Although GSK3 is part of both the WNT and the PI3K/AKT pathways, the two pathways are actually

independent.[117,118] Other regulatory pathways involving chromatin remodelers or non-coding RNA for instance, are important for self-renewal of mouse ESCs.[102]

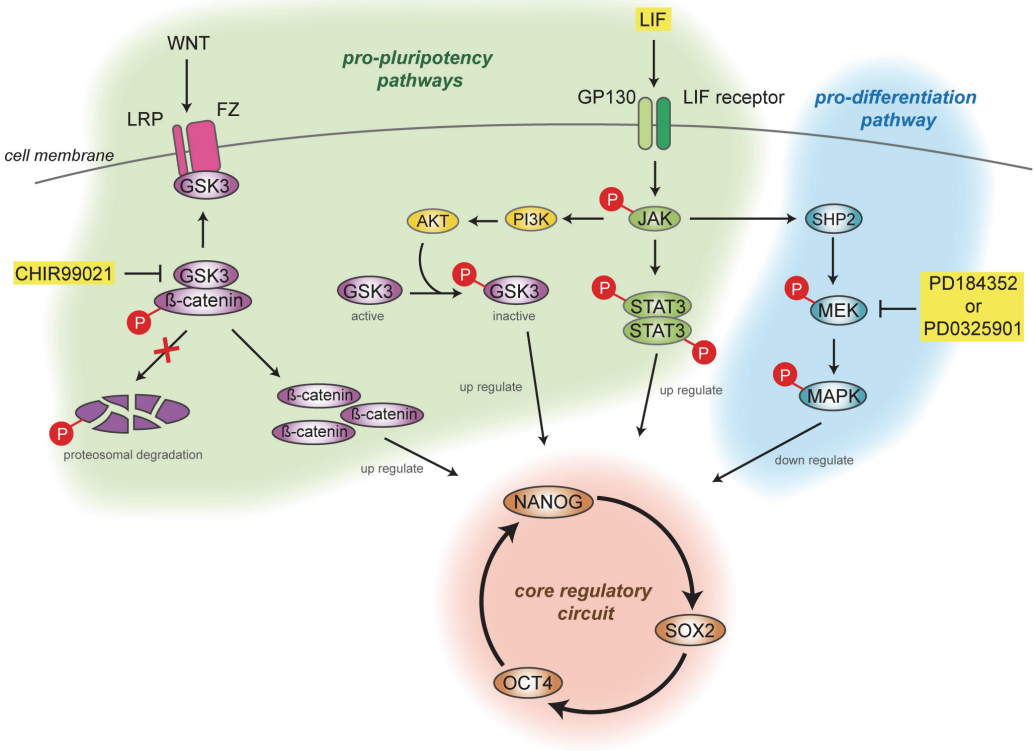

Figure 6. Regulatory networks controlling pluripotency maintenance and lineage commitment. Upon binding to its receptors, LIF activates three major intracellular pathways: the JAK/STAT3 pathway (in green), the PI3K/AKT pathway (in yellow) and the MAPK/Erk pathway (in blue). The canonical WNT pathway is also implicated in pluripotency maintenance (in purple). The core regulatory circuit is composed of NANOG, SOX2 and OCT4 forming an auto-regulatory feedback loop (in brown). Inhibitors of GSK3 and MEK (depicted as yellow boxed text) can be used as supplements in the culture medium in addition to LIF.

To sustain pluripotency of mouse ESCs, two small molecule inhibitors may be added in a serum-free medium in addition to LIF. This culture medium is known as 2i/LIF. The GSK3 inhibitor called CHIR99021 has IC_{50} in the range of 5 - 10 nM for mouse GSK3α and GSK3ß (Figure 7. A).[119] Treatment with CHIR99021 increases cytosolic level of ß-catenin,[119,120] which suggests that this molecule triggers the WNT pathway. However, this inhibitor might also block GSK3 in the PI3K/AKT pathway. The second small molecule supplemented in 2i/LIF medium is either of the two following MEK inhibitors: PD184352 is used as supplement together with

CHIR99021 and a FGF receptor inhibitor; PD0325901 is a more potent MEK inhibitor compared to PD184352 and is used in combination with CHIR99021 (Figure 7. B and C). Inhibition of MEK1 and MEK2 blocks the pro-differentiation MAPK/Erk pathway. Initially, the GSK3 and MEK inhibitors were used without LIF to show that ESC derivation and self-renewal can be achieved by blocking differentiation without the need of any pluripotency activating stimuli.[121] The two inhibitors are now widely used in combination with LIF to enhance pluripotency.[122] Mouse ESCs cultured in LIF containing medium present heterogeneous morphologies and gene expression patterns,[123–125] whereas these two parameters are more homogenous for cells cultured in 2i/LIF.[126–128]

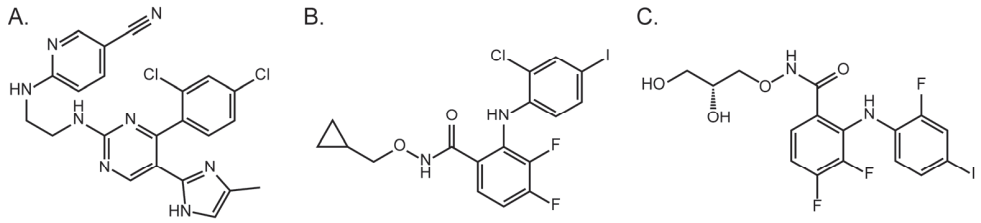

Figure 7. Chemical structures of CHIR99021 (A), PD184352 (B) and PD0325901 (C).

1.2.2 Polycomb and Trithorax complexes in pluripotent stem cells

Signalling cascades and external factors involved in pluripotency and self-renewal were detailed in chapter 1.2.1. This chapter concerns the nuclear and chromatin-associated components regulating development and ESC proliferation. Compared to somatic cells, stem cells have a permissive transcription landscape and a highly dynamic and open chromatin.[129–131] Two families of evolutionarily conserved regulators of gene expression are indispensable for proper embryonic development: Polycomb group (PcG) and Trithorax group (TrxG) proteins.[132–134] PcG and TrxG proteins were first identified as key repressors and activators, respectively, of the homeotic genes known as *Hox* genes in *Drosophila*.[135,136] In yeast, mammals as well as *Drosophila*, PcG and TrxG proteins are organized in several types of multimeric protein complexes with different enzymatic activities towards chromatin.

Polycomb Repressive Complexes

In mammals, PcG proteins form two distinct complexes: Polycomb repressive complexes 1 and 2 (PRC1/2), which have different subunit compositions and enzymatic activities.[137] PRC2 catalyses the installation of mono-, di- and trimethyl groups on lysine 27 of histone H3.[138,139] H3K27me3 is associated with gene silencing.[140,141] H3K27me2 is localized at intergenic regions, whereas H3K27me1 is localized at active gene bodies.[142] PRC2 complex contains the H3K27-specific methyltransferase EZH2 (or its homolog EZH1) comprising the catalytic SET domain. The other core components are the proteins EED, SUZ12 and RBAP46/48.[132,133] The WD40 repeat protein EED is essential for PRC2 activity.[143] This subunit was reported to bind to repressive marks, including H3K27me3, therefore leading to allosteric activation of PRC2.[144] This finding provides a mechanism for local spreading of H3K27me3. SUZ12 was demonstrated to be important for complex stability as well as activity[145,146] and forms with RBAP46/48 the minimal module for nucleosome binding.[143]

Additional non-core proteins such as JARID2, AEBP2 and PCL proteins can be associated with PRC2 and modulate its activity and/or targeting.[133] JARID2 was reported to play an important role in PRC2 recruitment to target genes.[147,148] The zinc finger protein AEBP2 was reported to stimulate PRC2 activity[145] and PCL proteins modulate PRC2 targeting and function.[149–151]

PRC1 has a more heterogeneous composition compared to PRC2. Its catalytic subunit is RING1A or RING1B, H2AK119-specific E3 ubiquitin ligases. H2AK119Ub was shown to be involved in transcriptional repression of target genes[152] and RING1A/B were found to have an essential role in repressing developmental regulators in mouse ESCs.[153] The PRC1 complexes can be classified as canonical (containing PCGF4/2, CBX2/4/6/7/8, PHC1/2/3 and RING1A/B) or non-canonical (containing RYBP, KDM2B, PCGF1 and RING1A/B) depending on their subunit compositon.[133] The non-canonical PRC1 subunit KDM2B is a H3K4- and H3K36-specific demethylase containing a CXXC domain, which binds to unmethylated CpG DNA. KDM2B was proposed to recruit non-canonical PRC1 to unmethylated CpG islands.[154–156] The canonical PRC2 subunit PCGF4 enhances

ubiquitin ligase activity,[157] while PHC proteins promote PRC1 clustering though polymerization.[158] Protein Pc, the *Drosophila* counterpart of CBX proteins, binds to H3K27me3 through its chromodomain, therefore recruiting canonical PRC1 to H3K27me3-marked genes (Figure 8).[139] The human PRC1-associated CBX proteins have lower affinity and specificity for H3K27me3, therefore questioning the validity of this recruitment mechanism for mammals.[159] Observations of PRC1-mediated H2AK119 ubiquitination independently of PRC2 further challenge this hierarchical model.[160,161] More recently, it was proposed that non-canonical PRC1-dependent H2AK119 ubiquitination leads to PRC2 binding and subsequent H3K27 methylation (Figure 8).[162,163] Supporting this model, PRC2 complex containing AEBP2 and JARID2 was found to bind H2AK119Ub nucleosomes and this interaction promotes PRC2-mediated H3K27 methylation.[164]

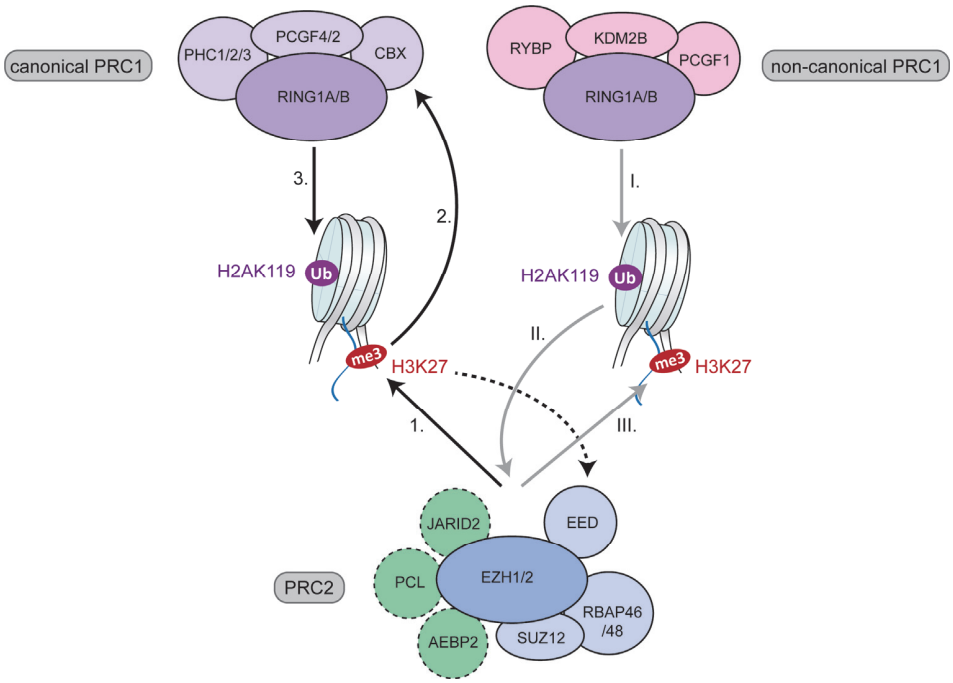

Figure 8. Interplay between PRC2, canonical and non-canonical PRC1 complexes. The classical model involves methylation of H3K27 by PRC2 (black arrow n°1.), local spreading by EED-mediated recruitment of PRC2 to H3K27me3 nucleosomes (black dotted arrow), canonical PRC1 recruitment via CBX proteins, which bind to H3K27me3 (black arrow n°2.) and subsequent H2AK119 ubiquitination by canonical PRC1 RING1A/B (black arrow n°3.). The alternative model entails H2AK119 ubiquitination by non-canonical PRC1 RING1A/B (grey arrow n°I.), recruitment of PRC2 to H2AK119Ub nucleosomes (grey line n°II.) and subsequent H3K27 methylation by PRC2 (grey arrow n°III.)

MLL complexes

The Trithorax group is a heterogeneous family of proteins including chromatin modifiers and remodelers associated with active gene transcription.[165] This group encompasses the NURF and SWI/SNF complexes as well as the COMPASS family. In yeast, SET1 is the TrxG protein responsible for methylation of H3K4. SET1 was found within a multiprotein complex called COMPASS (complex of proteins associated with SET1). In *Drosophila*, the COMPASS family is composed of 3 multimeric complexes, SET1, Trithorax (Trx) and Trithorax-related (Trr). In mammals, this family contains 6 members (two homologs of each *Drosophila* complex) called MLL complexes (Table 1).[133] The catalytic subunit of mammalian MLL complexes consists of one of the SET-domain containing proteins with a H3K4-specific methyltransferase activity, SET1A/B or MLL1/2/3/4.[166,167] All 3 H3K4 methylation states are associated with active gene transcription.[168] H3K4me2 and H3K4me3 are mainly found next to transcription start sites (TSSs), whereas H3K4me1 is located at enhancers.[168,169] WDR5, RBBP5, ASH2L and DPY30 (WRAD) together with the catalytic subunit form the core MLL complexes.[133] WDR5 is important for the stability of the complex.[170] Allis and coworkers proposed that this subunit binds to H3K4me2 nucleosomes, therefore serving as a recognition module.[171] However, another study showed that the WDR5 protein does not interact with H3K4 and is not able to discriminate between unmodified, mono-, di- or tri-methylated H3 tail.[172] ASH2L was reported to contain a DNA-binding motif suggesting a recruitment role for this subunit.[173,174] ASH2L together with RBBP5 and the SET-containing protein MLL1 were proposed to form a joint catalytic centre for methylation of H3K4.[175] RBBP5 is essential for complex stability and was proposed to undergo a phosphorylation switch controlling core complex formation and activity.[176] As for DPY30, it was shown to modulate MLL-mediated H3K4 methylation and to be crucial for ESC differentiation.[177] Additionally to WRAD, other proteins can associate with SET proteins such as Menin, which modulates activity and targeting of MLL1/2;[178] the H3K27-specific demethylase UTX, which coordinates H3K27me3 demethylation with methylation of H3K4 at TSS of active genes;[179] and the unmethylated CpG-binding protein CFP1,

which regulates target sites of SET1A/B.[180] MLL complexes have non-redundant roles. Indeed, SET1A/B are considered as the default H3K4 methyltransferases and deposit di- as well as tri-methyl marks mainly at active promoters.[181] MLL1/2 appear to be more specific for *Hox* genes.[181] MLL2 was shown to deposit H3K4me3 at bivalent promoters (detailed in 1.2.3).[182,183] MLL3/4 install H3K4me1 at poised enhancers.[184] Importantly, MLL3/4 were found to be part of a macromolecular complex called ASCOM (ASC-2 complex), which acts as a transcriptional coregulator.[185]

Mammalian TrxG complex	**Subunit composition (non-exhaustive)**	**Main function and redundancy**
SET1A	SET1A, WDR5, ASH2L, RBBP5, DPY30, CFP1	Default H3K4 methyltransferase, deposits di- and tri-methyl at active promoters. Redundant with SET1B
SET1B	SET1B, WDR5, ASH2L, RBBP5, DPY30, CFP1	Default H3K4 methyltransferase, deposits di- and tri-methyl at active promoters. Redundant with SET1A
MLL1	MLL, WDR5, ASH2L, RBBP5, DPY30, Menin	Deposits mono-, di- and tri-methyl on H3K4. Non-redundant with other MLL complexes
MLL2	MLL2, WDR5, ASH2L, RBBP5, DPY30, Menin	Deposits mono-, di- and tri-methyl on H3K4. Among its targets are bivalent developmental genes. Non-redundant with other MLL complexes
MLL3	MLL3, WDR5, ASH2L, RBBP5, DPY30, UTX	Deposits H3K4me1 at poised enhancers. Can be part of ASCOM complex. Redundant with MLL4
MLL4	MLL4, WDR5, ASH2L, RBBP5, DPY30, UTX	Deposits H3K4me1 at poised enhancers. Can be part of ASCOM complex. Redundant with MLL3

Table 1. Subunit composition and functions of mammalian MLL complexes.

It is worth noting that the nomenclature of proteins MLL2 and MLL4 can be confusing in the literature.[186] MLL2 is sometimes named MLL4 and inversely. In this work, MLL2 corresponds to the Trithorax homolog KMT2B, whereas MLL4 corresponds to the Trithorax-like homolog KMT2D.

PcG and TrxG protein regulate gene expression

In mouse, homozygous gene deletion of any of the 6 TrxG proteins (SET1A/B, MLL1/2/3/4) leads to developmental failure and death of the embryo, highlighting the crucial role of MLL complexes in development.[187–190] The same phenotype is observed for double knockouts of EZH2, EED, SUZ12 or RING1B, demonstrating that PcG proteins are essential for proper embryonic development.[146,191–194] In mouse ESCs, the two PcG complexes are dispensable for pluripotency maintenance but required for proper differentiation.[195]

H3K27me3 is widely distributed in the genome and spans large genomic regions, whereas H3K4me3 forms smaller peaks. The high abundancy of H3K27 methylation states (~ 20 % H3K27me1, ~ 50 % H3K27me2 and ~ 10 - 20 % H3K27me3 in mouse ESCs[196]) is consistent with the fact that H3K27 methylation is associated with reversibly silenced regions (facultative heterochromatin) unlike the DNA- and H3K9-methylation, which impose a more static gene repression at centromeres and telomeres (constitutive heterochromatin).

Originally discovered at *Hox* genes in *Drosophila*, it is currently the consensus that PcG and TrxG proteins are found genome-wide in mammals and are associated with gene repression and activation, respectively. However, the way in which these complexes lead to gene expression control is more elusive. It was shown that PRC1 imposes gene repression via several mechanisms: it blocks RNA Polymerase (Pol) II,[197] inhibits the nucleosome remodeling complexes SWI/SNF[198,199] and induces chromatin compaction.[200–202] Within PRC2, the catalytic subunit EZH2 binds to DNA methyltransferases and thus could serve as a recruitment platform for DNA methylation-mediated repression.[203] As for TrxG proteins, several links were found between active gene transcription, MLL complexes and H3K4me3. TFIID, part of the RNA Polymerase II pre-initiation complex, contains the protein TAF3, which has a PHD finger binding specifically to H3K4me3.[63] This mark is also recognized by the PHD finger of BPTF, a subunit of the remodeling complex NURF.[56,204] The *Drosophila* ASH2 protein was demonstrated to affect the phosphorylation status of RNA Polymerase II, therefore influencing its activity.[205]

Several studies suggest that PcG and TrxG complexes establish and maintain repressive and active transcriptional states by installing histone PTMs that inhibit binding and/or activity of the opposite complex(es). Indeed, PRC2 binding and activity on H3K27 are inhibited by active chromatin marks such as H3K4me3[206,207] and H3K27ac.[208–211] In the same way, the repressive chromatin mark H3K27me3 inhibits binding and activity of the SET1A complex.[212]

Nuclear organization of chromatin into subcompartments

One aspect that is believed to be crucial for effective activation and silencing is the sequestration of genes into dedicated subcompartments.[213] Analysis of chromatin contacts by chromatin conformation capture (3C) techniques revealed chromosomal regions within which sequences preferentially contact each other.[214,215] These so called topologically associated domains (TADs) were proposed to organize the genome according to gene expression state and histone modifications.[214–218] Such architectural features are believed to keep genes in a local environment enriched in activation- or repression-associated proteins. In *Drosophila*, PcG proteins are found in discrete foci, named Polycomb bodies where genes are silenced.[219,220] Genomic regions that are distant or even located on different chromosomes were found to associate into Polycomb bodies.[221,222] In mammalian cells, the existence of Polycomb bodies is not yet confirmed. PRC1 was proposed to function as a "3D organizer" and to spatially constraint genes in a silenced network.[223] Polymerization of PHC2, a core component of PRC1, plays an essential role in PRC1 clustering as well as in robust gene silencing and might be key to the organization of silenced genes into Polycomb bodies.[158] Similarly, transcriptionally active genes are organized into functional compartments called transcription factories.[224,225] Intra- and interchromosomal interactions occur within sets of active genes, therefore forming transcription foci.[226,227] Moving out of the factories switches off gene transcription, whereas induction provokes relocation of a gene into a transcription factory.[226,228] This demonstrates the intimate link between active transcription and location in such subcompartments. Cook and coworkers counted more than 2'000 extranucleolar RNA Pol II and III transcription factories per human nucleus.[229,230] They also

measured a diameter of ~ 70 - 80 nm for each factory and calculated that each factory contains between 20 and 30 active RNA Pol II. A transcription factory can be described as a subnuclear compartment where several RNA Pol II units transcribe distal genes as well as genes located on different chromosomes.

How TrxG and PcG proteins are recruited to their target genes is a longstanding question. In *Drosophila*, Polycomb and Trithorax responsive elements (PREs and TREs) are genomic regions that recruit Pc and Trx complexes, respectively. For PREs, this recruitment is thought to take place via interaction with transcription factors such as PHO, which contains a sequence-specific DNA-binding domain.[231–233] However, no mammalian PREs and TREs have been described until now. CpG islands are overrepresented in PcG targets and were proposed to contribute to PcG recruitment.[234] As previously mentioned, some subunits of MLL and Polycomb complexes are suspected to have a role in recruitment (CFP1, WDR5, ASH2L for MLL complexes and KDM2B, JARID2, PCL proteins for Polycomb complexes). Finally, non-coding RNA was suggested to recruit and modulate PcG and TrxG protein complexes.[235–240]

1.2.3 Bivalent chromatin decorates developmental genes in embryonic stem cells

H3K4me3 and H3K27me3 are associated with opposite gene expression states and are present at different genomic locations in somatic cells. However, these two marks are not mutually exclusive. In 2006 and 2007, several groups found H3K4me3 and H3K27me3 to coexist at promoters of key developmental genes (mainly TF genes) in mouse as well as in human ESCs.[241–244] These so called bivalent domains, defined as large H3K27me3 regions harbouring smaller H3K4me3 regions, are not only found at key developmental genes but also at genes controlling morphogenesis and encoding cell surface receptors.[245] Bivalent genes are repressed or expressed at low levels and bivalency resolves upon differentiation, leaving the gene marked with either H3K4me3 or H3K27me3 or none of them (Figure 9).[241,245,246] Therefore bivalent chromatin was proposed to keep developmentally

important genes silenced but poised for activation in pluripotent stem cells. Bivalent domains were also detected in more differentiated states and in cancer cells, although at lower levels.[245–256] The correlation between bivalency and pluripotency led to the hypothesis that bivalent chromatin reflects lineage potential and keeps the plasticity (or "stemness") of ESCs.[245] However, the function(s) of bivalent domains is (are) still a subject of debate. In 2013, close to 4'000 and 3'000 genes were reported to be bivalent in human and mouse ESCs, respectively and more are discovered every year.[257] These numbers represent ~ 13 % and ~ 20 % of proteins coding genes in mouse and human, respectively.[258] Ku et al. mapped the occupancy of Polycomb complexes genome-wide by ChIP-seq in human and mouse ESCs and classified bivalent domains into two functionally distinct categories according to their PRC1 occupancy.[259] The authors found that PRC2 occupies essentially all bivalent genes, whereas PRC1 occupies a conserved subset of them. The PRC1-positive bivalent genes are more efficient at retaining H3K27me3 upon differentiation, cover larger genomic regions and are more enriched in TF genes, compared to the PRC1-negative bivalent genes.

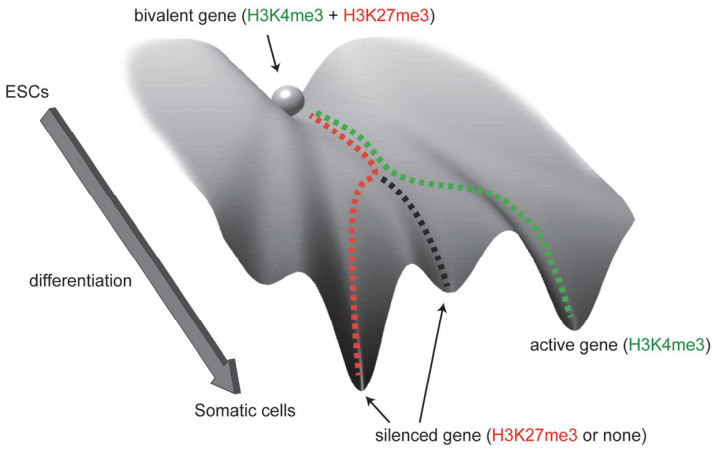

Figure 9. Schematic view of epigenetic bivalency. In pluripotent stem cells, such as ESCs, the coexistence of H3K4me3 and H3K27me3 represses key developmental genes (lineage-specific genes). After differentiation, bivalent genes remain marked with either H3K4me3 (rendering gene transcription active) or H3K27me3 or none of them (both keep genes repressed).

Bivalent marks are present in an asymmetric fashion on nucleosomes

The discovery of bivalent chromatin raised the question of how H3K4me3 and H3K27me3 are distributed on nucleosomes: Are H3K4me3 and H3K27me3

present on the same nucleosomes (bivalent nucleosomes) or on different nucleosomes? Do H3K4me3 and H3K27me3 reside on the same H3 or on opposite H3 molecules? Are H3K4me3 and H3K27me3 randomly distributed or one conformation is preferred? To answer these questions, several techniques were used and are detailed below.

ChIP is a widely used method to map histone modifications enrichment genome-wide.[260] In this technique, fixed cells are sonicated to fragment chromatin and the protein (or mark) of interest is then immunoprecipitated. The recovered DNA can be analysed by various techniques such as quantitative polymerase chain reaction (qPCR), micro-chip or sequencing. Most of the bivalent domains were discovered by ChIP-based methods and declared 'bivalent' when both H3K4me3 and H3K27me3 were detected. Although this technique is informative about enrichment of single histone modifications, ChIP data have to be interpreted carefully when considering combinations of histone modifications. Indeed, two marks can appear as if they are enriched at the same loci but might originate from different cell subpopulations or from different alleles. Sequential ChIP (or Re-ChIP) overcomes this limitation, by immunoprecipitating the chromatin recovered from a first IP. Bernstein et al. used Re-ChIP to show the coexistence of H3K4me3 and H3K27me3 at the *Irx2* promoter.[241] The second reason why conventional ChIP data are complex to interpret when considering a pair of marks is because it yields relative and not quantitative enrichments. Bivalent domains are declared 'bivalent' when levels of H3K4me3 and H3K27me3 are above certain thresholds. These thresholds might be set at different values in different research groups rendering impossible the comparison of ChIP data coming from different sources. In 2015, Grzybowski et al. developed a ChIP-seq-based method in which semi-synthetic barcoded nucleosomes are used as internal standards, enabling to measure actual histone modification densities.[261] With this method, the authors found the sum of H3K4me3 and H3K27me3 densities being superior to 100 % for transcriptionally silent genes. This confirms the coexistence of H3K4me3 and H3K27me3 on the same nucleosomes at ~ 35 % of all genes in mouse ESCs (considering protein coding genes[258], 35 % represents ~ 7'900 genes). Furthermore, the authors classified bivalent nucleosomes into two categories: the

canonical bivalency, in which case genes are silent, contain between 70 - 100 % H3K27me3 and 1 - 60 % H3K4me3 near their TSS, whereas a new class of formally bivalent and highly transcriptionally active genes are characterized by nearly 100 % H3K4me3 and between 10 - 40 % H3K27me3. This finding suggests that the dominant modification controls the transcriptional state.

Knowing that H3K4me3 and H3K27me3 are indeed present on the same nucleosomes, different conformations are possible: symmetric (with both H3 copies carrying K4me3 and K27me3), *cis* asymmetric (with one H3 molecule unmodified and the other H3 carrying K4me3 as well as K27me3) and *trans* asymmetric bivalent nucleosomes (with K4me3 on one H3 copy and K27me3 on the other H3 tail). Several methods were used to obtain information about the conformation of bivalent nucleosomes. In 2012, Voigt et al. used an immunoaffinity purification method on native mononucleosomes followed by MS to show that H3K4me3 and H3K27me3 i) coexist on the same nucleosomes and ii) are present on different H3 molecules.[246] This *trans* asymmetric conformation (Figure 10. A) was confirmed in 2016 by Sen et al.[262] More recently, Shema et al. adopted a single molecule approach using total internal reflection fluorescence (TIRF) microscopy and fluorescently labelled antibodies to analyse mononucleosomes isolated from mouse ESCs.[263] The authors showed that 0.5 % of nucleosomes in mouse ESCs are truly bivalent with both marks on the same nucleosomes (considering a nucleosome repeat length of ~ 190 bp, this represents ~ 147'000 nucleosomes).[2,258] They also demonstrated that the majority (94 %) of bivalent nucleosomes are *trans* asymmetric. Contrary to Voigt et al., Shema et al. found that bivalent nucleosomes modified on the same tail (symmetric or *cis* asymmetric) do actually exist and represent 6 % of bivalent nucleosomes.

In summary, bivalent domains have the following characteristics:

1. The quantities of H3K27me3 and H3K4me3 are 70 - 100 % and 1 - 60 %, respectively.[261] This implies that some nucleosomes in bivalent domains carry only H3K27me3 (in a symmetric or asymmetric fashion).
2. H3K4me3 and H3K27me3 are present on the same nucleosomes in a *trans* asymmetric fashion.[246,261,263]
3. A small portion of bivalent nucleosomes have the two marks on the same

histone H3 (either in a *cis* asymmetric or in a symmetric configuration).[263]

In view of these findings, a bivalent promoter can be depicted as a genomic region containing a majority of *trans* asymmetric bivalent nucleosomes, some H3K27me3 nucleosomes and rare symmetric or *cis* asymmetric bivalent nucleosomes (Figure 10. B).

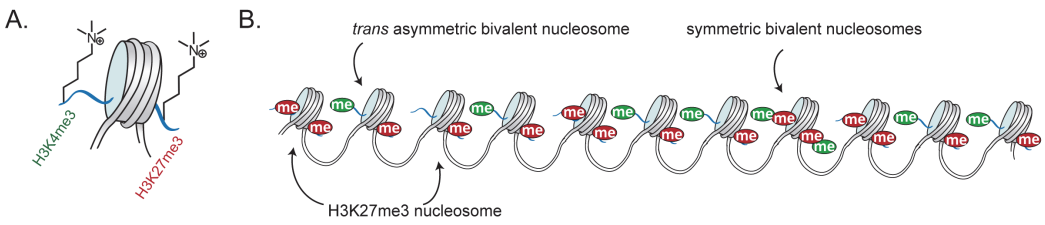

Figure 10. Bivalent marks are found asymmetrically in mouse ESCs. A. Schematic view of a *trans* asymmetric bivalent nucleosome. B. Schematic view of the nucleosome composition of a bivalent domain, containing mainly *trans* asymmetric bivalent nucleosomes, some H3K27me3 nucleosomes (symmetric or asymmetric) and a few nucleosomes with K4me3 and K27me3 on the same H3 tail (depicted here as symmetric bivalent nucleosome). "me" represents trimethylation at H3K4 and H3K27 in green and red, respectively.

Little is known about the 3D organization of bivalent domains. One can speculate that they are gathered into subnuclear compartments, similarly to Polycomb bodies and transcription factories, where they might be coregulated by Polycomb and Trithorax proteins.

Installation, regulation and resolution of bivalent domains

Bivalent marks are not randomly distributed since they are present on the same nucleosomes and for the majority of nucleosomes, on opposite H3 tails. This suggests mechanisms that write, maintain and erase bivalent marks in a controlled manner. Concerning regulation and function(s) of bivalent domains, the following interrogations remain to be answered:

1. How do PRC2 and MLL complexes write bivalent marks at promoters of developmental genes? Which mark is deposited first?
2. How is bivalency maintained in ESCs and resolved upon differentiation? What mechanism controls whether a bivalent gene resolves to active or repressed state?

3. Do bivalent domains have biological function(s)? Is asymmetry critical for its function(s)? How does asymmetry influence the binding and enzymatic activity of chromatin effectors?

Currently, these questions have only preliminary and sparse answers (Figure 11). *De novo* appearance of bivalent domains is observed in several cases such as reprogramming of somatic cells into iPSCs,[264] mitosis-to-meiosis during germ line production[265,266] and even during differentiation into adult tissue cells.[247,249] On the other hand, differentiation of ESCs witnesses an important decrease in the number of bivalent genes.[241,245,246] Moreover, bivalent domains are dynamically regulated during the cell cycle in pluripotent cells with an increased abundance in G1 phase.[267] These observations suggest a tight regulation of the installation, maintenance and removal of bivalent marks. PRC2 is responsible for installing H3K27me3 and occupies nearly all bivalent regions.[259] MLL2 was identified as the H3K4 methyltransferase responsible for installing H3K4me3 at bivalent genes in ESCs.[182,183,267] PRC2 activity is inhibited by the presence of H3K4me3 on the same tail (*cis* inhibition), whereas SET1-like H3K4 methyltransferases binding and activity are inhibited by the presence of H3K27me3 on the same tail.[206,207,212] These findings are in line with the fact that bivalent marks are present in majority in *trans* conformation on nucleosomes. However, how MLL2 and PRC2 are recruited to bivalent genes is unclear. Lynch et al. proposed that the signal for chromatin bivalency is encoded into the local DNA sequence.[268] Indeed, bivalent promoters are enriched in CpG dinucleotides and the majority of bivalent genes (93 % in stem cells) overlaps with CpG islands (CGIs),[245,247,269] proposed to be necessary and sufficient to recruit PRC2.[234] However, not all CGIs/promoters are bivalent and bivalency is also found at intragenic CGIs,[270] at gene bodies and upstream of TSSs.[249,262] Therefore, CGIs are only a partial answer to the question of recruitment. MicroRNA,[271] crosstalks with H2A.Z or other histone modifications[248,262,269] and unmethylated CpGs[272,273] are potential regulatory mechanisms for installation and maintenance of bivalent marks. Concerning the resolution of bivalency, the H3K27-specific demethylase UTX was proposed to play a role in the activation of bivalent genes during differentiation of ESCs.[274] Downregulation of LSD1 during

differentiation might represent an additional mechanism for the activation of bivalent genes.[275]

Whether bivalent domains have a biological function remains unclear. First of all, co-occupancy of PcG and TrxG proteins was observed not only in mammals but also in *Drosophila*,[276,277] suggesting that bivalent/poised state is a conserved feature among species. Two opposite views have emerged concerning a potential function of bivalent domains. The above-mentioned hypothesis proposed by Bernstein et al. suggests that bivalency keeps developmental genes silent but poised for activation. This model is supported by the resolution of bivalency for the majority of bivalent genes upon differentiation[241,245,246] and by the presence of bivalent chromatin at developmental genes in germ lines.[37,278–280] Furthermore, poised RNA Pol II (phosphorylated at serine 5) was detected at bivalent regions[281] and was shown to be important for proper differentiation,[282] also supporting the poised state hypothesis. However, despite the global decrease in the number of bivalent genes in adult tissues and somatic cells compared to ESCs, some genes acquired *de novo* bivalency in adult tissues, suggesting an alternative role for bivalent domains.[249] Bivalency was proposed to have a protective role from aberrant DNA hypermethylation, a hallmark of cancer.[265,269,283,284] On the other hand, several studies suggest that bivalent state does not have a particular function. Maupetit-Méhouas et al. state that bivalent chromatin is the default state of inactive CGIs/promoters.[284] Lynch et al. suggest that bivalency is the result of the competition between Polycomb silencing and transcriptional activation.[268] This view is consistent with the observation that bivalent regions with H3K4me3 > H3K27me3 have in general higher expression levels than bivalent regions with H3K4me3 < H3K27me3.[261,269] Finally, loss of H3K4me3 at bivalent promoters in MLL2$^{-/-}$ mESCs does not prevent activation of these genes upon retinoic acid induced differentiation, arguing against the model in which the bivalent state controls gene expression.[182,183] Although the poised gene expression hypothesis seems to be the most widely accepted among scientists, some evidence argue against this model.

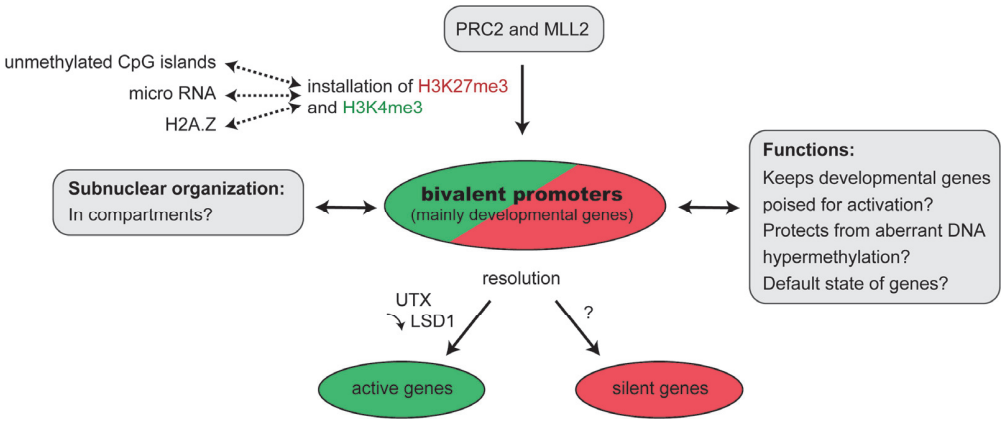

Figure 11. Overview of potential function(s), subnuclear organization and mechanisms regulating installation and resolution of bivalent domains. PRC2 and MLL2 install H3K27me3 and H3K4me3, respectively, at promoters of developmental genes. Upon resolution, bivalent genes are converted to either active or repressed genes. Dashed arrows indicate elements that might influence the formation of bivalent domains. Whether bivalent chromatin have a biological function and how it is organized in the nucleus are still unclear.

