Menin/MLL complexes in chromatin regulation

Rick van Nuland

ISBN/EAN: 978-90-393-6012-5

Printed by Proefschriftmaken.nl, Uitgeverij BOXpress

Cover: ‘nucleosome binding through multivalent interactions’
Adapted nucleosome structure from PDB:1KX5

Printing of this thesis was financially supported by Ipsen Pharmaceutica BV

Menin/MLL complexes in chromatin regulation

Menin/MLL complexen in chromatine regulatie
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
donderdag 24 oktober 2013 des middags te 12.45 uur

door

Rick van Nuland
geboren op 26 juni 1985
te Anna Paulowna

Promotor: Prof.dr. H.Th.M. Timmers

Voor Eric, Bastard forever

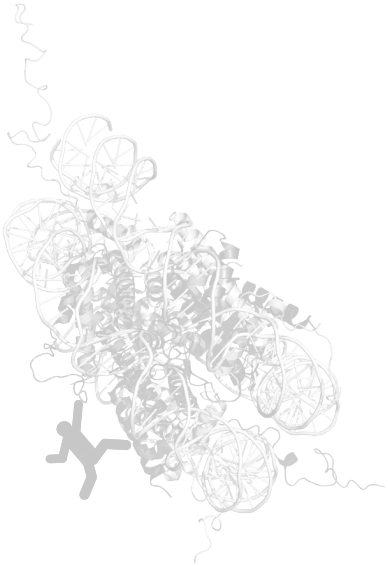
Table of contents

List of abbreviations used

ac	Acetylation
Bod1	Bi-orientation defected protein 1
BRE	B-recognition element
ChIP	Chromatin immunoprecipitation
CTD	C-terminal domain
DNMT	DNA methyltransferase
DPE	Downstream promoter element
EcR	Endysone receptor
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
ER	Estrogen receptor
FDR	False discovery rate
FIHP	Familial isolated hyperparathyroidism
FRAP	Fluorescence recovery after photobleaching
HAT	Histone acetyltransferase
HDAC	Histone de-acetylase
HMT	Histone methyltransferase
IBD	Integrase binding domain
iBAQ	Intensity based absolute quantification
ILV	Isoleucine, leucine and valine
INR	Initiator element
JNK	Jun kinase
K_D	Dissociation constant
KMT	Lysine methyltransferase
K_{off}	Dissociation rate
Ledgf	Lens epithelial derived growth factor
LFQ	Label free quantification
LOH	Loss of heterozygozity
LSD	Lysine specific demethylase
MBD	Methyl binding domain
MBM	Menin binding motif
me	Methylation
MEF	Mouse embryonic fibroblast
MEN1	Multiple endocrine neoplasia type 1
MLA	Methyl lysine analog
MLL	Mixed lineage leukemia

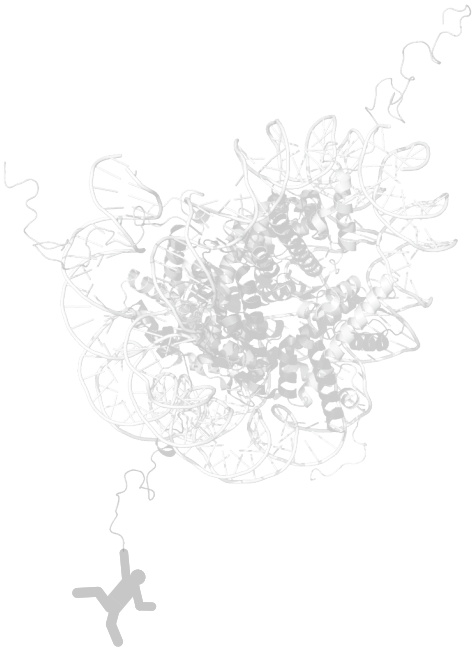
NCP	Nucleosome core particle
NDR	Nucleosome depleted region
NMR	Nuclear magnetic resonance
NR	Nuclear hormone receptor
Pa1	Ptip associating protein 1
ph	Phosphorylation
PHD	Plant homeodomain
PIC	Pre-initiation complex
Pol II	RNA polymerase II
PPAR	Peroxisome proliferator activated receptor
PRC(1/2)	Polycomb repressive complex 1 or 2
PRMT	Protein arginine methyltransferase
PTM	Post-translational modification
RXR	Retinoid X receptor
SAM	*S*-adenomethionine
SET	Su(var)3-9, enhancer of zeste, tritorax
SHL	Super helical loop
SPRY	Spla and ryanodine receptor
Taf	TFIID associated factor
TBP	TATA binding protein
TK	Thymidine kinase
Trx	Trithorax
TSS	Transcription start site
ub	Ubiquitinated
VDR	Vitamin D receptor
WD	Tryptophan-aspartate
WH	Winged helix
Win	Wdr5 interaction motif
WRAD	Wdr5-Rbbp5-Ash2l-Dpy30 complex
WT	Wild type

Modifications on histones are given by the histone, the site of modification, the type and the degree of the modification. Tri-methylation at lysine 4 on histone H3 is therefore referred to as H3K4me3 and so on. Proteins names generally start with a capital followed by lower case. Protein complexes names are always given in full capitals for clarity.



Chapter 1

General introduction



1

Genome organization and chromatin modifications

The human genome consists of about $3x10^9$ basepairs of DNA that encode for an estimated 22.000 distinct genes. This two meters of DNA is fitted into an approximately 6 µm nucleus by packaging and folding the DNA fiber into chromatin. The repetitive unit of this chromatin fiber is the nucleosome. Nucleosomes consist of approximately 146-bp of DNA, which is wrapped 1.7 times around an octamer of histone proteins. Nucleosomes are spaced by an average 50-bp of linker DNA (Kornberg 1974). Canonical histone octamers consist of two copies of the canonical histones H2A, H2B, H3 and H4. Histones contain a central histone fold domain by which H2A and H2B or H3 and H4 form dimers inside the nucleosome. The presence of several positively charged amino acids interspersed by small residues especially on histone H3, allows for an extended unstructured conformation of the N-terminus. As such, the first 39 amino acids of H3 protrude out of the nucleosome structure. Similar to H3, the N-terminus of H4 and both the N- and C-termini of histone H2A and H2B are highly flexible and also extend out of the nucleosome (Luger et al. 1997). A fifth histone protein, histone H1, binds to the linker DNA and promotes higher order chromatin structure and compaction (Izzo et al. 2008).

Histone variants

Histone proteins are highly conserved. Nevertheless several histone variants are present in mammals that can replace the canonical histones. In metazoans three main classes of histone H3 variants can be distinguished, the replication dependent H3.1 and H3.2, the replication-independent H3.3 and the centromeric variant Cenp-A. Distinct functions have been attributed to these variants. For example, histone H3.3 is deposited specifically at active genes, promoters and regulatory elements by the histone chaperone Daxx (Elsaesser et al. 2010). Additionally, the histone H2A variant H2A.X is detected specifically at sites of DNA damage where it is highly phosphorylated. On the other hand, the histone H2A.Z variant is found enriched around 5'-ends of genes and enhancers where it lowers the stability of the H2A/H2B dimer in the nucleosome, potentially being important for transcription regulation (Wyrick and Parra 2009).

Nucleosome remodeling

The complex of histones and DNA is not static and is continuously being remodeled by ATP dependent enzymes that lower the affinity of histones for DNA. The human genome encodes 27 ATP dependent chromatin remodelers that fall into four major classes, the Swi/Snf, Iswi, Ino80 and Chd families, each of which have distinct cellular functions in either gene repression or activation (Hargreaves and Crabtree 2011). Chromatin remodeling enzymes can slide nucleosomes either into or away from regulatory regions, thereby masking or opening up binding sites for DNA binding factors. Genome wide mapping of nucleosome positions revealed

that nucleosomes are not randomly distributed (Rhee and Pugh 2011). Nucleosomes at the start and end of genes are usually well positioned yet this regular spacing fades out into the body of the gene. At many genes a nucleosome depleted region (NDR) is found upstream of the first nucleosome. This is in part due to a disfavored spacing of AA/TT dinucleotides found frequently enriched at promoter regions. Based on different algorithms, DNA sequence defines approximately 50% of the nucleosome positions in the compact yeast genome (Jiang and Pugh 2009). ATP dependent remodeling and the DNA sequence context together are required for the NDR. These sites usually contain regulatory sequences and transcription factor binding sites. Similar to promoters, enhancer elements are also generally depleted for nucleosomes and therefore hyper-sensitive to DNaseI treatment (Kulaeva et al. 2012).

Histone modifications

The flexible tails as well as the globular domain of histones are subjected to several post-translational modifications (PTM) such as acetylation (ac), methylation (me), phosphorylation (ph) and ubiquitination (ub) (Kouzarides 2007). An overview of the reported modifications found on histone H3 is depicted in Figure 1.

Figure 1. Modifications on histone H3. Schematic representation of all published modifications on histone H3. For simplicity only methylation, acetylation and phosphorylation are indicated. The globular part of histone H3 starts around residue 38 and is indicated. Numbering is relative to the N-terminus. Please note that the translation initiating methionine is removed in cells.

Lysine acetylation

Histone acetylation at lysine residues is generally associated with promoters and enhancer sites where it positively correlates with transcriptional activation. Acetylation neutralizes the positive charge of the lysine residue and this was initially thought to be the major cause of reduced chromatin compaction. More recent work showed that acetylated lysines recruit protein complexes that function in chromatin regulation directly. Histone acetyltransferases (HAT) enzymes catalyze the transfer of an acetyl moiety from acetyl coA to the ε-amino group of a specific lysine residue. Several families of HATs have been identified including the Gcn5 related, Myst and Cbp/p300 families. HATs are usually part of large co-factor complexes such as the Gcn5 protein in the SAGA complex. These HAT enzymes are opposed by the action of histone deacetylases (HDAC) that can remove the acetyl group. HDACs are found in four different classes: the NAD^+ dependent Sirtuins (class III) and the three classes of Zn^{2+} dependent HDAC I, II and IV families (Kim and Yang 2011). Lysine acetylation was initially identified on transcriptional co-regulators such as histones and transcription factors. Recent unbiased approaches revealed that the mammalian acetylome consist of approximately 1.750 nuclear

and cytoplasmic proteins (Choudhary et al. 2009). The functional relevance of lysine acetylation on many of these proteins however remains to be elucidated.

Arginine methylation

Methylation can be found on both arginine and lysine residues. Arginine can be modified with one or two methyl groups, the latter either symmetrically or asymmetrically. Protein arginine methyltransferases (PRMT) transfer the methyl group of *S*-adenomethionine (SAM) to the arginine amino group. This group of enzymes consists of three major classes based on their structural properties. Arginine methylation has different outputs dependent on the substrate, the site and the modifications in the vicinity. Asymmetric di-methylation at arginine 2 on histone H3 (H3R2) for example negatively affects the binding to and modification of the adjacent histone H3 lysine 4 (H3K4) thereby repressing transcription (Klose and Zhang 2007). On the other hand, symmetric H3R2 di-methylation correlates with active chromatin (Migliori et al. 2012). Additionally, non-histone substrates like Cbp/p300, Cpg-1 and Nip45 can be methylated by the Carm1 PRMT, modulating their transcriptional activity (Bedford and Clarke 2009).

Lysine methylation

Methylation of lysines is also highly site specific and can be present as one, two or three methyl groups on every residue (me1/2/3). Lysine methyltransferases (KMT) generally contain a catalytic domain (SET: Su(var)3-9, Enhancer of zeste, Trithorax) that transfers the methyl group of SAM to the substrate lysine. The exception is the non-SET domain KMT Dot1l which modifies H3K79 in the globular part of histone H3 (Min et al. 2003a). Whereas histone acetylation correlates with transcription activation, lysine methylation can act in either repression or activation, dependent on the site and extent of modification (Black et al. 2012). One of the best studied methylations is present on lysine 4 in H3. H3K4me3 is almost exclusively found in promoter regions of active or poised genes, where it generally correlates with the transcription levels (Vermeulen and Timmers 2010; Eissenberg and Shilatifard 2010). H3K4 methylation is mediated by the multisubunit SET1/MLL complexes (discussed in detail below). Whereas H3K4me3 peaks at the first nucleosome downstream of the transcription start site (TSS) (referred to as +1 nucleosome), the intermediate methylation states show a different distribution, with H3K4me1 peaking specifically at enhancer sites (Heintzman et al. 2009). An average active gene profile is depicted in Figure 2. The difference in the distribution of the marks is not exclusive as mono-methylation can also be detected at promoters (Calo and Wysocka 2013). Additionally, the intermediate states are found more broadly across active chromatin (Barski et al. 2007).

The identification of the lysine specific demethylases (LSD) and the related Jumonji proteins revealed the dynamic and reversible nature of lysine methylation (Kooistra and Helin 2012). Histone demethylase enzymes can discriminate between different lysine residues and the

methylation grade. Demethylases for H3K4me are generally associated with transcriptional repression by acting at the promoter. However, recent work has shown that their demethylation activity is also required to keep the H3K4me1 stable at enhancer sites and thereby keeping them active (Outchkourov et al. 2013).

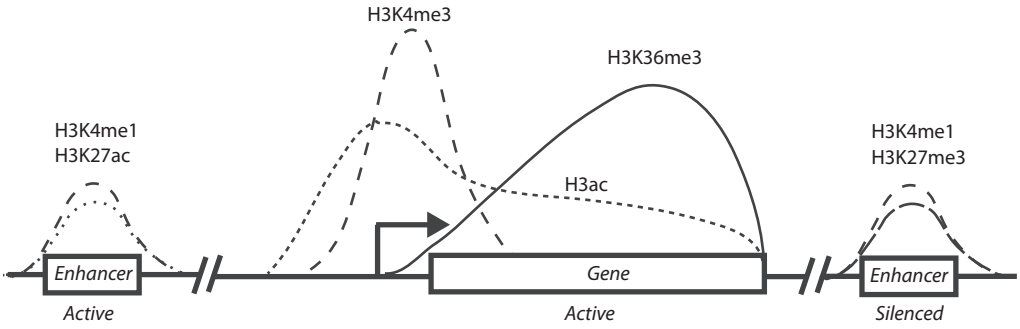


Figure 2. Average gene profile of an active gene. Active genes are generally marked by the presence of H3K4me3 at the promoter, H3K36me3 in the gene body and a more dispersed acetylation pattern. On enhancers, H3K4me1 is present together with H3K27me3. Once activated, repressive H3K27me3 is replaced by H3K27ac.

Whereas H3K4me3 correlates with an active or poised gene promoter, methylation of lysine 36 on histone H3 (H3K36me) marks active transcription in the bodies of genes. In yeast, the Set2 methyltransferase is responsible for this modification. This enzyme binds to the elongating RNA polymerase II (Pol II). When a gene is transcribed, H3K36me3 is deposited. HDAC complexes can recognize this mark and subsequently the removal of acetylation establishes a repressive environment, preventing cryptic transcription initiations in gene bodies (Li et al. 2007). H3K36me in mammals is more complex and implicated in splicing, response to DNA damage and polycomb recruitment (Wagner and Carpenter 2012). In mammals H3K36me1 and –me2 is deposited by a family of three proteins (Nsd1, Nsd2 and Nsd3). H3K36me1 and –me2 are again a substrate for the human Set2 protein that coverts them to H3K36me3 (Li et al. 2009). The Nsd proteins are highly similar but not redundant in their action. Translocations of *Nsd2* specifically are associated with multiple myeloma whereas mutations in *Nsd1* are found in several developmental disorders as well as in cancer (Morishita and di Luccio 2011), underscoring the importance of H3K36me for cellular function.

Methylation of lysine 9 on histone H3 (H3K9) is found at repressed genes (Zeng et al. 2010). H3K9me recruits heterochromatin protein Hp1 and its binding results in the local compaction of the chromatin fiber. This modification can be found in large blocks mainly at telomeres, at centromeres and on loci in the nuclear periphery. Compaction of chromatin has major effects on the accessibility of the DNA double helix for transcription factors and it is thought that this is the main mechanism used to by cells to silence large segments in the genome. H3K27me3 is also found in repressed chromatin however is deposited by the Polycomb repressive complex

2 (PRC2) at promoters and in the bodies of genes. Initially identified in flies as a repressor of homeotic gene transcription, PRC2 is essential for proper development and the shutdown of developmental genes at the right time and place. Despite the fact that H3K27me3 has been studied for many years, surprisingly little is known about the molecular mechanisms underlying the repressive nature of this modification (Sparmann and van Lohuizen 2006; Margueron and Reinberg 2011).

Phosphorylation

Histone phosphorylation on serine and threonine residues is largely restricted to mitosis. Phosphorylation of residues 3,10 and 28 on histone H3 (H3T3, H3S10 and H3S28, respectively) are the well studied and implicated in mitotic progression as well as transcription regulation. The mitotic kinase Aurora-B is able to phosphorylate H3S10, which prevents the binding of Hp1 to the neighboring H3K9me3 and heterochromatin spreading (Duan et al. 2008). A similar phospho/methyl switch is observed for the repressive Polycomb mediated H3K27me, for which effector proteins were shown to be dissociating by adjacent phosphorylation at H3S28 (Lau and Cheung 2011). Additionally, H3T3ph is conducted by the Haspin kinase (Dai et al. 2005) and is found along chromosome arms where it temporarily ejects the basal transcription factor TFIID from H3K4me3 decorated promoters during mitosis (Varier et al. 2010). During metaphase, H3T3ph concentrates at the inner centromere, regulating and localizing the Chromosomal passenger complex through direct binding of the Bir domain of Survivin to the phosphorylated threonine (Wang et al. 2010a). Additional to mitotic phosphorylation, DNA damage induces heavy phosphorylation of the variant H2A.X histone, marking the site of damage (Sawicka and Seiser 2012).

Ubiquitination

Histones can also be covalently decorated with a small (7 kDa) ubiquitin moiety. This modification is deposited on lysine residues by an E3 ligase enzyme together with a conjugating E2 enzyme and a ubiquitin activating ATP dependent E1 protein. Generally ubiquitination is found as poly-ubiquitin chains and it targets proteins for degradation by the proteasome. Mono-ubiquitination however serves as a signaling modification. On mammalian histones mono-ubiquitin is detected at both H2AK119 (H2Aub) and H2BK120 (H2Bub) (Weake and Workman 2008).

H2Aub is mediated by the Polycomb repressive complex 1 (PRC1) E3 ligase complex. H2Aub and PCR1 localize together on the inactive X chromosome and to repressed promoters of Polycomb target genes. H3K27me3 by the PCR2 complex is upstream of H2Aub and it is proposed that H2Aub affects higher order chromatin structures and H3K4me by unknown mechanisms (Vidal 2009). At least three deubiquitinating enzymes are found that can remove the H2Aub. However it is not known how these factors are targeted and regulated in cells (Atanassov et al. 2011).

In humans H2Bub is deposited by the Rnf20/Rnf40 E3 ligases that share sequence homology with yeast Bre1. The SAGA subunit Usp22 is a specific proteolytic enzyme that can dynamically remove ubiquitin groups from H2B. Additional E3 ligases and deubiquitinating enzymes have been described but their function *in vivo* is not well established (Vissers et al. 2008). Work in both mammalian cells and yeast has revealed crosstalk between H2Bub and the deposition of H3K4me. The human SET1 complex subunit Wdr82 was implicated in this crosstalk by direct binding to H2Bub (Lee et al. 2007). Very recently two independent groups proposed additional mechanisms for yeast Set1 and human MLL1. Using an *in vitro* system Kim *et al.* showed that the SET domain of yeast Set1 is critical for this crosstalk. Conformational changes in the ubiquitinated nucleosome potentially make it a better substrate for methylation by the SET domain (Kim et al. 2013). Interestingly, H2Bub also promotes H3K79me, possibly through a similar mechanism. On the other hand Wu *et al.* could show that the Ash2l subunit of MLL1 binds directly to H2Bub thereby stabilizing the KMT complex on nucleosomes. Interestingly, although Ash2l is a general subunit of all SET1/MLL complexes, this type of regulation is not observed for the MLL3 complex (Wu et al. 2013).

DNA methylation

Additional to histone modifications, the methylation and hydroxymethylation of DNA also plays an important role in the regulation of transcription and other chromatin related processes. Approximately 80% of human genes have a markedly high CG dinucleotide content in their promoter regions. These regions of several hundreds base pairs (CpG islands) are implicated in transcriptional regulation. Cytosines in a CpG context can be methylated by three conserved enzymes (DNMT) and further oxidized to hydroxymethyl cytosine by the family of Tet enzymes. CpG methylation correlates with transcriptional repression but is relatively low in normal cells. Conversely in many cancer cells, methylation of tumor suppressor genes is frequently observed. DNA modifications have the ability to repel DNA binding factors as was reported for CXXC type zinc finger domains, but can also recruit methyl binding domain (MBD) containing proteins (Spruijt et al. 2013). The function of hydroxymethylation is far from understood, yet it is suggested to be an intermediate in the DNA demethylation pathway (Smith and Meissner 2013).

Chromatin states

The concerted action of DNA and histone modifications allows for the classification of loci based on their modification status. Classically, the state of chromatin compaction was subdivided in silent heterochromatin and active euchromatin. Classic heterochromatin can be visualized using microscopy as intense staining areas in the nucleus, corresponding to telomeres and satellite repeat sequences. Facultative heterochromatin however is present at loci that are differentially expressed between cell types. The inactive X chromosome in females is the best example of this

(Bickmore and van Steensel 2013).

The development of genome wide mapping of histone and DNA modifications as well as other chromatin proteins resulted in a more complex classification referred to as the five colors of chromatin. Two types of active chromatin can be identified based on the presence of active modifications and proteins. These types differ by the presence or absence of H3K36me (yellow and red). Repressed loci are marked by Hp1 binding (green) but are distinct from the also repressed Polycomb bound loci (blue). Additionally, large blocks of facultative repressed chromatin (black) were identified (Filion et al. 2010). Together these data illustrate that the genome is organized in blocks of chromatin modifications, that correlate to the transcriptional state of the underlying DNA. This classification is not absolute and depends on the modifications and proteins that are included in the analysis. Ultimately the chromatin signature of a locus can help in understanding the relation between the modifications and its transcriptional state.

Translating the histone code

In order to translate the chromatin signature into a functional output, modifications are recognized by effector proteins through specialized domains that directly bind to the modified residue (Yun et al. 2011). Methylated lysines are recognized by domains that bind to the modified histone tail using a hydrophobic pocket build up of several aromatic residues. These domains include Chromo-, PWWP-, Chromobarrel-, Tudor- and Plant homeodomains (PHD) (Musselman et al. 2012). Based on the size and composition of the pocket, reader domains can accommodate a non-, mono-, di- or tri-methylated lysine. Additionally, the residues surrounding the lysine determine the binding specificity of these modules. Chromodomains preferentially bind to tri-methylated lysines as was shown for Hp1 binding to H3K9me3 (Nielsen et al. 2002) and Polycomb binding to H3K27me3 (Min et al. 2003b). Zinc dependent PHD fingers found in Taf3, Bptf and the Ing proteins are typical H3K4me3 binding modules as they can only accommodate three additional histone H3 N-terminal residues in their pocket (Shi et al. 2006; Li et al. 2006; van Ingen et al. 2008). Addition of N-terminal residues as well as modifications on H3T3 and H3R2 inhibit the binding of these domains to H3K4me3 peptides. PWWP domains contain a Pro-Trp-Trp-Pro motif and were identified as specific binding modules for H3K36me and H4K20me, despite a low affinity for methylated peptides compared to other reader modules (Vezzoli et al. 2010; Qiu et al. 2012). We show in chapter 4 that the positive surface of these domains however allows the PWWP domain to interact with DNA, thereby increasing the affinity for nucleosomes drastically (Eidahl et al. 2013).

Acetylated lysines can be recognized by Bromodomain containing proteins. Bromodomains usually contain two tyrosine residues that contact the acetylated lysine and generally bind with a low affinity to histone peptides. Bromodomains are frequently found in tandem, for example in the TFIID subunit Taf1, thereby increasing the affinity for double-acetylated histones considerably. The spacing between the two domains also determines its specificity as it dictates

the spacing between the acetylated residues that can be bound (Filippakopoulos et al. 2012).

The Bir domain of Survivin, binds specifically to H3T3ph but does not contain a hydrophobic pocket. In this case, the binding is established by a small positively charged surface on the protein. Alternatively, phosphorylated residues can be recognized by BRCT domains found in the DNA damage protein Mdc1, that can bind to H2A.XS139ph and in the MLL3/4 subunit Ptip (Yun et al. 2011).

An additional level of complexity is established by the combination of reader domains within a complex, such as the Taf1 and Taf3 subunits of TFIID. Multiple domains can also co-occur within a single polypeptide. Tandem domains allow for a high degree of specificity in recognizing and translating the histone code either on the same or on different histone tails. Good examples are the Brpf1 and Brpf2 proteins that contains in a Bromo-, a PHD- and two PWWP domains allowing for combinatorial binding to different modifications at the same time.

SET1 methyltransferase complexes

Several different chromatin regulatory complexes associate with H3K4 methylated chromatin through direct interactions. H3K4me3 in yeast is deposited exclusively by the Set1 enzyme, as deletion of the gene results in complete loss of mono-, di- and tri-methylation (Briggs et al. 2001; Roguev et al. 2001; Nagy et al. 2002; Miller et al. 2001). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. At least six different homologs are identified in mammals: Set1a, Set1b and Mll1-4 (Shilatifard 2012), that can bind to shared and specific subunits (Table 1). Set1 related KMTs display only a very weak methyltransferase activity on their own and require the interaction with additional subunits for maximal activity (Dou et al. 2006). The highly conserved catalytic SET domain is located in the extreme C- termini of these proteins and only has exo-methyltransferase activity as addition of N-terminal residues on histone H3 prevents its activity (Takahashi et al. 2011).

Core SET1/MLL complex

The Set1 related methyltransferase enzymes all bind to a conserved complex of Wdr5-Rbbp5-Ash2l and Dpy30, the mammalian homologs of Csp30, Csp50, Csp60 and Csp25 (referred to as WRAD), that are essential for proper methylation by the catalytic subunit (Ernst and Vakoc 2012). The function of these proteins however is far from understood. The trypthophan-aspartate (WD) repeat protein Wdr5 can bind to the Wdr5-interaction (Win) motif present in the catalytic subunit close to the SET domain. Previously it was reported that Wdr5 can bind directly to the tail of histone H3 (Wysocka et al. 2005). These observations are strongly challenged by crystallographic studies from independent groups showing that this binding pocket also binds to the Win motif of the Set1/Mll proteins (Song and Kingston 2008; Patel et al. 2008). Affinity measurements also demonstrated that a histone peptide is competed out by a peptide of the

Win domain. Wdr5 itself binds to Rbbp5 through a V-shaped groove that is distinct from the Mll binding pocket (Avdic et al. 2011). Crystallographic information on Rbbp5 is lacking but mutational analysis showed that Rbbp5 binds to Wdr5 via a valine-aspartate-valine motif. An adjacent binding motif on Rbbp5 was identified that interacts with the SPla and RYanodine Receptor (SPRY) domain of Ash2l (Chen et al. 2012). Dpy30 is present at least as a dimer and interacts with the winged helix (WH) domain of Ash2l through its four-helix bundle domain (Chen et al. 2012; Patel et al. 2011). Whereas the functions of Rbbp5 and Dpy30 are very much unclear, it appears that Ash2l is important for the binding of the complex to DNA and H2Bub (Wu et al. 2013). Ash2l also harbors a DNA binding motif that is structurally related to the forkhead box transcription factors by which it can directly interact with DNA (Chen et al. 2011; Sarvan et al. 2011). Mutations in the DNA binding motif impairs the methylation of target genes and their transcription.

The WRAD complex increases the catalytic activity of the SET domain dramatically both *in vivo* and *in vitro*. It has been proposed that the WRAD module slightly molds the catalytic site of the SET domain thereby stimulating its methyltransferase activity (Patel et al. 2011). This is in agreement with the electron microscopy (EM) structure of the yeast Set1 complex in which the Wdr5 and Rbbp5 homologs are positioned close to the catalytic domain (Takahashi et al. 2011). It has been reported that the interactions between Set1/Mll and the WRAD module components are pairwise. Mll binds directly to Wdr5, Wdr5 to Rbbp5, Rbbp5 to Ash2l and Ash2l to Dpy30. Nevertheless, others have shown that even without Wdr5, the other WRAD components are able to stimulate the activity of Mll1 in an *in vitro* assay (Cao et al. 2010; Avdic et al. 2011). This apparent discrepancy could be dependent on the different partial methyltransferase fragments used in these assays.

SET1A and SET1B

The nearest homologs of the yeast Set1 in mammals, Set1a and Set1b, only contain a SET domain and a RNA recognition motif. Both proteins strongly interact with the conserved Wdr82 and Cfp1 subunits. Wdr82 is essential for global tri-methylation in mammals, demonstrating that the SET1A/B complexes are responsible for the bulk of H3K4me3 (Wu et al. 2008; Lee and Skalnik 2008). On the other hand, MLL1-4 do not function globally in tri-methylation but only at specific genomic sites. The SET1A/B subunit Cfp1 contains a CXXC zinc finger that specifically binds to non-methylated CpG stretches (Tate et al. 2010). In yeast, deletion of this subunit results in a dramatic decrease in H3K4me3 levels with only little effect on the intermediate states. Notably, yeast does not have DNA methylation. Elegant experiments by Bird and colleagues showed that insertion of a non-methylated CpG island into the mouse genome results in the recruitment of Cfp1 and consequently H3K4me3 deposition on the surrounding chromatin (Thomson et al. 2010). These results indicate that DNA methylation can directly affect H3K4me. In chapter 3 we describe a novel subunit, specific for the SET1B complex that was previously linked to

kinetochores and chromosome alignment during mitosis (Porter et al. 2007). The function of this protein in H3K4 methylation however remains to be studied.

Yeast	*Fly*	*Mammals*	*Specificity*	*Function*
Set1	Set1	Set1a		HMT
		Set1b		HMT
	Trx	Mll1		HMT
		Mll2		HMT
	Trr (Lpt)	Mll3		HMT
		Mll4		HMT
Cps60	Ash2	Ash2l	shared	DNA binding
Cps50	Rbbp5	Rbbp5	shared	
Cps40	Cxxc1	Cfp1	Set1a/b specific	CpG binding
Cps35	Wdr82	Wdr82	Set1a/b specific	Ubiquitin recognition
Cps30	Wds	Wdr5	shared	H3 tail recognition
Cps25	Dpy30	Dpy30	shared	
Cps15		Bod1	Set1b specific	
	Mnn1	Menin	Mll1/2 specific	TF binding
		Psip1	Mll1/2 specific	H3K36 binding
	Utx	Utx	Mll3/4 specific	H3K27me demethylase
		Ncoa6	Mll3/4 specific	NR co-activator
	Ptip	Ptip	Mll3/4 specific	
	Pa1	Pa1	Mll3/4 specific	

Table 1. SET1/MLL subunits. Conservation of SET1/MLL subunits in yeast (*S. cerevisiae*), Fly (*D. melanogaster*) and human (*H. sapiens*) and their specificity. The molecular function of the individual subunits is mentioned when descibed in literature.

MLL1 and MLL2

Mll1 and Mll2 are the human homologs of Drosophila *trithorax* (*Trx)* and differ from the Set1a and Set1b proteins by their domain structure and interactors (Shilatifard 2012). *Trx* was identified in a genetic screen controlling body segmentation in flies through the regulation of genes in the *Hox* cluster. This function is conserved, illustrated by the homeotic phenotype of *Mll1* knockout mice (Eissenberg and Shilatifard 2010). *Mll1* was first identified as a frequent translocation partner in Mixed Lineage Leukemia (MLL) hematopoietic cancers (Meyer et al. 2009). The oncogenic function of Mll1 will be discussed in more detail below. Mll1 and Mll2 both contain a series of DNA binding AT-hooks followed by a CXXC domain similar to the SET1A/B subunit Cfp1. Additionally Mll1 and Mll2 contain four PHD fingers of which PHD3 in Mll1 is reported to bind H3K4me3. This interaction is important for maintaining and propagating of the methyl mark (Chang et al. 2010). Conversely, PHD3 is also reported to interact with the RNA recognition motif of Cyp33, switching Mll1 to a repressor by proline-isomerization (Wang et al. 2010b). The 500 kDa Mll1 protein is produced from a single mRNA molecule and proteolytically processed into a 320 kDa N- and a 180 kDa C-terminal part by Taspase (Capotosti et al. 2007). Taspase also processes the Mll2 protein into two smaller fragments.

The MLL1 and MLL2 complexes were shown to interact with Menin and Psip1. Psip1/Ledgf/p75 is a highly expressed protein that contains a N-terminal PWWP domain that interacts with H3K36me, a DNA binding AT-hook and a C-terminal Integrase binding domain (IBD). The IBD functions as a protein-protein interaction domain for several factors such as Menin, Cdca7l and the Cdc7 kinase. Additionally, upon HIV infection the viral integrase can binds to this IBD domain. This interaction directs the integration of viral DNA to actively transcribed chromatin illustrating the importance of chromatin binding by Psip1 (Ferris et al. 2010). A short isoform of Psip1, p52, lacks the IBD but contains an additional eight amino acids. These residues are shown to be required for the interaction with the splicing machinery (Pradeepa et al. 2012).

The Menin tumor suppressor protein was identified in MLL1/2 and shown to be important for the recruitment of the complex to promoters of the *Hox* locus, cell cycle inhibitor genes and nuclear receptor (NR) targets (Menin function will be discussed in more detail below) (Thiel et al. 2012). Menin can interact with the N-terminal domain of the Mll1/2 proteins via a Menin binding motif (MBM) (Yokoyama et al. 2004). The interaction of Psip1 to Mll is completely dependent on Menin, indicating a linear interaction. Recent crystallographic data however, showed that Psip1 could also interact directly with Mll1 (Yokoyama and Cleary 2008; Huang et al. 2012).

Menin and the MEN1 syndrome

Inactivating germ line mutations in the Menin coding tumor suppressor gene *MEN1* are found in approximately 1 out 30.000 people. Upon loss of heterozygozity (LOH) patients develop multiple endocrine neoplasia type 1 (MEN1) (Chandrasekharappa et al. 1997). MEN1 patients display tumors in the main endocrine organs (pancreas, parathyroid, pituitary). Besides the familial MEN1 syndrome, 44% of sporadic pancreas islet tumors were shown to have a mutation in the *MEN1* gene (Jiao et al. 2011) underscoring the importance of Menin in neuroendocrine tissues. Homozygous deletion of *Men1* in mice is embryonic lethal. Heterozygous mice however develop tumors of endocrine origin making this a useful MEN1 model (Crabtree et al. 2001). A high incidence of tumors in non-endocrine tissues was also reported in mouse models (Gracanin et al. 2009).

Menin is a 610 amino acid polypeptide and was recently shown to consist of 22 helical structures that form a central pocket in which an Mll1 peptide can bind (Huang et al. 2012; Murai et al. 2011). Interestingly, MEN1 associated mutations are not restricted to the binding pocket but are scattered over the protein. The tumor suppressor function is, at least in part, mediated through lowering H3K4me3 and consequently reducing the expression of the cell cycle inhibitors $p27^{KIP1}$ and $p18^{INK4C}$. Reduced expression of these proteins accelerates growth in cells lacking functional Menin (Milne et al. 2005; Karnik et al. 2005). Moreover, Menin can bind to sequence specific transcription factors such as nuclear hormone receptors and thereby recruiting histone methyltransferase activity to target genes (Dreijerink et al. 2009b; 2006).

Menin was initially described as an interactor for JunD, a member of the AP1 transcription factor complex (Agarwal et al. 1999). Mll1/2 and JunD both bind to the same pocket on Menin and were shown to do so in a mutually exclusive manner (Fig. 3)(Huang et al. 2012). Members of this class of transcription factors can form homo- as well as heterodimers by their basic-leucine zipper domains and can bind to specific DNA elements to regulate transcription. JunD protects cells against senescence and apoptosis whereas other AP1 complex members are associated with several types of cancer. Menin serves as a repressor of JunD mediated transcription by inhibiting the phosphorylation of JunD by Jun kinase (JNK), as was illustrated using reporter assays (Agarwal et al. 2003; Huang et al. 2012). Still, the role of JunD in MEN1 related tumors and the target genes for this complex remain to be elucidated.

Over the years an entire collection of proteins have been linked directly, such as JunD and NFκB (Agarwal et al. 1999; Heppner et al. 2001), or indirectly such as Ezh2 (Thiel et al. 2013), to Menin function. For many of these interaction the molecular mechanisms and implications for MEN1 however are lacking (Dreijerink et al. 2009a). Menin is ubiquitously expressed but heterozygous deletion results in specific tumors in the neuroendocrine tissues (Bertolino et al. 2003). Therefore the tissue specific tumor suppressive function of Menin appears to depend on the cell type specific interactions (Gracanin et al. 2009). Additionally, Menin interacts with

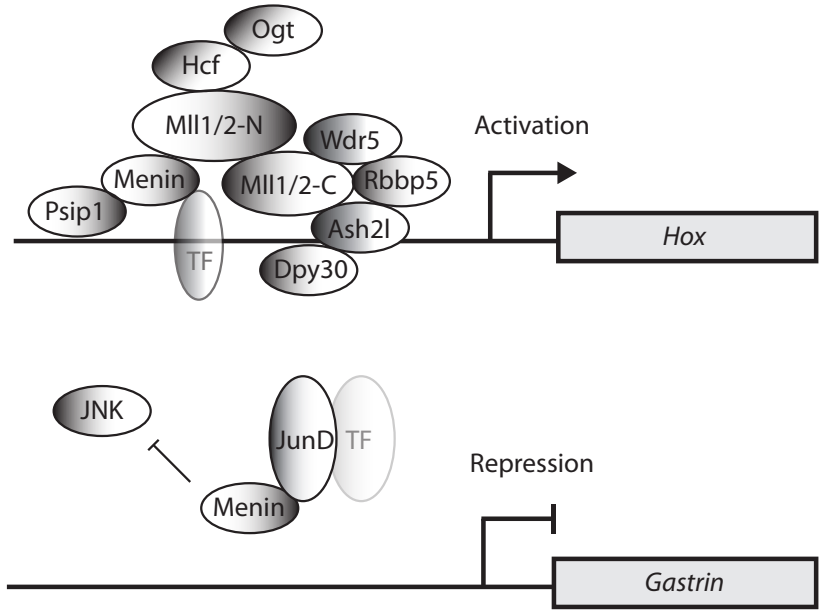


Figure 3. Functional Menin complexes. (Upper panel) Menin interacts with the MLL1/2 complexes and bridges histone methyltransferase activity to sequence specific transcription factors (TF). Menin serves as an transcriptional activator for homeotic genes in this context. (Lower panel) Menin can bind the JunD transcription factor. In complex with Menin, phosphorylation of JunD by the Jun kinase (JNK) is inhibited, thereby repressing JunD mediated transcription.

Mll1 oncogenic fusion proteins and was shown to be essential for oncogenic transformation in cells carrying such fusions (discussed in detail below). Thus, Menin acts as a tumor suppressor in normal tissues but in context of an Mll1 fusion, it functions as an oncogene.

Mixed Lineage Leukemia

Translocation of the *MLL1* gene, located at 11q23, results in acute myeloid and lymphoid leukemia. *Mll1* translocations are found usually in infants and the prognosis is generally very poor despite intense research efforts (Marschalek 2010). All fusions truncate the Mll1 protein, which leads to a loss of the C-terminus such that methyltransferase activity, PHD fingers and WRAD interactions are lacking. Several studies have shown that fusion partners of Mll1 are involved in transcriptional regulation and elongation (Meyer et al. 2009). ELL, the first Mll1 translocation partner described, is involved in increasing the catalytic rate of the RNA polymerase II enzyme. The Mll1 fragment misrecruits ELL activity to MLL1 target genes thereby increasing the transcription rate of these genes (Mohan et al. 2010). The misregulation of MLL1 target genes like *HoxA9* drives oncogenic transformation in these cells (Ayton and Cleary 2003). Other fusion partners include the HAT Cbp/p300 and the elongation factor Af9. Psip1 and Menin are both shown to be essential for oncogenic transformation of Mll- fusions (Yokoyama and Cleary 2008; Yokoyama et al. 2005). Interestingly, the Mll fusions lack methyltransferase activity nonetheless rely on the wild type allele for efficient transformation (Thiel et al. 2010).

MLL3 and MLL4

Next to *SET1A/b* and *MLL1/2*, the mammalian genome encodes for two additional H3K4me3 methyltransferase genes, the *MLL3* and *MLL4* genes respectively. These *Drosophila Trr* homologs are non redundant with Mll1/2 and serve a distinct set of genes. *Trr* knockout flies do not display homeotic phenotypes and knockout of the MLL3/4 subunit Ptip in mammalian cells has only minor effects on *Hox* gene expression (Wang et al. 2009; Sedkov et al. 1999). *Trr* was implicated in the regulation of nuclear receptor mediated transcription as was shown for the ecdysone receptor (EcR) (Sedkov et al. 2003). In mammals, Mll4 can directly bind to the estrogen receptor (ER) using its LXXLL motifs and regulates the expression of target genes (Mo et al. 2006). Recently it was demonstrated that *Drosophila Trr* acts primarily at enhancer sites by mediating H3K4me1 (Herz et al. 2012). Interestingly, MLL3/4 complexes are able to fully methylate nucleosomes *in vitro*, indicating that active demethylation is important to maintain the H3K4me1 state at enhancers. In light of this the presence of the Jumonji protein Utx, that demethylates the repressive H3K27me, offers interesting models (Agger et al. 2007). At enhancers, H3K27 is acetylated upon activating stimuli (Creyghton et al. 2010). As methylation and acetylation cannot occur on the same lysine, demethylation by Utx is required before acetylation can be put on. The nuclear receptor co-activator Ncoa6 associates with MLL3/4 complexes and is suggested to be important for integrating steroid signaling and NR target

gene regulation by H3K4me (Kawagoe et al. 2012). Additionally to Utx and Ncoa6, the Mll3/4 proteins binds to the BRCT domain containing Ptip and its partner Pa1 (Cho et al. 2007). The functions of Ptip and Pa1 in the MLL3/4 complex remains largely unknown, despite their high stoichiometric appearance (demonstrated in chapter 3).

Gene expression and transcription regulation

Chromatin modifications can have direct effects on transcription by influencing the initiation and elongation of Pol II. All protein coding genes are transcribed by this enzyme. RNA polymerases incorporate nucleotides to the growing RNA molecule using DNA as a template. Basal transcription factors are recruited to the promoter and build the pre-initiation complex (PIC). PIC formation occurs as a stepwise recruitment and initiates with the binding of the TFIID complex. Subsequently TFIIA, TFIIB and TFIIF are recruited. Once the PIC is set up, Pol II is engaged followed by additional factors like TFIIE and TFIIH. Pol II features a C-terminal domain (CTD) that consists of multiple heptapeptide repeats. Once Pol II initiates, serine 5 of this repeat gets phosphorylated by the Cdk7/cyclinH kinase in TFIIH (Buratowski 2009). Two pausing factors, Nelf and Dsif keep Pol II in a paused state downstream (+20 to +50-bp) of the TSS (Adelman and Lis 2012). Upon kinase action by the P-TEFb complex, CTD serine 2 (CTD S2ph) and Dsif get phosphorylated and subsequently Pol II is released from the promoter. The CTD S2ph modification on Pol II can recruit enzymatic activities such as the Set2 H3K36me KMT that modifies chromatin following the elongating polymerase (Hsin and Manley 2012). During elongation CTD S5ph is actively removed by CTD phosphateses. When Pol II reaches the 3'-end of a gene it terminates and is available for new rounds of initiation. Pol II pausing for longer periods is observed for genes that require immediate response to stimuli or stress, such as the *HSP70* gene in *Drosophila* (Adelman and Lis 2012).

Open chromatin structures promote transcription from cryptic promoters or in the opposite direction to create divergent transcripts. Frequently these mRNA molecules are immature and degraded by the action of the nuclear exosome. In some instances however, these non-coding RNA molecules are stabilized and function in the regulation of transcription. In yeast it was shown that some cryptic anti-sense transcripts can regulate the expression of the sense mRNA in *cis* (Smolle and Workman 2013).

Promoters and enhancers

Initiation of transcription is normally directed to core promoters. Generally these sites overlap with the TSS. Promoter sequences contain several regulatory DNA elements. The knowledge on DNA elements in promoters results to a large extent from studies using model core promoters derived from *Drosophila*. These promoters differ substantially from 'normal' promoters where

they display a large heterogeneity in promoter DNA sequences. Mammalian promoters can be divided in three main classes based on the sequence elements and the chromatin context. Class I promoters generally contain a conserved TATA sequence and have a low CpG content. These promoters are mainly found at tissue specific genes, that generally have a single TSS despite the dispersed nucleosome positioning. A second class (Class II) associates with ubiquitously expressed genes and tend to contain only a short CpG island that overlaps with the TSS. Class II promoters have a more structured nucleosome positioning and NDR. These genes generally have multiple start sites and lack a strong TATA box element. A third class (Class III) is less well defined and contains of the developmentally regulated genes. These genes feature the same motif composition and multiple start sites as class II, yet are distinct by the large CpG islands that extend into the gene bodies. Other elements that can be found at promoters are less well defined. The B recognition element (BRE) can be found in close proximity of the TATA box and is bound by the TFIIB protein. The initiator element (INR) overlaps with the TSS of the gene and is reported to recruit the Taf1/2 subunits of TFIID. Additional well studied features like the downstream promoter elements (DPE) are considered specific for flies. Typically, class I promoters have a sharp H3K4me3 peak at the +1 nucleosome, whereas class II and III have a more spread modification pattern (Lenhard et al. 2012). Class III promoters can also feature a bivalent state of active H3K4me3 and repressive H3K27me3 modifications. The existence and importance of these bivalent promoters however is under debate.

Sequence specific transcription factors integrate external signals such as growth factor signaling and steroid hormones into a transcriptional response. There are multiple different classes of transcription factors that can bind on their own, or in combination with other transcription factors to the DNA. Transcription factor binding sites are found at distal enhancers and are marked by DNAseI sensitivity and H3K4me1 but can also be located close to a gene. Transcription factors recruit co-factors such as the HAT enzyme Cbp/p300 that can further activate the enhancer. When activated, enhancers can loop to promoter regions. These loops are stabilized by the Ctcf insulator protein and the mediator co-activator complex (Kulaeva et al. 2012). Loop formation between an enhancer and the promoter mediates the activation of the gene through the action of the basal transcription machinery.

TFIID and TBP

The basal transcription factor complex TFIID comprises of approximately 13 different TFIID associated factors (Tafs) and the TATA binding protein (TBP). TFIID was first isolated from the phosphocellulose fractions of HeLa nuclear extracts eluting at high (1 M) salt concentrations. Several activities are attributed to this large complex such as core promoter binding, histone acetyltransferase- and kinase activity (Thomas and Chiang 2006). A subset of Tafs (Taf4, Taf5, Taf6, Taf9 and Taf12) is present in two copies and forms the symmetric core of the complex. Single copies of the other Tafs (Taf1, Taf2, Taf3, Taf7, Taf11, Taf13 and TBP) break this symmetry

and together form a clamp shaped complex (Bieniossek et al. 2013). Interestingly Taf9 and Taf10 are also found together with Taf like proteins (Taf5l, Taf6l) in the related human SAGA complex (Timmers and Tora 2005).

Within the PIC there are designated modules that can bind to the specific DNA sequences that were discussed before. One major contributor to PIC binding to DNA is TBP. This horseshoe shaped molecule is part of several transcription regulatory complexes and can bind to a stretch of TATA(A/T)A(A/T)(A/G) bases located just upstream of the TSS of many genes (TATA box element). Recent data derived from high-resolution chromatin immunoprecepitation (ChIP) in yeast revealed that genes with a degenerated TATA sequence in fact are highly enriched for the Taf1 TFIID subunit. Taf1 localizes upstream of the TSS and generally has a nucleosome positioned right in front. In contrast, TATA containing promoters display a lower nucleosome occupancy and TBP binding partially overlaps with the +1 nucleosome (Rhee and Pugh 2012). Very recently it was shown that on a *Drosophila* model promoter the structure of the TFIID complex changes by the interactions with both TFIIA and DNA (Cianfrocco et al. 2013). Furthermore, the binding of TFIID to the TATA element depends completely on the action of TFIIA, arguing for a stepwise locking of TFIID on the different promoter elements. Next to the TATA sequence, the basal machinery interacts with the core promoter through interactions of Taf6/9 with the DPE and Taf1/2 with the INR (Thomas and Chiang 2006).

Taf1, the largest subunit of the TFIID complex contains a double Bromodomain that can directly interact with acetylated histone tails (Jacobson et al. 2000). Using quantitative proteomics the TFIID complex was also found to bind to H3K4me3 peptides and this binding is enhanced by acetylation on adjacent H3 residues (Vermeulen et al. 2007) (summarized in Figure 4). The

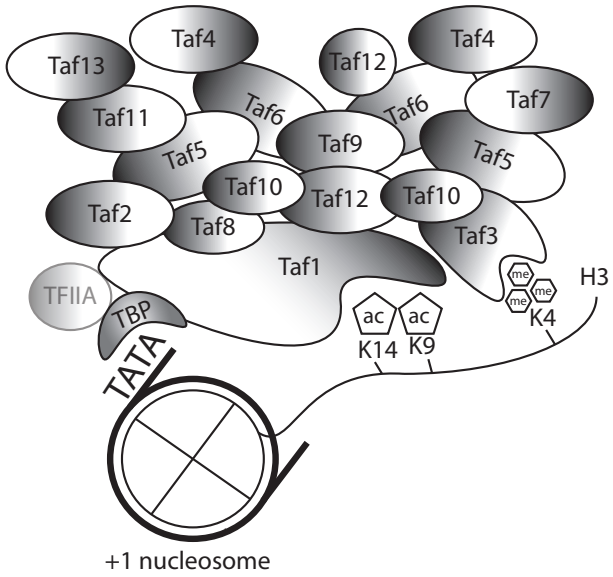


Figure 4. TFIID binding to multiple promoter features. Schematic representation of the TFIID complex binding to an active promoter nucleosome at the +1 position. TFIID associates with the TATA element through its TBP subunit. TBP/TATA-binding is stabilized by the TFIIA complex. Taf1 associates with acetylated lysines on the histone tails via its double Bromodomains. The PHD domain of Taf3 can bind directly to H3K4me3 chromatin.

