Structural and functional studies of a histone deacetylase-mediated gene silencing in Arabidopsis

LI JIKUN (B. Sc)
A0066424H

A thesis submitted to Department of Biological Sciences

The National University of Singapore

In partial fulfillment for the Degree of Master of Sciences

2013

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this thesis.

This thesis has also not been submitted for any degree in any university previously.

LI JIKUN

10 July 2013

Acknowledgement

I would like to express my greatest gratitude to my supervisors, A/P He Yuehui and A/P Adam Yuan for giving me an opportunity to work on this interesting project and giving me guidance in the field of Structural Biology and Plant Biology. They have been instrumental in providing me much technical and moral support throughout my post-graduate studies. I would like to express my appreciation to all my co-workers in the lab for the kind encouragement, constant help and support.

Special thanks to my thesis committee members Prof Yu Hao, A/P J. Sivaraman and Dr Lin Qingsong for their constructive criticism and advice during my Qualifying Examination and the preparation of this thesis.

Table of Contents

Summary

In *Arabidopsis thaliana*, HISTONE DEACETYLASE 6 (HDA6) plays a crucial role in gene silencing. To further understand the molecular mechanisms underlying HDA6-mediated gene silencing, I first analysed HDA6 structure by modelling, based on the human HISTONE DEACETYLASE 2 (HDAC2) crystal structure. This allowed me to visualize the interaction surface of HDA6 by mapping the residues directly involved in ligand interaction. I also analysed and identified several interaction partners of HDA6 via yeast two-hybrid assays and further characterised its known partners including FVE and MULTICOPY SUPRESSOR OF IRA (MSI5). Structural modelling of MSI1 and FVE was contrasted against the solved structure of p55 to shed light onto why FVE and MSI5 are partners with HDA6 within a histone deacetylase (HDAC) complex but not MSI1, which is a component of the PRC2 complex. I finally analysed the genetic interaction of *HDA6* with one of its partners in gene silencing and found that these two genes act largely non-additively to silence their targets.

List of Tables

List of Figures

List of abbreviations

3D	three-dimensional
AG	AGAMOUS
AGL	AGAMOUS-LIKE
AGO4	AGONAUTE4
AP1	APETALA 1
BLAST	Basic Local Alignment Search Tool
CAF-1	CHROMATIN ASSEMBLY FACTOR-1
CASP	Critical Assessment of protein Structure Prediction
cDNA	complementary DNA
CMT3	CHROMOMETHYLASE 3
CO	CONSTANS
co-IP	co-immunoprecipitation
Col-WT	Columbia wild type
DCL-3	DICER-LIKE 3
DMSO	dimethyl sulfoxide
DNMT	DNA methyltransferase
dsRNA	double stranded RNA
DTT	dithiothreitol
E(z)	Enhancer of zeste
E. Coli	Escherichia coli
EDTA	ethylene-diamine-tetra-acetate
EMF	EMBRYONIC FLOWER
Esc	EXTRA SEX COMBS

FIE	FERTILIZATION-INDEPENDENT ENDOSPERM
FIS	FERTILIZATION-INDEPENDENT SEED
FLC	FLOWERING LOCUS C
FLD	FLOWERING LOCUS D
FOG-1	FRIEND OF GATA 1
FRI	FRIGIDA
FT	FLOWERING LOCUS T
GA	gibberellin
GI	GIGANTEA
GSH	Reduced glutathione
GSSG	Oxidized glutathione
GST	glutatione S-transferase
HDAC	histone deacetylase complex
HMM	hidden Markov models
HMT	Histone Methyltransferase
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth
Ler	Landsberg erecta
MCS	multiple cloning site
MES	2-(N-morpholino)ethanesulfonic acid
MET1	METHYLTRANSFERASE 1
mRNA	messenger RNA
MS	Murashige and Skoog
MSI	MULTICOPY SUPRESSOR OF IRA

NAD	nicotinamide-adenine dinucleotide
N-CoR	NUCLEAR RECEPTOR COREPRESSOR
NMR	nuclear magnetic resonance
NuRD	NUCLEOSOME REMODELLING AND HISTONE DEACETYLASE
PBS	phosphate buffered saline
PcG	POLYCOMB GROUP
PCR	polymerase chain reaction
PDB	protein data bank
PHD-PRC2	PLANT HOMEODOMAIN / POLYCOMB REPRESSIVE COMPLEX 2
PMSF	phenylmethylsulfonylfluoride
PRC	POLYCOMB REPRESSIVE COMPLEX
PSI-BLAST	Position-Specific Iterated BLAST
PVDF	polyvinylidene difluoride
RbAp	RETINOBLASTOMA ASSOCIATED PROTEIN
RBBP	RETINOBLASTOMA BINDING PROTEIN
RdDM	RNA-directed DNA Methylation
RdRP	RNA-dependent RNA polymerase
RPD3	REDUCED POTASSIUM DEPENDENCY
RT-PCR	reverse transcription polymerase chain reaction
SCOP	Structural Classification of Proteins
SD	short day
SDS	sodium dodecylsulphate

SDS-PAGE	SDS polyacrylamide gel electrophoresis
siRNA	small interfering RNA
SMRT	SILENCING MEDIATOR FOR RETINOID AND THYROID
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
Su(z)12	SUPPRESSOR OF ZESTE 12
SUVR5	SUPPRESSOR OF VARIEGATION 3-9-RELATED 5
TASSER	Threading/ASSembly/Refinement
TE	transposable elements
TEK	TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK
Tris	tris (hydroxymethyl)-aminomethane
VRN2	VERNALIZATION 2
Y2H	yeast two-hybrid

Chapter 1 Literature Review

1.1 Introduction on structural biology

Life is extremely complex and in order to understand life, we need to investigate at multiple levels. Biology came a long way from the early days of looking at cells under the microscope since they were viewed as the most basic building blocks of life. Through investigation of cells on how they behave and organise themselves, we hoped to understand life at the multicellular level. However, reality proved to be much more complicated and people begin looking into cells for chemical pathways and individual molecules to answer some really difficult questions.

Structural biology is a fairly newcomer in the field of biology where it seeks to understand basic biology at fundamental chemistry and physics at the molecular level. This goes further in depth beyond the cellular level where the interactions between the individual molecules manifest themselves into biological pathways we observe. Critically, behaviour and structure of proteins in 3-Dimension is key to understanding them at a molecular level due to an inseparable link between structure and function.

Crucial to understanding elemental and intrinsic biological functions and characteristics of reproduction, metabolism, mutations and specificity of cells, structural biology can also help us conceptualize how cells organise and communicate in order to form tissue, and subsequently into the higher level of

organs and eventually the entire organism in terms of complex multicellular eukaryotes. While the three-dimensional (3D) structures of all biological molecules are important, it is that of proteins that garnered the most attention as they provided the most challenges and answers to that of biological functions.

A simple way to describe proteins is that they are chains of amino acids chemically bonded together through what we call peptide bonds. Proteins come in all shapes and sizes with a huge array of functions such as fibrous proteins that provide structural integrity, membrane proteins that are largely involved in cellular transport while globular proteins are involved in various chemical processes.

As a linear polypeptide, it is largely uninteresting and serves neither specific function nor take part in any reactions of significance. Only when proteins are properly folded into their specific 3D structures, upon which they can carry out their specific functions. For centuries, the actual structures of these proteins have stumped biologists and chemists since the most powerful light microscopes are unable to shed any light on them visually. The first hints about protein structures begin to surface only when the late 1950s for which Kendrew and Perutz would eventually be duly awarded the Nobel Prize in Chemistry in 1962 "for their studies of the structures of globular proteins".

1.2 Protein folding

In order to carry out their biological functions, proteins must be folded into specific spatial conformations. Ultimately, proteins derived their 3D structure

through folding in space from their primary sequences. We approach protein folding by seeking to understand the "folding code" and the "folding pathway", both of which describes the mechanistic question of primary amino acid sequence determining its native 3D structure and the kinetic question of the route taken by a protein achieve its final native structure respectively.

The composition of different permutations of amino acid residues differentiates one protein from the others and it has been deduced that the "folding code" must be embedded within its primary sequence. The unique amino acid side-chains of each of the twenty naturally occurring amino acids would provide folding instructions in the form of inter and intra molecular interactions such as hydrophobic interactions, Van der Waals-interactions, electrostatic interactions and hydrogen bonding. The "folding code" argues that each amino acid acts as a single "instructional unit" with its identity conferred by the nature of its side-chain and that the location of each amino acid residue in the primary sequence of a protein is equivalent to the logical placement of a statement among others in a computer source code. In this analogy, the "execution" of the folding code will result in the folding of the polypeptide chain into its native structure based on inter and intra molecular forces determined by the primary amino acid sequence. Famous experiments performed on ribonuclease (RNase) by Anfinsen and colleagues showed that the fully denatured RNase could spontaneously refold in vitro, indistinguishable from native proteins in terms of enzymatic activity (Anfinsen et al., 1961; White, 1961; Haber and Anfinsen, 1962).

1.3 Introduction on Robetta, Phyre2 and Pro-sp3-TASSER modelling

There are many tools freely available on the web for academic use for the purpose of structural modelling and predictions. These tools have proved to be invaluable to biologists who seek to uncover the potential structures of their protein of interest which structures have yet to be solved via protein crystallization or NMR. With the ever enhancements to their prediction algorithms, these tools would be of greater importance as techniques and computing power improves over time. The bi-annual Critical Assessment of protein Structure Prediction (CASP) experiment for protein structure prediction seeks to test the structure prediction methods from different institutions and groups so as to conduct an independent assessment of the state of the art in protein structure modelling to the research community and software users. Consistently among the top performers are Robetta hosted at the Baker lab, Phyre 2 hosted at Imperial College, London and Pro-sp3-TASSER hosted at Georgia Tech.

The Robetta full-chain protein structure prediction server hosted at http://robetta.bakerlab.org/ is one of the best publicly available protein structure prediction services (Aloy et al., 2003; Kim et al., 2004). In modelling protein structure, the program first searches for structural homologs using BLAST, PSI-BLAST, and 3D-Jury, then parses the target sequence into its individual domains, or independently folding units of proteins, by matching the sequence to structural

families in the Pfam database. Domains with structural homologs then follow a "template-based model" (i.e., homology modelling) protocol.

The Baker laboratory's in-house alignment program, K*sync, produces a group of sequence homologs, and each of these is modelled by the Rosetta de novo method to produce a decoy (possible structure). The final structure prediction is selected by taking the lowest energy model as determined by a low-resolution Rosetta energy function. For domains that have no detected structural homologs, a de novo protocol is followed in which the lowest energy model from a set of generated decoys is selected as the final prediction. These domain predictions are then connected together to investigate inter-domain, tertiary-level interactions within the protein. Finally, side-chain contributions are modelled using a protocol for Monte Carlo conformational search (Chivian et al., 2005).

The Phyre and Phyre2 servers hosted at Imperial College, London predict the 3D structure of a target protein based on homology modelling via sequence alignment, where a related sequence of known structure is used as a template (Kelley and Sternberg, 2009). From the currently existing database of proteins with known structures in Structural Classification of Proteins (SCOP) supplemented by new depositions in Protein Data Bank (PDB), the primary sequence of these known proteins are extracted and submitted to Position-Specific Iterated BLAST (PSI-BLAST) to gather sequence homologs against a non-redundant sequence database which would be used to create a large database of Hidden Markov Models (HMM).

When the input of the primary sequence of the query protein is submitted to the server, it would search a non-redundant sequence database to look for homology using five iterations of PSI-BLAST which would provide an accurate secondary structure prediction and powerful sequence profile. The results of this alignment is then converted into a Hidden Markov Model (HMM) which captures the mutational propensities at each position in the protein to derive an evolutionary fingerprint. The query HMM is then scanned against the HMM database using a profile–profile alignment algorithm for which the alignments will be ranked and scored. Based on the structures of the top ten highest scoring alignments from the HMM database as a template, full 3D models of the query protein would be created.

Pro-sp3-TASSER hosted at Georgia Tech is a modification on the treading step of the Threading/ASSembly/Refinement (TASSER) method to derive an algorithm that combines successful approaches such as meta-servers, coupling alignment with quality assessment, and iterative refinement (Zhou et al., 2009; Zhou and Skolnick, 2009).

Step A consist of threading and alignment by the newly developed PRO-SP3 threading method; step B is the generation of an ensemble of models by short TASSER simulations; step C is the selection of top models with TASSER-QA from the ensemble, followed by full TASSER refinement and final model selection (**Figure 1**).

The PRO-SP3 threading algorithm consists of five threading scores that align the target sequence to templates which were subsequently ranked independently with scores taken directly or modified from SP3 (Zhou and Zhou, 2005) and PROSPECTOR_3 (Skolnick et al., 2004).

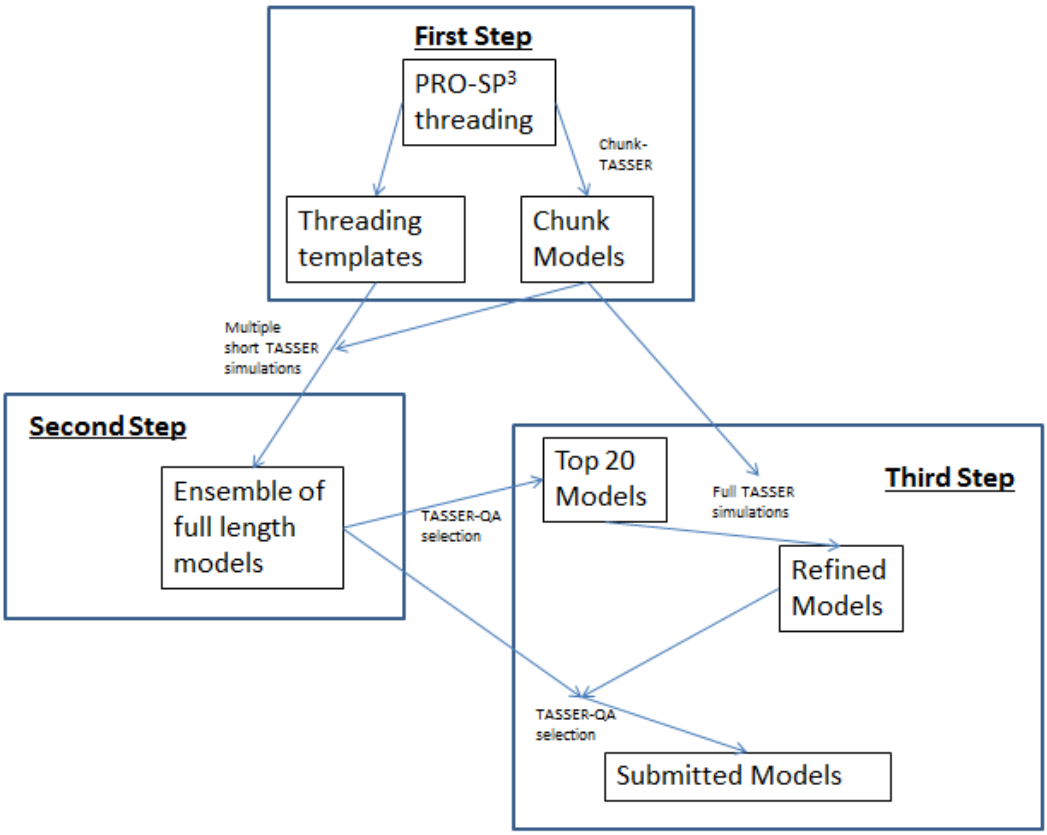

Figure 1. An overview of Pro-sp3-TASSER. This figure is adapted from (Zhou et al., 2009).

1.4 Epigenetical gene silencing

Epigenetics is a relatively new area in the field of biology. Despite the understanding of DNA structure and the genetic code, many biological mechanisms still elude the understanding of biologists as the conventional dogma failed to answer these questions. One example would be the basis of cellular differentiation where all the cells in the body contain the same DNA and yet have such diverse structure and function. Epigenetical controls provided answers on how gene expressions can be regulated on a molecular level, controlling spatial and temporal expression of different genes using a wide range of mechanisms.

This complex action of gene silencing requires the coordination of many key players including DNA methylases, histone deacetylases and histone methyltransferases. Chromatin modifications such as nucleosome remodelling, DNA methylation and modifications to histones can result in changes in the chromatin structure, upon which gene expression can be regulated (Henikoff and Shilatifard, 2011).

Among many developmental processes, expression for genes that regulate flower development is under epigenetical regulations. Polycomb Repressive Complex 2(PRC2) subunits CURLY LEAF (CLF), EMBRYONIC FLOWER (EMF) 2 and FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) repress the expression of *FLOWERING LOCUS C (FLC)*. CLF is a homolog of Enhancer of zeste [E(z)] which repress *AGAMOUS (AG)*, *FLC* and *FLOWERING LOCUS T*

(FT) via Histone 3 Lysine 27 trimethylation (H3K27me3) to prevent H3K4me3 and thus controlling flowering (Goodrich et al., 1997; Schubert et al., 2006). By repressing *FLC* and *FT*, the two key floral transition integrator and regulator, PRC2 complexes are critical in the control of flowering in *Arabidopsis* (Jiang et al., 2008).

Epigenetic regulations of transposable elements (TEs) and TE-like repeat sequences is very important as the failure of TE silencing can result in rearrangements and mutations of genes due to TE insertions, compromising genomic integrity alongside the temporal and spatial expression of genes (Gazzani et al., 2003; Liu et al., 2004; Law and Jacobsen, 2010). One key protein that participates in silencing TEs and TE-like sequence containing genes, such as FLC and FWA, is the AT-hook DNA binding protein, TRANSPOSABLE ELEMENT SILENCING VIA AT-HOOK (TEK) (Soppe et al., 2000; Gazzani et al., 2003; Liu et al., 2004). TEK directly binds to an FLC-repressive regulatory region and the silencing repeats of FWA and associates with Arabidopsis homologs of the Retinoblastoma-associated protein 46/48 (RbAp46/48), FVE and MSI5, which mediate histone deacetylation (Ausin et al., 2004; Gu et al., 2011). The absence of TEK activity would result in increased histone acetylation, reduced H3K9me2 and DNA hypomethylation in TE and TE-like repetitive DNA sequences (Zaratiegui et al., 2007; Saze and Kakutani, 2011), which ultimately leads to the upregulation of FLC and FWA as well as TE reactivation and subsequently, transposition(Xu et al., 2013a; Xu et al., 2013b).