In conclusion, bivalency is an intriguing combination of histone PTMs associated with opposite transcriptional states. Present at high levels in pluripotent cells, it decorates inactive key developmental genes. Although researchers have reached an agreement on the *trans* asymmetric conformation of bivalent nucleosomes, their biological function is still a subject of debate. Several questions about bivalent domains remain open. In particular, how these domains are organized in the 3D space within the nucleus is uncertain. How bivalent domains are dynamically regulated during cell cycle and during cellular differentiation is also unknown. In the current work, we designed probes targeting bivalent marks in order to answer these questions. To test *in vitro* the binding of these probes to bivalent PTMs, we needed access to histones with site-specific methylation marks. To this end, we used expressed protein ligation and synthetic peptides.

1.3 Protein engineering to introduce modifications in nucleosomes

1.3.1 Site-specific introduction of modifications in proteins

Having access to large amounts of homogeneously labelled or modified proteins is central to study their function. For example, labelling proteins with small molecule fluorophores is crucial for many microscopic applications.[285] Introduction of unnatural amino acids (Uaas) such as crosslinkers can give insights about a protein's interactome.[286,287] Introducing PTMs in a site-specific manner into proteins enables to investigate their biological effects as well as to develop probes that target modified proteins. Multiple methods are available to site-specifically introduce non-proteinogenic amino acids into proteins: cysteine-directed conjugation, genetic code expansion (such as amber codon suppression), native chemical ligation (NCL) and its variant, expressed protein ligation (EPL) (Figure 12). NCL and EPL are described in 1.3.2.

First described by Schultz's and Chamberlin's groups in 1989, genetic code expansion involves the development of an orthogonal aminoacyl-tRNA synthetase/tRNA pair.[288,289] In this method, the Uaa is encoded by the amber codon UAG such that an engineered tRNA synthetase loads the Uaa on the CUA-codon tRNA. In a second step, the ribosomal machinery translates the mRNA into protein by incorporating the Uaa in place of the amber codon (Figure 12. A). This technique can be applied in bacteria, yeast as well as mammalian cells and a large variety of non-proteinogenic amino acids can be introduced into proteins. The same Uaa can be incorporated at multiple sites within one protein.[290] Genetic code expansion was adapted to four-base codons to introduce multiple Uaas in proteins in an orthogonal manner.[291,292] However, the main challenge is to reach orthogonality with the 20 natural aminoacyl-tRNA synthetase/tRNA pairs. One of the drawback of this technique is the lack of modularity. Indeed, for every Uaa of interest, a new pair of aminoacyl-tRNA synthetase/tRNA needs to be developed. Pyrolysyl-tRNA synthetase/tRNA pairs were developed to introduce acetyllysine as well as more exotic PTMs in H3.[293–295] Precursors of mono- and dimethyllysine can be

incorporated into H3 by amber codon suppression and subsequently converted by chemical means (Figure 12. B).[296–298] This method can also be used to introduce lysines with a thiol moiety, subsequently ligated with a ubiquitin thioester, yielding a native isopeptide bond (Figure 12. B).[299]

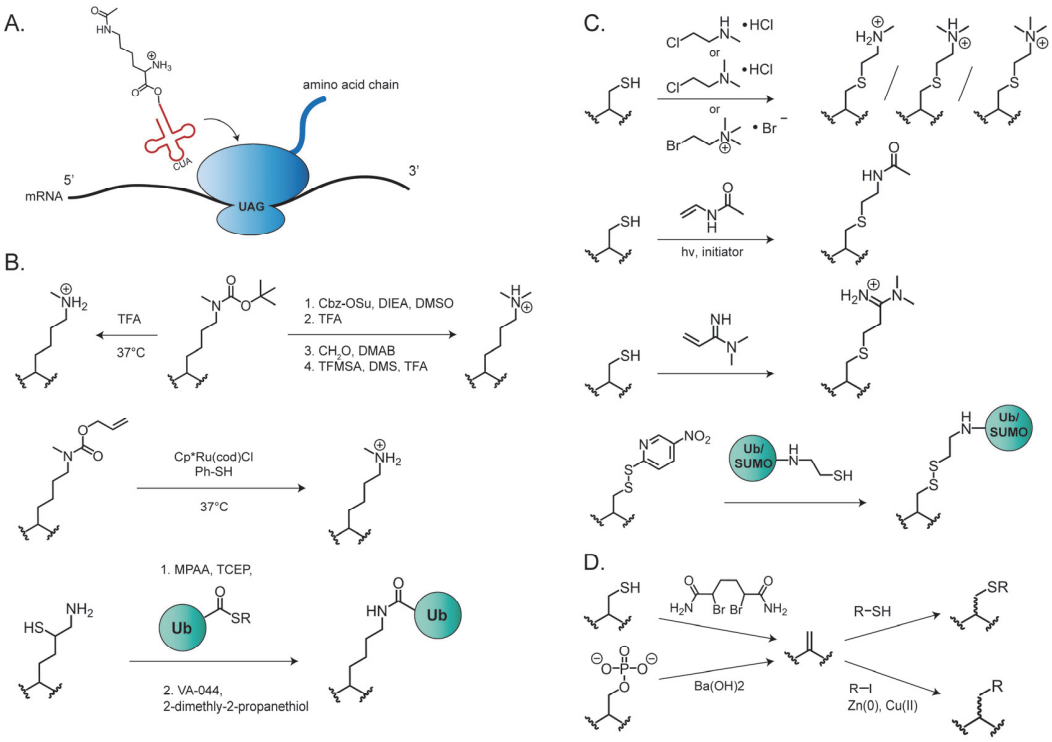

Figure 12. Site-specific introduction of Uaas in proteins and more specifically of PTMs into histones. A. Amber codon suppression can be used to produce acetyllysine histones. B. Precursors can be incorporated by amber codon suppression into histones and then converted to mono- or dimethyllysine. δ-thiol lysine can be incorporated into proteins and subsequently ligated with ubiquitin thioester resulting in an isopeptide bond. C. Cysteine can be converted to a variety of histone PTMs analogues as well as S-S linked ubiquitin. D. Dehydroalanine can be used to convert cysteine or phosphoserine into a variety of histone PTMs.

Cysteine-directed conjugation makes use of the nucleophilicity of the sulphur atom to site-specifically label proteins carrying one or several cysteine mutation(s). This method is straightforward, therefore enabling large amounts of modified proteins to be synthesized. However, these reactions are not traceless since the sulphur atom replaces the native γ-carbon atom of the side chain. Concerning histone PTMs, cysteine-directed conjugations were applied to obtain methyl- and acetyllysine analogues as well as methylarginine analogues (Figure 12. C).[300–302]

This method was also used to produce disulfide linked ubiquitinated or sumoylated histones (Figure 12. C).[48,303,304] Finally, dehydroalanine represents an alternative to obtain modified proteins.[305] Dehydroalanine can be obtained either from cysteine (incorporated by conventional mutagenesis) or from phosphoserine (incorporated by amber codon suppression) and subsequently converted to histone PTMs or histone PTMs mimics (Figure 12. D).[306–308]

1.3.2 Expressed protein ligation to introduce modifications in histones

NCL consists in ligating an N-terminal cysteine peptide with a C-terminal thioester peptide, yielding a native amide bond. This traceless ligation is used in EPL to join a peptide with a recombinant protein. EPL is a method of choice to obtain milligram amounts of site-specifically modified histones. Typically, one of the two pieces to ligate is a synthetic peptide obtained by solid phase peptide synthesis (SPPS), whereas the other piece is recombinantly expressed.

Solid phase peptide synthesis

SPPS was developed in the group of Robert Bruce Merrifield in the 60s and is based on the successive addition of amino acids on a solid support (Figure 13).[309,310] In SPPS, the peptide is synthesized from the C- to the N-terminus. To avoid multiple additions of the same amino acid on the peptide, amino acid's α-amino group are protected by *tert*-butoxycarbonyl (Boc) or fluorenylmethyloxycarbonyl (Fmoc) protecting groups. The α-carboxy group of each amino acid is activated prior to coupling with the peptide. For this purpose, various coupling reagents are available, the most common being *N,N'*-dicyclohexylcarbodiimide (DCC), HBTU and HATU. The Hünig's base *N*-ethyldiisopropylamine (DIPEA) is used during activation. Unreacted reagent is washed away and α-amino group deprotection is performed (with piperidine for Fmoc group or with trifluoroacetic acid (TFA) for Boc group). Reactive side chains carry orthogonal protecting groups, which are removed upon cleavage from the resin (with TFA for Fmoc SPPS or with

hydrofluoric acid (HF) for Boc SPPS). Originally, Boc was used as α-amino protecting group. However, anhydrous HF requires specific handling. For this reason, the milder Fmoc strategy is generally preferred. Automation of SPPS allows to synthesize peptides with minimal handling.

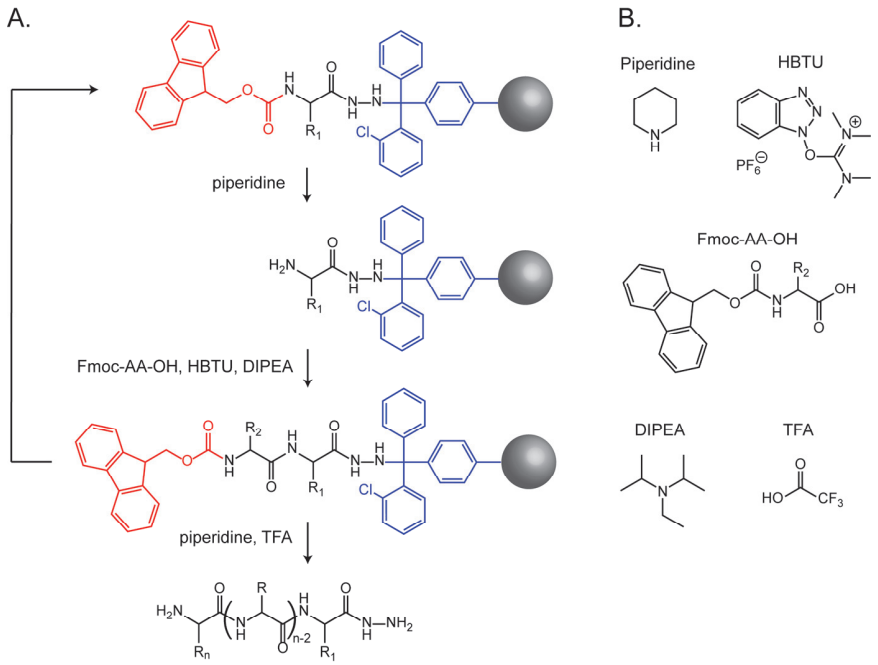

Figure 13. Fmoc SPPS with hydrazine 2-chlorotrityl resin results in peptide hydrazide. A. Each SPPS cycle starts with Fmoc deprotection by piperidine treatment, followed by reaction with Fmoc-protected pre-activated amino acid. Following piperidine deprotection of the last added amino acid, resin cleavage and side chain deprotection is performed by reacting with TFA. Fmoc group and chlorotrityl linker are indicated in red and blue, respectively. B. Structures of reagents used for Fmoc SPPS.

A wide variety of non-proteinogenic amino acids can be incorporated into peptides by SPPS. However, full length proteins are often required to study their functions. With an efficiency of ~ 99 % for each coupling steps, the length limit of SPPS is ~ 50 - 60 amino acids. For instance, the theoretical yield for the synthesis of a peptide consisting of 50 residues is 61 % (considering a 99 % yield for each cycle). Ligating two (or more) peptides represents a strategy to synthesize full length proteins.

Native chemical ligation

Several chemical strategies exist to join two unprotected peptides and obtain a native amide bond: Staudinger ligation,[311,312] KAHA ligation,[313,314] salicylaldehyde ester- mediated ligation[315] and NCL. NCL joins two peptides in aqueous conditions and produces a native peptide bond.[316,317] In NCL, the amino terminal peptide (N-fragment) harbours a C-terminal thioester, which reacts in a *trans*-thioesterification reaction with the N-terminal cysteine of the carboxyl terminal peptide (C-fragment) (Figure 14). The formed thioester then undergoes an S to N acyl shift, therefore yielding a native amide bond. Thioester peptides can be produced by Boc SPPS via benzylic thiols or alkyl thioesters.[318–320] Fmoc SPPS also enables to obtain C-terminal thioester peptides by using a diamino benzoic acid (Dbz) linker or by producing N-methyl cysteine peptides or bis(2-sulfanylethyl) amido (SEA) peptides as surrogates.[321–323] An alternative strategy to produce thioester peptides by Fmoc SPPS consists in converting *in situ* a hydrazide moiety to an azide by addition of sodium nitrite ($NaNO_2$) followed by addition of a small molecule thioester, such as 2-mercaptoethanesulfonic acid sodium (MESNa) or 4-mercaptophenylacetic acid (MPAA).[324,325] The C-terminal hydrazide group can be easily obtained via SPPS by reacting 2-chlorotrityl resin with hydrazine prior to adding the first amino acid (Figure 13).[326]

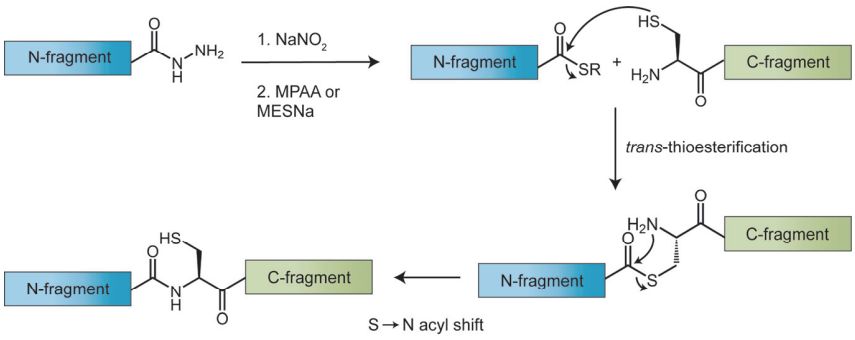

Figure 14. Mechanism of native chemical ligation. The C-terminal thioester on the N-fragment is obtained from a hydrazide moiety by reaction with $NaNO_2$ and a thiol (MPAA or MESNa). The thioester is then reacted with the N-terminal cysteine of the C-fragment in a *trans*-thioesterification. The formed thioester subsequently undergoes S to N acyl shift, yielding a native peptide bond between the N- and C-fragments.

Expressed protein ligation

Ligating two synthetic peptides via NCL limits the length of the ligation product to ~ 120 amino acids. However, most protein sequences have more residues. Successive NCLs can be performed to produce fully synthetic proteins. However, EPL represents a powerful alternative to obtain full length modified proteins. First described by Muir et al., EPL makes use of recombinant protein expression and ligates two fragments in the same way than NCL.[327] In EPL, either the C- or the N-fragment is a recombinant protein and the other fragment is a synthetic peptide that carries single or multiple Uaa(s). This high modularity enables the installation of modifications both at the N- or C-terminal end of proteins.

In case the modification to incorporate is close (< 60 residues) to the N-terminus in the amino acid sequence, the N-fragment with C-terminal thioester is obtained from a synthetic peptide hydrazide and the C-fragment is obtained recombinantly as a fusion protein with a cleavable tag at the N-terminus, such as SUMO (Figure 15. A).[328–330] Upon SUMO cleavage by the SUMO protease ULP1, the N-terminal cysteine is obtained. Other tags and their corresponding proteases were also used to produce recombinant proteins with N-terminal cysteine: factor Xa,[331] tobacco etch virus (TEV),[332] thrombin[333] and $3C^{pro}$ proteases.[334] In case the modification to introduce is close to the C-terminal (< 60 residues), the C-fragment containing the N-terminal cysteine is produced by SPPS and the N-fragment is recombinantly expressed as a fusion protein with an intein (Figure 15. B).[327] Inteins are self-excising protein domains found in various organisms, which ligate the flanking polypeptides in a process called protein (*cis*-)splicing.[335,336] Inteins also exist as split unstructured polypeptides that assemble and refold prior to (*trans*-) splicing.[337,338] The first step in splicing is an N to S acyl shift (Figure 15. B). Engineered (split) inteins, which are unable to proceed with the following steps of splicing, can be reacted with an exogenous thioester to form the N-fragment with a C-terminal thioester.[327,339–341] A split intein column can be used for one-step purification and thioester formation.[342]

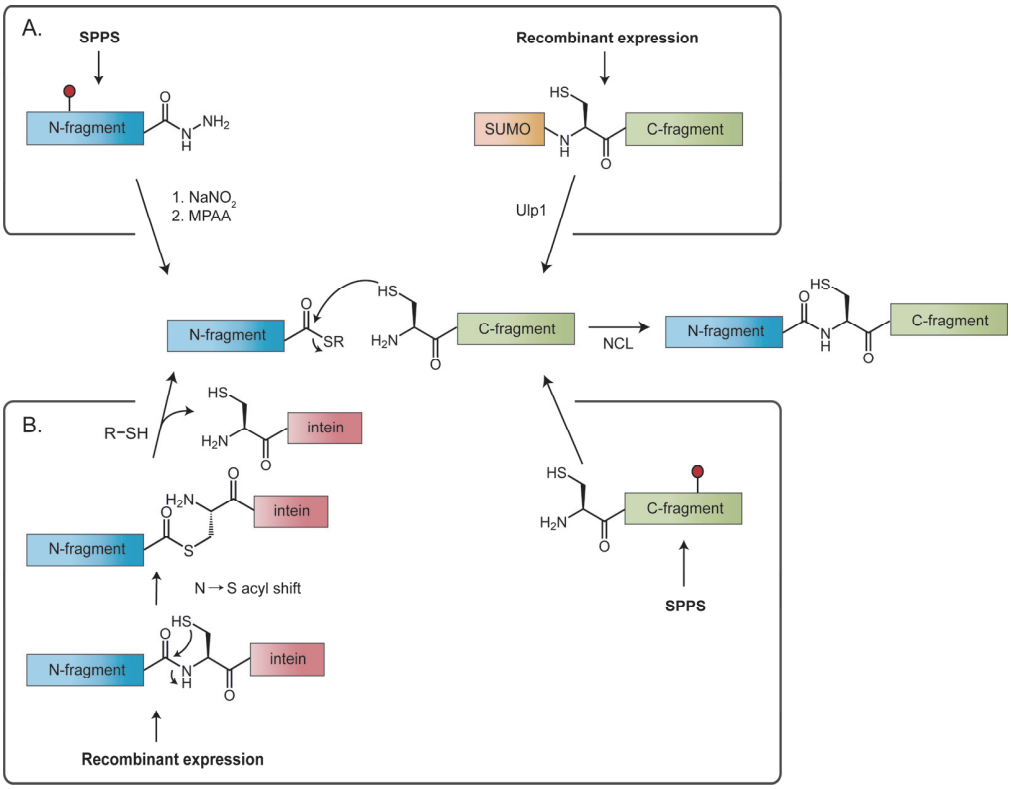

Figure 15. Expressed protein ligation is a semi-synthesis strategy that combines NCL and recombinant protein expression to obtain full length modified proteins. A. The N-fragment harbouring the modification (displayed as a red circle) is obtained by SPPS as a peptide hydrazide and is converted in situ to a thioester. The N-terminal cysteine of the C-fragment is obtained upon cleavage of the SUMO tag with ULP1. Subsequent ligation reaction yields an N-terminally modified full length protein. B. The C-fragment carrying the modification is produced by SPPS with an N-terminal cysteine. The N-fragment is expressed as a fusion protein with an intein. After N to S acyl shift, the intein thioester is reacted with an exogenous thioester to yield the N-fragment thioester. Subsequent ligation reaction leads to a C-terminally modified full length protein.

Ligation and desulfurization

The product of NCL contains a cysteine residue at the junction point. Due to scarcity of this amino acid in proteins, ligation methods at non-cysteine junction points have emerged and desulfurization represents the most widespread.[343,344] Desulfurization of cysteine results in alanine, whereas desulfurization of ß- or γ-thiol amino acids results in a variety of proteinogenic amino acids, such as valine,[345,346] phenylalanine,[347] glutamine,[348] threonine,[349] leucine,[350,351] glutamate,[352] asparate,[353] tryptophan,[354] arginine,[355] proline[356–358] as well as isopeptide bonds on lysine.[359,360] Desulfurization was originally performed with Pd/Al_2O_3 or Raney Nickel.[343,344]

However, Pd/Al_2O_3 desulfurization can cause degradation of protecting groups and Raney Nickel reaction can lead to epimerization of secondary alcohols.[361,362] For these reasons, a milder metal-free radical-based desulfurization was developed (Figure 16).[363] According to Wan & Danishefsky, the radical desulfurization starts with the removal of a hydrogen atom from cysteine by reaction with the radical initiator VA-044, resulting in a sulphur radical (Figure 16. A). In a second step, the phosphorus atom of tris(2-carboxyethyl)phosphine (TCEP) abstracts the sulphur atom from cysteine, yielding an alkyl radical. This alkyl radical is proposed to propagate the chain reaction by removing an H from other ligation products (Figure 16. B). Finally, alkyl radical yields alanine in place of the original cysteine (Figure 16. C).

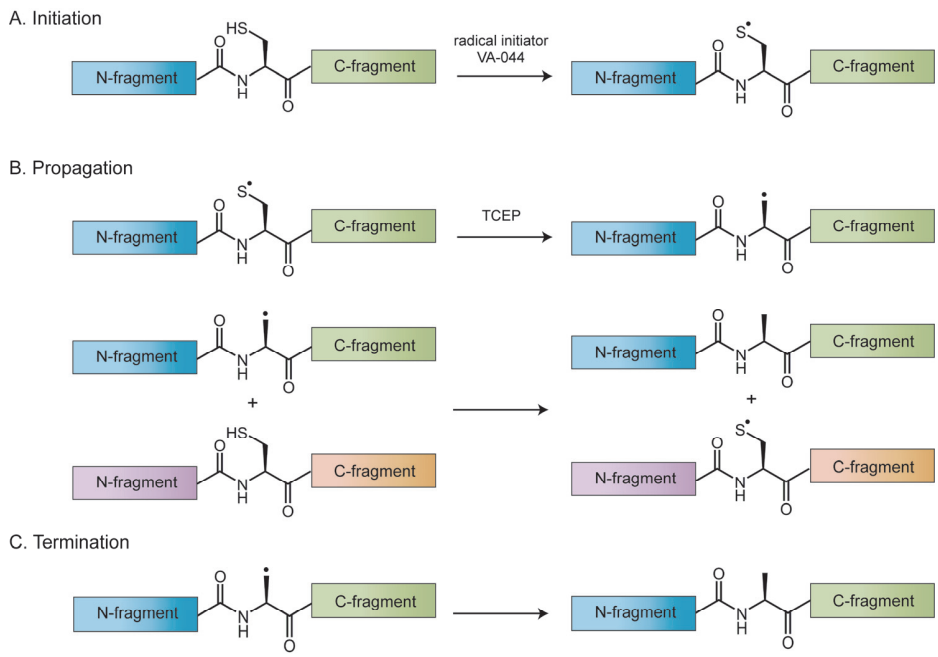

Figure 16. Mechanism of radical-based desulfurization of cysteine. A. Initiation: the radical initiator VA-044 removes H from the sulphur atom of cysteine. B. Propagation: TCEP abstracts the sulphur atom, resulting in the formation of an alkyl radical. Alkyl radicals propagate a chain reaction by removing H atoms for other cysteines. C. Termination: alanine is generated from the alkyl radical.

One-pot ligation/desulfurization reactions

In order to avoid a loss of material due to intermediate purification, one-pot ligation and desulfurization reactions would be advantageous. Typical thiols used to facilitate NCL reaction are MESNa or MPAA. The aryl thiol MPAA is incompatible

with one-pot ligation/desulfurization because it acts as a radical scavenger.[364,365] MESNa allows one-pot ligation/desulfurization; however, its kinetics is slower than MPAA.[366] Therefore, 2,2,2-trifluoroethanethiol (TFET) was proposed as efficient thiol additive that enables one-pot ligation/desulfurization.[367] Nevertheless, TFET-thioester peptides are prone to hydrolysis. Methyl thioglycolate (MTG) represents an alternative to this issue.[368]

EPL for semi-synthesis of modified histones

EPL is a method of choice for introducing PTMs in histones since lots of PTMs are located close to the N-terminus. For instance, EPL was used to produce semi-synthetic H3 or H4 with lysine acetylation,[49,73,369] serine phosphorylation[370] or lysine trimethylation on the N-terminal tail.[73,212,369] To introduce modification(s) in the middle of a protein (i.e. more than ~ 60 residues away from the C- or N-terminus in the amino acid sequence) or for introducing multiple modifications both at the C- and N-terminus, a three-piece ligation strategy can be adopted.[371,372] For instance, a three-piece approach employed a photolytic thiol-bearing ligation auxiliary to introduce ubiquitin on the ε-amino group of lysine in H2B or H2A, yielding a native isopeptide bond.[373–375]

The semi-synthetic modified histones can thus be refolded together with the other core histones, resulting in modified octamers. These octamers can then be used for nucleosome reconstitution with suitable DNA molecules such as the 601 strong nucleosome positioning sequence.[376] By constructing adapted DNA pieces, oligonucleosomes (array of multiple nucleosomes separated by linker DNA) can be reconstituted. Combining EPL with DNA ligation at non-palindromic overhangs, heterotypic nucleosome arrays can be assembled from a set of differently modified nucleosomes.[377–379] Our group recently developed a chemical route to produce asymmetrically modified nucleosomes, i.e. with sister histones harbouring different modifications.[207] Additionally, DNA barcoding allowed the creation of a nucleosome library for which each member displays a distinct combination of histone PTMs.[380] These techniques enable the access to various chromatin templates and help to address important biological questions, such as the functions and mechanisms of

chromatin-associated proteins,[379] the effects of histone PTMs on nucleosome stability and chromatin compaction[49,375] intra- and internucleosomal histone modification crosstalks,[207,212,370,373,374,380] combinatorial readout of histone PTMs[73] and chromatin effector kinetics.[75]

1.4 Aims

Histone modifications work in synergy to install chromatin states. Bivalent chromatin is the combination of H3K4me3, deposited by MLL complexes, and H3K27me3, installed by PRC2. Bivalency is proposed to keep key developmental genes in a poised state in ESCs. The subnuclear distribution of this combinatorial chromatin state in live cells as well as its dynamics is currently unknown due to a lack of suitable imaging techniques. Furthermore, how bivalent chromatin is established by PcG and TrxG proteins is unclear. In the present work, we aimed to address these questions by engineering genetically encoded sensors for bivalent chromatin and by assessing the influence of H3K4me3 and H3K27me3 on the activity of the H3K4-specific methyltransferase SET1B.

1. Currently, no live cell imaging method exists for combinations of histone marks. This prompted us to develop genetically encoded probes based on naturally occurring reader domains and apply this strategy for bivalent chromatin, a histone PTM pattern of high biological interest. After testing the probes with reconstituted nucleosomes containing site-specifically modified histones, we aimed to observe the subnuclear localization of bivalent domains in live stem cells. Furthermore, we tested the influence of epigenetic drug treatments on bivalent domains.

2. To investigate the influence of histone PTM crosstalks on the installation of H3K4 methylation, we tested the H3K4 methyltransferase activity of SET1B on symmetrically modified H3K27me3 and asymmetrically modified H3K4me3 nucleosomes.

Accomplishment of these goals yielded insights about the subnuclear organization of bivalent chromatin and about the crosstalk between H3K4me3 and H3K27me3. The realization of project 1 revealed that bivalent domains are organized

into local clusters. In project 2, we uncovered a positive feedback mechanism by which H3K4me3 stimulates SET1B activity. We expect that our work will promote future research perspectives concerning combinatorial chromatin states.

Chapter 2. Development and evaluation of genetically encoded probes selective for bivalent histone marks

The work described in this chapter is the preprint version of the article entitled "Engineered multivalent sensors to detect coexisting histone modifications in living stem cells" accepted in Cell Chemical Biology. The following persons contributed to this project: Aurore M.-F. Delachat, Nora Guidotti, Andreas L. Bachmann, Antonio C. A. Meireles-Filho, Horst Pick, Carolin C. Lechner, Cédric Deluz, Bart Deplancke, David M. Suter and Beat Fierz.

Aurore M.-F. Delachat and Beat Fierz designed the study. Aurore M.-F. Delachat performed the cloning and recombinant expression of the different probe variants as well as the *in vitro* binding assays with nucleosomes and peptides. Aurore M.-F. Delachat, Carolin C. Lechner and Nora Guidotti made the modified peptides, histones and octamers used in the different binding assays. Aurore M.-F. Delachat and Horst Pick performed the immunofluorescence experiments. Horst Pick performed the confocal imaging of HEK cells and helped with the mESC culture and imaging. Aurore M.-F. Delachat performed the confocal imaging of ESCs, the local maxima count and the Western Blot (WB) of acid-extracted histones. Aurore M.-F. Delachat and Antonio C. A. Meireles-Filho performed the ChIP experiments.

2.1 Project background and outline

Mapping the genomic location of PTM patterns is crucial to understand chromatin function and organization. As discussed in 1.2.3, ChIP-based methods enable to map histone PTMs genome-wide and to detect the coexistence of multiple PTMs (Re-ChIP). ChIP can also be used to obtain quantitative information with a single-nucleosome precision.[261] Additionally, immunoprecipitation of mononucleosomes can be coupled with MS or single-molecule TIRF microscopy to have insights about nucleosome symmetry.[246,262,263] Mapping genomic location of histone marks in a combinatorial manner is essential to unravel the function of chromatin states, but detecting subnuclear organization of PTM patterns is also of

high importance. Immunofluorescence coupled with confocal microscopy is widely used to visualize subcellular distribution of protein PTMs at the single cell level. However, this technique does not enable to prove the coexistence of multiple PTMs in close proximity as the focal volume is typically ~ 250 nm × 500 nm. To circumvent this limitation, Hattori et al. developed a single cell imaging method based on proximity ligation assay.[381] The authors used a pair of modification-specific antibodies as templates for rolling circle amplification that takes place only when the two histone PTMs of interest are in close proximity. Coupled with confocal microscopy, this method allows the detection of two different histone marks in close vicinity (~ 30 nm) and therefore the observation of the subnuclear organization of chromatin states. In a different approach, antibody fragments targeting two different epitopes were used to construct bispecific detection reagents for immunohistochemistry.[382]

The above-mentioned techniques rely on the availability and specificity of antibodies (or antibody fragments) for the PTMs of interest. These criteria are not always straightforward to attain. Furthermore, antibody-based methods require sample fixation and permeabilisation, rendering impossible live cell imaging. Currently, imaging strategies to visualize patterns of histone PTMs in living cells are very limited. However, such methods exist for single histone marks. For instance, fluorescently labelled antibody fragments loaded in live cells localize at modified histones, enabling to visualize their subnuclear organization.[383,384] Genetically encoded sensors are more suitable for long term tracking and for living organisms. Genetic fusions of a fluorescent protein (FP) with an antibody fragment were used to follow histone acetylation or methylation in live cells.[385–387] The use of naturally occurring reader domains to design sensors targeting histone modifications represents an interesting alternative to antibodies. Fluorescence resonance energy transfer (FRET) sensors using bromo-, chromodomains and 14-3-3(τ) were used to follow levels of histone acetylation, methylation and phosphorylation, respectively.[388–393] Reader domains were also used in bioluminescence or bimolecular fluorescence complementation (BiFC) approaches.[394–396] Recently, bromo- and chromodomains were used in tandem repeats separated by flexible

linkers to target histone acetylation and methylation, respectively.[397,398] In these studies, the authors highlighted the enhanced affinity for the PTM of interest in the case of tandem reader domains compared to single reader domain constructs.

The lack of live cell imaging method for chromatin states prompted us to set up a modular strategy for the development and optimization of genetically encoded probes targeting combinations of histone PTMs. We took advantage of the abundance of reader domains in nature to design fusion constructs that bind selectively to a pattern of histone marks. By including a FP in the fusion proteins, we used direct fluorescence as readout for microscopic purposes. We designed probes to target one of the most intriguing pattern of histone PTMs, bivalent chromatin.

First of all, we designed several variants for which different parameters such as linker length were varied. The selectivity of recombinant sensors was assessed *in vitro* with binding assays using reconstituted nucleosomes with semi-synthetic methylated histones. In this step, the best performing variant was selected and its affinity for modified histone peptides was then quantified. Following this validation step, the recombinant sensor was used to visualize bivalent chromatin in mouse ESCs. The change of subnuclear distribution of the sensor upon reader domain mutations confirmed the selectivity for bivalent chromatin *in vivo*. Finally, the probe was used to observe the influence of epigenetic small molecule modulators on bivalent domains.

2.2 Results

2.2.1 Design of variants

As described in 1.1.2, individual reader domains have mild affinities for their target histone mark, whereas multivalent chromatin effectors reach higher affinities for PTM patterns. We sought to use this multivalency effect to envision bivalent chromatin probes as engineered proteins composed of two reader domains, one specific for H3K4me3 and one specific for H3K27me3 (Figure 17. A). We named these sensors chromatin sensing multivalent probes, abbreviated cMAPs.

Additionally to two reader domains, cMAPs contain a FP as a reporter as well as nuclear localization signals (NLS) (one at the N- and one at the C-terminus) and purification/detection tags (FLAG tag at the N-terminus, His_6 and HA tags at the C-terminus). FP and reader domains are connected by linkers into one single fusion protein.

Additionally to using direct fluorescence as readout, we aimed to develop split engineered sensors for BiFC applications.[399] For this reason, Venus was chosen for its good performance in BiFC assays[400] and was initially placed in the middle of the construct (with flanking reader domains). Furthermore, Venus is characterized by a fast chromophore maturation.[401] Among the known natural reader domains for H3K4me3 and H3K27me3, we carefully chose well characterized ones. The chromodomain (CD) of the *Drosophila* polycomb (Pc) protein binds to H3K27me3 peptide with a measured affinity of 5 µM (Figure 17. B, left panel).[62] Crystal structures revealed that tryptophan and tyrosine residues form an aromatic cage to accommodate the hydrophobic charged ammonium group.[61,62] Importantly, Pc CD discriminates between H3K27me3 and H3K9me3 with a 25-fold selectivity.[62] Concerning H3K4me3-specific reader domains, the plant homeodomain (PHD) of the human transcription initiation factor TFIID subunit 3 (TAF3) binds to H3K4me3 peptide with an affinity of 160 nM (Figure 17. B, right panel).[63] Crystal structure indicates also an aromatic cage as the trimethyllysine binding site.[402] In order to equilibrate the binding affinities between the two reader domains, another PHD was chosen with a K_D in the same range than the one of Pc CD; the mouse ING2 PHD binds H3K4me3 peptide with a K_D of 1.5 µM.[403] From chain statistics, an average distance of 4 nm can be estimated for H3K4me3 and H3K27me3 occurring on the same histone tail, and a distance of 10 nm when they reside on different H3 tails (Figure 17. C). As H3 tails are unstructured and highly flexible, these distances are subjected to high variations. Based on these considerations, we designed linkers that have an appropriate length to span the distance between the two target PTMs but that are not excessively long, which would result in a poor gain in affinity for the multivalent effect.

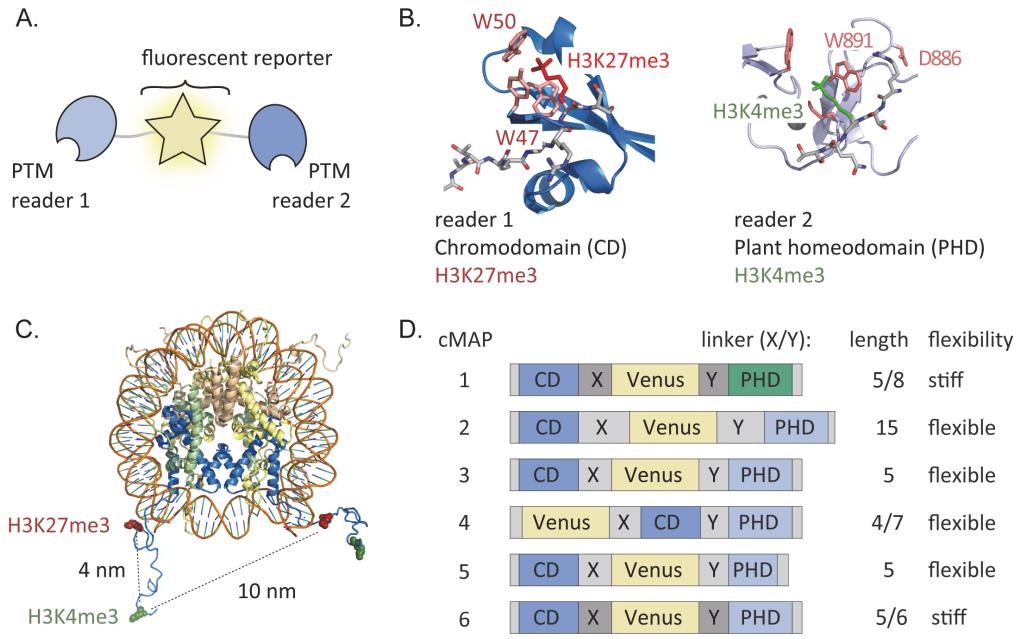

Figure 17. Design of the sensors cMAPs. A. Scheme of cMAPs general structure. B. Crystal structures of the reader domains. Ribbon representation of Pc CD (PDB: 1PDQ) and TAF3 PHD (PDB: 5C13) in complex with H3K27me3 and H3K4me3 peptides (shown as sticks), respectively. H3K27me3 and H3K4me3 residues are displayed in red and green, respectively. The residues displayed in pink form the binding site. C. Ribbon representation of a bivalent nucleosome (PDB: 1KX5) with estimations of the distances between the residues H3K4 and H3K27. H3, H4, H2A and H2B are represented in blue, green, yellow and orange, respectively. D. Domain arrangement of cMAP1-6. Linker length and flexibility are indicated for each variant.

