Investigation of Proteins Responsible for the Establishment and Recognition of Prominent Lysine Modifications

Von der Fakultät 4: Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von
Raluca-Maria Tămaş
aus Cluj-Napoca, Rumänien

Hauptberichter: Prof. Dr. Albert Jeltsch
Mitberichter: Prof. Dr. Dieter H. Wolf
Tag der mündlichen Prüfung: 19.09.2014

Institut für Biochemie der Universität Stuttgart
2014
Eidesstattliche Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbst verfasst und dabei keine anderen als angegebenen Quellen und Hilfsmittel verwendet habe.

Stuttgart, den 17.06.2014

Raluca-Maria Tămaș
Contents

Acknowledgements ..................................................................................................................... VII

List of Publications ..................................................................................................................... IX

Publications resulting from work during the doctoral studies: ........................................ IX

Publications resulting from work previous to the doctoral studies: ................................ IX

Zusammenfassung ..................................................................................................................... XI

Abstract ..................................................................................................................................... XV

List of Abbreviations ................................................................................................................ XVII

1. Introduction ........................................................................................................................... 1

   1.1. Epigenetics ................................................................................................................... 1

   1.2. Organization and Structure of Chromatin ...................................................................... 2

   1.2.1. Epigenetic signals ................................................................................................... 2

   1.2.2. DNA methylation .................................................................................................... 3

   1.2.3. Noncoding RNA .................................................................................................... 4

   1.2.4. Histone post-translational modifications .................................................................. 4

   1.3. Lysine Acetylation ........................................................................................................ 9

   1.4. Lysine Methylation ....................................................................................................... 10

   1.4.1. Histone 3 lysine 4 methylation (H3K4me) ............................................................. 11

   1.4.2. Histone 3 lysine 36 methylation (H3K36me) ........................................................ 12

   1.4.3. Histone 3 lysine 9 methylation (H3K9me) ............................................................. 14

   1.4.4. Histone 3 lysine 27 methylation (H3K27me) ........................................................ 15

   1.5. Histone Lysine Methyltransferases ............................................................................. 17

   1.5.1. The methylation reaction ....................................................................................... 17

   1.5.2. The catalytic SET domain ..................................................................................... 17

   1.5.3. The SET domain HKMT families ......................................................................... 20
1.6. Investigation of PKMT non-histone targets ................................................................. 31

1.7. Reading Domains ........................................................................................................ 32

1.7.1. Methyl Lysine Readout......................................................................................... 33

1.7.2. Acetyl Lysine Readout ......................................................................................... 35

2. The Aim of the Presented Study .................................................................................... 39

2.1. Specific Goals and Achievements of Each Project ......................................................... 39

2.1.1. Yeast-3-Hybrid Protein-Protein Interaction Study for the Identification of Histone Lysine Methyltransferase Target Interactors ......................................................... 39

2.1.2. Identification of the Specific Targets of Histone Tail Reading Domains ............ 39

2.1.3. Determination of the Substrate Specificity of PKMTs and Search of Putative Non-Histone Targets ................................................................................................................. 40

2.1.4. Study of the influence of the SUV39H1 chromodomain on the catalytic activity of the enzyme ................................................................. 41

3. Materials and Methods .................................................................................................. 43

3.1. Yeast-Three-Hybrid Investigations .............................................................................. 43

3.1.1. Preparation of the Yeast-3-Hybrid Vector Constructs ......................................... 43

3.1.2. Transformation and Mating of Yeast ..................................................................... 46

3.1.3. Protein Extraction and Immunoblotting of Yeast Extracts ................................. 46

3.1.4. Small Scale His-tagged Protein Purification ......................................................... 47

3.1.5. Large scale native His-tag purification from yeast cells ........................................ 48

3.2. The PHF1 Tudor Domain Binds to Trimethylated Histone 3 Lysine 27, With a Strong Preference for the Histone H3T Variant ......................................................... 49

3.2.1. Fluorescence anisotropy measurements .............................................................. 49

3.2.2. Cell culture ........................................................................................................... 49

3.2.3. Native Histone pulldown assay ............................................................................ 49

3.3. Characterization of the Binding Specificity of the Chromodomains of CBX2 and CBX7 Proteins ................................................................................................................................. 51
3.3.1. Site directed mutagenesis, expression and purification ................................................. 51

3.3.2. Application of MODified™ Histone Peptide Arrays for the analysis of the binding specificity of the CBX2 and CBX7 chromodomains .......................................................... 52

3.3.3. Application of peptide SPOT arrays containing methyllysine analogs for the analysis of the binding specificity of the CBX2 and CBX7 chromodomains ........................................ 53

3.3.4. Cell culture .................................................................................................................. 53

3.3.5. Far western blot analysis of the binding specificity of the CBX7 wild type and mutated chromodomain to recombinant histones and native mononucleosomes........... 53

3.3.6. Fluorescence anisotropy measurements .................................................................... 54

3.3.7. Native mononucleosomal pulldowns with domains (CIDOP) or antibody (ChIP), coupled to quantitative PCR (qPCR) experiments ................................................................. 54

3.4. Investigation of the Binding Specificities of the MLL2 (KMT2D) and MLL3 (KMT2C) PHD Fingers .................................................................................................................. 56

3.4.1. Cloning, expression and purification ........................................................................... 56

3.4.2. Native histone pulldown with recombinant GST-tagged MLL2 PHD 3-5 and MLL3 PHD 4-6 58

3.4.3. Site directed mutagenesis .......................................................................................... 59

3.5. Study of the Specificity of Various SET Domain Methyltransferases ....................... 60

3.5.1. Cloning, expression and purification of the SET-domains ........................................ 60

3.5.2. In vitro methyltransferase assay with recombinant H3 substrate ............................. 61

3.5.3. Peptide SPOT array synthesis .................................................................................... 61

3.5.4. Methylation of peptide SPOT arrays for the determination of the specificity profile of SET domain methyltransferases ................................................................. 62


3.6.1. Cloning and site directed mutagenesis of the SUV39H1 constructs ...................... 63

3.6.2. Protein expression and purification ......................................................................... 64
3.6.3. Application of MODified™ Histone Peptide Arrays for the analysis of the binding specificity of the SUV39H1 chromodomain ................................................................. 65
3.6.4. Fluorescence anisotropy measurements for the characterization of the binding specificity of the SUV49H1 chromodomain ......................................................... 66
3.6.5. Far western blot analysis of the binding specificity of the SUV39H1 wild type chromodomain to native mononucleosomes ................................................................. 66
3.6.6. Native nucleosome pulldown with SUV39H1 chromodomain ...................... 66
3.6.7. In vitro methyltransferase assay with recombinant H3 substrate............... 67
3.6.8. In vitro methyltransferase assay with reconstituted chromatin............... 67
3.6.9. In vitro methyltransferase assay using peptide arrays containing two different peptides in one SPOT ........................................................................................................... 67
3.6.10. Cell culture .................................................................................................................. 68
3.6.11. SUV39H1 localization studies in SUV39H dn MEF cells ......................... 68

4. Results ................................................................................................................................ 71
4.1. Yeast-3-Hybrid Approach to Identify PKMT Target Binding Partners .............. 71
4.2. The PHF1 Tudor Domain Binds to Trimethylated Histone 3 Lysine 27, With a Strong Preference for the Histone H3T Variant ................................................................. 88
4.3. Characterization of the Binding Specificity of the Chromodomains of CBX2 and CBX7 Proteins ......................................................................................................................... 92
4.4. Investigation of the Binding Specificities of the MLL2 (KMT2D) and MLL3 (KMT2C) PHD Fingers ......................................................................................................................... 103
4.5. Study of the Specificity of Various SET Domain Methyltransferases ............... 120

5. Discussion ............................................................................................................................... 143
5.1. Yeast-3-Hybrid Studies for the Identification of PKMT Target Binding Partners in a Methylation-Dependent Approach ............................................................................. 143
5.2. The PHF1 Tudor Domain Binds Specifically to H3K27me3, and Shows Comparable Binding to H3K36me3 and H3K27me3 Modified Histones .......................................................... 149

5.3. The Chromodomains of CBX7 and CBX2 Bind Specifically to H3K27me3 and Serve as Markers of the Modification at Genome Level........................................................................... 151

5.4. The PHD Fingers of MLL2 and MLL3 Show Preference for H4 Peptides Acetylated at Lysine 16 in Association with Other Modifications........................................................................ 155

5.5. The SET Domains of MLL2 and SET1A Present Substrate Specificity Preferences That Will Be Used in the Identification of Their Non-Histone Targets........................................ 159

5.6. The SUV39H1 Chromodomain Binds Specifically to H3K9me3 and Seems to Inhibit the Catalysis of the SUV39H1 SET Domain under the Investigated Conditions.................. 163

Conclusion.................................................................................................................................................. 169

References..................................................................................................................................................... 171

Curriculum Vitae.......................................................................................................................................... 191


Acknowledgements

I would like to thank everybody who has supported me in completing this work. In this respect, I am first of all thankful to my supervisor, Prof. Dr. Albert Jeltsch, who has offered me support, excellent guidance and many great suggestions during my entire thesis work and also during the writing process. I am also grateful for being offered the opportunity to be a part of this working group, and for having the chance to work on such interesting project over the past few years. In addition, I would also like to thank Prof. Dr. Albert Jeltsch for introducing me to the field of epigenetics during my Bachelor Degree studies, and for awakening my interest in this subject, an interest that has only deepened with the passing of time.

I am also thankful to Prof Dr. Dieter Wolf, who has kindly agreed to be the second principal co-referee for my PhD Thesis.

I would also like to thank Prof. Dr. Bernhard Hauer, Prof. Dr. Stephan Nußberger, Prof. Dr. Peter Scheurich and Prof. Dr. Christina Wege for accepting to read and review my PhD Thesis.

I am also grateful to all our collaborators: Prof. Dr. Yoichi Shinkai for providing the SUV39H1 inactive mutant construct; Prof. Dr. Xiaodong Cheng for providing the SUV39H1 SET construct; Prof. Dr. Danesh Moazed for providing the Clr4 protein; Prof. Dr. Wolfgang Fischle for providing the reconstituted chromatin; and Prof. Dr. Thomas Jenuwein for providing the SUV39 dn MEF cells.

I am very thankful to Prof. Dr. Arunkumar Dhayalan for his supervision during my Master studies, which was essential for my success as a PhD student. I am also thankful for all his help with the initial preparations on the Yeast-3-hybrid and the SUV39H1 projects.

I am thankful to Dr. Srikanth Kudithipudi for synthesizing the peptide SPOT arrays, for providing the purified SUV39H1 SET protein and for all the valuable project-related discussions.

I would also like to thank Dr. Ina Kycia for giving me a chance to get involved in the PHF1 project, Goran Kungulovski for the involvement in the CIDOP-seq project, and Sara Weirich for her cooperation in the SET domain specificity project.

I am also grateful to all the present and previous lab members, but especially to Dr. Arunkumar Dhayalan, Dr. Ina Kycia, Dr. Renata Jurkowska, Dr. Srikanth Kudithipudi, Dr Tomek Jurkowski,
Goran Kungulovski and Cristiana Lungu for all the discussions and suggestions regarding my work.

I would like to thank Sandra, Regina and Dragica, for all the technical support, which made my work possible. I am also grateful to Elisabeth, Pavel and Mrs Geiger for all the help regarding the organization and formalities for the thesis submission.

I am also thankful to the Wilhelm Sander-Stiftung for the financial support.

In particular I am very grateful to Sandra, for being also a very pleasant company, especially during the time spent together in Bremen in the first half of 2012. Also during that time, I found moral support in Ute Claus and Elisa Brambilla, to which I am also grateful.

I am especially thankful to Ina, who was a very good friend both during the time spent in Bremen, and also once I joined the lab again in Stuttgart.

I would like to thank all my lab friends: Ina, Miruna, Rebekka, Sara, Cristiana, Renata, Goran, Max, Denis, Tomek and Razvan, for creating a very welcoming and warm environment both at the workplace and also outside the working hours.

Last but not least, I am very grateful to all my family members, for all their support, positive thoughts and encouragements. Most of all, I am very grateful to my boyfriend, Radu, who was a permanent source of support, patience and love.
List of Publications

Publications resulting from work during the doctoral studies:

Kungulovski G, Henry C, Ycia I, Reinhard R, Tamas R, Labhart P, Jurkowska RZ, Jeltsch A, (2014), Application of Histone Modification Specific Interaction Domains as an Alternative to Antibodies, Genome Research, manuscript accepted for publication. (I performed the site directed mutagenesis and purification of CBX7 mutants, and the specificity analysis of the CBX7 wild type and mutants by the Celluspots and peptide SPOT array binding assays, fluorescence depolarization binding assays and far western analysis. I also performed CBX7 CIDOP-qPCR and anti-H3K27me3 ChIP-qPCR experiments.)

Kycia I, Kudithipudi S, Tamas R, Kungulovski G, Dhayalan A, Jeltsch A, (2014), The Tudor domain of the PHD finger protein 1 is a dual reader of lysine trimethylation at lysine 36 of histone H3 and lysine 27 of histone variant H3t, Journal of Molecular Biology 426(8):1651-60. (I performed the fluorescence anisotropy binding assays for the wild type and mutant PHF1 protein with the H3TK27me3 and H3.1K27me3 peptides, and the native histones pulldown experiment.)

Publications resulting from work previous to the doctoral studies:


Zusammenfassung


Zusätzlich, wurde in diese Arbeit die Substratspezifität zweier Methyltransferasen untersucht, die eine SET Domäne enthalten und Histon 3 Lysin 4 (H3K4) methylieren. Beide Methyltransferasen binden unabhängig voneinander an denselben Koaktivator Komplex, welcher COMPASS genannt wird. In dieser Arbeit konnte gezeigt werden, dass SET1A, eine H3K4 trimethyl-Transferase, nur im COMPASS Komplex aktiv ist. Es konnte allerdings gezeigt werden, dass diese PKMT einige Sequenzen bevorzugt methyliert, welche nicht in Histon 3 vorkommen, was eine Suche nach neuen Nicht-Histon Substraten nahelegt.


Schließlich wurde die Bindungsspezifität der Chromodomäne der SUV39H1 Methyltransferase untersucht. SUV39H1 ist verantwortlich für die Histon 3 tri-Methylierung am Lysin 9 (H3K9me3), woraus eine Genrepression und Stilllegung von Heterochromatin resultiert. Es konnte gezeigt werden, dass die Chromodomäne von SUV39H1 spezifisch mit H3K9me3 interagiert und dass die Bindung der Chromodomäne zu seinem Targetpeptid zu einer Inhibiton der katalytischen Aktivität des Enyzms unter den verwendeten in vitro Bedingungen führt.
Abstract

Histone post-translational modifications influence chromatin architecture, either by direct effects on the interaction between histones and DNA, or indirectly, by serving as docking places for regulatory proteins, which bind through conserved functional domains termed “reading” domains. Different combinations of histone modifications define various chromatin states, each of which being associated with a particular set of regulatory enzymes. Lysine methylation is an important histone post-translational modification, which can occur at various positions in histones, with different roles in epigenetic regulation. This mark is generally established by SET domain Protein Lysine Methyltransferases (PKMTs). Recently, PKMTs have been reported to also methylate numerous non-histone substrates, which subsequently recruit so-called “reading” domains. These domains specifically interact with the methylated lysine in a sequence context-dependent manner. In this work, I tried to establish a Yeast-3-Hybrid method for the identification of methylation-dependent interactors of methylated non-histone proteins. For validation, I attempted to test the interaction between reported PKMT substrates fused to the Gal4-DNA-Binding Domain and methyl-“readers” fused to the Gal4-Activation Domain in yeast, either in the presence or absence of the corresponding PKMTs. Later in the project the known “reading” domains would be replaced by a library of human cDNA, in order to search for novel “readers” of protein lysine methylation marks.

Additionally, this work presents the study of the substrate specificities of two SET domain methyltransferases responsible for the methylation of histone 3 lysine 4 (H3K4), which are mutually exclusive members of the same coactivator complex, the human COMPASS. In this study, SET1A, an H3K4 trimethylase, was shown to be active only as part of the core COMPASS complex. This PKMT proved to have a higher preference for some sequences other than histone 3, justifying a search for novel non-histone substrates. MLL2, a member of the mixed lineage leukemia (MLL) family, responsible for H3K4 monomethylation, revealed stimulation of activity when part of the core COMPASS complex, and showed some differences in the substrate specificity when acting alone, compared to the complex. The search for non-histone protein substrates is in progress for SET1A/COMPASS, and also MLL2 alone and within the complex.

The targeting of most PKMTs is achieved with the help of histone modification “reading” or DNA-binding domains. The binding specificity of the PHD finger “reading” domains of MLL2,
and its paralog MLL3, was investigated during this doctoral study. Although most of the PHD fingers did not bind to histone tails, the MLL2 PHD 3-5 group of domains and the MLL3 PHD 4-6 group of domains bound specifically to modified histone tail peptides. Preference towards both histone 3 (H3) and histone 4 (H4) was identified and the strongest binding was seen on H4 peptides containing acetylation at lysine 16 together with multiple acetylations or methylations. This finding suggested recruitment to active chromatin, which is enriched in acetylation marks, but the specificity needs to be further confirmed and characterized in more detail.

I also investigated the histone binding specificity of PHF1, a member of the Polycomb Repressive Complex 2. This complex is responsible for developmental gene repression by the trimethylation of histone 3 lysine 27 (H3K27me3). The tudor domain of PHF1 showed preferred binding to its target, H3K27me3 in the sequence context of testis-identified histone variant H3T, in comparison to the canonical histone H3.1. The specificity for the same mark and histone variant was also identified for the chromodomain of the Polycomb Repressive Complex 1 member, CBX7, while the chromodomain of its paralog, CBX2, did not show discrimination between the histone variants, although it presented the same specificity towards the H3K27me3 mark. We propose that the discrimination between histone variants is a unique feature of some “reading” domains, and the role of this particular function needs to be elucidated. Moreover, the H3K27me3-specific CBX7 chromodomain was used as a tool in the validation of new methods developed by Kungulovski et al., 2014, with the purpose of replacing antibodies raised against specific histone modifications in adaptations of several antibody-based assays.

Finally, this PhD work also presents the binding specificity of the chromodomain of the SUV39H1 methyltransferase. SUV39H1 is responsible for histone 3 lysine 9 trimethylation (H3K9me3), and the consequent gene repression and silencing of heterochromatin. I showed that the chromodomain of SUV39H1 bound specifically to H3K9me3, and binding of the chromodomain to its target peptide seemed to inhibit the catalytic activity of the enzyme in our in vitro conditions.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>53 (pBridge-53)</td>
<td>p53 cloned as Gal4-DNA-BD in the pBridge vector</td>
</tr>
<tr>
<td>AbA</td>
<td>Aureobasidin A</td>
</tr>
<tr>
<td>Ac</td>
<td>Acetyl</td>
</tr>
<tr>
<td>Acinus</td>
<td>Apoptotic chromatin condensation inducer in the nucleus</td>
</tr>
<tr>
<td>AdoHcy (SAH)</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>AdoMet (SAM)</td>
<td>S-adenosyl-L-homocysteine</td>
</tr>
<tr>
<td>AF4</td>
<td>AF4/FMR2 family member 4</td>
</tr>
<tr>
<td>AKAP6</td>
<td>A-kinase anchor protein 6</td>
</tr>
<tr>
<td>ANK</td>
<td>Ankyrin repeats</td>
</tr>
<tr>
<td>ASC-2/NCOA6</td>
<td>Nuclear receptor coactivator 6</td>
</tr>
<tr>
<td>ASCOM</td>
<td>ASC-2 complex</td>
</tr>
<tr>
<td>Ash1</td>
<td>Achaete-scute homolog 1</td>
</tr>
<tr>
<td>ASH1L</td>
<td>Achaete-scute homolog 1-like</td>
</tr>
<tr>
<td>ASH2</td>
<td>Achaete-scute homolog 2</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>AWS</td>
<td>Associated with SET domain</td>
</tr>
<tr>
<td>BAH</td>
<td>Bromo-adjacent homology</td>
</tr>
<tr>
<td>BPTF</td>
<td>Bromodomain and PHD finger-containing transcription factor</td>
</tr>
<tr>
<td>CBP/p300</td>
<td>CREB-binding protein/p300 acetyltransferase</td>
</tr>
<tr>
<td>CBX</td>
<td>Chromobox</td>
</tr>
<tr>
<td>CD</td>
<td>chromodomain</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-dependent kinase</td>
</tr>
<tr>
<td>CDYL1</td>
<td>Chromodomain Y-like protein</td>
</tr>
<tr>
<td>CENP-A</td>
<td>Histone H3-like centromeric protein A</td>
</tr>
<tr>
<td>CENPC1</td>
<td>Centromere protein C</td>
</tr>
<tr>
<td>CHD3</td>
<td>Chromodomain-helicase-DNA-binding protein 3</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>Chromatin immunoprecipitation and sequencing</td>
</tr>
<tr>
<td>chromodomain</td>
<td>CHRomatin Organization MOdifier domain</td>
</tr>
<tr>
<td>Clr4</td>
<td>Cryptic loci regulator 4</td>
</tr>
</tbody>
</table>
CpG  Cytosine phosphate guanine site
CXXC1 (CFP1)  CXXC-type zinc finger protein 1
DAPI  4',6-diamidino-2-phenylindole
DF31  Decondensation factor 31, isoform A
DNA  Deoxyribonucleic acid
DNMT  DNA methyltransferase
DPY30  Protein dpy-30 homolog
DSB  Double stranded break
E(Z)  Enhancer of Zeste
E2F1  Transcription factor E2F1
Eaf3  ESA1-associated factor 3
ENL  YEATS domain-containing protein 1
ER  Estrogen receptor
ERV  Endogenous retroviral regions
ESC  Embryonic stem cell
EZH1/EZH2  Enhancer of zeste 1/2
FYRC  Phenylalanine-Tyrosine rich domain C-terminal
FYRN  Phenylalanine-Tyrosine rich domain N-terminal
G9a (EHMT2)  Euchromatic histone-lysine N-methyltransferase 2
Gal4-DNA-BD  Gal4 DNA binding domain
Gal4-AD  Gal4 activation domain
GLP  G9a-like protein
GNAT  Guanine nucleotide-binding protein G(t)
H1  Histone 1
H2A  Histone 2A
H2A.Bbd  Histone 2A variant
H2A.X  Histone 2A Barr body-deficient variant
H2A.Z  Histone 2A variant Z
H2AK119  Histone 2A lysine 119
H2B  Histone 2B
H2BK120  Histone 2B lysine 120
H3  Histone 3
H3.1  Histone 3 variant 1
H3K14  Histone 3 lysine 14
H3K27  Histone 3 lysine 27
H3K36  Histone 3 lysine 36
H3K4  Histone 3 lysine 4
H3K79  Histone 3 lysine 79
H3K9  Histone 3 lysine 9
H3R26  Histone 3 arginine 26
H3T  Histone 3 variant T
H4  Histone 4
H4K12  Histone 4 lysine 12
H4K16  Histone 4 lysine 16
H4K20  Histone 4 lysine 20
H4K5  Histone 4 lysine 5
H4K8  Histone 4 lysine 8
H4R19  Histone 3 arginine 19
HAT  Histone acetyltransferase
HCF-1  Host cell factor 1
HDAC  Histone deacetylase
HEK293  Human embryonic kidney 293 cells
HepG2  Human liver hepatocellular carcinoma G2 cells
hG9a  Human G9a
HKMT  Histone lysine methyltransferase
HMG  High mobility group
hMOF  human Males absent On the First
hMutSα  DNA mismatch repair protein Msh6
Hox  homeobox
HP1  Heterochromatin protein 1
ICF  Immunodeficiency, Centromere instability and Facial anomalies syndrome
IRF1  Interferon regulatory factor 1
ISWI  imitation SWI (*Drosophila melanogaster*)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{cat}$</td>
<td>Reaction rate</td>
</tr>
<tr>
<td>$K_d$</td>
<td>Binding constant</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Michaelis-Menten constant</td>
</tr>
<tr>
<td>Lam (pBridge-Lam)</td>
<td>Lamin cloned in pBridge vector</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>macroH2A</td>
<td>Histone 2A macro variant</td>
</tr>
<tr>
<td>MBT</td>
<td>Malignant brain tumor</td>
</tr>
<tr>
<td>me</td>
<td>Methyl</td>
</tr>
<tr>
<td>Me1</td>
<td>Monomethyl</td>
</tr>
<tr>
<td>Me2</td>
<td>Dimethyl</td>
</tr>
<tr>
<td>Me2a</td>
<td>Asymmetric dimethylation</td>
</tr>
<tr>
<td>Me2s</td>
<td>Symmetric dimethylation</td>
</tr>
<tr>
<td>Me3</td>
<td>Trimethyl</td>
</tr>
<tr>
<td>MeCP2</td>
<td>Methyl-CpG-binding protein 2</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse embryonic fibroblasts</td>
</tr>
<tr>
<td>mESC</td>
<td>Mouse embryonic stem cells</td>
</tr>
<tr>
<td>Metnase (SETMAR)</td>
<td>SET domain and mariner transposase fusion gene-containing protein</td>
</tr>
<tr>
<td>mG9a</td>
<td>Mouse G9a</td>
</tr>
<tr>
<td>MINT</td>
<td>Neuronal Munc18-1-interacting protein 1</td>
</tr>
<tr>
<td>MLA</td>
<td>Methyllysine analog</td>
</tr>
<tr>
<td>MLL</td>
<td>Mixed lineage leukemia</td>
</tr>
<tr>
<td>MLL1/KMT2A</td>
<td>Mixed lineage leukemia 1/ Lysine methyltransferase 2A</td>
</tr>
<tr>
<td>MLL2/KMT2D</td>
<td>Mixed lineage leukemia 2/ Lysine methyltransferase 2D</td>
</tr>
<tr>
<td>MLL3/KMT2C</td>
<td>Mixed lineage leukemia 3/ Lysine methyltransferase 2C</td>
</tr>
<tr>
<td>MLL4/KMT2B</td>
<td>Mixed lineage leukemia 4/ Lysine methyltransferase 2B</td>
</tr>
<tr>
<td>MLL5/KMT2E</td>
<td>Mixed lineage leukemia 5/ Lysine methyltransferase 2E</td>
</tr>
<tr>
<td>MMR</td>
<td>Mismatch repair</td>
</tr>
<tr>
<td>MORN</td>
<td>Membrane Occupation and Recognition Nexus</td>
</tr>
<tr>
<td>MPP8</td>
<td>M-phase phosphoprotein 8</td>
</tr>
<tr>
<td>MSL</td>
<td>Male specific lethal</td>
</tr>
<tr>
<td>MYST</td>
<td>MOZ, YBF2/SAS3, SAS2 and TIP60 protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>N/A</td>
<td>Not applicable</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end joining</td>
</tr>
<tr>
<td>NR</td>
<td>Nuclear receptor</td>
</tr>
<tr>
<td>NSD1</td>
<td>Nuclear SET domain-containing protein 1</td>
</tr>
<tr>
<td>NSD2</td>
<td>Nuclear SET domain-containing protein 2</td>
</tr>
<tr>
<td>NSD3</td>
<td>Nuclear SET domain-containing protein 3</td>
</tr>
<tr>
<td>NTD</td>
<td>N-terminal domain</td>
</tr>
<tr>
<td>NURD</td>
<td>Nucleosome Remodeling Deacetylase</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical density at 600 nm</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>p53</td>
<td>Cellular tumor antigen p53</td>
</tr>
<tr>
<td>PcG</td>
<td>Polycomb Group</td>
</tr>
<tr>
<td>PCL</td>
<td>Polycomb-like</td>
</tr>
<tr>
<td>PEV</td>
<td>Position effect variegation</td>
</tr>
<tr>
<td>PHD</td>
<td>Plant homeodomain</td>
</tr>
<tr>
<td>PHF1</td>
<td>PHD finger protein 1</td>
</tr>
<tr>
<td>PHF19</td>
<td>PHD finger protein 19</td>
</tr>
<tr>
<td>PIC</td>
<td>Pre-initiation complex</td>
</tr>
<tr>
<td>PKMT</td>
<td>Protein lysine methyltransferase</td>
</tr>
<tr>
<td>PPARBP</td>
<td>Peroxisome proliferator-activated receptor-binding protein</td>
</tr>
<tr>
<td>PRC1</td>
<td>Polycomb repressive complex 1</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb repressive complex 2</td>
</tr>
<tr>
<td>PRDM</td>
<td>Protein arginine demethylase</td>
</tr>
<tr>
<td>PRDM9</td>
<td>Protein arginine demethylase 9</td>
</tr>
<tr>
<td>Protein MCM10</td>
<td>Protein MCM10 MCM10 minichromosome maintenance homolog</td>
</tr>
<tr>
<td>Protein MCM10</td>
<td>MCM10 minichromosome maintenance homolog</td>
</tr>
<tr>
<td>Protein MCM10</td>
<td>deficient 10 homolog</td>
</tr>
<tr>
<td>PR-SET7/SET8</td>
<td>Histone-lysine N-methyltransferase SETD8</td>
</tr>
<tr>
<td>PSC</td>
<td>Posterior sex combs</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>PWWP</td>
<td>Pro-Trp-Trp-Pro motif-containing domain</td>
</tr>
<tr>
<td>RAR</td>
<td>Retinoic acid receptor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RBBP5</td>
<td>Retinoblastoma-binding protein 5</td>
</tr>
<tr>
<td>RE</td>
<td>Repeat element</td>
</tr>
<tr>
<td>RIZ</td>
<td>Retinoblastoma protein-interacting zinc finger protein</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNAP II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RPD3</td>
<td>Histone deacetylases Rpd3</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>Sac7d</td>
<td><em>Sulfolobus acidocaldarius</em> DNA-binding protein 7d</td>
</tr>
<tr>
<td>SANT</td>
<td>Swi3, Ada2, N-Cor, TFIIB domain</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SET</td>
<td>Su(var)3-9 (Suppressor of variegation 3-9), E(z), and Trx (Trithorax) domain</td>
</tr>
<tr>
<td>SET1A</td>
<td>SET domain-containing protein 1A</td>
</tr>
<tr>
<td>SET1B</td>
<td>SET domain-containing protein 1B</td>
</tr>
<tr>
<td>SET2L</td>
<td>SET domain-containing protein 2L</td>
</tr>
<tr>
<td>SET7/9</td>
<td>SET domain-containing protein 7/9</td>
</tr>
<tr>
<td>SETD2</td>
<td>SET domain-containing protein 2</td>
</tr>
<tr>
<td>SETDB1/2</td>
<td>SET domain bifurcated 1/2</td>
</tr>
<tr>
<td>SIR2</td>
<td>NAD-dependent histone deacetylases Silent information regulator 2</td>
</tr>
<tr>
<td>SIRT1</td>
<td>NAD-dependent histone deacetylases sirtuin-1</td>
</tr>
<tr>
<td>SMYD2</td>
<td>SET and MYND domain-containing protein 2</td>
</tr>
<tr>
<td>Sso7d</td>
<td><em>Sulfolobus solfataricus</em> DNA-binding protein 7d</td>
</tr>
<tr>
<td>Su(var)3-9</td>
<td>Protein suppressor of variegation 3-9</td>
</tr>
<tr>
<td>SUV39H1</td>
<td>Human Protein suppressor of variegation 3-9 homolog 1/2</td>
</tr>
<tr>
<td>SUV4-20H1/2</td>
<td>Human Protein suppressor of variegation 4-20 homolog 1/2</td>
</tr>
<tr>
<td>Swd3</td>
<td>Set1 complex component swd3</td>
</tr>
<tr>
<td>T (pGADT7-T)</td>
<td>SV40 large T antigen cloned in pGADT7</td>
</tr>
<tr>
<td>TAF10</td>
<td>Transcription initiation factor TFIID subunit 10</td>
</tr>
<tr>
<td>TAF3</td>
<td>Transcription initiation factor TFIID subunit 3</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid</td>
</tr>
<tr>
<td>TFIID</td>
<td>TATA-box-binding protein</td>
</tr>
<tr>
<td>Trr</td>
<td>Trithorax-related</td>
</tr>
<tr>
<td>Trx</td>
<td>Trithorax</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription termination site</td>
</tr>
<tr>
<td>TTK</td>
<td>Phosphotyrosine picked threonine-protein kinase</td>
</tr>
<tr>
<td>ub</td>
<td>ubiquitination</td>
</tr>
<tr>
<td>UHRF1</td>
<td>Ubiquitin-like PHD and RING finger domain-containing protein 1</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>WDR5</td>
<td>WD repeat-containing protein wdr-5</td>
</tr>
<tr>
<td>WDR82</td>
<td>WD repeat-containing protein wdr-82</td>
</tr>
<tr>
<td>WIZ</td>
<td>Widely-interspaced zinc finger-containing protein</td>
</tr>
<tr>
<td>ZDH8</td>
<td>Zinc finger DHHC domain containing protein 8</td>
</tr>
<tr>
<td>ZZZ3</td>
<td>the ZZ-type Zinc finger-containing protein 3</td>
</tr>
</tbody>
</table>
1. Introduction

1.1. Epigenetics

The term “epigenetics” was first used in 1942 by the British scientist Conrad Waddington to describe “the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being” (Waddington, 1942). The word contains the prefix “epi-”, which in Greek means “above”, suggesting that it refers to more than just traditional Mendelian genetics. Waddington set the basis for the study of heritability which ventures beyond the limits of genetics. In the 1950’s, Waddington perfected his theory by coining the term “epigenetic landscape” (Figure 1) to describe the way an embryonic stem cell differentiates into various cell types, like a marble rolling down a slope, with the surrounding environment influencing the choice of valleys that it can roll into (Waddington, 1957).

There has been much advancement in the field of epigenetics since the time of Waddington, and the definition of epigenetics has also acquired some significant changes. In 2009, the broader scientific community accepted a consensus definition of the term “epigenetics”. It describes an epigenetic trait as a “stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence”. This trait can be passed on both to daughter cells in the same organism (mitosis) and also to the next generations (meiosis) (Berger et al., 2009).

Epigenetics explains the process of development in multicellular organisms, in which all the cells contain the same genetic material (DNA), but acquire individual functional identities by committing to various differentiation states. Therefore, cells can have disparate, but stable
gene expression profiles and completely different functions, giving rise to all the tissue types and organs of a living organism (Goldberg et al., 2007).

Some representative epigenetic phenomena, which were previously biological conundrums from the point of view of genetics, include the position effect variegation (PEV) in *Drosophila*, in which the expression of a gene is governed by its local environment (Aagaard et al., 1999); paramutation in maize, a situation in which one allele causes heritable changes in another one (Patterson et al., 1997); or imprinting of maternal or paternal loci happening in mammals (Weaver & Bartolomei, 2014).

### 1.2. Organization and Structure of Chromatin

Inside the nuclei of eukaryotic cells, DNA is folded and condensed more than 10⁶-fold with the help of histone and non-histone proteins. The resulting entity has been termed chromatin (Margueron et al., 2005). The smallest structural unit of chromatin is the nucleosome, which comprises 147 base pairs of DNA, wrapped 1.7 times around a histone core (Kouzarides, 2007). The histones are organized in an octamer, consisting of a tetramer formed by two histone 3 (H3) and two histone 4 (H4) molecules and two histone 2A (H2A) and histone 2B (H2B) heterodimers. The histones, being highly basic proteins, tightly package the DNA, but their N-terminal tails are protruding out of the nucleosome and are accessible to other proteins (Jenuwein & Allis, 2001). Consecutive nucleosomes arrange themselves in a fiber with a diameter of roughly 11 nm, named “beads-on-a-string” fiber due to its appearance in electron micrographs (Margueron et al., 2005; Trojer & Reinberg, 2007). Only in the presence of histone 1 (H1) and at physiological salt conditions, these fibers have the inherent propensity to twist and fold into compacted filaments with a diameter of 30 nm (Jenuwein & Allis, 2001; Robinson & Rhodes, 2006). The organization of chromatin inside the nucleus is guided by so-called epigenetic signals, marks of epigenetic nature, which are responsible for defining the transcriptional state of the DNA around them.

#### 1.2.1. Epigenetic signals

The most important epigenetic signals are DNA methylation, histone post-translational modifications (PTMs) and noncoding RNAs. These factors can influence either directly or indirectly the overall chromatin structure and have been shown to have essential roles in processes such as gene transcription, DNA repair, DNA replication, chromosome organization,
X chromosome inactivation, imprinting, suppression of transposable element mobility, cellular differentiation and development (Kouzarides, 2007; Tollefsbol, 2011).

All the processes mentioned above are essential and therefore, problems occurring at the level of the epigenetic players may result in diseases. In cancer for example, the cancer progression or metastasis is governed by epigenome transformations at the tumor site (Feinberg, 2004). In addition to cancer, dysfunctional epigenetic processes are responsible for diseases such as the genomic imprinting disorders Angelman syndrome and Prader-Willi syndrome, the autoimmune diseases Type I Diabetes or Rheumatoid Arthritis, the immune disease called ICF syndrome (Immunodeficiency, Centromere instability and Facial anomalies syndrome) or neurological disorders, such as Alzheimer’s diseases, ATRX syndrome or multiple sclerosis (Jiang et al., 2004).

1.2.2. DNA methylation

DNA methylation is a heritable, stable and essential mark which facilitates epigenetic regulation (Goldberg et al., 2007). It is probably the chromatin modification that has been most thoroughly studied so far. In mammals, the methyl group is transferred to the 5-position of cytosines present in CpG dinucleotides, by specific enzymes termed DNA methyltransferases (DNMTs), and with the use of the universal methyl group donor S-adenosyl-L-methionine (SAM) (Jurkowska et al., 2011). DNA methylation occurs at 60-80% of the CpG sites in the human genome, and is associated with transcriptional repression when it occurs at promoter CpG islands. In mammals, the DNA methylation pattern is established and maintained by DNMTs, which can be classified according to their function. Two out of four members of the DNMT family, namely DNMT3A and DNMT3B are de novo methyltransferases, responsible for the establishment of methylation during germ cell development and at the preimplantation stage of embryogenesis (Denis et al., 2011; Jin & Robertson, 2013; Jurkowska et al., 2011). DNMT3L, in contrast to the other family members, is enzymatically inactive, but contributes to the activation of DNMT3A. The DNMT3A/3L dimer is required for the establishment of proper imprinting during gametogenesis. Both DNMT3A and DNMT3B are responsible for the methylation of different sets of repeat elements (REs), DNMT3B being the one silencing minor satellite repeats at pericentromeric chromatin (Jurkowska et al., 2011). DNMT1 is the maintenance methyltransferase, which uses hemi-methylated DNA as a substrate in order to preserve the methylation pattern during DNA replication. Through the
maintenance mechanism, the integrity of the methylation pattern is preserved in the somatic differentiated cells (Denis et al., 2011). Although there are clear differences between the functions of the maintenance and de novo methyltransferases, it has become evident that some functional overlap does occur (Jurkowska et al., 2011).

DNA methylation plays important roles in gene regulation, chromatin organization during embryogenesis and gametogenesis (Goll & Bestor, 2005), X chromosome inactivation, silencing of repetitive elements and of the centromeres, and mammalian imprinting (Yang & Kuroda, 2007). DNMTs are therefore crucial for genomic integrity, as they are responsible for stabilizing the genome, and in particular the repetitive sequences (Jin & Robertson, 2013).

1.2.3. Noncoding RNA

Noncoding RNAs are responsible for the control of multiple epigenetic phenomena (Bernstein & Allis, 2005). RNA involvement has been reported, for example, in dosage compensation in Drosophila and in mammals through the rox and XIST RNAs respectively. In addition, RNA has been proven to be involved in the silencing of genes and repetitive sequences by either posttranscriptional or transcriptional RNA interference (RNAi). The noncoding RNAs that work via transcriptional RNAi mechanisms are more epigenetic, in the sense that they cause long-term silencing that is inheritable through cell division (Goldberg et al., 2007).

1.2.4. Histone post-translational modifications

The amino termini or “tails” of the histone proteins are easily accessible for modifying enzymes, and are subject to various post translational modifications, such as acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines and threonines, or ubiquitination of lysines (Margueron et al., 2005). An extra degree of complexity is added by the fact that lysine methylation at the ε-amino group can occur at three different levels: mono-, di- and trimethylation, while arginine methylation at the guanidine-ε-amino group can be either mono- or dimethylation, which in turn is either symmetric or asymmetric (Bannister & Kouzarides, 2011; Margueron et al., 2005), (Figure 2).
The effects of the histone PTMs can be direct, as in the case of acetylation or phosphorylation, which can cause the disruption of contacts between the DNA and the histone core and result in an “opening” of chromatin and subsequent transcriptional activation. More indirect effects result from marks such as specific methylation, which serve as recruiters for different chromatin modifying enzymes (Goldberg et al., 2007). Histone PTMs can also function in the opposite way, as blockers of the binding of certain histone tail interactors (Bannister & Kouzarides, 2011; Fischle et al., 2005).

PTMs can occur at numerous residues of histones, offering the possibility of many different combinations of these modifications occurring on the same histone tail. These combinations could have specific functions and the sum of all modification patterns which encode for a particular chromatin state has been termed “histone code” (Jenuwein & Allis, 2001).

Histone modifications, in contrast to DNA methylation, can be much more dynamic and rapidly changing. For example, acetylations, methylations or phosphorylations can either appear or disappear in minutes after the arrival of the stimulus at the cell surface (Kouzarides, 2007). When it comes to comparing histone modification stabilities among themselves, methyl
groups, particularly methyllysines, based on thermodynamic principles, have a significantly lower turnover than phosphoryl and acetyl groups (Jenuwein & Allis, 2001).

Histone modifications can be divided in two categories according to their function: those that are correlated with transcriptional activation, and those that are correlated with transcriptional repression. The modification itself does not dictate the effect on transcription, but rather the combination of the modification and the position where it occurs encodes the full meaning of the mark. For example, trimethylation of lysines is associated with active genes if it occurs at histone 3 lysine 4 (H3K4), or histone 3 lysine 36 (H3K36), but is a silencing mark when it occurs at lysine 9 of histone 3 (H3K9), lysine 27 of histone 3 (H3K27), or lysine 20 of histone 4 (H4K20) (Kouzarides, 2007).

Lysine methylations are stable epigenetic marks introduced by histone lysine methyltransferases (HKMTs). Despite their stability, they can become actively removed by histone demethylases (Bannister & Kouzarides, 2011).

Lysine acetylation is one of the most investigated marks and is almost completely associated with active transcription. The neutralization of the amino group charge from the lysine side chain by the addition of the acetyl group results in the weakening of the interaction with DNA and the unfolding of chromatin (Bannister & Kouzarides, 2011).

Ubiquitination can be set, among others, at two critical positions: H2A lysine 119 (K119) and H2B lysine 120 (K120) in humans. The first modification is a repressive one, and is introduced by PRC1 (Polycomb Repressive Complex 1). The latter modification has been associated with transcriptional elongation. Ubiquitination is also reported to be involved in the response to UV-induced DNA damage (Wang et al., 2006).

1.2.4.1. Chromatin states

Gene expression is considered to be primarily dependent on the status of chromatin. Transcriptionally active chromatin, called euchromatin, was first defined as the part of the nucleus where the nucleic acid staining is less dense (Heitz, 1929). The more loose structure of euchromatin permits access to the DNA and therefore facilitates gene expression (Nelson & Monteggia, 2011).

Euchromatin is generally characterized by high levels of RNA transcription, high RNA polymerase occupancy, and the presence of histone acetylation and H3K4me2/3, especially at
promoter regions and transcription start site (TSS). There is also an anticorrelation of H3K9me2/3 and H3K27me3 marks with euchromatic loci. Chromatin regulators representative for active chromatin are histone deacetylases (HDACs), such as RPD3 and SIR2, remodelers such as DF31, a protein which causes chromatin decondensation, or transcriptional coactivator complexes, such as the Set1 complex (Filion et al., 2010).