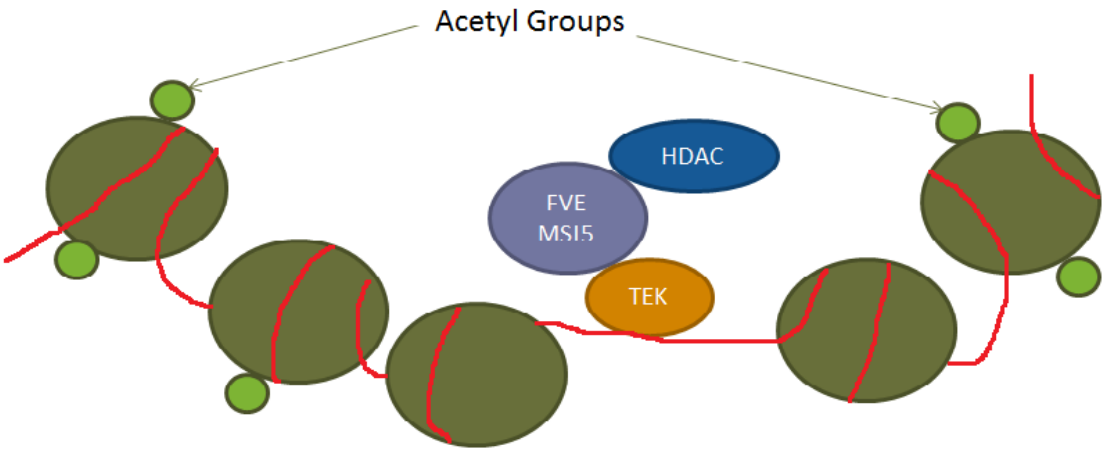

Figure 2. Binding of TEK on targets affect chromatin conformation and its association with FVE/MSI5-containing histone deacetylation complex results in repressive modification and subsequently, gene silencing. This figure is adapted from (Xu et al., 2013b).

Epigenetical changes are heritable changes to gene expression that does not involve any changes to the DNA sequence. Among the many epigenetical controls of gene expression, common mechanisms involved are DNA methylation, chromatin modifications, loss of imprinting and non-coding RNA (Hamilton, 2011). Epigenetic regulation has long-term and wide-ranging effects such as transcriptional regulation, cell cycle progression, and developmental events on top of other biological processes.

1.5 Histones and chromatin

There are two main classes of mammalian histone deacetylase complexes (HDAC) based on their sequence similarity to that of yeast HDACs (**Figure 3**). Class I HDACs includes HDAC 1, 2, 3, 8 and 11 while class II includes HDAC 4, 5, 6, 7, 9 and 10. Class I HDACs contain a well-conserved catalytic domain (Khochbin and Wolffe, 1997). HDAC1 and HDAC2 were known to be important components of two multi-protein complexes known as Sin3/HDAC and nucleosome remodelling and histone deacetylase complex (NuRD)/Mi2/NRD (Knoepfler and Eisenman, 1999; de Ruijter et al., 2003) while HDAC3 appears to be a nuclear receptor co-repressor (Ahringer, 2000). Studies have also suggested that HDAC3 is a member of the stable core of the Silencing Mediator for Retinoid and Thyroid (SMRT) receptors and/or Nuclear Receptor Co-Repressor (N-CoR) complexes (Huang et al., 2000; Li et al., 2000; Urnov et al., 2000; Wen et al., 2000).

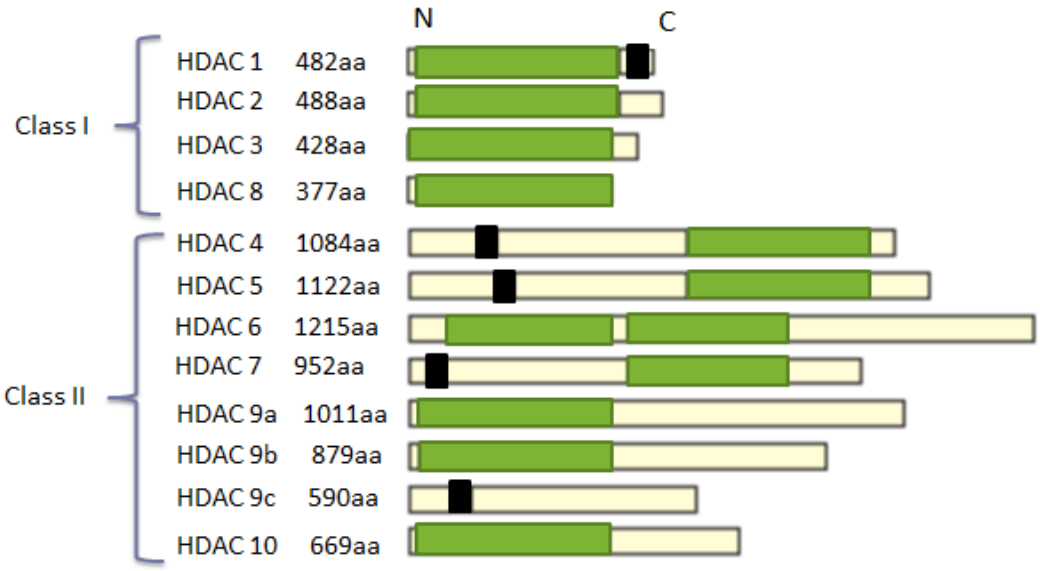

Figure 3. A schematic representation of Class I and Class II Histone Deacetylases isoforms. The black regions depict a Nuclear Localisation Signal with N and C denoting the N and C terminus respectively. This figure is adapted from (de Ruijter et al., 2003).

HDAC2 is ubiquitously expressed in different tissues and while it lacked a DNA binding domain (Brunmeir et al., 2009; Ma and Schultz, 2013), it plays a key role in the deacetylation of lysine residues at the N-terminal regions of core histones (H2A, H2B, H3 and H4) through the formation the core histone deacetylase complex with RBBP4 and RBBP7. Removal of this acetyl group is broadly associated with gene repression whereby HDAC2 forms transcriptional repressor complexes via interaction with proteins such as zinc-finger transcription factors. HDAC2 also interacts in the late S-phase of DNA-replication with DNA methyltransferase1 (DNMT1) in the other transcriptional repressor complex composed of CHROMATIN ASSEMBLY FACTOR-1 (CAF1).

In order for HDAC2 to play its multiple roles in different transcriptional regulation activities such as cell cycle progression, cell differentiation, apoptosis and tissue specifications, it requires other components of the HDAC complex to bind to histones or chromatin. There are multiple complexes of such nature and each are identified by the specific components associated with them and the different chromatin modifications catalysed (**Figure 4**).

One such protein is FRIEND OF GATA 1 (FOG-1) which is an essential cofactor in many complexes such as that of NuRD, Histone methyltransferases (HMT) and HDACs. FOG-1 forms heterodimers with the GATA family of transcription factors which would bind to the chromatin via its zinc finger domains while the N-terminal of FOG-1 recruits the various transcription regulatory complexes. This heterodimer formed by FOG-1 can either promote or

repress transcriptional activities, depending on the cell type or the promoter region it is bound to.

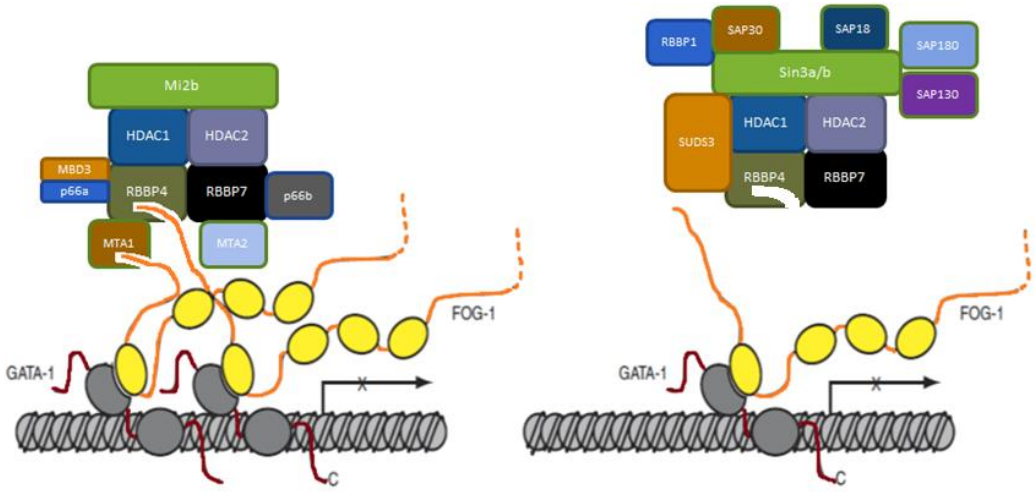

Figure 4. A schematic model of how HDAC complexes can be bound to the chromatin via different co-factors. Different HDAC complexes are composed of a variety of proteins and adapt different specificity. This figure is adapted from (Lejon et al., 2011).

One of the fundamental epigenetical controls involves chromatin remodelling brought about by chemical modifications to histones. In eukaryotes, histones are proteins found in eukaryotic cell nuclei that packs and orders the DNA into the basic subunit of the chromatin called nucleosome (Luger et al., 1997). Nucleosomes are packaged into successively higher order structures, allowing for the compaction and storage of genetic material. Without histones, the unwound DNA in chromosomes would be very long (about 1.8 meters), in contrast to the 90 micrometres of chromatin found in cell nuclei (**Figure 5**).

The key component of the eukaryotic nucleosome is a histone octamer consisting of a tetramer derived from two sets of H3/H4 heterodimer bound to two H2A/H2B dimers (Luger et al., 1997) around which 146 base pairs of DNA are wrapped, forming the nucleosome . Long N-terminal tails projects out of the hydrophobic histone core would be subjected to extensive covalent modifications at multiple sites and although the histone core could also be modified, the modifications are not as extensive (Strahl and Allis, 2000).

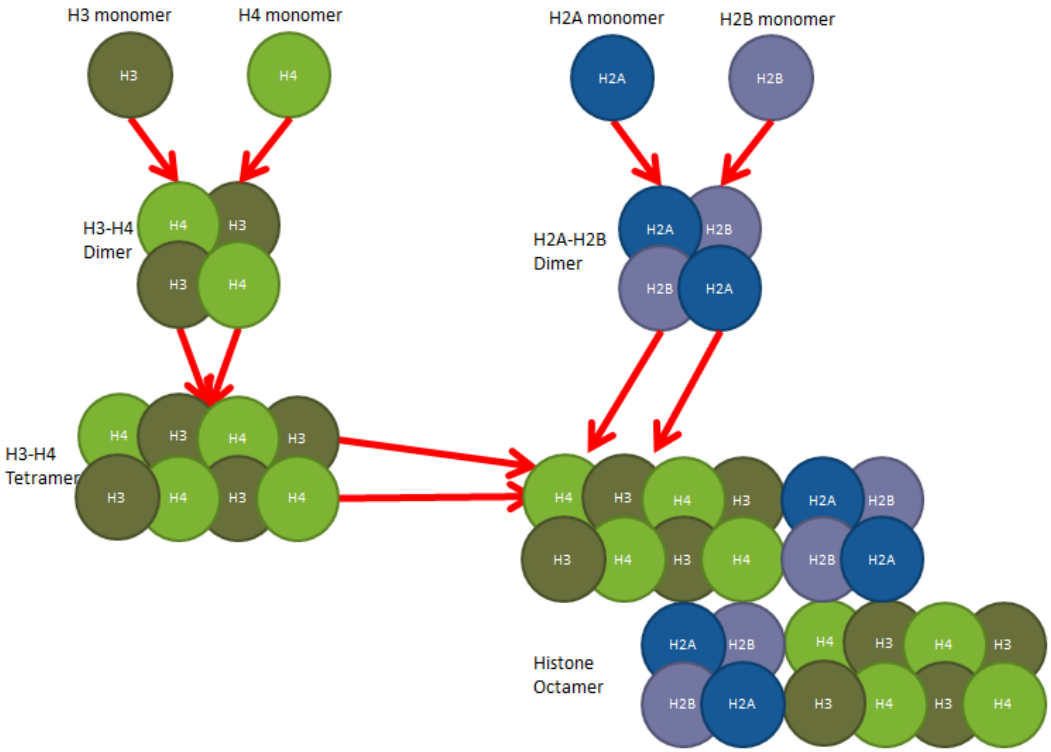

Figure 5. Organisation of histones subunits. Histones organise themselves from individual monomers into a functional octamer of 8 histone protein cores and packs the DNA to form nucleosomes.

Common post-translational modifications on the histone tail include methylation, acetylation, phosphorylation, ubiquitination and SUMOylation which would alter the interaction of the histone with the DNA and other nuclear proteins. Such modifications alters chromatin structure and affects the access of transcription factors and other nuclear proteins to DNA and allowed processes such as gene regulation, DNA repair and chromosome condensation to take place. As summarised in **Table 1**, the general effects of mono-methylation and acetylation leads to gene activation while di- and tri-methylation leads to either gene activation or repression, depending on the specific histone as well as the specific residue on the histone tail where the modification took place (Barski et al., 2007; Benevolenskaya, 2007; Koch et al., 2007; Steger et al., 2008; He, 2009; Rosenfeld et al., 2009).

Table 1. Some common histone modifications and their effects on the gene loci. The list is non-exhaustive and describes the most expected outcome and is not definitive.

Type of modification	Histone						
	H3K4	H3K9	H3K14	H3K27	H3K79	H4K20	H2BK5
mono-methylation	activation	activation		activation	activation	activation	activation
di-methylation		repression		repression	activation		
tri-methylation	activation	repression		repression	activation repression		repression
acetylation		activation	activation				

Epigenetic regulation processes of at DNA and histone levels are closely related (Chen and Pikaard, 1997; Li et al., 2002). DNA methylation can directly or indirectly trigger repressive histone modification at the target loci via histone deacetylation of H3K9 and H3K14, H3K9 dimethylation (H3K9me2) and H3K27 monomethylation (H3K27me1) (Zilberman et al., 2003; Wierzbicki et al., 2009). Synergistically, the H3K9 methyltransferase that catalyses dimethylation of H3K9 is recognized and bound by Chromomethylase 3(CMT3) which helps to maintain CHG (where H = A, T or C) methylation (Ebbs and Bender, 2006; Law and Jacobsen, 2010).This coupled action of DNA methylation together with repressive histone modifications helps establish a silenced heterochromatin, a critical gene regulatory process in development.

1.6 Gene regulation by histone deacetylation

There are three distinct families of HDACs in mammals, comprising a group of about 20 proteins (Taunton et al., 1996). The *Arabidopsis* HDA6 belongs to the Class I HDACs which bears the greatest similarities in amino acid sequence to the yeast Reduced Potassium Dependency (Rpd3) protein. The Class II HDACs, despite sharing sequence similarities to the Class I proteins, are grouped together based on their homology to the yeast Histone deacetylase 1 (Hda1) protein. The Class III HDACs are distinctly different in sequence and function to the HDACs in classes I and II, as they require the cofactor nicotinamide-adenine dinucleotide (NAD) for enzymatic activity and are insensitive to the chemical inhibitors such as butyrate, trichostatin and trapoxin. HDAC proteins are not restricted to the nuclear compartment, and indeed several non-histone substrates have been

identified for both NAD-dependent and independent enzymes (North et al., 2003; Zhang et al., 2003).

HDA6 is the *Arabidopsis thaliana* homolog of the human *HDAC2* and plays an important role in the transcriptional repression of the target gene when tethered to the target promoter. On top of deacetylation of lysines of H3 and H4, it is also required for cytosine methylation in transgenes and silenced ribosomal RNA (rRNA) genes, since histone deacetylation in the region takes place prior to DNA methylation (Aufsatz et al., 2002; He et al., 2009; Earley et al., 2010).

Histone acetylation takes place when the epsilon amino group of lysine residues of the histone tail is modified by the addition of an acetyl group by histone acetyltransferases (HAT) and reciprocally, the removal of the acetyl group by HDACs (**Figure 6**). Acetylation and transcription activation are closely coupled since the identification of HAT enzymes that function in concert with protein complexes to activate transcription (Roth et al., 2001; Carrozza et al., 2003). Conversely, the repression of transcription involves the recruitment of HDAC enzymes to specific genes (Urnov et al., 2001). Acetylation of the histone loosens histone-DNA contacts enabling gene transcription factors and other DNA modifying enzymes access to the DNA strands for transcription and other enzymic activities (Lee et al., 1993). In addition, the acetylated histones helps to recruit transcriptional activators containing an acetyl-lysine binding module called the bromodomain (Jacobson et al., 2000) among other protein complexes involved in transcription regulation.

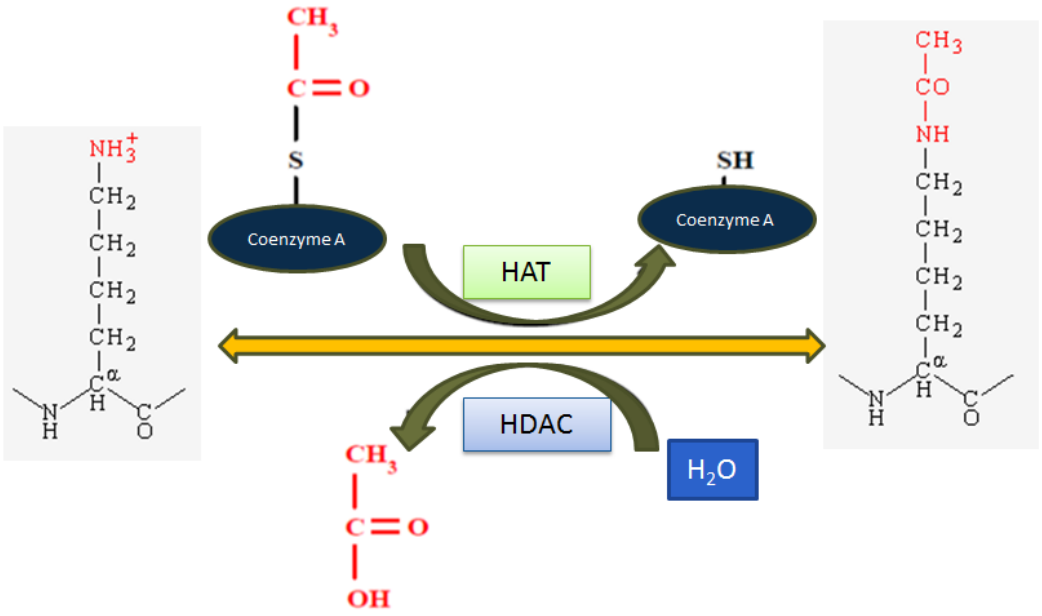

Figure 6. Schematic for the equilibrium of acetyl modification on histones. The opposing enzymic activity by of histone acetyltransferases (HAT) and deacetylases (HDAC) maintains equilibrium of the histone acetylation state of the acetyl groups on Lysine of histones.

