

**Deciphering the role of PRC2
accessory proteins in promoting
cold-induced epigenetic
switching in *Arabidopsis thaliana***

Mathias Langkilde Nielsen

2021

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

University of East Anglia

John Innes Centre

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current U.K. Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

All cells in multicellular organisms contain the same genetic material. Cell differentiation and adaptation to environmental signals require the accurate switching of gene expression states and the maintenance of those states through cell division, even in the absence of the initial signal. Polycomb complexes play an important role in these switching and epigenetic silencing mechanisms.

The *Arabidopsis thaliana* developmental regulator *FLOWERING LOCUS C* (*FLC*) has emerged as an excellent system to dissect Polycomb Repressive Complex 2 (PRC2) regulation. *FLC* is expressed as plants germinate in autumn to prevent premature flowering. The prolonged cold of winter epigenetically silences *FLC*, aligning flowering with the following spring. The epigenetic silencing involves a cold-induced PRC2 switch and deposition of the repressive chromatin mark H3K27me3 in the *FLC* nucleation region. This switch requires the core PRC2 and two accessory proteins VIN3 and VRN5, but the molecular mechanism underlying the switch and the subsequent memory of cold exposure remain largely unknown.

The work in this thesis aimed to understand the role of VIN3 and VRN5 in the repression of *FLC*. In the first Results chapter, the mechanism of VIN3 binding at *FLC* is studied; VIN3 was shown to contain a domain that can associate with nucleic acids. The second Results chapter dissects the function of the conserved C-terminal domain of VIN3 and VRN5; this domain mediates protein-protein interactions between VIN3 and VRN5 and adopts a novel polymerization fold. In the third Results chapter, the local chromatin conformation at *FLC* is studied; the expression state of *FLC* was found to correlate with distinct 3D conformation. The 3' end of *FLC* containing a promoter for *FLC* antisense transcripts is involved in setting this 3D conformation. Overall, the work advances our understanding of the molecular function of PRC2 accessory proteins in epigenetic silencing.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

ABSTRACT	3
ACKNOWLEDGEMENTS	5
CONTENTS	7
ABBREVIATIONS	12
LIST OF FIGURES	14
LIST OF TABLES	17
1 GENERAL INTRODUCTION	18
1.1 CELLULAR MEMORY	19
1.1.1 <i>Chromatin – central for cis memory</i>	21
1.1.1.1 Chromatin – Active or repressive	22
1.2 <i>FLC</i> – A MODEL FOR EPIGENETIC MEMORY	24
1.2.1 <i>Vernalization</i>	24
1.2.2 <i>Molecular events during FLC repression</i>	25
1.3 POLYCOMB REPRESSION	28
1.3.1 <i>Polycomb Repressive Complex 2</i>	29
1.3.1.1 Accessory proteins make PRC2 specific	32
1.3.1.1.1 PRC2.1 – The PCL containing PRC2 complex	33
1.3.1.1.2 PRC2.2 – The JARID2/AEBP2 PRC2 complex	34
1.3.1.2 Arabidopsis PRC2 proteins	36
1.3.1.2.1 The vernalization PRC2 complex	36
1.3.1.2.2 Arabidopsis accessory proteins	37
1.3.2 <i>Polycomb Repressive Complex 1</i>	41
1.4 RECRUITMENT OF POLYCOMB	42
1.4.1 <i>PhoRC</i>	43
1.4.2 <i>Recruitment – a Rethink</i>	44
1.4.3 <i>Recruitment of PRC2 in Arabidopsis</i>	45
1.4.4 <i>RNA in Polycomb recruitment</i>	47
1.4.4.1 The role of RNA in the association of PRC2 with <i>FLC</i>	47
1.5 MEMORY IN 3D	50
1.5.1 <i>Arabidopsis chromatin loops</i>	52
1.6 THESIS OUTLINE	53
1.7 STATEMENT OF COLLABORATIONS	55
2 VIN3 RECRUITMENT TO THE NUCLEATION REGION	56
2.1 INTRODUCTION	56

This work was supported by the UKRI Biotechnology and Biological Sciences Research Council Norwich Research Park Biosciences Doctoral Training Partnership

Acknowledgements

Firstly, I would like to express my deepest gratitude and appreciation to all the members of the Dean lab over the last four years. Thank you for all the help and support during my PhD, I feel extremely privileged to have been given the chance to work in the same lab as so many extremely talented scientists. A special thank you to Yusheng Zhao, who took me under his wing and helped and supported me throughout my entire PhD.