Euchromatin has been divided according to epigenetic mark and regulator distributions in two different types, which can in turn be divided in multiple subdomains (Ernst et al., 2011; Filion et al., 2010). One type, which contains genes with a broad expression pattern over many embryonic tissues and stages in development, is less occupied by proteins and is characterized by the presence of the activating mark H3K36me3 in the bodies of the actively transcribed genes. H3K36me3 has been shown to confer stability to nucleosomes by suppression of nucleosome turnover, and therefore prevents cryptic transcription inside gene bodies (Zentner & Henikoff, 2013). The other type of euchromatin is characterized by the presence of many regulatory proteins, most probably because it contains tissue-specific genes. This type of chromatin is also the first one to be replicated during S-phase. The high protein occupancy of these euchromatic domains suggests that they serve as regulatory hubs, generally required for tissue specific expression pathways (Filion et al., 2010).

On the other hand, highly condensed chromatin, or heterochromatin, remains transcriptionally inert and is visible in the interphase nucleus as heterochromatic foci, or as the inactive X chromosome, or “Barr body”, in the female mammalian cells (Jenuwein & Allis, 2001). Heterochromatin can further be divided in a few different categories according to their function and the epigenetic marks and regulators contributing to their formation.

Regions of the genome such as satellite repeats or transposable elements, found at telomeres, centromeres and knobs, are silenced throughout the cell cycle and are termed “constitutive” heterochromatin. This type of deeply silenced chromatin is supported by promoter DNA methylation, and also by the repressive histone modifications H3K9me3 and H4K20me3. Classic pericentric heterochromatin markers are SUV39H1 and HP1 proteins (Grewal & Jia, 2007).

The second type of heterochromatin is the “facultative” one, which is found at developmentally regulated loci, and which can be activated in response to cellular signals (Grewal & Jia, 2007). These heterochromatic domains are characterized by H3K27
methylations, and correspond to Polycomb Group (PcG) regulated chromatin (Filion et al., 2010).

A third type of heterochromatin, which comprises approximately 50% of the entire heterochromatin, was more recently identified as the late-replicating peripheral chromatin associated with the nuclear lamina (Filion et al., 2010). This general classification of heterochromatin can be further extended to multiple subdivisions based on finer differences in state (Ernst et al., 2011).

Chromatin dynamics can be affected both by direct interference to the structure by histone marks, or by recruitment of chromatin remodelers, which change the accessibility of chromatin in an ATP-dependent manner, perhaps by removing or sliding nucleosomes (Goldberg et al., 2007).

**1.2.4.2. Histone variants**

In addition, different histone variants can be incorporated into nucleosomes, in a replication-dependent, replication- and cell cycle-independent, or tissue-specific manner. The so-called “canonical histones” are replication-dependent and are encoded in humans in three large gene clusters, containing 55, 6 and 3 genes respectively, with a total of 10-20 genes encoding each canonical histone (Marzluff et al., 2002). They are highly expressed during the S phase. All the other types are transcribed from low-copy or single genes and are incorporated outside of replication (Chen et al., 2013). So far, no variants have been identified for histone H4 and H2B, but both H2A and H3 show multiple variants. The H2A family is the broadest, and contains, in addition to the canonical H2A, also variants such as H2A.Z, macroH2A, H2A.Bbd and H2A.X. All these proteins address various functions, ranging from double strand break repair (H2A.Z) to X chromosome inactivation and repression (macroH2A). H3 has several very similar variants, such as H3.1, H3.2, and H3.3, with only few amino acid exchanges. In contrast, the H3 variant CENP-A, which is present at the centromeres, shows high sequence difference (Chen et al., 2013).

H3 is also present as tissue specific variants, like histone H3T. This is a histone variant with very high expression level in the testis, which is incorporated into the nucleosomes during early spermatogenesis. Although the H3T histones are also produced in somatic cells, they are thought to represent a small proportion of the bulk chromatin. So far, H3T have not been
reported to have any function in somatic cells (Tachiwana et al., 2010). H3T differs from the canonical H3.1 in only four residues: Val24 in H3T is replaced by Ala in H3.1; Met71 in H3T is Val71 in H3.1; Ser98 in H3T is replaced by Ala in H3.1; and Val111 is Ala111 in H3.1. The most outstanding property of nucleosomes containing H3T is instability, with a rapid exchange rate compared to H3.1 histones. The instability of the nucleosomes is thought to be caused by Val111, and partially by Met71, by weakening the interaction between the H3T/H4 tetramer and the H2A/H2B dimers (Tachiwana et al., 2010).

1.3. Lysine Acetylation

Lysine acetylation, the first histone modification identified, seems to have a rather cumulative effect of charge neutralization caused by multiple acetylations on the histone tail, and is the only modification which is not so specific with respect to the residue position where it occurs. Individual lysine mutations do not influence the overall effect of lysine acetylation on chromatin structure (Zentner & Henikoff, 2013). Most probably for the same reason, the enzymes responsible for the acetyl transfer, histone acetyl transferases (HATs), show a low substrate specificity. The acetylation mark is removed by a class of enzymes termed histone deacetylases (HDACs). Both types of enzymes may also act on non-histone targets. They are normally associated with actively transcribed loci, where both acetylation and deacetylation events need to occur rapidly during transcription, to facilitate the formation of the Polymerase Initiation Complex (PIC), and the nucleosome eviction in the path of the elongating polymerase, followed by a rapid reassembly of chromatin after a round of transcription (Zentner & Henikoff, 2013). As mentioned, histone acetylation can either destabilize the nucleosome structure and reduce nucleosome occupancy at the TSS, or in some cases act as recruiter of chromatin remodeling factors to active promoter regions (Horikoshi et al., 2013).

Nuclear protein acetyltransferases are divided into three major families: MYST, GNAT and CBP/p300. Most of their known substrates are lysines in the histone tails, but acetylation has also been identified in the histone core domain, at the K56 position for example (Kouzarides, 2007).

H4K16 acetylation is one of the most investigated acetylation marks, because it has a strong effect in the creation of open chromatin, and also acts as a switch for the transition from silenced to transcriptionally active chromatin, both in yeast and in humans (Horikoshi et al., 2013). The primary HAT responsible for the H4K16 acetylation is hMOF (human Males absent
On the first, a member of the MYST HAT family. This protein was first identified in Drosophila as a part of the male specific lethal (MSL) complex, responsible for X chromosome dosage compensation. The role in dosage compensation for the H4K16ac mark is not conserved in mammals. However, MOF was subsequently identified to occupy active promoters genome-wide, as a member of the NSL (non-specific lethal) complex (Horikoshi et al., 2013).

The MOF acetyltransferase was shown to be essential for mammalian development. Absence of Mof in mice results in peri-implantation embryonic lethality, while the outcome of MOF conditional deletion in human Purkinje cells is blebbing and cell death (Horikoshi et al., 2013). Evidence suggests that the underlying reasons are the decrease in the DNA damage response, and the loss of genome stability. In Drosophila, H4K16ac is responsible for the inhibition of the chromatin remodeler ISWI, and this might explain why this mark seems to have an independent role during transcription in comparison to the other H4 lysine acetylations (K5, K8, or K12). Therefore, the loss of the mark may result in more serious consequences on the structure of chromatin (Zentner & Henikoff, 2013). Shogren-Knaak et al. reported that H4K16ac inhibits the formation of the 30-nm fiber, by disturbing the interaction between unmodified H4K16 and an acidic patch on H2A. The result of the acetylation is chromatin decondensation (Shogren-Knaak et al., 2006). This finding was confirmed by the fact that the H4K16 deacetylase SirT2 induces chromatin condensation (Kouzarides, 2007).

At genome level in HEK cells, H4K16ac is enriched in regions close to the chromosome ends, and depleted at centromeric regions. In particular, only 18% of the H4K16ac peaks were annotated within gene-poor regions, while the other 82% were within genes. Out of the latter, roughly half were in introns, 10% were at TSS and another 10% were at the transcription termination sites (TTS), while only a negligible number of binding sites was registered in exons. H4K16ac sites which corresponded to transcription factor binding sites were reported to be found at the promoters of genes that are involved in proliferation, cell cycle and apoptosis (Horikoshi et al., 2013).

1.4. Lysine Methylation

Lysine methylation is probably the most investigated histone modification, since its discovery in 1964 by K. Murray (Murray, 1964). The lysine residue can exist either unmodified or in three sequential modification states: mono-, di- and trimethylation (Figure 3). The mark can occur on at least 12 positions in the histone tails (Figure 2), and therefore, it has the potential to
offer a large variety of combinations that might implicate different outcomes on the chromatin structure and function.

Figure 3 - Possible Lysine Methylation States. Lysine residues in histones can be found in four different states: unmodified, mono-, di-, and tri-methylated. Protein lysine methyltransferases (PKMTs) are responsible for adding the methyl groups, while Lysine demethylases (KDMs) are the enzymes that remove the mark. The image was adapted from AtdBio, 2014.

Histone lysine methylations have a central role in the organization of chromatin, although it is not clear yet whether they are a direct cause of the chromatin states, or whether they occur after DNA methylation in order to confer more variability and complexity to shape the chromatin state. The addition of the methyl groups most probably always has indirect effects on the chromatin architecture. The methyl group is small and it does not neutralize the charge of the lysine residue, and therefore does not directly result in condensation or decondensation of chromatin. Rather, it serves as a docking site for regulatory proteins (Bannister & Kouzarides, 2005). For example, the enzyme SUV39H1 is a so called “writer” of histone 3 lysine 9 trimethylation (H3K9me3) marks, while the structural protein HP1 represents the “reader” of this mark by binding specifically to it and causing downstream effects that lead to heterochromatin formation and stabilization (Goldberg et al., 2007).

1.4.1. Histone 3 lysine 4 methylation (H3K4me)

One well investigated histone lysine methylation is H3K4me3, found at the promoters of activated genes, in a conserved manner from yeast to higher eukaryotes (Shilatifard, 2008). In Drosophila, H3K4me3 is set by the Trithorax complex, which is thought to work as an antagonist of the Polycomb complex, responsible for the H3K27me3 establishment (Zentner & Henikoff, 2013). The H3K4 methylation is achieved by the COMPASS (Complex Proteins Associated with Set1) methyltransferase complex in S. cerevisiae, and by COMPASS-like homologous complexes in mammals (Krogan et al., 2002). Although there is only one H3K4 methylating enzyme in yeast, Set1, there are more than six Set1-related proteins in mammals,
such as MLL1 to 4 or SET1A and SET1B. However, only SET1A/B and MLL1 (also known as KMT2A) and MLL4 (also known as KMT2B) introduce trimethylation of the lysine, while the MLL2 (also known as KMT2D) and MLL3 (also known as KMT2C) enzymes are H3K4 monomethylases (Cheng et al., 2014). The yeast Set1 is only enzymatically active within the complex and is able to introduce up to three methyl groups per substrate lysine (Shilatifard, 2008). Also conserved from yeast to human is the dependency of H3K4me3 on the presence of H2B monoubiquitination at lysine 123 in yeast and lysine 120 in humans (Nguyen & Zhang, 2011).

The most prominent function of the H3K4me3 mark is the activation of transcription. This is achieved through the capability to recruit the basal transcription complex TFIIID, which interacts with H3K4me3 through the plant homeodomain (PHD) finger of its component, TAF3. This explains the association with the RNA Pol II that is phosphorylated at the serine 5 of its C-terminal domain, which is the polymerase form present at the TSS (Pérez-Lluch et al., 2011).

H3K4me3 is also implicated in splicing, through the recruitment of CHD1 and its interactors, which are part of the spliceosome. The mark has an important function in V(D)J recombination in lymphocytes, as it recruits RAG2 (Recombination Activating Gene 2), a protein which has a function in antigen-receptor gene assembly (Shilatifard, 2008).

The monomethylation of H3K4 has different functions than its trimethylation. H3K4me1 is a marker of active enhancers (Hu et al., 2013) and is also found at the promoters of conditionally repressed inducible genes, such as the genes responsible for muscle contraction and development, stress, immune and inflammatory response (Cheng et al., 2014).

1.4.2. Histone 3 lysine 36 methylation (H3K36me)

The H3K36me3 mark is found within gene bodies, and reports show that the localization of the mark is tightly dependent on the activity of the spliceosome. The beginning of the H3K36me3 domain on the gene bodies correlates with the 3’end of the first intron, and keeps increasing until the 3’end of the gene (Kim et al., 2011).

This mark is associated with the elongating RNA Pol II, in its serine 2 phosphorylated form. In budding yeast, H3K36me3 is able to prevent cryptic transcription by recruiting Eaf3, which interacts with the Rpd3S deacetylase complex. In this setup, the histones become deacetylated after the RNA Pol II passes through them and returns chromatin in a stable
euchromatic state, preventing any transcription initiation events inside the body of the genes (Joshi & Struhl, 2005). In *S. cerevisiae*, the methyltransferase responsible for the methylation of H3K36 is Set2, which interacts with the C-terminal domain of the elongating RNA Pol II (Lee & Shilatifard, 2007). In the absence of Set2, unmodified H3K36 and histones acetylated by Rpd3S accumulate on open reading frames (ORFs), causing spurious transcription initiation inside the gene bodies. It has been shown that, in yeast, co-transcriptional acetylation is partly accomplished through histone exchange in coding regions, and the role of H3K36me3 is to prevent the interaction between the H3 proteins and their chaperones, and consequently preclude the incorporation of new acetylated histones inside the ORFs (Venkatesh et al., 2012).

In higher eukaryotes, there are several H3K36 methyltransferases, including the Set2 homolog SETD2, ASH1L, and also NSD1, NSD2, HYPB (Yuan et al., 2013), SMYD2 (Zhang et al., 2013), and PRDM9 (Eram et al., 2014).

An important role of H3K36 methyltransferases is to antagonize the silencing marks set by the Polycomb group complexes. It has been shown that H3K27me3, set by the Polycomb Repressive Complex 2 (PRC2) does not affect the methylation of H3K36, although these two marks are mutually exclusive (Yuan et al., 2011). However, it seems that monoubiquitination of histone H2A by the other Polycomb Repressive Complex, PRC1, inhibits the enzymatic activities of multiple H3K36 methyltransferases, resulting in separation of the domains occupied by H3K27me3/H2AK119ub and H3K36me3. Reciprocally, H3K36me2/3 inhibits the activity of PRC2 (Yuan et al., 2011).

The trimethylation of H3K36 is also reported to be important in DNA mismatch repair (MMR), by directly recruiting the protein hMutSα, a mismatch recognition protein. During G1 and early S phase of the cell cycle, there is abundant H3K36me3 present on chromatin, in order to ensure the presence of hMutSα once mismatches start to appear on the newly replicated DNA. In the absence of the H3K4 methyltransferase SETD2, cells present a phenotype characteristic for MMR deficiency, with microsatellite instability and high mutation frequency (F. Li et al., 2013).

Recently, it has been shown that H3K36me2 accumulates at the site of double stranded breaks (DSBs) in the DNA, and is essential for DNA repair via non-homologous end joining (NHEJ). The methyltransferase responsible for the dimethylation of H3K36 at these sites is Metnase (also
called SETMAR), a DNA DSB repair component. The H3K36me2 mark is responsible for the recruitment and stabilization of other components of the DNA repair complexes, and the degree of H3K36me2 correlates to the efficiency of the repair (Fnu et al., 2011).

1.4.3. Histone 3 lysine 9 methylation (H3K9me)

Methylation of the lysine 9 residue in histone H3 is mostly associated with transcriptional repression. However, the three methylation states of H3K9 contribute to the complexity of the “histone code”, as they have very distinct localizations, different functions and are also set by several different enzymes. From a global perspective, the mono-, di-, and trimethyl H3K9 are part of separate chromosomal domains, which can be easily distinguished either by their nuclear localization patterns or by the replication timings of the chromatin they occupy (Wu et al., 2005). The formation of the different chromatin organization domains has been proven to be critical for phenomena such as imprinting, chromosome condensation and segregation, and gene dosage compensation (Wu et al., 2005).

Mono- and dimethylated H3K9 shows a dispersed and punctate staining in the murine nucleus, and is excluded from the nucleoli and the dense chromatin areas, identified as DAPI-dense heterochromatic foci. The trimethylated form, on the other hand, is enriched within pericentric heterochromatin, overlapping perfectly with the DAPI-dense regions. These localization patterns reveal that H3K9me3 is a marker of heterochromatin, while H3K9me1/2 are found in more euchromatic areas (Rice et al., 2003). The mutually exclusive localization of the marks suggests the distinct functions they fulfill. At a closer look, mono- and dimethylated H3K9 also show different nuclear localizations. The first one is found in the interior of the nucleus in many punctate foci that correspond exclusively to DNA synthesis sites during early S phase. The modification is excluded from the nuclear periphery, the nucleoli and the associated chromocenters. Dimethylated H3K9 localizes more towards the periphery of the nucleus and is lined around the nucleoli, but also forms nuclear foci that are somewhat larger than the H3K9me1 foci. The dimethylated mark is part of chromatin which is replicated throughout the entire S phase, representing an intermediate form between mono- and trimethylated domains. H3K9me3 enriched chromatin replicates late, and is almost exclusively located within pericentric heterochromatin (Wu et al., 2005).

The methyltransferases responsible for the methylation of H3K9 are also distinct both in accordance with the different methylation states produced, and with the methylation state
preferred for the substrate lysine. So far in mammals, at least five methyltransferases are known to be active at the H3K9 position: SUV39H1/2, G9a/GLP, and SETDB1 (Wu et al., 2005). In mice, it was shown that the first two, Suv39h1/2, are responsible for 75% of cellular H3K9me3, while G9a is responsible for 50-60% of the cellular pool of H3K9me1/2 (Peters et al., 2003), and particularly all H3K9me2, and partially H3K9me1 (Rice et al., 2003). Rice and colleagues (2003) also proved that Suv39h1 and Suv39h2 are both sufficient and also required for H3K9 trimethylation at constitutive heterochromatin.

More recently, it was shown through ChIP-Seq (Chromatin Immunoprecipitation Sequencing) with H3K9 methylation specific antibodies that H3K9me2 and H3K9me3 signals in stretches that span 10 kilobases (kb) around the TSS of genes are higher in silent regions compared to transcriptionally active ones. However, H3K9me1 seems to be a marker for active chromatin, being situated in higher levels at the promoters of active genes (Barski et al., 2007).

1.4.4. Histone 3 lysine 27 methylation (H3K27me)

Facultative heterochromatin is a silenced state which is more readily able to be switched on, in comparison to the constitutive heterochromatin. Here, repression is characterized by the presence of the H3K27me3 mark, which is set by the PcG complex PRC2. Facultative heterochromatin is representative for the silencing of Hox gene clusters in Embryonic Stem Cells (ESCs), imprinting and the inactivation of the X chromosome (Völkel & Angrand, 2007).

PRC2 is composed of several proteins, including an H3K27 methyltransferase, represented by E(Z) (Enhancer of Zeste) in *Drosophila* or its mammalian homologs EZH1 and EZH2. It has been proven that PRC1, which contains an E3 ligase component responsible for the ubiquitination of H2AK119, is the primary cause of the repression activity. The exact process of this transcriptional silencing effect mediated by the H2A ubiquitination remains elusive (Völkel & Angrand, 2007). In contrast, ubiquitination of the H2B histone serves as a stimulator of H3K4me3 appearance, while the deubiquitination of this residue promotes H3K27me3-dependent silencing (Schuettengruber et al., 2007).

H3K27me3 shows localization to some restricted genomic areas, but the most prominent distribution is over large chromosomal domains, which can sometimes span over hundreds of kilobases. The broader of these domains represent the Hox genes. It is possible that one of the roles of the mark is to ensure epigenetic inheritance during cell division, when all the PcG
proteins are stripped of the replicating DNA, by promoting a rapid recruitment of the repressive complexes to their targets (Schuettengruber et al., 2007). It has been reported that genes silenced by PcG organize in so called “PcG bodies”, subnuclear silencing compartments that appear as small nuclear speckles devoid of RNA Pol II, which gradually become reduced as development progresses (Schuettengruber et al., 2007).

The homeotic genes that are silenced with the help of H3K27me3 are mostly involved in development, organogenesis and morphogenesis. They are clustered together and, during the process of development, the genes in the clusters become activated in the order of their physical location within the cluster. This type of consecutive activation of genes was termed temporal collinearity. During the mouse embryonic development, it was observed that the temporal collinearity of gene activation is concomitant with the disappearance of H3K27me3 marks. Towards the end of embryonic development, the genes that stay silenced show even higher levels of H3K27me3 to ensure strong silencing (Soshnikova & Duboule, 2009).

H3K27me3 also decorates the repressed alleles of imprinted genes, together with H3K9me2, and the loss of the mouse PRC2 member Eed results in loss of genomic imprinting (Völkel & Angrand, 2007). Another important function of the H3K27me3 mark is in the silencing of the female X-chromosome. The X inactivation process occurs in mouse embryos in two different situations: the imprinted inactivation of the paternal X chromosome in early development, followed by the random inactivation of one X chromosome later in development. The noncoding RNA (ncRNA) Xist is expressed from the X chromosome about to be silenced, and coats it. At this point, H3K4 is hypomethylated, and the H3K27me3 and H2AK119ub1 are enriched. At later stages of the silencing process, H3K27me3 levels get reduced, while macroH2A becomes incorporated in the inactive X chromosome (Xi), and DNA methylation of the promoters of Xi genes is established (Völkel & Angrand, 2007).

When it comes to the PcG repression, regions containing H3K27me3 reflect a relatively stable repressed state. However, when H3K4me3 is also present in the same region, “bivalency” is established. The two opposing marks keep the genomic loci in a “poised” state, ready for transcriptional activation or repression in response to environmental stimuli (Blomen & Boonstra, 2011; Ernst et al., 2011). Only a small fraction of the bivalent genes keep their state during differentiation, while the rest will rapidly turn into monovalent regions, preserving
either one or the other of the modifications (either H3K27me3, or H3K4me3) (Blomen & Boonstra, 2011).

1.5. Histone Lysine Methyltransferases

All the lysine methylation marks presented earlier, such as the methylation of H3K4, H3K36, H3K9 or H3K27, together with many other histone and non-histone methylation marks, are set by enzymes called methyltransferases. Because most of these enzymes are not exclusively methylating histones and because they are specific for lysine substrates, they have been termed protein lysine methyltransferases (PKMTs).

1.5.1. The methylation reaction

PKMTs catalyze the methylation of lysine residues using S-adenosyl-L-methionine (AdoMet or SAM) as a cofactor. The lone electron pair of the lysine side chain nitrogen atom performs a nucleophilic attack on the methyl group bound to the sulphonium cation of SAM, resulting in the transfer of the methyl group to the substrate lysine and the production of S-adenosyl-L-homocysteine (AdoHcy or SAH) from the cofactor (Figure 4)(Copeland et al., 2009).

![Figure 4 - Graphic representation of the lysine methylation reaction by PKMTs. The ε-amide acts as a nucleophile, as its lone electron pair attacks SAM, and leads to the formation of a covalent link to the released methyl group. The drawing was adapted from Copeland et al., (2009).](image)

1.5.2. The catalytic SET domain

All protein lysine methyltransferases, except for the seven-β-strand DOT1 family of H3K79 methyltransferases, contain a conserved domain, termed SET, which is responsible for their catalytic activity. The name SET comes from the initials of the first three proteins, from
*Drosophila*, in which the SET domain was identified: Su(var)3-9 (Suppressor of variegation 3-9), E(z), and Trx (Trithorax) (Dillon et al., 2005). The SET domain was first described in 1998 (Jenuwein et al., 1998) as a 130 amino acid conserved motif which folds in a unique way when compared to all other SAM-dependent methyltransferases. In most cases, the SET domain is packed together with an N-terminal pre-SET (or N-SET), and a C-terminal post-SET (or C-SET) (Figure 5). The sequence of the two SET adjacent domains is not conserved within the SET domain family of enzymes (Qian & Zhou, 2006).

The pre-SET domain (yellow, Figure 5A) has a role in stabilizing the structure of the core SET domain through direct interaction with it. In the pre-SET of the SUV39 family of proteins, there are 9 conserved cysteines separated by variable numbers of residues and coordinating three zinc ions (Dillon et al., 2005).

![Figure 5](image)

**Figure 5 - Representation of the structure of two SET domain examples in ribbon form.** A. The SET domain of the *Neurospora crassa* DIM-5, a typical SET1 family methyltransferase. B. The SET domain of the orphan methyltransferase SET7/9. Structural components are color coded. The core SET domain is green, and the pseudo-knot is highlighted in purple. The post-SET, and non-canonical C-SET respectively, are grey. The pre-SET and corresponding N-SET are yellow. The H3 peptide is shown in red, the cofactor AdoHcy is blue, and the zinc ions are colored spheres. All subunits are labeled and indicated by arrows. The image was taken from Dillon et al., 2005.

The SET domain (green, Figure 5) probably evolved by the duplication of a three-stranded β-sheet structure, which is folded into three different sheets gathered around a knot-like structure, in which the C-terminus of the domain is passed through a short loop formed by a preceding fragment from its sequence (purple, Figure 5). Through this “pseudo-knot” formation, two motifs, which are the most conserved in the SET domain, are brought in proximity to each other, giving rise to the active site, at the meeting point between the cofactor and the substrate binding pockets (Taylor et al., 2003). The biggest difference
between SET domain methyltransferases and all other SAM-dependent methyltransferases is the arrangement of the binding pockets for the substrate and the cofactor. These two binding sites are found on opposite faces of the domain, but they are connected through a hydrophobic channel, which passes through the entire core of the SET domain. In this setup, the substrate lysine side chain passes through the hydrophobic channel, meeting the SAM methyl group in the active site found at the core of the domain. Because of this catalytic site organization, if the space inside the active site permits it, the enzyme is able to be processive, doing several rounds of methylation without substrate dissociation, simply by replacing the SAH with SAM at the cofactor binding site (Trievel et al., 2002).

The post-SET domain (grey, Figure 5) folds onto the main SET domain surface and completes its active site, constructing a hydrophobic channel (Qian & Zhou, 2006). Most of the SET domain methyltransferases, such as the SUV39, SET1 and SET2 families, have three conserved cysteine residues, which are essential for the enzymatic activity (Figure 5, A). There are also examples, such as SET7/9, which do not possess pre- and post-SET domains rich in cysteines, but rather have a C-terminal alpha helix packed on the core SET active site in order to complete the access channel for the substrate lysine and to obtain a more ordered 3D structure (Figure 5, B) (Dillon et al., 2005; Qian & Zhou, 2006).

In general, SET domain HKMTs do not show much sequence similarity, and many of the residues in the active sites are not conserved, contributing to the large variety of specificities that these enzymes can have. In some cases, the SET domain itself contains an insertion, termed i-SET, which can be very different both in length and amino acid sequence. This part of the domain seems to have an important role in target specificity determination, as it was shown to interact with the substrate peptide. Peculiarly though, some SET domain methyltransferases have the same target specificity, such as H3K4 peptides for MLL1 and SET7/9, but their i-SET sequence shows no similarity, suggesting convergent evolution of these fragments (Qian & Zhou, 2006). Some SET domains have been shown to methylate multiple targets, and it has been inferred that the interior of the hydrophobic pockets for the substrate binding and the residues of the active site are responsible for defining a consensus motif that the enzyme prefers to methylate (Chuikov et al., 2004). The study of the specificity of such SET domain methyltransferases is one of particular importance, as it has the potential to unravel
many novel PKMT-dependent regulatory pathways by the identification of methyltransferase targets based on their substrate sequence specificities.

1.5.3. The SET domain HKMT families

So far, seven families of SET domain methyltransferases have been defined: SUV39, SET1, SET2, EZH, RIZ/PRDM and SMYD, while others are orphan members of the SET domain superfamily. The proteins in a family show a higher degree of sequence similarity both surrounding and inside the SET domain (Dillon et al., 2005).

1.5.3.1. The SUV39 family

The name of the family comes from the first members characterized, the *D. melanogaster* Su(var)3-9, which came up in a screen for PEV mutants as a repressor protein (Tschiersch et al., 1994), together with its homologs from mouse (Suv39h2), human (SUV39H1) and *S. pombe* (Clr4). In mammals, SUV39H1 has a paralog, SUV39H2, which presents 55% amino acid sequence identity to SUV39H1, and which is thought to have appeared by a recent gene duplication event. Although the enzymes show sequence similarity within the SET domains, the protein sequence presents only 30% identity with the *D. melanogaster* Su(var)3-9, and 27% with the *S. pombe* Clr4. All the previously mentioned PKMTs are responsible for the formation of heterochromatin by the setting of H3K9me3 marks (Dillon et al., 2005). The function of SUV39H1 and H3K9me3 in the formation of heterochromatin is described in more detail in sections 1.2.4 and 1.4.3.

The SUV39H1 protein is 412 amino acids long, and it contains the catalytic domain at the C-terminus, while the N-terminus serves for the regulation of its activity (Figure 6). A stretch of 44 amino acids close to the N-terminus is responsible for the interaction with HP1, which in turn binds specifically and strongly to H3K9me3, the mark SUV39H1 sets (Melcher et al., 2000). The targeting of SUV39H1 is also independent on HP1, with the help of its N-terminal chromodomain (CHRomatin Organization MOdifier domain), which recognizes the same mark (Wang et al., 2012).

All targeting mechanisms for SUV39H1 seem to promote spreading of the H3K9me3 mark in a seemingly continuous feedback loop, but it is not clear how the enzyme is recruited to sites of initiation of methylation. It is believed that both the chromodomain of Suv39/Clr4 and that of HP1/Swi6/Chp2 recognize H3K9me, but they do not compete in binding, as the enzyme
chromodomain prefers H3K9me3, while HP1, a more abundant protein inside heterochromatin, does not discriminate between the abundant H3K9me2 and H3K9me3 (Al-Sady et al., 2013).

<table>
<thead>
<tr>
<th>SUV39 family</th>
<th>MLL2 (5537 aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLP (1298 aa)</td>
<td>SET1A (1707 aa)</td>
</tr>
<tr>
<td>G9a (1210 aa)</td>
<td>SET1B (1966 aa)</td>
</tr>
</tbody>
</table>

**Figure 6 - Structural features of the PKMTs studied in the present work.** The black boxes group the enzymes in their respective families. ANK - ankyrin repeats; PHD finger – Plant homeodomain finger; HMG box - high mobility group box; FYRN - Phe/Tyr-rich domain N-terminal; FYRC - Phe/Tyr-rich domain C-terminal; RRM - RNA recognition motif; PWWP - Pro-Trp-Trp-Pro motif domain; AWS – Associated With SET domain; SANT – Swi3, Ada2, N-Cor, TFIIB domain; MORN –Membrane Occupation and Recognition Nexus. Not drawn to scale. The figure was produced with VectorNTI Advance™ 11.0.
For the yeast Clr4, it has also been shown that the binding of the protein’s chromodomain to the H3K9me3 mark causes enhancement of the catalytic activity (Al-Sady et al., 2013). Since Clr4 and the human SUV39H1 show only 27% identity, and because the function of this enzyme is essential for survival, it would be interesting to find out if these observations are also valid in humans.

When discussing mobility of the protein inside the nucleus and within heterochromatin, it was shown that in comparison to HP1, a large fraction of about 20-40% of SUV39H1 is immobile at the site of pericentric heterochromatin, and may play a structural role there. Using FRAP (Fluorescence Recovery After Photobleaching) experiments, HP1 showed a very dynamic half-life on heterochromatin, of about 4.2 seconds, and complete recovery within 1 minute, while the dynamic fraction of SUV39H1 had a recovery half-life of 19 seconds (Krouwels et al., 2005).

In Drosophila, Su(var)3-9 also gets recruited to chromatin by interacting with histone H1, which is mostly present in condensed heterochromatin. The binding of H1 does not stimulate the enzymatic activity of the protein, but is rather hypothesized to facilitate the methylation by positioning the enzyme in the context of the nucleosome substrate. It appears that Su(var)3-9 and H1 contribute together to the silencing of transposable elements (TEs) and TE-like sequences (Lu et al., 2013). However, the interaction between Su(var)39 and the H1 protein has not been characterized. Moreover, the interaction still awaits confirmation in the case of the human homolog SUV39H1, together with an identification of the interacting regions between the two binding partners.

In metazoans, there are several PKMTs that use H3K9 as substrate, and their functions have been split according to the preferred product methylation state. So far, the enzymes discussed are involved in heterochromatin formation, but the euchromatin H3K9 methyltransferases G9a and its paralog G9a-like protein (GLP) are also part of the SUV39 family of methyltransferases (Figure 6). There are no corresponding orthologs for the latter two enzymes in yeast, frog or worm, and the closest protein in Drosophila shows only 20% identity to human G9a and 18% identity to GLP. G9a and GLP are very similar in their catalytic C-terminal part, and show most differences on the N-terminal side. A GLP ortholog in zebrafish also shows only 45% identity to the human protein (Dillon et al., 2005).

Although G9a and GLP have been shown to have the same substrate specificity and are active independently in vitro, they seem to have non-redundant functions in vivo. Knockout of each
of the enzymes results in severe reductions of H3K9me1 and H3K9me2 levels (Shinkai & Tachibana, 2011). The explanation for these peculiar effects was found in 2005, when Tachibana et al. showed that G9a and GLP have the ability to form homo- and heterodimers, by interaction of their SET domains. In vivo however, only formation of heteromeric complexes was detected (Tachibana et al., 2005). WIZ (Widely-interspaced zinc finger-containing protein), a multiple zinc finger protein, was identified to bind with high affinity to the G9a/GLP dimer, and probably this three member complex is the most stable form in which G9a and GLP are present in cells (Tachibana et al., 2005). Another interactor of G9a is HP1, which binds to the auto-methylated N-terminal part of G9a (Rathert et al., 2008).

Just as the SUV39H1 protein, G9a and GLP also contain additional functional domains. In the case of G9a and GLP, they are ankyrin repeats (Figure 6), which recognize the marks set by the enzymes: H3K9me1/2. This recognition facilitates the spreading of the mark (Collins & Cheng, 2010).

The SUV39 family of protein lysine methyltransferases contains one more pair of enzymes which show sequence similarity, and both methylate H3K9: SETDB1 and SETDB2 (SET domain bifurcated 1 and 2) (Dillon et al., 2005). The first one di- and trimethylates class I and II ERVs (endogenous retroviral regions) in mESCs and plays a critical role in stem cell maintenance (Karimi et al., 2011). SetDB1 occupies a subset of bivalent genes which are also H3K9-methylated, and its absence results in destabilizing the mES cell state (Bilodeau et al., 2009). In zebrafish, Setdb2 has been shown to be repressing fibroblast growth factor 8, participating in cilia formation and left-right asymmetry (Xu et al., 2010).

1.5.3.2. The SET1 family

The SET1 family of proteins contains several members, ranging from yeast to humans. Some of the members are the Drosophila TRX and its mammalian homologs, MLL1 (KMT2A, encoded by chromosome 11q23 locus) and MLL4 (KMT2B, encoded by chromosome 19q13.1 locus), the TRR (Trithorax related) from flies together with its mammalian homologs, MLL2 (KMT2D, encoded by chromosome 12q13.12 locus) and MLL3 (KMT2C, encoded by chromosome 7q36.1 locus), and also the yeast Set1 protein with its homologs, the fly Set1, or the mammalian SET1A (also known as SETD1A or SET1) and SET1B (also known as SETD1B or SET1L). All of the SET1 family members act on the same substrate and set the methylation of H3K4. They also have the SET domain C-terminally placed, followed by a cysteine motif
containing post-SET (Völkel & Angrand, 2007). Although these proteins are part of the same family, their actual sequence homology does not reach over 30% in any case (Ansari et al., 2009).

**The MLLs**

The Mixed Lineage Leukemia (MLL) proteins, as their name states, are frequently involved in several aggressive leukemias, such as acute myelogenous leukemia (AML) or acute lymphoblastic leukemia (ALL) (Ansari et al., 2009). These diseases normally occur because of translocations in the MLL genes, in which most often the N-terminal parts of MLL1 become fused to the C-terminus of another protein, such as AF4 or ENL (Hess, 2004). MLL2 and MLL3 are also often involved in pancreatic carcinoma, glioblastomas and solid tumors (Ruault et al., 2002).

The four MLLs are very large proteins, with amino acid sequences of a few thousand residues, and the general organization of the MLL1/MLL4 and MLL2/MLL3 pairs are very similar, though not identical. MLL1 and MLL4 contain a CXXC domain in the N-terminal regulatory part, which could recruit them to unmethylated DNA target sites (Birke et al., 2002). They also each contain four plant homeodomain (PHD) fingers and one bromodomain, conserved motifs known to facilitate protein-protein interaction, such as binding to modified histone tails. The FYRN (Phe/Tyr-rich domain N-terminal) and FYRC (Phe/Tyr-rich domain C-terminal) domains interact with each other, and serve for the formation of heterodimers from the MLL fragments once the enzyme gets cleaved during its maturation by the protease Taspase1 (Hsieh et al., 2003). MLL1, 3 and 4 also contain a series of N-terminal AT-hooks, which are probably involved in their localization to target genomic loci. The MLL2 and MLL3 enzymes contain seven, and eight PHD fingers, respectively, which probably serve for the regulation of interactions with histones and other binding partners. However, the function of most of these domains is largely unknown. So far, the functions of the two enzymes have been reported as redundant (Hu et al., 2013). If this were truly the case, why would there be differences in the regulatory domains of the enzymes, such as the additional PHD finger in MLL3? The investigation of the PHD fingers of MLL2 and MLL3 could shed light on the fine differences between the functions of the two proteins. Interestingly, the enzymes also contain the FYRN/FYRC domains, but they are situated adjacent to each other, not separately as in MLL1/MLL4 (Figure 6) (Ali et al., 2014; Patel et al., 2014).
In yeast, the only H3K4 methyltransferase is Set1, and it is active only when it acts in complex with several other proteins forming together the so called COMPASS multiprotein-subunit complex. It has been shown that the same kind of complexes is formed also in mammals around either of the enzymes MLL1, MLL2, MLL3, MLL4, SET1A or SET1B. They all share a set of common complex members: ASH2 (Absent, Small or Homeotic 2), WDR5 (WD40 repeat containing protein 5), DPY30 (Dosage compensation protein dpy-30 homolog) and RBBP5 (retinoblastoma binding protein 5) (Hu et al., 2013). The core complex members do not seem to affect the localization of the enzymes, but they are rather involved in regulating their catalytic activity (Ansari et al., 2009). The additional members added to the complexes divide the functions of the different enzymes in the cells (Table 1).

One of the essential functions of the MLL proteins is the activation of the evolutionarily conserved homeobox containing genes, which play critical roles in embryonic development. MLL1 and MLL4, just like the Drosophila Trx, are involved in the maintenance, but not initiation, of the Hox gene expression patterns, in order to dictate the positional tissue differentiation in complex organisms. In accordance to this, knockout of MLL results in embryonic lethality (Ansari et al., 2009).

Table 1 - Composition of the human COMPASS and COMPASS-like co-activator complexes. All the core complex members, WDR5, RBBP5, DPY30 and ASH2, are the same. The additional factors result in the formation of three types of complex, each containing one methyltransferase: SET1A/SET1B, MLL1/MLL4 or MLL2/MLL3 (Hu, Garruss, et al., 2013).

<table>
<thead>
<tr>
<th>MTase</th>
<th>hCOMPASS</th>
<th>hCOMPASS-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core complex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTase</td>
<td>SET1A</td>
<td>SET1B</td>
</tr>
<tr>
<td>MTase</td>
<td>MLL1 (KMT2A)</td>
<td>MLL4 (KMT2B)</td>
</tr>
<tr>
<td>MTase</td>
<td>MLL3 (KMT2C)</td>
<td>MLL2 (KMT2D)</td>
</tr>
<tr>
<td>Additional members</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional members</td>
<td>HCF1</td>
<td>CXXC1</td>
</tr>
<tr>
<td>Additional members</td>
<td>WDR82</td>
<td>Menin</td>
</tr>
<tr>
<td>Additional members</td>
<td>ASC-2</td>
<td>UTX</td>
</tr>
<tr>
<td>Additional members</td>
<td>PTIP</td>
<td>PA1</td>
</tr>
</tbody>
</table>
The MLL2 and MLL3 methyltransferases show considerable, but not complete, overlap between their tissue expression patterns. MLL2 is expressed in most adult tissues, including various hematopoietic cells, but expression is not detected in the liver (Data retrieved from gene expression databases: Bgee ID O14686, CleanEx ID HS_MLL2, Geneinvestigator ID O14686). MLL3 expresses in high levels in the ovary and testis, followed by brain and liver. In addition, MLL3 is expressed in heart, lung, kidney, peripheral blood, placenta and fetal thymus. Lack of expression was observed in fetal liver and skeletal muscles (Data retrieved from gene expression databases: Array Express ID Q8NEZ4, Bgee ID Q8NEZ4, CleanEx ID HS_MLL2, Geneinvestigator ID Q8NEZ4). This suggests that maybe MLL2 and MLL3 are not completely redundant, as in some tissues only one of the two is expressed.

MLL2 and MLL3 are involved in nuclear receptor (NR)-based signaling. They are the catalytic units of the Trithorax-related (Trr) complex ASCOM (ASC-2 complex). ASCOM is composed of either MLL2 or MLL3, the core complex members, ASC-2 (Activating signal cointegrator 2, also known as NCOA6), PTIP, PA1 and UTX (Hu et al., 2013). ASC-2 is involved in the signaling pathways of many nuclear receptors, such as retinoic acid receptors (RAR) or estrogen receptors (ER). It is thought that the MLL2 and MLL3 ASCOM complexes are not identical in composition, and that the enzymes have redundant, but still essential functions in nuclear receptor induced H3K4 methylation.

The MLL1 and MLL4 proteins are also involved in NR pathways. Both proteins, and also their interactor, Menin, contain LXXLL ER-interacting motifs, and have been shown to be involved in the regulation of estrogen responsive genes (Ansari et al., 2009). In addition, Menin, part of the MLL1 or MLL4 complex, is involved in the regulation of the CDK inhibitors p27 and p18 (Hughes et al., 2004). In addition, the MLL proteins MLL1 and MLL4 also have crucial roles for the cell cycle progression. MLL1 was shown to activate the promoters of cyclins E1 and E2. It is also known that MLL1 and MLL4 interact with distinct subsets of E2F proteins, transcription factors which activate cell cycle regulatory players, such as cyclins (Takeda et al., 2006).

Recent results have uncovered a role of MLL4 in the H3K4 trimethylation of bivalent genes in embryonic stem cells. Depletion of MLL4 was shown to result in loss of bivalency of genes involved in differentiation, such as all four Hox gene clusters, by depletion of their H3K4me3 enrichment. The effect of the MLL4 knock out was restricted to the above-mentioned bivalent loci, and did not have a major effect global H3K4 methylation levels. Although MLL1 and MLL4
are very similar, their functions are divided, also by the fact that MLL4 is expressed very early in embryonic development, while MLL1 gets expressed in a mid-to-late stage of development, including hematopoiesis (Hu et al., 2013).

The other two MLLs, 2 and 3, have been involved in gene repression at promoters characterized by the modification set: H3K4me1, H3K27me3 and H4K20me1. Although so far H3K4 methylation states have been exclusively associated with transcriptional activation, it seems that monomethylation introduced by MLL2/3 at certain promoters can cause a context-dependent repression of the corresponding genes. This is a particular situation, in which the genes responsible for muscle development are silenced in myoblasts, but expressed in myotubes. Cheng et al (2014) suggest that the H3K4me1 at the TSS of these genes marks them for conditional repression, and correlate the mark with high nucleosomal occupancy at these promoters. It appears that in myoblasts, the MLL2/MLL3 enzymes set the H3K4me1 and cause the subsequent silencing of the muscle development genes, while in myotubes, the MLL1/4 and SET1 proteins are recruited to these loci and cause activation of transcription by H3K4 trimethylation (Cheng et al., 2014). The function of these two enzymes is still rather unclear, as the mark they establish seems to be associated both with active enhancers and with conditionally repressed genes. Therefore, the study of the targeting and substrate specificity of MLL2 and MLL3 is critical.