We first designed cMAP1 and cMAP6 where Venus is flanked by Pc CD at its N-terminus and by ING2 and TAF3 PHD at its C-terminus, respectively (Figure 17. D). The alpha helical linker between the PHD and bromodomain of BPTF was previously demonstrated to be required for optimal bivalent binding to H3K4me3/H4K16ac nucleosomes.[404] Therefore, we thought to use polyproline stretches as linkers between Venus and the reader domains for cMAP1 and cMAP6. In aqueous solvents, polyproline stretches adopt a left-handed helical conformation (called PolyProline II) with a 3-fold symmetry. Therefore a polyproline linker fixes the orientation of the two linked domains to ~ 120 ° between each other and decreases greatly the flexibility of the protein.

Flexible linkers composed of GGGGS repeats are widely used in protein engineering.[382,387–390,398] We then derived cMAP2 and cMPA3 from the cMAP6 construct replacing the polyproline linkers by 3 and 1 GGGGS repeats, respectively.

We then speculated that tandem reader domains (only separated by a short flexible linker) might bind to their target with higher affinity compared to when they are spaced with Venus. Thus, we designed cMAP4 in which the order of the domains is permuted to have Pc CD and TAF3 PHD only separated by a 7 amino acid flexible linker.

An analysis of cMAP3 sequence with a nucleolar localization sequence prediction tool revealed a potential nucleolar localization signal at the C-terminus of TAF3 PHD (Figure 18. A).[405] In order to avoid any influence of this amino acid stretch in the subnuclear localization of the sensor, we created cMAP5 in which 9 residues at the N-terminus of TAF3 PHD were deleted. Analysis of cMAP5 sequence with the same tool revealed a prediction score below the critical threshold for the entire sequence (Figure 18. B). Importantly, structural analysis of the TAF3 PHD suggests that the 9 deleted residues are not critical for H3K4me3 binding.[402]

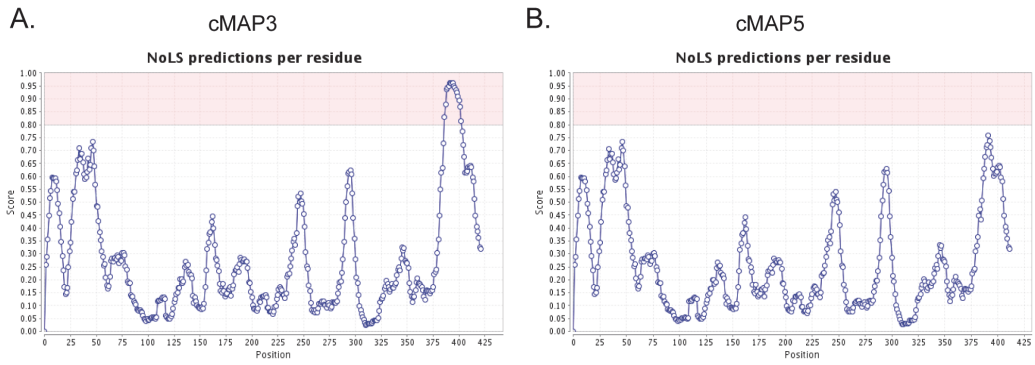

Figure 18. Scores for predicted nucleolar localization sequence for cMAP3 (A) and cMAP5 (B) obtained on http://www.compbio.dundee.ac.uk/www-nod/

Several characteristics were varied in the 6 cMAP constructs described here: length and stiffness of linkers, domain organization, origin and length of PHD. By varying these different parameters, we aspired to optimize the selectivity of cMAPs for bivalent nucleosomes (i.e. carrying both H3K4me3 and H3K27me3) versus monovalent nucleosomes (i.e. carrying either H3K4me3 or H3K27me3). To test the

binding and selectivity of these 6 constructs, we performed *in vitro* binding assays with recombinant cMAPs and a nucleosome library composed of differently modified semi-synthetic histones.

2.2.2 Recombinant expression and purification of the probes

In order to test their binding towards bivalent marks, cMAP1-6 were recombinantly expressed. The constructs were cloned in bacterial vectors and expressed in *E. Coli* by IPTG induction. cMAP proteins were then purified by Ni-NTA affinity purification. Importantly, the resin-bound proteins were washed with a buffer containing 1M NaCl to eliminate genomic DNA. SDS PAGE analysis showed that the purified proteins were at least 95 % pure from proteinic contaminants (Figure 19. A). Proteins concentration was determined by UV-vis spectroscopy from the Venus absorption peak (Figure 19. B). 260 / 280 nm ratios between 0.73 and 0.96 were measured, indicating a negligible contamination by nucleic acid. Yields between ~ 1 and ~ 23 mg of protein per litre culture were obtained. Following buffer exchange, Ni-NTA purified cMAP proteins were used for binding assays.

A deeper analysis of cMAP2, 3 and 6 was performed by subjecting proteins to anion exchange chromatography (AIEX). A gradually increasing salt concentration (from 50 mM to 1 M) was used for elution with a constant pH at 7.8. At this pH value, cMAP2, 3 and 6 are slightly positively charged, as their pI is 8.06. As expected, proteins eluted in the flow-through (between 0 and 7 mL elution volume) (Figure 19. C). Two smaller peaks appeared at elution volumes ~ 27 mL and ~ 34 mL. After analysis by UV-vis spectroscopy (Table 2) and SDS PAGE (Figure 19. C), we determined that the peak at ~ 27 mL elution volume contained mainly cMAP protein, whereas the peak at ~ 34 mL elution volume contained mainly nucleic acid. This AIEX analysis revealed that cMAPs have a low but non-negligible affinity for DNA. cMAP2 was further subjected to size exclusion chromatography (SEC) and the fractions were analysed by SDS PAGE (Figure 19. D). The obtained single peak indicated that the protein was monomeric.

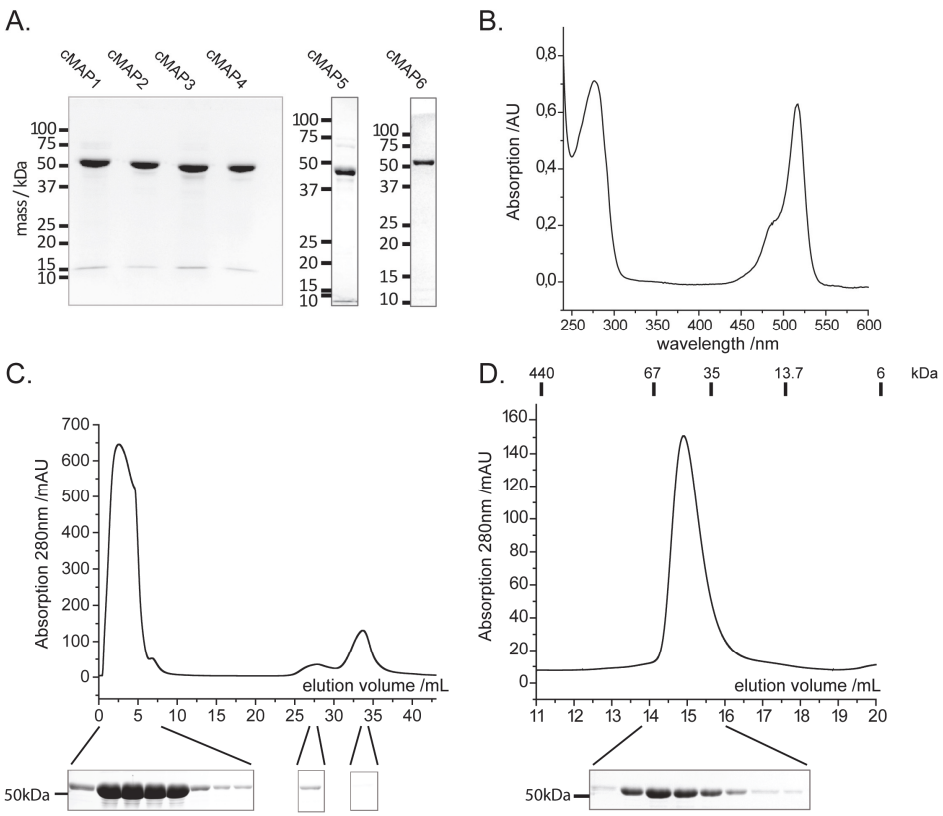

Figure 19. Purification of cMAPs. A. SDS PAGE analysis of cMAP1-6 purified by Ni-NTA affinity column. B. Representative UV-vis absorption spectrum of cMAP proteins. C. AIEX trace of cMAP2 and SDS PAGE analysis of the corresponding fractions. D. Gel filtration trace of cMAP2 and SDS PAGE analysis of the corresponding fractions.

	Ratio 260 / 280	**Absorption at 515 nm**	**Main component**
Peak between 0 and 7 mL elution volume	0,54	Yes	cMAP
Peak at ~ 27 mL elution volume	0,67	Yes	cMAP
Peak at ~ 34 mL elution volume	1,88	No	Nucleic acid

Table 2. UV-vis spectrometric analysis of the AIEX fractions of cMAP2 purification.

With pure recombinant cMAPs in hands, we then synthesized modified histone peptides and full length histones via EPL.

2.2.3 Production of nucleosomes using site-specifically modified, semi-synthetic histone proteins

Site-specifically modified histone peptides and full length histone proteins were synthesized via EPL. Histone peptides were used for quantitative measurement of affinities, whereas full length modified histones were used for nucleosome reconstitution further employed in binding assays (described in 2.2.4).

Peptides **1b** (H3(1-14)K4me3), **1d** (H3(1-28)K27me3) and **1e** (H3(1-28)K4me3K27me3) corresponding to residues 1-14 or 1-28 of human H3.1 with lysine 4 or/and 27 trimethylated were synthesized on hydrazine 2-chlorotrityl resin by SPPS using the Fmoc protection strategy. The peptides hydrazide **1b**, **1d** and **1e** were purified by RP-HPLC on a preparative scale and characterized by analytical RP-HPLC and ESI-MS (Figures 20 - 22, Table 3).

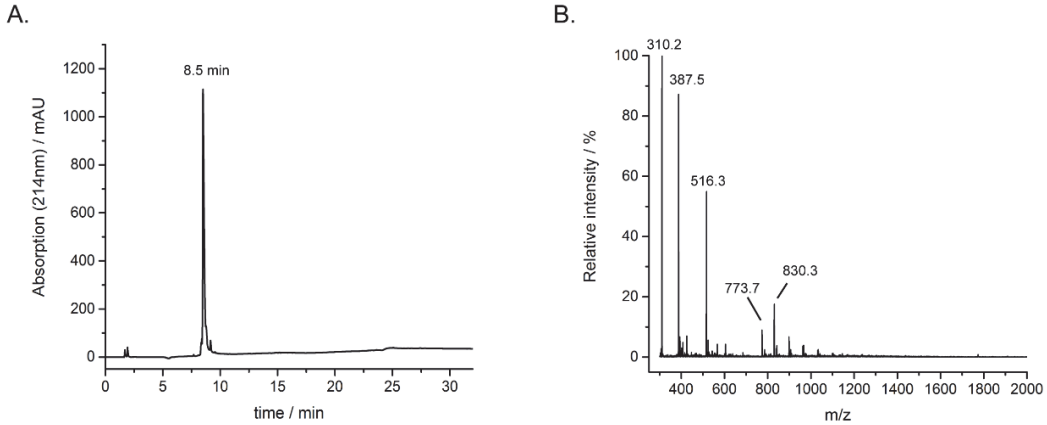

Figure 20. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **1b** (H-ARTK(me3)QTARKSTGGK-NH-NH_2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 1545.9 Da, found: 1545.8 Da, 310.2 $[M+5H]^{5+}$, 387.5 $[M+4H]^{4+}$, 516.3 $[M+3H]^{3+}$, 773.7 $[M+2H]^{2+}$, 830.3 $[M+TFA+2H]^{2+}$.

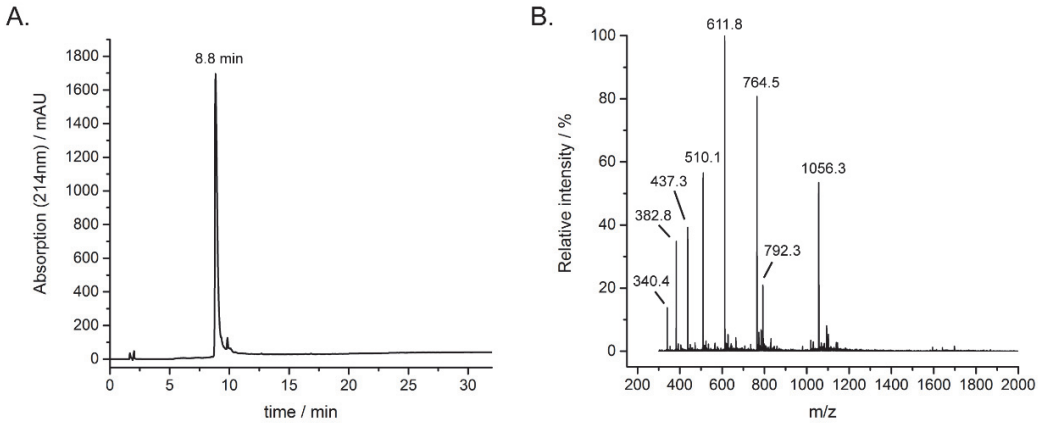

Figure 21. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **1d** (H-ARTKQTARKSTGGKAPRKQLATKAARK(me3)S-NH-NH_2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 3053.8 Da, found: 3054.3 Da, 340.4 $[M+9H]^{9+}$, 382.8 $[M+8H]^{8+}$, 437.3 $[M+7H]^{7+}$, 510.1 $[M+6H]^{6+}$, 611.8 $[M+5H]^{5+}$, 764.5 $[M+4H]^{4+}$, 792.3 $[M+TFA+4H]^{4+}$, 1056.3 $[M+TFA+3H]^{3+}$.

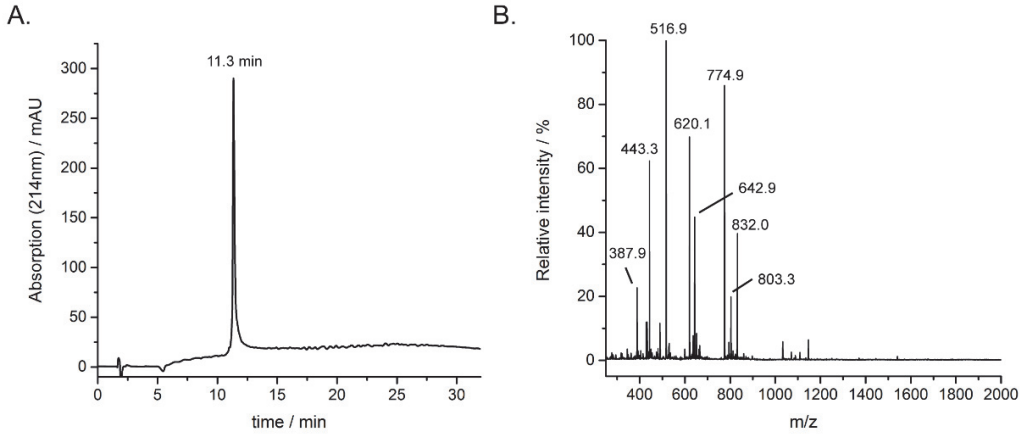

Figure 22. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **1e** (H-ARTK(me3)QTARKSTGGKAPRKQLATKAARK(me3)S-NH-NH_2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 3095.9 Da, found: 3095.6 Da, 387.9 $[M+8H]^{8+}$, 443.3 $[M+7H]^{7+}$, 516.9 $[M+6H]^{6+}$, 620.1 $[M+5H]^{5+}$, 642.9 $[M+TFA+5H]^{5+}$, 774.9 $[M+4H]^{4+}$, 803.3 $[M+TFA+4H]^{4+}$, 832.0 $[M+2TFA+4H]^{4+}$.

Name	Sequence	MW calculated (Da)	MW observed (Da)	Yields
1b	H-ARTK(me3)QTARKSTGGK-NH-NH_2	1545.9	1545.8	24 %
1d	H-ARTKQTARKSTGGKAPRKQLATKAARK(me3)S-NH-NH_2	3053.8	3054.3	38 %
1e	H-ARTK(me3)QTARKSTGGKAPRKQLATKAARK(me3)S-NH-NH_2	3095.9	3095.6	15 %

Table 3. Peptides **1b**, **1d** and **1e** used for Expressed Protein Ligation to produce site-specifically modified histone peptides and full length histone proteins.

Synthesis of modified full length histones **3b** (H3K4me3), **3d** (H3K27me3) and **3e** (H3K4me3K27me3) was performed by ligation of peptides **1b**, **1d** or **1e** with truncated versions of H3 histone protein, followed by desulfurization. For this purpose, N-terminal truncated variants of human H3.1_C110A variant **2** (H3(Δ1-14, A15C)) and **2'** (H3(Δ1-28, A29C)) were recombinantly expressed as N-terminal fusions to a His_6-SUMO tag. Following Ni-NTA affinity purification, refolding and cleavage of the His_6-SUMO tag from the His_6-SUMO-H3(Δ1-14, A15C) and His_6-SUMO-H3(Δ1-28, A29C) fusion proteins were achieved by dialysis with recombinant SUMO protease ULP1. Histones **2** and **2'** were purified by RP-HPLC on a preparative scale and characterized by analytical RP-HPLC and ESI-MS (Figures 23 - 24).

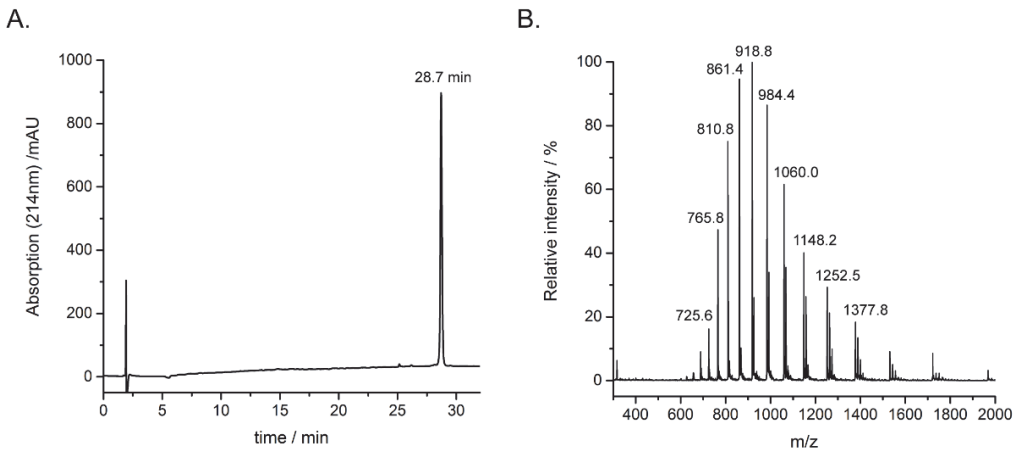

Figure 23. RP-HPLC chromatogram (A) and ESI-MS (B) of purified N-terminal truncated histone **2** H3(Δ1-14, A15C). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 13769 Da, observed: 13767 Da, 725.6 $[M+19H]^{19+}$, 765.8 $[M+18H]^{18+}$, 810.8 $[M+17]^{17+}$, 861.4 $[M+16H]^{16+}$, 918.8 $[M+15H]^{15+}$, 984.4 $[M+14H]^{14+}$, 1060.0 $[M+13H]^{13+}$, 1148.2 $[M+12H]^{12+}$, 1252.5 $[M+11H]^{11+}$, 1377.8 $[M+10H]^{10+}$.

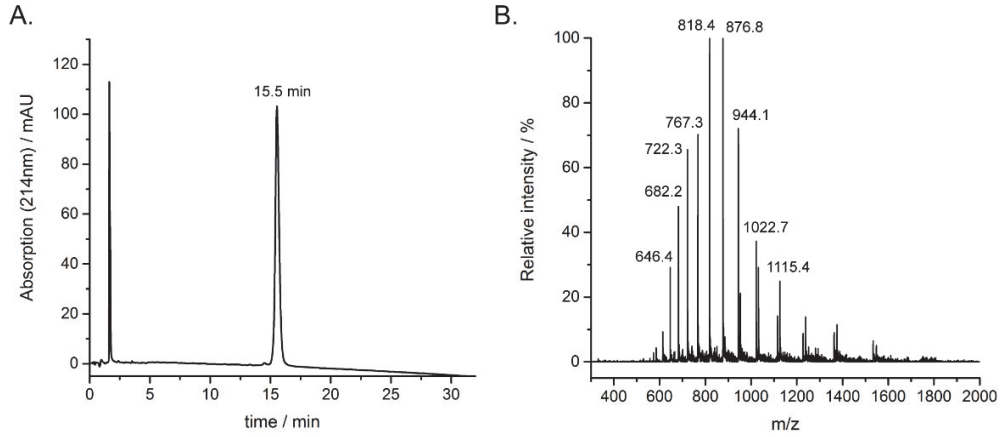

Figure 24. RP-HPLC chromatogram (A) and ESI-MS (B) of purified N-terminal truncated histone **2'** H3(Δ1-28, A29C). Gradient 47 % - 62 % solvent B in 30 min. Calculated mass: 12261 Da, observed: 12261 Da, 646.4 $[M+19H]^{19+}$, 682.2 $[M+18H]^{18+}$, 722.3 $[M+17H]^{17+}$, 767.3 $[M+16H]^{16+}$, 818.4 $[M+15H]^{15+}$, 876.8 $[M+14H]^{14+}$, 944.1 $[M+13H]^{13+}$, 1022.7 $[M+12H]^{12+}$, 1115.4 $[M+11H]^{11+}$.

Ligation and subsequent desulfurization of truncated histone **2** with peptide **1b** yielded full length modified histone **3b**. In the same way, ligation and desulfurization of truncated histone **2'** with peptide **1d** and **1e** yielded full length modified histones **3d** and **3e**, respectively. Histones **3d** and **3b** were synthesized by standard EPL. In a typical ligation reaction, between 1 and 2 µmol of H3 tail peptide

hydrazide was dissolved in an acidic buffer to a final concentration of 2.5 mM. The peptide solution was kept at - 10 °C and $NaNO_2$ was added to a final concentration of 18 mM. After mixing, the peptide was incubated at - 10 °C for 20 min, yielding the peptide-azide. Typically, 0.6 µmol of truncated recombinant H3 protein was dissolved in a buffer with 0.2 M MPAA to a final concentration of 1 mM. The molar ratios peptide / protein was 4:1. The protein in MPAA ligation buffer was added to the peptide-azide and the pH was adjusted to 7.0. The reaction was monitored by analytical RP-HPLC. After 1 h, TCEP was added to a final concentration of 20 mM. When the reaction had reached completion, ligation product was purified by semi-preparative RP-HPLC. The ligation product was characterized by analytical RP-HPLC and ESI-MS. The obtained protein was then desulfurized to convert the cysteine into alanine at the ligation site. For this purpose, the protein was dissolved in a buffer containing 250 mM TCEP to a final concentration of 0.5 mM. Glutathione and VA-044 radical starter were added to final concentrations of 40 mM and 20 mM, respectively. The pH was adjusted to 6.5 and the reaction was allowed to proceed at 37 °C. The reaction was monitored by analytical RP-HPLC and ESI-MS. When the reaction had reached completion, desulfurization product was purified by semi-preparative RP-HPLC. The purified protein was characterized by analytical RP-HPLC and ESI-MS (Figures 25 – 26, Table 4).

For histones **3e** and **3b**, a previously described one-pot ligation-desulfurization strategy was used.[207,367] Briefly, between 1 and 2 µmol of H3 tail peptide hydrazide was dissolved in ligation buffer to a final concentration of 10 mM and the pH readjusted to 3.0. The peptide solution was kept at - 20 °C and $NaNO_2$ was added to a final concentration of 15 mM. After mixing, the peptide was incubated at - 20 °C for 5 min. Subsequently, TFET was added to a final concentration of 1 M and the pH of the peptide solution was adjusted to 6.8. The peptide was incubated for 10 min at RT and transferred to a tube with lyophilized, truncated recombinant H3 protein. The molar ratio peptide / protein was 7:1 or 5:1. A buffer containing 0.5 M TCEP was added to a final concentration of 10 mM TCEP and the reaction was allowed to proceed at 25 °C under Argon atmosphere. The reaction was monitored by analytical RP-HPLC. When the reaction had reached completion, desulfurization

was performed in the same tube without prior purification of the ligation product. A buffer containing 0.5 M TCEP was added to a final concentration of 0.25 M TCEP in the reaction solution. Then, glutathione and VA-044 radical starter were added to final concentrations of 40 mM and 20 mM, respectively. The reaction was incubated at 42 °C and monitored by analytical RP-HPLC and ESI-MS. When the reaction had reached completion, desulfurization product was purified by semi-preparative RP-HPLC. The purified protein was characterized by analytical RP-HPLC and ESI-MS (Figure 27, Table 4).

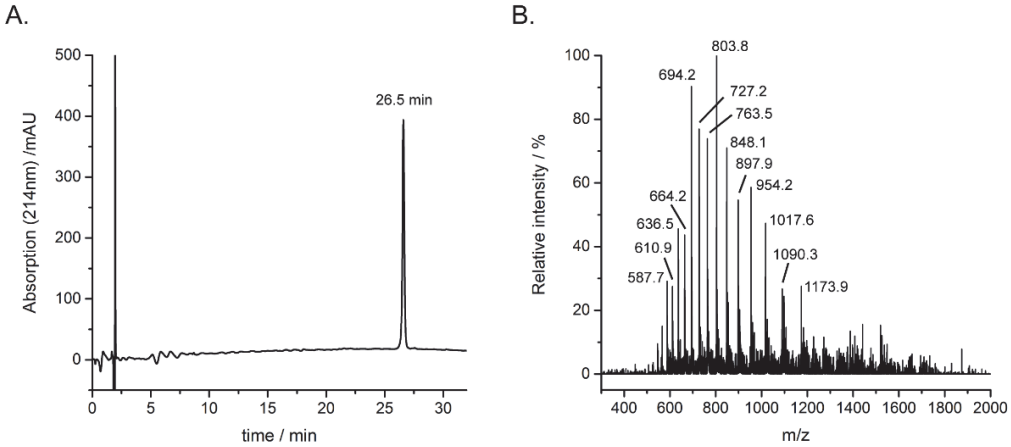

Figure 25. RP-HPLC chromatogram (A) and ESI-MS (B) of purified histone **3b** H3K4me3. Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 15251 Da, observed: 15250 Da, 587.7 $[M+26H]^{26+}$, 610.9 $[M+25H]^{25+}$, 636.5 $[M+24H]^{24+}$, 664.2 $[M+23H]^{23+}$, 694.2 $[M+22H]^{22+}$, 727.2 $[M+21H]^{21+}$, 763.5 $[M+20H]^{20+}$, 803.8 $[M+19H]^{19+}$, 848.1 $[M+18H]^{18+}$, 897.9 $[M+17H]^{17+}$, 954.2 $[M+16H]^{16+}$, 1017.6 $[M+15H]^{15+}$, 1090.3 $[M+14H]^{14+}$, 1173.9 $[M+13H]^{13+}$.

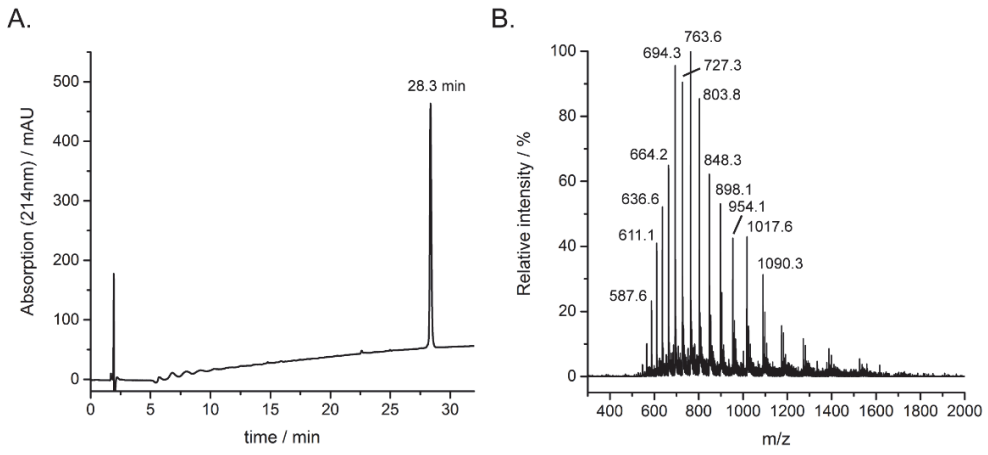

Figure 26. RP-HPLC chromatogram (A) and ESI-MS (B) of purified histone **3d** H3K27me3. Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 15251 Da, observed: 15252 Da, 587.6 $[M+26H]^{26+}$, 611.1 $[M+25H]^{25+}$, 636.6 $[M+24H]^{24+}$, 664.2 $[M+23H]^{23+}$, 694.3 $[M+22H]^{22+}$, 727.3 $[M+21H]^{21+}$, 763.6 $[M+20H]^{20+}$, 803.8 $[M+19H]^{19+}$, 848.3 $[M+18H]^{18+}$, 898.1 $[M+17H]^{17+}$, 954.1 $[M+16H]^{16+}$, 1017.6 $[M+15H]^{15+}$, 1090.3 $[M+14H]^{14+}$.

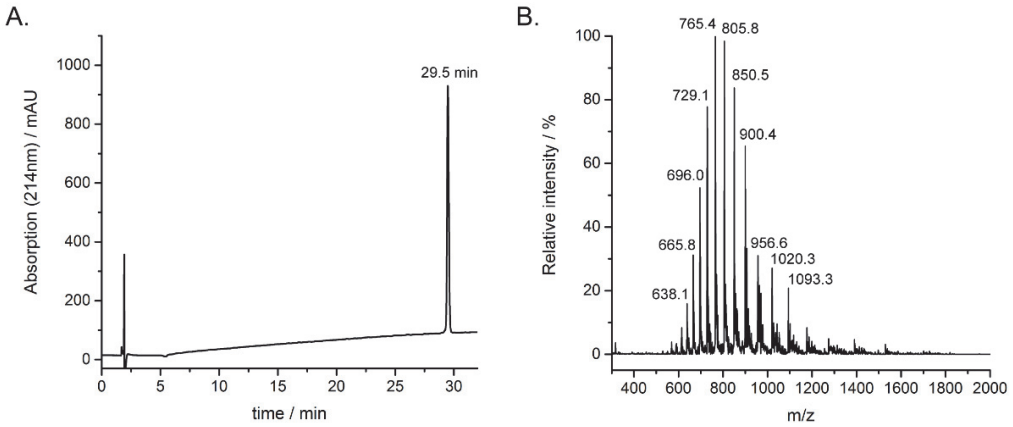

Figure 27. RP-HPLC chromatogram (A) and ESI-MS (B) of purified histone **3e** H3K4me3K27me3. Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 15293 Da, observed: 15290 Da, 638.1 $[M+24H]^{24+}$, 665.8 $[M+23H]^{23+}$, 696.0 $[M+22H]^{22+}$, 729.1 $[M+21H]^{21+}$, 765.4 $[M+20H]^{20+}$, 805.8 $[M+19H]^{19+}$, 850.5 $[M+18H]^{18+}$, 900.4 $[M+17H]^{17+}$, 956.6 $[M+16H]^{16+}$, 1020.3 $[M+15H]^{15+}$, 1093.3 $[M+14H]^{14+}$.

Name	Modification	Synthesis method	Yield
3b	H3K4me3	Standard	12 %
3b	H3K4me3	One-pot	33 %
3d	H3K27me3	Standard	26 %
3e	H3K4me3 and H3K27me3	One-pot	28 %

Table 4. Modified, semi-synthetic histones H3 proteins.

For the synthesis of histone **3c** (H3K9me3), refer to Kilic et al., *Nat. Commun.*, 2015.[75]

In order to refold modified octamers with histones **3b**, **3d** and **3e**, the core histones H2A, H2B and H4 are required. Furthermore, for refolding of unmodified octamer, unmodified H3 is needed. Human wild-type H4, H3_C110A, H2A and H2B were then recombinantly expressed and purified by cation exchange. Octamer refolding was performed by dissolving stoichiometric quantities of lyophilized core histones in a denaturing buffer followed by dialysis in a 2 M NaCl containing buffer. The scale was 100 - 500 µg of histone H3 with an overall protein concentration of 74 µM. The obtained crude octamers were purified by SEC (Figure 28. A). Fractions containing pure octamers were pooled and concentrated to the range of 30 - 60 µM. SDS PAGE analysis indicates equal amounts of each core histone (Figure 28. B).

Several types of binding assays were performed with cMAP proteins and nucleosomes (described in 2.2.4). For some of them, unlabelled reconstituted nucleosomes were used. To this end, unlabelled 601 DNA was obtained by digestion of a plasmid containing 32 repeats of the 601 sequence separated by restriction sites. For one of the binding assays, biotin-labelled nucleosomes were required. The biotinylated 166 bp DNA fragment with a 19 bp linker prior to the strong nucleosome positioning sequence 601 was produced by PCR with a biotinylated primer.[73] Mono-nucleosomes were reconstituted either at a small scale by dilution or at a larger scale by dialysis.[406] In the first method, typically 10 or 20 pmol of 601 DNA were mixed with the respective histone octamers in 2 M NaCl in a volume of 10 µL. Following several additions of 10 mM Tris-HCl pH 7.6, the final concentration of NaCl was 0.2

M. In the second method, nucleosomes were reconstituted in 70 µl sample volume by mixing 60 pmol of DNA with the respective histone octamer at high salt conditions (2 M KCl) followed by gradient dialysis into low salt conditions (0.25 M KCl). Nucleosome concentrations were determined by UV quantification and their quality was assessed by native gel electrophoresis (Figure 28. C).

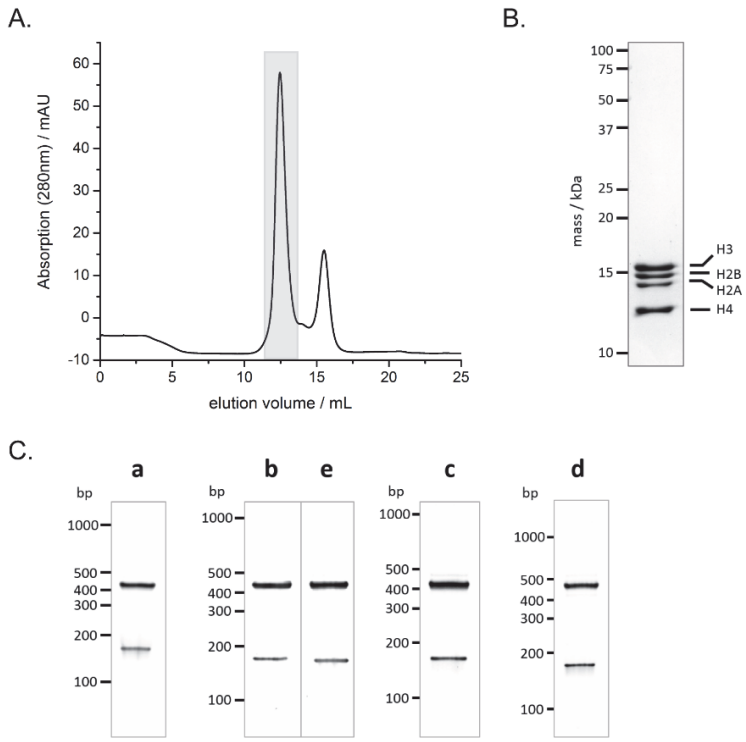

Figure 28. A. Gel filtration purification of H3K4me3 octamers. The shaded region indicates the fractions containing pure octamers. B. SDS PAGE analysis of purified H3K4me3 octamers. C. Native PAGE analysis of reconstituted biotinylated nucleosomes (**a**: unmodified nucleosomes; **b**: H3K4me3 nucleosomes; **c**: H3K9me3 nucleosomes; **d**: H3K27me3 nucleosomes; **e**: H3K4me3K27me3 nucleosomes)

With the goal of reconstituting *trans* bivalent nucleosomes, we mixed histones **3b** and **3d** in equimolar ratio during octamer refolding and used these octamers for nucleosome reconstitution. Histones **3b** and **3d** are then statistically distributed and these nucleosomes (called **e'**) consist of 50 % of *trans* bivalent nucleosomes (H3K4me3/H3K27me3), 25 % of H3K4me3 nucleosomes and 25 % of H3K27me3 nucleosomes. To compare **e'** nucleosomes with monovalent nucleosomes, we reconstituted **b'** and **d'** nucleosomes for which we mixed

unmodified histone H3 and H3K4me3 (**3b**) or H3K27me3 (**3d**) histones in equimolar ratios. Table 5 summarizes the modifications carried by the different nucleosomes used in binding assays.

Nucleosomes	Modifications on the two H3 copies
a	Unmodified/unmodified
b	H3K4me3/H3K4me3
c	H3K9me3/H3K9me3
d	H3K27me3/H3K27me3
e	H3K4me3K27me3/H3K4me3K27me3
b'	2:1:1 ratio of H3K4me3/unmodified; H3K4me3/H3K4me3; unmodified/unmodified
d'	2:1:1 ratio of H3K27me3/unmodified; H3K27me3/H3K27me3; unmodified/unmodified
e'	2:1:1 ratio of H3K4me3/H3K27me3; H3K4me3/H3K4me3; H3K27me3/H3K27me3

Table 5. Reconstituted nucleosomes employed in binding assays.

