Recognition of post-translational histone modifications by antibodies and epigenetic reading domains

by

Ina Bock

A Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biochemistry

Approved, Thesis Committee

Prof. Dr. Albert Jeltsch
(Professor of Biochemistry, Jacobs University Bremen)

Prof. Dr. Andreas Ladurner
(Professor of Biochemistry, Ludwig Maximillians University of Munich)

Prof. Dr. Matthias Ullrich
(Professor of Microbiology, Jacobs University Bremen)

Date of Defense: November 17th, 2011

School of Engineering and Science
Jacobs University Bremen, Germany
List of publications


Acknowledgements

I would like to express my gratitude to all those who supported me in completing this thesis. In this respect, I would like to thank first of all my supervisor, Prof. Dr. Albert Jeltsch, whose excellent guidance, support and valuable suggestions helped me in the entire research process and during the writing of this thesis. I also want to thank him for the opportunity to work in his lab and for introducing me to the exciting field of Epigenetics, for providing me with such interesting projects and for all the patience throughout the last three years.

I am thankful to Prof. Dr. Andreas Ladurner and Prof. Dr. Matthias Ullrich for being co-referees of my PhD thesis.

I would like to thank Dr. Philipp Rathert and Dr. Ole Brandt for their collaboration in the antibody project.

I would like to thank all my co-workers and former lab members for their help, support, interest and valuable advice. In this respect, I would like to thank especially Dr. Arunkumar Dhayalan, Dr. Srikanth Kudithipudi, Dr. Arumugam Rajavelu, Dr. Sergey Ragozine, Dr. Tomek Jurkowski, Dr. Renata Jurkowska and Raluca Tamas for all of the valuable discussions and ideas for improvement of experiments.

Many thanks are due to Srikanth for doing the peptide synthesis and thereby contributing so much to my work.

I would like to thank Qazi for the good collaboration regarding the MLL project.

Many thanks are due to Sandra, who did not only give excellent technical support for all the lab matters, but also is a delightful person to talk to.

I would like to thank Martin, Razvan, Suneetha, Pavel, Raghu, Nemanja, Goran, and Cristiana for not only being likable co-workers, but for creating such a warm atmosphere in the lab.

I am thankful to Raluca not only for scientific discussions, but also for becoming a friend within the last three years.

Last, but not least, I would like to thank my family and friends for all the support and patience. Many thanks are due to my sister Michaela, and my friends Bella, Ela, Dani, Ela, Meike, Anne, Brigitte, Nadine and Angela. I also would like to thank my former host family for always being there for me and my boyfriend, Mark, for all the support, patience, love, and especially for being exactly the way he is.
Abstract

Molecular epigenetics is described as the study of heritable changes in gene function without alterations in DNA sequence. Epigenetic phenomena are regulated by three interconnected players: DNA methylation, post-translational histone modifications and non-coding RNAs, which are involved in the regulation of many cellular processes including cellular differentiation and transcription. Post-translational histone modifications either directly modulate chromatin structure or they can serve as binding signals for reading domains, which recognize the modifications and mediate most of the biological functions of these modifications.

In the present PhD work, we have introduced Celluspots peptide arrays as a tool for the initial screening of putative reading domains interacting with a diverse set of post-translational histone modifications in one experiment in competition. With this tool, we studied the binding specificities of antibodies and reading domains with known as well as unknown primary targets. Furthermore, we studied differences and similarities in substrate recognition of two histone lysine methyltransferases, MLL3 and MLL1, from the same protein family.

Celluspots peptide arrays were first validated with antibodies directed towards post-translational histone tail modifications. In this study, we found that most of the antibodies bound well to the modification they had been raised for, but some failed. Some antibodies showed high cross-reactivity and most of the antibodies were inhibited by additional modifications close to the primary one. Furthermore, the comparison of the specificity profiles of antibodies, which had been raised for the same modification, revealed that the binding profiles sometimes differed greatly. Therefore, we did not only validate the method with this approach, but we further introduced Celluspots peptide arrays as a good tool for the quality control of epigenetic antibodies.

Additionally, we applied Celluspots peptide arrays for the investigation of the specificity of the interaction of reading domains with known substrate specificity with histone peptides. The results that we obtained with this approach agreed with literature concerning the primary targets of the reading domains, but we also obtained previously unknown information concerning the influence of secondary modifications for the binding affinity to the primary targets of these reading domains.

After validation of Celluspots peptide arrays, we screened approximately 20 reading domain candidates and proceeded with the most promising candidates for further analysis. One such candidate is a human polycomb group protein, which was shown to associate with the core components of the Polycomb repressive complex 2. The Polycomb repressive complex 2 is the major histone 3 lysine 27 methyltransferase-containing complex, which tri-methylates this lysine residue. Tri-methylated histone 3 lysine 27 is a post-translational modification, which is
associated with transcriptional repression of developmental genes, especially \textit{HOX} genes. We found that this polycomb group protein recognizes this histone tail modification in a histone variant specific manner. This histone variant is thought to be exclusively expressed in the mammalian testis and we propose that this polycomb group protein is involved in targeting the complex to this histone variant tri-methylated at lysine 27 in the human testis.

The histone lysine methyltransferases of the mixed lineage leukemia (MLL) protein family mono-, di- and tri-methylate histone 3 lysine 4. We studied the substrate specificity for two members of this family, MLL3 and MLL1, and revealed that the preferred substrate sequences are different. With these differences in substrate recognition we searched for non-histone targets and found similar, but also different non-histone targets at the peptide level. This is the first time that differences in substrate recognition was observed for MLL methyltransferases, which do not depend on differing MLL-complex members, which are known to be involved in targeting the methyltransferases to their gene targets.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ac</td>
<td>acetylation</td>
</tr>
<tr>
<td>ADD</td>
<td>ATRX-DNMT-DNMT3L</td>
</tr>
<tr>
<td>domain</td>
<td>(PHD finger like domain)</td>
</tr>
<tr>
<td>AdoHcy</td>
<td>S-adenosyl-L-homocysteine</td>
</tr>
<tr>
<td>AdoMet</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>ALL</td>
<td>Acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>AML</td>
<td>Acute myeloid leukemia</td>
</tr>
<tr>
<td>ASC-2</td>
<td>Activating signal cointegrator-2</td>
</tr>
<tr>
<td>ASCOM</td>
<td>ASC-2 complex</td>
</tr>
<tr>
<td>ASH2</td>
<td>Achaete-scute homologue 2</td>
</tr>
<tr>
<td>53BP1</td>
<td>p53 binding protein 1</td>
</tr>
<tr>
<td>BRD2</td>
<td>Bromo domain containing protein 2</td>
</tr>
<tr>
<td>CBP/p300</td>
<td>CREB binding protein</td>
</tr>
<tr>
<td>Chromo</td>
<td>Chromatin organization modifier domain</td>
</tr>
<tr>
<td>Dnmt3a</td>
<td>DNA methyltransferase 3a</td>
</tr>
<tr>
<td>EED</td>
<td>Extra-embryonic endoderm</td>
</tr>
<tr>
<td>EPC1</td>
<td>Enhancer of polycomb</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of Zeste 2</td>
</tr>
<tr>
<td>GLP</td>
<td>G9a-like protein</td>
</tr>
<tr>
<td>GNAT</td>
<td>Gcn5-related N-acetyltransferases</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone acetyltransferase</td>
</tr>
<tr>
<td>HCF1</td>
<td>Host cell factor 1</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>hDPY30</td>
<td>Human dpy-30-like protein</td>
</tr>
<tr>
<td>H1</td>
<td>Histone 1</td>
</tr>
<tr>
<td>H1.bK26</td>
<td>Histone variant 1.b lysine 26</td>
</tr>
<tr>
<td>H2A</td>
<td>Histone 2A</td>
</tr>
<tr>
<td>H2AK119</td>
<td>Histone 2A lysine 119</td>
</tr>
<tr>
<td>H2B</td>
<td>Histone 2B</td>
</tr>
<tr>
<td>H2BK120</td>
<td>Histone 2B lysine 120</td>
</tr>
<tr>
<td>AdoMet</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>AdoHcy</td>
<td>S-adenosyl-L-homocysteine</td>
</tr>
<tr>
<td>AML</td>
<td>Acute myeloid leukemia</td>
</tr>
<tr>
<td>ASH2</td>
<td>Achaete-scute homologue 2</td>
</tr>
<tr>
<td>53BP1</td>
<td>p53 binding protein 1</td>
</tr>
<tr>
<td>BRD2</td>
<td>Bromo domain containing protein 2</td>
</tr>
<tr>
<td>CBP/p300</td>
<td>CREB binding protein</td>
</tr>
<tr>
<td>Chromo</td>
<td>Chromatin organization modifier domain</td>
</tr>
<tr>
<td>Dnmt3a</td>
<td>DNA methyltransferase 3a</td>
</tr>
<tr>
<td>EED</td>
<td>Extra-embryonic endoderm</td>
</tr>
<tr>
<td>EPC1</td>
<td>Enhancer of polycomb</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of Zeste 2</td>
</tr>
<tr>
<td>GLP</td>
<td>G9a-like protein</td>
</tr>
<tr>
<td>GNAT</td>
<td>Gcn5-related N-acetyltransferases</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone acetyltransferase</td>
</tr>
<tr>
<td>HCF1</td>
<td>Host cell factor 1</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>hDPY30</td>
<td>Human dpy-30-like protein</td>
</tr>
<tr>
<td>H1</td>
<td>Histone 1</td>
</tr>
<tr>
<td>H1.bK26</td>
<td>Histone variant 1.b lysine 26</td>
</tr>
<tr>
<td>H2A</td>
<td>Histone 2A</td>
</tr>
<tr>
<td>H2AK119</td>
<td>Histone 2A lysine 119</td>
</tr>
<tr>
<td>H2B</td>
<td>Histone 2B</td>
</tr>
<tr>
<td>H2BK120</td>
<td>Histone 2B lysine 120</td>
</tr>
</tbody>
</table>

**Histone Modifications**

<table>
<thead>
<tr>
<th>Histone</th>
<th>Lysine Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3</td>
<td>3 lysine 4</td>
</tr>
<tr>
<td>H3.1</td>
<td>canonical 3.1</td>
</tr>
<tr>
<td>H3t</td>
<td>Testis specific 3t</td>
</tr>
<tr>
<td>H3K4</td>
<td>3 lysine 9</td>
</tr>
<tr>
<td>H3K27</td>
<td>3 lysine 27</td>
</tr>
<tr>
<td>H3K36</td>
<td>3 lysine 36</td>
</tr>
<tr>
<td>H3K56</td>
<td>3 lysine 56</td>
</tr>
<tr>
<td>H3K79</td>
<td>3 lysine 79</td>
</tr>
<tr>
<td>H3R2</td>
<td>3 arginine 2</td>
</tr>
<tr>
<td>H3S10</td>
<td>3 serine 10</td>
</tr>
<tr>
<td>H3S28</td>
<td>3 serine 28</td>
</tr>
<tr>
<td>H3T3</td>
<td>3 threonine 3</td>
</tr>
<tr>
<td>H4</td>
<td>4 lysine 4</td>
</tr>
<tr>
<td>H4K5</td>
<td>4 lysine 5</td>
</tr>
<tr>
<td>H4K12</td>
<td>4 lysine 12</td>
</tr>
<tr>
<td>H4K16</td>
<td>4 lysine 16</td>
</tr>
<tr>
<td>H4K20</td>
<td>4 lysine 20</td>
</tr>
<tr>
<td>H4R3</td>
<td>4 arginine 3</td>
</tr>
<tr>
<td>HKDM</td>
<td>histone lysine demethylase</td>
</tr>
<tr>
<td>HKMT</td>
<td>histone lysine methyltransferase</td>
</tr>
<tr>
<td>HP1β</td>
<td>heterochromatin protein 1 beta</td>
</tr>
<tr>
<td>ING</td>
<td>inhibitor of growth</td>
</tr>
<tr>
<td>JAZF1</td>
<td>juxtaposed with another</td>
</tr>
<tr>
<td>JMJD2A</td>
<td>Jumonji domain containing protein 2A</td>
</tr>
<tr>
<td>L3MBTL1</td>
<td>human lethal 3 malignant brain tumor repeat-like protein</td>
</tr>
<tr>
<td>MBT</td>
<td>malignant brain tumor</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>me1</td>
<td>mono-methylation</td>
</tr>
<tr>
<td>me2</td>
<td>di-methylation</td>
</tr>
<tr>
<td>me3</td>
<td>tri-methylation</td>
</tr>
<tr>
<td>MLL</td>
<td>Mixed lineage leukemia</td>
</tr>
<tr>
<td>MPP8</td>
<td>M-phase phosphoprotein 8</td>
</tr>
<tr>
<td>MYST</td>
<td>MOZ; Ybf2/Sas3; Sas2; TIP60</td>
</tr>
<tr>
<td>NURF</td>
<td>Nucleosomal remodeling factor</td>
</tr>
<tr>
<td>PA-1</td>
<td>PTIP-associated 1</td>
</tr>
<tr>
<td>PcG</td>
<td>Polycomb group protein</td>
</tr>
<tr>
<td>PCL</td>
<td>Polycomblike protein</td>
</tr>
<tr>
<td>ph</td>
<td>phosphorylation</td>
</tr>
<tr>
<td>PHD</td>
<td>Plant homeodomain</td>
</tr>
<tr>
<td>PHF1</td>
<td>PHD finger protein 1</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>PR (PRDI-BF1 and RIZ homology) domain</td>
</tr>
<tr>
<td>PRE</td>
<td>Polycomb response element</td>
</tr>
<tr>
<td>PRMT</td>
<td>Protein arginine N-methyltransferases</td>
</tr>
<tr>
<td>PTIP</td>
<td>Pax2 transactivation domain-interacting protein</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>PWWP</td>
<td>Proline-Tryptophan-Tryptophan-Proline motif</td>
</tr>
<tr>
<td>RbAp46/48</td>
<td>Retinoblastoma-binding protein 46/48</td>
</tr>
<tr>
<td>RBBP5</td>
<td>Retinoblastoma binding protein 5</td>
</tr>
<tr>
<td>SET</td>
<td>Su(VAR)3-9; enhancer of zeste (E(Z)); trithorax (TRX)</td>
</tr>
<tr>
<td>SETDB1</td>
<td>SET domain bifurcated 1</td>
</tr>
<tr>
<td>SMYD</td>
<td>SET and MYND domain</td>
</tr>
<tr>
<td>STAT5</td>
<td>Signal transducer and activator of transcription 5</td>
</tr>
<tr>
<td>SUMO</td>
<td>Small ubiquitin-like modification</td>
</tr>
<tr>
<td>SUV39H1</td>
<td>Suppressor of variegation 3-9 homolog 1</td>
</tr>
<tr>
<td>SUZ12</td>
<td>Suppressor of Zeste 12</td>
</tr>
<tr>
<td>ub1</td>
<td>mono-ubiquitylation</td>
</tr>
<tr>
<td>UTX</td>
<td>Ubiquitously transcribed TPR protein transcribed on the X-chromosome</td>
</tr>
<tr>
<td>WDR5</td>
<td>WD repeat containing protein 5</td>
</tr>
<tr>
<td>YY1</td>
<td>Transcription factor Ying Yang 1</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Epigenetics

1.2 Organization and Function of Chromatin

1.3 Post-translational Histone Modifications

1.3.1 Lysine Acetylation

1.3.2 Lysine Methylation

1.4 Histone Lysine Methyltransferases

1.5 Reading Domains

1.5.1 Readout of Acetyllysine

1.5.2 Readout of Methyllysine

2 The Aim of the present study

2.1 Specific goals and achievements of the project

2.1.1 Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays

2.1.2 Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails

2.1.3 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

2.1.4 Similarities and differences in substrate specificity of MLL3 and MLL1 of the mixed lineage leukemia family of methyltransferases