**SET1A/B**

The human SET1A and SET1B are homologs of the single H3K4 methyltransferase in budding yeast, Set1, showing identity to it of 35%, and 37%, respectively. Among themselves, they present approximately 39% identity, while their SET and post-SET domains alone are 85% identical and 97% similar (Figure 6). The pre-SET domains of SET1A and SET1B are also 46% identical. The two enzymes show similarities also in their regulatory N-terminal regions (Lee et al., 2007). Both SET1A and SET1B only show activity within the COMPASS complex, together with the already discussed core complex members, and also share two Set1/COMPASS-specific interactors, WDR82 and CXXC1 (or CFP1), while HCF-1 is a unit shared also with the COMPASS-like complexes of MLL1 and MLL4 (Table 1) (Hu et al., 2013). It has been shown in yeast that the human RBBP5 homolog, Swd1, and the human WDR5 homolog, Swd3, are required for the transition from di- to trimethylation by the Set1 complex. WDR5 recognizes H3K4me2, and is important for the targeting of the complex (Wysocka et al., 2005).
The catalytic activity of the enzymes is affected by an autoinhibitory domain that is found in the central part of the Set1 sequence (Soares et al., 2014). In addition, H3K4 methylation is dependent upon H2BK120 monoubiquitination, a mark that is recognized by WDR82. The interplay between SET1 complex activity and H2B ubiquitination appears to also be dependent upon the integrity of the SET1 pre-SET domain (Kim et al., 2013). In contrast to the other subunits, CFP1 was reported to play an inhibitory role for the SET1A and SET1B complexes.

In yeast, the regulation of Set1 levels has been shown to be achieved through preferential degradation of Set1 which is in complex with RNA Pol II or with H3K4me in order to suppress random and promiscuous H3K4 trimethylation along the length of the gene, and keep the trimethylation domains restricted around the TSS. Set1 is tethered to gene promoters by specific regulators, which prolong its stay to facilitate complete methylation of the lysine 4 residues (Soares et al., 2014).

Pertaining to their nuclear localization, the two complexes appear in euchromatic nuclear speckles that do not co-localize, suggesting that they have non-redundant functions in the cells, by targeting different sets of genes for activation (Lee et al., 2007). In addition to their ubiquitous expression, SET1A and SET1B also seem to have a more widespread role in histone H3 lysine 4 trimethylation, as depletion of these complexes by WDR82 knockdown results in a major reduction of H3K4me3 levels, although the MLL1 to 4 are not affected (Wu et al., 2008).

There is no explanation so far for the partial non-redundancy of the two SET1A and SET1B enzymes. It could probably be caused either by differential targeting through “reading” domains or indirectly through interaction partners, or by different substrate specificity that would potentially result in the methylation of alternate targets, in addition to H3K4. It is also unknown how this specificity compares to the MLL enzymes, for example.

1.5.3.3. **The SET2 family**

The SET2 family is composed of several proteins: three closely related members, named Nuclear receptor binding SET domain proteins 1, 2 and 3 (NSD1, NSD2, and NSD3); the *Drosophila* protein Ash1 and its homolog ASH1L, the methyltransferase SETD2, and the uncharacterized protein SET2L. The SET2 family is characterized by the presence of the SET and the post-SET domains, preceded by an N-terminal AWS (Associated With SET) motif.
The AWS is also a cysteine-rich domain and is probably functionally similar to the pre-SET domain of the SUV39 family. This family of methyltransferases presents various histone lysine methylation activities. SETD2, a huntingtin interacting protein, has been reported to methylate H3K36 (Sun et al., 2005), while NSD1, 2 and 3 are also active on H3K36 (Rayasam et al., 2003). ASH1 methylates H3K4 and H3K9, and also H4K20 (Beisel et al., 2002). These PKMTs associate in multi-protein complexes that function as coactivators and corepressors involved in numerous cellular processes. NSD1, 2 and 3 have nonredundant functions during development. NSD3 (Figure 6) contains several chromatin binding domains in addition to the SET domain. These motifs recruit the protein to unmodified H3K4 and H3K9me3 tails (He et al., 2013). NSD1 has been shown to specifically methylate some non-histone targets in addition to histone H3 (Kudithipudi et al., 2014). However, no similar investigations have been done for NSD2 and NSD3.

1.5.3.4. The EZH family

The EZH family contains the Polycomb group protein E(Z) and its homologs. In humans, these are the two proteins EZH1 (Figure 6) and EZH2. The EZH family members contain a SET domain N-terminal cysteine rich region, showing some similarity to pre-SET and AWS motifs in the other families of protein methyltransferases. In Drosophila, E(Z) associates with ESC and SU(Z)12 as part of the repressive complex PRC2, responsible for H3K27 methylation. The Polycomb complexes are important for initiating the formation of H3K27me3-containing repressive domains, and the regulation of developmental gene expression (Mousavi et al., 2012). The human homolog EZH2, is additionally also responsible for the methylation of H1K26, which has a transcriptional repressive function (Völkel & Angrand, 2007).

Both EZH1 and EZH2 function together in introducing the complete H3K27 methylation pattern, being responsible for ESC identity and pluripotency. The two enzymes are part of different varieties of the PRC2 complex, and their functions in ESCs are non-redundant. Surprisingly, the PRC2-EZH1 complex seems to have the capacity to condense chromatin independent of the catalytic activity of EZH1 (Margueron et al., 2008). The expression levels of the two proteins are inversely correlated throughout the development process, with EZH2 level high during embryonic development, and EZH1 expressed mostly in adult tissues (Margueron et al., 2008). More recently, EZH1 has been located at genomic loci that associate with active transcription, H3K4me3, and RNA Pol II presence (Mousavi et al., 2012). Moreover,
EZH1 is able to only partially compensate for EZH2 loss during ESC differentiation (Shen et al., 2008). The depletion of EZH1 does not affect global methylation levels of H3K27 in ESCs, partially due to its low expression in this cell type. EZH1 seems to have a more restricted presence, as it targets only a subset of the PRC2-EZH2 genes (Margueron et al., 2008). However, EZH1 depletion during differentiation resulted in a reduction of RNA Pol II levels inside the gene bodies and delayed activation of genes responsible for differentiation. When EZH2 is depleted, EZH1 is able to fully compensate for its loss at H3K27 monomethylation level (Mousavi et al., 2012). More and more studies are starting to uncover the mysterious roles of the more weakly active EZH family member, EZH1. Of course, the reason for the lower level of EZH1 activity could be that histones may not be its primary substrate, and that its function could partially be independent of H3K27 methylation. However, there has not been any publication focusing on this aspect.

1.5.3.5. **The RIZ/PRDM family**

The PRDM family comprises proteins that contain a SET-like domain with some critical sequence changes in SET conserved motifs. The PR/SET domain has been reported to show activity in a few cases, such as PRDM2/RIZ1, responsible for H3K9 methylation (Kim et al. 2003), or the PRDM9 mouse homolog, responsible for H3K4 methylation (Hayashi et al., 2005). Several of the PRDM family members contain Kruppel-type zinc fingers (of the C2H2 type), which suggests their direct sequence-dependent recruitment to DNA (Völkel & Angrand, 2007).

1.5.3.6. **The SMYD family**

The SET and MYND-domain containing proteins (SMYD) family has five members, each containing the zinc finger of MYND type. The MYND domain offers specific interaction with DNA. An example is SMYD3, a H3K4 methyltransferase that also interacts with RNA Pol II, contributing to H3K4 methylation spreading during transcription elongation (Völkel & Angrand, 2007).

1.5.3.7. **Orphan methyltransferases**

The last class of methyltransferases contains SET domain proteins, which lack a post-SET domain. This group contains the histone 4 lysine 20 di- and trimethyltransferases SUV4-20H1/2, the H4K20 monomethylase PR-SET7/SET8, the monomethyltransferase SET7/9, and MLL5 (Völkel & Angrand, 2007). SET7/9 (Figure 6) has important roles in DNA damage repair
(DDR) and genome stability (Chuikov et al., 2004) and has been reported to be involved in the transcriptional activation of the collagenase (Martens, Verlaan, Kalkhoven, & Zantema, 2003) and insulin (Francis et al., 2005) genes. SET7/9, originally thought to methylate H3K4, has been shown to be inactive on nucleosomal substrates (Huang & Berger, 2008). MLL5 (Figure 6), although a member of the MLL family based on overall sequence similarities, is characterized as orphan due to the lack of a Post-SET domain. MLL5 has essential functions in cell cycle, homeostasis and survival. The protein is recruited to H3K4me3 marks by its PHD finger, and loses the binding if threonine 3 or 6 are phosphorylated, releasing the protein during mitosis (Ali et al., 2013). The activity of this MLL family member had not been investigated in detail.

1.6. Investigation of PKMT non-histone targets

In general, many histone lysine methyltransferases have been shown to also be involved in methyllating other proteins. In humans, there are 56 SET-domain proteins identified and so far all the ones characterized are lysine methyltransferases (Clarke, 2013). Most of these enzymes are thought to be histone lysine methyltransferases on the basis of sequence similarities between SET domains (Clarke, 2013), many of them have been shown to also methylate non-histone targets, while others seem to be exclusively non-histone methyltransferases, like SET7/9 (Huang & Berger, 2008).

The methylation of the tumor suppressor p53 by SET7/9 was the first non-histone methylation performed by a PKMT ever to be reported. The mark at a specific C-terminal lysine of p53 causes the protein’s activation (Chuikov et al., 2004). The oncogene p53 in its methylated form is bound by Tip60, which subsequently acetylates it (Kurash et al., 2008). SMYD2 or Set8 can also methylate p53 at two different lysine positions, causing its inactivation (Huang et al., 2006; Shi et al., 2007). SET7/9 methylates several other non-histone targets to regulate their functions: E2F1 (E2F transcription factor 1), SIRT1 and SUV39H1 (Wang et al., 2013). In the case of SUV39H1, the methylation by SET7/9 results in severe downregulation of the catalytic activity of the enzyme, and subsequent chromatin relaxation and genome instability, finally inhibiting cell proliferation (Wang et al., 2013). The methylation of TAF10 by SET7/9 results in recruitment of RNA Pol II to TAF10-targeted genes and their activation (Rathert et al., 2008). Recently, the nuclear targets of SET7/9 were extended to several more non-histone targets: AKAP6, CENPC1, MeCP2, MINT, PPARBP, ZDH8, IRF1, Cullin1 and TTK, and also histones H2A and H2B (Dhayalan et al., 2011).
Also the methyltransferases G9a and GLP are not exclusively specific for H3K9, but rather have a target Arg-Lys motif, with the arginine residue unmethylated. The enzymes have been shown to methylate also other histone and non-histone targets, such as H1 and H3K27, or respectively the non-histone proteins CDYL1 (Chromodomain Y-like protein 1), WIZ, ACINUS (Apoptotic chromatin condensation inducer in the nucleus), and the G9a and GLP N-terminal domains. HP1-β was shown to recognize the methylated non-histone target sequences at peptide level (Rathert et al., 2008). Interactions with reading domains such as this one, if valid in vivo, might serve to direct the targets to specific locations in the nucleus. In the case of CDYL1, the methylation by G9a affects the CDYL1 chromodomain ability to bind to H3K9me3 peptides, as the CDYL1 internal methylation competes with the H3 peptide for binding into the aromatic pocket (Rathert et al., 2008).

Another reported non-histone methylation event performed by G9a is the dimethylation of K44 of mouse Dnmt3a or K47 of human DNMT3A, which can be bound by the chromodomain of the MPP8 protein. MPP8 is also known to bind to the automethylation site of GLP, and to form dimers, which could result in a bridging function of MPP8 between the G9a/GLP complex and DNMT3A, resulting in the formation of a strong silencing complex (Chang et al., 2011).

The rapid identification of many PKMT non-histone targets raises the need for efficient ways to determine the functions of these methylation events. Proteins are normally methylated on their surface, and therefore the methyl groups are not likely to interfere much with protein structure. It can be assumed that the methylated target lysine will serve either to recruit interactors or to disrupt the binding of interactors, therefore acting as a switch between two different conditions. So far, the study of the roles of PKMT target methylations has not been done via high throughput screens, but rather in more directed, hypothesis-directed manners, which are more likely to fail. Therefore, a high throughput approach would be very useful, if it were developed.

1.7. Reading Domains

The protein domains that are recruited specifically to particular PTMs on histones or other proteins are termed mark “readers”. The marks serve as scaffolds for their corresponding readers, which in turn recruit so called “effectors”, proteins that establish different chromatin states and gene transcription states. The binding of reading domains to histone PTMs has been involved, in addition to gene transcription regulation, in DNA replication, repair and
recombination processes (Patel & Wang, 2013). The readers contain conserved domains responsible for the PTM binding capacity, and are often part of multi-protein complexes, which are directed by the reader to target chromatin loci. Reading domains can be structurally categorized in several different groups, such as the domains of the Royal family (chromododomains, tudor, PWWP and MBT domains), plant homeodomain (PHD) fingers, or bromodomains (Kouzarides, 2007).

1.7.1. Methyl Lysine Readout

Members of the “Royal family”, a large superfamily of homologous conserved domains, perform most of the methyllysine readout. The Tudor domains, named the same as the Drosophila protein where they were first identified, show close similarity to plant Agenet domains and also to MBT (malignant brain tumor) domains. At the same time, MBT domains are similar to chromododomains and PWWP (Pro-Trp-Trp-Pro motif) domains, suggesting that all these separate types of domains are divergent branches of the same superfamily (Maurer-Stroh et al., 2003). Other domains, mostly associated with methylated lysine or arginine binding, are the PHD fingers, while BAH (bromo-adjacent homology) domains have been reported to bind only methylated lysines (Patel & Wang, 2013).

The tudor domain is conserved and found in various species. In mammals it is present in around 30 proteins. This domain typically comprises ~60 amino acids, folded into a barrel structure composed of 4-5 β-strands. A subset of tudor domains has been assigned to bind to methylated arginines, and they are shown to be involved in processes such as germ cell development, alternative splicing or RNA metabolism. The second subset of Tudor domains comprises methyllysine-binding domains, which poses an aromatic cage to interact with the methylated substrate. Examples of studied Tudor domains include those of UHRF1, or the proteins PHF19 (PHD finger protein 19) and PHF1 (PHD finger protein 1) (Lu & Wang, 2013).

The chromodomain was first identified and characterized in the HP1 protein, which contains one chromo- and one chromoshadow domain, connected by a hinge region. The chromodomain presents a histone-like fold, having some similarity to the archaeal DNA-binding proteins Sac7d/Sso7d (Lomberk et al., 2006). The domain is composed of approximately 50 amino acids that fold into a short three-stranded anti-parallel β-sheet, with an α-helix running across it. Inside the hydrophobic core there are several conserved amino
acids, among which the aromatic residues that form the hydrophobic cage responsible for the methyllysine binding (Brehm et al., 2004) (Figure 7).

![Figure 7 - Crystal structure representations in ribbon form of the chromodomains of SUV39H1 (PDB ID: 3MTS), MPP8 (PDB ID: 3R93) and HP1 (PDB ID: 1KNE). The secondary structures are labeled. The figure was taken from Wang et al., 2012.]

The chromodomain of HP1 is highly specific for H3K9me3 tail binding, with a dissociation constant (Kd) of 2.5 μM (Jacobs & Khorasanizadeh, 2002). The peptide binds in a hydrophobic groove on the surface of the chromodomain, and the interaction of the residues forming the groove with the H3 residues determines the specificity of the domain. The trimethylated lysine is accommodated in between three aromatic residues and stabilized by cation-π interactions (Jacobs & Khorasanizadeh, 2002). In humans, there are three different paralogs of HP1: α (CBX5), β (CBX1) and γ (CBX3), members of the group of chromobox proteins (CBX). Although highly conserved, HP1α, β and γ do not have identical functions. hHP1α and hHP1β are associated with heterochromatin, and contributing to its condensation, while hHP1γ is present in euchromatin, and is involved in RNA processing and transcription elongation (Canzio et al., 2014).

The second identified chromodomain was that of the Polycomb group protein PC2 from Drosophila (Kouzarides, 2007). The mammalian homologs of PC2 are the Cbx (Chromobox) proteins. Cbx2, Cbx4, Cbx6, Cbx7 and Cbx8 have been reported to associate with the PRC1 complex. At any one time, only one of the Cbx proteins can be associated with the PRC1 complex, and each influences the function of the complex during pluripotency and differentiation in a different manner. In mouse ESC2, the maintenance of pluripotency is
achieved by Cbx7, which promotes the PRC1 complex-dependent repression of pro-differentiation genes, while Cbx2 and Cbx4 direct lineage commitment by simply changing the targeting of the PRC1 complex (Morey et al., 2012). Although the chromodomains of these proteins have been studied extensively, the binding specificities are in some cases not completely clear. For example, Cbx7 is reported to bind strongly not only to H3K27me3, the same mark set by the complex it belongs to, but also to H3K9me3, a mark associated with constitutive heterochromatin, where PRC1 is normally not targeted (Bernstein et al., 2006).

Chromodomains have been identified in many other proteins, such as the heterochromatic methyltransferase SUV39H1 (Wang et al., 2012), the ATPase CHD1 (Pray-Grant et al., 2005), or the H3K9me3-binder MPP8 (M-phase phosphoprotein 8), which is involved in epithelial-to-mesenchymal transition (EMT) (Kokura et al., 2010).

The PHD finger is a conserved motif formed by two anti-parallel β-sheets followed by at least one short α-helix, and folded into a cross-braced topology. The structure coordinates two zinc ions with the help of a Cys4-His-Cys3 motif, which stabilizes the overall structure. The first investigated PHD fingers were those of BPTF (Bromodomain and PHD finger-containing transcription factor) of the NURF remodeling complex, and of ING2 (Inhibitor of growth 2). Just like for the chromodomains, PHD fingers form a similar aromatic pocket for the binding of the methylated lysine, while the peptide lies in a hydrophobic groove on the PHD finger surface. The aromatic cage can differ, both in the residues that are part of it, and in their number. For example, the aromatic cage of BPTF is composed of four aromatic residues, while that of JARID1A is formed of only two perpendicular Trp residues (Patel & Wang, 2013).

Most of the nuclear epigenetic regulators contain reading domains, and sometimes, depending on the protein size, one protein can encompass a large number and variety of domains. These conserved domains serve, either individually or in associations among domains, to interact with other proteins or with DNA. The study of the binding specificity of reading domains is essential for the identification of the corresponding protein functions and the identification of the regulatory networks to which they belong.

1.7.2. Acetyl Lysine Readout

Another relevant lysine PTM is acetylation, which can occur at many different lysines concomitantly and is less position specific than methylation. The bromodomain was the first
identified histone lysine acetylation “reader”. The domain is found in proteins that are part of chromatin remodeling and coactivator complexes, acetyltransferases, methyltransferases or helicases. The first structure solved was that of PCAF (p300/CBP-associated factor) bound to H4K8ac, and it showed that bromodomains fold into a left-handed, antiparallel, four-helical bundle, having the long loops which connect the helices packed against the structure, to form the peptide binding hydrophobic pocket (Dhalluin et al., 1999). The binding constants of bromodomains towards their target marks span a wide range, from values like 3 µM, to as high as 350 µM (Jacobson, 2000). In many cases, the bromodomain is found in tandem repeats, such as in the case of TAF1 of TFIID, which binds to poly-acetylated H4 peptides, at K5, K8, K12 and K16 (Jacobson, 2000).

More recently, tandem PHD fingers have been shown to be also capable of acetyllysine binding. The DPF3b protein, a member of the BAF chromatin remodeling complex, contributes to the activation of genes involved in heart and muscle development (Lange et al., 2008). It has two adjacent PHD fingers at the C-terminus, which pack against each other in a face-to-back conformation, with the H3 peptide bound across both domains. PHD1 is responsible for the H3K14ac interaction, while PHD2 binds to the N-terminus of the peptide (Zeng et al., 2010) (Figure 8, A).

The acetyl binding pocket is present on a different position on the surface than the pockets which are normally responsible for methylated lysine binding in PHD fingers (Patel & Wang, 2013). Residues 11-16 of the peptide bind in an elongated hydrophobic groove, which is perpendicular to the PHD1 β-sheets. The acetylated lysine is accommodated in a hydrophobic pocket, formed of Phe 264, Leu 296, Trp 311 and Ile 314, while the Asp 263 carbonyl oxygen and the Cys 313 sulfhydryl are in the range of a hydrogen bond distance to the acetyl amide of H3K14ac. The acyl chain of the acetylated lysine is stabilized by the interaction with an Arg289 side chain. Mutation of Phe 264 or Asp 263 to alanine resulted in loss of specificity for H3K14ac, while binding to the H3 peptide is kept (Zeng et al., 2010) (Figure 8, A).

The N-terminal part of the peptide forms a β-strand that is anti-parallel to the PHD2 β-sheet. The peptide is bound in a bent conformation, which causes the side chains of H3K4 and H3K9 to be brought together. Because of steric hindrance, the peptide containing methylated lysine 4 loses binding affinity (Figure 8, B). DPF3b was also able to bind in a similar manner to H4
peptides, H4K16ac and H4 with acetylated N-terminus, but with lower affinity (Zeng et al., 2010).

A second tandem PHD finger pair binding to acetylated lysine was identified recently, and its structure was solved this year (Dreveny et al., 2014) (PDB ID: 4LLB). MOZ (Monocytic leukemia zinc finger protein)/MYST3 is a MYST type acetyltransferase important in hematopoietic stem cell differentiation and self-renewal (Thomas et al., 2006). The N-terminal tandem PHD fingers of MOZ form a globular domain, with an N-terminal α-helix from the PHD1 folding onto PHD2. PHD2 is homologous with the majority of PHD fingers, but PHD1 has a more distinct sequence, somewhat closer to the BAF complex members DPF1, 2 and 3 (Dreveny et al., 2014).

This double PHD finger bound strongly to unmodified and H3K14ac peptides, and was unaffected by the presence of methylation at H3K9, but did not tolerate methylation of H3K4. The H4 peptide was also bound, but much weaker, and, in contrast to DPF3b, this took place only in the absence of acetylation marks. The H3 peptide adopts an α-helical conformation, from residues 4 to 11, with half of the side chains sticking out of the structure (Glu 5, Thr 6, Lys 9 and Ser 10), while the others project towards the interior of the domain (Lys 4, Ala 7, Arg 8 and Thr 11) (Figure 9, A). For this reason, methylation of H3K4 precludes binding, while the H3K9 residue not only does not influence binding, but is also exposed to the catalytic domain of MOZ, which acetylates it.
The H3K14ac binding pocket is situated in PHD1, and is composed of several hydrophobic amino acids (Ser 210, Phe 211, Asn 235, Leu 242, Trp 257, Cys 259 and Ile 260), having a weak positive electrostatic potential, which prevents the binding of unmodified K14 (Figure 9, B). The acetyl group of Lys 14 forms a hydrogen bond to the backbone amine of Ile 260 (Dreveny et al., 2014).

**Figure 9 - The structure of the MOZ/MYST3 double PHD fingers.** A. Surface representation of the PHD1-PHD2 structure colored according to the electrostatic potential (red – negative; blue – positive). The H3 peptide (red ribbon) adopts a helical shape. H3 residues are labeled in black. B. The H3K14ac peptide binding to MOZ, in close-up. The H3 peptide and residues are represented in red. The MOZ interaction pockets are represented by the blue crescents, with the residues involved in hydrogen bonding labeled in dark blue, and the ones involved in hydrophobic interactions labeled in light blue. The image was taken from Dreveny et al., (2014).

DPF3b and MOZ are so far the only PHD fingers identified to bind to acetylated lysines. However, there are many more PHD fingers from various nuclear proteins whose binding specificities have not been investigated or identified yet. Most probably, more of these tandem acetyllysine-binding PHDs will be identified, especially now that the pioneer examples have been found.
2. The Aim of the Presented Study

The present doctoral thesis work shows a look into the vast subject of lysine post-translational modifications, in particular investigating lysine methylation and acetylation, some enzymes responsible for the setting of the methylation marks, and some reading domains which recognize acetylated and methylated lysines, together with possible functions of the PTMs, and their readers and writers in the cellular environment.

2.1. Specific Goals and Achievements of Each Project

2.1.1. Yeast-3-Hybrid Protein-Protein Interaction Study for the Identification of Histone Lysine Methyltransferase Target Interactors

In recent years, multiple non-histone targets of PKMTs have been identified, and the methylations serve different functional purposes. Rathert and colleagues, (2008), and Dhayalan et al., (2011) identified several non-histone targets of G9a and SET7/9, respectively, but the functions of these target methylations has remained elusive. The aim of the current project was to identify interactors of these non-histone targets, and possibly identify interactions that are methylation-specific through a Yeast-3-Hybrid method. The work presented here shows the efforts to set up and to validate the novel method.

2.1.2. Identification of the Specific Targets of Histone Tail Reading Domains

The presence of chromatin reading domains in the structure of nuclear proteins is very common. These play essential roles in protein localization (Collins et al., 2008; Dhayalan et al., 2010; Fischle et al., 2005), or protein-protein interaction (Chang et al., 2011; Wang et al., 2010). The identification of the binding specificities of reading domains is essential for the understanding of protein functions within the intricate epigenetic regulatory networks.

The PHF1 Tudor domain binding to H3TK27me3 and H3K36me3

PHF1 (PHD finger protein 1) is a member of the PRC2 complex, and is known to bind to the H3K27 methyltransferase EZH2 via its two PHD fingers (O’Connell et al., 2001). In this study, the binding of the Tudor domain of PHF1 to its target peptides, H3K36me3 and H3K27me3 (the latter in the context of the histone variant H3T), was characterized. This investigation was part of a broader study of the function of PHF1 tudor domain, published in Kycia et al., 2014.
The CBX2 and CBX7 chromodomains binding to H3K27me3

CBX2 and CBX7 are two of the Drosophila PC homologs belonging to the PRC1 complex in human cells. CBX7 is highly expressed in ESCs, where it is essential for their self-renewal, Xi and the inhibition of differentiation (O’Loghlen et al., 2012). In ESCs, CBX7 is responsible for the negative regulation of the transcription of its orthologs, among which CBX2, which can replace it in the PRC1 complex and promote differentiation (O’Loghlen et al., 2012). CBX7 and CBX2 each contain a chromodomain that binds specifically to H3K27me3 (Bernstein et al., 2006).

In this work the binding specificity of the two chromodomains to H3K27me3 was confirmed and further characterized, to include also the differential preference between CBX2 and CBX7 with respect to binding to the H3.1 or H3T histone variants. In addition, the CBX2 and CBX7 chromodomains were used as tools for the validation of a novel concept based on the use of reading domains as alternatives to histone PTM-specific antibodies (Kungulowski et al, 2014).

The PHD fingers of MLL2 (KMT2D) and MLL3 (KMT2C)

The H3K4me1 methyltransferases MLL2 and MLL3 show very close resemblance both in the sequence of their SET domains, and also in their general domain organization and structure (Figure 6). Although differentially expressed, the two methyltransferases still show considerable overlap between their tissue expression patterns.

The aim of the project was to investigate the histone tail peptide binding of the MLL2 and MLL3 PHD fingers, and the MLL2 PHD 3-5 triplet and MLL3 PHD 4-6 triplet appeared to prefer binding to acetylated H4 (11-30) peptides sometimes associated with lysine or arginine methylations.

2.1.3. Determination of the Substrate Specificity of PKMTs and Search of Putative Non-Histone Targets

The SET domains, including their adjacent pre-SET, AWS and/or post-SET domains, of several PKMTs were cloned, purified as GST-fused recombinant proteins, and tested for methyltransferase activity. The catalytic domain of MLL2 presented strong methylation activity towards its known target, H3K4. The activity of the catalytic domains of SET1A and MLL2 was also investigated in the context of the COMPASS core complex (in the presence of WDR5, ASH2L, and RBBP5), in which the enzymes both successfully methylated H3K4. The
initial aim of this project was to identify the substrate specificity of the active enzymes, and based on it, to find their putative non-histone targets.

2.1.4. Study of the influence of the SUV39H1 chromodomain on the catalytic activity of the enzyme

The regulatory N-terminal part of SUV39H1 (Figure 6) contains a chromodomain which tethers the enzyme to H3K9me3. In this work, the characterization of the binding specificity of the SUV39H1 chromodomain to the H3K9 methylated tail was performed. Furthermore, the putative role of the chromodomain in the regulation of the enzyme activity was studied.
3. Materials and Methods

3.1. Yeast-Three-Hybrid Investigations

3.1.1. Preparation of the Yeast-3-Hybrid Vector Constructs

For the yeast-3-hybrid experiments, the Matchmaker™ Gold Yeast-Two-Hybrid System from Clontech (Cat. No. 630489) was used in the three-hybrid variety offered by the company. In this setup, the methyltransferase (MTase) and its target were cloned in the same vector, the pBridge vector (Clontech, PT3212-5, Cat. No. 630404), which replaces the pGBK7 (Clontech, Cat. No. 630443, 630489).

The methyltransferase catalytic SET domain was cloned in the Multiple Cloning Site II (MCS II) of the pBridge vector (Figure 10, B), using the NotI and BglII restriction sites. The methyltransferases used in this study were the human G9a enzyme (residues 931 – 1210), the murine G9a enzyme (residues 984 – 1263), the murine GLP enzyme (residues 1002 – 1295) and the full length murine Set7 enzyme (residues 1 – 366). The methyltransferase targets were cloned as Gal4-DNA binding domain (DNA-BD) fusions inside the MCS I of the pBridge vector, using EcoRI and BamHI restriction sites for the G9a/GLP targets and BamHI and PstI, or EcoRI and BamHI restriction sites for the Set7 targets. The MTase targets were cloned to express protein domains, not full length proteins.

The human and murine G9a, and murine GLP targets were the following: human CDYL (NCBI accession number NP_004815.3, residues 64 – 283), WIZ (NCBI accession number AAI36642.1, residues 179 – 359), human G9a N-terminal domain (NCBI accession number AAH18718.1, residues 112 – 352), and murine Dnmt3a (residues 1 – 218, 1 – 274).

The Set7 targets were cloned as the following fragments: human AKAP6 (NCBI accession number NP_004265.3, residues 485 – 776), human CENP1 (NCBI accession number NP_001803.2, residues 279 – 474), human IRF1 (NCBI accession number NP_002189.1, residues 71 – 240), human MeCP2 (NCBI accession number NP_004983.1, residues 201 – 486), human MINT (NCBI accession number NP_055816.2, residues 1994 – 2281), human PPARBP (NCBI accession number NP_004765.2, residues 847 – 1084), and human ZDH8 (NCBI accession number NP_001171953.1, residues 227 – 400).
Figure 10 - Maps of the pGADT7 Gal4 AD-fusion vector and the pBridge Gal4-DNA-BD fusion vector. A. Example of the pGADT7 vector map, with HP1β chromodomain cloned in the multiple cloning site as Gal4-AD-fusion. B. Example of the pBridge vector map, with the PKMT target Dnmt3a (1-274) cloned with an additional C-terminal His-tag in MCS I, and G9a SET (984-1263) cloned in MCS II. The images were made using SnapGene 2.2.2.
The Gal4 activation domain (AD) was fused to the methyl-lysine binders HP1β (NCBI accession number NP_006798.1) full length (residues 1 – 185) and chromodomain (residues 19 – 71), and MPP8 chromodomain (NCBI accession number NP_059990.2, residues 58 – 112). The methyl-lysine binders were cloned in the pGADT7 vector (Clontech, Cat. No. 630442, 630489, 630491), using the EcoRI and BamHI restriction sites (Figure 10, A).

The pBridge vector backbone was modified in two cases, to include the full length ADH1 promoter (1-1472) instead of the original truncated version (1074 – 1470). In one case, the empty pBridge vector was used, and in the other, it was the pBridge-murine-G9a vector. The full length promoter was amplified by Polymerase Chain Reaction (PCR) from the pGADT7 vector. The truncated ADH1 promoter was excised from the pBridge vector backbone by PCR, resulting in linearized vector. The two fragments were fused by the Gibson assembly method (Gibson et al., 2009), using the Gibson Assembly Master Mix according to the instructions provided by the supplier (New England Biolabs, Cat. No. E2611L). The newly created vector is referred to as pBridge-ADH1-FL (Figure 11).

Figure 11 – Vector map of the modified pBridge vector, containing the ADH1 full length promoter for the expression of the Gal4-DNA-BD fusion protein. WIZ (179-359), the methylation target of G9a, was cloned in the MCS I, to be expressed as Gal4-DNA-BD fusion. A His-tag was introduced via PCR at the C-terminus of WIZ before cloning the gene into the vector. The mouse G9a (984-1263) was cloned in the MCS II. The image was made using SnapGene 2.2.2.
As described above, the G9a target WIZ was introduced in the MCS I of either pBridge empty vector or pBridge containing murine G9a in the MCS II (pBridge-mG9a), with the reverse primer containing also the sequence for a His\textsubscript{6}-tag, and using the EcoRI and BamHI restriction sites. The resulting constructs are referred to as pBridge-ADH1-FL-WIZ-His\textsubscript{6} and pBridge-ADH1-FL-mG9a-WIZ-His\textsubscript{6}.

The integrity of all the constructs was confirmed by sequencing.

3.1.2. Transformation and Mating of Yeast

In accordance with the Yeastmaker\textsuperscript{TM} Yeast Transformation System 2 User Manual (Clontech, Protocol No. PT1172–1, Version No. PROY3570), the plasmids containing the DNA-BD fusions were transformed into the MATa Y2HGold yeast strain (Clontech, Cat. No. 630498), while the AD fusion plasmids were transformed into the MATα Y187 yeast strain (Clontech, Cat. No. 630457) using a lithium acetate-based method. For the transformations, the small-scale protocol was performed. The bait plasmids were tested for autoactivation and toxicity, as described in the Matchmaker\textsuperscript{TM} Gold Yeast-Two-Hybrid System User Manual (Clontech, Protocol No. PT4084–1, Version No. PR912676). The mating of the two strains was achieved in accordance to Section VI – Control Experiments of the same protocol, both with the Clontech control plasmids, and with the individually designed plasmids instead of the company controls.

3.1.3. Protein Extraction and Immunoblotting of Yeast Extracts

Transformed yeast cells were grown in selective media specific to the transformed plasmid until they reached an OD\textsubscript{600} of 1-2 in 1.5 mL culture. The Y2HGold strain transformed with pBridge vector was grown in SD-Trp media, synthetically defined (SD) medium containing drop-out mix (DO) without tryptophan (Trp), and the Y187 strain transformed with pGADT7 vector was grown in SD-Leu media. The cells were harvested by centrifugation at 4000 revolutions per minute (rpm) for 1 minute. After the supernatant was discarded, the pellet was resuspended in 1 mL water and centrifuged again. The pellet was resuspended again in 1 mL of water, to which 150 µL Lysis Buffer (1.86 M NaOH, 1 M β-mercaptoethanol) were added. The sample was incubated on ice for 10 minutes, followed by the addition of 150 µL 50% trichloroacetic acid (TCA) and another 10 minutes of incubation. The precipitated proteins were pelleted by centrifugation at 13000 rpm for 10 minutes. The supernatant was discarded, and 100 µL of acetone were added to the pellet. The sample was centrifuged once more at
13000 rpm for 1 minute, and the pellet was dried at 50°C for one hour. Once the pellet was completely dry, it was resuspended in 40 µL SDS-PAGE running buffer and 10 µL 5x protein loading dye (10% w/v SDS, 10 mM DTT or β-mercaptoethanol, 20% v/v glycerol, 0.2 M Tris pH 6.8, 0.05% w/v bromophenol blue). The samples were boiled for 10 minutes, then stored at -20°C until they were loaded on a 12% Sodium Dodecyl Sulfate – Polyacrylamide Gel (SDS-PAGE). In the case of the Y2HGold cells transformed with pBridge-ADH1-FL-WIZ-His6 and pBridge-Adh1-FL-murine-G9a-WIZ-His6, the cultures were grown to an OD600 of 1 in 5 mL instead of 1.5 mL culture volume. Also, after the pellet was dry, the samples were resuspended in 80 µL Laemmli Buffer and 20 µL 5 x protein loading dye. After the separation on SDS-PAGE, the samples were blotted using a Trans-Blot Turbo Transfer System (BioRad) and subsequently stained with Ponceau stain (0.1% (w/v) Ponceau S, 5% acetic acid). After the destaining with water, the nitrocellulose membrane (Whatmann) containing the transferred proteins was blocked by overnight incubation at 4°C in TTBS buffer (10 mM Tris/HCl pH 7.5, 0.05% Tween-20 and 150 mM NaCl) with an addition of 5% non-fat dried milk. Then, the blot was washed three times for 10 minutes with TTBS buffer, and a 1 hour incubation at room temperature was performed with tag-specific or modification-specific antibodies: mouse anti HA-tag (Covance, Cat. No. MMS-101P, diluted 1:10000), mouse anti-Gal4 DNA-BD (Abcam, # ab135397, diluted 1:1000), mouse anti-His-tag (Sigma/GE Healthcare, # 27-4710-01, diluted 1:6000), mouse anti-tetra His (Quiagen, Cat. No. 34670, diluted 1:3000) antibodies diluted in TTBS buffer containing 1% non-fat dried milk. In the case of the anti-tetra His antibody, blocking and antibody incubations were performed in TTBS with 3% Bovine serum albumin (BSA) instead of the non-fat dried milk. Next, the membrane was washed three times 10 minutes with TTBS and incubated for 1 hour at room temperature with horseradish peroxidase conjugated secondary antibody: sheep anti-mouse antibody (GE Healthcare, # NA931, dilution 1:5000) in TTBS buffer containing 1% non-fat dried milk. Finally, the membrane was washed twice for 10 minutes with TTBS and submerged in ECL developing solution (Thermo Fisher Scientific). Images were either captured on Hyperfilm™ high performance autoradiography film (GE Healthcare), or using the Fusion-SL Advance™ (PeqLab).

3.1.4. Small Scale His-tagged Protein Purification

Yeast cells were grown in SD-/Trp or SD/-Met/-Trp until they reached an OD600 of 0.5. After harvesting approximately 450 x 10⁷ cells by centrifugation and washing them once with water, the pellet was stored at -80°C for at least 24 hours. On the purification day, the cells were
thawed on ice, resuspended in 600 µL Buffer A (6M Guanidinium-HCl, 100 mM Na2HPO4/NaH2PO4, pH 8.0, 10 mM Imidazole, 250 mM NaCl, 0.5% NP40), and mixed with 300 µL glass beads. After 10 minutes of vortexing at 4°C, the beads and cell debris were pelleted by centrifugation at 11000 x g for 10 minutes at 4°C. For each 350 µL of supernatant, 100 µL of Ni2+-NTA beads were added. The tubes were kept for two hours at 4°C with continuous rotation. After the incubation, the beads were washed twice with 500 µL Buffer A, by centrifugation. For the removal of the guanidinium chloride and protein precipitation, ethanol precipitation was used. For every 50 µL of sample, 450 µL of ice-cold ethanol (100%) was added and kept for at least 2 hours at -20°C. The sample was centrifuged at maximum speed for 15 minutes, then the supernatant was removed carefully and discarded. The pellet was washed with cold 90% ethanol, vortexed and repelleted for 5 minutes at full speed. The pellets were then dried for 20 minutes at 65°C. The pellet was then resuspended in 50 µL of Urea Buffer (8M Urea, 40 mM Tris, 0.1 M EDTA, 5% SDS, 0.05% Bromophenolblue, pH 6.8) and boiled for 5 minutes, followed by a centrifugation step at 2500 rpm for 3 minutes. The supernatant was kept and loaded on the final SDS-PAGE gel as control. The beads were then washed twice with Buffer A and three times with Buffer T (50 mM Na2HPO4/NaH2PO4, pH 8.0, 250 mM NaCl, 20 mM Imidazole, 0.5% NP40). The beads were resuspended in 100 µL Urea Buffer containing 250 mM imidazole and boiled for 15 minutes at 95°C. The beads were pelleted by centrifugation and the supernatant containing His-tagged protein was stored at -20°C until it was loaded on an SDS-PAGE gel. Immunoblotting was done according to section 3.1.3.

3.1.5. Large scale native His-tag purification from yeast cells

Cells were grown in liquid culture (SD/-Trp media) until they reached an OD600 of 0.5. A total of 3 x 10^10 cells were used as starting material for the purification. After harvesting, the cells were washed once with water, and the pellet was then frozen at -80°C at least overnight. On the day of the purification, the cells were thawed on ice, and resuspended in 20 mL of Lysis buffer (50 mM HEPES, 25 mM Imidazole, 250 mM NaCl, 1 mM DTT, 10% glycerol, 1 mM EDTA, 1 mM PMSF) containing Protease Inhibitor Cocktail per every 5 ml of pellet. The cells were disrupted mechanically using a French Press system (Thermo IEC OMFA078A, Laboratory Press), by four passes of each cell extract through the press at 20000 PSI. The complete lysis was verified by microscopy. The extracts were then centrifuged at 4°C, at 18000 rpm for 90 minutes, and the supernatants were loaded on Ni2+-NTA columns for affinity chromatography. The columns were washed with 150 mL of Lysis buffer, and then elution was done in four 500
µL samples with Elution buffer (50 mM Hepes, 300 mM imidazole, 250 mM NaCl, 1 mM DTT, 10% glycerol). The samples were subsequently passed through two rounds of dialysis, first in Dialysis Buffer I (50 mM HEPES, 150 mM NaCl, 1 mM DTT, 10% glycerol), and then in Dialysis Buffer II (50 mM HEPES, 150 mM NaCl, 1 mM DTT, 50% glycerol). The obtained purified protein samples were loaded on SDS-PAGE and the immunoblotted as described in section 3.1.3.

3.2. The PHF1 Tudor Domain Binds to Trimethylated Histone 3 Lysine 27, With a Strong Preference for the Histone H3T Variant

3.2.1. Fluorescence anisotropy measurements

The experiments were carried out according to the methods section of the publication which incorporates them (Kycia et al., 2013). The purified wild type and W41A and Y47A mutant PHF1 tudor domains were provided by Dr. Ina Kycia in interaction buffer (20 mM HEPES pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT and 10% glycerol). The fluorescence anisotropy assay was performed with the same buffer conditions, by titrating 100 µM peptide with increasing concentrations of protein domain. The measurements were recorded using a Jasco FP-8300 spectrofluorometer. The purified peptides (H3TK27me3 (18-34), H3.1K27me3 (18-34)) were fluorescein-isothiocyanate-coupled, ordered from Intavis AG (Köln, Germany). The data were fitted to a binary binding equilibrium model using the Microsoft Office Excel Solver module, and the binding constants were derived from these fits, as described previously (Dhayalan et al., 2011).

3.2.2. Cell culture

HEK293 cells were grown at 37°C and 5% CO₂ in Dulbecco’s modified Eagle’s medium with 5% fetal bovine serum until they reached approximately 90% confluence.