PHD finger of Taf3 mediates this interaction by direct high affinity binding to the methylated lysine (van Ingen et al. 2008). Interestingly in yeast, Taf3 lacks a PHD finger domain despite the presence of H3K4me3 at active gene promoters. Additionally, under steady state conditions, mutation of lysine 4 on histone H3 or the removal of the methyltransferase in yeast has no effect at the basal level of mRNA expression (Margaritis et al. 2012). On the other hand, recent work also showed that H3K4me3 can directly affect transcription using an *in vitro* transcription system (Lauberth et al. 2013). Intriguingly, abolishment of the H3K4me3 binding only affects a subset of genes that lack a clear TATA element. These observations illustrate that this 'basal' transcription complex uses distinct binding modules to serve different genes (Fig. 4).

Outline of the thesis

Methylation of histone H3 at lysine 4 by SET1/MLL complexes marks the promoters of active genes and is important for recruitment of the basal transcription machinery to chromatin. Misregulation of gene expression programs, through mutations in the enzymes and co-factors involved in chromatin regulation can result in the development of diseases such as the MEN1 syndrome. In this thesis we aimed to understand the molecular function of the Menin/MLL complexes and their interacting proteins. Furthermore we studied the mechanisms by which proteins and protein complexes are recruited to modified chromatin.

The Menin/MLL complexes are recruited to gene promoters through the interaction with sequence specific transcription factors. In **chapter 2** we demonstrate that the vitamin D nuclear receptor interacts with the Menin/MLL complex and that its target genes are misregulated in MEN1 tumors. These results indicate that the loss of Menin/MLL function on vitamin D receptor targets can contribute to tumorgenesis in MEN1.

The SET1/MLL complexes consist of multiple subunits of which the molecular functions are far from understood. In **chapter 3** we applied an unbiased mass spectrometry based approach in which we differentially purify the complexes and determine the stoichiometry of the interacting proteins. These results improve our understanding of the molecular interactions within these complexes. Our approach led to the identification of a novel SET1B subunit and revealed that many of the SET1/MLL interacting proteins are sub-stoichiometric relative to the enzymatic core components.

In **chapter 4** we describe a detailed study on the MLL2 specific subunit Psip1. This sub-stoichiometric MLL component and co-factor for the HIV integrase carries a PWWP domain, that was previously shown to be essential for its function. The role of this domain in Psip1 however remained to be elucidated. We demonstrate that the Psip1-PWWP domain can bind to H3K36me chromatin and that this binding is strongly enhanced by electrostatic interactions with the nucleosomal DNA. Based on our data we now understand the specificity of PWWP domains for histone methylation sites close to the nucleosomal core.

The recruitment of the basal transcription machinery to target genes also depends on several interactions with DNA and histone modifications. We aimed to study the concerted action of these interactions using *in vitro* reconstituted designer nucleosomes, as is described in **chapter 5**. We demonstrate that TFIID binds directly to the nucleosome particle through multivalent synergistic interactions with H3K4me3, H3K14ac and specific DNA elements.

Taken together, in this thesis we unravel parts of the complex molecular interactions that play a role in chromatin regulation. These results improve our understanding on the molecular pathways that ultimately result in a transcriptional output. Furthermore, alterations in these pathways can seriously affect cell homeostasis and eventually lead to diseases such as cancer.

Chapter 2

Regulation of vitamin D receptor function in MEN1 related parathyroid adenomas

Koen M.A. Dreijerink, Radhika A. Varier, Rick van Nuland, Roel Broekhuizen, Gerlof D. Valk, Jacqueline E. van der Wal, Cornelis J.M. Lips, J. Alain Kummer and H.Th. Marc Timmers

Mol Cell Endocrinol. 2009 Dec 10;313(1-2):1-8

Abstract

Multiple endocrine neoplasia type 1 (MEN1) is heriditary syndrome characterised by the occurrence of parathyroid, gastroenteropancreatic and pituitary tumors. The *MEN1* gene product, Menin, co-activates gene transcription by recruiting histone methyltransferases for lysine 4 of histone H3 (H3K4). We investigated whether in MEN1 tumors global changes in H3K4 tri-methylation (H3K4me3) occur or whether alterations in gene expression can be observed. By immunohistochemistry we found that global levels of H3K4me3 are not affected in MEN1 related parathyroid adenomas. Menin can interact directly with the vitamin D receptor (VDR) and enhance the transcriptional activity of VDR. Messenger RNA levels of VDR target genes *CYP24* and *KLK6* were significantly lower in MEN1 parathyroid adenomas compared to normal tissue. Thus, aberrant gene expression in MEN1 tumors is not caused by lower global H3K4me3, but rather by specific effects on genes that are regulated by Menin-interacting proteins, such as VDR.

Introduction

Multiple endocrine neoplasia type 1 (MEN1) is an inherited syndrome that is characterised by the occurrence of tumors of the parathyroid glands, the pancreas and duodenum, the pituitary gland, the adrenal glands as well as neuroendocrine tumors of the thymus, lungs and stomach, often at a young age (Brandi et al. 2001). MEN1 is caused by germ line mutations in the *MEN1* gene (Chandrasekharappa et al. 1997). To date, more than 450 different germ line mutations have been identified, most of which are clearly inactivating (Lemos and Thakker 2008). The *MEN1* gene is a tumor suppressor gene: loss of the wild type allele is required for a cell to become a tumor cell (Larsson et al. 1988). The *MEN1* gene is expressed ubiquitously and encodes the protein Menin. Menin is localised in the nucleus and can take part in many cellular processes such as regulation of gene transcription, DNA repair and DNA replication (Lemos and Thakker 2008). Insight into Menin function was obtained by the identification of interacting proteins, which indicate a role in histone modification.

In each cell nucleus, roughly two metres of DNA are packaged into small units called nucleosomes. These nucleosomes consist of the core histone proteins (H2A, H2B, H3 and H4). Menin can influence gene transcription by regulating post-translational modification of the tails of these histone proteins. Menin has been shown to repress gene transcription by attracting histone deacetylase activity (Kim et al. 2003). More recently, Menin was found to be an integral component of complexes that contain members of the mixed-lineage leukaemia (MLL) family (Hughes et al. 2004; Yokoyama et al. 2004). MLL proteins possess methyltransferase activity specifically directed at lysine 4 of histone H3 (H3K4), and especially tri-methylation of this residue (H3K4me3) (Ruthenburg et al. 2007). Menin is part of the MLL1 and MLL2, but not of the MLL3 or MLL4 complexes (Lee et al. 2006). The Menin-MLL1 histone methyltransferase (HMT) complex was found to be important for β-catenin regulated transcription of the *c-Myc* gene and is stabilised by the chromatin associated protein lens epithelium derived growth factor (Ledgf) or Psip1 (Sierra et al. 2006; Yokoyama and Cleary 2008). The Menin-MLL1 complex can also activate the expression of several other genes involved in cell proliferation such as the *CDKN2C* and *CDKN1B* cyclin dependent kinase inhibitor genes and genes involved in cell differentiation such as several homeoboxdomain (*Hox*) genes (Yokoyama et al. 2004; Hughes et al. 2004; Milne et al. 2005). It has been shown that loss of Menin can lead to reduced H3K4me3 at these specific target genes (Karnik et al. 2005; Dreijerink et al. 2006). To date, no study has addressed global H3K4me3 levels in MEN1 related tumors.

The receptor for activated vitamin D ($1,25(OH)_2D_3$; calcitriol) is a member of the nuclear hormone receptor family. In the presence of ligand, the vitamin D receptor (VDR) can bind to vitamin D responsive DNA elements and regulate transcription of target genes. In parathyroid cells, $1,25(OH)_2D_3$ inhibits parathyroid hormone gene transcription and parathyroid hormone secretion (Demay et al. 1992). Sequences in the human parathyroid hormone gene that bind

the 1,25-dihydroxyvitamin D3 receptor and mediate transcriptional repression in response to 1,25-dihydroxyvitamin D3. Furthermore, $1,25(OH)_2D_3$ negatively regulates parathyroid cell proliferation (Bikle 2009). Low $1,25(OH)_2D_3$ levels are associated with secondary hyperparathyroidism in patients with renal failure. Several reports suggest that aberrant VDR function can contribute to parathyroid adenoma formation. VDR null or 1α-hydroxylase null (unable to synthesize $1,25(OH)_2D_3$) mice develop parathyroid hyperplasia (Bouillon et al. 2008). VDR gene polymorphisms are associated with primary hyperparathyroidism and alterations of VDR mRNA and protein levels have been reported to occur in parathyroid adenomas (Carling et al. 1995; 2000; Sudhaker Rao et al. 2000). Parathyroid adenomas are the most common manifestation of MEN1, with a penetrance of almost 100%. Moreover, germ line mutations of the *MEN1* gene have been found in patients with familial isolated hyperparathyroidism (FIHP) (Miedlich et al. 2001).

We have previously reported that Menin can co-activate nuclear receptor mediated gene transcription (Dreijerink et al. 2006). Reduction of Menin levels led to decreased expression of estrogen receptor alpha (ERα) and peroxisome proliferator-activated receptor gamma (PPARγ) target genes (Dreijerink et al. 2006; 2009). To investigate how loss of Menin function can lead to aberrant gene expression in MEN1 tumors, we first determined global H3K4me3 staining in parathyroid tumors. To investigate whether expression of specific Menin-HMT target genes is affected in MEN1 tumors, we analysed the interaction between Menin and VDR and the expression of VDR target genes in MEN1 parathyroid adenomas.

Results

Analysis of global H3K4me3 levels in MEN1 versus sporadic parathyroid adenomas

Observations that Menin is an integral part of HMT complexes for H3K4 suggest that aberrant gene expression in MEN1 tumors can either be the result of a general loss of H3K4me3 activity or of target gene specific effects. To determine whether global H3K4me3 levels are affected in MEN1 tumors, we performed immunohistochemical analysis of sections of sporadic parathyroid adenomas and parathyroid tumors from MEN1 patients. Whereas MEN1 tumors invariably contain mutations in the *MEN1* gene, only ~20% of the sporadic parathyroid tumors harbor lesions in *MEN1* (Heppner et al. 1997).

As expected we found that compared to normal parathyroid tissue, Menin levels were reduced in all four MEN1 tumors. In contrast, Menin expression was not affected in most sporadic adenomas with only two samples showing a partial Menin down regulation (Table 1). However, we found no global loss of H3K4me3 levels. Total histone H3 levels that served as an internal control were also found to be equal among the samples as expected (Table 1). Interestingly, in one sporadic adenoma (Table 1: sample S6), we found areas within the tumor that showed complete loss of Menin expression while H3K4me3 staining was similar throughout the tumor

(Fig. 1). Together, these observations indicate that global H3K4me3 is not affected in MEN1 related parathyroid adenomas.

Menin can co-activate VDR mediated transcription via a direct ligand-independent interaction

Vitamin D signalling is an important regulatory mechanism in the parathyroid gland. As MEN1 patients frequently develop parathyroid adenomas, we were interested to study whether Menin can regulate VDR function. For this purpose, luciferase reporter assays were performed. Cos7 cells were transiently transfected with increasing amounts of MEN1 expression vector together with the expression vector for a fusion protein containing the ligand-binding domain of VDR fused to the DNA binding domain of the yeast Gal4 activator. A thymidine kinase (TK) promoter-luciferase construct with five Gal4 binding sites was used as a reporter gene. Similar to our previous observations, overexpression of Menin increased ligand-induced transcription to almost threefold, depending on the amount of Menin that was co-transfected (Fig. 2A) (Dreijerink et al. 2006). VDR binds to its response elements as a heterodimer with the retinoid X receptor (RXR) (Carlberg et al. 1993). To determine if Menin can also interact with RXR, additional luciferase experiments were performed. We found that Menin can indeed also co-activate RXRα mediated transcription in the presence of ligand (Fig. 2B).

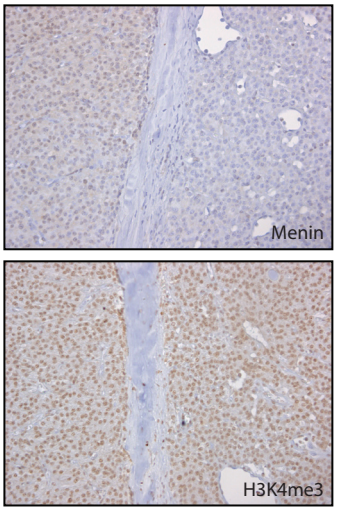


Figure 1. Loss of Menin expression in a part of a sporadic parathyroid adenoma does not lead to reduced H3K4me3. Immunohistochemistry was performed on sections of a sporadic parathyroid adenoma using antibodies directed at Menin (upper panel, compare left and right) and H3K4me3 (lower panel, compare left and right). Total histone H3 levels did not differ between the Menin positive and negative areas (not shown).

Tissue	Menin	H3K4me3	H3
N1	+	+	+
N2	+	+	+
S1	+	+	+
S2	+	+	+
S3	+	+	+
S4	+	+	+
S5	+	+	+
S6	±	+	+
S7	±	+	+
M1	±	+	+
M2	±	+	+
M3	-	+	+
M4	-	+	+

Table 1. Immunohistochemical analysis of sporadic (S) and MEN1 related (M) parathyroid adenomas, compared to normal parathyroid tissue (N).

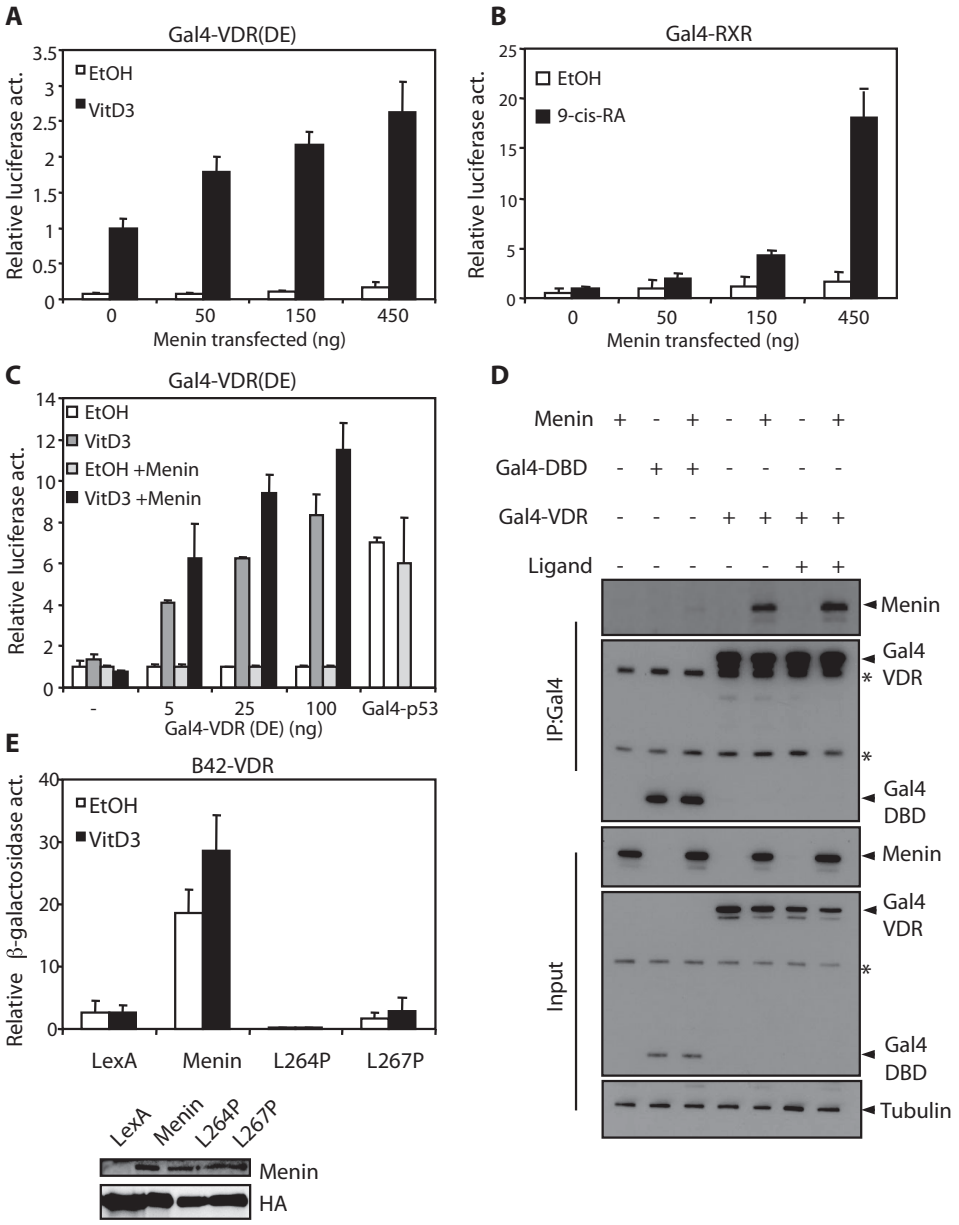


Figure 2. Menin can co-activate VDR and RXR mediated transcription and interact directly with VDR. (A) Increasing amounts of a MEN1 expression construct were co-transfected together with a Gal4-VDR(DE) construct and a Gal4-TK-driven firefly luciferase reporter vector in Cos7 cells. Cells were treated with 100 nM $1,25(OH)_2D_3$ or vehicle for 24 hours. A CMV promoter-driven Renilla luciferase construct was co-transfected for normalisation. Results are shown as Firefly/Renilla luciferase ratios. Error bars indicate SD. (B) Menin was co-transfected together with a Gal4-RXRα construct in luciferase assays similar to panel A. Cells were treated with 1 μM 9-cis-retinoic acid or vehicle for 24 hours. (C) Luciferase assays using Gal4-VDR(DE) and Gal4-TK luciferase constructs were performed in MEN1 /- MEFs that had either been transduced with a control or with a MEN1 expression virus. Gal4 and Gal4-p53 were used as controls. Error

In addition to overexpression experiments we investigated the impact of loss of Menin on VDR function. To this end we established derivatives of *MEN1* /- mouse embryonic fibroblasts (MEFs) stably transduced with retroviruses containing *MEN1* or control sequences. We conducted luciferase assays in these cells using increasing amounts of the Gal4-VDR expression plasmid and the TK promoter luciferase construct. Luciferase assays using expression constructs for Gal4 or Gal4-p53, which have not been reported to interact with Menin, were included as controls. Expression levels of the Gal4 constructs were equal between the cell lines as assessed by immunoblotting (data not shown). These transient luciferase assays showed that ligand dependent VDR function was attenuated in the *MEN1* /- cells compared to the *MEN1* re-expressing cells (Fig. 2C). As expected Gal4 and Gal4-p53 dependent transcription did not differ between these cell lines. These results indicate that loss of Menin results in lower VDR activity.

To investigate whether Menin and VDR can interact directly, we performed co-immunoprecipitation experiments in mammalian cells. Menin and Gal4-VDR expression constructs were co-transfected and cells were treated with $1,25(OH)_2D_3$ or vehicle. Fig. 2D shows that Menin is co-immunoprecipitated with VDR both in the presence and absence of ligand. To strengthen these observations and to study the effects of disease-related *MEN1* gene mutations on the interaction yeast two-hybrid experiments were performed using expression constructs for B42-fused VDR and LexA-coupled Menin. Figure 2E confirms that Menin and VDR can interact in a ligand-independent manner. Two patient derived *MEN1* gene mutations L264P (from a MEN1 family) and L267P (from a FIHP family) caused a clear reduction of the yeast two-hybrid interaction between Menin and VDR (Poncin et al. 1999). In conclusion, Menin can co-activate VDR mediated transcription, which most likely involves a direct interaction between Menin and VDR proteins.

Expression of VDR target genes in MEN1 parathyroid adenomas

Parathyroid adenoma is the most common manifestation of MEN1. VDR has an important role in parathyroid function. Having established that Menin can act as a VDR co-activator and that *MEN1* gene mutations can lead to disruption of the Menin-VDR interaction, we reasoned that in parathyroid adenomas from MEN1 patients VDR target gene expression may be affected. The only known target gene for VDR in parathyroid adenomas is the $1,25(OH)_2D_3$ inactivation enzyme 25-hydroxyvitamin D_3-24-hydroxylase (*CYP24*) (Correa et al. 2002). A large set of VDR

bars indicate SD. EV: empty vector. (D) Co-immunoprecipitation of Gal4-VDR and Menin. 293T cells were transiently transfected with Menin expression constructs and Gal4-VDR (in the presence or absence of 100 nM $1,25(OH)_2D_3$) or Gal4-DBD control and Gal4 antibodies were used for immunoprecipitation. Immunoblot analysis of input and anti-Gal4 immunoprecipitates are shown. Asterisks indicate non specific bands detected by the Gal4-antibody. (E) Yeast two-hybrid experiments were performed using a B42 fusion proteins with hVDR and LexA-proteins fused to wild type Menin or two disease-related MEN1 mutations L264P and L267P. Yeast cells were treated with 100 nM $1,25(OH)_2D_3$ or vehicle for 12 hours. Relative β-galactosidase activities are shown. Error bars indicate SD. Immunoblot analysis of yeast lysates shows equal expression of the LexA-Menin fusion proteins and of B42-HA-VDR.

target genes was identified in a large-scale microarray study in SCC25 oral tumor cells (Wang et al. 2005). We screened inducibility of these genes by 1,25$(OH)_2D_3$ in MCF-7 cells, a cell type in which Menin was demonstrated to be important for nuclear receptor function (Dreijerink et al. 2006). We found that besides *CYP24*, one other gene called kallikrein 6 (*KLK6*) could be induced (data not shown). *KLK6* is a serine protease that is expressed in the parathyroid gland (Petraki et al. 2001). Two other possible target genes that could not be induced in MCF-7 cells, semaphorin 3B (*SEMA3B*) and arachidonic acid lipoxygenase 5 (*ALOX*) were also included in the analysis, as well as the *VDR* gene itself and the *PTH* gene, which is negatively regulated by 1,25$(OH)_2D_3$ (Demay et al. 1992). We found that mRNA levels of *CYP24* were lower but not significantly reduced in sporadic parathyroid adenomas compared to normal parathyroid tissue (P=0.09) (Fig. 3A). However, in the MEN1 parathyroid adenomas the presence of *CYP24* and *KLK6* mRNA is significantly lower (P=0.03, P=0.04 respectively) (Fig. 3A). In this context, the *SEMA3*, *ALOX* , *PTH* and *VDR* mRNA levels did not differ (Fig. 3B).

The *CDKN2C* and *CDKN1B* cyclin dependent kinase inhibitor genes encoding $p18^{INK4c}$ and $p27^{Kip1}$ are well established Menin target genes (Milne et al. 2005). Cyclin dependent kinase inhibiting

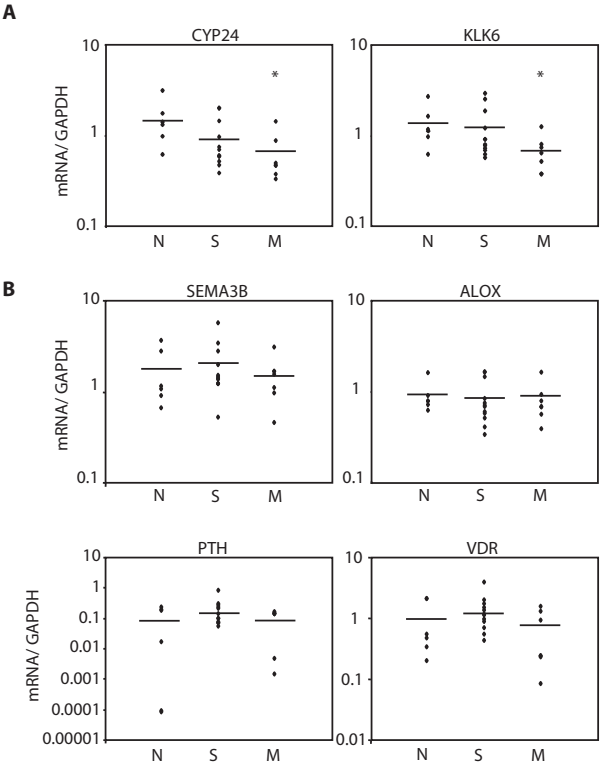


Figure 3. Analysis of expression of VDR target genes in 5 normal parathyroid tissues (N), 12 sporadic adenomas (S) and 7 MEN1 related adenomas (M). (A) Expression of CYP24 and KLK6 gene mRNA relative to GAPDH mRNA levels was assessed by quantitative RT-PCR. Means are indicated by horizontal bars. * indicates statistical significance (P<0.05) compared to normal parathyroid tissue using a two-sided student's t-test. (B) Messenger RNA levels of SEMA3B, ALOX, PTH and VDR.

proteins exert anti-proliferative effects by blocking cyclin dependent kinase regulated cell cycle progression. These genes have all been reported to be regulated by VDR. $1,25(OH)_2D_3$ was shown to induce $p18^{INK4c}$ protein levels in a squamous cell tumor cell line (Gedlicka et al. 2006). Saramaki *et al.* showed that VDR can be present on the *CDKN1A* (encoding $p21^{Waf1/Cip1}$) promoter on VDR responsive elements (Saramäki et al. 2006). Cheng *et al.* have reported that VDR can bind to the *CDKN1B* promoter via the Sp1 transcription factor (Cheng et al. 2006). Previously, Correa Buchwald *et al.* reported reduced *CDKN2*C, *CDKN1A* and *CDKN1B* mRNA levels in sporadic parathyroid adenomas (Buchwald et al. 2004). We also performed RT-PCR analysis of mRNA levels of these CDKI genes in our series of sporadic and MEN1 parathyroid adenomas. We found that *CDKN2C* levels were not altered, *CDKN1A* mRNA was expressed significantly lower in sporadic tumors (P=0.004) and *CDKN1B* was significantly reduced both in sporadic and in MEN1 related parathyroid adenomas compared to normal parathyroid tissue (P=0.004 and P=0.01)(Fig. 4).

Taken together, we found *CYP24* and *KLK6* mRNA levels are reduced in parathyroid adenomas in MEN1 patients. *CDKN1A* mRNA was lower in sporadic adenomas and *CDKN2C* mRNA levels were attenuated in both sporadic and MEN1 related adenomas.

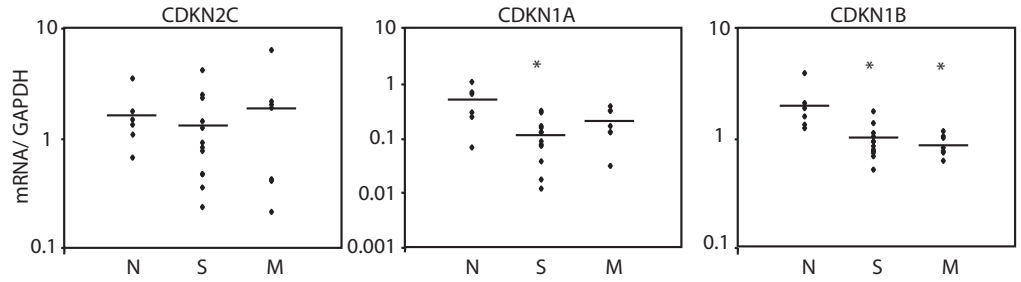


Figure 4. Messenger RNA levels of cyclin dependent kinase inhibitor genes differ between normal parathyroid tissue and adenomas. In 5 normal parathyroid samples (N), 12 sporadic parathyroid adenomas (S) and 7 MEN1 related adenomas (M), mRNA levels of the CDKI genes CDKN2C, CDKN1A and CDKN1B were measured using quantitative RT-PCR relative to GAPDH. * indicates statistical significance (P<0.05) as in figure 3.

Discussion

In this report, we show that global H3K4me3 levels are not affected in MEN1 parathyroid adenomas. Further, our results show that Menin can act as a co-activator for VDR mediated transcription. Messenger RNA levels of several VDR target genes are reduced in MEN1 patient derived parathyroid adenomas.

Loss of *MEN1* does not affect global H3K4me3

Global loss of H4K20 tri-methylation and H4K16 acetylation has been observed in various tumors (Fraga et al. 2005). Distinct global alterations of histone modifications were reported to

be prognostically relevant in prostate carcinomas (Seligson et al. 2005). H3K4 tri-methylation levels however, were not assessed in these studies. Since there are many methyltransferase complexes, we hypothesized that it is not likely that in the absence of Menin in a cell global loss of H3K4 tri-methylation occurs (Ruthenburg et al. 2007). Indeed, transient knockdown of Menin by siRNA oligos in MCF-7 breast cancer cells did not lead to loss of H3K4me3 (Dreijerink et al. 2006). However, in pancreatic islets from *MEN1*+/- mice, Karnik *et al.* showed reduced total H3K4me3 levels compared to *MEN1*+/+ mice (Karnik et al. 2005). Our results indicate that loss of Menin expression does not affect global H3K4me3 staining in parathyroid tumors. Although these effects could be tissue specific, we believe that it is more likely that the changes in gene expression which have been observed *in vitro* and in MEN1 tumors are the result of gene specific actions of Menin. This notion is supported by a recent report that showed MEN1 tumor specific alterations in *Hox* gene expression compared to non-MEN1 parathyroid adenomas (Shen et al. 2008). This target gene specificity may be the result of the affinity of the Menin HMT complex for other transcriptional regulators such as nuclear receptors, b-catenin or other chromatin binding factors such as Psip1 (Sierra et al. 2006; Yokoyama and Cleary 2008).

Menin co-activates VDR mediated transcription

We have shown previously that Menin can stimulate ERα and PPARγ mediated transcription and that Menin is important for recruiting methyltransferase activity to ERα target genes (Dreijerink et al. 2006; 2009). In this report, both in overexpression expression experiments and in *MEN1* -/- MEFs, we show that Menin can also co-activate VDR mediated transcription. Although we did not analyse the mechanism for VDR co-activation in much detail, but based on analogy with ERα and PPARγ regulation we suggest that recruitment of Menin by the activated VDR is responsible for elevated H3K4me3 levels of VDR target promoters. Interestingly, both immunoprecitation experiments and yeast two-hybrid studies show that the interaction between Menin and VDR is largely ligand-independent similar to the interaction between Menin and PPARγ. In contrast, Menin acts in a ligand dependent fashion with ERα and RXRα (Dreijerink et al. 2006; 2009). Further structural analysis is required to be able to determine the exact nature of the Menin-nuclear receptor interaction.

Altered expression of VDR target genes in MEN1 parathyroid adenomas

We analysed a series of MEN1 patient derived parathyroid adenomas to find that expression of two VDR target genes *CYP24* and *KLK6* was significantly lower in MEN1 tumors. In sporadic adenomas slightly lower levels of *CYP24* mRNA (but not significant P=0.09) were observed. Previously, Correa *et al.* reported a significant reduction in *CYP24* mRNA levels in sporadic parathyroid adenomas (Correa et al. 2002). The size of our series may have been a limiting factor for showing reduced *CYP24* in the sporadic adenomas. Therefore, reduced expression of VDR target genes may not be restricted to MEN1 parathyroid tumors. As *MEN1* gene mutations

are also found in about 20% of sporadic parathyroid adenomas, MEN1 related and sporadic adenomas may share a common pathway of development (Heppner et al. 1997). A large study aimed to define mRNA expression profiles that could differentiate between several types of parathyroid adenomas showed no differences between sporadic and MEN1 related tumors (Haven et al. 2004). However, the *CYP24*, *KLK6* and CDKI genes were not included in that analysis (Carola Haven, Paul Eilers and Hans Morreau, personal communication).

Interestingly, expression of the *PTH* gene that is negatively regulated by VDR was not altered in the adenomas. It is known that Menin can be involved in TGF-β mediated down regulation of PTH production in cell lines, which makes it even more unlikely that the activating Menin-HMT complex is present on the *PTH* gene promoter (Sowa et al. 2004).

There have been conflicting reports on the expression of VDR in parathyroid adenomas. Initial reports showed that vitamin D uptake was higher in parathyroid adenomas compared to cases of secondary hyperplasia and that VDR levels were similar in adenomas compared to normal tissue (Korkor 1987; Trydal et al. 1992). More recent studies showed reduced levels of VDR mRNA in parathyroid adenomas and reduced immunostaining (Carling et al. 2000; Sudhaker Rao et al. 2000). In our series, mean VDR mRNA levels (as determined by quantitative RT-PCR) did not differ significantly between the three groups. Nevertheless, expression of the VDR target genes *CYP24* and *KLK6* was found reduced in MEN1 adenomas. We analysed mRNA levels of the several CDKI genes. In sporadic adenomas we found reduced expression of *CDKN1A* and *CDKN1B*. In MEN1 adenomas, we found only reduced expression of *CDKN1C.* These findings are in accordance with a recent report by Correa Buchwald *et al.*, who found that in sporadic parathyroid adenomas *CDKN1C* was significantly down regulated (Buchwald et al. 2004). Although the CDKI genes have been described as VDR targets, in our MCF-7 system expression of these genes was not induced after a six hour stimulation with $1,25(OH)_2D_3$ (data not shown). It thus remains unclear if the reduced transcription of these genes is related to altered VDR activity in parathyroid adenomas.

These are the first data to support a role for aberrant nuclear receptor function in MEN1 tumor formation. Profiling of nuclear receptor expression in other MEN1 lesions will lead to better understanding of MEN1 tumorigenesis. As nuclear receptors are potential targets for novel drugs, restoration of nuclear receptor function or of histone modifications may prove beneficial for future therapeutic strategies for MEN1 patients.

Materials & Methods

Patient material

Normal parathyroid tissue and parathyroid adenoma samples were obtained from the Department of Pathology/UMCU biobank and from the Department of Pathology/UMCG and used in accordance with the hospital scientific committee regulations, the Declaration of

Helsinki, and the code "Proper Secondary Use of Human Tissue" as installed by the Federation of Biomedical Scientific Societies (www.federa.org/?s=1&m=78&p=&v=4).

Plasmids and mutagenesis

Construction of vectors pXJ440hVDR, pXJ440hVDR(DE), Gal4-RXRα, Gal4-DBD, pEG202NLS-Menin and pBabeHygroMenin has been published (Lavigne et al. 1999; Pijnappel et al. 1993; Zwartjes et al. 2004; Hughes et al. 2004; Dreijerink et al. 2006). The Menin expression vector pCDNA3.1M+ was a kind gift from G. Weber. The LexA-Menin L264P and L267P mutations were introduced by site-directed mutagenesis using the Quikchange procedure. All constructs were verified by DNA sequence analysis of the complete Menin cDNA sequence. The B42-hVDR expression plasmid was constructed in the pJG4-5 vector by PCR amplification of the VDR cDNA, from vector pXJ440hVDR using oligos hVDR-EcoR1-F: GATCATCCGAATTCATGGAGGCAATGGCGGCCA and hVDR-Xho1-R: GATCATCCCTCGAGTCAGGAGATCTCATTGCCAAA, and inserted using *Eco*RI and *Xho*I.

Antibodies and immunoblotting

Antibodies used were: α-Menin (Bethyl, A300-105A), α-H3K4me3 (Abcam, ab-8580), α-histone H3 (Abcam, ab-1791), α-Gal4 (Santa Cruz, SC-510) and α-HA (Roche, 3F10). Immunoblotting was carried out as described before (Dreijerink et al. 2006).

Cell lines, stable and transient protein expression

Cos7 (African green monkey kidney) cells and 293T (human embryonic kidney) cells were routinely cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin. Generation of *MEN1*-/- (MEN1T/T) mouse embryonic fibroblasts (MEFs) has been described (Bertolino et al., 2003). MEFs were maintained in DMEM, containing 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin and 100 mM β-mercaptoethanol. For Menin re-expression studies, *MEN1*-/- MEFs were infected with viruses obtained from Phoenix cells transiently transfected with pBabeHygro or pBabeHygroMenin. Cells were grown in medium containing 500 μg/ml hygromycin and monoclonal cell lines were isolated.

Luciferase reporter experiments in Cos7 cells were carried out as described (Dreijerink et al. 2006). In summary, DNA (750 ng per well, in a 12-well format) was transfected using FuGene 6 reagent (Roche). Transfection mixtures consisted of 200 ng luciferase reporter, 25 ng pCMV-*Renilla*, 5 ng pXJ440hVDR(DE), 25 ng Gal4-RXRa, and supplemented with a maximum of 520 ng pCDNA3.1M+ and/or empty pCDNA3 plasmid. 24 hrs after transfection, the medium was changed to medium containing 100 nM $1,25(OH)_2D_3$ (in ethanol), 1 μM 9-cis-retinoic acid (in ethanol) or vehicle. Luciferase and *Renilla* activities were measured after 24 hours. Luciferase assays in MEFs were carried out using Effectine transfection reagent (Qiagen). Transfection

mixtures (325 ng per well, in a 12-well format) consisted of 200 ng luciferase reporter, 25 ng pCMV-Renilla, Gal4 DNA, or increasing amounts of pXJ440hVDR(DE), or 50 ng Gal4-p53, supplemented with pCDNA3 plasmid. 24 hours after transfection, the medium was changed to medium containing vehicle or 100 nM $1,25(OH)_2D_3$. 24 hours later, cells were lysed and luciferase and *Renilla* activities were measured. To be able to compare the results from the two cell lines, results are shown as relative luciferase activities corrected for the ligand-independent signal at the appropriate concentration of Gal4-VDR, to show the fold induction. The Gal4-p53 result was corrected for the Gal4 control experiment.

Immunohistochemistry

Formalin-fixed (10% (v/v) formalin for 18 hours), paraffin-embedded tissues from 2 normal parathyroid glands, 7 sporadic (non-familial) and 4 parathyroid adenomas from MEN1 patients were used. All tissues were sampled from surgical specimens within 2 hours after resection. Immunohistochemical staining was essentially performed as previously described using rabbit polyclonal antibodies against Menin (dilution 1:100), against H3K4me3 (1:7,500) and against histone H3 (1:15,000) (Strik et al. 2002). The optimal dilution for each antibody was determined by end-point titrations on tissue slides of human tonsil. To ensure specific staining the highest dilution was chosen for each antibody, which still yielded a clear nuclear staining. The tissue sections were subjected to antigen retrieval by boiling in 10 mM sodium citrate buffer pH 6 for 15-20 minutes in a microwave oven. Bound antibodies were visualized with 3,3'-diaminobenzidine (0.1 mg/ml in 0.02% H_2O_2) as the chromogen. Negative control slides were stained with non-immune rabbit IgGs.

Normal, diminished or loss of Menin, H3K4me3 and Histone H3 staining was determined independently by two experienced pathologists comparing staining intensity of neoplastic cells within the adenoma to the non-neoplastic neuroendocrine cells present in the same slide (often at the rim of the lesion) and to other non-neoplastic cells within the lesion (endothelial and inflammatory cells). The intensity of the staining was scored as: similar (+), weaker (+/-), or negative (-) when the majority of cells (>75%) showed such staining intensity.

Co-immunoprecipitations

293T cells in 60 mm dishes were transiently transfected with 1 µg of pCDNA3.1M+ or pCDNA3 and 2 µg of pXJ440hVDR or 6 µg of Gal4-DBD. After 24 hours the medium was changed to medium containing vehicle or 100 nM $1,25(OH)_2D_3$. After another 24 hours cells were lysed in buffer A (50 mM Tris-HCl pH 8, 10% glycerol, 100 mM NaCl, 10 mM $MgCl_2$, 0.5 mM PMSF, 1% protease inhibitor cocktail (Sigma), 1 mM NaF, 1 mM Na_3VO_4, and 0.5 mM DTT). Dynabeads Protein A were incubated with 2 µg of anti-Gal4 antibody in buffer A for 1 hr at 4˚C and then washed 3 times in buffer A before incubating with lysate for 3 hrs at 4˚C. Subsequently the beads were washed 3 times in buffer A and eluted in sample buffer and proteins retained by the

anti-Gal4 antibody visualised by immunoblot analysis.

Yeast two-hybrid analysis

EGY48 cells were transformed with the B42-VDR construct and the indicated LexA-Menin constructs. Cells were grown overnight at 30˚C in 2% galactose-1% sucrose containing SC medium lacking the appropriate amino acids and in the presence of vehicle or 100 nM $1,25(OH)_2D_3$. Lysates were prepared and the LacZ activity was determined by a liquid β-galactosidase assay as described previously (Albert et al. 2002).

Analysis of VDR target gene expression

Frozen sections from 6 normal parathyroid tissue samples, 12 sporadic parathyroid adenomas and 7 parathyroid adenomas from MEN1 patients were used for this analysis. This set was unrelated to the paraffin samples used for immunohistochemistry. After homogenization, RNA from human samples was isolated using the QIAzol lysis reagent (Qiagen). 200 ng of total RNA was used for cDNA synthesis using oligo(dT). Messenger RNA levels were analyzed by quantitative PCR on a Chromo4-equipped PCR cycler (Bio-Rad) and normalized against a standard reference cDNA. To ensure identical PCR reaction conditions, for every gene, all PCR reactions from the normal and adenoma samples were performed simultaneously on one 96-well plate. For non-intron spanning primer pairs, PCR reactions were performed on RNA that had not been treated with reverse transcriptase as a control. Primer sequences used are: hCYP24-F: CTACCGCAAAGAAGGCTACG, hCYP24-R: TCATCACTTCCCCTGGTTTC, hKLK6-F: GATGGTGGTGCTGAGTCTGA, hKLK6-R: GAGGTGTAGAGGGCAGCTTG, hSEMA3B-F: ATGCACCAGTGGGTGTCATA, hSEMA3B-R: TTGTACATGAGGGGGTGGTT, hALOX-F: TCATCGTGGACTTTGAGCTG, hALOX-R: GACAATCTTGTTGGCCAGGT, hPTH-F: TAGCTCCCAGAGATGCTGGT, hPTH-R: AGCTTTGTCTGCCTCTCCAA, hVDR-F: CCAGTTCGTGTGAATGATGG, hVDR-R: AGATTGGAGAAGCTGGACGA, hCDKN2C-F: ACGTCAATGCACAAAATGGA, hCDKN2C-R: CGAAACCAGTTCGGTCTTTC, hCDKN1A-F: GGAAGACCATGTGGACCTGT, hCDKN1A-R: GGCGTTTGGAGTGGTAGAAA, hCDKN1C-F: CCGGCTAACTCTGAGGACAC, hCDKN1C-R TTGCAGGTCGCTTCCTTATT, hGAPDH-F: TCTGGTAAAGTGGATATTGTTG, hGAPDH-R: GATGGTGATGGGATTTCC. Mean mRNA levels in sporadic and MEN1 tumors were compared to normal parathyroid samples using a two-sided student's t-test. $P<0.05$ was considered statistically significant.

Acknowledgements

We thank the members of the Timmers laboratory, especially F. van Werven, for advice. We thank the UMC-U Biobank (A. Gijsbers-Bruggink) for assistance in collecting human samples. *MEN1-/-* MEFs were kindly provided by C. Zhang (University of Lyon, France). We are grateful to G. Folkers (Utrecht University, The Netherlands), G. Mengus (IGBMC, Strasbourg, France) and G. Weber (University of Tours, France) for providing plasmids and to C. Haven, P. Eilers and H. Morreau, (Leiden University Medical Center, The Netherlands) for sharing microarray data.

This work is supported by the Netherlands Organization for Health Research and Development (ZonMw; AGIKO 920-03-231 to CL, HT and KD) and the Netherlands Proteomics Center (to HT and KD).

Chapter 3

Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes

Rick van Nuland[#], Arne H. Smits[#], Paschalina Pallaki, Pascal W.T.C. Jansen, Michiel Vermeulen and H.Th. Marc Timmers

[#] *These authors contributed equally*

Mol Cell Biol. 2013 May;33(10):1067-77

Abstract

Methylation of lysine 4 on histone H3 (H3K4) at promoters is tightly linked to transcriptional regulation in human cells. At least six different COMPASS-like multi-subunit (SET1/MLL) complexes have been described that contain methyltransferase activity towards H3K4, but a comprehensive and quantitative analysis of these SET1/MLL complexes is lacking. We applied label free quantitative mass spectrometry to determine the subunit composition and stoichiometry of the human SET1/MLL complexes. We identified both, known and novel, unique and shared, interactors and determined their distribution and stoichiometry over the different SET1/MLL complexes. In addition to being a core COMPASS subunit, the Dpy30 protein is a genuine subunit of the NURF chromatin remodeling complex. Furthermore, we identify the Bod1 protein as a discriminator between the SET1B and SET1A complexes and we show that the H3K36me-interactor Psip1 preferentially binds to MLL2 complex. Finally, absolute protein quantification in crude lysates mirrors many of the observed SET1/MLL complex stoichiometries. Our findings provide a molecular framework to understand the diversity and abundance of the different SET1/MLL complexes, which together establish the H3K4 methylation landscape in human cells.

Introduction

The basic repeating unit of chromatin in eukaryotic cells constitutes of ~147 base pairs of DNA wrapped around an octamer of histone proteins to form the nucleosome core particle (Luger et al. 1997). These histone proteins are subject to post-translational modifications (PTMs), such as methylation, acetylation, phosphorylation and ubiquitination (Jenuwein and Allis 2001; Berger 2007). In human cells, nucleosomes and their PTMs are involved in regulation of virtually all DNA associated processes such as transcription, replication and response to DNA damage (Fischle et al. 2003; Kouzarides 2007). Methylated lysines and arginines are known to recruit effector proteins to specific genomic loci to impose their specific regulatory function upon the underlying DNA (Jenuwein and Allis 2001). Methylation on histone H3 at lysine 4 (H3K4) is conserved from yeast to humans and is tightly linked to the transcription of genes by RNA polymerase II (Bernstein et al. 2005). Whereas tri-methylation of H3K4 (H3K4me3) primarily marks promoters of actively transcribed genes, mono-methylation (H3K4me1) in combination with H3K27 acetylation has recently been established as a hallmark of active enhancers (Heintzman et al. 2007; 2009). H3K4me3 can be recognized by PHD finger containing proteins such as the Bptf subunit of the NURF chromatin remodeling complex and the Taf3 subunit of the basal transcription factor TFIID, thereby recruiting the basal transcription machinery to activated promoters (Vermeulen et al. 2007; Wysocka et al. 2006).

In yeast cells, the Set domain-containing protein Set1p is the only methyltransferase for H3K4. Set1p together with other proteins (Cps25/Sdc1p, Cps30/Swd3p, Cps35/Swd2p, Cps40/Spp1p, Cps50/Swd1p, Csp15/Shg1p and Cps60/Bre2p) assembles into the Set1/COMPASS complex (Miller et al. 2001). In contrast, higher eukaryotes contain at least six COMPASS-like complexes with H3K4 methyltransferase activity. These complexes are distinguished by six different catalytic Set domain proteins (Set1a, Set1b, Mll1, Mll2, Mll3, Mll4; named HMTs for histone methyltransferases hereafter) (Shilatifard 2012). The SET1A and SET1B complexes are responsible for maintaining global levels of H3K4me3 (Wu et al. 2008), whereas complexes with the mixed lineage leukemia proteins (Mll1-4) display gene specificity. Interestingly, the Trr protein of fruit flies (and by analogy mammalian Mll3/4) has recently been found to be critical for H3K4me1 (Herz et al. 2012). *SET1/MLL* gene deletion studies in mice revealed diverse non-overlapping phenotypes, which indicate that these genes perform non-redundant functions during development (Eissenberg and Shilatifard 2010). The six distinct SET1/MLL complexes share a conserved core consisting of Wdr5, Rbbp5, Ash2l and Dpy30 (named WRAD hereafter). This WRAD module can associate with the catalytic subunit and has been implicated in regulating its enzymatic activity (Patel et al. 2011; Steward et al. 2006; Dou et al. 2006). The specific subunits such as the Menin subunit of MLL1/2 and the Ptip subunit of MLL3/4, however, have been shown to direct these distinct complexes to certain genomic loci (Wang et al. 2009).

The exact subunit composition and abundance of the SET1/MLL complexes in human cells

is unknown at present, which complicates assessment of their contributions in establishing and maintaining methylation of H3K4. Recent developments in label free quantitative mass spectrometry based interaction proteomics (Hubner et al. 2010) and the application of a novel method for quantifying the stoichiometry of these interactions (Smits et al. 2012), enables a careful analysis of protein complex composition in a quantitative manner. Here we provide the first comprehensive and unbiased analysis of the six different SET1/MLL complexes in human cells. Single-step affinity purification of shared and unique subunits of the different complexes from nuclear extracts reveals a high degree of heterogeneity in the subunit composition. We find that the WRAD core subunits Dpy30 and Wdr5, but not Ash2l or Rbbp5, are present in other large protein complexes. Additionally, we propose that Bod1 is the human homolog of yeast Shg1p. Furthermore, Bod1 and Psip1 bind selectively to the SET1B and MLL2 complexes, respectively.

Results

In order to obtain quantitative information on the different human COMPASS-like complexes, individual subunits of these complexes were tagged with GFP, allowing for single-step affinity purification and identification of complexes in a single liquid chromatography combined with tandem mass spectrometry (LC-MS/MS) run (Hubner et al. 2010; Vermeulen et al. 2010). We made use of HeLa cell lines in which expression of the tagged protein can be induced upon doxycycline addition (Tighe et al. 2008). Nuclear extracts obtained from these cells and from the parental cell line (WT) were subjected to GFP affinity purification in triplicate as described (Hubner et al. 2010), followed by direct on-bead digestion and mass spectrometry analysis on an LTQ-Orbitrap-Velos instrument. Once the interactors were identified, their relative stoichiometry was determined as described (Smits et al. 2012)(Fig. 1A). To allow comparison of stoichiometries between different experiments, all data were normalized to the total amount of catalytic subunit in that specific experiment and consequently the stoichiometry is expressed relative to the total amount of HMT. We used this workflow to analyze the shared SET1/MLL complex subunits (Ash2l, Rbbp5, Wdr5 and Dpy30), as well as two specific subunits for MLL1/2 (Menin and Psip1), MLL3/4 (Pa1 and Ptip) and SET1A/B (Wdr82 and Cfp1).