3 Results and Discussion

3.1 Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays

3.2 Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails

3.3 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

3.3.1 The role of PHF1 in polycomb group silencing, DNA repair and disease

3.3.2 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

3.3.3 The EED WD40 repeats do not discriminate between H3t and H3.1

3.3.4 Discussion and Conclusion
3.4 Similarities and differences in substrate specificity of MLL3 and MLL1 of the mixed lineage leukemia family of methyltransferases 36
3.4.1 The role of MLL family members in HOX gene activation, nuclear-receptor mediated gene activation and in disease 36
3.4.2 Similarities and differences in substrate specificity of MLL3 and MLL1 38
3.4.3 Discussion and Conclusion 43

4 Materials and Methods 45
4.1 Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays 45
4.2 Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails 45
4.3 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t 45
4.3.1 Cloning, site directed mutagenesis, expression and purification 45
4.3.2 Synthesis of Celluspots peptide arrays and peptide SPOT arrays 46
4.3.3 Conversion of cysteine into tri-methyllysine analogs 47
4.3.4 Interaction of GST fusion protein domains with post-translational histone tail modifications on peptide arrays 47
4.3.5 Circular dichroism spectroscopy 47
4.3.6 Peptide pulldown experiments 47
4.3.7 Equilibrium peptide binding experiments 48
4.3.8 Pulldown experiments with native histones 49
4.3.9 Cell culture, transfection and confocal fluorescence microscopy 49
4.4 Similarities and differences in substrate specificity of MLL3 and MLL1 of the mixed lineage leukemia family of methyltransferases 50
4.4.1 Cloning, expression and purification 50
4.4.2 Synthesis of peptide SPOT arrays 50
4.4.3 Methylation of peptide SPOT arrays 50

5 References 52
1 Introduction

1.1 Epigenetics

The term epigenetics comprises two words: "epi-", which means above or beyond and "genetics", which stands for inheritance of variation in living organisms. Epigenetic phenomena involve heritable changes in gene function without alterations in DNA sequence [Bonasio et al., 2010]. A more general definition was also proposed in 1996, in which epigenetics is termed as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" [Russo et al., 1996]. Epigenetic signals include DNA methylation, histone post-translational modifications and non-coding RNAs, which interplay with each other and with regulatory proteins, to define the chromatin structure of a gene and its transcriptional activity. Various cellular processes are regulated by these three key players including transcription, DNA replication, DNA repair, cellular differentiation, development, and suppression of transposable element mobility. Most epigenetic signals are reversible, which allows cells to adapt and respond to environmental stimuli. However, epigenetic marks can also have long term effects, for example DNA methylation and histone acetylation are involved in learning and memory formation [Nelson and Monteggia, 2011]. Furthermore, epigenetics plays a crucial role in maintaining normal development, because aberrant placement of epigenetic marks is involved in causing many diseases. These diseases are for example cancer, neurological disorders (e.g. Alzheimer’s disease, Multiple Sclerosis or ATRX syndrome) or autoimmune diseases (e.g. Rheumatoid arthritis or Diabetes type I) [Portela and Esteller, 2010]. Cancer was originally thought to be caused by mutated genes, but later on it was found that epigenetic alterations often accompany genetic mutations that cause cancer [Brower, 2011]. Common epigenetic changes in cancer cells are hypermethylation of tumor suppressor genes accompanied by high levels of silencing histone modifications, while a global loss of DNA methylation is observed [Portela and Esteller, 2010].

1.2 Organization and Function of Chromatin

In the nucleus of eukaryotes, DNA is packaged into chromatin. The fundamental unit of chromatin is the nucleosome. One nucleosome is composed of a histone protein octamer with two copies each of the core histones H3, H4, H2A and H2B with approximately 146 base pairs (bp) of DNA wrapped 1.7 times around it [Luger et al., 1997] (Figure 1).
The histones H3 and H4 interact through a 4-helix bundle from both H3 histones and thereby form a H3-H4 tetramer, while H2A and H2B form two dimers [Luger et al., 1997]. Binding of the histone octamer to DNA occurs primarily at the DNA phosphodiester backbones.

Easily accessible, transcriptionally active chromatin is formed by repeating nucleosomal units connected with linker DNA, which appear like “beads on a string” also known as the 11 nm fiber [Trojer and Reinberg, 2007] (Figure 2). Transcriptionally active chromatin is also called euchromatin. This definition is based on the density of staining of the nucleic acid in micrographs [Heitz, 1929]. Promoter regions of transcribed genes are mainly devoid of nucleosomes. Repositioning of nucleosomes is therefore essential for both the compaction into higher order chromatin, also called heterochromatin, as well as for transcription in active genes of euchromatic regions. ATP-dependent chromatin remodeling complexes are indispensable for movement or exchange of nucleosomes, e.g. nucleosomes containing histone variants executing different functions [Hargreaves and Crabtree, 2011].

Figure 1: Illustration of one nucleosome core particle. 146 bp doublestranded DNA (green and brown) are wrapped around the eight histone proteins (H3: blue, H4: green, H2A: yellow, H2B: red) [adapted from Luger et al., 1997].

Figure 2: Chromatin Organization. Chromatin is easily accessible in the 11 nm fiber. Further compaction is achieved by the incorporation of Histone H1 leading to the formation of a 30 nm fiber. Metaphase chromosomes establish the highest level of chromatin condensation [adapted from Trojer and Reinberg, 2007].
Further compaction of DNA is achieved by incorporation of the linker Histone H1 leading to a 30 nm fiber and higher order chromatin [Robinson et al., 2006] (Figure 2). Heterochromatin is divided into two groups: facultative and constitutive heterochromatin. Facultative heterochromatin is found in genomic regions, where developmental genes are differentially expressed during embryogenesis and differentiation and which then become silenced. In contrast, constitutive heterochromatin contains permanently silenced regions like centromeres and telomeres [Grewal and Jia, 2007]. Heterochromatin is crucial for the structure of chromatin. This structure compacts the genetic information, but also protects certain chromosomal areas like the telomeres. Additionally, centromeric heterochromatin formation is required for the proper segregation of chromosomes during mitosis [Pidoux and Allshire, 2005]. The original definition of euchromatin and heterochromatin has evolved over the years. Now euchromatin is further defined as gene-rich, transcriptionally active and hyperacetylated, while heterochromatin is gene-poor, transcriptionally inactive, hypoacetylated and hypermethylated [Trojer and Reinberg, 2007; Kouzarides, 2007]. Furthermore, euchromatic genes are replicated early on, whereas heterochromatic genes are replicated later. In Drosophila this general classification into euchromatin and heterochromatin was further diversified into five principal chromatin types [Filion et al., 2010]. Out of these five chromatin types two can be characterized as euchromatic and three as different heterochromatic types. This study shows that the classification of chromatin into euchromatin and heterochromatin is very simplistic and that there are probably distinct subtypes of chromatin in other organisms as well.

Chromatin structure and function is further regulated by the exchange of canonical histones with histone variants [Talbert and Henikoff, 2010]. There is an increasing list of histone variants of H3, H2A, H2B and H1 histones. Non-canonical histone variants differ in their primary amino acid sequence and have various roles in processes including transcription initiation and termination, DNA repair, meiotic recombination or sperm chromatin packaging [Talbert and Henikoff, 2010]. One major difference between canonical and non-canonical histones (also called replacement histones) is that the canonical histones are expressed in
the S-Phase of the cell cycle, while replacement histones can be expressed and incorporated into the nucleosome throughout the whole cell cycle [Loyola and Almouzni, 2007].

Some of the variants are directly implicated in chromosome condensation like MacroH2As, which are for example enriched on the silenced female Xi chromosome in humans. Other variants can be essential for kinetochore assembly during mitosis like the human CENP-A, which is a highly divergent H3 variant and shares only 46% identity with the canonical H3.1 [Talbert and Henikoff, 2010; Loyola and Almouzni, 2007]. There are also lineage-specific histone variants, which are specialized for distinct functions in particular tissues. For example, one of these variants is the testis specific histone variant H3t (Figure 3). This variant is supposed to be restricted to mammals and it is thought to be exclusively expressed in testis [Schenk et al., 2011]. The amino acid sequence of H3t differs from H3.1 in four amino acids: A24V, V71M, A98S and A111V (Figure 3). M71 and V111 in H3t are implicated in forming less stable H3-H4 tetramers, which is important for rapid exchange of histones with protamines during spermatogenesis [Tachiwana et al., 2010].

1.3 Post-translational Histone Modifications

Histones are mainly globular, but their N-terminal tails protrude from the nucleosome and can form contacts with the DNA, with adjacent nucleosomes or can interact with other proteins. The N-terminal histone tails and a few residues within the histone cores are known to be modified with various post-translational modifications (PTMs) at many different sites [Bannister and Kouzarides, 2011; Kouzarides, 2007] (Figure 4). These PTMs can influence the structure of chromatin either directly by neutralizing or adding charge to the amino acid side chains thereby loosening or forming contacts with the surrounding DNA of the nucleosome or indirectly as a binding signal for effector proteins also called reading domains (discussed in 1.5). Histone PTMs are very dynamic, that means not all modifications occur at the same time at the same place. Some sites can be modified with more than one PTM, therefore these modifications are mutually exclusive since they cannot appear at the same time (Figure 4). Most histone PTMs are reversible. Depending on the stimuli in the nucleus, amino acids are modified by enzymes which are setting the mark and can be converted back into their original state by enzymes removing the mark.

The most diversely modified amino acid is lysine, which can be acetylated, mono-, di- or trimethylated, but also can be modified with larger modifications like ubiquitin or small ubiquitin-like modification (SUMO). Lysine histone acetylation is, in general, associated with transcriptional activation, while lysine methylation can be either an activating mark or a repressing mark depending on the site and the number of methyl groups added (further discussed in 1.3.1 and 1.3.2).
Figure 4: Post-translational histone modifications. The main post-translational histone modifications of H2A, H2B, H3.1 and H4 are shown in this figure: acetylation in blue (A), methylation in red (M), phosphorylation in yellow (P) and ubiquitylation in green (U). The gray numbers under the modified amino acids indicate their position in the sequence [adapted from Portela and Esteller, 2010].

Ubiquitylation of histones is not as well characterized due to the size of 76 amino acids of ubiquitin, but two known sites lie within the histone cores of H2A and H2B. Mono-ubiquitylation of H2BK120 (H2BK120ub1) is involved in transcriptional initiation and elongation [Lee et al., 2007; Kim et al., 2009], while H2AK119ub1 is linked with transcriptional repression [Wang et al., 2004] (Figure 4). Sumoylation was found on all four core histones and is thought to act by antagonizing lysine acetylation and ubiquitylation. Hence, it is implicated in transcriptional repression [Bannister and Kouzarides, 2011].

Arginine can also be methylated, but in contrast to lysine it can be only mono- or di-methylated, which can occur in symmetrical/asymmetrical manner. Arginine methylation, mediated by protein arginine N-methyltransferases (PRMTs), is associated with both transcriptional activation and transcriptional repression. The deimination of arginine converts arginine to citrulline, thereby neutralizing the positive charge of arginine [Bannister and Kouzarides, 2011]. For a long time citrullination was thought to be the counter-reaction to arginine methylation, but more recently it was shown that the jumonji protein JMJD6 can demethylate mono- and di-methylated H3R2 and H4R3 (H3R2me1/me2 and H4R3me1/me2) [Chang et al., 2007].

Another important modification is histone phosphorylation. Phosphorylation occurs on serines, threonines and tyrosines mainly on the N-terminal tails, but a few phosphorylation sites also lie within the histone cores. The role of histone phosphorylation during transcription is not yet fully elucidated, but it is an important modification for DNA repair and for chromatin condensation [Kouzarides, 2007]. Upon DNA damage the histone variant γ-H2AX becomes phosphorylated in mammalian cells and therefore serves as one of the earliest signals in DNA damage control. During Mitosis Aurora B kinase phosphorylates H3S10 in order to decondensate chromatin, because this phosphorylation leads to the release of the Heterochromatin protein 1 beta (HP1β), which is involved in chromatin compaction, from the
adjacent H3K9me3 [Fischle et al., 2005] (Figure 4). In addition, the phosphorylation of H3T3 mediated by Haspin kinase is indispensable for normal metaphase chromosome alignment [Dai et al., 2005].

Other reported histone PTMs include proline isomerization, ADP ribosylation and the more recently discovered histone modifications β-N-acetylglucosamine and histone tail clipping [Bannister and Kouzarides, 2011]. Mono- and poly-ADP ribosylation takes place on glutamate and arginine residues, though there is not much known about the function of this modification [Hassa et al., 2006].

So far, many non-histone proteins were reported to be modified with the sugar β-N-acetylglucosamine on the side chains of serine and threonine and not long ago, the histones H2A, H2B and H4 were discovered to be also modified with this sugar [Sakabe et al., 2010].

1.3.1 Lysine Acetylation

The N-terminal histone tails are lysine-rich and many lysine residues on all four core histones as well as the linker histone H1 are reported to be acetylated under certain conditions (Figure 4). Lysine acetylation is regulated by two counteracting enzyme families: histone acetyltransferases (HATs) and histone deacetylases (HDACs) [Bannister and Kouzarides, 2011]. HATs transfer an acetyl-group of the cofactor Acetyl-CoA to the ε-amino group of the lysine side chain thereby generating acetyllysine (Figure 5). The unmodified lysine side chain is positively charged and has the potential to interact with the negatively charged DNA phosphodiester backbone. Upon addition of the acetyl-group, this electrostatic interaction is weakened by the neutralization of the positive charge of lysine. Thus, this modification has a huge impact on the structure of chromatin.

HAT enzymes can be categorized into the two main classes: type-A and type-B. The type-B HATs are mainly found in the cytoplasm, while the type-A enzymes show primarily a nuclear localization. The eukaryotic type-A HATs count three main families: GNAT, MYST and CBP/p300. These HATs are present in many large multiprotein complexes and acetylate many lysine residues in the N-terminal histone tails.

Figure 5: Schematic illustration of the catalytic mechanism of the reversible transfer of the acetyl-group to the lysine side chain mediated by HATs and retracted by HDACs [adapted from Seet et al., 2006].
In general, they show low specificity, because they acetylate more than one site. In mammals, HDACs are categorized into three classes: Class I and II deacetylases and class III, which are NAD-dependent enzymes of the Sir family [Kouzarides, 2007]. Like HATs, HDACs also perform their deacetylase function mainly in multiprotein complexes. Part of these complexes can be (co-)activators in the case of HAT-containing complexes or (co-)repressors in regard to HDAC multiprotein complexes, which recruit the complexes to their target sites.

Newly synthesized histones are rapidly acetylated at H4K5, H4K12 and some H3 lysines in the N-terminal tails by type-B HATs in the cytosol in many eukaryotes [Sobel et al., 1995; Parthun, 2007]. The acetylation pattern seems to be important for the recognition by chaperones, which assemble the new histones with replicated DNA [Shahbazian and Grunstein, 2007]. Additionally, it was shown that H3K56 in the histone core is also acetylated after histone expression and this acetylated residue is important for proper nucleosome assembly, but H3K56ac has additional functions as well e.g. in DNA repair [Xu et al., 2005; Tjeertes et al., 2009; Das et al., 2009]. Immediately after the assembly of the nucleosome, the histones are deacetylated. However, acetylation is not only important for correct nucleosome assembly, but also has various roles in transcription, DNA repair, DNA replication and chromosome decondensation. Acetylated H4K16 was discovered to have a special role in chromosome decondensation. Unmodified H4K16 interacts with an acidic patch on H2A thereby promoting nucleosome compaction. In vitro it was shown that the acetylation of H4K16 does not inhibit the assembly of the histone octamer, but prevents further compaction into a 30 nm like fiber [Shogren-Knaak et al., 2006].