3.2.3. Native Histone pulldown assay

Native histones from HEK293 cells were isolated by acid extraction using a previously described method (Shechter et al., 2007). For the GST-pulldown experiment, 20 µL of glutathione sepharose 4B resin (Amersham Biosciences) were washed two times with pulldown interaction buffer (25 mM Tris-HCl pH 8.5 adjusted at 4°C, 100 mM KCl, 5 mM MgCl₂, 200 µM PMSF, 0.1% NP-40, 10% glycerol). All the wash steps were done by centrifugation at 2000 rcf for 2 minutes. Then, 7 µg of GST-tagged Tudor domain and 30 µg of native histones were incubated separately with washed beads in interaction buffer containing 5% BSA, under
rotation for 1.5 hours at 4°C. After the incubation, the beads were pelleted by centrifugation. The supernatant resulting from the incubation of the GST-fused domains was discarded, and the supernatant resulting from the pre-clearing of the histones was added to the sample containing the sepharose-bound GST-fused domains. The samples were incubated overnight at 4°C under rotation. Next, the beads were washed 5 times with washing buffer (25 mM Tris–HCl pH 8.0 at 4°C, 500 mM KCl, 5 mM MgCl2, 200 μM PMSF, 0.1% NP-40 and 10% glycerol), and once with the pulldown interaction buffer. The beads were pelleted by centrifugation at 10000 rpm for 10 minutes. The supernatant was discarded and replaced by 30 µL of 5 x SDS-PAGE loading dye diluted in SDS-PAGE running buffer, and the samples were incubated at 100°C for 10 minutes. After centrifugation, the obtained supernatants, together with a 10% native histone input sample were loaded on a 16% SDS-PAGE gel, separated by electrophoresis and transferred to a nitrocellulose membrane (Whatman). The membrane was blocked by overnight incubation at 4 °C with TTBS buffer (10 mM Tris–HCl (pH 7.5), 0.05% Tween-20 and 150 mM NaCl) containing 5% non-fat dried milk powder. The rest of the procedure was carried on at room temperature. First, the membrane was washed three times for 5 minutes with TTBS, and then incubated for one hour with TTBS buffer containing 1% non-fat milk powder and the rabbit anti-H3K27me3 antibody (Active Motif #39155, diluted 1:1000). After three washes with TTBS, the secondary antibody incubation (anti-rabbit antibody from GE Healthcare, #NA934V, diluted 1:2500) was also performed for one hour in TTBS with 1% non-fat dried milk. Next, the membrane was washed three more times with TTBS and then submerged in ECL developing solution from Thermo Fischer Scientific. The images were recorded using the Fusion-SL-3500.WL (PeqLab), with an exposure time of 10 minutes. The membrane was then washed with TTBS and then stripped according to the mild stripping protocol from Abcam (Western blot membrane stripping for restaining protocol). The verification of the successful stripping procedure was done by ECL detection with an exposure time of 30 minutes. The membrane was washed again with TTBS, then the membrane was blocked again overnight, and the procedure was repeated as described for the detection with anti-H3K27me3 antibody. This time, the primary antibody used was rabbit anti-H3K36me3 antibody (Abcam #ab9050, diluted 1:1000). The secondary antibody was the same as before. The quantification of the bands was done using the FusionCapt Advance Solo 4 software associated with the PeqLab detection system. The background subtraction was done using a rolling ball algorithm.
3.3. Characterization of the Binding Specificity of the Chromodomains of CBX2 and CBX7 Proteins

3.3.1. Site directed mutagenesis, expression and purification

The pGEX-6P-2 (GE Healthcare) constructs containing GST-fused CBX7 (res. 8-62) or CBX2 (res. 9-66) cloned using the BamHI/XhoI restriction enzymes were obtained from Dr. Ina Kycia. Site directed mutagenesis to introduced the CBX7 F11A, W32A and V13A and the CBX2 W36A mutations was performed using PCR-megaprimers as described previously (Jeltsch & Lanio, 2002). Together with the desired mutation, silent mutations introducing restriction sites were designed as part of the mutagenesis primers. Colony screening for positive mutant chromodomain constructs was done using PCR amplification of the insert and digestion with the appropriate enzyme. Mutagenesis was confirmed by DNA sequencing.

The obtained pGEX-6P-2 plasmids were transformed into *E. coli* BL21 cells (Novagen) and were overexpressed in Luria Bertani (LB) medium. The cells were grown at 37°C until they reached an OD<sub>600</sub> ~ 0.6, and then induced with 1 mM isopropyl-β-D-thiogalactoside and grown overnight at 22°C. The cells were harvested by centrifugation, washed once with STE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, and 100 mM NaCl) and the cell pellet was stored overnight at -20°C.

For the purification procedure, the *E. coli* cell pellet was thawed on ice, and resuspended in Sonication buffer (20 mM HEPES pH 7.5, 0.5 M KCl, 0.2 mM DTT, 1 mM EDTA and 10% glycerol). The resuspended cells were lysed by sonication and the clear extract was obtained by centrifugation at 18000 rpm for 90 min at 4°C. The clear cell lysate was loaded on an affinity chromatography column containing Glutathione Sepharose 4B resin (Amersham Biosciences). The column was washed with 150 mL of Sonication buffer and then the purified GST-fused

![Figure 12 - Quality of the purified GST-fusion CBX2 and CBX7 wild type and mutant proteins by SDS-PAGE and Coomassie staining.](image-url)
protein was eluted using Sonication buffer containing 40 mM reduced glutathione. The resulting eluted protein fraction was dialyzed first for two hours in 20 mM HEPES pH 7.5, 0.2 M KCl, 0.2 mM DTT, 1 mM EDTA and 10% glycerol, and then in 20 mM HEPES pH 7.5, 0.2 M KCl, 0.2 mM DTT, 1 mM EDTA and 60% glycerol overnight at 4°C. The protein purity was assessed by Coomassie stained SDS-PAGE gels (Figure 12). CBX2, both wild type and mutant, runs slightly lower than the expected size.

3.3.2. Application of MODified™ Histone Peptide Arrays for the analysis of the binding specificity of the CBX2 and CBX7 chromodomains

The MODified™ Histone Peptide Arrays were prepared by Active Motif (Cat. No. 13001) using the Celluspots method as described (Winkler et al., 2009). The binding experiments done with GST-tagged reading domains were performed as previously described (Bock et al., 2011). The peptide arrays were initially blocked overnight at 4°C in TTBS with 5% non-fat dried milk, and then washed twice for 5 minutes with TTBS. The rest of the procedure was carried out at room temperature. The third wash was done with Array Interaction Buffer (20 mM Hepes pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT and 10% glycerol). Next, the GST-fused chromodomain was incubated with the peptide array in Array Interaction Buffer for 1.5 hours. In the case of CBX7 wild type, the final concentration of the reading domain in the assay was 0.01 µM, while the concentrations used for CBX2 wild type, and the CBX7 mutants was 0.1 µM. After the incubation, the arrays were washed once with Array Interaction Buffer, and twice with TTBS. Then, the peptide array was incubated for 1 hour with primary goat anti-GST antibody (GE Healthcare #27-4577-01, diluted 1:5000), in TTBS containing 1% non-fat dried milk. The array was next washed for three times with TTBS, and then incubated for 1 hour with the secondary antibody, anti-goat antibody conjugated with horseradish peroxidase (Invitrogen #81-1620, diluted 1:12000) in TTBS with 1% non-fat dried milk. After that, the array was washed another three times with TTBS and submerged in ECL solution (Thermo Fischer Scientific) for detection. Detection was done either using the Fusion-SL-3500.WL (PeqLab) or by incubation with Hyperfilm™ high performance autoradiography films (GE Healthcare). The exposure time was 10 minutes on film with CBX7 wild type, 15 minutes on film with CBX2 wild type, and 30 minutes with Fusion-SL-3500.WL for all the mutants. The exposition times with the autoradiography film and the Fusion-SL-3500.WL are comparable. The intensities of the spots were recorded using the Array Analyze Software, which also calculated the specificity factor.
for each modification, by dividing the average of the intensities of all the spots containing the modification to the average of the intensities of all the spots not containing the modification.

### 3.3.3. Application of peptide SPOT arrays containing methyllysine analogs for the analysis of the binding specificity of the CBX2 and CBX7 chromodomains

Peptide arrays were synthesized by Dr. Srikanth Kudithipudi using the SPOT-method as described previously (Frank, 2002). An array contained four spots: H3T (21-35) wild type, H3T (21-35) K9C, H3.1 (21-35) wild type, H3.1 (21-35) K9C. The cysteines were then converted by Dr. Ina Kycia to N-methylated aminoethylcysteine, which are methyllysine analogs, according to the protocol published previously (Simon et al., 2007).

The binding assays with the CBX2 and CBX7 wild type chromodomains were performed as described in section 3.3.2. The SPOT array was used instead of the MODified™ Histone Peptide Array. The proteins were used at the same concentrations as in 3.3.2. The ECL-submerged peptide SPOT arrays were incubated with autoradiography film (GE Healthcare), for 30 seconds in the case of CBX7, and 5 minutes in the case of CBX2.

### 3.3.4. Cell culture

HepG2 cells were grown at 37°C and 5% CO₂ in Dulbecco’s modified Eagle’s medium with 10% fetal bovine serum until they reached approximately 75% confluence.

### 3.3.5. Far western blot analysis of the binding specificity of the CBX7 wild type and mutated chromodomain to recombinant histones and native mononucleosomes

Native histones were isolated from HepG2 cells by acid extraction, as described in Shechter et al., (2007). Native mononucleosomes were extracted from HepG2 cells according to a previously described method (Brand et al., 2008). A total amount of 5 μg of recombinant H3 and H4 histones histones and 5 μg nucleosomes were loaded and run on an SDS-PAGE gel and then transferred to a nitrocellulose membrane (Whatman). The membrane was blocked overnight at 4°C with TTBS containing 5% non-fat dried milk, and the same procedure as for the MODified™ Histone Peptide Array was used (Section 3.3.2). A concentration of 0.1 μM of the CBX7 wild type, and two mutants was used for the incubation with the blot. The detection was done by exposure to autoradiography film (GE Healthcare), for 5 minutes in the case of CBX7 wild type, and for 10 minutes in the case of CBX7 F11A and W32A.
3.3.6. Fluorescence anisotropy measurements

The experiments were carried out according to the methods section 3.2.1. The purified wild type CBX2 and CBX7, together with the CBX7 F11A and W32A mutants and the CBX2 mutant W36A were purified as described in section 3.3.1, but dialyzed the second time in interaction buffer (20 mM HEPES pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT and 10% glycerol). The titration of 100 nM peptide was done with increasing concentrations of recombinantly purified protein. The purified fluorescein-isothiocyanate-coupled peptides (H3TK27me3 (18-34), H3.1K27me3 (18-34) and H3K9me3 (1-19)) were provided by Intavis AG (Köln, Germany).

3.3.7. Native mononucleosomal pulldowns with domains (CIDOP) or antibody (ChIP), coupled to quantitative PCR (qPCR) experiments

Native HepG2 mononucleosomes were isolated as described previously (Brand et al., 2008) by digestion of the nuclei with micrococcal nuclease. After the digestion, the samples were centrifuged at 13000 x g for 10 minutes, and the collected supernatant contained the soluble nucleosomal fraction, which was frozen with liquid nitrogen and stored at -80°C until further need.

For each pulldown sample 20 μL of glutathione sepharose 4B beads (GE Healthcare) or protein G magnetic Dynabeads (Invitrogen) were prepared by washing three times with 1 mL DP buffer (16.7 mM Tris-Cl, 167 mM NaCl, 1.1% Triton X-100, 1.2 mM EDTA and protease inhibitors) and the buffer was exchanged by centrifugation at 2000 x g for 2 minutes and the removal of the supernatant. The beads were used to pre-clear the native chromatin (60 μg for CBX2 and 30 μg for CBX7 and the H3K27me3 antibody, based on DNA absorption at 260 nm), for 1 hour at 4°C with rotation, in DP buffer.

Next, the pre-cleared nucleosomal solution was removed from the beads and incubated in solution with the GST-tagged protein (60 μg CBX2 or 30 μg CBX7) or anti-H3K27me3 antibody (Active Motif, #39155) (used according to the recommendations of the manufacturer) in the same buffer overnight at 4°C with rotation.

The following day, the complexes formed by the domain/antibody with chromatin were incubated with the corresponding beads, either glutathione sepharose 4B (GE Healthcare) or protein G magnetic Dynabeads (Invitrogen) for 1.5 hours at 4°C with rotation. Next, the beads were washed once with Low Salt Buffer (20 mM Tris-Cl, 150 mM NaCl, 1% Triton X-100, 0.1% SDS and 2 mM EDTA), once with High Salt Buffer (20 mM Tris-Cl, 500 mM NaCl, 1% Triton X-
100, 0.1% SDS and 2 mM EDTA), once with LiCl buffer (10 mM Tris-Cl, 250 mM LiCl, 1% NP-40, 1% DOC and 1 mM EDTA) and twice with TE buffer. Each wash involved a 10 minute incubation step with rotation at 4°C, a centrifugation step at 2000 x g for 2 minutes, and discarding the supernatant. Elution of the complexes was done with Elution buffer (50 mM Tris-Cl, 50 mM NaCl, 1 mM EDTA, 1% SDS) for 45 minutes with rotation at room temperature, followed by Proteinase K digestion and recovery of DNA using the ChIP DNA purification columns from Active Motif.

All the quantitative PCR experiments were performed using a CFX96 Real-Time detection system (Bio-Rad). The reaction mixes were done with SsoFast EvaGreen supermix (Bio-Rad). A standard curve representing known percentage of input was generated in order to estimate the percentage of precipitated DNA and the efficiency of each primer set was verified. The primer sequences that were used in this study are shown in Table 2.

Table 2 - Sequences of the primers used for the qPCR experiments

<table>
<thead>
<tr>
<th>Target Modification</th>
<th>Gene Name</th>
<th>Sequence of qPCR primers</th>
</tr>
</thead>
</table>
| H3K9me3             | Sat α     | FP 5`-ATC GAA TGG AAA TGA AAG GAG TCA-3’  
|                     |           | RP 5’-GAC CAT TGG ATG ATT GCA GTC A-3’  |
|                     | Gene desert chr. 12 | FP 5`-GCT GTT ACT TTT TAC AGG GAG TTT TTA-3’  
|                     |           | RP5’-ATA AAG CAG GTA AAG GTC CAT ATT TC-3’  |
| H3K27me3            | Hox11     | FP 5’-TTT CCA CAG CCT TTG CAG GCC-3’  
|                     |           | RP 5’-TCC AGG CTG CAA GAA GAA GCG GAG-3’  |
|                     | WNT2      | FP 5’-GTT GCA AGG AAA TTA CAG GGC-3’  
|                     |           | RP 5’-TCC ATC TGC CGA CTG TTG TTG G-3’  |
|                     | CAV1      | FP 5’-AAACGGACCCCTAAACACC-3’  
|                     |           | RP 5’-TTGGTGCCTTGGCTTACC-3’  |
|                     | Serping1  | FP 5’-CTT CTG CCC AGG ACC TGT TA-3’  
|                     |           | RP 5’-GCT GAG AAG GCC TGG TAG AG-3’  |
|                     | MyoD      | FP 5’-CCG CCT GAG CAA AGT AAA TGA-3’  
|                     |           | RP 5’-GGC AAC CGC TGG TTT G-3’  |
|                     | EN1       | FP 5’-CAG AGG CCA GGA TCG CAT -3’  
|                     |           | RP 5’-GCT AGG GAG AGC AAA CCC G-3’  |
| H3K36me3            | VEGFA     | FP 5’-GCC CTA ACC CCA GCC TTT GTT T-3’  
|                     |           | RP 5’-GTA CGT ATC GTG TCT TAT CAG TCT TTC C-3’  |
|                     | PABPC1    | FP 5’-CATCACTCCCAAGAAAATTAGTGGAGT-3’  
|                     |           | RP 5’-CACAGTCTCTAAAGCCCAA-3’  |
3.4. Investigation of the Binding Specificities of the MLL2 (KMT2D) and MLL3 (KMT2C) PHD Fingers

3.4.1. Cloning, expression and purification

The PHD fingers of MLL2 (KMT2D, RefSeq ID: NP_003473.3) and MLL3 (KMT2C, RefSeq ID: NP_733751.2) were amplified from HEK293 cDNA by PCR and cloned as described in Table 3 and Table 4 in the pGEX-6P-2 vector (GE Healthcare), using the BamHI and XhoI restriction sites, with the exception of MLL3 PHD0, MLL3 PHD1, MLL3 PHD3 and MLL3 PHD 1-3, which were cloned using the EcoRI and XhoI restriction sites. In the case of some domains, boundaries were optimized due to low concentration after purification procedure. The boundary changes are also documented in Table 3 and in Table 4.

Table 3 - The domain borders designed for the cloning of the MLL2 PHD fingers

<table>
<thead>
<tr>
<th>Domain</th>
<th>Initial boundaries</th>
<th>Redesigned boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHD0-2</td>
<td>Ser 169- Glu 341</td>
<td>Ser 169 – Ser 330</td>
</tr>
<tr>
<td>PHD1-2</td>
<td>Glu 225 – Glu 341</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD3-5</td>
<td>Asp 1358 – Leu 1572</td>
<td>Gln 1377 – Leu 1572</td>
</tr>
<tr>
<td>PHD0</td>
<td>Ser 169 – Glu 226</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD1</td>
<td>Glu 225 – Lys 281</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD2</td>
<td>Glu 272 – Ser 330</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD3</td>
<td>Asp 1358 – Ser 1437</td>
<td>Gln 1377 – Ser 1437</td>
</tr>
<tr>
<td>PHD4</td>
<td>Cys 1424 – Gln 1492</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD5</td>
<td>Ala 1502 – Leu 1572</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD6</td>
<td>His 5059 – Lys 5139</td>
<td>Leu 5089 – Glu 5145</td>
</tr>
</tbody>
</table>
Table 4 - The domain boundaries designed for the cloning of the MLL3 PHD fingers

<table>
<thead>
<tr>
<th>Domain</th>
<th>Initial boundaries</th>
<th>Redesigned boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHD0-3</td>
<td>Ser 281 – Glu 532</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD1-3</td>
<td>Ser 338 – Glu 532</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD4-6</td>
<td>Asp 951 – Asp 1149</td>
<td>Asn 956 – Asn 1144 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asn 956 – Asp 1149 (2)</td>
</tr>
<tr>
<td>PHD0</td>
<td>Ser 281 – Arg 337</td>
<td>Ser 281 – Gln 333</td>
</tr>
<tr>
<td>PHD1</td>
<td>Ser 338 – Gln 392</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD2</td>
<td>Glu 387 – Arg 444</td>
<td>N/A</td>
</tr>
<tr>
<td>PHD3</td>
<td>Lys 431 – Glu 532</td>
<td>Glu 463 – Asp 525</td>
</tr>
<tr>
<td>PHD4</td>
<td>Asp 951 – Glu 1011</td>
<td>Asn 956 – Glu 1011</td>
</tr>
<tr>
<td>PHD5</td>
<td>Glu 1006 – Tyr 1075</td>
<td>Glu 1006 – Thr 1063</td>
</tr>
<tr>
<td>PHD6</td>
<td>Gln 1072 – Asp 1149</td>
<td>Ser 1083 – Asp 1149 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ser 1083 – Asn 1144 (2)</td>
</tr>
<tr>
<td>PHD7</td>
<td>Lys 4401 – Glu 4515</td>
<td>Gln 4459 – Glu 4515</td>
</tr>
</tbody>
</table>

The success of the cloning was verified by insert release using restriction digestion, and then by sequencing. The constructs were overexpressed and purified as described in section 3.3.1, with one change. The overexpression was performed in the presence of 50 μM ZnSO₄, because the PHD fingers contain Zn²⁺ ions in their structure. The quality of the protein was assessed by SDS-PAGE and Coomassie staining (Figure 13, A and B).
3.4.2. Native histone pulldown with recombinant GST-tagged MLL2 PHD 3-5 and MLL3 PHD 4-6

HEK cells were cultured according to section 3.2.2. Native histones were isolated using an acid-based extraction method described previously (Shechter et al., 2007). The same pulldown...
protocol as described in the case of PHF1 tudor domain was used (section 3.2.3). Both for MLL2 PHD 3-5 and for MLL3 PHD 4-6, 30 μg were used, while GST was used in excess. Also, 30 μg of native histones were used. The histones were pre-cleared in interaction buffer containing 5% BSA, while the GST-tagged domains and GST were incubated with the glutathione sepharose 4B beads (GE Healthcare). Afterwards, the histones were added to the domains that were bound to the beads, and incubated overnight at 4°C with rotation. The washing steps were done with the same 100 mM KCl-containing interaction buffer. The samples were run on SDS-PAGE gels and transferred to nitrocellulose membranes, followed by antibody stainings. In this case, first, the staining for H4 pan acetyl marks was performed. The first antibody, rabbit anti-H4 pan acetyl (Active Motif #39243) was diluted 1:3000. The secondary antibody for all the primary antibodies used in this experiment was anti-rabbit coupled to horseradish peroxidase (GE Healthcare, #NA934V, diluted 1:3000). After detection, and before another staining was done, the blot was stripped according to the Abcam mild antibody stripping protocol (Western blot membrane stripping for restaining protocol), and the success of the stripping was verified by incubation of the blot with ECL and image detection for 30 min to 1 hour. Next, H4K16ac detection was tested with primary rabbit anti-H4K16ac antibody (Millipore 06-761, diluted 1:1:1000), followed by the same secondary antibody as before. After detection and stripping, H3K9me3 was detected with rabbit anti-histone 3 (tri methyl K9) antibody (Abcam, # ab71999), used in 1:3000 dilution. Detection in all of the cases was done with the Fusion-SL-3500.WL (PeqLab), with exposure times of 2-5 minutes using the ECL Plus reagents from Perkin Elmer (Product number NEL103001EA). The quantitative analysis of the lanes was done using the FusionCapt Advance Solo 4, with background subtraction by a rolling ball mechanism.

3.4.3. Site directed mutagenesis

The site directed mutagenesis of the constructs containing MLL2 PHD 3-5 fragment sequence was performed according to (Jeltsch & Lanio, 2002) using a PCR-megaprimer containing the mutation in a rolling PCR reaction with the wild type vector as the template. Silent mutations in the primer were used to introduce or remove restriction sites, in order to be able to discriminate between wild type and mutant plasmids in the colony screening, done by PCR followed by restriction digestion and agarose gel electrophoresis. Positive clones had the plasmid purified and sent for sequencing. Site directed mutagenesis was successful for all the mutants, except for the Y1404A, which contained additional mutations and deletions, and for
the V1425A mutant, which contained a double V1425A/E1431A. The latter was still used further in the experiments, as both the mutations were designed for the same purpose. The GST-tagged mutant proteins were expressed in *E. coli* BL21 cells and purified as described in section 3.3.1.

### 3.5. Study of the Specificity of Various SET Domain Methyltransferases

#### 3.5.1. Cloning, expression and purification of the SET-domains

The DNA fragments corresponding to the SET domain, including all the SET-domain associated conserved domains, such as the Pre-SET, Post-SET or AWS domains, were amplified from total cDNA of HEK293 cells. The SET domains cloned were the following: EZH1 607 – 747 (NCBI RefSeq ID: NP_001982), MLL2 residues 5346-5537 (NCBI RefSeq ID: NP_003473.3), SET1A residues 1567-1707 (NCBI RefSeq ID: NP_055527.1), SET1B residues 1783-1923 (NCBI RefSeq ID: NP_055863.1), NSD3 residues 1077 – 1299 (NCBI RefSeq ID: NP_060248.2) and MLL5 residues 312-463 (NCBI RefSeq ID: NP_061152.3). The amplified DNA fragments were introduced into the multiple cloning site of the pGEX-6P-2 vector (GE Healthcare) by restriction digestion with BamHI and XhoI, with the exception of SET1A, for which EcoRI and XhoI were used, and subsequent ligation. Colony screening was done by PCR amplification of the vector MCS with vector specific primers, and analysis of the fragment size by agarose gel electrophoresis. Positive clones were confirmed by insert release with specific restriction enzymes, and after that, by sequencing. The constructs were then transformed into *E. coli* BL21 cells and overexpressed and purified according to section 3.3.1, with the overexpression performed in media containing 50 μM ZnSO₄, as Zn²⁺ ions are needed for the proper enzymatic activity. The quality of the purifications was checked by SDS-PAGE stained with Coomassie (Figure 14).
3.5.2. *In vitro* methyltransferase assay with recombinant H3 substrate

All the recombinant H3.1 methylation reactions were performed in the same methylation buffer (50 mM Tris-HCl pH 9.0, 20 mM KCl, 10 mM MgCl$_2$, 5 mM DTT). In a 40 μL reaction volume containing 1.5 μg H3.1 and 0.76 μM tritium-labeled SAM (2.7 TBq/mmol; 51 Perkin Elmer), protein SET domains were added to a final concentration between 5 and 10 μM. The reactions were incubated for 3-4 hours at the specified temperature (either 37°C or 25°C). In the case of the methylation reaction performed by Sara Weirich, with the MLL3 SET, MLL2 SET, SET1A SET and SET1B SET in the presence or absence of complex, the buffer conditions were different (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 5 mM MgCl$_2$ and 3 mM DTT) and the reaction was performed at 25°C. For the samples that contained the complex, WDR5, RBBP5 and ASH2L were present in equimolar amounts. After the methylation reaction was incubated, 10 μL of 5x SDS-PAGE loading dye were added to the reaction mix, and the samples were boiled for 5 minutes. Next, they were loaded on a 16% SDS-PAGE gel. After the electrophoresis, the gel was washed once with water, and then soaked in Amplify NAMP100V solution (GE Healthcare) for 40 minutes. Then, the gel was dried with vacuum at 65°C, and then incubated with Hyperfilm™ high performance autoradiography films (GE Healthcare) at -80°C in the dark, for 1 to 10 days, depending on the signal intensity.

3.5.3. Peptide SPOT array synthesis

The peptide arrays used in this project were synthesized using the SPOT method (Frank, 2002) by Dr. Srikanth Kudithipudi. The success of the synthesis was confirmed by bromophenol blue staining of the membranes.
3.5.4. Methylation of peptide SPOT arrays for the determination of the specificity profile of SET domain methyltransferases

The same protocol was followed for the methylation of peptide SPOT arrays and of MODified Histone Peptide Array. The membranes were initially equilibrated in methylation buffer (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 5 mM MgCl₂ and 3 mM DTT) for 10 minutes, at room temperature, on a shaker. Then, the buffer was replaced with the reaction mix, containing 0.38 μM SAM (2.7 TBq/mmol; 51 Perkin Elmer) and between 3.8 μM of MLL2 SET or 8 μM SET1A SET, in the same buffer conditions. The methylation reaction was incubated for 2 hours at room temperature with gentle shaking. Then, the membrane was washed 5 times for 5 minutes with Wash buffer containing 100 mM Ammonium bicarbonate and 10% SDS. Next, the membrane was incubated for 5 minutes in Amplify NAMP100V solution (GE Healthcare) and then incubated with high performance autoradiography films (GE Healthcare) at -80°C in the dark, for 1 day in the case of MLL2 alone, 8 hours in the case of MLL2 in complex, and 18 hours in the case of SET1A. After the films were developed and scanned, the images were analyzed with the Phoretix Array software. The background was subtracted using the rectangle method, in which a rectangle containing only background is selected (an area on the membrane without any spots). Then, the intensities of the spots were quantified and retrieved in Microsoft Office Excel format. Further normalization was achieved by presenting all data relative to the minimum and the maximum intensity spot in each array.

\[
\text{Spot intensity (normalized)} = \frac{(\text{Spot intensity (raw)} - \text{Max intensity})}{(\text{Spot intensity (raw)} - \text{Min intensity})}
\]

The specificity profile was presented as an array where each spot was represented by one box in the same spatial arrangement as the peptide spot array. The boxes were color coded by conditional forming with triple color scale in Microsoft Office Excel. The bar diagram representing the specificity profile of each enzyme was analyzed as described previously (Rathert et al., 2008, Supplementary material). The incubation times and the concentrations of SET domains used for this experiment were chosen such that there would not be full turnover of the substrate, and therefore the intensity of each spot reflects the rate of the enzyme on that particular substrate. The assay offers the advantage that the substrates are methylated in competition, meaning that all the substrates are methylated in the same conditions. The relative rates of the methylation reaction correspond to \( \frac{K_{\text{cat}}}{K_d} \) rates for each peptide, a ratio that is considered an established parameter for enzyme specificity. The
discrimination factor for peptide recognition that is plotted in the bar diagram of the specificity profile was calculated to represent the relative contribution of each amino acid at each position (Rathert et al., 2008, Supplementary material). A discrimination factor of 5, for example, at a given position, signifies that the peptide which contains that amino acid at that site is methylated 5 fold faster than the average of all other peptides carrying all the other amino acids at that site. In the case that an amino acid could not be replaced with any other one, such as the case of K4, which is the target lysine, the discrimination factor would be calculated by division by 0 (or a value close to 0), resulting in error or a very high number. Therefore, an arbitrary value was given for the appropriate representation of these irreplaceable residues in the bar graph. For MLL2 alone, the discrimination factor for K4 was 30, for MLL2 in complex it was 35, and for SET1A it was 20. In the case of SET1A, R2 was also irreplaceable, and its discrimination factor was given an arbitrary value of 20, just like for the K4 position.


3.6.1. Cloning and site directed mutagenesis of the SUV39H1 constructs

The human SUV39H1 (NCBI RefSeq ID: NP_001269095.1), both as full length and chromodomain fragments residues 43-95 or 43-106 were amplified from HEK293 cell cDNA by PCR with specific primers, and cloned into pGEX-6P-2 by restriction digestion with BamHI and XhoI. Colony screening was performed by amplification by PCR of the MCS of the vector with vector-specific primers, and verification of the sizes of DNA fragments amplified by agarose gel electrophoresis. Positive clones were checked again by insert release after restriction enzyme digestion, and confirmed by sequencing. Site directed mutagenesis to introduce the Y67A and W64A mutations both in the full length and in the chromodomain (43-95) was done by site directed mutagenesis using PCR-megaprimers containing the desired mutation and additionally silent mutations either introducing or erasing restriction sites, to allow for screening. Mutants were screened by PCR amplification of the insert, and subsequent restriction digestion to differentiate between the wild type and mutant inserts. The results were also confirmed by sequencing. For the cellular localization studies, the full-length SUV39H1 inserts, both wild type and the two mutants, were subcloned in the pECFP-C1 and pEYFP-C1 vectors (Clontech Laboratories Inc.) using the XhoI and SalI restriction sites. The
screening was done as described previously, and the results were confirmed by sequencing. The YFP fused - full length SUV39H1 H324L inactive mutant mammalian construct was kindly provided by Prof. Dr. Yoichi Shinkai (Kyoto University).

3.6.2. Protein expression and purification

The constructs were transformed in *E. coli* BL21 cells and overexpressed as described previously in section 3.3.1, with some changes in the case of the full length enzyme. The full length GST-tagged SUV39H1 was expressed in *E. coli* BL21 (DE3) pLysS cells, induced at OD$_{600}$ of 0.9 and expressed overnight at 28°C. Expression in a 2 L culture was performed in the case of SUV39H1 full length proteins, compared to a standard 500 mL culture. Purification was performed according to the previously described protocol (section 3.3.1). The purified proteins were checked on SDS-PAGE gels stained with Coomassie (Figure 15). The concentration of the full length SUV39H1 enzyme was between 4 and 10 μM after purifications.

The SUV39H1 full length protein was also checked by western blot as described previously (section 3.2.3), with goat anti-GST antibody (GE Healthcare #27- 4577- 01, diluted 1:5000) as the primary antibody, and anti-goat antibody coupled to Alkaline Phosphatase (Dianova, Cat no. 305-055-003, diluted 1:10000). For detection, 5-bromo-4-chloro-3-indolyl phosphate (BCIP, X-phosphate, XP)/nitro blue tetrazolium chloride (NBT) (Roche) was used. The membrane was submerged in the BCIP/NBT solution (prepared according to the manufacturer’s instructions) for approximately 30 minutes, then the membrane was washed with water and the image was scanned (Figure 15, B).

The GST-tagged SUV39H1 SET domain (residues 82-412) was purified by Dr. Srikanth Kudithipudi, from a pGEX-2T construct kindly provided by Prof. Dr. Xiaodong Cheng, from the Emory University School of Medicine, Atlanta, USA. The concentration of this protein was 23 μM (Figure 15, C). The full length Clr4 protein was kindly offered by the laboratory of Prof. Dr. Danesh Moazed (Department of Cell Biology, Harvard Medical School, Boston, USA).
3.6.3. Application of MODified™ Histone Peptide Arrays for the analysis of the binding specificity of the SUV39H1 chromodomain

The binding assay with the SUV39H1 chromodomain (res. 43-95) was performed according to section 3.3.2, with one modification. The concentration of both the SUV39H1 chromodomain wild type and of the Y47A mutant was 1 μM. In the case of the SUV39H1 wild type chromodomain, the array submerged in ECL reagents (Thermo Fischer Scientific) was exposed to the autoradiography film for 20 minutes. The Y67A mutant was exposed using the Fusion-SL-3500.WL for 30 minutes.

Figure 15 - Images of the SUV39H1 proteins on SDS-PAGE gels stained with Coomassie or western blot for GST-tag detection. The expected protein sizes are marked on the gels with black asterisks. A. Purified SUV39H1 chromodomain (CD), with two different boundaries: 43-95 and 43-106. B. On the left, Coomassie staining of SUV39H1 full length wild type protein, run next to the same Prestained marker as on the right, where the western blot image of the same protein is shown. Detection is performed against the GST tag with secondary antibody coupled to AP, and color reaction upon incubation with BCIP/NBT. C. SDS-PAGE image stained with Coomassie, comparing the SUV39H1 full length wild type purified protein with the SUV39H1 SET domain protein. Recombinant H3 is also loaded on the left lane, next to the marker. D. Comparison of the SUV39H1 full length wild type with the SUV39H1 full length Y67A mutant, both as GST-fusions, on SDS-PAGE gels stained with Coomassie.
3.6.4. Fluorescence anisotropy measurements for the characterization of the binding specificity of the SUV49H1 chromdomain

The experiments were carried out as described in the methods section 3.2.1. The purified wild type and mutant SUV39H1 chromdomain (43-95) were purified as described in section 3.3.1, but dialyzed the second time in anisotropy assay interaction buffer (20 mM HEPES pH 7.5, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT and 10% glycerol). A concentration of 100 µM of each peptide was titrated with increasing concentrations of recombinantly purified protein. The purified fluorescein-isothiocyanate-coupled peptides (H3K9me3 (1-19) and H3K9me3/S10P (1-19)) were provided by Intavis AG (Köln, Germany).

3.6.5. Far western blot analysis of the binding specificity of the SUV39H1 wild type chromdomain to native mononucleosomes

Native nucleosomes were extracted from HEK293 cells according to a previously described method (Henikoff et al., 2009). The nucleosome insoluble fraction obtained from approximately 3 million cells in the 150-600 mM salt extraction sample were loaded and run on an SDS-PAGE gel and then transferred to a nitrocellulose membrane (Whatman). The membrane was blocked overnight at 4°C with TTBS containing 5% non-fat dried milk, and the same procedure as for the MODified™ Histone Peptide Array was used (Section 3.3.2). A concentration of 0.1 µM of the SUV39H1 chromdomain wild type, or GST protein, was used for the incubation with the blot. Immunodetection against the GST-tag was performed as described previously (Section 3.3.2). The detection was done by exposure to autoradiography film (GE Healthcare), for 2 minutes.

3.6.6. Native nucleosome pulldown with SUV39H1 chromdomain

The insoluble nucleosomal fraction was obtained by 100-600 mM salt extraction as described previously (Henikoff et al., 2009) from approximately 4 x 10⁵ HEK293 cells for each pulldown sample. The pulldown was performed as described in the native histone pulldown protocol (section 3.2.3), with some minor changes. The SUV39H1 chromdomain and GST were incubated with the beads at a concentration of 0.1 µM, in 300 µL of pulldown interaction buffer containing 5% BSA. At the washing steps, three different salt concentrations were used. For one set of samples, washing was done with pulldown interaction buffer with 100 mM KCl, just as in section 3.2.3, for another set of samples, the buffer contained 200 mM KCl, while for
the last set of samples, the buffer contained 300 mM KCl. After the samples were run on SDS-PAGE gel, the gel was only stained with Coomassie and visualized.

3.6.7. *In vitro* methyltransferase assay with recombinant H3 substrate

The methyltransferase assay was performed as described previously in section 3.5.2, with minor changes. Due to low concentrations of purified full length enzymes, only approximately 180 nM SUV39H1 wild type and Y67A mutant were used in the reaction, while the SUV39H1 SET domain, which is very active, was used in approximately 30 nM concentration. The methylation reaction was done in methylation buffer (50 mM Tris pH 8.5, 5 mM MgCl₂ and 4 mM DTT) at 25°C for 6 hours. The autoradiography films were exposed to the dried SDS-PAGE gels containing the tritium labeled methylated H3 protein for 1 week.

3.6.8. *In vitro* methyltransferase assay with reconstituted chromatin

The reconstituted chromatin was kindly provided by Prof. Dr. Wolfgang Fischle (Max Planck Institute for Biophysical Chemistry, Göttingen). The 12 x 200 x 601 tamplate was used for the construction of the oligonucleosomes, according to the previously published method (Huynh et al., 2005). The unmodified reconstituted chromatin consisted of oligonucleosomes formed of octamers and DNA in a 1:1 ratio, diluted in 10 mM EDTA pH 7.5, 20 mM NaCl. The modified reconstituted chromatin consisted of DNA and octamers in a 1:1 ratio, where 50% of the octamers contained H3K9me3 and 50% of the octamers contained unmodified octamers. Nucleosome amounts were compared to 1.5 μg of recombinant H3 by SDS-PAGE, so that comparable amounts of H3 substrate would be present in the assay, as in the previously described methyltransferase assay with recombinant H3 substrate. The methylation reaction was carried out according to the description in section 3.5.2 and section 3.6.7. The autoradiography films were exposed to the dried gels for 1 week before being developed.

3.6.9. *In vitro* methyltransferase assay using peptide arrays containing two different peptides in one SPOT

Peptide arrays containing either one peptide or a mixture of two different peptides in one spot were custom ordered and provided by Active Motif. The methylation of the arrays was performed according to the protocol described previously (section 3.5.4). The buffer used for the equilibration of the membrane and the methylation reaction contained 50 mM Tris-HCl pH 8.5, 5 mM MgCl₂ and 4 mM DTT. The reaction took place at room temperature, with gentle shaking. Approximately 200 nM SUV39H1 wild type and mutant enzymes and 1 μM of Clr4
were used in each assay, with film exposure times of 1 week for the wild type and mutant SUV39H1, and 6 hours in the case of Clr4.

### 3.6.10. Cell culture

Immortalized SUV39H dn MEFs, kindly given by Prof. Dr. Thomas Jenuwein (Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany), were grown in DMEM supplemented with 10% FBS, L-glutamine, sodium pyruvate, non-essential amino acids and 100 µM β-mercaptoethanol.

### 3.6.11. SUV39H1 localization studies in SUV39H dn MEF cells

SUV39H dn MEF cells were grown in 6 well plates containing 22 mm glass cover slips until they reached 30-40% confluency, and then they were transfected using the Xfect™ mESC Transfection Reagent (Clontech Laboratories Inc., Cat. No. 631320). The transfection was carried out following the manufacturer-provided instructions. In the case of the single transfections, 1.5 µg of each plasmid were transfected per well, while for co-transfections, 3 µg of total DNA was transfected per well, containing a mixture of 1.5 µg of each transfected plasmid. The cells were then grown for one or two days after the transfection, and then they were fixed and prepared for visualization. First, the cells were washed three times for 5 minutes with PBS (1 mM KH₂PO₄ pH 7.4, 155 mM NaCl, 3 mM Na₂HPO₄-7xH₂O) at 37°C. Then the cells were fixed using 3.7% paraformaldehyde in PBS, for 12 minutes at room temperature. After the fixation, the wells were washed 3 times for 5 minutes at room temperature, and if no antibody staining was necessary, the permeabilization and antibody incubation steps would be skipped. For the cells that were stained with anti-H3K9me3 antibody, permeabilization was done at this step with 0.2% Triton-X-100 in PBS for 10 minutes at 37°C. After another three washing steps, the cells were incubated with blocking solution containing 3% non-fat dried milk in PBS for 60 minutes at 37°C. Next, the wells were washed twice for 5 minutes with PBS containing 0.1% non-fat dried milk, and the cover slips were transferred to a moisturized incubation chamber. Each slide was incubated with 250 µL PBS with 0.1% nonfat dried milk containing rabbit anti-H3K9me3 (Abcam, ab8898, diluted 1:2500). The incubation was done overnight at 4°C. After that, the cover slips were washed 4 times for 5 minutes with PBS containing 0.1% non-fat dried milk, and then the secondary antibody incubation was done in the same buffer conditions, at 37°C for 60 minutes, with goat anti-rabbit Chromeo642 antibody (Active Motif, Cat. No. ATM15044, diluted 1:2000). After another 3 washes with PBS,
the cover slips were ready for mounting, using Mowiol. The Mowiol was allowed to dry for two hours, and then the microscopy slides were cleaned with ethanol and stored at 4°C. The microscopy images were recorded using a Zeiss LSM 510 Meta instrument (software version 3.0) and 40x objective. An argon laser line at 458 nm was used to excite EYFP fluorescence and an LP 475 filter was used for image recording. In the case of CFP fluorescence, the argon laser 514 nm was used for excitation and an LP 530 was used for image recording. When the images contained both YFP and Chromeo 642 fluorophores, YFP was excited with the argon laser at 514 nm and BP530-600 filters were used for detection, while the Chromeo 642 fluorophore was excited at 633 nm and detected with the LP 650 filter.
4. Results

4.1. Yeast-3-Hybrid Approach to Identify PKMT Target Binding Partners

Given that so many PKMTs have been proven to be involved in the methylation of various other substrates in addition to histones, it is imperative to get a better understanding of what each of these methylation events signals. Moreover, once the PKMT non-histone targets are identified, the function of the methylation signal is not trivial to unravel. There are two main hypotheses for the effects of the modification: direct effects on the structure and/or activity of the target protein; or indirect effects caused by changes in the interactome of the target protein. One way to test these possibilities is either to overexpress or to deplete the PKMT and observe the resulting consequences (Chuikov et al., 2004; Huang et al., 2006; Kurash et al., 2008; Wang et al., 2013). With respect to gain or loss of interactions due to the methylation, pulldowns of either endogenous or recombinantly expressed PKMT targets can be performed for the identification of interactors in any of the situations. Such was the case of the double interaction of the MPP8 dimers with the automethylated GLP on one side, and the Dnmt3aK44me2 on the other side, resulting in the formation of a tetramer (Chang et al., 2011). However, the limiting step for this approach is the identification of interactors, which can show low expression in cells. Many nuclear proteins with functions in chromatin regulation have a high turnover rate and low overall abundance, and they will escape identification through mass spectrometry approaches (Rathert et al., 2008).

Yeast-2-Hybrid is a method used commonly for identification of target protein binding partners. The method offers the possibility to test interaction of the target protein with a protein library for the identification of novel interactors. However, there has been no attempt to use yeast hybrid screening to identify the interactors of the already identified PKMT target proteins, especially not in a methylation-dependent manner.

A Yeast-3-Hybrid system (Clontech) is commercially available, but needs optimization for this particular purpose. In theory, a screening method can be designed, in which the three members of the Yeast-3-Hybrid system would be the PKMT, its non-histone target fused to the GAL4 DNA-binding domain (DNA-BD) representing the “bait”, and a library of human proteins fused to the GAL4 activation domain (AD) representing the “prey”. The library can be
replaced by a set of known reading domains that might bind methylated non-histone proteins (Figure 16).

Figure 16 - Schematic representation of the Yeast-3-Hybrid system approach developed here to identify methyl-specific interactions between the PKMT non-histone targets and putative interactors. There are two possible types of methylation-dependent interaction: A. The PKMT target only interacts with its binding partner if it is methylated. In this case, activation of the reporter genes will only occur in the presence of the PKMT. B. The PKMT target loses binding to its partner due to the methylation. This results in activation of the reporter genes in the absence of the PKMT, and their repression when the MTase is expressed.

