Structure-Function analysis of a novel DNA-binding domain of the histone acetyltransferase MOZ/KAT6A

Jonas Schönfeld

Student-ID: 4342810

Thesis submitted to the University of Nottingham

for the degree of Master by Research

September 2019

Abstract

In this study, we focussed on the poorly characterised NEMM domain comprised within the N-terminal 200 amino acids of the Monocytic Leukemia Zinc Finger Protein (MOZ/KAT6A). MOZ is an important developmental regulator and involved in several diseases such as neurodevelopmental disorders and cancers. The NEMM domain of MOZ was recently found to be essential for induction of acute myeloid leukemia (AML) by the pathological fusion protein MOZ-TIF2. Structural predictions suggest that this region forms a potential tandem winged-helix domain (WHD) as found in a subfamily of DNA-binding proteins. Consistent with this, EMSA experiments indicated that the NEMM formed complexes with random dsDNA fragments *in vitro*. Mapping analysis showed that the first predicted winged-helix domain (pWHD1) preferentially binds dsDNA sequences containing CpG motifs, whereas the preference for pWHD2 was not defined. However, the data suggest that the tandem WHDs may form a composite DNA interaction-surface. Preliminary mutational analysis defined a lysine-rich stretch as necessary for DNA-recognition by pWHD1. Large-scale purification of MOZ pWHD1 free from contaminating DNA was also optimised, generating high yields of homogenous protein for protein crystallisation trials.

In addition, mutations that disable the DNA-binding activity of MOZ or the function of the DPF and MYST domains were generated by site-directed mutagenesis in YFP-tagged full length MOZ. These were assessed in rescue experiments using CRISPR-Cas9 edited HEK293 cells depleted for MOZ expression. Even though preliminary data suggests functional impairment of mutated MOZ, the method was limited by uneven expression levels of mutant and WT constructs. Further experiments are required to determine if there is instability of mutant proteins, or if expression can be normalised to WT MOZ in future studies.

In summary, this study has uncovered a potential tandem winged helix domain at the N-terminus of the MOZ acetyltransferase that has DNA-binding activity and a preference for CpG motifs. This constitutes a novel function of MOZ which has not been observed previously. Optimisation of protein purification and preliminary mutagenesis data will aid the full characterisation of this domain in future studies.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor David M Heery for his guidance during the project. I am grateful for his continuous motivation, expert advice and the excellent opportunities he provided me with to develop my scientific abilities. I can't thank him enough for the support and help I got throughout my year in the UK. A big thank you also goes to Dr Hilary Collins who was always there to assist me with technical guidance and helped me through difficulties both in and outside the lab.

Thanks to everyone in the Gene Regulation & RNA Biology group who were welcoming and supportive throughout my master's year. I would like to especially mention Chris, Aalam and Barbara for their advice, support and friendship.

Thank you to my parents for all their love, support and encouragement through everything I have accomplished. Finally, thank you Lioba for always being there for me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

µl	Microlitre
µm	Micrometer
5mC	5-methylcytosine
6xHis-tag	6x Histidin-tag
A	Alanine
AA	Amino acid
ac	Acetylation
Acetyl-CoA	Acetyl-CoenzymeA
ADARI	Adenosine Deaminase 1
AML	Acute Myeloid Leukemia
AML1	Acute Myeloid Leukemia Protein 1
APC/C	Anaphase-Promoting Complex/Cyclosome
ATP	adenosine triphosphate
B	β-strand
BRE	TFIIB recognition elements
BRPF1	Bromodomain PHD-Finger Protein 1
BSA	Bovine Serum Albumin
bu	butyrylation
Cas9	CRISPR associated protein 9
CBP	CREB Binding Protein
CGI	CpG islands
ChIP	Chromatin-Immunoprecipitation
CNC	Cranial Neural Crest Cells
Co-IP	Co-Immunoprecipitation
CPC	Caionic-Polar-Cationic clip motif
cr	Crotonylation
CREB	cAMP Response Element Binding Protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CV	Column Volumes
Cysteine	C
DBD	DNA-binding Domain
ddH2O	double distilled water
DGS	DiGeorge Syndrome
DMEM	Dumbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNA-PAGE	Deoxyribonucleic Acid Polyacrylamide Gel Electrophoresis
DNMT	DNA Methyltransferases
dNTP	Deoxynucleotide
DPE	Downstream Promoter Elements
DPF	Double PHD Fingers

dsDNA	double-stranded DNA
DTT	Dithiothreitol
E	Glutamate
EB	Elution Buffer
ECACC	European Collection of Cell Cultures
EDTA	Ethylendiaminetetraacetic acid
eGFP	Enhanced Green Fluorescent Protein
ELP3	Elongator Complex Protein 3
EMSA	Electrophoretic Mobility Shift Assay
ERα	Estrogen Receptor α
ESA1	Essential Sas2-related Acetyltransferase 1
ESC	Embryonic Stem Cell
eYFP	Enhanced Yellow Fluorescent Protein
FCS	Fetal Calf Serum
FOX	Forkhead Box Protein
Q	Glutamine
G	Glycine
GCN5	General Control Non-depressible 5
GDD	Global Developmental Delay Syndrome
gDNA	Genomic DNA
gnat	Gcn5 *N*-acetyltransferases
GTF	General Transcription Factor
h	Hour
H	α-helix
H1,2A 2B,3,4	Histone 1,2A,2B,3,4
HAT	Histone Acetyltransferase
HAT1	Histone Acetyltransferase 1
HBO1	HAT Bound to ORC1
HEK 293	Human Embryonic Kidney cells 293
HEPES	4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid
HF	High Fidelity
HMG	High-Mobility Group proteins
hMOF	Human Males-absent On The First
HOX	Homeobox protein
HPA2	HRP Associated 2
HSC	Hematopoietic Stem Cell
HTH	Helix-Turn-Helix motifs
I	Isoleucine
IMAC	Immobilized Metal Ion Affinity Chromatography
ING5	INhibitor of Growth 5
INR	Initiator motif
IP	Immunoprecipitation

IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Lysine acetyltransferase
KANSL1	KAT8 Regulatory NSL Complex Subunit
KAT	Lysine Acetyltransferase
KDM4C	Lysine-specific Demethylase 4C
L	Loop
L	Leucine
L	Litre
LB	Luria-Bertani medium
LXXLL	leucine-rich helix
M	Methionine
MAFK	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog K
MBD	Methyl-CpG-binding domains
me	Methylation
me	methylation
MEIS1	Myeloid Ecotropic Viral Integration Site 1
min	Minute
ml	Millilitre
MLL1	Mixed Lineage Leukemia 1
MORF	MOZ-Related Factor
MOZ	Monocytic Zinc Finger Protein
MPL	Myeloproliferative leukemia protein
MPO	Myeloperoxidase
MRD32	Mental Retardation Autosomal Dominant 32
mRNA	Messenger RNA
MTE	Motif Ten Element
MYST	MOZ, Ybf2/Sas3, Sas2, TIP60
NCOA2	Nuclear Receptor Coactivator 2 (TIF2)
NCOA3	Nuclear Receptor Coactivator3
NEB	New England Biolabds
NEMM	N-terminal part of Enok, MOZ and MORF
NF-kB	Nuclear Factor Kappa-light-chain-enhancer of activated B cells
Ni	Nickel
NLS	Nuclear Localisation Signal
nm	Nanometer
NMR	Nuclear Magnetic Resonance
O/N	Over Night
OD	Optical density
P	Proline
p300	adenoviral E1A-binding protein of 300 kDA
PBS	Phosphate buffered saline

PBST	PBS containing 0.1% Tween
PCAF	GCN5 p300/CBP Associated Factor
PcG	Polycomb Group
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PHD	Plant Homeodomain
Phyre2	Protein Homolog/analogy Recognition Engine V 2.0
PIC	Transcription Preinitiation Complex
PIP5K1A	Phosphatidylinositol 4-Phosphate-5 kinase 1A
PMSF	PhenylMethaneSulfonylFluoride
Pol II	RNA-polymerase II
pr	Propionylation
PRC	Polycomb Repressive Complexes
PRMT1	Protein Arginine N-methyltransferases 1
PTM	Posttranslational Modifications
PU.1	PU box-binding protein1
pWHD	Predicted Winged Helix Domain
Q	Glutamine
qPCR	Quantitavie polymerase chain reaction
R	Arginine
RA	Retinoic Acid
RING	Real Interesting New Gene
RIPA	RadioImmunoprecipitation Assay Buffer
RNA	Ribonucleic Acid
rpm	Rounds Per Minute
RPMI	Roswell Park Memorial Institute medium
RT	Room Temperature
RUNX1	Runt-related Transcription Factor 1
RUNX2	Runt-related Transcription Factor 2
S	Serine
s	Second
Sas3/Ybf2	Something About Silencing 3
SCF	Skp, Cullin, F-box containing complex
SDM	Site-directed mutagenesis
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SelB	Selenocysteine Elongation Factor B
Sem	Standard error of the mean
sgRNA	single-guide RNA
SMC5/6	Structure Maintenance of Chromosomes 5/6
SRC1	Steroid Receptor Coactivator
ssDNA	single-stranded DNA

SUMO	Small Ubiquitin-related Modifier
SWI/SNF	Mating-type Switching/Sucrose Fermentation
T	Threonine
T/L	Turn or Loop
TA	Transcriptional Activator
TAD	Topologically Associated Domains
Taq	Thermus Aquaticus
TBE buffer	Tris-borate-EDTA buffer
TBP	TATA-BOX-Binding Protein
TBX	T-box DNA-binding Transcription Factor
TE	Tris-EDTA
TEMED	N,N,N',N'-Tetramethyl ethylenediamine
TF	Transcription Factor
TFII	Transcription Factor II
TIF2	Transcriptional Intermediary Factor 2 (NCOA2)
Tip60	HIV Tat-interacting 60 kDa protein
TSS	Transcription Start Site
TSS	Transcription Start Site
UTR	Untranslated Region
UV-light	Ultraviolet light
V	Valine
VSD	Ventricular Septal Defects
W	Wing
WB	Western Blot
WDR5	WD Repeat domain 5
WHD	Winged Helix Domain
WT	Wild-type
Y	Tyrosine
YFM	eYFP-FLAG-MOZ
YHMT2	eYFP-HA-MOZ-TIF2

Chapter 1: Introduction

1.1 Transcriptional regulation

Both multicellular and unicellular organisms are defined by their genetic information within the genomic DNA (deoxyribonucleic acid), which encodes all RNAs (ribonucleic acids) and proteins essential for life. Regulated expression of genes is essential for the ability of cells to grow and proliferate, perform specialised functions, respond to exogenous changes or control metabolism. In the development of multicellular organisms, coordinated expression of unique transcriptomes is required to produce differentiated cells and tissues with specialised functions. The organisation of genes in eukaryotic cells is different from prokaryotes and involves protein coding sequences (exons) that are interrupted by non-coding introns. This enables expression of multiple products from a single gene through alternative splicing. Transcriptional regulation is highly complex and is controlled by a variety of mechanisms that involve epigenetic modifications, transcription, processing, transport, translation and degradation of RNA as well as post-translational protein modifications and protein turnover (**Figure 1.1**; Haberle and Stark, 2018). Various proteins are essential for correct gene expression. DNA is tightly packaged into nucleosomes and chromatin remodelling protein influence the structure and accessibility of DNA for the transcription machinery. Sequence specific DNA-binding transcription factors (TFs) direct transcriptional activation to specific promoters, where they recruit additional regulators such as co-factors and chromatin remodelling proteins. General transcription factors (GTFs) and RNA (ribonucleic acid) polymerase then assemble to start transcription (Vaquerizas et al., 2009). The fundamental mechanisms involved in transcriptional activation will be discussed in the following chapters.

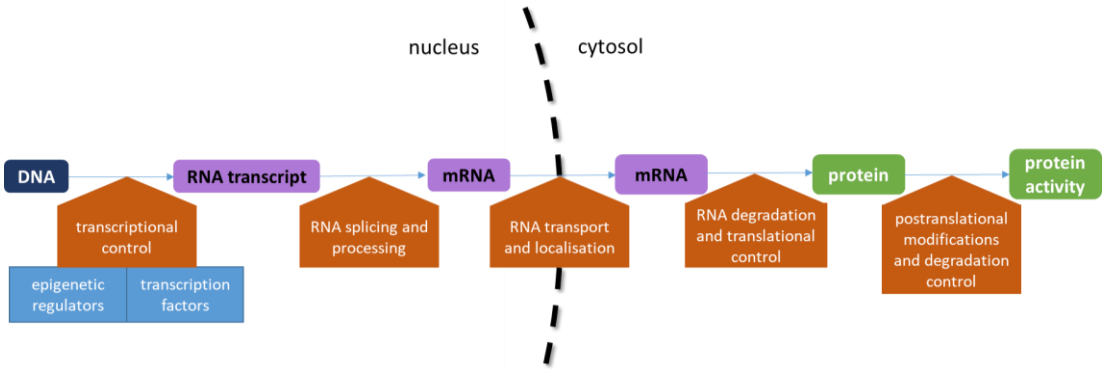


Figure 1.1 Schematic of regulatory steps of eukaryotic gene expression.

Gene expression is controlled by multiple steps that regulate DNA transcription, mRNA processing, transport, degradation and translation as well as protein activity. Figure adapted and modified after Alberts, 2015.

1.1.1 Eukaryotic gene promoters

RNA-polymerase II (Pol II) transcribes all protein-coding messenger RNAs (mRNA) as well as many non-coding RNAs such as microRNA and small nuclear RNA. Generally, transcription is initiated at the 5' end of a gene at the transcription start site (TSS). The region approximately 50 bp up-stream and 50 bp down-stream is referred to as the core promoter region where the transcription preinitiation complex (PIC) is assembled by GTFs. This region can contain several sequence motifs with fixed positions relative to the TSS. TATA-box motifs share the consensus sequence TATAWAWR (W = A/T; R = C/G) and are usually located around 30 bp upstream of the human TSS (Haberle and Stark, 2018). These sequences bind the TATA-box Binding Protein (TBP), a GTF that is a key component of the RNA Pol II holoenzyme and initiates the assembly of the PIC (Tora and Timmers, 2010). Binding of TBP bends the promoter DNA and gives directionality of transcription (Burley and Roeder, 1996). TATA-box sequences are widely conserved from yeast to humans. However, most promoters lack TATA-boxes and locate in CpG Islands (CGIs) (Sandelin et al., 2007). TATA-independent promoters still require TBP for PIC assembly, but they are more likely to lack directionality and can initiate transcription in both sense and antisense direction, which may have regulatory importance (Rhee and Pugh, 2012).

Another promoter element is the Initiator (INR) motif. The INR encompasses the TSS and is generally defined as a purine with a loose pyrimidine cap spanning the +1 TSS. INR motifs can accompany TATA-boxes or substitute them in TATA-less promoters. Downstream promoter elements (DPE) are often present in promoters lacking a TATA-box indicating that they could be associated with specific groups of genes. Other factors like the motif ten element (MTE) or Transcription Factor II B (TFIIB) recognition elements (BREs) are also regularly found in core promoters. (Zhovmer et al., 2010; Haberle and Stark, 2018).

In addition to defined sequence motifs, GC-rich sequences termed CGIs are often present in promoter regions of vertebrates. These sequences have a high density of CpG dinucleotides and are typically sites of transcription initiation. Half of the CGIs were found to contain TSSs and are located at promoters. The other half is referred to as orphan CGIs. Many orphan CGIs are located at transcriptional units and initiate transcription, but they often act in a tissue-specific manner. Many CGI promoters lack TATA-boxes but still recruit Pol II to initiate transcription with the help of other factors. Evidence suggests that CGI-chromatin is less stable compared to non-CGI chromatin which makes DNA in this region easier to access. CGI density was also found to correlate with increased H3K4 tri-methylation as well as H3 and H4 acetylation which also relate to transcriptionally active chromatin (see later). In addition, H3K36 di-methylation was depleted, a modification that inhibits transcriptional activation (Deaton and Bird, 2011).

Methylation of DNA via 5-methylcytosine (5mC) is an enzymatically imposed form of epigenetic regulation that is normally associated with gene repression or silencing. CGIs are hotspots for methylation and 5mC can be subsequently recognised by Methyl-CpG binding domains (MBDs) that assemble gene repression complexes to impose specific histone methylation marks or deacetylate histones. Furthermore, CpG motifs are often cognate binding sites and methylation can then block binding of proteins such as TFs (Mitton and Guzman, 2012). Gene silencing by methylation is mediated by DNA Methyltransferases (DNMTs). It plays an essential role in normal

development and induces stable gene suppression in several key processes such as X-chromosome silencing and silencing of repetitive element transcription as well as transposition (Jin et al., 2011). Furthermore, CGI methylation is relevant in several cancers where promoter regions of tumour suppressors get methylated.

The Polycomb-Repressive Complexes (PRC) 1 and 2 silence CGI promoters at target genes, such as *Homeobox* (*HOX*) gene clusters, by mediating H3K27 tri-methylation which was suggested to inhibit transcriptional elongation (Deaton and Bird, 2011) and also monoubiquitylate H2AK119 (Wang et al., 2004). Polycomb group (PcG) complexes have strong links to cancer as disruption of epigenetic processes can cause malignant cellular transformations. Several target genes of PcG complexes might also undergo hypermethylation under disease conditions which contributes to cancer development (Jin et al., 2011). Recently it was shown that PRC2, in addition to its epigenetic reader domains, possesses a DNA-binding winged helix domain (WHD, Chapter 1.1.3) that specifically recognizes unmethylated CpG motifs (Li et al., 2017).

In addition to control at the promoter regions, many *cis*-regulatory sequences are spread over long stretches of DNA. Enhancers can activate gene transcription over long genomic distances which requires chromatin folding to bridge the interaction with promoter regions. Furthermore, chromatin-remodelling is required to increase the accessibility of DNA for the transcription machinery (Chapter 1.1.1 and 1.1.5). It is known, that enhancers are activated by binding of TFs which then recruit chromatin-modifying enzymes and additional coactivator proteins to facilitate initiation of transcription. Large mediator complexes are then required to regulate interactions between activator proteins, GTFS and Pol II (**Figure 1.2**; Robson et al., 2019). The GTFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIG) assemble with Pol II in the PIC and recruit Pol II to the promoter. TBP of the TFIID complex binds to the TATA-box of the core-promoter and recruits additional TBP-associated factors. In addition, TBP is also required for PIC assembly at TATA-less promoters. TFIID also selectively binds H3K4me3 and enables recognition of the chromatin state by PIC. The interaction with DNA is stabilised by TFIIA and TFIIB which bind DNA as well as TBP.

TFIIB binds to BRE (TFIIB recognition elements) in close proximity up- or downstream of the TATA box and also interacts with Pol II at the catalytic region and was suggested to have a regulatory role. TFIIF binds Pol II and was suggested to recruit Pol II to the PIC. It enhances elongation and supports Pol II escape from the promoter. TFIIE supports binding of TFIIG to the complex and stimulates ATPase activity of TFIIG, which acts as a helicase and melts the DNA-strands. This enables Pol II to start transcription which facilitates phosphorylation of the C-terminal domain. Pol II is then released from the core promoter and binding GTFs and proceeds to elongate the mRNA transcript. This mechanism was extensively studied for TATA-dependent promoters but many other promoters containing INR in addition or as substitution for TATA-boxes are expected to be recognized in different mechanisms, maybe by other members of the TFIID complex (Tora and Timmers, 2010; Zhovmer et al., 2010, Vannini and Cramer, 2012; Rhee and Pugh, 2012; Danino et al., 2015).

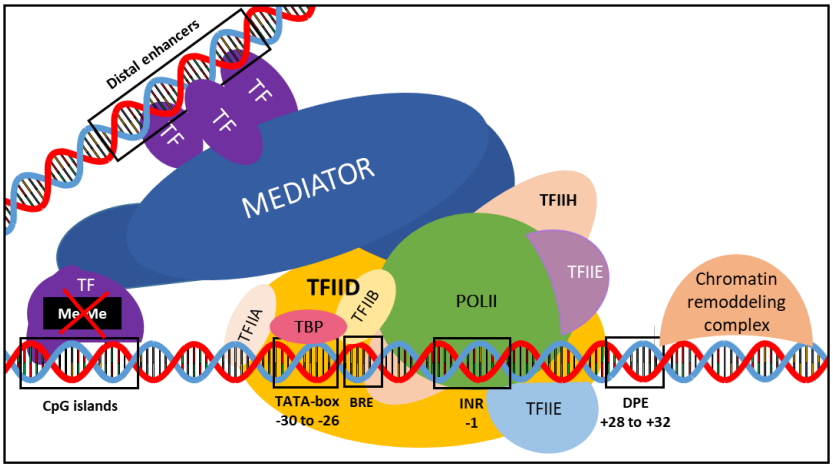


Figure 1.2 Schematic of transcription initiation

TFs, mediators, chromatin remodelling complexes, GTFs and POL II assemble to form complexes at the unmethylated CGIs, the TATA-box, BRE and INR close to the TSS as well as *cis*-regulatory enhancer regions to initiate transcription . Figure adapted and modified after Danino et al., 2015.

1.1.2 DNA-binding transcription Factors

In addition to GTFs (Chapter 1.1.1), other TFs regulate activity of individual genes. TFs are defined as proteins capable of binding DNA in a sequence specific manner and regulating transcription. The respective *cis*-regulatory sequences may be located at promoter or enhancer regions (Fulton et al., 2009; Lambert et al., 2018). TFs often function as homo- or heterodimers recognising palindromes or repeat sequence motifs. This permits cooperative interaction with DNA and high selectivity for target sequences with some TFs showing up to 1,000-fold preferences for their cognate DNA sequence. Knowledge of the recognized sequence is essential to identify potential binding sites and to determine the function of TFs. Accordingly, TFs are mainly characterised by their DNA-binding domains (DBDs). However, even homologous DBDs do not necessarily infer similarities in recognized sequences or functionality (Lambert et al., 2018). Sequence analysis of the human genome predicted that three major types of DBDs are characteristic for over 80% of TFs: C_2H_2 zinc fingers (675 TFs), homeodomains (257 TFs) and helix-loop-helix motifs (87 TFs) (Vaquerizas et al., 2009). TFs cooperate to aid specificity in DNA-binding and effector functionality. Protein-protein interactions can stabilise DNA-binding by providing multiple contacts to DNA and may also impact sequence preferences. Protein-protein interaction can also be mediated by TFs binding at DNA sequences in close proximity. In other cases, binding of one TF can modulate the shape of DNA which promotes binding of other TFs (Lambert et al., 2018).

Traditionally, TFs were classified as either repressors or activators. For example, some TFs act by steric hindrance to block binding of activator proteins. However, the majority of eukaryotic TFs act by recruitment of co-factors, which are components of co-repressor or co-activator complexes. These multi-subunit protein complexes provide diverse protein docking domains, allowing them to contact multiple TFs bound to promoters/enhancers simultaneously. The formed complexes facilitate recognition of specific DNA sequences as well as modifications of DNA and histones. They also act in chromatin remodelling which impacts availability of DNA

for the transcription machinery (Lambert et al., 2018). An example is the recruitment of histone acetyltransferases such as General Control Non-depressible 5 (GCN5) and adenoviral E1A-binding protein of 300 kDA (p300)/Creb Binding Protein (CBP) by multiple TFs (Lambert et al., 2018). These enzymes acetylate lysine sidechains present in histone tails or may also acetylate TFs themselves to regulate their function (Heery and Fisher, 2007). Changes on chromatin state and DNA-packaging then promote further recruitment of the mating-type switching/sucrose fermentation (SWI/SNF) complex which remodels nucleosomes to enable Pol II binding and initiation of transcription (Lambert et al., 2018).

Thus, DNA-binding TFs often act as pioneer factors that recruit co-activator, co-repressor and remodelling complexes to promoter and enhancer regions to regulate gene expression. Their presence in multi-subunit protein complexes with varying partners suggests that the impact of TFs on transcriptional activation can be very diverse and is strongly dependent on recruitment of co-factors as well as the cellular context (Lambert et al., 2018).

1.1.3 Winged-helix domains

The Winged-helix domains (WHDs), also known as forkhead domains, are a subtype of Helix-Turn-Helix motifs (HTH) which are widespread DNA-binding domains mainly known from TFs. WHDs were first discovered in the TF FOrkhead boX protein (FOX) A3, also known as hepatocyte factor 3γ, a member of the FOX family of TFs. In general, WHDs are characterised by three α-helices (H), a turn/loop (T/L), three β-strands (B) and two characteristic loops, “wings” (W/L), that are organised as H1-B1-H2-T/L-H3-B2-W1/L-B3-W2/L. Some proteins contain additional elements and/or lack the W2 loop (Harami et al., 2013).

Studies on various WHDs in complex with DNA have shown that sequence-specific recognition of double-stranded (ds) DNA is achieved by interaction of the helix H3 with the mayor groove of DNA. This mechanism is conserved in the FOX family and various other transcription factors. In addition to this interaction, some

proteins also recognize the minor groove of DNA via the wings W1 and/or W2 (Harami et al., 2013; Lalmansingh et al., 2012). Even though most canonical winged-helix proteins use similar mechanisms involving the helix H3 and wings to bind DNA, they can discriminate between very specific DNA sequences. Residues adjacent to helix H3 were suggested to play an important role in modulating sequence specificity (Gajiwala and Burley, 2000). Other proteins differ from this interaction profile and dsDNA recognition is mediated by the wing W1 binding the major groove of DNA while helix H3 interacts with the minor groove. Other secondary structure components can also be used to interact with the phosphate DNA backbone (Boura et al., 2007). The individual residues participating in DNA-binding can widely vary. Most protein-DNA interactions are mediated by polar sidechains that either directly contact DNA or use bridging water molecules (Gajiwala and Burley, 2000). DNA-binding WHDs can also develop very specific functional adaptations (Harami et al., 2013).

Versatile roles of WHDs have been discovered in transcriptional complexes. Subunits of RNA polymerases I, II and III complexes have conserved WHDs with high flexibility suggesting a role in establishment of stable DNA-binding during transcription initiation. WHDs can also be involved in DNA-strand separation or ligation as well as DNA-opening during transcription (Teichmann et al., 2012). A tandem WHD is present in the Pol I subcomplex A49-34.5, but only the second WHD was suggested to interact with DNA via positively charged surface residues (Geiger et al., 2010). The WHDs of two FOXP proteins were suggested to form “domain-swapped dimers”. The helices H1 and H3 of one monomer fold onto the helix H3 of the second monomer which is resembled the other way around. Each helix H3 then interacts with separate DNA-molecules to bridge interactions (Lalmansingh et al., 2012).

In addition to their function as classic TFs, FOXA proteins also function as pioneer factors. FOXA1 is able to bind DNA through its WHD while the C-terminal transcriptional activator (TA) domain binds core histones H3 and H4. The TA-domain

interrupts internucleosomal interactions and opens up nucleosomal arrays while the DBD is required to target specific genes. This was shown in *in vitro* nucleosome mapping studies in the absence of ATP-dependent chromatin remodelling enzymes which indicates that FOXA1 has an intrinsic chromatin remodelling function (Cirillo et al., 2002). To allow pioneer function and remodel chromatin, FOXA has to compete with linker histones. The WHD of FOXA1 is similar to the globular domains of linker histones H1 and 5 but lacks residues that are required for nucleosome array compaction (Clark et al., 1993; Lalmansingh et al., 2012). Therefore, competition for binding sites and displacement of linker histones could reduce compaction of chromatin and open up DNA. Comparative studies for the FOXO family suggest similar functions in chromosomal remodelling (Lalmansingh et al., 2012).

Linker histones H1 and H5 are composed of a central globular domain flanked by N- and C-terminal tails that are unstructured in the absence of nucleosomes. Linker histones are not located in the nucleosome core but play an important role in the formation of the higher chromatin structure. The globular domain contains a WHD that independently binds to nucleosomes *in vitro* and is assumed to have general preference for AT-rich DNA sequences (Lyubitelev et al., 2016; Cui and Zhurkin, 2009). WHDs from FOX proteins show structural conservation with the DBD of linker histones H1 and H5, although the sequence similarity is weak (Clark et al., 1993). However, the previously described WHD-TFs possess dipole moments and only have one DNA-binding site supported by additional weaker contacts. In contrast, the entire surface of linker histones is positively charged and has at least two DNA-binding sites. The geometry of the chromatin influences the DNA-binding mode and different linker histone variants may contact DNA in different manners. Several studies highlight different DNA-binding mechanisms for linker histones. It is generally accepted that helix H3 of the globular domain directly contacts nucleosomal DNA. It was considered that linker DNA is contacted by residues in H2 and H3, the loop L1 between helices H1 and H2 as well as the C-terminal β-strands and W1 and W2 loops (Lyubitelev et al., 2016). One model based on point mutation analysis suggests that

interactions with linker DNA are formed by residues of the loop L1 between helices H1 and H2 as well the C-terminal β-strands while H3 binds nucleosomal DNA (Zhou et al., 2013). Another model based on X-ray analysis proposed that the β-strand and helix H2 bind to nucleosomal DNA while helix H3 and the loop L1 between helices H1 and H2 contact linker DNA (Zhou et al., 2015). A similar prediction was also made in a computer-based model with only small differences in the interacting residues (Cui and Zhurkin, 2009). In addition to the globular core region, the C-terminal tail also provides positively charged residues and forms a secondary structure upon DNA-binding that might support binding-stability after initial DNA-recognition by the globular domain (Lyubitelev et al., 2016).

In addition to their versatility in DNA-binding, WHDs were also shown to be involved in RNA-binding. The Selenocysteine elongation factor B (SelB) contains 4 WHDs and recognizes mRNA in a sequence-independent manner. The recognition is very specific for a hairpin-tip structure and is mediated by WHD3 and 4 in a unique binding-mechanism (Selmer and Su, 2002). Furthermore, it was suggested that exposed patches of hydrophobic residues could form protein-protein interactions which is an essential function in many TFs. Protein-protein interaction mediates recruitment of co-factors and stabilises DNA-binding. In protein complexes such as the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex (Zheng et al., 2002) and the Anaphase-Promoting Complex/Cyclosome (APC/C) (Alfieri et al., 2016; Zhang et al., 2016), protein-protein interaction can be established by WHDs.

The mentioned examples show that WHDs are extremely versatile structures that are involved in many mechanisms ranging from recognition of diverse DNA and RNA sequences to protein-protein interactions.

1.1.4 Histones and their post-translational modifications

Histones are essential for the packaging of eukaryotic DNA within the cell nucleus. The human cell contains approximately two meters of diploid DNA whereas the nucleus has a diameter of only six μm which requires DNA to be compacted with the help of histones and none-histone proteins forming the chromatin structure. The first packaging units are core nucleosome particles, which contain 147 bp of DNA wrapped in two folds around an octamer of core histones (**Figure 1.3**). The histone octamers are formed by two molecules each of histone H2A, H2B, H3 and H4. Repeating units of nucleosomes are separated by 10-60 bp of linker DNA and form an array resembling "beads-on-a-string" with a diameter of around 10 nm. Linker histones H1/5 arrange the further condensation in thicker ~30 nm fibres by internucleosomal interactions. Further compaction steps *in vivo* can form 100-400 nm thick fibres of interphase and metaphase chromosomes (Peterson and Laniel, 2004).

The core histones are generally synthesised at higher levels during S-phase as they are essential for DNA packaging after DNA replication. They interact via a "hand-shake motif". Two H3/H4 dimers form a tetramer that associates with two H2A/H2B dimers in the presence of DNA. Multiple hydrophobic, hydrogen and electrostatic bonds allow the precise formation of the nucleosome complex. These interactions are facilitated by a central domain in the C-terminal part of the globular histone structure. Furthermore, each histone has an N-terminal "tail" consisting of 20-35 amino acids that is rich in basic residues such as lysine and protrude from the nucleosome surface. These tails are essential for the folding of nucleosomal arrays and impact the overall chromatin structure (Marmorstein and Zhou, 2014).

A variety of proteins are required to regulate chromatin. These include proteins that recognize DNA and histones, histone chaperones, pioneer TFs, chromatin-remodelling proteins as well as posttranslational modification enzymes that add various chemical groups to histones and the DNA (Marmorstein and Zhou, 2014; Haberle and Stark, 2018). Histone tails are subject to different post-

translational modifications, including acetylation (ac) and methylation (me) of lysine (K) and arginine (R), phosphorylation of serine (S) and threonine (T) as well as ubiquitylation and sumoylation of lysine residues (examples shown in **Figure 1.3**). In addition, lysine residues can be mono-, di-, or trimethylated and arginine residues can accept one or two methyl groups. Recently, mass spectrometry-based studies identified novel non-acetyl histone acylations such as crotonylation (cr), butyrylation (bu) and propionylation (pr) (Huang et al., 2014; Huang et al., 2015; Xiong et al., 2016). The combination of different histone marks critically affects the condensation status of chromatin and correlates with particular biological functions, such as transcriptional activity, DNA-repair, DNA-replication, recombination and X-chromosomal inactivation. For example, tri-methylation of H3K4 and acetylation of H3K27 correlate with transcriptional activation (Goll and Bestor, 2002; Peterson and Laniel, 2004; Haberle and Stark, 2018).

Various proteins like acetyltransferases and methyltransferases mediate these modifications (“writers”). Other enzymes like deacetylases and phosphatases act as “erasers” and can remove histone modifications. Several protein domains like PHD fingers, bromodomains and chromodomains (”readers”) recognize specific modifications and guide proteins to target sites (Glatt et al., 2011; Bannister and Kouzarides, 2011; Brent and Marmorstein, 2014).

Other than histone modifications, DNA can also be directly methylated as discussed in Chapter 1.1.1.

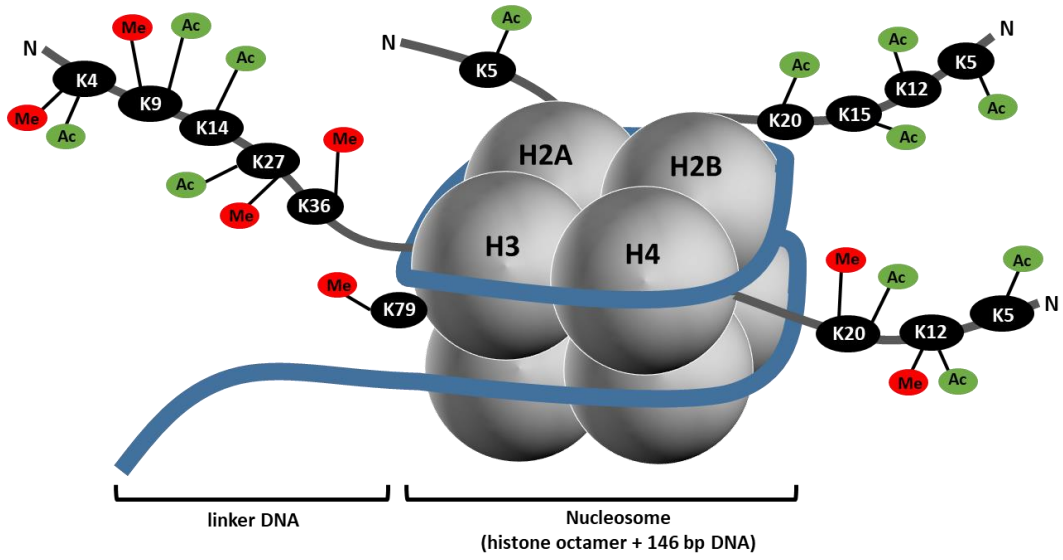


Figure 1.3 Schematic of the assembly of the nucleosome and histone modifications

Two of each core histones H2A, H2B, H3 and H4 assemble to octamers forming nucleosomes that are connected by linker DNA. A selection of regularly methylated (Me) and acetylated (Ac) histone lysine residues (K) as reported by the literature are indicated (modified after Sheikh and Akhtar, 2018; Lawrence et al., 2016).

1.1.5 Lysine acetyltransferases

Previously, proteins that acetylate lysine residues were referred to as Histone acetyltransferases (HATs). HATs are defined as enzymes that acetylate core histones and were first mentioned in 1964 (Pogo et al., 1966). However, because many of these enzymes also target none-histone proteins, a more general nomenclature was recently proposed referring to them as Lysine acetyltransferases (KATs; Allis et al., 2007). KATs form a diverse enzyme superfamily that is evolutionary conserved across eukaryotes from yeast to humans (Allis et al., 2007). They are mainly found in the nucleus, where they target histones as well as a variety of TFs and are essential for the regulation of gene expression. However, TFs are also located outside of the nucleus where they regulate protein activity in processes like metabolism and cell migration and are furthermore involved in different diseases (Lee and Workman, 2007; Close et al., 2010).

KATs typically act in multi-subunit protein complexes with a variety of different enzymes. This diversity contributes to the unique function of KAT complexes and their impact on histone marks (Marmorstein and Zhou, 2014; Heery and Fisher, 2007).

So far, 37 human KATs were identified and the histone acetylating members were divided in three main groups based on their catalytic domains: The Gcn5 *N*-acetyltransferases (GNAT) family including proteins such as GCN5 p300/CBP Associated Factor (PCAF) and ELongator complex protein 3 (ELP3); the MYST family which will be described in more detail below and adenoviral E1A-binding protein of 300 kDA (p300) and Creb Binding Protein (CBP) which do not share a true consensus KAT domain (Lee and Workman, 2007; Sheikh and Akhtar, 2018).

In general, KATs transfer an acetyl group from acetyl-CoenzymeA (CoA) to the ε-acetyl-L-lysine residues. In line with their structural differences, the acetylation domains of different families utilize different catalytic mechanisms (Marmorstein and Zhou, 2014). Members of the MYST family use a ping-pong like mechanism involving multiple sites and transfer steps while the GNAT family utilizes a sequential ordered mechanism with a direct transfer of the acetyl group from Acetyl-CoA to the substrate lysine residues (Heery and Fisher, 2007). Although KAT domains can target very specific residues, their substrates often overlap and can vary depending on associated interaction partners. KATs act in various multi-subunit complexes and their interacting proteins are important modulators of catalytic activity as well as substrate specificity. Respectively, the diversity of interaction partners and assembled complexes also indicates versatile functions of single KAT proteins. To identify correct genomic regions, reader domains like chromodomains, bromodomains, tudor domains, PHD fingers and WD40 repeats are necessary. These domains recognize histones and can bind to specific regions, e.g. bromodomains bind to acetylated lysine residues and chromodomains recognize methylated lysine tails. Their targets alter through different proteins and these domains are therefore important to recruit KAT complexes to the correct genomic regions. For example, the

chromodomain of heterochromatin protein 1 recognizes H3K9me while the chromodomain of polycomb binds H3K27me. Based on the role of reader domains, other histone modifications, e.g. methylation and sumoylation, affect the recruitment of KATs and regulate histone acetylation status (Lee and Workman, 2007). Autoacetylation is another tool to regulate KAT proteins. Rtt 109, p300/CBP and MYST proteins have been shown to inhabit autoacetylation sites which are essential to regulate catalytic activity and are highly conserved in the subfamilies (Marmorstein and Zhou, 2014). Furthermore, many KATs also acetylate non-histone proteins and regulate their activity, such as Acute Myeloid Leukemia protein1 (AML1), p53, Cohesin and Tubulin (Carlson and Glass, 2014). None-histone targets also include proteins important for DNA replication and recombination as well as DNA repair (Lee and Workmann, 2007).