Mutations of *HDA6* caused defects in transgene silencing as well as inability of the cell to maintain repression of silenced genes in *Arabidopsis* (Aufsatz et al., 2002; Probst et al., 2004). Furthermore, loss of function of *HDA6* leads to histone hyperacetylation in the nucleolus organizer regions containing the rDNA repeats and reduced CG and CHG methylation at target gene promoters (Wu et al., 2008; Earley et al., 2010). However, loss of *HDA6* functions under normal growth conditions does not greatly affect the development of the plant except for a slight delay in flowering, suggesting that *HDA6* regulation of the chromatin may be very specific (Probst et al., 2004).

Besides its role in histone deacetylation, HDA6 is also a key component of the RNA-directed DNA Methylation pathway (RdDM) in *Arabidopsis*. Briefly, RdDM utilises small double stranded RNA(dsRNA) generated from viral replication, products of RNA-dependent RNA polymerase (RdRP) protein such as RDR2 and transcribed inverted repeats which would be processed by Dicer-like 3(DCL-3) then incorporated into Agonaute4 (AGO4) to serve as sequence-specific guides during RdDM by pairing with complementary DNA targets or nascent scaffold RNAs from the DNA targets (Zilberman et al., 2003; Xie et al., 2004). This guiding complex would then recruit Domains Rearranged Methyltransferase (DRM2), the plant homolog of the mammalian DNA methyltransferase (DNMT3), and other methyltransferases to catalyse de novo DNA methylation (Aufsatz et al., 2002; Lippman and Martienssen, 2004; Kanellopoulou et al., 2005; Matzke and Birchler, 2005; Zhang and Zhu, 2011). This guidance by small interfering RNA (siRNA) in RdDM will result in a

reversible yet inheritable transcriptional repression of the target gene locus by ensuring preservation of specific CHH methylation patterns in daughter cells (Zhang and Zhu, 2011). *HDA6* is also essential for maintenance of silenced state as reactivation of RdDM-silenced promoters are observed in *hda6* mutants despite the continuous presence of the RNA-silencing signal (Aufsatz et al., 2007). It has also been suggested that HDA6 recognize the partial methylation at most cytosine residues resulting from the de novo step of RdDM where it then recruits DNA Methyltransferase 1(MET1) and CMT3 for the traditional CG and CNG (where N is any base) maintenance activities to reinforce preferentially C(N)G methylation (Bartee et al., 2001; Kishimoto et al., 2001; Lindroth et al., 2001; Papa et al., 2001).

HDA6 also plays a role in RNA-mediated heterochromatin formation, mainly at repeats near the centromeres, transposons and intergenic regions (Lippman and Martienssen, 2004). The *hda6* mutants have reduced H3K9me2 marks suggesting that *HDA6* is essential for siRNA-dependent heterochromatin formation as well as silencing of heterochromatic transposons. This deacetylation is also needed for subsequent methylation by HMTs (Lippman et al., 2003). Conversely, HDA6 catalysed histone deacetylation is also important for cytosine methylation in transgenes and endogenous rRNA genes (Aufsatz et al., 2002). For efficient gene silencing, DNA methylation alone would not suffice. As such, the recruitment of HDA6 in the form of a HDAC complex for histone deacetylation to bring about efficient gene repression via chromatin modification is critical (Aufsatz et al., 2007).

Heterochromatin constitutes the transcriptionally inactive state of the genome and in eukaryotes and is generally characterized by epigenetic marks such as DNA methylation and histone modifications like H3K9 methylation. DNA methylation and H3K9me2 in the chromatin are usually closely linked spatially and together, they create a mutually self-reinforcing and stable state of chromatin. However, H3K9me2 modifications can also be established without the presence of DNA methylation (Caro et al., 2012) . One such example is the Suppressor of variegation 3-9-related protein 5(SUVR5), which has three C2H2 zinc fingers that binds specifically to DNA sequences that map to gene promoters. Its Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain also confers SUVR5 methyltransferase activity needed for the dimethylation of histones. Indeed, the non-functional mutant of SUVR5 displays reduced dimethylation of H3K9 and H3K27 and hyperacetylation of histone H4 within the FLC locus and shows a moderate delayed flowering, of which these are known targets of the HDAC complex(Schmitges et al., 2011; Caro et al., 2012). The recruitment of HDAC complex to establish a gene-silencing state without DNA methylation can be achieved via SUVR5 and its interactions with co-factors within the HDAC complex.

FWA is a homeodomain-containing transcription factor that delays floral transition and in wild-type *Arabidopsis.* It is only expressed in the female gametes and extra-embryonic endosperm tissue from the maternal mRNA and is silenced in the sporophyte (Soppe et al., 2000; Kinoshita et al., 2004). The ectopic expression of *FWA* in the sporophyte would cause a late-flowering phenotype by

repressing *FT* activity. In wild-type *Arabidopsis, FWA* expression is repressed by repressive histone modifications and cytosine methylation in its 5' region (Soppe et al., 2000; Kinoshita et al., 2004; Lippman et al., 2004). As such, if there are defects or mutations in the repressive histone modifications and cytosine methylation pathways such as histone deacetylation, *FWA* silencing cannot take place and can be ectopically expressed. This would give rise to epi-alleles of *fwa* where *FWA* is expressed even though there are no mutations in the *FWA* gene itself.

The nature of *FWA* silencing and the late flowering phenotype observed during *FWA* upregulation made it a very useful tool for the investigation into abrogation of epigenetical gene silencing such as defects in de novo DNA methylation and maintenance of histone deacetylation.

1.7 Orthologs of Nurf55 and RBBP4 in *Arabidopsis* are MSI proteins

Nurf55 (p55) of *Drosophila* is homologous to RbAp48/46 in human and the MSI family of proteins in *Arabidopsis*. The *Drosophila* PRC2, which consist of four subunits consists of the four subunits E(z) , Su(z)12 (Suppressor of zeste 12), Esc (Extra sex combs), and Nurf55, forms the functional core of the Polycomb group repression machinery. Correspondingly, the *Arabidopsis* PRC2 complex consist of MEDEA (MEA), FIE, FERTILIZATION INDEPENDENT SEED (FIS) 2 and MSI1 (Kohler et al., 2003; Wang et al., 2006).

Nurf55 is key and non-catalytic subunit of PRC2 and other chromatin-modifying complexes such as HAT1 and CAF-1. Importantly, Nurf55 forms complexes with histones H3 and H4 to act as a recruiting factor of their complexes so that the complex is able to tether to their substrate nucleosomes. This affinity of Nurf55 to histones can also allow it to act as a histone chaperone that deposit H3/H4 heterodimers during replication-dependent and replication-independent nucleosome assembly (Tagami et al., 2004; Furuyama et al., 2006). Crucially, the binding sites for the histone H4 are overlapping with that of Su(z)12, demonstrating the versatility of Nurf55 in its multiple roles as components in a range of complexes as well as the variety of substrates it can bind with (Nowak et al., 2011). Su(z)12 is a key component of the PRC2 complex and is required during the first 6 hours of embryogenesis to establish the repressed state of homeotic genes throughout development. While the specific function of Su(z)12 is not well-documented yet, it is strongly conserved across many species and its zinc-finger domain at the C-terminal may be involved with chromatin silencing in conjunction with noncoding RNAs, using its zinc-finger domain to bind the RNA molecule (Rinn et al., 2007).

Both RBBP4 (RbAp48) and RBBP7 (RbAp46) are key components NURF, HDAC co-repressor complexes and are found in two distinct EZH2/EED complexes. RBBP4 and RBBP7 also promote transcriptional repression as an integral subunit of PRC2-like complexes that catalyse H3K27me3 through the histone methyltransferase activity is associated with the SET domain of E(z).Although the human RBBP4 and RBBP7 are closely related and can be

found in the same complex, RBBP4 and RBBP7 are found exclusively in CAF-1 and HAT1 respectively. The MSI family of proteins are homologous to RBBP; with MSI1sharing more sequence similarity to RBBP4 and that FVE is more similar to RBBP7.

MSI1 is a component of the CAF-1 complex (with FAS1 and FAS2) and the FIS complex (with MEA, FIE, MIS1 and possibly FIS2). MSI1 plays a role in the recruitment of chromosomal DNA into heterochromatic chromocenters and binds directly to helix 1 of histone H4 and also interacts with EMF1 and VRN2.

FVE has been known to interact with the plant Rb homolog and FVE mutations lead to reduced PRC2 catalysed H3K27me3 repression on *FLC* locus. FVE forms a complex with HDA6 to form a HDAC complex that mediated histone deacetylation and cytosine methylation.

1.8 MSI family of genes are important histone chaperones

Critical to chromatin modifying protein complexes are histone chaperones that aid the assembly of nucleosomes during replication and the recruitment of histone-modifying enzymes. These histone chaperones are also critical components of many chromatin-associated complexes, where they often act as key co-factors by repositioning nucleosome cores, effectively remodelling the chromatin structure.

The histone-binding proteins Retinoblastoma Associated proteins found in humans are highly homologous WD40-repeat proteins and were first identified in mammalian cells as the tumour-suppressor Rb-binding proteins (Loyola and

Almouzni, 2004).The human RETINOBLASTOMA ASSOCIATED PROTEIN 48 (RbAp48), also known as RBBP4 is closely related to RbAp46 (RBBP7) with up to 90% residue identity and these two proteins are key components of many protein complexes and could sometimes be found together in the same complex). However RBBP4 is found exclusively in the evolutionarily conserved CAF-1 complex while RBBP7 is found exclusively in HAT complex with Histone acetyltransferase type B catalytic subunit (HAT1) (Parthun et al., 1996; Verreault et al., 1998). Since the human RbAp46/48 is an integral subunit of Class I HDAC co-repressor complexes (Zhang et al., 1999; Fleischer et al., 2003; Murzina et al., 2008), it is likely that *Arabidopsis* HDA6 also forms a HDAC complex with other WD40-repeat proteins such as FVE (MSI4) and MSI5 similar to that of human RbAp46/48 in a Class I HDAC co-repressor complex (Gu et al., 2011).

The WD40 domain can be found in a number of eukaryotic proteins as adaptor/regulatory modules covering a wide variety of functions including signal transduction, transcription regulation, cell cycle control, autophagy, apoptosis, pre-mRNA processing and cytoskeleton assembly (Smith et al., 1999; Li and Roberts, 2001). These repeats are 40 residues long and usually contain the G-H dipeptide 11-24 residues from its N-terminus and end the W-D dipeptide at its C-terminus. Between GH and WD lies a conserved core serving as a stable propeller-like platform stably or reversibly protein interaction.

As visualized in **Figure 7**, several of these domains would combine to form a circularised beta-propeller structure with usually 6 to 8 blades where each blade is composed of a four-stranded anti-parallel b-sheet (Murzin, 1992). Each

WD40 sequence repeat forms the first three strands of one blade and the last strand in the next blade where the last C-terminal WD40 repeat completes the blade structure of the first WD40 repeat to create the closed ring propeller-structure. Residues on the top and bottom surface of the propeller are proposed to act as a site for protein-protein interaction, and proteins containing WD40 repeats are known to serve as platforms for the assembly of protein complexes or mediators of transient interplay among other proteins (Li and Roberts, 2001). The specificity of the interactions is determined by the sequences outside the repeats themselves.

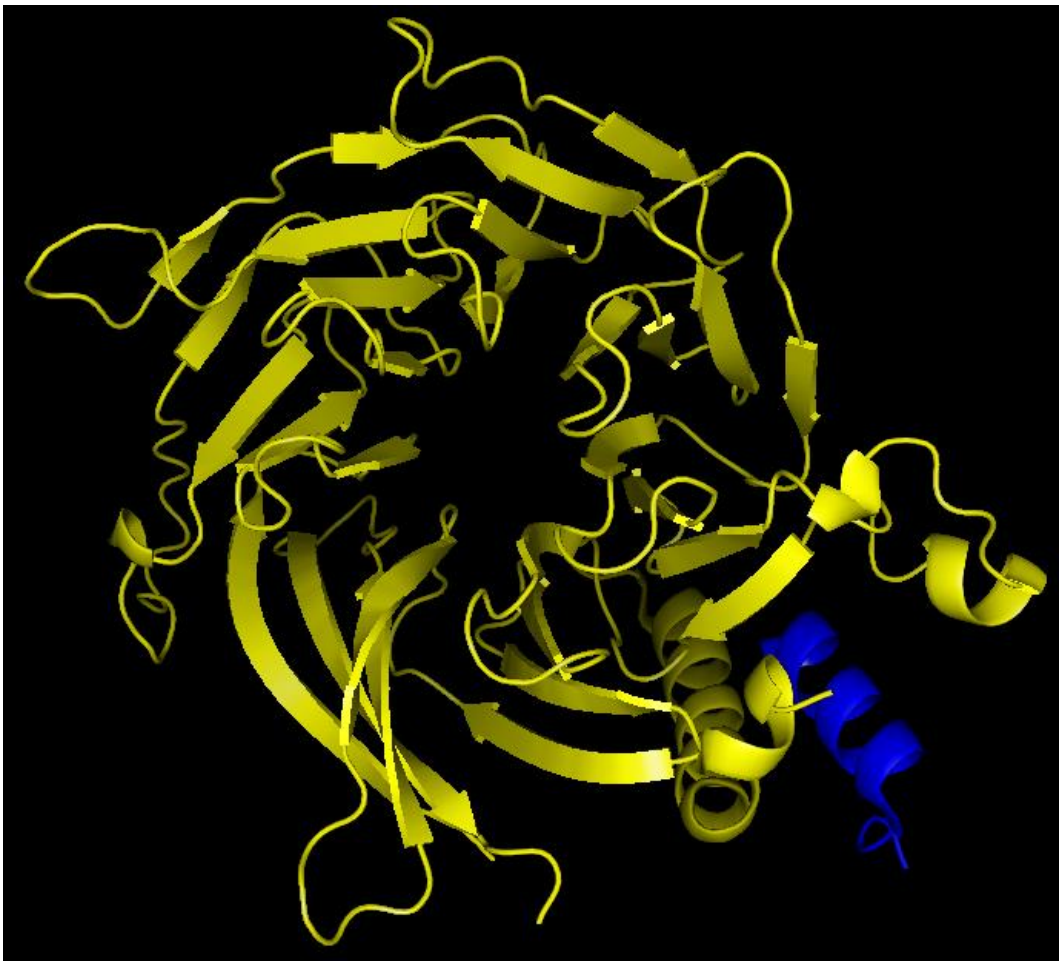

Figure 7. A typical circularised beta-propeller structure of a WD40 motif. The RbAp46 protein is in yellow while the histone H4 is in blue. Image is based on structure 3CFS (Murzina et al., 2008).

Importantly, both RBBP4 and RBBP7 are key components of HDAC co-repressor complexes which would typically cause transcriptional repression of target genes via deacetylation of core histones (Zhang et al., 1999; Nicolas et al., 2000; Korenjak et al., 2004). Both RBBP4 and RBBP7 are found in two distinct Enhancer of Zeste Homolog 2 and Embryonic Ectoderm Development (EZH2/EED) complexes responsible for the methylation of either histone H3K27 or H1K26 to which also represses the target regulatory genes epigenetically (Kuzmichev et al., 2004). Additionally, both RBBP4 and RBBP7 are components of the ATP-dependent Nucleosome Remodelling Factor (NURF), also involved in transcriptional regulation (Martinez-Balbas et al., 1998; Barak et al., 2003). RbAp46/48 also promotes transcriptional repression as an integral subunit of the evolutionarily conserved PRC2-like complexes that catalyse H3K27 trimethylation (He et al., 2003; Ausin et al., 2004; Doyle and Amasino, 2009). The solved crystal structure RbAp46 bound to histone H4 suggest that the RbAp46/48-containing complexes interact with histone substrates via either RbAp46 or RbAp48 interface (Murzina et al., 2008).

MSI1 is a core histone-binding subunit of several HDAC and CAF complexes and is involved in chromatin modifications for DNA replications and transcription repressions. It is found in complexes with histone deacetylases, histone acetyl transferases, the nucleosome remodelling factor NURF and HMTs (Hennig et al., 2003). MSI1 is not functionally redundant with the other MSI proteins and functions in multiple chromatin modifying complexes.

It is a component of the CAF-1 complex and the FIS complex , a polycomb group protein complex which is required to maintain the transcriptionally repressive state of homeotic genes that includes floral homeotic genes (Kaya et al., 2001; Hennig et al., 2003; Kohler et al., 2003; Guitton et al., 2004). MSI1 plays a role in the recruitment of chromosomal DNA into heterochromatic chromocenters and binds directly to helix 1 of histone H4, a region that is not accessible when H4 is in chromatin suggesting that conformational change must take place prior to binding.

The key player the flowering regulatory network is *FLC*, of which its importance results in it being a target and targeting other floral promoters or repressors. *FLC* expression suppresses the expression of *FT* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1).* The integration of signals from these 3 genes would eventual activate genes like *LEAFY* and *APETALA 1(AP1)* that promotes vegetative meristems into floral meristems (Kardailsky et al., 1999; Abe et al., 2005; Searle et al., 2006; Turck et al., 2008)

Involved in vernalization, MSI1 is part of the plant homeodomain / polycomb repressive complex 2 (PHD-PRC2) complexes with Polycomb group protein VERNALIZATION 2(VRN2), EZA1 and FIE which mediates H3K27 trimethylation (Kaya et al., 2001; Kohler et al., 2003). Playing an important role in meristem identity and maintenance through the regulation of gene expression of floral homeotic genes, MSI1 also interacts with CLF and EMF1, both of which are also subunits of the PRC2 complex (Hennig et al., 2003; Calonje et al., 2008). However, MSI1 does not repress the expression of FLC (Bouveret et al., 2006).