In transcription, histone acetylation plays a very important role. The promoter regions of transcribed genes are hyperacetylated at many N-terminal histone tails with a low level of acetylation found in the gene body. Nonetheless, long-range histone acetylation is also found (over several kilobases) in transcribed eukaryotic regions, which can include individual genes, but also entire gene-clusters [Calestagne-Morelli and Ausió, 2006]. For example, long-range hyperacetylation is observed on developing tissues in the Hox gene clusters (homeotic genes) or on the growth hormone gene cluster in humans [Calestagne-Morelli and Ausió, 2006].

Besides histone acetylation, there are many non-histone proteins, which are regulated by acetylation [Choudhary et al., 2009]. Many of these proteins are also involved in the regulation of DNA replication, repair or transcription, which further exemplifies the importance of this PTM.
1.3.2 Lysine Methylation

In contrast to histone lysine acetylation, methylation does not alter the charge of the lysine side chain. Therefore, this modification has no influence on electrostatic interactions with the surrounding DNA. Nevertheless, histone lysine methylation is required for many biological processes including heterochromatin formation, regulation of transcription or X-chromosome inactivation [Martin and Zhang, 2005].

Histone lysine methylation is dynamically regulated by lysine methyltransferases (HKMTs) (further discussed in 1.4) and lysine demethylases (HKDMs) [Allis et al., 2007]. Methylation takes place on all four core histones (Figure 4), but five lysine residues on H3 and H4 are the most thoroughly studied histone lysine methylation marks and are implicated in the above-mentioned processes by providing binding sites for reading domains of effector proteins that mediate downstream effects. These methylated lysine residues are H3K4, H3K9, H3K27, H3K36 and H4K20 (Figure 6). In general, methylation of H3K4 and K36 are associated with transcriptionally active genes, while the methylation of H3K9, K27 and H4K20 are implied in the regulation of gene silencing. However, many of these methylation sites have several functions, which seem to have opposing roles [Berger, 2007]. Therefore, most methylation sites cannot be strictly categorized in transcriptionally activating or repressing marks, but rather need to be evaluated in the context with other post-translational histone modifications and other epigenetic marks. For instance, H3K9me3, a modification strongly associated with transcriptional repression, is also found in the body of active genes [Mikkelsen et al., 2007], whereas if H3K9me3 is found in promoter regions it correlates with transcriptional repression [Barski et al., 2007; Mikkelsen et al., 2007].

![Figure 6: Major lysine methylation sites on the N-terminal histone tails of H3 and H4. The numbers refer to the methylated amino acid residue on the histones. The color code shown on the left side describes the functions for each state of the mono-, di- and tri-methylated lysine residues [adapted from Mosammaparast and Shi, 2010].](image)
Nonetheless, H3K4me3 is predominantly found on hyperacetylated H3 tails in the 5' end of active genes [Garcia et al., 2007; Liang et al., 2004]. H3K36me3 is a repressing mark located in the body and the 3' end of active genes (Figure 6) [Mikkelsen et al., 2007; Barski et al., 2007; Guenther et al., 2007]. In eukaryotes, H3K36 methyltransferases are associated with the elongating form of RNA Polymerase II and this methylation mark in turn recruits transcriptional repressors, which function to suppress internal initiation of transcription [Klose and Zhang, 2007]. Moreover, the methylation state of one lysine residue can have different functions. For example, H3K4me1, but not H3K4me3 or me2, are found at enhancer elements (Figure 6) [Heintzman et al., 2009].

H3K9me3 and H4K20me3 are both important modifications associated with constitutive heterochromatin, while H3K27me3 is considered a hallmark of facultative heterochromatin. The formation of constitutive heterochromatin involves the recognition of repetitive DNA elements by the RNAi machinery [Grewal and Jia, 2007]. Histone-modifying enzymes like HKMTs or HDACs are recruited in both cases to set repressive methylation marks and to remove the activating acetyl-groups. HP1β binds via its Chromo domain to H3K9me3 modified histone tails and interacts with the H3K9me3 methyltransferase SUV39H1. This interaction prompts further methylation of H3K9 on the histone tails of neighboring nucleosomes. Therefore, this interplay of SUV39H1 with HP1β may induce the spreading of heterochromatin in the surrounding sequences. Boundary elements with distinct histone modification patterns prevent the spreading of heterochromatin into adjacent euchromatic regions [Grewal and Jia, 2007].

H3K27me3 is enriched on silenced developmental genes like the four Hox gene clusters. Silencing of developmental genes in embryonic stem (ES) cells is required for maintaining pluripotency [Hawkins et al., 2011]. In ES cells some extended regions are enriched with H3K27me3, but also H3K4me3. These regions are termed as “bivalent domains”, since they harbor activating and repressing marks at the same time [Bernstein et al., 2006]. These bivalent domains are transcriptionally silenced, showing that the repressing mark overrules the activating mark in this case. Upon differentiation these regions remain either the activating H3K4me3 mark or the repressing H3K27me3 modification. Therefore, these “bivalent domains” are interpreted as poised for activation or repression after differentiation [Bernstein et al., 2006]. Another epigenetic phenomenon where H3K27me3 plays an important role is the X-chromosome inactivation in females [Plath et al., 2003]. In mammals, dosage compensation is accomplished by transcriptional silencing of one of the X-chromosomes [Augui et al., 2011]. In female somatic cells, initiation and maintenance of the silenced X-chromosome Xi is established by the RNA component Xist. Xist is transcribed from the Xi and coats the entire chromosome. Other epigenetic modifications on the Xi
include H4 hypoacetylation [Jeppesen and Turner, 1993], enrichment of variant histone macroH2A and DNA methylation [Brockdorff, 2002].

1.4 Histone Lysine Methyltransferases

All known histone lysine methyltransferases (HKMTs), with the exception of one enzyme, contain a so-called SET domain composed of 130 amino acids, harboring the methyltransferase activity. The SET domain was first identified in Drosophila and is highly conserved from yeast to human and is termed after its founding members: Su(VAR)3-9, enhancer of zeste (E(Z)) and trithorax (TRX) [Jenuwein et al., 1998]. The only known non-SET domain containing enzyme is Dot1 (Dot1L in humans), which methylates H3K79 in the core of histone H3 [Feng et al., 2002].

Up to date, more than 50 SET domain-comprising proteins with predicted or verified methyltransferase activity towards histone tails have been identified in humans [Upadhyay and Cheng, 2011]. Most of these SET domain-containing methyltransferases can be categorized into six subfamilies: SET1, SET2, SUV39, EZH, SMYD and PRDM. The SET domain itself is structurally characterized by a series of β-strands and an α-helix, which together form a loop adopting a knot-like structure [Jacobs et al., 2002]. The sequences preceding and following the SET domain (pre- and post-SET) are important for stabilizing the structure and are indispensable for the catalytic activity. Most pre- and post-SET domain regions are Cysteine-rich and are dependent on binding of Zn$^{2+}$-ions for the catalytic activity of the SET domain [Jacobs et al., 2002; Min et al., 2002]. Another conserved region within the SET domain is the co-factor S-adenosyl-L-methionine (AdoMet) binding site. It has been proposed, that within the catalytic pocket, a conserved tyrosine residue deprotonates the ε-amino group of the lysine site chain.

![Figure 7: Mechanism of the methyl-group transfer to lysine residues mediated by SET domain containing methyltransferases. A conserved tyrosine residue within the catalytic pocket deprotonates the ε-amino group of lysine thereby prompting a nucleophilic attack of lysine on the methyl-group of AdoMet (or SAM). Methyllysine and AdoHcy are the products of this catalytic reaction [adapted from Wood and Shilatifard, 2004].](image-url)
This deprotonation prompts a nucleophilic attack of the lysine on the methyl-group of AdoMet resulting in the methylated lysine residue and S-adenosyl-L-homocysteine (AdoHcy) as byproduct (Figure 7) [Wood and Shilatifard, 2004].

In humans, among the best characterized HKMTs are the H3K9 methyltransferases SUV39H1, G9a, G9a-like protein (GLP) and SETDB1. SUV39H1 tri-methylates H3K9 and is involved in the spreading of H3K9me3 in the formation of heterochromatin [Grewal and Jia, 2007], while G9a and GLP are both mono- and di-methyltransferases and mainly methylate H3K9 in euchromatic regions [Tachibana et al., 2002]. The mixed lineage leukemia (MLL) enzymes 1-5 from the Trithorax group, SET1 and SET7/9 are classified as H3K4 methyltransferases [Hublitz et al., 2009]. H3K27 methylation is accomplished by the polycomb group enzyme Enhancer of Zeste 2 (EZH2). EZH2 is the catalytic subunit of the Polycomb repressive complex 2 (PRC2), which is crucial for HOX gene silencing and X-chromosome inactivation [Margueron and Reinberg, 2011]. Moreover, many of these HKMTs methylate lysine residues in non-histone proteins as well [Rathert et al., 2008; Dhayalan et al., 2011a].

1.5 Reading Domains

Histone tail PTMs are recognized and bound i.e. “read” by protein modules termed as effectors or “reading domains”, which mediate most of the biological functions of these modifications [Taverna et al., 2007; Yun et al., 2011]. Many of the reading domains are part of proteins which also carry other functional domains like enzymatic activities. Moreover, multiprotein complexes often comprise various proteins with different reading modules, working in concert in fulfilling their biological functions. The binding affinities of reading domains for their targets have been shown to be relatively weak with $K_D \sim 10^{-4}-10^{-6}$ M [Daniel et al., 2005], but multidentate interactions lead to avidity effects, thereby potentially causing a much stronger association with chromatin.

In general, reading domains supply accessible surfaces i.e. cavities or surface grooves to harbor PTMs. They distinguish different modifications and also recognize the status of the modification e.g. in the case of mono-, di- or tri-methylation. Additionally, reading domains interact with the surrounding amino acid residues in order to differentiate the sequence context [Yun et al., 2011].

1.5.1 Readout of Acetyllysine

The classical acetyllysine reading module is the Bromo domain. The Bromo domain is an approximately 110 amino acid residues long domain folded into a left-handed antiparallel four
α-helical bundle with a hydrophobic binding pocket located between two loops of the α-helices [Dhalluin et al., 1999]. Bromo domains are frequently found in nuclear HATs such as yeast Gcn5p (GCN5L in humans), which was the first identified Bromo domain containing protein [Owen et al., 2000], or chromatin remodeling complexes. Moreover, many proteins harbor multiple Bromo domains i.e. tandem or double Bromo-modules like TAF1 (formerly TAF\(_{1250}\), Figure 8A), the largest subunit of the Transcription Factor TFIID, which preferentially bind multiple acetylated histone peptides. TAF1 binds preferentially to H4K5ac-K12ac [Jacobson et al., 2000].

More recently, a tandem PHD finger of human DPF3b was recognized as a novel acetyllysine reading domain [Lange et al., 2008; Zeng et al., 2010]. Before this discovery many PHD fingers or PHD finger-like reading domains were identified as either binding modules for methylated or unmodified lysine residues.

1.5.2 Readout of Methyllysine

Binding modules recognizing methylated lysine residues are a much more diverse and large group in comparison to acetyllysine binders. This group includes Chromo, PHD, Tudor, and PWWP domains, WD40, MBT and Ankyrin repeats. Overall these reading domains share similar recognition modes for the same methylation status [Yun et al., 2011].

In general, aromatic residues in the binding pocket contact the methalammonium moiety of methyllysine, thereby forming a so-called aromatic cage. Based on the number of aromatic residues, the aromatic cages are classified as half or full aromatic cages. However, the number of aromatic residues does not correlate with the methyl status of the lysine residue. Reading domains with preference for mono- and di-methyl-groups usually have narrow cavities, which deny access of the bulkier tri-methylated lysine residues. Tri-methyl binders mostly have wider binding pockets and form contacts to the methyl-groups via cation-π and van der Waals interactions [Yun et al., 2011].

The chromatin organization modifier domain (Chromo domain) belongs to the Royal superfamily and comprises about 50 amino acids. Members of the Royal family of reading domains have a characteristic fold of 4 β-sheets which fold into an “incomplete β-barrel” [Taverna et al., 2007]. The methylated histone peptide inserts into this structure and thereby completes the barrel. The Chromo domain family consists of more than 120 identified members and their best characterized member is HP1β, which binds to H3K9me3 at pericentric heterochromatin (Figure 8B; see 1.3.2). Other Chromo domains were shown to bind to H3K27me3 and H3K4me3.
For example, the Chromo domain of Polycomb protein (Pc, known as CBX in mammals), a member of the Polycomb repressive complex 1 (PRC1), targets H3K27me3 [Margueron and Reinberg, 2011], while the double Chromo domains of the Chromo helicase DNA binding protein CHD1 bind to H3K4me3 [Flanagan et al., 2005].

Tudor domains also belong to the Royal superfamily and share some sequence similarity with Chromo domains. The Tudor domain folds into a β-sandwich (β1-β2-β3) flanked by one α-helix. For instance, the Jumonji domain containing protein 2A (JMJD2A) comprises two Tudor domains (double Tudor domain) and was shown to bind to several methylated lysine residues on the histone tails: H4K20me3, H4K20me2, H3K4me3, H3K4me2 and H3K9me3 [Kim et al., 2006; Huang et al., 2006; Lee et al., 2008a]. The protein also harbors JmjN and JmjC domains, which are essential for its lysine demethylase activity. Thus, this is another example of a chromatin-associated protein with reading domains and enzymatic activity.

PWPP domains are Royal domain superfamily members, which bind with their Proline-Tryptophan-Tryptophan-Proline motifs to chromatin. This motif is found in more than 60 eukaryotic proteins and is approximately 135 amino acids in length. The structure of the domain adopts an N-terminal barrel-like five stranded structure with a C-terminal five-helix bundle [Qiu et al., 2002]. For example, the PWPP domain of DNA methyltransferase 3a (Dnmt3a) interacts with the repressing H3K36me3 mark and thereby contributes to the targeting of the methyltransferase to chromatin carrying that mark [Dhayalan et al., 2010].

Some other Royal superfamily members recognize lower methylation states (me1 or me2). The malignant brain tumor (MBT) repeats composed of ~ 70 amino acid residues are transcriptional repressors which are frequently perturbed in hematopoietic malignancies [Taverna et al., 2007]. For example, the human lethal 3 malignant brain tumor repeat-like
protein (L3MBTL1) recognizes H4K20me1, H4K20me2, H1.bK26me1 and H1.bK26me2 [Trojer et al., 2007].

Another important binding module is the plant homeodomain (PHD) finger with the motif Cys$_4$-His-Cys$_3$, which coordinates two zinc ions [Bienz, 2006]. PHD fingers are about 50 amino acids long and were mostly found to target H3K4me3 and/or H3K4me2 like the human inhibitor of growth (ING) family members or BPTF. BPTF is the largest subunit of the nucleosomal remodeling factor (NURF) ATP-dependent chromatin remodeling complex and a transcriptional activator [Mizuguchi et al., 1997].

The cysteine-rich ATRX-DNMT-DNMT3L (ADD) domain is a PHD finger-like domain, which was shown to preferentially bind to unmodified H3K4. These kinds of readers do not have real binding pockets, but rather display binding surfaces, which interact with the unmodified lysine residue through hydrogen bonds [Yun et al., 2011]. The absence of the activating H3K4me3 mark recruits the ADD domain of Dnmt3a to the preferential sides for DNA methylation [Zhang et al., 2010].

Moreover, WD40 repeats also belong to the group of methyllysine binders. This reading domain is characterized by a β-propeller fold creating a central channel, which is capable of accommodating modified lysine residues [Taverna et al., 2007]. For instance, the WD40 repeat comprising protein Extra-embryonic endoderm (EED), which is part of the H3K27-targeting complex PRC2, was shown to preferentially bind to the repressive marks H3K9me3, H3K27me3, H1K26me3 and H4K20me3, but not to the activating marks H3K4me3, H3K36me3 and H3K79me3 (Figure 8C) [Margueron et al., 2009; Xu et al., 2010]. Crystal structures of the EED WD40 repeats in complex with the respective methylated peptides revealed that EED has two small hydrophobic pockets, which only accommodate small amino acids at the -2 and +2 position (Kme3 = 0 position) [Xu et al., 2010]. The three activating marks all have larger amino acid residues in these positions and are therefore excluded from the binding pocket.