In our setup, the PKMT and its non-histone target are expressed from the same vector (pBridge), while the library-AD fusions are carried by another vector. Two different strains of *Saccharomyces cerevisiae*, Y2H Gold, of mating type MATα, and Y187, of mating type MATα, are transformed with the PKMT/PKMT Target-DNA BD construct, and the Reading domain/or Library-AD fusion construct, respectively. Transformants are selected using synthetically defined (SD) growth media selection (e.g. SD/-Trp for the bait, and SD/-Leu for the prey) and then mated in media that is selective for both plasmids (e.g. SD/-Leu/-Trp) (Figure 17). After the mating is allowed for a few hours in liquid culture, the yeast is plated on double, triple or quadruple selective media, which tests for the activation of one to four of the reporter genes in the system (e.g. SD/-Leu/-Trp/-His/X-α-Gal selects for presence of both plasmids, and also activation of the HIS3 and MEL1 reporter genes, respectively) (Figure 17). After at least two rounds of re-streaking and selection of positive blue colonies, the plasmids are isolated and sent for sequencing in order to identify the bait-interacting proteins (Figure 17).
In order to verify the validity of the assay, control experiments were performed with the full length HP1β and the chromodomains of HP1β and MPP8, cloned as Gal4-AD fusions, instead of the human protein library. This strategy was based on previous results showing that HP1β chromodomain binds strongly to methylated peptides from the G9a non-histone targets CDYL1, WIZ and G9a N-terminal domain (NTD) (Rathert et al., 2008), and MPP8 binds \textit{in vivo} to the G9a-methylated NTD of Dnmt3a (Chang et al., 2011).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure17.png}
\caption{Schematic representation of the main steps in the Yeast-3-Hybrid assay procedure. The baits are cloned into the vector and transformed in the Y2H strain. Preliminary tests for autoactivation and toxicity are done. The expression of the baits can be verified. The transformed yeast culture carrying the bait is added to a culture of Y187 strain transformed with the prey library. The mating is allowed over several hours, then cells are plated on multiple plates containing selective media: DDO/X/A (Double Drop-Out SD/-Trp and SD/-Leu to select for the two plasmids; X – X-α-Gal to show blue colonies where interaction happens; A – Aureobasidin A antibiotic, to select against the colonies where interaction does not take place). Finally, after 2-3 days, blue colonies are picked and re-streaked on more selective media: QDO/X/A (Quadruple drop-out media selecting for the two plasmids and for the activation of two reporter genes, X-α-Gal and Aureobasidin A). The verified blue colonies can then be picked, and the plasmids can be isolated and sent for sequencing for the identification of the interacting prey protein. The image was adapted from the Matchmaker\textsuperscript{TM} Gold Yeast-2-Hybrid System User Manual (Clontech Laboratories Inc., 2012).

Once all the constructs were successfully cloned, the first step in the optimization of the method was the efficient transformation into the desired yeast strains and the check for autoactivation of the reporter genes or toxicity caused by the presence of the construct. In
the Y2H Gold strain, AUR1-C reporter gene encodes for inositol phosphoryl ceramide synthase, which provides resistance to the drug Aureobasidin A (AbA'). In addition, HIS3 encodes for an enzyme involved in the biosynthesis of histidine, which is otherwise not produced by the Y2H Gold strain. MEL1 encodes for α-galactosidase, an enzyme responsible for the cleavage of X-α-Gal, resulting in a blue reaction product (Clontech Laboratories Inc., 2012).

A basic verification that the system was working was to perform a test mating procedure with yeast transformed with the control constructs provided by the company. These controls are designed for the Yeast-2-Hybrid interaction study, and therefore the pGBK7 vector was used instead of pBridge. The pGBK7 vector does not contain the second multiple cloning site, which was used in this study for the introduction of the PKMT. The positive control is based on the strong interaction between the murine p53 protein (53), which was fused to the Gal4-DNA-BD, with the SV40 large T antigen (T), which was fused to the Gal4-AD. As a negative control, the protein Lamin (Lam) fused to the Gal4-DNA-BD was provided (Clontech Laboratories Inc., 2012) (Figure 18).

![Figure 18 - Mating of the Yeast-2-Hybrid positive controls provided by Clontech Laboratories Inc. pGBK7-53 encodes for Gal4-DNA-BD-p53 fusion, which interacts with Gal4-AD-SV40 T antigen, encoded by pGADT7-T. SD/-Trp selects for presence of pGBK7 vector. SD/-Leu selects for presence of pGADT7 vector. SD/-Leu-Trp selects for both vectors. SD/-Leu-Trp/X-α-Gal/A selects for interaction between the prey and bait. The colonies where interaction occurs are blue.](image)

In the case of every mating experiment shown in this work, the mated culture was always plated on several different selection media. For control, SD/-Leu, SD/-Trp and SD/-Leu-Trp served to test for the presence of all the vectors, individually and together. The addition of selection such as X-α-Gal, AbA antibiotic, or absence of histidine (SD/-His) in the media was meant to identify positive interaction. In addition, induction of the PKMT expression was ensured by the lack of methionine in the media (SD/-Met) because the PKMT expression is driven from the methionine inducible MET25 promoter. Combinations of these selection criteria were used for the identification of positive colonies, as described in Table 5.
Table 5 – Overview of the selection criteria employed in this study, together with the media used for each type of selection and the phenotype of the positive yeast colonies in each case.

<table>
<thead>
<tr>
<th>Selection Criterion</th>
<th>Selection Media</th>
<th>Positive Interaction Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBridge vector presence</td>
<td>SD/-Trp</td>
<td>Growth selection (positive colonies grow)</td>
</tr>
<tr>
<td>(Gal4-DNA-BD fusion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pGADT7 vector presence</td>
<td>SD/-Leu</td>
<td>Growth selection (positive colonies grow)</td>
</tr>
<tr>
<td>(Gal4-AD-fusion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Bait” and “prey” interaction</td>
<td>SD/X-α-Gal</td>
<td>Color selection (positive colonies are blue)</td>
</tr>
<tr>
<td></td>
<td>SD/-His</td>
<td>Growth selection (positive colonies grow)</td>
</tr>
<tr>
<td></td>
<td>SD/A</td>
<td>Growth selection (positive colonies grow)</td>
</tr>
</tbody>
</table>

The Yeast-2-Hybrid positive control (Figure 18) revealed only strongly blue colonies on the X-α-Gal plate. In order to have a proper technical control for the Yeat-3-Hybrid experiments, the company-sent controls were re-cloned in the pBridge vector, and the mating experiment was repeated following the same procedure. In this case as well, the positive control showed exclusively blue colonies on the X-α-Gal plates.

Moreover, the cells grew on media lacking histidine, a sign of positive interaction (Figure 19, A). The negative control cells presented only white cells for the color selection, and did not grow on SD/-His media (Figure 19, B). This final technical control confirmed the method worked for standard, strongly interacting proteins.

The growth of Y2H Gold cells on plates containing 70 ng/ml AbA (Clontech Laboratories Inc., 2012) was tested after transformation with different pBridge vector constructs. In all cases, colony growth was not inhibited by the presence of the antibiotic (data not shown). As a result of this ubiquitous autoactivation response, this method for selecting positive interaction was not employed further.
In an initial round of testing, Y2H Gold cells were transformed with the pBridge vector constructs by the lithium acetate method and plated on medium selecting for the presence of the plasmid (SD/-Trp), and the presence of bait and prey interaction (X-α-Gal). This latter selection media is used as a control, to prevent the use of constructs, which show autoactivation by the bait alone.

It was revealed that none of the three G9a targets were toxic to the yeast cells, because growth on single selection media was not precluded, compared to the vector control (Figure 20, A-D).

However, the G9aNTD showed evident autoactivation of the MEL1 gene promoter and subsequent appearance of blue colonies (Figure 20, B). All the other pBridge constructs, containing the same PKMT targets, but human G9a SET domain and GLP SET domain instead of the mouse G9a SET, were also tested in the same manner, and the results were identical.
The G9aNTD target was still used in further mating tests, but its autoactivation capacity was kept in mind for the analysis of the results. The HP1-β full length construct was not well tolerated by the Y187 yeast cells. After plating the transformed cells without any dilution, growth was restricted to a few small colonies at least three days after transformation (Figure 20, E, right). To obtain more colonies for the mating experiments, these colonies were re-streaked on SD/-Leu plates selecting for the pGADT7 plasmid, so that sufficient material would be available (Figure 20, E, left). The growth of the Y187 cells transformed with pGADT7 carrying either HP1β or MPP8 chromodomains (Figure 20, F) was unaffected when compared with wild type yeast.

The two transformed strains were then used in small scale mating experiments. The plates testing for reporter gene activation are shown in Figure 21. The induction of the PKMT expression, caused by the lack of methionine in the media, did not have any effect on the outcome of the interaction testing. In this case, only color selection was used for initial testing of interaction. For the negative control, where the PKMT G9a and the target were not present, as well as for the WIZ and CDYL1 targets, no blue colonies were found (Figure 21, left side). In the case of the target G9aNTD, most colonies were light blue, both in the presence and in the absence of the methyltransferase induction (Figure 21). The same experiment was repeated.
with the chromodomain of HP1-β instead of the full length protein. The outcome was identical (data not shown).

The initial mating experiments were intended to be used as positive controls for the Yeast-3-Hybrid system, but in our experimental setup, no interaction was detected between the “bait” and the “prey” proteins. Consequently, optimization of the setup was sought by introducing alternative Yeast-3-Hybrid members, such as additional Gal4-AD-fused reading domains, additional PKMT targets and alternate PKMTs suited to our purpose.
In vitro, the SET domain of G9a is less active than the SET domain of GLP. For this reason, GLP SET was cloned into the pBridge constructs, and the mating experiments were repeated to check whether the increased activity would result in a higher degree of target methylation and subsequent reporter gene activation. In this experiment, media depleted of histidine provided additional selection for positive colonies (Figure 22, third and fourth columns). However, the methylation targets WIZ and CDYL1 showed no signs of interaction on selective media (Figure 22, second and third rows, respectively). G9aNTD and HP1-β showed positive interaction regardless of the presence of the PKMT.

Although the background autoactivation levels were high for this construct, there was still a difference observed between the two methyltransferases, GLP and human G9a. In the case of the latter, the blue color was weaker, and of similar intensity in all selection media (Figure 22, fifth row). In the case of G9aNTD expressed together with GLP, the intensity of the blue color was higher than in the absence of GLP, especially on the plates where –His nutritional selection occurs (Figure 22, fourth row). The result suggested that methylation-dependent interaction
did not occur, but rather the activation of the reporter genes was caused by the inherent activating property of the G9aNTD-DNA-BD bait protein. The methyltransferase of choice seemed to have a role in the intensity of the autoactivation, even in the samples where the PKMT was repressed. This phenomenon will be investigated and discussed later.

Next, HP1 was replaced with MPP8 chromodomain, to test if maybe the choice of reading domain was not appropriate for in vivo studies. MPP8 is known to bind not only to H3K9me3, but also to the trimethylated non-histone target of G9a/GLP, Dnmt3aNTD (N-terminal part of Dnmt3a) (Chang et al., 2011). We first tried mating of yeast expressing MPP8 chromodomain with yeast transformed with the G9a targets WIZ, CDYL1 and G9aNTD (Figure 23).
The chromodomain of MPP8 did not change the outcome of the interaction assay. Interaction was detected only in the case of the G9aNTD, which again showed high autoactivation capacity. However, the methyltransferase of choice was again proven to be important for the result, as GLP induced a much higher intensity of blue color production than human G9a (Figure 23, fourth and fifth rows). In this case as well, the effect was present also in the repressed state of the MET25 promoter. This issue will be addressed later in this section.

In addition, the G9a targets WIZ, CDYL1 and G9aNTD, although shown to be bound by HP1 \textit{in vitro} (Rathert et al., 2008), have not been reported to be bound by MPP8. Therefore, the lack of interaction detection might be due to a real lack of interaction.

To get a better pair of interactors, the mouse Dnmt3a NTD was cloned into the pBridge vector in two different lengths, as a 218 amino acids long (Dnmt3a 1-218) designed by Dr. Renata Jurkowska, and a 274 amino acids long (Dnmt3a 1-274) (Chang et al., 2011) domain. The interaction of the Dnmt3a domain methylated by G9a/GLP with the MPP8 chromodomain has been reported to occur \textit{in vivo} (Chang et al., 2011), not just via peptide assays \textit{in vitro}. Despite
this documented interaction, the binding could not be confirmed through the Yeast-3-Hybrid approach (Figure 24). In Figure 24, the diploid cells were all white, and did not grow on media without histidine. Interaction of the two Gal4 domains did not seem to occur. The results were not influenced by the presence of the mouse G9a methyltransferase (Figure 24).

Although the transformations and mating experiments with the interaction partners provided by the Clontech Laboratories Inc Company were verified successfully, it was not possible to do the same verification with the constructs destined to be used for the screening. Therefore, a thorough analysis of the expression of all the system components was performed, in order to identify where the problem was.

The reading domain in the pGDAT7 vector and the PKMT in the pBridge vector are HA-tagged. This makes the detection of the protein from yeast whole cell extracts easy. Both HP1β and MPP8 chromodomains were visible on western blots stained with antibody against the HA tag (Figure 25, A). The band corresponding to MPP8 was relatively weak, but this might be attributed to technical reasons, because the Ponceau stain is also weaker on this lane. Additionally, untransformed yeast control (data not shown), did not contain any signal.

Similarly, transformed Y2H Gold cell extracts showed clear signal for the mouse G9a, the human G9a, and GLP, also on immunoblots stained for HA-tag detection (Figure 25, B). In the case of the PKMT detection, the first problem was identified. Although the samples were grown either in absence or presence of inhibiting amount of methionine, the PKMT was still expressed. The amount of the PKMT was indeed lower in most of the samples that were exposed to methionine in the media, but inhibition was often not complete (Figure 25, B).
In the case of mG9a and hG9a, methionine did have an effect, although it was a highly variable one. For hG9a/WIZ and mG9a/CDYL1, the repression seemed efficient. For the G9aNTD constructs, in all three cases, the methyltransferase was expressed at lower levels than in the case of the constructs containing the other targets, but weak repression of mG9a and hG9a was still observed. The result for hG9a/CDYL1 was most probably a technical failure, maybe at the level of loading the SDS-PAGE gel or the transfer to the blot membrane.

Out of all the mG9a and hG9a constructs, the least repression was noticed in the case of mG9a/WIZ. The constructs containing GLP as the PKMT showed almost no visible methionine-induced repression. The problem of the leaky MET25 promoter was not pursued any further, because a much more serious issue was identified and characterized.
Detection with antibody against the Gal4-DNA-BD did not show the presence of any of the PKMT targets (Figure 25, C). Proteins expressed from a truncated ADH1 promoter, such as the PKMTs in this study (Figure 16, B), might be below the lower range of Gal4-antibody detection (MacDonald, 2001, page 46).

The initial approach to solve the PKMT target detection problem was to introduce a His6 tag at the C-terminus of one of the targets, Dnmt3a (1-274). The target was expressed in Y2H Gold cells, having either induction (SD/-Trp/-Met) or repression (SD/-Trp) of the PKMT expression.  

**Figure 26 - Attempts to detect His6-tagged PKMT targets using immunoblotting (IB).** SD/-Trp is media selecting for the presence of the expression plasmid. The PKMT expression is induced in this media. SD/-Trp/-Met is media which selects for the expression plasmid and represses the expression of the PKMT. The stars show the height of the expected protein. P.P. is the purified protein sample. Ponceau stain is provided for comparison. A. Yeast total protein extraction samples from cells expressing His6-tagged PKMT target. The first lane represents a positive control for the antibody staining, a recombinantly expressed and purified His6-tagged protein. The negative control is represented by extract from cells transformed with empty vector. B. Result of small scale His6-tagged protein purification under denaturing conditions. No signal is detected in the purified protein (P.P.) lanes at the expected size. C. Result of large scale His6-tagged protein purification under native conditions. FT – flow through. W – wash fraction. Ctrl No His6 – Cells transformed with pBridge construct carrying the target without His tag. D. Yeast total protein extraction samples from cells transformed with modified pBridge vector containing His6-tagged WIZ expressed from the full length ADH1 promoter. mG9a signifies that the construct also encodes for PKMT. Growth was done in inhibiting conditions for the PKMT.
After whole protein extraction by precipitation with 50% TCA and resuspension under denaturing conditions, the blotted samples showed no signal at the right size of the Gal4-DNA-BD-fused target (Figure 26, A).

The total protein extraction procedure was performed on cultures of a total OD$_{600}$ of 4.5. An OD$_{600}$ value of 1 is generally approximated to $3 \times 10^7$ cells (Day et al., 2004), and therefore, the starting material used for the yeast protein extraction corresponded to approximately $13.5 \times 10^7$ cells.

Next, a small scale affinity purification was performed using approximately $450 \times 10^7$ cells as starting material. The protein precipitation was done using ethanol at low temperatures, followed by denaturation in urea buffer. The purification was done using Ni$^{2+}$-NTA beads and elution was performed with imidazole, under denaturing conditions. One band was detected at approximately 70 kDa in all purification and wash samples, regardless of the growth media (Figure 26, B). This band was previously identified as a yeast endogenous protein, because it was present in cells transformed with empty pBridge vector as well. The band did not correspond to the Gal4-DNA-BD-Dnmt3a (1-274)-His$_6$ fusion, but was in the range of 10 kDa higher size. The wash fractions showed a few extra protein bands (Figure 26, B).

Using the same His$_6$-tagged target construct, a native large scale purification was done, starting from approximately $3 \times 10^{10}$ cells. As control, cells containing the same construct without a His$_6$ tag were used. The flow through and wash did not show any prominent signals, suggesting that all His$_6$-tagged proteins were immobilized on the column (Figure 26, C). There was a nicely purified unspecific band which ran at approximately 70 kDa, the same size as in Figure 26, B. The Dnmt3a-His$_6$ purification lane presented a few very weak lower molecular weight bands, but their size did not seem to correspond to the requirements. However, literature search revealed that the Dnmt3a N-terminal domain migrates on SDS-PAGE gels differently than expected judging by its size (Chang et al., 2011). To avoid future confusion, the next experiment was performed with His$_6$-tagged WIZ protein.

The last attempt to identify the PKMT targets in the yeast cell lysate was based on the supposition that the amount of target in the cells was too low for detection by antibodies, either due to low transcription levels, due to degradation of the DBA-BD fusion, leading to a short protein half-life. To address the first issue, the truncated ADH1 promoter was exchanged with the full length ADH1 promoter from pGADT7, to ensure higher levels of constitutive
target expression (Figure 17). The replacement was achieved in pBridge vector either containing no PKMT, or containing mouse G9a. The constructs were transformed in Y2H Gold cells, and the whole protein TCA extraction was performed (Figure 26, D). The yeast strains transformed with constructs containing the full length ADH1 promoter presented stunted growth. In addition, staining against the His$_6$-tag showed no difference between signals of cells containing empty vector or cells expressing the His$_6$-tagged target protein (Figure 26, D). In conclusion, after the investigation of the expression of all the members of the Yeast-3-Hybrid system, it appeared that the PKMTs and the AD-fusion proteins were present, while the PKMT targets fused to Gal4-DNA-BD were absent. The lack of the PKMT targets, if confirmed, would explain why all the control experiments with the reading domains and the conditionally methylated PKMT targets did not succeed.

Because the Yeast-3-Hybrid system was commercially available, it was very surprising that the preparatory stage was so complicated, and finally also unsuccessful. Because the failure of the positive controls that I had prepared was so unpredictable, additional preparatory work was carried out in parallel of the testing of controls. One main purpose of the presented Yeast-3-Hybrid project was to identify novel mono-methyllysine reading domains that would interact with the non-histone targets of the histone H3 lysine 4 monomethylase SET7/9. The constructs containing SET7/9 SET domain as the PKMT and its targets – AKAP6, CENP1, PPARBP, ZDH8, MeCP2, MINT and IRF1 had been prepared during the system validation experiments, and were also tested for autoactivation and toxicity, in anticipation of the Yeast-3-Hybrid interaction screening.

The initial plan was to optimize the experimental setup with the G9a/G9a target constructs, for which there were available reading domains to confirm that target methylation conferred the pairs of interacting proteins with the ability to activate the reporters. S. cerevisiae contains endogenous Set7, which might interfere with the methylation-dependent screen. The plan had been to produce Y2H Gold Set7 knock out (KO) mutants and transform them with the pBridge-SET7/9 constructs.

In the case of the library strain, a Set7 KO strain of Y187 would have been produced, and the Clontech Laboratories Inc. protocol and kit – Make Your Own “Mate & Plate™” Library System (Clontech Laboratories Inc., 2012) – would have been used instead of the ready-made Y187 libraries destined to be used for the G9a/GLP targets.
Only the testing of autoactivation and toxicity was performed for the SET7/9 targets. Three out of seven targets – AKAP6, PPARBP and IRF1 – showed autoactivation of the MEL1 promoter. AKAP6 showed weakly blue colonies, while PPARBP and IRF1 presented strong autoactivation (Figure 27). None of the transformants grew on SD/-His plates (Figure 27). Because the mating experiments with our project-related positive controls were not successful, the SET7/9 experiments were discontinued at this point.

The project did not develop according to the original plan, most probably because the company-provided Yeast-3-Hybrid system approach was not suited for this particular purpose. The interaction in the case of our internal positive controls was not observed, despite the use of different target/reader pairs. Therefore, optimization of the method is still required before it can be used for the identification of novel readers of methylated PKMT targets.

Figure 27 - Test for autoactivation and toxicity of the SET7/9 target constructs. The constructs were transformed in Y2H Gold cells and plated on media selecting for the plasmid (SD/-Trp), and two different selection media for verifying if autoactivation of reporters occurs (X-α-Gal and SD/-His). Each row on the left side shows the three different media selections, followed by the name of the target. Each row on the right side shows the name of the target followed by the three images of the selection.
4.2. The PHF1 Tudor Domain Binds to Trimethylated Histone 3 Lysine 27, With a Strong Preference for the Histone H3T Variant

The Polycomb Repressive Complexes, PRC1 and PRC2, are essential for the setting and the maintenance of facultative heterochromatin, and are responsible for the repression of developmentally regulated genes. PRC2 is the complex responsible for the methylation of H3K27, and has roles in germline development, X-chromosome inactivation, differentiation and stem cell pluripotency (Cao & Zhang, 2004). PRC2 contains several members, among which the PHF1 protein, which is important for Hox gene silencing and efficient H3K27 trimethylation (Sarma et al., 2008).

The PHF1 protein contains two PHD fingers and one tudor domain, which have potential to serve in protein-protein interaction. In yeast, the PHD fingers of the PHF1 homolog (PCL) were shown to interact with the enhancer of zeste protein (O’Connell et al., 2001).

The three individual domains of PHF1 were cloned and purified by Dr. Ina Kycia as GST-fusion proteins, and were subsequently tested for binding specificity to Celluspots peptide arrays (MODified™ Histone Peptide Array by Active Motif) (experiments performed by Dr. Ina Kycia, data not shown).

Only the tudor domain showed binding to the peptide spots, and consequently this reading domain was thereon investigated. The results of the study were presented in the publication (Kycia et al., 2013). My project was a part of the above-mentioned article, and involved the characterization of the peptide binding affinity and specificity of PHF1 tudor to H3K27me3 via fluorescence depolarization measurements and histone pulldown assays.

The binding affinity of the recombinant PHF1-Tudor-GST fusion protein to fluorescently labeled peptides was determined through fluorescence anisotropy measurements (Figure 28). Two binding pocket mutants, Y47A and W41A, were kindly provided by Dr. Ina Kycia, to be used as controls in the following experiments. The peptides used in the fluorescence anisotropy experiments contained the sequences surrounding the H3K27 position of either the canonical histone H3.1 or the testis-identified variant H3T, which has a valine at position 24 instead of the Ala24 present in H3.1 (Tachiwana et al., 2010). First, the binding of PHF1 tudor domain to the two different histone variant tails containing K27me3 was assessed.
(Figure 28, A). Secondly, the binding to the H3TK27me3 peptide by the wild type, compared to the two mutant PHF1 tudor domains, was recorded (Figure 28, B).

The wild type PHF1 tudor domain preferred H3TK27me3 peptides over H3.1K27me3 and also H3T unmodified peptides, showing that the domain was specific both for the particular histone variant and the appropriate modification. The binding affinity of PHF1 to H3TK27me3 was 6-fold stronger than to H3.1 with the same modification, while the binding to H3T without any modification was negligible (Table 6).

The recombinantly expressed mutant PHF1-tudor-GST fusion proteins, Y67A and W41A, were tested along with the wild type domain in the fluorescence anisotropy assay. Upon mutation of tyrosine 47 and tryptophan 41 to alanine (Y47A and W41A), the binding affinity to H3TK27me3 was completely lost (Figure 28, B, Table 6). The titrations of the H3TK27me3 peptides with mutant PHF1 tudor domains in solution (Figure 28, B) were similar to the titrations of the wild type domain binding to the unmodified H3T (Figure 28, A).

Literature sources report that PHF1 might actually be an H3K36me3 reader (Muselman et al., 2012; Qin et al., 2013). In the published work (Kycia et al., 2013), binding to H3K36me3-containing spots was also observed on the MODified™ Histone Modification Array. First, the fact that the H3K36 peptides are hydrophobic often results in unspecific binding to the peptides. In addition, the expectation that PHF1 would bind to H3K27me3 rather than another mark, together with the doubts regarding the H3K36me3-binding, prompted us to concentrate only on the study of H3K27me3 binding.
Figure 28 - The binding of PHF1 tudor wild type and mutants to peptides. The results were obtained by fluorescence anisotropy measurements by titration of peptide (100 nM) with increasing concentrations of PHF1 tudor domain. The figure shows averaged values of 3-5 repeats for each titration. The error bars represent the standard error of the mean. A. Binding of the wild type PHF1 tudor domain to H3K27me3, H3.1K27me3 and unmodified H3T peptides. B. Binding to the H3K27me3 peptide of the wild type PHF1 tudor domain and the two PHF1 tudor mutants. The binding constants corresponding to the presented titrations are shown in Table 6.
Table 6 - Binding constants ($K_d$) characterizing the interaction affinity of the wild type and mutant PHF1 tudor domains to various fluorescently labeled peptides. The values were calculated from the data sets presented in Figure 28, which were obtained using fluorescence depolarization measurements.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide</th>
<th>$K_d$ [µM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHF1-Tudor WT</td>
<td>H3T K27me3</td>
<td>25.8 ± 1.9</td>
</tr>
<tr>
<td></td>
<td>H3.1 K27me3</td>
<td>142 ± 19</td>
</tr>
<tr>
<td></td>
<td>H3T unmodified</td>
<td>&gt; 300</td>
</tr>
<tr>
<td>PHF1-Tudor Y47A</td>
<td>H3T K27me3</td>
<td>&gt; 300</td>
</tr>
<tr>
<td>PHF1-Tudor W41A</td>
<td>H3T K27me3</td>
<td>&gt; 300</td>
</tr>
</tbody>
</table>

Proof of binding to H3K27me3 and H3K36me3 peptides via fluorescence depolarization assays was shown. Although the affinity to H3K36me3 in vitro was 1.8-fold higher than to H3TK27me3 ($K_d = 14.4 ± 2.6$ µM), we hypothesized that, in vivo, the recruitment to H3TK27me3 would be more relevant, and I tried to show that, in a more native assay condition, H3K27me3 is preferred over the K36me3, or at least equally preferred to it. Histones isolated from HEK293 (human embryonic kidney 293) cells were pulled down with GST-tagged wild type and Y47A mutant PHF1 tudor domain. The resulting samples were run on an SDS-PAGE gel, which was blotted and stained with antibody against H3K27me3, stripped and re-stained with anti-H3K36me3 antibody (Figure 29). The results of the antibody stainings were comparable, showing that both H3K27me3 and H3K36me3 marks were enriched in the wild type lane, while the mutant bound much weaker to the modified native histones (Figure 29).

The wild type PHF1 tudor domain presented more than 3-fold enrichment for the H3K27me3 mark in comparison to the Y47A mutant (Figure 29, Table 7). In contrast, PHF1 tudor brought about less than two-fold enrichment compared to the mutant stained with H3K36me3-specific antibody (Figure 29, Table 7). The quantification of the pulldown showed clearly that the Y47A
binding pocket mutant more strongly affected the binding of the H3K27me3, than that of the H3K36me3 peptide. This suggested that H3K36me3 binding is not so much dependent on the intact binding pocket, and might actually be a more unspecific interaction. The result of the pulldown might have been affected by the tendency of native histones to aggregate, and cause sometimes carryover of unspecific signals together with the specific ones.

Table 7 – Quantification of the band intensities representing the histone mark enrichment resulted from the individual pulldowns with PHF1 wild type tudor domain and PHF1 Y47A tudor mutant. The western blot images of each antibody staining from Figure 29 were quantified and the estimation for each lane was presented as relative value to a total input of 1.0.

<table>
<thead>
<tr>
<th>Quantified mark</th>
<th>Input lane</th>
<th>PHF1 WT lane</th>
<th>PHF1 Y47A lane</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K27me3</td>
<td>1.0</td>
<td>0.64</td>
<td>0.18</td>
</tr>
<tr>
<td>H3K36me3</td>
<td>1.0</td>
<td>0.99</td>
<td>0.56</td>
</tr>
</tbody>
</table>

The function of the PHF1 is still not completely clear, although much investigation has been taking place on this subject. Interestingly, this tudor domain binds to H3K27me3, with a preference for the H3T variant. The binding of the reading domain to the same mark as the complex it is part of sets is not a new concept, but is rather common among epigenetic players, ensuring the spreading of the mark on neighboring nucleosomes. The affinity towards H3K36me3 is not so easy to explain, and if it does really happen in vivo, a switch between H3K27me3 and H3K36me3 preference might be occurring with the aid of unidentified factors, to fine tune the function of the PHF1 protein.

4.3. Characterization of the Binding Specificity of the Chromodomains of CBX2 and CBX7 Proteins

Embryonic stem cell pluripotency and differentiation are tightly regulated processes that involve the critical roles of the Polycomb- and Trithorax-related proteins, responsible for the repression and activation of genes, respectively. PRC1 is responsible for the ubiquitination of H2AK119, which is essential for efficient transcriptional repression (Schuettengruber et al., 2007). In Drosophila, the PRC1 complex comprises Polycomb (Pc), Posterior sex combs (Psc), and the E3 ubiquitin ligase RING. Mammals have several homologs for each of these proteins, resulting in a large variety of PRC1 complexes (Morey & Helin, 2010). Polycomb is responsible for PRC1 targeting in flies, as it binds specifically to H3K27me3, a histone mark associated with Polycomb group repression.
In mammals, there are five Pc homologs: CBX2, CBX4, CBX6, CBX7 and CBX8. The binding of all the polycomb group chromobox (Cbx) proteins has been investigated in detail in several instances already, both for the mouse (Bernstein et al., 2006) and the human homologs (Kaustov et al., 2011). The binding constants presented in literature suggest that the murine Cbx2 binds approximately equally to H3K9me3 and H3K27me3 peptides of the H3.1 variant, while mouse Cbx7 binds two-fold stronger to H3K9me3, compared to H3.1K27me3 (Bernstein et al., 2006). In humans, CBX7 shows very similar binding constants towards H3K9me3 peptides, while CBX2 presents only relatively weak binding (K_d = 185 ± 20) towards H3.1K27me3, and no binding towards H3K9me3 (Kaustov et al., 2011).

After identifying the surprising variability in affinity of the PHF1 tudor domains towards the H3.1 and H3T variants of histone H3, I tried to test whether this preference was a unique property, or whether other reading domains, such as the chromodomains of CBX2 and CBX7, could also sense the sequence difference at the K27 position. Part of the work presented here was included in a manuscript (Kungulovski et al., 2014).

The wild type CBX2 and CBX7 chromodomains, cloned and purified by Dr. Ina Kycia, were initially tested for binding to the MODified™ Histone Peptide Arrays (Figure 30). Similarly to literature reports, the CBX7 chromodomain preferred binding to H3K9me3 peptides, followed by H3K27me3 peptides. H3K9me2 and H3K27me2 were also bound, but much weaker (Figure 30). The preference between the two positions, K9 and K27, was reversed in the case of CBX2, which bound much stronger to H3K27me3 spots than to H3K9me3 spots (Figure 30). CBX2 was also able to bind very weakly to dimethylated H3K27 and H3K9 (Figure 30). Both chromodomains presented reduced binding capacity in the case of serine or threonine phosphorylations in the proximity of the target lysine (Figure 30).
Next, the two domains were tested on cellulose membranes with synthesized peptide spots containing unmodified and methyllysine analogs (MLAs) of H3K27me3 peptides from both H3.1 and H3T histone variants (Figure 31).

CBX2 and CBX7 showed preference towards the methylated H3T variant, and presented no binding to unmodified peptides of each variant (Figure 31). Once the differential recognition

---

**Figure 30 - Analysis of CBX2 and CBX7 chromodomains binding to modified histone peptide spots immobilized on nitrocellulose membranes.** Two replicates of the binding assay are presented. The replicate on the left is annotated for easier interpretation of the results. CBX2 (above) bound specifically to H3K27me3 peptide spots (purple) and weaker to H3K9me3 peptide spots (red). CBX7 (below) preferred binding to H3K9me3 peptides, followed by H3K27me3 peptides. For both CBX2 and CBX7, dimethylated H3K9 (blue) and H3K27 (orange) was much weaker bound than its trimethylated counterpart. Inhibition was observed when phosphorylation occurred at H3S10/T11 (green arrow), and H3S28 (red arrow).

Next, the two domains were tested on cellulose membranes with synthesized peptide spots containing unmodified and methyllysine analogs (MLAs) of H3K27me3 peptides from both H3.1 and H3T histone variants (Figure 31).

CBX2 and CBX7 showed preference towards the methylated H3T variant, and presented no binding to unmodified peptides of each variant (Figure 31). Once the differential recognition

---

**Figure 31 - CBX2 and CBX7 chromodomains prefer binding to the histone 3 variant H3T.** The membranes contain four peptide spots, with residues 21-35 of histone 3. The first two peptides represent the H3T variant, which contains a V at position 24. The first peptide is unmodified, and contains a lysine at position 27. The second H3T peptide contains a trimethyllysine analog (converted cysteine, C-me3) at position 27. The last two peptides contain the H3.1 sequence, with an alanine at position 24. The third peptide contains unmodified H3.1 peptide, with a native lysine at position 27, while the fourth peptide contains a trimethyllysine analog at position 27 (C-me3). A. CBX7 shows strong preference for the H3T methylated peptide, while no binding is observed for the H3.1 methylated peptide. B. CBX2 binds to both methylated peptides, showing stronger binding to the H3T variant. Neither of the chromodomains binds to unmodified H3 variant peptides.
of the histone variants was observed at peptide level, just like in the case of PHF1 tudor domain, the investigation was continued by production of methyllysine binding cage mutants.

Reported mutations of the aromatic cage of CBX7 (F11A and W32A) (Li et al., 2010) were introduced using site-directed mutagenesis of the wild type constructs. A third cage mutant, V13A, was designed using the solved structure of mouse Cbx7 bound to the H3K27me3 peptide (PDB ID: 2KVM) (Yap et al., 2010). The crystal structure of CBX2 bound to the H3K27me3 peptide (PDB ID: 3H91) (Kaustov et al., 2011) was used to design the W32A mutant. The purified GST-fusion mutant proteins were tested on the MODified™ Histone Peptide Arrays, and all but CBX7 V13A showed loss of binding to the target peptides (Figure 32).

![Figure 32: Results of CBX2 and CBX7 mutants binding to modified histone tail peptide arrays. CBX7 F11A and W32A show unspecific binding on the entire modified histone tail array. CBX7 F11A still keeps a very weak H3K27me3 specificity (three stronger intensity spots). CBX7 V13A did not lose binding ability. CBX2 W32A completely lost binding to histone tails. Merged images show a merge between the luminescence signal and the ambient light image of the array.

The fact that the V13A mutant still bound to the peptide array was not entirely surprising, because the amino acid exchange from alanine to valine represents a very mild alteration, although it is located very close to the aromatic cage.
The dissociation constants of the binding to the two histone variants with different modifications were obtained from the titration experiments measured using fluorescence depolarization assays (Figure 33, both figures). CBX7 bound to H3K9me3 with an intermediate affinity between the H3TK27me3 and the H3.1K27me3 peptides (Figure 33, A).
Figure 34 - CBX7 wild type and F11A, W32A and V13A mutants binding to fluorescently labeled H3T K27me3 (A), H3.1 K27me3 (B) and H3K9me3 (C), peptides recorded by fluorescence anisotropy measurements. The experiments were carried out by titration of peptide (100 nM) with increasing concentrations of CBX7 chromodomain. The average of at least three experiments for each titration is shown. Individual points represent averaged experimental data. For each titration, the data were fitted to a theoretical trendline using a binary-binding equilibrium fit. The error bars represent the standard error of the mean. For some points, the error bars are too small to be distinguished from the point markers.
The CBX7 mutants showed different degrees of binding impairment, documented in Figure 34. The W32A mutant showed the strongest effect, causing almost complete loss of binding to target peptides (Figure 34). Mutation of phenylalanine 11 resulted in weakened binding to histone peptides, but the effect was not as drastic as the W32A mutation. The V13A mutation only reduced the affinity approximately 3-fold compared to the wild type domain (Figure 34, Table 8), confirming the result of the peptide array (Figure 30). Due to inefficient inhibition of binding capabilities, the V13A CBX7 mutant was not used in any further experiments.

The other Pc homolog, CBX2 wild type bound with similar affinity to both histone variants of the H3K27me3 peptide, and only negligible binding was observed to the H3K9me3 peptide. The CBX2 mutant, W36A, showed reduced binding to the target peptide, H3.1K27me3 (Figure 33, B).

Table 8 - Binding constants (K_d) for the wild type and mutant CBX7 and CBX2 chromodomains towards various peptides.
The binding titration experiments used for the calculation of the constants were obtained using fluorescence anisotropy. The value is accompanied by the standard error of the mean. N/A is written in the case where the binding affinity was not measured.

<table>
<thead>
<tr>
<th>Domain</th>
<th>K_d H3t K27me3 [µM]</th>
<th>K_d H3K9me3 [µM]</th>
<th>K_d H3.1 K27me3 [µM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBX7 WT</td>
<td>2.7 ± 0.5</td>
<td>14.2 ± 0.7</td>
<td>18.5 ± 1.0</td>
</tr>
<tr>
<td>CBX7 F11A</td>
<td>84.0 ± 3.3</td>
<td>&gt; 500</td>
<td>&gt; 500</td>
</tr>
<tr>
<td>CBX7 W32A</td>
<td>&gt; 500</td>
<td>&gt; 500</td>
<td>&gt; 500</td>
</tr>
<tr>
<td>CBX7 V13A</td>
<td>8.8 ± 0.4</td>
<td>65.0 ± 4.4</td>
<td>61.5 ± 6.0</td>
</tr>
<tr>
<td>CBX2 WT</td>
<td>39.6 ± 4.4</td>
<td>&gt; 200</td>
<td>54.6 ± 0.5</td>
</tr>
<tr>
<td>CBX2 W36A</td>
<td>N/A</td>
<td>N/A</td>
<td>&gt; 500</td>
</tr>
</tbody>
</table>

The two chromodomains were further used in work that was partially published (Kungulovski et al., 2014). An aim of the project was to find ways to efficiently replace histone-binding antibodies with more reliable, and cost-efficient reagents. Nowadays, the investigation of histone post-translational modifications mostly relies on the use of antibodies that detect them, either for immunohistochemistry, for immunoprecipitations (IPs), or for
immunodetection on western blots. So far, antibodies are the only mediators for the study of the genome wide distribution of histone marks with single locus resolution.

However, antibodies are not the ideal tools for several essential reasons. Firstly, their quality varies considerably from batch to batch, especially in the case of polyclonal antibodies (Bock et al., 2011) and the quality controls performed by companies are often not rigorous enough. In addition, histone tails are normally covered with multiple post-translational modifications, while the antibodies are commonly raised against single modifications. As a result, false negative results appear due to preclusion of antibody interaction by additional bulky groups. Last, but not least, antibody production involves a complicated and costly process. In the manuscript by Kungulovski et al (2014), histone reading domains are proposed as alternatives to histone tail-specific antibodies. The recombinantly expressed, GST-fusion proteins can be cost-savingly produced in large amounts. The chromatin readers chosen for this project were subjected to several quality control tests normally used for the confirmation of antibody quality.

First, for a simple confirmation that the domains bound to modified, native nucleosomes, a far western blot analysis was performed (Figure 35). Recombinant histones and native HepG2 cell nucleosomes were transferred to nitrocellulose membranes and the chromodomain was

![Figure 35 - Far western blot experiment showing the binding of CBX7 WT (A) and F11A (B) and W32A (C) to either recombinant histones H3 and H4, or native mononucleosomes purified from HepG2 cells. The detection was accomplished using immunodetection of the chromodomain GST tag.](image-url)
incubated with the membrane to allow it to bind to its target modification. Next, primary anti-GST antibody and secondary antibody incubations followed.

Figure 36 – Comparison of the CIDOP-qPCR results for CBX7 and CBX2 with the ChIP-qPCR results for the H3K27me3 antibody. The amplicons used can be grouped in three different sets: associated with H3K27me3, associated with H3K9me3 and commonly enriched in H3K36me3 marks. All obtained results are presented as percentage of input. The values represent average of two biological repeats performed in triplicate technical repeats. The error bars represent the standard error of the mean. In the case of the reading domains, the experiment was performed both with the wild type and with the mutant as a negative control.
CBX7 wild type showed strong binding to native mononucleosomes isolated from HepG2 cells (Figure 35, A). CBX7 mutants showed loss of specific binding to the native mononucleosomes (Figure 35, B and C).

To further investigate the CBX2 and CBX7 chromodomains capacities to replace antibodies, chromatin interacting domain precipitation (CIDOP) assays were performed in order to specifically precipitate native chromatin according to the reading domain’s target histone modification. This novel, native ChIP-like approach is subject of a manuscript (Kungulovski et al., 2014) already submitted for publication. In the CIDOP procedure, chromatin was isolated from HepG2 cells, and then was digested with Micrococcal nuclease to create a homogenous mononucleosome preparation. The nucleosomes were then precipitated with the GST-tagged reading domains using Glutathione sepharose beads. After several rigorous washing steps, the proteins were cleaved with Proteinase K and the DNA was purified. For comparison to antibodies, CIDOP and ChIP were performed in parallel with the same experimental conditions. The recovered DNA was initially analyzed by quantitative PCR (qPCR) performed at specific loci, which were chosen based on their histone PTM status, according to the ENCODE ChIP-seq datasets for HepG2 cells (The ENCODE Project Consortium, 2012). Later, the DNA recovered from CIDOP/ChIP was subjected to whole genome analysis by next-generation sequencing (CIDOP-seq or ChIP-seq).

In the case of the chromodomains of CBX2 and CBX7, the CIDOP-qPCR results are shown in Figure 36, C and A, respectively. The experiments were performed in parallel with the ChIP-pPCR with anti-H3K27me3 antibody (Figure 36, B). CBX7 wild type showed strong pulldown of up to approximately 5% of the input of the H3K27me3-associated genes, while the mutant showed minimal pulldown of these loci. The degree of enrichment depended on the particular locus, with strongest enrichment at the Hox11 locus, intermediate values for the Serping1 locus, and relatively low enrichment (~1%) at the MyoD amplicon (Figure 36, A). Even for the lower values such as the pulldown of the MyoD locus, the results were highly reproducible. CBX7 wild type showed no binding to H3K9me3- or H3K36me3-containing selected amplicons (Figure 36, A).

The enrichment pattern was very similar to that of the anti-H3K27me3 antibody, which showed pulldown of up to approximately 4% of the input for the Hox11 locus and approximately 2% precipitation of the Serping1 locus (Figure 36, B), comparable to the CBX7
wild type results. However, the antibody presented a 2-fold increase in the pulldown of the MyoD locus (Figure 36, B). A striking difference between the chromodomain of CBX7 and the antibody was that the latter also specifically pulled down the satellite alpha repeat locus (Figure 36, A and B). This suggests that the antibody had high cross-reactivity with the H3K9me3 mark at this locus, compared to CBX7.