In contrast, FVE is directly involved in *FLC* repression to promote flowering (Ausin et al., 2004). FLC, a MADS-domain transcription factor and potent floral repressor, acts to repress the developmental transition to flowering in A*rabidopsis*. *FVE* and *MSI5* are two of the five *Arabidopsis thaliana MSI* genes, orthologs of the human *Retinoblastoma Binding Protein 4 (RBBP4)*. These are important histone chaperones with key roles in protein complexes responsible for chromatin structure modifications especially those that are involved in methylation, acetylation or deacetylation the H4 histone tail (Murzina et al., 2008).

FVE has been known to interact with the plant Rb homolog but the biological significance of this interaction remains largely unknown (Ausin et al., 2004). FVE is known to control flowering by repressing the expression of *FLC*, the central floral repressor for which *FVE* loss-of-function would result in a late flowering phenotype (Ausin et al., 2004). Increased histone acetylation at the *FLC* locus is also observed in *fve* mutants, indicating that *FVE* represses *FLC* expression via chromatin deacetylation (He et al., 2003; Ausin et al., 2004). Furthermore, *FVE* mutations lead to reduced PRC2 catalysed H3K27me3 repression on *FLC* locus (Doyle and Amasino, 2009).

In contrast, the biological functions and interactions of MSI5 are much more elusive until recently when it has been demonstrated that MSI5 acts in partial redundancy with FVE in *FLC* chromatin repression (Gu et al., 2011). Together with HDA6 deacetylase, FVE and MSI5 form a HDAC complex that cause deacetylation of the target histone and the subsequent transcriptional silencing at the target loci. As such, FVE and MSI5 are speculated to play

important roles in the RNA-directed chromatin silencing of various gene loci in plants.

Both FVE and MSI5 are important for CHH methylation and maintenance of symmetric cytosine methylation of endogenous RdDM target loci as well as the establishment of cytosine methylation of the previously unmethylated sequences. Observations in *fve* and *msi5* mutants include histone hyperacetylation, loss of cytosine methylation and transcriptional reactivation at the RdDM target. This suggests that MSI5/FVE-HDA6 HDAC complex-mediated histone deacetylation creates a repressive chromatin environment that promotes DRM2 catalysed cytosine methylation at previously unmethylated sequences (Gu et al., 2011).

MSI5 acts redundantly with FVE in RNA-directed gene silencing as key components of a HDAC complex comprising of various enzymes including HDA6 on targets such as development genes as well as transposable and repetitive elements (Gu et al., 2011).

1.9 Objectives of this study

The first objective of this study is to investigate the structural and functional features of HDA6 by comparing its model as predicted using Robetta with that of the human HDAC2, its homolog which structure has been solved with X-ray crystallography.

Secondly, we sought to identify and further characterize HDA6 partners to understand how HDA6 mediates gene silencing in *Arabidopsis*.

Chapter 2 Materials and Methods

2.1 Structural Modelling and PyMOL visualization

2.1.1 PyMOL

PyMOL (Schrodinger, 2010) is a powerful and comprehensive molecular visualization product for rendering and animating 3D molecular structures. It requires the input of atomic coordinates to a PDB file that can be downloaded from the PDB database or generated from various structure modeling softwares. Educational-Use-Only PyMOL is available for download at http://pymol.org/educational/

2.1.2 Robetta

The Robetta Full-chain Protein Structure Prediction Server is hosted by the Bakerlab of Molecular Engineering and Sciences at University of Washington. Robetta uses the Rosetta software package and is accessed at http://robetta.bakerlab.org/

2.1.3 Pro-sp3-TASSER

The Pro-sp3-TASSER Protein Structure Prediction tool is hosted by the Skolnick Research Group of Center for the Study of Systems Biology at Georgia Tech. It is freely available to academic users and not-for-profit institutions and can be found accessed at http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/index.html

2.1.4 Phyre2

Phyre2 (Protein Homology/analogY Recognition Engine V2.0) is hosted by the Structural Bioinformatics Group, Imperial College, London. It can be accessed at http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index

2.2 Sequence alignment with Clustal Omega

Clustal Omega is a multiple sequence alignment program for proteins that uses seeded guide trees and HMM profile-profile techniques to generate alignments (Goujon et al., 2010; Sievers et al., 2011). It is hosted by European Molecular Biology Laboratory- European Bioinformatics Institute (EMBL-EBI) and it is freely available for use at http://www.ebi.ac.uk/Tools/msa/clustalo/

2.3 ImageJ

The ImageJ program is a Java-based open source image enumeration software (Schneider et al., 2012) that can calculate area and pixel value statistics of user-defined selections and create density histograms and line profile plots. It can run as an online applet and can be freely downloaded at http://rsb.info.nih.gov/ij/download.html.The DNA gel photos are captured in grayscale for processing by ImageJ.

2.4 Plant Materials and Growth Conditions

Arabidopsis thaliana hda6/axe1-5 (Wu et al., 2008), pTEK::TEK-YFP(Xu et al., 2013b) were described previously. SUPPRESSOR OF VARIEGATION 3-9-RELATED 5 (*suvr5)* is a T-DNA insertion line obtained from the SALK Institute Genomic Analysis Laboratory (SALK_026224).

Columbia wild-type and transgenic *Arabidopsis* were grown on soil at 22 ℃ under long days (16 h light/8 h dark) or short days (16 h dark/8 h light). Seeds were sterilized by washing with 70% ethanol then rinsed with 90% ethanol and left to dry on filter papers before they were sequentially sowed in Petri dishes containing autoclaved Murashige and Skoog (MS) agar medium, which was adjusted to pH 5.8. The plates were maintained in a tissue culture room under LDs (16 h light/ 8 h dark). Basta selection was conducted twice within 10-20 days after seed germination to screen transgenic plants. Seeds were stratified on soil at 4 ℃ for 4-5 days before being transferred to a growth room in order to ensure synchronized germination.

To measure flowering time, seeds were kept for 2 to 3 days at 4 ℃ and then moved to growth chambers under LD (16 h light/8 h dark) at 22 ℃. This time was defined as day 0. Plants were grown on plates for 1 week in LD and then transferred to soil. To determine flowering time, the number of rosette and cauline leaves was counted at bolting.

2.5 Plant transformation

Transgenic plants were generated by the floral dip method for Agrobacterium-mediated infiltration.(Clough and Bent, 1998). Agrobacteria was cultured in LB media at 28 ℃ overnight and then spun down at 3500rpm for 15min to collect the cell pellets. The collected cells were then resuspended in 5% sucrose with 0.015% Silwet L-77 and 0.05% MES at pH5.7. *Arabidopsis* flower buds were dipped in this Agrobacteria preparation for around 1 minute then

placed in the dark for 24hours. Following this, the plants are grown under normal growth conditions until the seeds are harvested.

2.6 Plasmid construction

For pSUVR5::SUVR5-FLAG construction, the full length *SUVR5* genomic fragment which sequence at 5' included its promoter then downstream to its genomic coding region at 3' excluding its STOP codon. The 3' end was fused in frame with three copies of the FLAG tag for gateway cloning.

2.7 RNA isolation

Total RNAs were extracted from aerial parts of 9 to 10-day-old seedlings grown in soils or half-MS media under LDs using Qiagen RNAeasy plus mini kit according to the manufacturer's instructions.

2.8 Analysis of *FWA* Transcripts by RT-PCR

Total RNAs were used as templates to synthesize cDNAs by the M-MLV reverse transcriptase (Promega) *Tubulin 2 (TUB2)* was used as a loading control. cDNAs of *FWA* were amplified in a 20-μL volume with 33 and 37 cycles of 94 ℃ (30 s), 60 ℃ (30 s), and 72 ℃ (30 s). cDNAs of *Tubulin2* were amplified in a 20-μL volume with 24 cycles of 94°C (30 s), 60°C (30 s), and 72°C (30 s).

Primers for *FWA* and *TUB2* are as below:

FWA Forward: AGGCAATACTGGTGGAGGATGTCTACTG
FWA Reverse: GACAATAGTATGAGCCATGAGTGTCTCGAC
TUB2 Forward: ATCCGTGAAGAGTACCCAGAT
TUB2 Reverse: AAGAACCATGCACTCATCAGC

2.9 Yeast two-hybrid screening

HDA6 was used to screen the *Arabidopsis* cDNA library via the yeast two-hybrid system. Yeast transformation and library screening were performed following the procedures described in the user manuals (Yeastmaker™ Yeast Transformation System 2 and Matchmaker™ Gold Yeast Two-hybrid System, Clontech).

The full-length coding sequence of *HDA6* was cloned into the pGBKT7 vector (Clontech) while full length *MYB5*, *NYE1* and *RAD51C1* were cloned into the pGADT7 vector (Clonetech). Plasmids were introduced into the yeast strain AH109 and yeast cells were spotted on the SD media without leucine, tryptophan, histidine and adenine according to the manufacturer's instructions (Clontech).

Chapter 3 Results

3.1 Structural Modelling of HDA6 using Robetta

The Ginzu domain prediction used by Robetta for HDA6 modelling uses the structure '3MAX' as the reference parent. This file deposited in the PDB is the result of the solved crystal co-complex of HDAC2 and N-(4-aminobiphenyl-3-yl) benzamide (Bressi et al., 2010) expressed using a Baculovirus system. The primary sequence of human HDAC2 is the closest match to that of HDA6 among solved structures in the PDB database.

Human HDAC2 is a homolog of *Arabidopsis* HDA6 and its active site consists of, a lipophilic 'tube' which leads from the surface to the catalytic site, and a 'foot pocket' immediately adjacent to this catalytic machinery (**Figure 10** and **Figure 11**). Alignment of the protein sequence using Clustal Omega for HDAC2 and HDA6 revealed that the residues providing the catalytic surfaces are invariantly conserved (**Figure 9**).

A total of five models were predicted and model 1 was chosen since it has the lowest energy. The model with the lowest energy is the modelled structure with the most stable configuration within the calculated force field.

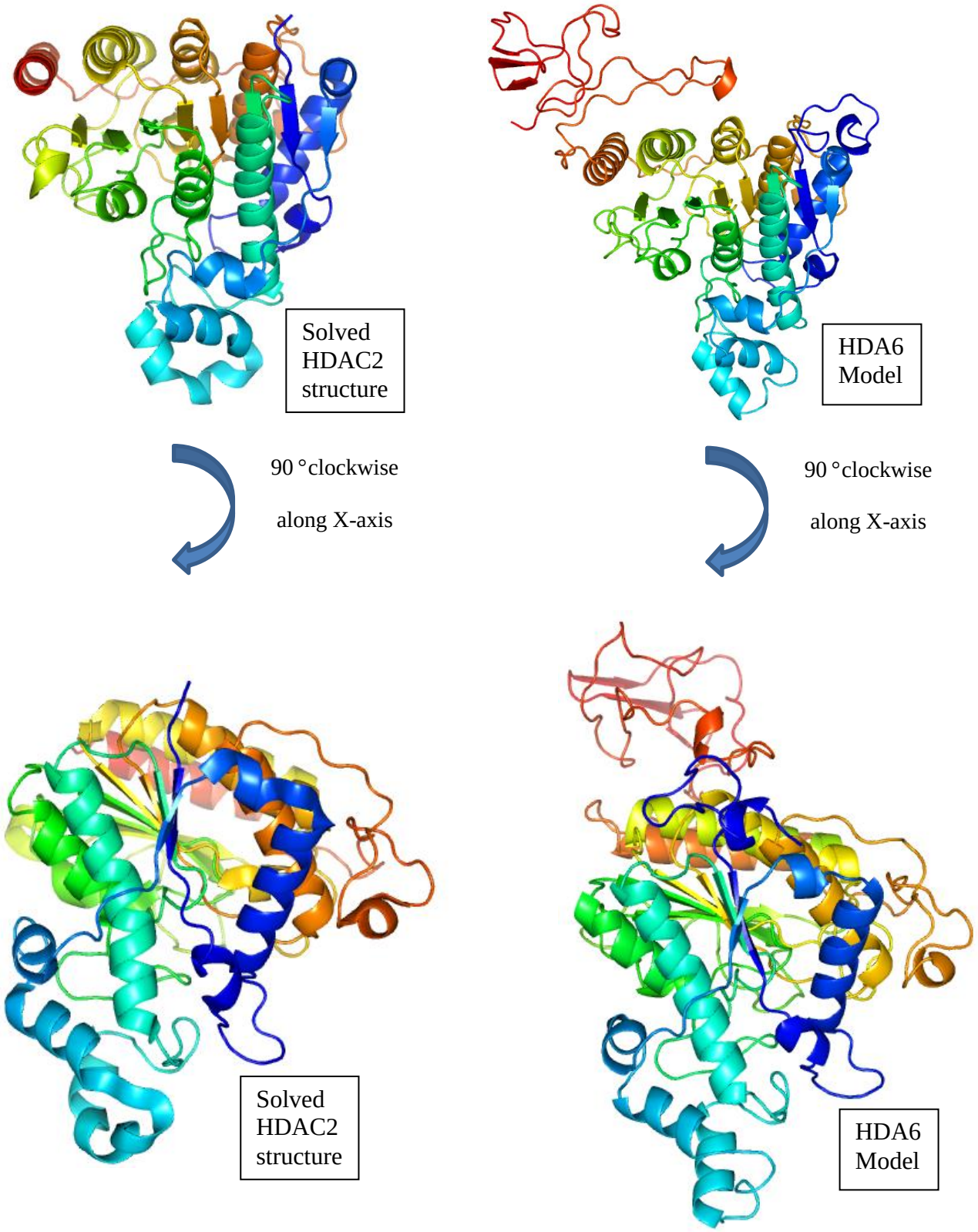

Figure 8. Crystal structure of HDAC2 on the left and the structure of HDA6 predicted by Robetta on the right. View from a 90 °clockwise rotation along X-axis on the bottom. While there are differences between the two structures, the key helices, sheets and folds of the core are conserved.

Figure 9. Alignment of protein sequences for **HDA6** (gi|15242626|ref|NP_201116.1) against **HDAC2** (gi|293336691|ref|NP_001518.3|). The specific residues involved in the formation of the active site critical to the catalytic activities of HDAC2 are highlighted in yellow and are all invariantly conserved in HDA6.

Table 2. Residues contributing to surface of active sites identified in the crystal structure of HDAC2 are invariantly conserved in HDA6.

	HDAC2	HDA6
lipophilic 'tube'	Gly154, Phe155, His183, Phe210, Leu276	Gly161, Phe162, His190, Phe216, Leu283
'foot pocket'	Tyr29, Met35, Phe114, Leu144	Tyr34, Met40, Phe121, Leu151

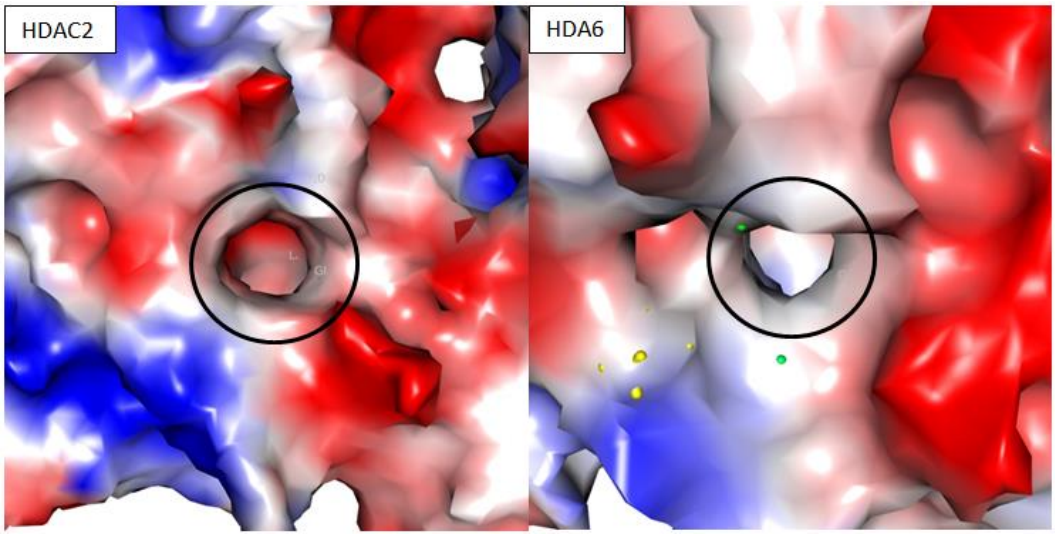

Figure 10. Electrostatic map on the lipophilic 'tube' of HDAC2 and HDA6. The lipophilic 'tube' entrance into the active site of HDAC2 (left) and HDA6 (right) are circled in black. The electrostatics colour code corresponds to positive charges (blue), negative charges (red) and hydrophobic residues (white). In both HDAC2 and HDA6, the interaction surface of the lipophilic 'tube' is largely hydrophobic and negatively charged and would bind strongly to the positively charged N-terminal tail of the Histone H3.

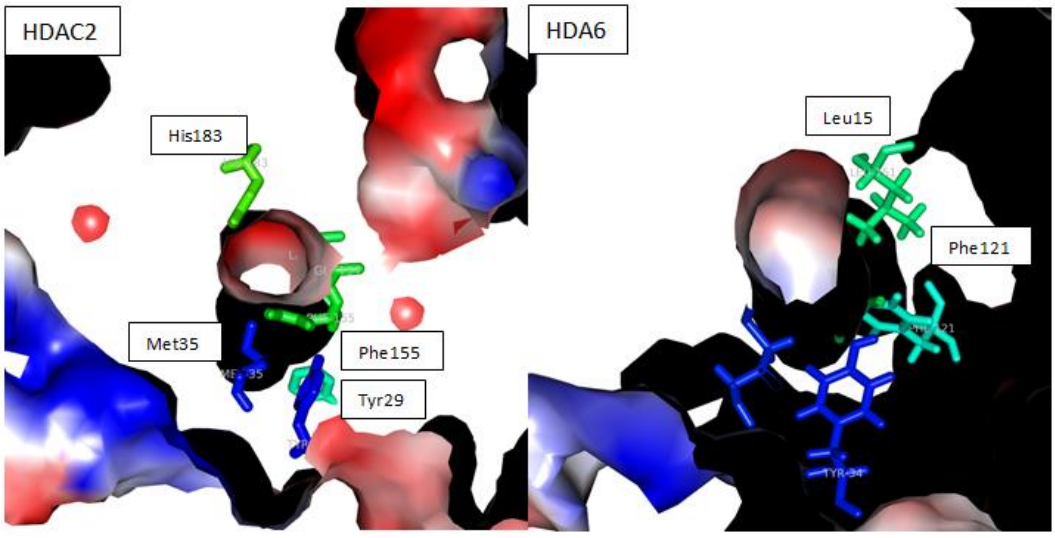

Figure 11. View of the lipophilic 'tube' of HDAC2 (left) and HDA6 (right) with the electrostatic charge of the protein surface removed. The residues involved in the formation of the surface of the lipophilic 'tube' are labelled and represented as sticks.