SET1/MLL core subunit interactions

Rbbp5 and Ash2l are part of the stable core of the human SET1/MLL complexes (Fig. 1B,C). Ash2l, Rbbp5, Wdr5 and Dpy30 have previously been described to form the WRAD complex independently of the catalytic HMT subunit (Patel et al. 2011; 2009; Dou et al. 2006; Steward et al. 2006). Purification of Ash2l or Rbbp5 revealed that Ash2l, Rbbp5 and Dpy30 are more abundant than the combined Set1/Mll proteins (Fig. 1B,C). These observations suggest that a

proportion of the WRAD module is not associated with HMT activity. The exception to this is the Wdr5 subunit, which is present in a 1:1 ratio with the HMTs (Fig. 1B,C). This is in agreement with previously reported direct interactions between Mll1 and Wdr5 (Trievel and Shilatifard 2009). These observations suggest the existence of a subcomplex consisting of Rbbp5, Ash2l and Dpy30. Alternatively, this putative RAD module is stabilized by overexpression of one of its components.

Wdr5 plays an important role in self-renewal and reprogramming and many of these functions are attributed to its role in SET1/MLL (Wysocka et al. 2005; Ang et al. 2011). In addition, Wdr5 has been found in other complexes including the ATAC histone acetyltransferase complex (Trievel and Shilatifard 2009; Spedale et al. 2012). We found Wdr5 to interact, either directly

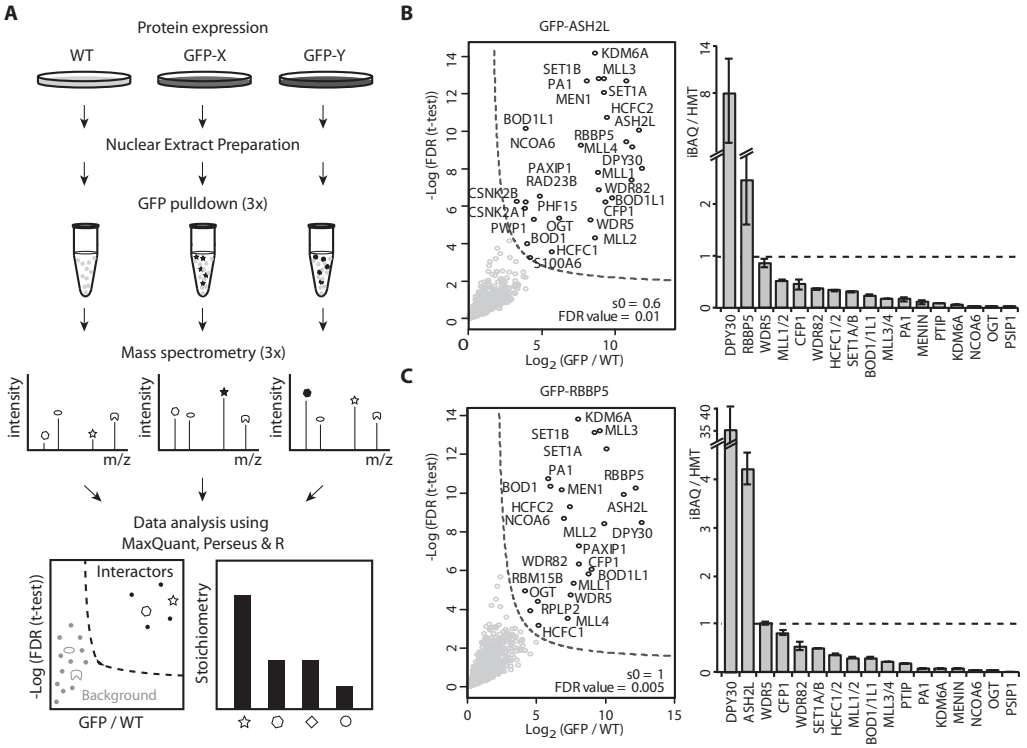


Figure 1. Experimental workflow. (A) Nuclear extracts were prepared from HeLa wild type or GFP-fusion protein expressing cells. GFP pull downs were performed in triplicate and analyzed separately by mass spectrometry. Raw data were analyzed by MaxQuant and specific interactors were selected from background using label free quantification in Perseus. iBAQ intensities were used to calculate the relative abundance of interaction partners. (B, C) Identification of interacting proteins for Ash2l (B) and Rbbp5 (C) by volcano plots (left panel) and the stoichiometry (>0.01) of these interactors presented by bar graphs (right panel). In the volcano plots, the ratio of GFP over WT in label free quantification are plotted against the -log10 of the false discovery rate (FDR) calculated by a permutation based FDR adapted t-test. Significant outliers are labeled. Bar graphs indicate the stoichiometry of interacting proteins (indicated at bottom) relative to Set1/Mll proteins. Dashed line indicates a ratio to the total Mll/Set1 protein of 1. Error bars indicate the standard deviations of the biochemical triplicate for each experiment.

or indirectly, with almost 200 different proteins (Fig. 2A). As expected (Trievel and Shilatifard 2009), amongst these are subunits of the ATAC, NSL, HBO1 and anaphase promoting complexes. Additionally, INO80 and TFIID subunits were identified as interactors. These complexes were not identified using the other (Ash2l, Rbbp5 and Dpy30) core subunits as baits indicating that they are exclusive for Wdr5 (Song and Kingston 2008). Interestingly, intensities of the different identified complexes are comparable, indicating that Wdr5 is equally distributed over several chromatin associated protein complexes. This suggests that Wdr5 is more abundant in cells than the other WRAD members. To investigate this, we performed intensity based absolute quantification (iBAQ) of HeLa nuclear extract (Schwanhäusser et al. 2011) (Table I). This analysis resulted in the identification and absolute quantification of ~4800 proteins. As expected, Wdr5 is ~10 fold more abundant than Rbbp5 and Ash2l (Table I). This supports the idea that Wdr5 is a universal hub in chromatin and transcription regulatory pathways (Trievel and Shilatifard 2009).

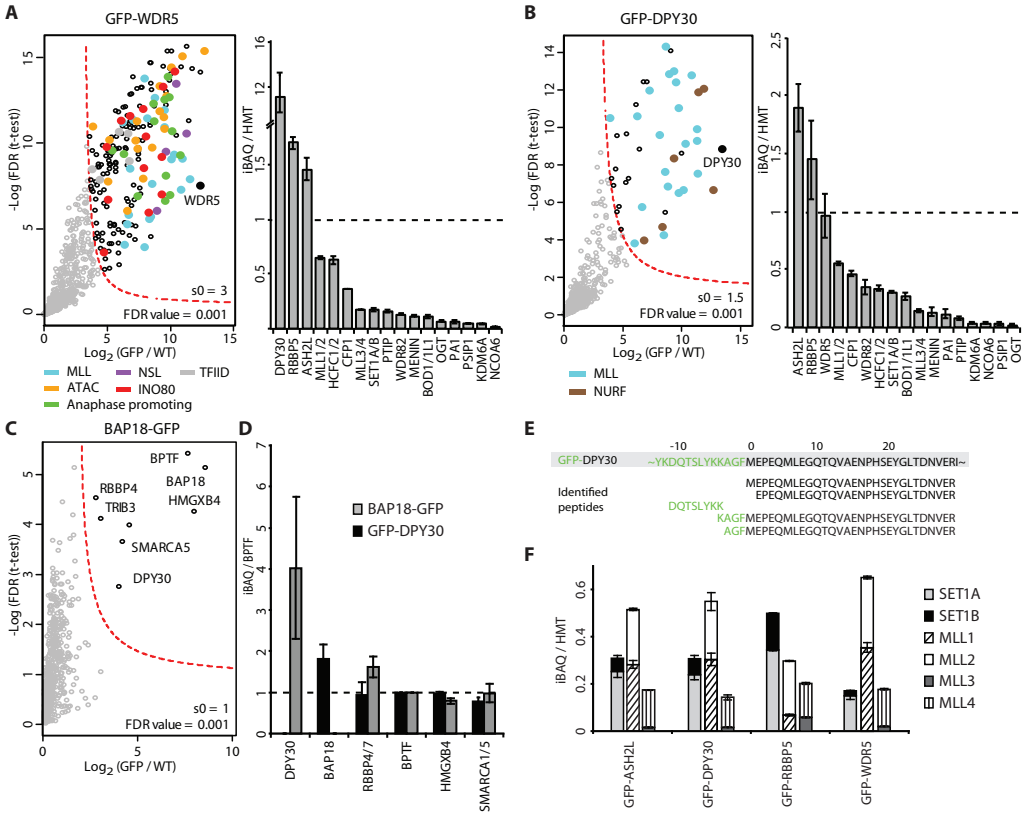


Figure 2. Wdr5, Dpy30 and Bap18 purifications. (A, B) Volcano plot of Wdr5 and Dpy30 interacting proteins (as indicated above the left panels). Coloring of significant outliers based on GO classification (cellular compartment). Bar graphs of the stoichiometry (>0.01) of interacting proteins (right panels). Layout is as in Figure 1. (C) Volcano plot of Bap18 interacting proteins. Layout is as in Figure 1. (D) Stoichiometry of Dpy30 and Bap18 bound NURF subunits relative to Bptf. (E) Peptide sequences of Dpy30 N-terminal peptides identified in the Dpy30 pull down. (F) Stoichiometry of HMTs in WRAD pull downs.

Absolute protein quantification of HeLa nuclear extract

Protein IDs	Protein names	Amount (fmoles / 1 mg)
O75475	PSIP1	68047.34
P61964	WDR5	8870.15
B4DIS3	DPY30	5799.25
Q6UXN9	WDR82	4908.15
Q15291	RBBP5	894.73
Q9UBL3	ASH2L	596.69
Q8NFC6	BOD1/1L1	201.70
Q9P0U4-2	CFP1	189.01
O00255-2	MEN1	137.71
O15047/Q9UPS6	SET1A/B	133.68
Q6ZW49	PTIP	58.14
Q9BTK6	PA1	7.31
E9PQG7/Q9UMN6	MLL1/2	1.91
O14686-3/Q8NEZ4-3	MLL3/4	1.62
O15047	SET1A	132.68±7.0
O14686-3	MLL4	1.44±0.069
E9PQG7	MLL1	1.27±0.070
Q9UPS6	SET1B	1.00±0.48
Q9UMN6	MLL2	0.64±0.060
Q8NEZ4-3	MLL3	0.18±0.0024

Table 1. Absolute protein quantification of SET1/MLL subunits in HeLa nuclear extracts. Given in fmoles/1mg extract as measured by iBAQ quantification.

Strikingly, the WRAD members are at least four times more abundant than the sum of all HMT subunits. This is in line with findings by others (Dou et al. 2006; Steward et al. 2006) and by us (Fig. 1 and 2) that Wdr5, Rbbp5, Ash2l and Dpy30 may form various subcomplexes.

Dpy30 isolation led to the co-purification of SET1/MLL components, amongst which Ash2l showed the highest stoichiometry, which suggests a direct interaction between Ash2l and Dpy30 (Fig. 2B). Surprisingly, all known subunits of the NURF chromatin remodeling complex co-purified with Dpy30 (Fig. 2B). This indicates that Dpy30 is also an integral component of the NURF complex. Due to its small size (11 kDa), this protein may have been missed in previous, gel based analyses of the NURF complex (Barak et al. 2003). Roughly equal amounts of HMT and NURF complexes were present in the Dpy30 pull downs. Bap18 is a core component of NURF (Vermeulen et al. 2010) and was used to validate the Dpy30-NURF interaction (Fig. 2C). We determined the stoichiometry of the NURF subunits in the Bap18 and Dpy30 pull downs and found that Bap18 is present in two copies relative to the other subunits (Fig. 2D). By analogy we propose that Bap18 might serve as the direct anchor for Dpy30 in NURF. In the GFP-Dpy30 purifications we identified tryptic peptides from the junction of GFP-Dpy30 fusion protein and from N-terminus of endogenous Dpy30. This indicates co-purification of GFP tagged and endogenous Dpy30, which reveals the existence of Dpy30 multimers (Fig. 2E). Notably, Dpy30 is 6-10 times more abundant in HeLa nuclear extracts compared to Rbbp5 and Ash2l (Table I).

Relative abundance of the SET1/MLL complexes

Analysis of the shared subunits of the SET1/MLL complexes revealed that all six HMTs (Mll1-4 and Set1a/b) are present in a complex with the WRAD module (Fig. 1B,C and 2A,B). These

homologous HMTs have evolved from a common ancestor and consequently share a number of tryptic peptides. To determine the exact stoichiometry of each HMT complex, HMTs that share tryptic peptides are collapsed into a single stoichiometry value. This analysis revealed that the MLL1/2 complexes are most abundant (on average 50% of all complexes), whereas SET1A/B and MLL3/4 accounts for 32% and 18% of the total pool of WRAD-bound HMT, respectively (Fig. 2F). Based on the intensity of the unique peptides for each HMT we estimated their relative abundance (Fig. 2F (stacking of bars)). Whereas the Mll1 and Mll2 proteins are present in roughly equal amounts (both ~25%), Mll3 is fairly low compared to Mll4 (~3% versus ~15%, respectively). The same holds true for Set1a, which is more abundant than Set1b (~25% and ~7%, respectively; Fig. 2F).

The relative presence of the enzymatic components in the SET1/MLL complexes could be regulated post-translationally or a direct effect of differential protein expression. Interestingly, Set1a is the most abundant HMT as indicated by its absolute protein abundance in HeLa nuclear extract (Table I), which is not reflected in its degree of integration in SET1/MLL complexes (32% of all complexes). MLL1/2 and MLL3/4 complexes have similar abundance (1.9 and 1.6 fmol/mg, respectively), whereas their degree of interaction with WRAD is not equal (50% versus 18%). Therefore, the integration of these enzymatic components is not strictly determined by their absolute abundance but probably due to the presence of certain HMTs (like Mll3/4 and Set1a) in other complexes lacking the WRAD proteins.

MLL1/2 specific interactors

Menin acts as a tumor suppressor in the neuroendocrine MEN1 tumor syndrome, but is an essential cofactor for the oncogenic activity of rearranged *MLL* (Karnik et al. 2005; Hughes et al. 2004). Menin binds to Mll1/2 and the transcription factor JunD in a mutually exclusive manner (Huang et al. 2012; Hughes et al. 2004; Agarwal et al. 1999). Purification of the core MLL subunits revealed a 4- to 5-fold lower abundance of Menin compared to Mll1/2, indicating that not all MLL1/2 complexes contain Menin (Fig. 1B,C and 2A,B). On the other hand the fraction of Menin, which is associated with Mll1/2, binds to the WRAD complex with an stoichiometry similar to 1:1/1/1/6 (Menin:Wdr5/Rbbp5/Ash2l/Dpy30) (Fig. 3A). Amongst the other Menin interacting proteins, JunD was highly enriched but in a lower stoichiometry than MLL1/2 complex subunits. Intriguingly, the JunD dimerization partners, c-Fos, Fos1l, Atf2/7, and Atf3, were also identified in this experiment (Fig. 3A,C). This suggests that Menin interacts with dimers of JunD and other AP-1 transcription factors. Additionally, Menin was previously shown to localize to chromatin after UV radiation, which may depend on these AP-1 transcription factors (Farley et al. 2006). We also identified the nucleotide excision repair proteins Rad23a/b as novel Menin interactors, which may hint for a function for Menin in this pathway. Notably, the RNA binding protein Taf15 was also identified, which might link Menin to other non-MLL related functions.

Psip1 (Ledgf/p75) is well studied as the interaction partner for the HIV-integrase (Marshall et

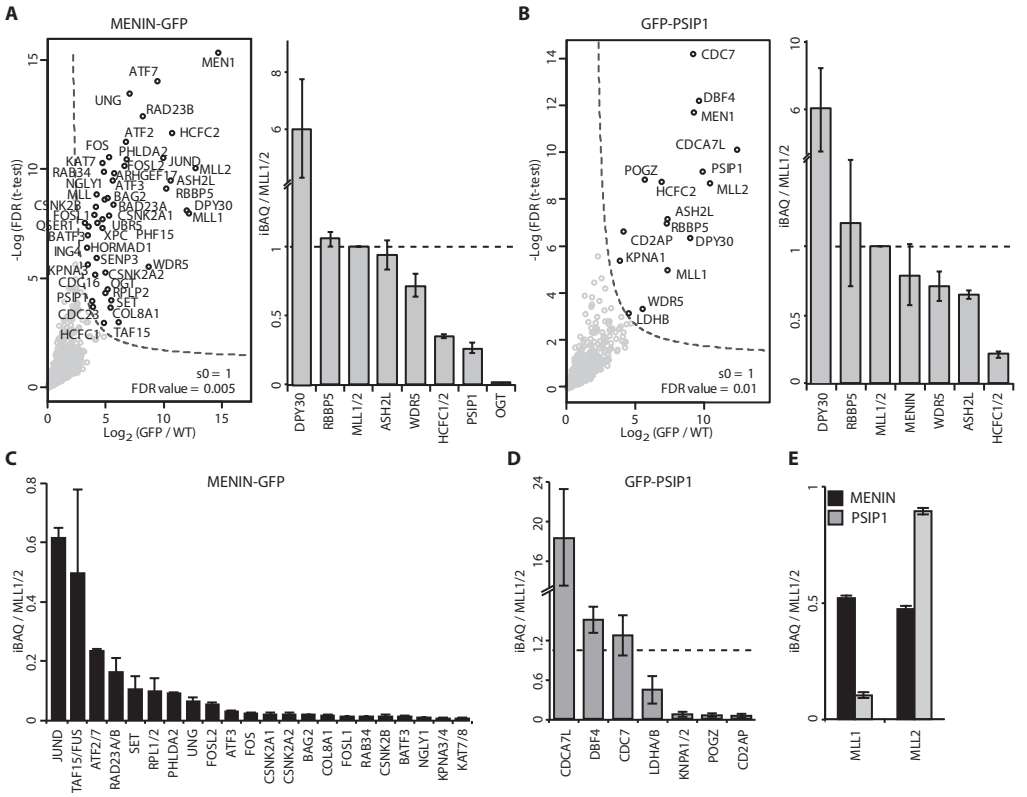


Figure 3. Menin and Psip1 purifications. (A, B) Volcano plot of Menin and Psip1 interacting proteins (as indicated above the left panels). Bar graphs of the stoichiometry (>0.01) of interacting SET1/MLL complex subunits (right panels). Layout is as in Figure 1. (C, D) Stoichiometry of specific interactors for Menin (C) and Psip1 (D). (E) Relative amounts of Mll1 and Mll2 proteins associating with Menin or Psip1.

al. 2007) and was more recently also implicated in MLL linked leukemia (Yokoyama and Cleary 2008). Like Menin, Psip1 does not interact with the Set1a/b or Mll3/4 methyltransferases. Analysis of the stoichiometry of Psip1-bound MLL subunits resulted in an identical picture as when Menin was used as bait (Fig. 3A,B). As the Menin protein has a 1:1 stoichiometry relative to Mll1/2 in the GFP-Psip1 purifications, it is likely that Menin and Psip1 only interact in the context of Mll1/2. On the other hand, Menin pull down indicated that Psip1 is only present in ~25% of the Menin-MLL1/2 complexes. This low abundance is supported by the analysis of WRAD subunits as baits, in which Psip1 was not identified as an interactor (Fig. 1B,C and 2A,B), together indicating that Psip1 is a sub-stoichiometric component in this complex. Interestingly, Psip1 enriches mostly Mll2, whereas Menin binds equal amounts of the Mll1 and Mll2 proteins (Fig. 3E). This indicates preferential binding of Psip1 to the MLL2 complex.

Our analyses confirmed Cdca7l/Jpo2 as a Psip1 interactor (Maertens et al. 2006). We find that Psip-Cdca7l is 18-fold more abundant than the Psip-MLL1/2 complexes (Fig. 3D). This is confirmed by the absolute protein quantification dataset, which shows that Psip1 is the most

abundant SET1/MLL interactor with a 100-fold higher abundance than Rbbp5 and Ash2l (Table I). As reported previously, Psip1 interacts with Dbf4 and Cdc7, which likely form a complex, and with the Pogz transcription factor (Hughes et al. 2010; Bartholomeeusen et al. 2009). These interactions are of a lower abundance than Mll1/2. Together, these observations reveal that only ~5% of Psip1 is present in MLL complexes, and that the bulk of Psip1 performs other functions.

Hcfc1 and -2 were previously identified as members of the MLL1 complex, but they are also linked to the SET1 complexes (Wysocka et al. 2003). Our experiments validate these proteins as sub-stoichiometric interactors for SET1A/B and MLL1/2 and provide no evidence for an interaction of Hcfc1/2 with the MLL3/4 complexes (Fig. 3A,B and 4A,B).

MLL3/4 specific interactors

Ptip was previously identified as a *bona fide* subunit of the MLL3/4 complexes and found to be associated with the 53bp1 protein involved in the DNA damage response pathway (Manke et al. 2003; Cho et al. 2007). The interaction with Mll3/4 was confirmed (Fig. 4A) and Ptip bound MLL3/4 complexes seem to have an 1:1/1/1/3 (Ptip:Wdr5/Rbbp5/Ash2l/Dpy30) stoichiometry. Interestingly, Dpy30 is approximately present as a trimer compared to the hexamer in MLL1/2. Other known MLL3/4 subunits, the histone H3K27 demethylase Kdm6a/Utx and the nuclear receptor co-activator NcoA6 are present in only ~19% and ~30% of Ptip pull downs relative to Mll3/4, respectively (Fig. 4A,B). This indicates that these subunits are genuine sub-stoichiometric components of the Ptip-containing MLL3/4 complexes.

Notably, Ptip-53bp1 complexes are more abundant than Ptip-HMT complexes (Fig. 4C). The GFP-Ptip purification (Fig. 4C) also identified Dclre1C, Mdc1 and Blm, which are all implicated in the DNA damage response pathway (Thompson 2012). These proteins were not identified previously as Ptip interactors, which may indicate that they interact via 53bp1. The zinc finger transcription factor Zbtb2 has been implicated in the p53 pathway (Jeon et al. 2009), and it was found as a strong Ptip interactor. The Zbtb2-Ptip interaction seems independent from the SET1/MLL complexes, because Zbtb2 was not identified in WRAD purifications (Fig. 4C). Furthermore, the Smad2/3 proteins interacted with Ptip. Interestingly, a role in TGFβ signaling via interaction with Smad proteins was previously described for the Xenopus Ptip homolog Swift (Shimizu et al. 2001). The novel interaction found with the zinc finger Znf639 may hint to a role for this protein in localizing Ptip and its associated proteins to specific genomic loci. It is important to note that these interactions were not found with Pa1 or other SET1/MLL proteins and could therefore also be linked to the 53bp1 DNA damage pathway or to novel functions of the Ptip protein.

Ptip-associated (Pa1) protein interacts with the MLL3/4 complexes with similar stoichiometry as Ptip, 1:1/1/1/3 (Pa1:Wdr5/Rbbp5/Ash2l/Dpy30) (Fig. 4B). Moreover, a larger proportion (~2 fold) of Ptip compared to Mll3/4 is found in these experiments. Conversely, Pa1 was 2-fold more abundant relative to Mll3/4 in the GFP-Ptip analysis (Fig. 4A). Ptip and Pa1 are found with a lower stoichiometry than Mll3/4 in the pull downs of the WRAD components, which

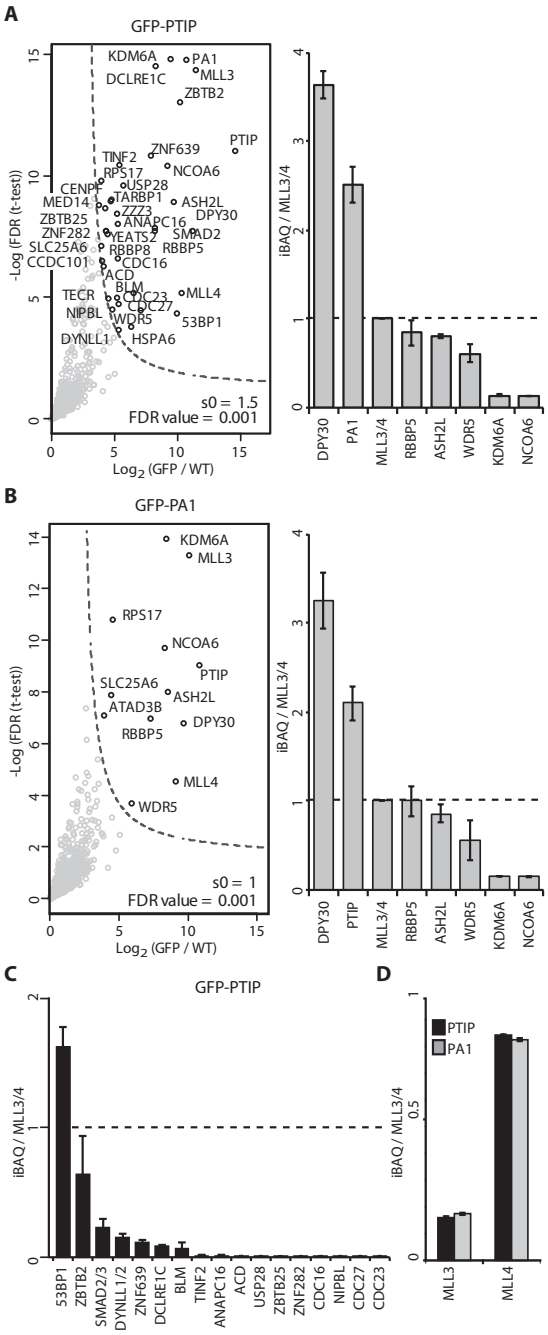


Figure 4. Ptip and Pa1 purifications. (A, B) Volcano plot of Ptip and Pa1 interacting proteins (as indicated above the left panels). Bar graphs of the stoichiometry (>0.01) of interacting SET1/MLL complex subunits (right panels). Layout is as in Figure 1. (C) Stoichiometry of specific interactors for Ptip. (D) Relative amounts of Mll3 and Mll4 proteins associating with Ptip or Pa1.

makes a Ptip/Pa1 dimer within the MLL3/4 complex unlikely (Fig. 1B,C and 2A,B). Instead, these data suggest that Ptip and Pa1 form a complex outside of the MLL3/4 complex. Analysis of the distribution of Mll3 and Mll4 proteins in the Ptip and Pa1 pull downs, show similar Mll3 and Mll4 distributions as found in the WRAD purifications, indicating no preference of Ptip and Pa1 for either of these HMTs (Fig. 4D).

SET1A/B specific interactors

The beta-propeller protein Wdr82 is part of the SET1A/B complexes and directly involved in regulation global levels of H3K4me3 (Lee et al. 2007; Wu et al. 2008). Wdr82 has also been found to interact with the PP1 phosphatase complex (Lee et al. 2010). Wdr82 is a stoichiometric component of the SET1A/B complexes based on pull downs of Ash2l, Rbbp5 and Dpy30 (Fig. 1A,B and 2B). Purification of Wdr82 indicated stoichiometric amounts of Cfp1 and Set1a/b and slightly lower amounts of the WRAD components (Fig. 5A). Again, Dpy30 is present as a trimer. The Wdr82 purification also identified the Bod1 and Bod1l1 proteins as novel interactors, which we studied further (see below). As expected (Lee et al. 2010) we also identified the whole PP1 protein phosphatase complex (Tox4, Pp1a/b/c/, Ppp1R10), which was present at a ~10-fold higher abundance than the Set1a/b proteins (Fig. 5C). In addition we identified several novel Wdr82-interacting proteins, which are in much lower abundances than PP1 and SET1A/B complexes (Fig. 5C).

Cfp1 purification resulted in SET1A/B complex members and appears to be an exclusive and specific SET1A/B subunit (Fig. 5B). Similar to Wdr82 purification, the Bod1 and Bod1l1 proteins were identified as interactors, which suggests that they are genuine interactors of the SET1A/B complexes. Analysis of Cfp1 stoichiometry in WRAD pull downs indicated that Cfp1 binds with a stoichiometry of approximately 1.6 relative to the total HMT subunit, 1.6:1/1/1/2 (Cfp1:Wdr5/Rbbp5/Ash2l/Dpy30). This may indicate a potential dimerization of Cfp1 in the SET1A/B complexes.

Bod1 and Bod1l1 are human paralogs of yShg1p and exclusive to SET1B

The bi-orientation defected 1 protein (Bod1) is important for proper chromosome segregation during mitosis (Porter et al. 2007). Bod1 and the highly similar Bod1l1 were identified in our experiments to interact with all members of the WRAD complex (Fig. 1B,C and 2A,B). Purification of the specific SET1A/B subunits Cfp1 and Wdr82 revealed that Bod1 and Bod1l1 are exclusively present in SET1A/B complexes, as we find specific and shared peptides for both proteins (Fig. 5A,B).

To investigate this further we constructed a GFP-Bod1 expressing cell line and analyzed its interactors. Interestingly, Set1b was the only HMT interactor (Fig. 5E). Estimation of the relative abundance of Set1a and Set1b in the Bod1, Cfp1 and Wdr82 pull downs, confirmed that Bod1

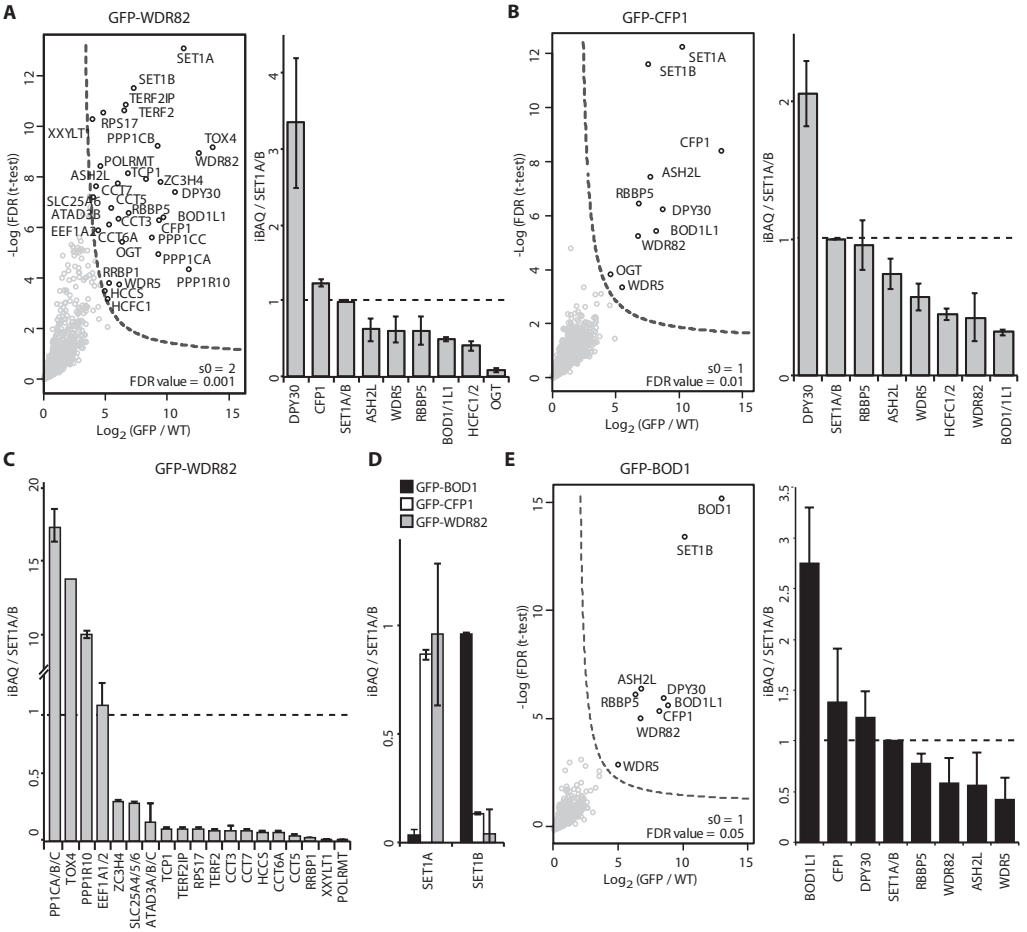


Figure 5. Wdr82, Cfp1 and Bod1 purifications. (A, B) Volcano plot of Wdr82 and Cfp1 interacting proteins (as indicated above the left panels). Bar graphs of the stoichiometry (>0.01) of interacting SET1/MLL complex subunits (right panels). (C) Stoichiometry of specific interactors for Wdr82. (D) Relative amounts of Set1a and Set1b proteins associating with Ptip, Pa1 or Bod1. (E) Volcano plot of Bod1 interacting proteins with significant outliers indicated (left panel). Bar graphs of the stoichiometry (>0.01) of interacting proteins (right panel). Layout is as in Figure 1.

exclusively interacts with the SET1B complex (Fig. 5D). In contrast, Cfp1 and Wdr82 interact mostly, but not exclusively, with Set1a and the ratio of Set1a/Set1b being similar to the ratios in the WRAD purifications (Fig. 5D and 2F).

The Bod1 pull down identified unique peptides for Bod1l1 with higher iBAQ intensity than the other SET1B complex members, which may indicate that Bod1 and Bod1l1 could interact outside of the SET1B complex. Alignment of the Bod1 homologous sequences from human, fruit fly and zebrafish revealed a high degree of similarity with the *S. cerevisiae* protein Shg1p (data not shown). Intriguingly, Shg1p (a.k.a. Csp15) is part of the yeast COMPASS complex. Shg1p interacts with RNA, which affects yeast SET1 complex assembly (Halbach et al. 2009;

Dehé et al. 2006). Based on the identification of Bod1 in SET1B and the high sequence similarity we propose that Bod1 and Bod1l1 are the higher eukaryotic paralogs of yeast Shg1p.

Discussion

Recent developments in quantitative mass spectrometry allowed an unbiased, quantitative and comprehensive interaction study on the six COMPASS-like H3K4 methyltransferase complexes in humans (Smits et al. 2012). We identified many known interactors illustrating the high quality of our data sets and calculated their stoichiometries. The ten SET1/MLL pull downs were clustered based on the identified interactors and presented as a heat-map reflecting their stoichiometry (Fig. 6A and summarized in Fig. 6B). Our approach indicates a high degree of diversity in the peripheral subunits of the different HMT complexes.

The smallest core subunit of the SET1/MLL complexes, Dpy30, shows an interesting behavior. It was shown to interact with Ash2l as a dimer (Patel et al. 2009). However, we observe a stoichiometry of ~6 in Mll1/2, whereas Mll3/4 and Set1a/b bind ~3 molecules of Dpy30 per complex (Fig. 6B). This argues for additional binding sites on the Mll and Set proteins or on the peripheral subunits. Dpy30 is only 11 kDa, therefore the quantification is based on a limited number of tryptic peptides, leading to the observed higher standard deviation. We also identified all members of the NURF complex (Barak et al. 2003) in the GFP-Dpy30 pull down and we identified Dpy30 in the pull down with the Bap18 subunit of NURF (Fig. 2C). These findings imply that Dpy30 is a genuine subunit of NURF and that its function should not be interpreted solely in the context of HMT complexes (Jiang et al. 2011).

The same holds true for Wdr5, which was identified as a central hub in chromatin associated complexes. Since Wdr5 was described as a SET1/MLL complex member, many Wdr5 interactors have previously been linked to the SET1/MLL complexes, such as the INO80, TFIID and NSL complexes (Trievel and Shilatifard 2009). This annotation is still present in gene ontology databases. These interactions are, however, not found in the other core SET1/MLL subunit pull downs (Rbbp5, Ash2l or Dpy30). This indicates that rather than SET1/MLL complex interactors, these protein complexes should be annotated as exclusive Wdr5 interactors. It was shown that Wdr5 binds to the tail of histone H3 and that this may facilitate methylation by HMTs (Dharmarajan et al. 2012; Song and Kingston 2008). Similarly, H3 tail presentation by Wdr5 could be relevant for the binding or modification by the complexes identified as Wdr5 interactors here.

Our results address the question of the limiting components for the assembly of SET1/MLL complexes in cycling HeLa cells. With the exception of Set1a, the HMTs are present in low amounts in crude nuclear extracts. The WRAD subunits and specific interactors like Wdr82, Psip, Menin and Bod1 are more abundant. These observations support the existence of (W)

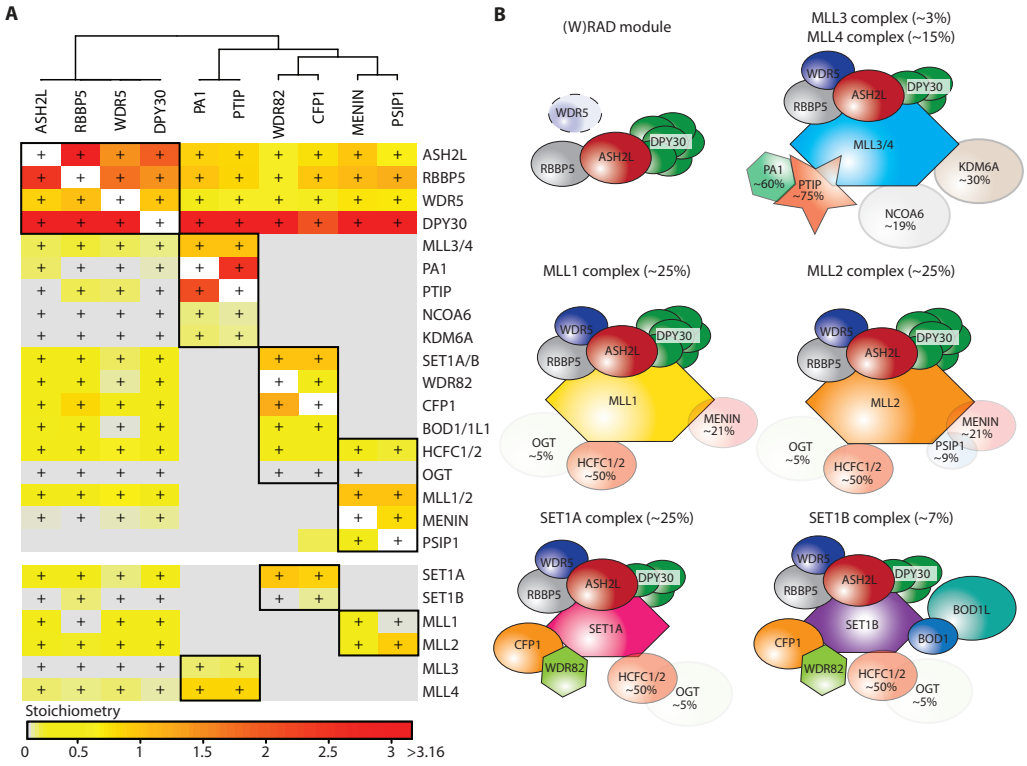


Figure 6. SET1/MLL interactome. (A) Hierarchical clustering of overlapping interactors identified for Wdr5, Rbbp5, Ash2L and Dpy30 pull downs. Bait proteins are indicated on top and identified interactors on the right. Coloring (ascending from grey to red via yellow and orange) is based on the stoichiometry of the interactor in that pull down, relative to the total amount of Set1a/b and Mll1-4 proteins found for that bait. Significant outliers are indicated (+) for each bait protein. (B) Summary of the composition and relative amounts of the SET1/MLL complexes determined in this study. Dpy30 is present as a multimer dependent on the complex context. Proteins that have a stoichiometry of <1 are displayed transparent and their relative abundance is indicated.

RAD subcomplexes lacking the HMT subunit (Steward et al. 2006). Furthermore, our unbiased approach shows that the complex specific subunits can also be part of other complexes, like Dpy30 in NURF, Wdr82 in the PP1 phosphatase (Lee et al. 2010) and Psip in the Cdc7/Dbf4/Cdca7l complexes (Maertens et al. 2006; Hughes et al. 2010). These observations raise questions regarding the regulation of these different interactions during the cell cycle or upon cellular stress. Interestingly, different components of the MLL1 complex display differential chromatin binding during mitosis (Blobel et al. 2009).

Whereas Menin binds Mll1 and Mll2 with equal affinity, Psip1 shows a striking preference for Mll2, under the used conditions in HeLa cells (Fig. 6B). This conflicts with an earlier protein-protein interaction study using fragments of these proteins (Huang et al. 2012). Although Psip1 and Menin are sub-stoichiometric MLL1/2 subunits, both proteins are essential for oncogenic transformation by Mll1 fusions (Yokoyama and Cleary 2008; Yokoyama et al. 2005). Our observation that Psip is mainly a MLL2 complex member may imply that the function of Psip1

during Mll1 fusion mediated oncogenesis is dependent on MLL2 function as much as it depends on the wild type MLL1 complex (Thiel et al. 2010).

Bod1 has been linked previously to chromosome segregation during mitosis (Porter et al. 2007). The phenotype of Bod1 depletion and the association of the other SET1B complex member Wdr82 with PP1, of which inactivation leads to a mitotic arrest (Fernandez et al. 1992), hints to a link between the SET1B complex, histone methylation and mitotic progression. In yeast, Set1p was reported to inhibit aurora kinase function during chromosome segregation, which involves methylation of the Dam1p kinetochore protein (Zhang et al. 2005). Based on these observations, we speculate that the SET1B complex plays a similar role during mitosis in mammalian cells.

The absolute protein quantification data of the nuclear extract are in good (but not perfect) agreement with the observed stoichiometries of SET1/MLL complex subunits. The higher abundance of WRAD subunits supports existence of free (W)RAD subcomplexes (Steward et al. 2006). Additionally, the finding that Wdr5 and Dpy30 are subunits of other chromatin regulatory complexes is mirrored by their higher abundance in nuclear extracts. This suggests that absolute protein quantifications could be predictive for multimerization and/or association with multiple complexes. This concept could be applied to predict composition of protein complexes in tissue samples for which insufficient material is available for biochemical purification.

Taken together, our quantitative mass spectrometry analysis of the major H3K4 methyltransferase complexes in human cells, revealed both known and novel, unique and shared, interactors and determined their distribution and stoichiometry over the different SET1/MLL complexes. The comprehensive and quantitative mapping of subunit composition and abundance provides a molecular framework to understand the diversity and contributions of the SET1/MLL complexes in establishing H3K4 methylation patterns.

Materials & Methods

Plasmids and cell culture

The ORF of the bait protein was amplified by PCR using the relevant human cDNA constructs and introduced into pDONR2.1. The DNA sequence of the amplified ORF was verified and introduced into a GATEWAY-compatible pCDNA5/FRT/TO. All proteins except for Menin were tagged by GFP at the N-terminus. cDNA constructs for Wdr5, Dpy30, Rbbp5 and Ash2l (short isoform) were kindly provided by dr. Ali Shilatifard and Bod1 cDNA was a kind gift from dr. Jason Swedlow. Human Pa1 and Cfp1 and mouse Wdr82 and Ptip cDNAs were obtained from Source Bioscience (Cfp1: IRATp970F0412D, Wdr82: IRCKp5014J0617Q, Pa1: IRAUp969E1119D, Ptip1: IRAV9968G04124D). Stable doxycycline inducible cell lines were created by transfecting pCDNA5/FRT/TO and pOG44 into HeLa FRT cells carrying the TET repressor using polyethyleneimine followed by antibiotic selection. Cells were grown in DMEM with high glucose supplemented

with pen/strep and L-Glutamine (all LONZA) under blasticidin and hygromycin B selection. Protein expression was induced by addition of 0.5 µg/ml doxycycline to the culture media 16 hours prior to cell harvesting. Expression of the proper-sized GFP-fusion protein was validated by immunoblotting and by fluorescence microscopy. Bap18-GFP bacterial artificial chromosomes were stably transfected in HeLa cells and selected using geneticin (G418, Gibco).

GFP affinity purification and sample preparation

Extract preparation (Dignam et al. 1983) and affinity purifications using GFP-beads (Hubner et al. 2010) were performed essentially as described before. Briefly, nuclei were isolated and nuclear extracts were prepared using hypotonic lysis. Purifications for GFP lines and WT HeLa cells were performed in triplicate using 1 mg of nuclear extract per purification and GFP binder beads (CHROMOTEK) in 20 mM HEPES-KOH pH 7.9, 20% glycerol, 300 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP-40, 0.5 mM DTT and complete protease inhibitors (Roche). All purifications included 50 µg/ml ethidium bromide to suppress DNA mediated interactions. After 2.5 hours incubation at 4°C the beads were extensively washed and on-bead digestion was performed using trypsin (Promega). After desalting and concentration on StageTips the peptides were applied to online nanoLC-MS/MS, using a 120 min acetonitrile gradient (5.6 - 76%). Mass spectra were recorded on an LTQ-Orbitrap-Velos mass spectrometer (Thermo) selecting the 15 most intense precursor ions of every full scan for fragmentation.

Data analysis

Raw data were analyzed using MaxQuant 1.3.0.5, with label free quantification (LFQ), match between runs (between triplicates) and the iBAQ algorithm enabled (Cox and Mann 2008). The identified proteins were filtered for known contaminants and reverse hits, as well as hits without unique peptides. Protein interactor identification was done as described (Hubner et al. 2010). In short, the normalized mass spectrometric intensities (LFQ intensities) were compared between the GFP tagged and control sample, using an adapted permutation based false discovery rate (FDR) *t*-test in Perseus (MaxQuant package). The threshold for significantly identify interactors is both based on the FDR and the ratio between GFP and control sample. This threshold is empirically optimized for each experiment.

The stoichiometry was determined for significant interactors as previously (Smits et al. 2012). Here, mass spectrometric intensities were normalized for the theoretical number of observable peptides by the iBAQ algorithm. Thereby, the normalized intensities (iBAQ intensities) give a measure for protein abundance and can be directly compared between proteins. The iBAQ intensities measured in the control sample indicate the amount of background binding and were therefore subtracted from the iBAQ intensities obtained in the GFP sample. Finally, the corrected iBAQ intensities were scaled to the total amount of Set1/Mll subunit, thereby

allowing direct comparison of the different purifications. The average and standard deviation of the resulting stoichiometries were calculated per triplicate.

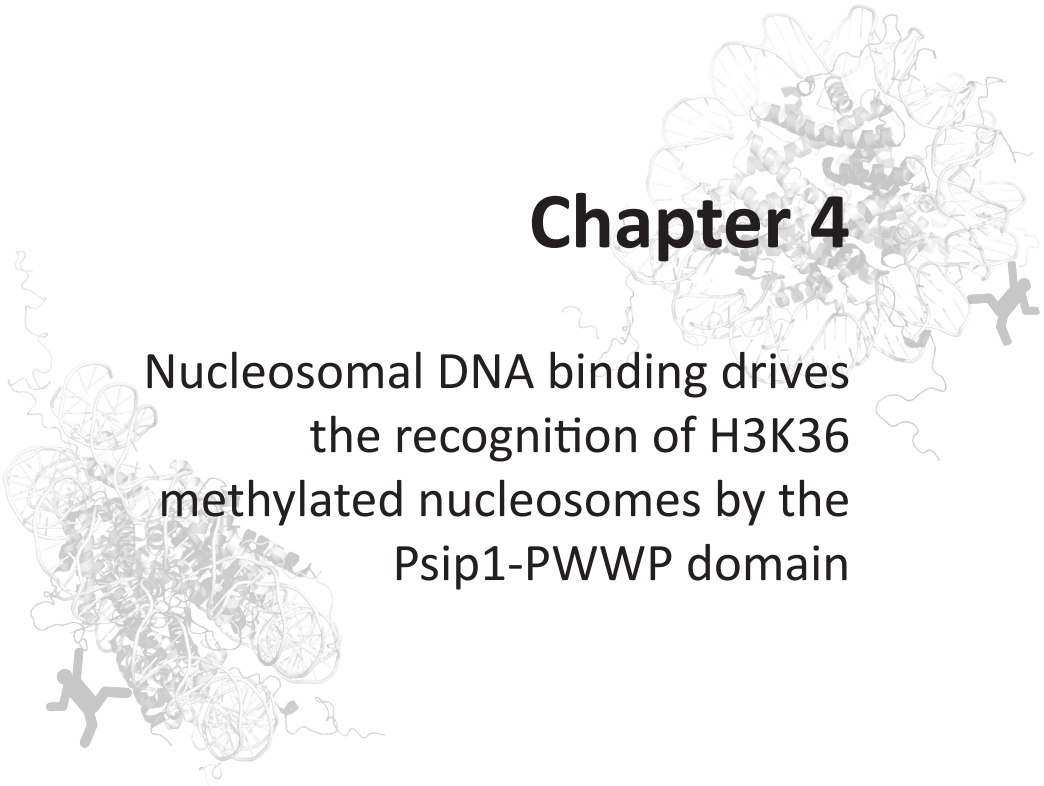
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaíno et al. 2013) with the dataset identifier PXD000172.

Absolute protein quantification in HeLa nuclear extract

Nuclear extracts from the HeLa Kyoto cell line were subjected to absolute protein quantification using iBAQ and a universal protein mix standard (UPS2), as described before (Schwanhäusser et al. 2011). Briefly, 3 µg of UPS2 was added to 10 µg of HeLa nuclear extract, followed by FASP digestion (Wiśniewski et al. 2009). In parallel, 100 µg of HeLa nuclear extract was applied to FASP followed by strong anion exchange (SAX) chromatography resulting in 8 fractions. After purification on stage tips, all 9 samples were separately measured by LC-MS/MS over a 4 hour acetonitrile gradient (5.6% - 76%). Raw data were analyzed by MaxQuant 1.3.0.5 with iBAQ quantification enabled. A linear regression curve was made between the known UPS2 concentrations (log scale) and the measured iBAQ intensities for these proteins (log scale), which was used to extrapolate the absolute protein numbers of all measured proteins in this sample. Next, linear regression of these absolute protein abundance and their iBAQ intensities of these in the SAX dataset, allowed to quantify the absolute amounts for all proteins found in the large SAX dataset.