With this PhD project I have had the honour of being part of the newly established collaborations between Caroline Dean and Mariann Bienz. I feel incredibly privileged to have been able to join this collaboration as an inexperienced student.

I am forever grateful for the introduction to the world of structural biology and biochemistry that has come with this collaboration. A very special thank you to Marc, Elsa and Anna for this collaboration.

I could not write this section without expressing my deepest appreciation and thankfulness to my fellow PhD students Jade and Svenja, as well as Yusheng, Anna, Marc and Tricia, for their enormous support for my well-being over the years; without their support this thesis would never have happened.

A special thank you to my final-year housemate Basti for his support during the very difficult final year of my PhD. I would also like to extend my sincere appreciation to my “Twitter-friend” Tiago who, even though we have not been able to meet in real life because of COVID, has been there for me every day over the last year.

Finally, I need to express my gratefulness for having such an amazing scientific mentor as Caroline Dean as my supervisor during my PhD. No words exist to say how much I have learned over my four years in the lab about how to be a great scientist.

Mathias Langkilde Nielsen
John Innes Centre, Norwich
2021

2.2	RESULTS	61
2.2.1	<i>The VEL proteins contain additional domains</i>	61
2.2.2	<i>Structural determination of the PhVIN3 composite domain</i>	65
2.2.3	<i>Composite domain of VIN3 associates with nucleic acid</i>	66
2.2.3.1	VIN3-CD associates with DNA non-specifically <i>in vitro</i>	66
2.2.4	<i>The CD associates with nucleic acids generally in vitro</i>	68
2.2.4.1	RNA	68
2.2.4.2	Different nucleic acids	69
2.2.5	<i>The atypical WH/Bromo-like domain is required for binding</i>	70
2.2.6	<i>ChIP-seq</i>	72
2.2.7	<i>The VEL proteins interact with other transcription regulators</i>	72
2.2.7.1	VIN3 associates with other PRC1-related proteins	76
2.2.8	<i>VAL1 is not required for VIN3 association with FLC</i>	78
2.3	DISCUSSION	78
2.3.1	<i>Several chromatin-interacting domains could enhance the avidity of VIN3 to chromatin</i>	79
2.3.1.1	The atypical VIN3 plant homeo-domain	80
2.3.1.2	A conserved 4-helix bundle could link VIN3 to nucleic acid	81
2.3.2	<i>VIN3 recruitment</i>	82
2.3.2.1	VAL1 – the Pho equivalent	83
2.3.2.2	Paused RNAPII as a signal for VIN3-PRC2 recruitment?	85
2.3.2.3	Open Chromatin mediates PcG association	87
2.3.3	<i>VEL proteins working upstream of PRC2</i>	87
2.4	SUMMARY	88
3	A NOVEL POLYMERIZATION DOMAIN INVOLVED IN THE REPRESSION OF FLC	90
3.1	INTRODUCTION	91
3.2	RESULTS	92
3.2.1	<i>The VEL domain is a four-helical domain distinct from other known polymerization domains</i>	92
3.2.1.1	Screens revealed crucial residues for VEL-VEL interaction	96
3.2.1.2	The VEL domain a novel polymerization domain	98
3.2.2	<i>The VEL domain is required for VIN3-VRN5 interaction</i>	101
3.2.2.1	The FNIII domain is not required for VEL-VEL interaction	103
3.2.2.2	The linker region in VIN3 is required for nuclear localization	106
3.2.3	<i>VIN3 head mutation R556D in Arabidopsis</i>	107
3.2.4	<i>VIN3 forms distinct nuclear foci</i>	109
3.3	DISCUSSION	110