All of these described reading domains and different ways of binding partner recognition add greatly to the regulation of the diverse chromatin functions.
2 The Aim of the present study

In the present PhD work, we have developed a tool for the initial screening of putative reading domains interacting with a diverse set of histone tail PTMs in one experiment in competition. We first validated the method with antibodies directed towards histone tail PTMs. Second, we used the same tool for the interaction with reading domains of known substrate specificity further validating the tool. Third, we screened approximately 20 reading domain candidates and proceeded with the most promising candidates for further analysis. Additionally, we studied two enzymes (MLL3 and MLL1) from the same methyltransferase family and tried to characterize differences in the substrate specificity and tried to find different non-histone targets for the two different enzymes.

2.1 Specific goals and achievements of the project

2.1.1 Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays

We used Celluspots peptide arrays containing many modified histone tail peptides in many different combinations for the specificity analysis of 36 commercial antibodies directed towards histone tail PTMs. By this approach we did not only validate the method, but we further introduced Celluspots peptide arrays as a good tool for the quality control of epigenetic antibodies.

2.1.2 Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails

We also applied Celluspots peptide arrays as a tool for the screening of reading domain candidates interacting with histone tail PTMs. For this purpose, seven already characterized reading domains (the HP1β and MPP8 Chromo domains, JMJD2A and 53BP1 Tudor domains, Dnmt3a PWWP domain, Rag2 PHD finger and BRD2 Bromo domain) were cloned, overexpressed and purified, and were tested on the peptide arrays. The results that we obtained with this approach agreed with literature concerning the primary targets of the reading domains. Furthermore, we were also able to obtain additional previously unknown information concerning the influence of secondary modifications for the binding affinity to the primary targets.
2.1.3 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

The PHD finger protein 1 (PHF1) is a nuclear protein that has been reported to be a member of the polycomb group protein (PcG) family. PHF1 is known to interact via its two PHD fingers with the catalytic subunit Enhancer of Zeste 2 (EZH2) of the Polycomb repressive complex 2 (PRC2) in humans. This interaction leads to increase in the tri-methyl activity for H3K27 of EZH2. No role of the PHF1 Tudor domain has been reported so far. In this study, we showed that the PHF1 Tudor domain recognizes H3K27me3 of the testis specific histone variant H3t. Furthermore, we analyzed a possible biological role for this protein.

2.1.4 Similarities and differences in substrate specificity of MLL3 and MLL1 of the mixed lineage leukemia family of methyltransferases

The mixed lineage leukemia (MLL) histone lysine methyltransferase family comprises of five members (MLL1-5). MLL1-4 have been reported to mono-, di- and tri-methylate H3K4 when assembled in multiprotein complexes. MLL1 shows strongest homology to MLL2, while MLL3 is most similar to MLL4 of this family and MLL5 appears to lack methyltransferase activity. In this study, MLL3 was the main analyzed methyltransferase and MLL1 was used as a reference enzyme. We investigated differences in the substrate specificity by employing a radioactive enzymatic assay, which was already well established in our workgroup. In this assay an H3 tail peptide array was methylated by MLL3 and MLL1, respectively, in the presence of radioactively labeled S-adenosyl-L-methionine. This peptide array comprises 420 peptide spots with each spot carrying a single amino acid exchange for the wildtype sequence of the H3 tail against each of the 20 natural amino acids. Incorporated radioactivity for each peptide spot corresponds to the substrate specificity of the enzyme. In the present study, this assay revealed differences in the H3K4 substrate specificity of MLL3 and MLL1. The arginine in position 8 of the H3 tail was much more stringently recognized by MLL3 than by MLL1. With the substrate specificity profile of MLL3, we searched for potential nuclear non-histone targets containing the target sequence using the scansite database (http://scansite.mit.edu/). With this search, potential non-histone targets were selected. These targets were synthesized on cellulose membrane as peptide arrays and were methylated by both enzymes in separate experiments. Several different targets for both enzymes were recognized at the peptide level. These results show first differences of the two enzymes, but further analysis of the substrate specificity at protein level and investigation of a biological role for these different substrates is needed.
3 Results and Discussion

One of the key players in molecular epigenetics is post-translational histone tail modifications which are involved in the regulation of many biological processes. Histone tail PTMs can be either directly involved in structure formation of chromatin like in the case of lysine acetylation or they can serve as binding signals for reading domains, which recognize the histone tail PTMs and mediate the downstream signal. The diversity of histone tail PTMs and the different kinds of reading domains that recognize these modifications has led to an ever growing interest in the identification of interacting pairs: modified histone tails and reading domain. Likewise, enzymes which modify histone tails are important regulators of epigenetic processes. Many of the initially identified histone methyltransferases or acetyltransferases are now also characterized as non-histone modifying enzymes. Some families of methyltransferases have several enzymes with the same specificity for certain residues on the histone tails. Therefore, it is important to understand why the nuclear machinery needs so many enzymes with one and the same substrate.

For the analysis of the specificity of reading domains with histone tail PTMs a very large pool of different substrates is needed. These substrates can be synthesized as separate peptides, which is expensive and each interaction with peptides would be performed as single experiments, though possibly in parallel. A more convenient and less expensive way is therefore the analysis of different substrates in competition in one experiment only. For this purpose we established Celluspots peptide arrays as a tool for the specificity analysis of antibodies and for the screening of reading domains. Antibodies and reading domains, which are already described in literature, were used to establish this technique. Afterwards, approximately 20 reading domain candidates were cloned, overexpressed, purified and tested for interaction with histone tail PTMs on the peptide arrays. One such promising candidate (the Tudor domain of PHF1) was selected for further analysis.

3.1 Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays

Antibodies against modified histone tails are central research reagents in chromatin biology and molecular epigenetics. The specificity of antibodies used in epigenetic research as well as other research areas is an important issue, because many research results are based on the accuracy of these reagents [Egelhofer et al., 2011, Bordeaux el al., 2010]. Moreover, the locus specific investigation of histone tail PTMs in chromatin relies on a single method – the specific interaction of modified histone tails with antibodies [Mendenhall and Bernstein, 2008; Lennartsson and Ekwall, 2009]. In general, there are no standardized methods or universally
accepted guidelines for the quality control of antibodies [Bordeaux et al., 2010]. In particular, for the specificity analysis of histone tail antibodies there are not only standardized methods missing, but especially a cost-efficient method is lacking due to the enormous number of different histone tail PTMs.

In the present study, Celluspots peptide arrays were applied for the specificity analysis of 36 commercial antibodies directed towards modified histone tails from different suppliers. Here, the results are only briefly described, for details refer to the publication of this work [Bock et al., 2011a]. The arrays contained 384 peptides from eight different regions of the N-terminal tails of histones, viz. H3 1-19, 7-26, 16-35 and 26-45, H4 1-19 and 11-30, H2A 1-19 and H2B 1-19, featuring 59 post-translational modifications in many different combinations (Figure 1). The reliability of the method was validated by several controls. Internal duplicates of the array, which ensures reproducible peptide spotting, showed that antibody binding results were highly reproducible. Binding to arrays from independently synthesized peptides was also very similar. Furthermore, the method was validated by reproducing some of the results using purified peptides. Finally, the identity of representative peptides synthesized in parallel and cleaved off from the matrix was confirmed by mass spectrometry.

From all these observations, we conclude that Celluspots peptide arrays are reliable tools for screening of antibody specificity.

Overall, most of the antibodies tested in this study bound well to the PTM, for which they had been raised for. However, with few exceptions, the analysis revealed previously unknown details of the binding properties that are important for the interpretation of experimental results. Some antibodies showed lost or reduced interaction with the predicted peptides in the presence of a second modification, others showed high cross-reactivity with modifications, for which they were not raised. Furthermore, specificity profiles for antibodies directed toward the same modification sometimes were very different. The entire details of this study are described in the publication [Bock et al., 2011a].

**Figure 1:** Design of the peptide array used in this study. The image obtained with an antibody directed towards H3K4me1 is used for illustration.
3.2 Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails

The diversity of different histone PTMs and the various modification sites on the histone tails have led to the proposal of a so-called “histone code” [Jenuwein and Allis, 2001; Strahl and Allis, 2000]. In this hypothesis, the histone PTMs serve as binding platforms, which recruit epigenetic reading domains and thereby mediate most of their biological functions. So far, multiple conserved reading domain families and their binding partners have been identified. However, the biological consequences of histone PTMs are based on the chromatin and cellular context as well [Berger, 2007].

In the present study, we applied the Celluspots peptide arrays (Figure 1) in order to validate them as a tool for the study of putative reading domains and to analyze the impact of secondary modifications on the binding affinity of epigenetic reading domains in respect to their primary specificity. Seven known epigenetic reading domains were tested on the peptide arrays, viz. the HP1β and MPP8 Chromo domains, JMJD2A and 53BP1 Tudor domains, Dnmt3a PWWP domain, Rag2 PHD finger and BRD2 Bromo domain. In general, the binding results agreed with literature data with regard to their primary targets. Nevertheless, in almost all cases we obtained additional new information concerning the influence of secondary modifications surrounding the target modification. Some modifications prompted a stronger binding affinity for a few of the reading domains, but some modifications occluded binding of the reading domains.

We conclude that Celluspots peptide arrays are powerful screening tools for studying the specificity of putative reading domains binding to modified histone peptides. Furthermore, they are valuable tools in analyzing the influence of secondary modifications for the binding behavior of reading domains. This analysis will lead to a better understanding of the biological signaling processes mediated by reading domains.

The results are described in detail in the publication [Bock et al., 2011b].

In collaboration with our workgroup, Celluspots peptide arrays were successfully introduced into the market as MODified™ Histone Peptide Array by Active Motif.
3.3 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

3.3.1 The role of PHF1 in polycomb group silencing, DNA repair and disease

The polycomb group (PcG) proteins, first identified in *Drosophila*, are key regulators of the homeotic genes (*HOX*), but are also important for the regulation of transcription of other genes, X-chromosome inactivation, cell fate transitions, tissue homeostasis, and tumorigenesis [Gieni and Hendzel, 2009; Sparmann and van Lohuizen, 2006]. The transcriptional repression of *HOX* genes mediated by PcG proteins is antagonized by the transcriptional activation of these genes by proteins belonging to the trithorax group (TrX). The absence of either group of proteins results in deregulation of the *HOX* genes with effects ranging from mild to severe developmental defects [Ringrose and Paro, 2004; Schwartz and Pirrotta, 2007]. In humans, there are two different kinds of Polycomb repressive complexes with homology to the *Drosophila* protein complexes: PRC1 and PRC2. The four core components of PRC2 are the H3K27-methyltransferase Enhancer of Zeste 2 (EZH2) and the non-catalytic components Suppressor of Zeste 12 (SUZ12), Extra-embryonic endoderm (EED) and the Retinoblastoma-binding protein 46/48 (RbAp46/48) [Simon and Kingston, 2009]. All of these four components are indispensable for the general function of the complex. SUZ12 is crucial for the minimal activity of PRC2 *in vitro* and for genome-wide H3K27me2 and H3K27me3 *in vivo*, while EED is required for all H3K27 methylation states [Cao and Zhang, 2004; Pasini et al., 2004; Montgomery et al., 2005]. Further regulation is established by alternative translation start sites of EED generating four distinct isoforms of EED, which have different functionality [Kuzmichev et al., 2004]. For example, some PCR2 variants with certain EED isoforms are reported to methylate the linker histone H1 *in vitro* [Kuzmichev et al., 2004; Kuzmichev et al., 2005]. Besides the four core components, there are additional PcG proteins, which further regulate the activity of the complex and are involved in the silencing of different targets.

How exactly the PRC2 complex is recruited to its target genes is still not fully understood. In *Drosophila*, there are so called Polycomb response elements (PREs), DNA sequence elements, which are proposed to directly or indirectly interact with PcG proteins [Simon and Kingston, 2009]. Mammalian PREs were also identified more recently, but the interaction depends on the transcription factor Ying Yang 1 (YY1), which associates with PRC2, but only a small fraction of PRC2 is found at YY1 response elements [Sing et al., 2009; Woo et al., 2010; Ku et al., 2008]. Therefore, it was proposed that other mechanisms are also involved in the targeting of the PRC2 complex like the association with other PcG proteins [Margueron and Reinberg, 2011].
One of these additional PcG proteins is the human PHD finger protein 1 (PHF1), which is a homologue of the Polycomblike protein (PCL) in *Drosophila*. PCL was shown to interact with E(Z) via its two PHD fingers by yeast two-hybrid and *in vitro* binding assays [O'Connell et al., 2001]. In the same study, the association of PCL with other core components of PRC2 was detected by co-immunoprecipitation. The two PHD fingers of PHF1 (Figure 2A) share 34% identity with the PCL PHD fingers and were also shown to interact with EZH2 in yeast two-hybrid experiments [O'Connell et al., 2001]. Thus, this interaction is conserved between flies and humans. More recently two independent studies showed that a fraction of PRC2 complex associates with PHF1 in HeLa cells [Sarma et al., 2008; Cao et al., 2008]. Sarma et al. further studied the impact of PHF1 on HOX gene silencing and found that the knockdown of PHF1 promotes decrease in the H3K27me3 levels accompanied by an increase in H3K27me2 in promoter regions of some HOX genes (*HOXA6*, *HOXA9* and *HOXA11*) resulting in the deregulation of these genes. From these results they concluded that PHF1 stimulates the trimethyl-activity, but does not change substrate specificity of EZH2 [Sarma et al., 2008]. Similarly, Cao et al. showed that PHF1 stimulates the activity of PRC2 *in vitro*. Knockdown of mPcl1, the closest murine homologue of PHF1, in NIH3T3 fibroblasts and a murine testis cell line revealed its important function for the suppression of transcription of certain HOX genes. mPcl1 was first identified in male germ cells [Kawakami et al., 1998] and the comparison of the knockdown of mPcl1 in fibroblasts with a testis cell line showed a stronger impact on the deregulation of HOX genes in the testis cell line demonstrating a tissue specific function of mPcl1 [Cao et al., 2008].

Besides its role in PcG gene silencing, PHF1 is also reported to be recruited to DNA double-strand breaks after irradiation in HeLa cells [Hong et al., 2008]. In the same study PHF1 was shown to interact with Ku70/Ku80 and some other proteins which are involved in DNA damage response pathways like p53, suggesting a role of PHF1 in genome maintenance processes as well. Furthermore, PHF1 deregulation was detected in a rare uterine sarcoma the endometrial stromal sarcoma [Micci et al., 2006; Chiang et al., 2011]. In these kinds of tumors, a specific rearrangement of the PHF1 gene was found with either the juxtaposed with another zinc finger (JAZF1) gene or the enhancer of polycomb (EPC1) gene leading to two zinc finger fusion genes. The rearrangement of those genes brings the PHF1 promoter under the control of the promoter of the other fusion partner in both cases [Micci et al., 2006]. The impact of this kind of PHF1 deregulation is unknown.