Acknowledgements

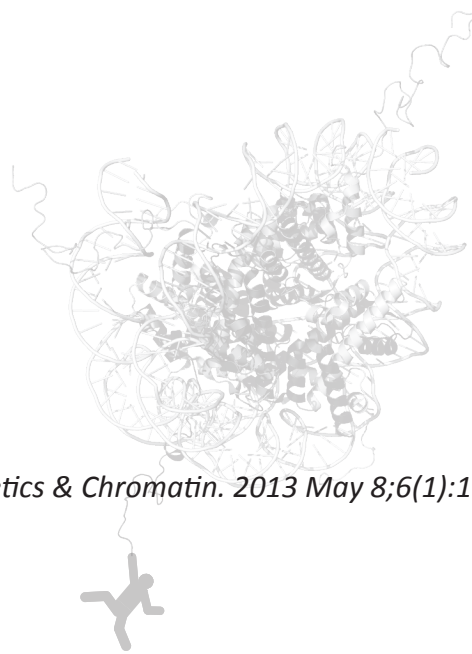
We thank dr. Steven Taylor for providing the Hela FRT cells, dr. Tony Hyman and dr. Ina Poser for providing the BAC cell line for Bap18, dr. Ali Shilatifard for the Ash2l, Rbbp5, Dpy30 and Wdr5 cDNA constructs and dr. Jason Swedlow for the Bod1 cDNA construct. This work was supported by the Netherlands Organization for Scientific Research (NWO) through a TOP grant (#700.57.302) to H.Th.M.T. and a VIDI to M.V., the Dutch Cancer Society (KWF) and by the Netherlands Proteomics Centre.



Chapter 4

Nucleosomal DNA binding drives the recognition of H3K36 methylated nucleosomes by the Psip1-PWWP domain

Rick van Nuland, Frederik M.A. van Schaik, Marieke Simonis, Sebastiaan van Heesch, Edwin Cuppen, Rolf Boelens, H.Th. Marc Timmers and Hugo van Ingen



Epigenetics & Chromatin. 2013 May 8;6(1):12

Abstract

Recognition of histone modifications by specialized protein domains is a key step in the regulation of DNA mediated processes like gene transcription. The structural basis of these interactions is usually studied using histone peptide models, neglecting the nucleosomal context. Here, we provide the structural and thermodynamic basis for the recognition of histone H3 lysine 36 methylated (H3K36me) nucleosomes by the Psip1-PWWP domain, based on extensive mutational analysis, advanced nuclear magnetic resonance (NMR), and computational approaches. The Psip1-PWWP domain binds H3K36me3 peptide and DNA with low affinity, through distinct, adjacent binding surfaces. PWWP binding to H3K36me nucleosomes is enhanced ~10,000-fold compared to a methylated peptide. Based on mutational analyses and NMR data, we derive a structure of the complex showing that the PWWP domain is bound to H3K36me nucleosomes through simultaneous interactions with both methylated histone tail and nucleosomal DNA**.** Concerted binding to the methylated histone tail and nucleosomal DNA underlies the high affinity, specific recognition of H3K36me nucleosome by the Psip1-PWWP domain. We propose that this bipartite binding mechanism is a distinctive, general property in the recognition of histone modifications close to the nucleosome core.

Introduction

Chemical modifications of nucleosomes, the complex of DNA and histone proteins that packages the eukaryotic genome, are key in the regulation of transcription, maintenance of genomic integrity, chromosome condensation and segregation (Fischle et al. 2003). Modifications such as acetylation or methylation of lysine residues of histone proteins can serve to recruit effector proteins to specific genomic sites (Kouzarides 2007). Methylation of lysine 36 in histone H3 (H3K36me) is conserved from yeast to human and is predominantly associated with actively transcribed chromatin (Barski et al. 2007). H3K36me has been implicated in diverse processes including splicing, DNA repair, repression of cryptic transcription and histone exchange (Wagner and Carpenter 2012). Recently, PWWP domains have been identified as H3K36me recognition domains, based on histone peptide interaction studies (Vermeulen et al. 2010). PWWP domains feature an aromatic cage as in other royal Tudor family proteins (Maurer-Stroh et al. 2003), that bind the methylated lysine side chain via cation-π interactions (Hughes et al. 2007). Interestingly, interaction studies have shown that PWWP domains bind methylated H3K36 histone tail peptides with very low affinity (Vezzoli et al. 2010; Wu et al. 2011), in stark contrast with the high affinity recognition of tri-methylated lysine 4 of H3 (H3K4me3) by PHD fingers (Li et al. 2006; Peña et al. 2006).

Here, we address the structural basis for H3K36me recognition by PWWP domains in the nucleosomal context. Unlike other methylated lysines in the unstructured N-terminal tail of histone H3, K36 is located at the point where the H3 tail protrudes from the nucleosome core (Luger et al. 1997). Since PWWP domains were previously also implicated in DNA binding (Lukasik et al. 2006; Laguri et al. 2008), we hypothesized that the close proximity of the nucleosomal DNA may critically contribute to binding affinity and/or specificity of PWWP domains for H3K36me. We concentrate on the PWWP domain containing protein Psip1, as its association with naked and chromatinized DNA has been the focus of previous studies (Llano et al. 2006; Turlure et al. 2006) and it was recently identified as a H3K36me interacting protein using synthetic peptides (Pradeepa et al. 2012). Psip1 (Ledgf/p75) was first isolated as an transcriptional co-activator (Ge et al. 1998) and tethers the HIV integrase to active host chromatin dependent on its PWWP domain (Ferris et al. 2010; Gijsbers et al. 2011). Psip1 is an essential subunit of the MLL complex in MLL oncogenic transformation via *HOX* gene regulation (Yokoyama and Cleary 2008), and is implicated in the homologous recombination pathway for DNA repair (Daugaard et al. 2012).

Using various experimental approaches, we show that concerted, low affinity interactions of the Psip1-PWWP domain with both nucleosomal DNA and methylated histone tail result in specific and high affinity binding to H3K36 methylated nucleosomes. During the final stages of our work, a similar conclusion was reached in another study (Eidahl et al. 2013). Our study underscores this notion with a NMR analysis of the PWWP-nucleosome complex, a structural model of the complex based on experimental interaction data and an extensive *in vitro* and

in vivo validation. Finally, based on a comparison with other PWWP domains and H3K36me binding modules, we propose that the bipartite recognition of methylated histone tail and nucleosomal DNA is a general feature of H3K36me recognition.

Results

H3K36 methylation dependent nucleosome binding by Psip1-PWWP

To characterize the interaction of the Psip1-PWWP domain with H3K36me nucleosomes, binding assays with immobilized GST-PWWP and mono-nucleosomal fractions from wild type or mutant *S. cerevisiae* were performed. Elimination of H3K36 methylation by deletion of the *SET2* histone methyltransferase gene or alanine substitutions of histone H3K36 strongly reduced the binding of nucleosomes to the PWWP domain (Fig. 1A). In contrast, loss of H3K4 methylation in a Δ*set1* strain or a H3K4A substitution had no effect on PWWP binding, but as expected eliminated the binding to the Taf3-PHD finger (Vermeulen et al. 2007).

To determine the contribution of residues neighboring H3K36, mono-nucleosomes carrying point mutations for residues 31-39 in histone H3 were used. Of these only V35A, K36A and K36R affect PWWP binding (Fig. 1B), suggesting limited involvement of the H3 amino acid sequence around the K36 methylation site in determining binding specificity. The specific interaction of the PWWP domain with mono-, di-, and tri-methylated H3K36 was confirmed using biotinylated H3 tail peptides (Fig. 1C and S1A) and mutation of W21 in the hydrophobic pocket of the domain completely abolished binding even in context of full length Psip1 (Fig. S1).

Next, immobilized GST-PWWP was used in binding to mono-nucleosomes prepared from mammalian cells. The bound fractions were enriched for H3K36me3 and H3K36me2 modifications whereas they showed little enrichment for H3K79me3 and H3K4me3. Comparable results were obtained using an extended fragment of the PWWP domain including the flanking AT-hook region (Fig. 1D) or full length Psip1 protein (Fig. S1B).

Adjacent PWWP surfaces bind weakly to H3K36me3 peptides and DNA

In order to address the structural basis for its interaction with H3K36 methylated nucleosomes, we solved the solution structure of the Psip1-PWWP domain (Fig. 2A,B, S2 and Table S1). As an Hdgf related PWWP domain (Wu et al. 2011), the core is formed by a five-stranded β-sheet core onto which two α-helices are packed. Psip1-PWWP residues M15, Y18, W21, and F44 form an aromatic cage acceptor for a methylated lysine side chain. Notably, the surface of the Psip1-PWWP domain is rich in basic residues that may interact with nucleosomal DNA (Fig. 2B). As a first step in dissecting the contributions of histone tail and nucleosomal DNA in the Psip1-PWWP–nucleosome interaction, NMR titration experiments were performed using a methylated histone peptide and a 10-bp DNA fragment.

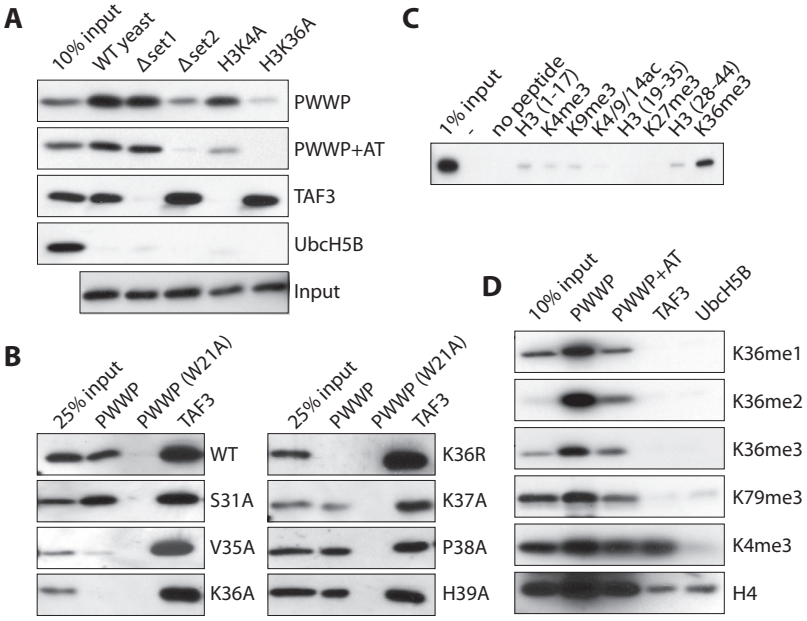


Figure 1. H3K36 methylation dependent nucleosome binding by the Psip1-PWWP domain. (A) Immunoblot analysis of GST pull downs with the indicated GST-fusion proteins incubated with mono-nucleosomes extracted from (mutant) yeast strains probed for histone H3. (B) GST pull downs with nucleosomes isolated from indicated (mutant) yeast strains. (C) Immunoblot analysis of pull downs with GST-PWWP lysates and the indicated biotinylated histone H3 peptides probed for GST. (D) GST pull down assay for GST-PWWP (PWWP), GST-PWWP including flanking AT-hook (PWWP+AT) or control proteins (Taf3, UbcH5B) were incubated with HeLa nucleosomes. Eluted proteins were detected by immunoblots with the indicated antibodies.

Addition of a H3K36me3 peptide (res. 28-41) resulted in clear chemical shift changes for the backbone amide resonances of residues around the aromatic cage and strand β4 of the PWWP domain (Fig. 2C,D left panels). The affinity for the H3K36me3 peptide is very low with a K_D of 17 mM (Fig. 2F left panel). In part, this may be due to the relatively closed conformation of the binding pocket when compared to crystal structures of related PWWP domains bound to methylated peptides (Fig. S2C). Strikingly, this very low affinity interaction is completely dependent on the methylation of H3K36, as no changes in chemical shift upon addition of non-methylated H3K36 peptide were observed, even at 11 mM of peptide (Fig. S3).

Addition of a dsDNA fragment that was previously suggested as a substrate for Psip1-PWWP (Singh et al. 2006) resulted in chemical shift changes for a distinct set residues localizing to a single basic patch on the PWWP surface (Fig. 2C and D, right panel). A K_D of 150 ± 50 μM was determined, indicating a low binding affinity for DNA (Fig. 2F, right panel). Imino-proton resonances of DNA base pairs did not change in chemical shift or relative intensity during the titration, suggesting that the interaction lacks sequence specificity (Fig. S4).

The DNA and histone interaction surfaces of the PWWP domain are adjacent and overlap

with areas of positive or negative potential, respectively (Fig. 2E). This arrangement provides an excellent electrostatic match to the negative phosphate backbone of the DNA and positive histone tail in the context of the nucleosome.

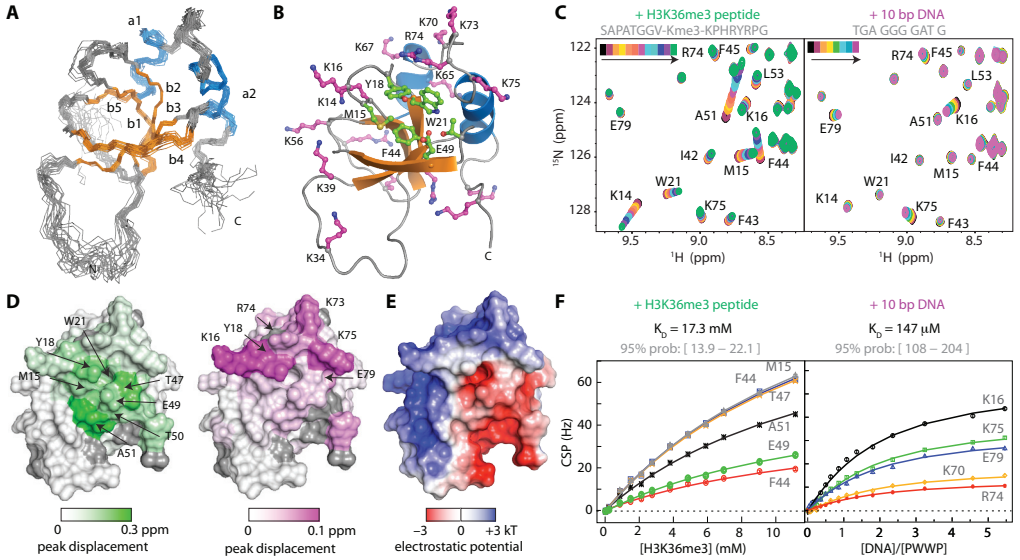


Figure 2. Interaction with H3K36me3 methylated peptides and dsDNA. (A) Ensemble of 20 best solution structures of Psip1-PWWP domain; orange – β-sheet; blue – α-helix; gray – other. (B) Aromatic cage (green) and basic residues (magenta) shown as balls-and-sticks. Residues that have been mutated in this study are labeled. (C) Sections of the 2D ^{1}H-^{15}N HSQC spectrum of PWWP during the titration with H3K36me3 peptide (left panel) and dsDNA fragment (right panel). Free PWWP spectrum in black; resonances of interest are labeled. (D) Interaction surfaces for the histone peptide (left) and dsDNA (right), coded on the van der Waals surface. Grey is used for residues w/o data; residues with shifts larger than 10% trimmed mean + 2 σ are labeled. (E) Electrostatic potential on the solvent accessible surface color (F) Chemical shift perturbation derived binding curves (symbols) including best fits (solid lines) for H3K36me3 peptide (left panel) and dsDNA (right panel).

Aromatic cage and basic surface determine binding specificity and affinity

In order to study the interaction of Psip1-PWWP with H3K36me3 in the nucleosomal context, we made use of the methylated lysine analogue (MLA) approach (Simon et al. 2007), in which a cysteine is alkylated to result in a thioether mimic of the methylated lysine (referred to as H3K$_C$36me3). Nucleosomes reconstituted from recombinantly expressed histones including H3K$_C$36me3 are able to bind endogenous Psip1 from a HeLa nuclear extract in a pull down assay (Fig. 3A). In a control experiment, a H3K$_C$36me3 peptide and its native counterpart were found to bind the Psip1-PWWP domain to a similar degree in a peptide pull down assay (Fig. 3B). Furthermore, comparison of NMR titration experiments of PWWP domain with either a native H3K36me3 or a H3K$_C$36me3 peptide showed that peptide binding affects the same PWWP residues in a highly similar manner (Fig. 3C). Based on the observed binding curve, we estimated a K_D value of 11 mM for the H3K$_C$36me3 peptide (95% confidence interval: 6< K_D <26

mM), comparable to the K_D of the native peptide. This correspondence increases further when restricting the fit to the same titration interval, in which case the best-fit K_D is 13 mM for the native H3K36me3 peptide (95% confidence interval:9< K_D <22 mM) (Fig. 3D).

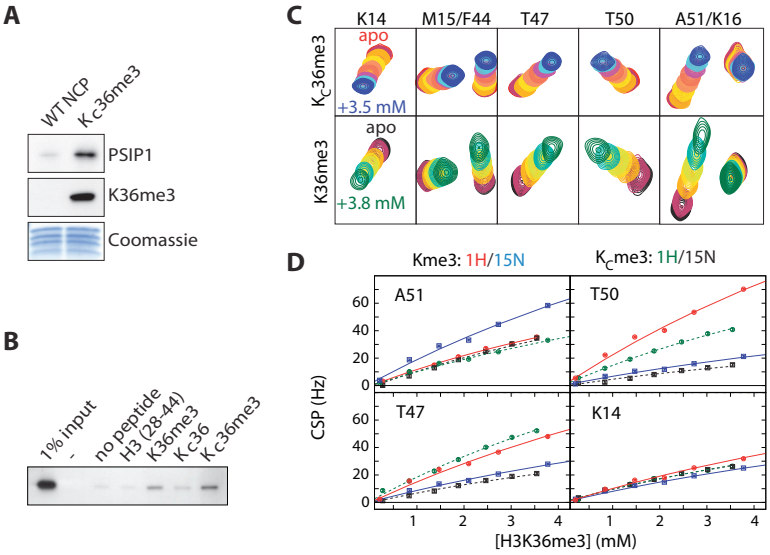


Figure 3. H3K$_C$36me3 MLA nucleosomes and peptides bind the Psip1-PWWP domain with comparable affinities as their native counterparts. (A) Recombinant MLA nucleosomes bind to endogenous Psip1. Nucleosomes were immobilized on beads using biotinylated 601 DNA and incubated with HeLa nuclear extract. Bound proteins were analyzed by immunoblot with the indicated antibodies and by coomassie blue staining to control for the amount of histones. (B) Psip1-PWWP binds preferentially to H3K36me3 and H3K$_C$36me3 peptides. Indicated peptides were biotinylated and immobilized, incubated with GST-PWWP lysates and analyzed by immunoblot with GST antibodies. (C) H3K36me3 (lower panels) and H3K$_C$36me3 peptides (upper panels) bind with similar affinity to same pocket in Psip1-PWWP. An overlay of sections of the 2D ^{1}H-^{15}N HSQC spectrum is shown for each NMR titration point. Assignments of resonances of interest are indicated on top; color-coding is indicated in the figure. (D) Chemical shift perturbation (CSP) derived binding curves and best fits for H3K36me3 and H3K$_C$36me3 peptide. Only points up to 4 mM of peptide were included in the fit. Curves for a selection of residues are shown; color-coding is indicated in the figure.

Having established the validity of the MLA approach, an extensive mutational analysis of the PWWP domain was carried out to dissect the contributions of the histone tail and DNA interaction surfaces to the binding affinity and specificity for methylated nucleosomes using electrophoretic mobility shift assays (EMSA). Wild type GST-PWWP preferentially binds to methylated over non-methylated nucleosomes (Fig. 4A). A residual band shift observed in case of non-modified nucleosomes suggests that Psip1-PWWP can also associate non specifically to nucleosomes, at least under the low ionic strength conditions of the EMSA experiment. Alanine substitutions of the aromatic cage residues M15 and Y18 strongly reduce specificity towards H3K$_C$36me3 nucleosomes as shown from the similar patterns observed for modified and non-modified nucleosomes (Fig. 4A). Notably, their effect on binding affinity is much less pronounced. Likewise, alanine mutation of two residues flanking the other side of the aromatic

cage (H48 and E49) results predominantly in a loss of specificity (Fig. S5A). Alanine mutations of W21 and F44 that are both part of the aromatic cage showed no binding whatsoever (Fig. 4A). These mutations may interfere with proper folding of the domain as they are part of the hydrophobic core of the protein and suggested by their reduced soluble expression levels.

Mutations in the putative nucleosomal DNA binding surface show a markedly different result. The K70A mutant binds with a lower affinity to nucleosomes, but retains preference for the modified nucleosomes. Strikingly, alanine substitutions of the solvent exposed R74 abolished the interaction with nucleosomes (Fig. 4A). In contrast, the K34A mutant showed a comparable binding pattern as wild type PWWP (Fig. 4A). Nearly all charge mutants showed severely reduced binding, but retained specificity for $H3K_C36me3$ nucleosomes (Fig. S5A). Of these, K39 (shown in Fig. 4A) and K56 were initially not found to interact with a DNA fragment in the NMR titration experiment, suggesting that binding to nucleosomal DNA involves a larger interaction surface.

To investigate the *in vivo* relevance of our findings we examined the contribution of the PWWP domain to the mobility and distribution of Psip1 in cells. To this end WT and mutant PWWPs were introduced in the context of full length Psip1 fused to GFP and expressed in HeLa cells (Fig. S6). The nuclear mobility of the GFP-Psip1 proteins was measured by fluorescence recovery after photobleaching (FRAP). HeLa cells expressing NLS-GFP or histone H2B-GFP were included as highly mobile and immobile controls, respectively (Fig. 4B left panel). FRAP curves for GFP-Psip1 are indicative of transient chromatin binding (Fig. 4B green curve), consistent with previously published results (Hendrix et al. 2011). Disruption of domain integrity (W21A), as well as mutations in the aromatic cage (M15A) or the DNA interaction surface (K70A, R74A), resulted in faster recovery of fluorescence after bleaching (Fig. 4B center and right panels), demonstrating the requirement of both DNA and histone tail interaction surfaces for stable association with chromatin.

To examine the effect of PWWP mutations on the genomic distribution of Psip1, the GFP-Psip1 cell lines were used for chromatin immunoprecipitation followed by high throughput sequencing (ChIP sequencing). To correlate binding of the Psip1 proteins to H3K36 methylation, genes were divided into two groups: high H3K36me3 (red) or low H3K36me3 (blue) containing genes (Fig. S7). Wild type Psip1 protein was selectively enriched on high H3K36me3 genes (Fig. 4C and S7B), while aromatic cage mutant M15A showed no enrichment, in accordance with our *in vitro* data. Mutation of K70 did not significantly affect the genomic distribution of Psip1 in correspondence with the mild effect on the affinity for nucleosomes of this mutant in EMSA.

Concerted binding of methylated histone tail and nucleosomal DNA

The PWWP–$H3K_C36me3$ nucleosome interaction was further analyzed using state-of-the-art solution NMR tailored for large supramolecular complexes. Recently, it was shown that a comprehensive characterization of protein–nucleosome interactions can be obtained (Kato et al. 2011) using methyl group based NMR (methyl TROSY) (Tugarinov et al. 2003; Zhou 2004),

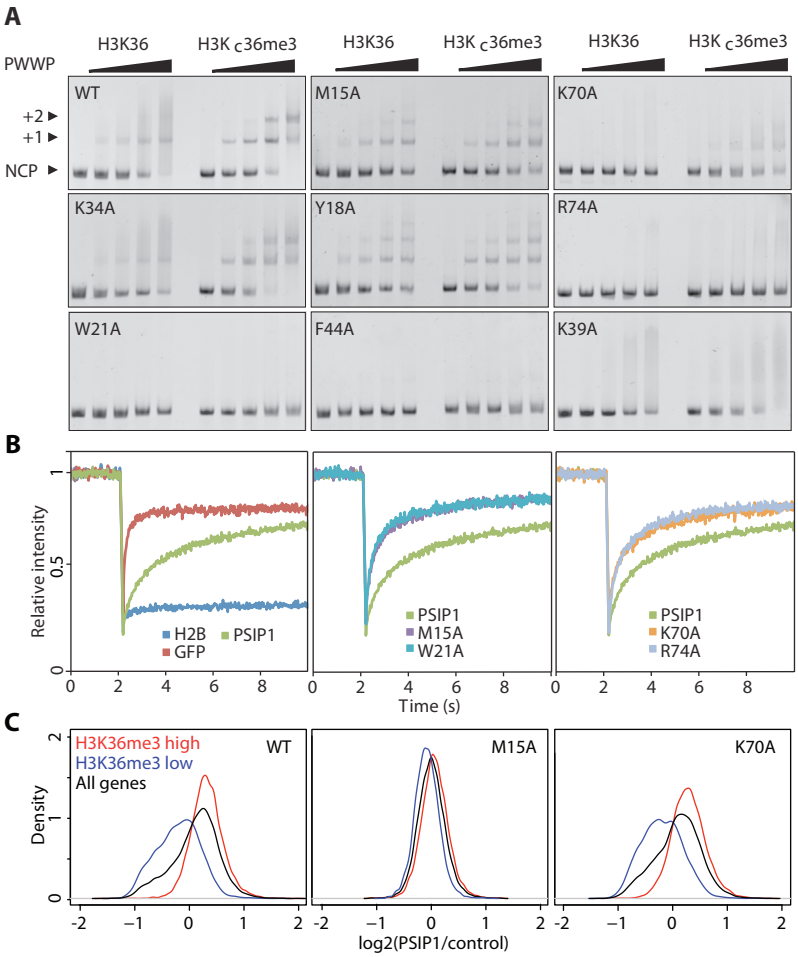


Figure 4. Importance of the aromatic cage and basic surface of the PWWP domain in binding affinity and specificity. (A) Titration of non-modified (H3K36) or modified nucleosomes (H3K$_c$36me3) with 0, 0.5, 1, 2 and 3 molar equivalents of the indicated GST-PWWP protein added and analyzed by staining EMSA gels with ethidium bromide. The free nucleosome particles (NCP) or complexes of either one (+1) or two (+2) PWWPs bound to nucleosome and the PWWP mutant are indicated. (B) FRAP-recovery curves for GFP-Psip1, indicated mutants or controls H2B-GFP or NLS-GFP. An average of 10 cells is presented. (C) Density plots of ChIP analysis are shown for the ratios Psip1 (or mutant)/control for all genes (black), genes enriched for H3K36me3 (red) and genes not enriched for H3K36me3 (blue).

in which only the histone methyl groups of isoleucine, leucine and valine (ILV) residues are observed.

Here, we used MLA nucleosomes with ILV methyl group labeled histone H3 (spectrum in Fig. S8A) and monitored the H3V35 methyl groups to probe the Psip1-PWWP interaction in a site specific manner. Notably, the tri-methyl lysine mimic itself is not observable in these experiments as it is not isotope labeled. In the unbound state, the methyl group resonances of H3V35 (as well as H3L20) are very intense compared to those from the nucleosome core (Fig. S8B), reflecting the highly dynamic nature of the N-terminal tail. Addition of ILV-labeled PWWP

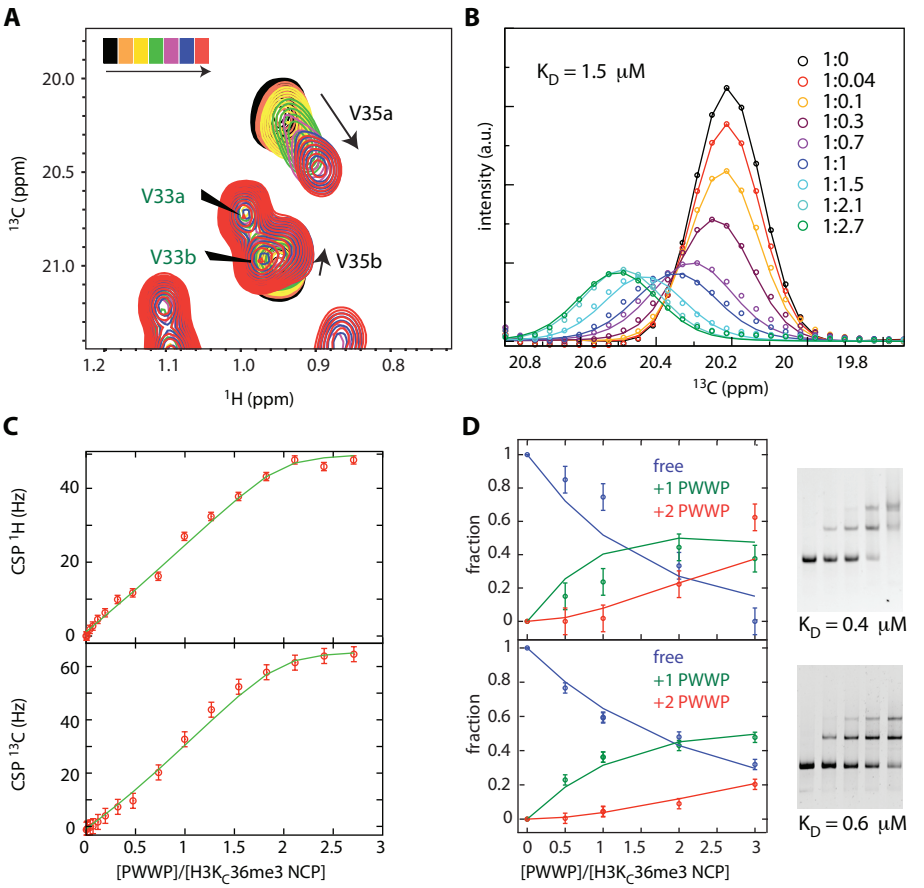


Figure 5. Specific and high affinity interaction of Psip1-PWWP with H3K$_C$36me3 nucleosomes. (A) Sections of 2D ^{13}C-^{1}H methyl TROSY spectra focusing on the resonances of the H3V35 methyl groups. Black spectrum corresponds to free nucleosomes and red to 1:2.7 molar ratio. PWWP ILV methyl groups are indicated with green labels. H3V35a/b refer to either of the V35 γ1/γ2 methyl groups. (B) Experimental (points) and fitted (solid lines) line shapes of H3V35a resonance, taken through the peak maximum in the ^{13}C dimension, at the indicated molar ratios of nucleosome:PWWP. Best fit K_D is 1.5 µM (95% probability K_D<8 µM) and K_{off} is 500 s-1 (95% probability 400< K_{off} <1000 s-1). (C) Line shape derived binding curve, highlighting the saturation of the binding sites. Change in H3V35a peak position is plotted as a function of PWWP domain added; points – experimental values, lines – fitted values. (D) EMSA based binding curves and fits for two independent titrations of H3K$_C$36me3 nucleosomes with GST-PWWP (left), along with the ethidium stained gels (right). Blue/green/red - integrated density of free/+1 PWWP/+2 PWWP nucleosome band. Estimated K_D's are indicated.

domain resulted in a clear change in peak position of the methyl groups of H3V35, without affecting other methyl groups in the nucleosome core or the H3 tail (Fig. 5A and S8A). Binding of Psip1-PWWP also causes a local loss of flexibility of the H3 N-terminal tail around H3K36, as indicated by the sharp decrease in H3V35 peak intensity in the bound state, but not that of H3L20 (Fig. 5B and S8B). For Psip1-PWWP, the methyl group of I42, close to the aromatic cage, showed a distinct change in chemical shift during the titration experiment (Fig. S8A). Together, these changes demonstrate the specific interaction between the Psip1-PWWP aromatic cage

and the methylated H3 tail in the nucleosomal context.

Based on fitting the experimental line shapes of the H3V35a methyl group to a 1:2 (nucleosome:PWWP) binding model, we find that the dissociation constant of Psip1-PWWP binding to the $H3K_C36me3$ side chain within the nucleosome is ~1.5 μM (Fig. 5B, 5C and S9). The interaction is highly dynamic: the dissociation rate (k_{off}) is ~500 s^{-1}, corresponding to a lifetime of the complex ($1/k_{off}$) of ~1-2 ms. The affinity found here at physiological ionic strength is comparable to the K_D value estimated from the gel shift assay (~0.5 μM), recorded at lower ionic strength and temperature (Fig. 5D). Strikingly, the affinity of Psip1-PWWP for methylated nucleosomes is four orders of magnitude higher than for a methylated peptide (K_D 17 mM) and two orders higher than for isolated DNA (K_D 150 μM). The enhanced affinity is due to simultaneous binding of both methylated histone tail (see chemical shift perturbations of H3V35, Fig. 5B) and nucleosomal DNA (see the binding defects of DNA interaction surface mutants, Fig. 4A). The magnitude of such enhancement in binding affinity upon linking of two binding sites cannot simply be predicted from the affinities for the isolated binding sites, as it depends crucially on the relative orientation of the linked sites and the length and flexibility of the linker (Zhou 2004; Jencks 1981). Following the framework of Zhou (Zhou 2004), the enhancement may be expressed in the form of an effective concentration given by $(K_{D,tail} \times K_{D,DNA})/ K_{D,nucleosome}$, which in our case evaluates to $1.5\times10^{-4} \times 1.7\times10^{-2} / 1.5\times10^{-6}$ = 1.7 M. This enhancement value is significantly higher than typical values in the mM range as found for linked DNA binding domains or bivalent pharmaceuticals (Zhou 2004; Rees et al. 2004). This suggests that there are limited entropic losses and structural rearrangements upon binding. Thus, our data indicate that both the DNA and histone interaction surfaces of Psip1-PWWP domain combine in a concerted manner to result in high affinity binding to H3K36me3 nucleosomes.

Structure of Psip1-PWWP – H3K36me3 nucleosome complex

Our experimentally observed chemical shift perturbations and mutational analysis were used to derive a structural model for the Psip1-PWWP–H3K36me3 nucleosome complex using the docking program HADDOCK (Dominguez et al. 2003). In order to reliably sample a large conformational space for the flexible H3 tail, the flexible multi-domain docking protocol (Karaca and Bonvin 2011) was used. In addition, a large surface of DNA around the H3 tail exit site was systematically explored to sample all possible interaction sites on the nucleosomal DNA (Fig. S10). After clustering and cross-validation of the final ensemble of solutions (Fig. S11, Table S2), we find one cluster of solutions that is in agreement with the cross-validation data, with the lowest energy structure shown in Figure 6A. The PWWP domain bridges the two DNA gyres around super helical location (SHL) -1/+6, right at the H3 tail entry/exit site. There is an excellent electrostatic match with the nucleosomal DNA, while at the same time the H3K36me3 side chain is snugly captured by the aromatic cage (Fig. 6B,C). Notably, the H3V35 methyl groups are more than 6 Å away from ILV methyl groups in the PWWP domain in the model, consistent with

the absence of intermolecular NOEs (data not shown).

PWWP-DNA contacts made by the basic residues are mainly to the phosphate or sugar backbone (Fig. 6D). The sequence specific contacts of residue R74 seen in the lowest energy-structure are not conserved in the cluster of solutions. Overall, residues K73, R74 and K75 bind the DNA non specifically around SHL +6, while K39 and K56 interact with the other DNA gyre at SHL –1. Residue K16 and K14 sit in between the two gyres and can interact with either. Notably, residues K67 and K70 do not mediate intermolecular interactions in many structures of this cluster, which may reflect the relatively minor binding defects of their alanine mutants in EMSA (see also Table S2).

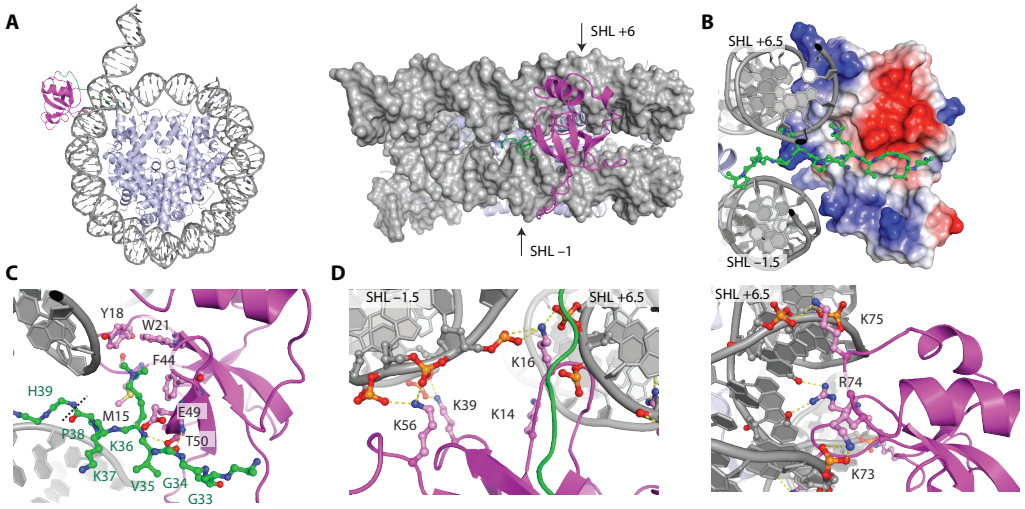


Figure 6. Structural model of Psip1-PWWP–H3K36me3 nucleosome complex. (A) Lowest energy structure of the cluster that is in agreement with the cross-validation mutation data. (B) Detailed view showing H3 residues 31-38 as balls-and-sticks, the DNA in cartoon representation and the van der Waals surface of the PWWP domain color coded by electrostatic potential. (C, D) Detailed view on the intermolecular interactions, focusing on the K36me3 recognition by the aromatic cage (C) and the DNA contacts mediated by K14, K16, K39 and K56 (D, left) and K73, R74 and K75 (D, right). Hydrogen bonds are shown as dashed yellow lines. Color-coding: magenta – PWWP domain, green – H3 tail, grey – DNA. PWWP residues are labeled in black, H3 residues in green. The dashed line in (C) indicates the peptide bond at which the H3 tail was cut in the flexible multi-domain docking protocol.

Discussion

Here we determined the molecular basis of H3K36me nucleosome recognition by the Psip1-PWWP domain. We show that the interaction with nucleosomal DNA is responsible for a ~10,000-fold enhancement in binding affinity into an *in vivo* relevant range. A similar conclusion was reached in the work of Eidahl et al. from a comparison of an NMR based estimated binding affinity for H3K36me3 peptides and a pull down assay based measurement of the affinity for

$H3K_C36me3$ nucleosomes (Eidahl et al. 2013). While full length Psip1 contains additional DNA binding domains (Turlure et al. 2006; Tsutsui et al. 2011), disruption of the PWWP basic surface markedly reduces *in vivo* chromatin binding ability of full length Psip1 as shown in this work and previously (Hendrix et al. 2011). Moreover, mutations in the DNA interaction surface were previously shown to result in a dramatic reduced HIV-infectivity in cells (Shun et al. 2008), underscoring the functional significance of the bipartite nucleosome binding for HIV integration and other Psip1 dependent cellular processes.

Most H3K36me3 binding proteins depend on a PWWP domain for proper chromatin binding, despite their low affinity for methylated peptides. Therefore, nucleosomal DNA binding may be a general driving force for the recognition of H3K36me3 nucleosomes *in vivo*. Superposition of homologous PWWP domains onto the structural model of the complex shows that these domains all share similar configurations of an aromatic cage and basic patches that potentially enable concerted binding to both methylated H3K36 side chain and nucleosomal DNA (Fig. 7A). This holds true both for the close homolog Hdgf2 PWWP domain, and more distantly related PWWP domains of Msh6, Dnmt3b, Brpf1 and Whsc1l1. These proteins are involved in the DNA damage response, DNA and histone methylation and acetylation. Moreover, the bipartite binding mode observed for Psip1 may extend beyond PWWP domains to other domains that bind H3K36me as well as to the recognition of other histone modifications close

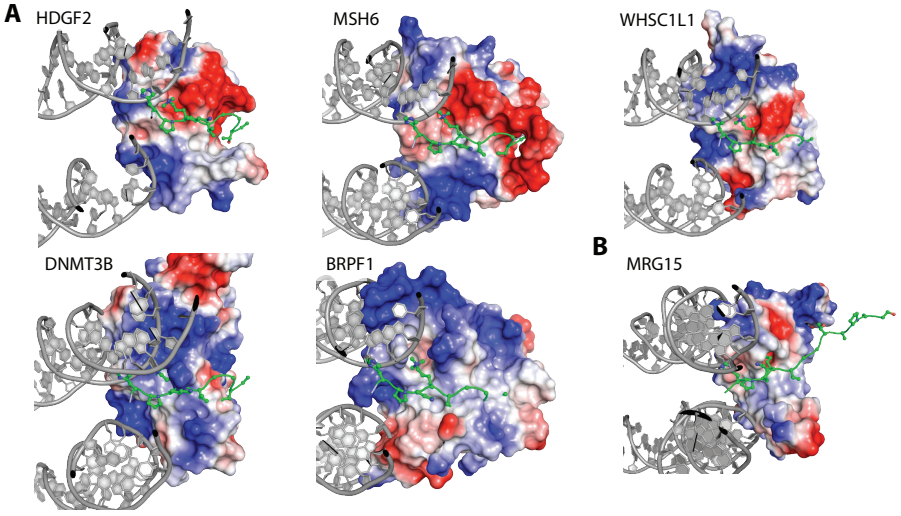


Figure 7. Conservation of bipartite recognition of H3K36me nucleosomes. (A) The PWWP domains of Hdgf2 (3QBY), Msh6 (2GFU), Dnmt3b (1KHC), Brpf1 (2X4X) and Whsc1l1 (2DAQ) were aligned with the Psip1-PWWP domain in our model of the PWWP–H3K36me3 nucleosome complex. Detailed view showing H3 residues 31-38 as balls-and-sticks, the DNA in cartoon representation and the van der Waals surface of the PWWP domain color coded by electrostatic potential. (B) The chromodomain of Mrg15 (2F5K) was manually docked on the H3K36me3 nucleosome model, positioning the aromatic cage around H3K36me3 and orienting its basic patches towards the DNA, while minimizing interatomic clashes.

to the nucleosomal DNA, such as H4K20me. The Mrg15 chromodomain that binds H3K36me also shows appropriately configuration of an aromatic cage and basic regions (Fig. 7B). Similarly, another chromodomain and PWWP domain that bind H4K20me have been shown to have a distinct basic binding surface for DNA (Kim et al. 2010; Qiu et al. 2012). In contrast, PHD finger domains that specifically bind to H3K4me3 at the N-terminus of H3, have an extensive negative surface potential (van Ingen et al. 2008) and are therefore likely repelled from the nucleosomal DNA.

In conclusion, we propose that recognition of H3K36 methylated chromatin not only occurs through the methylated lysine side chain and its amino acid sequence context, but also through the nucleosomal DNA. We propose that this mechanism also applies to the recognition of other modifications close to the nucleosome core, such as H4K20me and H3K79me. This mechanism testifies to the fact that recognition of histone modifications relies on the binding to modified histone residues embedded in the chromatin fiber. Just as histones are not merely packaging material in chromatin, the nucleosomal DNA is not inert in the readout of the epigenetic modifications.

Materials & Methods

Protein expression and purification

The PWWP domain of human Psip1 (3-100) or the PWWP+AT (3-207) were expressed as GST-fusions in BL21-DE3 or Rosetta 2 bacterial strains at 37°C in either LB medium or M9 minimal medium with $^{15}NH_4Cl$ and/or ^{13}C-glucose. The protein was purified by binding to a Glutathione agarose (GA) column (Sigma) and eluted with 50 mM reduced glutathione (Sigma). After thrombin digestion, PWWP was purified over a Sephadex-75 (HiLoad 16/60) column in buffer A (50mM Tris-HCl pH 7, 100 mM KCl, 1 mM DTT, 1 mM EDTA, 0.5 mM PMSF and protease inhibitors), applied to a MonoS HR5/5 in buffer A and eluted using a linear gradient (100 to 1000 mM KCl).

Drosophila histones were expressed, purified and alkylated as previously described (Dyer et al. 2004; Simon et al. 2007). Histones used for NMR studies were produced in M9 minimal medium containing desired isotopes. Methyl labeling of Ile-d1-[$^{13}CH_3$] and Val/Leu-[$^{13}CH_3$, $^{12}CD_3$] (ILV-labeling) followed the procedure of Tugarinov (Tugarinov et al. 2003).

Antibodies and plasmids

α-GST (SC), α-Psip1 (Bethyl A300-848A), α-H3K4me3 (Abcam ab8580), α-H4 (Upstate 07-108), α-H3K79me3 (Abcam ab2621), α-H3K36me1 (Abcam ab9048), α-H3K36me2 (Cellsignaling 9758), α-H3K36me3 (Abcam ab9050) and α-GFP (gift from Geert Kops) were used for ChIP and immunoblotting.

All GST fusions were cloned into pRPN265NB. Psip1 cDNA was introduced into pEGFP_C using the Gateway system (Invitrogen). All point mutations were created using site directed mutagenesis (Stratagene). Stable GFP tagged Psip1 HeLa lines were created by cloning Psip1 into pCDNA.5/FRT/TO (Invitrogen) and subsequent recombination into HeLa FRT cells carrying the Tet repressor for inducible expression (Tighe et al. 2008).

Nucleosome and peptide pull downs

Mono-nucleosomes were extracted from HeLa or yeast cells by MNase treatment of lysed cells as previously described (Vermeulen et al. 2007). Histone H3 mutants were selected from a mutant library (Dai et al. 2008). GA beads (Sigma) were covered with GST-fusion proteins, mixed with mono-nucleosomes and washed. Eluted proteins were analyzed by immunoblotting. HeLa mono-nucleosomes were incubated with premixed GFP-fusion protein and GFP binder beads (CHROMOTEK) and analyzed in a similar way. Peptide pull downs were performed as described previously (Vermeulen et al. 2007).

Nucleosome reconstitution and EMSA

The 601 DNA 'Widom' template was amplified using PCR, purified using anion exchange chromatography and used for reconstitution using salt-gradient deposition. Nucleosomes were incubated with GA purified GST-PWWP protein in 0.2X TBE and analyzed by native 5% 60:1 acryl-bis gel electrophoresis. Either 1.5 or 3 pmol of nucleosome was used in all experiments and up to 3 molar equivalents of protein in 8 μL load volume. All steps were performed at 4°C. Gels were stained with ethidium-bromide and analyzed on a Gel-Doc XR+ system (Bio-Rad). If applicable, band densities corresponding to free, singly and doubly bound nucleosomes were quantified using ImageJ software package and subsequently fit together to a 2:1 binding model using in-house written MatLAB routine (MATLAB version 7.13.0, The MathWorks Inc., 2011).

Strip-FRAP

FRAP studies were performed using a Zeiss 510 META confocal LSM as previously described (de Graaf et al. 2010). GFP protein expression was induced with 0.5 μg/ml doxycycline for 5 hours.

Chromatin immunoprecipitation

Chromatin preparation and ChIP were performed essentially as described (de Graaf et al. 2010; Mokry et al. 2010). Libraries were sequenced on AB/SOLiD 5500XL, producing 48-bp reads. Sequencing reads were mapped with Burrows-Wheeler Aligner (BWA-0.5.8c) (settings: –c –l 25 –k 2 –n 10) (Li and Durbin 2009). As a gene set the known protein coding genes as annotated in Ensembl 67 were used (http://www.ensembl.org). The number of reads mapped to each gene

was normalized to the total number of reads mapping inside genes per sample. A separation of H3K36me3 enriched and non-enriched was made based on the density plot of the read density. Genes were filtered to have at least 50 sequencing reads in the GFP tagged Psip1 ChIP sequencing data. All plots were created using the R package (http://www.r-project.org/).

NMR samples

Samples used for assignment and structure calculation contained ca. 1 mM PWWP domain in 90/10% H2O/D2O with 20 mM NaPi buffer at pH 6.2. Interaction studies were done at 0.3 mM PWWP in 20 mM NaPi pH 7 with 100 mM NaCl. ILV-labeled $H3K_c36me3$ nucleosome sample contained 116 μM nucleosome in 20 mM NaPi pH 7 with 100 mM NaCl.

Peptides were extensively lyophilized and dissolved in NMR buffer to a stock concentration of ~110 mM. Cysteine peptides were alkylated according to the MLA protocol (Simon et al. 2007) and purified using a Sephadex G-10 column followed by cation exchange chromatography. The purity of the peptide was confirmed by NMR. Annealed DNA oligos (Eurogentec, Belgium) were lyophilized and dissolved in NMR buffer to a stock concentration of 11.5 mM. Titration of $H3K_c36me3$ nucleosomes was done using a PWWP stock of 1.28 mM.

Psip1-PWWP structure determination

NMR experiments for assignment, and structure calculation of the Psip1-PWWP domain were carried out at 293K on a 600 or 750 MHz Bruker Avance II spectrometer. Processing was done using the NMRPipe package (Delaglio et al. 1995). Spectra were analyzed using Sparky (Goddard and Kneller, UCSF). Backbone assignments were obtained using MARS (Jung and Zweckstetter 2004) based on HNCACB and CBCACONH spectra. Side chain resonances were assigned using CCH-TOCSY, CBHD and NOESY spectra. Overall assignment completeness was 97.1% for all non-labile protons. Backbone dihedral angle restraints were derived using TALOS+ (Shen et al. 2009). Distance restraints were derived from ^{13}C and ^{15}N-edited 3D NOESY spectra (mixing time 120 ms). The NOE cross peaks were assigned and converted into distance restraints using CYANA 3.0 (López-Méndez and Güntert 2006; Herrmann et al. 2002). First, 10 ensembles of 100 structures were calculated by using CYANA using different random number seeds. Out of the 10 resulting distance restraint lists, only the restraints that were reproduced in all cases were retained to produce a final restraint list. This final list was then used to calculate 100 structures in CNS 1.2 (Brünger et al. 1998), which were subsequently refined in explicit water by using the RECOORD protocol (Nederveen et al. 2005). The final ensemble contained the 20 lowest-energy structures, contained neither distance violations > 0.5 A˚, nor dihedral angle violation > 5°, and was validated by using the iCing validation suite (Doreleijers et al. 2012).