3.3.1	<i>Multimerization of PRC2 through domain swapping – A role for the FNIII domain?</i>	113
3.3.2	<i>Avidity through polymerization</i>	116
3.4	SUMMARY	117
4	CHROMATIN CONFORMATION AT FLC	119
4.1	INTRODUCTION	119
4.1.1	<i>Chromatin loops involved in FLC regulation</i>	121
4.1.1.1	The active gene loop	121
4.1.2	<i>The repressive intragenic loop</i>	121
4.1.3	<i>Chromatin conformation capture</i>	123
4.2	RESULTS	124
4.2.1	<i>At least two chromatin loops form at FLC</i>	124
4.2.2	<i>Ectopic expression of COOLAIR disrupts gene loop formation</i>	128
4.2.3	<i>Role of COOLAIR promoter/3' FLC sequences in gene loop formation</i>	130
4.2.4	<i>Alternative loop formation through high antisense transcription</i>	132
4.3	INTERNAL GENE LOOP DYNAMICS	133
4.4	DISCUSSION	134
4.4.1	<i>RNAPII driving gene loop formation</i>	135
4.4.2	<i>The COOLAIR promoter/3' end of FLC is required for gene loop formation</i>	136
4.4.3	<i>Additional chromatin loops form at FLC</i>	137
4.4.4	<i>Internal loop associated with Polycomb repression?</i>	138
4.5	SUMMARY	139
4.6	OUTLOOK	140
5	TRANSGENIC LINES	142
5.1	INTRODUCTION	142
5.2	RESULTS AND DISCUSSION	143
5.2.1	<i>Characterization of the genetic backgrounds</i>	143
5.2.1.1	<i>vin3-1</i>	143
5.2.1.2	<i>vrn5-8</i>	145
5.2.2	<i>Design of constructs</i>	145
5.2.3	<i>Mutations to address the functional relevance of polymerization of the VEL domain</i>	147
5.2.4	<i>Mutational analysis of the in planta function of the composite domain</i>	152
5.2.5	<i>Mutational analysis of the in planta function of the FNIII domain</i>	156
5.2.6	<i>Analysis of interference of the fluorescence tag</i>	159

5.2.7	Overview of transgenic lines	160
5.3	SUMMARY	164
6	GENERAL DISCUSSION	167
6.1	THE VEL PROTEINS – PART OF A PLANT PHORC?	168
6.2	PROTEIN MEMORY	170
6.3	AVIDITY AS A GENERAL CONCEPT	172
6.4	THE VEL PROTEINS BEYOND FLOWERING TIME CONTROL	173
6.5	CONCLUSIONS	174
7	MATERIALS & METHODS	176
7.1	MATERIALS	176
7.1.1	<i>Plant material</i>	176
7.1.2	<i>Bacterial strains</i>	177
7.1.2.1	Escherichia coli	177
7.1.2.2	Agrobacterium tumefaciens	177
7.1.3	<i>Yeast strains</i>	178
7.1.4	<i>Plasmids</i>	178
7.1.5	<i>Antibodies</i>	179
7.1.6	<i>Restriction enzymes</i>	180
7.1.7	<i>Standard solutions</i>	180
7.2	METHODS	180
7.2.1	<i>Plant growth</i>	180
7.2.2	<i>RNA extraction</i>	181
7.2.3	<i>Dnase digestion</i>	181
7.2.4	<i>Complementary DNA (cDNA) synthesis</i>	182
7.2.5	<i>Quantitative PCR (qPCR) analysis</i>	182
7.2.6	<i>Cloning</i>	183
7.2.7	<i>Seamless megaprimer cloning</i>	183
7.2.8	<i>Generation of binary pSLJ clones by LR reaction</i>	184
7.2.9	<i>Triparental mating</i>	184
7.2.10	<i>Glycerol stocks</i>	184
7.2.11	<i>Floral dip of Arabidopsis</i>	185
7.2.12	<i>Selection of transformed Arabidopsis</i>	185
7.2.13	<i>Copy number analysis</i>	185
7.2.14	<i>Yeast transformation</i>	186
7.2.15	<i>Yeast-2-hybrid</i>	186
7.2.16	<i>Isolation of plasmid from yeast</i>	187
7.2.17	<i>Transformation of bacteria</i>	187