It is generally assumed that histone acetylation promotes gene expression as active promoters are often highly acetylated. Addition of a neutral acetyl group to the negatively charged lysine residues of histones is assumed to reduce the histone-DNA interaction and therefore loosen up the chromatin structure. This destabilisation of the internucleosomal contacts and loosening of the nucleosomal structure enables the recruitment of the transcriptional machinery to activate gene expression (Heery and Fisher, 2007; Sheikh and Akhtar, 2018).

As KATs are essential regulators of biological processes such as cell-cycle progression, DNA repair, cell differentiation and hormone signalling, they are also indicated in several human diseases such as cancer and developmental disorders. The impact of KATs on transcription makes them important players in oncogenic processes because their activity can dramatically impact tumour growth or suppression (Heery and Fisher, 2007). Genes encoding KATs like p300/CBP, Mixed Lineage Leukemia 1 (MLL1/KMT2A) and MOZ are involved in reciprocal translocations to produce oncogenic fusion proteins leading to mis-targeted acetylation that can drive leukemogenesis. MLL-fusions can increase histone acetylation of targeted genomic regions while MOZ-fusions change gene expression profiles and block

normal differentiation of myeloid progenitors resulting in Acute Myeloid Leukemia (AML) (Di Croce, 2005; Heery and Fisher, 2007). Mis-regulated histone acetylation is also associated with inflammatory lung diseases and metabolic diseases. Furthermore, KATs have been shown to be essential for viral replication e.g. in HIV and fungal infections (Cereseto et al., 2005; Heery and Fisher, 2007; Marmorstein and Zhou, 2014).

Due to their association with different diseases, inhibition of KATs could form an attractive therapeutic target especially in cancers. The structural and mechanistic differences of KAT domains as well as reader domains and interaction sites could be exploited for highly specific drug targeting (Heery and Fisher, 2007). And indeed, recently two inhibitors that compete with acetyl-CoA-binding were shown to have high potency and selectivity for inhibition of MOZ/MORF activity associated with MYC-induced lymphoma in mouse models (Baell et al., 2018)

1.2 The MYST family

Members of the MYST family, named after MOZ, Ybf2/Sas3 (Something About Silencing 3), Sas2 and Tip60 (HIV Tat-interacting 60 kDa protein), are characterized by a conserved KAT (MYST/HAT) domain that consists of a zinc finger and an acetyl-CoA binding motif. Five MYST members are known in humans: TIP60, MOZ, its paralog MORF (MOZ-Related Factor), HBO1 (HAT Bound to ORC1) and hMOF (human Males-absent Of the First), also known as KAT5, KAT6A, KAT6B, KAT7 and KAT8. The family was divided into three subgroups: TIP60 and MOF, MOZ and MORF, and HBO1 (Thomas and Voss, 2007; Yang, 2015). Structural analysis and enzymatic characterisation revealed that the MYST acetylation reaction is catalysed by a ping-pong mechanism (Yan et al., 2002).

Typical for KAT proteins, members of the MYST family have been assigned a variety of functions. The Hbo1 subgroup was linked to initiation of DNA replication and transcriptional suppression. TIP60 and MOF proteins are assumed to function in cell cycle control, regulation of apoptosis, DNA repair and transcriptional regulation

both as suppressor and activator (Thomas and Voss, 2007). The role of MOZ will be described in the following chapters.

1.2.1 The MOnocyctic Leukemia Zinc Finger Protein (MOZ)

The MOnocytic Leukemia Zinc Finger Protein MOZ (also KAT6A or MYST3) is a histone acetyltransferase and a member of the MYST family (**Figure 1.4**). It was first discovered in fusion genes between *KAT6A* and *CBP* which are generated by chromosomal translocations t(8;16)(p11;p13). The protein is encoded by the *KAT6A* gene which is located on chromosome 8p11.2 and is composed of 18 exons (Borrow et al., 1996). Three validated splicing forms described in the NCBI databases show differences in the 5' untranslated region (UTR), but all produce the same protein containing 2004 amino acids with a molecular weight of ~ 225 kDa. MOZ was shown to catalyse the acetylation of the lysine residues K9 and K14 of histone H3 *in vitro* and *in vivo* (Holbert et al., 2007; Voss et al., 2009; Dreveny et al., 2014). Additionally, MOZ acetylates H3, H2A and H2B *in vitro* (Carlson and Glass, 2014).

The N-terminus of MOZ contains the NEMM domain (N-terminal part of Enok, MOZ and MORF) including the H1/5 domain, a Double PHD Finger domain (DPF), the MYST/HAT (KAT) domain and an acidic region. The C-terminal part of MOZ has regions rich in serine (S), proline (P), glutamine (Q) and methionine (M). Relatively little is known about the function of the C-terminus, but it was suggested that the serine and methionine (SM) rich region has transcriptional activation potential (**Figure 1.4**; Yang, 2004; Yang and Ullah, 2007)

The role of the NEMM domain is not fully understood. The C-terminal part of the domain shows weak sequence similarity with the linker histones H1 an H5 and has been associated with nuclear localization as well as transcriptional co-activation due to the similarities to histone H5 (Kitabayashi et al., 2001; Katsumoto et al., 2008)

The DPF domain consists of two plant homeodomain (PHD) fingers. PHD fingers are conserved C3HC4 motifs that are common in nuclear proteins which regulate chromatin status and transcription. The MOZ DPF domain facilitates binding

to the histone-tail of H3 and regulates substrate specificity of MOZ. Interactions of the MOZ DPF with H3 peptides induce the formation of α-helical structures of the histone tail at $H3_{4-11}$ and $H3_{17-25}$ (Dreveny et al., 2014; Xiong et al., 2016). This enables recognition of different post-translational histone H3 modifications which change the binding ability of the MOZ DPF. Acetylation of H3K14, a mark that correlates with transcriptional activation, promotes additional contacts of the DPF to H3 and strengthens their interaction. As shown in **Figure 1.4 (bottom)**, structural analysis by our group showed that H3K14ac directly binds to a pocket of the MOZ DPF domain while H3K9 is oriented away from the domain and might present itself for catalytic activity of the MYST domain. In line with these findings, tri-methylation of H3K9 does not impact binding of the DPF (Dreveny et al., 2014). More recently, the MOZ DPF domain was identified as a reader of H3K14 crotonylation with a four- to eight-fold higher affinity for H3K14cr compared to H3K14ac. Co-localization and ChIP-qPCR studies showed that the MOZ-DPF is essential for the localization of MOZ at H3K14cr sites and the recruitment to promoter regions of MOZ target genes like *HoxA5* and *HoxA9*. These promoter regions also show higher levels of H3K14cr compared to other genes (Xiong et al., 2016).

The MYST domain acts as the catalytic unit of MOZ and has strong similarity to other members of the MYST family. The domain is composed of eight α-helices and nine β-strands folded into two β-sheets enclosed by α-helical fold. The domain inhabits the catalytic residues E680 and C646 in analogous positions to Esa1 which indicates a similar catalytic reaction as described above (Holbert et al., 2007). The N-terminal segment of the MYST domain forms a TFIIIA type zinc finger fold around one bound zinc ion which is required for KAT activity. The C-terminal segment contains a helix-turn-helix motif. Both motifs are known from eukaryotic TFs. It was demonstrated that in addition to its KAT activity, the MOZ MYST domain can bind cruciform DNA and interacts with DNA in the context of assembled nucleosomes (Holbert et al., 2007).

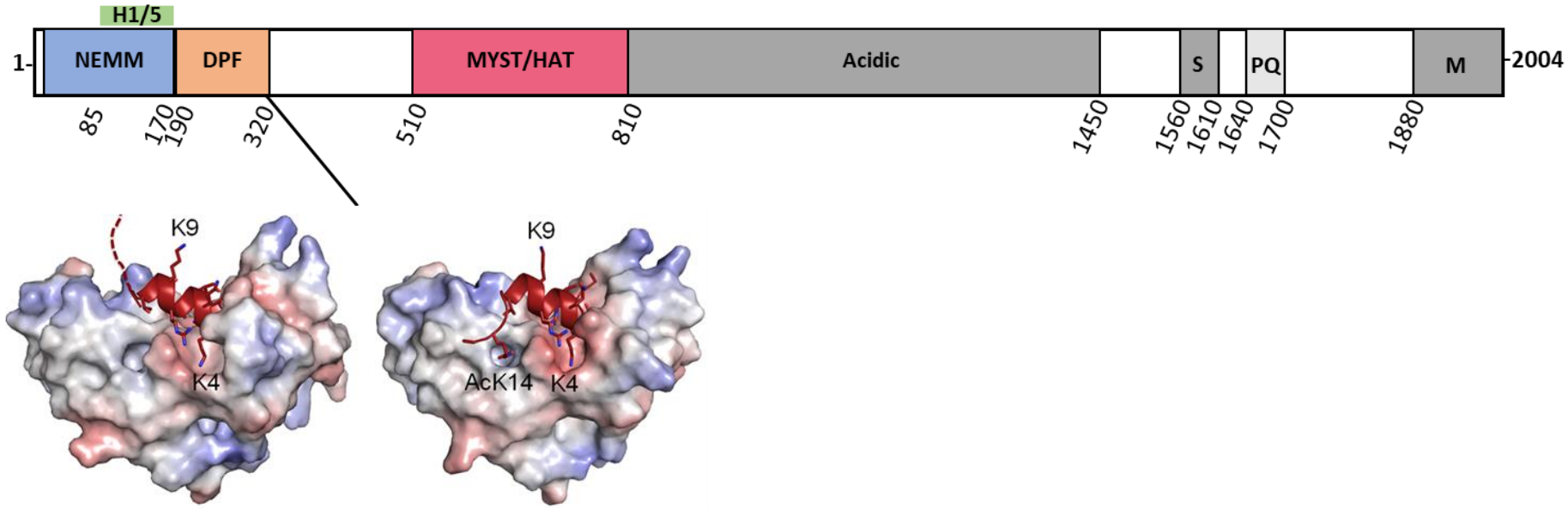


Figure 1.4 Domain organisation and structural formation of MOZ

Top: Schematic of the domain organisation of human MOZ, the H1/5 region maintained in the NEMM domain is depicted at the top. **Bottom:** Structure of the DPF domain in complex with the H3 tail (left) and with acetylated H3K14 (right). Binding induces the alpha helical fold of the tail (Image of the DPF was provided by David Heery, modified after Dreveny et al., 2014). NEMM= N-terminal part of Enok, MOZ and MORF; H1/5= linker histones H1- and H5 like domain; DPF= Double PHD finger domain; MYST/HAT= catalytic histone acetyltransferase domain (KAT domain); Acidic= glutamate/aspartate-rich region; S, PQ, M= serine-, proline/glutamine-, methionine-rich regions.

1.2.2 MOZ acts in multi-subunit protein complexes

Similar to other histone acetyltransferases, MOZ interacts with a variety of proteins forming multi-subunit protein complexes (examples are shown in **Figure 1.5**). Different interacting proteins can impact the KAT activity and which histone tails are targeted by MOZ. Therefore, these complexes may promote different functions of MOZ.

MOZ was found in complexes with the tetrameric INhibitor of Growth 5 (ING5) which facilitate acetylation of H3K14. ING5 is a member of the tumour suppressor ING family which is essential for histone acetylation but does not show any intrinsic enzymatic activity. ING5 may function by recruitment of complexes to specific chromatin regions through their PHD fingers (Doyon et al., 2006; Carlson and Glass, 2014).

BRomodomain PHD-Finger protein 1 (BRPF1) was identified to bridge the association of MOZ with ING5 and EAF6 and similar findings were made for other BRPF proteins and the MOZ orthologue MORF. The MYST domain of MOZ interacts with the BN1 and BN2 subdomains of BRPF1 (Mazumder and Heery, unpublished). Structural analysis suggests that MOZ and MORF complexes are analogous to HBO1-ING4/5 and TIP60-ING3 complexes. Multiple epigenetic reader domains of BRPF1, such as a PWWP domain, a bromodomain as well as a double PHD and a zinc finger assembly, can recognize and bind to specific histone modifications on H2, H3, and H4 (Carlson and Glass, 2014). Association with BRPF1 promotes a drastic increase of the KAT activity of the MOZ MYST domain, both, by direct interaction and recruitment of ING5 which also acts as an activator (Ullah et al., 2008).

It is expected that the combination of different epigenetic reader domains of BRPF1 and ING5 guide the MOZ KAT complex to chromatin substrates where MOZ acetylates different lysine targets. Carlson and Glass suggest a positive feedback loop where the MOZ complex is first recruited by ING5 to H3K4me3 sites of promoter regions. The MOZ DPF domain and the PWWP domain of BRPF1 strengthen the interaction with chromatin and MOZ acetylates different residues on histones H3 and

H4, such as H3K9, H3K14, H4K4, H4K8, H4K12 and H4K16. This increases the affinity of the MOZ DPF domain as well as the bromodomain of MOZ to acetylated lysine residues which further promotes spreading of chromatin acetylation and transcriptional activation (Carlson and Glass, 2014).

In vitro, MOZ interacts with AML1 and promotes expression of haematopoietic genes. Surprisingly, the MYST domain of MOZ is not required to facilitate transcriptional activation in these complexes (Kitabayashi et al., 2001). In addition, MOZ interacts with PU.1, a TFs which is essential for HSC maintenance and differentiation (Katsumoto et al., 2006). Another interaction partner is MLL1 which is a histone methyltransferase that di- and trimethylates H3K4. In these complexes, MOZ also associates with WD Repeat domain 5 (WDR5) which stabilizes the binding of MLL complexes to promoter sites. MOZ-MLL complexes cooperate to regulate expression of the *Hox* genes *HoxA5*, *HoxA7* and *HoxA9* through epigenetic modifications (Paggetti et al., 2010). Symplekin was also found to interact with MOZ in MOZ-MLL complexes where it prevents their recruitment to *HoxA9* promoter sites (Largeot et al., 2013). MOZ was associated with several other proteins including RUNT-related transcription factor 2 (RUNX2), MAFK (v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog K), and p53 (Kindle et al., 2005; Collins et al., 2006; Perez-Campo et al., 2013).

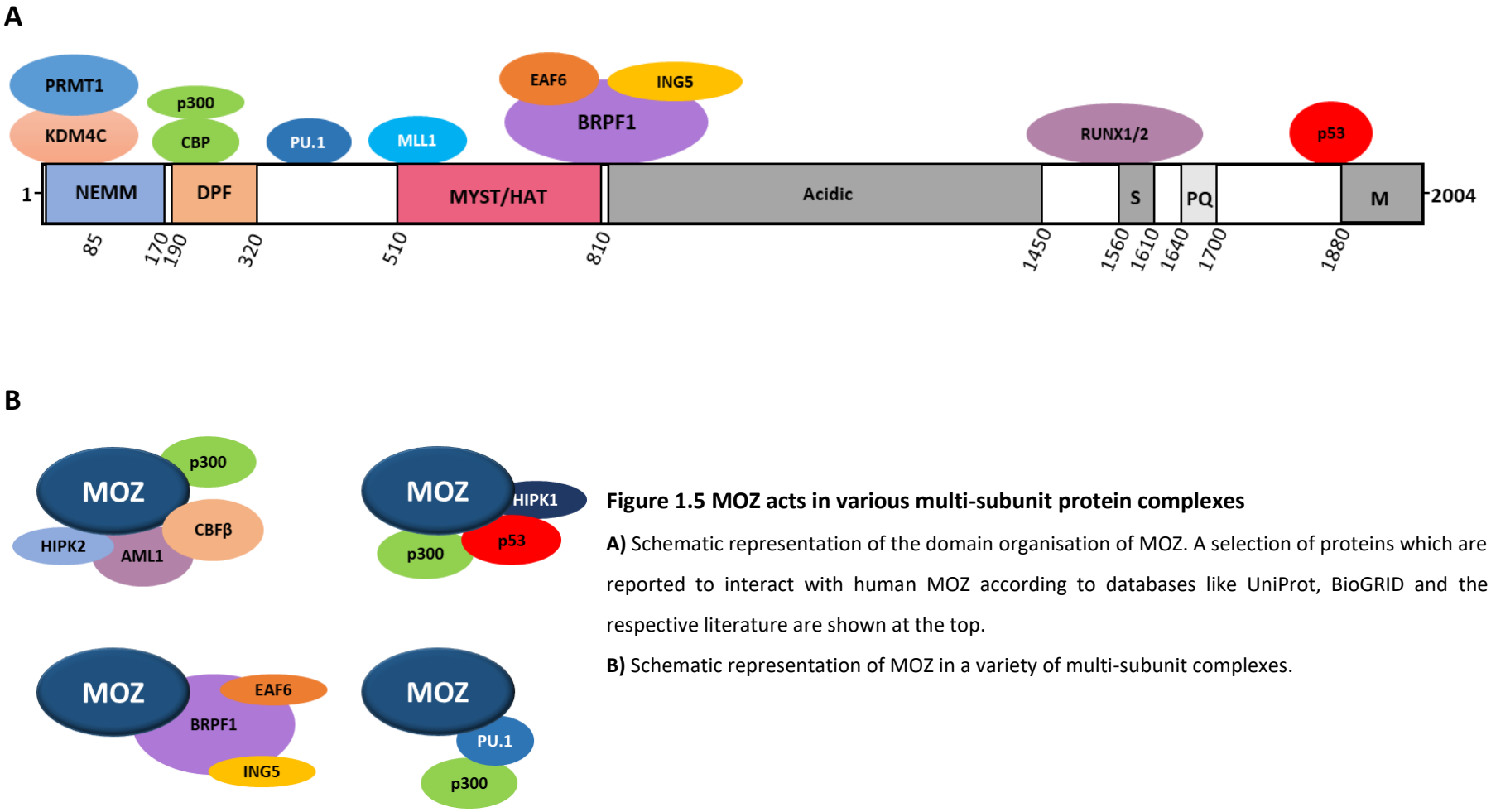


Figure 1.5 MOZ acts in various multi-subunit protein complexes

A) Schematic representation of the domain organisation of MOZ. A selection of proteins which are reported to interact with human MOZ according to databases like UniProt, BioGRID and the respective literature are shown at the top.

B) Schematic representation of MOZ in a variety of multi-subunit complexes.

1.2.3 MOZ in development and haematopoiesis

MOZ is associated with a variety of developmental processes, mainly through regulation of *Hox* genes which are known as critical regulators of body segment identity during development.

Zebrafish studies reported that MOZ-deficient mutants lack *HoxA2a* and *HoxB2b* expression which causes developmental defects in the facial region. MOZ-dependent *Hox* gene expression is essential for correct skeletogenesis of cranial neural crest cells (CNCs) (Miller et al., 2004; Crump et al., 2006). MOZ-deficient mouse embryos show an anterior homeotic transformation of the nervous system and axial skeleton in the cervical and thoracic region. In these embryos, H3K9 acetylation is especially reduced at *Hox* gene promoters, causing a downregulation of *Hox* gene expression. Based on these studies, MOZ plays different roles in the developmental in mice and zebrafish and seems to be required even more for correct development in mice (Voss et al., 2009).

MOZ and BRPF2 are upregulated during differentiation of mouse embryonic stem cells (ESCs) and interact to mediate proper H3K14 acetylation. The BRPF2-MOZ complex is required for correct differentiation of ESCs as lack of either protein resulted in severe defects in RA-induced differentiation (Cho et al., 2016).

MOZ in complex with ING5 was reported to regulate transcriptional activation of *T-box DNA-binding Transcription Factor 1* and *5* (*TBX1* and *5*) by acetylation of H3K9 at the associated promoter regions. Loss of MOZ in mice dramatically reduces expression of *Tbx1* and causes a phenotype similar to the DiGeorge syndrome (DGS) with ventricular septal defects (VSDs) (Voss et al., 2012; Vanyai et al., 2015).

Throughout the lifespan of an organism cells of the blood and immune system have to be replenished. HSC proliferation, maintenance and differentiation into different committed progenitors have to be strictly regulated at the transcriptional level. KATs, such as p300, CBP, HBO1 and MOZ are important players in haematopoiesis and regulate blood cell development (Sun et al., 2015).

MOZ was shown to be essential for maintenance of hematopoietic stem cells and correct differentiation of myeloid progenitors (Katsumoto et al., 2006). Mice deficient in MOZ are embryo lethal at developmental stage E15 with an undersized liver. During this embryonic stage the liver is responsible for production of hematopoietic cells and $MOZ^{-/-}$ mice showed a profound reduction of HSCs and hematopoietic progenitor cells, indicating an essential role of MOZ in maintenance of HSCs. However, lineage commitment of hematopoietic progenitors was not affected (Katsumoto et al., 2006; Thomas et al., 2006). The expression of important regulators of HSC maintenance and activity, such as *HoxA9* or the proto oncogenes c-*Mpl* (*Myeloproliferative leukemia protein*) and *c-Kit*, was significantly reduced in $MOZ^{-/-}$ liver cells indicating a combined effect of downregulation of different factors. This might be caused by reduced activity of AML1 and PU.1 which require MOZ as a transcriptional co-activator. Furthermore, T-cell production in the thymus was reduced while the few cells present were still able to undergo normal maturation (Katsumoto et al., 2006). MOZ is also required for the expansion of B-cell progenitors, most likely by transcriptional activation of *Myeloid Ecotropic Viral Integration Site 1* (*Meis1*) and *HoxA9* in co-operation with MLL (Sheikh et al., 2015 (2)).

Hematopoietic stem cells are able to reconstitute the hematopoietic system of lethally irradiated mice (Kondo et al., 2003). Recipients of embryonic fetal liver cells of $MOZ^{-/-}$ mice did not survive beyond day 28, as cells lacking MOZ expression were not able to reconstitute haematopoiesis demonstrating the essential role of MOZ (Thomas et al., 2006). A recent study showed that loss of MOZ in adult mice causes a rapid loss of HSCs which were also not reconstituted throughout the mice's lifespan. However, the absence of HSCs did not affect the viability or health of the mice. Surprisingly, mice lacking HSCs maintained blood and showed normal hematocrit as well as erythrocyte morphology suggesting that long-term haematopoiesis can be maintained by self-renewal of early progenitor cells without the involvement of HSCs (Sheikh et al., 2016).

Overall, these studies show that MOZ plays various important roles in development and stem cell maintenance by mediating transcriptional activation. The precise function of MOZ is highly context-specific and dependent on the cell type and development stage (Sheikh and Akhtar, 2018).

1.2.4 MOZ in disease

As described above, MOZ is essential in several developmental processes during embryogenesis, *Hox* gene regulation, and stem cell maintenance. Thus, it is not surprising that mutations and miss-regulation of MOZ are indicated in different diseases linked to the proteins' role in developmental processes.

Recent studies in the last 5 years have identified several KAT6A variants that can be causative for a global developmental delay syndrome (KAT6A GDD, also Mental Retardation autosomal Dominant 32 (MRD32); e.g. Arboleda et al., 2015; Tham et al., 2015; Satoh et al., 2017; Murray et al., 2017; Zwaveling-Soonawala et al., 2017). These studies identified KAT6A patients that display intellectual disabilities and delay of important developmental steps such as the first social smile, unaided sitting, independent walking or the ability to speak. The KAT6A variants found in GDD are *de novo* nonsense or missense mutations that are mainly located in the acidic region of MOZ (Exon17/18) and cause los of the C-terminal part. One mutation was also found in MYST domain and one was located in Exon 3 (**Figure 1.6**; Sheikh and Akhtar, 2018).

Mutations and miss-regulation in epigenetic regulators are common in different types of leukemia and lymphomas. Several known chromosomal translocations in KAT genes, such as MOZ an CBP/p300, cause the formation of fusion proteins inducing haematological malignancies. Secondary mutations of KATs also contribute to worse disease outcomes in different cancers (Sun et al., 2015; Sheikh et al., 2015 (1)). Chromosomal translocations of MOZ lead to the formation of fusion proteins which were found to be causative for Acute Myeloid Leukemia (AML). In this heterogeneous hematopoietic stem cell disorder, myeloid cells do not differentiate

and consequently blast cells accumulate while the granulocyte lineage is absent. MOZ was found in different fusions e.g. with CBP, p300, as well as the Nuclear Receptor Coactivator 2 (NCOA2/TIF2), 3 (NCOA3), LEUTX (**Figure 1.6**; Katsumoto et al., 2008; Chinen et al., 2014). In general, breakpoints are located in the acidic region of MOZ and the N-terminus with the NEMM-, H1/5 like-, DPF- and MYST domain are maintained while the C-terminus is replaced by a fusion partner. The fusion partners are required for transformation as loss of the C-terminus alone is not sufficient to induce transformation (Deguchi et al., 2003).

Using deletion-constructs it was shown that the DPF domain is not required for transformation in MOZ-TIF2 fusions. The MYST domain of MOZ and the CID domain of TIF2 are sufficient to activate oncogenic activity (Deguchi et al., 2003). Our group identified a conserved leucine-rich helix (LXXLL motif) in Steroid Receptor Coactivator (SRC1) and TIF2 that is required for CBP and p300 binding (Sheppard et al., 2001; Matsuda et al., 2004; Waters et al., 2006) and was found to be essential for the leukemogenicity of MOZ-TIF2 in cell and animal studies (Sheppard et al., 2001; Deguchi et al., 2003; Kindle et al., 2005). CBP-recruitment by MOZ-TIF2 could serve as a constitutive co-activator at MOZ target genes or recruit and modulate additional factors (Deguchi et al., 2003). Another study demonstrated that the first 79 amino acids of MOZ are required to induce leukemic cell transformation while the H1/5 region as well as the PHD fingers were not essential (Cheung et al., 2016). It was demonstrated that MOZ-TIF2 mutants with a deletion of the first 79 amino acids failed to induce transformation, opposed to internal deletions of the H1/5 region or the DPF domain which did not compromise transformation of HSCs. This was first associated with interaction of MOZ with the Protein arginine N-methyltransferase 1 (PRMT1) which was mapped to be dependent on the first 79 amino acids of MOZ. However, covalently fusing PRMT1 to the deletion constructs did not rescue leukemogenesis in *in vitro* mouse models indicating that PRMT1 recruitment is necessary but not sufficient for leukemic transformation. Consequent analysis found that interaction of MOZ-TIF2 with the Lysine Demethylase 4C (KDM4C) was also

dependent on the first 79 amino acids of MOZ. In contrast to PRMT1, KDM4C was found to be essential for leukemia development in MOZ-TIF2 fusions, but the exact relevance of the first 79 amino acids of MOZ was not further tested. The authors suggested that recruitment of KDM4C mediates removal of H3K9 tri-methylation at gene loci such as *HoxA9*, which enables consequent acetylation and transcriptional activation by MOZ (Cheung et al., 2016). Interestingly, the NEMM domain as well as the H1/5 domain of MORF were also maintained in fusions of MORF with KANSL1 (KAT8 Regulatory NSL Complex Subunit) causing retroperitoneal leiomyoma which supports the functional relevance of the N-terminus of MOZ/MORF in disease and most likely also in none-disease contexts (Panagopoulos et al., 2015).

Like MOZ, MOZ-TIF2 also forms stable complexes with BRPF1. It was suggested that BRPF1 is required to localize MOZ-TIF2 at *Hox* promoters where MOZ-dependent acetylation upregulates *Hox* expression (Shima et al., 2014).

The chromosomal translocation *t*(8;16)(p11;p13) leads to MOZ-CBP fusions and subsequent leukemia formation. It was suggested that impairment of AML1-induced transcriptional activation as well as interaction with Nuclear Factor Kappa-light-chain-enhancer of activated B cells (NF-kB) could be critical in leukemogenesis. The N-terminus of MOZ can also be fused to p300 in *t*(8;16)(p11;q13) translocations which leaves the KAT domains of both proteins intact. Abnormal histone acetylation caused by deregulated KAT activity could play an essential role in leukemogenesis (Sun et al., 2015).

In addition to hematologic malignancies, MOZ mutations are also relevant in oesophageal adenocarcinoma (Dulak et al., 2013) and MOZ up-regulation was found among top-ranking targets in different cancers and suggested to be an important promoter of the oncogenic state (Yang, 2015; Sheikh and Akhtar, 2018). MOZ is highly expressed in p53 gain-of-function tumours which increases global histone acetylation (Zhu et al., 2015). Furthermore, MOZ is localized in the chromosome 8p11-p12 locus which is amplified in 12-15% of breast cancers and increased MOZ expression was linked to worse predictions. It was suggested that overexpression of MOZ promotes

Estrogen Receptor α (*ERα*) acetylation and upregulation, which is known as a major driver in breast cancers (Turner-Ivey et al., 2014; Yu et al., 2017).

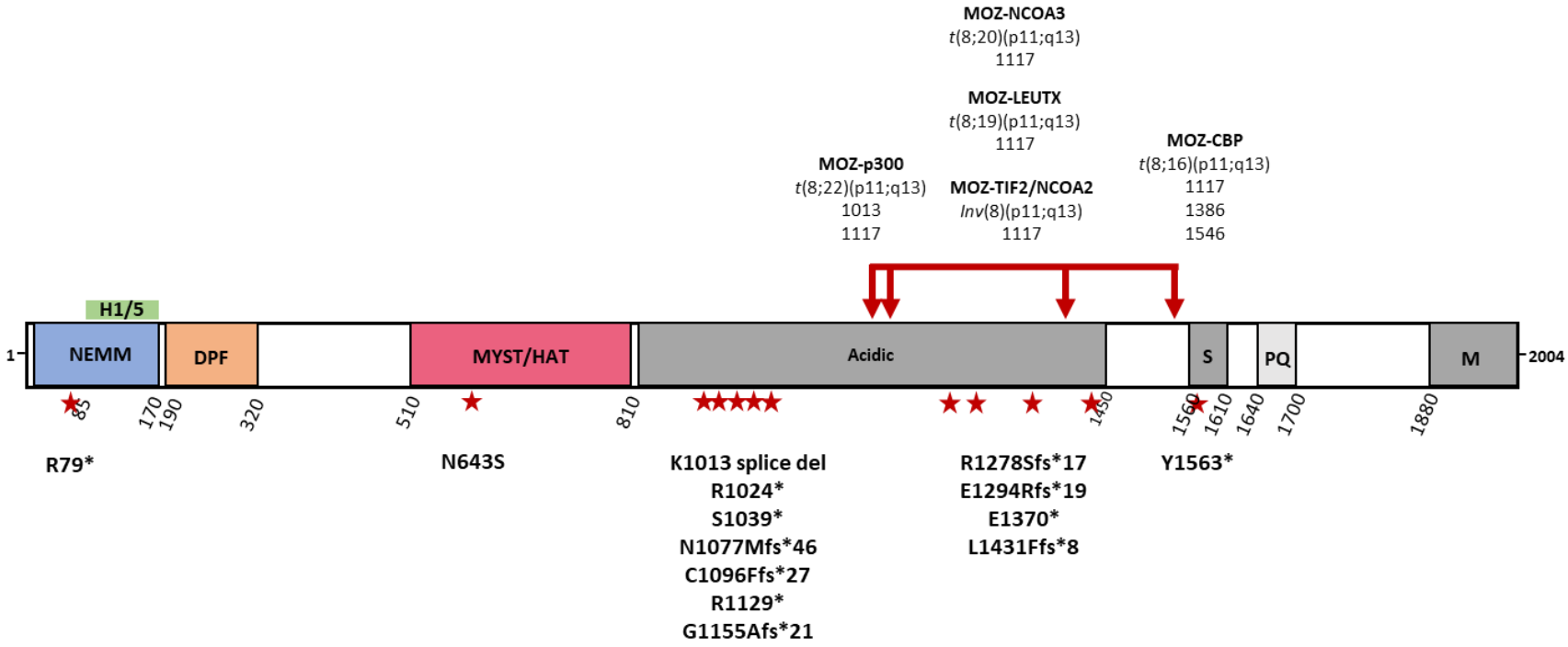


Figure 1.6 Schematic representation of the KAT6A/MOZ gene and disease related breakpoints and mutations

Locations of breakpoints associated with AML are indicated by red arrows (top). Predicted MOZ mutations originating from d*e novo mutations* in the KAT6A gene as reported in the literature, associated with Global Developmental Delay are indicated by red stars (bottom). Figure adapted and modified after Whitchurch and Heery, unpublished and Yang, 2015.

1.3 Project Aims and Objectives

Taken together, the *KAT6A* gene and its encoded protein MOZ play important roles in gene regulation during development and are involved in a variety of diseases. MOZ and its orthologue MORF are expressed in early stem cell progenitors and are required for long term renewal and maintenance of stem cells. Dysfunctional expression of MOZ in oncogenic fusion proteins in leukemias as well as aberrant and upregulated expression in other cancers indicate the therapeutic importance of MOZ. Characterisation of the MYST domain of MOZ was already exploited to develop small molecule inhibitors that are effective in a mouse model of B-cell lymphoma. In addition, investigation of its association with neurodevelopmental disorders has just begun and further research will be required to help these patients and their families understand their condition and its clinical progression.

Past research was strongly focussed on the N-terminus of MOZ because it contains the characteristic MYST (KAT) domain and is especially relevant in AML. While the MYST domain and the DPF domain are structurally and functionally well understood, relatively little is known about the NEMM and H1/5 domain. However, it was shown in previous studies that this region is essential in MOZ-TIF2 induced AML (Cheung et al., 2016) and retroperitoneal cancer (Panagopoulos et al., 2015). Given the importance of the N-terminal region of MOZ in AML, further investigation could help to reveal and define novel functions of MOZ in development and disease.

This study aimed to further investigate the role of the NEMM and H1/5 domain in regards of structure and function. First, structural predictions with Phyre2 and RaptorX software as well as sequence analysis were used to predict possible functions of the region. Following *in silico* analysis, potential functions were investigated and defined using polypeptide-based *in vitro* assays as well as cell-based models.

Chapter 2: Materials and Methods

2.1 Source of Materials

General laboratory supplies were analytical grade and purchased from Fisher Chemicals or Sigma Aldrich unless otherwise stated. All solutions were prepared in double deionised water or RNase free water if required. Deionised water was purified by passage through a Purite select/Neptune water purification system. The pH of all solutions was maintained and adjusted using a pH meter (Jenway 3150) with known pH standards. Sterilisation was achieved by autoclaving for 20 min at 120 °C or filtration through a 0.22 micron filter. Phosphate buffered saline (PBS) was prepared using PBS tablets supplied by OXOID Ltd.

Molecular size markers (1 kb Plus DNA ladder) and dNTPs were acquired from New England Biolabs (NEB) or Invitrogen Life Science. Oligonucleotides were purchased from Sigma-Aldrich. Enzymes were purchased from NEB and Roche.

SDS-PAGE and western blot transfer apparatus and Nitrocellulose transfer membrane were purchased from Bio-rad Laboratories. Acrylamide mix 30% (w/v) was obtained from National Diagnostics. Protein standards were purchased from NEB. HisPur™ Ni-NTA Resin and HisPur™ Cobalt Resin were acquired from Thermo Fisher. 5 ml HiTrap® Chelating HP and 5ml HiTrap® Heparin HP columns, Vivaspin™ concentrator spin columns as well as thrombin enzyme were obtained from GE Healthcare. Lysozyme was purchased from Pharmacia Biotech.

2.1.1 Bacterial culture reagents

The *Escherichia coli* (*E. coli*) strain DH5α (Hanahan, 1983) was used for molecular cloning, DNA manipulations and plasmid expression and BL21-DE3 cells were used for protein expression, purchased from Stratagene. Tryptone and yeast extract for bacterial growth medium were purchased from OXOID Ltd.

2.1.2 Tissue culture reagents

Supplements for tissue culture were obtained from Lonza Biowhittaker® and plasticware was acquired from Helena Biosciences. HEK293 cells lines were purchased from the European Collection of Cell Cultures (ECACC).

2.2 Bacterial preparation and culture

2.2.1 Composition of solutions and media used in bacterial procedures

Luria-Bertani (LB) medium: 1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract at pH 7.0. For preparation of LB agar plates 2% (w/v) bacteriological agar was added. Melted agar medium was supplemented with the required antibiotic and 20-25 ml poured into 10 cm petri-dishes. Plates were stored at 4 °C for up to 1 month.

TBF1 buffer: (30 mM KOAc, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol, pH 5.8).

TBF2 buffer: (10 mM MOPS pH 6.5, 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol).

Ampicillin: 1000x stock solution: 100 mg/ml in ddH_2O.

Kanamycin: 1000x stock solution: 50 mg/ml in ddH_2O.

2.2.2 Culturing of *Escherichia coli* strains

Single colonies from agar plates were used to inoculate liquid LB-cultures containing the appropriate antibiotics which were grown O/N at 37 °C while shaking at 220 rpm in cornical flasks or universal tubes (Bibby Sterilin, Staffordshire, UK).

2.2.3 Long-term storage of bacteria

For long-term storage, glycerol stocks consisting of a 75% to 25% mixture of bacteria in LB medium and glycerol were prepared and stored at -80 °C.

2.2.4 Preparation of competent *Escherichia coli* cells

Competent *E. coli* cells were prepared using a protocol based on Promega protocols and application guides, derived from a modified and less toxic version of

the Hanahan protocol (Hanahan, 1983). DH5α cells were freshly streaked on LB-agar plates and a single colony was picked to inoculate a 5 ml overnight culture and grown at 37 °C, 220 rpm. The overnight culture was used to inoculate a 250 ml LB culture supplemented with 20 mM $MgSO_4$. This culture was grown until the optical density $(OD)_{600}$ reached 0.4. Cells were pelleted by centrifugation at 4,000 rpm for 10 min at 4 °C. Cells were resuspended in 100 ml ice cold TBF1 buffer, incubated on ice for 5 min and again pelleted by centrifugation. The bacteria were resuspended in 6 ml ice cold TBF2 buffer and incubated on ice for 60 min. Finally, the cells were aliquoted, snap frozen in liquid nitrogen and stored at -80 °C.

2.2.5 Transformation of *Escherichia coli* cells

Heat shock protocols were used to transform plasmid DNA in DH5α and BL21-DE3 strains (Hanahan et al., 1991). 50 µl of competent cells and 1-100 ng plasmid DNA were thawed on ice, mixed and incubated for 30 min. Cells were heat shocked at 42 °C for 45 sec to 90 sec (depending on the size of the construct) followed by a 2 min incubation on ice. Afterwards, 500 µl LB medium was added and the cells were incubated at 37 °C, 220 rpm for 1 h. After centrifugation at 7200 rpm for 2 min, 400 µl media was removed, the cells resuspended in the remaining media and spread on LB agar plates containing the required antibiotic. Depending on the plasmid and transformation efficiency, 10-100% of bacteria were plated. The plates were incubated O/N at 37 °C.

2.3 Molecular biology techniques

2.3.1 Small-scale plasmid purification from bacterial cells

5-10 ml *E. coli* cultures were grown O/N in LB media supplemented with the appropriate antibiotic at 37 °C, 220 rpm and harvested by centrifugation. Plasmid DNA was extracted using the NucleoSpin® Plasmid kit (Macherey-Nagel) following the manufacturer's instructions.