### 3.3.2 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

The Tudor domain and the PHD fingers of PHF1 were cloned, overexpressed and purified as Glutathione S-transferase (GST) fusion proteins. Subsequently, they were tested for
interaction with histone tail PTMs on the Celluspots peptide arrays (Figure 2). The two PHD fingers of PHF1 cloned as separate GST fusion protein domains and in combination both did not give rise to any binding signal (data not shown). GST alone also did not bind to the Celluspots peptide array (data not shown). Analysis of the binding specificity of the Tudor domain revealed that the domain bound preferentially to H3K27me3-, H3K36me3- and a few H4K20me3-modified peptides (Figure 2C). Binding to H4K20me3 occurred only in the presence of the secondary modifications H4K12ac-K16ac. The recognition of this particular combination was also observed for several other reading domains tested on the Celluspots peptide arrays in our workgroup (data not shown) and for that reason it is possible that this is an artifact of the array. Therefore, the interaction with this particular peptide modification was not further analyzed. The H3K27me3 peptides also containing S28ph as secondary modification (Figure 2C, indicated by red arrows) were not recognized by the PHF1 Tudor domain, showing that this might be another reading domain, which is regulated by a so-called methyl/phospho-switch. The HP1β Chromo domain was shown to be released from H3K9me3-modified histone tails upon phosphorylation of the adjacent H3S10 during the M-phase of the cell cycle [Fischle et al., 2005].

Figure 2: Analysis of the binding specificity of the PHF1 Tudor domain to modified histone tail peptide arrays. A) Schematic illustration of the PHF1 protein. The protein comprises of one N-terminal Tudor domain (green) and two PHD fingers (red). B) Coomassie stained SDS gel of purified wildtype and variant PHF1 Tudor domain GST fusion proteins. The calculated molecular weights of the GST fusion wildtype protein and variants are about 34 kDa. (MW Marker: Molecular Weight protein Marker) C) PHF1 Tudor domain (c = 0.1 µM) on Celluspots peptide array. Peptide spots are annotated on the left copy of the duplicate. The color code is described on the right side of the image. The red arrows indicate unbound peptide spots carrying one of the target modifications and secondary inhibiting modifications, which are specified on the right side of the image.
It is possible that the PHF1 Tudor domain is regulated in a similar way by phosphorylation of H3S28, but a possible role of this secondary modification was not further analyzed and is therefore only hypothetical. Binding to H3K27me3- and H3K36me3-modified peptides was rather surprising for a PcG protein, since the two modifications are reported to be antagonistic in PRC2-mediated methylation [Yuan et al., 2011] and the amino acids surrounding the modified lysine residues differ greatly from each other (Figure 3A). Therefore, we investigated the possibility of two binding pockets within the Tudor domain of PHF1. For this reason, peptides were synthesized on cellulose membrane by the SPOT method [Frank, 2002] containing analogs of tri-methyllysine at the K27 and K36 position and one unmethylated control peptide (Figure 3B). For this purpose, a protocol for the site-specific installation of methyllysine analogs was employed after peptide synthesis, converting cysteine residues into methyllysine analogs [Simon et al., 2007]. The PHF1 Tudor domain did not bind to the unmethylated lysine peptides (spot no. 1), while the second and the fourth peptide spots (K27me3- and K27me3-K36me3 analogs respectively), were almost equally bound by the reading domain (Figure 3B). Only weak binding was observed for the H3K36me3 analog peptides (spot no. 3). From these results, we concluded that H3K27 is the primary target of the PHF1 Tudor domain and that the binding to H3K36me3 on the Celluspots peptide array might have been an artifact.

Next we analyzed the PHF1 gene expression pattern on the gene investigator website (https://www.genevestigator.com/) and found that PHF1 is expressed mainly in some brain tissues and in the testis (Figure 4). The mammalian testis is known to have more histone variants than most other tissues including a testis specific histone variant H3t. The histone variant H3t differs in the K27 region by one amino acid from the canonical H3.1, which is valine instead of alanine in position 24. Therefore, we were interested in studying a possible H3t specific readout of the PHF1 Tudor domain, since both are highly expressed in testis.

Figure 3: Analysis of the existence of one or two binding pockets within the PHF1 Tudor domain. A) Amino acid sequences of H3K27me3 and H3K36me3 (H3.1) employed on the Celluspots peptide arrays. B) Analysis of K27me3/K36me3 preference of the PHF1 Tudor domain (c = 0.3 µM) to investigate the possibility of two binding pockets within the Tudor domain. The peptide spots 1-4 are annotated on the right side of the image.
In order to analyze an H3t specific interaction, peptide arrays were synthesized containing peptide spots with the H3t sequence and the H3.1 sequence respectively, which were unmethylated control peptide spots and peptide spots containing H3K27me3 analogs (Figure 5A). Analysis of a histone variant specific context revealed that the Tudor domain of PHF1 bound much stronger to H3K27me3 analog peptides in H3t (spot no. 2), than to the corresponding methylated lysine analog in H3.1 (spot no. 4) (Figure 5A). Furthermore, we wanted to determine the amino acid residues, which are important for target recognition within the H3K27 region. Peptide arrays were synthesized containing single amino acid exchanges for each position surrounding the target lysine. Amino acid exchange from valine to serine in position 24 (spot no. 3) completely abolished binding of the reading domain to the H3K27me3 analog, while the exchange from alanine to serine in position 25 (spot no. 4) and arginine to alanine in position 26 (spot no. 5) reduced the binding signal (Figure 5B). Amino acid exchanges in the positions 23 and 28-31 (spot nos. 2 and 6-9) had little effect on the binding specificity of the Tudor domain of PHF1. This result shows that position 24 in the K27 region is important for the recognition of the PHF1 Tudor domain for its target lysine. Furthermore, one can conclude that the Tudor domain reads the amino acid residues 24-27 of the histone variant H3t.
Figure 5: Analysis of a histone variant specific readout of the Tudor domain of PHF1. “Kme3” refers to tri-methylated lysine analog. A) Binding specificity of the PHF1 Tudor domain (c = 20 nM) to the histone variant H3t in comparison to the canonical H3.1. The peptide spots 1-4 are described below the image. B) Amino acid exchanges for the H3K27 sequence of H3t. Peptide spot number 1 contains peptides with the H3t wildtype sequence. Peptide spots 2-9 carry amino acid exchanges indicated in red letters within the sequence and with red arrows. The upper row contains peptides with tri-methyllysine analogs and the lower row contains peptides with unmethylated lysines. The amino acids in positions 24-26 are important for the recognition of H3K27me3 by the PHF1 Tudor domain (c = 2 nM). No interaction was observed of the PHF1 Tudor domain with unmethylated lysine residues.

After identifying the primary histone target, we further wanted to determine the binding pocket of the Tudor domain of PHF1. In the protein data bank (pdb), there is a solution structure available for the PHF1 Tudor domain without peptide (pdb entry 2E5P).

Figure 6: Superposition of the PHF1 Tudor domain (pdb entry 2E5P) (green) and the UHRF1 tandem Tudor domains (pdb entry 3DB3) (red) in complex with an H3K9me3 peptide (orange) using Deep View Swiss PDB viewer 3.7. A) Superposition from two different sides. B) Identified aromatic (W41, Y47 and F65) and acidic (D67) amino acids which are in close proximity of the tri-methylated lysine residue.
This structure was superimposed with the crystal structure of one of the tandem Tudor domains of the E3 ubiquitin-protein ligase UHRF1 in complex with an H3K9me3 peptide (pdb entry 3DB3) in order to identify an aromatic cage within the Tudor domain (Figure 6A). This superposition revealed the three aromatic residues W41, Y47, F65 and one acidic residue D67 in close proximity to the tri-methylated lysine residue, which could be involved in forming an aromatic cage or in the formation of hydrogen bonds between the acidic residue and the tri-methyllysine group (Figure 6B). Subsequently, the four amino acids in the Tudor domain of PHF1 were mutated to alanine and correct mutagenesis for all constructs was verified by sequencing. PHF1 Tudor domain GST fusion variants were overexpressed and purified. All purified variants showed similar purity in comparison to the wildtype protein domain (Figure 2B) and correct protein folding of the variants was further determined by circular dichroism spectroscopy (Figure 7A). All spectra are very similar, indicating that the secondary structure composition and folding of all proteins is alike.

**Figure 7:** Analysis of protein folding and binding specificity of the PHF1 Tudor variants. A) Circular Dichroism spectra of purified wildtype and all variant PHF1 Tudor domain GST fusion proteins. The spectra of the variant proteins were scaled to the wildtype spectrum (by up to 28%) to correct for concentration differences and to allow better comparison. The color code for each spectra is described on the right side of the CD spectra. B) PHF1 Tudor W41A (c = 1 µM) on Celluspots peptide array. C) PHF1 Tudor Y47A (c = 0.5 µM) on Celluspots peptide array. D) PHF1 Tudor F65A (c = 0.5 µM) on Celluspots peptide array. E) PHF1 Tudor D67A (c = 0.5 µM) on Celluspots peptide array. F) PHF1 Tudor W41A (c = 1 µM) on K27me3-K36me3 analogs peptide array.
Subsequently, all Tudor domain variants were tested on the Celluspots peptide arrays (Figure 7B-E). None of the four variants bound to any of the modified histone tail peptides on the array even though the protein concentration was 10-fold higher in the case of the W41A variant and 5-fold higher in the case of the three other variants and the exposition time was longer than in the experiment with the wildtype protein. The dark spots at the edges of the images show the outline of the peptide array and were not visible after shorter exposition times (data not shown). These results were a first indication for the importance of these amino acid residues within the Tudor domain for H3K27me3 recognition. The W41A variant was also tested on the H3K27me3-H3K36me3 analog peptide array previously applied for the investigation of the possibility of two binding pockets within the Tudor domain. The W41A variant lost its binding preference for the H3K27me3 analog, instead the variant protein bound approximately equally to all four peptide spots including the unmethylated peptide (spot no. 1) (Figure 7F). This result further demonstrates the importance of this residue in trimethyllysine target recognition. The initial binding specificity results of the PHF1 Tudor domain and its variants on Celluspots peptide arrays and cellulose membranes were further confirmed by peptide pulldown experiments. Purified biotinylated peptides were immobilized on streptavidin beads and were subsequently incubated with the PHF1 Tudor domain or a GST control. After several washing steps, the bound proteins were eluted from the beads and separated on SDS gel with subsequent immunoblotting.

Figure 8: Analysis of the binding specificity of PHF1 Tudor domain by peptide pulldown and equilibrium peptide binding experiments. A) Immunoblot of peptide pulldown with PHF1 Tudor domain. An anti-GST antibody was used for detection. From left to light: Molecular Weight protein Marker, 5% Input of PHF1 Tudor domain (34 kDa), GST control (27 kDa) with H3K27me3 peptide (H3t), PHF1 Tudor domain with purified biotinylated peptides: H3(1-19), H3K4me3, H3(18-34), H3K27me3 (H3t). B) Equilibrium peptide binding experiments of Tudor wildtype and its variants W41A and Y47A with purified fluorescently labeled peptides. Color code: PHF1 Tudor wildtype with H3K27me3 (H3t) in red, wildtype with H3K27me3 (H3.1) in blue, wildtype with H3K27me0 (H3t) in green, W41A variant with H3K27me3 (H3t) in yellow, and Y47A variant with H3K27me3 (H3t) in pink.
An antibody directed towards GST was used for detection (Figure 8A). GST alone did not pull down the primary target H3K27me3 peptide (H3t) of the Tudor domain, validating the Tudor domain specific interaction with the peptide. Three additional control peptides were also included in this assay. An H3K4me3 peptide, which was not a target identified on the Celluspots peptide array, was selected because of its transcriptional activating function in comparison to the transcriptional repressing function of H3K27me3 and two unmodified peptides for the K4 and K27 regions (H3(1-19) and H3(18-34)). The strongest signal was observed for the Tudor domain's primary target, but a weak signal was also observed for the H3K4me3 peptide and the H3(18-34) peptide (Figure 8A). In the first case, the Tudor domain might still interact with the tri-methylated lysine residue; in the second case, the interaction could depend on the surrounding amino acids of the K27 sequence. In both instances, the signal observed for the primary target was much stronger than for the control peptides. This further supports the initial binding results.

We then determined the dissociation constants of the PHF1 Tudor domain in complex with H3K27me3 (H3t), H3K27me3 (H3.1) and H3K27me0 (H3(18-34)) (H3t) purified fluorescently labeled peptides by fluorescence depolarization (Figure 8B). The dissociation constants of the two Tudor variants W41A and Y47A in complex with H3K27me3 (H3t) were also determined by the same method. The dissociation constants of the Tudor variants F65A and D67A were not analyzed in this case, because the protein concentrations of the two variants were reduced by approximately 5-fold in comparison to the wildtype and the other two variants and for a good comparison similar protein concentrations are needed. The dissociation constant (K_D) of the PHF1 Tudor domain with H3K27me3 (H3t) was determined as 55.6 ± 5.2 µM (Figure 8B and Table 1). Binding to the H3K27me3 (H3.1) and the unmodified H3K27 (H3.t) peptide was approximately 4-fold and >5-fold weaker, respectively. The Tudor variants W41A and Y47A showed a reduced binding affinity to H3K27me3 (H3t) by 5-fold and >5-fold, respectively.

**Table 1:** Dissociation constants of PHF1 Tudor wildtype or variants with H3K27me3 (H3t), H3K27me3 (H3.1) or H3K27me0 (H3(18-34)) (H3t) fluorescently labeled purified peptides determined by fluorescence depolarization.

<table>
<thead>
<tr>
<th>Protein Peptide</th>
<th>K_D value (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHF1 Tudor H3K27me3 (H3t)</td>
<td>55.6 ± 5.2</td>
</tr>
<tr>
<td>PHF1 Tudor H3K27me3 (H3.1)</td>
<td>206 ± 13.4</td>
</tr>
<tr>
<td>PHF1 Tudor H3K27me0 (H3t)</td>
<td>&gt;300</td>
</tr>
<tr>
<td>PHF1 Tudor W41A H3K27me3 (H3t)</td>
<td>286 ± 3</td>
</tr>
<tr>
<td>PHF1 Tudor Y47A H3K27me3 (H3t)</td>
<td>&gt;300</td>
</tr>
</tbody>
</table>

The error margins represent the standard error of the mean (SEM) of independently repeated measurements.
Figure 9: Pulldown of native histones isolated from rat sertoli testis cells (SER-W3) with PHF1 Tudor domain and Tudor variants. A) Coomassie stained SDS gel of isolated native histones with the histone type indicated on the right side of the gel. B) Coomassie stained SDS gel loaded with the pulldown samples. From left to right: Molecular Weight protein Marker, 10% Input of SER-W3 native histones, GST control (27 kDa), PHF1 Tudor (34 kDa), W41A (34 kDa) and Y47A (34 kDa), PHF1 Tudor domain, GST control and histones H3 and H4 positions in the gel are indicated on the right side. C) Analysis of the pulldown experiment. The H3 and H4 bands for all pulldown samples were quantified by using Image J software. PHF1 Tudor domain was normalized to 1 (100%) and all other samples were normalized accordingly to PHF1 Tudor domain.

All experiments showed a strong preference of the PHF1 Tudor domain for the histone variant H3t tri-methylated in position 27. H3t is thought to be exclusively expressed in the mammalian testis [Schenk et al., 2011]. Therefore, the rat sertoli testis cell line SER-W3 was used as a model organism for the following experiments. First, native histones from SER-W3 cells were isolated by acid extraction [Shechter et al., 2007]. These native histones were then used in a pulldown experiment with the wildtype, the variants W41A and Y47A, and GST as control. In this pulldown, Glutathione Sepharose beads were incubated with the GST fusion proteins and unbound protein was removed by washing. Free binding surface of the beads was blocked by incubation with BSA. Blocking was followed by pulldown of the native histones. Weakly bound proteins were removed by washing several times with high salt buffer. The bound proteins were eluted from the beads and loaded on a SDS gel, separated by electrophoresis and the gel was stained with coomassie (Figure 9B). All four GST-tagged proteins were detected with equal amounts in the SDS gel (protein bands at 27 and 34 kDa). No pulldown of native SER-W3 histones was observed with the GST control, while the wildtype Tudor domain pulled down native H3 and H4 histones. H3 and H4 form relatively stable tetramers within the nucleosomes, therefore it is possible that these tetramers are still associated with each other after histone isolation and are therefore pulled down together. These complexes dissociate after boiling in SDS containing loading buffer and appear separated in the coomassie stained SDS gel.