7.2.17.1	Heat-shock transformation of <i>E. coli</i>	187
7.2.17.2	Electroporation of <i>Agrobacterium</i>	187
7.2.18	<i>Confocal microscopy</i>	188
7.2.19	<i>Protein expression in E. coli</i>	188
7.2.20	<i>Protein purification</i>	188
7.2.20.1	Purification of His-tagged proteins	188
7.2.20.2	Purification of GST-tagged proteins	189
7.2.21	<i>Gel filtration</i>	189
7.2.22	<i>Electrophoretic mobility shift assay (EMSA)</i>	189
7.2.23	<i>RNA synthesis</i>	190
7.2.24	<i>Crosslinked nuclear immunoprecipitation mass spectrometry</i>	190
7.2.25	<i>Preparation of gel slices for Trypsin digestion.</i>	191
7.2.26	<i>Chromatin Conformation Capture (3C)</i>	192
7.2.26.1	3C quantification	193
7.2.27	<i>Protein chromatin immunoprecipitation</i>	194
7.2.27.1	ChIP-Seq	195
7.2.28	<i>Transient expression in N. benthamiana</i>	195
7.2.29	<i>Light microscopy</i>	196
7.2.30	<i>Co-IP</i>	196
7.2.30.1	Immunoblot	196
7.2.31	<i>Preparation of competent cells</i>	196
7.2.31.1	Chemically competent <i>E. coli</i> cells	196
7.2.31.2	Electrocompetent <i>Agrobacterium</i> cells	197
8	REFERENCES	198

Abbreviations

3C	Chromatin conformation capture
3D	Three dimensional
Arabidopsis	<i>Arabidopsis thaliana</i>
ASAP	Apoptosis- and splicing-associated protein
bp	base pair(s)
CD	Composite domain
CGI	CpG island
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation sequencing
Co-IP	Co-immunoprecipitation
Col-0	Columbia 0
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
FLC	Flowering Locus C
FNIII	Fibronectin type III
FRI	FRIGIDA
H 1/2A/2B/3/4	Histone 1/2A/2B/3/4
H2Ub	Ubiquitinated Histone 2
H3K4	Histone 3 lysine 4
H3K9	Histone 3 lysine 9
H3K9me2	Dimethylated H3K9
H3K27	Histone 3 lysine 27
H3K27 me2/me3	Di-/tri-methylated H3K27
H3K36	Histone 3 lysine 36
H3K36 me2/me3	Di-/tri-methylated H3K36
H4R3	Histone 4 arginine 3
HDAC	Histone deacetylase
HEK	Human embryonic kidney
Hi-C	High-throughput 3C
IP-MS	Immunoprecipitation mass spectrometry
lncRNA	Long non-coding RNA

mRNA	Messenger RNA
ncRNA	Non-coding RNA
PcG	Polycomb-group
PHD	Plant homeodomain
PhoRC	Pho repressive complex
PR-Dub	Polycomb repressive deubiquitinase
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PRE	Polycomb response element
RIF	Relative interaction frequencies
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
SF²	San Feliu-2
SAM	Sterile alpha motif
SEC	Size-exclusion chromatography
SEC-MALS	SEC- multiangle light scattering
TAD	Topological associating domain
TSS	Transcription start site
TTS	Transcriptional termination site
TU	Transcriptional unit
WH	Winged-helix
Y2H	Yeast-2-hybrid
ZnF	Zinc finger

List of figures

Figure 1-1 Different mechanisms of cellular memory	20
Figure 1-2 Histones as information carriers	23
Figure 1-3 Expression and chromatin dynamics during vernalization	26
Figure 1-4 Polycomb Repressive Complex 2 (PRC2) core proteins	31
Figure 1-5 The Arabidopsis VEL family	39
Figure 1-6 Model of PhoRC acting as a platform for PRC1/PRC2 recruitment through head-to-tail polymerization	44
Figure 1-7 Two models for the function of chromatin loops	51
Figure 2-1 Bioinformatically identified PRE cis motifs at Arabidopsis <i>FLC</i>	59
Figure 2-2 Secondary structure prediction of the N-terminal region of VIN3	62
Figure 2-3 Exotic VIN3 CDs	64
Figure 2-4 The composite domain of VIN3 mainly presents a positive surface	65
Figure 2-5 AtVIN3 CD associates with dsDNA	67
Figure 2-6 AtVIN3 CD binding to RNA	69
Figure 2-7 AtVIN3 CD binding to nucleic acid	70
Figure 2-8 The WH/Bromo-like domain is required for nucleic acid binding	71
Figure 2-9 VIN3-GFP protein-ChIP in <i>vall-2 FRI</i> background	78
Figure 2-10 Hypothetical model for recruitment of VIN3, and later PRC2, to the nucleation region.	89
Figure 3-1 Alignment of the VEL domain in the Arabidopsis VEL proteins	93
Figure 3-2 Gel filtration of the VEL domain	94
Figure 3-3 Yeast-2-hybrid screen for interaction mutants	97
Figure 3-4 SEC-MALS of VIN3 VEL domain mutants	98
Figure 3-5 VEL domain structure	99