Beside binding assays with nucleosomes, we measured affinities of cMAP3 for monovalent and bivalent H3 histone peptides (described in 2.2.4). To this end, we synthesized peptides **5b** (H3(1-35)K4me3 V35Y), **5d** (H3(1-35)K27me3 V35Y) and **5e** (H3(1-35)K4me3K27me3 V35Y). Here, we used the V35Y mutation to be able to quantify peptide concentration by UV-vis spectrometry. Since we had access to peptides **1b**, **1d** and **1e**, we adopted a ligation/desulfurization strategy with the adaptor peptides **4** (H3(15-35) A15C,V35Y) and **4'** (H3(29-35) A29C,V35Y). Peptides **4** and **4'** correspond to residues 15-35 or 29-35 of human H3.1 with N-terminal alanine (15 or 29) mutated to cysteine and valine 35 mutated to tyrosine. These peptides were synthesized on rink amide resin by SPPS, using the Fmoc protection and HBTU activation strategy. Crude peptides were purified by RP-HPLC

on a preparative scale. Peptides **4** and **4'** were then characterized by analytical RP-HPLC and ESI-MS (Figure 29 - 30, Table 6).

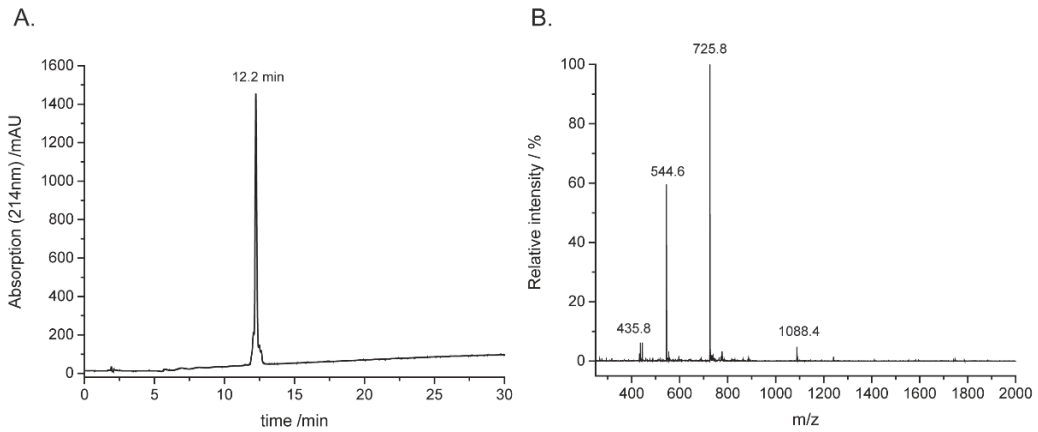

Figure 29. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **4** (H-CPRKQLATKAARKSAPATGGY-NH2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 2174.6 Da, found: 2174.4 Da, 435.8 $[M+5H]^{5+}$, 544.6 $[M+4H]^{4+}$, 725.8 $[M+3H]^{3+}$, 1088.4 $[M+2H]^{2+}$.

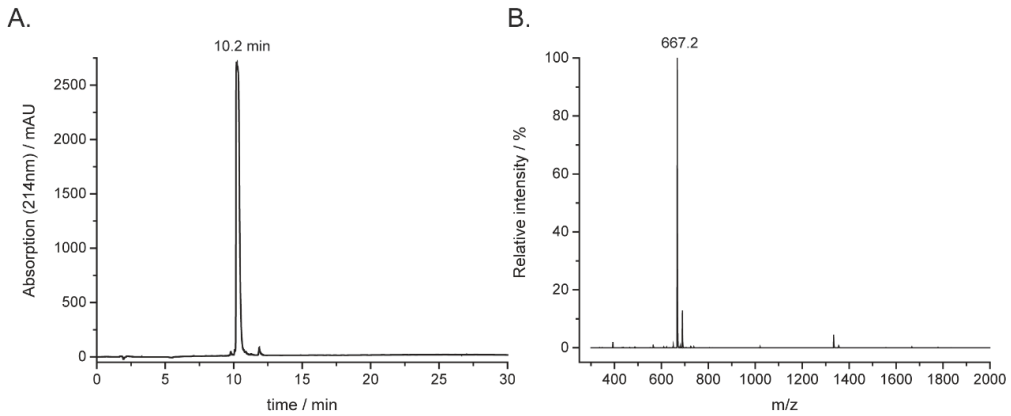

Figure 30. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **4'** (H-CPATGGY-NH2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 666.8 Da, found: 667.2 Da, 667.2 $[M+1H]^{1+}$.

Name	Sequence	MW calculated (Da)	MW observed (Da)	Yields
4	H-CPRKQLATKAARKS APATGGY-NH2	2174.6	2174.4	45 %
4'	H-CPATGGY-NH2	666.8	667.2	27 %

Table 6. Peptides **4** and **4'** used for Expressed Protein Ligation to produce peptides **5b**, **5d** and **5e**.

For peptides **5b**, **5d** and **5e**, the previously described one-pot ligation-desulfurization strategy was used. Briefly, between 1 and 2 µmol of H3 tail peptide hydrazides (**1b**, **1d** and **1e**) were converted to the TFET thioesters and, after ligation to N-terminal cysteine peptides **4** or **4'**, the cysteine at the ligation site was desulfurized as described above. When the reaction had reached completion, the peptides were purified by semi-preparative RP-HPLC. The purified peptides were characterized by analytical RP-HPLC and ESI-MS (Figure 31 - 33, Table 7).

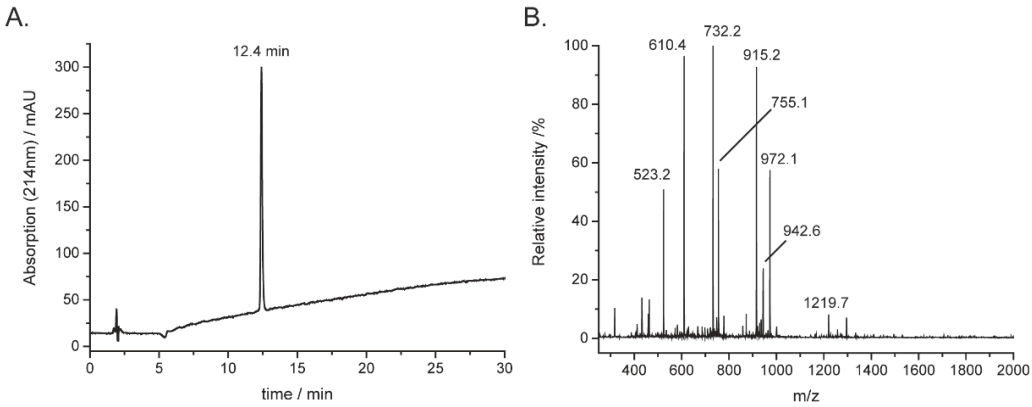

Figure 31. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **5b** (H-ARTK(me3)QTARKSTGGKAPRKQLATKAARKSAPATGGY-NH2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 3657.3 Da, found: 3656.6 Da, 523.2 $[M+7H]^{7+}$, 610.4 $[M+6H]^{6+}$, 732.2 $[M+5H]^{5+}$, 755.1 $[M+TFA+5H]^{5+}$, 915.2 $[M+4H]^{4+}$, 942.6 $[M+TFA+4H]^{4+}$, 972.1 $[M+2TFA+4H]^{4+}$, 1219.7 $[M+3H]^{3+}$

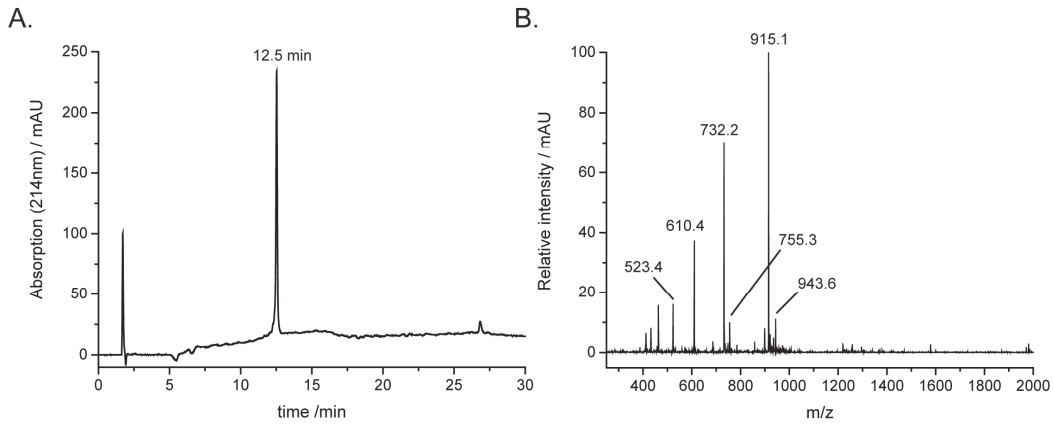

Figure 32. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **5d** (H-ARTKQTARKSTGGKAPRKQLATKAARK(me3)SAPATGGY-NH2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 3657.3 Da, found: 3656.4 Da, 523.4 $[M+7H]^{7+}$, 610.4 $[M+6H]^{6+}$, 732.2 $[M+5H]^{5+}$, 755.3 $[M+TFA+5H]^{5+}$, 915.1 $[M+4H]^{4+}$, 943.6 $[M+TFA+4H]^{4+}$.

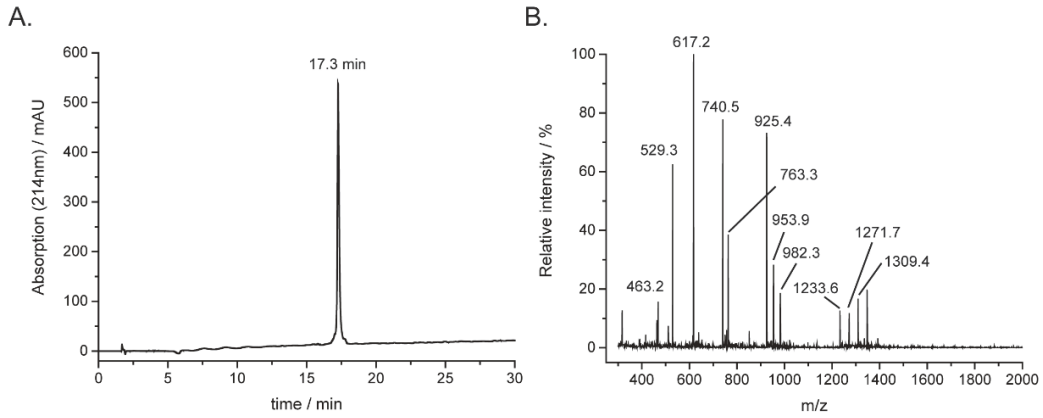

Figure 33. RP-HPLC chromatogram (A) and ESI-MS (B) of purified peptide **5e** (H-ARTK(me3)QTARKSTGGKAPRKQLATKAARK(me3)SAPATGGY-NH2). Gradient 0 % - 70 % solvent B in 30 min. Calculated mass: 3699.3 Da, found: 3697.6 Da, 463.2 $[M+8H]^{8+}$, 529.3 $[M+7H]^{7+}$, 617.2 $[M+6H]^{6+}$, 740.5 $[M+5H]^{5+}$, 763.3 $[M+TFA+5H]^{5+}$, 925.4 $[M+4H]^{4+}$, 953.9 $[M+TFA+4H]^{4+}$, 982.3 $[M+2TFA+4H]^{4+}$, 1233.6 $[M+3H]^{3+}$, 1271.7 $[M+TFA+3H]^{3+}$, 1309.4 $[M+2TFA+3H]^{3+}$.

Name	Sequence	MW calculated (Da)	MW observed (Da)	Yields
5b	H-ARTK(me3)QTARKSTGGKAPRKQLATKAARKSAPATGGY-NH2	3657.3	3656.6	54 %
5d	H-ARTKQTARKSTGGKAPRKQLATKAARK(me3)SAPATGGY-NH2	3657.3	3656.4	46 %
5e	H-ARTK(me3)QTARKSTGGKAPRKQLATKAARK(me3)SAPATGGY-NH2	3699.3	3697.6	57 %

Table 7. Peptides **5b**, **5d** and **5e**.

These libraries of modified nucleosomes and peptides enabled us to characterize cMAPs *in vitro*.

2.2.4 *In vitro* evaluation of probes binding and selectivity

To evaluate binding of cMAP1-6 to H3K4me3 and H3K27me3 as well as to assess their preference for bivalent versus monovalent and unmodified nucleosomes/histone peptides, several binding assays were performed. First, nucleosome pulldowns were performed by incubating modified or unmodified nucleosomes with immobilized cMAPs and by detecting the retained nucleosomes with anti-H3 WB (Figure 34. A). With this method, the affinity of cMAP2, 3, 5 and 6 were tested on nucleosomes **a**, **b** and **d** (Figure 34. B and C). All four probes retain efficiently methylated nucleosomes (**b** and **d**), whereas no binding was observed for unmodified nucleosomes (**a**), confirming that CD and PHD are folded and recognize their target PTM.

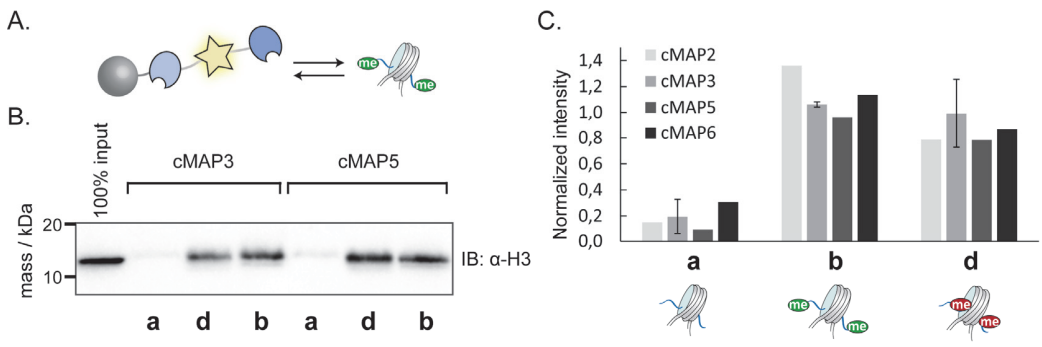

Figure 34. Pulldown of modified and unmodified nucleosomes using immobilized cMAPs. A. Schematic view of the experiment. B. Pulldown of nucleosomes **a**, **d** and **b** with cMAP3 and cMAP5. Nucleosome retention is visualized by anti-H3 WB. Here, 5 pmol of nucleosomes were incubated with 20 pmol immobilized cMAP3. C. Pulldown of nucleosomes **a**, **b** and **d** with cMAP2, 3, 5 and 6. Upper panel: Relative WB band intensities were quantified and normalized. For cMAP3, band intensities were averaged over 2 independent replicates. Error bars: standard deviation. Lower panel: schematic view of nucleosomes **a**, **b**, **d**.

To assess the preference for bivalent nucleosomes, cMAP binding was tested in an electromobility shift assays (EMSA) with nucleosomes **a**, **b'**, **d'**, **e'** and 601 DNA (Figure 35). We tested the feasibility of this assay with cMAP3. All nucleosomes (Figure 35. A) as well as DNA (Figure 35. D) were shifted by cMAP3 in the range of 100 - 200 nM, indicative of a relatively high unspecific binding to DNA. This assay does not allow to detect the binding preference for monovalent (**b'**, **d'**) versus unmodified (**a**) nucleosomes, probably due to DNA binding (compare **b'** and **d'** lanes with **a** lanes in Figure 35. A). However, a slight but noticeable difference of affinity is visible between bivalent (**e'**) and monovalent/unmodified nucleosomes (**a**, **b'**, **d'**) (compare band intensity of unbound nucleosomes, at ~ 450 bp, in **e'** lanes with **a**, **b'**, **d'** lanes in Figure 35. A; arrows indicate unbound nucleosomes at 220 nM cMAP3). Indeed, at 220 nM cMAP3, **e'** nucleosomes are almost completely shifted, whereas **b'**, **d'** and **a** nucleosomes are only partially shifted. By plotting the unbound nucleosome band intensities as a function of cMAP3 concentration and fitting, apparent K_{DS} showed a 1.7 and 1.6-fold preference for bivalent nucleosomes (**e'**) over H3K27me3 (**d'**) and H3K4me3 (**b'**) nucleosomes, respectively (Figure 35. B).

The affinity difference of 1.6-fold for **e'** over **b'** nucleosomes was maintained for 3 independent triplicates (Figure 35. C).

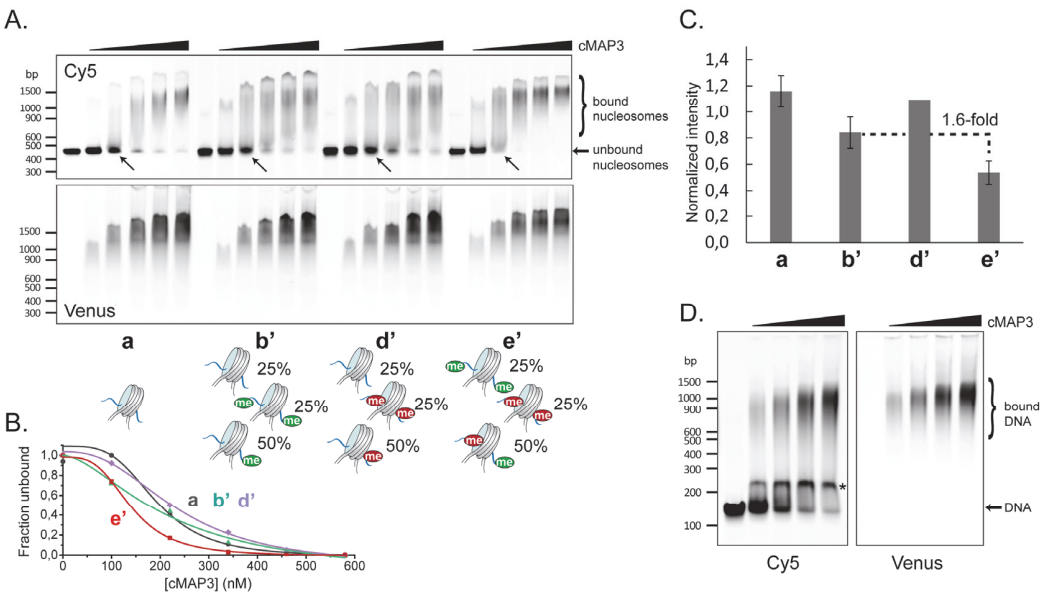

Figure 35. EMSA of cMAP3. A. EMSA of cMAP3 with Cy5-labeled nucleosomes. From left to right, 0, 100, 220, 340, 460, 580 nM of cMAP3 were mixed with nucleosomes **a**, **b'**, **d'** or **e'** and analysed by native PAGE. Upper panel: Cy5 fluorescence. Arrows indicate unbound nucleosomes with 220 nM protein. Middle panel: Venus fluorescence. Lower panel: schematic view of nucleosomes **a**, **b'**, **d'**, **e'**. B. Relative intensity of the unbound nucleosome band (at ~ 450 bp) as a function of cMAP3 concentration. Fit curves: binding isotherms with K_Ds of 192 nM (**a**), 218 nM (**b'**), 237 nM (**d'**), 140 nM (**e'**). C. Band intensities of unbound nucleosomes (at 220 or 150 nM protein concentration) was quantified, normalized and, for **a**, **b'** and **e'**, averaged over 3 independent triplicates. D. EMSA of cMAP3 with 601 Cy5-labeled 601 DNA. From left to right, 0, 100, 150, 200, 250 nM of cMAP3 was mixed with DNA and analysed by native PAGE. Left panel: Cy5 fluorescence. Right panel: Venus fluorescence. *: this band is due to a defect of the native gel.

Gel shift assays indicate that cMAP3 binds bivalent nucleosomes with higher affinity than monovalent nucleosomes. However, this assay does not allow to clearly characterize the multivalent effect because of unspecific binding to DNA. For this reason, we performed cMAP pulldowns by incubating cMAP proteins with immobilized unmodified and modified nucleosomes (Figure 36. A). The retained cMAPs was detected by anti-FLAG WB after extensive washing. This way, unspecific DNA binding does not interfere with the readout of the experiment. After checking the homogeneous loading of nucleosomes on beads (Figure 36. B), the

affinity of cMAP1-4 were tested with nucleosomes **a**, **b**, **d** and **e**. Additionally, the affinity of cMAP3 was tested for nucleosomes **c** (H3K9me3). For cMAP3, no binding was detected to H3K9me3 (**c**) nucleosomes, indicating that CD is specific for H3K27me3 (Figure 36. C). cMAP1 showed non-negligible binding to unmodified nucleosomes (**a**), which might be explained by the propensity of polyproline type II left-handed helices to form DNA-binding motifs (Figure 36. D).[407,408] cMAP1-4 bind nucleosomes containing either H3K4me3 or H3K27me3; however, only cMAP3 and 4 showed increased retention on the bivalent nucleosomes (**e**) (Figure 36. D). This result demonstrates that short and flexible linker sequences are required for efficient multivalent binding. Among the two best performing variants, we chose cMAP3 to continue our investigation because of its ability to be easily transformed into a BiFC split probe. cMAP3 exhibits an enhanced retention of 2.2- and 2.6-fold on H3K4me3K27me3 (**e**) over H3K4me3 (**b**) and H3K27me3 (**d**) nucleosomes, respectively.

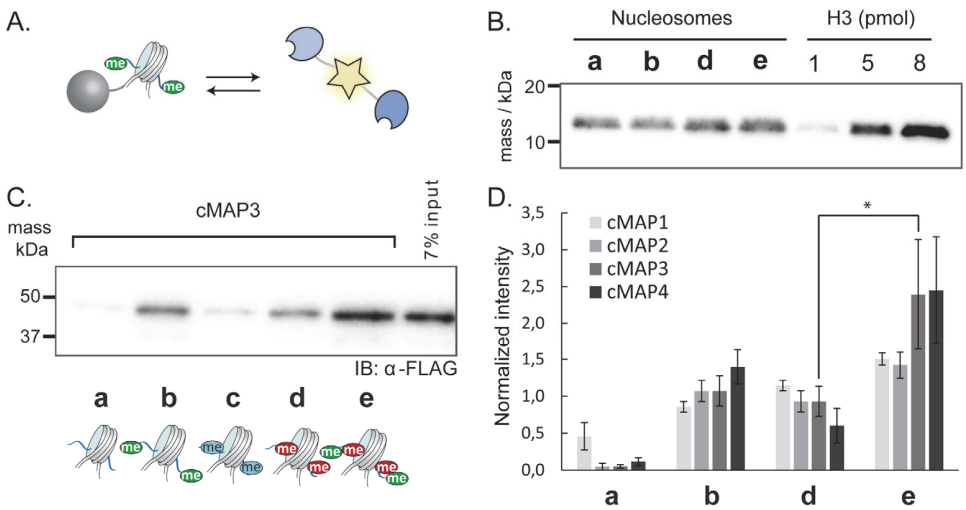

Figure 36. Pulldown of cMAPs using immobilized nucleosomes. A. Schematic view of the experiment. B. Homogeneous nucleosome loading on beads. Nucleosomes are detected by anti-H3 (C-term) Western Blot. C. Pulldown of cMAP3 with nucleosomes **a**, **b**, **c**, **d** and **e**. Upper panel: Protein retention is visualized by anti-FLAG WB. Lower panel: schematic view of nucleosomes **a**, **b**, **c**, **d** and **e**. D. Pulldown of cMAP1-4 with nucleosomes **a**, **b**, **d** and **e**. Relative WB band intensity were quantified, normalized and averaged over 3 or 4 independent replicates (3 replicates for cMAP1,2,4 and 4 replicates for cMAP3). Error bars: standard deviation. *: Student T test $p < 0.05$.

This pulldown experiment gave qualitative but no quantitative information about cMAP3 selectivity for bivalent marks. Therefore, we used microscale thermophoresis (MST) to measure binding affinities of cMAP3 for peptides **5b**, **5d** and **5e**.[409,410] MST is based on the physical principle of thermophoresis, which is the direct movement of particles in a temperature gradient. This phenomenon depends on changes in the hydration shell, charge or size of molecules. In a typical MST experiment, a titration serie is prepared by mixing increasing concentrations of an unlabelled molecule with constant amounts of a fluorescently labelled binding partner. Upon binding, one (or several) of the three above-mentioned parameters changes and this change can be detected by measuring fluorescence while applying a temperature gradient. MST is an immobilization-free technique, which requires picomolar amounts of the fluorescently labelled compound. This technique was chosen here for the low quantities of labelled compound required, for its easy sample preparation and its rapidity of execution. We unsuccessfully attempted to measure the affinity of cMAP3 with Cy5-labeled nucleosomes as well as 601 DNA using MST. Therefore, MST measurements were performed with cMAP3 and histone peptides **5b**, **5d** and **5e** (Figure 37). Titrating the peptides, Venus fluorescence was measured as a function of time upon application of a temperature gradient for 30 s (Figure 37. A). The difference in fluorescence between the few seconds before switching on the laser and the few seconds before the laser is turned off was plotted as a function of peptide concentration (Figure 37. B). Binding isotherms indicate a ~ 50-fold enhancement of affinity for the bivalent peptide **5e** compared to monovalent peptides **5b** and **5d** (Table 8).

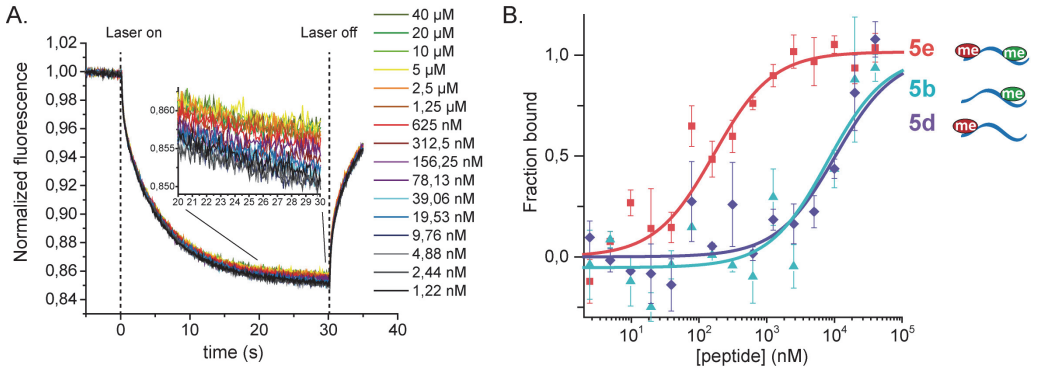

Figure 37. Binding affinities of cMAP3 for histone peptides **5b**, **5d** and **5e** determined by MST. A. Normalized fluorescence profile for peptide **5e**. Legend indicates peptide concentrations. Inset: zoom between 20 and 30 s. B. Binding isotherms for peptides **5b**, **5d** and **5e**. Error bars: standard deviation, n = 2.

Peptides	K_D (µM)
5b	7,9 ± 3,0
5d	8,5 ± 2,5
5e	0,165 ± 0,044

Table 8. Binding affinities of cMAP3 for peptides **5b**, **5d** and **5e** determined by MST.

Together, these results show that cMAP3 preferentially bind to bivalent versus monovalent targets (nucleosomes and peptides). For cMAP3, cMAP pulldown indicates a 2 to 3-fold selectivity for bivalent (**e**) versus monovalent (**b**, **d**) nucleosomes, whereas MST shows a ~ 50-fold enhanced binding for bivalent (**5e**) versus monovalent (**5b**, **5d**) peptides. This difference might be explained by a reduced accessibility of the H3 tail within nucleosomes.

With a selective sensor in hands, we then investigated the subnuclear localization of cMAP3 in living cells.

2.2.5 Visualization of bivalent domains in live cells

In vitro binding assays showed cMAP3 selectivity for bivalent marks. With this functional probe, we proceeded with the visualization of bivalent chromatin in living ESCs. First of all, we checked the global nuclear distribution of H3K4me3 and

H3K27me3 in mESCs by immunofluorescence imaging using modification-specific antibodies (Figure 38). A granular pattern that covers the whole nucleus was observed for H3K4me3. H3K27me3 signal was enriched at the periphery of the nucleus. These results are in line with previously reported distributions for these marks.[381]

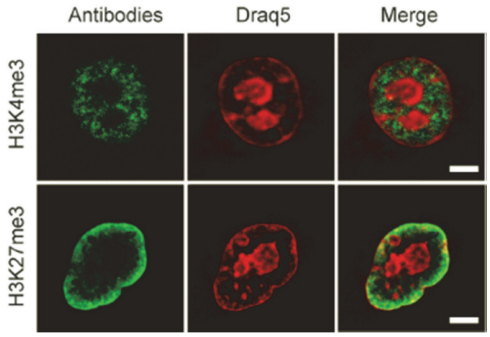

Figure 38. Imaging of H3K4me3 and H3K27me3 by immunofluorescence in fixed mESCs. Draq5 was used as a far-red DNA stain (scale bar: 5 µm).

We then performed ChIP in mESCs (Figure 39) to confirm the coexistence of H3K4me3 and H3K27me3 at the promoters of 5 genes previously described as bivalent (*Irx2*, *Olig1*, *HoxC5*, *HoxB13*, *Gata-4*).[241,246] As expected, these 5 genomic regions carried both marks. We also analysed 4 genes previously reported to have only H3K4me3 (*Pou5f1*, *Polm*, *Gapdh*, *Tcf4*) and one gene that carries only H3K27me3 (*HoxA3*).[241,246] As expected, *Pou5f1*, *Polm*, *Gapdh* and *Tcf4* have a high H3K4me3 signal and a low H3K27me3 signal, whereas *HoxA3* has a high H3K27me3 signal and a low H3K4me3 signal.

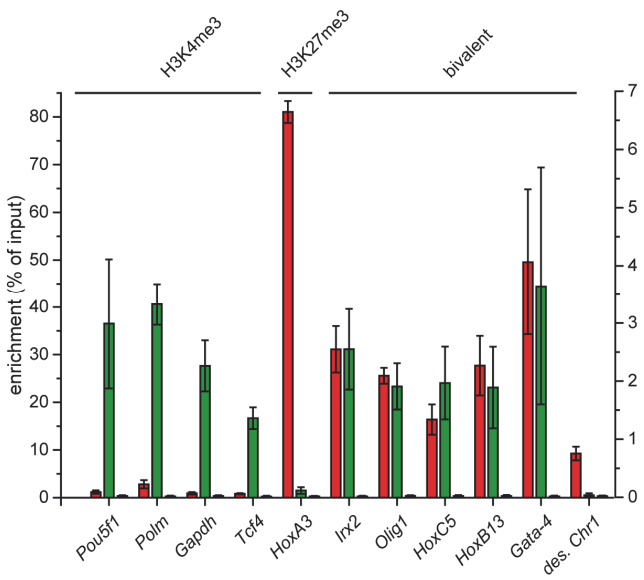

Figure 39. ChIP signals for H3K4me3 (green bars) and H3K27me3 (red bars) at 11 genomic regions in mESCs. The left axis corresponds to H3K4me3 signal and the right axis to H3K27me3 signal. The negative control (IgG) is represented on the right axis (blue bars).“des. Chr1” is a desert region on chromosome 1 and is used here as a negative control. ChIP signals were averaged over 3 independent replicates. Error bars: standard deviation.

When imaging transiently transfected cMAP3 in mouse ESCs using confocal fluorescence microscopy, fluorescent foci were observed throughout the nucleus (Figure 40. A, upper panel). Using an automated procedure (described in 2.2.7), we quantified the number of foci per nucleus for cells with similar cMAP3 expression levels, excluding nucleoli where FP-fusion proteins tend to accumulate (29.2 ± 1.8 foci / nucleus, error: standard error of the mean, Figure 40. B). The localization of cMAP3 foci did not follow the distribution of H3K27me3 heterochromatin but showed a distinct pattern.

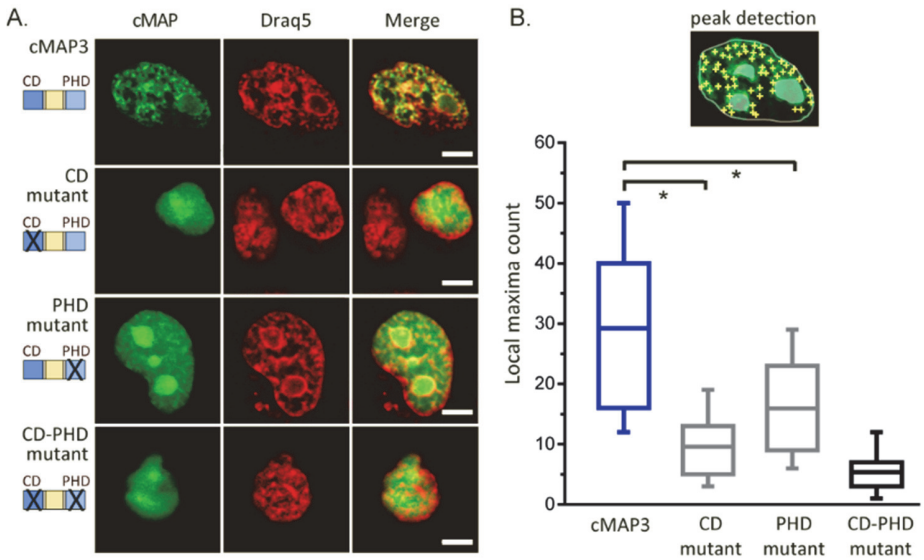

Figure 40. Visualizing bivalent chromatin domains in ESCs. A. Localization of cMAP3 and indicated mutant proteins in mouse ESCs (scale bar: 5 µm). B. Quantification of fluorescent foci in mouse ESCs for cMAP3 and reader domain mutants. The number of fluorescent spots detected was counted for each nucleus using an automated peak-finding protocol (number of cells analysed: cMAP3, n = 79; CD mutant, n = 81; PHD mutant, n = 41; CD-PHD mutant, n = 54) (Line: mean, Box: 25-75 % of the data, whiskers: 10-90 % of the data, Student T test * $P < 0.05$).

To corroborate reader domain dependent localization, we then introduced point mutations in the CD and PHD. When mutating two key Trp residues to Ala in the CD (W47A and W50A, Figure 17. B), binding to methylated H3K27 is abolished.[139] Introducing these mutations into cMAP3 indeed diminished subnuclear localization and produced homogenous staining of the whole nuclei, demonstrating that the PHD by itself is insufficient to result in sensor accumulation (9.6 ± 0.7 foci / nucleus, Figures 40. A and B). Similarly, mutations of an Asp and Trp residue in the PHD to Ala (D886A and W891A, Figure 17. B) result in complete loss of H3K4me3 recognition.[63] Similarly, these mutations reduced the formation of fluorescent foci significantly (15.9 ± 1.3 foci / nucleus, Figures 40. A and B), demonstrating that both reader domains are required for efficient chromatin localization. Finally, a cMAP3 construct with both CD and PHD mutated resulted in a largely homogenous nuclear distribution (5.4 ± 0.6 foci / nucleus, Figures 40. A and B).

Bivalent chromatin was found at about 35% of all genes in mESCs,[261] whereas it is present in much lower quantity in more differentiated cell lines.[241,381] We therefore transiently transfected HEK293 cells with cMAP3. As expected, the sensor adopted a homogeneous distribution in the differentiated cells (Figure 41).

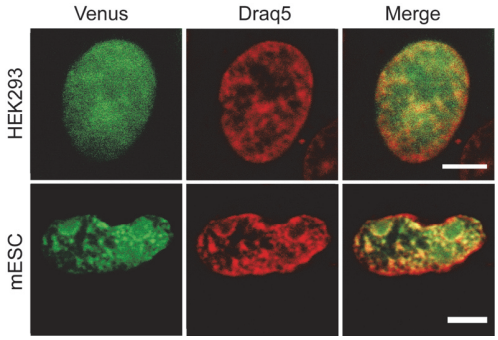

Figure 41. cMAP3 localization in HEK293 cells and mESCs (scale bar: 5 µm).

Together, these results show that cMAP3 has a bivalent domain-dependant subnuclear localization in mouse ESCs and that cMAP3 can be employed as a bivalent chromatin sensor in live cells.

2.2.6 Modulation of the epigenetic state with small molecules

Having validated cMAP3 specificity, we employed the probe to monitor how small molecule epigenetic modulators change the chromatin state in ESC. We selected a range of available compounds (Table 9) and treated mESCs using published protocols.

Compound	Enzyme(s) inhibited	Conditions	Observed phenotypic change
UNC1999	EZH1 & EZH2	7 days, 5 µM	None
SAHA	Class I & II histone deacetylases	72 hours, 5 µM	Apoptosis
		5 hours, 2.5 µM	None
Tranylcypromine (TCP)	LSD1	72 hours, 10 µM	None
(+)-JQ1	Bromodomains of BET family (Mainly BRD4)	72 hours, 500 nM	Slower growth and differentiation
Garcinol	Histone acetyltransferases p300 & PCAF	24 hours, 20 µM	Apoptosis

Table 9. mESC treatments with epigenetic small molecule modulators.

Focusing on epigenetic modulators that did not show phenotypic changes, we investigated the effects of suberoylanilide hydroxamic acid (SAHA), a histone deacetylase (HDAC) inhibitor;[411] tranylcypromine (TCP), an inhibitor of the H3K4 specific lysine-specific demethylase 1 (LSD1);[412] and UNC1999, a highly potent and selective inhibitor of the H3K27 specific lysine methyltransferases (KMT) enhancer of zeste homolog 1 and 2 (EZH1 and EZH2) (Figure 42. A).[413] Treatment of mESCs with SAHA (at a concentration of 2.5 µM for 5 h) or TCP (at a concentration of 10 µM for 3 days) did not result in a statistically significant change in cMAP3 binding patterns and number of fluorescent foci (38.2 ± 1.5 and 40.5 ± 2.0 foci / nucleus, respectively, Figure 42. B and C). This reveals that inhibition of histone deacetylases or LSD1 has a mild influence on bivalent chromatin. In contrast, treatment with 5 µM of UNC1999 for 7 days resulted in clear change of cMAP3 localization (Figure 42. B and C), without phenotypic changes in the ESC. Imaging cMAP3 in mouse ESCs treated with UNC1999 compared to the untreated control group revealed a strong reduction in the number of fluorescent foci in the treated group (11.1 ± 1.1 foci / nucleus, Figure 42. C), demonstrating that bivalent marks are largely abolished by UNC1999 action.