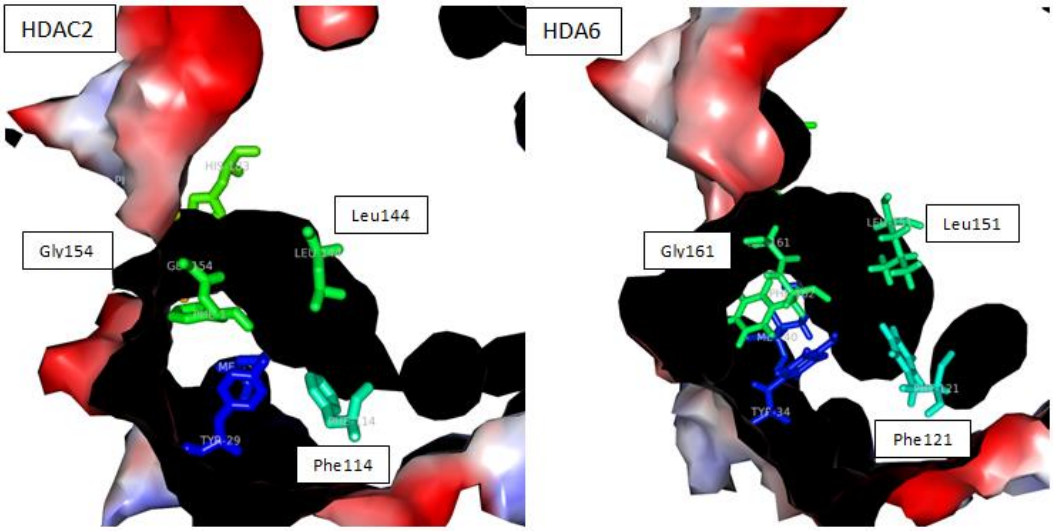

Figure 12. View of the lipophilic 'tube' rotated 90 °clockwise along the X-axis from **Figure 11** for HDAC2 (left) and HDA6 (right). The electrostatic charge surface is removed to allow unobstructed view of the lipophilic 'tube'.

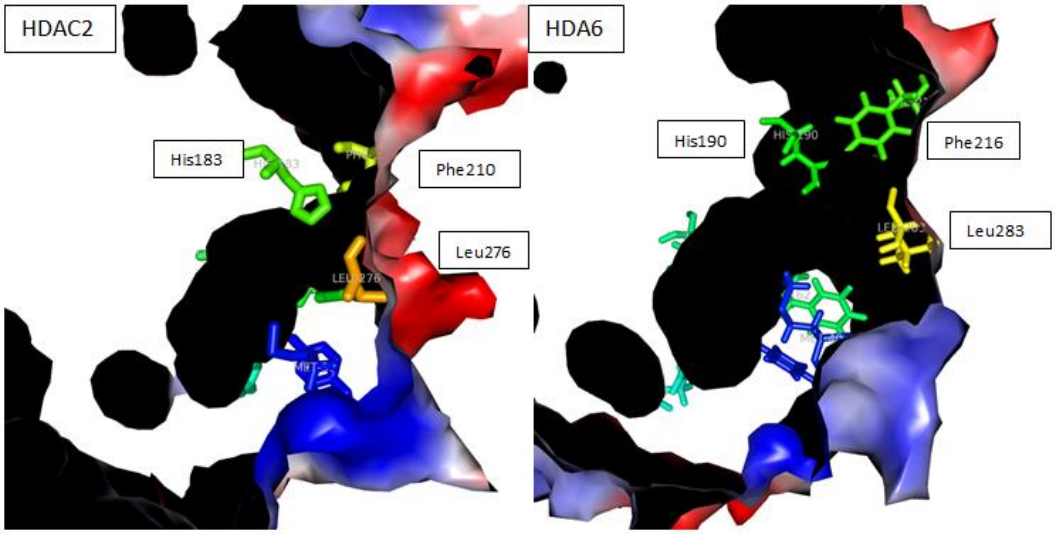

Figure 13. View of the lipophilic 'tube' rotated 90 °anti-clockwise along the X-axis from **Figure 11** for HDAC2 (left) and HDA6 (right). The electrostatic charge surface is removed to allow unobstructed view of the lipophilic 'tube'.

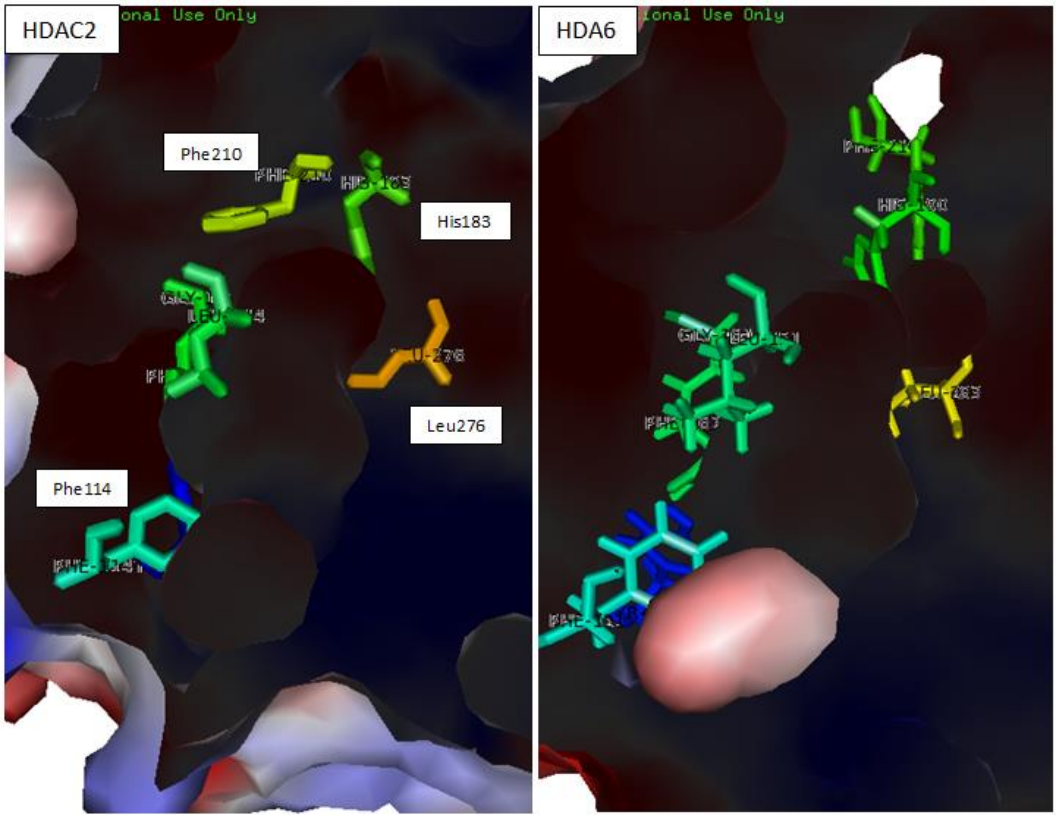

Figure 14. View of the lipophilic ‘tube’ rotated 180 °along the X-axis from **Figure 11** for HDAC2 (left) and HDA6 (right). The electrostatic charge surface is removed to allow unobstructed view of the lipophilic ‘tube’.

The catalytic site for the deacetylase activity is a lipophilic ‘tube’ which leads from the surface into a ‘foot pocket’ nestled deep within the protein. Both active sites are very similar electrostatically, where they are negatively charged at the ‘entrance’ of the entrance, hydrophobic in the middle and positively charged at the ‘end’ of the pocket (**Figure 10**). This negatively charged lipophilic ‘tube’ binds strongly to the positively charged N-terminal tail of the Histone H3 in place to allow the deacetylation of the acetylated lysine residues.

Corresponding residues mapped onto HDA6 that provided the surface for the lipophilic ‘tube’ and ‘foot pocket’ for HDAC2 occupied very similar positions spatially in the protein (**Figure 11** to **Figure 14**). The shape and size of the lipophilic ‘tube’ and ‘foot pocket’ created by key catalytic residues for both HDAC2 and HDA6 are also very similar which is critical to allow the binding of the N-terminal tail of the Histone H3.

3.2 Sequence alignment of Nurf55, RBBP4 and MSI proteins

Using Clustal Omega, we aligned the primary sequence of *Drosophila melanogaster* Nurf55 (a subunit of CAF-1) (sp|Q24572|CAF1_DROME) and *Homo sapiens* RBBP4 (sp|Q09028|RBBP4_HUMAN) against the *Arabidopsis thaliana* MSI1 (gi|15237140|ref|NP_200631.1|), MSI2 (gi|15227294|ref|NP_179269.1|), MSI3 (gi|15236251|ref|NP_195231.1|), MSI4 (gi|30680701|ref|NP_565456.2|) and MSI5 (gi|79490015|ref|NP_194702.2|) (**Figure 15**).

Drawing information from six publications, we identified specific residues in Nurf55 and RbAp46/48 that directly interacted with their ligands and they were highlighted in different colours and numbered accordingly, as represented in **Table 3.** Based on the alignment data, these highlighted residues are juxtaposed against the Arabidopsis *MSI* family of proteins.

The alignments revealed that there is close homology between Nurf55 and RBBP4 with MSI1/2/3 in contrast to FVE and MSI5. This difference can be accentuated by highlighting the residues conserved in MSI1/2/3 but not in FVE/MSI5 in grey and highlighting residues conserved in all five MSI proteins in brown.

Residues highlighted under the numbers 1 to 4 are involved in a key interaction surface of Nurf55 and RbAp46/48 that interacts with both histone H4 and Su(z)12. In contrast, residues highlighted under the numbers 5 and 6 are involved in another interaction surface of Nurf55 and RbAp46/48 that interacts with both Histone H3 and FOG-1.

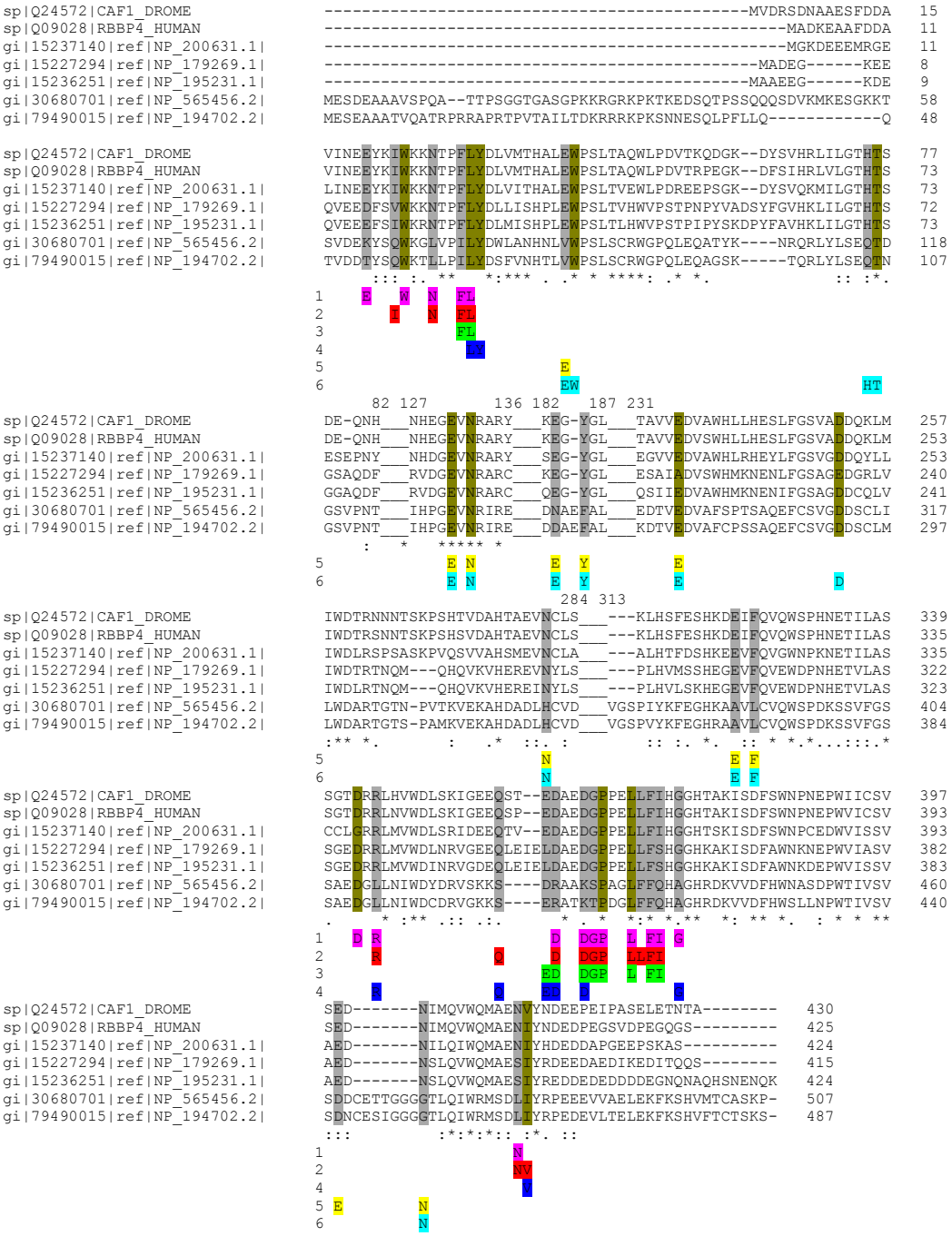

Figure 15. Sequence alignment of *Drosophila melanogaster* Nurf55 and *Homo sapiens* RBBP4 against *Arabidopsis thaliana* MSI family of proteins. The highlighted residues correspond to the specific residues of Nurf55 and RbAp46/48 that interacted directly with its ligands. Colour and number coding representing the different publications are found in **Table 3**. *Drosophila melanogaster* **Nurf55** (a subunit of CAF-1) (sp|Q24572|CAF1_DROME) and *Homo sapiens* **RBBP4** (sp|Q09028|RBBP4_HUMAN) against the *Arabidopsis thaliana* **MSI1** (gi|15237140|ref|NP_200631.1|), **MSI2** (gi|15227294|ref|NP_179269.1|), **MSI3** (gi|15236251|ref|NP_195231.1|), **MSI4** (gi|30680701|ref|NP_565456.2|) and **MSI5** (gi|79490015|ref|NP_194702.2|).

Table 3. References from where amino acid residues were identified from and highlighted in **Figure 15**.

Number	Colour	PDB id	PDB Title	Ligand	Reference
1	Pink	2YB8	CRYSTAL STRUCTURE OF NURF55 IN COMPLEX WITH SU(Z)12	Su(z)12	(Schmitges et al., 2011)
2	Red	2XYI	CRYSTAL STRUCTURE OF NURF55 IN COMPLEX WITH A H4 PEPTIDE	Histone H4	(Nowak et al., 2011)
3	Green	3CFS	Structural basis of the interaction of RbAp46/RbAp48 with histone H4	Histone H4	(Murzina et al., 2008)
4	Blue	3C9C	Structural Basis of Histone H4 Recognition by p55	Histone H4	(Song et al., 2008)
5	Yellow	2XU7	STRUCTURAL BASIS FOR RBAP48 BINDING TO FOG-1	FOG-1	(Lejon et al., 2011)
6	Teal	2YB8	CRYSTAL STRUCTURE OF NURF55 IN COMPLEX WITH SU(Z)12	Histone H3	(Schmitges et al., 2011)

3.3 Structural Modelling of MSI1 and FVE using Phyre2 and Pro-sp3-TASSER

The solved crystal structure of NURF55 complexed with Histone H4 (PDB id: 2XYI) (Nowak et al., 2011) was analysed using PyMOL (Schrodinger, 2010) to visualize the histone H4 binding groove of Nurf55. These residues that formed the interaction surfaces in this histone H4 groove is highlighted in red and denoted "2" in **Figure 15** and **Table 3**

To map the residues of Nurf55 directly interacting with the H4 ligand onto MSI1 and FVE, we first aligned the corresponding sequences as detailed in **Figure 15**. Next, structural modelling for MSI1 and FVE was performed in order to visualize the two proteins in 3D. To do so, we used Phyre2 and Pro-sp3-TASSER, two web-based services for protein structure prediction that are free for non-commercial use.

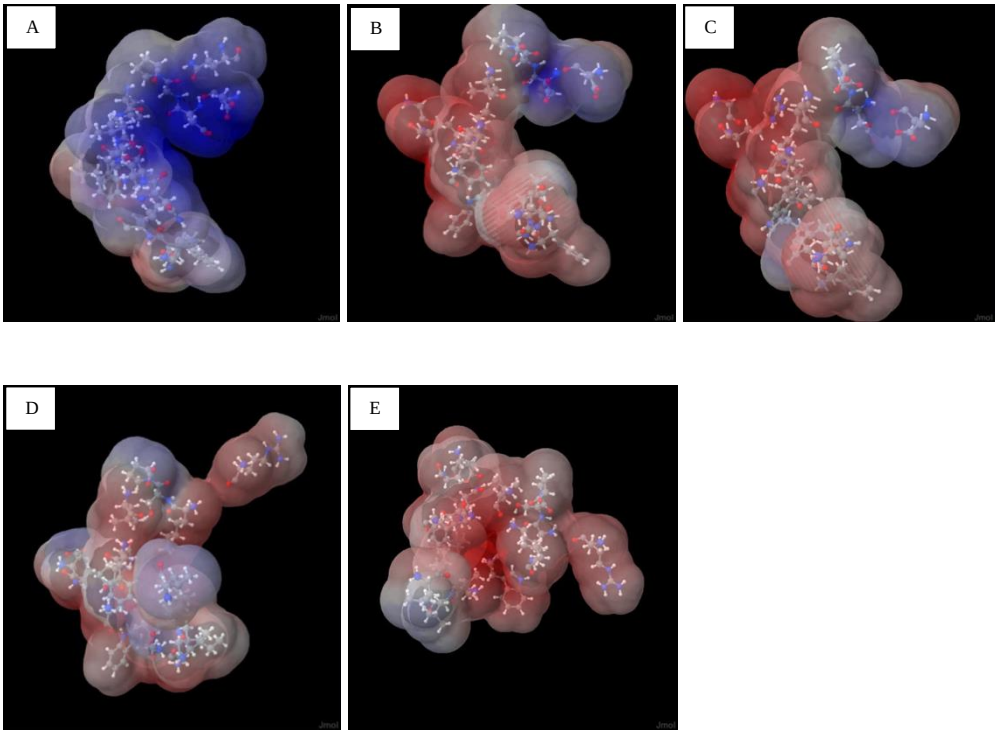

Figure 16. Comparison of electrostatic surface of the histone H4 groove of Nurf55 with models of MSI1 and FVE. Electrostatic surface is based on the residues of Nurf55 directly interacting with the ligand histone H4. (A) The solved crystal structure of Nurf55 complexed with Histone H4 (PDB id: 2XYI). (B) and (C) are predicted models of MSI1 by Phyre2 and Pro-sp3-TASSER respectively while (D) and (E) are predicted models of FVE by Phyre2 and Pro-sp3-TASSER respectively.