Figure 3-6 The VEL domain is required for interaction between VIN3 and VRN5 in <i>N. benthamiana</i>	102
Figure 3-7 Point mutations do not block interaction of VIN3 and VRN5	103
Figure 3-8 VIN3 FNIII domain is not required for VIN3-VRN5 co- immunoprecipitation	104
Figure 3-9 FNIII mutations do not affect co-immunoprecipitation of VIN3 with VRN5	105
Fig. 3-10 Subcellular localization of WT and mutant GFP-VIN3	106
Figure 3-11 FLC vernalization responds in VIN3_R556D	108
Figure 3-12 VIN3-GFP forms small foci in root nuclei	110
Figure 3-13 Speculative model for the function of VEL polymerization	113
Figure 3-14 PRC2 dimerization through domain swapping	115
Figure 4-1 Chromatin loop dynamics at <i>FLC</i>	122
Figure 4-2 Principle of Chromatin Conformation Capture (3C)	124
Figure 4-3 Chromatin loop at <i>FLC</i> analysed by TaqIa	125
Figure 4-4 Chromatin loop at <i>FLC</i> analysed with NlaIII	126
Figure 4-5 Additional interaction regions detected by NlaIII-3C	127
Figure 4-6 Gene loop analysis in ColFRI and <i>ntl8-D3 FRI</i>	129
Figure 4-7 Schematic illustration of the 3' end of chimeric <i>FLC</i> constructs	130
Figure 4-8 Gene loop in <i>TEX2.0</i> and <i>FLCΔCOOLAIR</i>	131
Figure 4-9 Investigation of alternative loops in <i>ntl8-D3</i>	132
Figure 4-10 Chromatin loops at <i>FLC</i> assayed with DpnII	134
Figure 5-1 Characterization of <i>vin3-1</i>	144
Figure 5-2 Genomic <i>VIN3</i> and <i>VRN5</i> structure	146
Figure 5-3 Alignment of the VEL domain	149
Figure 5-4 Conserved positive patches in the CD in the VEL proteins	153
Figure 5-5 FNIII putative hinge region for domain swapping	157

**Figure 5-6 Subcellular localization of VIN3- Δ FNII-eGFP in stable
Arabidopsis transgenics**

159

List of tables

Table 1-1 Potential PRC2 subcomplexes in Arabidopsis	41
Table 2-1 VIN3 and VNR5 robustly associated with VEL-PRC2	74
Table 2-2 VIN3 and VRN5 associated with TOPLESS and members of the TOPLESS interactome	75
Table 2-3 VIN3 associates with splicing factors	76
Table 2-4 VIN3 associated with PRC1 and PR-DUB proteins	77
Table 5-1 Generation of VIN3 VEL domain <i>in planta</i> mutations	148
Table 5-2 Overview of stable transgenic lines generated to address the <i>in vivo</i> function of the composite domain	155
Table 5-3 Overview of constructs relevant for addressing the role of the FNIII domain	158
Table 5-4 VIN3 stable transgenic lines	161
Table 5-5 VRN5 stable transgenic lines	162
Table 5-6 Domain boundaries used for the stable transgenic lines	163
Table 5-7 Nuclear localization of VIN3 and VRN5 in <i>N. benthamiana</i> and <i>A. thaliana</i>	164
Table 7-1 Arabidopsis lines used in this thesis	176
Table 7-2 List of used antibodies	179
Table 7-3 List of enzymes used for 3C	180

Chapter 1

1 General Introduction

In this chapter, I will introduce the central concepts of chromatin and Polycomb-mediated gene regulation in plants and other eukaryotes that are relevant to the main research topic of this thesis. The first part is a general introduction to transcriptional memory and the role of chromatin. The second part continues with an example where these mechanisms have been studied in detail. This is the topic of my thesis work; the epigenetic switching of the *Arabidopsis* developmental regulator *FLOWERING LOCUS C*. The introduction ends with a more detailed discussion of Polycomb mediated gene regulation, highlighting the similarities between Polycomb systems in animals and plants. These have become even more obvious with the work conducted in this thesis.