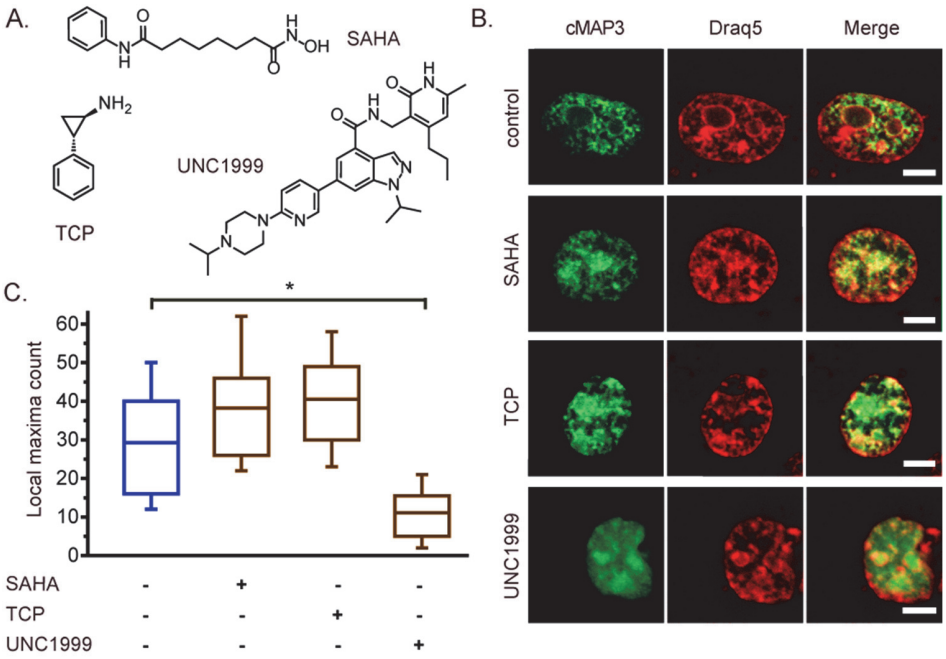

Figure 42. Small molecule disruption of bivalent domains. A. Chemical structures of employed epigenetic modulators. B. Localization of cMAP3 in mouse ESCs untreated or treated with SAHA, TCP or UNC1999 (scale bar: 5 µm). C. Quantification of fluorescent foci per nucleus for cMAP3 in mouse ESCs untreated or treated with SAHA, TCP or UNC1999 (number of cells analysed: untreated, n = 79; SAHA treated, n = 74; TCP treated, n = 51; UNC1999 treated, n = 44) (Line: mean, Box: 25-75 % of the data, whiskers: 10-90 % of the data, Student T test * $P < 0.05$).

To corroborate the loss of bivalency, we extracted histones from UNC1999 treated and untreated cells. We then assessed H3K27me3 and H3K9me3 levels by WB (Figure 43. A). Indeed, H3K27me3 levels were strongly reduced due to inhibition of EZH2, consistent with the situation in cells carrying a knockout mutation in the EZH2 methyltransferase.[414] Loss of H3K27me3 was further confirmed by immunofluorescence staining (Figure 43. B). However, H3K9me3 level was not significantly changed by UNC1999 action.

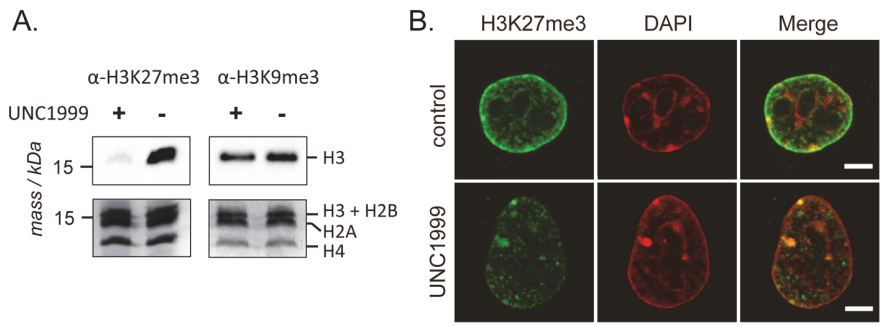

Figure 43. UNC1999 treatment decreased the global level of H3K27me3 without changing the level of H3K9me3. A. Anti-H3K27me3 and anti-H3K9me3 WB of acid-extracted histones from mESCs untreated and treated with UNC1999. Upper panels: WB; Lower panels: Coomassie Blue stained gel. B. H3K27me3 immunofluorescence staining of mouse ESCs untreated and treated with UNC1999 (scale bar: 5 µm).

Together, these results show that cMAP3 is sensitive to changes in histone PTM patterns and can thus be employed to monitor the effect of epigenetic modulators on chromatin states at the single cell level.

2.2.7 Validation of the procedure for local maxima count

The local maxima analysis of images of treated or untreated mESCs transfected with cMAP3 or reader domain mutants was performed with ImageJ software (Figure 44). The following macro was applied to all images:

```
run("Select All");
run("Smooth", "stack");
run("Find Maxima...", "noise=130 output=Count exclude");
run("Find Maxima...", "noise=130 output=[Point Selection]");
```

First, images were smoothed. Then the number of local maxima was counted with a noise tolerance of 130. Finally, the local maxima overlapping with nucleoli were manually subtracted.

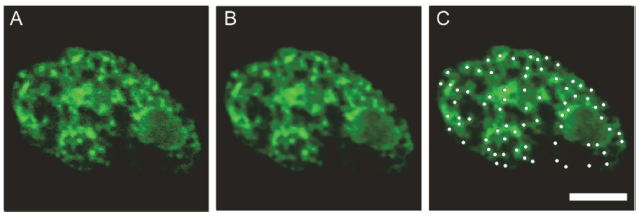

Figure 44. Image treatment for local maxima count. A) Original image of cMAP3 transiently transfected in mESCs. B) Image in panel A smoothed. C) Image in panel B with 74 local maxima analysed by ImageJ with a noise tolerance 130. Maxima are represented as white dots (scale bar: 5 µm).

To investigate the dependence of the number of foci on the noise tolerance, we additionally performed the local maxima count with a noise tolerance of 110 on the images of mESCs transfected with cMAP3 and reader domain mutants as well as UNC1999 treated cells transfected with cMAP3 (Figure 45).

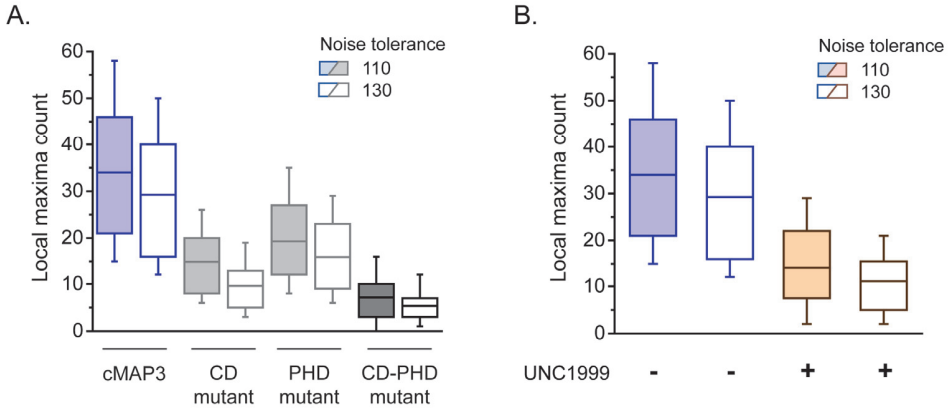

Figure 45. Dependence of the local maxima count on the noise tolerance. A. Local maxima count for cMAP3 and the three reader domain mutants in mESCs with a noise tolerance of 110 and 130. B. Local maxima count for cMAP3 in ESCs untreated or treated with UNC1999 with a noise tolerance of 110 and 130.

For a noise tolerance of 110, local maxima counts were only slightly higher compared to a noise tolerance of 130. This shows the mild influence of this criterion on the evaluation of probe subnuclear localization. Importantly, the local maxima count difference between WT cMAP3 and mutants and between treated and untreated cells remains statistically significant at a noise tolerance of 110. The method used here to quantify fluorescence granularity in nucleus is therefore valid to quantify the effect of reader domain mutations and of epigenetic drug treatments.

2.3 Discussion and conclusion

The subnuclear organization and dynamics of chromatin states is not well understood, as tools for their detection in live cells are lacking. Here, we have developed a general pipeline for the design and *in vitro* validation of genetically encoded multivalent probes for patterns of histone PTMs. We applied this strategy for the stem cell specific bivalent chromatin state and named these probes cMAPs. cMAPs were engineered from a FP and a set of reader domains found in naturally occurring chromatin effector proteins. Several versions of cMAP were designed varying the following parameters: linker length and stiffness, domain organization, origin and length of the PHD. The binding of these variants was tested in a pulldown assay using a library of reconstituted nucleosomes carrying defined PTMs. Among the 4 tested variants, two (cMAP3-4) showed an increased affinity for bivalent compared to monovalent nucleosomes. This demonstrates that two reader domains with modest affinities for their cognate histone PTM, when connected with appropriate linker sequences, result in an increased affinity upon multivalent binding. The strategy used here consisted in creating a library of probes with different characteristics and assessing their selectivity for the desired target with a suitable binding assay. This approach is highly modular. In this case, probes were engineered with reader domains specific for H3K4me3 and H3K27me3. However, the strategy is applicable to any PTM pattern of interest by choosing other PTM-specific reader domains. For instance, the coexistence of H3K27me3 and H3K36me3 on nucleosomes in an asymmetric fashion was observed and is poorly understood.[246] The development of sensors for this pattern could help elucidate its function(s) and dynamics.

After selection of cMAP3 as the most suited variant, we transiently transfected this probe in live mouse ESCs. Imaging by fluorescence confocal microscopy enabled the observation of discrete subnuclear foci, revealing the organization of bivalent chromatin into local clusters. The number of foci was significantly decreased for cMAP3 mutants in which key residues of the binding sites were changed to alanine. This result demonstrates that both reader domains play an

important role in the probe distribution. Moreover, cMAP3 enabled direct monitoring of the loss of bivalency upon treatment with a small molecule epigenetic modulator. Precisely, a reduction of fluorescent foci was observed upon inhibition of the catalytic subunits of PRC2 with the small molecule UNC1999, whereas no significant change in cMAP3 distribution of was measured upon inhibition of histone deacetylases or the histone demethylase LSD1.

The organization of bivalent chromatin into discrete and local clusters resembles Polycomb bodies or transcription factories and suggests that bivalent promoters are brought together into "bivalent bodies" where genes might be regulated together and not individually. This hypothesis would have to be confirmed by further experiments. For example, CRISPR-Cas9 could be potentially used to include or excise a bivalent promoter from a precise genomic location and eventual relocation of the gene in or out of a "bivalent body" could be imaged combining cMAP3 (to image bivalent chromatin) with fluorescent *in situ* hybridization (FISH) (to image a precise genomic location).

cMAP3 thus provides a novel platform to monitor bivalent chromatin dynamics in live cells. The developed sensor also represents a mean to target bivalent loci and modulate them by recruiting enzymatic activity to this specific chromatin state, enabling to rewrite the histone code.

Chapter 3. Installation and removal of a histone modification associated with active gene transcription

Chromatin effectors called writers and erasers dynamically regulate chromatin states and gene expression by catalysing the enzymatic addition and removal of histone PTMs, respectively. Many diseases are associated with misregulation of writers or erasers. Among the most studied and complex histone marks are methyl groups on lysines. Methylated lysine comes in three flavours: monomethyl-, dimethyl- and trimethyllysine. Depending on the methylation state, the histone and the residue, methyllysine can be associated with active gene transcription or gene repression. Methyltransferases and demethylases can be specific for one or two methylation states, whereas others can catalyse the installation and removal of all three states, respectively.[45] The mechanisms of recruitment and regulation of these enzymes are of high interest to the scientific community and will help to find answers to disorders associated with malfunctions of these proteins. This chapter focuses on the H3K4-specific demethylase LSD1 and the family of MLL H3K4-specific methyltransferases. H3K4 methylation is associated with active gene transcription and was proposed to counteract Polycomb repression.

3.1 The demethylase LSD1 binds to unmodified nucleosomes

3.1.1 Background

Methyl histone PTMs were thought to be irreversible before the discovery of the first histone demethylase LSD1 (Lysine-specific histone demethylase 1A, also called KDM1A, AOF2, KIAA0601 or BHC110) in 2004.[415] LSD1 is a nuclear homolog of amine oxidases that catalyses the removal of methyl groups from H3K4me1/2 (but has no activity on H3K4me3).[415] In this oxidation reaction, flavin adenine dinucleotide (FAD) is used as cofactor and molecular oxygen is used to regenerate FAD (Figure 46).[416] The reaction produces formaldehyde and H_2O_2 as byproducts.

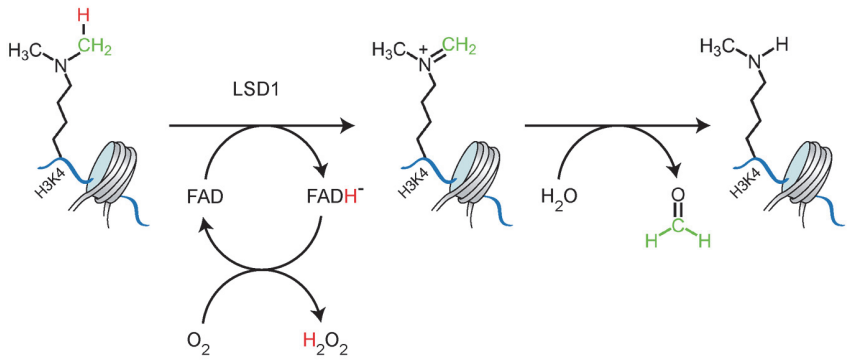

Figure 46. Mechanism of FAD-dependent H3K4me2 demethylation by LSD1.

LSD1 is a transcriptional repressor crucial for embryonic development and implicated in cancer.[417,418] This demethylase is involved in the maintenance of pluripotency in stem cells and in the regulation of bivalent genes.[275] Although its specificity for H3K4 is well established, it was suggested that LSD1 is also a H3K9-specific demethylase.[419] Therefore, LSD1 could be a transcriptional repressor or coactivator depending on the demethylated residue on H3. Further studies showed demethylase activity on other protein substrates, such as p53,[420] DNMT1[421] and MYPT1.[422] LSD1 has not only multiple substrates but also multiple binding partners. Indeed, CoREST and HDACs are generally associated with LSD1.[423–426] The protein BHC80 was also detected in the LSD1 complex.[427] Lan et al. discovered that the PHD finger of BHC80 binds specifically to H3K4me0 histone peptides with a K_D of 33 μM and this interaction is abrogated by methylation of H3K4.[66] The authors showed by RNAi and ChIP studies that LSD1 and BHC80 are dependent on one another to repress target genes by H3K4-demethylation. They proposed a recruitment role for BHC80: BHC80 binding to H3K4me0 nucleosomes would maintain LSD1 at its target loci to prevent re-methylation by H3K4-specific methyltransferases. They also suggested that BHC80 recruits LSD1 to H3K4me0 nucleosomes such that LSD1 demethylates neighbouring nucleosomes. This type of positive feedback mechanism is observed for PRC2 as well (mentioned in chapter 1). On the other hand, other studies showed by competitive assays and SPR that LSD1 itself binds to unmodified H3 peptide and to full length unmodified histone H3.[428,429] This is supported by the spacious catalytic site of the amine oxidase domain (AOD), which lacks specificity for the methylation states of H3K4.[430] Its multiple substrates and

associated proteins suggest that LSD1 may have broad chromatin-associated functions and probably complex regulatory mechanisms.

Several structures of LSD1 were reported after its discovery as demethylase.[430–432] LSD1 contains an N-terminally unstructured region, dispensable for its activity,[428] followed by a Swi3p, Rsc8 and Moira (SWIRM) domain, important for protein stability[430,431] and enzymatic activity[430] (Figure 47). Although its sequence is similar to DNA-binding modules, the SWIRM domain was shown to have no interaction with DNA *in vitro*.[431] At the C-terminus, the tower domain formed by an antiparallel coiled-coil of two alpha helices is inserted into the AOD (Figure 47). The tower domain is crucial for demethylase activity[430,431] and mediates association with CoREST.[431] The active site of the AOD is composed of a large catalytic cavity.[430,431] Adjacent to the active site and in close proximity to the SWIRM domain, a surface groove is proposed to accommodate the H3 tail.[430,432]

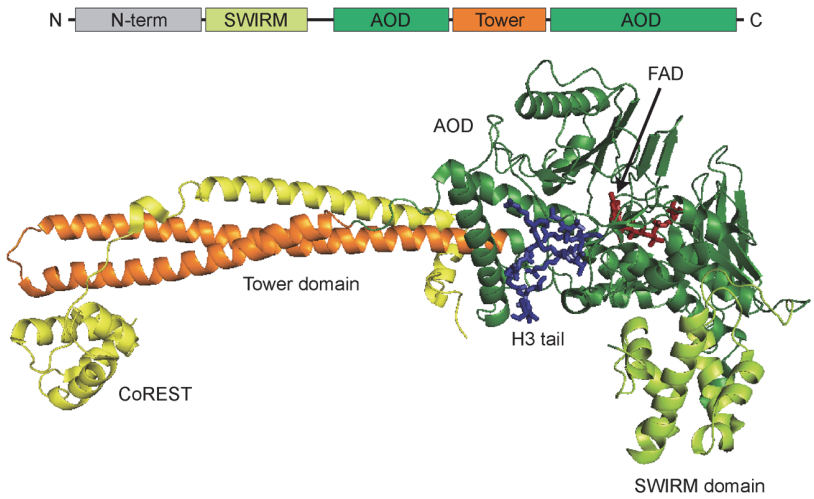

Figure 47. Structure of LSD1. Upper panel: Domain organization of human LSD1 (isoform 2). Lower panel: ribbon representation of LSD1 (171-836) in complex with CoREST (308-440) (displayed in yellow), the H3 tail (1-16)K4M (displayed in blue) and FAD (displayed in red) (PDB: 2V1D). The SWIRM domain, the tower domain and the AOD are represented in light green, orange and dark green, respectively.

3.1.2 Results and discussion

The data described here is in the frame of a collaboration with the laboratory of Prof. Champak Chatterjee (University of Washington, USA). This collaborative

work is the subject of a manuscript in preparation with the following contributors: Abhinav Dhall, Caroline E. Weller, Patrick Shelton, Aurea Chu, Aurore M.-F. Delachat, Beat Fierz and Champak Chatterjee. Abhinav Dhall performed LSD1 expression and the EMSA. Aurore M.-F. Delachat prepared the reconstituted nucleosomes and performed the MST assay.

Knowing that both BHC80 PHD and LSD1 bind unmodified H3 peptides, we aimed to decipher the importance of each binding event for LSD1 recruitment to nucleosomes. For this purpose, we developed a binding assay using MST to quantify LSD1 binding to unmodified nucleosomes (Figure 48. A). Recombinant histones and Cy5-labeled DNA containing the strong nucleosome positioning sequence 601 were used to reconstitute nucleosomes. His_6-tagged human full-length LSD1 protein was recombinantly expressed and purified by Ni-NTA affinity purification. Protein binding to unmodified mononucleosomes was first checked by EMSA (Figure 48. B). Then titration series were prepared with increasing concentrations of unlabelled LSD1 (between 8 nM and 131.7 µM) and a constant concentration of 10 nM of labelled nucleosomes. Cy5 fluorescence was measured as a function of time upon application of a temperature gradient (Figure 48. C). Binding isotherms were obtained from the fluorescence measurements and indicated a K_D of 5.98 ± 1.66 µM and a Hill coefficient n of 1.49 ± 0.54 for full length LSD1 binding to unmethylated nucleosomes (Figure 48. D).

Additionally, we showed that full length LSD1 binds to 601 DNA with a K_D of 127 nM (Dhall et al., Manuscript in preparation). As LSD1 binds to both 601 DNA and unmodified histone H3, one can speculate about the contribution of each binding events to the total binding affinity. We demonstrated that LSD1 binding to 601 DNA is mainly electrostatic, since this interaction is greatly diminished at high salt concentration (Dhall et al., Manuscript in preparation). Here the MST measurement was done with 150 mM NaCl. At this concentration, the contribution of DNA is probably limited. The reported affinities of LSD1 for unmodified H3 peptide are K_i = 1.8 µM[428] and K_i = 1.77 µM[429]. As for full length H3, Burg et al. reported a K_i of 18.9 nM and a K_D of 9.02 nM. These affinity constants are 3.3-fold and more than 600-fold lower, respectively, compared to the K_D of 5.98 µM for nucleosomes

measured by MST. However, any comparison between our measurement and previously reported affinities has to be cautiously interpreted for several reasons. First of all, our measurement was performed with 150 mM NaCl, whereas the cited binding assays were done without salt. LSD1 binding to H3 peptides was shown to be electrostatic.[428,429,432] Therefore, the strength of the interaction is subjected to high differences upon changes in the salt concentration. Secondly, LSD1 used in our study was full length, whereas Burg et al. and Forneris et al. used truncated LSD1 (LSD1 Δ1-150 for Burg et al. and LSD1 Δ1-184 for Forneris et al.). Despite being unstructured, the N-terminal region of LSD1 might modulate binding to nucleosomes.

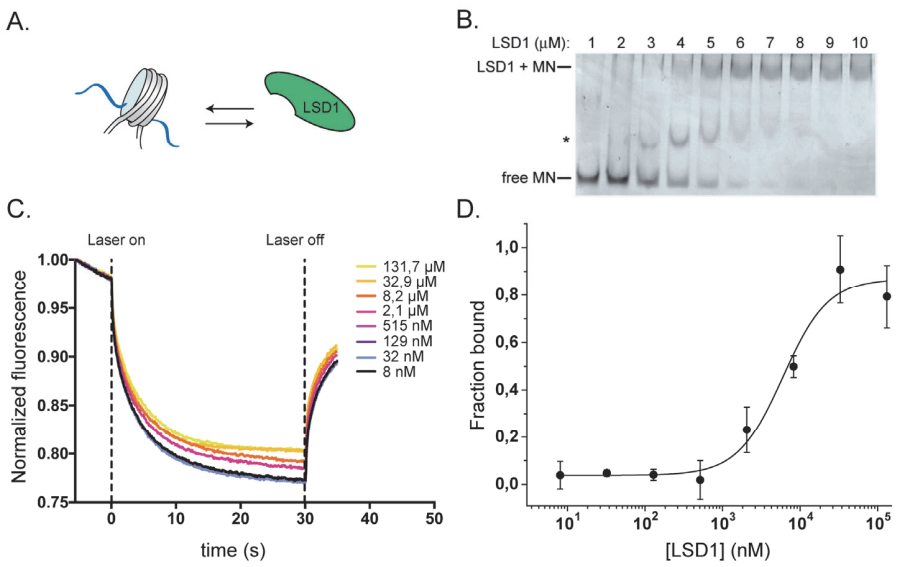

Figure 48. Binding affinity of LSD1 for unmodified nucleosomes. A. Scheme of the interaction quantified by MST. B. EMSA of unmodified mononucleosomes incubated with the indicated concentration of LSD1. Mononucleosomes were visualized by native PAGE. An asterisk indicates sub-saturated LSD1-bound species. Adapted from Dhall et al., Manuscript in preparation. C. Normalized fluorescence profiles. Legend indicates peptide concentrations. Adapted from Dhall et al., Manuscript in preparation. D. Binding isotherm for full length LSD1 and unmodified nucleosomes determined by MST. Error bars: standard deviation, n = 4.

Compared to the affinity of BHC80 for unmodified H3 peptide, the affinity of LSD1 for unmodified nucleosomes measured here is ~ 5-fold stronger. This difference calls into question the biological importance of the interaction of BHC80

PHD finger with unmodified H3 peptide. As an alternative model to the one proposed by Lan et al., we propose that LSD1 binds to its own product, unmodified H3K4 nucleosomes, and this interaction maintains LSD1 at its target genes (Figure 49). Additionally, BHC80 is proposed to have the role of fine tuning this positive feedback mechanism by enhancing binding to unmethylated H3K4, potentially on the same nucleosome (Figure 49. B, lower panel).

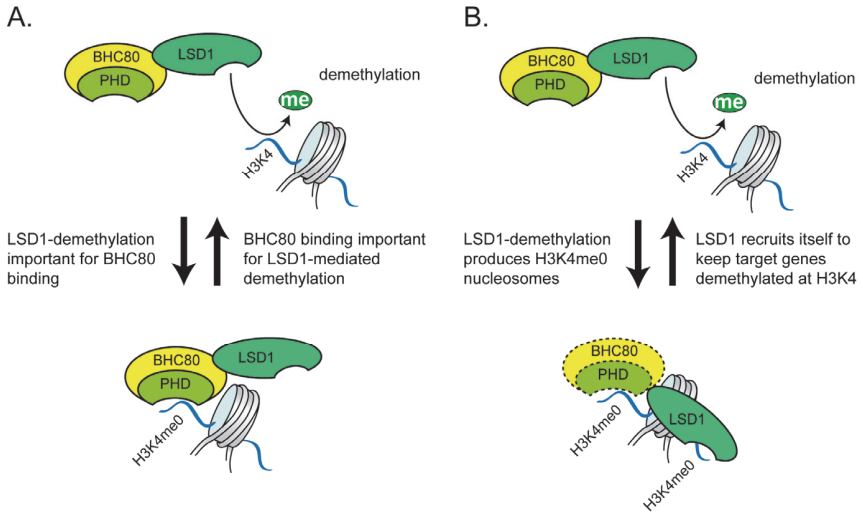

Figure 49. Models for LSD1 recruitment. A. Model proposed by Lan and coworkers.[66] B. New model in which LSD1 binds its product, H3K4me0 nucleosomes, through interaction with nucleosomal DNA and unmodified H3 tail. Dotted circles: fine tuning role for BHC80 and its PHD finger.

3.1.3 Conclusion

Previous studies showed the binding of LSD1 to H3K4me0/1/2/3 peptides. However, no report demonstrated the binding of LSD1 to unmodified nucleosomes. Here, we used MST to measure a K_D of 5.98 ± 1.66 μM for the interaction of full length LSD1 with unmodified nucleosomes containing recombinant histones. Knowing that the AOD domain (together with tower domain) is responsible for binding to 601 DNA (Dhall et al., Manuscript in preparation) and that the catalytic site is large and not specific for one H3K4 methylation state, one can speculate that the observed binding phenomena is the result of AOD domain binding to both nucleosomal DNA as well as unmodified H3 tail. To confirm this, binding of truncated LSD1 proteins (lacking individually the N-terminal domain, the SWIRM

domain or the AOD/Tower domains) to nucleosomes will be tested by EMSA. To know the contribution of DNA binding to the total binding, MST will be used to measure affinity of full-length LSD1 to 601 DNA. The model proposed in Figure 49. B will be further tested by comparing nucleosomal affinities of BHC80 and BHC80-LSD1 complex with the nucleosomal affinity of LSD1 by itself.

3.2 Influence of bivalent marks on the activity of the methyltransferase SET1B

3.2.1 Background

Chromatin effectors are regulated by various mechanisms, among which histone modification crosstalks represent a complex and unclear regulatory layer. MLL complexes are involved in crosstalks with several histone modifications including their own product, H3K4 methylation. The third PHD finger of the N-terminal part of MLL1 protein was demonstrated to bind to H3K4me3.[433] Furthermore, ASCOM, SET1A, WDR5 and RBBP5 were shown to bind to H3K4me3.[171,212] These observations suggest a positive feedback loop for MLL complexes. H2BUb is required for yeast SET1-mediated H3K4 methylation.[434] PRC2 is inhibited by active chromatin marks[206,207,435] and, in the same way, MLL complexes have reduced binding and activities on substrates containing the repressive mark H3K27me3.[212] This observation leads to the questioning of how H3K4me3 and H3K27me3 can be deposited on the same nucleosomes, forming bivalent chromatin.

In 2012, an MS-based immunoprecipitation method revealed that nucleosomes exist in the cell in both symmetric and asymmetric conformations, i.e. with different modifications on sister histones.[246] The authors revealed that bivalent marks are found in an asymmetric fashion on mononucleosomes. Since then, several questions arose: are asymmetrically modified nucleosomes actively or passively produced by chromatin modifier enzymes? Does asymmetry have a biological role? And if yes, which one(s)? In 2016, colleagues developed a modular and traceless

strategy to synthesize asymmetrically modified semi-synthetic nucleosomes using EPL.[207] In this study, symmetric and asymmetric nucleosomes were used to perform histone methyltransferase (HMTase) assay with PRC2. Assays with asymmetric H3K27me3 nucleosomes showed that PRC2 is activated by its own product, H3K27me3, when present on the opposite H3 tail (*trans* stimulation). Assays with symmetric and asymmetric H3K4me3 nucleosomes revealed that PRC2-mediated installation of H3K27me3 is inhibited by the active chromatin mark only when present on the same H3 molecule (*cis* inhibition). This suggests an active mechanism for the appearance of bivalent nucleosomes, i.e. PRC2 favours the installation of H3K27me3 on the tail opposite to H3K4me3. Assays with *trans* asymmetric bivalent nucleosomes revealed that the *trans* H3K27me3 activation can partially override the *cis* H3K4me3 inhibition. This finding together with the observation that bivalent marks are present mainly in an asymmetric fashion in the cell[246,263] suggests that PRC2 activity at bivalent genes has to be restrained.

3.2.2 Results and discussion

The data presented here is in the frame of a collaboration with the laboratory of Prof. Jaehoon Kim (KAIST, South Korea). Jaehoon Kim provided the MLL complexes. Aurore M.-F. Delachat prepared reconstituted nucleosomes and performed the HMTase assays. Carolin C. Lechner provided the asymmetrically modified H3K4me3 and symmetrically modified H3K27me3 octamers.

The use of chemically synthesized asymmetrically modified nucleosomes gave useful insights about the influence of pre-existing H3K4me3 and/or H3K27me3 on the activity of PRC2. However, little is known about the activity of MLL complexes on asymmetrically modified nucleosomes. Here we took advantage of the access to asymmetrically modified nucleosomes to perform HMTase assays with MLL complexes. MLL complexes catalyse the installation of methyl groups on lysine using S-adenosylmethionine (SAM) as cofactor (Figure 50. A). The six human MLL complexes were obtained by purification of FLAG-tagged SET1 or MLL proteins from stably transfected insect cells, which allows to have a defined subunit

composition for the complexes. MLL1-4 complexes contained the catalytic subunit (containing pre-SET, SET and post-SET domains), WDR5, RBBP5, ASH2L and DPY30. The SET1A/B complexes contained the catalytic subunit (containing n-SET, SET and post-SET domains), WDR5, RBBP5, ASH2L, DPY30 and CFP1. The activity of the purified complexes were tested on unmodified and H2BUb nucleosomes in HMTase assays (Figure 50. B). MLL3 and MLL4 showed no detectable activity on both substrates. Similar to the yeast SET1 complex,[434] human SET1A/B and MLL1-2 showed a H2BUb-dependent activity. In yeast, the H2BUb dependency was attributed to the n-Set domain, which was present in the human proteins used here. Among the four active complexes in our hands, SET1B showed the greatest activity and a mild dependency on H2BUb. These two characteristics are desired when studying the influence of H3K4me3 and H3K27me3 individually on the enzymatic activity. Therefore, we chose SET1B to perform HMTase assays with SAM labelled with tritium (^{3}H-SAM). The ^{3}H-methyl group deposited on H3 by SET1B were detected by fluorography after SDS PAGE and by scintillation counting.

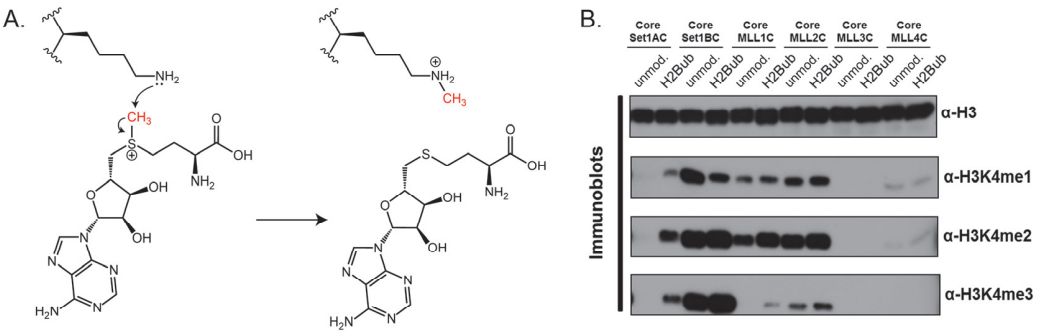

Figure 50. MLL complexes catalyse the installation of methyl groups on H3K4. A. Mechanism of SAM-dependent lysine methylation by MLL complexes. B. Activity of the six MLL complexes on unmodified or H2B ubiquitinated mononucleosomes. H3K4 methylation was detected by immunoblot with H3K4me1/2/3-specific antibodies. Data by Jaehoon Kim's lab.

First, we compared SET1B activity on symmetrically modified H3K27me3 nucleosomes and unmodified nucleosomes (Figure 51). Enzymatic reactions performed with either 0.2 or 2 pmol of SET1B resulted in similar signals for H3K27me3 compared to unmodified nucleosomes after 2 hours (Figure 51. A and

B). Kinetic analysis revealed similar methylation levels comparing unmodified and H3K27me3 nucleosomes at all time points (Figure 51. C). These results showed that H3K27me3 has no significant influence on SET1B complex activity in these experimental conditions. A similar experiment was performed by Kim et al. with human SET1A complex and revealed a clear inhibitory effect of H3K27me3 on the activity of the methyltransferase.[212] Despite the fact that the authors used SET1A, whereas we used SET1B, different methods were employed to express the enzymatic complexes. Kim et al. purified human SET1A complex from a FLAG-tagged subunit (CFP1, SET1A or WDR82) transfected in human cells. Therefore, this SET1A complex was composed of the full-length catalytic subunit and an undetermined number of endogenous subunits. We purified human SET1B complex from FLAG-tagged SET1B protein stably transfected in insect cells. Consequently, the SET1B complex used here contains a truncated catalytic subunit (residue 1672 to 1966 contains the n-SET, SET and post-SET domains) and a defined set of other subunits. The absence of H3K27me3 inhibition of SET1B complex observed here together with the clear inhibition of SET1A complex by H3K27me3 in Kim et al. revealed that the element responsible for this histone crosstalk resides in the difference in complex composition. Supporting this hypothesis, several studies showed that different subunit composition or different lengths of the catalytic subunit influence methyltransferase activity as well as sensitivity to other histone PTMs.[171,181,433,434]

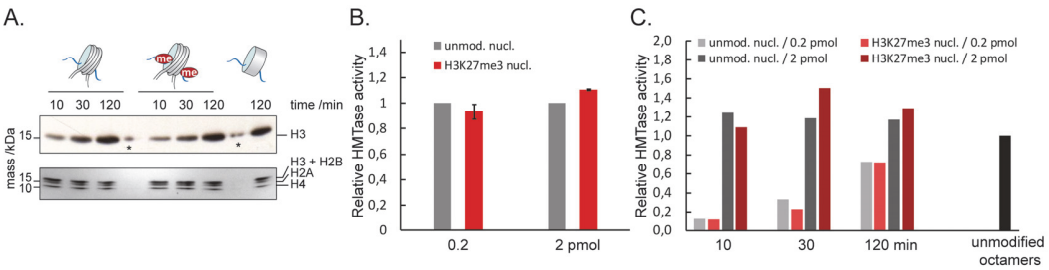

Figure 51. H3K27me3 does not influence SET1B activity on H3K4. HMTase assays with SET1B complex, unmodified or H3K27me3 nucleosomes as substrates and ^{3}H-SAM as cofactor. A. Histones were separated by SDS PAGE, stained with coomassie blue (lower panel) and probed for ^{3}H-methyl incorporation by fluorography (upper panel). 0.2 pmol of SET1B was used. Stars indicate leakage of the sample in the neighbouring lane. B. and C. Enzymatic reactions were subjected to scintillation counting. For panel B, 0.2 or 2 pmol of SET1B was used and the reaction was stopped after 2 hours. Signals are normalized to "unmodified nucl.". Error bars: standard deviation, n = 2. For panel C, 0.2 or 2 pmol of SET1B was used and the reactions was stopped after 10, 30 or 120 min. Signals are normalized to "unmodified octamers". n = 1.

We then decided to investigate how the presence of K4me3 on one H3 tail influences SET1B activity on the opposite H3 molecule. HMTase assays were performed with ^{3}H-SAM on asymmetric H3K4me3 (asH3K4me3) and unmodified nucleosomes. A 1.3-fold increased signal was measured for asH3K4me3 compared to unmodified nucleosomes after 2 hours reaction (Figure 52). Kinetic analysis indicates that this increase is also measured after 10 and 30 minutes. H3K4me3 therefore enhances SET1B activity in *trans*.