The CIDOP-qPCR experiment performed with the other chromobox protein, CBX2, also showed specific pulldown of loci assigned as H3K27me3 targets, but the overall low binding affinity of CBX2 to the target H3K27me3 peptide was reflected in the degree of pulldown. The chromatin precipitation experimental procedure requires harsh wash conditions, which also decrease the CBX2 pulldown efficiency. Noteworthy is the fact that the pulldown of H3K27me3-rich loci was lost with the W32A mutant (Figure 36, C). The amounts of DNA enriched from the input by CBX2 pulldown were at the lower limit of the detection capabilities, having values in the range of 1% or less of input, and showing relatively high variability in the obtained results (Figure 36, C). Despite its weak binding to H3K27me3-containing histone tails, CBX2 showed a clear pulldown of H3K27me3-specific loci. However, CBX2 was not the optimal candidate in this study, because its performance in the CIDOP-qPCR experiment did not compare to the strength and reliability offered by the antibody.

Given the strong and H3K27me3-specific binding of CBX7, this chromodomain was further used in the next step of the project, the CIDOP-seq. The next-generation sequencing was performed by collaborators, Chisato Henry and Paul Labhart, while the analysis of the data was done by Goran Kungulovski and the before mentioned collaborators. The CIDOP-seq and ChIP-seq experiments showed strong overlap (93%) between the H3K27me3 antibody and the reading domain at whole genome level in all H3K27me3 intervals (Kungulovski et al., 2014, Figure 5E-F in the manuscript).

The clarification of the binding specificity of CBX7 was achieved by far western blot analysis of the binding of CBX7 chromodomain to nucleosomes from HepG2 cells depleted of H3K27me3 by cultivation in the presence of an EZH2 inhibitor, and to histones isolated from H3K9me3-depleted Suv39h double knockout MEF cells. CBX7 binding was unaffected by H3K9me3 depletion in MEF cells, while binding was strongly reduced when EZH2 H3K27 methyltransferase was inhibited (Kungulovski et al., 2014, Figure 5B-C in the manuscript). Goran Kungulovski performed these final experiments.
All the results presented above point out to a single-modification, H3K27me3-specific binding of both CBX2 and CBX7, despite the additional evidence of H3K9me3-binding at peptide level. In addition, both domains showed pulldown of H3K27me3-associated loci, although the binding affinity and, consequently, also the loci enrichment compared to input was higher in the case of CBX7. The latter chromodomain not only showed strong H3K27me3-pulldown at specific loci, but also closely reproduced the results obtained with the anti-H3K27me3 antibody. Moreover, CBX7 lacked the antibody-specific cross-reactivity with the H3K9me3 marks at the satellite alpha repeat locus. Finally, CBX7 was also successfully used in CIDOP-seq experiments, where its occupancy correlated very well with the antibody inside the H3K27me3 intervals.

4.4. Investigation of the Binding Specificities of the MLL2 (KMT2D) and MLL3 (KMT2C) PHD Fingers

Another important epigenetic mark is the methylation of histone 3 at position 4 (H3K4). A family of enzymes that are responsible for the introduction of the methylation at H3K4 is the MLL family. Two of its members, MLL1 (KMT2A) and MLL4 (KMT2B), introduce H3K4 trimethylation, an activating mark essential for activation of transcription. Another two members, MLL2 (KMT2D) and MLL3 (KMT2C), are only H3K4 monomethylases and are required for setting this mark at enhancers (Hu et al., 2013). The structural similarity between MLL2 and MLL3 is clearly reflected in the organization of their individual plant homeodomains (PHDs), which appear clustered in three groups along the sequence length (Figure 37, A). MLL2 contains seven PHD fingers, out of which three are clustered towards the N-terminus, another three are towards the middle of the sequence, while the last PHD finger is next to other conserved domains, including the SET domain, in the C-terminal part of the enzyme (Figure 37, A). MLL3 contains eight PHD fingers, which are organized the same as MLL2, with the one additional domain located in the N-terminal cluster (Figure 37, A). The similarity between the proteins goes even deeper, to the sequence of the PHD fingers, which are pairwise highly similar (Figure 37, B and C).

In this project, in order to study the chromatin targeting of the two enzymes, the PHD fingers of MLL2 and MLL3 were cloned as GST fusions both individually, and also in groups, based on the structural clustering mentioned above. Some of the domains purified well on the first trial, but for most of them, boundary optimization was performed in order to increase their
purification quality and yield. Once the GST-fusion protein domains were purified and their size and purity was confirmed by SDS-PAGE (Figure 13), they were all tested for binding to the modified histone tail binding arrays.

**Figure 37 – A.** Schematic representation of the structure and domain organization of MLL2 and MLL3. The panel is taken from Figure 6, from the introduction, section 1.5.3. B. Sequence alignment of all the PHD fingers of MLL2 and MLL3. The highly conserved, zinc coordinating residues are highlighted with orange color. Moderately conserved residues are yellow, and weakly conserved ones are blue. The image was adapted from (Ali et al., 2014). C. Phylogenetic tree created with ClustalW2 – Phylogeny, using the neighbor-joining method, based on the sequence alignment of the PHD fingers of MLL2 (KMT2D) and MLL3 (KMT2C). The phylogenetic tree branches are in cladogram view. The numbers on the right represent the amount genetic change.
In the case of MLL2, most of the PHD fingers showed weak or no binding to the MODified™ Histone Peptide Arrays (Figure 38). PHD3, PHD4 and PHD6 of MLL2 showed unspecific binding over the entire peptide array, while all the rest of the single domains showed only very weak binding or no binding at all (Figure 38).

The groups of domains were also tested, and PHD 0-2 and PHD 1-2 of MLL2 presented no binding to modified histone peptides. For some domains, their low concentration was a limitation for the assay, as it precluded trials with increased protein concentration, such as in the case of MLL2 PHD5, MLL2 PHD 0-2 or MLL2 PHD 1-2 (Figure 38).

Interestingly, PHD 3-5 of MLL2 showed specific binding to the histone tail peptide array, both to histone 3 and histone 4. The triple domain was recruited to H3 (16-35) peptides (Figure 39, A and C), and binding was dependent on PTMs, as the unmodified peptide shows no binding (Figure 39, C, blue arrow).

Taking into consideration the peptides with one single PTM, the strongest binding occurred at H3K27me3 (Figure 39, C, red box with coordinates K2 on the array), followed by other K27 modifications, such as dimethylation (Figure 39, C, blue box, array coordinates K1) or acetylation (Figure 39, C, dark blue box, array coordinates K3), and by weak binding to the H3R26me2a/s (Figure 39, C, black arrows, array coordinates J21-22). Binding to the H3 (res.16-35) peptide was stimulated when mono-, di-, trimethylated or acetylated K27 was
associated with symmetrically or asymmetrically dimethylated H3R26 (Figure 39, C, array coordinates K5 to K14). Binding to the K27-containing peptide was inhibited by the phosphorylation of serine 28, which is adjacent to the central lysine 27 (Figure 39, C, green arrows).

The domain also bound specifically to the H3 (26-45) peptides, with different modification states of the lysine 36 (Figure 39, A). The unmodified and monomethylated peptide was bound

Figure 39 - Analysis of the specific binding of MLL2 PHD 3-5 to modified histone peptides. MODifiedTM Histone Peptide Arrays were used to test the ability of the domains to bind to histone tails with various combinations of PTMs. Relevant peptide modifications are labeled according to the legends. A. The magnified left panel shows the array annotated with the peptides and their boundaries, indiscriminate of PTMs. The annotation is shown as a colored line above the peptide spots. MLL2 PHD 3-5 binds to peptides from H3 and H4. The arrays are provided as duplicates, and the non-annotated duplicate is shown in the upper right panel. B. Magnification of the H3 (26-45) peptides. C. Magnification of the H3 (16-35) peptide containing area, with annotations of relevant modifications. D. Magnification of the H4 (1-19 and 11-30) peptides area, with annotations of the two peptides and the relevant PTMs.
more weakly, while the dimethylated, trimethylated and acetylated H3K36 peptide was bound stronger (Figure 39, A).

The PHD 3-5 fragment of MLL2 also showed strong and specific binding to modified H4 peptides. However, binding to unmodified H4, residues 1-19, was not recorded, while binding to unmodified H4, residues 11-30, was weak (Figure 39, D, blue arrows, under the yellow line and the green line, respectively). Recognition of the N-terminal H4 (1-19) peptide was only achieved in two cases, when triple acetylation occurred at K8, K12 and K16 (Figure 39, red box under yellow line), and when the peptide was fully acetylated, at K5, K8, K12 and K16 (Figure 39, D, blue box under the yellow line). Interestingly, the triple acetylation of K5, K8 and K16 did not cause binding, suggesting that the position of the acetylations in the sequence is important (Figure 39, D, red arrow).

The binding of MLL2 PHD 3-5 was overall stronger to the H4 (11-30) peptides (Figure 39, stretch of peptides below the green line). Spots M18-N6 contain only single modifications of the H4 (11-30) peptide, and the triplet domain preferred binding to three single modifications: H4K16ac (Figure 39, D, magenta box, coordinate M20), and H4R19 symmetric and asymmetric dimethylation (Figure 39, black arrows, coordinates M23-24). Peptides with coordinates N7-N19, found in the magenta box, contain H4K16ac among other modifications, and were strongly bound by the PHD finger group.

The strongest signals on the H4 peptide were observed in peptides where H4K16ac and H4R19me2a/s are combined (Figure 39, D, black arrows inside magenta box, coordinates N10-11). The set of black arrows stretching from row N to row O contains peptides with either symmetric or asymmetric dimethylation of R19, which were also preferentially bound by MLL2 PHD 3-5 (Figure 39, D). The triple acetylated H4 (11-30) peptide at K12, K16 and K20 was also bound by the MLL2 PHD 3-5 (Figure 39, D, green box inside the magenta box).

The detailed analysis of the binding specificity of the MLL2 PHD 3-5 domain to peptide posts suggests that there could be three different pockets, one for each peptide stretch, H3 (16-35), H3 (26-45) and H4 (11-30). Also, it might be that the binding to the H4 (11-30) peptide is achieved through an acetyllysine-binding pocket, or a pocket accommodating both arginine dimethylation and an adjacent lysine acetylation.

The same type of investigation was performed in the case of the MLL3 PHD fingers. Just as for MLL2, all of the single PHD fingers showed either unspecific binding, such as MLL3 PHD2, or
very weak or no binding at all (Figure 40). MLL3 PHD7 and the group MLL3 PHD 0-3 were not purified successfully, and were therefore not included in the subsequent experiments.

The MLL3 PHD 4-6 triplet, which showed also sequence-based similarity to MLL2 PHD 3-5 (Figure 37, B and C), presented a similar binding pattern to that of MLL2 PHD 3-5 (Figure 39). The triplet was cloned for boundary optimization with two different sets of boundaries, and both purified domains had high purity and high concentrations (Figure 13). On the glass arrays, both showed the same specificity, but the MLL3 PHD4-6 (1) (res. 956 – 1144) showed overall stronger binding to the peptide spots and from this point onwards, this was the domain used for further experiments and analysis (Figure 41, A).

The annotated result showed an almost identical binding pattern to that of MLL2 PHD 3-5 (Figure 41, B). However, the binding intensities were slightly changed. The MLL3 PHD 4-6
seemed to prefer binding to the H3K36 peptides in comparison to H3K27 peptides, which were very weakly bound (Figure 41, B, black line for the H3K27 peptides and C, for H3K36 peptides).

The H3 (26-45) peptides were bound equally, without any clear discrimination according to the modification states of K36 (Figure 41, C).

**Figure 41 – Analysis of the specific binding of MLL3 PHD 4-6 to modified histone peptides.** MODified™ Histone Peptide Arrays were used to test the ability of the domains to bind to histone tails with various combinations of PTMs. Relevant peptide modifications are labeled according to the legends in each panel. A. (1) and (2) are two variations of the PHD 4-6 triplet, cloned with different boundaries (see section 3.4.1) B. The magnified left panel shows the left side of the MLL3 PHD 4-6 (1) array annotated with the peptides and their boundaries, indiscriminate of PTMs. The annotation is shown as a colored line above the peptide spots. MLL3 PHD 4-6 bound to peptides from H3 and H4. The non-annotated duplicate of MLL3 PHD 4-6 (1) is shown in the upper right panel. B. Magnification of the H3 (26-45) peptides. C. Magnification of the H4 (1-19 and 11-30) peptides area, with annotations of the two peptides and the relevant PTMs.
Binding to the N-terminal H4 peptide (1-19) was almost undetectable. The only peptide weakly bound was the fully acetylated one (Figure 41, D, blue box under the yellow line). Triple acetylations of the H4 peptide (1-19) were not bound (Figure 41, D, red arrow and red box).

The MLL3 PHD 4-6 bound specifically to modified H4 (11-30) peptides (Figure 41, B and D, peptide stretches under the green line), to which it showed the strongest interaction.

In particular, the H4K16ac (Figure 41, D, magenta box under the green line, coordinates M20) and the two dimethylated R19 modifications (Figure 41, D, black arrows under the green line, coordinates M23-24) were the three single modification peptides which were specifically bound. The stretch of peptides with the highest intensity signal was contained in the magenta box (Figure 41, D, coordinates N7-N19), and all peptides carried H4K16ac in combination with other modifications. The three spots in the cyan box contain additionally to K16ac, also acetylated H4K12 and either mono-, di-, or trimethylated H4K20. The last spot in the magenta stretch is additionally labeled with a green box, and contains a tripled of acetylated lysines, at positions 12, 16 and 20. This spot was also strongly bound by MLL3 PHD 4-6 (Figure 41, D).

Just as MLL2, the triple PHD finger of MLL3 bound strongly also to the stretch of peptides containing H4R19 asymmetrically and symmetrically dimethylated together with other PTMs (Figure 41, D, set of black arrows that stretches from N20 to O3).

Table 6: Ponceau and IB

<table>
<thead>
<tr>
<th></th>
<th>Ponceau</th>
<th>IB</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>10% input</td>
<td>GST MII2 PHD3-5</td>
</tr>
<tr>
<td>GST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MII2 PHD3-5</td>
<td></td>
<td>MII3 PHD4-6</td>
</tr>
<tr>
<td>MII3 PHD4-6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 42 - Native histone pulldown with GST-fused MII2 PHD 3-5 and MII3 PHD 4-6. The triplet PHD fingers pull down native HEK histones. GST is used as a negative control. On the right, immune staining of the same western blot with different antibodies. The enrichment relative to the input is shown under each band.

Taken together, the peptide array binding results of both the MLL2 and MLL3 PHD triplets showed the most intense binding to peptides from histone 4, especially when H4K16ac was present in association with other acetylations or even arginine dimethylations. An initial
experiment to test the validity of the peptide binding results in a more native setting was a native histone pulldown with histones isolated from HEK293 cells (Figure 42).

The GST-pulldown samples were analyzed on a western blot with antibodies against H4 pan-acetyl, H4K16ac and H3K9me3 marks. Enrichment was observed for both MLL2 and MLL3 fragments, compared to input and GST control, with the highest fold enrichment for multi-acetylated H4 tails (Figure 42). However, enrichment of H3K9me3 marks was not in accordance with the peptide array result (Figure 39, Figure 41), and was also rather counterintuitive, as acetylations are activating marks, while H3K9me3 is a repressive mark. The observation of pulldown of H3K9me3-rich peptides might be explained by the natural tendency of native histones to aggregate, which might cause carryover of unspecific signal together with the specifically pulled down histones.

Regardless of the modifications they prefer, the triplets of PHD fingers PHD 3-5 of MLL2 and PHD 4-6 of MLL3 were able to bind histone tails. Moreover, the glass array screening suggests that they prefer H4 peptides with acetylation marks. The next step in the characterization of the domains was to find the binding pocket for the acetylated lysine and disrupt it to prove its function. The search for the residues involved in the acetyllysine interaction is not trivial for several reasons. First, there are only two reported PHD fingers that bind to acetyllysine, and they do so in tandem arrangement of two PHD fingers. Second, the presence of three PHD fingers complicates the search for the binding pocket, because it could be found either in one of the individual PHD fingers, or at the interface between two PHD fingers. So far, no PHD finger triplet structure has been solved, so it is unknown in which alternate way they might assemble together.

So far, mutants were designed and produced only for MLL2 PHD 3-5. Because the PHD finger group of MLL3 is so similar to the one in MLL2, the design of the MLL3 mutants will be easier once the target amino acids are found in MLL2 PHD 3-5.

For the identification of the residues which might have a role in the H4K16ac interaction, analysis was done on each PHD finger separately. Because none of the studied PHD fingers have solved structures, a protein BLAST (Basic Local Alignment Search Tool) provided by NCBI (National Center for Biotechnology Information) was done with the sequence of each of the MLL2 PHD 3, 4 and 5, as query. The search revealed the existing human PHD fingers with the closest sequence similarity to the MLL2 domains and with solved crystal or NMR structures.
These structures served further as templates to replace the unknown structure of the target MLL2 PHD fingers in a structural superposition to the known acetyl binders DPF3b (PDB ID: 2KWN) and MOZ (PDB ID: 4LLB).

Both DPF3b and MOZ are tandem PHD fingers, but the superposition of the structures was done only with the PHD finger responsible for the acetylysine binding. The iterative magic fit option was used in Swiss-PdbViewer V4.1.0 to best fit backbone atoms of the domains. Once this was achieved with a root mean square error (RMS) lower than 2 Å, the residues of the template domain which had their side chains in proximity of a few Angstroms from the acetylysine were identified and then traced back to the MLL2 PHD domain by direct sequence alignment. In this way, residues that might serve to form the binding pocket of the target lysine could be determined and mutagenized. An overview of the identification of residues for mutagenesis is shown in Table 9.

The BLAST search with the sequence of MLL2 PHD3 resulted in the identification of two structures that showed high similarity to MLL2 PHD3, and could be used as templates for the structure alignments: the PHD domain of UHRF2 (PDB ID: 2E6S) and PHD1 of the tandem PHD structure of MLL3 PHD1-2 (PDB ID: 2YSM). They were superimposed on the DPF3b PHD1 and MOZ/MYST3 PHD1.

Table 9 - Overview of the process of identification of the MLL2 PHD 3-5 binding pockets. The templates are the PHD fingers identified by MLL2 PHD sequence BLAST and used for structural alignment with acetyllysine-binding PHD fingers. The templates are needed to replace the MLL2 PHD fingers because they have unsolved structures.

<table>
<thead>
<tr>
<th>Target MLL2 PHD finger</th>
<th>Template used for structural alignment (identified by protein BLAST)</th>
<th>Acetyllysine PHD domain structure with bound peptide used for structural alignment</th>
<th>Corresponding figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLL2 PHD3 (1379-1431)</td>
<td>MLL3 PHD1 (344-389) UHRF2 PHD (341-395)</td>
<td>DPF3b PHD1 (261-320)</td>
<td>Figure 43, A</td>
</tr>
<tr>
<td></td>
<td>UHRF2 PHD (341-395)</td>
<td>DPF3b PHD1 (261-320)</td>
<td>Figure 43, B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MOZ/MYST3 PHD1 (194-270)</td>
<td>Figure 43, C</td>
</tr>
<tr>
<td>MLL2 PHD4 (1429-1475)</td>
<td>DPF3b PHD1 (261-320)</td>
<td>N/A (direct sequence alignment)</td>
<td>Figure 44, A</td>
</tr>
<tr>
<td></td>
<td>UHRF1 PHD (300-364)</td>
<td>DPF3b PHD1 (261-320)</td>
<td>Figure 44, B</td>
</tr>
<tr>
<td></td>
<td>UHRF1 PHD (300-364)</td>
<td>MOZ/MYST3 PHD1 (194-270)</td>
<td>Figure 44, C</td>
</tr>
<tr>
<td>MLL2 PHD5 (1506-1557)</td>
<td>MLL1 PHD3 (1564-1628)</td>
<td>DPF3b PHD1 (261-320)</td>
<td>Figure 45, A</td>
</tr>
<tr>
<td></td>
<td>MLL1 PHD3 (1564-1628)</td>
<td>MOZ/MYST3 PHD1 (194-270)</td>
<td>Figure 45, B</td>
</tr>
</tbody>
</table>
Identification of the MLL2 PHD3 acetyllysine-binding pocket

Figure 43 – Identification of the putative MLL2 PHD3 acetyllysine-binding pocket. Superpositions of template PHD finger structures to PHD domains reported to bind to acetyllysine. Magic iterative fit performed in Swiss-PdbViewer v4.1.0, for the backbone atoms only. The number of atoms involved in the fit and the RMS in Å are noted in black. The left panels show a complete superposition of the two domains in ribbon form. The template domain replacing MLL2 PHD3 is green, while the domain binding acetyllysine is blue. The histone peptide is shown in orange sticks, with the acetyllysine residue cyan. The right panel shows a magnification of the acetyllysine putative pocket of the template domain, with only the residues found close to the acetyllsine show in green stick representation. The lower panel shows an alignment of MLL2 PHD3 with the template domain. The residues corresponding to the putative acetyllysine binding pocket are highlighted with a red asterisk and black frames. A. Superposition of MLL3 PHD1 (res. 344–389) and DPF3B PHD1 (261–320) bound to H4K16ac peptide. B. Superposition of UHRF2 PHD (res. 341–395) and DPF3B PHD1 (res. 261–320) bound to H4K16ac peptide. C. Superposition of UHRF2 PHD (res. 341–395) and MOZ/MYST3 PHD1 (res. 194–270) bound to H3K14ac peptide.
In the case of the template MLL3 PHD 1-2, sequence alignment analysis revealed that PHD1 is more similar to MLL2 PHD3, so MLL3 PHD1 structure was used in the structure superpositions to DPF3b PHD1 (Figure 43, A). After the superposition of MLL3 PHD1 and DPF3b PHD1, three residues had side chains near the acetyllysine: Val 346, Leu 371 and Pro 386, which correspond to MLL2 residues Val 1382, Val 1409 and Val 1425, respectively (Figure 43, A; Table 10).

From the fitting of the UHRF2 PHD structure to the DPF3b PHD1 structure, five residues which might form the acetyl binding pocket were identified: Val 349, Tyr 369, Asn 375, Cys 389 and Pro 390, corresponding to Val 1382, Tyr 1404, Asn 1410, Cys 1424 and Val 1425 in the third PHD domain of MLL2 (Figure 43, B; Table 10). The UHRF2 PHD template was also fitted to the first PHD domain of MOZ, to retrieve some more possible residues that might influence acetyllysine binding: Val 367, Tyr 369, Cys 389 and Pro 390. These amino acids were traced back through the alignment with MLL2 PHD3 to Glu 1402, Tyr 1404, Cys 1424 and Val 1425, respectively (Figure 43, C; Table 10).

All three structural superpositions for the identification of the putative MLL2 PHD3 acetyllysine binding pocket were similarly close, with RMS values ranging from 1.49 Å to 1.52 Å (Figure 43, A).

The PHD4 of MLL2 (res. 1429-1475) shows high similarity to the PHD finger of UHRF1 (PDB ID: 3ZVY) and to the PHD fingers of DPF3b itself. Since DPF3b PHD1 is one of the domains that actually bind to acetyllysines, a direct sequence alignment was used in order to identify MLL2 PHD4 residues corresponding to the residues of DPF3b that interact with H4K16ac and H3K14ac (Figure 44, A). It revealed that Glu 1431 and Val 1432 of MLL2 match to Asp 263 and Phe 264 of DPF3b, respectively, which directly interact with the acetyl group of the target lysine (Zeng et al., 2010). Ile 1450, Leu 1457, Trp 1469, Cys 1471 and Lys 1472 of MLL2 are corresponding to Arg 289, Leu 296, Trp 311, Cys 313 and Ile 314 of DPF3b, which form the hydrophobic pocket for the lysine side chain binding (Figure 44, A, Table 10).

Next, the UHRF1 PHD domain structure was fitted to both PHD1 of DPF3b (Figure 44, A), and PHD1 of MOZ (Figure 44, B), to get more possible hits of acetyllysine proximal residues. Out of the superposition with DPF3b, Leu 320, Phe 340, Trp 358, Cys 360 and Pro 361 of UHRF1
were selected, and these led back to Val 1432, Tyr 1452, Trp 1469, Cys 1471 and Lys 1472 of MLL2 (Figure 44, B, Table 10).

Identification of the MLL2 PHD4 acetyllysine-binding pocket

| A. Direct sequence alignment
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DPF3b PHD1 (res. 261-320)</td>
</tr>
</tbody>
</table>

| B. UHRF1 PHD as template for MLL2 PHD4
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DPF3b PHD1 (261-320)</td>
</tr>
<tr>
<td>UHRF1 PHD (300-364)</td>
</tr>
<tr>
<td>144 backbone atoms</td>
</tr>
</tbody>
</table>

| C. UHRF1 PHD as template for MLL2 PHD4
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MOZ/MYST3 PHD1 (194-270)</td>
</tr>
<tr>
<td>UHRF1 PHD (300-364)</td>
</tr>
<tr>
<td>152 backbone atoms</td>
</tr>
</tbody>
</table>

Figure 44 - Identification of the putative MLL2 PHD4 acetyllysine-binding pocket. Superpositions of template PHD finger structures to PHD domains reported to bind to acetyllysine. A. Direct sequence alignment between DPF3b PHD1 (res. 261-320) and MLL2 PHD4 (res. 1429-1475), based on their sequence similarity. The putative acetyllysine-interacting residues are highlighted with black frames and red asterisks. B and C. Magic iterative fit performed in Swiss-PdbViewer V4.1.0, for the backbone atoms only. The number of atoms involved in the fit and the RMS in Å are noted in black. The left panels show a complete superposition of the two domains in ribbon form, with the UHRF1 PHD template domain, replacing MLL2 PHD4, in green, and the domain binding acetyllysine in blue. The histone peptide bound by the acetyllysine binding domain is shown in orange sticks, with the acetyllysine residue cyan. The H3 peptide bound by UHRF1 PHD is shown in magenta. The right panel shows a magnification of the acetyllysine putative pocket of the template UHRF1 domain, with only the residues found close to the acetyllysine show in green stick representation. The lower panel shows an alignment of MLL2 PHD4 with the template domain. The residues corresponding to the putative acetyllysine binding pocket are highlighted with a red asterisk and black frames. B. Superposition of UHRF1 PHD (res. 261-320) and DPF3b PHD1 (261-320) bound to H4K16ac peptide. C. Superposition of UHRF1 PHD (res. 261-320) and MOZ/MYST3 PHD1 (res. 194-270) bound to H3K14ac peptide.
Fitting the UHRF1 PHD structure to MOZ PHD1 was better than the fit to DPF3b, with a RMS of 1.13 Å compared to 1.74 Å. The superposition of MOZ resulted in identification of four relevant residues: Phe 340, Trp 358, Cys 360 and Pro 361, which correspond to Tyr 1452, Trp 1469, Cys 1471 and Lys 1472 of MLL2 PHD4 (Figure 44, C, Table 10).

The fifth PHD finger of MLL2 showed close similarity to PHD3 of MLL1 enzyme (PDB ID: 2KYU) in a BLAST search using MLL2 PHD5 (res. 1506-1557) sequence as query. The fitting of MLL1 PHD3 to DPF3b and MOZ was accurate, with RMS values of 1.43 Å and 1.52 Å, respectively (Figure 45). In the superposition with DPF3b, five residues of MLL1 PHD3 were shown to be close enough to the acetyllysine to have possible roles in interacting with it: Pro 1570, Leu 1571, Glu 1600, Cys 1624 and Val 1625. These amino acids correspond to positions 1508 (Pro), 1509 (Ile), 1535 (Glu), 1553 (Cys) and 1554 (Val) of MLL2 PHD5 (Figure 45, A, Table 10). The superposition between MLL1 PHD3 and MOZ PHD1 identifies four MLL1 potential acetyllysine interacting residues: Pro 1570, Val 1595, Tyr 1622 and Cys 1624, which are represented by Pro 1508, Phe 1551, Cys 1553 and Met 1530 of MLL2 (Figure 45, B, Table 10).

All the results of the superpositions are summarized in Table 10. The residues that were identified as putative acetyllysine interactors from the MLL2 PHD domains and which appeared more than once as results of the superpositions were considered more likely to be hits and taken into consideration for mutagenesis. Also residues such as glutamate or aspartate were preferentially chosen due to the known importance of the Asp 263 of DPF3b in interacting with the acetyl group of the acetyllysine (Zeng et al., 2010). Although in each PHD finger a cysteine residue was found in the results batch, corresponding to the acetyllysine interacting Cys 313 of DPF3b, the cysteines could not be considered for mutagenesis, because they are essential for the structural integrity of the PHD fingers, as they are residues coordinating the zinc ion. For each PHD finger, two amino acids were chosen for initial mutagenesis out of all the residues identified previously. The mutagenesis was performed to test whether exchanges of the residues to alanine affect the histone peptide binding of the domains. The cloning and purification of the mutants was successful in the case of MLL2 PHD 3-5 E1431A, Y1452A, W1469A, I1509A, M1530A and E1535A. The cloning of MLL2 PHD 3-5 Y1404A was unsuccessful at the stage of sequencing and will be repeated.
In the cloning of the MLL2 PHD 3-5 V1425A mutant, sequencing revealed the supervising incorporation of a second, E1431A mutation.

The following experiments were performed with all seven mutants that were purified successfully, including the PHD 3-5 V1425A/E1431A double mutant.
Table 10 - Overview of the residues identified via structural superimpositions and alignments to the DPF3b and MOZ tandem PHD fingers.

<table>
<thead>
<tr>
<th>MLL2 PHD domain</th>
<th>MLL2 PHD residue</th>
<th>Template residues from superpositions</th>
<th>Superposition/alignment partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLL2 PHD3</td>
<td>Val 1382</td>
<td>Val 349 (UHRF2)</td>
<td>UHRF2 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Val 346 (MLL3 PHD1)</td>
<td>MLL3 PHD1 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Glu 1402</td>
<td>Val 367 (UHRF2)</td>
<td>UHRF2 vs. MOZ</td>
</tr>
<tr>
<td></td>
<td>Tyr 1404</td>
<td>Tyr 369 (UHRF2)</td>
<td>UHRF2 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Val 1409</td>
<td>Val 1409 (MLL3 PHD1)</td>
<td>MLL3 PHD1 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Asn 1410</td>
<td>Asn 375 (UHRF2)</td>
<td>UHRF2 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Cys 1424</td>
<td>Cys 389 (UHRF2)</td>
<td>UHRF2 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Val 1425</td>
<td>Pro 390 (UHRF2)</td>
<td>UHRF2 vs. DPF3b</td>
</tr>
<tr>
<td>MLL2 PHD4</td>
<td>Glu 1431</td>
<td>Asp 263 (DPF3b)</td>
<td>MLL2 PHD5 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Val 1432</td>
<td>Leu 320 (UHRF1)</td>
<td>UHRF1 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Ile 1450</td>
<td>Arg 289 (DPF3b)</td>
<td>MLL2 PHD5 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Tyr 1452</td>
<td>Phe 340 (UHRF1)</td>
<td>UHRF1 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Leu 1457</td>
<td>Leu 296 (DPF3b)</td>
<td>MLL2 PHD5 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Trp 1469</td>
<td>Trp 358 (UHRF1)</td>
<td>UHRF1 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Cys 1471</td>
<td>Cys 360 (UHRF1)</td>
<td>UHRF1 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Lys 1472</td>
<td>Pro 361 (UHRF1)</td>
<td>UHRF1 vs. DPF3b</td>
</tr>
<tr>
<td>MLL2 PHD5</td>
<td>Pro 1508</td>
<td>Pro 1570 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Ille 1509</td>
<td>Leu 1571 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. MOZ</td>
</tr>
<tr>
<td></td>
<td>Met 1530</td>
<td>Val 1595 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Glu 1535</td>
<td>Glu 1600 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Phe 1551</td>
<td>Tyr 1622 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. MOZ</td>
</tr>
<tr>
<td></td>
<td>Cys 1553</td>
<td>Cys 1624 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. DPF3b</td>
</tr>
<tr>
<td></td>
<td>Val 1554</td>
<td>Val 1625 (MLL1 PHD3)</td>
<td>MLL1 PHD3 vs. DPF3b</td>
</tr>
</tbody>
</table>
The domains were purified at the same time with the MLL2 PHD3-5 wild type fragment, and a gel picture of their preparation is shown in Figure 46.

The binding of the domains was tested with the peptide array binding assay, using the MODified™ Histone Peptide Arrays. However, all the mutated proteins showed the same binding specificity as the wild type domain (Figure 47). Therefore, at this point, the acetyllysine binding pocket has not been identified yet.

Once the binding pockets of the three domains are identified, the same procedure can be repeated to discover the binding pockets of the MLL3 PHD 4-6 domains, which showed binding...
preferences similar to MLL2 PHD 3-5 on the peptide arrays. To sum up, most of the cloned and purified PHD fingers of MLL2 and MLL3 did not bind histone peptides.

However, the two PHD triplets, MLL2 PHD 3-5 and MLL3 PHD 4-6 showed strong and specific binding to several histone peptides, both from H3 and from H4. These PHD finger groups seemed to have similar preferences, towards H3 (16-35), H3 (26-45) and H4 (11-30). The strongest binding was observed on the H4 (11-30) peptides, especially when they were containing H4K16ac, together with other acetylations, lysine methylations or arginine dimethylations. In order to identify the binding pocket for acetyllysine, a comparison was made to other acetyllysine-binding PHD fingers and a first round of mutagenesis was performed on MLL2 PHD 3-5 to pinpoint the acetyllysine-interacting residues. The mutants did not lose binding to the target peptides, and therefore, the next step is to change the mutagenesis approach. The project is still in full development.

4.5. Study of the Specificity of Various SET Domain Methyltransferases

SET domain methyltransferases have been so far identified as protein lysine methyltransferases only. They have essential roles in the regulation of many epigenetic pathways, by the methylation of histones and multiple non-histone targets as well. In humans, there are 56 identified SET domain methyltransferases, and the majority of them have histone substrates (Clarke, 2013). However, rapidly increasing numbers are proving to be also able to methylate other proteins. For this reason, the study of the substrate sequence specificity of these PKMTs is important, in order to help with the identification of novel SET domain methyltransferase non-histone targets. This study attempted to investigate the specificities of several PKMTs, namely SET1A, SET1B, MLL2, MLL5, NSD3, and EZH1.

Initially, the catalytic domains, including the SET domain and all other associated domains, such as Pre-SET, Post-SET or AWS domains, were cloned into expression vectors as GST-fusions. Subsequently, their activity was tested on recombinant histone 3, in an in vitro methylation reaction performed in the presence of radioactively labeled SAM. Once the activity of the enzyme was validated at this level, the SET domain was further investigated to determine its substrate sequence preference. The enzymatic activity assay based on tritium-containing methyl group transfer used for this purpose was a method already established in our group (Rathert et al., 2008). This assay involves the methylation of a modified H3 peptide array using radioactively labeled SAM as cofactor. The array comprises...
240 peptide spots, with each spot containing peptides with one single amino acid replacement in the wild type sequence, against each of 18 amino acids (cysteine and tryptophan were omitted due to low peptide synthesis yield). From now on, this array will be referred to as specificity profile array. Differential intensity of the methylation signal relative to the wild type sequence defines the specificity profile of the enzyme. After summing up all the assay data, a putative specificity profile is defined for the enzyme and this is used as input for the search of novel enzyme targets in the Scansite3Beta database (http://scansite3.mit.edu). Nuclear proteins that have peptide sequences corresponding to the specificity motif can be then tested for methylation by the PKMT, first at peptide level, and then at protein level.

The initial methylation assay with recombinant H3.1 histone as substrate was performed at 37°C as described in section 3.5.2. After only one day of exposure of the dried SDS-PAGE gel to the autoradiography film, strong signal was detected for the methylation of H3.1 by the MLL2 SET domain (Figure 48, A).

No other the SET domain purified presented any methylation activity (Figure 48, A and B). Extending the exposure time to one week did not change the result (data not shown). In the case of the SET domains that were inactive in the initial conditions, the temperature of the reactions was lowered to 25°C, which resulted in weak H3.1 methylation signal in the case of SET1B after 10 days of film exposure, and no signal in all other cases (Figure 48, C). For the optimization of the reaction conditions, literature searches for different buffer compositions were performed, and the reactions were repeated as follows: EZH1 SET methylation with

![Methylation reaction: 37°C for 2 hours](image)

**Figure 48 - In vitro methyltransferase assay with radioactive SAM as cofactor and recombinant H3.1 as substrate.** A and B. The first attempt to methylate H3.1 using the SET domains of EZH1, MLL2, MLL5, SET1B, SET1A and NSD3. The reaction was incubated for 2 hours at 37°C. The autoradiography film was exposed to the dried SDS-PAGE gel for 1 day. C. Repetition of the methylation reaction at 25°C with the enzymes that showed no activity, with otherwise the same conditions as in A and B. The film was exposed to the gel for 10 days.

121
methylation conditions from Margueron et al., (2008), MLL5 SET domain reaction with buffer reported in Fujiki et al., (2009), NSD3 SET methylation with the conditions presented in Kim et al., (2006) and SET1A and SET1B SET domains as described in Wysocka et al., (2003). Still, no signal was detected in any of the cases for the H3.1 methylation after a week of film exposure (data not shown).

The MLL2 SET domain, which showed high methylation activity in the \textit{in vitro} recombinant histone methylation assay, was next tested on a small peptide spot array containing unmodified peptides centered on prominent histone modification sites: H3K4/K9 (H3 peptide, residues 1-20), H3K27 (H3 peptide, residues 15-35), H3K36 (H3 peptide, residues 28-48), H3K79 (H3 peptide, residues 67-87) and H4K20 (H4 peptide, residues 10-30). As previously reported in literature (Ansari et al., 2009), MLL2 SET methylated specifically the peptide containing H3K4 (Figure 49, A).

Both MLL2 SET and also the other two domains which showed very weak activity, SET1A and SET1B, were tested by Sara Weirich in the presence of the core complex members WDR5, ASH2L and RBBP5. MLL3, for comparison, and the core complex members were kindly provided by Sara Weirich. The methylation reaction was performed at 25°C and the film was exposed for 1 day.

MLL2, just like its relative MLL3, cloned and purified by Sara Weirich, showed weak activity alone, compared to much stimulated activity within the complex (Figure 49, B). SET1A alone presented no methylation activity under the assay conditions, while SET1A in complex presented strong methylation of H3.1 (Figure 49, B). SET1B showed no activity, either alone.
or in complex (Figure 49, B). The lack of methylation signal in the case of SET1B alone is explained by the fact that the film was exposed for only 1 day, in comparison to the 10 days exposure in the case of the result in Figure 48, C.

In order to study the specificity profile of MLL2 SET, both alone and together with the complex members, the peptide spot membrane methylation reaction was repeated on an array consisting of H3K4 wild type and modified peptides.

![Figure 50 - Methylation pattern resulting from the specificity profile array methylation with MLL3-SET (A), MLL2-SET (B) and MLL2 SET within the core complex (C).](image)

The specificity profile pattern of MLL2 SET alone was very similar, if not almost identical, to the specificity profile pattern of the single MLL3 SET on the specificity profile array (Figure 50). The specificity profile array analysis of MLL3 SET was performed and kindly provided by Sara Weirich. Both enzymes seemed to have specific preferences from A1 to R8 of the H3 peptide, while afterwards they did not seem to discriminate between residues (Figure 50).

For the analysis of the results, the intensity of the spots was recorded using the Phoretix™ Array software, and the same software was used for initial background subtraction using the rectangle method. This involved the selection of a rectangle in the image that contained only background signal. Always a rectangle containing no peptide spots was chosen. Next, the intensity of the spots was normalized to the maximum intensity and to the minimum intensity.
The analysis of the MLL2 SET methylating alone showed that some replacements, such as A1L, T6I, T6V, A7K, or A7M were favored by the SET domain, while most other replacements in the A1 to R8 positions caused reduction of MLL2 SET activity (Figure 51, A).

The MLL2 core complex favored replacements such as A1L, T3F or T3V, and disfavored many of the replacements in the A1 to R8 region. However, the selectivity of MLL2 was weaker in complex than alone. As one can observe, the complex showed fewer red boxes in the A1 to R8 columns (Figure 51, B) when compared to the same columns of the MLL2 SET alone array (Figure 51, A). It remains to be seen whether the MLL2 complex specificity is reproducible.
In order characterize the binding specificities of the enzyme in both situations, a specificity profile was derived from the analysis of the methylation arrays.

The discrimination factors of each amino acid from the H3 tail were calculated as described previously (Rathert et al., 2008), and were plotted in a bar diagram (Figure 52, A and B). The resulting specificity profile of the enzyme is written underneath the graph (Figure 52, A and B). The specificity profile of MLL2 is very similar to the previously determined MLL3 SET profile, in the unpublished work of Sara Weirich (data not shown). The MLL3 SET profile is: \([R]-2 \text{[ARCILMFTYV]}-1 \text{[K]}0 \text{[QNRHKMF]}+1 \text{[ACILSTV]}+2 \text{[A-Z]}+3 \text{[R]}+4\). In comparison to the MLL2 SET alone, MLL3 also prefers R and F at position +1, and S at position +2. The specificity profile array used for MLL3 contained also cysteine and tryptophan in the exchange analysis, and therefore, the results include cysteine as option at positions -1 and +2. The preference of MLL2 towards these amino acids at those particular positions cannot be stated presently.

Keeping in mind that the MLL2 SET alone and in complex showed visibly different specificities, a comparative analysis was achieved by presenting the ratio of the intensity values of the
complex versus those of MLL2 alone. Most of the amino acid exchanges appeared in a neutral (yellow) range, but some striking differences between the two MLL2 states were also observed (Figure 53). For example, exchanges such as A1D, A1Q, A1E, A1K, R2D, R2G, or T3D seemed disfavored by the complex compared to the MLL2 SET domain alone (Figure 53). In contrast, changing Q5 to L or F, or R8 to D, E or L resulted in higher stimulation of the complex compared to the MLL2 SET without complex partners (Figure 53).

At this point, the next step is the optimization of the substrate peptide for the specificity profile analysis. In order to validate the conclusions drawn from the comparison of the MLL2 methylation alone and in complex, a set of random peptides will be synthesized by peptide SPOT synthesis. The peptides will have their sequence originating from the wild type H3 with multiple replacements of residues according to the favored and disfavored exchanges of the complex relative to the lone MLL2 SET, observed in Figure 53. These peptides have already been designed and will be tested shortly.

MLL2 SET alone was already used for the methylation of a set of peptides from selected non-histone targets of MLL3, based on the specificity profile mentioned above. Because MLL2 and MLL3 SET domains had very similar specificities, it was possible that the targets of MLL3 would also become methylated at peptide level by MLL2.
MLL2 SET methylated five MLL3 target peptides other than H3: Chromodomain-helicase-DNA-binding Protein 3 (CHD3), Hepatocyte nuclear factor 4α, Protein MCM10 homolog, Zinc finger protein 862 and ZZ-type zinc finger-containing protein 3 (Figure 54). Out of these, only CHD3 and Protein MCM10 homolog were methylated in par with H3 (Figure 54).

**Figure 54** - Result of the methylation by MLL2 SET of an array containing peptides from putative non-histone targets of MLL3 SET, based on the MLL3 specificity profile. The methyltransferase assay was performed at room temperature. The film was incubated with the membrane for one week. The peptides that show methylation clearly above background level are numbered and identified in the legend.

MLL2 SET methylated five MLL3 target peptides other than H3: Chromodomain-helicase-DNA-binding Protein 3 (CHD3), Hepatocyte nuclear factor 4α, Protein MCM10 homolog, Zinc finger protein 862 and ZZ-type zinc finger-containing protein 3 (Figure 54). Out of these, only CHD3 and Protein MCM10 homolog were methylated in par with H3 (Figure 54).