Titration experiments and data analysis

Interaction studies of the Psip1-PWWP domain were carried out at 293K on a 600 MHz Bruker Avance II spectrometer Nucleosome spectra were recorded at 308K on a 900 MHz Bruker Avance III spectrometer with a TCI cryo-probe. Titration data were fitted using MatLAB scripts either using the fast-exchange assumption in case of fitting CSP derived binding curves or using explicit evaluation of the exchange matrix, and subsequent calculation of the FID in case of line shape fitting (see Supporting Materials in Kato et al. (Kato et al. 2011) for details).

Molecular graphics

All molecular graphics were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC). Electrostatic surfaces were calculated using the adaptive Poisson-Boltzman solver (Baker et al. 2001) and the AMBER force field.

Docking protocol Psip1-PWWP–nucleosome complex

We used our experimental chemical shift perturbation, transferred-NOESY and mutagenesis data, together with available literature data to create a structural model for the Psip1-PWWP–nucleosome complex with Haddock version 2.1 (Dominguez et al. 2003) and CNS 1.3 (Brünger et al. 1998; Brunger 2007). In what follows, we describe the docking procedure.

In short, the docking was divided in two stages: i) docking of the H3 N-terminal tail to the Psip1-PWWP domain guided by the chemical shift perturbation, transferred-NOESY and mutation data, and using homology derived interaction restraints from the homologous Brpf1-H3K36me3, Hdgf2-H4K20me3/H3K79me3 crystal structures; ii) docking of the PWWP-H3K36me3 complex to the nucleosome, again guided by the chemical shift perturbation and mutation data, together with restraints to enforce the covalent attachment of the H3 tail to the remainder of H3. This approach was based on the flexible multi-domain docking protocol described by Karaca et al. (Karaca and Bonvin 2011). It allows to dock the Psip1-PWWP domain efficiently to both the K36me3 side chain in the H3 tail and to the nucleosomal DNA, and at the same to sample a large conformational space for the flexible H3 tail. This procedure is described in details in the Supplementary material.

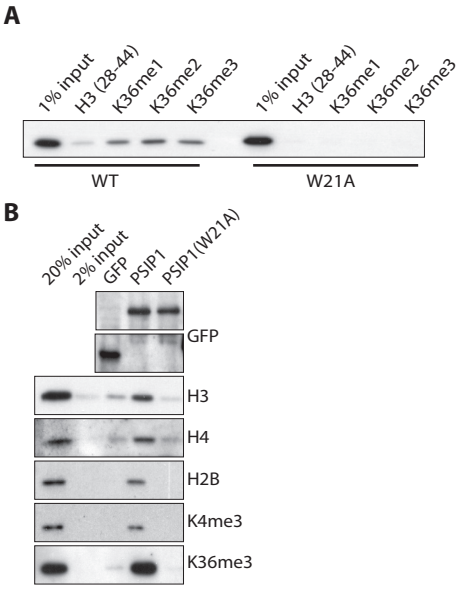
PDB ACCESSION CODES

The solution structure of the Psip1-PWWP domain is accessible from the Protein Data Bank, PDB-ID 3ZEH. The structural model of the H3K36me3-nucleosome–Psip1-PWWP domain complex is deposited under PDB-ID 3ZH1 and is also available from: www.nmr.chem.uu.nl/~hugo.

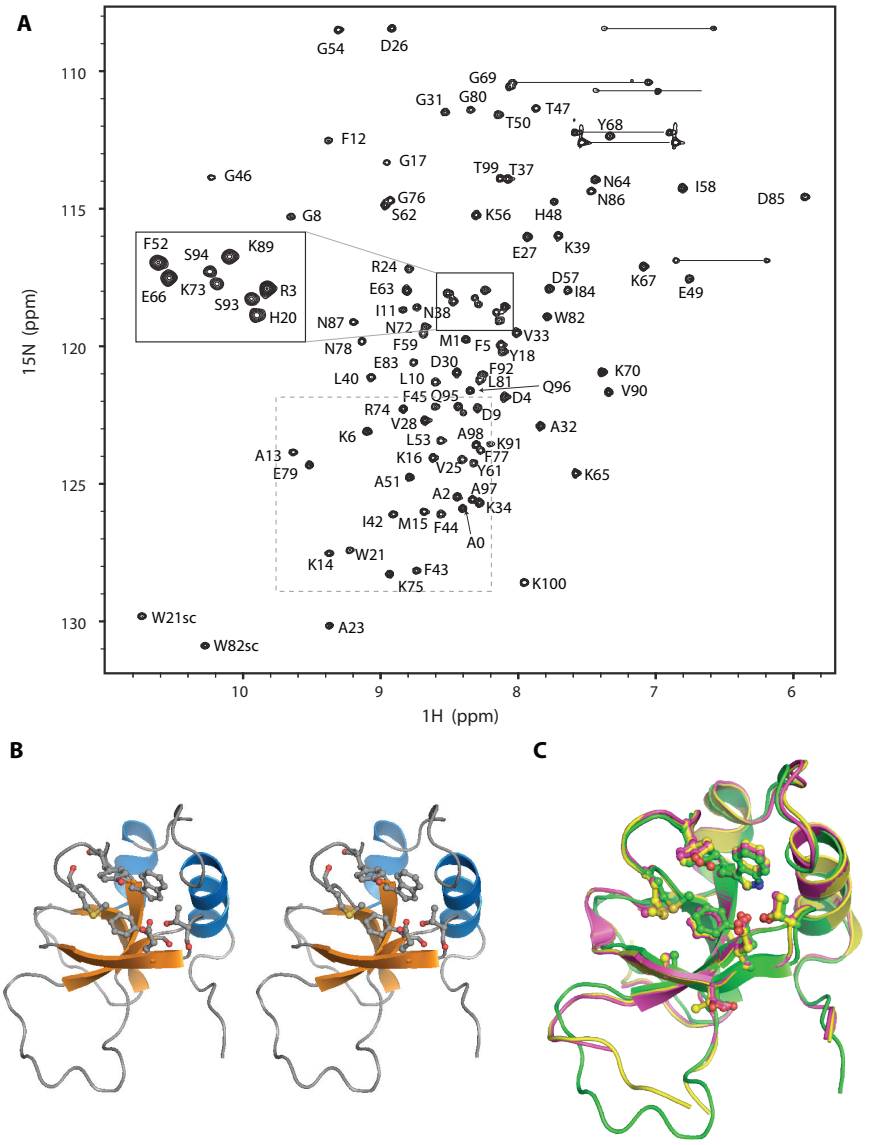
Acknowledgements

We thank dr. S. Taylor, dr. R. Roeder, dr. Y. Bai and dr. G. Kops for reagents, dr. P. de Graaf for help with FRAP and the HADDOCK Team (dr. A.S.J. Melquiond, dr. E. Karaca, dr. M. van Dijk and prof.dr. A.M.J.J. Bonvin) for valuable discussions on the docking. This work was supported by the Chemical Sciences division of Dutch Science foundation NWO through a Veni fellowship to H.v.I. (700.59.401), a NWO-Chemical Sciences TOP grant to H.Th.M.T. (700.57.302) and by the Netherlands Proteomics Centre, R.B. was supported by NWO-Chemical Sciences for the 900 MHz NMR and the TCI cryoprobe.

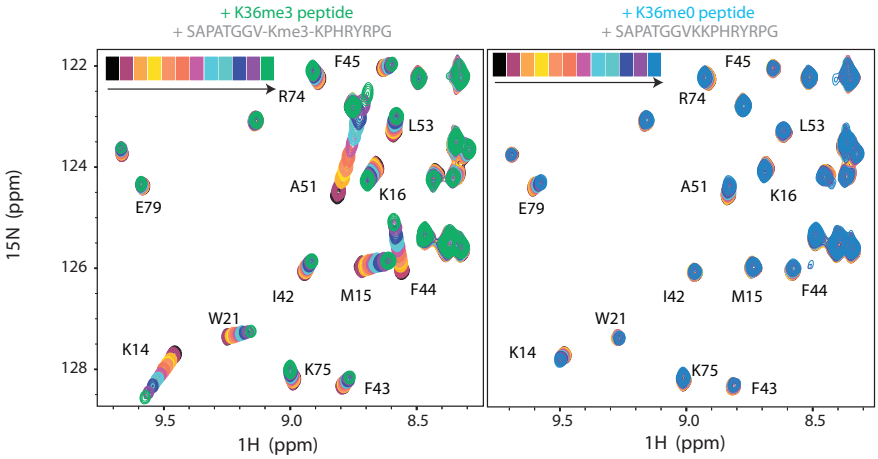
Supplementary figures



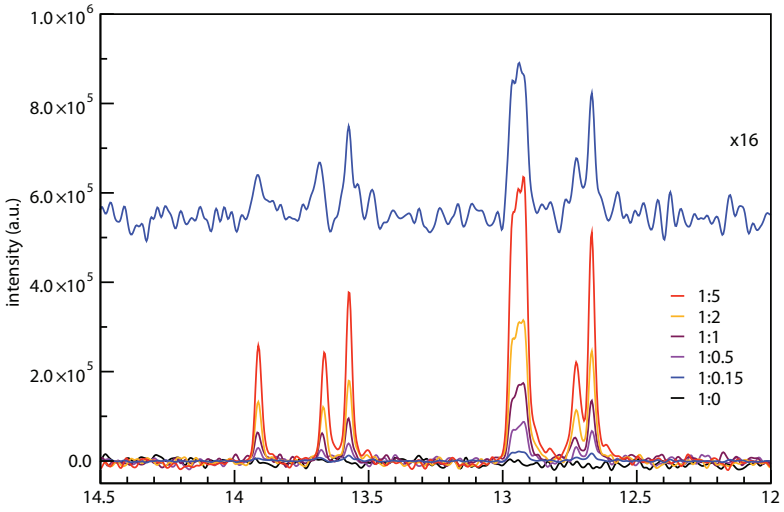
Supplementary Figure S1. Psip1 binds mono-, di- and tri-methylated peptides dependent on the PWWP domain. (A) Psip1-PWWP binds to H3K36me1, -me2 and -me3. Disruption of the PWWP domain abrogates the binding to H3 peptides. Immunoblot analysis of peptide pull downs with GST-PWWP or W21A mutant lysates and the indicated peptides were probed with GST antibodies. (B) Full length Psip1 enriched for H3K36 methylated nucleosomes and W21A mutation disrupts the interaction. GFP, Psip1-GFP or the W21A mutant were transiently expressed in 293T cells and bound to beads and incubated with mono-nucleosomes extracted from HeLa cells. Bound fractions were analyzed by immunoblotting with the indicated antibodies.



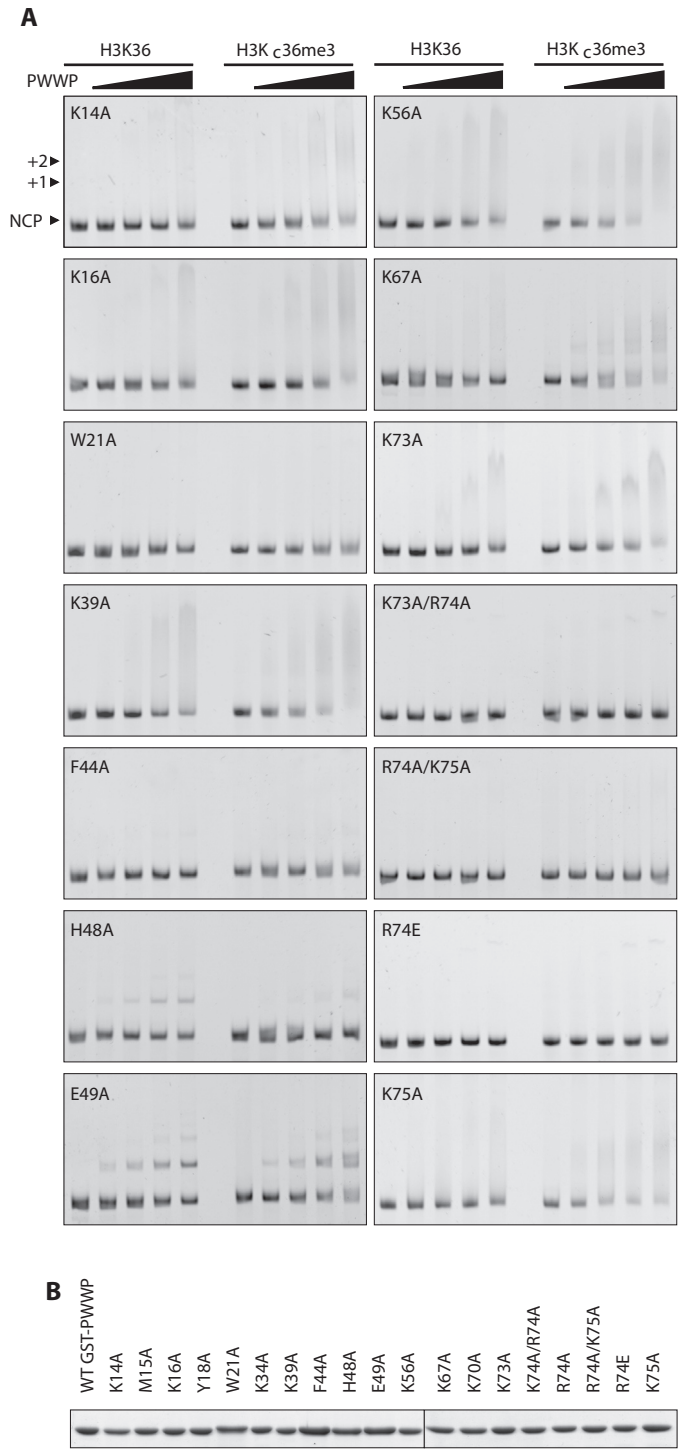
Supplementary Figure S2. Solution structure of Psip1-PWWP domain. (A) Backbone assignments of the Psip1-PWWP domain shown on the ^{15}N-^{1}H HSQC spectrum of free PWWP domain at 293K, pH 6.2, 600 MHz. The region in the solid box is expanded on the left; N and Q side chains resonances are indicated by the horizontal lines connecting the two resonances; indole side chain resonances of W are indicated by the subscript 'sc'. The region in the dotted box is shown in Figure 2 in the main text and Supplementary Figure S3. (B) Cross-eye stereo image of lowest energy structure; side chains of residues in the Kme3 binding pocket are shown as ball-and-stick. (C) Overlay of the unbound Psip1-PWWP domain (green) and the Hdgf2-PWWP domain in complex with H3K79me3 peptide (PDB: 3QJ6; yellow) and in complex with H4K20me3 peptide (PDB: 3QBY; magenta).



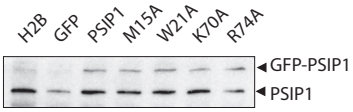
Supplementary Figure S3. Methylation of H3K36 is strictly required for the Psip1-PWWP–H3 tail peptide interaction. NMR titration results for methylated (left) and non-methylated (right) H3K36 peptide, showing an overlay of the spectra for each titration point. Sequences of the fragments that were used for titration are indicated at the top. Color-coding of the spectra is indicated at the top, the free PWWP spectrum is in black. Final concentration of peptide was 11 mM in both cases (molar ratio PWWP:peptide is 1:27). Assignments of resonances of interest are indicated.



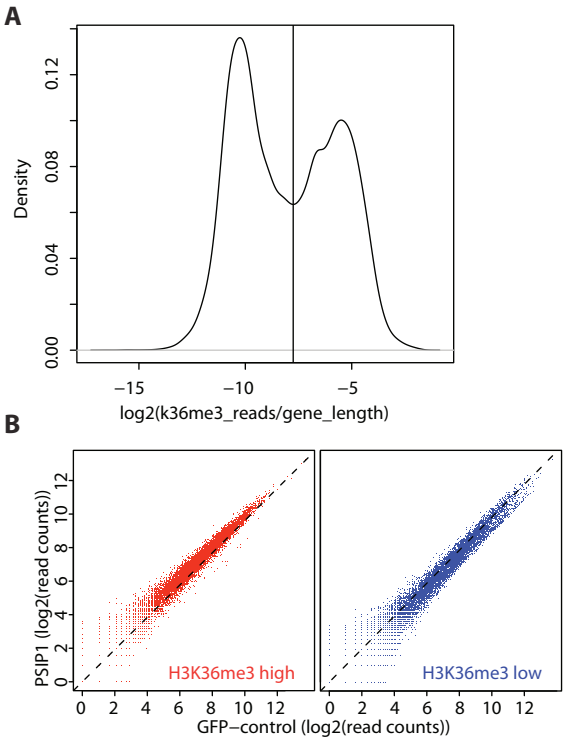
Supplementary Figure S4. Psip1-PWWP interacts with DNA in a non-sequence specific manner. Overlay of 1D spectra (recorded at 600 MHz) taken during the titration of SRE-DNA fragment to Psip1-PWWP domain, focusing on the DNA imino-protons. The protein:DNA molar ratio is indicated. In total 8 signals are visible from double-stranded DNA base pairs, which show no change in peak position, or relative intensity during the titration. The top spectrum in blue is amplified 16-fold to show that even at low protein:DNA ratio (1:0.15), where the DNA is mostly bound, the imino spectrum is the same as at high excess of DNA, where the DNA will be mostly free. This suggests that the PWWP-DNA interaction occurs through the phosphate backbone rather than through specific contact to DNA bases.



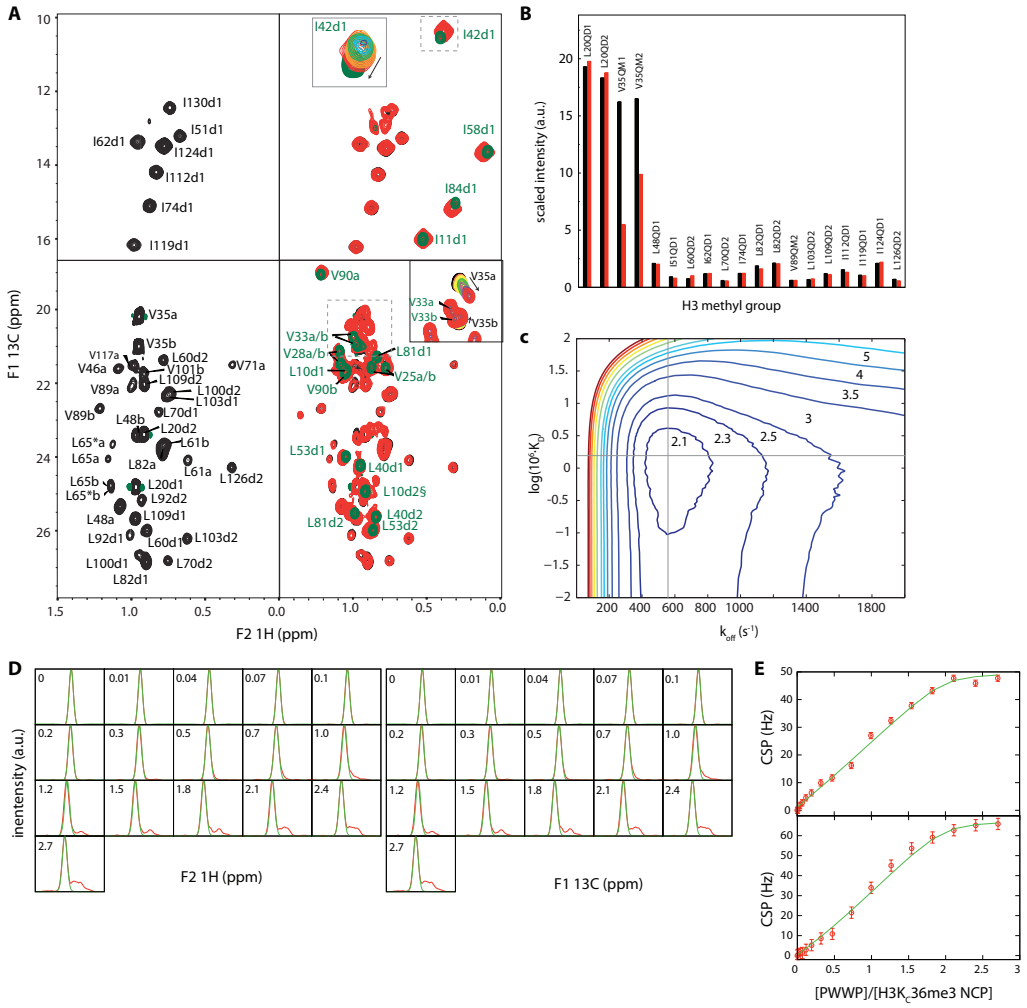
Supplementary Figure S5. Mutational analysis of PWWP aromatic cage and DNA interaction surface contribution to nucleosome binding affinity and specificity. (A) Native gel electrophoresis with GST-PWWP and wild type or H3K$_c$36me3 nucleosomes visualized by staining the gel with ethidium bromide. Titration of GST-PWWP protein with indicated nucleosomes were incubated in a 1:0, 1:0.5, 1:1, 1:2 and 1:4 (nucleosome:PWWP) ratios and loaded on gel. Lower bands are free nucleosome core particles (NCPs). Upper bands are complexes of nucleosomes with either one (+1) or two (+2) GST-PWWP molecules. Each PWWP mutant is indicated on the gel. (B) Comparable amounts of single step purified GST-PWWP mutants were loaded on gel and used for EMSA. Proteins were visualized by coomassie blue staining.



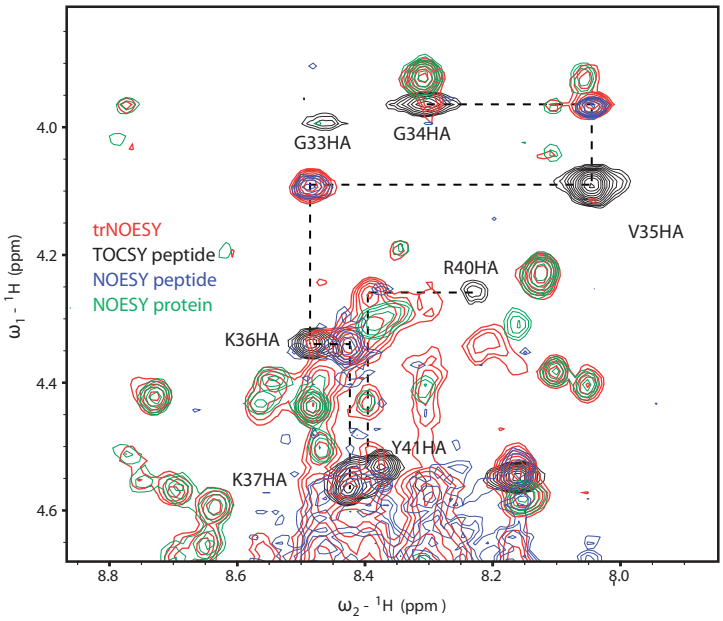
Supplementary Figure S6. Sub-endogenous GFP-Psip1 expression. Immunoblot analysis of HeLa FRT cells carrying GFP-Psip1 under a doxycycline inducible promoter using Psip1 antibodies. Cells were exposed to doxycycline for 4 hours, lysed in SDS sample buffer on gel. Lower band shows endogenous Psip1 expression, upper band GFP-fusion protein. GFP-fusions are indicated above the blot.



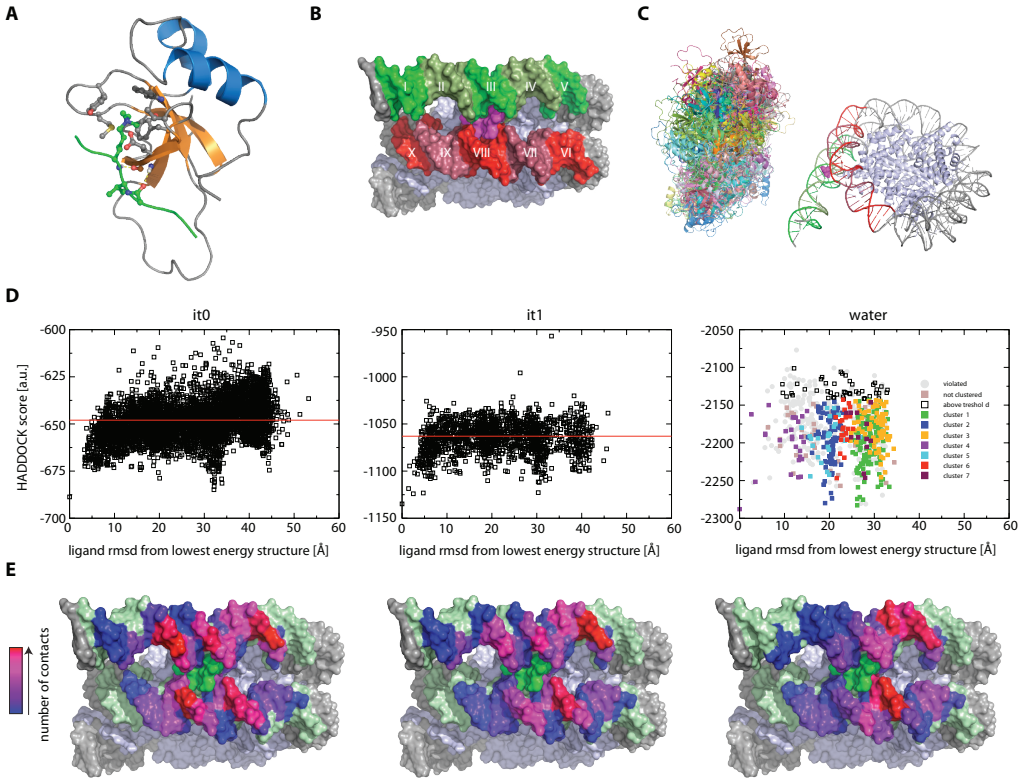
Supplementary Figure S7. Analysis of Psip1 and H3K36me3 ChIP sequencing. (A) H3K36me3 read density in genes shows a bi-modal distribution. ChIP sequencing was performed on H3K36me3. The number sequencing reads per gene was normalized to gene length (bp). The density plot of these normalized read counts is plotted. Genes are separated into H3K36me3 enriched and non-enriched using the minimum between the two peaks of the bimodal distribution as threshold (vertical line). (B) Psip1 is enriched on H3K36 tri-methylated genes. Data points represents read counts per gene. Genes were divides in low (blue) and high (red) H3K36me3 signal.



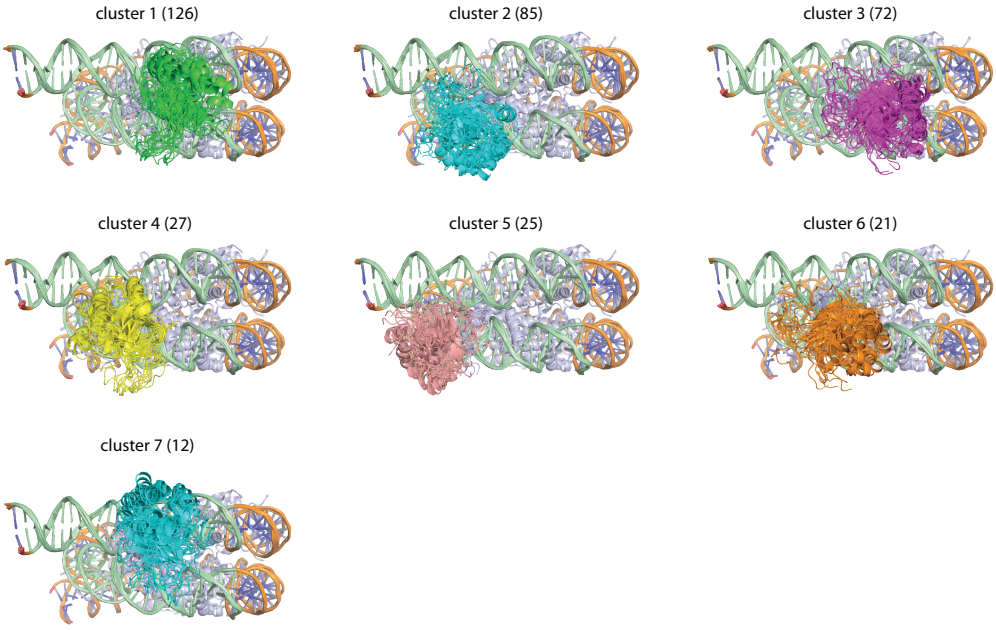
Supplementary Figure S8. Methyl TROSY spectra of H3K$_c$36me3 nucleosomes in free and PWWP bound forms (recorded at 900 MHz). (A) Methyl TROSY spectra of ILV labeled H3K$_c$36me3 nucleosomes in the apo state (black spectrum in left panel; including all H3 methyl group assignments); and overlay with PWWP bound state (red spectrum in right panel) and apo ILV-labeled PWWP (green spectrum in right panel). PWWP ILV residues are labeled in green in the overlay. The insets show an overlay of the titration spectra focusing on H3V35 and PWWP I42. Leu/Val methyl groups are labeled with their stereospecific assignments where available; otherwise the two methyl groups are arbitrarily assigned as "a" or "b". (B) Peak intensity of H3 methyl groups in unbound (black) and bound state (red); intensities are scaled such that median intensity for globular H3 residues is 1. Interestingly, the peak intensity of V35 remains ~5-fold higher compared to the nucleosome core, which suggests the presence of residual dynamics within the complex. (C) Reduced c^2-surface of line shape fitting of the H3V35a methyl group resonance as a function of k_{off} and K_D. Reduced c^2 values for the contour line are indicated. Best-fit values are indicated by the intersection of the grey lines (c^2_{red} = 1.99). F-statistic based critical value for 95% probability limit is 2.3. Data was fit to a 1:2 binding model, i.e. two PWWP molecules bind per nucleosome, one to each H3 tail, with each event independent from the other. Inclusion of cooperativity did not improve the quality of fit. (D) Individual line shape fits of the H3V35a resonance in the ^{1}H (left) and ^{13}C-dimension (right); molar equivalents of PWWP added are indicated; experimental line shape is shown in red, fitted line shape in green. The resonance is in fast-to-intermediate exchange ($\Delta\omega/k_{ex}$ = 0.4).



Supplementary Figure S9. Transferred-NOESY indicates an extended conformation for the H3 histone tail in complex with Psip1-PWWP. Overlay of 2D ^{1}H-^{1}H transferred NOESY spectrum recorded on a mixture of 1 mM H3K36me3 peptide and 0.1 mM PWWP domain with 2D TOCSY and NOESY spectra recorded on isolated peptide or protein domain. All spectra were recorded in 20 mM NaPi pH 7 and 100 mM NaCl at 25°C at 500 MHz with 100 ms mixing time. Strong αN connectivities in the transferred NOESY are observed for G34α-V35N, V35α-K36N, R40α-Y41N, as well as strong K37α-P38δ NOEs (not shown). These NOEs are consistent with an mostly extended conformation of residues 34-38 of the H3 peptide, as is also observed in crystal structures of methylated histones bound to PWWP domains (Vezzoli et al. 2010; Wu et al. 2011).



Supplementary Figure S10. Flexible multi-domain docking of Psip1–H3K36me3 nucleosome complex. (A) Lowest energy structure of H3 tail residues 28-38 bound to Psip1-PWWP domain obtained in the first docking stage. The aromatic cage residues of PWWP and residues G34-K36me3 of the H3 tail are shown in ball-and-sticks representation. The two enforced hydrogen bonds are indicated. (B) In the second stage the PWWP-H3K36me3 peptide complex is docked onto the nucleosome guided by DNA interaction restraints to one of the ten DNA patches indicated by roman numerals I-X and loose restraints between the H3 peptide and the H3 tail exit site, residue 39 shown in magenta. (C) Overlay of the first 100 starting conformations of the Psip1-PWWP domain relative to the nucleosome, in total 5500 random conformations were used. (D, E) Extent of conformational sampling during the second stage as judged by ligand rmsd (D) and R74 mediated DNA contacts (E). HADDOCK score and ligand rmsd for all solutions after rigid body docking (it0, 5500 structures), semi-flexible refinement (it1, 1500 structures) and refinement in explicit water (water, 750 structures). For subsequent semi-flexible refinement 1500 structures of the best 3000 solutions in it0 (below the red line) were chosen, by skipping every other structure. For final water refinement the 750 best structures in it1 were chosen (below the red line). The final ensemble consist of all structures with a proper peptide bond across the former cut-site, without violations of the experimental restraints and with proper Ramachandran plot quality for the H3 tail (filled squares, 404 structures). (E) Contact analysis for residue R74 of PWWP for all solutions after it0, it1 and for the final ensemble. The DNA surface is color coded according to the number of contacts between R74 and each DNA nucleotide. The passive DNA surface is shown in light green.



Supplementary Figure S11. Clustering analysis of final docking solutions. Superposition of the 10 best structures of for each of the 7 clusters found in the final ensemble of 404 solutions that are in agreement with experimental restraint data and have all backbone dihedral angles for residues 37-40 in H3 tail in core and allowed regions of the Ramachandran plot. Number of structures in each cluster is indicated between brackets.

A. Restraint information	
number of experimental distance restraints	1988
intra-residual/sequential/medium/long	451/612/316/609
TALOS derived dihedral angle restraints f/y	145
B. Average RMS deviation from experimental restraints	
All experimental distance restraints (Å)	0.017 ± 0.001
All dihedral angle restraints (°)	0.49 ± 0.05
C. Coordinate RMS deviation (Å)	
Average RMSD to mean	
Ordered backbone atoms	0.47 ± 0.30
Ordered heavy atoms	0.82 ± 0.51
Global backbone atoms	0.82 ± 0.75
Global all heavy atoms	1.19 ± 0.92
Pairwise RMSD	
Ordered backbone heavy atoms	0.85 ± 0.22
Ordered all heavy atoms	1.49 ± 0.23
Global backbone heavy atoms	1.67 ± 0.41
Global all heavy atoms	2.24 ± 0.42
D. iCing ROG score (red/orange/green %)	9/25/67
E. Ramachandran quality parameters (%)	
Residues in most favoured regions	84.2
Residues in allowed regions	15.2 1.6 ± 1.6
Residues in additionally allowed regions	0.3
Residues in disallowed regions	0.2
F. Average RMS deviation from current reliable structures (null deviation=1)	
Bond lengths	1.16 ± 0.01 0.73 ± 0.06 0.74 ± 0.13 0.79 ± 0.12
Bond angles	0.52 ± 0.02
Omega angle restraints	0.77 ±0.05
Side-chain planarity	1.06 ± 0.21
Improper dihedral distribution	0.90 ± 0.04
Inside/outside distribution	1.02 ± 0.02
G. Average deviation from current reliable structures (Z-scores, null deviation = 0)	
2nd generation packing quality	+5.8 ± 2.0
Ramachandran plot appearance	-2.7 ± 0.4
Chi-1/Chi-2 rotamer normality	-2.2 ± 0.7
Backbone conformation	-1.3 ± 0.4

Supplementary Table S1. Structural statistics for the free Psip1-PWWP domain.

[a] statistics are given for residues 1-93 of Psip1-PWWP, excluding the unstructured N-terminal cloning artifact. (residues -5 to 0) and the dynamically disordered C-terminal 7 residues (94-100) for which no NOEs were oberved.
Ordered regions are residues 5-30 and 38-89 of the PWWP domain.

res.[a]	intermolecular energy (s) in kcal/mol							g.s.[d]
	cl. 1[b] #126	cl. 2 #85	cl. 3[c] #72	cl. 4[d] #27	cl. 5[d] #25	cl. 6[d] #21	cl. 7[d] #12	
K16	**-147 (30)**	**-134 (26)**	-83 (40)	**-135 (31)**	**-114 (40)**	**-103 (45)**	-80 (37)	++
K67	-14 (31)	+1 (2)	-16 (35)	+1 (2)	-22 (37)	-4 (19)	-2 (4)	+
K70	-68 (36)	-51 (46)	-65 (54)	-45 (49)	-12 (28)	-86 (46)	-44 (48)	+
K73	-94 (46)	-77 (31)	-81 (37)	-81 (33)	-73 (34)	-62 (51)	**-102 (36)**	++
R74	-68 (31)	-80 (29)	-69 (34)	-95 (23)	-71 (26)	-99 (26)	-70 (23)	++
K75	-46 (41)	-92 (40)	**-97 (33)**	-83 (51)	-65 (49)	-61 (56)	-101 (42)	++
R3	-8 (27)	-	-	-	-8 (20)	-	-	n.d.
K6	-	-	-	-	-	-	-	n.d.
K14	-58 (52)	-80 (45)	-54 (51)	-48 (48)	**-29 (49)**	-38 (35)	-98 (23)	++
K34	-13 (36)	-10 (27)	-26 (43)	-7 (18)	-1 (2)	-3 (10)	-9 (31)	-
K39	-75 (54)	-60 (44)	**-13 (33)**	**-30 (41)**	-92 (46)	-	-	++
K56	-54 (44)	**-21 (43)**	**-10 (30)**	**-1 (3)**	-92 (40)	**-6 (21)**	**-20 (35)**	++
K65	-	-	-	-	-	-	-	n.d.
K91	-2 (4)	-1 (3)	-4 (16)	-1 (2)	-2 (3)	-1 (1)	-0 (2)	n.d.

Supplementary Table S2. Residue specific contribution to the intermolecular energy.

[a] active residues are shown in bold and have been explicitly restrained to contact the DNA in 50% of the docking solutions. Other residues have not been restrained.

[b] number of structures in each cluster indicated with '#'. Largest contribution to the intermolecular energy is shown in bold. Cells are filled in dark/light green if the residue makes a favorable contact (average + $2/1\sigma < 0$). Intermolecular energy contributions that conflict with the gel shift based mutation data is shown in bold-red.

[c] based on all structures in the cluster, otherwise based on best 50.

[d] ++/+/- = strong/moderate/no loss of binding affinity based on gel shift binding assay of alanine mutants (see Fig. 3 and Supplementary Fig. S5). n.d. is not determined.

Supplementary methods

Docking protocol Psip1-PWWP–nucleosome complex

We used our experimental chemical shift perturbation, transferred-NOESY and mutagenesis data, together with available literature data to create a structural model for the Psip1-PWWP–nucleosome complex with Haddock version 2.1 (Luger et al. 1997; Dominguez et al. 2003) and CNS 1.3 (Brünger et al. 1998; Brunger 2007). In what follows, we describe the docking procedure.

First, residues 31-38 of the H3 tail with a native tri-methylated K36 (M3L) were docked to the Psip1-PWWP domain, using the solution NMR structure ensemble containing 20 structures. To prepare these structures for docking, side chain dihedral χ_3 of M15 was adjusted to match the corresponding dihedrals found in the H3K36me3 bound Hdgf structures, resulting in a bound conformation of the binding pocket. The resulting structures were subsequently re-refined in explicit solvent using the default HADDOCK protocol. Based on pH titration data, H48 was set to be singly protonated (data not shown). The proton was placed at Nε2 to optimize the hydrogen-bonding network. The PWWP domain was set to be semi-flexible. The H3 tail peptide structure was taken from the crystal structure of the homologous PWWP domain of Brpf1 bound to a H3K36me3 peptide (PDB-ID: 2X4X (Vezzoli et al. 2010)), and was set to be fully flexible during docking. To increase sampling of the conformational space, the number of MD steps during the flexible simulated-annealing stage was doubled compared to their default values.

Docking was driven using ambiguous interaction restraints derived from chemical shift perturbation mapping and mutational analysis. Specifically, all PWWP residues with CSP larger than the 10% trimmed mean + 2σ in the me3-peptide titration and a surface accessible area of more than 33% or non-basic residues that led to loss of binding in the gel shift assay were defined as active resides. Neighboring solvent exposed residues were defined as passive residues. On the side of the H3 tail peptide only K36me3 was defined as an active residue and V35 and K37 were defined as passive residues. In total there were 12 ambiguous interaction restraints of which half were randomly removed for each calculated solution. In addition, ambiguous interaction restraints were defined between K36 methyl group protons and the aromatic cage residues of the PWWP domain (M15, Y18, W21 and F44), based on the total lack of chemical shift changes in the titration experiment using a non-methylated peptide. Backbone dihedral angle restraints consistent with an extended conformation (ϕ: -135±45° / ψ: +135±45°) were defined for residues 34-38 of the H3 peptide based on the transferred-NOESY data. Finally, homology based backbone hydrogen bond restraints were added between PWWP T50 and H3 tail K36 and G34. Equivalent hydrogen bonds are present in all Brpf1/Hdgf PWWP complexes with a tri-methylated lysine peptide.

In the rigid body docking phase, 200 solutions were calculated, of which the best 100 structures according to their HADDOCK score were refined in the (semi-)flexible phase, and further refined

in explicit solvent. The best 50 structures cluster within 4 Å heavy backbone atom interface rmsd for the ligand. The lowest energy structure is shown in Supplementary Figure S10A. All solutions were subsequently reduced to 11 representative structures using an all-atom rmsd based clustering with a cut-off of 2.6 Å, to be used as input for docking onto the nucleosome.

In the second stage, the PWWP-H3K36me3 complexes were docked onto the nucleosome guided by i) CSP and mutation derived ambiguous interaction restraints between basic residues of PWWP and the nucleosomal DNA around the H3 tail exit site; ii) restraints to enforce covalent attachment of the H3 tail in the PWWP complex to the remainder of H3 in the histone octamer. Docking was set up as a three-body docking, with separate chains for the nucleosomal DNA, the histone octamer with H3 up to residue 39, and the PWWP-H3K36me3 complex with H3 residues 28-38. The cut of histone H3 is at the point where it emerges from between the DNA helices, forming a natural point to allow for hinge like motions in the flexible multi-domain protocol. It should also be noted that the conformation of residue 38 and onwards in H3 are strictly conserved in all nucleosome structures, allowing an unambiguous definition of this hinge point. The starting conformation of octamer and DNA were derived from the crystal structure of *Dm.* nucleosome, PDB 2PYO (Clapier et al. 2008). Note that this DNA sequence corresponds to the α-Satellite rather than the 601 sequence used experimentally. Since recognition of DNA by both octamer and Psip1-PWWP domain is not sequence specific, this is not of concern. The DNA was further extended by adding a 10-bp straight B-DNA segment based on PDB-ID 355D and subsequent refinement. All lysine and arginine side chains in the PWWP domain were rebuilt by CNS, after removing all side chains atoms beyond the Cβ. This procedures breaks intra-molecular salt-brigdes within the PWWP domain, resulting exposed arginine and lysine side chains that are free to interact with DNA.

Ambiguous interaction restraints between PWWP and the nucleosomal DNA were derived from chemical shift perturbation mapping and mutational analysis. Specifically, basic PWWP residues were defined as active residues if i) their CSP was larger than the 10% trimmed mean + 2σ in the SRE-DNA titration, ii) their solvent accessible surface area of is more than 33%, and iii) their mutation resulted in a clear loss in binding affinity for K_c36me3 nucleosomes in the gel shift assay. These criteria resulted in 6 active residues for the PWWP domain (K16, K67, K70, K73, R74 and K75). In the absence of experimental data on the interaction site on the nucleosomal DNA, we scanned across the DNA around the H3 tail exit site (+/- 12-bp) by defining 10 sets of passive residues for the DNA, each 5-bp in length (see Supplementary Fig. S10B). For every docking solution calculated, ambiguous interaction restraints were defined between 3 randomly chosen active PWWP residues and one of the 10 passive DNA sets.

To ensure that the nucleosome remains intact during docking, the starting conformations were not randomized in the rigid body docking protocol. Rather, 5500 starting conformations were generated by randomly rotating the PWWP-H3K36me3 complex and placing it randomly in a box of dimensions 25x50x100 Å at ca. 50 Å from the nucleosome, oriented towards the H3 tail

exit site (see Supplementary Fig. S10C).

In the rigid body docking phase, 5500 solutions were calculated, i.e. 100 per combination of PWWP-H3K36me3 input structure and set of passive DNA residues. To allow sampling of large conformation space, a loose 12 Å unambiguous restraint between Cα of K36 and Cα of H39 was defined. All solutions were ranked according to their HADDOCK score, which is a weighted sum of van der Waals, electrostatic, desolvation, binding and restraint energies, and the buried surface area.

The extent of conformational sampling was checked by calculating the ligand rmsd of all solutions, where ligand rmsd is defined as the heavy atom backbone rmsd of the PWWP domain when superimposing the structures on the histone octamer (Supplementary Fig. S10D). The very high ligand rmsd of up to 55 Å shows that the PWWP domain occupies different positions on the nucleosomal DNA as anticipated. Notably, there are structures with low (favorable) scores over the whole range of ligand rmsd.

In addition, all DNA contacts (heavy atom distance <5 Å) made by basic PWWP residues were analyzed. Of all these residues, R74 makes the most contacts. In 89% of the solutions, R74 contacts a DNA residue, mostly the DNA right above and below the H3 tail exit site, but also up to 1 turn and more away from the exit point in this initial sampling phase (Supplementary Fig. S10E).

To ensure proper sampling of the conformational space also in the second stage of docking, we selected 1500 out of the best 3000 solutions according to their HADDOCK score for refinement in the (semi-) flexible phase (see also Supplementary Fig. S10D). Definition of interface residues in the PWWP-H3K36me3 complex was set to automatic. In addition residues 37-40 in the H3 tail were set to be fully flexible, all other atoms (including the DNA) were fixed. To favor solutions with small gaps across the cut-site of the H3 tail, a 1.2 Å unambiguous restraints for the peptide bond between residues 38 and 39 was enforced. To prevent that the H3 tail would be pulled out of the PWWP binding pocket, the interaction restraints from the first docking step were enforced as well.

The ligand rmsd over all solutions again shows that there are many favorable solutions with very different positions of the PWWP domain on the nucleosomal DNA. Residue R74 now shows a slightly more localized distribution of DNA contacts (Supplementary Fig. S10E). Note that at this point the gap between residues 38 and 39 varies between 3 and 20 Å.

In the final step, the best 750 solutions were refined in explicit water, and at the same time the covalent attachment of the two parts of the H3 tail was enforced by rebuilding it as a single chain. This allows the internal force field to recreate the peptide bond across 38 and 39 with the appropriate geometry. Residues 37-40 of H3 were set to be fully-flexible. In all structures, a proper peptide bond was recreated. Overall, there is less spread in the solutions, although still a considerable range in ligand rmsd values, up to 35 Å (Supplementary Fig. S10D). Solutions without hydrogen bond (>0.5 Å), dihedral angle (>5°), or unambiguous restraint (>0.5 Å)

violations and with H3 tail (residues 37-40) backbone dihedrals in the core and allowed regions of the ramachandran plot (as defined by PROCHECK (Laskowski et al. 1996)) were ordered according to their HADDOCK score, the 10% highest (unfavourable) energy structures were removed, resulting in a final ensemble of 404 structures (filled squares in Supplementary Fig. S10D, left panel). As can be seen in Supplementary Fig. S10D, this procedure also filters out some solutions with favorable HADDOCK scores, which does not contain an energy term for ramachandran plot quality. In these structures, the peptide bond could only be recreated at expense of very bad backbone dihedrals for the linking residues.

The final ensemble of 404 solutions was first subjected to a clustering analysis, using the ligand backbone rmsd as defined above. Using a cut-off of 10 Å with a minimum of 10 cluster members, 91% of all structures were clustered into 7 clusters. The superposition of the 10 best structures for each cluster in Supplementary Figure S11 shows the distinct translational and rotational positions of the PWWP domain on the nucleosomal DNA.

Next, we looked at the residue specific contributions to the intermolecular energies (sum of Vanderwaals and electrostatic terms) of all basic PWWP residues in the 50 best structures for the three largest clusters (Supplementary Table S2). Across all three clusters, active residues K16, K73, R74, and K75 make large contributions to the intermolecular energy, with K16 being overall most important. Residues K67 and K70 were defined as active but contribute little to the intermolecular energy. Both residues indeed show a smaller effect in gel shift compared to others and have least amount of chemical shift perturbation.

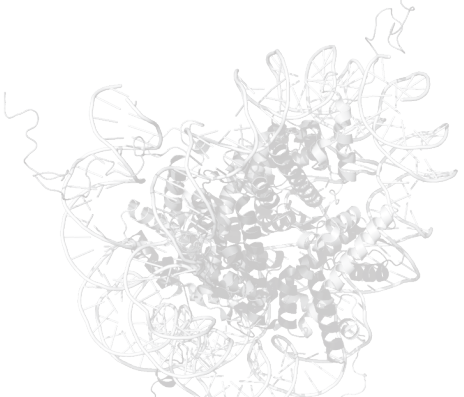
Analysis of the contacts made by the other basic residues provides the opportunity to cross-validate the docking solutions, as we have gel shift based mutation data for 4 of these residues, which were not included in the docking procedure. Cluster 3, 4, 6 and 7 are in conflict with the cross-validation data for both K39A and K56A mutations, while cluster 2 conflicts with the data for K56 and cluster 5 with that of K39. These clusters are therefore excluded from further consideration. The largest cluster of solutions, cluster 1, is in agreement with all cross-validation data and has been deposited in the PDB (3ZH1).

Chapter 5

Multivalent engagement of TFIID to nucleosomes

Rick van Nuland[#], Andrea W. Schram[#], Frederik M.A van Schaik, Pascal W.T.C Jansen, Michiel Vermeulen and H.Th. Marc Timmers

[#] These authors contributed equally



PLoS One. 2013 Sep 11;8(9):e73495

Abstract

The process of eukaryotic transcription initiation involves the assembly of basal transcription factor complexes on the gene promoter. The recruitment of TFIID is an early and important step in this process. Gene promoters contain distinct DNA sequence elements and are marked by the presence of post-translationally modified nucleosomes. The contributions of these individual features for TFIID recruitment remains to be elucidated. Here, we use immobilized reconstituted promoter nucleosomes, conventional biochemistry and quantitative mass spectrometry to investigate the influence of distinct histone modifications and functional DNA elements on the binding of TFIID. Our data reveal synergistic effects of H3K4me3, H3K14ac and a TATA box sequence on TFIID binding *in vitro*. Stoichiometry analyses of affinity purified human TFIID identified the presence of a stable dimeric core. Several peripheral Tafs, including those interacting with distinct promoter features, are substoichiometric yet present in substantial amounts. Finally, we find that the Taf3 subunit of TFIID binds to poised promoters in an H3K4me3 dependent manner. Moreover, the PHD finger of Taf3 is important for rapid induction of target genes. Thus, fine tuning of TFIID engagement on promoters is driven by synergistic contacts with both DNA elements and histone modifications, eventually resulting in a high affinity interaction and activation of transcription.