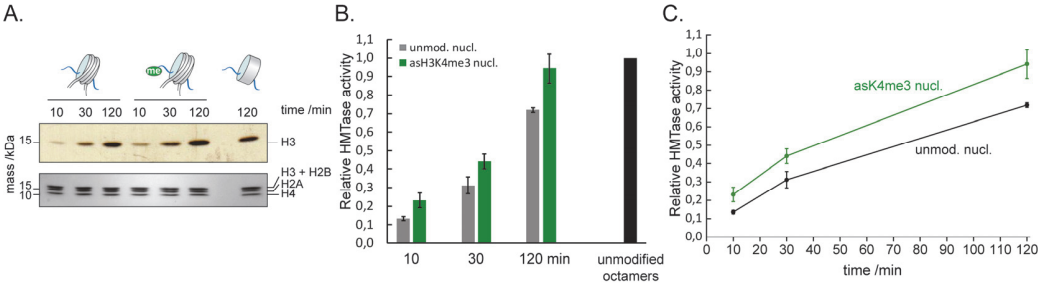

Figure 52. K4me3 on one H3 copy enhances SET1B enzymatic activity on the opposite H3 copy. HMTase assays with SET1B complex, unmodified or asymmetric H3K4me3 nucleosomes as substrates and ^{3}H-SAM as cofactor. A. Histones were separated by SDS PAGE, stained with coomassie blue (lower panel) and probed for ^{3}H-methyl incorporation by fluorography (upper panel). B and C. Enzymatic reactions were subjected to scintillation counting and signals normalized to "unmodified octamers". Data are represented as histogram (B) or as a function of time (C). Error bars: standard deviation, n = 2.

The interpretation of this result requires some considerations. First of all, asH3K4me3 nucleosomes contain one site available for SET1B-mediated methylation, whereas unmodified nucleosomes have two. Secondly, our experiments did not allow to differentiate between the three states of methylation. The signal therefore corresponds to the sum of all methyl groups installed on nucleosomes. We propose two hypothesis for the mechanism of this stimulation by H3K4me3 in *trans* (Figure 53). Hypothesis 1 involves a change of mono-/di-/tri-methyl ratios due to the presence of H3K4me3 on one tail. On unmodified nucleosomes, SET1B would deposit mainly monomethylation, whereas the equilibrium is shifted to the trimethyl state for asH3K4me3 nucleosomes. In this model, the ^{3}H signal with asH3K4me3 nucleosomes as substrates would be 1.5-fold higher than with unmodified substrates. This mechanism might involve allosteric activation of the enzyme by one or several of its subunit. Supporting this idea, WDR5 was shown to bind to H3K4me2/3 peptides and to H3K4me2 nucleosomes.[171] Moreover, this subunit was demonstrated to be responsible for the di- to tri-methyl conversion. Further supporting hypothesis 1, WDR5, RBBP5, ASCOM and human SET1A were shown to have an increased binding to H3 peptides when the latter is trimethylated at K4.[212] Hypothesis 2 involves a change in the kinetic of the enzyme. The presence of H3K4me3 would accelerate the installation of methyl groups on the opposite H3 molecule. These two hypotheses are not necessarily exclusive and other mechanisms might come into play. Furthermore, it is possible that the stimulation might be explained by a combination of hypotheses 1 & 2.

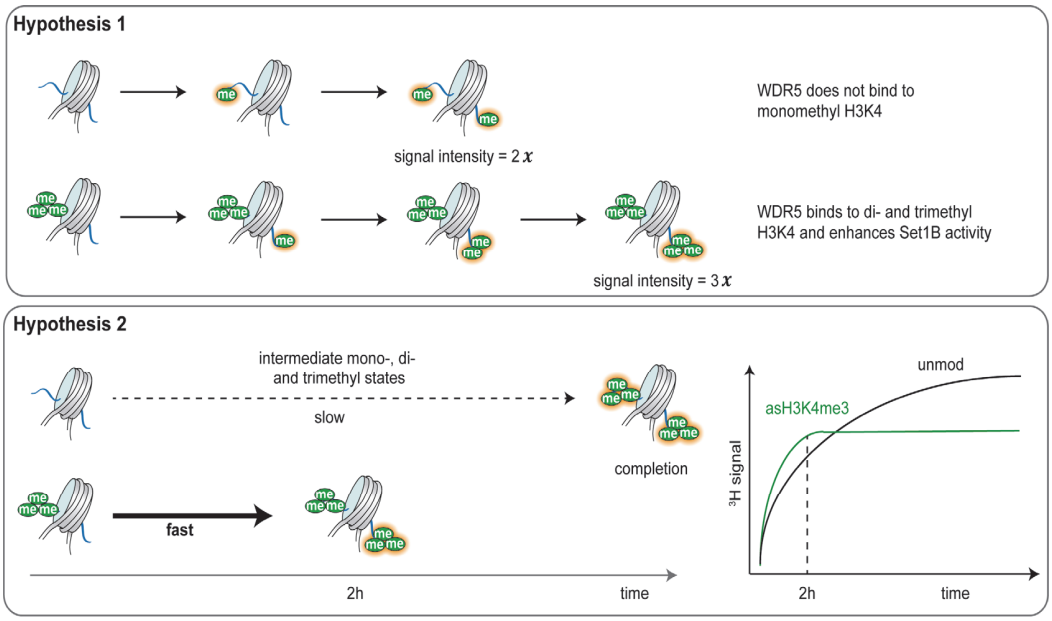

Figure 53. Hypothetical models for SET1B activation by H3K4me3 in *trans*. Pre-existing methyl groups at H3K4 are represented by green circles. ^{3}H-methyl groups at H3K4 are represented by glowing green circles. Here one methyl group is represent by one circle.

3.2.3 Conclusion

The access to chemically synthesized asymmetrically modified nucleosomes allowed us to study the influence of pre-existing marks on the activity of SET1B. A previous study indicated that H3K27me3 inhibits human SET1A complex.[212] Therefore, previous to testing if H3K27me3 could inhibit SET1B in *trans* (with asH3K27me3 nucleosomes), we performed HMTase assay with SET1B on symmetric H3K27me3 nucleosomes. Surprisingly, we found that human SET1B activity on H3K27me3 nucleosomes is similar to its activity on unmodified nucleosomes. This result showed that the element responsible for sensitivity to H3K27me3 is present in the complex used by Kim et al. but is absent in the complex used here. In order to have further insights about this, several experiments will be performed. HMTase assays with WB readout will enable us to identify changes of activity towards the three methylation states. To rule out if the N-terminal part of the catalytic subunit SET1B is responsible for this crosstalk, full-length SET1B complex will be used in HMTase assays with H3K27me3 and unmodified nucleosomes.

WDR5, RBBP5, ASCOM and SET1A have an enhanced binding to

H3K4me3 peptide versus unmodified peptide.[212] However, whether the active mark influences the activity of MLL complexes is unclear. Here, we showed that human SET1B activity is enhanced on asH3K4me3 nucleosomes compared to unmodified nucleosomes. To investigate the mechanism of this activation, several experiments will be performed. To test if binding of H3K4me3 induces a conformational change in the complex that enhances its activity, we will test HMTase activity on unmodified nucleosomes with and without adding H3K4me3 peptide in *trans*. A higher activity in presence of the peptide would suggest a conformational change within the enzyme upon binding to H3K4me3. Equal signals with and without the peptide added in *trans* would rather suggest that H3K4me3 on one tail serves as a recruitment platform and/or helps to position the enzyme on the nucleosomal substrate for increased activity on the unmodified H3 tail.

At completion, hypothesis 1 (Figure 53) involves a higher ^{3}H signal for asH3K4me3 substrates, whereas hypothesis 2 entails a 2-fold lower ^{3}H signal for asH3K4me3 substrates compared to unmodified nucleosomes. To test the validity of hypotheses 1 and 2, HMTase assays with WB readout as well as reactions for longer times will be performed.

Chapter 4. Materials and methods

4.1 Reagents and instrumentation

All solvents and reagents were purchased from commercial sources and used without further purification. Fmoc-Lys(Me3)-OH chloride was purchased from GL Biochem Ltd. (Shanghai, China). 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). O-(Benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HBTU) was purchased from Protein Technologies (Manchester, United Kingdom). N,N-Dimethylformamide (DMF), N,N-diisopropylethylamine (DIPEA) and piperidine were from Acros Organics (Geel, Belgium). Bovine serum albumin (BSA), D,L-Dithiothreitol (DTT), trifluoroacetic acid (TFA), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), sodium nitrite ($NaNO_2$), L-Glutathione reduced, Triton X-100, sodium fluoride (NaF), IGEPAL CA-630, β-mercaptoethanol, triisopropylsilane (TIS), 2,2,2- Trifluoroethanethiol (TFET), anti-Flag monoclonal M2 antibody mouse, anti-mouse IgG peroxidase conjugate, GMEM media, sodium pyruvate, PD184352, paraformaldehyde (powder), (±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carbo-xylic acid (Trolox) 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). Chemicals and solutions for preparation of SDS polyacrylamide gels (acrylamide, Precision Plus Protein™ All Blue Prestained Protein Standard), Immuno-Blot® polyvinylidene difluoride (PVDF) Membrane for Protein Blotting and Clarity™ Western ECL Substrate were purchased from Bio-Rad (Hercules, CA, USA). Slide-A-Lyzer™ dialysis cassettes, Slide-A-Lyzer™ Mini Dialysis buttons (3500 Da MWCO), Slide-A-Lyzer™ MINI dialysis devices (MWCO 10’000 Da) and protein G magnetic beads were from Thermo Scientific (Rockford, IL, USA). The centrifugal concentrators vivaspin 500 (10’000 MWCO, PES membrane) were purchased from Sartorius Stedim (Goettingen, Germany). The

anti-trimethyl-Histone H3 (Lys27) (α-H3K27me3, rabbit polyclonal) antibody used for immunofluorescence and Western Blot of acid-extracted histones as well as 2-mercaptoethanol and LIF for ESC culture were from Merck Millipore. UNC1999 was purchased from Sigma Aldrich. Ampicillin, phenylmethylsulfonyl fluoride (PMSF), lysozyme and isopropyl-β-D-thiogalactoside (IPTG) were purchased from BioChemica Applichem (Darmstadt, Germany). Kanamycin was purchased from Roth (Karlsruhe, Germany) or Acros (Geel, Belgium). All restriction enzymes were purchased from New England Biolabs (Bioconcept, Allschwil, Switzerland). 5% Criterion™ TBE Polyacrylamide gels were purchased from Bio-rad (Cressier, Switzerland). $ZnCl_2$ and H_2SO_4 were purchased from Fluka (Sigma-Aldrich, Taufkirchen, Germany). Glycerol was purchased from Fisher Chemicals (Zürich, Switzerland). The protease inhibitor cocktail cOmplete™ EDTA-free and X-tremeGENE 9 DNA transfection reagent were purchased from Roche (Switzerland). The Ni-NTA resin and QIAquick PCR Purification Kit were purchased from Qiagen (Hombrechtikon, Switzerland). Hydrazine monohydrate was purchased from Alfa Aesar (Heysham, UK). Dynabeads M-280 streptavidin and PureLink RNase A were purchased from Invitrogen (Thermo Scientific, Rockford, IL, USA). The IgG, anti-H3K4me3 and anti-H3K27me3 antibodies used for ChIP as well as the anti-Histone H3 antibody were from Abcam (Cambridge, UK). The anti-rabbit HRP conjugate was from Perkin Elmer (Zürich, Switzerland). MEM non-essential amino acids, L-glutamine, FBS embryonic stem cell-qualified, formaldehyde 16%, Methanol free and Alexa Fluor 647 antibody anti-rabbit antibody were purchased from LifeTechnologies (Thermo Scientific, Rockford, IL, USA). Penicillin/streptomycin was purchased from Bioconcept (Allschwil, Switzerland). GSK3 inhibitor XVI was from Calbiochem (San Diego, USA). Recombinant mouse E-cadherin Fc Chimera was from BioTechne AG (Zug, Switzerland). OptiMEM I 1x reduced serum medium was from Gibco (Thermo Scientific, Rockford, IL, USA). Draq5 was from Biostatus (Leicestershire, UK). Proteinase K was from Ambion (Thermo Scientific, Rockford, IL, USA).

Reaction vessels for manual peptide synthesis as well as the automated Tribute peptide synthesizer were from Protein Technologies Inc. Bacterial cells for

recombinant protein expression were grown in an HT infors AG incubator. E. coli cells were lysed by sonication using a Vibra-cell VCX 750 Sonics & Materials sonicator. Sedimentations were accomplished in an Avanti J20 XPI centrifuge and rotors (JA-12 and JA-8.1000) from Beckman Coulter. The Mini-Protean II system for SDS-PAGE and the Criterion blotter Transfer Cell for Western Blot were from Bio-Rad. Gels and blots were imaged using a ChemiDoc MP imaging system from Bio-Rad. Analytical RP-HPLC analysis was performed on an Agilent 1260 series instrument using an analytical Agilent Zorbax C18 column (150×4.6 mm, 5 µm particle size) at a flow rate of 1 mL/min. All RP-HPLC analyses were done with 0.1 % (v/v) TFA in H_2O (RP-HPLC solvent A) and 90 % acetonitrile and 0.1 % (v/v) TFA in H_2O (RP-HPLC solvent B) as mobile phases. Typically, a gradient from 0-70 % solvent A to solvent B over 30 min was used for analytical RP-HPLC analyses unless otherwise stated. Preparative RP-HPLC purifications were done on an Agilent 1260 preparative HPLC system with a preparative Agilent Zorbax C18 column (250×21.2 mm, 7 µm particle size) at a flow rate of 20 mL/min. Electrospray ionization mass spectrometric (ESI-MS) analysis was conducted on a LCQ Fleet Ion Trap mass spectrometer (Thermo fisher) interfaced with a Dionex Ultimate 3000 UPLC system or Shimadzu MS2020 single quadrupole instrument connected to a Nexera UHPLC system. Mass spectra were acquired by electrospray ionization in positive ion mode in the mass range of 300-2000 m/z. UV-Vis absorption measurements were carried out using an Agilent 8453 UV-Vis spectrophotometer. The Bioruptor sonicator used for ChIP was from Diagenode. The Fragment Analyzer used for ChIP was from Advanced Analytical. Microscale Thermophoresis measurements were performed on Monolith NT.115 (Nanotemper).

4.2 Production of nucleosomes using site-specifically modified, semi-synthetic histones

4.2.1 Solid phase peptide synthesis

2-Chlorotrityl resin was first reacted with hydrazine as previously described.[207] Briefly, 0.5g 2-chloro-trityl chloride resin was swollen in 3mL DMF for 15min and cooled to 4°C. A solution of 350µL DIPEA (2.01mmol, 3 eq.), 65µL hydrazine monohydrate (1.33mmol, 2 eq.) and 585µl DMF was added dropwise to the resin. The reaction was stirred for 1h at room temperature (RT).

For the peptides **1b** (H3(1-14)K4me3), **1d** (H3(1-28)K27me3) and **1e** (H3(1-28)K4me3K27me3), the sequences corresponding to residues 1 - 14 or 1 - 28 of human H3.1 with lysine 4 or/and 27 trimethylated were synthesized on this resin by SPPS using the Fmoc protection strategy. The peptide **1b** was synthesized manually and for the peptides **1d** and **1e**, serine 28 and trimethyllysine 27 were first manually coupled and then residues 1 to 26 were synthesized automated on the Tribute peptide synthesizer (PTI). HBTU was used for activation of all amino acid except trimethyl-lysine for which HATU was used. t-Butyl side chain protection was used for Serine and Threonine, t-Butyloxycarbonyl (Boc) side chain protection was used for Lysine, pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) side chain protection was used for arginine and trityl side chain protection was used for Glutamine. The pseudo proline Fmoc-Gln(Trt)-Thr(psiMe,-Mepro)-OH was used to install Q5 and T6 in peptides **1b**, **1d** and **1e**. For peptides **1d** and **1e**, Fmoc-Ala-Thr(psiMe, Mepro)-OH and Fmoc-Lys(Boc)-Ser(psiMe, Mepro)-OH were used to install A21-T22 and K9-S10, respectively. To improve synthesis yields, double couplings were performed where appropriate. Deprotection of the N-terminal Fmoc-group was achieved by treating the resin twice with 20 % (v/v) piperidine in DMF. The corresponding pre-activated amino acid (5 eq., activated with 4.8 eq. HBTU and 10 eq. DIPEA) was added to the resin and coupled for 30 min. Between coupling and deprotection steps, the resin was washed with DMF. After completion of peptide assembly, peptidyl-resins were washed with DMF, dichloromethane and methanol and dried under vacuum.

Following cleavage from the resin by a mixture of 95 % trifluoroacetic acid (TFA), 2.5 % water and 2.5 % triisopropylsilane (TIS), crude peptides were washed at least twice with cold diethyl ether and recovered by centrifugation. Crude peptides were dissolved in 50 % RP-HPLC solvent B and lyophilized. The peptides were purified by RP-HPLC on a preparative scale using linear gradients of 0 % - 15 % solvent B over 40 minutes for peptide **1b**, 5 % - 20 % solvent B over 45 minutes for peptide **1d** and 0 % - 40 % solvent B over 40 minutes for peptide **1e**. The peptides were characterized by analytical RP-HPLC and ESI-MS.

For peptides **4** (H3(15-35) A15C,V35Y) and **4'** (H3(29-35) A29C,V35Y), the sequences corresponding to residues 15 - 35 or 29 - 35 of human H3.1 with N-terminal alanine (15 or 29) mutated to cysteine and valine 35 mutated to tyrosine were synthesized on rink amide resin by SPPS, using the Fmoc protection and HBTU activation strategy. t-Butyl side chain protection was used for serine, threonine and tyrosine, Boc side chain protection was used for lysine, Pbf side chain protection was used for arginine and trityl side chain protection was used for glutamine and cysteine. The pseudo proline building block Fmoc-Ala-Thr(psiMe,-Mepro)-OH was used to install A21 and T22 in peptide **4**. To improve synthesis yields, double couplings were performed where appropriate. For Fmoc deprotection, amino acids pre-activation, coupling and peptide cleavage refer to the paragraph above about peptides **1b**, **1d** and **1e**.The crude peptides were purified by RP-HPLC on a preparative scale using linear gradients of 10%-35% solvent B over 40 minutes for peptide **4** and 0 % - 40 % solvent B over 40 minutes for peptide **4'**. The peptides were characterized by analytical RP-HPLC and ESI-MS.

4.2.2 Recombinant expression of truncated and wild type histones

Human wild type core histones H2A, H2B, H3_C110A and H4 were recombinantly expressed as described previously.[75] Briefly, histones were expressed in BL21 (DE3) pLysS cells from genes inserted into a pet15b plasmid. Cell cultures were grown at 37°C in LB media supplemented with 100 µg/mL ampicillin and 35 µg/mL chloramphenicol until an OD_{600} of 0.6. Protein expression was induced by

addition of IPTG to a final concentration of 0.5 mM. Expression of the protein was allowed to continue until 3 h post-induction. Cells were harvested, cell pellets were resuspended in lysis buffer (20 mM Tris, 1 mM EDTA, 200 mM NaCl, 1 mM -mercaptoethanol, 1x protease inhibitor cocktail, pH = 7.5). Cells were lysed by freeze-thawing and sonication. The inclusion body pellet was washed twice with 7.5 mL of histone lysis buffer with 1 % triton and once without triton. Histones were resolubilized in histone resolubilization buffer (6 M GdmCl, 20 mM Tris, 1 mM EDTA, 1 mM -mercaptoethanol, pH = 7.5), dialyzed into urea buffer (7 M urea, 10 mM Tris, 1 mM EDTA, 0.1 M NaCl, 5 mM -mercaptoethanol, pH = 7.5) followed by purification by cation exchange (using a HiTrap SP HP 5mL column).

N-terminal truncated variants of human H3.1_C110A variant **2** (H3(Δ1-14, A15C)) and **2'** (H3(Δ1-28, A29C)) were recombinantly expressed as N-terminal fusion to a His_6-SUMO tag in *E. coli* BL21 (DE3) pLysS cells. For truncated histone **2**, cells were grown in LB medium (supplemented with 50 µg/mL kanamycin and 35 µg/mL chloramphenicol) at 37 °C until reaching an OD_{600} of 0.6 - 0.8. Protein expression was induced by addition of IPTG to a final concentration of 0.5 mM. Expression of the protein was allowed to continue until 4 h post-induction. For truncated histone **2'**, cells were grown in autoinduction medium (1 % tryptone, 0.5 % yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 50 µM $FeCl_3$•$6H_2O$, 20µM $CaCl_2$, 10 µM $MnCl_2$•$4H_2O$, 10 µM $ZnSO_4$•$7H_2O$, 2 µM $CoCl_2$•$6H_2O$, 2 µM $CuCl_2$•$2H_2O$, 2 µM $NiCl_2$•$6H_2O$, 2 µM Na_2MoO_4•$2H_2O$, 2 µM Na_2SeO_3•$5H_2O$, 2 µM H_3BO_3) at 37 °C. Protein expression was allowed to proceed for 16 h. Subsequently, cells were harvested by centrifugation (4000 g, 4 °C, 20 min) and the cell pellet was resuspended in lysis buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, pH = 7.5). Cells were lysed by sonication and the lysate was centrifuged (15000 g, 4 °C, 30 min). The pelleted inclusion bodies were washed three times with lysis buffer containing 1 % Triton X-100 and twice with lysis buffer without detergent. Inclusion bodies were solubilized in Ni-NTA purification buffer A (6 M GdmHCl, 50 mM Tris-HCl, 100 mM NaCl, 5 mM imidazole, pH = 7.5), centrifuged (15000 g, 4 °C, 30 min) and applied to Ni-NTA

resin previously equilibrated with Ni-NTA purification buffer A. The protein was bound to the resin by gentle nutating overnight at 4 °C. The flow-through was collected and the resin was washed with 5 column volumes (CV) and 2 CV Ni-NTA purification buffer A followed by washing with 2 x 5 CV Ni-NTA purification buffer B (6 M Urea, 150 mM Tris-HCl, 150 mM NaCl, 50 mM imidazole, pH = 7.5). Finally, the protein was eluted with 4 x 1 CV Ni-NTA purification buffer C (6 M Urea, 150 mM Tris-HCl, 150 mM NaCl, 500 mM imidazole, pH = 7.5). Fractions were analysed by SDS PAGE and protein containing fractions were pooled.

The SUMO protease ULP1 required for cleavage of His_6-SUMO tag from the fusion protein was expressed from a pET30 plasmid in *E. coli* BL21 (DE3) cells. Cell cultures were grown in LB media supplemented with 100 µg/mL kanamycin. Protein expression was induced with 0.15 mM IPTG at OD_{600} of 0.6 and cell cultures were further incubated at 37 °C for 4 h. Harvesting and lysis of cells, isolation of inclusion bodies as well as purification by Ni-NTA affinity chromatography was done as described for His_6-SUMO-H3(Δ1-14, A15C) and His_6-SUMO-H3(Δ1-28, A29C) fusion proteins. Refolding and cleavage of His_6-SUMO from the His_6-SUMO-H3(Δ1-14, A15C) and His_6-SUMOH3(Δ1-28, A29C) fusion proteins was achieved by dialysis against 1 M Urea, 75 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, 25 mM L-arginine, 5 mM L-cysteine, pH = 7.5. The fusion protein and His_6-ULP1 were mixed in a mass ratio of 4:1 and dialyzed at 4 °C overnight. The cleavage reaction was checked for completeness by RP-HPLC and ESI-MS analysis. Truncated histone **2** and **2'** were purified from crude protein mixtures by preparative RP-HPLC with gradients of 30 % - 75 % solvent B in 30 min and 47 % - 62 % solvent B in 30 min, respectively. Collected fractions were analysed by analytical RP-HPLC and ESI-MS. Pure fractions were pooled, lyophilized and stored at - 20 °C. Typical yields were 5 mg of truncated histone **2** and 3mg of truncated histone **2'** per litre bacterial culture.

4.2.3 Ligation and desulfurization

Initially, histones **3d** (H3K27me3) and **3b** (H3K4me3) were synthesized by standard expressed protein ligation (EPL). In a typical ligation reaction, between 1

and 2 µmol of H3 tail peptide hydrazide was dissolved in ligation buffer (6 M GdmHCl, 0.2 M Na_2HPO_4, pH = 3.0) to a final concentration of 2.5 mM and the pH readjusted to 3.0. The peptide solution was kept at - 10 °C in an ice/NaCl bath and $NaNO_2$ was added dropwise from a stock of 200 mM to a final concentration of 18 mM. After mixing, the peptide was incubated at - 10 °C for 20 min, yielding to the peptide-azide. Typically, 0.6 µmol of truncated recombinant H3 protein was dissolved in mercaptoethanol acetic acid (MPAA) ligation buffer (6 M GdmHCl, 0.2 M Na_2HPO_4, 0.2 M MPAA, pH = 8.0) to a final concentration of 1 mM. The molar ratios peptide/protein was 4:1. The protein in MPAA ligation buffer was added to the peptide-azide and the pH was adjusted to 7.0. The reaction was monitored by analytical RP-HPLC. After 1 h, TCEP was added to a final concentration of 20 mM. When the reaction had reached completion, it was mixed with 1 volume of 20 % RP-HPLC solvent B and purified by semi-preparative RP-HPLC using a gradient of 20 % - 70 % HPLC solvent B in 40 min. The ligation product was characterized by analytical RP-HPLC and ESI-MS. The obtained protein was then desulfurized to convert the cysteine into alanine at the ligation site. The protein was dissolved in TCEP desulfurization buffer (6 M GdmHCl, 0.1 M Na_2HPO_4, 250 mM TCEP, pH = 6.5) to a final concentration of 0.5 mM. Glutathione and VA-044 radical starter were added from stock solutions to final concentrations of 40 mM and 20 mM, respectively. The pH was adjusted to 6.5 and the reaction was allowed to proceed at 37 °C. The reaction was monitored by analytical RP-HPLC and ESI-MS. When the reaction had reached completion, it was mixed with 1 Volume of 30 % RP-HPLC solvent B and purified by semi-preparative RP-HPLC using a gradient of 30 % - 70 %B in 40 min. The purified protein was characterized by analytical RP-HPLC and ESI-MS.

For histones **3e** (H3K4me3K27me3) and **3b**, a previously described one-pot ligation-desulfurization strategy was used.[207,367] Briefly, between 1 and 2 µmol of H3 tail peptide hydrazide was dissolved in ligation buffer to a final concentration of 10 mM and the pH readjusted to 3.0. The peptide solution was kept at - 20 °C in an ice / NaCl bath and $NaNO_2$ was added dropwise from a stock of 500 mM to a final concentration of 15 mM. After mixing, the peptide was incubated at - 20 °C for 5

min. Subsequently, TFET was added to a final concentration of 1 M and the pH of the peptide solution was adjusted to 6.8. The peptide was incubated for 10 min at RT and transferred to a tube with lyophilized, truncated recombinant H3 protein. The molar ratio peptide/protein was 7:1 or 5:1. TCEP buffer (0.5 M TCEP, 6 M GdmHCl, 0.2 M Na_2HPO_4, pH = 6.5) was added to a final concentration of 10 mM TCEP and the reaction was allowed to proceed at 25 °C under Argon atmosphere. The reaction was monitored by analytical RP-HPLC. When the reaction had reached completion, desulfurization was performed in the same 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 concentration of 0.25 M TCEP in the reaction solution. Then, glutathione and VA-044 radical starter were added from stock solutions to final concentrations of 40 mM and 20 mM, respectively. The reaction incubated at 42 °C and monitored by analytical RP-HPLC and ESI-MS. When the reaction had reached completion, it was mixed with 2 Volumes of RP-HPLC solvent A and purified by semi-preparative RP-HPLC using a gradient of 30 % - 70 % RP-HPLC solvent B in 45 min. The purified protein was characterized by analytical RP-HPLC and ESI-MS.

For peptides **5b** (H3(1-35)K4me3 V35Y), **5d** (H3(1-35)K27me3 V35Y) and **5e** (H3(1-35)K4me3K27me3 V35Y), the previously described one-pot ligation-desulfurization strategy was used.[207,436] Briefly, between 1 and 2 µmol of H3 tail peptide hydrazides (**1b**, **1d** and **1e**) were converted to the TFET thioesters and, after ligation to N-terminal cysteine peptides **4** or **4'**, the cysteine at the ligation site was desulfurized as described above. When the reaction had reached completion, the peptides were mixed with 2 Volumes of RP-HPLC solvent A and purified by semi-preparative RP-HPLC using a gradient of 0% - 40 % RP-HPLC solvent B in 40 min. The purified peptides were characterized by analytical RP-HPLC and ESI-MS.

4.2.4 Octamer refolding and nucleosome reconstitution

Lyophilized core histones were dissolved in unfolding buffer (6 M GdmHCl, 20 mM Tris-HCl pH = 7.5, 5 mM D,L-dithiothreitol) and mixed in stoichiometric

quantities (1 equivalent of H3 and H4 and 1.1 equivalent of H2A and H2B). The scale was 100 - 500 µg of histone H3 with an overall protein concentration of 74 µM. Histones were dialysed against 3 buffer changes of 3 h, 15 h and 3 h in refolding buffer (2 M NaCl, 10 mM Tris-HCl pH = 7.5, 1 mM EDTA, 1 mM D,L-dithiothreitol) in Slide-A-Lyzer™ dialysis cassettes. The obtained crude octamers were applied to a Superdex 200 10/300 column (GE Healthcare) and resolved in refolding buffer. Fractions containing pure octamers were pooled and concentrated in Vivaspin 500 centrifugal concentrator (10'000 MWCO) to the range of 30 - 60 µM. Octamers were stored at - 20 °C in 50 % v/v glycerol. Octamers were analysed by SDS PAGE.

For cMAP pulldown assays, a biotinylated 166 bp DNA fragment with a 19 bp linker prior to the strong nucleosome positioning sequence 601[437] was produced by PCR as described previously.[73] For nucleosome pulldown, unlabelled 601 DNA was obtained from a plasmid containing 32 repeats of the 601 sequence separated by restriction sites. Milligram amounts of this plasmid was prepared by *E. Coli* culture, processed with alkaline lysis and purified by size exclusion chromatography with a sepharose 6 Fast Flow column (GE Healthcare). Following digestion with EcoRV (New England Biolab), the 153bp 601 DNA was purified by PEG precipitation and concentrated to 1.5 - 2.5 µg/µL by ethanol precipitation. For EMSA with cMAP3 and for MST with LSD1 and unmodified nucleosomes, Cy-5 labelled 601 DNA was obtained by PRC.

Mononucleosomes were reconstituted either at a small scale by dilution or at a larger scale by dialysis as described previously.[406] Briefly, in the first method, typically 10 or 20 pmol of 601 DNA were mixed with the respective histone octamers in 2 M NaCl in a volume of 10 µL. Following several additions of 10 mM Tris-HCl pH = 7.6, the final concentration of NaCl is 0.2 M. In the second method, nucleosomes were reconstituted in 70 µl sample volume by mixing 60 pmol of DNA with the respective histone octamer at high salt conditions (2 M KCl, 10 mM Tris-HCl pH = 7.5, 1 mM EDTA) followed by gradient dialysis into low salt conditions (0.25 M KCl, 10 mM Tris-HCl pH = 7.5, 1 mM EDTA). Dialysis was carried out in Slide-A-Lyzer™ MINI dialysis devices using a two-channel peristaltic pump (1 mL/min) over 16 h at 4°C. Nucleosome concentrations were determined by UV

quantification and the quality of nucleosome reconstitution was assessed by native gel electrophoresis.

4.3 Cloning, expression and purification of multivalent sensors for bivalent chromatin

QuikChange II XL site-directed mutagenesis kit (Agilent), Gibson cloning[438] or classical (restriction enzymes) cloning was used to clone the 6 cMAP variants in pET28b(+) or pET15b bacterial vectors (Table 10). DH5α *E. coli* cells were used for all cloning steps. The Chromo-domain (CD) of the Drosophila Polycomb group protein Pc corresponds to entry P26017 on uniprot.org, residues 23 - 84. The Plant Homeo Domain (PHD) of the mouse inhibitor of growth protein 2 (ING2) corresponds to entry Q9ESK4 on uniprot.org, residues 207 - 281. The PHD of the human transcription initiation factor TFIID subunit 3 (TAF3) protein corresponds to entry Q5VWG9 on uniprot.org, residues 855 - 924. All primers and synthetic genes were obtained from Integrated DNA Technologies (IDT, Leuven, Belgium).

Variant	Cloning method	Primers/gene fragment	Template
cMAP1	Digestion by restriction enzymes and ligation	gBlock gene fragment encoding mouse ING2 PHD flanked by desired restriction sites (IDT)	cMAP6
cMAP2	--	gBlock gene fragment (IDT)	
cMAP3	Quikchange mutagenesis	5'-gcgggggtggtagcgtctcgaaaggaga-3' 5'-tctcctttcgagacgctaccacccccgc-3' 5'-catggacgaactgtataagggcggttacccggg-3' 5'-cccgggtaaccgcccttatacagttcgtccatg-3’	cMAP2
cMAP4	Quikchange mutagenesis & digestion by restriction enzymes and ligation	5’-accggcagcggacctcccccgc-3’ 5’-gcgggggaggtccgctgccggt-3’ 5’-taagcaaagcttggcagcggaga-3’ 5’-tgcttaggatccgctaccacccc-3’	cMAP6 & cMAP3
cMAP5	Quikchange mutagenesis	5'-tctgtccgaaatgtgcgaataaaggtggtag ccgtgct-3' 5’-agcacggctaccacctttattcgcacatttcg gacaga-3’	cMAP2
cMAP6	Quikchange mutagenesis	5'-gcggcacaccaagcaaacctcccccgccac ccgtctcgaaaggagagg-3' 5'-cctctcctttcgagacgggtggcgggggag gtttgcttggtgtgccgc-3' 5'-gcatggacgaactgtataagcccccgcctc caccgtacgtaatcagagacgagtg-3' 5'-cactcgtctctgattacgtacggtggaggcg ggggcttatacagttcgtccatgc-3'	cMAP2

Table 10. Primers used to clone cMAP constructs in bacterial vectors.

cMAPs were expressed in LB medium in *E. Coli* BL21 (DE3) by the addition of 0.5 mM IPTG and 10 µM $ZnCl_2$. In general, cells were induced at 37 °C for 4 h. For cMAP4, cells were induced overnight at 16 °C. Cell pellets were resuspended in

a lysis buffer consisting of a solution of 50 mM NaH_2PO_4 pH = 7.5, 300 mM NaCl, 10 mM imidazole, 1 % v/v IGEPAL CA-630, 10 % w/v glycerol, 10 µM $ZnCl_2$, 0.5 mM PMSF, 0.5 mM β-mercaptoethanol, 1x protease inhibitor cocktail. Cells were then lysed by addition of lysozyme to a final concentration of 1 mg/mL followed by sonication. The lysates were clarified by centrifugation and the supernatants were applied to Ni-NTA affinity purification. The resin-bound proteins were washed three times with a wash buffer consisting of 50 mM NaH_2PO_4 pH = 7.5, 1 M NaCl, 20 mM imidazole, 10 µM $ZnCl_2$, 0.5 mM β-mercaptoethanol. The proteins were eluted with a buffer consisting of 50 mM NaH_2PO_4 pH = 7.5, 150 mM NaCl, 500 mM imidazole, 10 µM ZnCl2, 1 mM β-mercaptoethanol.

For nucleosome pulldowns, cMAP2, 3, 5, 6 were dialyzed in 50 mM Tris-HCl pH = 7.4, 150 mM NaCl, 1 mM DTT, 10 µM $ZnCl_2$. For EMSA, cMAP2 was dialyzed in 50 mM Tris-HCl pH = 7.4, 150 mM NaCl, 1 mM DTT, 10 µM $ZnCl_2$, 5 % w/v glycerol. For cMAP pulldown assays, proteins were dialysed overnight into a buffer consisting of 50 mM Tris-HCl pH = 7.4, 150 mM NaCl, 1 mM β-mercaptoethanol, 10 µM $ZnCl_2$, 5 % w/v glycerol.

cMAPs were stored in 30 % glycerol at - 80°C or used freshly. cMAP2 was confirmed to be monomeric by gel filtration with an isocratic elution with a buffer consisting of 50 mM NaH_2PO_4, 50 mM NaCl, 1 mM DTT, pH = 7.8 (Superdex 200 10/300GL column GE Healthcare) after anion exchange chromatography (HiTrap Q FF column GE Healthcare).

4.4 *In vitro* binding assays

Nucleosome pulldown with immobilized cMAPs

20 µL of anti-FLAG M2 affinity gel (Sigma) (corresponding to 10 µL resin) are washed 2 x with 50 mM Tris-HCl pH = 7.4, 150 mM NaCl., 2 x with 0.1 M glycine pH = 3.5 and 4 x with PD buffer 1 (10 mM Tris-HCl pH = 7.4, 50 mM NaCl, 1 mM $MgCl_2$, 1 % v/v Triton X-100, 0.5 mM PMSF). Then, 1 or 20 µg of one of the cMAP variant was immobilized on the beads in 200 µL PD buffer 1 for 2 h at 4 °C. The resin was then washed 5 x with 200 µL PD buffer 1 to remove unbound cMAPs

and subjected to incubation with 3 or 5 pmol of mononucleosomes in 200 µL PD buffer 1 for 1 h or overnight at 4 °C. Subsequently, the resin was washed 3 or 4 x with 200 µL wash buffer (10 mM Tris-HCl pH = 7.4, 300 mM NaCl, 1 % Triton X-100, 0.5 mM PMSF) and boiled in 2 x Laemmli sample buffer. Beads were removed by filtration with Ultrafree-MC centrifugal units (Millipore) at 12'000 xg for 4 min. Samples were run on a 12 % SDS-PAGE followed by transfer on a PVDF membrane in transfer buffer (10 % methanol, 25 mM Tris, 200 mM Glycine) at 200 mA for 2 h at 4 °C. The blot was incubated in blocking solution (10 % non-fat milk in tris buffer saline (TBS) with Tween-20) for 2 h at RT and then incubated in blocking solution supplemented with anti-H3 rabbit polyclonal antibody (anti-H3 C-terminus, 1:1000) for 1 h at RT. After 3 washes with TBS-Tween, the blot was incubated in blocking solution supplemented with anti-rabbit HRP conjugate (1:2000) for 1 h at RT. After 3 washes with TBS-Tween, the blot was imaged after applying Clarity Western ECL substrates with Chemidoc MP imaging system Bio-Rad Gel imager.