**Figure 55** - Methylation of MODified™ Histone Peptide Array by the MLL2 SET complex. The minimized left and right panels show the two individual replicates, one with annotations and one without. Below, a magnification of the first replicate is presented. Methylation occurs at the peptides with H3K9me2 (blue boxes) and H3K9me3 (red boxes). The unmodified H3 peptide, the first peptide on the array, shows very weak, background level signal. The strongest peptides are the single H3K9me2 and H3K9me3 peptides (the first blue and red box, respectively). Inhibition of the methylation activity occurs when S10P, T11P, and R8Citr (green arrows) are present. Any modification of the H3K4, including mono- and dimethylation (red arrows) results in loss of stimulation activity.
The four proteins will be cloned and purified either as domains or full length proteins fused to GST, and the activity of MLL2 SET towards them will be verified. This will be done in parallel with a screen for non-histone targets based on the specificity profile of MLL2 SET itself, obtained by the search of nuclear proteins on ScanSite3Beta using the specificity profile as a query.

Finally, MLL2 SET was used, both alone and in complex, to methylate MODified Histone Peptide Arrays, in order to investigate whether the substrate modification status has any effect on the methylation activity. For the conditions used, there was no methylation signal visible for MLL2 SET alone, but the enzyme in complex did methylate a set of targets. The analysis of the spot intensities revealed that the enzyme within the core complex preferred methylating peptides that are di- or trimethylated at H3K9 (Figure 55). This methylation was inhibited by the presence of S10 or T11 phosphorylations (Figure 55). The H3K9me3-modified substrate methylation was also lost when K4 was modified, even when the modification was only mono- or dimethylation (Figure 55). The binding of trimethylated K9 at position +5 fits to the specificity profile of the MLL2 SET core complex. At this position, any residue was permitted, but mostly hydrophobic residues were preferred, such as leucine, methionine and phenylalanine (Figure 51, B). These are only preliminary results, which need confirmation through different assays. Work is still in progress on the MLL2 SET specificity project.

The SET1A SET domain was also active in the presence of the core complex members WDR5, PPARBP and ASH2L. The specificity profile array was methylated by the SET1A complex (Figure 56, A), and the data were analyzed in the same way as for the MLL2 SET domain. After background subtraction and normalization, the intensity values were represented as an array of boxes colored using a red-yellow-green scale (Figure 56, B). The color code emphasizes the fact that residues A1 to R8 were important for the recognition by SET1A complex (Figure 56).

Most of the replacements of the canonical histone sequence substrate resulted in a strong decrease or loss of activity, represented by red boxes in Figure 56, B. However, there were a few residues which were permitted as replacements, and in some cases even preferred. This is the case of A1I, A1L, T3I, T3L, T3M, T3F, T3Y and T3V (Figure 56, B). In position A7, most of the amino acids were accepted (Figure 56, B). It is obvious both from the analysis in Figure 56, B and also visible by eye from Figure 56, A, that there were three replacements which cause SET1A complex stimulation: A1L, T3F and T3Y.

128
These replacements were considered in the search of non-histone targets, by doing sequential restrictive specificity profile searches with each of these modifications as the only query for their position in the target profile sequence.

The specificity profile was derived from the processed data set, by the calculation of the discrimination factor for each amino acid replacement and the plotting of the results as a bar diagram (Figure 57). The specificity profile \( [A\text{ILPV}]-3/[R] -2/[A\text{TLILMFYV}]-1/[K]0/[Q]+1/[TILV]+2/[A-Z] \) (not P) +3/[RFY]+4 was used as query in the ScanSite3Beta database search, to identify putative non-histone targets. In addition, the restricted specificity profiles with only [L] at position -3, or only [F] at position -1, or [Y] at position -1, were also used as query for nuclear proteins on ScanSite3Beta.

All the results of the searches will be synthesized as peptide spots in an array for non-histone target methylation screening.

Figure 56 - Specificity profile array methylation by SET1A within the core complex. The first column contains only H3 wild type peptides (WT). Starting from the second column, each amino acid of the H3 peptide tail, starting from A1 and continuing until A15, is consecutively exchanged with all the amino acids labeled on the left on each row of the membrane. The target K4 is written in red in panel A. Image of the autoradiography film showing the source of the raw intensity data. The film was incubated with the membrane for two days. B. Analysis of the raw data after background subtraction and normalization from one experimental repeat. The wild type peptides are found in the first column and within the array, labeled with WT. The color scale on the right explains the coloring of the boxes. Red represents exchanges that are not favorable for the
The SET1A SET domain complex methylation specificity characterization and non-histone target identification is still in progress. The PKMTs MLL2/MLL3 and SET1A are all H3K4 methyltransferases, but their substrate recognition, although similar, was not identical (Table 11).

All three enzymes had a very strict requirement for an arginine at position -2, and were permissive for most residues at position +3 (Table 11). MLL2/MLL3 also strictly allowed only arginine at position +4, while SET1A also accepted F and Y in addition to R at this position (Table 11). At position -3, most small hydrophobic residues were accepted by all the SET domains. However, MLL2 also preferred arginine at this position (Table 11). Position -1 was also receptive of some hydrophobic and bulky amino acids. Arginine was additionally accepted by the MLL2/MLL3 pair compared to SET1A, while MLL2 also methylated when a glutamine was present at position -1 (Table 11). Position 2 accommodated most small amino acids, with the highest selectivity of the three SET domains observed in the case of SET1A, which only accepted threonine, isoleucine, leucine and valine at this position (Table 11). Finally, at position +1, SET1A was very strict, permitting methylation only if glutamine was found immediately after the target lysine. Both MLL2 and MLL3 accepted several bulky, mostly polar residues at position +1 (Table 11).
Table 11 - Overview of the specificity profiles of SET1A SET, MLL2 SET and MLL3 SET.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Preferred amino acids at each position of the substrate peptide, relative to the target lysine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3</td>
</tr>
<tr>
<td>SET1A</td>
<td>AILPV</td>
</tr>
<tr>
<td>MLL2</td>
<td>ARILTV</td>
</tr>
<tr>
<td>MLL3</td>
<td>A-Z</td>
</tr>
</tbody>
</table>

The identified specificity profiles give an indication of the variety of non-histone proteins which might still be unidentified substrates of these enzymes. This study will continue to identify and characterize these putative novel PKMT targets and the possible roles of the methylation marks introduced by the methyltransferases.

SUV39H1 is another SET domain protein lysine methyltransferase, with critical roles in the formation and maintenance of heterochromatin, by the methylation of lysine 9 of histone H3 (Peters et al., 2001). The enzyme contains C-terminally located Pre-SET, SET and Post-SET domains, and an N-terminal chromodomain (Figure 58).

The chromodomain was shown previously to stimulate in vitro the catalytic activity of SUV39H1 (Chin et al., 2006), but no follow up data was published. The purpose of the project presented here was to characterize the binding specificity of the SUV39H1 chromodomain, to confirm the stimulatory effects of the domain in vitro, and to make observations about the consequences of this stimulatory role in vivo.

For the investigation of the function of the SUV39H1 chromodomain, the domain was cloned with two sets of boundaries: SUV39H1 residues 43-95 (53 amino acids long) designed by Dr. Arunkumar Dhayalan, and SUV39H1 residues 43-106, based on Wang et al., (2012). Both protein domains were successfully purified as GST fusions (Figure 15), but only the first one was used in future experiments. The decision was made based on the protein concentration after purification. An aromatic cage mutant, Y67A, reported previously as essential for H3K9me3 binding (Chin et al., 2006), was cloned and purified along with the wild type, to serve as control in the binding assays.

For an initial comprehensive screening, the binding to the MODified™ Histone Peptide Arrays was assessed, and the resulting array image was annotated to facilitate its interpretation. As observed in Figure 59, A, the assay was highly reproducible on the two array duplicates.

The chromodomain of human SUV39H1 bound specifically to H3K9me3-containing peptides (Figure 59, A, red boxes). The binding to H3K9me2 was also observed, but was very weak,
comparative to background levels at most spots (Figure 59, A, blue boxes). A reduction of signal intensity was observed in the presence of H3S10P and H3T11P (Figure 59, A, green arrows).

The analysis of the spot intensities also showed a very strong specificity for the H3K9me3 modification, and no other modification appeared to be in any way preferred, judging by the specificity factors calculated for each modification present in the array (Figure 59, C).

Upon a close observation of the modified histone peptide array, there were three highly preferred peptide spots. The first one was the A15 peptide (Figure 59, A, red box), which contained only one modification, the trimethylated H3K9, which suggests that the reading domain preferred binding to histone tails lacking other modifications in the proximity of the target H3K9me3. Secondly, C13 was also highly enriched in signal (Figure 59, A, red box). This spot contained R8Citr adjacent to the H3K9me3 mark, a modification which neutralizes the charge of the arginine residue. The R8Citr appeared not to disturb, but rather to stimulate the chromodomain binding to its H3K9me3 target. Thirdly, D1, a peptide spot containing H3K9me3 together with H3K14 acetylation, was also strongly bound in comparison to the rest of the peptides (Figure 59, A, red box).

The integrity of the aromatic cage of the SUV39H1 chromodomain was critical to the specific histone tail binding capabilities of the domain, as proven by the inability of the SUV39H1 chromodomain Y67A mutant to bind to the histone peptide array (Figure 59, B).

The validation of the data obtained through the peptide spot binding assay was done with fluorescence anisotropy measurements. The binding of the SUV39H1 chromodomain in wild type and Y67A mutant form was recorded by titrations of fixed amounts of FITC-labeled H3K9me3 and H3K9me3/S10P peptides, by varying the chromodomain concentration in the assay.
The wild type domain bound to H3K9me3 specifically, with a binding constant ($K_d$) of approximately 100 $\mu$M (Figure 60, B). The affinity was strongly decreased in the case of the peptide which had an additional phosphorylation at serine 10 (Figure 60, A). In the same range was also the binding of the mutant chromodomain to the target H3K9me3 peptide (Figure 60, A and B).

**Figure 59** – Binding specificity investigation of the SUV39H1 chromodomain. A. Binding assay studying the interaction of the GST-tagged recombinant SUV39H1 chromodomain (CD) to immobilized peptides on a spot array (Active Motif) containing histone tail peptides with various PTMs. On the left, magnification of the left side of the array, signifying one repeat of the experiment. The array columns are labeled with numbers from 1 to 24, while the rows are labeled with letters from A to P. Relevant modifications are annotated according to the legend. On the right, minimized image of the second experimental repeat. B. Binding assay repeated with the SUV39H1 chromodomain Y67A mutant, in the same conditions as the wild type domain. C. Analysis of the binding specificity of the wild type CD resulting from the peptide array spot intensities. Bar graph representation of the specificity factors for the histone modifications which showed the ten highest specificity factors.
The binding to the H3K9me3 peptides was very specific on the histone tail arrays, but was not very strong ($K_d = 106 \, \mu M$). This fact prompted the supposition that the H3 tail methylated at position 9 might not be the primary target of the reading domain. To test this hypothesis in a very broad approach, chromatin was prepared according to Henikoff and colleagues (Henikoff et al., 2009) and the nucleosomes and tightly associated proteins extracted during the 150-600 mM salt extraction step were used in far western and SUV39H1 chromodomain pulldown experiments. This is the nucleosome fraction that is obtained as an insoluble pellet after a 150 mM salt extraction of mononucleosomes, and a subsequent extraction of the remaining chromatin using 600 mM salt buffer. The 150-600 mM chromatin fraction was loaded on an SDS-PAGE gel and then transferred to a nitrocellulose membrane that was used in a far western experiment, in which binding of the SUV39H1 chromodomain to the membrane-immobilized chromatin was investigated. GST was used as a negative control to the GST-

**Figure 60 – Histone peptide binding assays results for the SUV39H1 chromodomain (CD) wild type (WT) or Y67A mutant.** A. The titrations were performed in solution, with fluorescently labeled peptides (100 μM) titrated with increasing concentrations of chromodomain and the corresponding anisotropy changes were measured. The results represent the average of 2-3 independent experiments. The error bars represent the standard error of the mean. B. The binding constant values calculated from the data obtained in A from the fluorescence depolarization (FD) measurements. Values higher than 1 mM with large error margins were obtained for the peptides, which are very weakly bound.
tagged chromodomain. The reading domain showed strong binding to various bands on the blot, including bands corresponding to histone H3 and histone H1 (Figure 61, B). The GST protein showed complete absence of binding (Figure 61, A).

Figure 61 - Investigation of the SUV39H1 chromodomain to bind to non-histone targets in chromatin. A and B. Far western using 150-600 mM chromatin fraction blotted on nitrocellulose membrane and detected with GST-fused SUV39H1 chromodomain (B) or GST (A). Antibody detection against the GST tag served as a means of detection for the reading domain/target interaction. Ponceau staining and immunostaining signal on autoradiography film are provided in each case. The sizes of H3 and H1 are indicated by black lines at the corresponding height in panel B. C. Coomassie stained SDS-PAGE gel of samples retrieved from GST-pulldown experiments. 10% of the input was loaded for quantitative estimation of sample enrichment. GST and SUV39H1 chromodomain (SUV39) were treated in parallel in every case. Three different conditions of salt concentration (100 mM, 200 mM and 300 mM KCl) were used for the washing steps. The thick bands between 35 and 45 kDa correspond to the GST and GST-fused SUV39H1 chromodomain proteins.

To identify the proteins that interact with SUV39H1 chromodomain, the same chromatin fractions were used in a native GST-pulldown procedure, with different wash step stringencies: 100 mM salt, 200 mM salt and 300 mM salt conditions (Figure 61, C). Irrespective of the assay stringency, proteins were pulled down both in SUV39H1 sample and in the control GST sample, without any obvious difference (Figure 61, C). The strong carryover effect with the GST sample pulldown was caused by the fact that addition of the 150-600 mM chromatin fraction to the pulldown reaction buffer resulted in a visible precipitation of material. The experiment setup needs to be optimized in order to accommodate the solubility requirements of the chromatin fraction used.

The far western experiment confirmed that SUV39H1 chromodomain bound indeed to the H3 protein, among other putative interactors. From this point forward, the project was based on the investigation of the trimethylated H3K9, as a recruiting agent of the SUV39H1 chromodomain to chromatin. The H3K9me3 mark is also the mark that is set by the SUV39H1 PKMT. Therefore, the binding of the reading domain may be in close relation to the catalytic
activity of the enzyme. One hypothesis would be that the PKMT is recruited to H3K9me3 by its chromodomain and this recruitment results in methylation of neighboring nucleosomes, and consequent spreading of heterochromatin.

To test this hypothesis, full length SUV39H1 enzyme was cloned and purified as a GST-fusion (section 3.6.1). Site directed mutagenesis was performed to produce the Y67A mutant of the full length protein (section 3.6.1).

The expression and purification of the enzymes was established before the beginning of this doctoral work, but the proteins still purified with a large degree of degradation, and at low concentrations, of no more than 10 μM per preparation.

The overall activity of the enzymes on recombinant H3.1 tested by in vitro methylation assays with radioactive SAM as cofactor was very weak. The wild type only showed very weak methylation activity after one week of film incubation with the radioactive gel (Figure 62). The activity of the Y67A mutant enzyme was considerably stronger than that of the wild type, but still weaker than that of the truncated SUV39H1 which only contained the SET domain (Figure 62).

In a similar approach, the in vitro radioactivity-based methyltransferase assay was performed on reconstituted oligonucleosomes containing either unmodified recombinant histones or 50% unmodified and 50% H3Kc9 (MLA)-trimethylated histones. After the comparison and adjustment of the amounts of the two chromatin samples (Figure 63, A), methylation was performed with wild type and Y67A full length SUV39H1 proteins. The activity of the wild type protein was in general much weaker than that of the chromodomain mutant, although comparable amounts of enzymes were used in the reactions (Figure 63, B).
The presence of the trimethylated H3K9 analogs did not affect the methylation activity of the enzymes in any way, neither in the case of the wild type, nor in the case of the chromodomain mutant (Figure 63, B).

The effect of the chromodomain of SUV39H1 on its catalytic activity was studied with the help of one more approach. Peptide arrays were synthesized to serve as substrates for the SUV39H1 wild type and mutant enzymes. Each spot contained either one H3 peptide, unmodified or with different degrees of methylation at H3K9, or a mixture of two different peptides, the unmodified H3K9 peptide serving as substrate for the methyltransferase, mixed with H3 peptides mono-, di- or trimethylated at H3K9 (Figure 64). The purpose of the assay was to observe whether the binding of the chromodomain to the methylated H3 peptides from the mixed spots would stimulate or inhibit the activity of the PKMT recorded on those spots.

Both SUV39H1 wild type and Y67A methylated the unmodified and H3K9me1 peptides, while H3K9me2 was not suitable as substrate. Clr4, the yeast homolog of SUV39H1, was used here for comparison (Figure 64, C). It also preferred unmodified and monomethylated H3K9 as substrate for the methylation reaction, while the H3K9 dimethylation was much weaker. A slight preference for the H3K9me1 substrate peptides was observed for the Clr4 enzyme. Clr4 showed a generally higher activity, and the H3K9me2 methylation signal might be visible because of this higher signal, while in the case of the human SUV39H1 enzymes the H3K9me2 methylation might be below the detectable range. The control H3K9me3 peptides showed no methylation signal for any of the three enzymes.
All three proteins showed approximately the same methylation pattern of the peptide spots with the double modifications (Figure 64), suggesting no stimulatory effect of the chromodomain towards the catalytic activity. However, the enzymes did show very distinct levels of methylation activity: SUV39H1 wild type was weak, the chromodomain mutant was stronger, and Clr4 had very strong methylation activity.

Unexpectedly, in vitro, the chromodomain of SUV39H1 showed no indication that it might be a stimulating agent for the activity of the SET domain. In contrast, all the evidence points to an inhibitory role of the chromodomain. The project had been planned based on the reports of stimulation occurring via the chromodomain binding to its target. Although opposite effects could be observed in vitro in our setup, the investigation was directed in vivo, using mouse embryonic fibroblasts (MEFs) with SUV39H1/2 knocked out (Peters et al., 2001). These cells are depleted of the constitutive heterochromatin H3K9 methyltransferases and therefore lost the characteristic H3K9me3-containing heterochromatic foci. The chromatin dense spots stained by DAPI were still present in these cells, but they lacked the presence of the H3K9me3 modification inside the foci (Peters et al., 2001). The intention of the performed experiments was to follow the recovery of H3K9me3 upon transfection with full length SUV39H1 wild type or the chromodomain mutants Y67A and W64A fused to YFP or CFP.
Initially, fluorescently labeled HP1β was co-transfected with the SUV39H1 constructs in order to serve as H3K9me3 marker. In SUV39 double null (SUV39dn) MEFs, HP1β-YFP showed a dispersed localization over the entire nucleus, instead of the localization to the large heterochromatic foci present in wild type MEF cells (Peters et al., 2001). When co-transfection with SUV39H1 full length occurred, both HP1β-CFP and SUV39H1-WT-YFP were co-localized to punctate foci in the nucleus 24 hours after transfection (Figure 65). Co-localization of HP1β-CFP and SUV39H1-Y67A-YFP was also observed, but the cells also showed a diffuse localization of both proteins over the entire cell (Figure 65). The distributed nuclear localization of HP1β was also observed when single transfection was performed with HP1β-CFP instead of YFP.

However, a flaw was found in the strategy. The full SUV39H1 protein contains next to its chromodomain a specific motif responsible for the interaction to HP1β (Yamamoto & Sonoda, 2003). This motif is intact in both the wild type and the mutant Y67A and W64A proteins, and might affect the localization of SUV39H1 and HP1β without any role of the chromodomain in this localization.
An attempt to replace the HP1β protein in the experiment was the introduction of antibody staining against H3K9me3. MEF SUV39dn cells were transfected with SUV39H1-YFP constructs and grown for one or two days after transfection. The transfection rate for the SUV39 double null cells was usually lower than 5%, and cells overexpressing SUV39H1 constructs were forming aggregates over time and died within 2-3 days of the transfection.

The untransfected cells showed no signs of premature death. The WT MEFs showed antibody staining of the H3K9me3 mark arranged in large, bright and punctate foci inside the nucleus,

**Figure 66 – Observations of H3K9me3 mark recovery after transfection with various SUV39H1 mammalian expression constructs.** A. Anti-H3K9me3 antibody staining performed as control on SUV39H1 double null (dn) and wild type (WT) cells. The cells fixed 1 day after passaging. B. SUV39h double knockout MEFs transfected with SUV39H1-YFP mammalian expression constructs and stained with anti-H3K9me3 antibody after fixation, at one day after transfection.
while the SUV39h double null MEF cells showed weak, evenly distributed H3K9me3 staining throughout the nucleus (Figure 66, A).

Four different SUV39H1-YFP constructs were transfected into the SUV39h knockout cells: wild type, Y67A and W64A chromodomain mutants, and one full length SUV39H1 that is catalytically inactive (Muramatsu et al., 2013). One or two days after transfection, the cells were fixed, permeabilized, and then stained with anti-H3K9me3 antibody. The wild type and the chromodomain mutants showed localization in large foci inside the cell nucleus, while the inactive enzyme showed localization in very small, punctate foci within the nucleus (Figure 66, B). The inactive protein also showed many small cytoplasmic aggregates in all the transfected cells (Figure 66, B). The wild type and mutants also showed intracellular cytoplasmic aggregates in most of the cases, especially two days after transfection (data not shown). However, one or two days after the transfection, there was no recovery of the H3K9me3 pattern in either the wild type or the mutant and inactive samples (Figure 66, B).

As a result of the failure to rescue the phenotype and recover the H3K9me3 marks even in the case of the wild type SUV39H1 transfections, experiments could not be conducted as planned.

To this point, the specificity of the SUV39H1 chromodomain was determined at peptide level by two independent approaches, and also confirmed at a more native level via far western analysis of the 150-600 mM extracted chromatin fraction. This fraction is a mixture between highly transcribed, RNA Pol II-rich chromatin (Henikoff et al., 2009), with densely packed heterochromatin, rich in histone H1 (Figure 61, B). This suggests that the SUV39H1 chromodomain might have additional targets to the H3K9me3.

At the level of the full length enzyme, the chromodomain showed no stimulation of the enzyme activity with the introduction of H3K9me3 in the methylation reaction, in either of the experimental setups used. However, there was a clear difference in strength of enzymatic activity between the wild type SUV39H1, the chromodomain full length mutant SUV39H1, and the SET domain alone. The presence of the chromodomain seemed to have an inhibitory effect on the catalytic activity. The effect was partially relieved by the chromodomain mutation, while the complete loss of the domain in the case of the SET domain was highly stimulating. Unfortunately, no conclusions could be drawn at cellular level, because of the lack of H3K9me3 recovery over one or two days after transfection.
5. Discussion

The results shown in the previous chapter were obtained using a variety of proteins and methods involved in the field of epigenetics, but together give a comprehensive glimpse into the high degree of complexity of the epigenetic regulatory factor network, with its dynamic and rich interactome. PKMTs select their targets based on well-defined specificity profiles, and methylate them at specific lysine positions. Methylation-specific reading domains, such as HP1β and MPP8, in turn, bind the methylated PKMT targets. A high variety of such reading domains have been shown to bind to many different PTMs, among which methylation or acetylation of lysines, some in non-histone targets, but most found so far on histone tails. In this study, the binding specificity of several of these histone tail readers has been investigated: the tudor domain of PHF1, the chromodomains of CBX2 and CBX7, the PHD domains of the MLL2 and MLL3 methyltransferases, and the chromodomain of SUV39H1. The last three examples are reading domains part of multi-domain PKMTs that probably have a function in the direct or indirect regulation of the enzyme activity. In addition, this work presents an attempt to set up a method for the identification of new readers, specific for the methylated non-histone targets of PKMTs.

5.1. Yeast-3-Hybrid Studies for the Identification of PKMT Target Binding Partners in a Methylation-Dependent Approach

Recently, many PKMTs have been shown to catalyze the methylation of targets other than histone proteins. The methylation of lysines found on the histone tails has been studied extensively, and the role of each histone methylation event is relatively well defined. However, the functions of non-histone target methylations are largely unknown, and need to be investigated. The Yeast-3-hybrid method introduced in this work was destined to serve for the rapid identification of methylation-dependent interactors of PKMT targets. The method has not been successfully adapted yet for the desired purpose, but the approach is very straightforward and might become functional after some additional optimization. The method was shown to be functioning in the case of the positive and negative control pairs provided by the Clontech Laboratories Inc company. Therefore, the inability to show that the interaction occurs at the level of our reading domain/PKMT target is not a general technical failure of the system itself, but is rather an issue related to the validity of our internal control.
The main concern at this point is the fact that despite using the same vectors for the mating experiments, the Yeast-3-Hybrid system was able to show the interaction of p53 and the SV40 large T antigen, while the proof of positive interaction failed for all the combinations of PKMT target/reading domain pairs. The biggest technical flaw in the system was the inability to detect the Gal4-DNA-BD fusion proteins, either with anti-Gal4-DNA-BD antibody or with anti-His tag antibody after the introduction of the C-terminal His tag. One final experiment that is needed to elucidate the level where the problem occurs is the analysis of the expression of the company-provided p53-Gal4-DNA-BD fusion protein, expressed in the newly produced pBridge-53 construct in Y2H Gold cells. The detection can be tested with anti-Gal4-DNA-BD antibody, or if necessary, a His tag introduction could be introduced to improve the chances of detection. If the control p53-fusion protein is also not detectable, then the problem lies only at the level of the detection by tag-specific antibodies. If the p53 protein were to be detected, this would still not necessarily mean that the PKMT targets are definitely not present. It might be that the folding of the Gal4-DNA-BD with the fused protein domains cause occlusion of the antibody epitopes, and this might be relevant also in the case of the His-tag. It is however unlikely that such problems would appear in the case of all the tested baits. The expression of the bait proteins can also be checked at RNA expression level to analyze the level of mRNA corresponding to the PKMT target fusion proteins.

Arguing against the absence of the baits is the existence of signs of bait-induced autoactivation in the case of some of the constructs (Figure 20, Figure 27), suggesting that the Gal4-DNA-BD fusion proteins might be successfully expressed, but detection was unsuccessful. Another indication that the baits might be successfully expressed, is the fact that exchanging the truncated ADH1 promoter for the full length constitutive promoter resulted in stunted growth in the transformed yeast cells. The result implies that the increased PKMT target Gal4-DNA-BD fusion expression is deleterious to the cells and the short ADH1 promoter is better suited in the Yeast-3-Hybrid screening. A means to solve the detection problem would be to use specific antibodies, raised against the methylated and unmethylated PKMT target. In this way, it could be determined both if the targets are expressed in the cells and if the methylation occurs in the yeast system cellular environment. However, the production of protein and modification-specific antibodies is a complicated and expensive procedure, and this alternative would be considered only if other methods do not succeed.
If the presence of the bait proteins is not the true problem of the assay, there are several other possible reasons for the negative results obtained. In Yeast-2-Hybrid interaction experiments, there are several steps at which the system could be disturbed. It is essential that the Gal4 fusions are stable, have nuclear localization and are able to dimerize. The occlusion of the binding site either caused by the AD domain or the DNA-BD domain of Gal4 is one of the possible reasons for false negatives. To overcome such problems, the system could be switched to a LexA-VP16 Yeast-3-Hybrid system (MacDonald, 2001, p. 275). The LexA system was used successfully in a similar study (Weimann et al., 2013), in which the targets of PKMTs were identified by Yeast-2-Hybrid screening coupled to short-read second generation sequencing. However, this study lacked the extra complexity added by the presence of the third member, the PKMT in our Yeast-3-Hybrid system. Moreover, switching to vectors that permit C-terminal Gal4 domain fusions might be the trick to solve this problem (MacDonald, 2001, p. 43). However, there are no such vectors available for Yeast-3-Hybrid systems yet. If the lack of interaction is rather caused by improper folding of the PKMT targets, changing the boundaries of these protein fragments might improve the assay result (MacDonald, 2001, p. 43).

Literature sources state that generally, if a positive interaction fails to be shown in one Yeast-2-Hybrid system, it will most probably fail in all of the available systems (MacDonald, 2001, p. 47). Chances are that the levels of protein expression and the yeast strain of choice could give a higher range of sensitivity (MacDonald, 2001, p. 47). Therefore, since the increase of protein expression by the promoter exchange is not a viable option due to cell growth stunting, the use of different yeast strains could be an advisable option. In some cases, the detection of interaction presents directionality, meaning that the interacting partners only show reporter gene activation in one setup and not if the DNA-BD and the AD-domain inserts are reversed (Estojak et al., 1995). This hypothesis could also be tested by inversion of one of the pairs of interactors. This would be particularly useful in order to increase the concentration of PKMT target in the system, and improve the target methylation levels, as the AD-fusions are expressed from the ADH1 full length constitutive promoter. The readers, such as the HP1β chromodomain and the MPP8 chromodomain, would be expressed as the Gal4-DNA-BD domains.

Setting aside the problem of the non-detectable bait protein, the issue of the leaky PKMT expression also needs to be solved. According to literature, the expression of the PKMT from
the conditional MET25 promoter should be completely repressed by 1 mM methionine in the media (MacDonald, 2001, p. 280). In the case of the present study, the inhibition was not complete for cultures grown overnight in liquid culture containing 1 mM methionine. It might be that yeast colonies growing on 1 mM methionine media on agar plates show more inhibition. However, the leaky expression can be contained by gradually increasing the methionine concentration in the media until full PKMT repression is achieved. Although there was no sign of PKMT-dependent reporter gene activation, even in the cases of the baits that showed some autoactivation capacity, a difference was still observed between the mating experiments that were done in the presence of GLP, compared to hG9a. It was evident that the enzyme that showed higher activity in vitro, GLP, also induced stronger activation of the reporter genes in the cases of the experiments with autoactivating bait constructs. This suggests the possibility that the PKMT could indeed have an effect on the outcome of the interaction study.

The bait constructs that show autoactivation at the MEL1 promoter, responsible for the blue color production in the presence of X-α-Gal upon reporter gene activation, also presented signs of positive interaction in the small scale mating experiments. This is most probably caused by the intrinsic activating potential of the baits. In a Yeast-3-Hybrid screen, constructs such as the ones expressing the G9a target G9aNTD or the SET7/9 targets PPARBP or IRF1 would not be used, because they cause strong autoactivation of the reporter genes, and the resulting colonies are dark blue without the presence of the Gal4-AD interaction partner. On the other hand, weakly autoactivating constructs such as the one expressing the SET7/9 target AKAP6 could still be used in a screen, by considering as positives only the strongly dark blue colonies that grow fast on SD/-His media, and as negatives the light blue ones that grow slowly on SD/-His media (MacDonald, 2001, p. 125).

Another possible reason for the failed attempts to show reading domain interaction to PKMT targets is the fact that these are human or mouse proteins expressed in yeast, which might result in expression in inactive form of either of the proteins. This might change the interaction surface between the two proteins and cause loss of binding.

Another possibility would be a hypothesis related to the function of the methylations set on the non-histone targets. It might be that these modifications are meant to be signals for protein degradation, and would result in the destabilization of the bait-DNA-BD fusion
proteins. Since the MET25 promoter appears to be leaky, the permanent presence of the methyltransferase in the assays could result in a continuous cycle of expression and degradation of the PKMT target proteins. This could also explain why the expression of the targets has an effect on the yeast cell growth and reporter gene activation, but at the same time cannot be found by protein extraction and antibody detection. This hypothesis can be easily tested using pBridge vector that contains only Gal4-DNA-BD fusion, but no methyltransferase. However, the degree of leaky PKMT expression was variable, depending on the construct used, but these variations did not reflect in any way on the results of the mating experiments. In addition, in the case of WIZ, the expression of the WIZ-His6 protein was not detected even from a construct that did not contain any methyltransferase at the second multiple cloning site (Figure 26, D). This finding cannot be generalized for all the PKMT targets, but at least in the case of WIZ, the absence of detection signal is not caused by the activity of G9a.

Moreover, the interaction of the HP1β chromodomain with the G9a non-histone targets was not shown to happen in vivo at full length protein level, so the interaction observed in vitro might be too weak or not happening at all inside the cell. To support this statement, a study completed in 2013 by Liu and colleagues (Liu et al., 2013) performed a screen at whole proteome level by immobilizing the methyl-specific binding HP1β chromodomain to beads in order to capture methylated proteins from the HEK293T cell lysate. Using this method, 109 interactors were identified, but WIZ, CDYL1 and G9a were not among them. This result suggests that the G9a methyltransferase targets may not be interacting partners of HP1β, and better-suited interactor pairs should be chosen. Therefore, an optimal control would be the canonical G9a target, histone 3, which could be cloned as the Gal4-DNA-BD fusion. Histone 3 is the ideal “bait” candidate, not only because there is strong evidence of interaction between HP1β and H3K9me2/3, but also because it is a small, mostly unfolded protein, which would act as the perfect fusion protein, as lack of interaction detection could not be attributed to folding problems of the “bait” protein.

In addition, the interaction between endogenous MPP8 and the transiently expressed methylated N-terminal part of Myc-DNMT3A was shown to take place in vivo in HEK293T cells via immunoprecipitation experiments (Chang et al., 2011). The difference to the latter study was that in the present work, the mouse Dnmt3a was used instead of the human DNMT3A. The interaction between MPP8 and mDnmt3aK44me2 peptide was shown via ITC (isothermal
calorimetry) titrations, but was not shown in vivo (Chang et al., 2011). However, the interaction between MPP8 and the human DNMT3A NTD methylated at K47 was demonstrated. The sequences of the mouse and human DNMT3A spanning residues 1 to 274 are 90% identical, and at least 15 residues on each side of the target lysine are identical (with the exception of position -4 relative to the target K, where either an A or a T are present). The high similarity in sequences supports the hypothesis that MPP8 would interact with the mDNMT3a K44me also in vivo. The binding constant ($K_d$) of the interaction between MPP8 and the mDnmt3a K44me peptide was $\sim$ 12 $\mu$M (Chang et al., 2011), an affinity which should be detectable via a Yeast-3-Hybrid system, if all the expressed DNMT3a NTD proteins were methylated.

The pBridge-based Yeast-3-Hybrid system has been shown to be technically sound for genetic applications (Gordon & Buchwald, 2003). The system relies on the activation of four reporter gene promoters, and this represents a low number of Gal4-DNA binding sites to be occupied when the interaction takes place. Normally, the low levels of bait expression ensure that only specific interactions will result in target site binding. In this way, false positives are limited to a minimum. In our setup however, the low expression of the baits might be the limiting factor, because the positive selection depends not only on the interaction of the bait to the prey, but also on the efficiency of the methylation of the pool of targets. Therefore, the equilibrium has to be reached from two sides for the interaction and reporter gene activation to take place. On the one hand, this equilibrium depends on the Michaelis-Menten constant ($K_M$) characterizing the activity of the methyltransferase on its target substrate, and on the other hand it depends on the binding constant ($K_d$) characterizing the interaction between the methylated target and its reader. Adding up the two equilibria requirements, the chance that the interaction is detected at the concentrations available in this setup is considerably reduced.

Although the Yeast-3-Hybrid method described in this work is still facing a lot of open questions, there is still room for changes that might improve the system. For the future, an attempt to eliminate one of the limiting factors of the method would be the fusion between the PKMT and its target through a linker, to increase the chances that the methylation event occurs. However, this may require multiple attempts to clone these single-chain PKMT/PKMT target proteins, until an active form is developed. Consequently, increased size of the fusion protein might cause problems during expression in yeast, as yeast proteases tend to cleave
proteins larger than 60 kDa. An alternative screening method would be phage display, in which the fusion protein is expressed on the surface of phages multiplying inside bacterial hosts. The size of the protein might also be an issue in this system in addition to the introduction of the third member, the PKMT.

Presently, the first experiment to be done is the attempt to detect the p53-Gal4-DNA-BD “bait” protein provided by Clontech Laboratories Inc. as control. Second, H3 will be cloned as a Gal4-DNA-BD fusion in order to have a better internal control. Third, the Gal4 domains could be swapped, such that the PKMT target will be expressed from the pGADT7 vector as a Gal4-AD fusion, from the constitutive ADH1 promoter. This swap would ensure high enough concentrations of the PKMT target for efficient methylation. If the concentration of the reading domain would be too low for target interaction, the reader could be cloned into the newly designed pBridge vector with the full length ADH1 promoter, to increase its expression.

Finding a way to successfully screen for binding partners of non-histone targets based on their methylation status is becoming more and more important, as the number of identified PKMT non-histone targets is rapidly increasing, while the currently available methods for identifying the function of the methylations work only at an individual level and require a lot of time and effort. Although this project did not reach its purpose yet, it represents the first step towards a more efficient study of PKMT non-histone target identification.

5.2. The PHF1 Tudor Domain Binds Specifically to H3TK27me3, and Shows Comparable Binding to H3K36me3 and H3K27me3 Modified Histones

The work presented here is part of a more complex published investigation (Kycia et al., 2013). PHF1, a member of the multi-protein complex PRC2, was reported to influence the catalytic activity of the complex indirectly (Musselman et al., 2012; Sarma et al., 2008). Here, I investigated the role of one of the PHF1 conserved domains, the tudor domain. Previous papers investigating the binding of the PHF1 tudor domain only tested the binding to canonical histone peptides, and therefore concentrated on the affinity for H3K36me3, while the preference for H3K27me3 was overlooked (Cai et al., 2013; Musselman et al., 2012; Qin et al., 2013). In the present study, the PHF1 Tudor domain bound specifically to H3TK27me3 with a $K_d$ of ~ 26 μM, less than 2-fold higher than the binding constant to the H3K36me3 peptide
(Kycia et al., 2013). The binding to the canonical H3.1 variant with trimethylated K27 was 10-fold lower than to the H3K36me3 peptide (Kycia et al., 2013), (Table 6). The H3T variant, although initially identified in testis, where it is highly expressed, has been later found to be expressed in most tissue types (Tachiwana et al., 2010). The levels of its expression and incorporation still need to be clarified in the light of recent findings. The presence of the H3T variant at similar levels with the canonical H3.1 would be one way to explain the comparable GST-pulldown results (Figure 29). After the GST-pulldown with recombinant GST-tagged PHF1 tudor domain, staining with anti-H3K27me3 and anti-H3K36me3 antibodies showed very similar enrichment in both modifications, which could either be a reality, or an artifact caused by the aggregation of histones in the assay, resulting in carryover of non-specific interactions. However, reduced stability of the H3T nucleosomes would probably not be a feature that would promote high levels of H3T incorporation in somatic cells (Tachiwana et al., 2010). To settle these open questions, the solution would be the development of variant specific antibodies, to be able to differentiate between H3TK27me3 and H3.1K27me3-specific binding. It might also be the case that in vivo, the 3D nucleosome structure within chromatin would shift the preference of the tudor domain towards canonical H3.1 K27me3. This switch could also be influenced indirectly, by interaction with other chromatin regulators. Localization studies performed by Dr. Ina Kycia support the hypothesis that H3K27me3 is the main target of the domain, as the PHF1 protein shows complete co-localization with anti-H3K27me3 antibodies, and only partial co-localization with anti-H3K36me3 antibody staining (Kycia et al., 2013). The binding of PHF1 to H3K27me3 marks, not H3K36me3 marks, is also supported by the fact that the Drosophila homolog PCL (Polycomb-like), is reported to be an H3K27me3-reader (Cao et al., 2002; Min et al., 2003).

Presently, the function of the PHF1 tudor domain is controversial. It was reported that the activity of the PRC2 complex is stimulated by the presence of PHF1 (Sarma et al., 2008), while at the same time activating marks, including H3K36me2/3 inhibit the PRC2 complex (Schmitges et al., 2011). This data supports the idea that PHF1 is recruited to H3K27me3 primarily, resulting in the targeting and the stimulation of the PRC2 complex. In contrast, other studies suggested that the recruitment of PHF1 to H3K36me3-rich active gene bodies would serve to promote the spreading of facultative heterochromatin (Cai et al., 2013) and that PHF1 would also be involved in DNA damage repair, causing inhibition of the PRC2 complex activity (Musselman et al., 2012). All these contradicting results show that there are still open
questions about the exact function and binding specificity of the PHF1 protein. For the future, a much-awaited tool is the anti-H3TK27me3 antibody, which could at least partially resolve these issues. This body of work, together with the rest of the novel data in the publication (Kycia et al., 2013), puts one more piece of the puzzle in its right place.

5.3. The Chromodomains of CBX7 and CBX2 Bind Specifically to H3K27me3 and Serve as Markers of the Modification at Genome Level

In this work, I have investigated several reading domains involved in the targeting of the Polycomb Group repressor complexes. In addition to the tudor domain of PHF1, I also dove into the recruitment of PRC1, by studying two of its mutually exclusive components, CBX2 and CBX7. These homologs of the *Drosophila* Polycomb protein, were shown here to bind specifically to H3K27me3, especially when the modification was in the sequence context of the testis-abundant histone variant H3T. Surprisingly, CBX7 even showed higher preference for the H3T variant than for the H3K9me3 peptide and the H3.1K27me3, which are also preferentially bound. It seems that some, but not all H3K27me3-binding domains have the ability to discriminate between the sequences of the two histone variants H3.1 and H3T, although the difference at the peptide level is represented by one single amino acid exchange, A24V, respectively. This is a very subtle change between two small, hydrophobic side chains, but it seems to affect the binding of several histone reading domains. In the case of CBX7, and the earlier mentioned PHF1, the preference for H3T variant was strong, with more than 6-fold increase in binding affinity of CBX7 towards H3T, compared to H3.1. In contrast, CBX2 does not present a clear discrimination between the two variant peptides, having less than 2-fold difference in $K_d$ between the two (Figure 33). Therefore, the ability of reading domains to discriminate between histone variants is an outstanding feature, and not a ubiquitous property.

Another interesting observation was the dual specificity of both CBX2 and CBX7 chromodomains at peptide level. In the case of CBX2, the preference for H3K27me3 is evident, while in the case of CBX7, the preference for the H3K27me3 was proven *in vivo* (Kungulovski et al., 2014), although the H3K9me3 peptide was the main preference at peptide level. An explanation of this artifact would be the very similar sequence context of the H3K9 and the H3K27 peptides. Both contain the ARKS motif around the target lysine, and it might be that other parts of the histone tail, which are not part of the peptides used for the binding
experiments, are relevant for the chromodomains to discriminate between H3K27me3 and H3K9me3 histone tails. Another indication that only one of the two modifications is the specific target of CBX2 and CBX7 is the fact that both modifications are bound in the same pocket, because mutation of the aromatic cage results in loss of both interaction signals.

The binding constants obtained for the CBX2 and CBX7 chromodomains are within the ranges reported in literature (Table 12, Table 13). When compared to another study with human polycomb proteins, the values obtained in the present study were approximately 4 to 5-fold lower, both for the H3K9me3 peptide and the H3K27me3 peptide (Kaustov et al., 2011) (Table 12, Table 13). Although the obtained binding constants are not identical to the other studies, variability can occur due to slightly different boundaries of the chromodomain cloned, or the length of the peptide sequence used. For example, Kaustov et al., (2011), used shorter peptides in the titrations, which might explain the weak binding, despite the preservation of the order of preference for different modifications.

In the case of CBX7, the relative binding strength to the H3K9me3 compared to H3K27me3 peptides was similar in this work compared to the study by Kaustov et al., (2011), (Table 12). However, the preference for H3K27me3 compared to H3K9me3 was stronger in the data presented here than in the previous study. This could be explained by the difference in peptide size, as Kaustov et al. used 15 amino acid long peptides, compared to the 19 amino acids used in this work. This means that in the latter, K27me3 was further from the artificially placed N-terminus of the peptide, and therefore was given a more natural setting. The additional amino acids of the peptide used here might also be relevant in the specific recognition of H3K27me3, as the K9 and K27 residues are found in the same ARKS sequence context, and for the discrimination between the two peptides, the flanking sequences are relevant. The fact that the binding constant for the H3K27me3 interaction decreases with the increase of the peptide length supports the importance of the flank sequences. The mouse Cbx7 chromodomain showed approximately the same K_d values as the human CBX7 tested in this study (Bernstein et al., 2006). The H3K27me3 peptide used in this case was 17 amino acids long, confirming that the presence of these additional residues at the C-terminus of the peptide might increase the binding affinity of CBX7 to H3K27me3.
Table 12 - Comparison of the CBX7 binding constants (K<sub>d</sub>) obtained in this study with the K<sub>d</sub> values reported independently by others.