Introduction

RNA polymerase II (pol II) mediates the transcription of all protein coding genes in eukaryotic cells. Activation of transcription by sequence specific DNA binding transcription factors leads to recruitment of basal transcription factors to core promoters that together establish the pre-initiation complex (PIC) (Buratowski et al. 1989). PIC assembly is initiated by core promoter association of the TFIID complex, followed by the sequential binding of other basal factors and recruitment of pol II (Thomas and Chiang 2006). TFIID is a large complex and contains ~13 TBP associated factors (Tafs) and the TATA binding protein (TBP) (Sanders et al. 2002b). 9 of the 13 Tafs contain a histone fold dimerization domain that allows multiple pairwise interactions within the complex (Gangloff et al. 2001; Wright et al. 2006). TFIID adopts a clamp like shape that features a symmetrical core. The TFIID complex has been studied extensively in yeast using multistep affinity purified complexes. Coomassie staining based analysis of these complexes revealed that a subset of Tafs (Taf4, Taf5, Taf6, Taf9, Taf10, Taf11 and Taf12) are present in more than one copy (Sanders et al. 2002a). Recent work on reconstituted human TFIID confirmed these results and showed that upon addition of the Taf8/Taf10 dimer, a new surface is created that allows the assembly of single copies of the other Tafs to form a full TFIID complex (Bieniossek et al. 2013). Structural heterogeneity has been observed in TFIID preparations isolated from human cells and this was linked to a sub-stoichiometric Taf2 presence (Kaufmann and Smale 1994). Additionally, the binding of TFIID to DNA induces a structural rearrangement within the complex (Cianfrocco et al. 2013). The mechanism for this remains to be elucidated and it might be influenced by changes in TFIID composition.

Several subunits within TFIID can bind to specific DNA elements found at promoters. TBP interacts with the TATA element, which is found upstream of the transcription start site (TSS). In yeast as well as in mammals, only a subset of genes contains a high affinity TATA box sequence (FitzGerald et al. 2004; Bajic et al. 2006). Surprisingly, in yeast, TFIID association with promoters is inversely correlated with the presence of a consensus TATA sequence (Rhee and Pugh 2012). TBP association with TFIID and the TATA sequence is stabilized by binding of the TFIIA complex (Tan et al. 1996; Geiger et al. 1996). Additionally, Tafs1/2 interact with the initiator element (INR) (Chalkley and Verrijzer 1999) and Tafs6/9 can bind to a downstream promoter element (DPE) (Burke and Kadonaga 1996).

Chromatin has an important role in the regulation of transcription. The basic building block of chromatin is the nucleosome, comprised of an octamer of histone proteins. Post-translational modifications on the protruding tails of histones contribute to transcription regulation. Effector proteins that contain specific binding modules can recognize these chemical modifications and get recruited to genomic loci (Kouzarides 2007; Jenuwein and Allis 2001; Berger 2007). Tri-methylation of lysine 4 on histone H3 (H3K4me3) is associated with virtually all active and poised promoters both in yeast and in mammals (Barski et al. 2007; Heintzman et al. 2007).

Several H3K4me3 binding proteins have been identified, including the chromatin remodeler Bptf and the TFIID subunit Taf3. Binding of these proteins to H3K4me3 occurs through their plant homeodomain (PHD) fingers (Wysocka et al. 2006; Li et al. 2006; Vermeulen et al. 2007). Recently, it has been described that the Taf3-H3K4me3 interaction in mammals is required for PIC assembly on a selective group of genes which are mainly involved in the response to DNA damage (Lauberth et al. 2013). In addition to H3K4me3, promoter associated modifications include hyperacetylation on histone H3 and the presence of a specific histone variant H2A.Z, which replaces the canonical H2A (Li et al. 2005; Barski et al. 2007).

Here we show that the binding of TFIID to recombinant nucleosomes is synergistically enhanced by the presence of a TATA box in nucleosomes carrying $H3K_C4me3$ and H3K14ac. However, this binding is not affected by incorporation of the histone variant H2A.Z or the H3K27me3 repressive mark. To further dissect the biochemistry of TFIID and to investigate the requirements for TFIID binding to nucleosomes *in vitro*, we determined the stoichiometry of endogenous human TFIID. These experiments revealed that TFIID consists of a stable symmetric core and a number of peripheral sub-stoichiometric Tafs. Finally, we show that binding of Taf3 is enriched on 'poised' stress gene promoters containing H3K4me3 in a PHD finger dependent manner *in vivo*.

Results

TFIID binding to $H3K_C4me3$ nucleosomes

TFIID is a large protein complex containing various subunits that can interact with specific DNA elements and distinct histone modifications. Thus far, such interactions have mainly been studied using gel shift and peptide pull down assays (Buratowski et al. 1989; Workman and Roeder 1987; Burke and Kadonaga 1996; Vermeulen et al. 2007). Recently, several approaches have been developed to generate *in vitro* reconstituted nucleosomes containing specific histone modifications and DNA sequences (Simon et al. 2007; Shogren-Knaak et al. 2006). In combination with quantitative mass spectrometry, affinity purifications using such immobilized nucleosomes can reveal proteins and protein complexes that can specifically interact with these *in vitro* assembled modified nucleosomes species (Sanders et al. 2002a; Nikolov et al. 2011; Bartke et al. 2010). We applied a methyl lysine analog (MLA) approach to produce recombinant nucleosomes carrying a H3K4me3 mimic ($H3K_C4me3$) with the aim to use these as bait for affinity purifications in crude nuclear extracts. To validate our approach we first tested the interaction between the Taf3 PHD finger and different MLA peptides. As shown in Fig. 1A, the Taf3 PHD finger specifically binds to the histone H3 N-terminus containing the $H3K_C4me2$ and $H3K_C4me3$ modification analogs. This binding is specific and comparable to H3 peptides containing natural methylated lysines (H3K4me2 and H3K4me3). This indicates that the MLA approach can be used as a tool to study TFIID-nucleosome interactions.

Next, we reconstituted MLA containing histone octamers with the 'Widom' 601 sequence labeled with a biotin on the 5'-end. The 'Widom' 601 sequence was used to prevent unintentional sliding of the nucleosome and transcription factor binding. Furthermore, the 'Widom' 601 sequence allows for efficient reconstitution of nucleosomes. Reconstituted nucleosomes were immobilized on streptavidin conjugated magnetic beads and incubated with HeLa nuclear extract. To validate our assay we used western blotting to show the specific binding of the TFIID core subunit Taf5 to H3K$_C$4me3 containing nucleosomes. In contrast, Taf5 does not interact with unmodified or H3K$_C$36me3 marked nucleosomes, which validates the specificity of our approach (Fig. 1B).

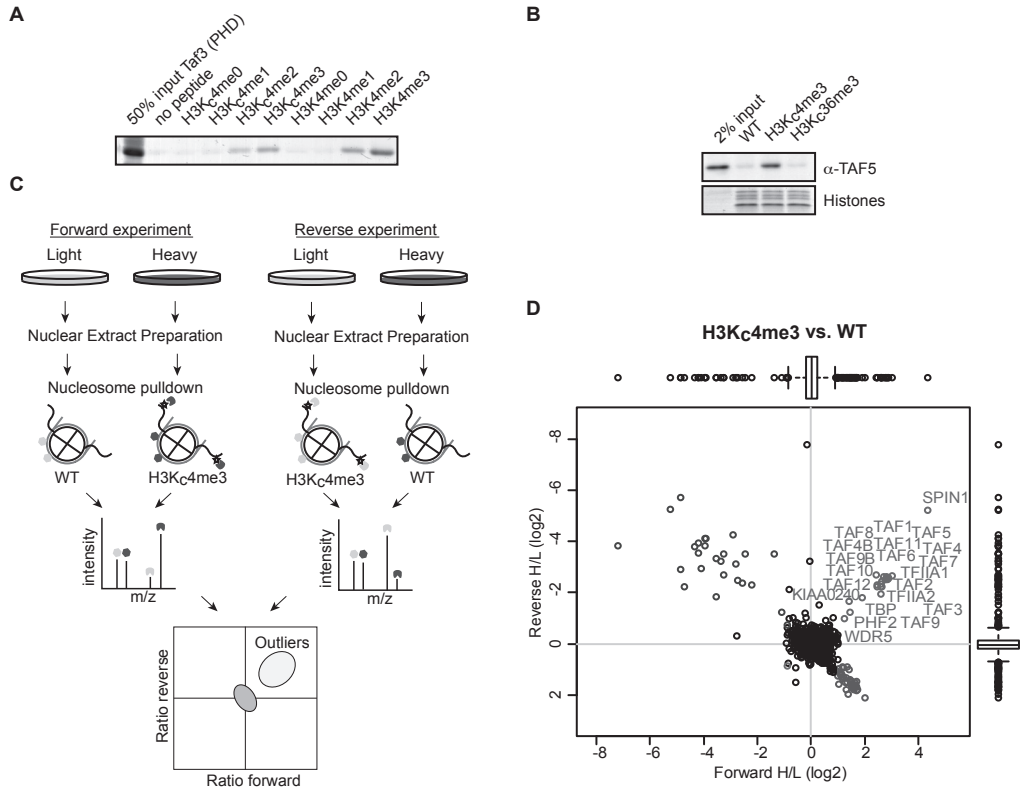


Figure 1. H3K$_C$4me3 nucleosomes bind endogenous TFIID and recombinant Taf3. (A) Pull down with the indicated biotinylated peptides using streptavidin coated beads incubated with GST-Taf3 PHD protein lysates. Proteins are visualized using Coomassie blue staining. (B) Immunoblot analysis of endogenous Taf5 binding to immobilized recombinant nucleosomes with the indicated MLA modification. Histones are visualized using Coomassie blue staining. (C) Workflow as applied for quantitative analysis of nucleosome interactors. In brief, heavy and light labeled extracts are used for pull down with immobilized differentially modified nucleosomes. Experiments are performed in label swap. Eluted proteins are measured using LC-MS/MS. Enriched proteins in both experiments are selected based on box plot statistics. (D) Scatter plot of SILAC ratios for H3K$_C$4me3 versus non-modified nucleosome interacting proteins. In the upper right corner significant outliers are depicted and labeled based on box plots analysis.

The H3K$_C$4me3 and unmodified control nucleosomes were then used for affinity purification in combination with SILAC labeled HeLa nuclear extracts. Quantitative mass spectrometry was applied to identify specific interactors in an unbiased manner (Vermeulen et al. 2010) (Fig. 1C). Nucleosomes with H3K$_C$4me3 showed enriched binding of all TFIID subunits and TBP (Fig. 1D). The SILAC ratio plots also reveal specific binding of TFIIA, which is known to functionally cooperate with TFIID during the early stages of PIC assembly. Several known H3K4me3 interactors were also identified, including Phf2 and Spin1 (Kaufmann and Smale 1994; Vermeulen et al. 2010; Bartke et al. 2010). In contrast, a number of known H3K4me3 interactors were not enriched in our experiments. This may be related to the use of the MLA instead of natural tri-methylated lysine, which can affect binding affinity. Indeed, although recombinant Sgf29 specifically interacts with H3K4me3 (Bian et al. 2011; Vermeulen et al. 2010), this protein does not bind to H3K$_C$4me3 peptides (data not shown). Interestingly, an uncharacterized protein (KIAA0240) was found to interact specifically with the H3K$_C$4me3 nucleosomes. This protein does not carry an annotated putative H3K4me3 interaction domain, indicating that it may interact with one of the H3K4me3 readers. In summary, these experiments reveal that a single histone modification (H3K4me3) contributes significantly to the overall affinity of TFIID for nucleosomes, despite the high basal affinity of the TBP subunit for DNA (Hahn et al. 1989).

TFIID binding to nucleosomes is enhanced by acetylation of K9/K14 and a TATA box and not disrupted by the presence of H3K27me3

The MLA approach can be used to study crosstalk between different chromatin modifications. One such crosstalk phenomenon has been described for embryonic stem cells, where H3K4me3 and H3K27me3 co-occur on silent but 'poised' developmentally regulated, bivalent genes (Voigt et al. 2012; Bernstein et al. 2006). We used both western blotting and quantitative mass spectrometry to study the interaction between TFIID and bivalent nucleosomes. As shown in Fig. 2A, the Taf3 PHD finger, which directly binds to H3K4me3, binds equally well to H3K$_C$4me3- and H3K$_C$4me3/H3K$_C$27me3 containing nucleosomes. In agreement with this, the TFIID complex was identified as a specific reader for H3K$_C$4me3/H3K$_C$27me3 marked nucleosomes, as revealed by quantitative mass spectrometry (Fig. 2B). Together, these results demonstrate that TFIID binding to H3K$_C$4me3 is not disrupted by the presence of H3K$_C$27me3.

Spin1 and KIAA0240 were again identified as specific interactors, as was TFIIA. Interestingly, Phf2 does not interact with bivalent nucleosomes but another PHD containing protein, Phf12, does. This protein is part of a complex containing the Jarid1A H3K4me demethylase enzyme (Vermeulen et al. 2010), which was not observed as a specific interactor in our experiments. Notably, experiments using nucleosomes containing only H3K$_C$27me3 did not yield significant interactors (data not shown).

The chromatin landscape around active gene promoters is characterized by the presence of several distinct features including the histone variant H2A.Z and acetylated histones H3 and

H4 (Juven-Gershon and Kadonaga 2010). Additionally, distinct DNA elements in the promoter region can contribute to PIC assembly. We investigated the contribution of these features to TFIID binding *in vitro* using recombinant nucleosomes (Fig. 2C).

Incorporation of the H2A.Z variant marks promoters and enhancers (Calo and Wysocka 2013). We first tested the effect of H2A.Z incorporation on TFIID binding to recombinant nucleosomes carrying H3K$_c$4me3 (Fig. 2D). This experiment revealed that the presence of unmodified H2A.Z

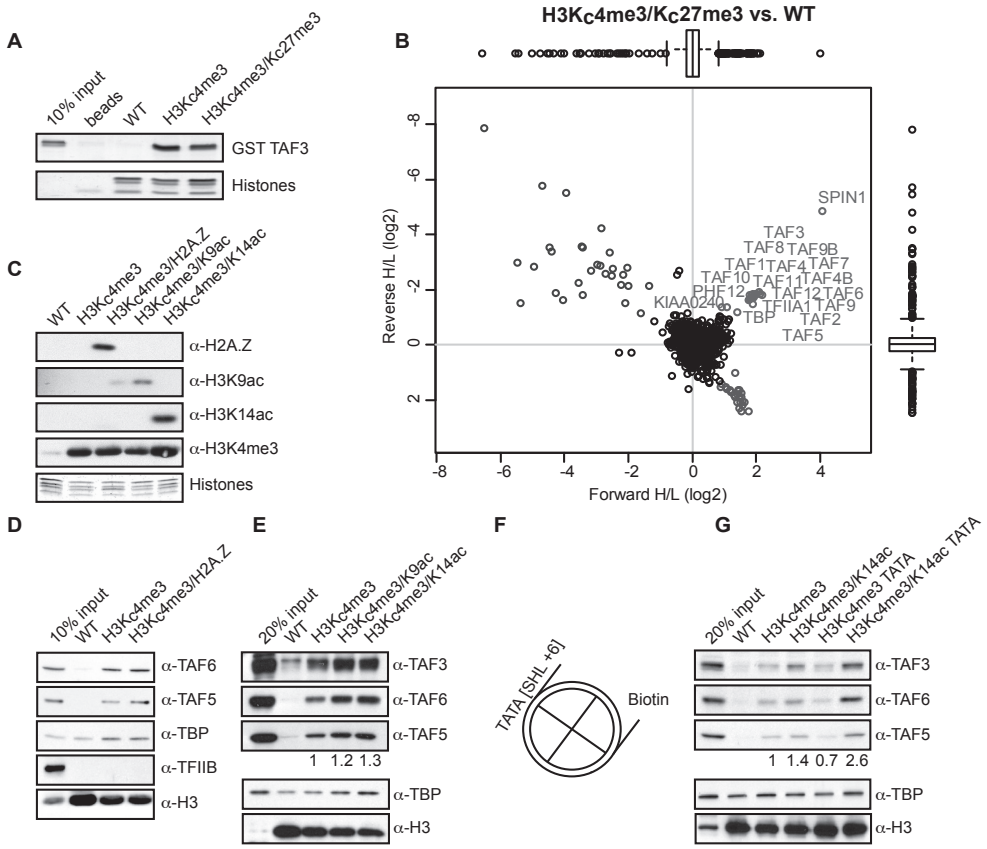


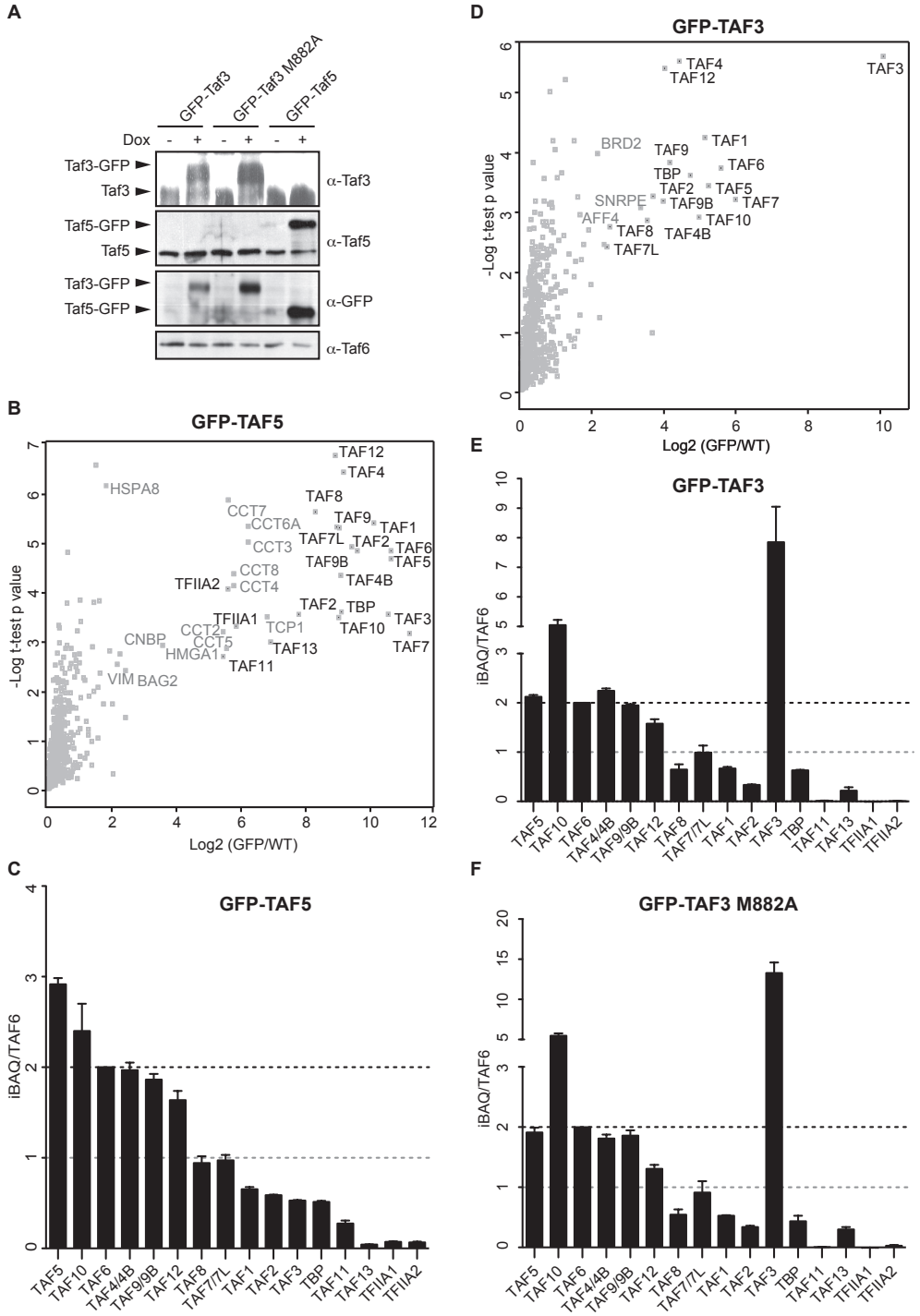
Figure 2. Presence of H2A.Z, H3K9/K14ac and a TATA sequence enhances binding of endogenous TFIID to recombinant nucleosomes. (A) Analysis of pull downs with recombinant nucleosomes immobilized on magnetic streptavidin coated beads and GST-Taf3 (PHD). Proteins are visualized by silver stain. (B) Scatter plot of SILAC ratios for H3K$_c$4me3/K$_c$27me3 versus non-modified nucleosome interacting proteins. In the upper right corner significant outliers are depicted and labeled in grey based on box plots analysis. (C) Immunoblot analysis of recombinant nucleosomes with the indicated antibodies showing the presence of modifications or variants. (D) TFIID binds to H3K$_c$4me3 nucleosomes independently of H2A.Z. Immunoblot analysis of eluted proteins using indicated antibodies. (E) TFIID binding is enhanced by histone H3 acetylation. Immunoblot analysis of eluted proteins using indicated antibodies. Taf5 antibody signal is quantified relative to the H3K$_c$4me3 pull down. (F) Schematic representation of the nucleosome with the introduced TATA sequence and biotin group indicated. (G) Combination of TATA DNA and H3K14 acetylation on H3K$_c$4me3 nucleosomes increases the interaction with TFIID. Immunoblot analysis of eluted proteins using indicated antibodies. Taf5 antibody signal is quantified relative to the H3K$_c$4me3 pull down.

in nucleosomes does not significantly influence TFIID binding *in vitro*. Next, we used an amber codon suppression approach to express recombinant histone H3 containing either acetylated K9 or K14 in bacteria (Neumann et al. 2009) and combined this with the aforementioned MLA approach. Using these doubly modified nucleosomes, we observed enhanced binding (1.2-1.4 fold) of TFIID to nucleosomes decorated with both H3K$_C$4me3 and H3K9ac or H3K14ac as compared to H3K$_C$4me3 alone (Fig. 2E,G). We find a comparable level of enhancement using either H3K9ac or H3K14ac combined with H3K$_C$4me3. These agonistic binding effects can be explained by the tandem bromodomain of Taf1, which was previously shown to interact with double acetylated histone H4 peptides (Jacobson et al. 2000). Furthermore, TFIID binding to histone H3K4me3 peptides was previously shown to be enhanced by additional H3K9 and H3K14 acetylation (Vermeulen et al. 2007). Unfortunately, efforts to express recombinant H3 bearing both H3K9 and H3K14 acetylation proved to be unsuccessful (data not shown).

We then set out to study the potential interplay between histone modifications and specific DNA elements in the regulation of TFIID binding to nucleosomes. To this end, recombinant nucleosomes were generated containing the classic 'Widom' 601 sequence, which carries a weak TATA sequence (GATATATAC) or a 601 variant carrying a consensus TATA sequence (TATATAAAAT) at super helical loop +6 (SHL +6) (Fig. 2F).

As shown in Fig. 2G, H3K$_C$me3 dependent TFIID binding is not potentiated in the presence of a consensus TATA sequence (Fig. 2G). However, when the consensus TATA DNA was used in combination with nucleosomes carrying both H3K$_C$4me3 and H3K14ac, binding was significantly enhanced as compared to nucleosomes carrying the weak TATA sequence or the methyl/acetyl combination (Fig. 2G). Together, these data reveal that diverse features including histone modifications and specific DNA elements affect TFIID binding to nucleosomes. Furthermore, the importance of a functional DNA element with regard to TFIID binding can be dependent on the modification state of the nucleosomes in *cis*, suggesting context dependent synergy. TBP binding itself seems not affected in these experiments, which may be explained by the nonspecific binding of TBP to unmodified nucleosomes. In these cases, TBP binding is TFIID independent (Fig. 2D,E and G). In these experiments, no competitor DNA was used, which provides a possible explanation for the observed TBP binding.

Figure 3. Stoichiometry analysis of endogenous TFIID. (A) Immunoblot analysis of GFP-Taf3, GFP-Taf3 M882A and GFP-Taf5 expression after 24 hours of doxycycline induction with the indicated antibodies (right). Endogenous proteins and GFP-fusions are indicated on the left. (B) Identification of interacting proteins for GFP-Taf5 by volcano plot. The ratio of identified proteins in all fusion lines over wild type in label free quantification are plotted against the -log10 of the false discovery rate (FDR) calculated by a permutation based FDR adapted t-test. In all experiments FDR was set to <0.05 and S0 = 1.5. Significant outliers are labeled. (C) Bar graphs indicate the stoichiometry of TFIID subunits (indicated at bottom) relative to Taf6. Black dashed line indicates a ratio to the total Taf6 protein. Error bars indicate the standard deviations of the technical triplicate. (D) Identification of interacting proteins for GFP-Taf3 by volcano plot. (E) Bar graphs indicate the stoichiometry of TFIID subunits (indicated at bottom) relative to Taf6 in GFP-Taf3 (E) and GFP-Taf3 M882A (F) purifications.


A
GFP-Taf3
GFP-Taf3 M882A
GFP-Taf5
Dox - + - + - +
Taf3-GFP
Taf3
α-Taf3
Taf5-GFP
Taf5
α-Taf5
Taf3-GFP
Taf5-GFP
α-GFP
α-Taf6
B
GFP-TAF5
-Log t-test p value
Log2 (GFP/WT)
TAF12
TAF4
HSPA8
TAF8
CCT7
CCT6A
TAF9
TAF1
TAF7L
CCT3
TAF2
TAF6
TAF9B
TAF5
CCT8
TFIIA2
CCT4
TAF4B
TAF2
TBP
TFIIA1
TAF3
CNBP
CCT2
TCP1
TAF10
CCT5
TAF13
TAF7
HMGA1
TAF11
VIM
BAG2
C
GFP-TAF5
iBAQ/TAF6
TAF5
TAF10
TAF6
TAF4/4B
TAF9/9B
TAF12
TAF8
TAF7/7L
TAF1
TAF2
TAF3
TBP
TAF11
TAF13
TFIIA1
TFIIA2
D
GFP-TAF3
-Log t-test p value
Log2 (GFP/WT)
TAF4
TAF12
TAF3
BRD2
TAF9
TAF1
TAF6
TBP
TAF2
TAF5
SNRPE
TAF9B
TAF7
AFF4
TAF10
TAF8
TAF4B
TAF7L
E
GFP-TAF3
iBAQ/TAF6
F
GFP-TAF3 M882A
iBAQ/TAF6

Stoichiometry determination of human endogenous TFIID

So far, we have shown that TFIID is recruited to immobilized nucleosomes containing histone modifications and a consensus TATA box in a synergistic manner. This implies that Taf1 and Taf3, which are the subunits binding acetylated and methylated lysines, as well as TBP, which binds TATA containing DNA, have to be present together in substantial amounts in the same complex. To determine the composition of endogenous human TFIID, we analyzed the stoichiometry using label free interaction proteomics combined with the iBAQ algorithm, which can be used to estimate relative protein abundance in a sample of interest (Smits et al. 2012; Hubner and Mann 2011). We generated a cell line containing a doxycycline inducible GFP-fusion of Taf5. As a control, wild type HeLa FRT cells were used (Fig. 3A). Nuclear extracts were made and these were subjected to single step GFP affinity purification in triplicate which was followed by on-bead trypsin digestion and LC-MS/MS analysis (Hubner et al. 2010).

Purification of GFP-Taf5 resulted in the identification of all known TFIID subunits (Fig. 3B). iBAQ based stoichiometry determination revealed the presence of a core complex containing a dimeric Taf6,4,9,10 and -12 module (Fig. 3C). Taf5 appears to be trimeric, which may be due to the moderate ~5 fold overexpression of the protein (Fig. 3A). Two Tafs, Taf8 and Taf7, are monomeric. The remaining TFIID subunits are substoichiometric, including Taf1, Taf3 and TBP (stoichiometry relative to Taf6 ~0.5). Taf11 and Taf13 are highly substoichiometric (~0.2 and 0.05, respectively). Thus, as expected, Taf1, Taf3 and TBP co-purify with core TFIID, although their stoichiometry is slightly lower compared to the dimeric core. This may hint towards the existence of distinct TFIID subcomplexes, each containing a specific set of peripheral Tafs. Alternatively, peripheral subunits may be partially dissociated from the core complex during affinity purification.

To further investigate this, we tagged and purified a peripheral TFIID subunit, Taf3. We also generated a stable cell line expressing a GFP tagged inducible Taf3 mutant, M288A, which cannot bind to H3K4me3. As was shown for Taf5, purification of GFP-Taf3 and GFP-Taf3 M882A resulted in the co-purification of all TFIID subunits (Fig. 3D,E and F). The stoichiometry determination for wild type and mutant Taf3 look almost identical, which illustrates the quality of the affinity purifications. Moreover, these data demonstrates that a mutated PHD finger is not affecting the integrity of the TFIID complex. In both purifications, the stoichiometry of Taf3 is significantly higher compared to the other Tafs, indicating that a proportion of Taf3 is not incorporated into TFIID. Nevertheless, we again identify a stable dimeric core. In addition, the stoichiometry of Taf10 exceeds the dimeric core, which implies the existence of a 'free' Taf3/Taf10 dimer. This is in agreement with the fact that Taf10 can associate with Taf3 as well as Taf8 through their respective histone fold domains (Demény et al. 2007; Leurent et al. 2002). Notably, relatively high amounts of Taf1 and TBP co-purified with GFP-Taf3 and GFP-Taf3 M882A (stoichiometry ~0.5 relative to the dimeric core) (Fig. 3E,F). This indicates that the lysine methyl , acetyl and TATA binding moieties co-exist within a single TFIID complex. These observations therefore

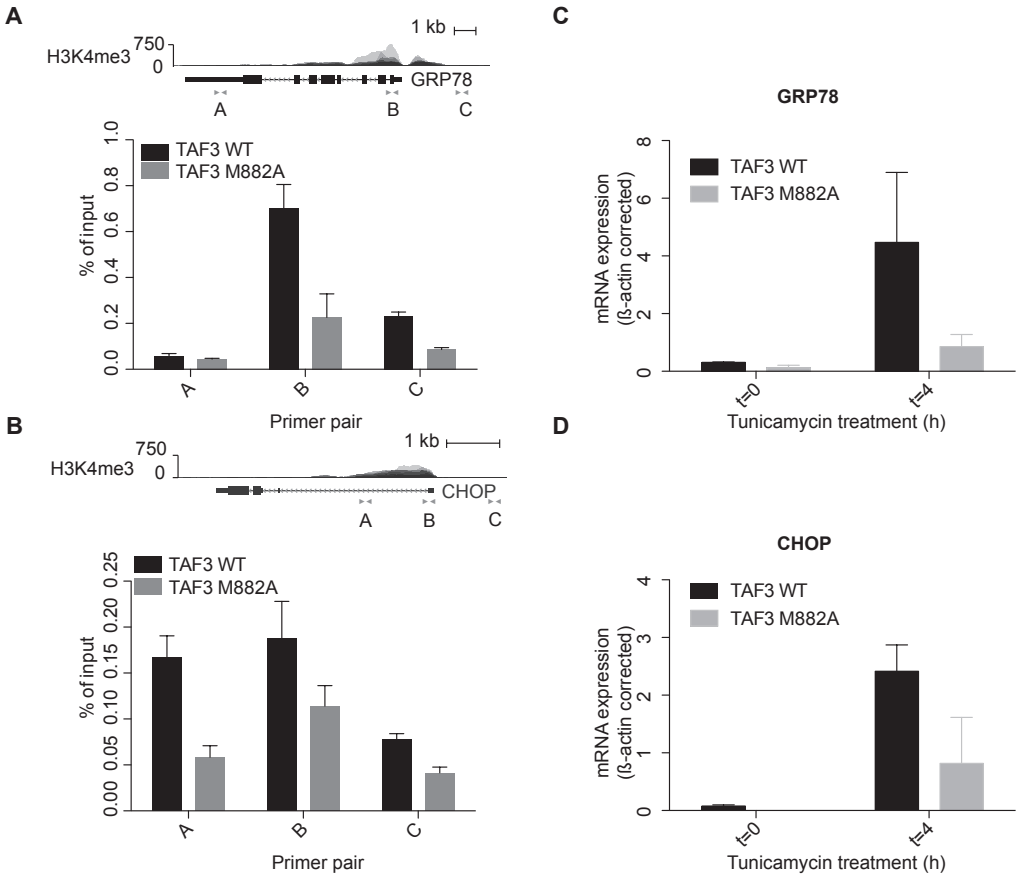


Figure 4. Taf3 binding to ER stress gene promoters is dependent on the PHD finger. (A) Overlay of ENCODE H3K4me3 profiles from seven human cell lines at the GRP78 locus (upper panel). ChIP analysis of GFP-Taf3 and GFP-Taf3 M882A at GRP78 with the indicated primer sets (lower panel). Standard deviations represent technical triplicates and similar results were observed in at least three independent experiments. (B) Similar labeling as in (A) but for the CHOP locus. (C and D) Analysis of mRNA expression levels of GRP78 and CHOP by quantitative RT-PCR. Levels were normalized to β-ACTIN and are presented as change compared to a control DMSO-treated sample. Samples were analyzed 4 hours after tunicamycin treatment. Standard deviations represent two biological duplicates.

substantiate our earlier results in which these three activities were found to act agonistically to anchor TFIID on 'active' promoter nucleosomes.

Taf3 requires its PHD finger for binding to H3K4me3 *in vivo*

We have demonstrated that incorporation of Taf3 lacking a functional PHD domain does not change the composition and stoichiometry of TFIID. Next, we wanted to determine how TFIID recruitment to target sites is affected by the absence of a functional PHD finger. Recent experiments in our lab have shown that H3K4me3 is present on ER stress responsive genes prior to stress, presumably to maintain these genes in a 'poised' chromatin state. Furthermore, in the

absence of Taf3, activation of ER stress genes such as *GRP78* and *CHOP* is impaired (Schram et al, 2013). Therefore, we used the GFP-Taf3 and GFP-Taf3 M882A cell lines to determine if the PHD-H3K4me3 interaction is important for the binding to and expression of these ER stress responsive genes *in vivo*. Mining published ENCODE ChIP sequencing data for H3K4me3 in different human cell lines revealed that H3K4me3 is found at the TSS of *GRP78* and *CHOP* in the absence of ER stress (Fig. 4A and B, upper panels). Scanning GFP ChIPs of the *GRP78* and *CHOP* loci revealed that Taf3 binding correlates well with the presence of H3K4me3 (Fig. 4A and B). As was observed for H3K4me3, Taf3 is present at these promoters prior to ER stress, which is indicative of a 'poised' state. Strikingly, the Taf3 mutant M882A, which can no longer bind to H3K4me3, shows impaired binding to the ER stress gene promoters (Fig. 4A and B). These results mirror recent findings showing that Taf3 is recruited to a specific subset of promoters enriched for DNA damage response related genes (Lauberth et al. 2013).

Next, the inducible GFP-Taf3 and GFP-Taf3 M882A cell lines were used to investigate the mRNA expression of ER stress target genes using the ER stress inducing agent tunicamycin. Strikingly, the induction of both *GRP78* and *CHOP* is impaired in the Taf3 mutant cell line compared to the Taf3 wild type (Fig. 4C and D). These results indicate that the interaction with H3K4me3 is required for the recruitment to and/or stabilization of Taf3 on ER stress gene promoters *in vivo*. Furthermore, this interaction is required for the rapid induction of ER stress responsive genes. Taken together, these *in vivo* observations strengthen our biochemical data and reveal that interactions with active histone modifications are relevant for TFIID binding to nucleosomes.

Discussion

The TFIID complex is important for the transcription initiation process and plays a major role in setting up the PIC at pol II promoters. Here, we have shown that synergistic effects of functional DNA elements and histone modifications mediate high affinity binding of TFIID to promoters.

Stoichiometry analyses of affinity purified TFIID complexes through a core (Taf5) and a peripheral (Taf3) subunit revealed the presence of a stable core complex consisting of two copies of Taf4, Taf5, Taf6, Taf9, Taf10 and Taf12. These results are in agreement with recent work by Berger and colleagues who used recombinantly expressed TFIID (sub)complexes for structural (cryo-EM) studies (Bieniossek et al. 2013). Outside of this stable symmetric core, peripheral Tafs appear to be present in substoichiometric amounts. These observations could be indicative of heterogeneity within holo TFIID complexes, as was suggested previously by Tora and colleagues (Müller et al. 2007). Each of these distinct holo TFIID complexes, all bearing a subset of peripheral Taf proteins, might serve its own specific target genes in a particular tissue or cell state. Indeed, individual Tafs and Taf variants have been shown to specify certain cell fates during development. Deletion of Taf7l in mice, for example, results in defective spermatogenesis (Cheng et al. 2007). It was also suggested that TFIID adopts different structural

configurations dependent on the subunit composition as incorporation of Taf4b induces an more open configuration compared to Taf4 containing complexes (Liu et al. 2008).

We observed a functional interplay between DNA elements and histone modifications on TFIID binding to promoter nucleosomes. Interestingly, the positive effect of a canonical TATA box was only apparent in the context of $H3K_c4me3$ and H3K14 acetylation. This result could indicate that acetylation of histone tails affects the accessibility of the TATA sequence for TBP binding. However, recent structural studies on TFIID binding to a TATA containing template showed that a structural rearrangement of Taf1/2 in TFIID can be induced by TATA DNA binding (Cianfrocco et al. 2013). These observations combined with our data suggest that structural changes in TFIID induced by interactions with DNA elements or histone modifications could result in the exposure of additional chromatin binding surfaces, which would potentiate the complex for multivalent engagement. Additionally, the spacing between the +1 nucleosome and the TATA element could affect TFIID binding to different promoters in a specific manner (Rhee and Pugh 2012).

We used ChIP experiments to show that Taf3 binds to the promoters of 'poised' stress genes. This binding is severely compromised when expressing a Taf3 mutant containing a point mutation (M882A), which cannot bind to H3K4me3. Roeder and colleagues recently showed that knock down of Taf3 only results in a minor change in global pol II dependent transcription. However, for the induction of early p53 response genes as well as ER stress genes, the interaction between Taf3 and H3K4me3 appears to be critical (Lauberth et al. 2013). Together these data illustrate that H3K4me3 binding by TFIID is only required when rapid induction of transcription is demanded. When unchallenged, acetyl and TATA binding can be sufficient for TFIID loading. These experiments from Roeder and colleagues, together with our data, further suggest that H3K4me3 can act either independently or cooperatively with a TATA box to regulate PIC formation and transcription. This, together with the fact that only ~10% of human pol II promoters contain a canonical TATA box, implies a certain degree of structural plasticity regarding TFIID engagement on different target genes. Interestingly, yeast TFIID lacks an H3K4me3 binding domain. Nevertheless, recent high resolution ChIP in yeast revealed that TFIID binding partially overlaps with the position of the +1 nucleosome (Rhee and Pugh 2012), indicating that TFIID can bind simultaneous to the nucleosome depleted region and to the first nucleosome. This observation is more pronounced on Taf1 depleted genes, indicating that SAGA and TFIID regulated genes are different in promoter architecture, at least in yeast (Basehoar et al. 2004). A systematic analysis of human promoter architecture and TFIID association however remains to be done.

Future experiments can be directed towards determining the exact position of TFIID subunits in promoter bound complexes. Furthermore, functional domain mapping experiments could be pursued to dissect the molecular mechanisms underlying the multivalent engagement of TFIID at various promoter nucleosomes. Additional stoichiometric analysis on nucleosome

bound TFIID using different promoter and/or enhancer nucleosomes for affinity purification could reveal the exact composition of the TFIID complex binding to these nucleosomes. In these experiments, we only made use of '601' based DNA sequences to avoid nucleosome sliding and transcription factor binding. Future experiments using promoter DNA sequences should reveal the contribution of additional DNA elements to TFIID binding. Finally, deciphering the genome wide profile of individual core and peripheral TFIID subunits in different cellular systems and stress conditions will increase our understanding regarding the assembly and composition of TFIID (sub)complexes and their role in the regulation of transcription initiation.

Materials & Methods

Plasmids and cell culture

The ORF of the bait protein was amplified by PCR using the relevant human cDNA constructs and introduced into pDONR2.1. The DNA sequence of the amplified ORF was verified and introduced into a GATEWAY compatible version of pCDNA5/FRT/TO essentially as described before (van Nuland et al. 2013a). Mouse Taf3 and mutant M882A were tagged by GFP at the N-terminus. Stable doxycycline inducible cell lines were created by transfecting pCDNA5/FRT/TO and pOG44 into HeLa FRT cells carrying the TET repressor using polyethyleneimine followed by antibiotic selection. Cells were grown in DMEM with high glucose supplemented with pen/strep and L-Glutamine (all LONZA) under blasticidin and hygromycin B selection. pRPN-mTaf3 (PHD) was described previously (Vermeulen et al. 2007). pDUET-H3K4C,K14X was derived by introducing K4C, C110A mutations and an amber codon at position 14, into Drosophila melanogaster histone H3. All histone H3 plasmids carried a C110A mutation. Amber codon histones were inserted into the pDUET plasmid using Nco1/Xho1 and transformed in bacteria that already carried pAC-KRS (kind gift of Robert Schneider). Point mutations in H3 were introduced using the Quickchange protocol (Stratagene) and verified by DNA sequencing. Other histone proteins were expressed from pET21b (gift from Y. Bai).

GFP affinity purification and sample preparation

Extract preparation (Dignam et al. 1983) and affinity purifications using GFP beads (Hubner et al. 2010) were performed essentially as described before. Briefly, nuclei were isolated using hypotonic lysis and nuclear extracts were prepared by using 420 mM NaCl Purifications for GFP lines and WT HeLa cells were performed in triplicate using 1 mg of nuclear extract per purification and GFP binder beads (CHROMOTEK) in 20 mM HEPES-KOH pH 7.9, 20% glycerol, 300 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP-40, 0.5 mM DTT and complete protease inhibitors (Roche). All purifications included 50 µg/ml ethidium bromide to suppress DNA mediated interactions. After 2.5 hours incubation at 4°C the beads were extensively washed

and on-bead digestion was performed using 0.1 µg trypsin (Promega).

Protein expression and nucleosome reconstitution

GST-mTaf3 PHD was expressed in E. coli strain BL21DE3 at 37°C in LB medium. Drosophila histones were expressed in E. coli strain BL21DE3 codon+ or Rosetta and prepared essentially as previously described (Dyer et al. 2004). For expression of acetylated histones 20 mM Nicotinamide and 10 mM N-acetyl L-Lysine (Sigma) was added to the cultures at OD_{600}=0.6 and protein expression was induced after 30 minutes using 0.5 mM IPTG as described before (Neumann et al. 2009). Histone H3K4C and derivatives were alkylated essentially as described before (Simon et al. 2007; van Nuland et al. 2013b). 167-bp DNA ('601 Widom' positioning sequence or TATA mutants) was produced by PCR amplification using one biotinylated primer, purification using DEAE anion exchange and ethanol precipitation. After octamer refolding, nucleosomes were reconstituted with the DNA using salt displacement.

Extract preparation and nucleosome pull downs

Hela S3 cells were cultured in large quantities using a bioreactor setup in MEM depleted from Lysine and Arginine supplemented with dialyzed FBS, Pen/Strep, L-Glutamine (all LONZA) and normal or $^{13}C^{15}N$- labeled arginine and lysine (Isotec). Nuclear extracts were prepared by isolating the nuclei and hypertonic lysis as described before (Vermeulen et al. 2007; Dignam et al. 1983). For nucleosome pull down assays magnetic Streptavidin beads (Sigma, MyOne) were coated with 130 pmol nucleosome and incubated in pull down buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.2 mM EDTA, 20% glycerol, 0.1% NP-40 and 1mM DTT) for one hour at 4 °C. After washing twice, 500 mg nuclear extract was added and beads were incubated for 2-3 hours rotating at 4°C. For mass spectrometry experiments, heavy labeled modified nucleosome pull downs and light controls were mixed at this point. Proteins were eluted from the beads in sample buffer after extensive washing and the bound proteins were analyzed by immunoblotting or processed for LC-MS/MS. Peptide pull down experiments were performed essentially as described (Vermeulen et al. 2007). Briefly, biotinylated peptides were alkylated as described before (Simon et al. 2007) and incubated with magnetic Streptavidin beads (Sigma, MyOne). After incubation and extensive washing beads were incubated with GST-Taf3 (PHD) lysate. Bound protein was visualized using Coomassie blue staining.

Mass spectrometry

Eluted proteins were separated on a SDS-PAGE gel and stained using Colloidal blue staining (Invitrogen). Lanes were sliced into eight pieces and samples were subjected to in-gel digestion with 0.1 µg trypsin (Promega) as described before (Vermeulen et al. 2007). Tryptic peptides were extracted from the individual gel slices and concentrated using stagetips with C18 material.

The peptides were applied to online nanoLC-MS/MS, using a 120 minutes acetonitrile gradient. Mass spectra were recorded on a LTQ-Orbitrap-Velos mass spectrometer (Thermo) selecting the 15 most intense precursor ions of every full scan for fragmentation. The data was analyzed using the Max-Quant software package (Cox and Mann 2008).

Chromatin immunoprecipitation

Cells were cross linked at 80-90% confluency using 1% paraformaldehyde in PBS for 10 min at room temperature. Reactions were quenched by addition of 125 mM glycine for 5 minutes on ice. After a cold PBS wash cells were scraped and collected by centrifugation (5 min, 400 g, 4°C). Pelleted cells were resuspended in ChIP lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 7.9, 1 mM DTT, 5 µM sodium butyrate (Merck) and complete protease inhibitors (Roche)) and disrupted by sonication (Bioruptor, Diagenode: seven cycles, 30 sec on/off, high setting) to produce an average DNA fragment size of ~400-bp. Samples were centrifuged (5 min, 200 g, 4°C) and supernatant collected. For immunoprecipitation, chromatin was diluted in IP buffer (0.5% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.9, 150 mM NaCl, 1 mM DTT, 5 µM sodium butyrate and complete protease inhibitors (Roche)), 1-5 ug antibody was added and rotated overnight at 4°C. Immunocomplexes were collected for 4 hrs at 4°C on protein A/G PLUS-agarose beads (Santa-Cruz), after o/n blocking in 1.5% fish gelatin and washing. Subsequently beads were washed four times at 4°C with wash buffer (0.25% NP-40, 0.05% SDS, 2 mM EDTA, 20 mM Tris-HCl pH 7.9, 250 mM NaCl, 5 µM sodium butyrate and complete protease inhibitors) and once with TE (10 mM Tris-HCl pH 6.8, 1 mM EDTA). Cross links of protein-DNA were reversed by overnight incubation at 65°C in 100 µl elution buffer (100 mM $NaHCO_3$, 1% SDS). Samples were treated with 1 mg/ml proteinase K (Roche) and 1 mg/ml RNase A for 2 hours at 37°C. DNA was purified using PCR purification kit (Qiagen) and amplified in a 25 µl reaction mixture (iQ SYBR green supermix (Biorad)) in a realtime PCR machine (CFX96, Biorad). Primer sequences are available upon request.

mRNA expression analysis

Total RNA was isolated using RNeasy kit (Qiagen) and cDNA was synthesized using the First-strand cDNA synthesis kit (Qiagen) both according to the manufacturers manual. Subsequently the cDNA was amplified in a 25 µl reaction mixture (iQ SYBR green supermix (Biorad)) in a realtime PCR machine (CFX96, Biorad). Primer sequences are available upon request.

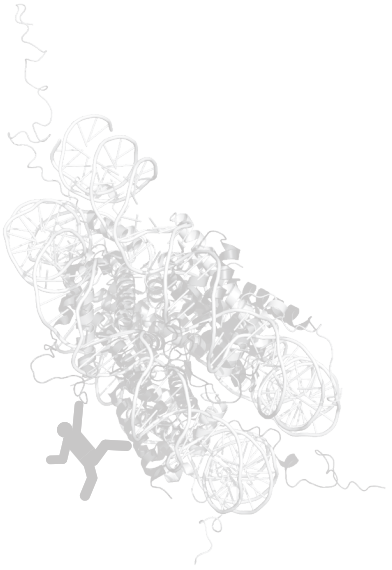
Antibodies

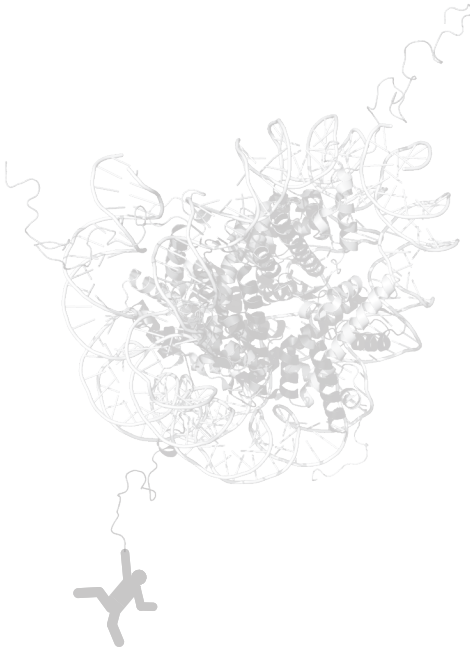
α-TBP (SL30) (gift from Henk Stunnenberg), α-Taf3 and α-Taf5 (obtained from Bob Roeder), α-Taf6 (25TA-2G7, Euromedex), α-TFIIB (Santa Cruz), α-H3 (Abcam Ab1791), α-H3K4me3 (home made), α-H3K9ac (Cell Signaling 9671), α-H3K14ac (Karmodiya et al. 2012) and H2A.Z (Abcam

Ab18263) were used for immunoblotting. For ChIP and immunoblotting α-GFP (gift from Geert Kops) was used. Quantification of antibody signals was performed using Adobe Photoshop.