EMSA with cMAP3 and nucleosomes

0.3 pmol of Cy5-labeled nucleosomes were mixed with various concentrations of cMAP3 in gel shift buffer (20 mM HEPES, pH = 7.5, 100 mM KCl, 0.05 % NP40, 0.1 mg/mL BSA) in a total volume of 10 µL and incubated for 30 min at 30 °C. For the gel shift with Cy5-labeled 601 DNA, conditions are identical except the concentration of BSA which was 0.5 mg/mL. 2 µL of sucrose 25 % was added in all samples and they were analysed by native gel electrophoresis.

cMAP pulldown with immobilized nucleosomes

200 µg Dynabeads M-280 streptavidin are washed with 1 mL PD buffer 2 (10 mM Tris-HCl pH = 7.5, 150 mM NaCl, 0.1 % v/v Triton X-100, 5 % w/v glycerol, 0.1 mg/mL BSA, 1 mM D,L-dithiothreitol). Then, 4 pmol biotinylated mononucleosomes were immobilized on the beads in 200 µL PD buffer 2 for 1 h at 4 °C. The resin was then washed twice with 500 µL PD buffer 2 to remove unbound nucleosomes and subjected to incubation with 30 pmol of one of the cMAP variants in 200 µL PD buffer 2 for 4 h at 4 °C. Subsequently, the resin was washed 6 x with

500 µL PD buffer 2 for 10 min at 4 °C and boiled in 2 x Laemmli sample buffer. Samples were run on a 12 % SDS-PAGE followed by transfer on a PVDF membrane in transfer buffer (10 % methanol, 25 mM Tris, 200 mM Glycine) at 200 mA for 2 h at 4 °C. The blot was incubated in blocking solution (10 % non-fat milk in TBS-Tween) for 2 h at RT and then incubated in blocking solution supplemented with anti-FLAG monoclonal M2 antibody mouse (1:1'000) for 1 h at RT. After 3 washes with TBS-Tween, the blot was incubated in blocking solution supplemented with anti-mouse peroxidase conjugate (1:12'000). After 3 washes with TBS-Tween, the blot was imaged after applying Clarity Western ECL substrates with Chemidoc MP imaging system Bio-Rad Gel imager. Each pulldown was performed in independent triplicates.

In order to take into account the variability in the amount of nucleosomes loaded on the beads, one of the two following experiments was performed. The anti-FLAG blot was stripped and an anti-H3 Western Blot was performed on the same membrane with anti-H3 rabbit polyclonal antibody (anti-H3 C-terminus, 1:1000) and anti-rabbit HRP conjugate (1:2000). The anti-FLAG bands were then normalized to the anti-H3 bands. Alternatively, the anti-H3 Western Blot was performed prior to the anti-FLAG blot and a corrected volume of the sample was loaded on the anti-FLAG SDS gel. The nucleosome loading on the beads was then found to be homogeneous (Figure 34. B).

MST with cMAP3 and modified histone peptides

cMAP3 binding to H3 peptides was determined by performing titrations using microscale thermophoresis for a readout of the interaction. Lyophilized peptides **5b**, **5d** and **5e** were each dissolved in MST buffer (50 mM Tris-HCl pH = 7.4, 150 mM NaCl, 10 µM $ZnCl_2$) and quantified by UV-VIS spectrophotometry using tyrosine absorption. A titration series was then performed by mixing recombinant cMAP3 with increasing concentrations of peptides **5b**, **5d** or **5e** in a total volume of 20 µL MST buffer containing Tween-20 to a final concentration of 0.05 %. The samples were centrifuged for 10 minutes to eliminate any aggregates. Samples were loaded into premium coated capillaries (Nanotemper) and binding was determined using a

detection LED power of 40 % and laser heating power (MST power) of 40 %. Experiments were performed in independent duplicates. The experimental data was averaged, normalized and analysed by nonlinear fitting using a single-site binding isotherm in Origin (OriginLab).

MST with LSD1 and unmodified nucleosomes

Titrations using microscale thermophoresis were done on a MonoLith NT.115 instrument equipped with green/red filters from NanoTemper technologies. LSD1 protein from a main stock was serially diluted in titration buffer (150 mM NaCl, 20 mM Tris-HCl, 5 % glycerol, pH = 7.8) to cover the concentration range 8 nM - 131 µM. Tween-20 was added to a final concentration of 0.05 %. Cy5-labeled mononucleosomes were added to a final concentration of 10 nM. After centrifugation for 10 min at 15,000 rpm, samples were loaded into premium coated capillaries and cap scanning done at 80 % LED power. Thermophoresis was done at 40 % IR laser power. Data was normalized and fitted to the Hill equation using Origin.

4.5 Mouse ESCs

4.5.1 Culture conditions

Mouse E14 ESCs were cultured in feeder-free conditions. Cells were grown on 0.1 % gelatin-coated Petri dishes in GMEM supplemented with MEM non-essential amino acids (consisting of glycine, L-alanine, L-asparagine, L-aspartic acid, L-glutamic acid, L-proline and L-serine all with a final concentration of 100 µM), sodium pyruvate at a final concentration of 2 mM, penicillin/streptomycin at final concentrations of 100 U/mL and 100 µg/mL, respectively, L-glutamine at a final concentration of 2 mM + 2-mercaptoethanol at a final concentration of 10 µM + 10 % FBS embryonic stem cell-qualified + 10 ng/mL LIF. The medium was supplemented with 3 µM GSK3 inhibitor XVI + 800 nM PD184352. Cells were cultured at 37 °C with 5 % CO_2.

For treatments with small molecule modulators, compounds were dissolved in DMSO at an appropriate concentration. The medium was then supplemented according to the conditions in Table 9.

4.5.2 Transient transfection and confocal microscopy

ESCs were plated on E-cadherin coated wells as previously described.[439] Between 4 and 5 h later, cells were transfected with plasmids pcDNA3.1 containing cMAP3 wild type or mutants under a cytomegalovirus (CMV) promoter using X-tremeGENE 9 DNA transfection reagent and OptiMEM I 1x reduced serum medium. Between 20 h and 24 h later, cells were analysed on a confocal microscope LSM 510 Meta (Zeiss) with Draq5 as DNA stain.

4.5.3 Immunofluorescence

Cells were plated on E-cadherin coated wells. Between 20 h and 24 h later, cells were fixed with 4 % paraformaldehyde for 15 min, washed 3 x with phosphate buffer saline (PBS) and permeabilised with 1 % Triton X-100 in PBS for 20 min. After washing 5 x with PBS, cells were incubated in blocking buffer (1 % BSA in PBS) for 30 min. Cells were then incubated in blocking buffer supplemented with rabbit polyclonal antibody against H3K27me3 (dilution 1:1000) for 1 h. Cells were washed 3 x with PBS for 5 min and incubated in blocking buffer supplemented with Alexa Fluor 647 antibody anti-rabbit antibody (1:1000) for 1 h. After washing 1 x with PBS for 5 min, cells were incubated in 4’,6-diamidino-2-phenylindole (DAPI) in PBS for 5 min. After a final wash in PBS for 5 min, cells were imaged with a confocal microscope LSM700 (Zeiss) in PBS supplemented with 1 mM of the antioxidant compound Trolox which serves as an antifading agent.

4.5.4 Western Blot of acid-extracted histones

Cells were washed with PBS, pelleted and incubated in hypotonic lysis buffer (10 mM Tris-HCl pH = 8.0, 1 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, Protease Inhibitor Cocktail cOmplete™ EDTA-free, 1 mM PMSF, 1 mM NaF) at 4 °C for 30 min. Nuclei were pelleted and incubated in 0.2 M H2SO4 for 2 h at 4 °C. After centrifugation, acid-extracted histones were dialysed against water with 2.5 % v/v acetic acid in Slide-A-Lyzer Mini Dialysis buttons (3'500 Da MWCO) overnight at 4 °C. The histone containing solution was removed from the dialysis device and lyophilized, redissolved in 50 µL water and run on a 12 % SDS PAGE. Proteins were transferred on a PVDF membrane in transfer buffer (10 % methanol, 25 mM Tris, 200 mM glycine) at 200 mA for 2 h at 4 °C. The blot was incubated in blocking solution (10 % non-fat milk in TBS-Tween) for 2 h at RT and then incubated in blocking solution supplemented with rabbit polyclonal antibody against H3K27me3 (1:1000) overnight at 4 °C. After washing 3 x with TBS-Tween, the blot was incubated with blocking solution supplemented with anti-rabbit HRP conjugate (1:2000) for 1 h at RT. After 3 washes with TBS-Tween, the blot was reacted with Clarity Western ECL Blotting substrates and imaged by Chemidoc MP imaging system Bio-Rad gel imager.

4.5.5 Chromatin immunoprecipitation

Growth and harvest

6 confluent plates of mouse E14 ESCs were cross-linked with 1 % formaldehyde (16 % stock solution, Methanol free) and incubated 10 min at RT. The fixation was quenched with 125 mM glycine. Cells were washed 2 x with 10 mL cold PBS and scraped off. After centrifugation, about 70 million cells were stored at - 80 °C.

Chromatin extraction

Chromatin preparation and immunoprecipitation was performed as described previously[440] with some modifications. Briefly, the cell pellet was resuspended in 10

mL cold buffer A (20 mM HEPES-KOH pH = 7.5, 10 mM EDTA, 0.5 mM EGTA, 0.25 % Triton-X-100, Protease inhibitor cocktail) and rotated at 4 °C for 10 min. After centrifugation, the pellet was washed with 10 mL cold buffer B (50 mM HEPES-KOH pH = 7.5, 150 mM NaCl, 1 mM EDTA, Protease inhibitor cocktail) at 4 °C for 10 min. After centrifugation, the pellet was resuspended in 1 mL cold buffer C (20 mM HEPES-KOH pH = 7.5, 1 mM EDTA, 0.5 mM EGTA, 0.05 % SDS, Protease inhibitor cocktail). Cells were sonicated using a Bioruptor sonicator for 20 x on high power (30 sec on / 30 sec off) at 4 °C. Sonication efficiency was checked by Fragment Analyzer. Sonicated chromatin was transferred to a 1.5 mL eppendorf and centrifuged at 14000 g/ 10 min/ 4 °C. The supernatant was collected and used for chromatin immunoprecipitation (100 µL per ChIP).

ChIP

40 µL of Protein G magnetic beads were used per IP. Beads were first washed three times with incubation buffer (10 mM Tris-HCl pH = 8.0, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.15 % SDS, 1 % Triton X-100, Protease inhibitor cocktail). Beads were then blocked overnight at 4 °C in 1.2 mL incubation buffer supplemented with 0.1 % BSA. In parallel, 1 µL of the sonicated chromatin was removed for the input sample while aliquots of 100 µL were incubated overnight at 4 °C with 4 µg of IgG, anti-H3K4me3 or anti-H3K27me3 in incubation buffer supplemented with 0.1 % BSA. After overnight incubation, blocked beads were added to each chromatin preparation and incubated for 2 h at 4 °C in a rotating wheel. Beads were then washed twice with 1 mL Wash Buffer 1 (10 mM Tris pH = 8.1, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % SDS, 0.1 % sodium deoxycholate, 1 % Triton X-100), once with 1 mL Wash Buffer 2 (10 mM Tris pH = 8.1, 500 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % SDS, 0.1 % sodium deoxycholate, 1 % Triton X-100), once with 1 mL Wash Buffer 3 (10 mM Tris pH = 8.1, 250 mM LiCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 % sodium deoxycholate, 1 % NP-40) and twice with 1 mL Wash Buffer 4 (10 mM Tris pH = 8.1, 1 mM EDTA, 0.5 mM EGTA). Co-immunoprecipitated DNA fragments were eluted from the beads in 400 µL elution Buffer (10 mM Tris pH = 7.5, 1 mM EDTA, 1 % SDS, 100 mM $NaHCO_3$, 200 mM NaCl). 10 µL of 20 mg/ml PureLink RNase A was added to DNA fragments and incubated for 30 min

37 °C. RNA-digested chromatin was incubated for 4 h at 65 °C with proteinase K to digest proteins and reverse crosslink, and the DNA was purified using the QIAquick PCR Purification Kit.

qPCR

Quantitative PCR was performed using a StepOnePlus™ System (Applied Biosystems) on 1 µL of IP, input which were mixed to 10 µL 2 X SYBR Select Master Mix (Applied Biosystems), 0.8 µL 10 µM forward / reverse primers, and 7.4 µL H_2O. The PCR program was performed as follows: 2 min at 50 °C; 10 min at 95 °C; and then 40 cycles of 95 °C for 3 sec and 60 °C for 30 sec. Primers listed in Table 11 were used for qPCR analysis.

Name	Forward primer	Reverse primer
Pou5f1	5'- ggc tct cca gag gat ggc tga g -3'	5'- tcg gat gcc cca tcg ca -3'
Polm	5'- tga cgg gca caa tta cac ca -3'	5'- aaa ggc ttc cgc gtc cta ga-3'
Gapdh	5'- ggg ttc cta taa ata cgg act gc -3'	5'- ctg gca ctg cac aag aag atg -3'
Tcf4	5'- cgg atg tga atg gat tac aat g -3'	5'- att gtt ctt cgg tct tgt tgg t -3'
HoxA3	5'- aat tac ctc cct gca tct caa a -3'	5'- tta tca gag cag acc cac aat g -3'
Irx2	5'- taa cac ggc ctg aaa tct tct c -3'	5'- gca tcc cac ttc tac agt cct c -3'
Olig1	5'- ggg tta cag gca gcc acc ta -3'	5'- atg cgg tgg aag agg atg ag -3'
HoxC5	5'- gta ctg cta cgg cgg att gg -3'	5'- tac ccc gtg gag aga gtt gg -3'
HoxB13	5'- ttc gag ctg gga gcg att ta -3'	5'- agc cga ggg tga ggg ttc ta -3'
Gata-4	5'- aag agc gct tgc gtc tct a -3'	5'- ttg cta gcc tca gat cta cgg -3'
des. chr1	5'- gga agc tgg ttt cac atg gt -3'	5'- ccc ttg att tct cga agg ag -3'

Table 11. List of oligonucleotides used for qPCR.

4.6 Histone methyltransferase assays

Asymmetrically modified H3K4me3 nucleosomes (one H3 unmodified and one H3 with K4me3) were obtained as previously described.[207] Briefly, a small peptide named lnc-tag containing an N-terminal cysteine with acetamidomethyl (Acm) protection and a TEV cleavage site was used for the traceless synthesis of asymmetrically modified octamers. N-terminal tail of H3 (with or without K4me3) with lnc-tag at the N-terminus were obtained by SPPS and subsequently ligated with

truncated H3. Following desulfurization, the N-terminal cysteine of both modified and unmodified H3 molecules were deprotected. The disulfide bond between unmodified H3 and H3K4me3 was done by activating one of the cysteine with 5,5′-Dithiobis(2-nitrobenzoic acid) (DTNB). Octamers were refolded using the H3 heterodimer, H4, H2A and H2B. Following nucleosome reconstitution, the lnc-tags were cleaved by TEV protease.

10 pmol nucleosomes were mixed with 0.2 or 2 pmol of SET1B complex and 1 µCi of ^{3}H-SAM in HMT buffer (50 mM Tris-HCl, pH = 8.5, 30 mM KCl, 5 mM $MgCl_2$, 4 mM DTT) in a total volume of 25 µL and incubated at 30 °C for 10 min, 30 min or 2 h. In parallel, 15 pmol octamers were mixed with 0.8 pmol of SET1B complex and 1 µCi of ^{3}H-SAM in the same buffer in a total volume of 25 µL and incubated at RT or 30 °C for 2 h. 20 µL of each samples was boiled in 5 x Laemmli sample buffer and run on a 12 % MINI-Protean TGX Precast gel (Bio-Rad). Following staining with coomassie blue and destaining, the gel was incubated in Amersham Amplify NAMP100V (GE healthcare) for 30 min and dried with DrygelSr. Slab gel dryer (Hoefer Scientific Instruments) for 2 h at 80 °C. Carestream Kodak BioMax MS film was exposed to the dry gel overnight at - 80 °C and the film was developed with the developer CURIX 60 (AGFA). The 5 µL left of the samples were spotted on P81 Ion Exchange Cellulose Chromatography Paper (Reaction Biology) and dried for 40 min at least. Following 3 washes with 50 mM $NaHCO_3$ pH = 9, filter papers were dried on DrygelSr. Slab gel dryer at 40 °C for 15 min. Filter papers were then mixed with 5 mL of Ultima Gold F scintillation cocktail (Perkin Elmer) and vortexed for 30 s. After 1 h incubation at RT, scintillation counting was performed with Tri-Carb 3100TR Perkin Elmer Liquid scintillation analyser.

Chapter 5. Literature

1. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* **389,** 251–60 (1997).
2. Widom, J. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. Natl. Acad. Sci.* **89,** 1095–1099 (1992).
3. Olins, A. L. & Olins, D. E. Spheroid Chromatin Units (ν Bodies). *Science* **183,** 330–332 (1974).
4. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.* **73,** 1897–1901 (1976).
5. Woodcock, C. L., Frado, L. L. & Rattner, J. B. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J. Cell Biol.* **99,** 42–52 (1984).
6. Bednar, J. *et al.* Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci.* **95,** 14173–14178 (1998).
7. Dorigo, B. *et al.* Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science* **306,** 1571–1573 (2004).
8. Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436,** 138–41 (2005).
9. Song, F. *et al.* Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science* **344,** 376–380 (2014).
10. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci.* **103,** 6506–6511 (2006).
11. Robinson, P. J. & Rhodes, D. Structure of the '30nm' chromatin fibre: A key role for the linker histone. *Curr. Opin. Struct. Biol.* **16,** 336–343 (2006).
12. 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* **160,** 1145–1158 (2015).
13. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357,** eaag0025 (2017).
14. Waddington, C. H. An introduction to modern genetics. *Introd. Mod. Genet.* (1939).
15. Holliday, R. Epigenetics: a historical overview. *Epigenetics* **1,** 76–80 (2006).
16. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **23,** 781–783 (2009).
17. Griffith, J. S. & Mahler, H. R. DNA Ticketing Theory of Memory. *Nature* **223,** 580–582 (1969).
18. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366,** 362 (1993).
19. Razin, A. & Cedar, H. DNA methylation and genomic imprinting. *Cell* **77,** 473–476 (1994).
20. Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* **115,** 2376–2418 (2015).
21. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128,** 693–705 (2007).

22. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379,** 131–7 (1996).
23. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the Mechanism of X Chromosome Inactivation. *Annu. Rev. Genet.* **36,** 233–278 (2002).
24. Cheung, P., Allis, C. D. & Sassone-Corsi, P. Signaling to Chromatin through Histone Modifications. *Cell* **103,** 263–271 (2000).
25. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28,** 1057–1068 (2010).
26. Bestor, T. H. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J.* **11,** 2611–2617 (1992).
27. Pradhan, S., Bacolla, A., Wells, R. D. & Roberts, R. J. Recombinant Human DNA (Cytosine-5) Methyltransferase I. EXPRESSION, PURIFICATION, AND COMPARISON OF DE NOVO AND MAINTENANCE METHYLATION. *J. Biol. Chem.* **274,** 33002–33010 (1999).
28. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99,** 247–257 (1999).
29. Bird, A. DNA Methylation de Novo. *Science* **286,** 2287–2288 (1999).
30. Gaydos, L. J., Wang, W. & Strome, S. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* **345,** 1515–1518 (2014).
31. Ragunathan, K., Jih, G. & Moazed, D. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348,** 1258699 (2015).
32. Audergon, P. N. C. B. *et al.* Restricted epigenetic inheritance of H3K9 methylation. *Science* **348,** 132–135 (2015).
33. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441,** 469–74 (2006).
34. Burton, N. O., Burkhart, K. B. & Kennedy, S. Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* **108,** 19683–19688 (2011).
35. Ashe, A. *et al.* piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of C. elegans. *Cell* **150,** 88–99 (2012).
36. Gatewood, J. M., Cook, G. R., Balhorn, R., Schmid, C. W. & Bradbury, E. M. Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. *J. Biol. Chem.* **265,** 20662–20666 (1990).
37. Brykczynska, U. *et al.* Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat. Struct. Mol. Biol.* **17,** 679–687 (2010).
38. Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14,** R47–R58 (2005).
39. Puschendorf, M. *et al.* PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat. Genet.* **40,** 411–420 (2008).
40. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23,** 314–318 (1999).
41. Greer, E. L. *et al.* Transgenerational Epigenetic Inheritance of Longevity in C. elegans. *Nature* **479,** 365–371 (2011).
42. Gu, S. G. *et al.* Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat. Genet.* **44,** 157–164 (2012).

43. Carone, B. R. *et al.* Paternally Induced Transgenerational Environmental Reprogramming of Metabolic Gene Expression in Mammals. *Cell* **143,** 1084–1096 (2010).
44. Öst, A. *et al.* Paternal Diet Defines Offspring Chromatin State and Intergenerational Obesity. *Cell* **159,** 1352–1364 (2014).
45. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **48,** (2012).
46. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–12 (2007).
47. Lu, X. *et al.* The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat. Struct. Mol. Biol.* **15,** 1122–1124 (2008).
48. Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat. Chem. Biol.* **7,** 113–119 (2011).
49. Shogren-Knaak, M. *et al.* Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* **311,** 844–847 (2006).
50. Carrozza, M. J., Utley, R. T., Workman, J. L. & Côté, J. The diverse functions of histone acetyltransferase complexes. *Trends Genet.* **19,** 321–329 (2003).
51. Hsu, J.-Y. *et al.* Mitotic Phosphorylation of Histone H3 Is Governed by Ipl1/aurora Kinase and Glc7/PP1 Phosphatase in Budding Yeast and Nematodes. *Cell* **102,** 279–291 (2000).
52. Seet, B. T., Dikic, I., Ming-Ming, Z. & Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7,** 473–83 (2006).
53. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21,** 564–578 (2011).
54. Li, Y. *et al.* Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol. Cell* **62,** 181–193 (2016).
55. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science* **295,** 2080–2083 (2002).
56. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442,** 91–95 (2006).
57. Dhalluin, C., Carlson, J. E., Zeng, L., He, C. & al, et. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491–6 (1999).
58. LeRoy, G., Rickards, B. & Flint, S. J. The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription. *Mol. Cell* **30,** 51–60 (2008).
59. Filippakopoulos, P. & Knapp, S. The bromodomain interaction module. *FEBS Lett.* **586,** 2692–2704 (2012).
60. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A. & al, et. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–4 (2001).
61. Min, J., Zhang, Y. & Xu, R.-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17,** 1823–1828 (2003).
62. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17,** 1870–1881 (2003).
63. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131,** 58–69 (2007).
64. Sanchez, R. & Zhou, M.-M. The PHD Finger: A Versatile Epigenome Reader. *Trends Biochem. Sci.* **36,** 364–372 (2011).

65. Zeng, L. *et al.* Mechanism and Regulation of Acetylated Histone Binding by the Tandem PHD Finger of DPF3b. *Nature* **466,** 258–262 (2010).
66. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448,** 718–722 (2007).
67. van Ingen, H. *et al.* Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3. *Structure* **16,** 1245–1256 (2008).
68. Hirota, T., Lipp, J. J., Toh, B.-H. & Peters, J.-M. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438,** 1176–80 (2005).
69. 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.* **8,** 983–994 (2007).
70. 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.* **14,** 1025–1040 (2007).
71. Wang, Z. & Patel, D. J. Combinatorial Readout of Dual Histone Modifications by Paired Chromatin-associated Modules. *J. Biol. Chem.* **286,** 18363–18368 (2011).
72. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. Structure and Function of a Human $TAF_{II}250$ Double Bromodomain Module. *Science* **288,** 1422–1425 (2000).
73. Ruthenburg, A. J. *et al.* Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell* **145,** 692–706 (2011).
74. Cowieson, N. P., Partridge, J. F., Allshire, R. C. & McLaughlin, P. J. Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis. *Curr. Biol.* **10,** 517–525 (2000).
75. Kilic, S., Bachmann, A. L., Bryan, L. C. & Fierz, B. Multivalency governs HP1α association dynamics with the silent chromatin state. *Nat. Commun.* **6,** ncomms8313 (2015).
76. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–5 (2000).
77. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293,** 1074–1080 (2001).
78. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28,** 817–825 (2010).
79. Filion, G. J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143,** 212–224 (2010).
80. Consortium, T. modENCODE *et al.* Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE. *Science* **330,** 1787–1797 (2010).
81. Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila. *Nature* **471,** 480–485 (2011).
82. Soloway, P. D. Analysis of Combinatorial Epigenomic States. *ACS Chem. Biol.* **11,** 621–631 (2016).
83. Su, Z. & Denu, J. M. Reading the Combinatorial Histone Language. *ACS Chem. Biol.* **11,** 564–574 (2016).
84. Smith, A. G. Embryo-Derived Stem Cells: Of Mice and Men. *Annu. Rev. Cell Dev. Biol.* **17,** 435–462 (2001).
85. Berdasco, M. & Esteller, M. DNA methylation in stem cell renewal and multipotency. *Stem Cell Res. Ther.* **2,** 42 (2011).
86. Pansky, B. *Review of Medical Embryology*. (McGraw-Hill, 1982).

87. Condic, M. L. Totipotency: What It Is and What It Is Not. *Stem Cells Dev.* **23,** 796–812 (2014).
88. Marikawa, Y. & Alarcón, V. B. Establishment of trophectoderm and inner cell mass lineages in the mouse embryo. *Mol. Reprod. Dev.* **76,** 1019–1032 (2009).
89. Kolios, G. & Moodley, Y. Introduction to Stem Cells and Regenerative Medicine. *Respiration* **85,** 3–10 (2013).
90. Raff, M. Adult Stem Cell Plasticity: Fact or Artifact? *Annu. Rev. Cell Dev. Biol.* **19,** 1–22 (2003).
91. Avasthi, S., Srivastava, R. N., Singh, A. & Srivastava, M. Stem cell: past, present and future--a review article. *Internet J. Med. Update* **3,** 22–31 (2008).
92. Bajada, S., Mazakova, I., Richardson, J. B. & Ashammakhi, N. Updates on stem cells and their applications in regenerative medicine. *J. Tissue Eng. Regen. Med.* **2,** 169–183 (2008).
93. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).
94. Zhou, H. *et al.* Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4,** 381–384 (2009).
95. Kim, D. *et al.* Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. *Cell Stem Cell* **4,** 472–476 (2009).
96. Hou, P. *et al.* Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds. *Science* **341,** 651–654 (2013).
97. Jordan, C. T., Guzman, M. L. & Noble, M. Cancer Stem Cells. *N. Engl. J. Med.* **355,** 1253–1261 (2006).
98. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).
99. Wang, J., Alexander, P. & McKnight, S. L. Metabolic Specialization of Mouse Embryonic Stem Cells. *Cold Spring Harb. Symp. Quant. Biol.* **76,** 183–193 (2011).
100. Savatier, P., Lapillonne, H., Jirmanova, L., Vitelli, L. & Samarut, J. Analysis of the Cell Cycle in Mouse Embryonic Stem Cells. in *Embryonic Stem Cells* 27–33 (Springer, Totowa, NJ, 2002). doi:10.1385/1-59259-241-4:27
101. Spivakov, M. & Fisher, A. G. Epigenetic signatures of stem-cell identity. *Nat. Rev. Genet.* **8,** 263–271 (2007).
102. Chen, C.-Y., Cheng, Y.-Y., Yen, C. Y. T. & Hsieh, P. C. H. Mechanisms of pluripotency maintenance in mouse embryonic stem cells. *Cell. Mol. Life Sci.* **74,** 1805–1817 (2017).
103. Morey, L., Santanach, A. & Croce, L. D. Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. *Mol. Cell. Biol.* **35,** 2716–2728 (2015).
104. Nichols, J. *et al.* Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* **95,** 379–391 (1998).
105. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126–140 (2003).
106. Mitsui, K. *et al.* The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113,** 631–642 (2003).
107. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24,** 372–376 (2000).

108. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. *STEM CELLS* **26,** 903–911 (2008).
109. Kalmar, T. *et al.* Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells. *PLOS Biol.* **7,** e1000149 (2009).
110. Hyslop, L. *et al.* Downregulation of NANOG Induces Differentiation of Human Embryonic Stem Cells to Extraembryonic Lineages. *STEM CELLS* **23,** 1035–1043 (2005).
111. Chew, J.-L. *et al.* Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Mol. Cell. Biol.* **25,** 6031–6046 (2005).
112. Rodda, D. J. *et al.* Transcriptional Regulation of Nanog by OCT4 and SOX2. *J. Biol. Chem.* **280,** 24731–24737 (2005).
113. Hirai, H., Karian, P. & Kikyo, N. Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor. *Biochem. J.* **438,** 11–23 (2011).
114. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688–690 (1988).
115. Hao, J., Li, T.-G., Qi, X., Zhao, D.-F. & Zhao, G.-Q. WNT/β-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev. Biol.* **290,** 81–91 (2006).
116. Kelly, K. F. *et al.* β-Catenin Enhances Oct-4 Activity and Reinforces Pluripotency through a TCF-Independent Mechanism. *Cell Stem Cell* **8,** 214–227 (2011).
117. Paling, N. R. D., Wheadon, H., Bone, H. K. & Welham, M. J. Regulation of Embryonic Stem Cell Self-renewal by Phosphoinositide 3-Kinase-dependent Signaling. *J. Biol. Chem.* **279,** 48063–48070 (2004).
118. Watanabe, S. *et al.* Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene* **25,** 2697–2707 (2006).
119. Bennett, C. N. *et al.* Regulation of Wnt Signaling during Adipogenesis. *J. Biol. Chem.* **277,** 30998–31004 (2002).
120. Cline, G. W. *et al.* Effects of a Novel Glycogen Synthase Kinase-3 Inhibitor on Insulin-Stimulated Glucose Metabolism in Zucker Diabetic Fatty (*fa*/*fa*) Rats. *Diabetes* **51,** 2903–2910 (2002).
121. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
122. Silva, J. *et al.* Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLOS Biol.* **6,** e253 (2008).
123. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–4 (2007).
124. Hayashi, K., Lopes, S. M. C. de S., Tang, F. & Surani, M. A. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell* **3,** 391–401 (2008).
125. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135,** 909–918 (2008).
126. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* **38,** 1027–1032 (2010).

127. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.* **13,** 838–845 (2011).
128. Marks, H. *et al.* The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149,** 590–604 (2012).
129. Meshorer, E. *et al.* Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. *Dev. Cell* **10,** 105–116 (2006).
130. Krejčí, J. *et al.* Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. *J. Cell. Physiol.* **219,** 677–687 (2009).
131. Efroni, S. *et al.* Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2,** 437–447 (2008).
132. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome Regulation by Polycomb and Trithorax Proteins. *Cell* **128,** 735–745 (2007).
133. Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352,** aad9780 (2016).
134. Geisler, S. J. & Paro, R. Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development* **142,** 2876–2887 (2015).
135. Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276,** 565–570 (1978).
136. Ingham, P. W. Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature* **306,** 591–593 (1983).
137. Aloia, L., Stefano, B. D. & Croce, L. D. Polycomb complexes in stem cells and embryonic development. *Development* **140,** 2525–2534 (2013).
138. Montgomery, N. D. *et al.* The Murine Polycomb Group Protein Eed Is Required for Global Histone H3 Lysine-27 Methylation. *Curr. Biol.* **15,** 942–947 (2005).
139. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298,** 1039–1043 (2002).
140. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional ONand OFF states by trxG and PcG proteins. *Genes Dev.* **20,** 2041–2054 (2006).
141. Schwartz, Y. B. *et al.* Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat. Genet.* **38,** 700–705 (2006).
142. Ferrari, K. J. *et al.* Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Mol. Cell* **53,** 49–62 (2014).
143. Nekrasov, M., Wild, B. & Müller, J. Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.* **6,** 348–353 (2005).
144. Margueron, R. *et al.* Role of the polycomb protein Eed in the propagation of repressive histone marks. *Nature* **461,** 762–767 (2009).
145. Cao, R. & Zhang, Y. SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. *Mol. Cell* **15,** 57–67 (2004).
146. Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23,** 4061–4071 (2004).
147. Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464,** 306–10 (2010).
148. Peng, J. C. *et al.* Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell* **139,** 1290–1302 (2009).

149. Cao, R. *et al.* Role of hPHF1 in H3K27 Methylation and Hox Gene Silencing. *Mol. Cell. Biol.* **28,** 1862–1872 (2008).
150. Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. *Mol. Cell. Biol.* **28,** 2718–2731 (2008).
151. Walker, E. *et al.* Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* **6,** 153–166 (2010).
152. Endoh, M. *et al.* Histone H2A Mono-Ubiquitination Is a Crucial Step to Mediate PRC1-Dependent Repression of Developmental Genes to Maintain ES Cell Identity. *PLOS Genet.* **8,** e1002774 (2012).
153. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135,** 1513–1524 (2008).
154. Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **1,** (2012).
155. He, J. *et al.* Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat. Cell Biol.* **15,** 373–384 (2013).
156. Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Mol. Cell* **49,** 1134–1146 (2013).
157. Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. *Mol. Cell* **20,** 845–854 (2005).
158. Isono, K. *et al.* SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Dev. Cell* **26,** 565–577 (2013).
159. Kaustov, L. *et al.* Recognition and Specificity Determinants of the Human Cbx Chromodomains. *J. Biol. Chem.* **286,** 521–529 (2011).
160. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25,** 3110–3122 (2006).
161. Tavares, L. *et al.* RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell* **148,** 664–678 (2012).
162. Blackledge, N. P. *et al.* Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* **157,** 1445–1459 (2014).
163. Cooper, S. *et al.* Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Rep.* **7,** 1456–1470 (2014).
164. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **21,** 569–571 (2014).
165. Schuettengruber, B., Martinez, A.-M., Iovino, N. & Cavalli, G. Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* **12,** 799–814 (2011).
166. Byrd, K. N. & Shearn, A. ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. *Proc. Natl. Acad. Sci.* **100,** 11535–11540 (2003).
167. Dou, Y. *et al.* Physical Association and Coordinate Function of the H3 K4 Methyltransferase MLL1 and the H4 K16 Acetyltransferase MOF. *Cell* **121,** 873–885 (2005).

168. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129,** 823–837 (2007).
169. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).
170. Steward, M. M. *et al.* Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat. Struct. Mol. Biol.* **13,** 852–4 (2006).
171. Wysocka, J. *et al.* WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development. *Cell* **121,** 859–872 (2005).
172. Couture, J.-F., Collazo, E. & Trievel, R. C. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat. Struct. Mol. Biol.* **13,** 698–703 (2006).
173. Sarvan, S. *et al.* Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat. Struct. Mol. Biol.* **18,** 857–859 (2011).
174. Chen, Y. *et al.* Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep.* **12,** 797–803 (2011).
175. Cao, F. *et al.* An Ash2L/RbBP5 Heterodimer Stimulates the MLL1 Methyltransferase Activity through Coordinated Substrate Interactions with the MLL1 SET Domain. *PLOS ONE* **5,** e14102 (2010).
176. Zhang, P. *et al.* A phosphorylation switch on RbBP5 regulates histone H3 Lys4 methylation. *Genes Dev.* **29,** 123–128 (2015).
177. Jiang, H. *et al.* Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. *Cell* **144,** 513–525 (2011).
178. Yang, Y. & Hua, X. In search of tumor suppressing functions of menin. *Mol. Cell. Endocrinol.* **265,** 34–41 (2007).
179. Issaeva, I. *et al.* Knockdown of ALR (MLL2) Reveals ALR Target Genes and Leads to Alterations in Cell Adhesion and Growth. *Mol. Cell. Biol.* **27,** 1889–1903 (2007).
180. Clouaire, T. *et al.* Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* **26,** 1714–1728 (2012).
181. Wu, M. *et al.* Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Mol. Cell. Biol.* **28,** 7337–7344 (2008).
182. Hu, D. *et al.* The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **20,** 1093–1097 (2013).
183. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141,** 526–537 (2014).
184. Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Mol. Cell. Biol.* **33,** 4745–4754 (2013).
185. Goo, Y.-H. *et al.* Activating Signal Cointegrator 2 Belongs to a Novel Steady-State Complex That Contains a Subset of Trithorax Group Proteins. *Mol. Cell. Biol.* **23,** 140–149 (2003).
186. Bögershausen, N., Bruford, E. & Wollnik, B. Skirting the pitfalls: a clear-cut nomenclature for H3K4 methyltransferases. *Clin. Genet.* **83,** 212–214 (2013).
187. Bledau, A. S. *et al.* The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* **141,** 1022–1035 (2014).
188. Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. J. & Korsmeyer, S. J. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378,** 505–8 (1995).