2.3.2 Large-scale plasmid purification from bacterial cells

250 ml LB media containing the appropriate antibiotic was inoculated with 5 ml overnight culture and grown O/N at 37 °C, 220 rpm. Cells were harvested by centrifugation and plasmid DNA was extracted using the QIAGEN Plasmid Maxi Kit following the manufacturer's instructions.

2.3.3 Ethanol precipitation

To concentrate DNA samples, 4 volumes of pure ethanol (>99%) and 0.1 volume of 3 M sodium acetate (NaOAc; pH 5.2) were added to the sample. After vigorous mixing and precipitation at -20°C for at least 1 h, the DNA was pelleted for 30 min at 13,000 rpm, 4 °C. To remove excess salt ions, the pellet was washed with 75% ethanol and the centrifugation repeated. The pellet was dried with an open lid under vacuum in a heated speed-vac (Savant DNA120 speed-vac) or in a heat block set at 55 °C followed by resuspension in sterile water.

2.3.4 Spectrophotometric quantification

The concentration of DNA was measured by OD measurement at 260 nm using the NanoDrop® ND-1000 UV-Vis Spectrophotometer. The OD_{260}/OD_{280} and OD_{260}/OD_{230} readings were also assessed to determine sample purity. To validate the results, plasmids DNA were restriction digested and compared to known molecular weight markers following gel electrophoresis.

2.3.5 Agarose gel electrophoresis

DNA was separated according to size on neutral agarose gels consisting of 0.8% - 1.2% agarose (w/v). Gels were prepared in 1x TBE buffer (40 mM Tris-base, 40 mM Boric Acid, 0.5 mM EDTA (pH 8.0)) containing 0.5 µg/ml ethidium bromide for DNA staining as described in standard protocols (Sambrook et al., 1989). DNA samples were mixed with 10X loading dye (39% (v/v) glycerol, 2% (v/v) 0.5 M ETDA, 0.25% (w/v) bromphenol blue, 0.25% (w/v) xylene cyanol FF) and loaded alongside

Invitrogen 1kb plus DNA ladder (0.2 µg/µl). Gels were run in Bio Rad Mini Sub® GT chambers for 30-60 min at 80-100 V. Bands were then visualised with an ultraviolet transilluminator (BioRad Gel Doc 2004) and analysed with BioRad Quantity One software.

2.3.6 Extraction of DNA from agarose gels

DNA was separated by agarose gel electrophoresis, bands were visualized using an ultraviolet light box and then cut out with a scalpel. DNA fragments were purified using the NEB Monarch® DNA Gel Extraction KIT according to the manufacturer's protocol.

2.3.7 DNA polyacrylamide gel electrophoresis (DNA-PAGE)

DNA-PAGE was used for separation of small DNA fragments (oligonucleotides) and Electrophoretic Mobility Shift Assays (EMSAs). 0.5x TBE, 6% - 8% polyacrylamide PAGE-gels were prepared with a Biorad apparatus according to **Table 1**. Gels were pre-run in 0.5x TBE buffer for 60-90 min at 4 °C and 100 V until the current was stable.

DNA samples were mixed with 10X loading dye (39% (v/v) glycerol, 2% (v/v) 0.5 M ETDA, 0.25% (w/v) bromphenol blue, 0.25% (w/v) xylene cyanol FF) or glycerol for a final concentration of 5% and loaded alongside Invitrogen 1kb plus DNA ladder (0.2 µg/µl). Gels were run in 0.5x TBE buffer for 60-90 min until the required separation was achieved.

Afterwards gels were stained with a 0.5 µg/ml ethidium bromide solution for 15 min while gently shaking. Bands were then visualised with an ultraviolet transilluminator (BioRad Gel Doc 2004) and BioRad Quantity One software.

Table 1 Composition for 0.6%, 0.7% and 0.8% DNA PAGE gels.

The volumes are stated for the production of one 10 cm gel.

% Acrylamide	6%	7%	8%
ddH_2O (ml)	6.9	6.57	6.2
5x TBE buffer (ml)	1.00	1.00	1.00
30 % (v/v) acrylamide mix (ml)	2.00	2.33	2.70
10% (w/v) APS (ml)	0.1	0.1	0.1
TEMED (ml)	0.001	0.001	0.001

2.3.8 Sequencing of DNA

Mini prepped DNA samples were sent for sequencing to Source Bioscience Ltd Nottingham, UK. Sequencing reactions were carried out by the company using Big Dye v3.1 chemistry and an AB13730xl automated sequencer.

2.3.9 Oligonucleotides

Oligonucleotides were designed using ApE, Snapgene® or Benchling software and purchased from Sigma Genosys as lypholised and desalted pellets. They were resuspended in sterile water according to the volume prescribed by the manufacturer to obtain a concentration of 100 µM and stored at -20 °C.

2.3.10 Annealing of single-stranded Oligonucleotides

Complementary oligonucleotides were mixed and incubated at 95 °C for 5 min. Afterwards, they were placed in a 95 °C water bath and let to cool at RT. Alternatively, a thermo cycler was used to heat the complementary oligonucleotides at 95 °C for 5 min followed by gradually cooling them to 20 °C (1 °C/1 min). Finally, all oligonucleotides were incubated at 4 °C for 1 h, aliquoted and stored at -20 °C.

2.3.11 Polymerase chain reaction (PCR)

PCR was used to amplify DNA fragments for construction of recombinant plasmids and genotyping. Reactions were carried out in an Applied Biosystems 2720 thermal cycler. PCR products were checked for the correct size by separation during agarose gel electrophoresis (Chapter 2.3.5).

Taq polymerase (Thermus aquaticus) was used to amplify DNA fragments from plasmid or gDNA for genotyping and colony PCR. *Taq* DNA polymerase (NEB) reactions were prepared as stated in **Table 2** and typically carried out as described in **Table 3.**

Phusion® High-Fidelity Polymerase (NEB) was used for cloning purposes and site-directed mutagenesis which require a polymerase with proofreading ability (3' to 5' exonuclease). Phusion PCR reactions were prepared as stated in **Table 4** and typically carried out as described in **Table 5** using "hot start" PCR.

Table 2 Standard Taq PCR reaction mix (25 µl)

2.5 µl	Standard *Taq* Reaction Buffer
0.4 µl	dNTPs (10 mM each of nucleotide)
0.5 µl	Primer *forward* (10 µM)
0.5 µl	Primer *reverse* (10 µM)
0.125 µl	*Taq* DNA Polymerase
15-150 ng	Template DNA
ad 25 µl	ddH_2O

Table 3 Standard Taq PCR reaction-PCR

	Temperature [°C]	Duration [s]	Number of cycles
Denaturation	95	30	1x
Denaturation	95	10	
Annealing	40-60	30	30x
Elongation	72	60s/kbp	
Elongation	72	300	1x
Cooling	8	∞	

Table 4 Standard Phusion-PCR reaction mix (50 µl)

10 µl	HF buffer 5x
1 µl	dNTPs (10 mM of each nucleotid)
2,5 µl	Primer *forward* (10 µM)
2,5 µl	Primer *reverse* (10 µM)
15-150 ng	Template DNA
1,5 µl	DMSO
0,5 µl	Phusion High-Fidelity DNA Polymerase
ad 50 µl	ddH_2O

Table 5 Standard Phusion-PCR reaction

	Temperature [°C]	Duration [s]	Number of cycles
Denaturation	98	30	1x
Denaturation	98	10	
Annealing	40-60	30	40x
Elongation	72	30s/kbp	
Elongation	72	480	1x
Cooling	8	∞	

2.3.12 Restriction digest

Restriction digest of plasmid DNA was used to generate compatible ends for cloning, to verify insertions and to check the quality of plasmids. Restriction sites were inserted during PCR. Buffers and reaction temperatures were used according to the manufacturer's protocols (Roche, NEB). Typically, digests were performed with 1 µg plasmid DNA and 2 units restriction endonuclease in a final volume of 50 µl containing 1x reaction buffer and were carried out at 37 °C for 2 h.

2.3.13 Site-directed mutagenesis (SDM)

To mutate or delete specific nucleotides in plasmid DNA, two complimentary oligonucleotides were designed that spanned the region to be mutated and contained the required base pair changes (for primers see appendix). Phusion-PCR reactions (Chapter 2.3.11) were carried out to incorporate the mutagenic primers in the plasmid DNA, resulting in nicked circular strands of mutated DNA. The reactions were carried out as described above with an input of 100 ng plasmid DNA and a prolonged extension time (depending on the construct) at 72 °C. To digest the none-mutated parental DNA after PCR the reaction was treated with 0.5 µl Dpn1 and incubated for 3 h at 37 °C. Reactions and controls were separated by agarose gel electrophoresis (Chapter 2.3.5) to check for complete digestion of the parental DNA. The PCR products were then transformed in DH5α, single colonies picked for small-scale plasmid preparations, which were then analysed by sequencing (Chapter 2.3.8).

2.3.14 Purification of genomic DNA from cell lines

Cells were grown as described in chapter 2.4 , harvested by centrifugation at 2000 rpm for 5 min and 2x washed with PBS. Genomic DNA (gDNA) was extracted using the Sigma GenElute Mammalian Genomic DNA Miniprep Kit following the manufacturers protocol.

2.3.15 RNA extraction and preparation of cDNA

Cells were grown as described in chapter 2.4 , harvested by centrifugation at 2000 rpm for 5 min, snap frozen in liquid nitrogen and stored at -80 °C. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Afterwards, reverse transcription was carried out with 500 ng of total RNA using the QuantiTect Reverse Transcription Kit (Qiagen).

2.3.16 Quantitative PCR analysis

Quantitative PCR (qPCR) was carried out with cDNA (Chapter 2.3.15) in 96-well plates or qPCR stripes (8 well) using 2x Brilliant II SYBR Green qPCR Master Mix (Agilent) in a Stratagene Mx3000P qPCR machine.

2.4 Cell culture

2.4.1 Maintenance of cell lines

The HEK293 human embryonic kidney cell line was grown in Dulbecco's Modified Eagle medium (DMEM), supplemented with 10% (v/v) heat inactivated FCS, 1% (v/v) 0.2 M L-glutamine and 1% (v/v) Penicillin/Streptomycin antibiotics. Adherent cells were grown in sterile plastics at 37 °C in a humified atmosphere containing 5% CO_2 and passaged at approximately 80% confluency for a maximum of 30 passages. For splitting cells, they were washed of in media, seeded into new flasks at a density of 1 x 10^5 cells per T75 flask and fresh medium was added.

2.4.2 Calcium phosphate-mediated transfection of HEK293 cells

Prior to transfection, HEK293 cells were seeded at a density of 1 x 10^6 cells per 10 cm dish in 10 ml culture medium or at a density of 2.5x 10^4 cells per well in 300 µl culture medium in Ibidi µ-Slide 8 well chambers. Cells were allowed to adhere for 24h and the medium was replaced 2-3 h prior to transfection. For each transfection a mixture was prepared as described in **Table 6.** The transfection mixture was added dropwise to a 2x hepes buffered saline (HBS) solution while vortexing. The

mixture was incubated for 20 min at RT and added dropwise to the cells. After 16 h of incubation the cells were washed three times with PBS and fresh media was added.

Table 6 Composition of transfection mixture for calcium phosphate-mediated transfection of HEK293 cells.

	Per 10 cm dish	Per Ibidi µ-Slide 8 well chambers
2 M $CaCl_2$	100 µl	3.75 µl
Plasmid DNA	14 µg	25 ng
Sterile ddH_2O	ad 800 µl	ad 30 µl

2.4.3 Cryopreservation for long-term storage of cells lines

$1x10^6$ Cells were harvested by centrifugation and washed 2x with 5 ml PBS. The cell pellet was resuspended in 400 µl of medium and mixed with 100 µl (10%) DMSO and 500 µl (50%) FCS in a cryovial. The suspension was gently mixed by pipetting, frozen at - 20 °C, placed at - 80 °C overnight and finally stored in liquid nitrogen.

2.4.4 Resuscitation of mammalian cells after cryopreservation

Cells in cryovials were thawed at 37 °C for 3 min. Cells were then transferred into T75 flasks and mixed with pre-warmed medium. After allowing the cells to recover overnight, the medium was changed to remove dead cells and debris. Cell lines were passaged at least three times before using them in further experiments.

2.4.5 Paraformaldehyde fixation and nuclear staining of cells

After transfection in Ibidi µ-Slide 8 well chambers, cells were gently washed twice with PBS and fixed by incubation in with 4% (w/v) paraformaldehyde for 10 min. Cells were washed 3x with PBS and stained with DRAQ-5 (concentration 1:2000 in PBS) for 5 min at RT. Cells were washed 2x with PBS and afterwards stored in PBS at 4 °C and protected from light.

2.4.6 Confocal imaging

Fluorescent cells were imaged using a Zeiss LSM510 Meta confocal microscope. Images were analysed and processed with LSM Zeiss Image Browser and ImageJ software.

2.5 Biochemical techniques

2.5.1 Compositions of solutions used in biochemical methods

Lysis buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 300 mM NaCl, 10 mM Imidazole pH 8.0; 1 mg/ml lysozyme.

Wash buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 300 mM NaCl, 20 mM Imidazole, pH 8.0.

Elution buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 300 mM NaCl, 300 mM Imidazole, pH 7.5.

Low-salt buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 300 mM NaCl pH 7.5.

High-salt buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 1.5 M NaCl pH 7.5.

Dilution buffer (for purification of His-tagged proteins): 20 mM Tris-HCL pH 7.5.

Low-salt heparin buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 175 mM NaCl pH 7.5.

High-salt heparin buffer (for purification of His-tagged proteins): 20 mM Tris-HCL, 1.5 M NaCl pH 7.5.

Phosphate MOZ-Lysis Buffer: 20 mM sodium phosphate (pH 7), 250 mM NaCl, 30 mM sodium pyrophosphate, 0.1% NP-40, 10 mM NaF, 0.1 mM Na3VO4, 1 mM PMSF, 1x Complete Protease Inhibitors EDTA-free cocktail.

SDS-PAGE loading buffer (4x): 62.5 mM Tris HCL pH 6.8, 40% glycerol, 2% SDS, 14.5 mM 2-β-mercaptoethanol and a trace of bromophenol blue.

Tris-glycine-SDS-PAGE running buffer (5x): 250 mM Tris base, 2 M glycine, 0.037% (w/v) SDS pH 8.3.

Brilliant Blue® staining solution: 0.25% brilliant blue® (w/v) in 45% methanol, 10% acetic acid and ddH_2O.

Destaining solution: 45 % methanol, 10 % acetic acid in ddH_2O.

5x Tris-glycine-SDS transfer buffer (for high molecular weight protein transfer): 48 mM Tris base, 380 mM glycine, 0.037 % (w/v) SDS, 20 % (v/v) methanol.

5x EMSA binding buffer: 50 mM Tris-HCL, 250 mM NaCl, 1 mM DTT, pH 8.

2.5.2 Preparation of whole cell extracts (for protein extraction of MOZ)

Cells were grown as described in chapter 2.4, harvested by centrifugation at 2,000 rpm for 5 min at 4 °C and washed twice with PBS. Cells were resuspended in appropriate volumes of Phosphate MOZ-lysis buffer, incubated for 1 h on ice, frozen at -80 °C overnight and thawed on ice again. The freeze-thaw-cycle was repeated two times. Samples were sonicated in a water bath sonicator to ensure complete lysis, shearing of DNA and to detach MOZ protein from chromatin (3x 20 sec on/30 sec off). Cellular debris was pelleted by centrifugation at 12,000 rpm at 4 °C for 15 min and the supernatant stored at -80 °C.

2.5.3 Preparation of histone extracts

Cells were harvested as described above and extracted using the EpiQuik Total Histone Extraction Kit according to the manufacturer's protocol.

2.5.4 Small-scale bacterial expression tests of His-tagged proteins

BL21-DE3 bacterial cells were freshly transformed and tested for expression within one week of transformation. Single colonies were inoculated into 2.5 ml of LB-Ampicillin medium and grown for 2 h at 37 °C and 220 rpm shaking. 2 ml were divided between two microcentrifuge tubes and one was induced with 25 µl of 20 mM IPTG (final concentration 0.5 mM) while 25 µl of water was added to the other tube. The

cultures were incubated for further 2 h at 37 °C and 220 rpm shaking, harvested by centrifugation for 5 min at 13,000 rpm, resuspended in 15 µl of 2x SDS loading buffer and boiled for 5 min at 100 °C. Induced and uninduced samples were run alongside on a SDS-PAGE gel followed by Coomassie brilliant blue staining to detect the presence of an induced band which indicates expression of the protein of interest.

The remaining 0.5 ml of bacterial culture were stored for later inoculation of large-scale cultures and preparation of glycerol stocks of well expressing colonies.

2.5.5 Large-scale bacterial expression of His-tagged proteins

Glycerol stocks or the remaining 0.5 ml of bacterial culture of well expressing colonies were used to inoculate 50 ml LB medium containing the required antibiotics and grown O/N at 37 °C and 220 rpm shaking. The following day, the O/N cultures were used to inoculate 75 ml – 2 L LB medium containing the required antibiotic in a final dilution of 1:100. The cultures were further grown at 37 °C and 220 rpm shaking until the cells reached the exponential phase at an OD_{600} = 0.6-0.7. Optionally, a sample was taken representing the uninduced fraction. After 15 min incubation on ice, IPTG was added to a final concentration of 0.6 mM and the cultures were incubated for further 4.5 h at 30 °C and 220 rpm shaking (conditions depended on the protein). Optionally, a sample was taken representing the induced fraction. Bacteria were harvested by centrifugation at 4,000 rpm for 20 min at 4 °C and the pellet stored at -20 °C or -80 °C until protein purification.

2.5.6 Small-scale purification of 6xHis-tagged proteins

For small-scale purifications of 6xHis-tagged proteins, 75-250 ml of bacterial cultures were used and purified with HisPur™ Ni-NTA Resin or HisPur™ Cobalt Resin (Thermo Scientific). Cells were resuspended in 4 ml lysis buffer and incubated for 30-60 min at 4 °C on a rolling shaker for digest by lysozyme. Cells were then sonicated for 5 min (10 s on/10 s off) and the cell debris cleared by centrifugation at 9,000 rpm for 30 min. 50 µl of Nickel-NTA or Cobalt beads were equilibrated 2x with 100 µl lysis

buffer. The supernatant was then transferred to a 15 ml falcon tube, mixed with the beads and incubated for 60 min – 16 h at 4 °C on a rolling shaker. The beads were harvested by centrifugation at 700 *xg* for 2 min at 4 °C and washed 5x with wash buffer. The protein was eluted from the beads with 50 µl elution buffer, this step was repeated two times and the supernatant containing the purified protein stored at - 80 °C. SDS-PAGE and a Bradford assays were used to check the quantity and quality of the purified protein. Optionally, 10 µl samples of the lysis, flow-through and wash fractions were taken for further analysis.

2.5.7 Large-scale purification of 6xHis-tagged proteins

For large-scale purifications of His-tagged proteins, 2 L of bacterial culture were used. Pellets were resuspended in 35 ml lysis buffer and incubated for 30 – 60 min at 4 °C on a rolling shaker. Cells were then sonicated for 6 min (20 s on/30 s off), this step was repeated until lysis and shearing of DNA was complete. The cell debris was then cleared by centrifugation at 9,000 rpm – 13,000 rpm for 60-90 min. The supernatant was passed through a 0.45 nm filter using a syringe and kept on ice until the column was prepared. A 5 ml HiTrap® Chelating HP column (GE Healthcare) with an ÄKTA-Start system was used for IMAC affinity purification. All steps were carried out at the recommended column flow rate of 5 ml/min and 4 °C. The column was loaded with 50 mM $NiSO_4$ and equilibrated with 5 column volumes (CV) lysis buffer. The column was loaded with the bacterial lysate, washed with 15 CV of wash buffer and the protein eluted in 6 CV of elution buffer and collected in 1.5 ml fractions. The fractions were analysed by Bradford assay and SDS-PAGE for protein quantity and quality. All samples were stored at - 80 °C.

For additional high-salt washes (discussed in Chapter 1) following the initial wash, the column was washed with a linear gradient of low- to high-salt buffer. The initial wash was then repeated until the conductivity reached a baseline level to remove excess salt from the column.

For additional purification using a HiTrap® Heparin HP column (GE Healthcare) (discussed in chapter 4), the product of the standard purification was diluted in dilution buffer to a final salt concentration of 175 mM NaCl and was directly loaded on the equilibrated heparin column. The column was washed with low-salt heparin buffer and eluted increasing the salt concentration from low- to high-salt heparin buffer through a linear gradient.

2.5.8 Protein concentration determination

Protein concentrations were determined using the Bradford assay with Bio-rad dye. Typical reactions were set up in duplicate with 800 µl ddH_2O, 2 µl sample and 200 µl Bradford reagent (final concentration of 20%). The reactions were thoroughly mixed and incubated at RT for 20 min. The absorbance was read at 595 nm. The protein concentration was calculated relative to diluted Bovine Serum Albumin (BSA) standards of known concentrations.

2.5.9 Sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

One-dimensional SDS-PAGE was used to analyse protein concentrations and cell lysates. Gels were prepared with the appropriate percentage of resolving gel (6 - 18%) depending on the molecular weight of the protein and 6% stacking gels as described below (**Table 7**). Samples were denatured for 5 min at 100 °C in SDS-PAGE loading buffer and loaded onto the gel alongside known protein standards (3 µl of Colour Prestained/Blue Protein Standard, Broad Range, NEB). Gels were run in minigel systems (Protean IV, Bio-Rad) in 1x Tris-glycine-SDS-PAGE running buffer at a constant voltage of 150 - 200 V for 1 – 2 h until the desired separation was reached.

Table 7 Composition of stacking and resolving gels for SDS-PAGE.

The volumes are stated for the production of one 5 ml resolving gel and a 2 ml stacking gel.

% Acrylamide	8%	15%	18%	Stacking
ddH_2O (ml)	2.3	1.1	0.6	1.4
30 % (v/v) acrylamide mix (ml)	1.33	2.50	3.00	0.33
1.5 M Tris pH 8.8 (ml)	1.3	1.3	1.3	-
1.0 M Tris pH 6.8 (ml)	-	-	-	0.25
10% (w/v) SDS (ml)	0.05	0.05	0.05	0.02
10% (w/v) APS (ml)	0.05	0.05	0.05	0.02
TEMED (ml)	0.003	0.002	0.002	0.002

2.5.10 Staining of SDS-PAGE gels

After running, gels were incubated in Brilliant Blue staining solution for 30 min, washed 2x in ddH_2O and incubated in destaining solution for 30 min twice. For complete destaining, gels were incubated overnight in ddH_2O under mild shaking.

2.5.11 Western Blotting

Proteins were separated with SDS-PAGE and transferred onto nitrocellulose membranes in wet transfer cassettes by electrophoresis in 1x Tris-glycine-SDS transfer buffer either for 1.5 h at RT and 60 V or O/N at 4 °C and 30 V. Membranes were stained in 1% Ponceau S solution to check for transfer efficiency. The membranes were blocked for 1 h at RT in blocking buffer and incubated with primary antibodies (**Table 8**) O/N at 4 °C. Membranes were washed three times for 5 min in 1x PBS containing 0.1% Tween 20 (PBST). Membranes were incubated with the appropriate secondary antibodies (HRP-conjugated antibodies, 1:2500 dilution) for 60 min at RT. Secondary antibodies were washed of three times for 5 min in 1x PBST. Protein-antibodies were detected by ECL chemiluminescence. Membranes were incubated with 400 µl developing solution which was then dried off and the

membrane covered in saran wrap. A luminescent image analyser (Fujifilm Las-4000) was used to detect the chemiluminescent signal.

Table 8 List of primary antibodies

Primary antibody	Dilution	Supplier
Anti-MOZ (29D2)	1:500	In house antibody
Anti-FLAG	1:500	Sigma-Aldrich (F3165)
Anti-H3K9ac	1:500	Abcam (ab10812)
Anti-H3K14ac	1:500	Upstate
Anti-H3K14cr	1:500	PTM BIO (PTM-535)
Anti-penta-His	1:500	Qiagen (34660
Anti-GFP	1:500	Abcam (ab6556)
Anti-H2A	1:500	Abcam (ab18255)

2.5.12 Electrophoretic mobility shift assay (EMSA)

EMSAs were used to analyse protein-DNA interactions. 270 ng PCR product, 4 µl Invitrogen 1kb plus DNA ladder or 100-200 pmol of Oligonucleotides were mixed with titrations of the assayed protein in binding buffer as described in **Table 9**. The reactions were incubated for 30 min at RT, mixed with 2 µl of 50% glycerol or 10x DNA loading dye and loaded against no protein controls onto 6-8% 0.5x TBE-polyacrylamide gels, which were pre-run for 60 min at 100 V. Gels were run in minigel systems (Protean IV, Bio-Rad) in 0.5x TBE buffer at 100 V for 60-90 min, stained with a 0.5 µg/ml EtBr solution for 15 min and imaged in an ultraviolet transilluminator (BioRad Gel Doc 2004) with BioRad Quantity One software.

Table 9 EMSA binding reaction mix (18 µl)

5x binding buffer	3.6 µl
DNA	Variable, 100 pmol to 750 ng
Protein	titration
ddH_2O	ad 18 µl

2.6 Bioinformatics

2.6.1 Sequence analysis

Protein and DNA sequences were analysed and aligned using the Expasy based software tool *MultiAlin* (Corpet, 1988), and Benchling software.

2.6.2 Protein structure predictions

Amino acid sequences of the MOZ N-terminus from different species as well as MORF were used as inputs for secondary structure predictions and 3D-modelling using the Protein Homolog/analogy Recognition Engine V 2.0 (Phyre2; Kelley et al., 2015) software and RaptorX software (Källberg et al., 2012).

2.6.3 Statistics

Generally, data are expressed as mean ± standard error of the mean (SEM) and Mann-Whitney-U was used for non-parametric statistical analysis after evaluation of sample size and Shapiro-Wilk test to determine normality of data. p-values < 5% were considered as statistical significant; *= $p<0.05$, **= $p<0.01$, ***= $p<0.001$.

2.6.4 Enrichment analysis

Data and analysis was provided by Cintia Monteiro, David Heery and Nigel Mongan. Analysis was performed with the following softwares: Molecular signature database (MSigDB; Subramanian et al., 2005; Liberzon et al., 2011), Motif based-sequence analysis tools (MEME Suite; Bailey et al., 2009; McLeay et al., 2010) and Regulatory sequence analysis tools (RSAT; Nguyen et al., 2018).

Chapter 3: Structural predictions and sequence analysis of the MOZ N-terminus

The aim of this project was to investigate the structure and function of the MOZ N-terminus. Structural predictions and sequence analysis were first used to gain a general understanding of the protein region and to estimate potential functions. The analysed regions spanned the first ~200 amino acids (AA) of MOZ and included the NEMM (1-170) domain comprising the H1/5 (85-170) domain, which was found to have weak similarity to linker histones H1 and H5 (Yang and Ullah, 2007), but excluded the DPF domain and further C-terminal regions of MOZ.

3.1 Sequence conservation, secondary structure predictions and 3D-models

The Expasy based software tool *MultiAlin*, which implements hierarchical clustering, was used for sequence alignments (Corpet, 1988). The first ~200 AA of human KAT6A/MOZ were aligned with MOZ proteins from different species as well as the human paralog KAT6B/MORF to analyse sequence conservation which might give an indication of regions of functional importance (**Figure 3.1 A**). Alignments showed very high sequence conservation of the first 92 AA across species, with >95% identity observed in mammals, reptiles and bird species. In addition, human KAT6A regions AA 12-35 and AA 62-75 were also highly conserved in fish and insects, but not in fungi. In contrast, the following region AA 93-200 shared only ~75% identity in mammals, reptiles and birds. Especially the regions AA 132-142 as well as AA 171-185 of human KAT6A were only weakly conserved. Human KAT6B/MORF also shows high sequence similarities with human MOZ for the first 79 AA (89%) but weaker identity with the following region. Overall, the conservation across evolution suggests that the MOZ N-terminus, especially the first part up to approximately AA 90, has structural or functional importance.

Structural predictions were implemented to investigate potential functions of the NEMM and H1/5 domains of MOZ. The Protein Homology/analogy Recognition Engine V 2.0 (Phyre2; Kelley et al., 2015) was used to create predictions of the secondary structure as well as a 3-D model of the first ~200 N-terminal amino acids of MOZ that did not include the DPF domain (AA 209-310) or other C-terminal components of MOZ. After analysis in “normal mode” created multiple hits with >90% confidence, “intensive mode" was used to create a combinatorial model based on the 20 top hits (**Figure 3.1 B, C, D**).

The first input was spanning the sequences 1-201 which included the NEMM domain containing the H1/5 domain. Several structures were predicted with a confidence of >90% that mainly included DNA-binding domains. The predictions indicated two separated domains of approximately AA 7-70 and 90-180. Five predictions also spanned the area approximately AA 7-130 (**Figure 3.1 B**). Based on this result, it could be estimated that two individual domains may be present in the MOZ N-terminus and consequently, both regions were subjected to Phyre2 separately for detailed analysis.

Predictions for the region spanning approximately AA 7-75 were generated based on 17 alignments with a score >90% (**Figure 3.1 C, D**). Regions N-terminal and C-terminal of this area were disordered, indicating that it was not possible to model them properly. The predicted structure contains three α-helices (H1, H2 and H3) connected by short loops (L1 and L2) followed by two β-strands (β1 and β2) which are connected by another loop (L3). Some models also included a small C-terminal α-helix (H4). The sequence 22-28 showed moderate disorder levels for the secondary structure and might be located in the loop between H1 and H2 instead of the H1 (**Figure 3.1 C**).

The second predicted structure was contained in the approximate region AA 90-166 (**Figure 3.1 C, D**). It showed a similar overall structure with three α-helices (H1, H2 and H3) connected by loops (L1 and L2) and one or two β-strands (β1 and β2) depending on the predictions which were also connected by a small loop (L3). This structure spanned the region AA 99-166 with N- and C- terminal ends again showing

high levels of disorder. Moderate disorders were also located at the beginning of H2 and H3.

Both structures were mainly predicted in alignment with DNA-binding domains and associated with 3-helical bundles from the linker histone H1/5 family in the winged-helix domain (WHD) superfamily with >90% confidence scores. Both were highly similar to the characteristic organisation of WHDs, H1-β1-H2-T/L-H3-β2-L/W1-β3-L/W2, and since most predictions were based on WHD motifs, the two predicted structures will be referred to as predicted Winged-Helix Domains 1 and 2 (pWHD1 and pWHD2) during this study. pWHD1 is approximately located in the region 7-75 and pWHD2 around 90-184.

In addition, RaptorX software (Källberg et al., 2012) was also used to generate template-based protein models which produced results similar to Phyre2. The secondary structure prediction matched those suggested by Phyre2 almost exactly with exception of single amino acid residues that were annotated differently and regions that showed higher disorder levels (**Figure 3.1 C**). Thus, the overall prediction of two winged helix folds in the MOZ N-terminus was supported by RaptorX analysis and also represented in 3D-models (**Figure 3.1 D**).

The majority of predictions for both domains were based on the globular domain of the linker histones H1 and H5 from human and yeast. A variety of structures were available for predictions in H1 and H5 (Ali et al., 2004, PDB-ID: 1UST; Cerf et al., 1994, PDB-ID: 1GHC; Eletsky et al., unpublished, PDB-ID: 2LSO; Ono et al., 2003, PDB-ID: 1UHM; Ramakrishnan et al., 1993, PDB-ID: 1HST; Zhou et al., 2015, PDB-ID: 4QLC; Bednar et al., 2017, PDB-ID: 5nl0) as well as the H1-like region of Heterochromatin Protein 1-binding protein 74 (HP1-BP74) (Hayashihara, et al., 2010, PDB-ID: 2RQP).

In addition, several pWHD1 predictions were based on alignments with transcriptional regulators that contain WHDs. One resulted from alignment with the WHD of Arabinose Repressor (AraR) from *Bacillus subtilis* (Jain and Nair, 2013, PD-ID: 4EGZ). Two predictions were based on WHDs from GntR (gluconate repressor)

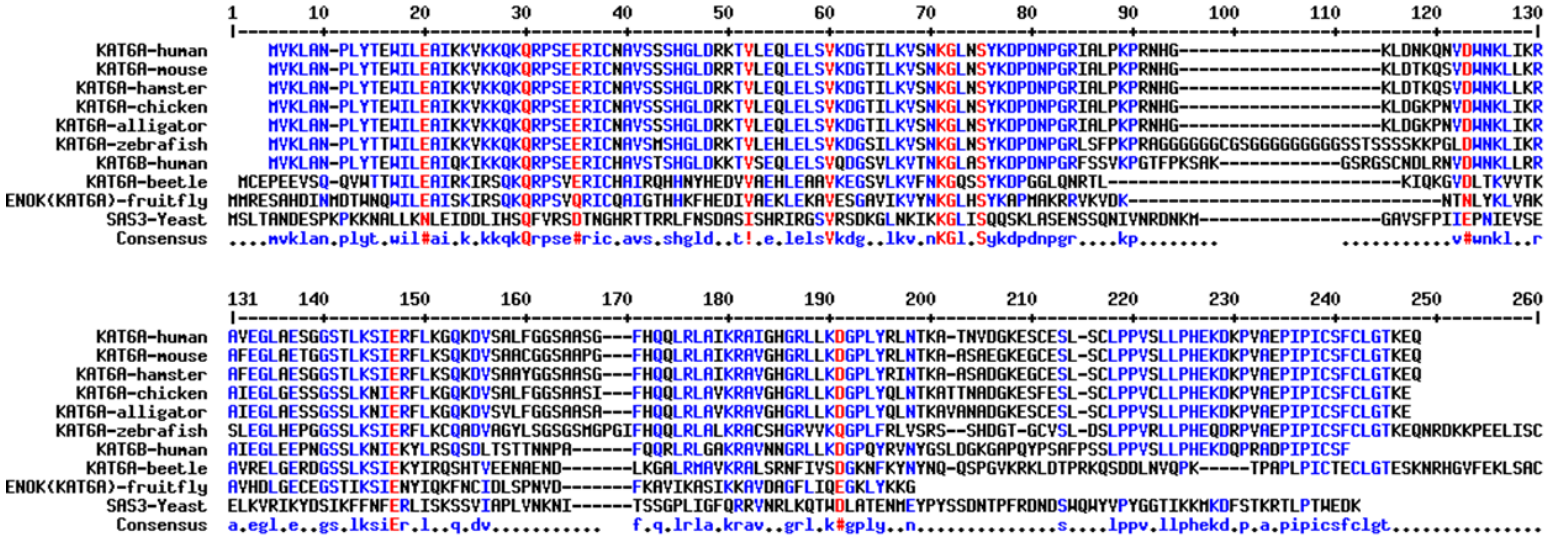
transcriptional regulators, a GntR-like bacterial transcription factor from *Listeria monocytogenes* (Osipiuk et al., unpublished; PDB-ID: 4R1H) and the transcriptional repressor DasR (Fillenberg et al., 2016; PDB-ID: 4ZS8). Other structures were based on WHD containing transcriptional regulators from *E. coli* (Brunzelle et al., unpublished; PDB-ID: 5KVR), *Bacillus subtilis* (Tan et al., unpublished; 3BWG), *Listeria monocytogenes* (Kim et al., unpublished; PDB-ID: 4HAM) and *Oenococcus oeni* (Zhang et al., unpublished; PDB-ID: 3BYZ6). Further alignments were generated with the WHD of the minichromosome maintenance protein from *Methanothermobacter thermautotrophicus* (Wiedemann et al., 2015; PDB-ID: 2MA3) and the DNA-binding WHD of *E. coli* Zur which is involved in zinc regulation (Gilston et al., 2014; PDB-ID: 4mtd).

pWHD2 showed additional sequence similarities with different members of the FOX protein family, the most prominent member of WHD-proteins formed by many TFs that contain a characteristic WHD. Predictions were based on several available structures from members of the family (e.g. Brent et al., 2008, PDB-ID: 3CO7; Weigelt et al., 2000, PDB-ID: 1E17; Tsai et al., 2006, PDB-ID: 2C6Y). Only the H1 and H2 of pWHD2 aligned with FOX proteins but not the H3 and wing motif which typically form the contacts to DNA in this protein family. Another prediction resulted from sequence alignment with a WHD of an ATPase from *Sulfolobus solfacaricus* (Xu et al., 2009, PDB-ID: 2FNA).

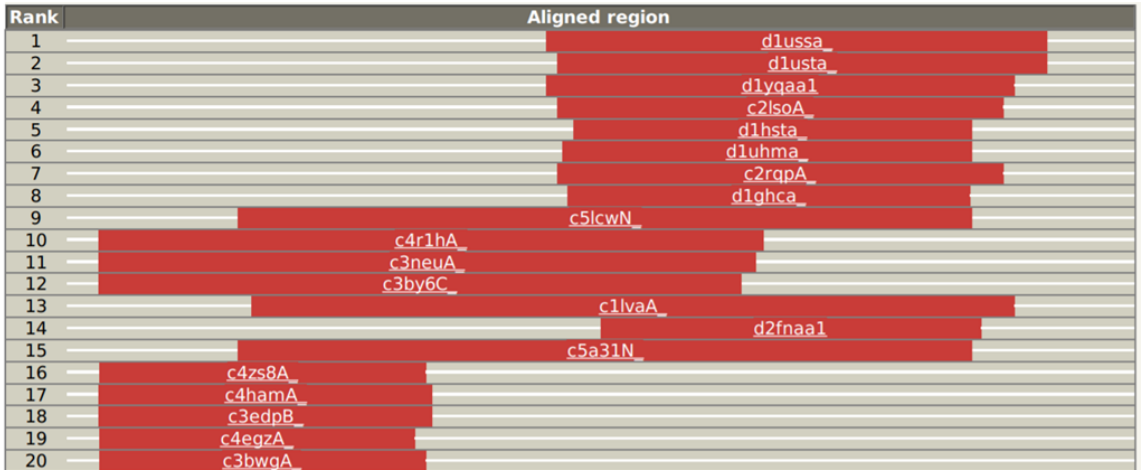
Five predicted structures were found to span the regions of both pWHD1 and pWHD2 (**Figure 3.1 B**). Two hits (AA 33-170) were based on the crystal structure of APC/C which is a ubiquitin ligase mainly involved in cell-cycle regulation. Structural similarity was shown for subunit 2, which contains a WHD that interacts with other proteins in holo-protein formation (Alfieri et al., 2016, PDB-ID: 5LCW; Chang et al., 2015, PDB-ID: 5A31) and aligned with pWHD2. Two predicted structures were again based on the proposed transcriptional regulators from *Listeria monocytogenes* (Osipiuk et al., unpublished; PDB-ID: 4R1H) and *Oenococcus oeni* (Zhang et al., unpublished; PDB-ID: 3bY6) which aligned with pWHD1. In this case, the predictions

extended the pWHD1 region and contained regions of the pWHD2. Interestingly, one prediction (AA 36-178) was based on the Selenocysteine elongation factor B (SelB) from *Moorella thermoacetica* which contains four WHDs in an L-shape with two of them aligning with pWHD1 and pWHD2. Both WHDs 3 and 4 were suggested to bind to mRNA (Selmer and Su, 2002; PDB-ID: 1LVA).