Acknowledgements

We thank members of the Timmers and Vermeulen lab for useful suggestions and discussions. We are very grateful to dr. Susan Kloet for critical reading of the manuscript and thank Arne Smits for advice on the stoichiometry determination. We acknowledge dr. Robert Schneider for reagents and discussions. Dr. Geert Kops and dr. Robert G. Roeder kindly provided antibodies. Work in the Timmers lab is supported by the Netherlands Proteomics Center. Work in the Vermeulen lab is supported by the Netherlands organization for scientific research (NWO-VIDI) and the Dutch Cancer Society.





Chapter 6

General discussion

In the work described in this thesis multiple approaches have been followed in order to deduce the dynamic interplay between histone modifications and transcription regulatory complexes. In this chapter, the general implications of our work for chromatin regulation are discussed and where applicable additional experiments as well as directions for further research are proposed. Furthermore, our findings on the interactions of the Menin tumor suppressor protein are discussed and placed in a clinical perspective.

Menin interactions in context

In this thesis results are presented which shed light on the function of the tumor suppressor Menin and the molecular interactions and processes this protein is involved in. In chapter 2 we show that Menin interacts with the vitamin D nuclear receptor and thereby regulates its transcriptional activity, also in the context of parathyroid tumors. Similar interactions were previously described with the PPARγ and ERα nuclear receptors (Dreijerink et al. 2006; 2009). Many identified Menin interactions appear to be transient and/or only present in certain cell types under specific conditions. This is probably why they were not identified in our unbiased quantitative mass spectrometry based interaction screen described in chapter 3. The direct interaction between Menin and ERα for example, is dependent on addition of the receptor ligand (Dreijerink et al. 2006). Moreover, estrogen responsive breast cancer cell lines which express high levels of ERα, were used in previous studies. In contrast, the HeLa cell lines that were used for our quantitative interaction studies, express ERα only at low levels. In chapter 3 we revealed that the majority of Menin interacts with the MLL1/2 methyltransferase complexes and the JunD transcription factor. Hence, Menin interactions apart from these stable 'core' binding partners will most likely be transient and dependent on the molecular context.

We observed in an unbiased manner that MLL1/2 and JunD are the prime Menin interactors in HeLa cells. These stable interactions were previously shown to be mutually exclusive and have opposing roles in transcription regulation (Yu et al. 1995; Huang et al. 2012). This could indicate that Menin acts as a dynamic switch. Nevertheless, it is unclear how and if Menin discriminates between these two interactions. It has been demonstrated that Menin interacts with short peptides of these proteins through a central binding groove. Post-translational modifications (PTM) on this interface could function to discriminate between JunD and MLL1/2 binding. Affinity purifications of Mll1/2 and/or JunD and analysis of the differential PTMs present on the associating Menin protein can possibly solve this issue. On the other hand, our absolute protein quantification data demonstrated that Menin is much higher expressed than Mll1 and Mll2 together. This leaves enough 'free' Menin to interact with JunD, despite the higher affinity for the Mll proteins (Huang et al. 2012). We also observed that only a small proportion of Menin is present in a complex. This free pool of Menin could serve as a scaffold for multiple transiently interacting factors. Interaction proteomics of Menin in different cell types and under different conditions is needed to determine these factors.

Although Mll1/2 and JunD are the Menin interactors with the highest stoichiometry, functionally not so much is known about the role of these proteins in the development of MEN1. Interestingly, heterozygous deletions of *Mll1* (Yu et al. 1995) or *JunD* (Thépot et al. 2000) do not mimic any of the observed tissue specific phenotypes of the *MEN1* heterozygous knockout mouse (Crabtree et al. 2001; Bertolino et al. 2003a). Menin heterozygous mice show a phenotype that closely resembles that of MEN1 patients, including multiple lesions in endocrine tissues upon loss of heterozygosity. One possibility for this discrepancy is the fact that Menin seems to act as a haploinsufficient tumor suppressor, since the loss of a single allele already induces faster proliferation of pancreatic islets (Lejonklou et al. 2012). In contrast, heterozygous deletions of *JunD* and *Mll1* or *Mll2* could result in sufficient functional complexes that can still suppress tumor formation in these mice.

The tumor suppressive function of Menin is generally attributed to its interaction with the MLL1 complex. This complex positively regulates the expression of the *CDKN1B* and *CDKN2C* (coding for $p27^{KIP1}$ and $p18^{INK4C}$, respectively) cell cycle inhibitor genes, through catalyzing H3K4me3 at their promoters (Karnik et al. 2005; Milne et al. 2005). Interestingly, double knockout mice for *CDKN1B* and *CDKN2B* develop a tumor phenotype that closely resembles the MEN1 mouse (Franklin et al. 2000). In humans however, little evidence exists for decreased $p18^{INK4C}$ or $p27^{KIP1}$ expression levels in MEN1 tumors (Lindberg et al. 2008). Additionally, we observed no clear differences in $p18^{INK4C}$ or $p27^{KIP1}$ expression in mouse embryonic fibroblasts derived from a Menin knockout mouse (data not shown). Genome wide chromatin immunoprecipitation (ChIP) approaches of Menin and its binding partners could help to dissect the indirect and direct targets of Menin. These targets may provide us with more insights in the mediators of tumor onset and progression in MEN1. Addressing these questions however requires a relevant MEN1 model cell system. Unfortunately, our efforts to derive tumor spheroids cultures from MEN1 patients have proven unsuccessful (data not shown). Nontheless, such approaches might hopefully provide us in the future with novel therapeutic strategies or biomarkers and ultimately improve patient care. Classifying MEN1 as an epigenetic disease through the misregulation of H3K4me3 at target genes opens novel opportunities. Recently inhibitors were developed, that target the Jarid1 family of H3K4me demethylases. Inhibition of H3K4 demethylation in MEN1 patients might help to restore the H3K4me balance at Menin targets that are downregulated as a consequence of Menin loss. This strategy however requires a detailed insight in the factors involved in maintaining the H3K4me balance in cells.

In the heterozygous MEN1 mouse, additional non-MEN1 related tumors are found that could be the result of a general cancer predisposition upon loss of *MEN1* (Bertolino et al. 2003b). Recently, Menin was also implicated as a tumor suppressor in melanomas through the misregulation of genes coding for proteins involved in DNA damage repair (Fang et al. 2013a). Moreover, *MEN1* was classified as a driver tumor suppressor gene based on genome sequencing studies (Vogelstein et al. 2013) underscoring the general tumor suppressive role of Menin. These

observations rationalize the efforts to understand Menin and its function at the molecular level.

Additional clinical, as well as molecular analysis of the relatively small MEN1 population is required to investigate the development and progression of MEN1. The relatively low numbers of patients with MEN1 (approximately 300 in The Netherlands) hampers large scale epidemiological studies. Nevertheless, major improvements are being made in systematically monitoring MEN1 patients. Initiatives such as the DutchMEN1 study group (Pieterman et al. 2012) can help to understand the molecular and clinical progression of the disease, genotype/phenotype correlations and improve future patient care.

In conclusion, our quantitative work on the molecular interactions of the Menin tumor suppressor indicate that misregulation of either JunD or MLL1/2 specific target genes is underlying the development of tumors in MEN1 patients. These pathways should therefore be the prime candidates to focus on in targeted drug development against MEN1 syndrome.

SET1/MLL, a multifaceted complex

In order to derive a detailed picture on the complexes that are involved in the deposition of H3K4me, we dissected the SET1/MLL complex composition by GFP tagging and mass spectrometry of eleven individual subunits as described in chapter 3. This revealed that the stable WRAD (Wdr5-Rbbp5-Ash2l-Dpy30) module, together with the enzymatic component form the core complex. Many associating proteins are present in lower stoichiometries indicating more transient interactions. This heterogeneous composition hints towards the existence of multiple functionally distinct subcomplexes, each of which harbors the WRAD-module as a stable enzymatic core. Alternatively, these WRAD module containing, substoichiometric assemblies may also represent inactive assembly intermediates or could be the result of the dissociation of peripheral subunits during sample preparation. It should be mentioned that the purifications described in chapter 3 were done using soluble nuclear extracts. Given their high affinity for chromatin, part of the SET1/MLL complexes may remain in the insoluble chromatin fraction during nuclear extraction. In light of this, it would be interesting to purify SET1/MLL complexes from the insoluble chromatin fraction using nucleases to solubilize chromatin bound complexes. Furthermore, there is evidence for differential chromatin binding of SET1/MLL subunits during the cell cycle (Blobel et al. 2009; Mishra et al. 2009). It would therefore be interesting to expand the stoichiometry analyses and investigate complex composition during different cell cycle stages, potentially revealing differential cell cycle dependent complex compositions.

Distinct SET1/MLL subcomplexes may have different genomic targets or cellular functions. One possibility is that Ptip-containing MLL3/4 complexes for example activate different genes compared to complexes lacking Ptip. Furthermore, Ptip has been shown to be important for the induction of H3K4me at sites of DNA damage (Muñoz and Rouse 2009; Daniel et al. 2010). Nonetheless, it is currently unclear how this links to activation of transcription. Additional knockdown and replacement studies, using chromatin binding mutants of SET1/MLL subunits,

are essential to examine the contribution of the different subunits and their domains in forming the H3K4 methylation pattern. Genome wide profiling of these substoichiometric SET1/MLL interactors is required to dissect their function and specificity. This could be particularly rewarding using an inducible setup in which transcriptional programs are activated such as hormone stimulation, induction of differentiation or various stress conditions rather than a static steady state condition such as proliferating populations of cancer cell lines. The catalytic subunits of the SET1/MLL complexes were not tagged in the described work due to their extreme molecular weight (186-593 KDa). As an extension of our work, endogenous tagging could be applied to investigate if these proteins are always in an active complex with WRAD or whether they could also be present in an inhibited state bound to other proteins. Such mechanism was observed for the Mll1 protein upon interaction with Cyp33 (Wang et al. 2010). RNA can also play an important role in the assembly of such multisubunit complexes (Halbach et al. 2009). Relevant in this regard is the fact that we included ethidium bromide during GFP affinity purifications to prevent indirect, nucleic acid mediated interactions. Potential RNA mediated interactions within the complex are also disrupted by ethidium bromide but whether this influences SET1/MLL assembly remains to be tested.

The presence of specific subunits in all SET1/MLL complexes analyzed in chapter 3 raised questions on the molecular assembly of the complexes. It is possible that specific subunits like Menin and Ptip localize to chromatin prior to the rest of the complex. These proteins would thereby sequentially recruit the methyltransferase activity to a locus. The development of several methods to study this sequential recruitment hypothesis such as the rapamycin induced targeting of factors to chromatin, can be a helpful tool in unraveling this (Hathaway et al. 2012). Additionally, tissue specific effects of the different complexes have to be addressed by using conditional deletion mouse models for the different enzymes and subunits. In conclusion, more systematic approaches have to be applied in order to dissect the functional aspects as well as the mechanism of recruitment of the SET1/MLL complexes.

Mll1 translocations are a major cause of infant leukemia. Detailed understanding of the mechanism of recruitment of Mll1 and its oncogenic fusions could result in the identification of potential drug targets (Deshpande et al. 2012). Recently compounds were identified that can inhibit the acetyl lysine binding of the BET Bromodomain containing Brd proteins, that associate with Mll1-fusion complexes. The inhibition of Brd2 and Brd4 was shown to block oncogenic transformation of leukemic cells (Dawson et al. 2011). Analogous to this, inhibition of MLL complex formation or chromatin recognition by using small molecule inhibitors against Menin or Psip1 might potentially give similar results (Shi et al. 2012).

Transcription regulation has been considered the major pathway in which H3K4me plays a role. Interestingly, we identified Bod1 as a novel subunit of the SET1B complex. Previously, this protein was shown to play a role in chromosome segregation (Porter et al. 2007; Compton 2007). It is interesting to note that chromatin at centromeres is highly enriched for the Cenp-A

variant of histone H3 interspersed with nucleosomes that carry H3K4me2 (Stimpson and Sullivan 2011). One could hypothesize that the Bod1 containing SET1B complexes are required for specific methylation at centromeres. Lack of this modification may underlie the observed mitotic defects in Bod1 knockdown cells (Porter et al. 2007). High resolution imaging of H3K4 methyltransferases at centromeres and/or analysis of the modification status of centromeric nucleosomes following Set1b knockdown could be pursued to answer this question. Alternatively, the effect of knockdown of the different KMT complexes on mitotic progression could give insights in the function of the active chromatin environment at these sites.

Psip1 and nucleosome recognition

The PWWP domain containing Psip1 was detected in substoichiometric amounts in MLL2 as described in chapter 3. On the other hand, Psip1 interacts with the transcription factor Cdca7l in an apparent high stoichiometry. Cdca7l is a low abundant protein and it appears that most, if not all molecules are in complex with the highly abundant Psip1. The function of this interaction however, remains unknown and requires more experiments. In addition, the Cdc7 kinase complex appears to interact with Psip1 in comparable stoichiometry as MLL1/2 but was not identified in the other purifications. Cdc7 can phosphorylate MCM complex subunits and thereby regulate G1/S transition and DNA replication (Montagnoli et al. 2002). DNA replication timing depends on the transcriptional activity of the locus (Dorn and Cook 2011). Highly transcribed (H3K36me marked) chromatin is replicated first and therefore it is possible that Psip1 targets the replication machinery to H3K36me marked loci. Experimental evidence for such a model however is lacking so far and knockdown of Psip1 did not result in clear cell cycle defects in a preliminary experiment (data not shown).

The PWWP domain of Psip1 targets it to H3K36me3 marked, actively transcribed chromatin as we demonstrate in chapter 4. Psip1 interacts with several proteins through its IBD domain among which are the HIV integrase and Menin (Yokoyama and Cleary 2008; Engelman and Cherepanov 2008). Integration of viral DNA in transcriptionally active loci could be beneficial with respect to the virus. At these loci, all requirements for transcription of viral DNA, such as Pol II and elongation factors are already present. The function of Psip1 in the MLL2 complexes nonetheless remains unclear, as well as chromatin binding of the MLL2 complex as a whole. H3K36me and H3K4me are generally considered to be present at different genomic locations, yet there are numerous loci where a significant overlap is present, especially around the +1 nucleosome position. These loci include the *HOX* genes, that are considered genuine MLL1/2 target genes (Yokoyama et al. 2004; Shilatifard 2012). Psip1 may therefore help to restrict the MLL2 complex to the +1 nucleosome position at these genes. This model could well only apply for a limited number of target genes, considering the low stoichiometry of Psip1 within the complex. The genome wide H3K4me distribution in cells where Psip1 is replaced with PWWP mutants, can potentially reveal the direct effect of Psip1 on H3K4me levels.

Recently, multiple PWWP domain proteins have been identified, that share the ability to associate with H3K36me3 marked chromatin (Vermeulen et al. 2010; Li et al. 2013; Fang et al. 2013b). These data validate our modeling approaches described in chapter 4, in which we propose that PWWP domains have a conserved function in nucleosome recognition by binding to the histone tail as well as to DNA. Such mechanisms may also apply for the recognition of other modifications close to the nuclear core such as H4K20me (Qiu et al. 2012). Furthermore, methylation in the globular core of the nucleosome particle, such as H3K79me, could well be recognized by a similar mechanism. So far, peptide based assays were not able to identify methylation specific interactions for this site (Vermeulen et al. 2010). Therefore the use of *in vitro* reconstituted nucleosomes could help to dissect the function and binding partners of this modification.

The use of nucleosomes in contrast to peptide based approaches, is well suitable for studying the combinatorial binding of protein complexes to different co-occurring modifications. Having said that, for detailed investigations on a single modification distal from the globular core, peptide based assays remain the desired method. This approach is much less labor intensive and easier to adapt and manipulate. In conclusion, one should consider the contribution of specific and non specific DNA binding in histone binding, as was also shown previously with respect to methylated and non-methylated DNA templates (Bartke et al. 2010).

Nucleosome recognition by TFIID

The deposition of H3K4me3 at promoters is important for the recruitment of the TFIID basal transcription factor complex (Vermeulen et al. 2007). Earlier, the Taf1 TFIID subunit was found to bind to acetylated histone tails through its double Bromodomain (Jacobson et al. 2000). We extended on these observations and performed interaction studies using nucleosomes carrying a combination of these modifications. We aimed to understand the multivalent engagement of TFIID to chromatin. To do so we constructed nucleosomes that contain methylated and acetylated lysines as wel as a TATA DNA sequence. The data presented in chapter 5 indicates that TFIID interacts synergistically with the different modifications and DNA sequences. This multivalent binding could potentially affect the dynamics of TFIID binding to different gene promoters in cells.

An interesting aspect of the TFIID complex was revealed when we analyzed its subunit stoichiometry. We found that TFIID, purified from HeLa cells, harbors a stable core consisting of two molecules of Taf4, Taf5, Taf6, Taf9, Taf10 and Taf12 in line with structural work form the Berger lab. Their structure shows that TFIID indeed consists of a symmetric core of dimers. Upon binding of the divergent Taf3/Taf10 and Taf8/Taf10 dimers, further assembly of single copy Tafs is induced in an asymmetrical fashion (Bieniossek et al. 2013). We show that the peripheral subunits are interacting with the core in a substoichiometric manner. This result illustrates that TFIID is highly dynamic in its nature. Mass spectrometry based studies on the

exchange of Tafs however, revealed only dynamic behavior of Taf12 and TBP in a Taf6 affinity purified complex under the used *in vitro* transcription assay conditions (N. Mischerikow, PhD thesis). Whether subcomplexes are present when TFIID is bound to chromatin remains open. In several reports on the structure of the TFIID complex, multiple specimens are observed for endogenous as well as for recombinant complexes, illustrating the flexibility of the complex structure (Berger et al. 2011). Interestingly, the TFIID complex was recently shown to be essential for pluripotency. Strikingly however, knockdown of some individual Tafs (Taf7 and Taf8) resulted in no phenotype, indicatory for a specific dependency on certain subcomplexes or subunits during reprogramming and differentiation, as was also suggested previously (Müller et al. 2007; Pijnappel et al. 2013). In light of this, Taf3 depletion only affects a subset of genes and Taf1 is only present at 75% of the transcribed genes. This argues for Taf1 and/or Taf3 independent transcription. Moreover, Taf1 also associates with a subset of silent promoters as was observed for Taf3 by us and others (Kim et al. 2005; Lauberth et al. 2013). These data indicate that TFIID can also interact with genes that are poised for rapid activation. The turnover of TFIID at these promoters might change upon activation and additional studies are required to pursue this further. It has been proposed that variations of TFIID might serve different promoters based on the DNA sequence motifs, nucleosome occupancy and histone modifications present (Müller et al. 2007). Bioinformatics analysis of promoter structure combined with global Taf binding profiles are essential to elucidate this complex question.

Recent structural work demonstrated that different modules in TFIID recognize distinct DNA elements in a core promoter (Cianfrocco et al. 2013). Binding of TBP and TFIIA to the TATA box element induced a structural rearrangement in the complex. Interestingly, the rearranged lobe consists of the Taf1/2 subunits. This rearrangement could potentially impact the histone tail binding of Taf1. Visualization of TFIID binding to the different nucleosomes that we describe in chapter 5 could therefore provide novel insights into the interplay between DNA and chromatin recognition by TFIID.

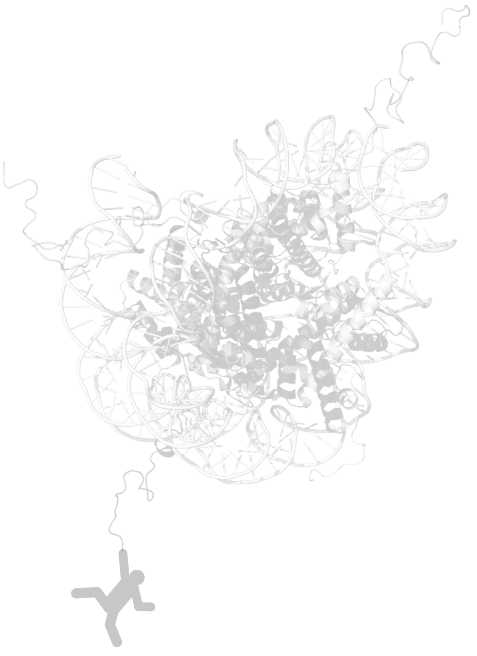
In this thesis we studied the direct effect of TFIID binding to H3K4me3 decorated nucleosomes, thereby ignoring the function of chromatin remodeling complexes. Interestingly, the Bptf subunit of the NURF remodeler complex also has the ability to bind directly to H3K4me3 and therefore competes with the TFIID complex at promoters (Li et al. 2006; Wysocka et al. 2006). The affinity for H3K4me3 of the Bptf PHD finger however is weaker than the affinity of the Taf3 PHD. Nevertheless, local H3K4me3 dependent remodeling of the promoter region could seriously affect TFIID localization in cells. The interplay between TFIID and chromatin remodeling complexes therefore requires further studies, for example to investigate TFIID binding in cells that lack remodeling activities.

In this thesis we only explored a subset of possible combinations of PTMs and DNA sequences. Thorough analysis of promoter architectures and the presence of chromatin modifications will bring additional possible combinations that are worth testing. To extend on this, the use

of nucleosome arrays with different modifications and DNA elements in diverse orientations provides a more detailed picture on the assembly of the PIC on designated templates. Additionally these 'designer' nucleosomes can be used to study other chromatin 'signatures' such as enhancer nucleosomes, that carry H3K4me1 and H3K27ac in combination with a transcription factor binding site containing sequence (Calo and Wysocka 2013). The use of nucleosomal arrays however, is technically challenging and should therefore be designed on a well defined chromatin template. One technical drawback of our approach is recognition of the methyl lysine analog (MLA). The affinity for this mimic needs to be determined empirically since some PHD fingers show a reduced binding compared to the natural histone peptide (H.T.M. Timmers, unpublished work). This effect could create a considerable bias when applying MLA for interaction studies.

In conclusion, in this thesis we used state-of-the-art methods to investigate the molecular interactions of complexes involved in chromatin and transcription regulation in great detail. One major question that remains is how chromatin modifying complexes such as SET1/MLL are recruited to their target genes. The identification of the vitamin D receptor as a direct Menin interactor demonstrates that the transient binding of sequence specific transcription factors to these complexes can contribute to this. Many chromatin modifiers are implicated in cancer and neurodegenerative diseases. At this point we are only at the start of designing drugs against these factors. A detailed understanding of the factors involved is therefore crucial and requires the use of quantitative methods as well as unbiased approaches. Finally, a general concept that appears is that interactions of proteins and protein complexes with chromatin are multivalent. This also illustrates that combinatorial therapies targeting the different interaction surfaces are required to inhibit these contacts. Our findings may therefore be of interest for the development of potential epigenetic drugs against leukemia and tumors in the MEN1 spectrum.

Bibliography



Adelman K, Lis JT. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**: 720–731.

Agarwal SK, Guru SC, Heppner C, Erdos MR, Collins RM, Park SY, Saggar S, Chandrasekharappa SC, Collins FS, Spiegel AM, et al. 1999. Menin interacts with the AP1 transcription factor JunD and represses JunD-activated transcription. *Cell* **96**: 143–152.

Agarwal SK, Novotny EA, Crabtree JS, Weitzman JB, Yaniv M, Burns AL, Chandrasekharappa SC, Collins FS, Spiegel AM, Marx SJ. 2003. Transcription factor JunD, deprived of Menin, switches from growth suppressor to growth promoter. *Proc Natl Acad Sci USA* **100**: 10770–10775.

Agger K, Cloos PAC, Christensen J, Pasini D, Rose S, Rappsilber J, Issaeva I, Canaani E, Salcini AE, Helin K. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**: 731–734.

Albert TK, Hanzawa H, Legtenberg YIA, de Ruwe MJ, van den Heuvel FAJ, Collart MA, Boelens R, Timmers HTM. 2002. Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. *EMBO J* **21**: 355–364.

Ang Y-S, Tsai S-Y, Lee D-F, Monk J, Su J, Ratnakumar K, Ding J, Ge Y, Darr H, Chang B, et al. 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**: 183–197.

Atanassov BS, Koutelou E, Dent SY. 2011. The role of deubiquitinating enzymes in chromatin regulation. *FEBS Lett* **585**: 2016–2023.

Avdic V, Zhang P, Lanouette S, Groulx A, Tremblay V, Brunzelle J, Couture J-F. 2011. Structural and biochemical insights into MLL1 core complex assembly. *Structure* **19**: 101–108.

Ayton PM, Cleary ML. 2003. Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. *Genes Dev* **17**: 2298–2307.

Bajic VB, Tan SL, Christoffels A, Schönbach C, Lipovich L, Yang L, Hofmann O, Kruger A, Hide W, Kai C, et al. 2006. Mice and men: their promoter properties. *PLoS Genet* **2**: e54.

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**: 10037–10041.

Barak O, Lazzaro MA, Lane WS, Speicher DW, Picketts DJ, Shiekhattar R. 2003. Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J* **22**: 6089–6100.

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837.

Bartholomeeusen K, Christ F, Hendrix J, Rain J-C, Emiliani S, Benarous R, Debyser Z, Gijsbers R, De Rijck J. 2009. Lens epithelium derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ. *J Biol Chem* **284**: 11467–11477.

Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. 2010. Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell* **143**: 470–484.

Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709.

Bedford MT, Clarke SG. 2009. Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33**: 1–13.

Berger I, Blanco AG, Boelens R, Cavarelli J, Coll M, Folkers GE, Nie Y, Pogenberg V, Schultz P, Wilmanns M, et al. 2011. Structural insights into transcription complexes. *J Struct Biol* **175**: 135–146.

Berger SL. 2007. The complex language of chromatin regulation during transcription. *Nature* **447**: 407–412.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, Gingeras TR, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**: 169–181.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315–326.

Bertolino P, Radovanovic I, Casse H, Aguzzi A, Wang Z-Q, Zhang C-X. 2003a. Genetic ablation of the tumor suppressor Menin causes lethality at mid-gestation with defects in multiple organs. *Mech Dev* **120**: 549–560.

Bertolino P, Tong W-M, Galendo D, Wang Z-Q, Zhang C-X. 2003b. Heterozygous Men1 mutant mice develop a range of endocrine tumors mimicking multiple endocrine neoplasia type 1. *Mol Endocrinol* **17**: 1880–1892.

Bian C, Xu C, Ruan J, Lee KK, Burke TL, Tempel W, Barsyte D, Li J, Wu M, Zhou BO, et al. 2011. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J* **30**: 2829–2842.

Bickmore WA, van Steensel B. 2013. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**: 1270–1284.

Bieniossek C, Papai G, Schaffitzel C, Garzoni F, Chaillet M, Scheer E, Papadopoulos P, Tora L, Schultz P, Berger I. 2013. The architecture of human general transcription factor TFIID core complex. *Nature* **493**: 699–702.

Bikle D. 2009. Nonclassic actions of vitamin D. *J Clin Endocrinol Metab* **94**: 26–34.

Black JC, Van Rechem C, Whetstine JR. 2012. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**: 491–507.

Blobel GA, Kadauke S, Wang E, Lau AW, Zuber J, Chou MM, Vakoc CR. 2009. A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol Cell* **36**: 970–983.

Bouillon R, Carmeliet G, Verlinden L, van Etten E, Verstuyf A, Luderer HF, Lieben L, Mathieu C, Demay M. 2008. Vitamin D and human health: lessons from vitamin D receptor null mice. *Endocr Rev* **29**: 726–776.

Brandi ML, Gagel RF, Angeli A, Bilezikian JP, Beck-Peccoz P, Bordi C, Conte-Devolx B, Falchetti A, Gheri RG, Libroia A, et al. 2001. Guidelines for diagnosis and therapy of MEN type 1 and type 2. Vol. 86 of, pp. 5658–5671.

Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, Winston F, Allis CD. 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. *Genes Dev* **15**: 3286–3295.

Brunger AT. 2007. Version 1.2 of the Crystallography and NMR system. *Nat Protoc* **2**: 2728–2733.

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica Section D, Biological crystallography* **54**: 905–921.

Buchwald PC, Akerström G, Westin G. 2004. Reduced p18INK4c, p21CIP1/WAF1 and p27KIP1 mRNA levels in tumours of primary and secondary hyperparathyroidism. *Clin Endocrinol (Oxf)* **60**: 389–393.

Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541–546.

Buratowski S, Hahn S, Guarente L, Sharp PA. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**: 549–561.

Burke TW, Kadonaga JT. 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA box-deficient promoters. *Genes Dev* **10**: 711–724.

Burke TW, Kadonaga JT. 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TafII60 of Drosophila. *Genes Dev* **11**: 3020–3031.

Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825–837.

Cao F, Chen Y, Cierpicki T, Liu Y, Basrur V, Lei M, Dou Y. 2010. An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain. *PLoS ONE* **5**: e14102.

Capotosti F, Hsieh JJ-D, Herr W. 2007. Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways. *Mol Cell Biol* **27**: 7063–7072.

Carlberg C, Bendik I, Wyss A, Meier E, Sturzenbecker LJ, Grippo JF, Hunziker W. 1993. Two nuclear signalling pathways for vitamin D. *Nature* **361**: 657–660.

Carling T, Kindmark A, Hellman P, Lundgren E, Ljunghall S, Rastad J, Akerström G, Melhus H. 1995. Vitamin D receptor genotypes in primary hyperparathyroidism. *Nat Med* **1**: 1309–1311.

Carling T, Rastad J, Szabó E, Westin G, Akerström G. 2000. Reduced parathyroid vitamin D receptor messenger ribonucleic acid levels in primary and secondary hyperparathyroidism. *J Clin Endocrinol Metab* **85**: 2000–2003.

Chalkley GE, Verrijzer CP. 1999. DNA binding site selection by RNA polymerase II Tafs: a Taf(II)250-Taf(II)150 complex recognizes the initiator. *EMBO J* **18**: 4835–4845.

Chandrasekharappa SC, Guru SC, Manickam P, Olufemi SE, Collins FS, Emmert-Buck MR, Debelenko LV, Zhuang Z, Lubensky IA, Liotta LA, et al. 1997. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* **276**: 404–407.

Chang P-Y, Hom RA, Musselman CA, Zhu L, Kuo A, Gozani O, Kutateladze TG, Cleary ML. 2010. Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL dependent gene transcription. *J Mol Biol* **400**: 137–144.

Chen Y, Cao F, Wan B, Dou Y, Lei M. 2012. Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30. *Cell Res* **22**: 598–602.

Chen Y, Wan B, Wang KC, Cao F, Yang Y, Protacio A, Dou Y, Chang HY, Lei M. 2011. Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding. *EMBO Rep* **12**: 797–803.

Cheng H-T, Chen J-Y, Huang Y-C, Chang H-C, Hung W-C. 2006. Functional role of VDR in the activation of p27Kip1 by the VDR/Sp1 complex. *J Cell Biochem* **98**: 1450–1456.

Cheng Y, Buffone MG, Kouadio M, Goodheart M, Page DC, Gerton GL, Davidson I, Wang PJ. 2007. Abnormal sperm in mice lacking the Taf7l gene. *Mol Cell Biol* **27**: 2582–2589.

Cho Y-W, Hong T, Hong S, Guo H, Yu H, Kim D, Guszczynski T, Dressler GR, Copeland TD, Kalkum M, et al. 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* **282**: 20395–20406.

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. 2009. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**: 834–840.

Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, Kadonaga JT, Nogales E. 2013. Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* **152**: 120–131.

Clapier CR, Chakravarthy S, Petosa C, Fernández-Tornero C, Luger K, Müller CW. 2008. Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **71**: 1–7.

Compton DA. 2007. Chromosome orientation. *J Cell Biol* **179**: 179–181.

Correa P, Segersten U, Hellman P, Akerström G, Westin G. 2002. Increased 25-hydroxyvitamin D3 1alpha-hydroxylase and reduced 25-hydroxyvitamin D3 24-hydroxylase expression in parathyroid tumors--new prospects for treatment of hyperparathyroidism with vitamin d. *J Clin Endocrinol Metab* **87**: 5826–5829.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.

Crabtree JS, Scacheri PC, Ward JM, Garrett-Beal L, Emmert-Buck MR, Edgemon KA, Lorang D, Libutti SK, Chandrasekharappa SC, Marx SJ, et al. 2001. A mouse model of multiple endocrine neoplasia, type 1, develops multiple endocrine tumors. *Proc Natl Acad Sci USA* **98**: 1118–1123.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**: 21931–21936.

Dai J, Hyland EM, Yuan DS, Huang H, Bader JS, Boeke JD. 2008. Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. *Cell* **134**: 1066–1078.

Dai J, Sultan S, Taylor SS, Higgins JMG. 2005. The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes Dev* **19**: 472–488.

Daniel JA, Santos MA, Wang Z, Zang C, Schwab KR, Jankovic M, Filsuf D, Chen H-T, Gazumyan A, Yamane A, et al. 2010. PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science* **329**: 917–923.

Daugaard M, Baude A, Fugger K, Povlsen LK, Beck H, Sørensen CS, Petersen NHT, Sorensen PHB, Lukas C, Bartek J, et al. 2012. Ledgf (p75) promotes DNA end resection and homologous recombination. *Nat Struct Mol Biol* **19**: 803–810.

Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan W-I, Robson SC, Chung C-W, Hopf C, Savitski MM, et al. 2011. Inhibition of BET recruitment to chromatin as an effective treatment for Mll fusion leukaemia. *Nature* **478**: 529–533.

de Graaf P, Mousson F, Geverts B, Scheer E, Tora L, Houtsmuller AB, Timmers HTM. 2010. Chromatin interaction of

TATA-binding protein is dynamically regulated in human cells. *J Cell Sci* **123**: 2663–2671.

Dehé P-M, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, Rodríguez-Gil A, Mkandawire M, Landsberg K, Shevchenko A, et al. 2006. Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem* **281**: 35404–35412.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**: 277–293.

Demay MB, Kiernan MS, DeLuca HF, Kronenberg HM. 1992. Sequences in the human parathyroid hormone gene that bind the 1,25-dihydroxyvitamin D3 receptor and mediate transcriptional repression in response to 1,25-dihydroxyvitamin D3. *Proc Natl Acad Sci USA* **89**: 8097–8101.

Demény MA, Soutoglou E, Nagy Z, Scheer E, Jànoshàzi A, Richardot M, Argentini M, Kessler P, Tora L. 2007. Identification of a small Taf complex and its role in the assembly of Taf-containing complexes. *PLoS ONE* **2**: e316.

Deshpande AJ, Bradner J, Armstrong SA. 2012. Chromatin modifications as therapeutic targets in MLL-rearranged leukemia. *Trends Immunol* **33**: 563–570.

Dharmarajan V, Lee J-H, Patel A, Skalnik DG, Cosgrove MS. 2012. Structural basis for WDR5 interaction (Win) motif recognition in human SET1 family histone methyltransferases. *J Biol Chem* **287**: 27275–27289.

Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**: 1475–1489.

Dominguez C, Boelens R, Bonvin AMJJ. 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**: 1731–1737.

Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CAEM, Stevens TJ, Vranken WF, Vriend G, Vuister GW. 2012. CING: an integrated residue based structure validation program suite. *J Biomol NMR* **54**: 267–283.

Dorn ES, Cook JG. 2011. Nucleosomes in the neighborhood: new roles for chromatin modifications in replication origin control. *Epigenetics* **6**: 552–559.

Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, Allis CD, Roeder RG. 2006. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* **13**: 713–719.

Dreijerink KMA, Lips CJM, Timmers HTM. 2009a. Multiple endocrine neoplasia type 1: a chromatin writer's block. *J Intern Med* **266**: 53–59.

Dreijerink KMA, Mulder KW, Winkler GS, Höppener JWM, Lips CJM, Timmers HTM. 2006. Menin links estrogen receptor activation to histone H3K4 tri-methylation. *Cancer Res* **66**: 4929–4935.

Dreijerink KMA, Varier RA, van Beekum O, Jeninga EH, Höppener JWM, Lips CJM, Kummer JA, Kalkhoven E, Timmers HTM. 2009b. The multiple endocrine neoplasia type 1 (MEN1) tumor suppressor regulates peroxisome proliferator-activated receptor gamma dependent adipocyte differentiation. *Mol Cell Biol* **29**: 5060–5069.

Dreijerink KMA, Varier RA, van Nuland R, Broekhuizen R, Valk GD, van der Wal JE, Lips CJM, Kummer JA, Timmers HTM. 2009c. Regulation of vitamin D receptor function in MEN1 related parathyroid adenomas. *Mol Cell Endocrinol* **313**: 1–8.

Duan Q, Chen H, Costa M, Dai W. 2008. Phosphorylation of H3S10 blocks the access of H3K9 by specific antibodies and histone methyltransferase. Implication in regulating chromatin dynamics and epigenetic inheritance during mitosis. *J Biol Chem* **283**: 33585–33590.

Dyer PN, Edayathumangalam RS, White CL, Bao Y, Chakravarthy S, Muthurajan UM, Luger K. 2004. Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth Enzymol* **375**: 23–44.

Eidahl JO, Crowe BL, North JA, McKee CJ, Shkriabai N, Feng L, Plumb M, Graham RL, Gorelick RJ, Hess S, et al. 2013. Structural basis for high affinity binding of Ledgf PWWP to mononucleosomes. *Nucleic Acids Res*.

Eissenberg JC, Shilatifard A. 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev Biol* **339**: 240–249.

Elsaesser SJ, Goldberg AD, Allis CD. 2010. New functions for an old variant: no substitute for histone H3.3. *Curr Opin Genet Dev* **20**: 110–117.

Engelman A, Cherepanov P. 2008. The lentiviral integrase binding protein Ledgf/p75 and HIV-1 replication. *PLoS Pathog* **4**: e1000046.

Ernst P, Vakoc CR. 2012. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief Funct Genomics* **11**: 217–226.

Fang M, Xia F, Mahalingam M, Virbasius C-M, Wajapeyee N, Green MR. 2013a. MEN1 is a Melanoma Tumor Suppressor that Preserves Genomic Integrity by Stimulating Transcription of Genes that Promote Homologous Recombination-Directed DNA Repair. *Mol Cell Biol*.

Fang R, Chen F, Dong Z, Hu D, Barbera AJ, Clark EA, Fang J, Yang Y, Mei P, Rutenberg M, et al. 2013b. LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. *Mol Cell* **49**: 558–570.

Farley SM, Chen G, Guo S, Wang M, A J, Lee F, Lee F, Sawicki M. 2006. Menin localizes to chromatin through an ATR-CHK1 mediated pathway after UV-induced DNA damage. *J Surg Res* **133**: 29–37.

Fernandez A, Brautigan DL, Lamb NJ. 1992. Protein phosphatase type 1 in mammalian cell mitosis: chromosomal localization and involvement in mitotic exit. *J Cell Biol* **116**: 1421–1430.

Ferris AL, Wu X, Hughes CM, Stewart C, Smith SJ, Milne TA, Wang GG, Shun M-C, Allis CD, Engelman A, et al. 2010. Lens epithelium derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc Natl Acad Sci USA* **107**: 3135–3140.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**: 212–224.

Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert J-P, Barsyte-Lovejoy D, Felletar I, Volkmer R, Müller S, Pawson T, et al. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**: 214–231.

Fischle W, Wang Y, Allis CD. 2003. Histone and chromatin crosstalk. *Curr Opin Cell Biol* **15**: 172–183.

FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C. 2004. Clustering of DNA sequences in human promoters. *Genome Res* **14**: 1562–1574.

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, et al. 2005. Loss of acetylation at Lys16 and tri-methylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* **37**: 391–400.

Franklin DS, Godfrey VL, O'Brien DA, Deng C, Xiong Y. 2000. Functional collaboration between different cyclin dependent kinase inhibitors suppresses tumor growth with distinct tissue specificity. *Mol Cell Biol* **20**: 6147–6158.

Gangloff YG, Pointud JC, Thuault S, Carré L, Romier C, Muratoglu S, Brand M, Tora L, Couderc JL, Davidson I. 2001a. The TFIID components human Taf(II)140 and Drosophila BIP2 (Taf(II)155) are novel metazoan homologues of yeast Taf(II)47 containing a histone fold and a PHD finger. *Mol Cell Biol* **21**: 5109–5121.

Gangloff YG, Romier C, Thuault S, Werten S, Davidson I. 2001b. The histone fold is a key structural motif of transcription factor TFIID. *Trends in Biochemical Sciences* **26**: 250–257.

Gangloff YG, Sanders SL, Romier C, Kirschner D, Weil PA, Tora L, Davidson I. 2001c. Histone folds mediate selective heterodimerization of yeast Taf(II)25 with TFIID components yTaf(II)47 and yTaf(II)65 and with SAGA component ySPT7. *Mol Cell Biol* **21**: 1841–1853.

Ge H, Si Y, Roeder RG. 1998. Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *EMBO J* **17**: 6723–6729.

Gedlicka C, Hager G, Weissenböck M, Gedlicka W, Knerer B, Kornfehl J, Formanek M. 2006. 1,25(OH)2Vitamin D3 induces elevated expression of the cell cycle inhibitor p18 in a squamous cell carcinoma cell line of the head and neck. *J Oral Pathol Med* **35**: 472–478.

Geiger JH, Hahn S, Lee S, Sigler PB. 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**: 830–836.

Gijsbers R, Vets S, De Rijck J, Ocwieja KE, Ronen K, Malani N, Bushman FD, Debyser Z. 2011. Role of the PWWP domain of lens epithelium derived growth factor (Ledgf)/p75 cofactor in lentiviral integration targeting. *J Biol Chem* **286**: 41812–41825.

Gracanin A, Dreijerink KMA, van der Luijt RB, Lips CJM, Höppener JWM. 2009. Tissue selectivity in multiple endocrine

neoplasia type 1-associated tumorigenesis. *Cancer Res* **69**: 6371–6374.

Hahn S, Buratowski S, Sharp PA, Guarente L. 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proc Natl Acad Sci USA **86**: 5718–5722.

Halbach A, Zhang H, Wengi A, Jablonska Z, Gruber IML, Halbeisen RE, Dehé P-M, Kemmeren P, Holstege F, Géli V, et al. 2009. Cotranslational assembly of the yeast SET1C histone methyltransferase complex. *EMBO J* **28**: 2959–2970.

Hargreaves DC, Crabtree GR. 2011. ATP dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* **21**: 396–420.

Hathaway NA, Bell O, Hodges C, Miller EL, Neel DS, Crabtree GR. 2012. Dynamics and memory of heterochromatin in living cells. *Cell* **149**: 1447–1460.

Haven CJ, Howell VM, Eilers PHC, Dunne R, Takahashi M, van Puijenbroek M, Furge K, Kievit J, Tan M-H, Fleuren GJ, et al. 2004. Gene expression of parathyroid tumors: molecular subclassification and identification of the potential malignant phenotype. *Cancer Res* **64**: 7405–7411.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type specific gene expression. *Nature* **459**: 108–112.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318.

Hendrix J, Gijsbers R, De Rijck J, Voet A, Hotta J-I, McNeely M, Hofkens J, Debyser Z, Engelborghs Y. 2011. The transcriptional co-activator Ledgf/p75 displays a dynamic scan-and-lock mechanism for chromatin tethering. *Nucleic Acids Res* **39**: 1310–1325.

Heppner C, Bilimoria KY, Agarwal SK, Kester M, Whitty LJ, Guru SC, Chandrasekharappa SC, Collins FS, Spiegel AM, Marx SJ, et al. 2001. The tumor suppressor protein Menin interacts with NF-kappaB proteins and inhibits NF-kappaB mediated transactivation. *Oncogene* **20**: 4917–4925.

Heppner C, Kester MB, Agarwal SK, Debelenko LV, Emmert-Buck MR, Guru SC, Manickam P, Olufemi SE, Skarulis MC, Doppman JL, et al. 1997. Somatic mutation of the MEN1 gene in parathyroid tumours. *Nat Genet* **16**: 375–378.

Herrmann T, Güntert P, Wüthrich K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**: 209–227.

Herz H-M, Mohan M, Garruss AS, Liang K, Takahashi Y-H, Mickey K, Voets O, Verrijzer CP, Shilatifard A. 2012. Enhancer-associated H3K4 mono-methylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev*.

Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**: 2119–2137.

Huang J, Gurung B, Wan B, Matkar S, Veniaminova NA, Wan K, Merchant JL, Hua X, Lei M. 2012. The same pocket in Menin binds both MLL and JUND but has opposite effects on transcription. *Nature* **482**: 542–546.

Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, Hyman A, Mann M. 2010. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J Cell Biol* **189**: 739–754.

Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, Lee JC, Hayes DN, Shanmugam KS, Bhattacharjee A, Biondi CA, et al. 2004. Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Mol Cell* **13**: 587–597.

Hughes RM, Wiggins KR, Khorasanizadeh S, Waters ML. 2007. Recognition of tri-methyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc Natl Acad Sci USA* **104**: 11184–11188.

Hughes S, Jenkins V, Dar MJ, Engelman A, Cherepanov P. 2010. Transcriptional co-activator Ledgf interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity. *J Biol Chem* **285**: 541–554.

Izzo A, Kamieniarz K, Schneider R. 2008. The histone H1 family: specific members, specific functions? *Biol Chem* **389**: 333–343.

Jacobson RH, Ladurner AG, King DS, Tjian R. 2000. Structure and function of a human TafII250 double bromodomain module. *Science* **288**: 1422–1425.

Jencks WP. 1981. On the attribution and additivity of binding energies. *Proc Natl Acad Sci USA* **78**: 4046–4050.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* **293**: 1074–1080.

Jeon B-N, Choi W-I, Yu M-Y, Yoon A-R, Kim M-H, Yun C-O, Hur M-W. 2009. ZBTB2, a novel master regulator of the p53 pathway. *J Biol Chem* **284**: 17935–17946.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**: 161–172.

Jiang H, Shukla A, Wang X, Chen W-Y, Bernstein BE, Roeder RG. 2011. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* **144**: 513–525.

Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schulick RD, Tang LH, Wolfgang CL, Choti MA, et al. 2011. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331**: 1199–1203.

Jung YS, Zweckstetter M. 2004. Mars - robust automatic backbone assignment of proteins. *J Biomol NMR* **30**: 11–23.

Juven-Gershon T, Kadonaga JT. 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* **339**: 225–229.

Karaca E, Bonvin AMJJ. 2011. A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure* **19**: 555–565.

Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. 2012. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13**: 424.

Karnik SK, Hughes CM, Gu X, Rozenblatt-Rosen O, McLean GW, Xiong Y, Meyerson M, Kim SK. 2005. Menin regulates pancreatic islet growth by promoting histone methylation and expression of genes encoding p27Kip1 and p18INK4c. *Proc Natl Acad Sci USA* **102**: 14659–14664.

Kato H, van Ingen H, Zhou B-R, Feng H, Bustin M, Kay LE, Bai Y. 2011. Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl based NMR. *Proc Natl Acad Sci USA* **108**: 12283–12288.

Kaufmann J, Smale ST. 1994. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev* **8**: 821–829.

Kawagoe J, Li Q, Mussi P, Liao L, Lydon JP, DeMayo FJ, Xu J. 2012. Nuclear receptor coactivator-6 attenuates uterine estrogen sensitivity to permit embryo implantation. *Dev Cell* **23**: 858–865.

Kim D, Blus BJ, Chandra V, Huang P, Rastinejad F, Khorasanizadeh S. 2010. Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat Struct Mol Biol* **17**: 1027–1029.

Kim G-W, Yang X-J. 2011. Comprehensive lysine acetylomes emerging from bacteria to humans. *Trends in Biochemical Sciences* **36**: 211–220.

Kim H, Lee J-E, Cho E-J, Liu JO, Youn H-D. 2003. Menin, a tumor suppressor, represses JunD mediated transcriptional activity by association with an mSin3A-histone deacetylase complex. *Cancer Res* **63**: 6135–6139.

Kim J, Kim J-A, McGinty RK, Nguyen UTT, Muir TW, Allis CD, Roeder RG. 2013. The n-SET Domain of Set1 Regulates H2B Ubiquitylation dependent H3K4 Methylation. *Mol Cell* **49**: 1121–1133.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436**: 876–880.

Klose RJ, Zhang Y. 2007. Histone H3 Arg2 methylation provides alternative directions for COMPASS. *Nature structural & molecular biology*, November.

Kooistra SM, Helin K. 2012. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* **13**: 297–311.

Korkor AB. 1987. Reduced binding of [3H]1,25-dihydroxyvitamin D3 in the parathyroid glands of patients with renal failure. *N Engl J Med* **316**: 1573–1577.

Kornberg RD. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**: 868–871.

Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693–705.

Kulaeva OI, Nizovtseva EV, Polikanov YS, Ulianov SV, Studitsky VM. 2012. Distant activation of transcription:

mechanisms of enhancer action. *Mol Cell Biol* **32**: 4892–4897.

Laguri C, Duband-Goulet I, Friedrich N, Axt M, Belin P, Callebaut I, Gilquin B, Zinn-Justin S, Couprie J. 2008. Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. *Biochemistry* **47**: 6199–6207.

Larsson C, Skogseid B, Oberg K, Nakamura Y, Nordenskjöld M. 1988. Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. *Nature* **332**: 85–87.

Laskowski R, Rullmann JA, MacArthur M, Kaptein R, Thornton J. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**.

Lau PNI, Cheung P. 2011. Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proc Natl Acad Sci USA* **108**: 2801–2806.

Lauberth SM, Nakayama T, Wu X, Ferris AL, Tang Z, Hughes SH, Roeder RG. 2013. H3K4me3 Interactions with Taf3 Regulate Preinitiation Complex Assembly and Selective Gene Activation. *Cell* **152**: 1021–1036.

Lavigne AC, Mengus G, Gangloff YG, Wurtz JM, Davidson I. 1999. Human Taf(II)55 interacts with the vitamin D(3) and thyroid hormone receptors and with derivatives of the retinoid X receptor that have altered transactivation properties. *Mol Cell Biol* **19**: 5486–5494.