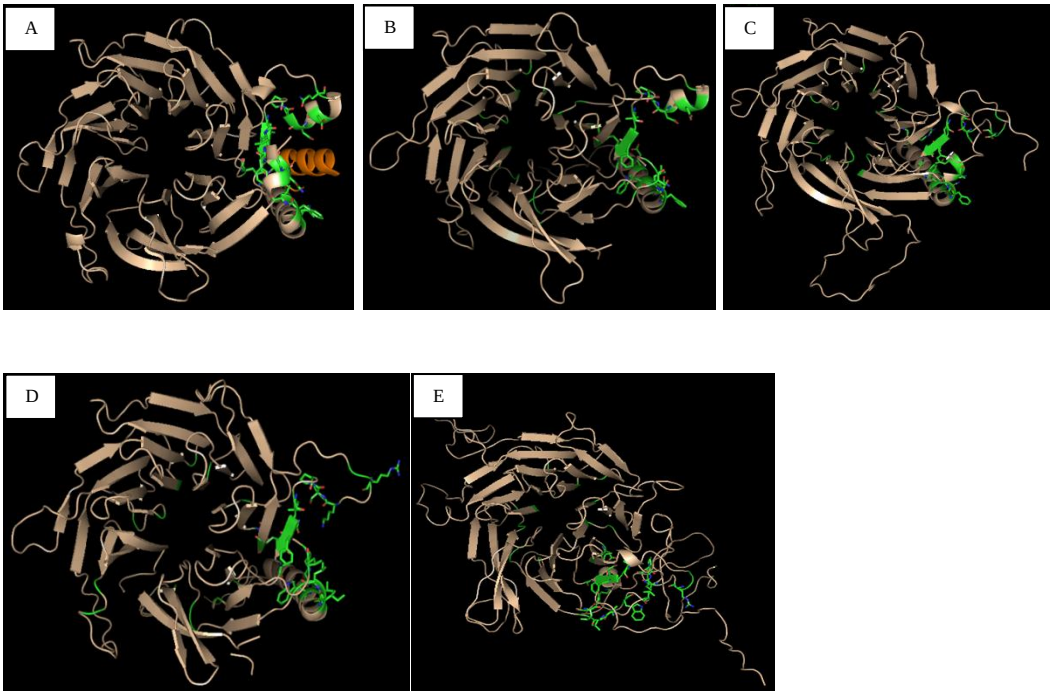

Figure 17 Comparison of cartoon representation of Nurf55 with models of MSI1 and FVE. Represented in green are the residues of Nurf55 that interact directly with the helix 1 of H4 ligand (orange helix). (A) The solved crystal structure of Nurf55 complexed with Histone H4 (PDB id: 2XYI) (B) and (C) are predicted models of MSI1 by Phyre2 and Pro-sp3-TASSER respectively while (D) and (E) are predicted models of FVE by Phyre2 and Pro-sp3-TASSER respectively. Residues highlighted in green for MSI1 and FVE models are based on alignment of residues of Nurf55 that interact directly with the helix 1 of H4 ligand.

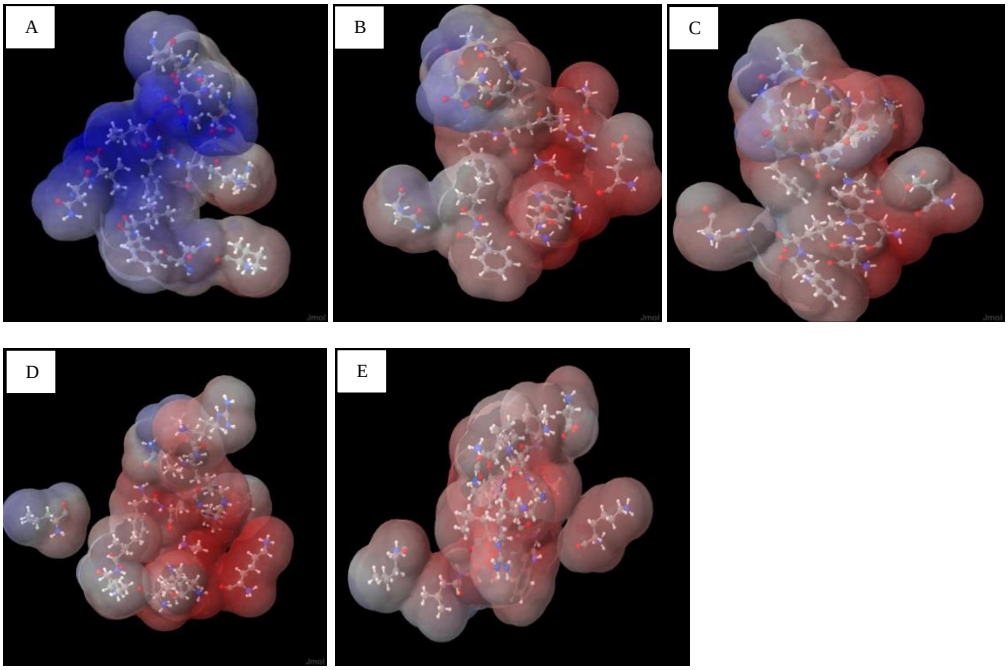

Figure 18. View of the electrostatic surface of the histone H4 groove of Nurf55 with models of MSI1 and FVE as shown in **Figure 16**, rotated 90 °clockwise about the X-axis. Molecular and electrostatic surface representation follows that of **Figure 16.** (A) The solved crystal structure of Nurf55 complexed with Histone H4 (PDB id: 2XYI). (B) and (C) are predicted models of MSI1 by Phyre2 and Pro-sp3-TASSER respectively while (D) and (E) are predicted models of FVE by Phyre2 and Pro-sp3-TASSER respectively.

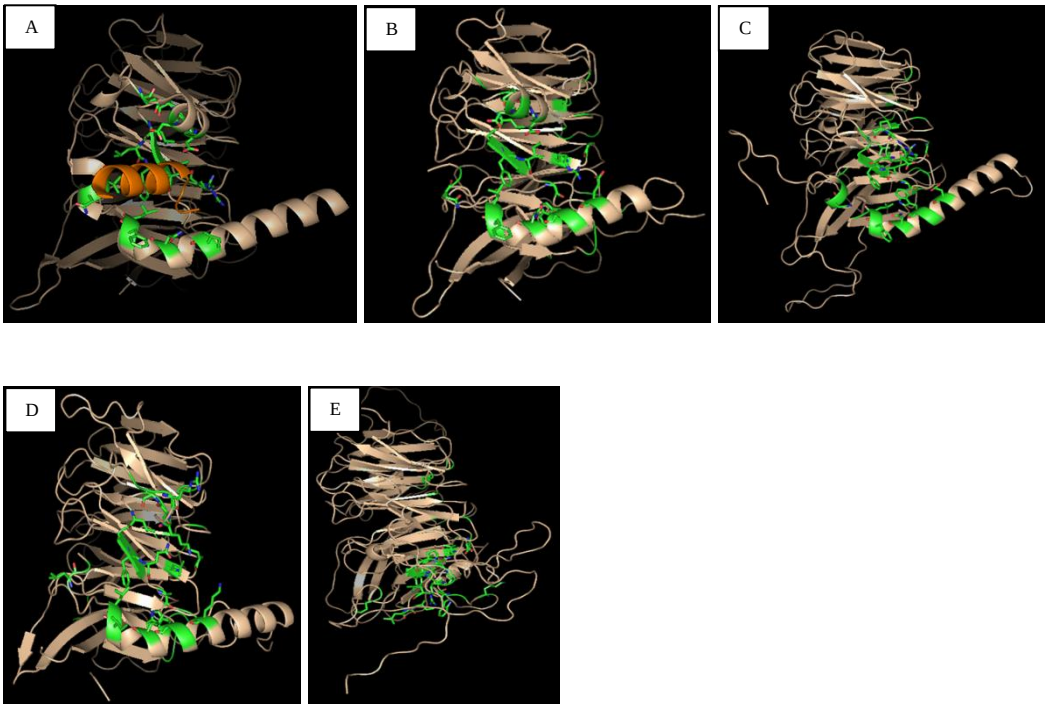

Figure 19 View of the cartoon representation of the histone H4 groove of Nurf55 with models of MSI1 and FVE as shown in **Figure 17**, rotated 90 °clockwise about the X-axis. Represented in green are the residues of Nurf55 that interact directly with the helix 1 of H4 ligand (orange helix). (A) The solved crystal structure of Nurf55 complexed with Histone H4 (PDB id: 2XYI) (B) and (C) are predicted models of MSI1 by Phyre2 and Pro-sp3-TASSER respectively while (D) and (E) are predicted models of FVE by Phyre2 and Pro-sp3-TASSER respectively. Residues highlighted in green for MSI1 and FVE models are based on alignment of residues of Nurf55 that interact directly with the helix 1 of H4 ligand.

The top model from Phyre2 modelling for both MSI1 and FVE used the histone-binding protein rbbp7 (3CFV) as its highest scoring template since it gave the highest sequence coverage and are modelled with 100% confidence.

Pro-sp3-TASSER modelling for MSI1 and FVE returned with five models each and the first model labelled number "1" (lowest energy) was chosen. The model with the lowest energy is the model with the most stable spatial arrangement of the protein within the designated parameters.

These predicted 3D models of MSI1 and FVE were visualized alongside the crystal structure of NURF55 complexed with Histone H4 (PDB id: 2XYI) (Nowak et al., 2011) for analysis of the 3D structure of the binding grooves and the spatial arrangement of the residues that interact with helix 1 of histone H4.

3.4 Screening the interacting protein of HDA6 from the *Arabidopsis* cDNA library

To find potential binding partners of HDA6, we conducted Y2H screens of HDA6 against a cDNA library of *Arabidopsis*. Full length *HDA6* was used to construct the bait plasmid pGBKT7-HDA6 (BD plasmid) and transformed into AH109 yeast cells. We tested for the toxicity and auto-activation of *HDA6* as bait in this Y2H system by comparing its growth against positive controls and negative controls (empty pGBKT7 vs. pGBKT7-HDA6), observing no significant difference colonies growth.

The pGBKT7-HDA6 AH109 cells were then mixed with a cDNA library of *Arabidopsis* in Y187 yeasts in a 2L flask and incubated at 30℃ for 20–24

hours, slowly shaking at 50 rpm to allow mating. After a visual check for zygotes under the microscope as an indication for successful mating, the cells are collected in a pellet and subsequently plated in double dropout (SD /−Leu/−Trp) plates for incubation at 30 ℃ for 5 days.

A total of 114 colonies were selected from the double dropout plates and they were inoculated overnight 2ml of double dropout media, each tube carefully numbered. 1 µl of the overnight culture is dripped on the triple dropout (SD/−His/−Leu/−Trp) plate in their correspondingly numbered grid. Similarly, another 1 µl from the same overnight culture is then dripped onto the exact corresponding numbered grid on the quadruple dropout (SD/−Ade/−His/−Leu/−Trp) plate (**Figure 20**).

The 15 colonies that grew in the both the triple and quadruple dropout plates as identified with their numbered grid were selected for plasmid isolation. The rescued plasmids were subsequently cloned into *E. coli* Top10 cells to be multiplied so that the plasmids can be sequenced to identify the gene encoding the interactive protein (AD plasmid).

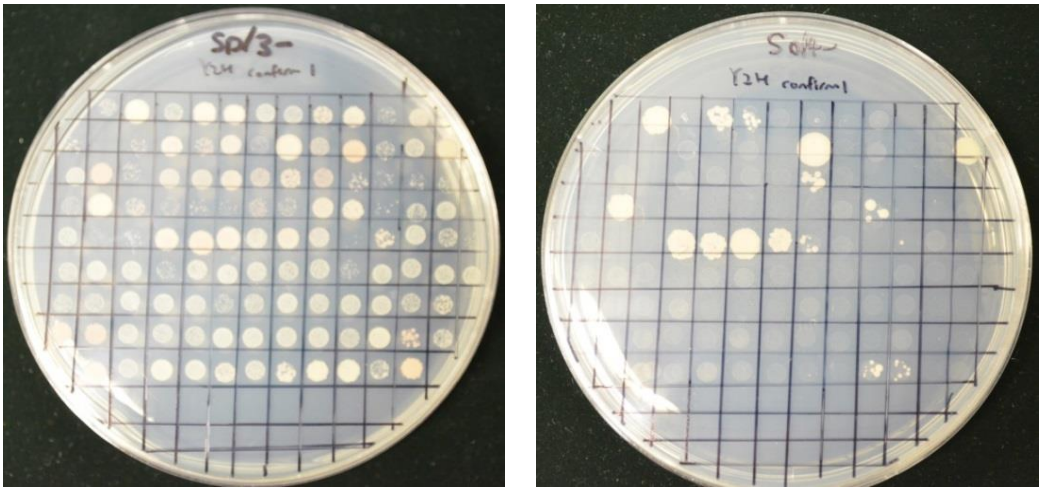

Figure 20. Y2H screening on triple and quadruple dropout plates. 1 µl of the overnight culture is dripped on the triple dropout plate in the corresponding numbered grid. Another 1 µl from the same overnight culture is then dripped onto the exact corresponding numbered grid on the quadruple dropout plates. The 15 colonies that grew in the both the triple and quadruple dropout plates were identified with their numbered grid. Numbers on the grids were removed before the pictures were taken in order to visualize colony growth.

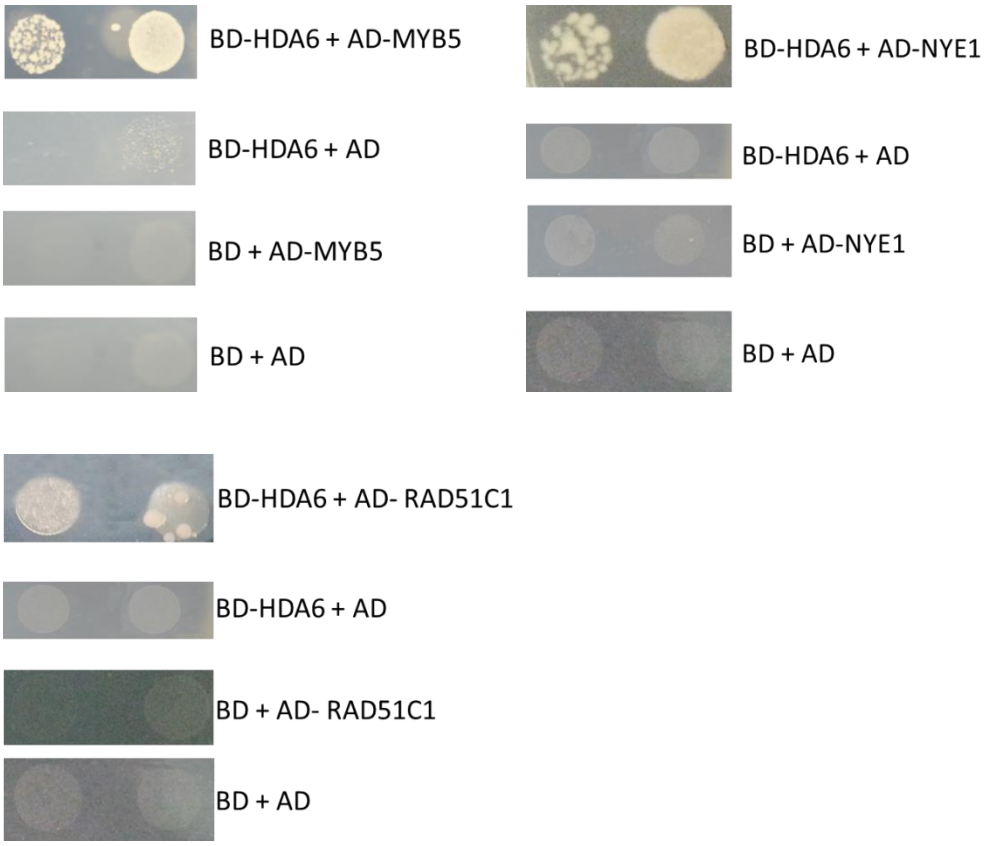

Figure 21. Confirmation of positive interactions to eliminate false positives of MYB5, NYE1 and RAD51C1with HDA6. Full-length HDA6 protein fused to the GAL4 DNA-Binding Domain (BD) was co-expressed with MYB5, NYE1 and RAD51C1 fused to the GAL4 Activation Domain (AD) in Y2HGold and plated on quadruple dropout plates. Absence of colony growth for Y2HGold cells with either empty BD and/or empty AD indicates genuine positives.

Among the 15 plasmids sequenced, *MYB5*, *NYE1* and *RAD51C1*were selected for further investigation. To eliminate false positive interaction of MYB5, NYE1 and RAD51C1with HDA6, full-length HDA6 protein fused to the GAL4 DNA-Binding Domain (BD) was co-expressed with MYB5, NYE1 and RAD51C1 fused to the GAL4 Activation Domain (AD) in Y2HGold and plated on quadruple dropout plates. Absence of colony growth for Y2HGold cells with either empty BD and/or empty AD indicates genuine positives (**Figure 21**).

The transcription repressor MYB5 (Gene ID: 820556) is a 249 residues protein that belongs to the *MYB* family of transcription regulators. It is expressed in trichomes and seeds, including the seed coat and acts as a negative regulator of trichome branching and plays a role in the correct formation of the seed coat and possibly the formation the underlying endosperm layers. It is also involved in seed coat and trichome development (Li et al., 2009).