1.1 Cellular memory

Multicellular organisms like animals and plants are made up of many different cell types that have various functions. All cells in an organism contain the same genetic information. The different behaviours of cells are therefore not due to different genetic material but rather differences in which genes are expressed and which are not. Distinct gene expression profiles in cells can be established as they respond to external stimuli, for example signals from other cells or physical factors. As cells differentiate during development, they acquire specific cell identities that become fixed and are remembered through cell division, even after the initial signal is no longer present (Bonasio et al., 2010). This constitutes a clear example of cellular memory. Another classic example is the memory of exposure to environmental cues. For example, exposure to external stresses that are inhibitory for normal growth in plants can be remembered, enabling a stronger and faster response the next time the plants experience the same stress (Ramirez-Prado et al., 2018). Since all cells contain the same genetic information, the memory cannot be held in the DNA sequence; instead, it is held at the epigenetic level. Epi, meaning above, describes the heritable differences in gene output that cannot be described by modification in the DNA sequence, as defined by Arthur Riggs and others (Clarke, 2004).

The transcriptional output of genes is regulated by transcription factors, regulators that affect, for example, the amount of engaging RNA Polymerase II (RNAPII), the release of RNAPII into active elongation, and ultimately the amount of mRNA that is produced from a specific gene. The mRNA can give rise to a protein that is able to promote its own expression, thereby establishing a feedback loop where the gene can remain expressed even after the initial stimuli is no longer present (Fig. 1-1). This kind of memory is referred to as trans-based memory because the memory is held in the trans-factors (Berry et al., 2015; Bonasio et al., 2010). While diffusible transcriptional factors are the classic example of trans-memory, recently a second kind of trans-memory has been described. Instead of

regulating the transcription process, the memory can be held in proteins that affect the behaviour of the produced protein. A typical example is the behaviour of prions. Prion proteins can adopt several conformations, one of which is a self-templating conformation that causes all new proteins to switch into that conformation, again establish a self-sustaining feedback loop that can be maintained in the absence of the initial signal (Fig. 1-1). Initially, prions were linked to diseases; however recently, protein-based epigenetics has gained increasing attention as an additional way to establish epigenetic inheritance (Harvey et al., 2018).

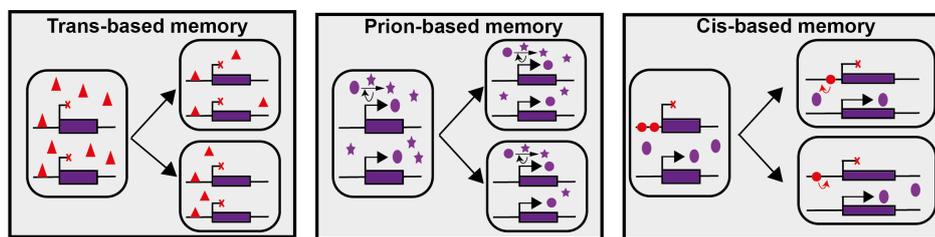


Figure 1-1 Different mechanisms of cellular memory. Trans-based memory with trans factor (triangles) as the information carrier. Prion-based memory with prion conformation (stars) as information carrier. Active genes give rise to proteins in the native conformation (purple circle), but the presence of a prion protein drives the conversion into the prion conformation (stars). Cis-based memory, where information is held locally at the locus, enabling the two alleles to be held in different expression states.

A characteristic of trans-based epigenetic memory is that it will affect the outcome of both alleles in a diploid organism. However, in some cases, alleles are maintained in opposite expression states. The most classic example of this is the inactivation of one of the X chromosomes in female mammals. Despite being in the same nucleus, one X chromosome is held in an active state while the other is in an inactive state. Therefore, the actual memory must be held at the individual chromosome or individual allele. Since the memory is held at the chromosome or gene, this form is referred to as cis memory (Fig. 1-1) (Berry et al., 2015; Bonasio et al., 2010). A key feature of cis epigenetic memory is that the two alleles in a diploid cell can be maintained in different states, as is the case for most genes on the X chromosome (Augui et al., 2011). While the actual establishing of the active or repressed state still relies on trans factor-like proteins, specific features in cis determine whether they bind or not. For example, during X

chromosome inactivation, the long non-coding RNA (lncRNA) XIST is expressed from, and only from, the X chromosome that is to be inactivated. XIST then recruits proteins that facilitate the repression (Brockdorff et al., 2020).