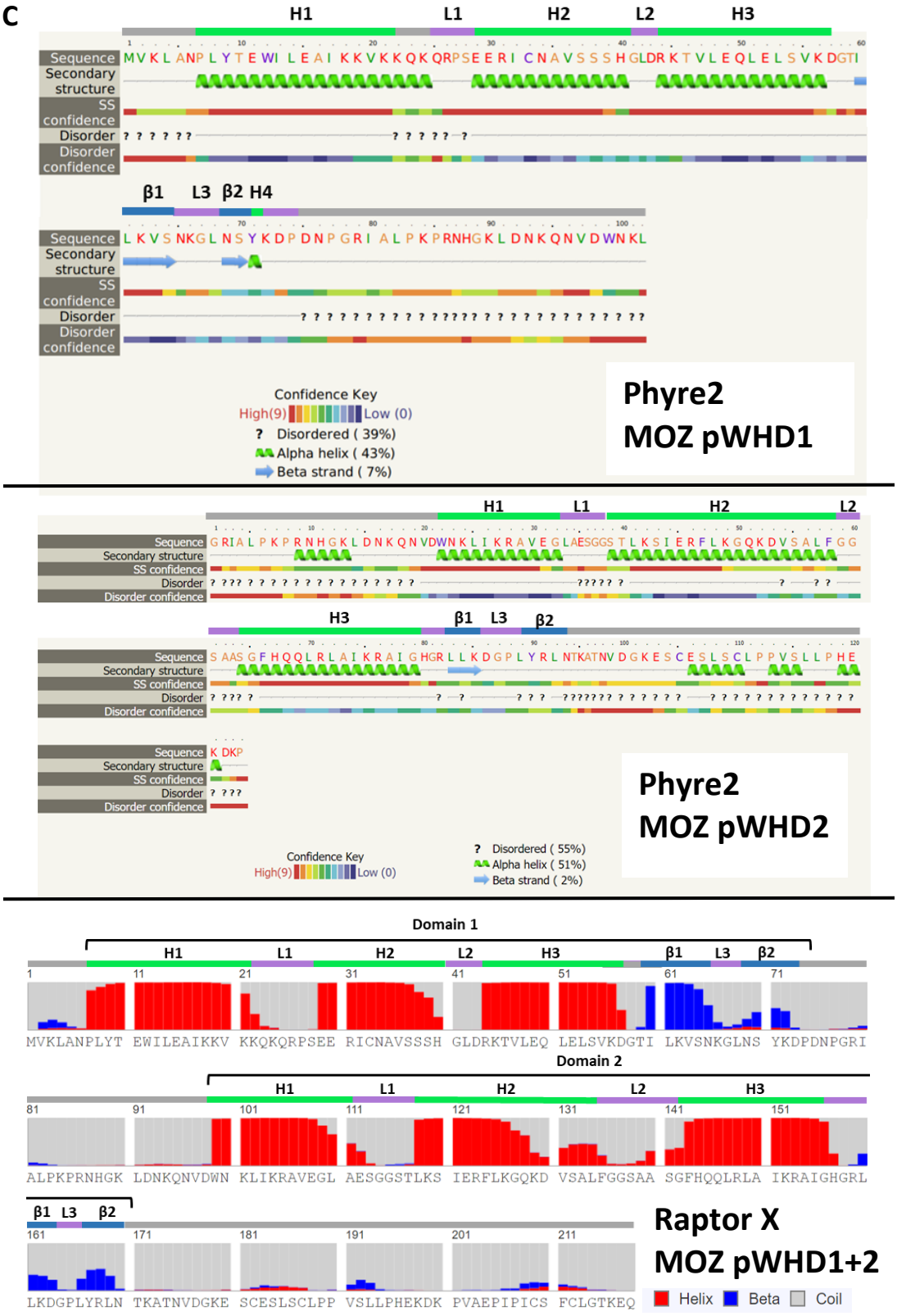
A


1 10 20 30 40 50 60 70 80 90 100 110 120 130
KAT6A-human MVKLAN-PLYTEWILEAIKKVKKQKQRPSEERICNAVSSSHGLDRKTVLEQLELSVKDGTILKVSNKGLNSYKDPDNPGRIALPKPRNHG----------KLDNKQNVDWNKLIKR
KAT6A-mouse MVKLAN-PLYTEWILEAIKKVKKQKQRPSEERICNAVSSSHGLDRRTVLEQLELSVKDGTILKVSNKGLNSYKDPDNPGRIALPKPRNHG----------KLDTKQSVDWNKLLKR
KAT6A-hamster MVKLAN-PLYTEWILEAIKKVKKQKQRPSEERICNAVSSSHGLDRKTVLEQLELSVKDGTILKVSNKGLNSYKDPDNPGRIALPKPRNHG----------KLDTKQSVDWNKLLKR
KAT6A-chicken MVKLAN-PLYTEWILEAIKKVKKQKQRPSEERICNAVSSSHGLDRKTVLEQLELSVKDGTILKVSNKGLNSYKDPDNPGRIALPKPRNHG----------KLDGKPNVDWNKLIKR
KAT6A-alligator MVKLAN-PLYTEWILEAIKKVKKQKQRPSEERICNAVSSSHGLDRKTVLEQLELSVKDGTILKVSNKGLNSYKDPDNPGRIALPKPRNHG----------KLDGKPNVDWNKLIKR
KAT6A-zebrafish MVKLAN-PLYTTWILEAIKKVKKQKQRPSEERICNAVSMSHGLDRKTVLEHLELSVKDGSILKVSNKGLNSYKDPDNPGRLSFPKPRAGGGGGGCGSGGGGGGGGGSSTSSSSKKPGLDWNKLIKR
KAT6B-human MVKLAN-PLYTEWILEAIQKIKKQKQRPSEERICHAVSTSHGLDKKTVSEQLELSVQDGSVLKVTNKGLASYKDPDNPGRFSSVKPGTFPKSAK----------GSRGSCNDLRNVDWNKLLRR
KAT6A-beetle MCEPEEVSQ-QVWTTWILEAIRKIRSQKQRPSVERICHAIRQHHNYHEDVVAEHLEAAVKEGSVLKVFNKGQSSYKDPGGLQNRTL----------KIQKGVDLTKVVTK
ENOK(KAT6A)-fruitfly MMRESAHDINMDTWNQWILEAISKIRSQKQRPSVQRICQAIGTHHKFHEDIVAEKLEKAVESGAVIKVYNKGLHSYKAPMAKRRVKVDK----------NTNLYKLVAK
SAS3-Yeast MSLTANDESPKPKKNALLKNLEIDDLIHSQFVRSDTNGHRTTRRLFNSDASISHRIRGSVRSDKGLNKIKKGLISQQSKLASENSSQNIVNRDNKM----------GAVSFPIIEPNIEVSE
Consensus ....mvklan.plyt.wil#ai.k.kkqkQrpse#ric.avs.shgld..t!.e.lelsVkdg..lkv.nKGl.Sykdpdnpgr....kp..... .....v#wnkl..r
131 140 150 160 170 180 190 200 210 220 230 240 250 260
KAT6A-human AVEGLAESGGSTLKSIERFLKGQKDVSALFGGSAASG---FHQQLRLAIKRAIGHGRLLKDGPLYRLNTKA-TNVDGKESCESL-SCLPPVSLLPHEKDKPVAEPIPICSFCLGTKEQ
KAT6A-mouse AFEGLAETGGSTLKSIERFLKSQKDVSAACGGSAAPG---FHQQLRLAIKRAVGHGRLLKDGPLYRLNTKA-ASAEGKEGCESL-SCLPPVSLLPHEKDKPVAEPIPICSFCLGTKEQ
KAT6A-hamster AFEGLAESGGSTLKSIERFLKSQKDVSAAYGGSAASG---FHQQLRLAIKRAVGHGRLLKDGPLYRINTKA-ASADGKEGCESL-SCLPPVSLLPHEKDKPVAEPIPICSFCLGTKEQ
KAT6A-chicken AIEGLGESSGSSLKNIERFLKGQKDVSALFGGSAASI---FHQQLRLAVKRAVGHGRLLKDGPLYQLNTKATTNADGKESFESL-SCLPPVCLLPHEKDKPVAEPIPICSFCLGTKE
KAT6A-alligator AIEGLAESSGSSLKNIERFLKGQKDVSVLFGGSAASA---FHQQLRLAVKRAVGHGRLLKDGPLYQLNTKAVANADGKESCESL-SCLPPVSLLPHEKDKPVAEPIPICSFCLGTKE
KAT6A-zebrafish SLEGLHEPGGSSLKSIERFLKCQADVAGYLSGSGSMGPGIFHQQLRLALKRACSHGRVVKQGPLFRLVSRS--SHDGT-GCVSL-DSLPPVRLLPHEQDRPVAEPIPICSFCLGTKEQNRDKKPEELISC
KAT6B-human AIEGLEEPNGSSLKNIEKYLRSQSDLTSTTNNPA------FQQRLRLGAKRAVNNGRLLKDGPQYRVNYGSLDGKGAPQYPSAFPSSLPPVSLLPHEKDQPRADPIPICSF
KAT6A-beetle AVRELGERDGSSLKSIEKYIRQSHTVEENAEND-------LKGALRMAVKRALSRNFIVSDGKNFKYNYNQ-QSPGVKRKLDTPRKQSDDLNVQPK-----TPAPLPICTECLGTESKNRHGVFEKLSAC
ENOK(KAT6A)-fruitfly AVHDLGECEGSTIKSIENYIQKFNCIDLSPNVD-------FKAVIKASIKKAVDAGFLIQEGKLYKKG
SAS3-Yeast ELKVRIKYDSIKFFNFERLISKSSVIAPLVNKNI------TSSGPLIGFQRRVNRLKQTWDLATENMEYPYSSDNTPFRDNDSWQWYVPYGGTIKKMKDFSTKRTLPTWEDK
Consensus a.egl.e..gs.lksiEr.l..q.dv..... f.q.lrla.krav..grl.k#gply..n.....s....lppv.llphekd.p.a.pipicsfclgt.....

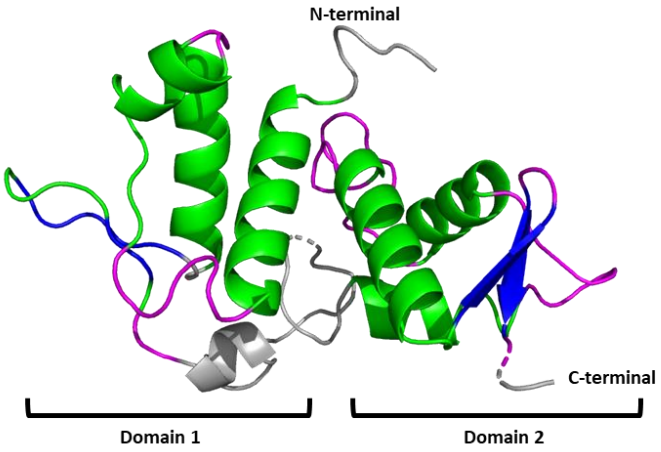
B


Rank
Aligned region
1 d1ussa_
2 d1usta_
3 d1yqaa1
4 c2lsoA_
5 d1hsta_
6 d1uhma_
7 c2rqpA_
8 d1ghca_
9 c5lcwN_
10 c4r1hA_
11 c3neuA_
12 c3by6C_
13 c1lvaA_
14 d2fnaa1
15 c5a31N_
16 c4zs8A_
17 c4hamA_
18 c3edpB_
19 c4egzA_
20 c3bwgA_

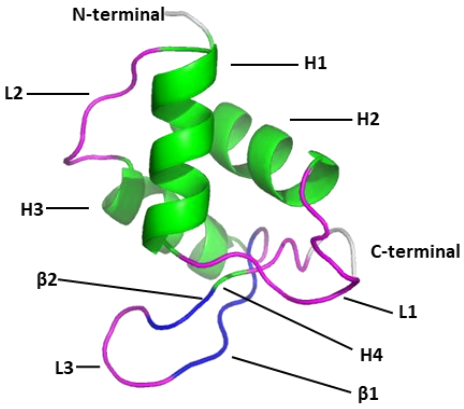
C

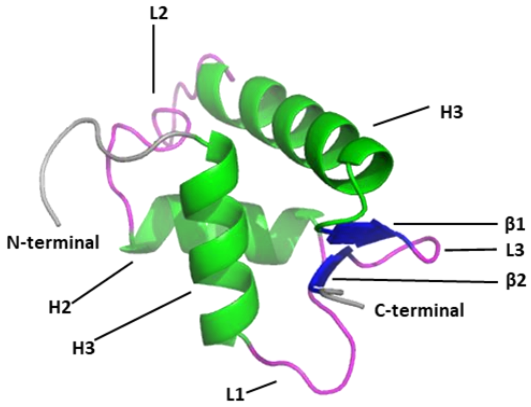

H1
L1
H2
L2
H3
Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence
β1
L3
β2
H4
Confidence Key
High(9) Low (0)
? Disordered (39%)
Alpha helix (43%)
Beta strand (7%)
Phyre2
MOZ pWHD1
? Disordered (55%)
Alpha helix (51%)
Beta strand (2%)
Phyre2
MOZ pWHD2
Domain 1
Domain 2
Raptor X
MOZ pWHD1+2
Helix
Beta
Coil

D **Phyre2** **MOZ pWHD1+2**



MOZ pWHD1 **MOZ pWHD2**





Raptor X **MOZ pWHD1+2**

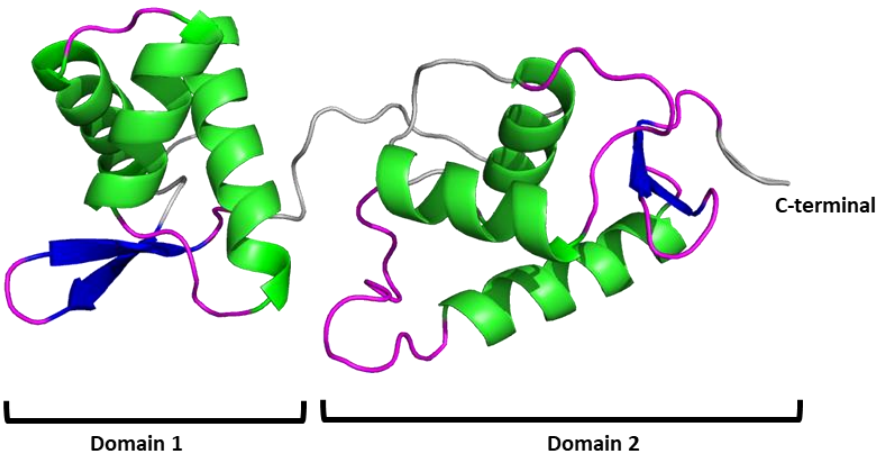


Figure 3.1 Sequence alignment and structural predictions of the MOZ NEMM and H1/5 domains

A) Sequence alignment of the first 201 AA of MOZ from different species and human MORF using *MultiAlin* (Corpet et al., 1988) software. **B)** Aligned regions of the MOZ-terminus by *Phyre2* (Kelley et al., 2015) for the first 201 AA of MOZ. **C)** Secondary structure predictions for pWHD1 and pWHD2 by *Phyre2* (top) and pWHD1+2 by RaptorX (bottom, Källberg et al., 2012). **D)** Predictions of the 3D-model for pWHD1+2, pWHD1 and pWHD2 by Phyre2 (top) and for pWHD1+2 by *RaptorX* (bottom).
Green=α-helix; Blue=β-strand; Magenta=loops; ?/grey=disordered prediction; C- and N-terminal ends are indicated. Highly disordered C- and N-terminal tails were clipped in 3-D models.

3.2 Summary

Structural predictions generated with Phyre 2 (Kelley et al., 2015) and RaptorX (Källberg et al., 2012) software indicated the presence of two potential WHDs in a tandem arrangement within the N-terminal 200 amino acids of MOZ. As described in Chapter 1.1.3, WHDs constitute a large family that appears to have evolved to have diverse functions. Proteins of this family predominantly act in DNA-binding but WHDs were also found to form complexes with RNA and to facilitate protein-protein interactions. The predicted domains in the NEMM region of human MOZ are located between residues 7-75 (pWHD1) and 99-166 (pWHD2), the latter of which corresponds to the H1/5 regions of MOZ. Both predicted MOZ WHD structures and particularly WHD1, show a high degree of evolutionary conservation suggesting that orthologous MOZ proteins share conserved molecular functions throughout the vertebrates.

While most WHD-containing proteins such as FOX family TFs typically contain a single domain, several studies in the literature have shown that some proteins contain two or more WHDs. For example, tandem WHDs have been noted in the RNA polymerase subunit RPC34 (Wei and Chen, 2018) and RPC62 contains 4 WHDs (Lefèvre et al., 2011). Tandem WH folds were also found in the transcriptional regulator TFIIEβ (Vannini and Cramer, 2012), SMC5/6 (Structure Maintenance of Chromosomes 5/6) proteins (Palecek and Gruber, 2015) and in the yeast replication

licensing factor Cdt1 (Khayrutdinov et al., 2009). RNA-binding SelB proteins contain 4 WHDs at their C-termini and two WH folds were also found in structure predictions of the MOZ N-terminus which supports the possibility of a MOZ tandem WHD (Selmer and Su, 2002).

Most of the structural suggestions for the MOZ pWHDs were related to DNA-binding domains and are based on alignments with the linker histones H1 and H5. As described above, linker histones bind DNA in various modes because they have to interact with nucleosomal DNA as well as linker DNA to form contacts between the nucleosomes. Because MOZ complexes act as KATs and function in chromatin modification and transcriptional regulation, they must have some residence time in association with nucleosomes and DNA at precise targets within the genome. In support of this, a motif related to binding of flexible cruciform DNA was already identified in the MYST domain of MOZ and associated with nucleosome binding and positioning of the domain (Holbert et al., 2007). Furthermore, our group previously described the binding of the MOZ DPF domain to histone H3 with a strong preference for H3K14ac (Dreveny et al., 2014) which was further supported by studies of another group (Xiong et al., 2016). Multiple binding modes are also typical for other epigenetic regulators, e.g. PHF1, a member of the PRC2 complex. The protein contains histone binding tudor and DPF domains as well as a WHD that binds to DNA (Li et al., 2017). In this context, it seems plausible that further contacts of MOZ with chromatin could be formed by additional DNA-binding domains located in close proximity to the DPF domain as indicated by the predicted WHDs. In addition, previous polypeptide-based studies in our group showed that constructs containing the DPF domain as well as pWHD1 and pWHD2 (MOZ 1-321) appeared to show enhanced interaction with core histones compared to the DPF (172-321) on its own (Deeves and Heery, unpublished). Based on this result and the structural predictions, it seems very likely that the N-terminus of MOZ contains additional domains within the first 200 amino acid residues that act in DNA-binding and may be involved in recruitment of MOZ to chromatin target sites. We therefore aimed to further

investigate the potential DNA-binding ability of the MOZ pWHD1 and pWHD2. However, it is also possible that the MOZ N-terminus has other or additional roles, e.g. in protein-protein interaction which would also be consistent with the association of MOZ in multi-subunit protein complexes and the reported involvement of the first 79 MOZ AA in interactions of MOZ-TIF2 with PRMT1 and KDM4C (Cheung et al., 2016).

Chapter 4: Expression and purification of recombinant MOZ protein

4.1 Expression vectors of the MOZ N-terminal regions

Having determined that the MOZ N-terminus may comprise two potential WHDs, we next set out to investigate the functions of these domains using recombinant proteins in *in vitro* assays. To facilitate this, the MOZ pWHDs were recombinantly expressed in *E. coli* and purified using Immobilized Metal Ion Affinity Chromatography (IMAC). Three constructs were used that produce protein polypeptides containing either pWHD1 (1-88), pWHD2 (79-201) or both domains pWHD1+2 (1-201) (**Figure 4.3 A**).

The expression vector pET-15b (Novagen) was used for all protein expressions (**Figure 4.1 B**). The protein of interest is fused to a 6xHistidine-tag (6xHis-tag) that enables IMAC because it interacts with divalent ions, such as nickel or cobalt, in a chelating manner and allows selective binding of the protein of interest. To elute the protein, high concentrations of imidazole can be used which competitively binds to divalent ions and causes the release of the 6xHis-tag from the resin (Cheung et al., 2012). A cleavage site allows removal of the tag from the protein by thrombin-digest after purification. Protein expression is controlled by a T7-promoter and inducible with Isopropyl fβ-D-1-thiogalactopyranoside (IPTG) using the lac operon. The lac repressor is constitutively expressed by the *lacI* gene, binds to the operator located downstream of the promoter sequence and blocks transcription of the gene of interest. IPTG acts similar to lactose and allosterically binds to the repressor which releases the repressor from the operator and thus enables expression. This system allows controlled expression of the protein of interest upon IPTG induction (Bell and Lewis, 2000).

Expression using T7-promoters requires the host to contain the prophage λDE3 sequence which encodes the T7 RNA polymerase under control of an IPTG-

inducible *lac*UV5-promoter (Baneyx, 1999). The protease deficient *E. coli* strain BL21-DE3 was used for protein expression in this project.

The required expression constructs were previously generated in our lab using restriction enzyme-based cloning with XhoI and BamHI. However, the purification still had to be established for this project

The expression vectors were freshly mini-prepped from *E. coli* DH5α, transformed into *E. coli* BL21-DE3 cells and positive clones selected using LB agar plates containing ampicillin.

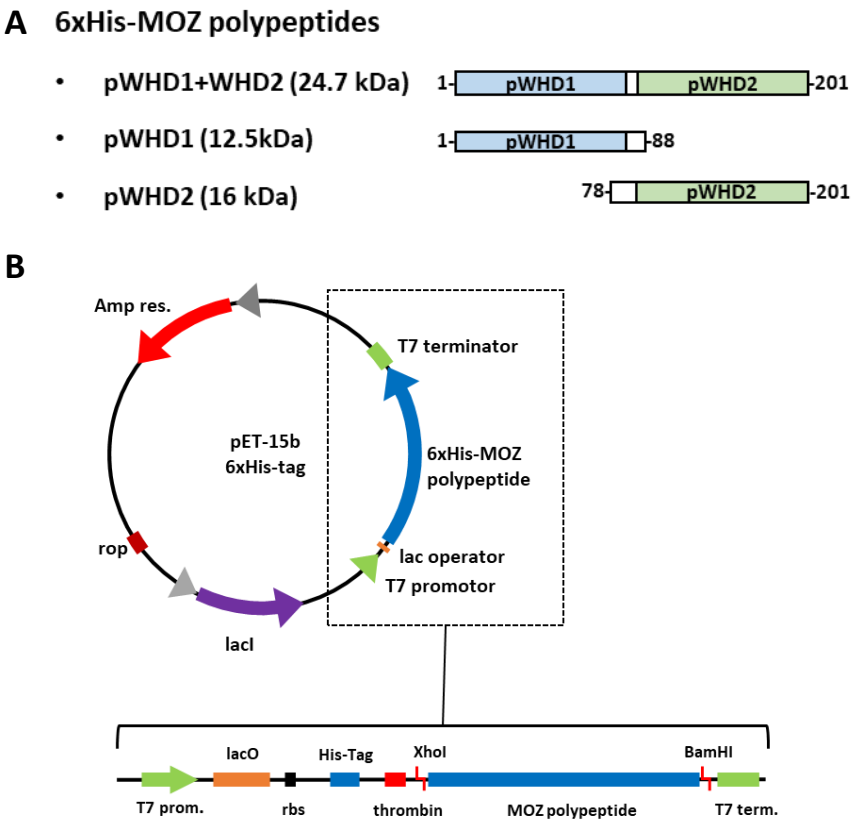


Figure 4.1 Schematic of 6xHis-MOZ polypetide expression vectors

A) Schematic of the expressed 6x-His-MOZ polypeptides containing the pWHD1, pWHD2 and pWHD1+2. **B)** The vector pET15-b (Novagen) was used in BL21-DE3 *E. coli* cells to express MOZ-polypeptides fused to an N-terminal 6xHis-tag. The vector contains a *lacI* gene and a lac operator to enable IPTG-induced protein expression by the T7 promoter. Ampicillin resistance was used for selection and a thrombin cleavage-site allows removal of the 6xHis-tag.

4.2 Small-scale expression and purification of MOZ-polypeptides

Small-scale expressions and purifications for the three MOZ 6xHis-pWHD constructs were optimised to the following protocol. Overnight cultures from freshly streaked colonies were used to inoculate 200 mL LB medium and grown at 37 °C until the cultures reached OD_{600} = 0.6 - 0.7. Cultures were then induced with 0.6 mM IPTG and incubated at 30 °C for 4.5 h. Cells were harvested by centrifugation, resuspended in lysis buffer (10 mM Tris-HCL, 300 mM NaCl, 10 mM Imidazole pH 8) and lysed by lysozyme digest and sonication (5x 10s on/10s off). The soluble protein was collected by centrifugation and the supernatant passed through a 0.45 nm filter. The protein lysates were incubated with equilibrated Ni-NTA resins (Thermo Scientific) and washed with wash buffer (10 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole pH 8) to remove unbound proteins before the 6xHis-tagged MOZ polypeptides were eluted in elution buffer (10 mM Tris-HCl, 300 mM NaCl, 300 mM Imidazole pH 7.5). The pH in the lysis and wash buffers was kept more basic to increase selectivity of the beads for the 6xHis-tag.

All three polypeptides were successfully expressed and purified as demonstrated by Coomassie staining and western blot (WB) (**Figure 4.2**). No commercial antibody was available for the polypeptides, so an anti-penta-His antibody was used for detection. As shown, 6xHis-pWHD2 generally yielded more product than the other constructs.

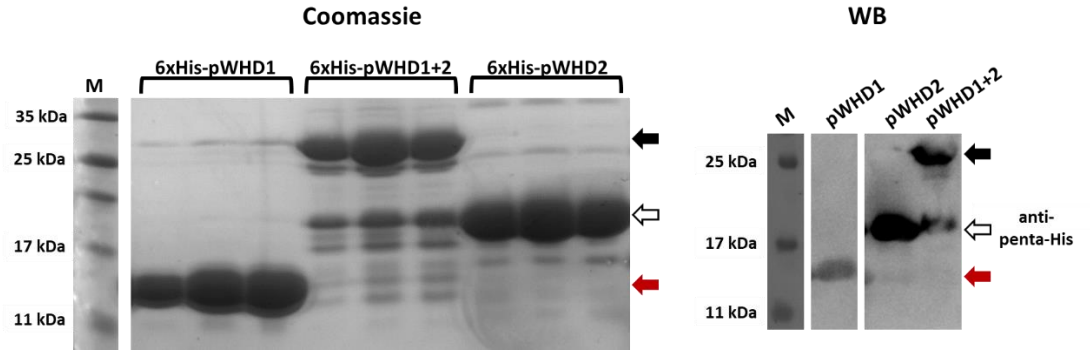


Figure 4.2 Small scale expression and purification of the 6xHis-MOZ polypeptides pWHD1, pWHD2, pWHD1+2

SDS-PAGE gel analysis by Coomassie staining (left) and western blot with anti-penta-His antibody (right) confirming purification of all 6xHis-MOZ polypeptides.

M=Marker; Arrows indicate the protein of interest. 10 µl purified product was used as inputs for Coomassie stains and 5 µl for WB.

4.3 Optimisation of large-scale purification of pWHD1

As described below, comprehensive analysis of the predicted domains in this study revealed the DNA-binding ability of the pWHD1 (Chapter 5). Additional biochemical methods such as Isothermal Titration Calorimetry (ITC) or X-ray crystallography could help to further characterise the domain. These experiments require high concentrations of homogenous protein which needs to be produced in large-scale. Even though these experiments exceeded the scope of this study and will have to take place in future projects, the purification process of the pWHD1 was still optimised.

For large-scale purifications, 2.5 L of bacterial culture were grown at 37°C until OD_{600}=0.6-0.7, induced with 0.6 mM IPTG and further incubated at 30 °C for 4.5 h. Cells were collected by centrifugation and lysed by lysozyme digest and sonication (5x 20s on/30 s off, prolonged when necessary). A HiTrap® Chelating 5 ml column (Amersham pharmacia biotech) in combination with an ÄKTA-Start system was used for large-scale protein purifications. The column was loaded with 50 mM NiSO4 and equilibrated with lysis buffer. The lysate was loaded onto the column, washed with

wash buffer and eluted in elution buffer. A strong peak for the elution was visible in the chromatogram and subsequent analysis with SDS-PAGE and Coomassie staining showed high concentrations of the protein of interest (**Figure 4.3 A**). While the first eluted fractions also contained background proteins, the product showed high homogeneity for later fractions. Bradford assays were used to determine the protein concentration with yields of around 45 mg of protein from 2.5 L of culture. This would be sufficient for downstream applications, but further analysis showed contamination of the protein product with DNA. Spectrophotometric analysis of the large-scale purification determined DNA concentrations of around 300 to 500 ng/µl with 260nm/280nm ratios around 1.9. Analysis by agarose gel electrophoresis and EtBr staining confirmed severe contamination with DNA. (**Figure 4.3 C**). Rather than DNA actually binding to the nickel-matrix, it appeared more likely that DNA co-purified because it was tightly bound to the protein of interest which shows affinity for DNA. This contamination was of no issue for the assays used in this study but would be problematic for more sensitive methods. After evaluation of different approaches, two methods were found to be sufficient for removal of contaminant DNA.

Heparin is a highly sulphated glycosaminoglycan that mimics the polyanionic structure of nucleic acids which allows DNA-binding proteins to bind. Generally, clusters with positively charges residues interact with heparin which also shows characteristics of a cation exchange resin (Bolten et al., 2018). In addition, different consensus sequences and motifs were identified that allow proteins to interact with heparin (Cardin and Weintraub, 1989; Sobel et al., 1992; Torrent et al., 2012). The MOZ pWHD1 was identified as a DNA-binding domain and the motif A-I-K-K-V-K-K-Q-K-Q-R-P contains a high amount of positively charged residues and resembles known heparin-interaction motifs suggesting that the protein would bind to the column. The protein-heparin interaction is most likely in competition with protein-DNA binding and should therefore remove remaining DNA from the protein of interest.

A 5 ml HiTrap® Heparin HP column was tested as a second purification step which typically requires low-salt conditions to allow binding of the protein of interest to the column since high ionic strength inhibits the interaction. The 6xHis-pWHD1 was first purified as described above using standard conditions with the nickel column. The product was then pooled and diluted with 20 mM Tris pH 7.5 to obtain a final salt concentration of 175 mM NaCl and loaded onto the heparin column which was equilibrated with low-salt heparin buffer (20 mM Tris, 175 mM NaCl pH 7.5). The column was then briefly washed with low salt buffer to remove any unbound protein. The protein of interest was eluted by increasing the salt concentration from 175 mM NaCl to 1.5 M NaCl through a linear gradient.

In addition to heparin, high salt concentrations can also be used to remove DNA. Typically, DNA binds to proteins or chromatography resins by interaction of negatively charged DNA components like the phosphate-backbone with positively charged residues, e.g. in hydrogen bonds. Salt ions (typically Na^+ and Cl^-) compete for these charged residues. With increasing ionic strength of the buffer more charged residues get covered by salt ions which destabilises DNA-binding and elutes DNA from the protein or column. Since the protein binds to the column in a chelating manner mediated by the 6xHis-tag, high-salt contents do not necessarily interrupt interaction of the protein of interest with the column resin (Vassylyeva et al., 2017).

Using higher salt concentrations (500 mM NaCl) in the lysis and wash buffer resulted in significant loss of the protein of interest (data not shown), which is a known problem from the literature (Vassylyeva et al., 2017). Consequently, a secondary wash step following the initial wash was implemented. In this step, the salt-concentration was increased from 300 mM NaCl to 1.5 M NaCl through a linear gradient followed by a repeat of the initial wash step to remove any excess salt from the column. The pH during the second wash step was reduced to 7.5 and the buffers contained no imidazole which reduces overall selectivity of the column and helps to avoid elution of the protein of interest.

Figure 4.3 B shows the direct comparison of the standard purification with an additional heparin purification as well as an additional high-salt wash using the same cell lysate from a 2.5 L culture that was divided between the different methods. SDS-PAGE gel analysis by Coomassie Blue staining confirmed that the protein of interest was maintained and successfully collected during heparin purification as well as high-salt washes. The overall yield in both methods was reduced compared to the standard purification but still sufficient for any downstream applications. However, the amount of purified protein was increased after high-salt washes compared to heparin purification. The products were analysed for DNA content by agarose gel electrophoresis and EtBr staining showing that the strong DNA-contaminations present in the standard purification were successfully removed during both steps (**Figure 4.3 C**). Spectrophotometric analysis also showed significant reduction of the DNA content in the 260 nm reading as well as the 260 nm/280 nm quotient after both heparin purification and high-salt washes which further confirms the absence of DNA (**Figure 4.3 C**). Even though the DNA content was slightly lower after heparin purification, there was no significant difference between heparin and high-salt purifications. In conclusion, both methods were found to be sufficient to remove bound DNA but high-salt washes generated higher amounts of purified protein and thus was used for further testing. The wash steps were prolonged in further purifications which finally produced a yield of approximately 30 mg of protein from a 2.5 L culture with high homogeneity (**Figure 4.3 D**). The increase in purity compared to the standard purification was most likely caused by the fact that other bacterial proteins bound to DNA were removed alongside DNA during the additional wash step.

Furthermore, it was tested whether the 6xHis-tag could be cleaved off using thrombin because removal of the flexible tag would increase the chance of crystal formation. Proteins were first concentrated in Vivaspin™ concentrator spin columns and the buffer exchanged to lysis buffer not containing any imidazole. The protein was then incubated with thrombin O/N at RT to cleave off the 6xHis-tag. The product was incubated with Ni-NTA beads which then only bound the tag but not the protein

of interest and allowed removal of the tag from the protein solution. Coomassie staining shows a product with lower molecular weight after incubation with thrombin which confirms removal of the tag (**Figure 4.3 E**). In this step, proteins were maintained while using Vivaspin™ protein concentrator spin columns for buffer exchange which can also be used to concentrate proteins e.g. for crystallography.

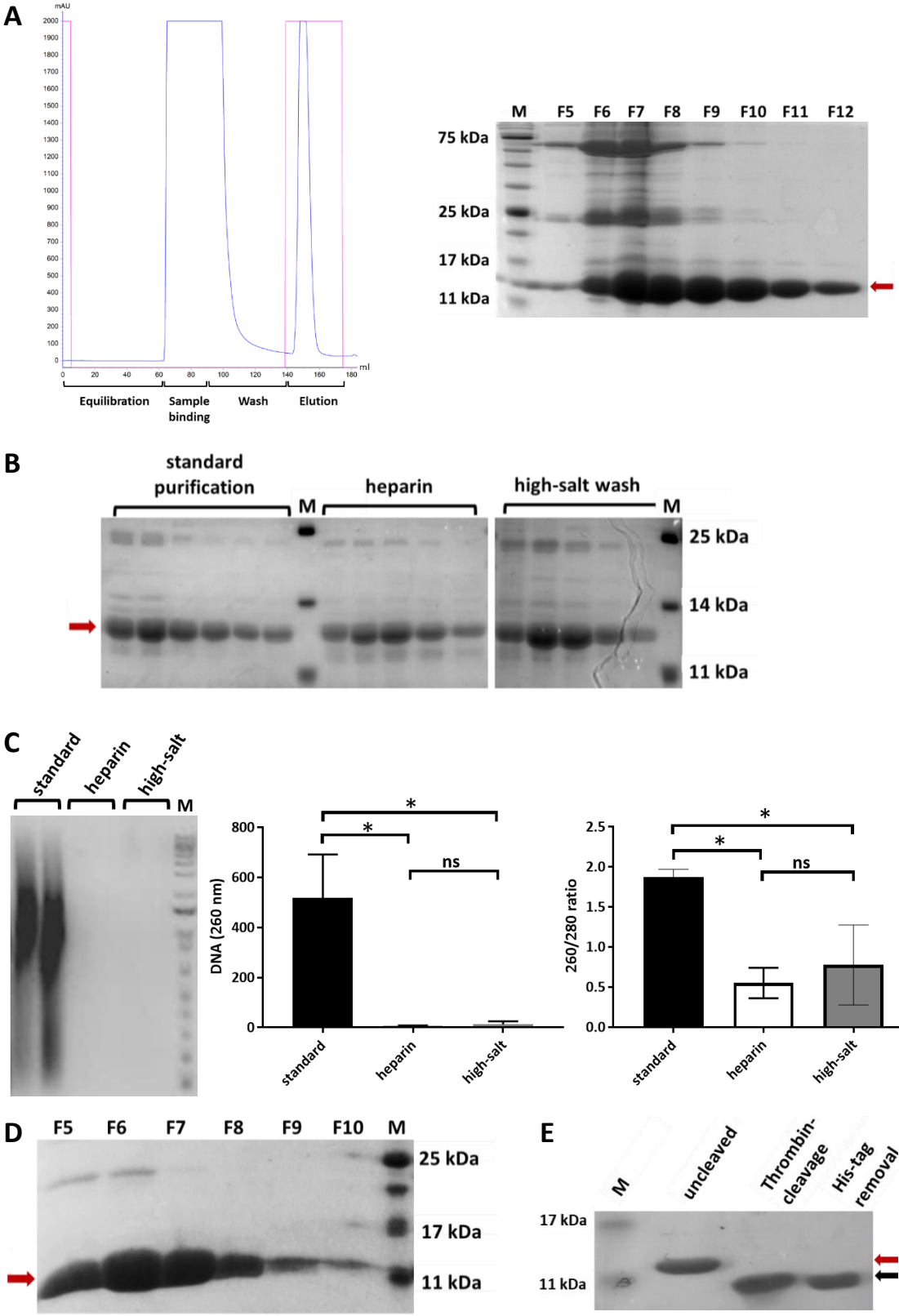

A
mAU
ml
Equilibration
Sample binding
Wash
Elution
M F5 F6 F7 F8 F9 F10 F11 F12
75 kDa
25 kDa
17 kDa
11 kDa
B
standard purification
heparin
high-salt wash
M
25 kDa
14 kDa
11 kDa
C
standard
heparin
high-salt
M
DNA (260 nm)
260/280 ratio
ns
D
F5 F6 F7 F8 F9 F10 M
25 kDa
17 kDa
11 kDa
E
M
uncleaved
Thrombin-cleavage
His-tag removal
17 kDa
11 kDa

Figure 4.3 Optimization of large-scale purifications for 6xHis-pWHD1

A) Standard large-scale purification of 6xHis-pWHD1, Left: Chromatogram (x-axis: UV-reading in mAU, y-axis: volume in ml), Right: SDS-PAGE gel analysis by Coomassie staining. **B)** SDS-PAGE gel analysis by Coomassie staining: Comparison of protein yields for 6x-His-pWHD1 after heparin or high-salt washes. **C)** Analysis of DNA content after different purification conditions by agarose gel electrophoresis and EtBr staining (left, 10 µl of two fractions resembling the highest protein concentrations were loaded) as well as photospectometer reading (middle and right); Statistical analysis was performed with Mann-Whitney-U test; *= $p<0.05$. **D)** SDS-PAGE gel analysis by Coomassie staining: Large scale purification of 6xHis-pWHD1 with an additional high salt wash yielding ~30 mg of protein from a 2.5 L culture. **E)** SDS-PAGE analysis by Coomassie staining confirming the successful cleavage of the 6xHis-tag from the pWHD1 by thrombin-digest.