Non-yellowing protein 1(NYE1) (Gene ID: 828391)is a 268 residues protein that is closely related to the stay-green protein and is upregulated during maximal senescence in *Arabidopsis* especially in leaves, suggesting that it plays a role in chlorophyll catabolism(Ren et al., 2007).

RAD51C DNA repair protein RAD51 homolog 3(Gene ID: 819136) is a 363 residues protein that encodes a protein involved in double stranded break repair via homologous recombination. It is a homodimer that interacts with XRCC3 (Osakabe et al., 2002) and is highly expressed in young flower buds and its expression is induced by genotoxic stress and DNA damage.

3.5 TEK forms a complex with SUVR5 in *Arabidopsis*

SUVR5 is able to recruit various histone-modifying to establish a state of gene repression without DNA methylation, a function that is common among repressive chromatin modifiers of large multi-protein complexes (Caro et al., 2012). A feature of these proteins is in their ability to bind to class I histone deacetylase enzymes such as HDAC1-3 (Fog et al., 2012).

Previous studies have established the interaction between TEK with HDA6 in the Class I HDAC complex, as well as that of FVE/MSI5 with HDA6 (Gu et al., 2011; Xu et al., 2013b), suggesting cooperative action of DNA methylation, histone deacetylation, and TEK gene silencing.

To investigate this link on the possible recruitment of the HDAC complex by SUVR5, we performed co-immunoprecipitation (co-IP) experiments to determine if TEK and SUVR5 interact in vivo. The SUVR5-FLAG-expressing plants were crossed to the TEK-YFP line and from the resulting F1 seedlings, total proteins were extracted for co-IP analysis.

Total protein extracts from seedlings of the TEK-YFP line (a negative control) and F1 of the doubly hemizygous SUVR5-FLAG and TEK-YFP were immunoprecipitated with anti-FLAG agarose. Subsequently, the precipitates were analysed by western blotting with anti-FLAG (recognizing SUVR5-FLAG) and anti-YFP (recognizing TEK-YFP). The input was the protein extract of doubly hemizygous SUVR5-FLAG and TEK-YFP (**Figure 22**). From this result, we infer that TEK is in a complex with SUVR5 in *Arabidopsis*.

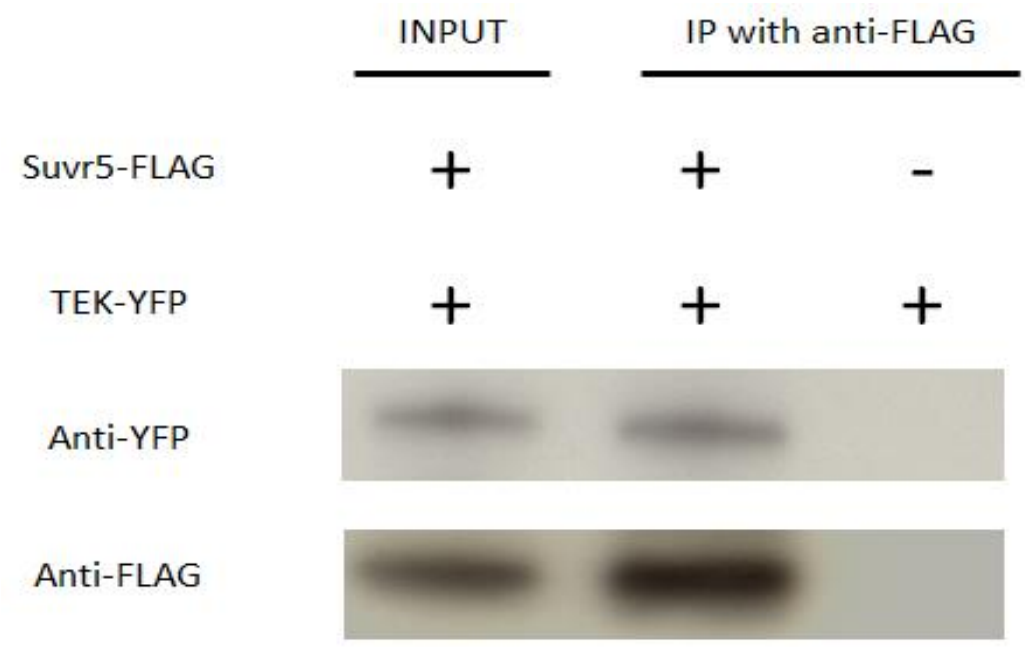

Figure 22. Co-immunoprecipitation of SUVR5 with TEK in seedlings. Total protein extracts from seedlings of the TEK-YFP line (a negative control) and F1 of the doubly hemizygous SUVR5-FLAG and TEK-YFP were immunoprecipitated with anti-FLAG agarose. The input was the protein extract of doubly hemizygous SUVR5-FLAG and TEK-YFP.

3.6 Genetic interaction of *HDA6* and *SUVR5*

The *FWA* gene locus is a known target of the histone deacetylase complex and is usually repressed by histone deacetylation to prevent its expression. De-repression of *FWA* can take place when the histone deacetylation pathway is affected and result in a late flowering phenotype. Total RNA were extracted for generating cDNA using RT-PCR. Using the cDNA as template, two rounds PCR with 33 and 37 cycles on *FWA* was performed (**Figure 23**).

By using the ImageJ software, we can compare the intensity of the DNA band from the *FWA* product semi-quantitatively (**Figure 24**). This method allowed the comparison of the intensity of the DNA bands with reference to the band with the highest intensity as 100%.

FWA mRNA levels from *suvr5* is about 30-20% (33 cycles and 37 cycles respectively) of the *hda6* mutants while the expression level of *FWA* mRNA is very similar for *hda6* and *hda6/suvr5* at 95%-90% (33 cycles and 37 cycles respectively).

Consistent with the non-additive effect of HDA6 and SUVR5 on *FWA* silencing, we found that the *hda6/suvr5* double mutants are phenotypically similar to the *hda6* mutant (**Figure 25** and **Figure 26**) while the late flowering effect is less pronounced in the *suvr5* mutant.

Figure 23. PCR on cDNA generated from total RNA extracted from Col wild type, *hda6*, *suvr5* and *hda6 suvr5* double mutants. *TUB2* products are the loading controls while PCR for *FWA* is performed for 33 and 37 cycles.

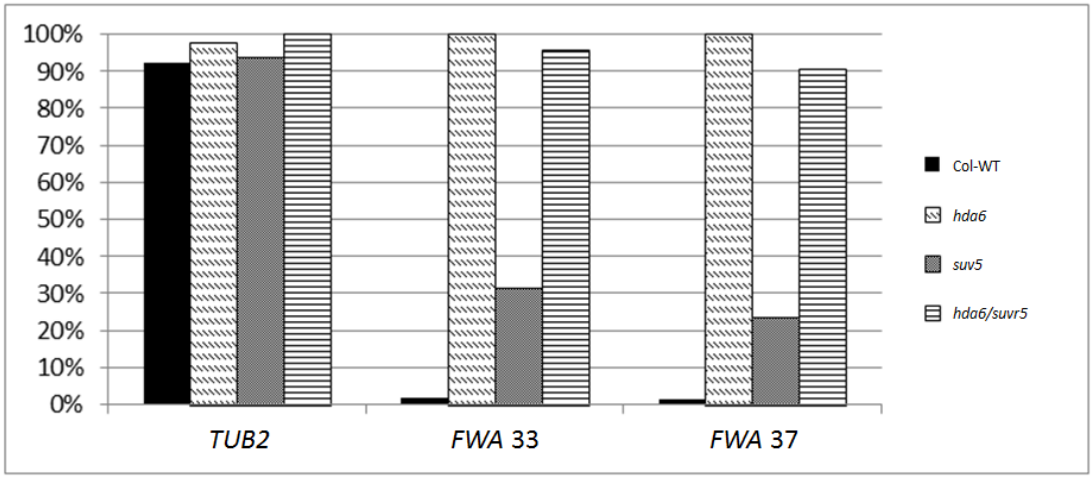

Figure 24. Intensity of DNA bands are compared using ImageJ. Reference to the band with the highest intensity is at 100%. *TUB2* is used as a loading control and two rounds of *FWA* with 33 and 37 cycles.

Figure 25. Late flowering phenotypes from *hda6* and *suvr5* mutants. Clock-wise from top-left: Col-WT, *hda6*, *hda6/suvr5* double mutant, *suvr5*. The *hda6/suvr5* double mutants are phenotypically similar to the *hda6* mutant while the late flowering effect is less pronounced in the *suvr5* mutant.

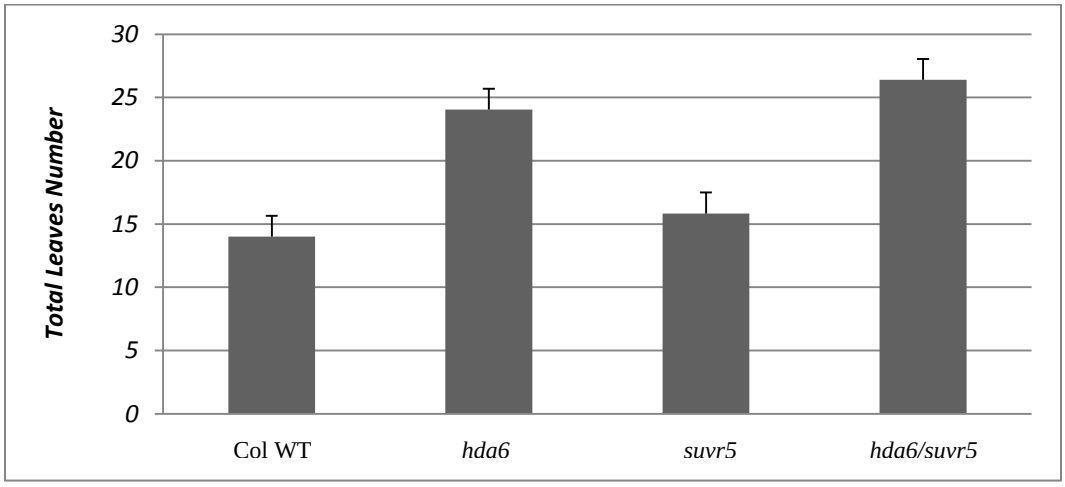

Figure 26. Total number of leaves counted at the first appearance of the floral meristems. Col WT = 14 (sd = 1.65145) ; *hda6* = 24.04 (sd = 2.13073) ; *suvr5* = 15.833 (sd = 1.6967) ; *hda6/suvr5* = 26.3846 (sd = 1.50214).

Chapter 4 Discussion

Juxtaposition of the modelled structure for HDA6 with the solved X-ray crystal structure of HDAC2 revealed much congruence, most strikingly that of the histone binding lipophilic 'tube' which leads from the surface to the catalytic site and the 'foot pocket' immediately adjacent to this catalytic machinery(Bressi et al., 2010) . The folds, sheets and helices are largely conserved and so are the key residues that interact directly with the ligand at the catalytic sites. Taken together, these characteristics suggest that *Arabidopsis* HDA6 is a histone deacetylase, with functional and structural similarities to its human homolog HDAC2. HDA6 is the key catalytic component of the *Arabidopsis* HDAC complex which is made up of other subunits in order to carry out its histone deacetylation activities (Probst et al., 2004; Aufsatz et al., 2007; Gu et al., 2011). The structural similarity of HDA6 and HDAC2 suggests that HDA6 would interact with *Arabidopsis* homologs of human RBBP4 and RBBP7 in the human HDAC complex. Investigations into HDA6 interaction with *Arabidopsis* MSI1 and FVE respectively would begin by making direct comparisons with that the spatial arrangement of human HDAC2 complex interacting with the chromatin via the histones H3/H4 heterodimers.

The *Drosophila* Nurf55 WD40 repeat protein is an integral subunit of the CAF-1 complex. WD repeat proteins related to Nurf55 are associated with histone deacetylase and histone acetyltransferases, leading to suggestions that Nurf55 and its homologs may function as a common platform for the assembly of protein complexes involved in chromatin metabolism (Martinez-Balbas et al., 1998). Nurf55 has two independent binding sites for interactions with the H3 tail and H4

helix 1 when bound to the H3/H4 heterodimer (Nowak et al., 2011; Schmitges et al., 2011).

Strikingly, although these two interaction surfaces are distinctly different spatially occupying different regions of the protein and that they interact with different sets of ligands, both sites are conserved in MSI1/2/3 but not FVE and MSI5 **(Figure 15)**. This strongly suggested that MSI1/2/3 is able to interact with both histone H4 and Su(z)12 non-simultaneously at one distinct site and interacting with Histone H3 and FOG-1 non-simultaneously at another site.

This feature is reminiscence of *Arabidopsis* MSI1, a core histone-binding subunit of PRC2 complexes that binds directly to helix 1 of histone H4. In addition, EMF2 and VRN2, both homologs of Su(z)12, interact with MSI1 but not with FVE, an curious feature as FVE is key component of HDAC complexes. Indeed, the human RBBP4 (homologous to MSI1) and RBBP7 (homologous to FVE) are distinguished from each other whereby RBBP4 is a component of the CAF-1complex while RBBP7 is found exclusively with HAT1 (Parthun et al., 1996; Verreault et al., 1998; Gu et al., 2011).

From **Figure 15**, it is observed that the residues responsible for binding to histone H4 and Su(z)12 overlaps, and that the binding site of histone H3 overlaps with FRIEND OF GATA protein 1 (FOG-1), in agreement with what was previously reported (Nowak et al., 2011). The identity of the residues interacting with FOG-1 and their spatial conformation are conserved in both RBBP4 and

RBBP7 supporting studies that shown FOG-1 interacts with both RBBP4 and RBBP7 in vitro (Hong et al., 2005; Lejon et al., 2011).

Notably, the binding site for H3/FOG-1 is very distinct from that of H4/Su(z)12. Indeed, the electrostatic regions in **Figure 16** and **Figure 18** and the cartoon illustration in **Figure 17** and **Figure 19** that represented the spatial arrangement of the residues that produced the histone H4 binding groove of Nurf55 is more similar to that of the model of MSI1 than FVE. Despite the high sequence homology, the structural and functional similarity to that of Nurf55 strongly favoured MSI1 over FVE.

WD40 proteins such as Nurf55 are largely non-catalytic and serve important structural functions by providing interaction surfaces in the formation of complexes (Martinez-Balbas et al., 1998; Li and Roberts, 2001). Their shape and size strongly suggested multiple interaction surfaces, tolerating interaction with a range of proteins and with multiple proteins simultaneously. As demonstrated by Nurf55, residues that interacts with histone H3 is positioned at the top of the β-propeller while histone H4 helix 1 and Su(z)12 binds to Nurf55 at the edge of the β-propeller(Lejon et al., 2011). In the PRC2 complex, both Nurf55 and Su(z)12 are required to tether the complex to nucleosomes and the absence of either component would abolish this binding (Nekrasov et al., 2005).

The histone H4 binding groove, which is structurally very similar in both Nurf55 and MSI1, also serves also as the binding site for Su(z)12. Through this interaction with Su(z)12, MSI1 becomes a component of the PRC2 complex. The

distinct difference of the histone H4 groove between MSI1 and FVE suggests that FVE is unable to interact directly with Su(z)12, illustrating why FVE is not part of the PRC2 complex (**Figure 16 to Figure 19**). As yet, it is unknown if Nurf55, or MSI1 of the PRC2 complex binds only Su(z)12 or both Su(z)12 and histone H4 at different stages of the catalytic process(Nowak et al., 2011). However, it is clear that MSI1 would play a crucial role in the recruitment of the H3/H4 heterodimers for the PRC2 and FIS complexes.

Nurf55binds directly to helix 1 of the histone fold of histone H4, a region that is not accessible when H4 is in chromatin (Nowak et al., 2011). The amino acid residues that formed the interaction surface of the histone H4 groove are found in the positions of 20 – 32, 343 – 378 and 418 – 419. From **Figure 15**, we can see that these residues are conserved in Nurf55, RBBP4, MSI1, MSI2 and MSI3 but not in FVE and MSI5. The non-conservation of these crucial residues concurs with the distinct difference in the shape and structure of this histone H4 groove.

The three novel partners of HDA6 identified via the Y2H assays MYB5, NYE1 and RAD51C will need to be investigated further to elucidate their functions and interactions with HDA6 and their possible roles in HDAC activities. MYB5 is a transcription regulator that can potentially form a complex with HDA6 which would carry out chromatin modifications via histone deacetylation. RAD51C protein is involved in the homologous recombination repair (HRR) pathway of double-stranded DNA breaks and could possibly work together with HDA6 to modify the chromatin to adopt a more open configuration in order for

DNA repair to take place. NYE1 is a poorly studied that is upregulated during maximal senescence in the *Arabidopsis* life cycle and is speculated to be involved in chlorophyll catabolic process. Intriguingly, this process does not seem to have much relation with HDA6 function such as chromatin modification and a more detailed study need to be performed in order to elucidate their potential interactions.

TEK is a nuclear matrix protein with AT-hook DNA binding motifs (Reeves and Nissen, 1990; Yasui et al., 2002; Cai et al., 2003) that represses the expression of various Transposable Elements (TE), the TE-like repeat-containing floral repressor gene *FLC* and *FWA* (Soppe et al., 2000; Gazzani et al., 2003; Liu et al., 2004). This repression of *FLC* and *FWA* by TEK to regulate floral transition is mediated via various chromatin structure modifications such as histone deacetylation and it binds to an *FLC*-repressive regulatory region and the silencing repeats of *FWA* through the association with FVE and MSI5, both of which are components of HDACs (Ausin et al., 2004; Gu et al., 2011; Xu et al., 2013b).

The histone-lysine N-methyltransferase SUVR5 is a plant Su(var)3–9 homolog with a SET histone methyltransferase domain, mediates H3K9me2 deposition and regulates gene expression in a DNA methylation–independent manner (Caro et al., 2012). Together with its binding partner LDL1/SWP1, it acts as a floral transition regulator via repression of *FLC* through H3K9me2 deposition in a DNA methylation–independent manner (Krichevsky et al., 2007).

The direct interaction between SUVR5 and TEK (**Figure 22**) led us to propose a model for the HDAC complex comprising of four proteins: HDA6, FVE/MSI5, TEK and SUVR5. This complex would regulate downstream genes by silencing via histone deacetylation and histone H3K9 methylation, of which would be important floral transition targets such as *FLC* and *FWA*.