189. Glaser, S. *et al.* The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics Chromatin* **2,** 5 (2009).
190. Lee, J.-E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* **2,** (2013).
191. O'Carroll, D. *et al.* The Polycomb-Group GeneEzh2 Is Required for Early Mouse Development. *Mol. Cell. Biol.* **21,** 4330–4336 (2001).
192. Schumacher, A., Faust, C. & Magnuson, T. Positional cloning of a global regulator of anterior–posterior patterning in mice. *Nature* **383,** 250–253 (1996).
193. Wang, J., Mager, J., Schnedier, E. & Magnuson, T. The mouse PcG gene eed is required for Hox gene repression and extraembryonic development. *Mamm. Genome* **13,** 493–503 (2002).
194. Voncken, J. W. *et al.* Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci.* **100,** 2468–2473 (2003).
195. Riising, E. M. *et al.* Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Mol. Cell* **55,** 347–360 (2014).
196. Peters, A. H. F. M. *et al.* Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Mol. Cell* **12,** 1577–1589 (2003).
197. King, I. F. G., Francis, N. J. & Kingston, R. E. Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro. *Mol. Cell. Biol.* **22,** 7919–7928 (2002).
198. Shao, Z. *et al.* Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. *Cell* **98,** 37–46 (1999).
199. Francis, N. J., Saurin, A. J., Shao, Z. & Kingston, R. E. Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol. Cell* **8,** 545–556 (2001).
200. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin Compaction by a Polycomb Group Protein Complex. *Science* **306,** 1574–1577 (2004).
201. Grau, D. J. *et al.* Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* **25,** 2210–2221 (2011).
202. Eskeland, R. *et al.* Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol. Cell* **38,** 452–464 (2010).
203. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439,** 871–4 (2006).
204. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442,** 86–90 (2006).
205. Pérez-Lluch, S. *et al.* Genome-wide chromatin occupancy analysis reveals a role for ASH2 in transcriptional pausing. *Nucleic Acids Res.* **39,** 4628–4639 (2011).
206. Schmitges, F. W. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* **42,** 330–341 (2011).
207. Lechner, C. C., Agashe, N. D. & Fierz, B. Traceless Synthesis of Asymmetrically Modified Bivalent Nucleosomes. *Angew. Chem. Int. Ed.* **55,** 2903–2906 (2016).
208. Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* **136,** 3131–3141 (2009).
209. Pasini, D. *et al.* Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.* **38,** 4958–4969 (2010).

210. Reynolds, N. *et al.* NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J.* **31,** 593–605 (2012).
211. Tie, F. *et al.* Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* **141,** 1129–1139 (2014).
212. Kim, D.-H. *et al.* Histone H3K27 Trimethylation Inhibits H3 Binding and Function of SET1-Like H3K4 Methyltransferase Complexes. *Mol. Cell. Biol.* **33,** 4936–4946 (2013).
213. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16,** 245–257 (2015).
214. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326,** 289–293 (2009).
215. Dixon, J. R. *et al.* Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **485,** 376–380 (2012).
216. Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148,** 458–472 (2012).
217. Kim, Y. J., Cecchini, K. R. & Kim, T. H. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci.* **108,** 7391–7396 (2011).
218. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515,** 402–405 (2014).
219. Saurin, A. J. *et al.* The Human Polycomb Group Complex Associates with Pericentromeric Heterochromatin to Form a Novel Nuclear Domain. *J. Cell Biol.* **142,** 887–898 (1998).
220. Buchenau, P., Hodgson, J., Strutt, H. & Arndt-Jovin, D. J. The Distribution of Polycomb-Group Proteins During Cell Division and Development in *Drosophila* Embryos: Impact on Models for Silencing. *J. Cell Biol.* **141,** 469–481 (1998).
221. Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M. & Cavalli, G. Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. *Genes Dev.* **17,** 2406–2420 (2003).
222. Vazquez, J., Müller, M., Pirrotta, V. & Sedat, J. W. The Mcp Element Mediates Stable Long-Range Chromosome–Chromosome Interactions in Drosophila. *Mol. Biol. Cell* **17,** 2158–2165 (2006).
223. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47,** 1179–1186 (2015).
224. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12,** 1059–1065 (1993).
225. Wansink, D. G. *et al.* Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell Biol.* **122,** 283–293 (1993).
226. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36,** 1065–71 (2004).
227. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42,** 53–61 (2010).
228. Osborne, C. S. *et al.* Myc Dynamically and Preferentially Relocates to a Transcription Factory Occupied by Igh. *PLOS Biol.* **5,** e192 (2007).
229. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription “factories9 in human nuclei. *J. Cell Sci.* **109,** 1427–1436 (1996).

230. Jackson, D. A., Iborra, F. J., Manders, E. M. M. & Cook, P. R. Numbers and Organization of RNA Polymerases, Nascent Transcripts, and Transcription Units in HeLa Nuclei. *Mol. Biol. Cell* **9,** 1523–1536 (1998).
231. Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M.-L. & Kassis, J. A. The Drosophila Polycomb Group Gene pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. *Mol. Cell* **1,** 1057–1064 (1998).
232. Fritsch, C., Brown, J. L., Kassis, J. A. & Muller, J. The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development* **126,** 3905–3913 (1999).
233. Mohd-Sarip, A., Venturini, F., Chalkley, G. E. & Verrijzer, C. P. Pleiohomeotic Can Link Polycomb to DNA and Mediate Transcriptional Repression. *Mol. Cell. Biol.* **22,** 7473–7483 (2002).
234. Mendenhall, E. M. *et al.* GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *PLOS Genet.* **6,** e1001244 (2010).
235. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression., Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation. *Nat. Nat.* **472, 472,** 120, 120–124 (2011).
236. Gomez, J. A. *et al.* The NeST Long ncRNA Controls Microbial Susceptibility and Epigenetic Activation of the Interferon-γ Locus. *Cell* **152,** 743–754 (2013).
237. Grote, P. *et al.* The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. *Dev. Cell* **24,** 206–214 (2013).
238. Brockdorff, N. Noncoding RNA and Polycomb recruitment. *RNA* **19,** 429–442 (2013).
239. Herzog, V. A. *et al.* A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nat. Genet.* **46,** 973–981 (2014).
240. Yang, Y. W. *et al.* Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency., Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *ELife ELife* **3, 3,** e02046–e02046 (2014).
241. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).
242. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8,** 532–8 (2006).
243. Zhao, X. D. *et al.* Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells. *Cell Stem Cell* **1,** 286–298 (2007).
244. Pan, G. *et al.* Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell* **1,** 299–312 (2007).
245. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
246. Voigt, P. *et al.* Asymmetrically Modified Nucleosomes. *Cell* **151,** 181–193 (2012).
247. Mohn, F. *et al.* Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Mol. Cell* **30,** 755–766 (2008).
248. Cui, K. *et al.* Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. *Cell Stem Cell* **4,** 80–93 (2009).
249. Weiner, A. *et al.* Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. *Nat. Biotechnol.* **34,** 953–961 (2016).

250. Kinkley, S. *et al.* reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in $CD4^+$ memory T cells. *Nat. Commun.* **7,** ncomms12514 (2016).
251. Jadhav, U. *et al.* Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* **165,** 1389–1400 (2016).
252. Lin, B. *et al.* Global analysis of H3K4me3 and H3K27me3 profiles in glioblastoma stem cells and identification of SLC17A7 as a bivalent tumor suppressor gene. *Oncotarget* **6,** 5369–5381 (2015).
253. Bapat, S. A. *et al.* Multivalent epigenetic marks confer microenvironment-responsive epigenetic plasticity to ovarian cancer cells. *Epigenetics* **5,** 716–729 (2010).
254. Rodriguez, J. *et al.* Bivalent domains enforce transcriptional memory of DNA methylated genes in cancer cells. *Proc. Natl. Acad. Sci.* **105,** 19809–19814 (2008).
255. Chapman-Rothe, N. *et al.* Chromatin H3K27me3/H3K4me3 histone marks define gene sets in high-grade serous ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumour cells. *Oncogene* **32,** 4586–4592 (2013).
256. Roh, T.-Y., Cuddapah, S., Cui, K. & Zhao, K. The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci.* **103,** 15782–15787 (2006).
257. Li, Q., Lian, S., Dai, Z., Xiang, Q. & Dai, X. BGDB: a database of bivalent genes. *Database* **2013,** (2013).
258. Kersey, P. J. *et al.* Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.* **44,** D574–D580 (2016).
259. Ku, M. *et al.* Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. *PLOS Genet.* **4,** e1000242 (2008).
260. Furey, T. S. ChIP-seq and Beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13,** 840–852 (2012).
261. Grzybowski, A. T., Chen, Z. & Ruthenburg, A. J. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Mol. Cell* **58,** 886–899 (2015).
262. Sen, S., Block, K. F., Pasini, A., Baylin, S. B. & Easwaran, H. Genome-wide positioning of bivalent mononucleosomes. *BMC Med. Genomics* **9,** 60 (2016).
263. Shema, E. *et al.* Single-molecule decoding of combinatorially modified nucleosomes. *Science* **352,** 717–721 (2016).
264. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
265. Lesch, B. J. & Page, D. C. Poised chromatin in the mammalian germ line. *Development* **141,** 3619–3626 (2014).
266. Sin, H.-S., Kartashov, A. V., Hasegawa, K., Barski, A. & Namekawa, S. H. Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biol.* **13,** 53 (2015).
267. Singh, A. M. *et al.* Cell-Cycle Control of Bivalent Epigenetic Domains Regulates the Exit from Pluripotency. *Stem Cell Rep.* **5,** 323–336 (2015).
268. Lynch, M. D. *et al.* An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.* **31,** 317–329 (2012).
269. Court, F. & Arnaud, P. An annotated list of bivalent chromatin regions in human ES cells: a new tool for cancer epigenetic research. *Oncotarget* **8,** 4110–4124 (2016).
270. Lee, S.-M. *et al.* Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes. *Proc. Natl. Acad. Sci.* **114,** E1885–E1894 (2017).

271. Graham, B. *et al.* MicroRNAs of the miR-290–295 Family Maintain Bivalency in Mouse Embryonic Stem Cells. *Stem Cell Rep.* **6,** 635–642 (2016).
272. Neri, F. *et al.* Dnmt3L Antagonizes DNA Methylation at Bivalent Promoters and Favors DNA Methylation at Gene Bodies in ESCs. *Cell* **155,** 121–134 (2013).
273. Kong, L. *et al.* A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG islands. *Nucleic Acids Res.* **44,** 8682–8692 (2016).
274. Dhar, S. S. *et al.* An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Res.* **44,** 3659–3674 (2016).
275. Adamo, A. *et al.* LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat. Cell Biol.* **13,** 652–659 (2011).
276. Beisel, C. *et al.* Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. *Proc. Natl. Acad. Sci.* **104,** 16615–16620 (2007).
277. Schwartz, Y. B. *et al.* Alternative Epigenetic Chromatin States of Polycomb Target Genes. *PLOS Genet.* **6,** e1000805 (2010).
278. Hammoud, S. S. *et al.* Distinctive Chromatin in Human Sperm Packages Genes for Embryo Development. *Nature* **460,** 473–478 (2009).
279. Erkek, S. *et al.* Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol.* **20,** 868–875 (2013).
280. Lesch, B. J., Dokshin, G. A., Young, R. A., McCarrey, J. R. & Page, D. C. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc. Natl. Acad. Sci.* **110,** 16061–16066 (2013).
281. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* **9,** 1428–35 (2007).
282. Ma, C. *et al.* Nono, a Bivalent Domain Factor, Regulates Erk Signaling and Mouse Embryonic Stem Cell Pluripotency. *Cell Rep.* **17,** 997–1007 (2016).
283. Bernhart, S. H. *et al.* Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Sci. Rep.* **6,** srep37393 (2016).
284. Maupetit-Méhouas, S. *et al.* Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. *Nucleic Acids Res.* **44,** 621–635 (2016).
285. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The Fluorescent Toolbox for Assessing Protein Location and Function. *Science* **312,** 217–224 (2006).
286. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2,** 261–7 (2005).
287. Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* **2,** 201–6 (2005).
288. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **244,** 182 (1989).
289. Bain, J. D., Diala, E. S., Glabe, C. G., Dix, T. A. & Chamberlin, A. R. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J. Am. Chem. Soc.* **111,** 8013–8014 (1989).
290. Huang, Y. *et al.* A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli. *Mol. Biosyst.* **6,** 683–686 (2010).
291. Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H. & Sisido, M. Incorporation of Nonnatural Amino Acids into Proteins by Using Various Four-Base Codons in an

Escherichia coli in Vitro Translation System. *Biochemistry (Mosc.)* **40,** 11060–11064 (2001).
292. Anderson, J. C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7566–7571 (2004).
293. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nϵ-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4,** 232–234 (2008).
294. Neumann, H. *et al.* A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Mol. Cell* **36,** 153–163 (2009).
295. Gattner, M. J., Vrabel, M. & Carell, T. Synthesis of ε- N -propionyl-, ε- N -butyryl-, and ε- N -crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun.* **49,** 379–381 (2013).
296. Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically Encoding Nϵ-Methyl-l-lysine in Recombinant Histones. *J. Am. Chem. Soc.* **131,** 14194–14195 (2009).
297. Nguyen, D. P., Alai, M. M. G., Virdee, S. & Chin, J. W. Genetically Directing ε-N, N-Dimethyl-l-Lysine in Recombinant Histones. *Chem. Biol.* **17,** 1072–1076 (2010).
298. Ai, H., Wook Lee, J. & G. Schultz, P. A method to site-specifically introduce methyllysine into proteins in E. coli. *Chem. Commun.* **46,** 5506–5508 (2010).
299. Virdee, S. *et al.* Traceless and Site-Specific Ubiquitination of Recombinant Proteins. *J. Am. Chem. Soc.* **133,** 10708–10711 (2011).
300. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128,** 1003–1012 (2007).
301. Li, F. *et al.* A Direct Method for Site-Specific Protein Acetylation. *Angew. Chem.* **123,** 9785–9788 (2011).
302. Le, D. D., Cortesi, A. T., Myers, S. A., Burlingame, A. L. & Fujimori, D. G. Site-Specific and Regiospecific Installation of Methylarginine Analogues into Recombinant Histones and Insights into Effector Protein Binding. *J. Am. Chem. Soc.* **135,** 2879–2882 (2013).
303. Chatterjee, C., McGinty, R. K., Fierz, B. & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **6,** 267–269 (2010).
304. Dhall, A. *et al.* Sumoylated Human Histone H4 Prevents Chromatin Compaction by Inhibiting Long-range Internucleosomal Interactions. *J. Biol. Chem.* **289,** 33827–33837 (2014).
305. 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.* **130,** 5052–5053 (2008).
306. Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chem.* **124,** 1871–1875 (2012).
307. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354,** aag1465 (2016).
308. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science* **354,** 623–626 (2016).
309. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).

310. Merrifield, R. B. Solid Phase Peptide Synthesis. II. The Synthesis of Bradykinin. *J. Am. Chem. Soc.* **86,** 304–305 (1964).
311. Nilsson, B. L., Kiessling, L. L. & Raines, R. T. Staudinger Ligation: A Peptide from a Thioester and Azide. *Org. Lett.* **2,** 1939–1941 (2000).
312. Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A "Traceless" Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2,** 2141–2143 (2000).
313. Bode, J. W., Fox, R. M. & Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and α-Ketoacids. *Angew. Chem.* **118,** 1270–1274 (2006).
314. Pusterla, I. & Bode, J. W. An oxazetidine amino acid for chemical protein synthesis by rapid, serine-forming ligations. *Nat. Chem.* **7,** 668–672 (2015).
315. Zhang, Y., Xu, C., Lam, H. Y., Lee, C. L. & Li, X. Protein chemical synthesis by serine and threonine ligation. *Proc. Natl. Acad. Sci.* **110,** 6657–6662 (2013).
316. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266,** 776–779 (1994).
317. Tam, J. P., Lu, Y.-A., Liu, C.-F. & Shao, J. Peptide synthesis using unprotected peptides through orthogonal coupling methods., Peptide synthesis using unprotected peptides through orthogonal coupling methods. *Proc. Natl. Acad. Sci. U. S. Am. Proc. Natl. Acad. Sci. U. S. Am.* **92, 92,** 12485, 12485–12489 (1995).
318. Hojo, H. & Aimoto, S. Polypeptide Synthesis Using the S-Alkyl Thioester of a Partially Protected Peptide Segment. Synthesis of the DNA-Binding Domain of c-Myb Protein (142–193)–NH2. *Bull. Chem. Soc. Jpn.* **64,** 111–117 (1991).
319. Lu, W., Qasim, M. A. & Kent, S. B. H. Comparative Total Syntheses of Turkey Ovomucoid Third Domain by Both Stepwise Solid Phase Peptide Synthesis and Native Chemical Ligation. *J. Am. Chem. Soc.* **118,** 8518–8523 (1996).
320. Canne, L. E., Walker, S. M. & Kent, S. B. H. A general method for the synthesis of thioester resin linkers for use in the solid phase synthesis of peptide-α-thioacids. *Tetrahedron Lett.* **36,** 1217–1220 (1995).
321. 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.* **47,** 6851–6855 (2008).
322. Ollivier, N. *et al.* Tidbits for the synthesis of bis(2-sulfanylethyl)amido (SEA) polystyrene resin, SEA peptides and peptide thioesters. *J. Pept. Sci.* **20,** 92–97 (2014).
323. Erlich, L. A., Kumar, K. S. A., Haj-Yahya, M., Dawson, P. E. & Brik, A. N-Methylcysteine -mediated total chemical synthesis of ubiquitin thioester. *Org. Biomol. Chem.* **8,** 2392–2396 (2010).
324. Fang, G.-M. *et al.* Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* **50,** 7645–7649 (2011).
325. Zheng, J., Tang, S., Qi, Y., Wang, Z. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **8,** 2483–95 (2013).
326. 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.* **2,** 315–318 (1996).
327. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95,** 6705–6710 (1998).

328. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5,** 75–86 (2004).
329. Zuo, X. *et al.* Enhanced Expression and Purification of Membrane Proteins by SUMO Fusion in Escherichia coli. *J. Struct. Funct. Genomics* **6,** 103–111 (2005).
330. Pentelute, B. L., Barker, A. P., Janowiak, B. E., Kent, S. B. H. & Collier, R. J. A Semisynthesis Platform for Investigating Structure−Function Relationships in the N-Terminal Domain of the Anthrax Lethal Factor. *ACS Chem. Biol.* **5,** 359–364 (2010).
331. 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.* **3,** 981–991 (1996).
332. Tolbert, T. J. & Wong, C.-H. New Methods for Proteomic Research: Preparation of Proteins with N-Terminal Cysteines for Labeling and Conjugation. *Angew. Chem.* **114,** 2275–2278 (2002).
333. Liu, D., Xu, R., Dutta, K. & Cowburn, D. N-terminal cysteinyl proteins can be prepared using thrombin cleavage. *FEBS Lett.* **582,** 1163–1167 (2008).
334. Busch, G. K. *et al.* Specific N-terminal protein labelling: use of FMDV 3C pro protease and native chemical ligation. *Chem. Commun.* **0,** 3369–3371 (2008).
335. Vila-Perelló, M. & Muir, T. W. Biological Applications of Protein Splicing. *Cell* **143,** 191–200 (2010).
336. Shah, N. H. & Muir, T. W. Inteins: nature's gift to protein chemists. *Chem. Sci.* **5,** 446–461 (2014).
337. Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Naturally Split Inteins Assemble through a "Capture and Collapse" Mechanism. *J. Am. Chem. Soc.* **135,** 18673–18681 (2013).
338. Mootz, H. D. Split Inteins as Versatile Tools for Protein Semisynthesis. *ChemBioChem* **10,** 2579–2589 (2009).
339. Xu, M. Q. & Perler, F. B. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15,** 5146–5153 (1996).
340. Chong, S. *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192,** 271–281 (1997).
341. Shah, N. H., Vila-Perelló, M. & Muir, T. W. Kinetic Control of One-Pot Trans-Splicing Reactions by Using a Wild-Type and Designed Split Intein. *Angew. Chem. Int. Ed.* **50,** 6511–6515 (2011).
342. Vila-Perelló, M. *et al.* Streamlined Expressed Protein Ligation Using Split Inteins. *J. Am. Chem. Soc.* **135,** 286–292 (2013).
343. 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.* **123,** 526–533 (2001).
344. Pentelute, B. L. & Kent, S. B. H. Selective Desulfurization of Cysteine in the Presence of Cys(Acm) in Polypeptides Obtained by Native Chemical Ligation. *Org. Lett.* **9,** 687–690 (2007).
345. Haase, C., Rohde, H. & Seitz, O. Native Chemical Ligation at Valine. *Angew. Chem. Int. Ed.* **47,** 6807–6810 (2008).
346. Chen, J., Wan, Q., Yuan, Y., Zhu, J. & Danishefsky, S. J. Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angew. Chem.* **120,** 8649–8652 (2008).

347. Crich, D. & Banerjee, A. Native Chemical Ligation at Phenylalanine. *J. Am. Chem. Soc.* **129,** 10064–10065 (2007).
348. Siman, P., Karthikeyan, S. V. & Brik, A. Native Chemical Ligation at Glutamine. *Org. Lett.* **14,** 1520–1523 (2012).
349. Chen, J., Wang, P., Zhu, J., Wan, Q. & Danishefsky, S. J. A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. *Tetrahedron* **66,** 2277–2283 (2010).
350. Harpaz, Z., Siman, P., Kumar, K. S. A. & Brik, A. Protein Synthesis Assisted by Native Chemical Ligation at Leucine. *ChemBioChem* **11,** 1232–1235 (2010).
351. Tan, Z., Shang, S. & Danishefsky, S. J. Insights into the Finer Issues of Native Chemical Ligation: An Approach to Cascade Ligations. *Angew. Chem. Int. Ed.* **49,** 9500–9503 (2010).
352. Cergol, K. M., Thompson, R. E., Malins, L. R., Turner, P. & Payne, R. J. One-Pot Peptide Ligation–Desulfurization at Glutamate. *Org. Lett.* **16,** 290–293 (2014).
353. Thompson, R. E., Chan, B., Radom, L., Jolliffe, K. A. & Payne, R. J. Chemoselective Peptide Ligation–Desulfurization at Aspartate. *Angew. Chem. Int. Ed.* **52,** 9723–9727 (2013).
354. Malins, L. R., Cergol, K. M. & Payne, R. J. Chemoselective sulfenylation and peptide ligation at tryptophan. *Chem. Sci.* **5,** 260–266 (2014).
355. Malins, L. R., Cergol, K. M. & Payne, R. J. Peptide Ligation–Desulfurization Chemistry at Arginine. *ChemBioChem* **14,** 559–563 (2013).
356. Shang, S., Tan, Z., Dong, S. & Danishefsky, S. J. An Advance in Proline Ligation. *J. Am. Chem. Soc.* **133,** 10784–10786 (2011).
357. Ding, H., Shigenaga, A., Sato, K., Morishita, K. & Otaka, A. Dual Kinetically Controlled Native Chemical Ligation Using a Combination of Sulfanylproline and Sulfanylethylanilide Peptide. *Org. Lett.* **13,** 5588–5591 (2011).
358. Townsend, S. D. *et al.* Advances in Proline Ligation. *J. Am. Chem. Soc.* **134,** 3912–3916 (2012).
359. Kumar, K. S. A., Haj-Yahya, M., Olschewski, D., Lashuel, H. A. & Brik, A. Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chem. Int. Ed.* **48,** 8090–8094 (2009).
360. Yang, R., Pasunooti, K. K., Li, F., Liu, X.-W. & Liu, C.-F. Dual Native Chemical Ligation at Lysine. *J. Am. Chem. Soc.* **131,** 13592–13593 (2009).
361. 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.* **37,** 2271–2274 (1996).
362. Node, M. *et al.* A Raney nickel — sodium hypophosphite combination system for reductive desulfurization without racemization of optically active secondary alcohol. *Tetrahedron* **53,** 12883–12894 (1997).
363. 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.* **46,** 9248–9252 (2007).
364. Siman, P. *et al.* Chemical Synthesis and Expression of the HIV-1 Rev Protein. *ChemBioChem* **12,** 1097–1104 (2011).
365. Moyal, T., P. Hemantha, H., Siman, P., Refua, M. & Brik, A. Highly efficient one-pot ligation and desulfurization. *Chem. Sci.* **4,** 2496–2501 (2013).

366. Johnson, E. C. B. & Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **128,** 6640–6646 (2006).
367. Thompson, R. E. *et al.* Trifluoroethanethiol: An Additive for Efficient One-Pot Peptide Ligation−Desulfurization Chemistry. *J. Am. Chem. Soc.* **136,** 8161–8164 (2014).
368. Huang, Y.-C. *et al.* Synthesis of l- and d-Ubiquitin by One-Pot Ligation and Metal-Free Desulfurization. *Chem. – Eur. J.* **22,** 7623–7628 (2016).
369. He, S. *et al.* Facile synthesis of site-specifically acetylated and methylated histone proteins: Reagents for evaluation of the histone code hypothesis. *Proc. Natl. Acad. Sci.* **100,** 12033–12038 (2003).
370. Shogren-Knaak, M. A., Fry, C. J. & Peterson, C. L. A Native Peptide Ligation Strategy for Deciphering Nucleosomal Histone Modifications. *J. Biol. Chem.* **278,** 15744–15748 (2003).
371. Cotton, G. J., Ayers, B., Xu, R. & Muir, T. W. Insertion of a Synthetic Peptide into a Recombinant Protein Framework: A Protein Biosensor. *J. Am. Chem. Soc.* **121,** 1100–1101 (1999).
372. Cotton, G. J. & Muir, T. W. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.* **7,** 253–261 (2000).
373. McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G. & Muir, T. W. Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **453,** 812–816 (2008).
374. McGinty, R. K. *et al.* Structure–Activity Analysis of Semisynthetic Nucleosomes: Mechanistic Insights into the Stimulation of Dot1L by Ubiquitylated Histone H2B. *ACS Chem. Biol.* **4,** 958–968 (2009).
375. 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.* **134,** 19548–19551 (2012).
376. 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.* **276,** 19–42 (1998).
377. Zheng, C. & Hayes, J. J. Intra- and Inter-nucleosomal Protein-DNA Interactions of the Core Histone Tail Domains in a Model System. *J. Biol. Chem.* **278,** 24217–24224 (2003).
378. Blacketer, M. J., Feely, S. J. & Shogren-Knaak, M. A. Nucleosome Interactions and Stability in an Ordered Nucleosome Array Model System. *J. Biol. Chem.* **285,** 34597–34607 (2010).
379. 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.* **12,** 188–193 (2016).
380. Nguyen, U. T. T. *et al.* Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. *Nat. Methods* **11,** 834–840 (2014).
381. Hattori, N., Niwa, T., Kimura, K., Helin, K. & Ushijima, T. Visualization of multivalent histone modification in a single cell reveals highly concerted epigenetic changes on differentiation of embryonic stem cells. *Nucleic Acids Res.* **41,** 7231–7239 (2013).
382. van Dieck, J. *et al.* Development of Bispecific Molecules for the In Situ Detection of Protein-Protein Interactions and Protein Phosphorylation. *Chem. Biol.* **21,** 357–368 (2014).

383. Hayashi-Takanaka, Y., Yamagata, K., Nozaki, N. & Kimura, H. Visualizing histone modifications in living cells: spatiotemporal dynamics of H3 phosphorylation during interphase. *J. Cell Biol.* **187,** 781–790 (2009).
384. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39,** 6475–6488 (2011).
385. Sato, Y. *et al.* Genetically encoded system to track histone modification in vivo. *Sci. Rep.* **3,** (2013).
386. Kimura, H., Hayashi-Takanaka, Y., Stasevich, T. J. & Sato, Y. Visualizing posttranslational and epigenetic modifications of endogenous proteins in vivo. *Histochem. Cell Biol.* **144,** 101–109 (2015).
387. Sato, Y. *et al.* A Genetically Encoded Probe for Live-Cell Imaging of H4K20 Monomethylation. *J. Mol. Biol.* **428,** 3885–3902 (2016).
388. Sasaki, K., Ito, T., Nishino, N., Khochbin, S. & Yoshida, M. Real-time imaging of histone H4 hyperacetylation in living cells. *Proc. Natl. Acad. Sci.* **106,** 16257–16262 (2009).
389. Ito, T. *et al.* Real-Time Imaging of Histone H4K12–Specific Acetylation Determines the Modes of Action of Histone Deacetylase and Bromodomain Inhibitors. *Chem. Biol.* **18,** 495–507 (2011).
390. Nakaoka, S., Sasaki, K., Ito, A., Nakao, Y. & Yoshida, M. A Genetically Encoded FRET Probe to Detect Intranucleosomal Histone H3K9 or H3K14 Acetylation Using BRD4, a BET Family Member. *ACS Chem. Biol.* **11,** 729–733 (2016).
391. Lin, C.-W., Jao, C. Y. & Ting, A. Y. Genetically Encoded Fluorescent Reporters of Histone Methylation in Living Cells. *J. Am. Chem. Soc.* **126,** 5982–5983 (2004).
392. Lin, C.-W. & Ting, A. Y. A Genetically Encoded Fluorescent Reporter of Histone Phosphorylation in Living Cells. *Angew. Chem. Int. Ed.* **43,** 2940–2943 (2004).
393. Kanno, T. *et al.* Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells. *Mol. Cell* **13,** 33–43 (2004).
394. Sekar, T. V., Foygel, K., Gelovani, J. G. & Paulmurugan, R. Genetically Encoded Molecular Biosensors To Image Histone Methylation in Living Animals. *Anal. Chem.* **87,** 892–899 (2015).
395. Sekar, T. V., Foygel, K., Devulapally, R. & Paulmurugan, R. Degron Protease Blockade Sensor to Image Epigenetic Histone Protein Methylation in Cells and Living Animals. *ACS Chem. Biol.* **10,** 165–174 (2015).
396. Vincenz, C. & Kerppola, T. K. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc. Natl. Acad. Sci.* **105,** 16572–16577 (2008).
397. Sanchez, O. F., Mendonca, A., Carneiro, A. D. & Yuan, C. Engineering Recombinant Protein Sensors for Quantifying Histone Acetylation. *ACS Sens.* **2,** 426–435 (2017).
398. Tekel, S. J., Vargas, D. A., Song, L., LaBaer, J. & Haynes, K. A. Tandem histone-binding domains enhance the activity of a synthetic chromatin effector. *bioRxiv* 145730 (2017). doi:10.1101/145730
399. Guidotti, N. Development and in vitro characterization of a genetically encoded split fluorescent probe targeting bivalent chromatin. (2014).
400. Kodama, Y. & Hu, C.-D. An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio. *BioTechniques* **49,** 793–805 (2010).

401. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
402. Kamps, J. J. A. G. *et al.* Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. *Nat. Commun.* **6,** ncomms9911 (2015).
403. Peña, P. V. *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442,** 100–103 (2006).
404. Ruthenburg, A. J. *et al.* Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell* **145,** 692–706 (2011).
405. Scott, M. S., Troshin, P. V. & Barton, G. J. NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC Bioinformatics* **12,** 317 (2011).
406. Dyer, P. N. *et al.* Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. *Methods Enzymol.* **375,** 23–44 (2003).
407. Hicks, J. M. & Hsu, V. L. The extended left-handed helix: A simple nucleic acid-binding motif. *Proteins Struct. Funct. Bioinforma.* **55,** 330–338 (2004).
408. Vlasov, P. K. *et al.* Left-handed helix of polyproline ii type in linker regions of DNA-binding proteins. *Biophysics* **53,** 663–664 (2008).
409. Zillner, K. *et al.* Microscale Thermophoresis as a Sensitive Method to Quantify Protein: Nucleic Acid Interactions in Solution. in *Functional Genomics* 241–252 (Springer, New York, NY, 2012). doi:10.1007/978-1-61779-424-7_18
410. Seidel, S. A. I. *et al.* Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **59,** 301–315 (2013).
411. Marks, P. A. & Breslow, R. Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* **25,** 84–90 (2007).
412. Liang, Y. *et al.* A Novel Selective LSD1/KDM1A Inhibitor Epigenetically Blocks Herpes Simplex Virus Lytic Replication and Reactivation from Latency. *mBio* **4,** e00558-12 (2013).
413. Konze, K. D. *et al.* An Orally Bioavailable Chemical Probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* **8,** 1324–1334 (2013).
414. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. *Mol. Cell. Biol.* **27,** 3769–3779 (2007).
415. Shi, Y. *et al.* Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119,** 941–953 (2004).
416. Forneris, F., Binda, C., Vanoni, M. A., Mattevi, A. & Battaglioli, E. Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. *FEBS Lett.* **579,** 2203–2207 (2005).
417. Wang, J. *et al.* Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **446,** 882–7 (2007).
418. Amente, S., Lania, L. & Majello, B. The histone LSD1 demethylase in stemness and cancer transcription programs. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1829,** 981–986 (2013).
419. Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437,** 436–439 (2005).
420. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. *Nature* **449,** 105–8 (2007).

421. Wang, J. *et al.* The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nat. Genet.* **41,** 125–129 (2009).
422. Cho, H.-S. *et al.* Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells. *Cancer Res.* canres.2446.2010 (2010). doi:10.1158/0008-5472.CAN-10-2446
423. Hakimi, M.-A. *et al.* A core–BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl. Acad. Sci.* **99,** 7420–7425 (2002).
424. Hakimi, M.-A., Dong, Y., Lane, W. S., Speicher, D. W. & Shiekhattar, R. A Candidate X-linked Mental Retardation Gene Is a Component of a New Family of Histone Deacetylase-containing Complexes. *J. Biol. Chem.* **278,** 7234–7239 (2003).
425. Humphrey, G. W. *et al.* Stable Histone Deacetylase Complexes Distinguished by the Presence of SANT Domain Proteins CoREST/kiaa0071 and Mta-L1. *J. Biol. Chem.* **276,** 6817–6824 (2001).
426. Shi, Y. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422,** 735–738 (2003).
427. Iwase, S. *et al.* Characterization of BHC80 in BRAF–HDAC complex, involved in neuron-specific gene repression. *Biochem. Biophys. Res. Commun.* **322,** 601–608 (2004).
428. Forneris, F., Binda, C., Vanoni, M. A., Battaglioli, E. & Mattevi, A. Human Histone Demethylase LSD1 Reads the Histone Code. *J. Biol. Chem.* **280,** 41360–41365 (2005).
429. Burg, J. M., Gonzalez, J. J., Maksimchuk, K. R. & McCafferty, D. G. Lysine-Specific Demethylase 1A (KDM1A/LSD1): Product Recognition and Kinetic Analysis of Full-Length Histones. *Biochemistry (Mosc.)* **55,** 1652–1662 (2016).
430. Stavropoulos, P., Blobel, G. & Hoelz, A. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol.* **13,** 626–632 (2006).
431. Chen, Y. *et al.* Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc. Natl. Acad. Sci.* **103,** 13956–13961 (2006).
432. Yang, M. *et al.* Structural Basis for CoREST-Dependent Demethylation of Nucleosomes by the Human LSD1 Histone Demethylase. *Mol. Cell* **23,** 377–387 (2006).
433. Chang, P.-Y. *et al.* Binding of the MLL PHD3 Finger to Histone H3K4me3 Is Required for MLL-Dependent Gene Transcription. *J. Mol. Biol.* **400,** 137–144 (2010).
434. Kim, J. *et al.* The n-SET Domain of Set1 Regulates H2B Ubiquitylation-Dependent H3K4 Methylation. *Mol. Cell* **49,** 1121–1133 (2013).
435. Yuan, W. *et al.* H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. *J. Biol. Chem.* **286,** 7983–7989 (2011).
436. Thompson, R. E. *et al.* Trifluoroethanethiol: An Additive for Efficient One-Pot Peptide Ligation−Desulfurization Chemistry. *J. Am. Chem. Soc.* **136,** 8161–8164 (2014).
437. Thåström, A. *et al.* Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences1. *J. Mol. Biol.* **288,** 213–229 (1999).
438. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
439. Nagaoka, M. *et al.* E-Cadherin-Coated Plates Maintain Pluripotent ES Cells without Colony Formation. *PLoS ONE* **1,** e15 (2006).
440. Pijnappel, W. W. M. P., Pijnappel, W. W. M. P., Bergsma, A. J. & Timmers, H. T. M. Chromatin Immunoprecipitation in mouse ES cells. *Protoc. Exch.* (2013). doi:10.1038/protex.2013.037

Chapter 6. CV

DELACHAT Aurore Marie-France

auroredelachat@gmail.com +41 78/634.25.24 Renens, Switzerland

EDUCATION

Diplomas

Aug. 2013 - Nov. 2017 **PhD** in Biochemistry
Ecole Polytechnique Fédérale de Lausanne (EPFL), LCBM
Under the supervision of Prof. Beat Fierz

Sept. 2011 - Mar. 2013 **Master of Science** in Molecular and Biological Chemistry
EPFL

Sept. 2008 - Jun. 2011 **Bachelor** of Chemistry and Chemical Engineering
EPFL

Projects

Sept. 2012 - Mar. 2013 **Master project** under the supervision of Prof. Hagan Bayley FRS (University of Oxford) and Dr. MER Ruud Hovius (EPFL)
Towards Reversible Self-Folding Droplet Networks

Jan. 2012 - Jun. 2012 **Semester project** under the supervision of Dr. MER Ruud Hovius (EPFL)
Solid-state nanopores: Analysis of colloid translocations, development of simultaneous electrical and optical recordings & formation of free-standing lipid bilayer

Feb. 2011 - Jun. 2011 **Project of Computational Chemistry** under the supervision of Prof. Anne-Clémence Corminboeuf (EPFL)
The Influence of Density Functional Approximations on the Energy Decomposition Analysis of Charge-Transfer Complexes

WORK EXPERIENCE

Nov. 2013 - Jan. 2017	**Various teaching** tasks as part of PhD studies Exercises and exams corrections, practical work in laboratories, supervision of students
2012	**Private tutor** in Mathematics for the company Aaacademia
Aug. 2011 - Sept. 2011	**Internship** in the laboratory of Prof. Jérôme Waser (LCSO, EPFL) *Pd(0/II)-catalyzed intramolecular oxy-alkynylation of olefins: preparation of relevant reagents and substrates*

PUBLICATIONS

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, **Manuscript accepted in Cell Chemical Biology**

Dhall A., Weller C. E., Shelton P., Chu A., **Delachat A. M.-F.**, Fierz B., Chatterjee C., Differential cross-regulation of LSD1 activity by histone sumoylation and acetylation, **Manuscript in preparation**

Steinmann S.N., Piemontesi C., **Delachat A.**, Corminboeuf C., Why are the Interaction Energies of Charge-Transfer Complexes Challenging for DFT?, *J. Chem. Theory Comput.,* **2012**, *8* (5), pp 1629-1640.
Available from: http://dx.doi.org/10.1021/ct200930x