10 µl of purified protein were loaded on SDS-PAGE gels. M=Marker; F=elution fraction; Red arrows indicate 6xHis-pWHD1; Black arrow marks pWHD1.

4.4 Summary

Polypeptides containing the MOZ pWHD1, pWHD2 and pWHD1+2 fused to a 6xHis-tag were successfully expressed in *E. coli* and purified in small-scale based on IMAC. The quantity and quality of small-scale purifications were sufficient to conduct the planned experiments to analyse the DNA-binding ability of the MOZ domains.

In addition, the 6xHis-pWHD1 was also produced in large-scale. Analysis of the purified protein showed high level contamination with DNA that most likely co-purified because it was tightly bound to the protein. Heparin as well as high-salt washes were found to efficiently remove bound DNA. Both methods produced sufficient quantities of protein, but higher yields were obtained from high-salt washes. The methods could also be combined, but implementation of both would be more time consuming and most likely reduce the overall protein yield. Thus, purification of 6xHis-pWHD1 was optimised using IMAC in combination with high-salt washes to yield approximately 30 mg of protein from 2.5 L of culture with high homogeneity. A second purification step using size-exclusion after thrombin cleavage could be of benefit because this would completely remove any additional proteins as well as 6xHis-pWHD1 protein that remained bound to DNA. The required column was

not available in our lab but will be implemented by the collaborators that will conduct further experiments. Crystallography trials with the MOZ pWHD1 will now be set up in collaboration with another group (Dr. Dreveny, University of Nottingham).

Chapter 5: Characterisation of the DNA-binding function of the MOZ N-terminus

Structural predictions and sequence analysis suggested the presence of two conserved regions comprised in the first 200 AA of MOZ that potentially fold to form a tandem WHD domain (Chapter 3). As described above, WHDs are versatile domains that mainly act as DNA-binding domains but can also function in RNA-binding or protein-protein interactions. However, since most predictions for the MOZ pWHDs were associated with proteins recognizing DNA, further analysis focused on this function. Having successfully used bacterial expression to purify pWHD1, pWHD2 and pWHD1+2 (Chapter 4), these were used to further explore the DNA-binding activity of MOZ in Electrophoretic Mobility Shift Assays (EMSAs). The speed at which molecules migrate during gel electrophoresis is mainly determined by their charge and size which can be used to separate free DNA from protein-bound DNA. When the protein of interest interacts with DNA, they form a complex with higher molecular weight than the corresponding free DNA which reduces their mobility in the gel. This results in reduction of free DNA band intensity and an upwards shift of the DNA-band after size-separation as illustrated in **Figure 5.1**. The charge of the complex is mainly determined by the negatively charged DNA-backbone and roughly remains the same within the protein-DNA complex and thus, only has minimal impacts on migration in the gel (Hellman and Fried, 2007; Holden and Tacon, 2011).

Both polyacrylamide and agarose gels can be used in EMSAs. Generally, polyacrylamide gels have higher densities and were found to provide better resolution for protein-DNA complexes especially with small DNA fragments and proteins with low molecular weight (Hellman and Fried, 2007). EMSAs in this study were carried out with relatively small polypeptides (12.5 kDa – 24.7 kDa) as well as some very short DNA-sequences (12 bp) and after testing different compositions of

both agarose gel and polyacrylamide gels, 7% polyacrylamide gels were found to be effective for our purposes.

Typically, nucleic acids in EMSAs are labelled with radioactive isotopes, tags like biotin or digoxygenin or fluorophores (Holden and Tacon, 2011). However, standard nucleic acid staining methods e.g. with EtBr can also be used to detect nucleic acids (e.g. Li et al., 2017), but they are not as sensitive as other labelling methods. For this project, the high concentrations of protein available allowed usage of high amounts of input DNA for EMSAs facilitating easy detection with EtBr staining and negating the requirement for radio isotopic labelling. Thus, EtBr detection was the method of choice in this project.

Briefly, a fixed amount of dsDNA probe was mixed with a titrated amount of the protein of interest and incubated for 30 min at RfT. The reaction-mix was then loaded on a pre-run DNA-PAGE gel, the DNA-bands were separated by electrophoresis, stained with EtBr and visualized with UV-light to detect DNA-shifts which indicate protein-DNA interaction.

It has to be considered that protein-DNA interactions *in vitro* and *in vivo* may also be impacted by binding conditions as well as other factors such as chromosome structure and co-factor proteins which can dramatically change the functionality of the protein.

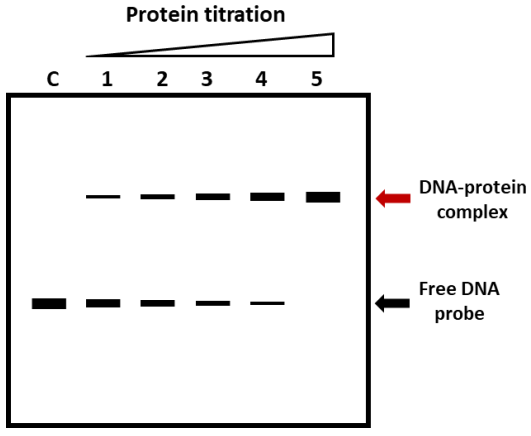


Figure 5.1 Idealised schematic of Electrophoretic Mobility Shift Assays (EMSA)

A titration of protein is mixed with DNA which causes a shift of the DNA band. With increasing amounts of protein more DNA is bound. Black arrow indicates the free DNA probe; Red arrow marks the DNA-protein complex.

5.1 General DNA-binding ability of the predicted WHDs

Although data from our lab as well as others indicates that MOZ can bind chromatin via interactions with histones and recruitment of MOZ to various promoter regions of target genes was previously demonstrated, literature searches revealed no information about a potential DNA sequence that might be recognized by MOZ. This is in part due to the lack of ChIP-seq data for this protein, as no ChIP grade antibodies are yet commercially available. Moreover, potential DNA-binding activity by the N-terminal region of MOZ has not been investigated. Therefore, deciding on what to use as potential DNA template was problematic. As a preliminary assay, it was first tested whether the MOZ pWHDs associate with dsDNA fragments present in commercial DNA-ladder (1kb Plus DNA Ladder, Invitrogen). The ladder provides random dsDNA sequences with various fragment lengths (100 bp - 15,000 bp) and could allow a first estimation of protein-DNA interaction. In EMSAs, normalised amounts of 6xHis-pWHD1, pWHD2 and pWHD1+2 (based on Bradford assays and confirmed by Coomassie staining; **Figure 5.2 A**) were tested and found to produce mobility shifts of bands present in the 1 kb+ DNA ladder (**Figure 5.2 B, C, D**). Increasing amounts of 6xHis-pWHD1 induced defined upwards shifts of both low and high MW bands (e.g. see 100 bp band). 6xHi pWHD2 produced a more smeared band shift profile at higher protein concentrations which was also observed for 6xHis-pWHD1+2. As a control, reduced shifts were observed if the protein preparations were denatured by heating 6xHis-pWHD1+2 to 80 °C prior to usage in the assay (**Figure 5.2 E**) which suggests that folding of modular domains is essential for DNA-binding ability. Thus, all three constructs appear to have DNA-binding ability, but these results may indicate different modes of action.

To get a clearer assessment on DNA-binding and to possibly define a specific DNA sequence as well as the DNA-binding mechanism, it was sought to use a single dsDNA probe of known sequence with a length between 100 and 200 bp. Single random DNA sequences were tested for this purpose. A 164 bp DNA probe available in our lab showed positive results in EMSAs. The sequence consists of a PCR generated

fragment comprising the first 46 codons of Phosphatidylinositol 4-Phosphate-5 kinase (PIP5K) 1A with flanking XhoI and BamHI restriction sites (**Figure 5.2 A**). The sequence was amplified in multiple standard Taq-PCRs and purified from an agarose gel to obtain purified DNA in high enough concentrations to be used in EMSAs.

The 164 bp DNA fragment showed clear DNA-shifts after incubation with the 6xHis-pWHD1 and 6xHis-pWHD1+2 polypeptides indicating that the DNA was bound (**Figure 5.2 F**). With increasing amounts of protein, up to four shifted DNA-bands were visible for the 6xHis-pWHD1 while the intensity of the free-DNA band was further reduced. The tandem module 6xHis-pWHD1+2 showed two clearly shifted DNA-bands and possibly one additional fainter band. The intensity of the free-DNA band was also reduced. The DNA-shifts produced by the combined 6xHis-pWHD1+2 construct were larger than for the 6xHis-pWHD1 alone with a 2.2 fold increase (as determined with ImageJ) which correlates with the approximate 2 fold larger MW of 6xHis-pWHD1+2. Unlike 6xHis-pWHD1 containing peptides, 6xHis-pWHD2 did not induce specific higher MW complexes, although depletion in the band intensity of the free DNA probe was observable which could suggest non-specific interactions (**Figure 5.2 F**). 6xHis-pWHD2 also caused shifts of the DNA ladder indicating that the domain does interact with DNA. As reported in the literature, DNA/RNA-binding domains and especially WHDs are versatile protein domains that recognize very different sequences and can also be specific for secondary structures. Moreover, different WHDs can form sequence specific contacts with the major groove of the double helix, or backbone contacts that are non-specific (Gajiwala and Burley, 2000). Thus, the tandem domains may function together as a single DNA recognition module in the context of the MOZ protein but contribute different sequence-specific or non-specific contacts in doing so. The observed multiple shifts of DNA bands are likely to represent the assembly of multiple 6xHis-pWHD1 or 6xHis-pWHD1+2 molecules per DNA probe. This may be due to the presence of multiple target sequences within the 164bp probe. Because the polypeptides are relatively small, multiple proteins could bind to the same DNA-strand without sterically hindering each other which would result in a

higher molecular weight of the complex. Alternatively, the pWHDs may form multimeric complexes due to protein-protein interactions or cooperative DNA binding.

Given the timeframe of this project it was decided to focus further investigation on the pWHD1 as interaction with the 164 bp DNA could allow to determine whether it binds a specific target DNA sequence and how binding is facilitated which will be described in the following chapters.

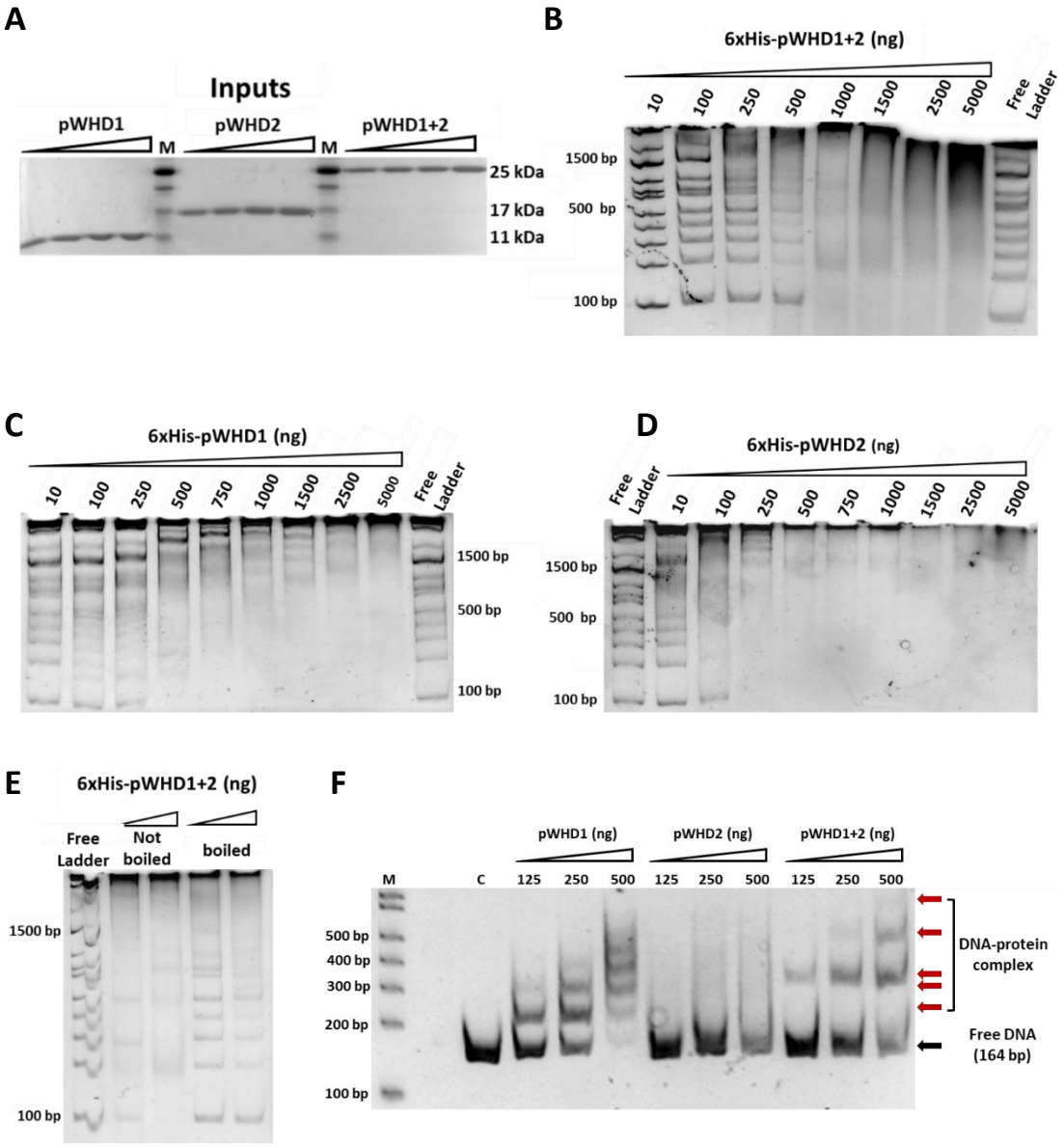


Figure 5.2 General assessment of the DNA-binding by MOZ pWHDs

A) SDS-PAGE analysis by Coomassie staining showing standardised protein inputs of the 6xHis-MOZ pWHDs for EMSAs. **B)** EMSA result of 6xHis-pWHD1+2 with DNA-ladder. **C)** EMSA result of 6xHis -pWHD1 with DNA-ladder. **D)** EMSA result of 6xHis-pWHD2 with DNA-ladder. **E)** EMSA result of boiled pWHD1+2 with DNA-ladder. **F)** EMSA results of 6xHis-pWHD1, pWHD2 and pWHD1+2 with a 164 bp DNA probe (750 ng input DNA).

M=marker; C=control; Black arrows mark free DNA; Red arrows indicate shifted DNA-protein complexes; protein inputs are indicated above in ng.

5.2 Deletion mapping of a specific DNA-recognition sequence of pWHD1

The MOZ pWHD1 was able to produce shifts of the 164 bp DNA probe which indicates that it can function autonomously as a minimal DNA-binding domain, at least *in vitro*. Binding of DNA may involve non-specific DNA contacts or recognition of specific sequence motifs embedded in the 164 bp probe which most likely contains multiple binding sites based on the observation of multiple shifts during EMSAs. Deletion mapping analysis was used to localise where binding of pWHD1 may occur on the probe.

The 164 bp DNA sequence was first divided into three parts. Oligonucleotides were designed to generate dsDNA probes corresponding to bases 1-55 (Probe A, 55 bp), 56-109 (Probe B, 54 bp) or 110-164 (Probe C, 55 bp) of the original probe (**Figure 5.3 A**). The complementary forward and reverse oligonucleotides were annealed and subjected to EMSAs.

As shown in **Figure 5.3 B**, 6xHis-pWHD1 produced clear shifts with Probe A as well as Probe B. At least two shifted bands were visible for Probe A which is in line with the previously observed multiple shifts using the parental 164 bp sequence. On the other hand, little or no complex formation was observed for Probe C. Even though free DNA of Probe C may have been slightly reduced, only little DNA smear was produced and no band shifts induced. This differential association with different probes suggests a sequence preference of 6xHis-pWHD1 rather than an unspecific, sequence-independent mode of DNA-binding. Comparison of Probe A, B and C revealed that the probes have different GC-contents. Probe A contains 64% GC-nucleotides, while B and C only contain 52% and 51%. Furthermore, Probe C is the only fragment that does not contain CpG dinucleotides. Probe B has two consecutive CpG motifs and Probe A, which produced multiple shifts, contains eight CpG dinucleotides.

To further investigate the potential recognition sequence, a series of 14 bp oligonucleotides were designed based on the original 164 bp probe. The 14mers spanned the regions 1-14 (14mer-A), 15-28 (14mer-B), 29-42 (14mer-C), 43-56

(14mer-D), 66-79 (14mer-E), which all contained CpG base-pairs but different bases accompanying the motif (**Figure 5.3 A**). This was also done to test whether 6xHis-pWHD1 could associate with small DNA probes and if shifts were detectable using the relatively low sensitivity of EtBr based EMSAs.

The EtBr staining of these gels produced DNA background that increased with higher protein titrations and remained relatively even for all tested DNA sequences (**Figure 5.3 C**). This was most likely caused by DNA contaminations present in the protein purification of the 6xHis-pWHD1 which were discussed above (Chapter 4). This contamination was not observed or had a much weaker impact in previous EMSAs. The DNA-fragments used in those assays were significantly bigger which increases the intensity of the EtBr staining and probably reduced the visible effect of background DNA. In contrast, the 14 bp oligonucleotides produce less intense bands and therefore required higher exposure times with UV-light for detection which may have made background DNA visible.

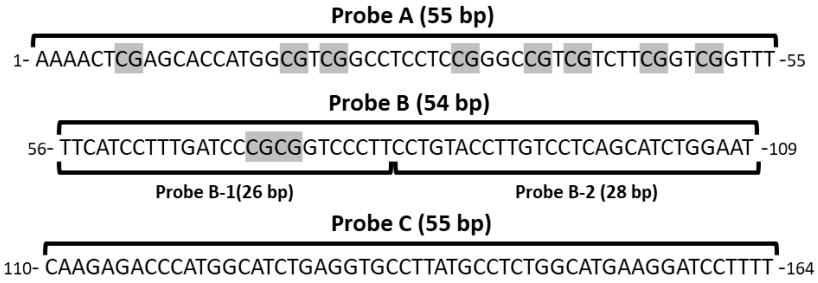
Even though background DNA smear was visible on the gels, interaction of the 6xHis-pWHD1 with the 14mers was still detectable because the contaminating DNA mainly consisted of bigger fragments and ran higher than the used probes. Decrease of the free DNA with titrations of protein was clearly visible and resulting shifted bands were detectable (**Figure 5.3 C**). This allowed to observe differences in the interaction of the 14mers with the 6xHis-pWHD1. 14mer-C and 14mer-D showed the strongest reduction in free DNA. 14mer-B and 14mer-D also showed interaction with pWHD1 which was reduced compared to the previously mentioned oligonucleotides. In contrast, the 14mer-A showed almost no reduction of free DNA. Interestingly, even though all sequences contained CpG dinucleotides, they were not evenly recognized by the 6xHis-pWHD1. This could either suggest that CpG motifs are not the defining motif recognized by the 6xHis-pWHD1 or that surrounding bases also impact the protein-DNA interaction. In contrast to the other 14mers, 14mer-A contains only one CpG motif which is also the only one flanked by a thymine as well as adenine. The DNA fragments 14mer-C and 14mer-D, which showed the strongest interaction,

contain two or more CpG dinucleotides which are flanked by other cytosines or guanines.

The results so far suggested that the MOZ pWHD1 may recognize sequences containing CpG motifs. This hypothesis was further tested with different oligonucleotides that were also used to confirm the recognition sequence of PHF1 (Li et al., 2017). One DNA fragment contained the consensus sequence recognized by all members of the FOX-family (WH-motif, after Biggs et al., 2001), one provided an AT-rich sequence (16mer-AT-rich) and two contained CG-rich sequences but only one of them comprised CpG dinucleotides (12mer-CpG) while the other only contained a GpC motif (12mer-GpC) (**Figure 5.3 A**). The 16mer-AT-rich DNA fragment and the consensus WH-motif showed little to no interaction with the 6xHispWHD1 in EMSAs as almost no reduction of free DNA and only very faint band shifts were observed (**Figure 5.3 D**). In contrast, 6xHis-pWHD1 produced clear shifts with the palindromic 12mer-CpG DNA and input DNA was completely bound indicating strong protein-DNA interaction. Removal of the CpG motifs from the sequence while maintaining the GpC motif (12mer-GpC) strongly reduced the ability of the 6xHis-pWHD1 to bind the DNA. These results suggest that the CpG dinucleotide, but not the GpC motif, is required for binding and forms a distinct recognition motif for the domain.

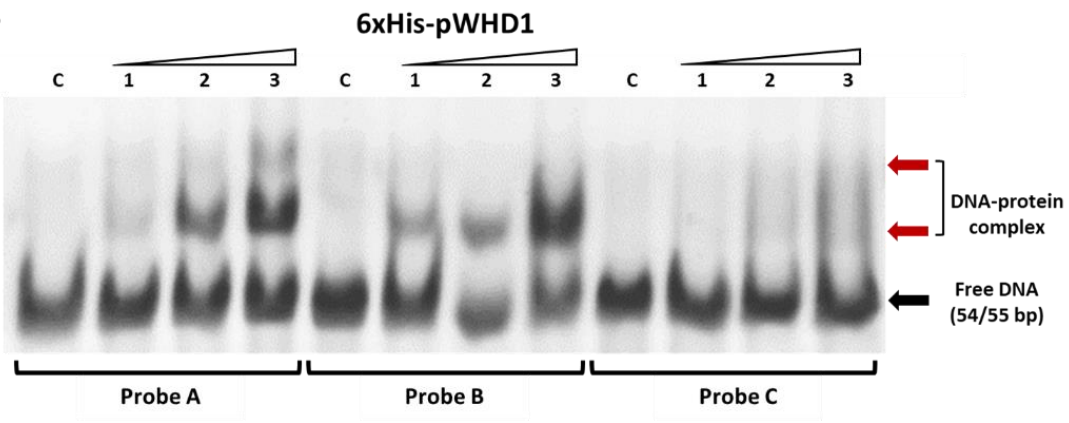
It was also tested whether 6xHis-pWHD1 recognizes single-stranded DNA (**Figure 5.3 E**). Incubation with only forward oligonucleotide of Probe B, which was not annealed to the reverse strand, only produced a small DNA-shift. In addition, the forward strand of Probe B was separately annealed with reverse oligonucleotides that only stretched either the first or second half of the sequence and thus should create DNA probes that are half double stranded and half single-stranded (Probe B-1 and B-2, **Figure 5.3 A**). Only the combination with the CpG motif present as dsDNA (Probe B-1) showed interaction with 6xHis-pWHD1. This further supports the findings that the 6xHis-pWHD1 preferentially binds dsDNA containing CpG sequences.

A

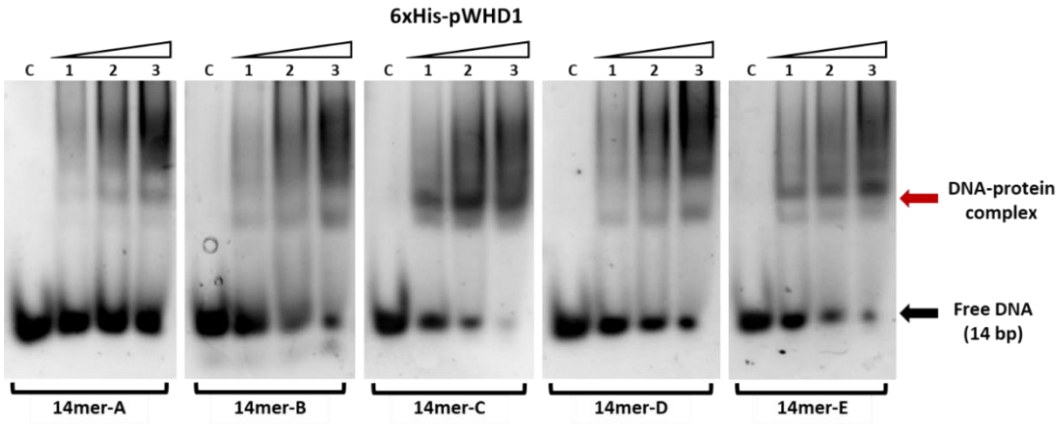


DNA name	DNA sequence
14mer-A	AAAACTCGAGCACC
14mer-B	ATGGCGTCGGCCTC
14mer-C	CTCCGGGCCGTCGT
14mer-D	CTTCGGTCGGTTTT
14mer-E	GATCCCGCGGTCCC
12mer-CpG	GGGCGGCCGCCC
12mer-GpC	GGGGGGCCCCCC
16mer-AT-rich	TTTTTATTAATAAAAA
13mer-WH-motif	CTATGTAAACAAC

B



C



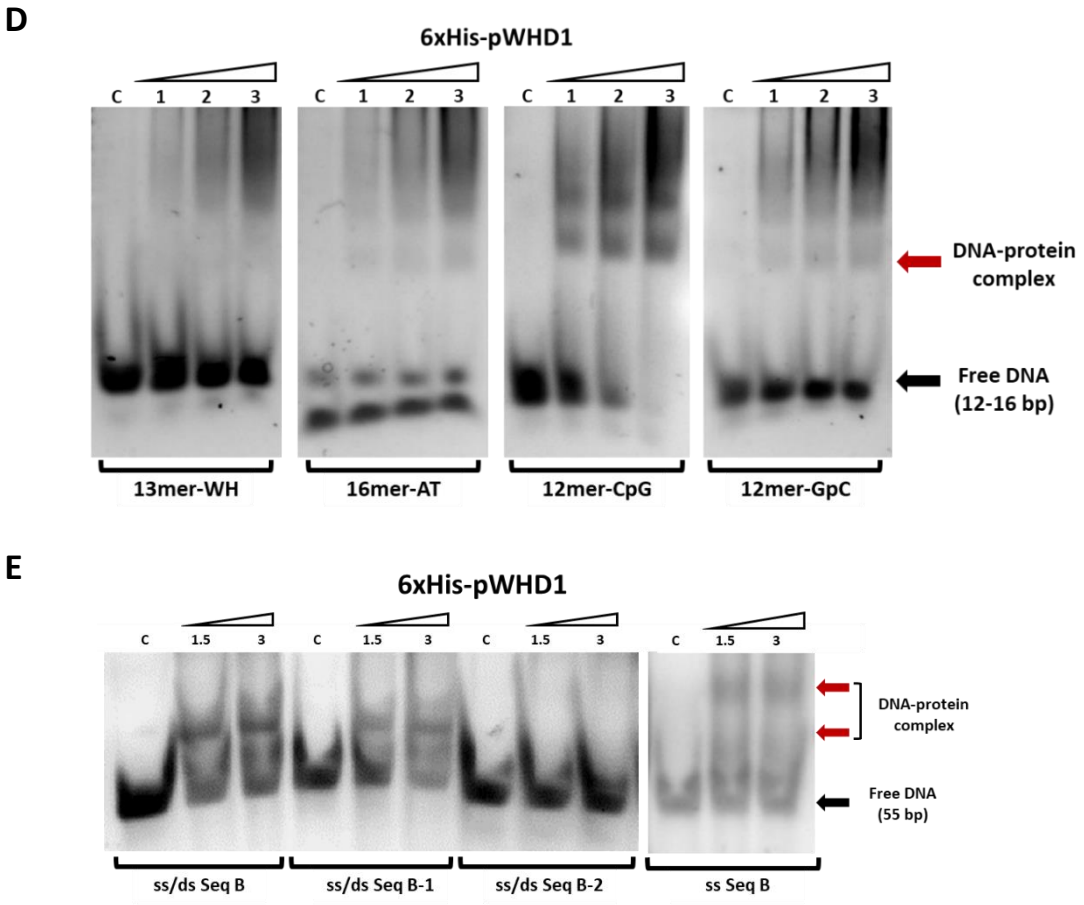


Figure 5.3 Binding analysis of pWHD1 with various DNAs

A) Sequences of DNAs used for EMSAs. The sequence of the 164 bp DNA fragment and boundaries of Probe A, B and C are shown at the top. The bottom table shows the sequences of 12mers, 14mer and 16mers. Only the forward sequence of each DNA-duplex is shown. **B)** EMSA result of 6xHis-pWHD1 with Probe A, B and C. **C)** EMSA result of 6xHis-pWHD1 with 14mers derived from the 164 bp DNA fragment. **D)** EMSA results of 6xHis-pWHD1 with different double stranded DNAs. **E)** EMSA result of 6xHis-pWHD1 with half ss/half dsDNA and ssDNA.

For B and E protein inputs in µg are indicated above and 200 pmol DNA input were used. For C and D protein to DNA molar ratios are indicated above and 100 pmol DNA input were used. Data shown are representative of at least two independent experiments. Black arrows mark free DNA probe; Red arrows indicate shifted DNA-protein complexes.

5.3 MOZ pWHD1+2 recognizes additional DNA-sequences

The polypeptide 6xHis-pWHD1+2 was also tested for binding affinity to Probe A, B and C (**Figure 5.4**). Similar to 6xHis-pWHD1 on its own, pWHD1+2 showed clear shifts with Probe A and B. In contrast to 6xHis-pWHD1, 6xHis-pWHD1+2 also produced a shifted band for Probe C, but this shift was much weaker compared to the probes A and B. This additional interaction of 6xHis-pWHD1+2 could result from interaction of the pWHD2 with DNA or also be related to a combined binding mechanism by pWHD1 and pWHD2.

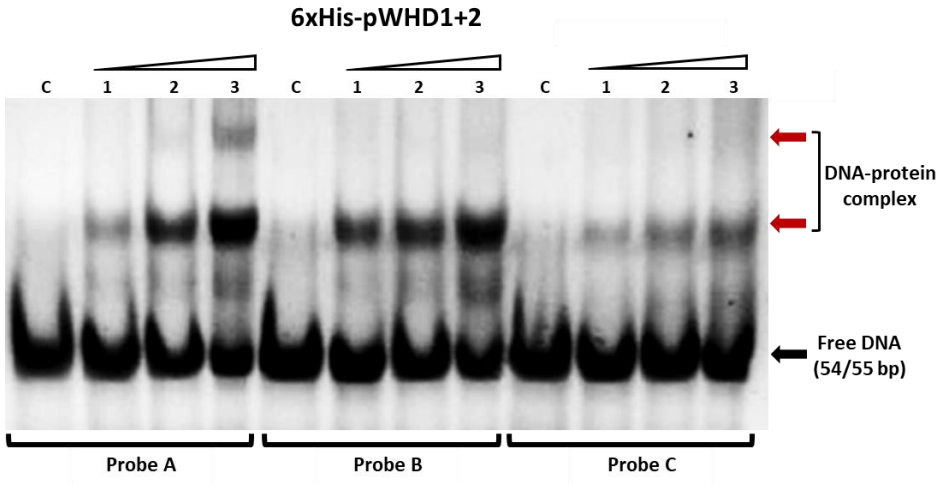


Figure 5.4 Binding analysis of MOZ pWHD1+2 with various dsDNA probes

EMSA result of 6xHis-pWHD1+2 with Probe A, B and C.

Protein inputs in µg are indicated above and 200 pmol DNA input were used. Data shown are representative of at least two independent experiments. Black arrows mark free DNA probe; Red arrows indicate shifted DNA-protein complexes.

5.4 6xHis-tag does not impact DNA-binding of MOZ-pWHDs

The polypeptides used in this study were purified using fusion-tags. 6xHis-tags were not previously reported to interact with DNA and are not expected to impact the DNA-binding mechanisms the MOZ pWHDs. However, to confirm that EMSA results are not impacted by interaction of the 6xHis-tag with DNA, thrombin was used to cleave the tag off the 6xHis-pWHD1 polypeptide (described in Chapter 4.3). The

cleaved products were tested in EMSAs with Probe B DNA to compare the DNA-binding ability of polypeptides with or without the tag. No changes were observable in the ability of the protein to interact with DNA (**Figure 5.5**). The distance of the shift was slightly reduced which is most likely caused by the slight reduction of the molecular weight of the protein after removal of the tag. This result confirms that the protein-DNA interaction relies only on the MOZ domain and is independent of the 6xHis-tag.

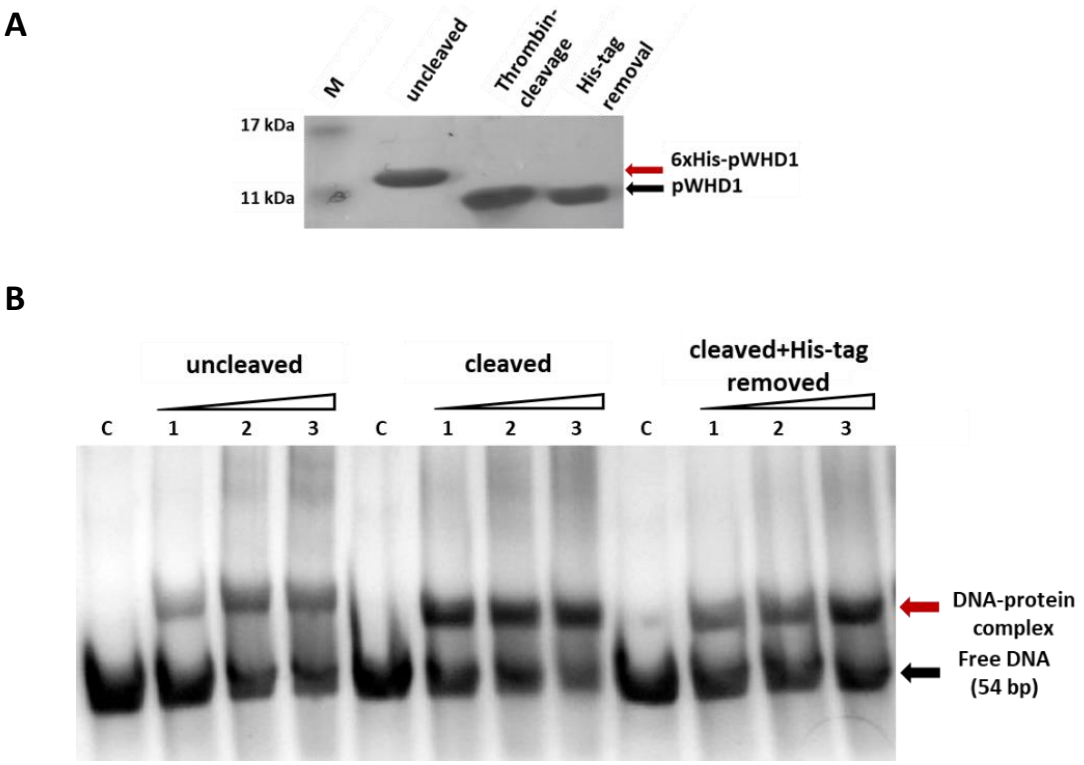


Figure 5.5 DNA-binding ability of pWHD1 is maintained after removal of the 6xHis-tag

A) SDS-PAGE analysis by Coomassie staining showing the inputs of 6xHis-pWHD1 before and after His-tag removal for EMSAs. B) EMSA showing that the removal of the 6xHis-tag had no effect on the DNA-binding ability of the pWHD1. Protein inputs in µg are indicated above, 200 pmol of DNA input were used. M=Marker; Arrows indicate the expected protein bands.

5.5 Mapping of the DNA-interaction surface of the MOZ pWHD1

The FOX protein family is the most prominent member of WHD containing proteins. Their canonical recognition mode involves contacts between helix H3 and the mayor groove of DNA, a mechanism that was also identified in many other WHD proteins. The C-terminal "wings" of several WHDs were also found to interact with DNA (Lalmansingh et al., 2012). These contacts are usually additional and directed to the minor groove of DNA, but can also replace the helix H3 interaction with the major groove (for a detailed description see Chapter 1.1.3). An example for this mechanism of binding is the WHD of PHF1, a member of the polycomb PRC2 complex. PHF1 specifically recognizes CpG motifs via a unique lysine rich stretch (IKKK/RK) located in the W1 loop while general DNA-binding is supported by additional contacts (Li et al., 2017). Another study of the PCG2 polycomb group protein, which also contains a WHD binding CGCG sequences, was found to mainly interact with DNA through two conserved glycine residues in its wing region as well as helix B. Additional contacts are made with an N-terminal loop (20 loop) as well as the helix (Liu et al., 2014).

Mapping of the specific sequence recognized by the MOZ pWHD1 during this study (Chapter 5.2) suggested a strong preference for CpG motifs whereas the domain was not able to bind the canonical target sequence recognized by FOX proteins. Thus, it seems plausible that the MOZ pWHD1 may bind via a similar mechanism to PHF1 or PCG2 as they all bind CpG dinucleotides. However, no obvious sequence similarities to either the PHF1 or PCG2 WHD were found in the wing region. However, the loop L1 connecting helices H1 and H2 of the MOZ pWHD1 contains a highly conserved IKKVKQKQR motif. This sequence shows similarity to the IKKK/RK DNA-recognition motif of PHF1 and also provides two glutamine residues comparable to the two glutamines that were found to be essential for DNA-binding in PCG2. Based on these sequence similarities and consensus in the recognition of CpG DNA motifs, it seems possible that the loop L1 of the MOZ-pWHD1 could play an important role in DNA recognition. As described in the introduction (Chapter 1.1.3), linker histones H1 and H5 make multiple contacts to DNA. While they bind nucleosomal DNA with

the helix H3 with mechanisms similar to most WHDs, it was also demonstrated that they contact linker DNA via the wing region as well as the loop L1 between helices H1 and H2. Most of these DNA-contacts are facilitated by positively charged surface residues (Zhou et al., 2013 and 2015; Lyubitelev et al., 2016). In addition, it was shown that histone H1 also recognizes nucleosomal DNA via C-terminal regions with lysine rich motifs ("ASAKKEK and KKPKAKKAVAT"; Zhou et al., 2013). Since most predictions for the MOZ pWHD1 were based on alignments with linker histones, this further supports the possibility of DNA recognition by the loop L1 region. However, this would still be an atypical mechanism and deviate from the DNA-binding mode of most WHDs which mainly recognize DNA via the helix H3 and wing regions. In addition, pWHD2 may also impact functionality of pWHD1.