Consistent with our model that HDA6 and SUVR5 act in the same complex, *suvr5* is non-additive with *hda6* on *FWA* mRNA levels (**Figure 23** and **Figure 24**). De-repression from *suvr5* is about 70-80% of the *hda6* mutants while the expression level of *FWA* mRNA is very similar (less than 10% difference) for *hda6* and *hda6/suvr5*. This is further supported by the measure of flowering time (**Figure 26**) where the double mutant of *hda6/suvr5* is only slightly stronger than that of the *hda6* mutant. In short, SUVR5 works together with HDA6 to silence *FWA* expression in *Arabidopsis*.

HDA6 is responsible for the deacetylation of lysine residues on the N-terminal of the core histones, playing a key role in epigenetic repression and transcriptional regulation. However, its histone deacetylase activities manifests only via large multiprotein complexes, for which we propose that one of these complexes would comprise of HDA6, FVE/MSI5, TEK and SUVR5. This complex would be critical in the epigenetical control of floral transition in *Arabidopsis* via transcriptional regulation of key floral genes such as *FWA*.

References

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science **309,** 1052-1056.

Ahringer, J. (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends in genetics : TIG **16,** 351-356.

Aloy, P., Stark, A., Hadley, C., and Russell, R.B. (2003). Predictions without templates: new folds, secondary structure, and contacts in CASP5. Proteins **53 Suppl 6,** 436-456.

Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proceedings of the National Academy of Sciences of the United States of America **47,** 1309-1314.

Aufsatz, W., Stoiber, T., Rakic, B., and Naumann, K. (2007). Arabidopsis histone deacetylase 6: a green link to RNA silencing. Oncogene **26,** 5477-5488.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. (2002). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. The EMBO journal **21,** 6832-6841.

Ausin, I., Alonso-Blanco, C., Jarillo, J.A., Ruiz-Garcia, L., and Martinez-Zapater, J.M. (2004). Regulation of flowering time by FVE, a retinoblastoma-associated protein. Nature genetics **36,** 162-166.

Barak, O., Lazzaro, M.A., Lane, W.S., Speicher, D.W., Picketts, D.J., and Shiekhattar, R. (2003). Isolation of human NURF: a regulator of Engrailed gene expression. The EMBO journal **22,** 6089-6100.

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

Bartee, L., Malagnac, F., and Bender, J. (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes & development **15,** 1753-1758.

Benevolenskaya, E.V. (2007). Histone H3K4 demethylases are essential in development and differentiation. Biochemistry and cell biology = Biochimie et biologie cellulaire **85,** 435-443.

Bouveret, R., Schonrock, N., Gruissem, W., and Hennig, L. (2006). Regulation of flowering time by Arabidopsis MSI1. Development **133,** 1693-1702.

Bressi, J.C., Jennings, A.J., Skene, R., Wu, Y., Melkus, R., De Jong, R., O'Connell, S., Grimshaw, C.E., Navre, M., and Gangloff, A.R. (2010). Exploration of the HDAC2 foot pocket: Synthesis and SAR of substituted N-(2-aminophenyl)benzamides. Bioorganic & medicinal chemistry letters **20,** 3142-3145.

Brunmeir, R., Lagger, S., and Seiser, C. (2009). Histone deacetylase HDAC1/HDAC2-controlled embryonic development and cell differentiation. The International journal of developmental biology **53,** 275-289.

Cai, S., Han, H.J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nature genetics **34,** 42-51.

Calonje, M., Sanchez, R., Chen, L., and Sung, Z.R. (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis. The Plant cell **20,** 277-291.

Caro, E., Stroud, H., Greenberg, M.V., Bernatavichute, Y.V., Feng, S., Groth, M., Vashisht, A.A., Wohlschlegel, J., and Jacobsen, S.E. (2012). The SET-domain protein SUVR5 mediates H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation-independent manner. PLoS genetics **8,** e1002995.

Carrozza, M.J., Utley, R.T., Workman, J.L., and Cote, J. (2003). The diverse functions of histone acetyltransferase complexes. Trends in genetics : TIG **19,** 321-329.

Chen, Z.J., and Pikaard, C.S. (1997). Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. Genes & development **11,** 2124-2136.

Chivian, D., Kim, D.E., Malmstrom, L., Schonbrun, J., Rohl, C.A., and Baker, D. (2005). Prediction of CASP6 structures using automated Robetta protocols. Proteins **61 Suppl 7,** 157-166.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. The Plant journal : for cell and molecular biology **16,** 735-743.

de Ruijter, A.J., van Gennip, A.H., Caron, H.N., Kemp, S., and van Kuilenburg, A.B. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. The Biochemical journal **370,** 737-749.

Doyle, M.R., and Amasino, R.M. (2009). A single amino acid change in the enhancer of zeste ortholog CURLY LEAF results in vernalization-independent, rapid flowering in Arabidopsis. Plant physiology **151,** 1688-1697.

Earley, K.W., Pontvianne, F., Wierzbicki, A.T., Blevins, T., Tucker, S., Costa-Nunes, P., Pontes, O., and Pikaard, C.S. (2010). Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. Genes & development **24,** 1119-1132.

Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. The Plant cell **18,** 1166-1176.

Fleischer, T.C., Yun, U.J., and Ayer, D.E. (2003). Identification and characterization of three new components of the mSin3A corepressor complex. Molecular and cellular biology **23,** 3456-3467.

Fog, C.K., Galli, G.G., and Lund, A.H. (2012). PRDM proteins: important players in differentiation and disease. BioEssays : news and reviews in molecular, cellular and developmental biology **34,** 50-60.

Furuyama, T., Dalal, Y., and Henikoff, S. (2006). Chaperone-mediated assembly of centromeric chromatin in vitro. Proceedings of the National Academy of Sciences of the United States of America **103,** 6172-6177.

Gazzani, S., Gendall, A.R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. Plant physiology **132,** 1107-1114.

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., and Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature **386,** 44-51.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic acids research **38,** W695-699.

Gu, X., Jiang, D., Yang, W., Jacob, Y., Michaels, S.D., and He, Y. (2011). Arabidopsis homologs of retinoblastoma-associated protein 46/48 associate with

a histone deacetylase to act redundantly in chromatin silencing. PLoS genetics **7,** e1002366.

Guitton, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U., and Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Development **131,** 2971-2981.

Haber, E., and Anfinsen, C.B. (1962). Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. The Journal of biological chemistry **237,** 1839-1844.

Hamilton, J.P. (2011). Epigenetics: principles and practice. Digestive diseases **29,** 130-135.

He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. (2009). NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. Genes & development **23,** 318-330.

He, Y. (2009). Control of the transition to flowering by chromatin modifications. Molecular plant **2,** 554-564.

He, Y., Michaels, S.D., and Amasino, R.M. (2003). Regulation of flowering time by histone acetylation in Arabidopsis. Science **302,** 1751-1754.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends in genetics : TIG **27,** 389-396.

Hennig, L., Taranto, P., Walser, M., Schonrock, N., and Gruissem, W. (2003). Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development. Development **130,** 2555-2565.

Hong, W., Nakazawa, M., Chen, Y.Y., Kori, R., Vakoc, C.R., Rakowski, C., and Blobel, G.A. (2005). FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. The EMBO journal **24,** 2367-2378.

Huang, E.Y., Zhang, J., Miska, E.A., Guenther, M.G., Kouzarides, T., and Lazar, M.A. (2000). Nuclear receptor corepressors partner with class II histone deacetylases in a Sin3-independent repression pathway. Genes & development **14,** 45-54.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science **288,** 1422-1425.

Jiang, D., Wang, Y., Wang, Y., and He, Y. (2008). Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components. PloS one **3,** e3404.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes & development **19,** 489-501.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation Tagging of the Floral Inducer FT. Science **286,** 1962-1965.

Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B., and Araki, T. (2001). FASCIATA genes for chromatin assembly factor-1 in arabidopsis maintain the cellular organization of apical meristems. Cell **104,** 131-142.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols **4,** 363-371.

Khochbin, S., and Wolffe, A.P. (1997). The origin and utility of histone deacetylases. FEBS letters **419,** 157-160.

Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic acids research **32,** W526-531.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science **303,** 521-523.

Kishimoto, N., Sakai, H., Jackson, J., Jacobsen, S.E., Meyerowitz, E.M., Dennis, E.S., and Finnegan, E.J. (2001). Site specificity of the Arabidopsis METI DNA methyltransferase demonstrated through hypermethylation of the superman locus. Plant molecular biology **46,** 171-183.

Knoepfler, P.S., and Eisenman, R.N. (1999). Sin meets NuRD and other tails of repression. Cell **99,** 447-450.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., James, K.D., Lefebvre, G.C., Bruce, A.W., Dovey, O.M., Ellis, P.D., Dhami, P., Langford, C.F., Weng, Z., Birney, E., Carter, N.P., Vetrie, D., and Dunham, I. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome research **17,** 691-707.

Kohler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. The EMBO journal **22,** 4804-4814.

Korenjak, M., Taylor-Harding, B., Binne, U.K., Satterlee, J.S., Stevaux, O., Aasland, R., White-Cooper, H., Dyson, N., and Brehm, A. (2004). Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. Cell **119,** 181-193.

Krichevsky, A., Gutgarts, H., Kozlovsky, S.V., Tzfira, T., Sutton, A., Sternglanz, R., Mandel, G., and Citovsky, V. (2007). C2H2 zinc finger-SET histone methyltransferase is a plant-specific chromatin modifier. Developmental biology **303,** 259-269.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. Molecular cell **14,** 183-193.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews. Genetics **11,** 204-220.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell **72,** 73-84.

Lejon, S., Thong, S.Y., Murthy, A., AlQarni, S., Murzina, N.V., Blobel, G.A., Laue, E.D., and Mackay, J.P. (2011). Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48.FOG-1 complex. The Journal of biological chemistry **286,** 1196-1203.

Li, D., and Roberts, R. (2001). WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cellular and molecular life sciences : CMLS **58,** 2085-2097.

Li, G., Hall, T.C., and Holmes-Davis, R. (2002). Plant chromatin: development and gene control. BioEssays : news and reviews in molecular, cellular and developmental biology **24,** 234-243.

Li, J., Wang, J., Wang, J., Nawaz, Z., Liu, J.M., Qin, J., and Wong, J. (2000). Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3. The EMBO journal **19,** 4342-4350.

Li, S.F., Milliken, O.N., Pham, H., Seyit, R., Napoli, R., Preston, J., Koltunow, A.M., and Parish, R.W. (2009). The Arabidopsis MYB5 transcription factor regulates

mucilage synthesis, seed coat development, and trichome morphogenesis. The Plant cell **21,** 72-89.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science **292,** 2077-2080.

Lippman, Z., and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. Nature **431,** 364-370.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS biology **1,** E67.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., and Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430,** 471-476.

Liu, J., He, Y., Amasino, R., and Chen, X. (2004). siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes & development **18,** 2873-2878.

Loyola, A., and Almouzni, G. (2004). Histone chaperones, a supporting role in the limelight. Biochimica et biophysica acta **1677,** 3-11.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389,** 251-260.

Ma, P., and Schultz, R.M. (2013). Histone deacetylase 2 (HDAC2) regulates chromosome segregation and kinetochore function via H4K16 deacetylation during oocyte maturation in mouse. PLoS genetics **9,** e1003377.

Martinez-Balbas, M.A., Tsukiyama, T., Gdula, D., and Wu, C. (1998). Drosophila NURF-55, a WD repeat protein involved in histone metabolism. Proceedings of the National Academy of Sciences of the United States of America **95,** 132-137.

Matzke, M.A., and Birchler, J.A. (2005). RNAi-mediated pathways in the nucleus. Nature reviews. Genetics **6,** 24-35.

Murzin, A.G. (1992). Structural principles for the propeller assembly of beta-sheets: the preference for seven-fold symmetry. Proteins **14,** 191-201.

Murzina, N.V., Pei, X.Y., Zhang, W., Sparkes, M., Vicente-Garcia, J., Pratap, J.V., McLaughlin, S.H., Ben-Shahar, T.R., Verreault, A., Luisi, B.F., and Laue, E.D. (2008). Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure **16,** 1077-1085.

Nekrasov, M., Wild, B., and Muller, J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO reports **6,** 348-353.

Nicolas, E., Morales, V., Magnaghi-Jaulin, L., Harel-Bellan, A., Richard-Foy, H., and Trouche, D. (2000). RbAp48 belongs to the histone deacetylase complex that associates with the retinoblastoma protein. The Journal of biological chemistry **275,** 9797-9804.

North, B.J., Marshall, B.L., Borra, M.T., Denu, J.M., and Verdin, E. (2003). The human Sir2 ortholog, SIRT2, is an NAD+-dependent tubulin deacetylase. Molecular cell **11,** 437-444.

Nowak, A.J., Alfieri, C., Stirnimann, C.U., Rybin, V., Baudin, F., Ly-Hartig, N., Lindner, D., and Muller, C.W. (2011). Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. The Journal of biological chemistry **286,** 23388-23396.

Osakabe, K., Yoshioka, T., Ichikawa, H., and Toki, S. (2002). Molecular cloning and characterization of RAD51-like genes from Arabidopsis thaliana. Plant molecular biology **50,** 71-81.

Papa, C.M., Springer, N.M., Muszynski, M.G., Meeley, R., and Kaeppler, S.M. (2001). Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. The Plant cell **13,** 1919-1928.

Parthun, M.R., Widom, J., and Gottschling, D.E. (1996). The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell **87,** 85-94.

Probst, A.V., Fagard, M., Proux, F., Mourrain, P., Boutet, S., Earley, K., Lawrence, R.J., Pikaard, C.S., Murfett, J., Furner, I., Vaucheret, H., and Mittelsten Scheid, O. (2004). Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. The Plant cell **16,** 1021-1034.

Reeves, R., and Nissen, M.S. (1990). The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. The Journal of biological chemistry **265,** 8573-8582.

Ren, G., An, K., Liao, Y., Zhou, X., Cao, Y., Zhao, H., Ge, X., and Kuai, B. (2007). Identification of a novel chloroplast protein AtNYE1 regulating chlorophyll degradation during leaf senescence in Arabidopsis. Plant physiology **144,** 1429-1441.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell **129,** 1311-1323.

Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., DeSalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. BMC genomics **10,** 143.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. Annual review of biochemistry **70,** 81-120.

Saze, H., and Kakutani, T. (2011). Differentiation of epigenetic modifications between transposons and genes. Current opinion in plant biology **14,** 81-87.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., Bunker, R.D., Wirth, U., Bouwmeester, T., Bauer, A., Ly-Hartig, N., Zhao, K., Chan, H., Gu, J., Gut, H., Fischle, W., Muller, J., and Thoma, N.H. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Molecular cell **42,** 330-341.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods **9,** 671-675.

Schrodinger, LLC. (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. The EMBO journal **25,** 4638-4649.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A., and Coupland, G. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes & development **20,** 898-912.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G.

(2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology **7,** 539.

Skolnick, J., Kihara, D., and Zhang, Y. (2004). Development and large scale benchmark testing of the PROSPECTOR_3 threading algorithm. Proteins **56,** 502-518.

Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. Trends in biochemical sciences **24,** 181-185.

Song, J.J., Garlick, J.D., and Kingston, R.E. (2008). Structural basis of histone H4 recognition by p55. Genes & development **22,** 1313-1318.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Molecular cell **6,** 791-802.

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.E., Chen, J., Lazar, M.A., Blobel, G.A., and Vakoc, C.R. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Molecular and cellular biology **28,** 2825-2839.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature **403,** 41-45.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell **116,** 51-61.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science **272,** 408-411.

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual review of plant biology **59,** 573-594.

Urnov, F.D., Wolffe, A.P., and Guschin, D. (2001). Molecular mechanisms of corepressor function. Current topics in microbiology and immunology **254,** 1-33.

Urnov, F.D., Yee, J., Sachs, L., Collingwood, T.N., Bauer, A., Beug, H., Shi, Y.B., and Wolffe, A.P. (2000). Targeting of N-CoR and histone deacetylase 3 by the oncoprotein v-erbA yields a chromatin infrastructure-dependent transcriptional repression pathway. The EMBO journal **19,** 4074-4090.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1998). Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. Current biology : CB **8,** 96-108.

Wang, D., Tyson, M.D., Jackson, S.S., and Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **103,** 13244-13249.

Wen, Y.D., Perissi, V., Staszewski, L.M., Yang, W.M., Krones, A., Glass, C.K., Rosenfeld, M.G., and Seto, E. (2000). The histone deacetylase-3 complex contains nuclear receptor corepressors. Proceedings of the National Academy of Sciences of the United States of America **97,** 7202-7207.

White, F.H., Jr. (1961). Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease. The Journal of biological chemistry **236,** 1353-1360.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature genetics **41,** 630-634.

Wu, K., Zhang, L., Zhou, C., Yu, C.W., and Chaikam, V. (2008). HDA6 is required for jasmonate response, senescence and flowering in Arabidopsis. Journal of experimental botany **59,** 225-234.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS biology **2,** E104.

Xu, Y., Gan, E.S., and Ito, T. (2013a). The AT-hook/PPC domain protein TEK negatively regulates floral repressors including MAF4 and MAF5. Plant signaling & behavior **8**.

Xu, Y., Wang, Y., Stroud, H., Gu, X., Sun, B., Gan, E.S., Ng, K.H., Jacobsen, S.E., He, Y., and Ito, T. (2013b). A matrix protein silences transposons and repeats through interaction with retinoblastoma-associated proteins. Current biology : CB **23,** 345-350.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature **419,** 641-645.

Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and gene silencing. Cell **128,** 763-776.

Zhang, H., and Zhu, J.K. (2011). RNA-directed DNA methylation. Current opinion in plant biology **14,** 142-147.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes & development **13,** 1924-1935.

Zhang, Y., Li, N., Caron, C., Matthias, G., Hess, D., Khochbin, S., and Matthias, P. (2003). HDAC-6 interacts with and deacetylates tubulin and microtubules in vivo. The EMBO journal **22,** 1168-1179.

Zhou, H., and Zhou, Y. (2005). Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins **58,** 321-328.

Zhou, H., and Skolnick, J. (2009). Protein structure prediction by pro-Sp3-TASSER. Biophysical journal **96,** 2119-2127.

Zhou, H., Pandit, S.B., and Skolnick, J. (2009). Performance of the Pro-sp3-TASSER server in CASP8. Proteins **77 Suppl 9,** 123-127.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science **299,** 716-719.