A much weaker pulldown was detected for the Tudor variants W41A and Y47A. The variants pulled only 30-40% of the histones in comparison to the wildtype protein (Figure 9C).
This result shows that these two aromatic residues within the Tudor domain are important for the interaction with native histones as well and that the wildtype domain indeed is capable of interacting not only with synthesized peptides, but also with native histones. Post-translational histone modifications of the pulled H3 histones were not analyzed in this case.

After this first indication for an interaction of the PHF1 Tudor domain with native histones from rat testis cells, we further analyzed a possible role of the aromatic cage of the domain in targeting the protein to H3t containing nucleosomes. For this purpose, the full length protein was cloned as YFP fusion and the four mutations within the Tudor domain were introduced as single mutations by mutagenesis into the full length protein as CFP fusions. All constructs were verified by sequencing. Wildtype YFP fusion protein and variant CFP fusion proteins were transfected into SER-W3 testis cells and mouse NIH3T3 fibroblasts as control cell line. The localization of the proteins in both cell lines was analyzed by confocal fluorescence microscopy (Figure 10). Transfection of both cell lines with the PHF1-YFP/CFP fusion
proteins was rather difficult due to toxicity of the constructs and low transfection rates. No differences were observed for the localization of PHF1 wildtype YFP fusion proteins in comparison to PHF1 variant CFP fusion proteins in NIH3T3 fibroblasts after single transfection (Figure 10A, PHF1-W41A-CFP is shown as example for the localization of PHF1 variants). Wildtype and variants both were localized within the whole cell (nucleus and cytosol).

In contrast, PHF1 wildtype YFP fusion protein was localized in the nucleus of SER-W3 cells, while the variant CFP fusion proteins showed nuclear, but also cytosolic localization (Figure 10B, PHF1-W41A-CFP is shown as example for the localization of PHF1 variants). These initial localization results indicate that the identified aromatic cage within the PHF1 Tudor domain, which we showed to recognize tri-methylated H3K27, is important for proper localization within testis cells.

### 3.3.3 The EED WD40 repeats do not discriminate between H3t and H3.1

EED, one of the core components of the PRC2 complex, is reported to target the complex to H3K27me3 modified histone tails via its WD40 repeats [Margueron et al., 2009]. Therefore, we analyzed the possibility that the WD40 repeats of EED also recognize H3K27me3 of the histone variant H3t, or if this histone variant specific readout is unique for the PcG protein PHF1. Therefore, the EED WD40 repeats were cloned as a GST fusion and the construct was verified by sequencing. Overexpression and purification of the protein was problematic and the protein yield after purification was relatively low. However, the protein concentration was sufficient to test the reading domain for interaction with post-translational histone tail modifications on the Celluspots peptide array (Figure 11A). Binding was observed to H3K27me3-modified peptides, although there was a lot of background binding detected to other modifications as well. The double modified peptide H3K9me3-K14ac was also a preferred substrate of the EED WD40 repeats. Furthermore, just like in the case of the PHF1 Tudor domain, the secondary modification S28ph abolished binding of the EED WD40 repeats to H3K27me3 as well, but a possible role of the inhibitory effect of S28ph for the binding to H3K27me3 was not studied. Except for H3K27me3, none of the other targets were further analyzed.

In order to analyze the possibility of a sequence specific readout of H3K27me3 in a histone variant context, we wanted to determine the dissociation constants of the EED WD40 repeats binding to H3K27me3 (H3t), H3K27me3 (H3.1) and H3K27me0 (H3(18-34)) (H3t) purified fluorescently labeled peptides by fluorescence depolarization (Figure 11C). Normally, in fluorescence depolarization experiments the protein concentration is increased until the signal reaches saturation.
Figure 11: Analysis of the binding specificity of the EED WD40 repeats. A) EED WD40 repeats (c = 10 nM) on Celluspots peptide array. Peptide spots are annotated on the left copy of the duplicate. The color code is described on the right side of the image. The red arrows indicate unbound peptide spots carrying one of the target modifications and secondary inhibiting modifications, which are specified on the right side of the image. B) Coomassie stained SDS gel of purified EED WD40 repeats GST fusion protein (MW = 69 kDa). C) Equilibrium peptide binding experiments of the EED WD 40 repeats and purified fluorescently labeled peptides. Color code: H3K27me3 (H3.1) in red, H3K27me3 (H3t) in blue, H3K27me0 (H3t) in green.

In this case this was not possible due to the low concentration of the purified EED WD40 repeats. Therefore, it was not possible to determine the dissociation constants. However, the binding curves for H3K27me3 (H3.1) and (H3t), respectively, were very similar, suggesting a non-histone variant specific readout. The binding affinity to unmodified H3K27 peptide was reduced, showing that binding was due to the tri-methylated lysine residue (Figure 11C).

The WD40 repeats of EED were reported to read the positions -2 till +2 (Kme3 = 0) of the transcriptionally repressing marks H3K27me3, H3K9me3, H4K20me3 and H1K26me1 [Xu et al, 2010], which means in the case of H3K27 that the domain recognizes the amino acids in position 25-29. In this study, we showed that the PHF1 Tudor domain reads the valine in position 24 in a K27me3 context of the histone variant H3t. The results obtained from equilibrium peptide binding experiments suggest that the EED WD40 repeats indeed do not recognize the position 24 in an H3K27me3 context, which is in concert with the study from Xu et al.
3.3.4 Discussion and Conclusion

In the present study, we analyzed the binding specificity of the PHF1 Tudor domain to post-translational modified histone tails and found that the Tudor domain binds specifically to H3K27me3 of the histone variant H3t. In previous studies, no histone peptide binding by the PHF1 reading domains (Tudor domain and the two PHD finger) was observed [Cao et al., 2008; Sarma et al., 2008]. This could be due to the fact that the interaction of the PHF1 Tudor domain with H3K27me3 modified histone tails in the canonical H3.1 is rather weak and a possible interaction of the Tudor domain of PHF1 with H3t was not analyzed in earlier studies. However, Cao et al. analyzed the function of mPcl1, the closest murine homologue of the human PHF1, in the murine testis, and they also speculated that the addition of other PcG proteins to the core components of PRC2 might have tissue specific roles. Here, we were able to show that PHF1, a protein highly expressed in certain brain tissues and the human testis, recognizes the residues 24-27 in H3K27me3 of the testis specific histone variant H3t in several in vitro experiments, using synthesized peptides on cellulose membranes, peptide pulldown and equilibrium peptide binding. Celluspots peptide arrays were used as a screening tool and prompted these additional experiments, providing another example that peptide arrays are a good screening tool for putative reading domains. We identified the tri-methyllysine binding pocket comprising aromatic residues and one acidic residue within the PHF1 Tudor domain, which are involved in the H3K27me3 recognition, by superposition with another Tudor domain in complex with an H3K9me3 peptide. Mutation of either of those residues resulted in reduced affinity for the target modification as observed in several experiments, for example in the equilibrium peptide binding experiments with the PHF1 Tudor variants W41A and Y47A. Additionally, we tested the interaction of native histones isolated from a rat testis cell line with the PHF1 Tudor domain and found that the Tudor domain is capable of pulling down native histones. The binding affinity of the Tudor domain variants W41A and Y47A for native histones was reduced by 60-70% in comparison to the wildtype Tudor domain, which further supports the initial results with modified peptides. We considered to confirm the preferred interaction of PHF1 Tudor domain with H3t by mass spectrometry, but the analysis of the H3 histone composition in the pulldown samples is rather difficult since it is not clear how much H3t is expressed in the testis cells. H3t is reported to be highly expressed in primary spermatocytes, but other publications mention the expression in other parts of the testis as well [Witt et al., 1996]. Therefore, it is not clear if the amount in the pulldown would be sufficient for mass spectrometry. Antibodies directed towards H3K27me3 were tested to evaluate if the antibodies recognize only H3.1 or H3t or both and we found that the antibodies recognized both sequences equally (data not shown). Therefore, a specific detection of H3t modified histone tails in comparison to H3.1 modified
histone tails is not possible with the currently available antibodies. Since the detection of H3K27me3-modified histone tails in H3t is rather difficult, we plan to perform pulldown experiments of native histones isolated from testis cells in comparison to native histones isolated from a control cell line, for example human embryonic kidney cells (HEK293), in order to show differences in the total H3 level after pulldown.

Even though the affinity of the PHF1 Tudor domain for H3K27me3 is relatively weak ($K_D = 55.6 \pm 5.2 \, \mu M$), it is still almost 2-fold stronger than the reported affinity of EED WD40 repeats for H3K27me3 ($K_D = 108 \pm 17 \, \mu M$, Xu et al., 2010), which was shown to be important for targeting of the PRC2 complex to H3K27me3-modified histone tails [Margueron et al., 2009; Hansen et al., 2008]. Therefore, the interaction of the PHF1 Tudor domain with H3K27me3 could also be important for targeting of the PRC2 complex to H3K27me3-modified histone tails in H3t in the testis.

Furthermore, we tried to determine the biological function for the recognition of H3K27me3 in H3t by PHF1 by performing localization studies in a rat testis cell line in comparison to murine fibroblasts. So far, we were only able to obtain preliminary results due to the toxicity of the constructs and a low transfection rate in these cell lines. Nonetheless, the preliminary results suggested that in the testis cell line the wildtype was localized in the nucleus, while the mutant proteins were localized in the nucleus and in the cytosol as well. In contrast, the wildtype and the variants both were localized in the nucleus and in the cytosol in the control cell line. From these results, one can conclude that the aromatic binding pocket within the PHF1 Tudor domain is important for the targeting to the H3K27me3 in H3t, which is primarily expressed in testis. After mutation of the aromatic or the acidic residue(s) within the binding pocket, the H3K27me3-specific interaction is lost. This interaction depends on H3t and this histone variant is most likely not expressed in the control cell line used in this study. Therefore, identical localization was observed for the wildtype and the variant proteins, because the target was missing. Thus, a biological role of PHF1 in targeting the PRC2 complex to H3K27me3 in H3t in the testis seems probable, but further experiments need to be conducted optimizing the transfection of testis cells with the PHF1 constructs in order to confirm the initial results and this hypothesis. Additional experiments with a human testis cell line are planned and a specific antibody-staining for H3K27me3 to analyze a possible co-localization with this mark as well, but at the moment an H3t specific staining of H3K27me3 is not possible, since there is currently no antibody available for this histone variant.

The function of this possible targeting of PRC2 by PHF1 to H3K27me3-modified histone tails in H3t in the human testis is difficult to determine, since there is only very little known about H3t in general and its special function in testis. Knowing the function of this histone variant will also shed light on the role of PHF1 in the human testis. During spermatogenesis histones are exchanged for protamines, which are capable of packing the DNA tightly [Balhorn et al.,
and only about 4% of the haploid sperm genome are reported to retain nucleosomes [Hammoud et al., 2009]. In this process of exchanging nucleosomes with protamines, it seems probable that the more stable canonical histones are first exchanged with the less stable H3t-H4 tetramers [Tachiwana et al., 2010] and that this exchange is followed by the exchange of H3t containing nucleosomes with protamines. Transcriptional regulation throughout this whole process is still essential. Therefore, it seems likely that a PcG protein, which is not a core component of PRC2, is involved in the targeting of the complex to H3K27me3-modified histone tails in H3t. This hypothesis needs to be further analyzed by additional experiments. One of these experiments would be the pulldown of PRC2 core members of human testis cells and to analyze the complex composition. With this experiment we want to show that PHF1 is associated with the core complex, but possibly we could also identify other PcG or non-PcG proteins that assemble with these complexes as well. In the past, PHF1 association with PRC2 was shown in HeLa cells [Sarma et al., 2008; Cao et al., 2008], but so far no localization of PHF1 with PRC2 is reported for testis cells.
3.4 Similarities and differences in substrate specificity of MLL3 and MLL1 of the mixed lineage leukemia family of methyltransferases

3.4.1 The role of MLL family members in HOX gene activation, nuclear-receptor mediated gene activation and in disease

In mammals, there are at least ten different H3K4 methyltransferases. These include, among others, the SET1-like methyltransferases SET1A, SET1B and the mixed lineage leukemia (MLL) family members MLL1-4 [Ruthenberg et al., 2007]. The fifth MLL family member MLL5 is reported to lack methyltransferase activity [Madan et al., 2009; Sebastian et al., 2009]. MLL1-4 are reported to mono-, di- and tri-methylate H3K4 and are implicated in the regulation of many biological processes including transcriptional activation of HOX but also other genes, cell cycle regulation, and development [Ansari and Mandal, 2010]. The originally in yeast identified ySet1-containing complex COMPASS is the founding member of a family of evolutionary conserved complexes that are collectively named COMPASS-like complexes.

The name for the mixed lineage leukemia (MLL) proteins stems from the finding that the MLL1 gene is involved in chromosomal rearrangements, which lead to a large number of different fusion genes. Expressed as part of different chimeric proteins, MLL1 is involved in acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) [Marschalek, 2010]. Although the MLL family comprises of five members, only MLL1 has been implicated in human leukemias [Liu et al., 2009].

The MLL family members are large proteins. For example, MLL1 comprises 3969 amino acids and MLL3 has 4911 amino acids, respectively (Figure 12A). MLL proteins harbor many different kinds of functional domains, including DNA-binding domains like AT-hooks and CXXC or reading domains like PHD fingers and Bromo domains and of course the catalytic activity, the SET domains as well as pre- and post-SET domains (Figure 12A). Some of these domains are involved in substrate recognition like the MLL1 CXXC domain [Cosgrove and Patel, 2010].

MLL1 is the closest homolog of MLL2 and respectively MLL3 is the closest homolog of MLL4. In this respect, the two subgroups constitute two different kinds of MLL multiprotein complexes (Figure 12B). Both COMPASS-like complexes contain the core components WD repeat containing protein 5 (WDR5), Retinoblastoma binding protein 5 (RBBP5), achaete-scute homologue 2 (ASH2) and dpy-30-like protein (hDPY30) (Figure 12B, shown in green), but also additional non-overlapping proteins [Dou et al., 2006; Steward et al., 2006].
Figure 12: Schematic illustration of MLL family members and multiprotein complexes. A) Schematic architecture of MLL1 and MLL3 proteins. Indicated are some DNA-binding domains, reading domains and the SET domains. B) Schematic illustration of COMPASS-like complexes. MLL proteins are indicated in red. Core components of both complexes are colored in green. MLL1/2 specific members are colored in yellow and MLL3/4 specific members are indicated in blue color [modified from Smith et al., 2011].

The MLL1/2-COMPASS-like complexes contain additionally the tumor suppressor menin, which was shown to be involved in the recruitment of the complex to HOX and other gene loci, and the host cell factor 1 (HCF1) (Figure 12B, shown in yellow) [Yokoyama et al, 2004]. The MLL3/4-COMPASS-like complexes, also called ASCOM (ASC-2 complex), are composed of the core components and the activating signal cointegrator-2 (ASC-2), Pax2 transactivation domain-interacting protein (PTIP), PTIP-associated 1 (PA-1) and the H3K27 demethylase ubiquitously transcribed TPR protein transcribed on the X-chromosome (UTX) (Figure 12B, shown in blue) [Goo et al., 2003; Lee et al., 2006].

The Mll1-complex, the murine ortholog of the human MLL1-complex, is reported to be required for the tri-methylation of H3K4 of <5% of promoters carrying this modification in mouse embryonic fibroblasts (MEFs) [Wang et al., 2009]. In concert with this finding is the observation that the majority of H3K4me3 in mammalian cells is mediated by SET1A and SET1B [Wu et al., 2008]. Many of these Mll1 dependent H3K4me3-modified promoters were determined to be promoters of HOX genes [Wang et al., 2009]. Also H3K4me3 in the HOX genes promoters was shown to also depend on menin, which is part of MLL1- and MLL2-complexes.