Mutational analysis was implemented to provide a first assessment of the DNA-binding surface of the MOZ pWHD1. Mutational disruption of single or multiple amino acids is typically used to map and/or confirm residues that are essential for protein function. If structural data such as DNA-bound protein structures are available, this can of course directly hint to essential residues. However, since no such structure was available for the MOZ N-terminus, we decided to focus mutational analysis on positively charged surface residues which are generally suggested by the literature to be especially important for DNA-binding and are the main target of mutational analysis in most studies related to WHDs (e.g. Li et al., 2017; Zhou et al., 2013 and 2015; Hayashihara et al., 2010). It has to be considered that the available structure is only predicted and that the actual borders of secondary structure elements as well as their orientation and flexibility can strongly vary from the prediction. However, this provides the best estimation of the structural location of single residues and finally, nine residues were selected for point mutations as shown in **Table 10** and **Figure 5.6.** The lysine rich stretch in the loop between helices H1 and H2 was targeted with substitution mutations of K18, K19, K21, K22 and K24. Furthermore, three mutations targeted the canonical DNA-recognition motifs and were located in the helix H3 (R44) and wing region (K62 and K72). The function of

charged residues like lysine and arginine is typically interrupted by replacement with uncharged alanine residues. This strategy is also found in most mutagenesis studies of WHDs that try to disrupt DNA-binding (e.g. Li et. al, 2017; Zhou et al., 2013 and 2015). However, even stronger interference can be achieved by replacement with residues of opposite charges, in this case with negatively charged glutamate or aspartate (e.g. Choi et al., 2017). To target the lysine rich stretch, K18 and K19 in pWHD1+2 were first changed to alanine. As these mutations showed no strong impact on DNA-binding function *in vitro* (see below), K19, K21, K22 and K24 were replaced by glutamates in 6xHis-pWHD1 both as a quadruple mutant and as single mutants. The quadruple mutant was also generated using alanine instead of glutamate substitutions. Because most studies that focused on the helix H3 and wing regions showed disruption of DNA-binding with lysine/arginine to alanine mutations, we also used this set of mutations for R44, K62 and K72.

Table 10 List of point mutations for 6xHis-pWHD1+2 and 6xHis-pWHD1

The table shows point mutations that were incorporated into the 6xHis-pWHD1+2 and 6xHis-pWHD1 using site-directed mutagenesis.

Construct	Predicted pWHD1 region	Position in MOZ
6xHis-pWHD1+2^{K18A}	Loop L1	18
6xHis-pWHD1+2^{K19A}	Loop L1	19
6xHis-pWHD1^{K18A}	Loop L1	18
6xHis-pWHD1^{K-E}	Loop L1	19, 21, 22, 24
6xHis-pWHD1^{K19E}	Loop L1	19
6xHis-pWHD1^{K21E}	Loop L1	21
6xHis-pWHD1^{K22E}	Loop L1	22
6xHis-pWHD1^{K24E}	Loop L1	24
6xHis-pWHD1^{K-A}	Loop L1	19, 21, 22, 24
6xHis-pWHD1^{R44A}	Helix H3	44
6xHis-pWHD1^{K62A}	Wing W1	62
6xHis-pWHD1^{K72A}	Wing W2	72

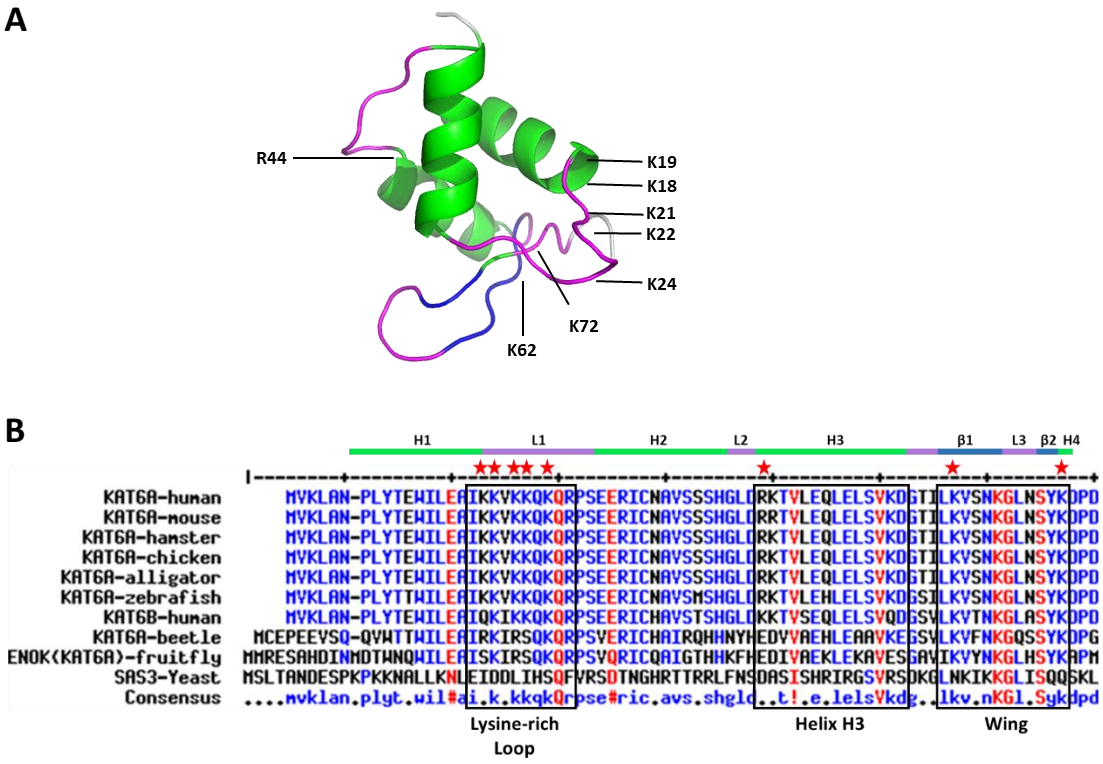


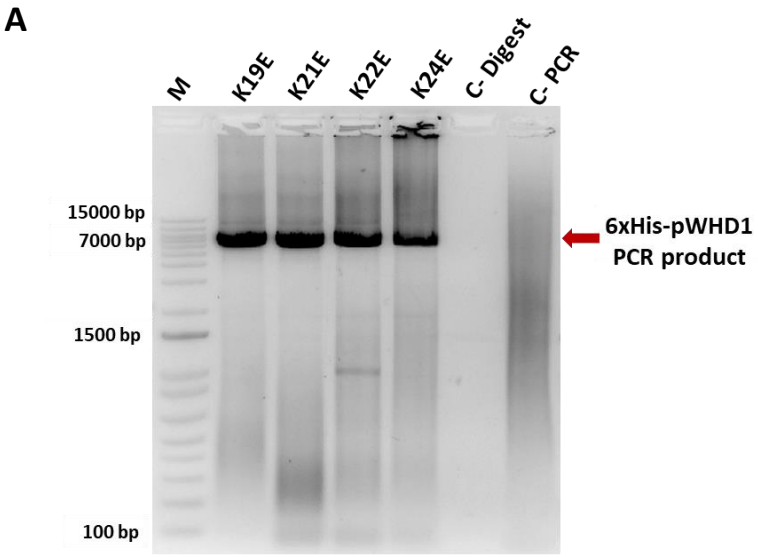
Figure 5.6 Positions of mutations in MOZ pWHD1

A) Predicted tertiary structure of the MOZ pWHD1 (residues selected for mutagenesis are labelled; Phyre2 prediction after Kelly et al., 2015). **B)** Sequence alignment of the MOZ pWHD1 from different species. Predicted secondary structure elements are shown on top and selected point mutations are marked with red stars. Green=α-helix; Blue=β-strand; Magenta=loops; grey=disordered prediction.

5.5.1 Site-directed mutagenesis of the MOZ N-terminus

Whole plasmid site-directed mutagenesis was used to create point mutations in the 6xHis-pWHD1 and 6xHis-pWHD1+2 expression vectors. First, primers were designed which contained the minimally necessary base-pair changes to introduce the desired codon changes in the plasmid DNA. PCR with a prolonged extension time (see Chapter 2.3.13) was then used to create full-length PCR products of the plasmids which contained the required mutations. The template DNA consisting of the original WT vector was eliminated with DpnI. This enzyme specifically digests methylated DNA and does not recognize DNA synthesised by PCR. Thus, the template plasmid, which

was methylated during production in bacteria, gets digested while the mutated PCR-product is maintained. An example in **Figure 5.7 A** shows that input plasmid, which was not amplified during PCR, was completely digested (C- Digest, left panel) while the PCR products remain intact (right panel). The PCR product was then transformed into DH5α and plasmids from single colonies were mini-prepped. The integration of the desired mutations listed in **Table 10** was confirmed by sequencing (**Figure 5.7 B**).


A
M
K19E
K21E
K22E
K24E
C- Digest
C- PCR
15000 bp
7000 bp
1500 bp
100 bp
6xHis-pWHD1
PCR product

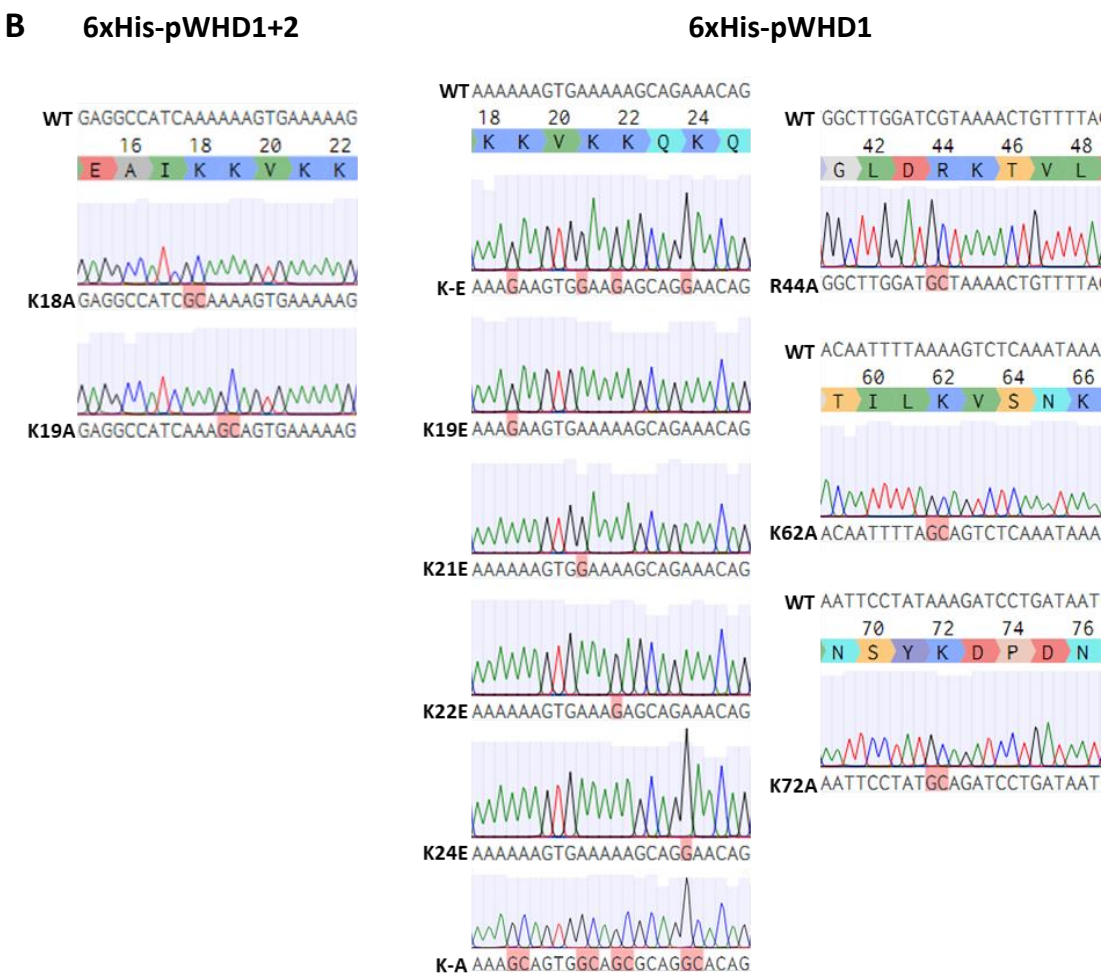

B
6xHis-pWHD1+2
6xHis-pWHD1
WT GAGGCCATCAAAAAAGTGAAAAAG
16 18 20 22
E A I K K V K K
K18A GAGGCCATCGCAAAAGTGAAAAAG
K19A GAGGCCATCAAAGCAGTGAAAAAG
WT AAAAAAGTGAAAAAGCAGAAACAG
18 20 22 24
K K V K K Q K Q
K-E AAAGAAGTGGAAGAGCAGGAACAG
K19E AAAGAAGTGAAAAAGCAGAAACAG
K21E AAAAAAGTGGAAAAGCAGAAACAG
K22E AAAAAAGTGAAAGAGCAGAAACAG
K24E AAAAAAGTGAAAAAGCAGGAACAG
K-A AAAGCAGTGGCAGCGCAGGCACAG
WT GGCTTGGATCGTAAAACTGTTTTA
42 44 46 48
G L D R K T V L
R44A GGCTTGGATGCTAAAACTGTTTTA
WT ACAATTTTAAAAGTCTCAAATAAA
60 62 64 66
T I L K V S N K
K62A ACAATTTTAGCAGTCTCAAATAAA
WT AATTCCTATAAAGATCCTGATAAT
70 72 74 76
N S Y K D P D N
K72A AATTCCTATGCAGATCCTGATAAT

Figure 5.7 Creation of 6xHis-pWHD1 and 6xHis-pWHD1+2 mutagenesis constructs

A) Agarose gel electrophoresis showing the PCR-product of different mutagenesis constructs after DpnI digest. The WT DNA (C- Digest) was completely digested while the mutated PCR products were maintained. . M=Marker; C- Digest= control digest of the input plasmid; C- PCR= PCR negative control. **B)** Sequencing electropherograms confirming the successful incorporation of different point mutations in the sequence of 6xHis-pWHD1+2 and 6xHis-pWHD1 plasmid DNA. The WT DNA and amino acid sequences are shown at the top. The DNA sequences and electropherograms of the mutants are shown below and the required codon changes for mutations are marked in red. Mutation K18A of 6xHis-pWHD1 is not displayed as sequencing results were similar to 6xHis-pWHD1+2.

5.5.2 Disruption of protein-DNA interaction in mutant constructs

After successful incorporation of the desired mutations in the 6xHis-pWHD1 vector, all mutants were expressed and purified in small-scale (**Figure 5.8**) as previously described (Chapter 4.2) The DNA-binding ability of the mutants was then assessed in EMSAs with the 54 bp Probe B DNA fragment which was clearly recognized by the WT 6xHis-pWHD1 as well 6xHis-pWHD1+2.

6xHis-pWHD1+2^{K19A} showed a slight decrease in complex formation, but neither mutation of K18 or K19 to alanine showed a strong disruption of the DNA-binding ability (**Figure 5.8 A, left**) as was observed in the literature for PHF1 (Li et al., 2017). However, it remained possible that pWHD2 was able to rescue mutational depletions of pWHD1 because pWHD2 may also contribute to DNA-binding ability (Chapter 5.1 and 1.1). Thus, for further assessment only the 6xHis-pWHD1 was used for domain specific mutational analysis.

Alanine mutation of K19 was not obtained in 6xHis-pWHD1 and substitution of K18 showed no impact on DNA-binding (**Figure 5.8 A, right**). For further testing, four lysine residues were individually or simultaneously mutated to glutamates, which is expected to generate a stronger disruption than replacement with alanine. Simultaneous mutation of the four lysine residues K19, K21, K22 and K24 (6xHis-pWHD1^{K-E}) to glutamates showed complete abolishment of DNA-binding (**Figure 5.8 B**). This result indicates that pWHD1 interact with DNA through the lysine rich stretch

located between helices 1 and 2 (**Figure 5.8**). Subsequently, the four lysine residues were mutated individually to further define the interaction interface (**Figure 5.8 B**). Substitution of the residue K24 with glutamate did not completely abolish the interaction with DNA like 6xHis-pWHD1$^{K\text{-}E}$, but no complex formation was observed and the reduction of free DNA was lower than for the WT indicating a disruption of protein-DNA interaction. Mutating K19 also showed small changes in DNA-binding as the shifted bands were not as strong as in the WT. Disruption of K21 and K22 had no visible effects on DNA-binding.

Because of the limited time frame of the project, the mutations in the helix H3 and wing region as well as the mutant 6xHis-pWHD1$^{K\text{-}A}$ were not tested for DNA-binding.

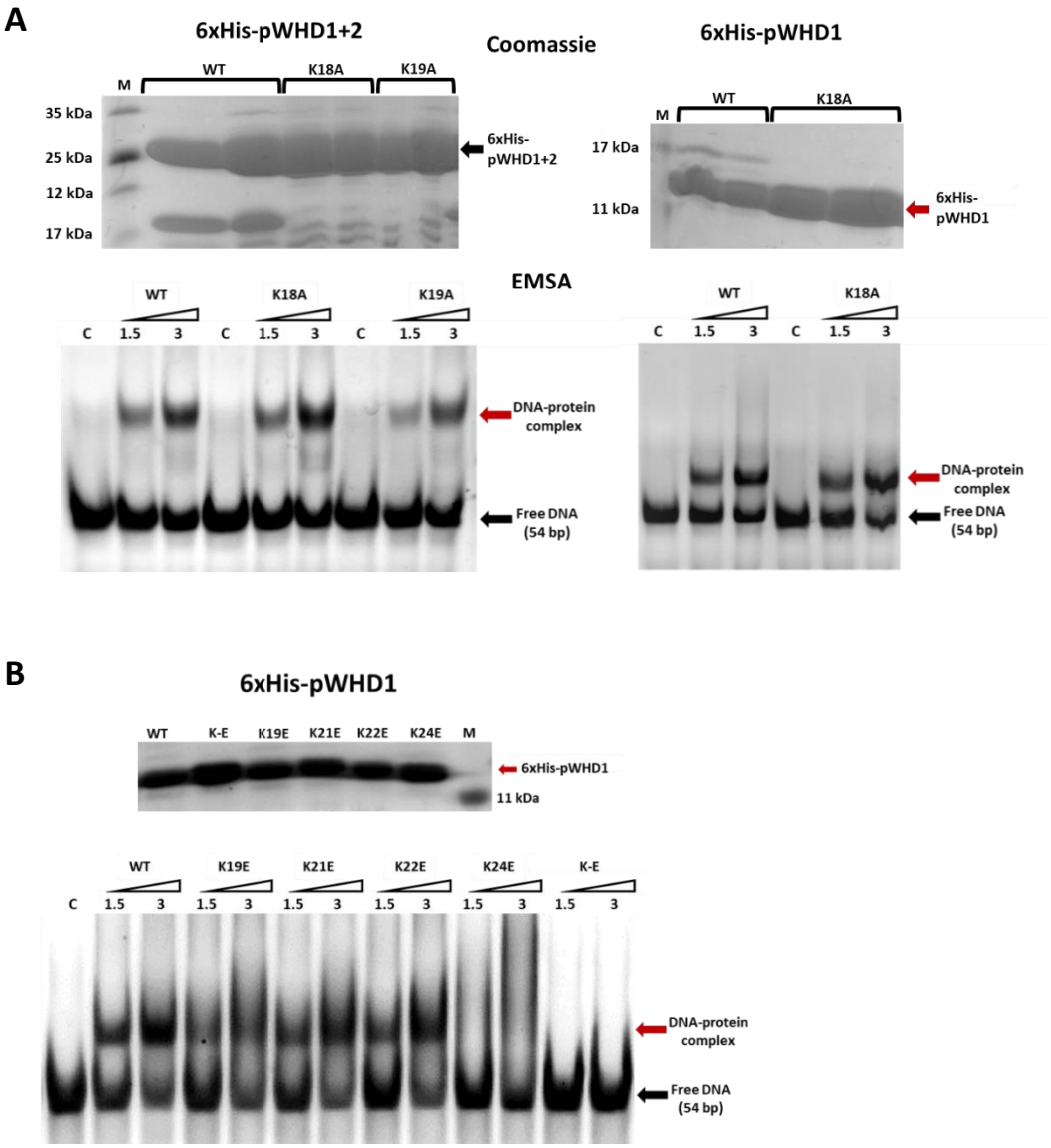


Figure 5.8 Mutational analysis of pWHD1 and pWHD2

A) Top: Expression of WT and mutant constructs of 6xHis-pWHD1+2 and 6xHis-pWHD1; 10 µl were loaded. Bottom: EMSA result of the binding of Probe B DNA with WT or mutant forms of 6xHis-pWHD1+2 and 6xHis-pWHD1. **B)** Top: SDS-PAGE analysis by Coomassie staining showing standardised inputs of 6xHis-pWHD1 mutants (5 µg input); Bottom: EMSA result of the binding of Probe B DNA with WT or mutant forms of 6xHis-pWHD1. Data shown are representative of at least two independent experiments.

Protein inputs in µg are indicated above and 200 pmol DNA input were used; M=Marker; Black arrows mark free DNA probe; Red arrows indicate shifted DNA-protein complexes.

5.6 Summary

Structural predictions indicated that the N-terminal region of MOZ is likely to contain two WHDs which were mainly aligned with DNA-binding proteins. Using recombinant proteins in EMSAs, it was demonstrated that both predicted MOZ WHDs incubated with random dsDNA fragments were able to form protein-DNA complexes suggesting that the domains are involved in dsDNA-binding. In addition, tests with pWHD1 suggested that the domain prefers dsDNA and has weaker affinity for ssDNA. Mapping analysis provided compelling evidence that pWHD1 acts as a minimal DNA-binding domain that appears to a have a strong preference for sequences containing CpG motifs. Furthermore, the results indicate that abundance of CpG dinucleotides and the sequence surrounding the motif also impact the binding-ability of pWHD1. Based on the 12mer-CpG probe that was clearly recognized by pWHD1, future studies could further define the recognition sequence by mutational analysis of bases flanking the CpG motif. So far, first assessments indicate that pWHD1 prefers CpG dinucleotides flanked by further C or G bases, but this will require further confirmation. CpG motifs are frequent targets of DNA methylation which is generally associated with gene silencing (Jones, 2012). Thus, it would also be interesting to investigate the ability of pWHD1 to bind hemi- or fully methylated DNA. Our lab previously performed RNA-seq analysis of CRISPR edited cell lines depleted in MOZ expression (Monteiro and Heery, see below Chapter 6). Interestingly, analysis of promoter regions of genes that were differentially expressed upon absence of MOZ identified enrichment of motifs that were rich in cytosines and provided CpG dinucleotides (**Figure 5.9**; analysis was performed by Cintia Monteiro, David Heery and Nigel Mongan, related software is cited in Chapter 2.6). These findings may suggest that regulation of target genes by full-length MOZ correlates with CpG content of promoter regions which would fit with the suggested sequence preference of pWHD1 for CpG motifs. However, other MOZ domains and co-factors, e.g. the PHD fingers, may also influence MOZ target sequences and this type of enrichment analysis is also not a truly compelling method to identify target motifs.

Implementation of additional methods could help to confirm and further define the recognition sequence. ITC would typically be used to determine the stoichiometry of protein-DNA complex formation (e.g. Li et al., 2017). Other methods such as DNA-foot printing (Hellmann and Fried, 2007), SELEX-Seq (Gu et al., 2013), DNA-microarray-based analysis (Li et al., 2007) and ChIP-Seq could be used to confirm the recognition sequence. Unfortunately, no suitable chip grade antibodies are currently available, therefore an alternative being considered for future studies in our lab is to use CRISPR Cas9 to add an N-terminal epitope tag e.g. FLAG to facilitate this. Exogenous expression of tagged MOZ may also be used for ChIP analysis.

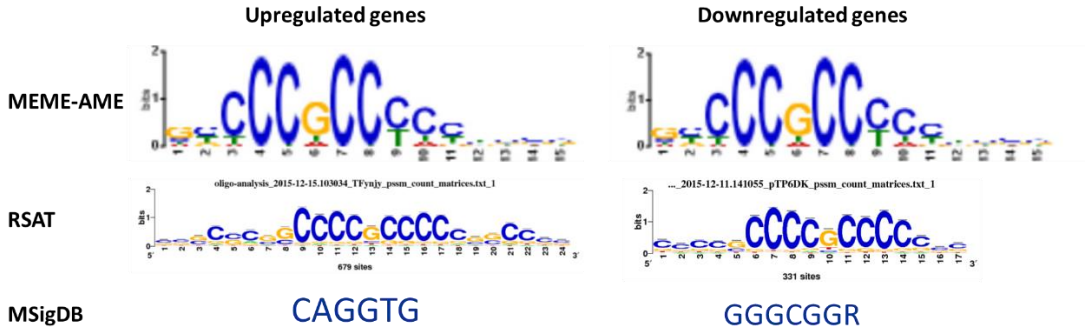


Figure 5.9 Enrichment analysis of differentially expressed genes in MOZ knockout cells

Analysis of regions upstream of the TSS of genes that were differentially expressed in HEK293 cells depleted in MOZ expression. Analysis was performed different software as indicated.

Information content (bits) is shown on y-axis, position on x-axis. Data provided by Cintia Monteiro, Nigel Mongan and David Heery, unpublished. Software tools: MSigDB, MEME-AME, RSAT as cited in Chapter 2.6.

In contrast to pWHD1, while pWHD2 did undergo association with the dsDNA probe as shown by depletion of the free probe, no sequence specificity was apparent for pWHD2 for any of the sequences tested. Similarly, the interaction with DNA-ladder generated a more smeared shift, indicating increase of the molecular weight of the complex, but in contrast to pWHD1 did not produce defined band shifts. It is therefore possible that pWHD2 forms non-specific contacts with DNA, e.g. via the

phosphate sugar backbone, or that it recognizes a sequence that was not yet tested during this study. The known sequence conservation and structural alignment of pWHD2 with linker histones H1/5 supports this point as histones H1/5 also have to bind various sequences. While linker histones are specific for the conformation of nucleosomes in combination with linker DNA (Lyubitelev et al., 2016) and it was suggested that they prefer AT-rich sequences, no distinct recognition sequence was yet defined as most studies rely on the "Widom" sequences (Cui and Zhurkin, 2009). Furthermore, DNA-recognition of WHDs can be dependent on specific DNA conformations or tertiary structures, binding conditions or additional co-factors. pWHD2 may also not mainly function in dsDNA-binding after all, but recognize ssDNA, RNA or other proteins.

Interestingly, the tandem module containing both pWHD1 and pWHD2 was able to bind DNA sequences that were not recognized by just pWHD1. These findings indicate that the domains may function cooperatively in DNA-binding and provide different modes of interaction with target DNAs. Examples of other proteins that contain two or more WHDs show diverse functionality between the domains. Rpc34, a subunit of RNA polymerase III, contains a tandem WHD in its N-terminal half. It was shown that the second WH fold strongly binds dsDNA and is important for DNA-melting while the combined module of both WHDs was suggested to interact with different subunits of Pol III during PIC assembly (Wei and Chen, 2018). Another Pol III subunit, hRPC 62, comprises four WHDs which were all suggested to be important for sequence independent recognition of dsDNA as well as protein-protein interaction (Lefèvre et al., 2011). A tandem WHD in SMC complexes forms homo- or heterodimers via the first WHD (Palecek and Gruber, 2015). Furthermore, WHDs are often connected by flexible linkers and are overall flexible domains which further promotes functional versatility, e.g. different subunits of RNA polymerases harbouring two or more WHDs were found to have multiple structural formations (Teichmann et al., 2012). Thus, pWHD1 and pWHD2 could be engaged in DNA-binding via a variety of mechanisms. The domains may interact to form an extended DNA-

binding surface. Moreover, binding of one domain could impact the protein fold to promote binding of the second domain. DNA may also be bend upon binding of one domain to make it available for the second domain. In cellular contexts, interaction with co-factors and packaging of DNA in nucleosomes may also impact the functionality of the MOZ pWHDs. These possibilities and potential functions of pWHD2 will be interesting to study in the future.

Mutational analysis of pWHD1 suggested that the lysine rich stretch in the loop L1 between helices H1 and H2 is required for DNA-binding. Simultaneous disruption of four lysine residues by replacement with glutamates completely abolished protein-DNA complex formation. Reversing the charge of four residues may also impact protein fold and disrupt binding in a non-specific manner, but strong mutations of multiple residues or deletion of whole secondary structures were also used in other studies to map the interaction surfaces of WHDs (e.g. Choi et al., 2017). The finding that single disruption of K24 also shows reduction in DNA-binding further supports the assumption that this region is essential for DNA-recognition. To further define this interaction site, the residues Q23, Q25 and R26 should also be tested. Most likely, multiple contacts are made by different residues to form the protein-DNA complex as multiple lysine, arginine and glutamine residues are present in this region of pWHD1. These amino acids were also found in other proteins to be able to engage in DNA-binding (Liu et al., 2014; Li et al., 2017; Choi et al., 2017). As other WHDs, including linker histones H1 and H5, interact with DNA through different secondary structure components, it seems likely that the MOZ pWHD1 also forms more contacts in addition to the loop L1 region. Whether the loop is important for specific interaction with CpG motifs or provides general DNA-contacts has to be investigated in the further studies. To provide a comprehensive analysis of the DNA-binding surface, future mutational studies should involve the helix H3 as well the C-terminal wing region as these are typically involved in DNA-binding in other winged helix proteins. In addition, resolving the crystal structure of pWHD1 and pWHD2

bound to DNA could answer outstanding questions about the domain fold and how the protein binds DNA.

Chapter 6: Functional analysis of MOZ domains in cell-based models

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas9 (CRISPR associated enzyme 9) edited HEK293 cell lines depleted in *KAT6A*/MOZ expression were previously generated in our lab (Monteiro, Collins, Whitchurch, Mazumder, Heery, unpublished). The CRISPR-Cas9 system is a component of the natural bacterial immune systems that has been modified for application in genome engineering and is now widely used for high-throughput genome editing. In molecular biology, small homologous RNA sequences are used as guide RNAs which form complexes with the Cas9 endonuclease and guide the protein to target genomic sequences to allow locus specific DNA-editing. Our group used Cas9 nickase enzyme, which requires two single-guide (sg) RNAs to target opposing strands in the target sequence, thus creating double-stranded DNA breaks with very low to negligible off target effects (Shen et al., 2014). Following single cell sorting, clonal expansion and genotyping, clones were selected that targeted Exon 3 of the *KAT6A* gene in HEK293 cells causing depletion of MOZ protein expression. In clone G10 disruptive deletions of Exon 3 were confirmed for all alleles by genotyping, RNA-Seq and western blots. Even though the ATG start codon remained intact and deletions were in frame, western blots confirmed significant loss of MOZ expression and no production of truncated product (**Figure 6.1**). A control HEK293 clone that was only treated with one sgRNA was confirmed to be wild type and used as positive control (AF10). The clone maintained both sequence integrity as well as MOZ protein expression. In line with the known function of MOZ as a histone acetyltransferase, analysis showed that acetylation levels of target residues such as H3K14 and H3K9 were strongly reduced in clone G10. Furthermore, RNA-sequencing (RNA-seq) revealed a set of differentially expressed genes. These included known MOZ targets like *Hox* genes but also a variety

of other genes such as components of the cell metabolism (manuscript in preparation).

Previous data from our lab showed that exogenous expression of MOZ effectively rescued the MOZ phenotype in clone G10. Expression of differentially expressed genes was restored to normal levels. We now wanted to test whether this model system could also facilitate the elucidation of specific functions of single MOZ domains by introduction of domain specific mutations. Differences in the rescue effect of WT MOZ and mutant MOZ proteins could indicate in which regard single domains may impact the functionality of MOZ in histone acetylation, DNA-binding or transcriptional control.

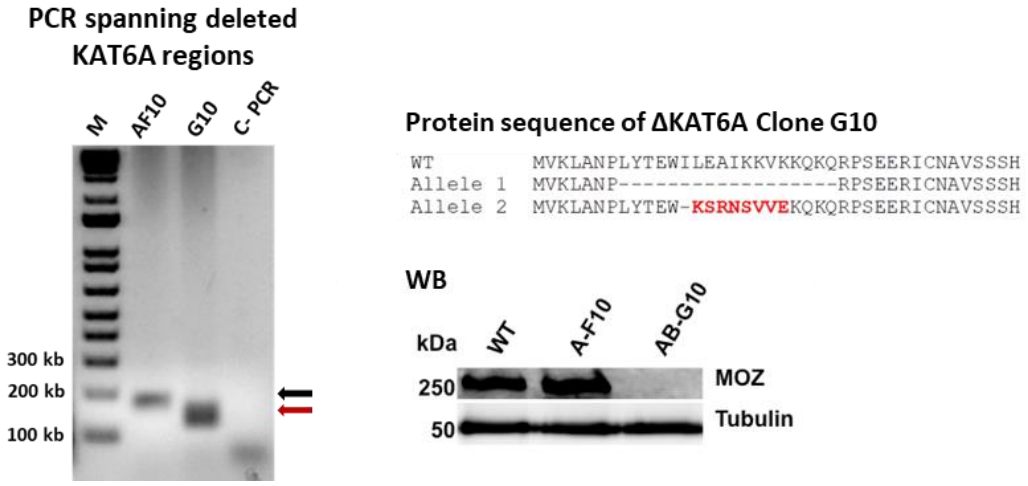


Figure 6.1 Genotyping of CRISPR edited HEK293 Δ KAT6A cells

PCR with genomic DNA using primers spanning the region around the *KAT6A* ATG start codon showed deletion of genomic regions in Clone G10 compared to the control AF10 (left) which was confirmed by sequencing (top right). Western blots showed significant loss of MOZ expression in Clone G10 (right bottom).

M=Marker; C- PCR= PCR negative control. Arrows mark expected bands. Sequencing and western blot data were generated and provided by Cintia Monteiro and David Heery.

The MYST (KAT) domain as well as the DPF domain are already well understood in regards of their structure and function. In addition, the results of the present study suggest that the N-terminus of MOZ is involved in DNA-binding. We therefore decided to focus our rescue experiments on these domains through introduction of point mutations that have been reported to abolish domain specific functions.

It was shown that the point mutations in the C2HC zinc finger (C543G) as well as in the acetyl-CoA binding motif (G657E) of the human MOZ MYST domain deplete KAT activity. Functionality of the domain was found to be required for the leukemogenic potential of MOZ-TIF2 fusion proteins. (Deguchi et al., 2003).

A recent study demonstrated that the MOZ DPF domain is a specific reader for H3K14 crotonylation as well as acetylation (Xiong et al., 2016). Crotonylation is a non-acetyl acylation mark that is particularly enriched at promoters and enhancers of transcriptionally active genes. The transcriptional activation effect may be even stronger for crotonyl modifications because of the increased rigidity and bulk of the group compared to acetyl (Sabari et al., 2015). Structural and mutational analysis showed that insertion of point mutations (S210M and D282A) in the MOZ DPF disrupt binding to H3K14cr (Xiong et al., 2016). *In vitro*, crotonyl binding-deficient full-length MOZ mutants show impaired co-localisation of MOZ with H3K14cr and reduced recruitment to promoter regions of MOZ target genes. It was also suggested that crotonylation is a better potentiator for transcriptional control by MOZ than acetylation (Xiong et al., 2016; Sabari et al., 2015). It was demonstrated that histone crotonylation and acetylation correlate with the level of available crotonyl/acetyl-CoA linking chromatin status and metabolic state of the cell (Sabari et al., 2015). Interestingly, recent results in our group using CRISPR edited cell lines depleted in MOZ expression suggest that MOZ is tightly involved in metabolic regulation as absence of MOZ causes mis-regulation of several metabolic genes (Monteiro and Heery, unpublished).

As described above, the first 79 MOZ amino acids in MOZ-TIF2 fusion proteins were found to be essential in leukemia development (Cheung et al., 2016). While exogenous expression of WT MOZ-TIF2 constructs induced cell transformation in HSCs, this was lost when MOZ-TIF2 was lacking the first 79 MOZ amino acids. Co-IP studies suggested that the N-terminal 79 MOZ amino acids are required for interaction with the lysine demethylase KDM4C as well as the arginine methyltransferase PRMT1 and suggested that this interaction is essential for leukemogenesis. Pharmacological inhibition of PRMT1 and KDM4C also disrupted cell transformation. However, fusing PRMT1 directly to these deletion constructs was not sufficient to rescue the leukemic potential (Cheung et al., 2016). Thus, the DNA-binding function of the MOZ N-terminus that was identified in this project may also be required. Alternatively, the pWHD1 might also have additional functions in protein-protein interaction. The NEMM domain of MORF, which shares >85% sequence identity with the MOZ pWHD1, is also maintained in fusions with KANSL1 in retroperitoneal leiomyoma further supporting potential functional relevance of the domain (Panagopoulos et al., 2015).

We selected different mutations to disrupt either the pWHD1 (NEMM), DPF or MYST (KAT) domain of MOZ in full-length YFP-FLAG-MOZ (YFM) and full-length YFP-HA-MOZ-TIF2 (YHMT2) constructs as listed in **Table 11** and shown in **Figure 6.2.** All constructs were previously cloned in the vector pEYFP-c1 (Clontech) by our group. It was previously demonstrated that eYFP-FLAG-tagged of MOZ and eYFP-HA-tagged MOZ-TIF2 function similary to WT MOZ and MOZ-TIF2 in terms of nuclear localisation, acetylation of H3 and AML1 dependent transcription in reporter assays (Kindle et al., 2005; Collins et al., 2006; Kindle et al., 2010).

Table 11 List of point mutations for YFP-FLAG-MOZ and YFP-HA-MOZ-TIF2.

The table shows the point mutations that were incorporated into the YFM and YHMT2 constructs using site-directed mutagenesis. YFM= eYFP-FLAG-MOZ; YHMT2= eYFP-HA-MOZ-TIF2.

Name of the construct	Position in MOZ	MOZ domain	Reference
YFM^{K18A}	18	pWHD1	Present study
YFM^{K19A}	19	pWHD1	Present study
YFM^{K-E}	19, 21, 22, 24	pWHD1	Present study
YFM^{S210M}	210	DPF	Xiong et al., 2016
YFM^{C233A}	233	DPF	Deeves, Heery unpub.
YFM^{D282A}	282	DPF	Xiong et al., 2016
YFM^{G657E}	657	MYST (KAT)	Deguchi et al., 2003
$YHMT2^{K19A}$	19	pWHD1	Present study
$YHMT2^{K-E}$	19, 21, 22, 24	pWHD1	Present study
$YHMT2^{S210M}$	210	DPF	Xiong et al., 2016
$YHMT2^{G657E}$	657	MYST (KAT)	Deguchi et al., 2003

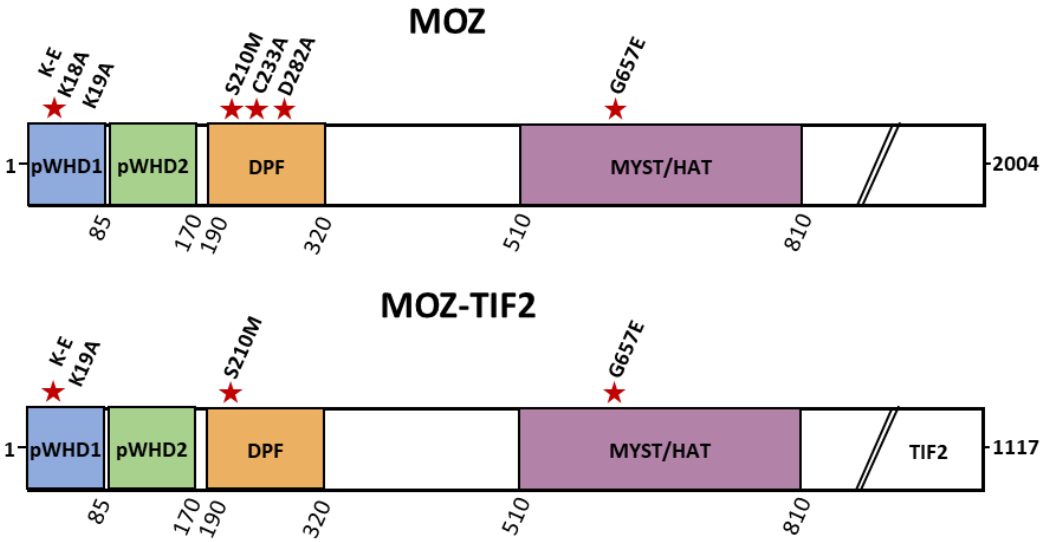


Figure 6.2 Schematic of positions of point mutations inserted into YFP-FLAG-MOZ and YFP-HA-MOZ-TIF2 constructs

Positions of point mutations that were incorporated into the YFM and YHMT2 constructs using site-directed mutagenesis are indicated by red stars. YFM= eYFP-FLAG-MOZ; YHMT2= eYFP-HA-MOZ-TIF2.