Lee J-H, Skalnik DG. 2008. Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol* **28**: 609–618.

Lee J-H, Tate CM, You J-S, Skalnik DG. 2007a. Identification and characterization of the human Set1b histone H3-Lys4 methyltransferase complex. *J Biol Chem* **282**: 13419–13428.

Lee J-H, You J, Dobrota E, Skalnik DG. 2010. Identification and characterization of a novel human PP1 phosphatase complex. *J Biol Chem* **285**: 24466–24476.

Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaumik SR, Shilatifard A. 2007b. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* **131**: 1084–1096.

Lee S, Lee D-K, Dou Y, Lee J, Lee B, Kwak E, Kong Y-Y, Lee S-K, Roeder RG, Lee JW. 2006. Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proc Natl Acad Sci USA* **103**: 15392–15397.

Lejonklou MH, Barbu A, Stålberg P, Skogseid B. 2012. Accelerated proliferation and differential global gene expression in pancreatic islets of five-week-old heterozygous Men1 mice: Men1 is a haploinsufficient suppressor. *Endocrinology* **153**: 2588–2598.

Lemos MC, Thakker RV. 2008. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat* **29**: 22–32.

Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* **13**: 233–245.

Leurent C, Sanders S, Ruhlmann C, Mallouh V, Weil PA, Kirschner DB, Tora L, Schultz P. 2002. Mapping histone fold Tafs within yeast TFIID. *EMBO J* **21**: 3424–3433.

Li B, Gogol M, Carey M, Lee D, Seidel C, Workman JL. 2007. Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* **316**: 1050–1054.

Li B, Pattenden SG, Lee D, Gutiérrez J, Chen J, Seidel C, Gerton J, Workman JL. 2005. Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc Natl Acad Sci USA* **102**: 18385–18390.

Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li G-M. 2013. The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutSα. *Cell* **153**: 590–600.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ. 2006b. Molecular basis for site specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**: 91–95.

Li Y, Trojer P, Xu C-F, Cheung P, Kuo A, Drury WJ, Qiao Q, Neubert TA, Xu R-M, Gozani O, et al. 2009. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem* **284**: 34283–34295.

Lindberg D, Akerström G, Westin G. 2008. Evaluation of CDKN2C/p18, CDKN1B/p27 and CDKN2B/p15 mRNA expression, and CpG methylation status in sporadic and MEN1 associated pancreatic endocrine tumours. *Clin Endocrinol (Oxf)* **68**: 271–277.

Liu W-L, Coleman RA, Grob P, King DS, Florens L, Washburn MP, Geles KG, Yang JL, Ramey V, Nogales E, et al. 2008. Structural changes in Taf4b-TFIID correlate with promoter selectivity. *Mol Cell* **29**: 81–91.

Llano M, Vanegas M, Hutchins N, Thompson D, Delgado S, Poeschla EM. 2006. Identification and characterization of the chromatin binding domains of the HIV-1 integrase interactor Ledgf/p75. *J Mol Biol* **360**: 760–773.

López-Méndez B, Güntert P. 2006. Automated protein structure determination from NMR spectra. *J Am Chem Soc* **128**: 13112–13122.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**: 251–260.

Lukasik SM, Cierpicki T, Borloz M, Grembecka J, Everett A, Bushweller JH. 2006. High resolution structure of the Hdgf PWWP domain: a potential DNA binding domain. *Protein Sci* **15**: 314–323.

Maertens GN, Cherepanov P, Engelman A. 2006. Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J Cell Sci* **119**: 2563–2571.

Manke IA, Lowery DM, Nguyen A, Yaffe MB. 2003. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* **302**: 636–639.

Margaritis T, Oreal V, Brabers N, Maestroni L, Vitaliano-Prunier A, Benschop JJ, van Hooff S, van Leenen D, Dargemont C, Géli V, et al. 2012. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet* **8**: e1002952.

Margueron R, Reinberg D. 2011. The Polycomb complex PRC2 and its mark in life. *Nature* **469**: 343–349.

Marschalek R. 2010. Mixed lineage leukemia: roles in human malignancies and potential therapy. *FEBS J* **277**: 1822–1831.

Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD. 2007. Role of Psip1/Ledgf/p75 in lentiviral infectivity and integration targeting. *PLoS ONE* **2**: e1340.

Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. 2003. The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in Biochemical Sciences* **28**: 69–74.

Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J, Ben Abdelali R, Macintyre E, De Braekeleer E, De Braekeleer M, et al. 2009. New insights to the MLL recombinome of acute leukemias. *Leukemia* **23**: 1490–1499.

Miedlich S, Lohmann T, Schneyer U, Lamesch P, Paschke R. 2001. Familial isolated primary hyperparathyroidism--a multiple endocrine neoplasia type 1 variant? *Eur J Endocrinol* **145**: 155–160.

Migliori V, Müller J, Phalke S, Low D, Bezzi M, Mok WC, Sahu SK, Gunaratne J, Capasso P, Bassi C, et al. 2012. Symmetric di-methylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol* **19**: 136–144.

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A. 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci USA* **98**: 12902–12907.

Milne TA, Hughes CM, Lloyd R, Yang Z, Rozenblatt-Rosen O, Dou Y, Schnepp RW, Krankel C, Livolsi VA, Gibbs D, et al. 2005. Menin and MLL cooperatively regulate expression of cyclin dependent kinase inhibitors. *Proc Natl Acad Sci USA* **102**: 749–754.

Min J, Feng Q, Li Z, Zhang Y, Xu R-M. 2003a. Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* **112**: 711–723.

Min J, Zhang Y, Xu R-M. 2003b. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**: 1823–1828.

Mishra BP, Ansari KI, Mandal SS. 2009. Dynamic association of MLL1, H3K4 tri-methylation with chromatin and Hox gene expression during the cell cycle. *FEBS J* **276**: 1629–1640.

Mo R, Rao SM, Zhu Y-J. 2006. Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. *J Biol Chem* **281**: 15714–15720.

Mohan M, Lin C, Guest E, Shilatifard A. 2010. Licensed to elongate: a molecular mechanism for MLL based leukaemogenesis. *Nat Rev Cancer* **10**: 721–728.

Mokry M, Hatzis P, de Bruijn E, Koster J, Versteeg R, Schuijers J, van de Wetering M, Guryev V, Clevers H, Cuppen E. 2010. Efficient double fragmentation ChIP sequencing provides nucleotide resolution protein-DNA binding profiles. *PLoS ONE* **5**: e15092.

Montagnoli A, Bosotti R, Villa F, Rialland M, Brotherton D, Mercurio C, Berthelsen J, Santocanale C. 2002. Drf1, a novel regulatory subunit for human Cdc7 kinase. *EMBO J* **21**: 3171–3181.

Morishita M, di Luccio E. 2011. Cancers and the NSD family of histone lysine methyltransferases. *Biochim Biophys Acta* **1816**: 158–163.

Muñoz IM, Rouse J. 2009. Control of histone methylation and genome stability by PTIP. *EMBO Rep* **10**: 239–245.

Murai MJ, Chruszcz M, Reddy G, Grembecka J, Cierpicki T. 2011. Crystal structure of Menin reveals binding site for mixed lineage leukemia (MLL) protein. *J Biol Chem* **286**: 31742–31748.

Musselman CA, Lalonde M-E, Côté J, Kutateladze TG. 2012. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**: 1218–1227.

Müller F, Demény MA, Tora L. 2007. New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *J Biol Chem* **282**: 14685–14689.

Nagy PL, Griesenbeck J, Kornberg RD, Cleary ML. 2002. A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. *Proc Natl Acad Sci USA* **99**: 90–94.

Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CAEM, Nabuurs SB, Güntert P, Livny M, Markley JL, Nilges M, et al. 2005. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**: 662–672.

Neumann H, Hancock SM, Buning R, Routh A, Chapman L, Somers J, Owen-Hughes T, van Noort J, Rhodes D, Chin JW. 2009. A method for genetically installing site specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol Cell* **36**: 153–163.

Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, Kouzarides T, Murzin AG, Murzina NV, Laue ED. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**: 103–107.

Nikolov M, Stützer A, Mosch K, Krasauskas A, Soeroes S, Stark H, Urlaub H, Fischle W. 2011. Chromatin affinity purification and quantitative mass spectrometry defining the interactome of histone modification patterns. *Mol Cell Proteomics* **10**: M110.005371.

Outchkourov NS, Muiño JM, Kaufmann K, van Ijcken WFJ, Koerkamp MJG, van Leenen D, de Graaf P, Holstege FCP, Grosveld FG, Timmers HTM. 2013. Balancing of Histone H3K4 Methylation States by the Kdm5c/SMCX Histone Demethylase Modulates Promoter and Enhancer Function. *Cell Rep*.

Patel A, Dharmarajan V, Cosgrove MS. 2008. Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem* **283**: 32158–32161.

Patel A, Dharmarajan V, Vought VE, Cosgrove MS. 2009. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **284**: 24242–24256.

Patel A, Vought VE, Dharmarajan V, Cosgrove MS. 2011. A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **286**: 3359–3369.

Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani OP, Zhao R, Kutateladze TG. 2006. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**: 100–103.

Petraki CD, Karavana VN, Skoufogiannis PT, Little SP, Howarth DJ, Yousef GM, Diamandis EP. 2001. The spectrum of human kallikrein 6 (zyme/protease M/neurosin) expression in human tissues as assessed by immunohistochemistry. *J Histochem Cytochem* **49**: 1431–1441.

Pieterman CRC, van Hulsteijn LT, Heijer den M, van der Luijt RB, Bonenkamp JJ, Hermus ARMM, Borel Rinkes IHM, Vriens MR, Valk GD, DutchMEN1 Study Group. 2012. Primary hyperparathyroidism in MEN1 patients: a cohort study with longterm follow-up on preferred surgical procedure and the relation with genotype. *Ann Surg* **255**: 1171–1178.

Pijnappel WW, Hendriks HF, Folkers GE, van den Brink CE, Dekker EJ, Edelenbosch C, van der Saag PT, Durston AJ. 1993. The retinoid ligand 4-oxo-retinoic acid is a highly active modulator of positional specification. *Nature* **366**: 340–344.

Pijnappel WWMP, Esch D, Baltissen MPA, Wu G, Mischerikow N, Bergsma AJ, van der Wal E, Han DW, Bruch HV, Moritz S, et al. 2013. A central role for TFIID in the pluripotent transcription circuitry. *Nature* **495**: 516–519.

Poncin J, Abs R, Velkeniers B, Bonduelle M, Abramowicz M, Legros JJ, Verloes A, Meurisse M, Van Gaal L, Verellen C, et al. 1999. Mutation analysis of the MEN1 gene in Belgian patients with multiple endocrine neoplasia type 1 and related diseases. *Hum Mutat* **13**: 54–60.

Porter IM, McClelland SE, Khoudoli GA, Hunter CJ, Andersen JS, McAinsh AD, Blow JJ, Swedlow JR. 2007. Bod1, a novel kinetochore protein required for chromosome biorientation. *J Cell Biol* **179**: 187–197.

Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. 2012. Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. *PLoS Genet* **8**: e1002717.

Qiu Y, Zhang W, Zhao C, Wang Y, Wang W, Zhang J, Zhang Z, Li G, Shi Y, Tu X, et al. 2012. Solution structure of the Pdp1 PWWP domain reveals its unique binding sites for methylated H4K20 and DNA. *Biochem J* **442**: 527–538.

Rees DC, Congreve M, Murray CW, Carr R. 2004. Fragment based lead discovery. *Nat Rev Mol Cell Biol* **3**: 660–672.

Rhee HS, Pugh BF. 2011. Comprehensive genome wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**: 1408–1419.

Rhee HS, Pugh BF. 2012. Genome wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**: 295–301.

Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF. 2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* **20**: 7137–7148.

Ruthenburg AJ, Allis CD, Wysocka J. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* **25**: 15–30.

Sanders SL, Garbett KA, Weil PA. 2002a. Molecular characterization of Saccharomyces cerevisiae TFIID. *Mol Cell Biol* **22**: 6000–6013.

Sanders SL, Jennings J, Canutescu A, Link AJ, Weil PA. 2002b. Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Mol Cell Biol* **22**: 4723–4738.

Saramäki A, Banwell CM, Campbell MJ, Carlberg C. 2006. Regulation of the human p21(waf1/cip1) gene promoter via multiple binding sites for p53 and the vitamin D3 receptor. *Nucleic Acids Res* **34**: 543–554.

Sarvan S, Avdic V, Tremblay V, Chaturvedi C-P, Zhang P, Lanouette S, Blais A, Brunzelle JS, Brand M, Couture J-F. 2011. Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain. *Nat Struct Mol Biol* **18**: 857–859.

Sawicka A, Seiser C. 2012. Histone H3 phosphorylation - a versatile chromatin modification for different occasions. *Biochimie* **94**: 2193–2201.

Schram AW, Baas R, Jansen PW, Riss A, Tora L, Vermeulen M, Timmers HT. 2013. A dual role for SAGA-associated factor 29 (SGF29) in ER stress survival by coordination of both histone H3 acetylation and histone H3 lysine-4 trimethylation. *PLoS ONE* **7**: e70035.

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.

Sedkov Y, Benes JJ, Berger JR, Riker KM, Tillib S, Jones RS, Mazo A. 1999. Molecular genetic analysis of the Drosophila trithorax-related gene which encodes a novel SET domain protein. *Mech Dev* **82**: 171–179.

Sedkov Y, Cho E, Petruk S, Cherbas L, Smith ST, Jones RS, Cherbas P, Canaani E, Jaynes JB, Mazo A. 2003. Methylation at lysine 4 of histone H3 in ecdysone dependent development of Drosophila. *Nature* **426**: 78–83.

Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK. 2005. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* **435**: 1262–1266.

Shen H-CJ, Rosen JE, Yang LM, Savage SA, Burns AL, Mateo CM, Agarwal SK, Chandrasekharappa SC, Spiegel AM,

Collins FS, et al. 2008. Parathyroid tumor development involves deregulation of homeobox genes. *Endocr Relat Cancer* **15**: 267–275.

Shen Y, Delaglio F, Cornilescu G, Bax A. 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**: 213–223.

Shi A, Murai MJ, He S, Lund G, Hartley T, Purohit T, Reddy G, Chruszcz M, Grembecka J, Cierpicki T. 2012. Structural insights into inhibition of the bivalent Menin-MLL interaction by small molecules in leukemia. *Blood* **120**: 4461–4469.

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Peña P, Lan F, Kaadige MR, et al. 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**: 96–99.

Shilatifard A. 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* **81**: 65–95.

Shimizu K, Bourillot PY, Nielsen SJ, Zorn AM, Gurdon JB. 2001. Swift is a novel BRCT domain coactivator of Smad2 in transforming growth factor beta signaling. *Mol Cell Biol* **21**: 3901–3912.

Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844–847.

Shun M-C, Botbol Y, Li X, Di Nunzio F, Daigle JE, Yan N, Lieberman J, Lavigne M, Engelman A. 2008. Identification and characterization of PWWP domain residues critical for Ledgf/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. *J Virol* **82**: 11555–11567.

Sierra J, Yoshida T, Joazeiro CA, Jones KA. 2006. The APC tumor suppressor counteracts beta-catenin activation and H3K4 methylation at Wnt target genes. *Genes Dev* **20**: 586–600.

Simon MD, Chu F, Racki LR, la Cruz de CC, Burlingame AL, Panning B, Narlikar GJ, Shokat KM. 2007. The site specific installation of methyl lysine analogs into recombinant histones. *Cell* **128**: 1003–1012.

Singh DP, Kubo E, Takamura Y, Shinohara T, Kumar A, Chylack LT, Fatma N. 2006. DNA binding domains and nuclear localization signal of Ledgf: contribution of two helix-turn-helix (HTH)-like domains and a stretch of 58 amino acids of the N-terminal to the trans-activation potential of Ledgf. *J Mol Biol* **355**: 379–394.

Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**: 204–220.

Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M. 2012. Stoichiometry of chromatin associated protein complexes revealed by label free quantitative mass spectrometry based proteomics. *Nucleic Acids Res*.

Smolle M, Workman JL. 2013. Transcription-associated histone modifications and cryptic transcription. *Biochim Biophys Acta* **1829**: 84–97.

Song J-J, Kingston RE. 2008. WDR5 interacts with mixed lineage leukemia (MLL) protein via the histone H3-binding pocket. *J Biol Chem* **283**: 35258–35264.

Sowa H, Kaji H, Kitazawa R, Kitazawa S, Tsukamoto T, Yano S, Tsukada T, Canaff L, Hendy GN, Sugimoto T, et al. 2004. Menin inactivation leads to loss of transforming growth factor beta inhibition of parathyroid cell proliferation and parathyroid hormone secretion. *Cancer Res* **64**: 2222–2228.

Sparmann A, van Lohuizen M. 2006. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**: 846–856.

Spedale G, Timmers HTM, Pijnappel WWMP. 2012. ATAC-king the complexity of SAGA during evolution. *Genes Dev* **26**: 527–541.

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F, et al. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**: 1146–1159.

Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A. 2006. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol* **13**: 852–854.

Stimpson KM, Sullivan BA. 2011. Histone H3K4 methylation keeps centromeres open for business. *EMBO J* **30**: 233–234.

Strik MC, Bladergroen BA, Wouters D, Kisiel W, Hooijberg JH, Verlaan AR, Hordijk PL, Schneider P, Hack CE, Kummer JA. 2002. Distribution of the human intracellular serpin protease inhibitor 8 in human tissues. *J Histochem Cytochem* **50**: 1443–1454.

Sudhaker Rao D, Han ZH, Phillips ER, Palnitkar S, Parfitt AM. 2000. Reduced vitamin D receptor expression in parathyroid adenomas: implications for pathogenesis. *Clin Endocrinol (Oxf)* **53**: 373–381.

Takahashi Y-H, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, Skiniotis G. 2011. Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci USA* **108**: 20526–20531.

Tan S, Hunziker Y, Sargent DF, Richmond TJ. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**: 127–151.

Tate CM, Lee J-H, Skalnik DG. 2010. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. *FEBS J* **277**: 210–223.

Thépot D, Weitzman JB, Barra J, Segretain D, Stinnakre MG, Babinet C, Yaniv M. 2000. Targeted disruption of the murine junD gene results in multiple defects in male reproductive function. *Development* **127**: 143–153.

Thiel AT, Blessington P, Zou T, Feather D, Wu X, Yan J, Zhang H, Liu Z, Ernst P, Koretzky GA, et al. 2010. MLL-AF9-induced leukemogenesis requires coexpression of the wild type Mll allele. *Cancer Cell* **17**: 148–159.

Thiel AT, Feng Z, Pant DK, Chodosh LA, Hua X. 2013. The trithorax protein partner Menin acts in tandem with EZH2 to suppress C/EBPα and differentiation in MLL-AF9 leukemia. *Haematologica* **98**: 918–927.

Thiel AT, Huang J, Lei M, Hua X. 2012. Menin as a hub controlling mixed lineage leukemia. *Bioassays* **34**: 771–780.

Thomas MC, Chiang C-M. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**: 105–178.

Thompson LH. 2012. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat Res* **751**: 158–246.

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, et al. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**: 1082–1086.

Tighe A, Staples O, Taylor S. 2008. Mps1 kinase activity restrains anaphase during an unperturbed mitosis and targets Mad2 to kinetochores. *J Cell Biol* **181**: 893–901.

Timmers HTM, Tora L. 2005. SAGA unveiled. *Trends in Biochemical Sciences* **30**: 7–10.

Trievel RC, Shilatifard A. 2009. WDR5, a complexed protein. *Nat Struct Mol Biol* **16**: 678–680.

Trydal T, Varhaug JE, Myking A, Aksnes L, Aakvaag A, Aarskog D. 1992. Serum vitamin D metabolites and calcitriol receptor concentration in parathyroid tissue in primary hyperparathyroidism. *Acta Endocrinol* **127**: 407–412.

Tsutsui KM, Tsutsui KM, Sano K, Sano K, Hosoya O, Hosoya O, Miyamoto T, Miyamoto T, Tsutsui K, Tsutsui K. 2011. Nuclear protein Ledgf/p75 recognizes supercoiled DNA by a novel DNA binding domain. *Nucleic Acids Res* **39**: 5067–5081.

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. 2003. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **125**: 10420–10428.

Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. 2006. A tripartite DNA binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of Ledgf/p75 with chromatin in vivo. *Nucleic Acids Res* **34**: 1653–1665.

van Ingen H, van Schaik FMA, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JAW, Liskamp RMJ, Timmers HTM, Boelens R. 2008. Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit Taf3. *Structure* **16**: 1245–1256.

van Nuland R, Smits AH, Pallaki P, Jansen PWTC, Vermeulen M, Timmers HTM. 2013a. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol Cell Biol*. **33**: 1067-1077.

van Nuland R, van Schaik FM, Simonis M, van Heesch S, Cuppen E, Boelens R, Timmers HM, van Ingen H. 2013b. Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the Psip1-PWWP domain. *Epigenetics Chromatin* **6**: 12.

Varier RA, Outchkourov NS, de Graaf P, van Schaik FMA, Ensing HJL, Wang F, Higgins JMG, Kops GJPL, Timmers HTM. 2010. A phospho/methyl switch at histone H3 regulates TFIID association with mitotic chromosomes. *EMBO J* **29**: 3967–3978.

Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, et al. 2010. Quantitative Interaction Proteomics and Genome wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **142**: 967–980.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WWMP, van Schaik FMA, Varier RA, Baltissen MPA, Stunnenberg HG, Mann M, Timmers HTM. 2007. Selective anchoring of TFIID to nucleosomes by tri-methylation of histone H3 lysine 4. *Cell* **131**: 58–69.

Vermeulen M, Timmers HTM. 2010. Grasping tri-methylation of histone H3 at lysine 4. *Epigenomics* **2**: 395–406.

Vezzoli A, Bonadies N, Allen MD, Freund SMV, Santiveri CM, Kvinlaug BT, Huntly BJP, Göttgens B, Bycroft M. 2010. Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat Struct Mol Biol* **17**: 617–619.

Vidal M. 2009. Role of polycomb proteins Ring1A and Ring1B in the epigenetic regulation of gene expression. *Int J Dev Biol* **53**: 355–370.

Vissers JH, Nicassio F, van Lohuizen M, Di Fiore PP, Citterio E. 2008. The many faces of ubiquitinated histone H2A: insights from the DUBs. *Cell Div* **3**: 8.

Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, et al. 2013. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **41**: D1063–9.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. 2013. Cancer genome landscapes. *Science* **339**: 1546–1558.

Voigt P, LeRoy G, Drury WJ, Zee BM, Son J, Beck DB, Young NL, Garcia BA, Reinberg D. 2012. Asymmetrically modified nucleosomes. *Cell* **151**: 181–193.

Wagner EJ, Carpenter PB. 2012. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* **13**: 115–126.

Wang F, Dai J, Daum JR, Niedzialkowska E, Banerjee B, Stukenberg PT, Gorbsky GJ, Higgins JMG. 2010a. Histone H3 Thr-3 phosphorylation by Haspin positions Aurora B at centromeres in mitosis. *Science* **330**: 231–235.

Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, et al. 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1 mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol* **29**: 6074–6085.

Wang T-T, Tavera-Mendoza LE, Laperriere D, Libby E, MacLeod NB, Nagai Y, Bourdeau V, Konstorum A, Lallemant B, Zhang R, et al. 2005. Large-scale in silico and microarray based identification of direct 1,25-dihydroxyvitamin D3 target genes. *Mol Endocrinol* **19**: 2685–2695.

Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, Patel DJ. 2010b. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC mediated repression. *Cell* **141**: 1183–1194.

Weake VM, Workman JL. 2008. Histone ubiquitination: triggering gene activity. *Mol Cell* **29**: 653–663.

Wiśniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat Methods* **6**: 359–362.

Workman JL, Roeder RG. 1987. Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* **51**: 613–622.

Wright KJ, Marr MT, Tjian R. 2006. Taf4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci USA* **103**: 12347–12352.

Wu H, Zeng H, Lam R, Tempel W, Amaya MF, Xu C, Dombrovski L, Qiu W, Wang Y, Min J. 2011. Structural and histone binding ability characterizations of human PWWP domains. *PLoS ONE* **6**: e18919.

Wu L, Lee SY, Zhou B, Nguyen UTT, Muir TW, Tan S, Dou Y. 2013. ASH2L Regulates Ubiquitylation Signaling to MLL: trans-Regulation of H3 K4 Methylation in Higher Eukaryotes. *Mol Cell* **49**: 1108–1120.

Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, Shilatifard A. 2008. Molecular regulation of H3K4 tri-methylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* **28**: 7337–7344.

Wyrick JJ, Parra MA. 2009. The role of histone H2A and H2B post-translational modifications in transcription: a

genomic perspective. *Biochim Biophys Acta* **1789**: 37–44.

Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W. 2003. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* **17**: 896–911.

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD. 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**: 859–872.

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, et al. 2006. A PHD finger of NURF couples histone H3 lysine 4 tri-methylation with chromatin remodelling. *Nature* **442**: 86–90.

Yokoyama A, Cleary ML. 2008. Menin critically links MLL proteins with Ledgf on cancer-associated target genes. *Cancer Cell* **14**: 36–46.

Yokoyama A, Somervaille TCP, Smith KS, Rozenblatt-Rosen O, Meyerson M, Cleary ML. 2005. The Menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**: 207–218.

Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, Kitabayashi I, Herr W, Cleary ML. 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with Menin to regulate Hox gene expression. *Mol Cell Biol* **24**: 5639–5649.

Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**: 505–508.

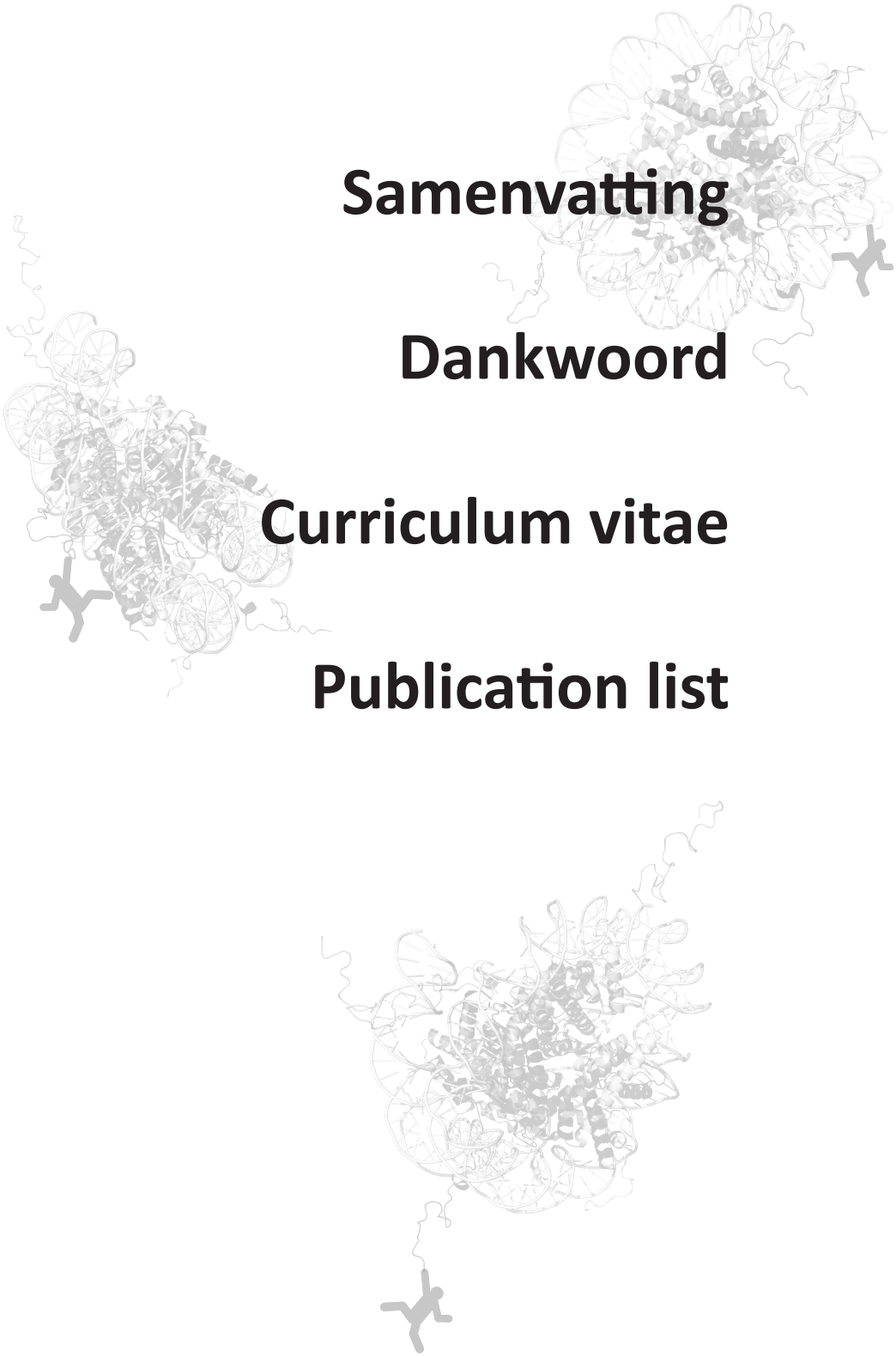
Yun M, Wu J, Workman JL, Li B. 2011. Readers of histone modifications. *Cell Res* **21**: 564–578.

Zeng W, Ball AR, Yokomori K. 2010. HP1: heterochromatin binding proteins working the genome. *Epigenetics* **5**: 287–292.

Zhang K, Lin W, Latham JA, Riefler GM, Schumacher JM, Chan C, Tatchell K, Hawke DH, Kobayashi R, Dent SYR. 2005. The Set1 methyltransferase opposes Ipl1 aurora kinase functions in chromosome segregation. *Cell* **122**: 723–734.

Zhou H-X. 2004. Polymer models of protein stability, folding, and interactions. *Biochemistry* **43**: 2141–2154.

Zwartjes CGM, Jayne S, van den Berg DLC, Timmers HTM. 2004. Repression of promoter activity by CNOT2, a subunit of the transcription regulatory Ccr4-not complex. *J Biol Chem* **279**: 10848–10854.



Samenvatting

Dankwoord

Curriculum vitae

Publication list

Nederlandse samenvatting voor niet-ingewijden

Een mens bestaat uit biljoenen individuele cellen. Bijna al deze cellen bevatten hetzelfde erfelijk materiaal (DNA), de ene helft komt van de moeder, de andere helft van de vader. Het DNA bestaat uit een lange code met een lengte van ongeveer twee meter en is onderverdeeld in aparte stukken (genen) die elk het bouwplan voor een onderdeel van de cel bevat (eiwit). Cellen binnen een organisme hebben uiteenlopende eigenschappen. Zo zorgen darmcellen voor de opname van voedingsstoffen en maakt het samentrekken van spiercellen bewegen mogelijk. Om deze functies goed uit te kunnen voeren bevatten de verschillende celtypen elk een ander pakket aan eiwitten. Dit betekent dus dat er in de ene cel andere genen gebruikt worden dan in de andere. Wanneer een cel te veel, te weinig, of bijvoorbeeld een gemuteerde vorm van een eiwit produceert kan dit leiden tot ziektes waaronder kanker.

Het aflezen van een gen uit de DNA-code gebeurt door een polymerase-eiwit. Dit eiwit maakt een tijdelijke kopie van het gen (mRNA). Dit proces wordt transcriptie genoemd. Vervolgens wordt dit mRNA gebruikt als sjabloon voor het eiwit waarvoor het gen codeert. Het wel of niet aflezen en kopiëren van een gen in mRNA is sterk gereguleerd. Ten eerste moet het polymerase herkennen waar het moet beginnen met het aflezen van het DNA. Als het te vroeg of te laat zou beginnen kan dit resulteren in foutieve eiwitten die bijvoorbeeld een ziekte zouden kunnen veroorzaken. De DNA-code voorafgaand aan elk gen bevat een stukje DNA dat het begin van het gen aangeeft (promoter). Ten tweede zorgen signalen van buiten de cel, zoals hormonen, ervoor dat DNA herkennende eiwitten (transcriptiefactoren) binden aan die promoter. Het binden van transcriptiefactoren aan het promoter-DNA veroorzaakt vervolgens een opeenvolging van moleculaire gebeurtenissen die uiteindelijk leiden tot transcriptie, een soort domino-effect. Elk van deze gebeurtenissen is gereguleerd om alles zorgvuldig te laten verlopen.

De twee meter DNA in de kern van een cel zit verpakt in eiwitpakketjes (histonen). Vergelijk het met een stuk draad (het DNA) om een spoel (de histoneiwitten). Dit samen wordt een nucleosoom genoemd (zie omslag). Het in meer of mindere mate ingepakt zijn van het DNA in nucleosomen zorgt voor variërende bereikbaarheid van het promoter-DNA voor transcriptiefactoren. Daarnaast kunnen nucleosomen ook informatie bevatten in de vorm van chemische modificaties die als een soort vlaggetjes aangeven wat er met de DNA-code gebeuren moet. Deze modificaties zorgen voor een extra laag in de regulatie van transcriptie. Het plaatsen van de modificaties gebeurt door enzymen (ook een soort eiwitten).

In dit proefschrift hebben wij gekeken naar een groep van eiwitten die de nucleosomen in de buurt van promoters markeert voor transcriptionele activatie (MLL-complexen). Het Menin-eiwit is onderdeel van zo een MLL-complex en is onder andere betrokken bij het plaatsen van histone modificaties op nucleosomen in de promoters van genen die actief moeten worden (aangeduid als H3K4me3). Het niet plaatsen van deze chemische vlaggetjes kan leiden tot een lagere productie van het eiwitproduct van dat specifieke gen. In patiënten met het Multiple

Endocrine Neoplasia type 1 (MEN1) syndroom mist het eiwit Menin door een genetische fout. Patiënten met een defect in het Menin-gen ontwikkelen tumoren in de bijschildklier, de alvleesklier en de hypofyse. In **hoofdstuk 2** van dit proefschrift laten we zien dat het Menin-eiwit bindt aan de vitamine D-receptor. De vitamine D-receptor is een transcriptiefactor die specifieke genen aan kan zetten. Vervolgens laten we zien dat Menin betrokken is bij het aanzetten van deze genen, waarschijnlijk door het markeren van de promoters met H3K4me3. Menin-verlies resulteert dus in het slecht functioneren van de vitamine D-receptor wat zou kunnen uitdraaien op tumorvorming in organen die belang hebben bij vitamine D-functies zoals de bijschildklier.

Het markeren van promoters doet Menin niet alleen. Het functioneert in het MLL-complex, een groepje sterk aan elkaar bindende eiwitten. Hoe sterk eiwitten aan elkaar binden geeft in zekere mate aan hoe belangrijk deze interacties zijn. In **hoofdstuk 3** hebben we gekeken naar Menin en de andere eiwitten die betrokken zijn bij het plaatsen van H3K4me3 op promoters. We hebben getracht op een kwantitatieve manier iets te zeggen over hoe vaak Menin nu eigenlijk met een ander eiwit interacteert. Vergelijk het met een sociaal netwerk. Wanneer een persoon vaak met een ander persoon omgaat is dit waarschijnlijk een belangrijke relatie. Als het slechts om een vage kennis gaat is dit waarschijnlijk minder belangrijk en zullen er ook minder vaak interacties zijn. In onze analyse zagen we dat Menin een grote vriendenkring heeft en dat de belangrijkste interacties plaatsvonden met de eiwitten betrokken bij het plaatsen van de H3K4me3-modificatie. Verder vonden we nieuwe eiwitten betrokken bij dit proces waarvan de functie in de toekomst verder onderzocht moet worden.

Een van de eiwitten die interacteert met Menin en ook een rol speelt bij het markeren van promoters is Psip1. In **hoofdstuk 4** hebben we gekeken naar de functie van dit eiwit. Psip1 speelt een belangrijke rol in HIV-infectie en bij de ontwikkeling van leukemie. We hebben laten zien dat Psip1 kan binden aan nucleosomen met een modificatie die gevonden wordt op genen die actief getranscribeerd worden (H3K36me3). Het gaat hierbij om een vergelijkbaar vlaggetje als H3K4me3, maar dan op een ander stuk van het histon H3-eiwit. Histonmodificaties worden herkend door eiwitten die beschikken over een modificatiebindend domein. Vergelijk het met een slot en een sleutel waarbij het domein als een slot past bij de juiste modificatie (de sleutel). Door middel van het maken van mutaties in het Psip1-eiwit laten we zien dat niet alleen de modificatie op het histon herkend wordt, maar ook het DNA dat gewikkeld zit om de histonen. Deze vinding verklaart waarom Psip1 eigenlijk maar heel zwak bindt aan het gemodificeerde histon maar veel beter als het verpakt zit in de vorm van een nucleosoom. Deze kennis is belangrijk met het oog op het ontwikkelen van remmers tegen HIV en leukemie die voorkomen dat Psip1 aan nucleosomen bindt.

Het herkennen van histonmodificaties is belangrijk om eiwitten betrokken bij transcriptie aan te trekken. De eiwitten van het TFIID-complex zijn essentieel voor het rekruteren van het polymerase-eiwit. Wij laten zien in **hoofdstuk 5** dat de modificaties op nucleosomen in samenspel met de DNA-code een belangrijke bijdrage leveren in het aantrekken van dit TFIID-

complex. Onze resultaten laten zien dat de samenstelling van de DNA-code en de verschillende combinaties van modificaties resulteert in een meer of minder sterke binding van TFIID aan de promoter. In dit hoofdstuk kijken we ook opnieuw naar de sterkte van relaties tussen eiwitten, in dit geval binnen het TFIID-complex. Dit laat zien dat er in TFIID waarschijnlijk een 'harde kern' van eiwitten is die mogelijk in sommige gevallen met de een, en in sommige gevallen met andere groepjes eiwitten optrekt. Onze proeven laten zien dat promotoren herkend kunnen worden op verschillende manieren en dat de herkenning van een combinatie van promotereigenschappen belangrijk is voor efficiënte transcriptie.

Samenvattend, in ons onderzoek hebben we laten zien dat het plaatsen van histonmodificaties belangrijk is voor het aanzetten van genen en mogelijke tumorvorming. Tevens laten we zien dat eiwitten betrokken bij het plaatsen van de modificaties in gevarieerde samenstellingen voorkomen. Bovendien hebben we uitgezocht dat niet slechts de histonmodificatie, maar ook de context (het DNA) betrokken is bij de binding van eiwitten in het geval van eiwitten zoals Psip1. Op een grotere schaal zien we dat het samenspel tussen verschillende modificaties en het DNA de binding van essentiële eiwitcomplexen kan beïnvloeden. Het begrijpen van de regulatie van transcriptie en de betrokken eiwitten is een essentieel onderdeel van het snappen hoe een cel zich gedraagt onder gezonde omstandigheden danwel bij ziektes zoals kanker. Gedetailleerde kennis van de moleculaire interacties die plaatsvinden binnen een cel zijn daarom belangrijk voor het ontwikkelen van nieuwe medicijnen en therapieën in de toekomst.

Dankwoord

Sommige eiwitten functioneren het best wanneer zij zich in een complex met anderen bevinden. Dit proefschrift had nooit tot stand kunnen komen zonder de vele interacties met anderen, soms direct en soms indirect, maar allemaal belangrijk op hun eigen unieke manier.

Om te beginnen: Marc, je hebt me vijf jaar geleden kunnen overtuigen om toch Amsterdam maar te verruilen voor Utrecht, iets waar ik geen spijt van heb. Als promotor, als mentor en vaak ook als technische vraagbaak heb ik enorm veel van je geleerd. We waren het niet altijd eens, maar altijd op een prettige en open manier. Je hebt me laten inzien dat wetenschap meer is dan een goede proef doen. Ik heb met verbazing gekeken naar dat glas dat bij jou altijd halfvol was en zal dit proberen mee te nemen in de toekomst. Het is enorm jammer dat je de laatste lootjes niet van dichtbij mee hebt kunnen maken, maar jouw opleiding heeft er mede voor gezorgd dat het uiteindelijk allemaal tot een goed einde komt.

Toch een beetje als baas(je) aan de zijlijn. Michiel, jouw kinderlijke enthousiasme voor wetenschap en je hebben zien ontwikkelen van postdoc tot baas zijn een inspiratie voor mij geweest. Dank voor alle adviezen, hulp met het afronden van dit proefschrift, meettijd en natuurlijk gezelligheid. Heel veel succes met het voortzetten van je groep in Nijmegen.

Hugo, wie had ooit kunnen voorspellen dat structuurbiologie nog eens leuk zou kunnen worden. Samen avondjes doortrekken om nucleosomen te bouwen of over elke zin in een manuscript te debatteren. Het was een superervaring om met jou te werken en ik hoop dat jouw talent behouden blijft voor de wetenschap.

Mijn paranimfen, Pascal, spilfiguur van de Vermeulen-groep en trouwe medeborrelaar. Bedankt voor alle wetenschappelijke en minder wetenschappelijke praat. Jouw inzet en verantwoordelijkheidsgevoel samen met de grappen en grollen maakten het superleuk binnen en buiten het lab. Jan, of het nu om fietsen, rennen, zwemmen of promoveren gaat, altijd een paar stappen op mij voor. Bedankt voor alle gezelligheid en vriendschap. Heren, bedankt dat jullie mijn paranimfen willen zijn.

Natuurlijk wil ik ook alle labgenoten bedanken. Dreé, zonder jou had ik hier niet eens beland. Het was heel fijn om samen alle ups en downs door te maken, gezellig biertjes te drinken en uiteindelijk toch nog samen aan een project te werken. Ik weet zeker dat je het op een nieuwe plek super gaat doen. Richard, bedankt voor alle hulp met biochemische experimenten en ondersteuning. Ik heb heel veel van je geleerd en zonder jouw inzet had dit boekje er heel anders uitgezien. Ik heb genoten van je humor en de samenwerking. Petra, bedankt voor alle hulp in het wegwijs worden op het lab, achter de microscoop en in de wetenschap. Ik hoop dat je snel met hernieuwd enthousiasme aan een nieuwe uitdaging kan beginnen. Het gaat je lukken! Hetty, het was een pittige tijd, maar ik ben blij dat je er weer bent. De nieuwe lichting, Maria, ik vind het heel knap hoe je in je eentje de gistexpertise in leven houdt. Het gaat goed

komen! Roy, projecten die anders lopen dan je van tevoren dacht horen er een beetje bij. Ik weet zeker dat met je nieuwe plannen alles op zijn pootjes terecht komt. Natuurlijk wil ik ook Betty, Sandy en Andrea bedanken voor de ondersteuning de afgelopen jaren. Alle oud-labgenootjes: Pim, Nikolay, GP, Marijke, Sjoerd, Folkert, Marcus en Radhika. Bedankt dat jullie mij op sleeptouw hebben genomen in de eerste jaren. Ik heb heel veel van jullie geleerd. Ik wil ook alle studenten bedanken met wie ik samen heb mogen werken. Charlotte, Erik en in het bijzonder Lina, bedankt voor jullie inzet. Soms vrees ik dat ikzelf meer geleerd heb van jullie, dan jullie van mij. Ik wens jullie het allerbeste in jullie carrière.

Natuurlijk wil ik ook alle andere mensen van de afdeling bedanken en dan in het bijzonder de Vermeulen-groep. Nelleke, je gestructureerdheid en inzet zijn bewonderenswaardig. Irem, Susan en Anneloes, bedankt voor de samenwerkingen en gezelligheid. Arne, het was gaaf om samen zo snel een project van de grond te tillen en af te ronden. Je ongeremde interesse en focus zijn motiverend voor iedereen in je omgeving. Uiteraard bedank ik natuurlijk alle huidige en vroegere collega's van de Holstege-, Bos-, Burgering-, en Kops-groep voor alle gezelligheid, sportieve en wetenschappelijke ondernemingen. En natuurlijk niet te vergeten, de niet-wetenschappelijke, maar net zo belangrijke, ondersteunende collega's op de afdeling, bedankt voor jullie hulp. Het maakte het leven op het lab er een stuk eenvoudiger op. Marieke en Sebastiaan, bedankt voor jullie hulp met het Psip1-project.

Leden van het kansloze AiO-clubje en "maar analist", Mottige, Talie, Assie, Dreé, Maaike en Pascal. Jullie hebben de laatste jaren tot een feest gemaakt. Bedankt voor de koffiemomentjes, borrels en de weekendjes weg. Echt heel erg jammer dat ik jullie feestjes ga missen, maar ik ga er van uit dat jullie er wel eentje extra voor mij drinken. Succes met jullie toekomst!

Ik wil ook de MEN1-groep bedanken. Koen, het is jammer dat we niet direct samen hebben kunnen werken, maar het was fijn om je altijd te kunnen raadplegen. Gerlof en Menno, de samenwerking was altijd erg prettig en ik denk dat we veel van elkaar hebben kunnen leren. Dit zal in de toekomst zeker zijn vruchten gaan afwerpen. Elfi, ik vind het bijzonder dapper van je om zo diep het MEN1-onderzoek in te durven duiken. Ik vond het erg leuk en leerzaam om samen over de meer klinisch georiënteerde projecten te praten en de eerste proeven te doen. Ik weet zeker dat je het uiteindelijk tot een mooi proefschrift gaat breien.

De leden van de beoordelings- en/of OiO-commissie: Wouter de Laat, Boudewijn Burgering, Inne Borel-Rinkes en Rolf Boelens, bedankt voor het lezen en beoordelen van dit proefschrift en/of jullie bijdrage aan de jaarlijkse evaluaties. Frank, ik ben je heel erg dankbaar voor het soepel inspringen tijdens het afronden van dit proefschrift en de hulp bij het aanvragen van de fellowships.

Het is soms lekker om even stoom af te blazen of om even totaal niet met wetenschap bezig te zijn. Aan een rots hangen, keihard op de pedalen of lekker eten en wijntjes drinken was voor mij een ideale uitlaatklep. Ik wil daarom alle vrienden en vriendinnen bedanken voor de steun en interesse gedurende de laatste vijf jaar.

Uiteraard wil ik ook mijn familie en schoonfamilie bedanken. Pap & Mam, het is fijn om een veilige thuishaven te hebben waar je altijd gesteund zal worden al is het onderzoek nog zo onbegrijpelijk. Bedankt daarvoor. Niels, Bart, Jorien en Alana, iedereen volgt zijn eigen weg en dat maakt jullie zulke mooie unieke mensen, bedankt voor de leuke momenten samen. Gerard, Corien, bedankt voor alle heerlijke vakanties, gezelligheid en de rustmomentjes!

Liefste Tes, ik mocht eigenlijk niets schrijven want: “wat heb ik er nou aan gedaan?”. Jouw steun, liefde en het mij gewoon even laten razen heeft me vaak door de minder leuke tijden gesleept. Het samen leuke dingen doen en genieten van het leven maakt alles er gewoon een stuk leuker op! Ik heb heel erg veel zin in de uitdagingen die ons te wachten staan en ben heel erg dankbaar en trots dat je deze samen met mij wilt aangaan.

Rick

Curriculum vitae

Rick van Nuland werd op 26 juni 1985 in Anna Paulowna geboren. In 2003 behaalde hij zijn VWO-diploma met het profiel Natuur & Gezondheid aan de Gemeenschappelijke Scholen Gemeenschap Schagen. In hetzelfde jaar begon hij aan zijn bacheloropleiding Biomedische Wetenschappen aan de Vrije Universiteit in Amsterdam die hij in 2006 afrondde. Vervolgens startte hij een masteropleiding in Biomolecular Sciences wederom aan de Vrije Universiteit Amsterdam. Tijdens deze masteropleiding heeft hij een onderzoeksstage gedaan aan de Universiteit van Stellenbosch in Zuid-Afrika onder begeleiding van Professor Jacky Snoep waar hij kinetische modellen voor gist heeft aangepast voor gebruik in het bestuderen van wijnfermentatie. Daarna heeft hij op het Nederlands Kanker Instituut in Amsterdam onder begeleiding van Sophia Bruggeman en Professor Maarten van Lohuizen gekeken naar de functie van het Ring1b polycomb-eiwit in stamcellen. Na het succesvol afronden van de masteropleiding is hij in september 2008 begonnen als Onderzoeker in Opleiding aan het Universitair Medisch Centrum Utrecht in het laboratorium van Professor Marc Timmers waarvan de resultaten zijn beschreven in dit proefschrift. Rick zal zijn carrière voortzetten als postdoctoraal onderzoeker in het lab van Professor Or Gozani op Stanford University, USA. Hij zal daar gaan werken aan epigenetische factoren die een rol spelen bij het opstarten van DNA-replicatie.

List of publications

van Nuland R*, Schram AW*, van Schaik FMA, Jansen PW, Vermeulen M, Timmers HT. Multivalent engagement of TFIID to nucleosomes. *PLoS One,* 2013 Sep 11;8(9):e73495 (*shared first author)

van Nuland R*, Smits AH*, Pallaki P, Jansen PW, Vermeulen M, Timmers HT. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol Cell Biol*. 2013 May;33(10):1067-77 (*shared first author)

van Nuland R, van Schaik FMA, Simonis M, van Heesch S, Cuppen E, Boelens, Timmers HT, van Ingen H. Nucleosomal DNA binding drives the recognition of H3K36 methylated nucleosomes by the Psip1-PWWP domain. *Epigenetics Chromatin.* 2013 May 8;6(1):12

Haanstra JR, Kerkhoven EJ, van Tuijl A, Blits M, Wurst M, **van Nuland R**, Albert MA, Michels PA, Bouwman J, Clayton C, Westerhoff HV, Bakker BM. A domino effect in drug action: from metabolic assault towards parasite differentiation. *Mol Microbiol.* 2011 Jan;79(1):94-108

Dreijerink KM, Varier RA, **van Nuland R**, Broekhuizen R, Valk GD, van der Wal JE, Lips CJ, Kummer JA, Timmers HT. Regulation of vitamin D receptor function in MEN1 related parathyroid adenomas. *Mol Cell Endocrinol.* 2009 Dec 10;313(1-2):1-8