In contrast to the MLL1/2-complexes, the MLL3/4-containing ASCOM complexes are reported to be implicated in the regulation of only a few HOX genes, but rather play important roles in nuclear receptor-mediated gene activation [Ansari and Mandal, 2010]. ASC-2 is a coactivator of many nuclear receptors and other transcription factors and plays a major role...
in ASCOM complexes in transcriptional activation [Mahajan and Samuels, 2005; Ansari and Mandal, 2010]. For example, MLL3 and MLL4 are involved in the tri-methylation of H3K4 in the promoter region of RAR-2, a gene which is activated in a retinoic acid-dependent manner. This transcriptional activation is mediated by ASC-2 [Lee et al., 2008b; Lee et al., 2006]. Moreover, MLL3- and MLL4-ASCOM complexes are reported to act as coactivators of the tumor suppressor gene p53. MLL3/4 are required for tri-methylation of H3K4 and thereby the activation of p53-target genes in response to the DNA damaging agent doxorubicin [Lee et al., 2009].

Besides the differing target genes and the distinct multiprotein complexes, some targets are also overlapping for the different MLL proteins. For instance, MLL1 and MLL3 were both shown to coordinate with estrogen receptors and thereby to be involved in estrogen-mediated HOXB9 expression [Ansari et al., 2011].

3.4.2 Similarities and differences in substrate specificity of MLL3 and MLL1

MLL3 and MLL1 both methylate H3K4 within target gene promoters, although they are members of different COMPASS-like complexes. Recruitment to different target genes is partially established through the recruitment by menin in the case of MLL1 or ASC-2 in the case of MLL3 or other members of the two different types of multiprotein complexes.

![Figure 13: MLL3 and MLL1 SET domains. A) Sequence alignment of the MLL3 and MLL1 SET domains including the pre- and post-SET domains obtained by ClustalW multiple alignment from BioEdit. Identical amino acids (46%) are indicated by the same highlighted colors. B) Coomassie stained SDS gels of purified GST fused MLL3 SET domain (Molecular Weight = 47.5 kDa) and MLL1 SET domain (Molecular Weight = 45 kDa). (Cloning, overexpression and purification including the coomassie stained SDS gel from MLL1 SET domain were carried out by Qazi Raafiq.) C) Methylation of a peptide array containing H3, H4, H2A and H2B tail peptides (peptide spots 1-10 are annotated on the right side of the image) by MLL3 SET domain. The methylated peptide spot number 1 corresponds to the H3 1-20 sequence.](image)
The two methyltransferases share 46% sequence identity within their SET domains, including pre- and post-SET domains (Figure 13A). Although MLL3 and MLL1 display strong homology within their SET domains, we were interested in studying differences in the substrate specificity, which are not dependent on other complex members, but rather emerge from the differences within the two SET domains itselfs. Therefore, MLL3 and MLL1 SET domains including pre- and post-SET domains were cloned, overexpressed and purified as GST fusion proteins (in the case of MLL1 these steps were carried out by Qazi Raafiq) (Figure 13B). In order to evaluate the methyltransferase activity, both enzymes were tested in the presence of the radioactively labeled cofactor S-adenosyl-L-methionine on histone tail peptide SPOT arrays including different H3, H4, H2A and H2B peptides synthesized on cellulose membrane [Frank, 2002]. MLL3 and MLL1, both methylated the H3 peptide containing the sequence 1-20 specifically (Figure 13C, the result is shown for MLL3 only). This result agrees with the reported methylation target H3K4 for both enzymes.

Next, we analyzed the substrate specificity for H3K4 for both enzymes by methylating H3 tail peptide SPOT arrays in the presence of the radioactively labeled cofactor S-adenosyl-L-methionine. In total, these arrays comprised 420 peptide spots with each spot carrying a single amino acid exchange for the wildtype sequence of the H3 tail against each of the 20 natural amino acids. Both enzymes methylated the target lysine in position 4 and methylation was abolished by exchange for any other amino acid except lysine in this position (Figure 14; in the case of MLL1 this experiment was conducted by Qazi Raafiq). Important for the target recognition were arginine in position 2 and 8, threonine in position 3 and 6, and glutamine in position 5 for both enzymes. For all of these amino acids, it was possible to exchange them against some of the 20 natural amino acids, but not against all of them. MLL3 recognized the arginine in position 8 much more stringently than MLL1 (Figure 14, indicated by the red ovals), because with MLL3 the radioactive signal observed for the peptide spot containing arginine 8 was by far the strongest in that column. Weaker signals were obtained for the exchange of arginine to asparagine, cysteine, glutamine, lysine, methionine, phenylalanine and tyrosine, and exchanges against any other amino acid either abolished or greatly reduced MLL3's methylation activity (Figure 14A). In contrast, MLL1 accepted almost any amino acid exchange in position 8 except for proline, serine, threonine and tryptophan (Figure 14B). The specificity profiles of MLL3 and MLL1 both were reproducible, with results obtained from at least two independent experiments each. We derived a specificity profile containing all amino acids which were accepted by MLL3 in the positions 2-8 within the H3 tail sequence. Since arginine in position 2 and 8 were by far the most preferred amino acids over any other amino acid, we selected arginine as the only possibility in these positions. The target lysine is designated as position 0 in the specificity profile. Amino acids preceding the target lysine are in the -1 and -2 position respectively.
Figure 14: Specificity profiles of MLL3 and MLL1 for the H3 tail amino acid sequence. The wildtype (WT) H3 (1-15) sequences are indicated above the specificity profiles with the target lysine (H3K4) highlighted in red. The amino acid exchanges are specified on the left side of the specificity profiles. Arginine in position 8 is indicated by a red oval in both specificity profiles. A) Specificity profile of MLL3. B) Specificity profile of MLL1 (carried out by Qazi Raafiq).

Amino acids following the target lysine are in the +1 to +4 position (arginine 8 corresponds to the +4 position). With the results from Figure 14A, we derived the following specificity profile: \([R]_2[ARCILMFTYV]_1[K]_0[NRHKMF]_1[ACILSTV]_2[A-Z]_3[R]_4. A-Z in the +3 position means that all 20 natural amino acids are allowed in that position. With this specificity profile, we searched the scansite database (http://scansite.mit.edu/) for possible nuclear non-histone targets for MLL3 and obtained 42 potential protein targets. The sequences for each candidate were synthesized as 15 amino acid long peptides with the target lysine in the middle of the sequences, unless the target lysine was at the N-terminus, which was the case for the H3 tail. H3 (1-15) and H3K4A (1-15) were included as positive and negative controls, respectively, at the beginning and at the end of the non-histone peptide arrays (Figure 15A-B, peptide spot no. 1A and 3C, 2A and 4C). The potential non-histone targets were methylated by MLL3 and MLL1 using radioactively labeled S-adenosyl-L-methionine as cofactor in separate experiments. The target specificity for both enzymes was analyzed manually and the signal intensity for each methylated peptide spot was determined by the programme Phoretix™ Array and compared with the methylation activity towards the H3K4 positive control (Figure 15C). The methylation activity of MLL3 and MLL1 respectively for H3K4 was normalized to 1 (100%) and the methylation activity for all non-histone targets was normalized with the H3K4 activity.
Figure 15: Analysis of the methylation of putative non-histone targets by MLL3 and MLL1. A) Methylation of potential non-histone targets by MLL3. B) Methylation of potential non-histone targets by MLL1 (the enzyme used in this experiment was provided by Qazi Raafiq). C) Bar diagram of the methylation of putative non-histone targets by MLL3 (red) and MLL1 (blue). Methylation activities were subtracted by the negative controls and were normalized to the corresponding H3K4 positive controls.

MLL3 methylated the H3K4 positive control peptides preferentially over all non-histone target peptides (Figure 15A and 15C). In contrast, MLL1 methylated the H3K4 positive controls, but also some of the putative non-histone targets with similar activity (Figure 15B-C). Non-histone targets, which were methylated by both enzymes, are the Zinc finger protein 862 (peptide spot no. 20B), the Signal transducer and activator of transcription 5A and 5B (STAT5; peptide spot no. 11B), the protein MCM10 homolog (peptide spot no. 16A), the homeobox protein Mohawk (peptide spot no. 19A) and the YY1-associated factor 2 (peptide spot no. 17B). Out of these non-histone targets, the Zinc finger protein 862 was the most preferred non-histone target of both MLL3 and MLL1. MLL1 methylated this peptide sequence with slightly higher activity and MLL3 methylated the same target with >60% in comparison to the methylation activity for H3K4.

Individual non-histone targets, which were methylated by MLL3 only, are the Chromodomain-helicase-DNA-binding protein 3 (CHD3; peptide spot no. 3A), Neurogenin-2 (peptide spot no. 3B) and the protein ELYS (peptide spot no. 6A) all with a relative activity between 30% and 20%.
MLL1 also methylated some putative non-histone targets, which were not methylated by MLL3. These proteins are the ZZ-type zinc finger-containing protein 3 (peptide spot no. 2C), the histone lysine methyltransferase MLL4 (peptide spot no. 20A), the Nuclear receptor subfamily 2 group F member 6 (peptide spot no. 7B), the Hepatocyte nuclear factor 4-alpha (peptide spot no. 10A), the Prolow-density lipoprotein receptor-related protein 1 (peptide spot no. 15A) and the 80 kDa MCM3-associated protein (peptide spot no. 17A). Out of these proteins, the ZZ-type zinc finger-containing protein 3, MLL4 and the Nuclear receptor subfamily 2 group F member 6 were all methylated by MLL1 with equal activity relative to H3K4.

After the analysis of the methyltransferase activity of MLL3 and MLL1 towards the putative non-histone targets, we further compared the ranked amino acids within the non-histone targets with the respective specificity profiles of the two enzymes in order to explain the different specificities towards the non-histone targets. However, this analysis did not reveal any conclusive explanation for the similarities and differences in the substrate specificities.

Table 2: Compilation of putative non-histone targets, which were methylated either by MLL3 or by MLL1 or by both enzymes. The number-letter code for the peptide number corresponds to the labeling of Figure 15A-B.

<table>
<thead>
<tr>
<th>Peptide No.</th>
<th>Protein</th>
<th>Peptide Sequence</th>
<th>Methylated by MLL3</th>
<th>Methylated by MLL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>Histone 3 (1-15)</td>
<td>ARTKQTARKSTGGKA</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2A</td>
<td>Histone 3 (1-15) K4A</td>
<td>ARTAQTARKSTGGKA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3A</td>
<td>Chromodomain-helicase DNA-binding protein 3</td>
<td>DGPRVR1KLKRGRPG</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>6A</td>
<td>Protein ELYS</td>
<td>DKQLR1KHKVRVRGR</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>10A</td>
<td>Hepatocyte nuclear factor 4-alpha</td>
<td>GRWRL1KQFRAGMK</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>15A</td>
<td>Prolow-density lipoprotein receptor-related protein 1</td>
<td>CNSRCKKTFRQCSN</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>16A</td>
<td>Protein MCM10 homolog</td>
<td>PALPRTR1VRATPKA</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17A</td>
<td>80 kDa MCM3-associated protein</td>
<td>CLGERLK1HLERLRS</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>19A</td>
<td>Homeobox protein Mohawk</td>
<td>NARRRL1KNTVQPD1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20A</td>
<td>Histone lysine methyltransferase MLL4</td>
<td>VQGPR1KHCHRHAAV</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3B</td>
<td>Neurogenin-2</td>
<td>ETVQRR1KTRRLKAN</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>7B</td>
<td>Nuclear receptor subfamily 2 group F member 6</td>
<td>COYCL1KKCFRVSQMR</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>11B</td>
<td>Signal transducer and activator of transcription 5A / 5B</td>
<td>MSLKR1KRADRRGAE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17B</td>
<td>YY1-associated factor 2</td>
<td>KTRPR1KNVDRSSAQ</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20B</td>
<td>Zinc finger protein 362</td>
<td>DGPR1KCVRPRPSI</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2C</td>
<td>ZZ-type zinc finger-containing protein 3</td>
<td>PVLKR1KCLRSEAP</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3C</td>
<td>Histone 3 (1-15)</td>
<td>ARTKQTARKSTGGKA</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4C</td>
<td>Histone 3 (1-15) K4A</td>
<td>ARTAQTARKSTGGKA</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
3.4.3 Discussion and Conclusion

In the present study, we showed that the MLL1 and MLL3 methyltransferases with highly homologous SET domains including pre- and post-SET domains are still divergent in their substrate recognition. Therefore, we conclude that target specificity of the methyltransferases MLL3 and MLL1 can be regulated on two different levels. On the one side they are recruited by other COMPASS-like complex members to their target genes, but on the other side the SET domains of the enzymes also discriminate the target sequences and are thereby directly involved in target recognition.

Analyzing the H3K4 target specificity by employing amino acid exchanges of the wildtype H3 tail sequence, we found that MLL3 and MLL1 prefer similar amino acids in the H3 tail sequence after amino acid exchanges in the positions -2 to +3 in the H3 tail sequence. However, the arginine in the +4 position of the H3 tail sequence was much more stringently recognized by the MLL3 SET domain than by the MLL1 SET domain. We derived a specificity profile for the -2 to +4 positions, which was used for the subsequent search for putative non-histone targets in the scansite database (http://scansite.mit.edu/). With this specificity profile we obtained 42 potential non-histone targets. MLL3 methylated eight out of these 42 candidates, although with less activity than H3K4. In contrast, MLL1 methylated eleven non-histone targets and out of these methylated four with similar activity as H3K4 and another four with relatively high activity. Some of these recognized and methylated non-histone targets were targets for both enzymes. However, some identified non-histone targets were individual targets of MLL3 and MLL1, respectively. At present the similarities and differences in the substrate specificity of both enzymes cannot be explained by the H3 tail specificity profiles.

Next we plan to derive a specificity profile for MLL1 and then search for putative non-histone targets with this specificity profile in the scansite database. These potential new non-histone targets would then be synthesized on cellulose membranes and methylated by both enzymes. This additional search for putative non-histone targets could reveal new insights of substrate recognition by the two SET domains.

The next steps after the analysis of the potential non-histone targets for MLL3 and MLL1 are cloning, overexpression and purification of these identified targets, followed by the methylation of these protein targets. We were able to identify several targets at the peptide level, but on peptides the potential target sequence is easily accessible. The methylated peptide sequences of the non-histone targets could be folded towards the inside of the proteins and therefore could be unaccessible for methylation by the enzymes. Once, non-histone targets methylated at the protein level are identified, the next step would be the
analysis of a possible biological role of this methylation. Therefore, all of the identified targets in this study are still only putative non-histone targets of MLL3 and MLL1, respectively. For the preferred non-histone target of both enzymes MLL3 and MLL1, the Zinc finger protein 862, there is not much known except that it may be involved in transcriptional regulation. For this protein, the analysis of a biological role of its methylation would therefore mean to also study the biological role of this protein itself. Out of the other weaker methylated non-histone targets, there are several interesting proteins, where a functional connection between these proteins and MLL3 or MLL1 seems possible. In this respect, the Chromodomain-helicase-DNA-binding protein 3 would be an interesting target for MLL3, since the methylation target sequence lies within an N-terminal PHD finger of the protein. Therefore, this methylation could have a role in regulating the transcriptional activity of CHD3. CHD3 is part of an ATP-dependent nucleosome remodeling complex and functions as transcriptional corepressor [Kunert and Brehm, 2009]. Another appealing non-histone target for MLL3 and MLL1 is the Signal transducer and activator of transcription 5A / 5B, which is involved in the growth hormone stimulation of production of insulin-like growth factor [Rosenfeld and Hwa, 2009]. Mainly MLL3, but also MLL1 were shown to be involved in hormone-stimulated gene expression [Ansari and Mandal, 2010]. Out of the individual MLL1 non-histone targets, the methyltransferase MLL4 from the same protein family would also be an interesting non-histone target for further studies.
4 Materials and Methods

4.1 Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays

Celluspots peptide arrays were provided by Intavis AG (Köln, Germany). They are now commercially available as MODified\textsuperscript{TM} Histone Peptide Array by Active Motif (Cat. No. 13001). The peptide arrays were prepared by the Celluspots method as described [Winkler et al., 2009].