6.1 Generation of MOZ and MOZ-TIF2 mutagenesis constructs

Site-directed mutagenesis was used to introduce point mutations in the constructs YFP-FLAG-MOZ and YFP-HA-MOZ-TIF2 as described above (Chapter 2.3.13 and 5.5.1). An example of successful PCR and DpnI digest for the YFM^{K-E} is shown in (**Figure 6.3 A**). Incorporation of all mutations listed in **Table 11** was confirmed by sequencing (**Figure 6.3 B**; only the sequencing results for YFM are shown). Integrity and quality of all plasmids was further checked by restriction digest with XhoI and EcoRV. Agarose gel electrophoresis confirmed the release of the expected 1539 bp fragment as exemplarily shown in **Figure 6.3 C** for the constructs YFM$^{WT/K-E/S210M/G657E}$.

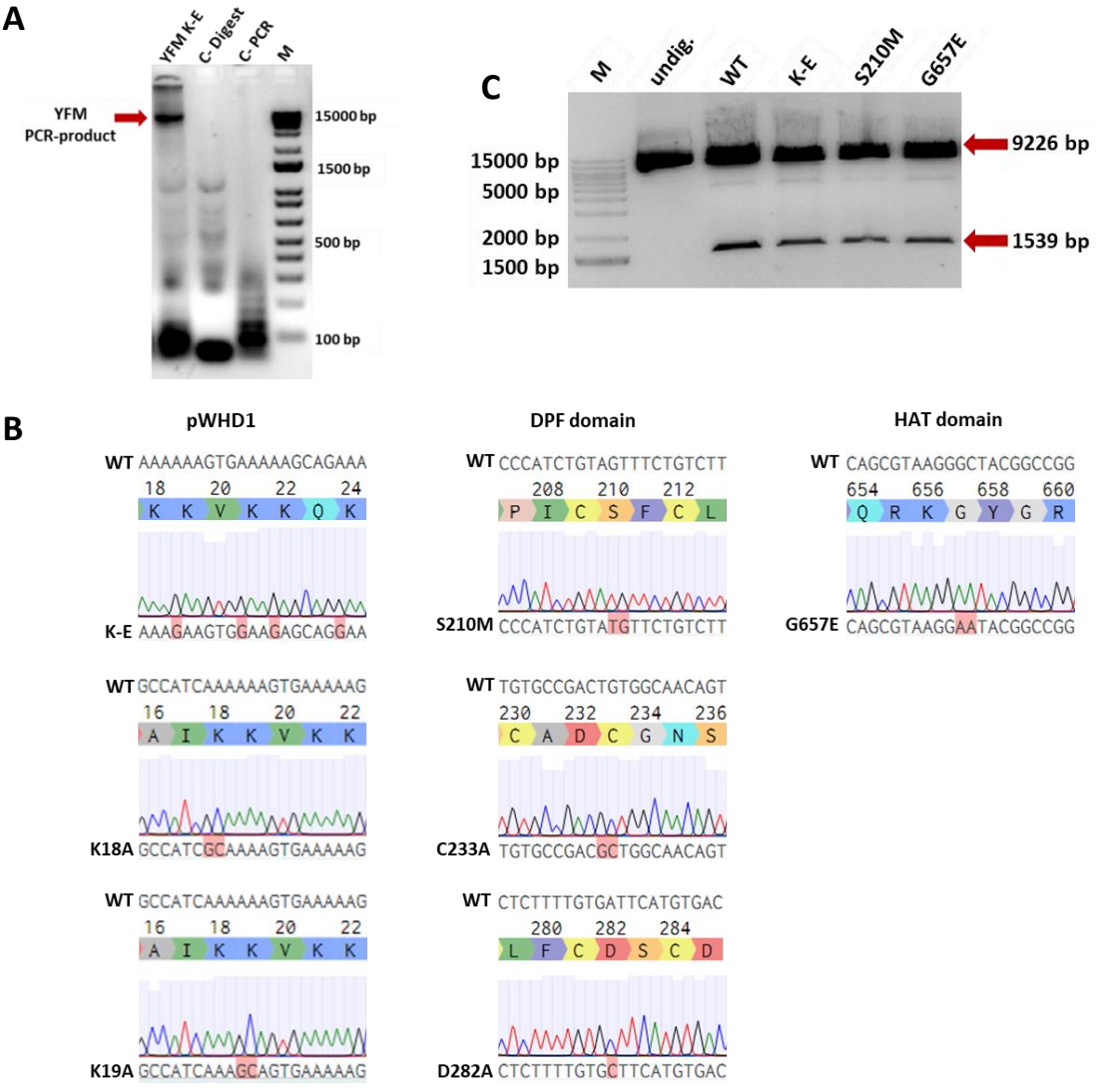


Figure 6.3 Generation of YFP-FLAG-MOZ and YFP-HA-MOZ-TIF2 mutagenesis constructs

A) Agarose gel electrophoresis showing the PCR-product of the YFM^{K-E} mutagenesis construct after DpnI digest. The WT DNA (control digest) was completely digested while the mutated PCR products were maintained. M=Marker; C- Digest= control digest of the input plasmid; C- PCR= PCR negative control PCR **B)** Sequencing electropherograms confirming the successful incorporation of different point mutations in the sequences of YFM plasmid DNA (Similar results were obtained for YHMT2, data not shown). The WT DNA and amino acid sequences are shown at the top. DNA sequences and electropherograms of the mutants are shown below, required codon changes for mutations are marked in red. **C)** Agarose gel after restriction digest of YFM constructs with XhoI and EcoRV. M=Marker; Arrows mark expected DNA bands.

6.2 Subcellular localisation of mutant MOZ and MOZ-TIF2

YFP-FLAG-MOZ and YFP-HA-MOZ-TIF2 mutants were first examined in regards of their subcellular localisation. HEK293 cells were seeded in Ibidi µ-Slide 8 well chambers and transfected with the full-length YFM and YHMT2 mutant constructs. At 24 h post transfection, cells were fixed in paraformaldehyde and visualised by confocal microscopy. Cell nuclei of chosen constructs were also stained with the DNA dye DRAQ-5.

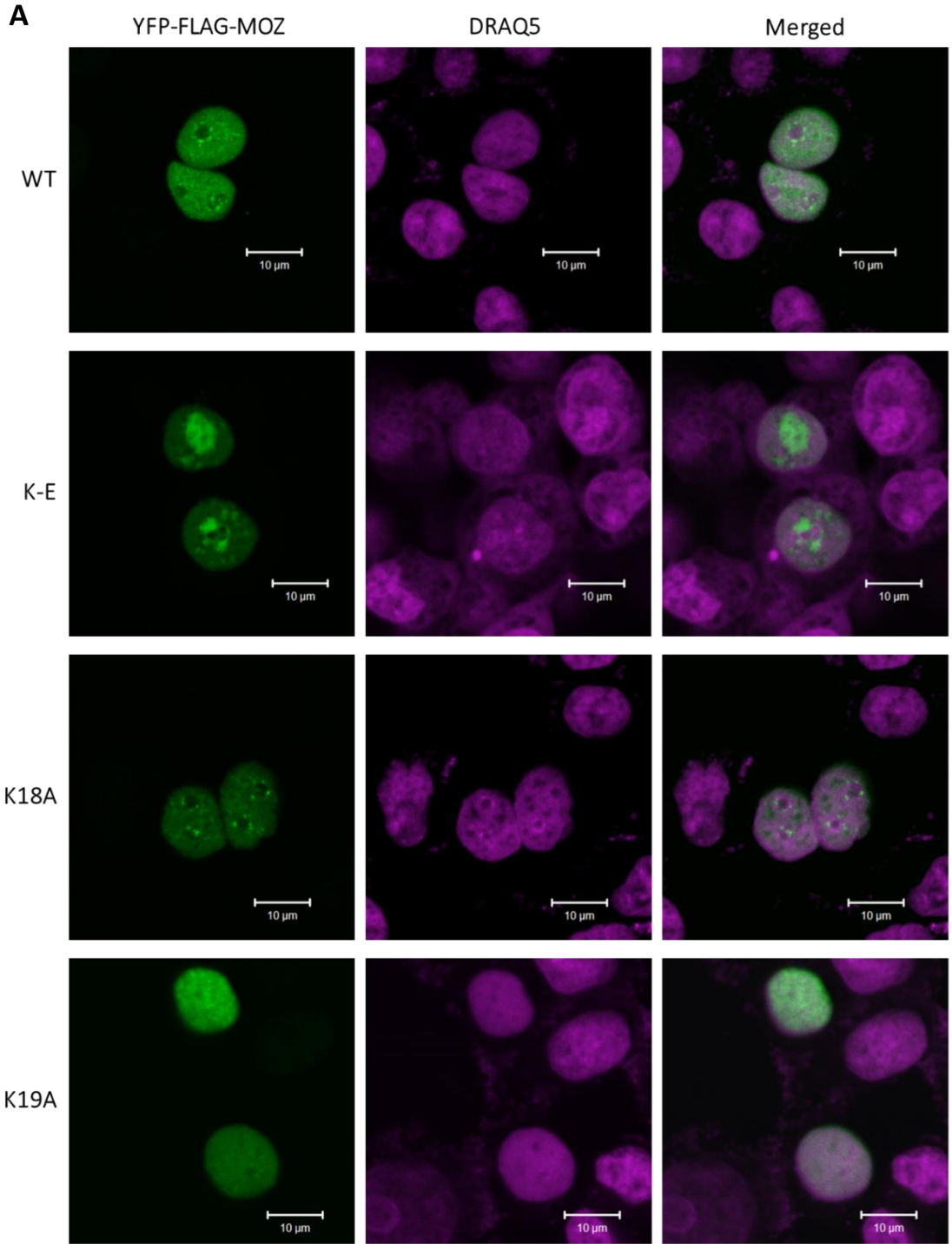
Figure 6.4 B shows the subcellular localisation of YFP-FLAG-MOZ constructs. In line with previous findings, the WT localised to the nucleus, as indicated by the overlap with DRAQ-5 staining, but was excluded from nucleoli, which were not stained by DRAQ-5. YFP-FLAG-MOZ was partially concentrated in subnuclear speckles which were frequently localised at the peripheries of nucleoli. The function of these speckles is not known, but they appear to form a distinct nuclear subdomain. Previous studies demonstrated that the speckles are independent from splicing speckles, Cajal bodies, paraspeckles and PML (promyelocytic leukemia protein) bodies (Kindle et al., 2005).

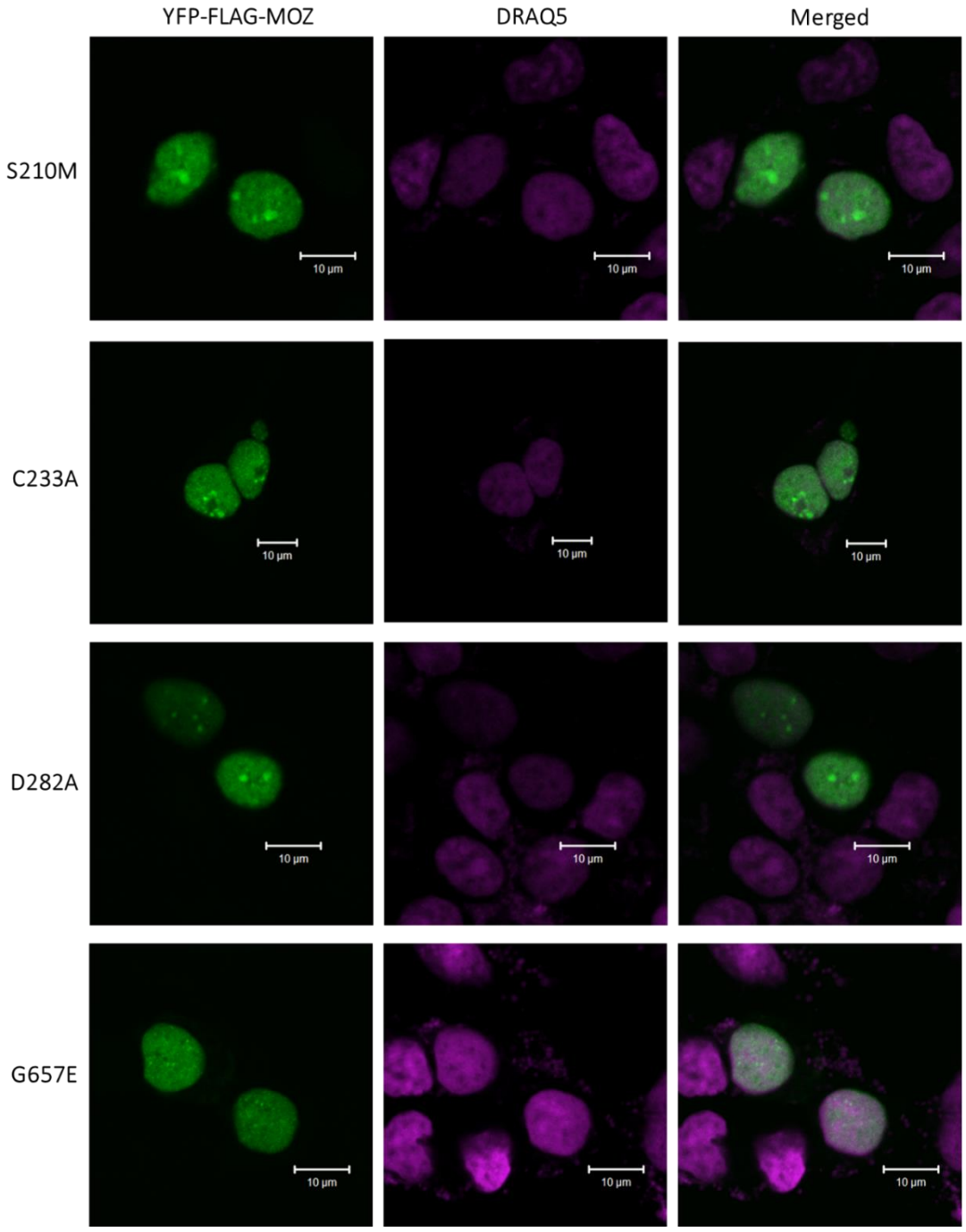
The mutations K18A in the pWHD1 and G657E in the MYST domain did not noticeably alter the subcellular localisation of YFP-FLAG-MOZ. However, this was expected after the substitution K18A also showed no effects in DNA-binding analysis (Chapter 5.5.2) and as G657E is not expected to impact binding to specific sites.

Constructs containing disruptions of residues that are required for recognition of H3K14cr marks (S210M and D282A) retained their even distribution in the nucleus and peri-nucleolar concentration in speckles. However, these speckles were increased in size compared to the WT. Similar results were found for the mutant C233A, which disrupts a zinc-binding motif essential for the folding of the DPF, but the size increase of MOZ foci was reduced compared to the other mutants. Substitution of K19 with alanine in pWHD1, which was shown to have partial effects on DNA-binding (Chapter 5.5), prevented formation of the typical MOZ speckles and the protein remained evenly distributed throughout the nucleus. The quadruple mutation (YFM^{K-E}), that was shown to completely disrupt DNA-binding ability of pWHD1, caused the formation of larger protein aggregates within the nucleus while YFM remained evenly distributed throughout the nucleus and excluded from nucleoli.

As shown in **Figure 6.4 B**, all YFP-HA-FLAG-MOZ-TIF2 displayed a similar distribution and localised within the nucleus. In line with previous literature findings, YHMT2 was concentrated around nucleolar regions and occasionally formed speckles similar to MOZ (Kindle et al., 2005). Not all mutant constructs were yet tested for their subcellular localisation but no alterations were found in mutant YHMT2 constructs so far.

Changes in subcellular localisation for constructs containing mutations in the DPF and pWHD1 domain could indicate that the protein may not be recruited to correct chromatin sites which may be investigated in future studies (discussed below). The formation of nuclear aggregates observed for YFM^{K-E} may indicate strong disruption of protein function and incorrect localisation which then causes aggregation. However, aggregation could also be caused by incorrect protein folding due to the incorporated mutations.


A
YFP-FLAG-MOZ
DRAQ5
Merged
WT
K-E
K18A
K19A
10 µm


YFP-FLAG-MOZ
DRAQ5
Merged
S210M
C233A
D282A
G657E
10 µm

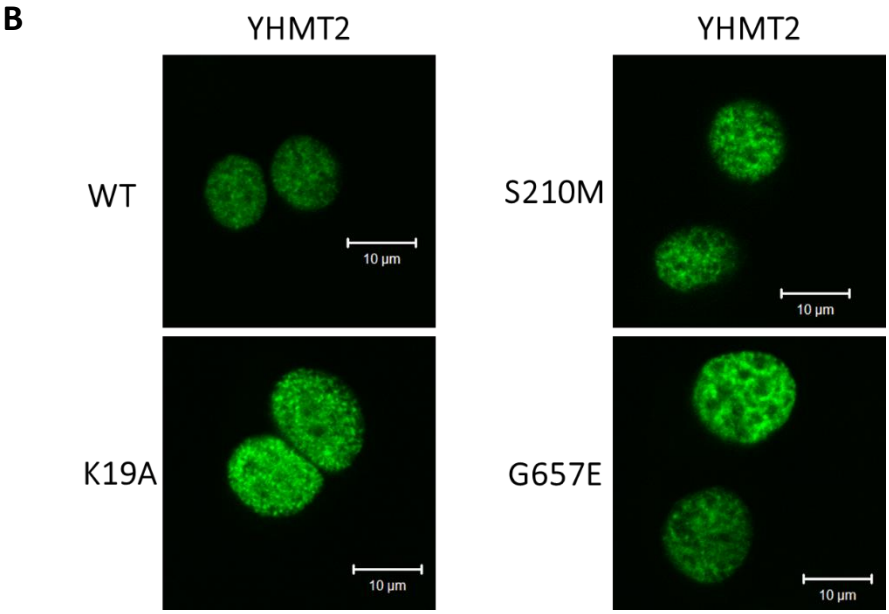


Figure 6.4 Subcellular localisation of MOZ and MOZ-TIF2 mutants

WT and mutant constructs of YFM **(A)** and YHMT2 **(B)** were expressed in HEK293 cells and imaged for YFP expression via confocal microscopy. DRAQ-5 was used as nuclear stain. White scale bar= 10 µm.

6.3 Rescue experiments in HEK293 CRISPR-Cas9 cell lines

As discussed in the introduction (Chapter 1.2), MOZ acetylates H3K9 and H3K14. Previous analysis showed, that these acetylation marks are reduced in clone G10. Through mutational analysis via rescue experiments, we now wanted to assess whether functionality of single domains is required to establish H3K9 and H3K14 acetylation. Furthermore, it was recently reported that the MOZ DPF recognizes H3K14 crotonylation (Xiong et al., 2016) and that MOF, another member of the MYST family, has crotonyltransferase activity (Liu at al., 2017). Thus, it was also tested whether depletion of MOZ or mutations impact H3K14 crotonylation. In addition, the mutant YFM$^{K\text{-}E}$ was tested for effects on the expression of genes that were previously found to be down regulated by depletion of MOZ in clone G10 in RNA-Seq analysis. Two genes of the *Hox* gene cluster, *HoxA5* and *HoxA13,* as well as one metabolic gene, *Succinate Dehydrogenase Complex Subunit D* (*SDHD*), were selected for testing.

The WT full-length construct YFP-FLAG-MOZ as well as the mutants YFM^{K-E}, YFM^{S210M} and YFM^{G657E} disrupting either the pWHD1, DPF or MYST domain were transfected in the MOZ depleted HEK293 clone G10. As controls, empty GFP vector (pVenus-N1) was expressed in the single guide control AF10 and clone G10. After optimisation of the transfection protocol to achieve high transfection rates, cells were seeded in 10 cm dishes and transfected on the following day with 14 µg of plasmid DNA via the calcium-phosphate method. Cells were harvested after further 48 h of incubation.

Western blot analysis using an anti-GFP antibody showed that the bulk expression levels mutant YFP-FLAG-MOZ in transfected cells were substantially lower compared to the WT construct (**Figure 6.5 A**). No significant differences in H3K14 crotonylation were observed in clone G10 compared to the control AF10. Rescue with different YFM constructs did also not induce strong alteration in crotonylation levels, only a small increase may be observed for WT YFM, suggesting that MOZ does not directly impact H3K14 crotonylation (**Figure 6.5 B**). As expected, clone G10 showed reduced acetylation of H3K14 (**Figure 6.5 B**). Exogenous expression of YFP-FLAG-MOZ rescued this effect and transfected cells even showed a slight increase in H3K14ac compared to the control AF10, perhaps due to higher expression than endogenous MOZ. The MOZ mutant constructs did not restore H3K14ac levels as effectively as the WT. While introduction of the YFM^{K-E} construct caused a slight rescue, acetylation levels appeared to be reduced for YFM^{S210M} and YFM^{G657E} compared to controls . H3K9 acetylation levels remained stable in G10 compared to the WT. Rescue with WT YFM further increased the levels of H3K9ac (**Figure 6.5 B**). In contrast to the findings for H3K14ac, all mutant YFP-FLAG-MOZ constructs also caused increase of H3K9 acetylation, but the effect was weaker compared to the WT.

Preliminary RT-qPCR analysis showed that exogenous expression of WT YFP-FLAG-MOZ caused a strong upregulation of *HoxA5, HoxA13* and SDHD genes compared to clone G10 that was transfected with GFP. This is consistent with the observed requirement of MOZ for expression of these genes in HEK293 cells as

detected by RNA-Seq (Monteiro et al., in preparation). The mutant construct YFM^{K-E} showed some rescue of gene expression but the effect was lower (**Figure 6.5 B**). However, this data requires more biological repeats and better experimental conditions as described below for validation.

However, because the western blots showed that mutant YFP-FLAG-MOZ were expressed at lower levels than the WT (**Figure 6.5 A**), it is not clear whether the effects on histone acetylation or gene expression are caused by disruption of domains or by lower levels of MOZ expression. The fact that H3K14ac levels appeared to be reduced rather than rescued after introduction of S210M and G657E compared to the control while both constructs were able to increase acetylation of H3K9 may indicate that the effect was caused by disruption of the DPF and MYST domain. However, these results require further validation and the method needs to be improved, but there was insufficient time to achieve this within the timeframe of the project.

The lower expression levels of mutant constructs may be caused by lower transfection efficiencies into HEK293 cells or reduced stability of either the produced mRNA or protein which could be related to reduced protein function. To investigate these possibilities and to potentially improve the method, cells were first sorted for GFP expression (Beckman Coulter MoFlo XDP) 48 h post transfection to ensure that a homogenous mass of transfected cells were used for all constructs. However, downstream experiments were not yet performed. Analysis of the sorted cells showed that the transfection efficiency of the WT constructs was indeed increased with approximately 44% compared to the mutant construct with around 19%-30% (**Figure 6.5 D, left**). In addition, the GFP expression levels were 6.5-10 fold higher for the WT compared to mutant constructs which showed relatively even expression levels (**Figure 6.5 D, right**). There is a number of possible explanations for these observations. The discrepancies in transfection efficiencies but especially in expression levels for different constructs explain the results of the previous western blot. These differences may be caused by lower exogenous mRNA levels produced

from the mutant constructs which could be analysed by RT-qPCR using MOZ specific primers. More likely, proteins were destabilised due to disruption of protein function by the integrated mutations. However, plasmid quality could also impact the expression efficiency. Even though all constructs were maxi-prepped in the same way and agarose gels showed no apparent differences in quality, it will be advisable to ensure freshly maxi-prepped plasmids with high quality for all constructs are used for future analysis.

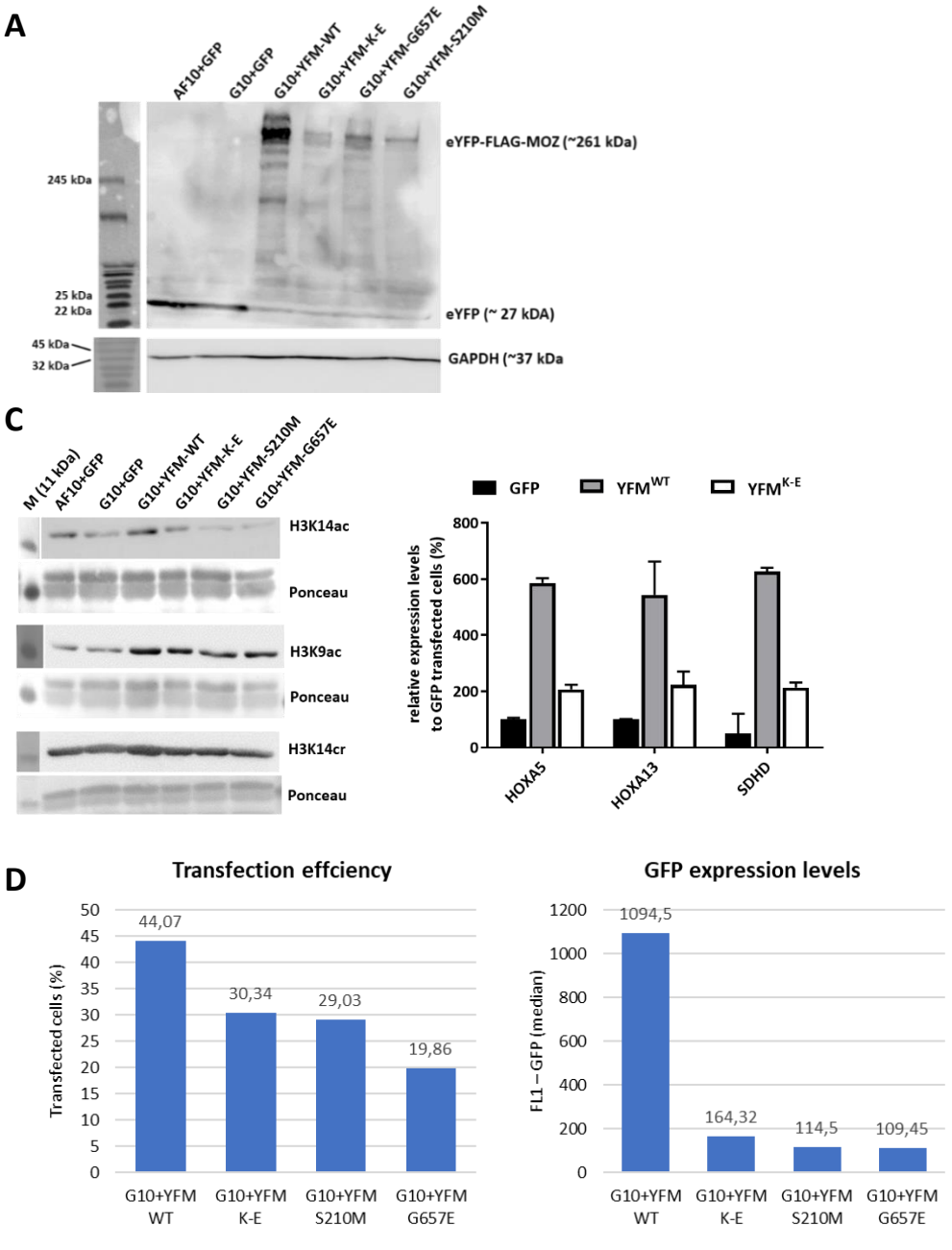


Figure 6.5 Rescue of histone acetylation and gene expression by exogenous MOZ expression

A) Western blot showing expression levels of exogenous WT and mutant YFM as well as GFP in the clones G10 and AF10 using anti-GFP antibody. **B)** Western blot showing the effect of exogenous expression of WT and mutant YFM on acetylation and crotonylation of histone H3 in clone G10 compared to GFP transfected controls using anti-H3K9ac, anti-H3K14ac and anti-H3K14cr antibodies. **C)** RT-qPCR-analysis for rescue effects of YFM^{WT} and $YFM^{K\text{-}E}$ in clone G10 assessing three genes that were previously shown to be downregulated; error bars= SEM. **D)** Analysis of transfection efficiency (left) and GFP expression (right) of YFM WT and mutant MOZ constructs by flow cytometry.
Ponceau stain and anti-GAPDH were used as loading controls.

6.4 Summary

A selection of substitution mutations were designed to functionally disable the pWHD1, DPF or MYST (KAT) domain of MOZ and to assess their specific functions in cell-based models. Two of the disruptions in the PHD fingers altered subcellular localisation of MOZ by causing an increase in the size of the nuclear speckles which were also observed for the WT. While previous studies were not able to identify the function of these speckles (Kindle et al., 2005), they appear to be similar to the punctate distribution of PcG proteins (Buchenau et al., 1998). PcG complexes such as PRC1 and 2 mainly mediate transcriptional repression via histone methylation. The PcG foci were shown to be located at sites of PcG-mediated silencing and identified as topologically associated domains (TADs). These are formed by nucleosome fibres folding into chromatin loops and resemble epigenetic domains characterised by specific combinations of chromatin marks. Intradomain chromatin contacts are formed at a much higher frequency within TADs and the structures were suggested to play important roles for PcG-mediated silencing e.g. of *HOX* gene clusters (Schuettengruber et al., 2017). Thus, it may be speculated that the nuclear speckles formed by MOZ also resemble domains with distinct chromatin marks that are important for epigenetic regulation.

Changes in the subnuclear localisation of DPF and pWHD1 mutants may indicate that recruitment to chromatin sites is impaired leading to perturbed

localisation. This would be in line with previous studies that demonstrated that the S210M and D282A mutations disrupt recruitment of MOZ to H3K14cr sites (Xiong et al., 2016). Thus, further microscopy-based analysis by immunostaining could reveal further discrepancies in recruitment to chromatin sites, e.g. H3K9/14ac or interaction with other proteins such as KDM4C, PRMT1 or BRPF1. However, formation of protein aggregates in the nucleus upon mutation of the pWHD1 could either be a sign of disrupted localisation and functionality of MOZ or indicate that simultaneous incorporation of the four mutations causes misfolding of the protein. Tests with additional constructs may clarify this effect. Different sets of mutations (e.g. K24E, Chapter 5.5.2) that also disrupt DNA-binding or deletion constructs that exclude the region of pWHD1 in full-length MOZ could be implemented for further analysis. Overall, nuclear localisation of MOZ proteins appears to be intact, although distribution in the nucleus is altered.

Rescue experiments using exogenous expression of MOZ to analyse domain specific functionality with mutagenesis constructs was limited by transfection and expression differences. The method will have to be optimised e.g. by ensuring similar plasmid qualities combined with cell sorting to generate even expression patterns for all constructs. The preliminary data acquired so far indicates that exogenous expression of WT MOZ sufficiently rescues defects in transcriptional activation and histone acetylation caused by CRISPR-mediated knockout of MOZ. The results may also indicate that H3K14ac is impaired by disruption of either the DPF or MYST domain while this effect was not as impactful for the pWHD1 mutant. As expected, H3K14 crotonylation was not impacted by exogenous MOZ expression. However, these results are only preliminary and require confirmation under better experimental conditions.

Chapter 7: Concluding remarks and future directions

The study of the histone acetyltransferase MOZ/KAT6A has both biological and clinical interest. MOZ is a key regulator of gene expression in hematopoietic stem cells and other tissues in both adult and developing organisms and is also involved in neurodevelopmental disorders and cancer. MOZ appears to function within numerous multi-subunit protein complexes acting as a co-factor and the catalytic subunit for acetylation of histones and other proteins. Earlier studies identified genes of the *HoxA* cluster as MOZ targets consistent with its role in regulation of development and tissue differentiation. From clinical perspective, *de novo* mutations of MOZ were recently linked to childhood-neurodevelopmental disorders associated with developmental delay. MOZ is also known for its links to cancer, in particular AML resulting from chromosomal abnormalities affecting the *KAT6A* gene. Given that previous studies demonstrated that the first 79 amino acids of MOZ located in the NEMM domain are essential for leukemogenesis in MOZ-fusion proteins (Cheung et al., 2016), this project focussed on characterising the NEMM domain comprised in the N-terminal 200 amino acids of MOZ and for the first time presented evidence that this region may represent a tandem winged-helix domain involved in DNA-binding.

Using recombinant proteins, our EMSA experiments indicate that both pWHD1 and pWHD2 bind dsDNA. Furthermore, it was demonstrated that the pWHD1 acts as a minimal binding domain that appeared to comprise sequence preference for CG-rich dsDNA containing CpG motifs. This finding is particularly interesting in the context of CpG-islands which are often present in promoter regions of genes with strong importance for transcriptional regulation. Motif analysis of genes that were differentially expressed after depletion of MOZ performed by our group also found enrichment of C-rich sequences comprising CpG-dinucleotides in MOZ target genes. Thus, future studies may further investigate the enrichment of MOZ at CpG promoter

sites through ChIP-Seq/qPCR analysis using WT and mutant forms of MOZ. Even though no sequence specific DNA recognition was found for the MOZ pWHD2 during this study, the results suggest that the domain is most likely also involved in DNA-binding. pWHD2 may stabilise DNA-recognition by non-specific interactions or recognize additional sequences that were not yet identified. The MOZ pWHDs could also cooperate as a tandem module and impact the DNA-binding mechanism of each other. The DNA-binding module comprised in the MOZ N-terminus may play an important role in recruitment of MOZ complexes to target promoter sites that are rich in CpG-Islands where it acetylates nucleosomes or recruits additional cofactors to modulate gene expression (**Figure 7.1**).

This study completed a first mutational assessment of the DNA-binding mechanism by the MOZ pWHD1 which indicated that a lysine-rich stretch located between helices 1 and 2 is important for DNA recognition. However, additional mapping supported by structural data e.g. by X-ray crystallography, will be required to comprehensively analyse the DNA-interaction surface of pWHD1 and pWHD2. Optimisation of large-scale purification during this project will enable further structural analysis to take place.

As previous results from our lab suggest that the tandem pWHDs increase affinity of MOZ to histones, further analysis of both the pWHDs and the DPF domain of MOZ will be required to unravel how recruitment of MOZ to target sites is facilitated in detail. In addition, the pWHD1 region of MOZ was found to be required for interaction with PRMT1 and KDM4C in MOZ-TIF2 fusion proteins which was directly linked to leukemogenic potential of the fusion. Both depletion of the MOZ region as well as pharmacological inhibition of PRMT1 and KDM4C disrupted cell transformation (Cheung et al., 2016). It remains to be established whether DNA-binding and cofactor recruitment of the MOZ pWHD1 are cooperative or mutually exclusive. It was previously demonstrated that the WHD in human RPC62 protein can mediate both DNA and protein binding (Lefévre et al., 2011). Alternatively, co-

localisation at genomic foci facilitated by pWHD1 could prompt complex formation between MOZ, PRMT1 and KDM4C.

Furthermore, the NEMM domain of MORF which shares >85% sequence identity with MOZ pWHD1 was also maintained in fusions with KANSL1 in retroperitoneal leiomyoma (Panagopoulos et al., 2015). The H1/5 region that resembles pWHD2 was also shown to be important for nuclear localisation of MOZ (Kitabayashi et al., 2001). Further *in vivo* analysis will be required to define the role of pWHD1 and pWHD2 in MOZ and MORF. So far, rescue experiments in this study using CRISPR edited cell lines to compare the impact of domain disruption on functions like histone acetylation or transcriptional activation by MOZ were limited by differences in the transfection efficiencies and expression levels between WT and mutant constructs. Whether exogenous expression of MOZ is limited because the protein is destabilised due to reduced functionality has to be determined in the future. It may also be possible that plasmid qualities impacted expression levels. It seems feasible that rescue experiments can be optimised in future projects and used to assess specific domain functions both for the normal MOZ protein as well as the disease related fusions. First preliminary results indicate that mutagenesis of the DPF and HAT domain directly affects acetylation levels of H3K14 and that MOZ generally does not impact crotonylation of H3K14 which is in line with literature findings (e.g. Deguchi et al., 2003; Xiong et al., 2016; Monteiro et al., in preparation)

As all mutant constructs localised to the nucleus, future experiments could investigate co-localisation of MOZ and MOZ-TIF2 mutants with target chromatin sites such as H3K9/14ac or interacting proteins e.g. KDM4C, PRMT1 or BRPF1. Furthermore, enrichment of WT and mutant MOZ at promoter sites of target genes such as *Hox* genes could be investigated by ChIP-qPCR using antibodies against the incorporated FLAG or YFP tag. Analysis of the DPF mutants already showed that co-localisation with H3K14cr and recruitment to promoters of target genes was impaired (Xiong et al., 2016). Thus, it would also be interesting to investigate the impact of depletion of pWHD1 either by mutagenesis or depletion constructs.

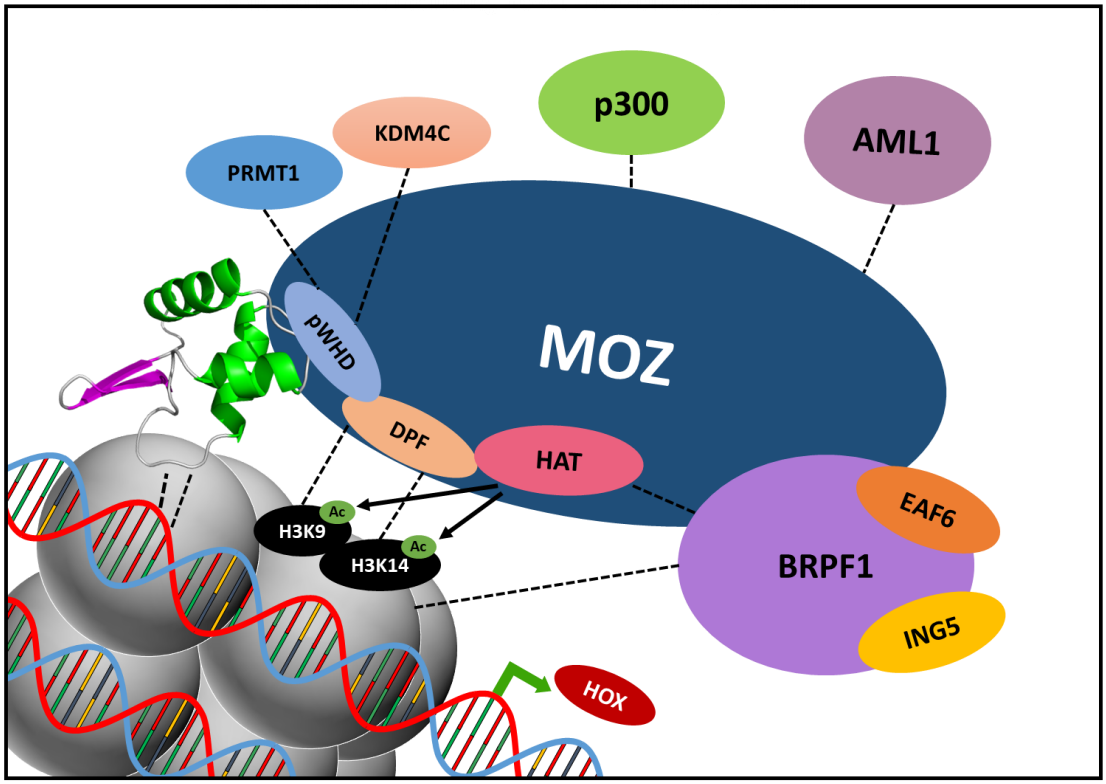


Figure 7.1 Schematic diagram showing potential functions of MOZ domains

Different readers domains such as the MOZ DPF guide MOZ complexes to chromatin where epigenetic modifications are facilitated. MOZ pWHDs may be involved in recruitment to target genes. Epigenetic modifications and co-factor recruitment modulate transcriptional activity e.g. of *Hox* genes.