<table>
<thead>
<tr>
<th>Reading domain</th>
<th>K&lt;sub&gt;d&lt;/sub&gt; H3K9me3 [µM]</th>
<th>K&lt;sub&gt;d&lt;/sub&gt; H3.1 K27me3 [µM]</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBX7 WT (human)</td>
<td>14.2 ± 0.7</td>
<td>18.5 ± 1.0</td>
<td>Present study</td>
</tr>
<tr>
<td>CBX7 WT (human)</td>
<td>55 ± 5</td>
<td>110 ± 17</td>
<td>(Kaustov et al., 2011)</td>
</tr>
<tr>
<td>Cbx7 WT (mouse)</td>
<td>12 ± 3</td>
<td>22 ± 5</td>
<td>Bernstein et al., 2006</td>
</tr>
</tbody>
</table>

The human CBX2 chromodomain literature data fit to the results obtained in this work, when considering the overall weaker binding caused by the shorter peptides used by Kaustov et al. (2011) (Table 13). In contrast to the human homolog, the mouse Cbx2 presented the same binding constant for both H3K9me3 and H3K27me3, compared to the human CBX2 proteins, which bind the H3K9me3 peptide too weakly to determine a K<sub>d</sub> value (Table 13). However, the specificity towards H3K9me3 was not observed in vivo, in localization studies (Bernstein et al., 2006).

Table 13 - Comparison between the CBX2 binding constants (K<sub>d</sub>) obtained in this study with the K<sub>d</sub> values reported independently by others.

<table>
<thead>
<tr>
<th>Reading domain</th>
<th>K&lt;sub&gt;d&lt;/sub&gt; H3K9me3 [µM]</th>
<th>K&lt;sub&gt;d&lt;/sub&gt; H3.1 K27me3 [µM]</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBX2 WT (human)</td>
<td>&gt; 200</td>
<td>54.6 ± 0.5</td>
<td>Present study</td>
</tr>
<tr>
<td>CBX2 WT (human)</td>
<td>&gt; 500</td>
<td>185 ± 20</td>
<td>(Kaustov et al., 2011)</td>
</tr>
<tr>
<td>Cbx2 WT (mouse)</td>
<td>41 ± 6</td>
<td>44 ± 5</td>
<td>Bernstein et al., 2006</td>
</tr>
</tbody>
</table>

The CBX2 and CBX7 proteins are essential players in the PRC1 complex, and play important roles in its targeting to specific target loci (Gil et al., 2004). The confirmation that the CBX2 and CBX7 proteins bind specifically to H3K27me3-containing peptides emphasizes the connection between the PRC1 complex, and its partner complex, PRC2, which is responsible for setting the H3K27me3 mark. CBX proteins all contain a single structured domain, the chromodomain, at the N-terminus. In addition, CBX2 contains an AT-hook DNA-binding motif, while all the other Pc homologs contain AT-hook-like motifs adjacent to the chromodomain. In the middle of the protein, a less conserved region is found, while at the C-terminus, there
is a polycomb repressor box (Senthilkumar & Mishra, 2009). Moreover, Cbx4, Cbx6, Cbx7 and Cbx8 have been shown to interact with RNA. The recruitment of PRC1 to its target loci is most probably a complex process governed by a mixture between the affinity of the binding to H3K27me3 tails, together with the targeting provided by the AT-hook of CBX2 and potentially the AT-hook-like motifs and the RNA binding of the other CBX proteins, and maybe also possible interactions mediated by the region of low conservation.

It has long been reported that both CBX7 and CBX2 bind H3K9me3 peptides as well as H3K27me3 peptides, but the reason for this double preference or the role of H3K9me3 towards the regulation of PRC1 was unclear (Bernstein et al., 2006; Kaustov et al., 2011). The present study (Kungulovski et al., 2014) clarifies this subject, by showing that the H3K9me3-binding is not ensuring the targeting of the CBX7 chromodomain to chromatin, while the H3K27me3 modification is the critical histone modification for the successful recruitment of the domain.

The strongly mono-specific precipitation of H3K27me3-associated loci by CIDOP, coupled to qPCR or next generation sequencing, makes CBX7 a solid candidate for a novel experimental tool that can replace anti-H3K27me3 antibodies in western blot and ChIP experiments. Although at peptide level CBX7 could only be used as a double mark reader, at the level of chromatin, either immobilized on a blot membrane or isolated in solution, CBX7 proved to be a reliable anti-H3K27me3 antibody replacement. The analysis of the pulldown at specific loci also pointed out that in some situations, the antibody showed crossreactivity with H3K9me3 (Figure 36), while CBX7 did not, promoting even more the reliability of the domain compared to the antibody. The generally lower binding constants of antibodies compared to the CBX7 domain could be compensated by increasing the concentration of the reading domain in reactions. This would not be an issue, because the CBX7 chromodomain purifies well, with little degradation and high concentrations, and the costs of producing recombinant proteins in bacteria is negligible compared to the financial investment and time required for antibody development and production. For this application however, CBX2 binding affinity was too low to confer the assay reliability. This shows one of the limitations to the pool of reading domains that can be chosen as antibody alternatives. However, recombinantly produced proteins offer the possibility of detection using common tags, and also leave room for engineering in order to optimize or change their specificities. All these numerous advantages promote the use of native and engineered reading domains instead of antibodies developed to recognize histone
post-translational modifications. Depending on the availability of reading domains and their specificities, the range of applications could be broadened, for example to reading domains that bind to specific methylated PKMT targets, the subjects of my first project.

5.4. The PHD Fingers of MLL2 and MLL3 Show Preference for H4 Peptides Acetylated at Lysine 16 in Association with Other Modifications

The SET1 family of methyltransferases, comprising SET1A, SET1B and the MLL enzymes (MLL1 to MLL4) is a group of highly specific H3K4 methyltransferases, generating mono-, di- or trimethylated H3K4 (Ansari et al., 2009). MLL1-4 are very large proteins with sizes between 300 and 600 kDa, associated with various other proteins in multisubunit complexes, such as the human COMPASS. These four MLL enzymes have complex domain organization and the function of many of these domains is yet to be elucidated.

Each of the MLL enzymes contains several PHD fingers, conserved domains known to have roles in protein-protein interaction (Musselman & Kutateladze, 2011). MLL1 has been so far investigated in most detail. Out of its four PHD fingers, PHD1 is reported to have an intramolecular interaction function with the fourth PHD finger, resulting in a unique structural fold (Yokoyama et al., 2011). The second PHD finger shows high similarity to the MLL4 PHD2, and both have been reported to show E3 ubiquitin ligase activity in the presence of CDC34, an E2-conjugating enzyme. At least in the case of MLL1, the ubiquitination activity occurs intramolecularly, and the MLL1 ubiquitination might have a role in its cell cycle-regulated degradation (Wang et al., 2012). PHD3 of MLL1 has been investigated in detail, as it was shown to bind both to H3K4me3, and also to the co-repressor nuclear cyclophilin Cyp33 (Wang et al., 2010).

The MLL2 and MLL3 proteins, which were investigated in this work, possess more PHD fingers than MLL1 and MLL4, but most of their functions are completely unknown. Wang et al. investigated the E3 ligase capabilities of all the MLL2 PHD fingers individually, and also of the MLL3 PHD 5, but no ubiquitin ligase activity was identified in any of the cases, despite the high sequence similarities between MLL1 PHD 2, MLL2 PHD4 and MLL3 PHD5 (Wang et al., 2012).

Others previously investigated the binding of all the PHD fingers of MLL2, and some of MLL3 to histone peptides without much success. In literature, the only binding specificity reported was for MLL2 PHD 3-5, which was shown to bind to unmodified H4 tail peptides, when they
were not symmetrically dimethylated at H4R3 (Dhar et al., 2012). Dhar et al. (2012) showed this binding specificity using peptide pulldown experiments. In the present study, the binding of both the MLL2 PHD 3-5 and the MLL3 PHD 4-6 was characterized. They had very similar preferences, to peptides from H3 and H4 tails, and bound especially strong to H4 peptides (11-30) which contained an acetylation mark at K16, particularly when combined with other acetylations and arginine dimethylation. In contrast to Dhar et al., in my experiments, the binding to the H4 (1-19) peptides was very weak compared to all the other specific interactions identified. In the paper by Dhar et al., (2012) they investigated the specificity of the MLL2 PHD 3-5 only in selected peptide pulldown experiments with a few peptides only. The peptides always contained no more than one modification, and no PTMs other than methylations were used. Therefore, compared to the literature results, the work shown here offers a much broader range of peptides and PTMs for the screening of the domain specificity. Also, providing all the putative substrates in competition, on one array, promoted the discrimination between highly preferred binding and weaker or more unspecific binding.

Another explanation for the weak binding to H4 peptides could be the length of the peptides used in each case. Dhar et al, (2012), used H4 (1-23), a slightly longer peptide than the N-terminal 19 residue H4 used on the MODified™ Histone Peptide Array. It might be that the addition of the four amino acids (20-23) to the C-terminus of the peptide would increase the binding affinity to it. In the present study, we identified much stronger binding to H4 (11-30), out of which in the Dhar et al. experiments they only spanned part of the N-terminal sequence (residues 11-23 part of the 1-23 peptide).

On the histone peptide array, at the level of the H4 (1-19), the binding was too weak to be able to confirm the inhibition by H4R3me2s. The H4 (11-30), which showed strong and specific binding, however, did not contain the arginine residue at position 3. Therefore, in our experimental setup, the influence of the R3me2s on the binding could not be investigated. Still, in the light of the peptide binding preferences revealed by MLL2 PHD 3-5, it is surprising that an arginine dimethylation would cause inhibition. My results showed that both R26me2a/s on the H3 (16-35) peptides and R19me2a/s on the H4 (11-30) peptides stimulated binding. However, in both cases, the symmetric and asymmetric arginine methylations did not seem to be differentiated by the domains.
Another inconsistency with the Dhar et al. (2012) study was the binding to H3 (16-35) and H3 (26-45) peptides with different modifications at K27 and K36 positions, which was observed on the modified histone peptide array, but not in the peptide pulldowns by Dhar and colleagues. Dhar et al. (2012) investigated the binding to the H3K27 and the H3K36 marks by using the same peptide, H3 (21-44) in peptide pulldown experiments. This setup is not ideal for the binding of H3K27, because the mark is located relatively close to the N-terminal truncation of the H3 tail, which offers suboptimal conditions for interaction, as in most cases, the residues surrounding the target lysine are also specifically recognized. They observed no binding to the H3K27-methylated peptide, and very weak pulldown of the H3K36 methylated peptides. On the peptide array binding assay presented in this work, MLL2 PHD 3-5 did not show strong binding to H3K27me1, me2 or me3 alone, but the additional presence of other modifications increased the binding to the H3 (16-35) peptide. In contrast, the H3K36-containing peptide from the array was bound by PHD 3-5 even in unmodified state. The binding to H3K36-containing peptides was not reproduced in the Dhar et al., (2012), study with the H4 (21-44) peptides. It is important to mention that multiple modifications were not investigated by Dhar et al., (2012).

Generally, the slightly different boundaries for cloning the PHD 3-5 of MLL2 might also cause slight changes in structure that would partially compromise the binding specificity of the domains in the case of Dhar et al., (2012).

Interestingly, the MLL3 PHD 4-6 triplet showed very similar binding to that of MLL2 PHD 3-5, and this is the first time that any kind of binding specificity was recorded for this set of PHD fingers. Judging by the sequence similarities between the PHD fingers of MLL2 and MLL3, it was expected that the PHD 3-5 and the PHD 4-6 tandem PHD fingers would have related functions, and this was confirmed by their binding specificities.

The only PHD fingers from MLL2 and MLL3 which have their structures solved are the tandem PHD fingers 1 and 2 of MLL3 (PDB ID: 2YSM). The structure of these domains, together with other tandem PHD known structures, such as the DPF3b (Zeng et al., 2010) or the MOZ PHD fingers, show the tendency of the domains to fold against each other, to form higher order 3D structures. Even in the case of the MLL3 PHD 1-2, these two domains are actually part of a four domain cluster, with PHD0 and PHD3 in immediate vicinity of the PHD 1-2 domains, and the folding of all the four PHD fingers cannot be predicted with the information available.
presently. The presence of the other two PHD fingers might completely alter the overall fold of the domains 1 and 2 relative to each other.

From the information gathered so far in this investigation, two hypotheses arise. First, it might be that each of the domains has a specific pocket, one for the H3 (16-35) peptide, one for the H3 (26-45) peptide, and the other for the H4 (11-30) peptide. In this case, identifying each pocket by site directed mutagenesis would result in loss of binding for the individual target peptide and not the other two peptides. If this were the case, the binding pockets would be in the core of each PHD finger. The mutagenesis performed in this study is based on this hypothesis. However, the designed mutants did not lose binding to any of the target peptides.

A possible explanation for this result would be that the mutation of the residues to alanine are too weak and do not result in a strong effect on the binding pocket. This question can be solved by introduction of stronger mutations, such as the replacement of the target residues by lysines. However, it might be that the binding pockets are not similar to the previously identified PHD fingers that bind acetylylsine. This leads us to the second hypothesis, that the binding pockets are not part of each PHD finger, but rather form only at the interface between two or even three PHD fingers. This possibility would be supported by the fact that the individually cloned PHD fingers do not show binding to the histone peptides in the array, and that the mutants of putative binding pockets found in the core of the PHD domains did not show any effect. In this latter case, the entire PHD group might have a single binding pocket for all three target peptides, and the disruption of the pocket would result in complete loss of binding to either of these peptides.

Another important question is whether the binding is specific to acetylated lysines alone, and in particular H4K16, or whether maybe the binding pocket also accommodates specifically dimethylated arginine as well.

Last but not least, most of the PHD fingers of MLL2 and MLL3 did not show any specific binding to histones. However, this does not mean that they do not have important functions. In order to shed some light on the roles of these domains as well, pulldowns or far western blot analyses with mammalian cell nuclear extracts could be performed in order to identify non-histone binding partners of these PHD fingers. Also, the interactions among the recombinantly expressed PHD fingers can be tested by GST pulldown assays in order to verify if any of the domains would be involved in intramolecular interactions. Very good candidates would be the
PHD6 of MLL2 and PHD7 of MLL3, which are located C-terminally, distant from all the other PHD fingers. Just as the interaction between PHD1 and PHD4 of MLL1 (Yokoyama et al., 2011), the interaction between two of the PHD fingers of MLL2 or MLL3 could also add structural complexity to the enzymes.

There are still many open questions regarding the binding partners and binding specificities of the PHD fingers of the MLL2 and MLL3 enzymes. The initial findings reported here are very promising and will be the basis of an in depth study of the function and regulation of the two MLL family H3K4 monomethyltransferases.

5.5. The SET Domains of MLL2 and SET1A Present Substrate Specificity Preferences That Will Be Used in the Identification of Their Non-Histone Targets

The field of protein methylation is very broad, and is expanding at a very rapid pace. The methylation of lysines and arginines in histones is essential for the epigenetic regulation of transcription and genome stability. Recently, numerous non-histone proteins have been shown to be specifically methylated by various methyltransferases, most of which having the property to methylate both histones and other proteins. So far, SET domain methyltransferases have been identified to function exclusively as lysine methyltransferases and non-histone lysine methylation has been reported to be involved in signaling pathways, such as the p53 tumor suppressor pathway (Shi et al., 2007).

The investigation of PKMT methylation targets is essential for the discovery of the intricate regulatory networks that govern the functioning of cells. The human COMPASS complex is essential for survival, due to its H3K4 methyltransferase activity. Gene activation is mostly achieved in humans by the two homologous PKMTs, SET1A and SET1B. They have a wide array of genomic targets to which they are specifically targeted in order to set promoter H3K4 trimethylation. It has been shown that their targets are non-overlapping, as the proteins showed different nuclear localization patterns (Lee et al., 2007). Although they show very high similarity, the enzymes differ in target site specificity, and this difference could be only one of several functional differences between the two enzymes. The results shown here revealed that both SET1A and SET1B were practically inactive when they were not part of the COMPASS complex. However, SET1A gained strong stimulation of activity in the presence of three of the
core complex members (WDR5, RBBP5 and ASH2L). The complete lack of activity of SET1B could either be explained by a technical failure, such as poor choice of boundaries resulting in loss of SET domain activity, or by the lack of some additional factors required for SET1B activation. It might be that in the presence of the full human COMPASS core complex, also including WDR82 and CFP1, SET1B would also become active.

The observation that SET1A is strongly active in complex, but very weakly alone, might be highly relevant. The inactivation of SET1A when it is not within the core complex could be essential in order to avoid spurious methylation when SET1A is not at its target loci. In this case, SET1A would be able to methylate non-histone targets if they were also recruited to the SET1A target gene promoters, or were direct interactors of SET1A or its complex members.

Interestingly, the specificity profile derived for the SET1A complex revealed that even one amino acid exchange in the histone H3 sequence can increase its methylation activity drastically, as in the case of A1L, T3F and T3Y. This result suggests that there might be target sequences which are preferred to the sequence of H3, so novel non-histone targets of SET1A are probably waiting to be discovered. One target, the heat shock protein 70 (HSP70), has already been identified by Cho et al., (2012). The methylation occurs at Lys 560 of the enzyme, in a [L]-3 [K]-2 [G]-1 [K]0 [I]+1 [S]+2 [E]+3 [A]+4 sequence context. The target is not a canonical one, judging by comparison to the derived specificity profile, but it does have a leucine at position -3, one of the three residues which showed significant stimulation of the activity by single amino acid exchange. The presence of the leucine at the -3 position could alter the SET domain binding to the adjacent residues, compared to the H3 peptide, and result in slight shifts in the specificity profile. Most probably there are many other SET1A putative targets yet to be found.

The other investigated methyltransferase, MLL2, also acts on the N-terminal H3 tail, to methylate the lysine 4 position. However, the sister methyltransferases MLL2 and MLL3 both only reach monomethylation of their substrate (Herz et al., 2012) and are responsible for the monomethylated H3K4 at enhancers and at the promoters of conditionally repressed inducible genes (Cheng et al., 2014).

Compared to SET1A/B, which methylated H3K4 at a genome-wide level, MLL2 and MLL3 have much more restricted target loci. This could be an explanation for the activity of the enzymes also in the absence of the complex members. The reduced availability of histone substrate
might have resulted in the development of more non-histone methylation targets. The ability to act both alone and within the complex, with slightly different specificities in each case, broadens the pool of putative non-histone targets. This also offers the chance of having binding partners independent of the human COMPASS members. In the absence of the complex members, the enzyme has a higher available binding surface, which might also increase the possibilities of interaction with novel substrates.

MLL2 bound to the complex showed lower overall selectivity for the residue exchanges in the substrate peptide. A possible explanation could be that MLL2 acts only on histone H3 when in complex, and in this setup, it is targeted by the other complex members to the target loci, where H3 is highly abundant. However, when MLL2 is alone, it might need to be much more specific in order to choose the correct non-histone targets from the pool of available substrates.

Although the complex was not so restrictive towards the substrate sequence, it still showed some important preference difference compared to the single MLL2 SET domain. These differences are definitely worth exploring further, as they might be relevant for the target selection process of the enzyme in the two situations.

The specificity of the MLL2 and MLL3 SET domains was not identical, but very similar. Therefore, if they were to be shown to act non-redundantly, the difference in function could not be explained by the specificity of the enzyme alone. It would probably be either a function of the regulatory domains, such as the PHD fingers, or a cumulative effect of both.

The MLL2 SET domain alone was shown to methylate a few non-histone targets at peptide level. If the methylations were confirmed at protein level, the investigation of their function would be interesting. CHD3 is a member of the NuRD deacetylase complex (Lemos et al., 2003). MLL2 has been shown to have repressive roles at the promoters of conditionally repressed inducible genes (Cheng et al., 2014), and it might be interesting to test whether the methylated CHD3 causes recruitment of NuRD to the MLL2 target promoters. The Hepatocyte nuclear factor 4α is a nuclear hormone receptor which acts as a transcription factor, controlling the transcription of a set of genes essential in the development of the liver, kidney and the intestine (Duda et al., 2004). The Protein MCM10 homolog, another putative non-histone target of MLL2, is a DNA replication initiation factor with roles in prevention of DNA damage during replication (Chattopadhyay & Bielinsky, 2007). In contrast to CHD3, the ZZ-
type Zinc finger-containing protein 3 (ZZZ3) is part of the ATAC histone acetyltransferase complex (Guelman et al., 2009). The methylation mark could have different functions on different targets. As presented in the previous chapter, the PHD domains of MLL2 might target it to acetylated histones H3 and H4, which are also the reported targets of the ATAC complex. The last identified putative target of MLL2 was the Zinc finger protein 862, which has not been studied so far.

Presently it is unclear whether the MLL2 and MLL3 enzymes are generally repressive when they monomethylate gene promoters, or if they rather have either activating or repressing functions depending on the target locus and the proteins responsible for their targeting. A very interesting observation was the preference of MLL2 to methylate H3 peptides containing H3K9me3. H3K9me3 is a repressing mark characteristic for constitutive repression of genes. So far, MLL2 has been reported to be associated with repression only in the case of conditionally repressed inducible genes, which are associated mostly with H3K4me1, set by MLL2/3 itself, and with H3K27me3. It might be that the MLL2 enzyme also has functions at constitutively silenced loci, or that the low levels of H3K9me3 at the inducible genes would be sufficient to boost enzymatic activity and cause rapid repression of the gene upon repressive signal arrival. There are many points in the story of the MLL2, and also MLL3 and SET1A methyltransferases, that are still unresolved, and their clarification is eagerly awaited.

Although the majority of the PKMTs have probably been identified already, it is becoming evident that we presently understand only the tip of the iceberg of their biology. Probably most of the methyltransferases have multiple targets and each of these targets is part of the multitude of regulatory pathways existing in the cells. Novel approaches, such as the identification of targets through peptide array methyltransferase assay screening, are making the discovery process faster, and are shedding light upon the extraordinary crosstalk between regulatory pathways and also between modification pathways (Clarke, 2013), to find out not only the function of the methylation itself, but also how different modifications on the same substrate influence each other.
5.6. The SUV39H1 Chromodomain Binds Specifically to H3K9me3 and Seems to Inhibit the Catalysis of the SUV39H1 SET Domain under the Investigated Conditions

The human SUV39H1 methyltransferase and its homologs have been studied extensively since the discovery of the *Drosophila* SU(VAR)3-9 (Aagaard et al., 1999; Tschiersch et al., 1994). SUV39H1 is essential in the initiation of heterochromatin formation at pericentromeres and telomeres (García-Cao et al., 2004; Peters et al., 2003).

At the beginning of this study, the specificity and function of the chromodomain of the human SUV39H1 had not been characterized. It had only been suggested that the domain could have a role in the enzyme catalytic activity, because deletion or mutation of the aromatic cage of the domain resulted in impaired activity (Chin et al., 2006).

Here, the binding specificity for H3K9me3-containing peptides was shown *in vitro*, both with peptides immobilized to a substrate and with fluorescently labeled peptides in solution. The binding to H3K9me3 peptides was highly specific, and was confirmed later, by the end of 2012, by the publication of a crystal structure of the SUV39H1 chromodomain together with fluorescence polarization titrations with H3K9me3-peptides (Wang et al., 2012). The results of Wang et al. (2012) are similar to what I observed as well. The H3K9me3-containing peptide was the most preferred, followed by H3K9me2 peptides. However, Wang and colleagues (2012) show 5-fold better binding for the H3K9me3-containing peptide. Because the buffer conditions they used were very similar, this discrepancy could be explained only by the different boundaries chosen for the chromodomain. In their study, Wang et al., (2012), showed that the N-terminal boundary of the chromodomain is crucial for its binding capability. In particular, the residue at position 43, a phenylalanine, is contributing to the aromatic cage, and is therefore essential for peptide binding. They proved this by showing that the truncated chromodomain used for the crystal structure, SUV39H1 (res. 44-106), completely lost binding to target peptides, when compared to the SUV39H1 (res. 42-100). The domain used in the present study, SUV39H1 (res. 43-95), contains the aromatic cage phenylalanine 43, but the fact that this is the first amino acid in the sequence might be the cause for the higher $K_d$ observed in the fluorescence anisotropy measurements shown in the present work.
In addition to what was also reported by Wang et al., (2012), the peptide array analysis also revealed an interesting stimulation of binding in the presence of the double modification H3K9me3 and H3K14ac. This preference was considered an artifact at that time, since H3K9me3 and H3K14ac were considered to be mostly mutually exclusive marks. However, in light of new information which show overlap of H3K9me3 with H3K14ac (Soldi & Bonaldi, 2013), this preference should not be overlooked. It might be that the chromatin loci containing this double mark could recruit SUV39H1 with higher preference, and independently of HP1.

In addition, the far western data performed on a protein-rich chromatin fraction suggests that the SUV39H1 chromodomain might also have roles in interaction with other proteins, and even with other histones. The signal occurring at the size of the H1 protein is particularly interesting because H1 was shown to recruit SU(VAR)3-9 to heterochromatin in Drosophila (Lu et al., 2013). Moreover, H1b has been shown to be methylated at lysine 26 by EZH2, which is also responsible for H3K27 methylation (Kuzmichev et al., 2004). Both H3K27 and H1bK26, and also H3K9 are all found in a ARKS sequence context, and it might be possible that the SUV39H1 chromodomain would also specifically bind to the methylated H1 histone. There are no H1 peptides present on the MODIFYed™ Histone Peptide Arrays, and therefore this possible target was not included in the initial screening. Interestingly, Lu et al., (2013), did not show the region responsible for the interaction with H1, but they did observe a stimulation of the enzyme activity on reconstituted chromatin containing H1 histones compared to chromatin lacking H1. Moreover, the stimulation did not occur when the histones were free in solution, and in a nucleosomal context (Lu et al., 2013). Therefore, adding H1 to the reconstituted nucleosome methylation assay presented in this work might result in similar outcomes. Consequently, there are still several open directions in the search for additional SUV39H1 chromodomain targets and possible ways to address the function of the chromodomain in regard to the catalytic activity.

The first indications that the chromodomain may have an involvement in enzyme catalysis as well appeared in 2006, when Chin et al. did a comprehensive in vitro study to characterize the catalytic activity of SUV39H1. They observed a much lower turnover rate \(K_{cat}\) of human SUV39H1 on H3 (1-19) peptides than the Drosophila SU(VAR)3-9, corresponing to 8, and 396 turnovers per hour, respectively. In addition, they also showed that SAM is the limiting factor in the methylation reaction, as the \(K_m\) for SAM was 12 \(\mu\)M when recombinant H3 was used as a substrate. In their reactions therefore, Chin et al., (2006) used a mixture of tritium labeled
and “cold” SAM adding up to 25 μM concentration. For the experiments shown in the present work, only radioactively labeled SAM was used, in a final concentration of only 0.76 μM, which might have partly accounted for the lower activity of the enzyme. In addition, the purification of the SUV39H1 enzyme was done using a baculovirus expression system in the case of Chin et al., (2006), which resulted in higher purity of the protein. Another factor that could have influenced the results of the work presented here was the high degree of degradation in the purified protein samples. As western blot analysis showed (Figure 15, B), GST-tagged N-terminal truncations were present, suggesting that truncations containing the chromodomain alone, without the C-terminal SET domain, might also be present in the preparations. The individual chromodomains binding to methylated histones in the reaction might have influenced the results of the methyltransferase assays. All the experiments shown by Chin et al., (2006) were in an in vitro setting, and the stimulation experiments showing the increased activity of the wild type full length SUV39H1, compared to the full length chromodomain mutants, was only shown on H3 (1-19) peptides, and not on recombinant H3.

However, a new study confirmed the stimulation in the case of the yeast SUV39H1 homolog, Clr4, in a more native setting (Al-Sady et al., 2013). They used reconstituted dinucleosomes, which contained one “substrate” nucleosome which contained no PTMs, and one “effector” nucleosome, which either had a trimethyllysine analog at position 9, or a K9R mutation. Using this substrate, they showed that the full length Clr4 was stimulated when methylation occurred on nucleosomes which contained the MLA, compared to the ones containing the K9R mutation. The stimulation was characterized by a 5-fold increase in the $K_{cat}/K_M$ ratio. This experiment was similar to the SUV39H1 methyltransferase assay performed in this work on reconstituted oligonucleosomes. However, I did not observe any stimulation between the unmodified and the 50% K9me3 (MLA) oligonucleosomes. This could be explained by the general setup of the experiment. The oligonucleosomal substrates contained several adjacent nucleosomes with random incorporation of the H3 proteins with the K9me3 analog. In the case of Al-Sady et al., (2013), the substrates are much better defined, with only one chromodomain target and one methyltransferase substrate present in one dinucleosome, and the enzyme would always bind to them in the same way. Moreover, Al-Sady et al. (2013) also suggest that the binding of the chromodomain happens after the binding of the substrate by the SET domain, and that the chromodomain interaction with H3K9me3 serves to align the SET domain active site residues in a conformation that promotes catalysis. In my experimental
setup however, the incorporation of the chromodomain target histone and methyltransferase substrate histones was random, and therefore some of the spatial orientations required for binding both targets by the same molecule might not always have been favorable. This would decrease the visible effect of the stimulation, if there indeed was one. In addition, the fact that the experiments were done with enzymes from different organisms should not be taken lightly. It might be that such stimulation is actually stronger in the case of Clr4 than in the case of the human SUV39H1.

In the present study, some interesting observations were made also at cellular level. The initial attempt to detect methylation recovery upon SUV39H1 transfection with co-transfected HP1β as H3K9me3 reader was a confirmation of results published previously (Krouwels et al., 2005; Lachner et al., 2001). HP1β was reported to interact with the SUV39H1 N-terminal part (Stewart et al., 2005), and since we wanted to study the differential recovery of H3K9me3 between the wild type and the chromodomain mutants, it was decided to use antibody detection in order to avoid any interference caused by this interaction. However, Krouwels et al., (2005) also observed recovery of the H3K9me3 using antibodies against the mark, while in our setup, we failed to identify it. It might either be that the time required for the recovery was longer than the time in which the cells died as a result of the transfections. Or, detection of the methylation could also vary according to the antibodies used in each study. Other studies of SUV39H1-dependent recovery of H3K9me3 were done with transient transfection of the SUV39H1-Myc mammalian constructs that were delivered through virus infection, and the analysis of recovery was performed after 3 days from transfection (Lachner et al., 2001). It is possible that the longer time after transfection resulted in higher levels of H3K9 trimethylation set by the transfected enzymes. Also, the tag that was used is much shorter than the YFP tag, and a smaller protein might also be easier to express in the cells, and maybe cause less cell death upon transfection and protein expression. Due to failure to see clear recovery of H3K9me3 by antibody detection, the SUV39H1 chromodomain function in vivo could not be properly investigated.

However, from our in vitro experiments we derived a hypothesis of the possible mode of action of the SUV39H1 chromodomain. In our setup, we observed that the SET domain is more active than the chromodomain mutant, which in turn is more active than the full length wild type enzyme. Therefore, it seems that binding of the chromodomain to its H3K9me3 target results in inhibition of the SET activity rather than stimulation. In vitro, upon binding of the
H3K9me3 peptide by the chromodomain, access of the substrate H3 tail to the SET domain might somehow be precluded. Upon chromodomain mutation, which decreased, but not completely abolished binding, judging by the anisotropy measurements, the inhibition caused by the chromodomain would be partially alleviated, as observed. Moreover, in the complete absence of the chromodomain, the enzyme is even more stimulated, which fits to the hypothesis. A model of internal competition of substrate binding does not explain the experimental results, as this model would imply that the chromodomain is competing with the enzyme for the binding to unmodified or monomethylated H3, to which the chromodomain showed very weak binding. In addition, the chromodomain binds specifically to H3K9me3, which is the product of the enzymatic reaction, and this fact would rather support a model of stimulation by prevention of product inhibition of the enzyme. Therefore, a model of allosteric inhibition would best fit to our experiments, as the enzyme activity is decreased upon binding of the chromodomain to its target H3K9me3 peptide, and this could be achieved by changes in the protein conformation that would result in the lowering of the enzymatic reaction rate. However, more experiments should be performed, both in vitro and in vivo, in order to confirm or disprove these findings.
Conclusion

The work presented in this doctoral thesis concentrated on the elucidation of several important questions in the field of epigenetics. Although most of the projects are still in progress, the results shown here brought some light into the investigated subjects.

First, the new method for the Yeast-3-Hybrid method destined for the identification of PKMT targets in a methylation-dependent manner is still being optimized. Although the company-provided controls were working, our internal controls proved to be unfit for the purpose, and new controls will be designed and tested. In addition, the Gal4-fusion domains will be switched, to increase the chances of PKMT target methylation when the target is highly expressed from the ADH1 full length promoter.

In addition, the binding specificities of several “reading” domains was thoroughly characterized. The PHF1 tudor domain was shown to bind similarly to H3K36me3 and H3K27me3 by native histone pulldown, raising more questions about the function of this protein. In addition, the PHF1 tudor domain, together with the chromodomain of CBX7, bound specifically to H3K27me3, with a strong preference towards the H3T variant of histone 3, compared to the canonical H3.1 variant. This was shown to be a unique feature, characteristic to these domains, and not shared by the chromodomain of CBX2, which was shown to bind specifically to H3K27me3 on both H3.1 and H3T variants. For the future, histone variant-specific anti-H3K27me3 antibodies are the tools which need to be developed in order to solve the questions related to the role of the preference towards the H3T tail in the case of some histone tail “reading” domains.

Moreover, the study of the specificity of the PHD fingers of MLL2 and MLL3 is still in full progress. The MLL2 PHD 3-5 fragment and the MLL3 PHD 4-6 fragment showed very similar binding preferences, to modified peptides from H3 and H4 tails. The domains show to be more specific for peptides from H4 (11-30), especially containing H4K16ac, combined with other acetylations, arginine dimethylations, or lysine methylations. It is still needed to determine the exact specificity of these domains, and in the process, to identify the residues responsible for the peptide interaction. The function of the MLL2/MLL3 enzymes is currently being revealed, and the influence of the PHD fingers on this function is essential to be found.
Also, the substrate specificity of the MLL2, and also of SET1A methyltransferases, is essential for the elucidation of their additional roles in the epigenetic regulatory network. The differences between the specificity of MLL2 when it catalyzes alone and when it does it within the COMPASS-like complex are important features that might bring new insights into the differential pool of substrates that the enzyme might have in these two conditions. Additionally, the specificity profile of SET1A, which is active only in the COMPASS complex, revealed some interesting preferences compared to the wild type H3 peptide, suggesting the existence of other SET1A substrates, which might be even preferred to the standard H3 substrate. For the future, a screen for possible non-histone targets of SET1A/COMPASS and MLL2 alone and within the COMPASS-like complex will be performed, and any identified targets will be characterized in detail.

Last but not least, the SUV39H1 chromodomain binding specificity to H3K9me3 was characterized in detail, and its function with respect to the catalytic activity of the enzyme was investigated. Despite previous reports of stimulation, there was only evidence of inhibition caused by the chromodomain in our *in vitro* methyltransferase assays. For the future, the specificity of the SUV39H1 chromodomain towards other possible substrates, such as histone H1, would be the first direction to visit, as the similarity of the sequence in H1K26me3 and our results showing the binding of the SUV39H1 chromodomain to proteins corresponding to the size of H1, are strong indications to follow this lead.

All in all, the projects presented in the doctoral thesis show promising results and are still continued presently. With every answered question, there are several more questions arising and needing investigation, as the situation always is in the natural sciences. With this work, I hope I have achieved to add at least a few more pieces of knowledge to the great and largely unknown masterpiece that is our natural world.
References


174


184


Web page references:


Array Express Database: http://www-test.ebi.ac.uk/, retrieved on 15.05.2014

Atdbio: http://www.atdbio.com/content/56/Epigenetics, retrieved on 05.05.2014

Bgee Database: http://bgee.unil.ch/bgee/bgee, retrieved on 15.05.2014

CleanEx Database: http://cleanex.vital-it.ch/, retrieved on 15.05.2014

Geneinvestigator Database: https://www.genevestigator.com/gv/, retrieved on 15.05.2014

Scansite3Beta Database: http://scansite3.mit.edu, retrieved on 16.05.2014
Curriculum Vitae

Ralucia-Maria Tamas

Address: Beim Wasserturm 4
71332 Waiblingen
Germany
Date of Birth: 04.11.1986

Education

10/2010 to date
Universität Stuttgart – Stuttgart, Germany
Graduate Studies – Pursuing a Doctoral Degree at the Institute of Biochemistry, under the supervision of Prof. Dr. Albert Jeltsch
Expected graduation – September 2014

09/2008 – 09/2010
Jacobs University – Bremen, Germany
Graduate Studies – Master of Science Degree in Molecular Life Sciences with specialization in Cellular and Molecular Biology and in Molecular Genetics (Final GPA 1.03; where 1.0 – excellent, 4.67 – failing)

09/2005 – 06/2008
Jacobs University – Bremen, Germany
Undergraduate studies – Bachelor of Science Degree in Biochemistry and Cell Biology (cumulative GPA 1.26; where 1.0 – excellent, 4.67 – failing)

09/2001 – 09/2005
“Silvania” National College – Zalau, Romania
High School – with focus on Mathematics and Informatics (Valedictorian – overall GPA 9.96, where 10 – excellent, lower than 5 – failing)

Relevant Courses

Graduate: Physiology of Eukaryotic Cells, Molecular Genetics, Cellular Biochemistry, Molecular Life Sciences Seminar, Molecular Epigenetics, Molecular and Cellular Neurobiology, Microbial Pathogenicity, Literature Course Molecular Immunology, Molecular Genetics, Cellular Biochemistry, Quantitative Analysis of Biochemical Experiments.

Undergraduate: Advanced Biochemistry and Molecular Biology, Advanced Cell Biology, Methods in Biology and Biochemistry and Cell Biology, Biomedicine and Infection biology, Microbiology, Neurobiology and Physiology, Biochemistry and Biotechnology, Current Topics in Molecular Life Sciences Seminar, Molecular Biology and Genomics, General Biology, Animal Senses and Behavior, General Organic Chemistry, General Inorganic Chemistry, General Biochemistry and Cell Biology, Biochemistry and Molecular Biology Lab Course, Microbiology Lab Course, Genetics Lab Course, Molecular Cell Biology Lab Course.

Internship Experience

Panum Institute (University of Copenhagen) – Copenhagen, Denmark
Internship – Worked on synoviocytes, fibroblast-like cells from rheumatoid joint tissue. The project involved testing the induction of caveolae internalization using either GM1 or CD13 as caveolar markers. The methods involved immunofluorescence microscopy, electron microscopy, sucrose gradient ultracentrifugation, SDS-PAGE, western blotting
Awards

05/2008 The Oxford University Press (OUP) Biosciences Prize 2008 – award offered to most accomplished graduating students by OUP in collaboration with the Biosciences Department faculty at Jacobs University, Bremen

09/2005 – 06/2007 Jacobs University – Bremen, Germany Member of the President’s List of the Academic Years 2005/2006, 2006/2007 and 2007/2008 – award given to all students with the annual GPA between 1.50 and 1.00

Technical Skills

PCR, Molecular cloning, Site directed mutagenesis, Protein overexpression and purification, Fluorescence depolarization measurements, Cell culture, Immunohistochemistry, Mammalian cell transfection, Fluorescence microscopy, Laser scanning microscopy, basic skills of Electron Microscopy (TEM), Binding assays using arrays with synthesized peptide spots, Enzyme kinetics (methylation) assays with radioactive substrate, Yeast-3-hybrid protein-protein interaction studies, protein pull-down experiments, Mammalian cell chromatin and histone isolation, ChiP-qPCR, SDS-Polyacrylamide Gel Electrophoresis, Agarose gel electrophoresis, Western blot, Far western blot, Thin Layer Chromatography, Surface Plasmon Resonance, Circular Dichroism (CD) measurements, basic MALDI-TOF experience.

Publications


Other Skills

Languages: Romanian – mother language; English – proficient; German – basic knowledge

Computer skills: Basic knowledge in programming (Pascal, some Visual Fox Pro), MS Word, MS Excel, MS Power Point

Personal Interests
I love animals and nature in general. I enjoy traveling, which gives me the opportunity to explore new places and meet new people. I also like biking and hiking, as they give me a chance to experience the beauties of nature. I enjoy music, dancing and sports, even more if all three are combined. I am also happy when I have the time to spend with my close friends and loved ones or to quietly enjoy a good book.

**Application of Histone Modification Specific Interaction Domains as an Alternative to Antibodies**

Goran Kungulovski¹, Ina Kycia¹,⁴, Raluca Tamas¹, Renata Z. Jurkowska¹, Srikanth Kudithipudi¹, Chisato Henry², Richard Reinhardt³, Paul Labhart², and Albert Jeltsch¹,*

¹Institute of Biochemistry, Stuttgart University, Pfaffenwaldring 55, 70569 Stuttgart, Germany; ²Active Motif 1914 Palomar Oaks Way, Carlsbad, 92008 CA, USA; ³Max-Planck-Genomzentrum Köln, Carl-von-Linné-Weg 10, 50829 Köln, Germany; ⁴present address: The Jackson Laboratory for Genomic Medicine, 263 Farmington Avenue, Farmington, CT 06030, USA

**Running title**: Application of HMIDs as an alternative to antibodies

**Key words**: Histone post-translational modification, chromatin binding domains, protein design, histone antibody, ChIP-seq

*Corresponding author:
Prof. Dr. Albert Jeltsch
Institute of Biochemistry
Faculty of Chemistry
University Stuttgart
Pfaffenwaldring 55
D-70569 Stuttgart
Germany
Phone: +49 711 685 64390
Fax: +49 711 685 64392
albert.jeltsch@ibc.uni-stuttgart.de
Abstract

Post-translational modifications (PTMs) of histones constitute a major chromatin indexing mechanism and their proper characterization is of highest biological importance. So far, PTM specific antibodies have been the standard reagent for studying histone PTMs despite some caveats such as lot-to-lot variability of specificity and binding affinity. Herein, we successfully employed naturally occurring and engineered histone modification interacting domains for detection and identification of histone PTMs and ChIP-like enrichment of different types of chromatin. Our results demonstrate that histone interacting domains are robust and highly specific reagents and can replace or complement histone modification antibodies. These domains can be produced recombinantly in E. coli at low cost and constant quality. Protein design of reading domains allows for generation of novel specificities, addition of affinity-tags and preparation of PTM binding pocket variants as matching negative controls, which is not possible with antibodies.

**Highlights**

- The PHF1 Tudor domain binds H3K36me3 and H3tK27me3 (on the histone variant H3t).
- It uses the same trimethyllysine binding pocket for the interaction with both peptides.
- PHF1 co-localizes with K27me3 in cells, but not with K36me3.
- Our data suggest that PHF1 binds to H3tK27me3 in human chromatin.

**Abstract**

PHF1 associates with the Polycomb repressive complex 2 and it was demonstrated to stimulate its H3K27-trimethylation activity. We studied the interaction of the PHF1 Tudor domain with modified histone peptides and found that it recognizes H3K36me3 and H3tK27me3 (on the histone variant H3t) and that it uses the same trimethyllysine binding pocket for the interaction with both peptides. Since both peptide sequences are very different, this result indicates that reading domains can have dual specificities. Sub-nuclear localization studies of full-length PHF1 in human HEK293 cells revealed that it co-localizes with K27me3, but not with K36me3, and that this co-localization depends on the trimethyllysine binding pocket indicating that K27me3 is an *in vivo* target for the PHF1 Tudor domain. Our data suggest that PHF1 binds to H3tK27me3 in human chromatin, and H3t has a more general role in Polycomb regulation.

DOI: 10.1016/j.jmb.2013.08.009