Antibody binding experiments were performed as described in the Materials and Methods section of the publication [Bock et al., 2011a], attached as Appendix Chapter 1 to this thesis.

4.2 Application of Celluspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails

Celluspots peptide arrays were provided by Intavis AG (Köln, Germany). They are now commercially available as MODified\textsuperscript{TM} Histone Peptide Array by Active Motif (Cat. No. 13001). The peptide arrays were prepared by the Celluspots method as described [Winkler et al., 2009].

Cloning, expression and purification of the reading domains and binding of the protein domains to peptide arrays were performed as described in the Methods section of the publication [Bock et al., 2011b], attached as Appendix Chapter 2 to this thesis.

Equilibrium peptide binding experiments of the MPP8 Chromo domain with purified fluorescently labeled peptides (H3K9me3 and H3K9me3-S10ph, both containing the H3 tail amino acids 1-19), purchased from Intavis AG (Köln, Germany), were performed as described [Dhayalan et al., 2011b].

4.3 The PHF1 Tudor domain binds to H3K27me3 of the histone variant H3t

4.3.1 Cloning, site directed mutagenesis, expression and purification

The sequences encoding the Tudor domain of human PHD finger protein 1 (PHF1) (residues 29-86; NCBI accession number NP_077084.1), the two PHD fingers of human PHF1 (residues 87-240) and the WD40 repeats of human Extra-embryonic endoderm (EED) (residues 77-441; NCBI accession number NP_003788.2) were amplified from cDNA derived from HEK293 cells and cloned as GST fusion proteins into the pGEX-6P-2 vector (GE Healthcare) using BamHI/XhoI restriction sites. The W41A, Y47A, F65A and D67A mutations
of the Tudor domain of PHF1 were introduced by using a PCR-megaprimer mutagenesis method as previously described [Jeltsch and Lanio, 2002]. Mutagenesis was confirmed by restriction marker site analysis and by DNA sequencing. Human full length PHF1 (residues 1-567) was amplified from PHF1 cloned in pFlag-CMV4 vector, obtained from Prof. Dr. Danny Reinberg (New York University School of Medicine, New York, USA), and subcloned in the mammalian pEYFP-C1 and pECFP-C1 expression vectors (Clontech) using XhoI/BamHI restriction sites. The PHF1 full length sequence in pECFP-C1 vector (Clontech) was used as template for introducing the single mutations W41A, Y47A, F65A and D67A by using a PCR-megaprimer mutagenesis method as previously described [Jeltsch and Lanio, 2002]. Mutagenesis was confirmed by restriction marker site analysis and by DNA sequencing.

The corresponding pGEX-6P-2 plasmids were transformed into E. coli BL21 cells (Novagen) and grown for overexpression in Luria Bertani medium at 37 °C to OD_{600} ~ 0.6. The cell cultures were then shifted to 22 °C for 15 min and induced overnight with 1 mM isopropyl β-D-thiogalactoside. The cells were collected and resuspended in 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 cell pellets were separated from the supernatant for 1 hour and 20 min at 20000 rpm. The supernatants were passed through Glutathione Sepharose 4B resin (Amersham Biosciences) for purification of the proteins and washed with the same buffer used for resuspension of the cells. The bound proteins were eluted with similar buffer containing 40 mM reduced glutathione and dialyzed in 20 mM HEPES pH 7.5, 0.2 M KCl, 0.2 mM DTT, 1 mM EDTA and 10% glycerol for 2-4 hours and then overnight in 20 mM HEPES pH 7.5, 0.2 M KCl, 0.2 mM DTT, 1 mM EDTA and 60% glycerol.

4.3.2 Synthesis of Celluspots peptide arrays and peptide SPOT arrays

Celluspots peptide arrays were provided by Intavis AG (Köln, Germany). They are now commercially available as MODified™ Histone Peptide Array by Active Motif (Cat. No. 13001). The peptide arrays were prepared by the Celluspots method as described [Winkler et al., 2009].

Peptide arrays were synthesized using the SPOT synthesis method [Frank, 2002] by Dr. Srikanth Kudithipudi. Each spot had diameters of 2 mm and contained approximately 9 nmol of peptide (Autospot Reference Handbook, Intavis AG). Successful synthesis of each peptide was confirmed by bromphenol blue staining of the membranes.
4.3.3 Conversion of cysteine into tri-methyllysine analogs

For the site specific installation of methyllysine analogs after peptide synthesis of the peptide SPOT arrays a method was applied converting a cysteine into a methyllysine analog as described [Simon et al., 2007].

4.3.4 Interaction of GST fusion protein domains with post-translational histone tail modifications on peptide arrays

Binding experiments of the GST fusion protein domains with Celluspots peptide arrays and with the peptide SPOT arrays containing the tri-methyllysine analogs were both carried out under the same conditions as described [Bock et al., 2011b]. Protein concentrations were applied on Celluspots peptide arrays as follows: PHF1 Tudor domain 0.1 µM, PHF1 Tudor domain W41A 1 µM, PHF1 Tudor domain variants Y47A, F65A and D67A 0.5 µM and EED WD40 repeats 10 nM. Protein concentrations of the PHF1 Tudor domain were applied on peptide SPOT arrays containing the tri-methyllysine analogs as follows: on H3K27me3/H3K36me3 membrane 0.3 µM and Tudor domain variant W41A 1 µM, on H3t/H3.1 specific readout membrane 20 nM and on H3K27me3 H3t amino acid exchanges membrane 2 nM.

4.3.5 Circular dichroism spectroscopy

Circular dichroism experiments were carried out at room temperature using 10 µM of the purified wildtype and all mutant PHF1 Tudor domain GST fusion proteins in a Jasco J-810 spectropolarimeter with a 0.1 mm cuvette in buffer containing 20 mM HEPES pH 7.5, 200 mM KCl, 6% glycerol, 0.2 mM DTT and 1 mM EDTA. The spectra of the variants were scaled to the wildtype spectrum (by up to 28%) to correct for concentration differences and allow better comparison.

4.3.6 Peptide pulldown experiments

For each pulldown sample 10 µl streptavidin beads (New England Biolabs) were washed with buffer containing 20 mM HEPES pH 7.5, 100 mM KCl and 10% glycerol and incubated with 20 µg of purified biotinylated peptides (H3K27me3 and H3K27unmod, both amino acids 18-34 with the sequence from H3t, H3K4me3 and H3K4unmod, both amino acids 1-19), purchased from Intavis AG (Köln, Germany), in the same buffer under rotation overnight at 4 ºC. Unbound biotinylated peptides were removed by washing three times with the same
buffer. 25 µg of PHF1 Tudor domain GST fusion protein or GST control, respectively, were added to the beads using the same buffer and were incubated under rotation for 3 h at 4 °C. Unbound PHF1 Tudor domain or GST control were removed by washing five times with similar buffer containing 300 mM KCl followed by one wash with similar buffer containing 100 mM KCl. Biotinylated peptides with bound protein domains were eluted from the beads with 30 µl 2x SDS-PAGE loading dye for 10 min at 100 °C and spun down. The supernatants of all pulldown samples and 5% input of PHF1 Tudor domain GST fusion protein were loaded on 16% SDS gel and separated by electrophoresis. The proteins were transferred onto nitrocellulose membrane (Whatman) by Western blotting. The nitrocellulose membrane with the transferred proteins was blocked by incubation in TTBS buffer (10 mM Tris/HCl pH 7.5, 0.05% Tween-20 and 150 mM NaCl) containing 5% non-fat dried milk at 4°C overnight, then washed three times with TTBS buffer, incubated with goat anti-GST antibody (GE Healthcare #27-4577-01) 1:5000 dilution in TTBS buffer containing 1% nonfat dried milk for 1 h at room temperature. Then, the membrane was washed three times with TTBS and incubated with horseradish peroxidase conjugated anti-goat antibody (Invitrogen #81-1620) 1:12000 in TTBS containing 1% non-fat dried milk for 1 h at room temperature. Finally, the membrane was washed four times with TTBS and submerged in ECL developing solution (Thermo Fisher Scientific) and images were captured on HyperfilmTM high performance autoradiography film (GE Healthcare) with exposure times of <1 min. The films were developed using AGFA Curix 60 developing machine.

4.3.7 Equilibrium peptide binding experiments

Peptide binding of the PHF1 Tudor domain wildtype and W41A and Y47A variants, as well as EED WD40 repeats was analyzed by fluorescence depolarization using a Varian Carry Eclipse fluorescence spectrophotometer as described [Dhayalan et al., 2011b]. Purified FITC-coupled peptides (H3K27me3 and H3K27unmod, both amino acids 18-34 with the sequence from H3t, and H3K27me3, amino acids 18-34 with the sequence from H3.1) were purchased from Intavis AG (Köln, Germany).

The data were fitted to a binary-binding equilibrium including a variable baseline (BL) and effect factor (F) to determine the equilibrium-binding constant using the Microsoft Excel Solver module.

\[
\text{Signal} = BL + F \times \text{Fraction bound}
\]

\[
\text{Fraction bound} = \frac{\frac{p}{2} - \sqrt{M}}{c_{\text{peptide}}}
\]

\[
p = \left( c_{\text{peptide}} + c_{\text{protein}} + \frac{1}{K_{\text{Ass}}} \right)
\]
4.3.8 Pulldown experiments with native histones

Native histones were isolated from the SER-W3 cell line using the acid extraction method as described [Shechter et al., 2007]. For the GST pulldown with native histones, 20 µl of Gluthathione Sepharose 4B resin (Amersham Biosciences) was washed once with histone interaction buffer (25 mM Tris/HCl pH 8.0, 100 mM KCl, 5 mM MgCl₂, 200 µM PMSF, 10% glycerol and 0.1% NP-40 equivalent) and was incubated in the same buffer with 7 µg of wildtype PHF1 Tudor domain GST fusion protein or W41A, Y47A variant Tudor domain GST fusion proteins or GST control under rotation for 1.5 h at 4 °C. The beads were washed twice with histone interaction buffer and blocked with histone interaction buffer containing 5% BSA under rotation for 3 h at 4 °C. Then the beads were washed once with histone interaction buffer and incubated with 11 µg of native histones in histone interaction buffer containing 1% BSA under rotation for 3 h at 4 °C. Finally, the beads were washed five times with wash buffer (25 mM Tris/HCl pH 8.0, 300 mM KCl, 5 mM MgCl₂, 200 µM PMSF, 10% glycerol and 0.1% NP-40 equivalent) and one time with similar buffer containing 100 mM KCl. The bound proteins were eluted from the beads with 30 µl 2x SDS-PAGE loading dye for 10 min at 100 °C and spun down. The supernatants of all pulldown samples and 10% input of native histones were loaded on 16% SDS gel and separated by electrophoresis. Subsequently, the proteins in the SDS gel were stained with Coomassie Brilliant Blue.

4.3.9 Cell culture, transfection and confocal fluorescence microscopy

SER-W3 cells were grown in Dulbecco’s modified eagle’s medium (1 g/L glucose) with 10% (v/v) fetal calf serum and 2 mM L-glutamine and NIH3T3 cells were grown in Dulbecco’s modified eagle’s medium (4.5 g/L glucose) with 10% (v/v) fetal calf serum and 2 mM L-glutamine at 37 °C in 5% (v/v) CO₂. The cells were seeded on cover slips and 1-2 x 10⁵ cells were transfected with the PHF1-YFP and PHF1-variant-CFP constructs in six well plates using FuGENE (Promega) and 1 µg of total plasmid DNA per well according to the manufacturer’s instructions. Two days after transfection, the cells were fixed with 4% (w/v) paraformaldehyde, chromatin was stained with DAPI and the cells were embedded with Mowiol (Carl Roth). Confocal fluorescent images were taken with a Zeiss LSM510 instrument (Carl Zeiss, Jena, Germany, software version 3.0) using a 63 x oil immersion objective for transfected SER-W3 cells and 40 x objective for transfected NIH3T3 cells. The Argon laser
line at 514 nm and 458 nm was used to excite YFP and CFP fluorescence, respectively, and a BP530-600 filter for YFP and a LP475 filter for CFP were used for image recording. Images of DAPI stained nuclei were visually inspected with the microscope, but were not recorded due to limitation of the Zeiss LSM510 instrument.

4.4 Similarities and differences in substrate specificity of MLL3 and MLL1 of the mixed lineage leukemia family of methyltransferases

4.4.1 Cloning, expression and purification

The sequence encoding the SET domain including the pre- and post-SET domain of the human mixed lineage leukemia 3 (MLL3) (residues 4733-4911; NCBI accession number NP_733751.2) was amplified from cDNA derived from HEK293 cells and cloned as GST fusion protein into the pGEX-6P-2 vector (GE Healthcare) using BamHI/NotI restriction sites. The sequence encoding the SET domain including the pre- and post-SET domain of the human mixed lineage leukemia 1 (MLL1) (residues 3812-3969; NCBI accession number NP_005924.2) was amplified from cDNA derived from HEK293 cells and cloned as GST fusion protein into the pGEX-6P-2 vector (GE Healthcare) using BamHI/XhoI restriction sites by Qazi Raafiq.

The expression and purification of the proteins were carried out as described in section 4.3.1. The expression and purification of the MLL1 SET domain was done by Qazi Raafiq.

4.4.2 Synthesis of peptide SPOT arrays

Peptide arrays were synthesized using the SPOT synthesis method [Frank, 2002] by Dr. Srikanth Kudithipudi. Each spot had diameters of 2 mm and contained approximately 9 nmol of peptide (Autospot Reference Handbook, Intavis AG). Successful synthesis of each peptide was confirmed by bromphenol blue staining of the membranes.

4.4.3 Methylation of peptide SPOT arrays

All peptide SPOT membranes (histone tail membrane, specificity profiles for H3K4 and putative non-histone targets membranes) were first incubated in equilibration buffer (50 mM Tris/HCl pH 8.0, 50 mM NaCl, 0.5 mM DTT) for 5 min at room temperature. Afterwards, the membranes were incubated with the SET domains in methylation buffer containing 50 mM Tris/HCl pH 8.0, 50 mM NaCl, 0.5 mM DTT and 0.38 µM [methyl-3H]-AdoMet (2.7 TBq/mmol;
Perkin Elmer) for 2-4 h at room temperature. Protein concentrations were applied as follows: MLL3 SET domain on histone tail membrane 7.8 µM, MLL3 SET domain on specificity profile for H3K4 membrane 6.6 µM, MLL3 SET domain on putative non-histone targets membrane 4.7 µM, MLL1 SET domain on specificity profile for H3K4 membrane 3 µM and MLL1 SET domain on putative non-histone targets membrane 1.5 µM. Methylation of the specificity profile for H3K4 membrane by MLL1 was carried out by Qazi Raafiq. Purified MLL1 SET domain was provided by Qazi Raafiq for the methylation of putative non-histone targets membrane. The membranes were washed four times with wash buffer (100 mM NH₄HCO₃ and 1% SDS) and incubated in Amplify NAMP100V solution (GE Healthcare) for 10 min. The membranes were incubated on HyperfilmTM high performance autoradiography films (GE Healthcare) in the dark for 15 h or 3-7 days (depending on the signal intensity) at -80 °C. The films were developed using AGFA Curix 60 developing machine. The methylation of the H3K4 specificity profile on peptide arrays by MLL3 and MLL1 was at least carried out twice and the methylation of the putative non-histone targets by MLL3 and MLL1 was done once for both enzymes. The signal intensity for each peptide spot for the putative non-histone target results were analyzed with the programme Phoretix™ Array.
5 References


Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM: Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science 2006; 312(5774):748-751.


Due to copyright issues the appendix is not attached to this thesis.