Appendix

8.1 Plasmids

List of plasmids used in this study

Name	Insert	Vector	Mutation
6xHis-MOZpWHD1+2	His-MOZpWHD1+2	pET-15b	
6xHis-MOZpWHD1+2^{K18A}	His-MOZpWHD1+2	pET-15b	K18A
6xHis-MOZpWHD1+2^{K19A}	His-MOZpWHD1+2	pET-15b	K19A
6xHis-MOZ pWHD1	His-MOZ pWHD1	pET-15b	
6xHis-MOZ pWHD1^{K18A}	His-MOZ pWHD1	pET-15b	K18A
6xHis-MOZ pWHD1^{K-E}	His-MOZ pWHD1	pET-15b	K19/21/22/24E
6xHis-MOZ pWHD1^{K19E}	His-MOZ pWHD1	pET-15b	K19E
6xHis-MOZ pWHD1^{K21E}	His-MOZ pWHD1	pET-15b	K21E
6xHis-MOZ pWHD1^{K22E}	His-MOZ pWHD1	pET-15b	K22E
6xHis-MOZ pWHD1^{K24E}	His-MOZ pWHD1	pET-15b	K24E
6xHis-MOZ pWHD1^{K-A}	His-MOZ pWHD1	pET-15b	K19/21/22/24A
6xHis-MOZ pWHD1^{R44A}	His-MOZ pWHD1	pET-15b	R44A
6xHis-MOZ pWHD1^{K62A}	His-MOZ pWHD1	pET-15b	K62A
6xHis-MOZ pWHD1^{K72A}	His-MOZ pWHD1	pET-15b	K72A
6xHis-MOZ pWHD2	His-MOZ pWHD2	pET-15b	
YFP-FLAG-MOZ	FLAG-MOZ	pEYFP-c1	
YFP-FLAG-MOZ^{K18A}	FLAG-MOZ	pEYFP-c1	K18A
YFP-FLAG-MOZ^{K19A}	FLAG-MOZ	pEYFP-c1	K19A
YFP-FLAG-MOZ^{K-E}	FLAG-MOZ	pEYFP-c1	K19/21/22/24E
YFP-FLAG-MOZ ^{S210M}	FLAG-MOZ	pEYFP-c1	S210M
YFP-FLAG-MOZ^{C230A}	FLAG-MOZ	pEYFP-c1	C230A
YFP-FLAG-MOZ^{D282A}	FLAG-MOZ	pEYFP-c1	D282A
YFP-FLAG-MOZ^{G657E}	FLAG-MOZ	pEYFP-c1	G657E
YFP-HA-MOZ-TIF2	HA-MOZ-TIF2	pEYFP-c1	
YFP-HA-MOZ-TIF2^{K-E}	HA-MOZ-TIF2	pEYFP-c1	K19/21/22/24E
YFP-HA-MOZ-TIF2^{K19A}	HA-MOZ-TIF2	pEYFP-c1	K19A
YFP-HA-MOZ-TIF2^{S210M}	HA-MOZ-TIF2	pEYFP-c1	S210M
YFP-HA-MOZ-TIF2^{G657E}	HA-MOZ-TIF2	pEYFP-c1	G657A

8.2 Oligonucleotides

Table 12 Oligonucleotides for site-directed mutagenesis

Name	Sequence
MOZ G657E-F	CAATACCAGCGTAAGGAATACGGCCGGTTTCTCATCGATTTC
MOZ G657A-R	GAAATCGATGAGAAACCGGCCGTATTCCTTACGCTGGTATTG
MOZ C233A-F	GAACTCATCTCCTGTGCCGACTGCTGGCAACAGTGGCCATCCATCC
MOZ C233A-R	GGATGGATGGCCACTGTTGCCAGCGTCGGCACAGGAGATGAGTTC
MOZ C307A-F	CCAAAAGGCATGTGGATAGCTCAAATATGTCGACCTAGG
MOZ C307A-R	CCTAGGTCGACATATTTGAGCTATCCACATGCCTTTTGG
MOZ S210M-F	GAACCAATCCCCATCTGTATGTTCTGTCTTGGTACAAAAG
MOZ S210M-R	CTTTTGTACCAAGACAGAACATACAGATGGGGATTGGTTC
MOZ D282A-F	GGATAACATGCTCTTTTGTGCTTCATGTGACCGAGGTTTTC
MOZ D282A-R	GAAAACCTCGGTCACATGAAGCACAAAAGAGCATGTTATCC
MOZ K18A-F	GTGGATTTTGGAGGCCATCGCAAAAGTGAAAAAGCAGAAAC
MOZ K18A-R	GTTTCTGCTTTTTCACTTTTGCGATGGCCTCCAAAATCCAC
MOZ K19A-F	GATTTTGGAGGCCATCAAAGCAGTGAAAAAGCAGAAACAGC
MOZ K19A-R	GCTGTTTCTGCTTTTTCACTGCTTTGATGGCCTCCAAAATC
MOZ K19/21/22/24E-F	GGAGGCCATCAAAGAAGTGGAAGAGCAGGAACAGCGTCCTTCAG
MOZ K19/21/22/24E-R	CTGAAGGACGCTGTTCCTGCTCTTCCACTTCTTTGATGGCCTCC
MOZ K18E-F	GATTTTGGAGGCCATCGAAAAAGTGAAAAAGCAG
MOZ K18E-R	CTGCTTTTTCACTTTTTCGATGGCCTCCAAAATC
MOZ K19E-F	GATTTTGGAGGCCATCAAAGAAGTGAAAAAGCAGAAACAGC
MOZ K19E-R	GCTGTTTCTGCTTTTTCACTTCTTTGATGGCCTCCAAAATC
MOZ K21E-F	CCATCAAAAAAGTGGAAAAGCAGAAACAGCGTC
MOZ K21E-R	GACGCTGTTTCTGCTTTTCCACTTTTTTGATGG
MOZ K22E-F	CAAAAAAGTGAAAGAGCAGAAACAGCGTC
MOZ K22E-R	GACGCTGTTTCTGCTCTTTCACTTTTTTG
MOZ K24E-F	GAAAAAGCAGGAACAGCGTCCTTC
MOZ K24E-R	GAAGGACGCTGTTCCTGCTTTTTC
MOZ K19/21/22/24A-F	GGAGGCCATCAAAGCAGTGGCAGCGCAGGCACAGCGTCCTTCAG
MOZ K19/21/22/24A-R	CTGAAGGACGCTGTGCCTGCGCTGCCACTGCTTTGATGGCCTCC
MOZ R44A-F	CCATGGCTTGGATGCTAAAACTGTTTTAG
MOZ R44A-R	CTAAAACAGTTTTAGCATCCAAGCCATGG
MOZ K62A-F	GATGGAACAATTTTAGCAGTCTCAAATAAAGGACTC
MOZ K62A-R	GAGTCCTTTATTTGAGACTGCTAAAATTGTTCCATC
MOZ K72A-F	GACTCAATTCCTATGCAGATCCTGATAATC
MOZ K72A-R	GATTATCAGGATCTGCATAGGAATTGAGTC

Table 13 List of oligonucleotides for RT-qPCR. Primers for RT-qPCR were previously designed and tested by C. Monteiro (Monteiro, 2018).

Name	Sequence
HoxA5-F	GCGCAAGCTGCACATAAG
HoxA5-R	CGGTTGAAGTGGAACTCCT
HoxA13-F	CCTCTGGAAGTCCACTCTGC
HoxA13-R	GGTATAAGGCACGCGCTTC
SDHD-F	CTTGCTCTGCATGGACTATT
SDHD-R	CCCCATGAACATAGTCAGTAACAA
GAPDH-F	AGGTGAAGGTCGGAGTCAAC
GAPDH-R	ATGACAAGCTTCCCGTTCT

References

Alberts, B. (2015). *Molecular biology of the cell* (Sixth edition. ed.). New York, NY: Garland Science, Taylor and Francis Group.

Alfieri, C., Chang, L., Zhang, Z., Yang, J., Maslen, S., Skehel, M., and Barford, D. (2016). Molecular basis of APC/C regulation by the spindle assembly checkpoint. *Nature, 536*(7617), 431.

Ali, T., Coles, P., Stevens, T. J., Stott, K., and Thomas, J. O. (2004). Two homologous domains of similar structure but different stability in the yeast linker histone, Hho1p. *Journal of molecular biology, 338*(1), 139-148.

Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., and Shiekhattar, R. (2007). New nomenclature for chromatin-modifying enzymes. *Cell, 131*(4), 633-636.

Arboleda, V. A., Lee, H., Dorrani, N., Zadeh, N., Willis, M., Macmurdo, C. F., Manning, M. A., Kwan, A., Hudgins, L., and Barthelemy, F. (2015). De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay. *The American Journal of Human Genetics, 96*(3), 498-506.

Baell, J. B., Leaver, D. J., Hermans, S. J., Kelly, G. L., Brennan, M. S., Downer, N. L., Nguyen, N., Wichmann, J., McRae, H. M., and Yang, Y. (2018). Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth. *Nature, 560*(7717), 253.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic acids research, 37*(suppl_2), W202-W208.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology, 10*(5), 411-421.

Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research, 21*(3), 381.

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A. R., Papai, G., Reymer, A., Syed, S. H., Lone, I. N., Tonchev, O., and Crucifix, C. (2017). Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. *Molecular cell, 66*(3), 384-397. e388.

Bell, C. E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. *Nature structural and molecular biology, 7*(3), 209.

Biggs Iii, W. H. (2001). Identification and characterization of members of the FKHR (FOX O) subclass of winged-helix transcription factors in the mouse. *Mammalian Genome, 12*(6), 416-425.

Bolten, S. N., Rinas, U., and Scheper, T. (2018). Heparin: role in protein purification and substitution with animal-component free material. *Applied microbiology and biotechnology, 102*(20), 8647-8660.

Borrow, J., Stanton, V. P., Andresen, J. M., Becher, R., Behm, F. G., Chaganti, R. S., Civin, C. I., Disteche, C., Dubé, I., and Frischauf, A. M. (1996). The translocation t (8; 16)(p11; p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB–binding protein. *Nature genetics, 14*(1), 33.

Boura, E., Silhan, J., Herman, P., Vecer, J., Sulc, M., Teisinger, J., Obsilova, V., and Obsil, T. (2007). Both the N-terminal loop and wing W2 of the forkhead domain of transcription factor Foxo4 are important for DNA binding. *Journal of Biological Chemistry, 282*(11), 8265-8275.

Brent, M. M., Anand, R., and Marmorstein, R. (2008). Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure, 16*(9), 1407-1416.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D. J. (1998). The distribution of polycomb-group proteins during cell division and development in Drosophila embryos: impact on models for silencing. *The Journal of cell biology, 141*(2), 469-481.

Burley, S., and Roeder, R. (1996). Biochemistry and structural biology of transcription

factor IID (TFIID). *Annual review of biochemistry, 65*(1), 769-799.

Cardin, A. D., and Weintraub, H. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis: An Official Journal of the American Heart Association, Inc., 9*(1), 21-32.

Carlson, S., and Glass, K. C. (2014). The MOZ histone acetyltransferase in epigenetic signaling and disease. Journal of cellular physiology, 229(11), 1571-1574.

Cereseto, A., Manganaro, L., Gutierrez, M. I., Terreni, M., Fittipaldi, A., Lusic, M., Marcello, A., and Giacca, M. (2005). Acetylation of HIV-1 integrase by p300 regulates viral integration. *The EMBO journal, 24*(17), 3070-3081.

Cerfeletsky, C., Lippens, G., Ramakrishnan, V., Muyldermans, S., Segers, A., Wyns, L., Wodak, S. J., and Hallenga, K. (1994). Homo-and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of histone H5. *Biochemistry, 33*(37), 11079-11086.

Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H., and Barford, D. (2015). Atomic structure of the APC/C and its mechanism of protein ubiquitination. *Nature, 522*(7557), 450.

Cheung, N., Fung, T. K., Zeisig, B. B., Holmes, K., Rane, J. K., Mowen, K. A., Finn, M. G., Lenhard, B., Chan, L. C., and So, C. W. E. (2016). Targeting aberrant epigenetic networks mediated by PRMT1 and KDM4C in acute myeloid leukemia. *Cancer cell, 29*(1), 32-48.

Cheung, R. C. F., Wong, J. H., and Ng, T. B. (2012). Immobilized metal ion affinity chromatography: a review on its applications. *Applied microbiology and biotechnology, 96*(6), 1411-1420.

Chinen, Y., Taki, T., Tsutsumi, Y., Kobayashi, S., Matsumoto, Y., Sakamoto, N., Kuroda, J., Horiike, S., Nishida, K., and Ohno, H. (2014). The leucine twenty homeobox (LEUTX) gene, which lacks a histone acetyltransferase domain, is fused to KAT6A in therapy-related acute myeloid leukemia with t (8; 19)(p11; q13). *Genes, Chromosomes and Cancer, 53*(4), 299-308.

Cho, H. I., Kim, M. S., and Jang, Y. K. (2016). The BRPF2/BRD1-MOZ complex is involved in retinoic acid-induced differentiation of embryonic stem cells. *Experimental cell research, 346*(1), 30-39.

Choi, J., Bachmann, A. L., Tauscher, K., Benda, C., Fierz, B., and Müller, J. (2017). DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nature structural and molecular biology, 24*(12), 1039.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell, 9*(2), 279-289.

Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature, 364*(6436), 412.

Close, P., Creppe, C., Gillard, M., Ladang, A., Chapelle, J.-P., Nguyen, L., and Chariot, A. (2010). The emerging role of lysine acetylation of non-nuclear proteins. *Cellular and Molecular Life Sciences, 67*(8), 1255-1264.

Collins, H. M., Kindle, K. B., Matsuda, S., Ryan, C., Troke, P. J., Kalkhoven, E., and Heery, D. M. (2006). MOZ-TIF2 Alters Cofactor Recruitment and Histone Modification at the RARβ2 Promoter DIFFERENTIAL EFFECTS OF MOZ FUSION PROTEINS ON CBP-AND MOZ-DEPENDENT ACTIVATORS. *Journal of Biological Chemistry, 281*(25), 17124-17133.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic acids research, 16*(22), 10881-10890.

Crump, J. G., Swartz, M. E., Eberhart, J. K., and Kimmel, C. B. (2006). Moz-dependent Hox expression controls segment-specific fate maps of skeletal precursors in the face. *Development, 133*(14), 2661-2669.

Cui, F., and Zhurkin, V. B. (2009). Distinctive sequence patterns in metazoan and yeast nucleosomes: implications for linker histone binding to AT-rich and methylated DNA. *Nucleic acids research, 37*(9), 2818-2829.

Danino, Y. M., Even, D., Ideses, D., and Juven-Gershon, T. (2015). The core promoter: At the heart of gene expression. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1849*(8), 1116-1131.

Deaton, A. M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and development, 25*(10), 1010-1022.

Deguchi, K., Ayton, P. M., Carapeti, M., Kutok, J. L., Snyder, C. S., Williams, I. R., Cross, N. C., Glass, C. K., Cleary, M. L., and Gilliland, D. G. (2003). MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer cell, 3*(3), 259-271.

Di Croce, L. (2005). Chromatin modifying activity of leukaemia associated fusion proteins. *Human molecular genetics, 14*(suppl_1), R77-R84.

Doyon, Y., Cayrou, C., Ullah, M., Landry, A.-J., Côté, V., Selleck, W., Lane, W. S., Tan, S., Yang, X.-J., and Côté, J. (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Molecular cell, 21*(1), 51-64.

Dreveny, I., Deeves, S. E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H. M., and Heery, D. M. (2014). NAR Breakthrough Article: The double PHD finger domain of MOZ/MYST3 induces α-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. *Nucleic acids research, 42*(2), 822.

Dulak, A. M., Stojanov, P., Peng, S., Lawrence, M. S., Fox, C., Stewart, C., Bandla, S., Imamura, Y., Schumacher, S. E., and Shefler, E. (2013). Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature genetics, 45*(5), 478.

Fillenberg, S. B., Friess, M. D., Körner, S., Böckmann, R. A., and Muller, Y. A. (2016). Crystal structures of the global regulator DasR from Streptomyces coelicolor: implications for the allosteric regulation of GntR/HutC repressors. *PloS one, 11*(6), e0157691.

Fulton, D. L., Sundararajan, S., Badis, G., Hughes, T. R., Wasserman, W. W., Roach, J.

C., and Sladek, R. (2009). TFCat: the curated catalog of mouse and human transcription factors. *Genome biology, 10*(3), R29.

Gajiwala, K. S., and Burley, S. K. (2000). Winged helix proteins. *Current opinion in structural biology, 10*(1), 110-116.

Geiger, S. R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A. J., and Cramer, P. (2010). RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Molecular cell, 39*(4), 583-594.

Gilston, B. A., Wang, S., Marcus, M. D., Canalizo-Hernández, M. A., Swindell, E. P., Xue, Y., Mondragon, A., and O'Halloran, T. V. (2014). Structural and mechanistic basis of zinc regulation across the E. coli Zur regulon. *PLoS biology, 12*(11), e1001987.

Glatt, S., Alfieri, C., and Müller, C. W. (2011). Recognizing and remodeling the nucleosome. *Current opinion in structural biology, 21*(3), 335-341.

Goll, M. G., and Bestor, T. H. (2002). Histone modification and replacement in chromatin activation. *Genes and development, 16*(14), 1739-1742.

Gu, G., Wang, T., Yang, Y., Xu, X., and Wang, J. (2013). An improved SELEX-Seq strategy for characterizing DNA-binding specificity of transcription factor: NF-κB as an example. *PloS one, 8*(10), e76109.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature reviews Molecular cell biology, 19*(10), 621.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology, 166*(4), 557-580.

Hanahan, D., Jessee, J., and Bloom, F. R. (1991). [4] Plasmid transformation of Escherichia coli and other bacteria. In *Methods in enzymology* (Vol. 204, pp. 63-113): Elsevier.

Harami, G. M., Gyimesi, M., and Kovács, M. (2013). From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends in biochemical sciences, 38*(7), 364-371.

Hayashihara, K., Uchiyama, S., Shimamoto, S., Kobayashi, S., Tomschik, M.,

Wakamatsu, H., No, D., Sugahara, H., Hori, N., and Noda, M. (2010). The middle region of an HP1-binding protein, HP1-BP74, associates with linker DNA at the entry/exit site of nucleosomal DNA. *Journal of Biological Chemistry, 285*(9), 6498-6507.

Heery, D. M., and Fischer, P. M. (2007). Pharmacological targeting of lysine acetyltransferases in human disease: a progress report. *Drug discovery today, 12*(1-2), 88-99.

Hellman, L. M., and Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature protocols, 2*(8), 1849.

Holbert, M. A., Sikorski, T., Carten, J., Snowflack, D., Hodawadekar, S., and Marmorstein, R. (2007). The human monocytic leukemia zinc finger histone acetyltransferase domain contains DNA-binding activity implicated in chromatin targeting. *Journal of Biological Chemistry, 282*(50), 36603-36613.

Holden, N. S., and Tacon, C. E. (2011). Principles and problems of the electrophoretic mobility shift assay. *Journal of pharmacological and toxicological methods, 63*(1), 7-14.

Huang, H., Lin, S., Garcia, B. A., and Zhao, Y. (2015). Quantitative proteomic analysis of histone modifications. *Chemical reviews, 115*(6), 2376-2418.

Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D., and Zhao, Y. (2014). SnapShot: histone modifications. *Cell, 159*(2), 458-458. e451.

Jain, D., and Nair, D. T. (2012). Spacing between core recognition motifs determines relative orientation of AraR monomers on bipartite operators. *Nucleic acids research, 41*(1), 639-647.

Jin, B., Li, Y., and Robertson, K. D. (2011). DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes and cancer, 2*(6), 607-617.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews genetics, 13*(7), 484.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server.

Nature protocols, 7(8), 1511.

Katsumoto, T., Aikawa, Y., Iwama, A., Ueda, S., Ichikawa, H., Ochiya, T., and Kitabayashi, I. (2006). MOZ is essential for maintenance of hematopoietic stem cells. *Genes and development, 20*(10), 1321-1330.

Katsumoto, T., Yoshida, N., and Kitabayashi, I. (2008). Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis. *Cancer science, 99*(8), 1523-1527.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols, 10*(6), 845.

Khayrutdinov, B. I., Bae, W. J., Yun, Y. M., Lee, J. H., Tsuyama, T., Kim, J. J., Hwang, E., Ryu, K. S., Cheong, H. K., and Cheong, C. (2009). Structure of the Cdt1 C-terminal domain: conservation of the winged helix fold in replication licensing factors. *Protein Science, 18*(11), 2252-2264.

Kindle, K., Collins, H., and Heery, D. (2010). MOZ–TIF2-mediated destruction of CBP/p300 is blocked by calpain inhibitor 2. *Leukemia, 24*(7), 1359.

Kindle, K. B., Troke, P. J., Collins, H. M., Matsuda, S., Bossi, D., Bellodi, C., Kalkhoven, E., Salomoni, P., Pelicci, P. G., and Minucci, S. (2005). MOZ-TIF2 inhibits transcription by nuclear receptors and p53 by impairment of CBP function. *Molecular and cellular biology, 25*(3), 988-1002.

Kitabayashi, I., Aikawa, Y., Nguyen, L. A., Yokoyama, A., and Ohki, M. (2001). Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ–CBP fusion protein. *The EMBO journal, 20*(24), 7184-7196.

Kondo, M., Wagers, A. J., Manz, M. G., Prohaska, S. S., Scherer, D. C., Beilhack, G. F., Shizuru, J. A., and Weissman, I. L. (2003). Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annual review of immunology, 21*(1), 759-806.

Lalmansingh, A. S., Karmakar, S., Jin, Y., and Nagaich, A. K. (2012). Multiple modes of chromatin remodeling by Forkhead box proteins. *Biochimica et Biophysica*

Acta (BBA)-Gene Regulatory Mechanisms, 1819(7), 707-715.

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., and Weirauch, M. T. (2018). The human transcription factors. *Cell, 172*(4), 650-665.

Largeot, A., Paggetti, J., Broséus, J., Aucagne, R., Lagrange, B., Martin, R. Z., Berthelet, J., Quéré, R., Lucchi, G., and Ducoroy, P. (2013). Symplekin, a polyadenylation factor, prevents MOZ and MLL activity on HOXA9 in hematopoietic cells. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1833*(12), 3054-3063.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics, 32*(1), 42-56.

Lee, K. K., and Workman, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nature reviews Molecular cell biology, 8*(4), 284.

Lefèvre, S., Dumay-Odelot, H., El-Ayoubi, L., Budd, A., Legrand, P., Pinaud, N., Teichmann, M., and Fribourg, S. (2011). Structure-function analysis of hRPC62 provides insights into RNA polymerase III transcription initiation. *Nature structural and molecular biology, 18*(3), 352.

Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., and Bulyk, M. L. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature, 549*(7671), 287.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics, 27*(12), 1739-1740.

Liu, J., Huang, J., Zhao, Y., Liu, H., Wang, D., Yang, J., Zhao, W., Taylor, I. A., and Peng, Y.-L. (2014). Structural basis of DNA recognition by PCG2 reveals a novel DNA binding mode for winged helix-turn-helix domains. *Nucleic acids research, 43*(2), 1231-1240.

Liu, X., Wei, W., Liu, Y., Yang, X., Wu, J., Zhang, Y., Zhang, Q., Shi, T., Du, J. X., and Zhao, Y. (2017). MOF as an evolutionarily conserved histone

crotonyltransferase and transcriptional activation by histone acetyltransferase-deficient and crotonyltransferase-competent CBP/p300. *Cell Discovery, 3*, 17016.

Lyubitelev, A., Nikitin, D., Shaytan, A., Studitsky, V., and Kirpichnikov, M. (2016). Structure and functions of linker histones. *Biochemistry (Moscow), 81*(3), 213-223.

Marmorstein, R., and Zhou, M. M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol, 6*(7), a018762. doi:10.1101/cshperspect.a018762

Matsuda, S., Harries, J. C., Viskaduraki, M., Troke, P. J., Kindle, K. B., Ryan, C., and Heery, D. M. (2004). A conserved α-helical motif mediates the binding of diverse nuclear proteins to the SRC1 interaction domain of CBP. *Journal of Biological Chemistry, 279*(14), 14055-14064.

McLeay, R. C., and Bailey, T. L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics, 11*(1), 165.

Miller, C. T., Maves, L., and Kimmel, C. B. (2004). moz regulates Hox expression and pharyngeal segmental identity in zebrafish. *Development, 131*(10), 2443-2461.

Mitton, K. P., and Guzman, A. E. (2012). 5-Methylcytosine, a possible epigenetic link between ageing and ocular disease. *Experimental eye research, 96*(1), 2.

Monteiro, C. J. (2018). Function of MYST3 / MOZ: a histone acetyltransferase and chromatin regulator implicated in haematopoietic stem cell development and leukaemia. PhD thesis, University of Nottingham.

Murray, C. R., Abel, S. N., McClure, M. B., Foster II, J., Walke, M. I., Jayakar, P., Bademci, G., and Tekin, M. (2017). Novel causative variants in DYRK1A, KARS, and KAT6A associated with intellectual disability and additional phenotypic features. *Journal of pediatric genetics, 6*(02), 077-083.

Nguyen, N. T. T., Contreras-Moreira, B., Castro-Mondragon, J. A., Santana-Garcia, W., Ossio, R., Robles-Espinoza, C. D., Bahin, M., Collombet, S., Vincens, P., and

Thieffry, D. (2018). RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic acids research, 46*(W1), W209-W214.

Ono, K., Kusano, O., Shimotakahara, S., Shimizu, M., Yamazaki, T., and Shindo, H. (2003). The linker histone homolog Hho1p from Saccharomyces cerevisiae represents a winged helix–turn–helix fold as determined by NMR spectroscopy. *Nucleic acids research, 31*(24), 7199-7207.

Paggetti, J., Largeot, A., Aucagne, R., Jacquel, A., Lagrange, B., Yang, X., Solary, E., Bastie, J., and Delva, L. (2010). Crosstalk between leukemia-associated proteins MOZ and MLL regulates HOX gene expression in human cord blood CD34+ cells. *Oncogene, 29*(36), 5019.

Palecek, J. J., and Gruber, S. (2015). Kite proteins: a superfamily of SMC/kleisin partners conserved across bacteria, archaea, and eukaryotes. *Structure, 23*(12), 2183-2190.

Panagopoulos, I., Gorunova, L., Bjerkehagen, B., and Heim, S. (2015). Novel KAT6B-KANSL1 fusion gene identified by RNA sequencing in retroperitoneal leiomyoma with t (10; 17)(q22; q21). *PloS one, 10*(1), e0117010.

Perez-Campo, F. M., Costa, G., Lie-a-Ling, M., Kouskoff, V., and Lacaud, G. (2013). The MYST erious MOZ, a histone acetyltransferase with a key role in haematopoiesis. *Immunology, 139*(2), 161-165.

Peterson, C. L., and Laniel, M.-A. (2004). Histones and histone modifications. *Current Biology, 14*(14), R546-R551.

Pogo, B., Allfrey, V., and Mirsky, A. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America, 55*(4), 805.

Ramakrishnan, V., Finch, J., Graziano, V., Lee, P., and Sweet, R. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature, 362*(6417), 219.

Rhee, H. S., and Pugh, B. F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature, 483*(7389), 295.

Robson, M. I., Ringel, A. R., and Mundlos, S. (2019). Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Molecular cell, 74*(6), 1110-1122.

Sabari, B. R., Tang, Z., Huang, H., Yong-Gonzalez, V., Molina, H., Kong, H. E., Dai, L., Shimada, M., Cross, J. R., and Zhao, Y. (2015). Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Molecular cell, 58*(2), 203-215.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning: a laboratory manual*: Cold spring harbor laboratory press.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature reviews genetics, 8*(6), 424.

Satoh, C., Maekawa, R., Kinoshita, A., Mishima, H., Doi, M., Miyazaki, M., Fukuda, M., Takahashi, H., Kondoh, T., and Yoshiura, K.-i. (2017). Three brothers with a nonsense mutation in KAT6A caused by parental germline mosaicism. *Human Genome Variation, 4*, 17045.

Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome regulation by polycomb and trithorax: 70 years and counting. *Cell, 171*(1), 34-57.

Selmer, M., and Su, X. D. (2002). Crystal structure of an mRNA-binding fragment of Moorella thermoacetica elongation factor SelB. *The EMBO journal, 21*(15), 4145-4153.

Sheikh, B. (1), Phipson, B., El-Saafin, F., Vanyai, H., Downer, N., Bird, M., Kueh, A., May, R., Smyth, G., and Voss, A. (2015). MOZ (MYST3, KAT6A) inhibits senescence via the INK4A-ARF pathway. *Oncogene, 34*(47), 5807.

Sheikh, B. N., and Akhtar, A. (2018). The many lives of KATs—detectors, integrators and modulators of the cellular environment. *Nature reviews genetics*, 1.

Sheikh, B. N. (2), Lee, S. C., El-Saafin, F., Vanyai, H. K., Hu, Y., Pang, S. H. M., Grabow, S., Strasser, A., Nutt, S. L., and Alexander, W. S. (2015). MOZ regulates B-cell

progenitors and, consequently, Moz haploinsufficiency dramatically retards MYC-induced lymphoma development. *Blood, 125*(12), 1910-1921.

Sheikh, B. N., Yang, Y., Schreuder, J., Nilsson, S. K., Bilardi, R., Carotta, S., McRae, H. M., Metcalf, D., Voss, A. K., and Thomas, T. (2016). MOZ (KAT6A) is essential for the maintenance of classically defined adult hematopoietic stem cells. *Blood, 128*(19), 2307-2318.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., and Huang, X. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature methods, 11*(4), 399.

Sheppard, H. M., Harries, J. C., Hussain, S., Bevan, C., and Heery, D. M. (2001). Analysis of the steroid receptor coactivator 1 (SRC1)-CREB binding protein interaction interface and its importance for the function of SRC1. *Molecular and cellular biology, 21*(1), 39-50.

Shima, H., Yamagata, K., Aikawa, Y., Shino, M., Koseki, H., Shimada, H., and Kitabayashi, I. (2014). Bromodomain-PHD finger protein 1 is critical for leukemogenesis associated with MOZ–TIF2 fusion. *International journal of hematology, 99*(1), 21-31.

Sobel, M., Soler, D., Kermode, J., and Harris, R. (1992). Localization and characterization of a heparin binding domain peptide of human von Willebrand factor. *Journal of Biological Chemistry, 267*(13), 8857-8862.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., and Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545-15550.

Sun, X.-J., Man, N., Tan, Y., Nimer, S. D., and Wang, L. (2015). The role of histone acetyltransferases in normal and malignant hematopoiesis. *Frontiers in oncology, 5*, 108.

Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012). Structural and functional

aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription, 3*(1), 2-7.

Tham, E., Lindstrand, A., Santani, A., Malmgren, H., Nesbitt, A., Dubbs, H. A., Zackai, E. H., Parker, M. J., Millan, F., and Rosenbaum, K. (2015). Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. *The American Journal of Human Genetics, 96*(3), 507-513.

Thomas, T., Corcoran, L. M., Gugasyan, R., Dixon, M. P., Brodnicki, T., Nutt, S. L., Metcalf, D., and Voss, A. K. (2006). Monocytic leukemia zinc finger protein is essential for the development of long-term reconstituting hematopoietic stem cells. *Genes and development, 20*(9), 1175-1186.

Thomas, T., and Voss, A. K. (2007). The diverse biological roles of MYST histone acetyltransferase family proteins. *Cell Cycle, 6*(6), 696-704.

Tora, L., and Timmers, H. T. M. (2010). The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Trends in biochemical sciences, 35*(6), 309-314.

Torrent, M., Nogués, M. V., Andreu, D., and Boix, E. (2012). The “CPC clip motif”: a conserved structural signature for heparin-binding proteins. *PloS one, 7*(8), e42692.

Tsai, K.-L., Huang, C.-Y., Chang, C.-H., Sun, Y.-J., Chuang, W.-J., and Hsiao, C.-D. (2006). Crystal structure of the human FOXK1a-DNA complex and its implications on the diverse binding specificity of winged helix/forkhead proteins. *Journal of Biological Chemistry, 281*(25), 17400-17409.

Turner-Ivey, B., Guest, S. T., Irish, J. C., Kappler, C. S., Garrett-Mayer, E., Wilson, R. C., and Ethier, S. P. (2014). KAT6A, a chromatin modifier from the 8p11-p12 amplicon is a candidate oncogene in luminal breast cancer. *Neoplasia, 16*(8), 644-655.

Ullah, M., Pelletier, N., Xiao, L., Zhao, S. P., Wang, K., Degerny, C., Tahmasebi, S., Cayrou, C., Doyon, Y., and Goh, S.-L. (2008). Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Molecular and cellular biology, 28*(22), 6828-6843.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular cell, 45*(4), 439-446.

Vanyai, H. K., Thomas, T., and Voss, A. K. (2015). Mesodermal expression of Moz is necessary for cardiac septum development. *Developmental biology, 403*(1), 22-29.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nature reviews genetics, 10*(4), 252.

Vassylyeva, M. N., Klyuyev, S., Vassylyev, A. D., Wesson, H., Zhang, Z., Renfrow, M. B., Wang, H., Higgins, N. P., Chow, L. T., and Vassylyev, D. G. (2017). Efficient, ultra-high-affinity chromatography in a one-step purification of complex proteins. *Proceedings of the National Academy of Sciences, 114*(26), E5138-E5147.

Voss, A. K., Collin, C., Dixon, M. P., and Thomas, T. (2009). Moz and retinoic acid coordinately regulate H3K9 acetylation, Hox gene expression, and segment identity. *Developmental cell, 17*(5), 674-686.

Voss, A. K., Vanyai, H. K., Collin, C., Dixon, M. P., McLennan, T. J., Sheikh, B. N., Scambler, P., and Thomas, T. (2012). MOZ regulates the Tbx1 locus, and Moz mutation partially phenocopies DiGeorge syndrome. *Developmental cell, 23*(3), 652-663.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature, 431*(7010), 873.

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., and Carr, M. (2006). Structural diversity in p160/CREB-binding protein coactivator complexes. *Journal of Biological Chemistry, 281*(21), 14787-14795.

Wei, Y.-Y., and Chen, H.-T. (2018). Functions of the TFIIE-related tandem winged-helix domain of Rpc34 in RNA polymerase III initiation and elongation. *Molecular*

and cellular biology, 38(4), e00105-00117.

Weigelt, J., Climent, I., Dahlman-Wright, K., and Wikström, M. (2000). Letter to the Editor: 1H, 13C and 15N resonance assignments of the DNA binding domain of the human Forkhead transcription factor AFX. *Journal of biomolecular NMR, 17*(2), 181-182.

Wiedemann, C., Szambowska, A., Häfner, S., Ohlenschläger, O., Gührs, K.-H., and Görlach, M. (2015). Structure and regulatory role of the C-terminal winged helix domain of the archaeal minichromosome maintenance complex. *Nucleic acids research, 43*(5), 2958-2967.

Xiong, X., Panchenko, T., Yang, S., Zhao, S., Yan, P., Zhang, W., Xie, W., Li, Y., Zhao, Y., and Allis, C. D. (2016). Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. *Nature chemical biology, 12*(12), 1111.

Xu, Q., Rife, C. L., Carlton, D., Miller, M. D., Krishna, S. S., Elsliger, M. A., Abdubek, P., Astakhova, T., Chiu, H. J., and Clayton, T. (2009). Crystal structure of a novel archaeal AAA+ ATPase SSO1545 from Sulfolobus solfataricus. *Proteins: Structure, Function, and Bioinformatics, 74*(4), 1041-1049.

Yan, Y., Harper, S., Speicher, D. W., and Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nature structural and molecular biology, 9*(11), 862.

Yang, X., and Ullah, M. (2007). MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene, 26*(37), 5408.

Yang, X. J. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic acids research, 32*(3), 959-976.

Yang, X. J. (2015). MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease. *Biochim Biophys Acta, 1853*(8), 1818-1826. doi:10.1016/j.bbamcr.2015.04.014

Yu, L., Liang, Y., Cao, X., Wang, X., Gao, H., Lin, S.-Y., Schiff, R., Wang, X.-S., and Li, K. (2017). Identification of MYST3 as a novel epigenetic activator of ERα frequently amplified in breast cancer. *Oncogene, 36*(20), 2910.

Zhang, S., Chang, L., Alfieri, C., Zhang, Z., Yang, J., Maslen, S., Skehel, M., and Barford, D. (2016). Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature, 533*(7602), 260.

Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., and Pagano, M. (2002). Structure of the Cul1–Rbx1–Skp1–F box Skp2 SCF ubiquitin ligase complex. *Nature, 416*(6882), 703.

Zhou, B.-R., Feng, H., Kato, H., Dai, L., Yang, Y., Zhou, Y., and Bai, Y. (2013). Structural insights into the histone H1-nucleosome complex. *Proceedings of the National Academy of Sciences, 110*(48), 19390-19395.

Zhou, B.-R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T. S., and Bai, Y. (2015). Structural mechanisms of nucleosome recognition by linker histones. *Molecular cell, 59*(4), 628-638.

Zhovmer, A., Oksenych, V., and Coin, F. (2010). Two sides of the same coin: TFIIH complexes in transcription and DNA repair. *The Scientific World JOURNAL, 10*, 633-643.

Zhu, J., Sammons, M. A., Donahue, G., Dou, Z., Vedadi, M., Getlik, M., Barsyte-Lovejoy, D., Al-Awar, R., Katona, B. W., and Shilatifard, A. (2015). Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature, 525*(7568), 206.

Zwaveling-Soonawala, N., Maas, S. M., Alders, M., Majoie, C. B., Fliers, E., van Trotsenburg, A. P., and Hennekam, R. C. (2017). Variants in KAT6A and pituitary anomalies. *American Journal of Medical Genetics Part A, 173*(9), 2562-2565.