1.1.1 Chromatin – central for cis memory

The storage of memory in cis means that the memory element must be maintained at the specific allele. As described for X inactivation, this can be in the form of non-coding RNA (ncRNA) and localized proteins. RNA and proteins, together with DNA, make up the chromatin. The main protein component of the chromatin is nucleosomes, containing the core histone proteins H2A, H2B, H3 and H4 (Luger et al., 1997). The histone proteins make up an octamer with two H2A-H2B dimers and a H3-H4 tetramer. Approximately 147 base pairs (bp) of DNA wrap around the histone octamer to form the nucleosome, the core particle that constitutes the chromatin (Davey et al., 2002). Nucleosomes are formed every ~200 bp across the entire chromosome. Nucleosomes are connected with a stretch of histone-free “linker-DNA” often 20-80 bp long and the non-core histone H1; together they make up the overall conformation of the chromatin fibre (Perišić et al., 2010). Histones are highly basic proteins that allow the negative-charged DNA to closely associate. The packaging of DNA into nucleosomes makes the DNA less accessible for regulatory proteins, while the DNA outside of the nucleosome is more accessible. Therefore, the position of the nucleosome can be used to regulate access to the underlying DNA sequence. Histones contain a highly basic amino acid tail that extends from the core and can connect different nucleosomes (Luger et al., 1997). The linking of adjacent nucleosomes and the accessibility to other proteins make the histone tail an excellent hub for the regulation of chromatin. Histones and their tails can be modified by so-called writer proteins, which deposit chemical modifications like acetyl and methyl groups (Strahl and Allis, 2000). In addition, the modifications also serve as information carriers that can be read by reader proteins, in order to alter the chemical properties of the chromatin and thereby the

accessibility of the underlying DNA. Reader proteins often contain highly specific protein domains that recognize specific tail modifications (Musselman et al., 2012). In addition to the chemical modifications of histones and their tails, the 3D conformation of the chromatin is also involved in the regulation of gene expression. In metazoans, for example, chromatin comes together in the nucleus, despite being located along the chromatin fibre, to form Topologically associating domains (TADs) (Szabo et al., 2019).

1.1.1.1 Chromatin – Active or repressive

The different histone modifications play an important role in the transcriptional output from the underlying DNA sequence (Patel and Wang, 2013) and these characteristics separate chromatin into different types. The two main types of chromatin are hetero- and euchromatin. Heterochromatin constitutes the repressed part of the genome. It is formed in regions with repetitive sequences and with few genes, for example around the centromeres and telomeres. Heterochromatin can further be separated into constitutive heterochromatin, which contains stably repressed regions of the genome and is associated with H3K9 methylation. In contrast, facultative heterochromatin is often associated with H3K27 methylation and maintains the ability to convert to euchromatin (Trojer and Reinberg, 2007). Euchromatin contains most of the genes and is considered to be the active part of the genome, where the genes are actively expressed. Euchromatin is generally associated with H3K4 and H3K36 methylation as well as histone acetylation.

To fulfil the definition of ‘epigenetic’, the information must be heritable through cell division, meaning that the chromatin modifications need to be able to be restored as the DNA is copied and divided to the daughter cells. DNA is replicated semi-conservatively, with each daughter cell receiving one strand from the mother cell. This makes it easy to understand how methyl groups that are added to the DNA base cytosine can function as a carrier of epigenetic information, as each new daughter cell would contain

one strand with the information. However, the nucleosomes are temporarily disassembled as the replisome moves along the chromatin fibre, and re-deposited behind the replication machinery. New histones then fill in the remaining gaps on the new DNA strand. If the histone modifications are maintained, they can recruit read-write proteins that can re-establish the correct chromatin information (Fig. 1-2). For histones alone to carry the information correctly, this process must be tightly controlled by distributing old histones roughly 50:50 between the daughter cells and placing them correctly. For a long time it was assumed that this happened, but there was little experimental evidence. However, because of the central role of this process in epigenetic memory, a lot of effort has recently been put into improving our understanding of how a cell faithfully inherits histone modification information (Stewart-Morgan et al., 2020). The ability to accurately place histones behind the replication is called positional memory and it is key to maintaining the link between histone modifications and the underlying sequence. This is achieved through the histone chaperones, which are tightly linked to the replication machinery (Schlissel and Rine, 2019).

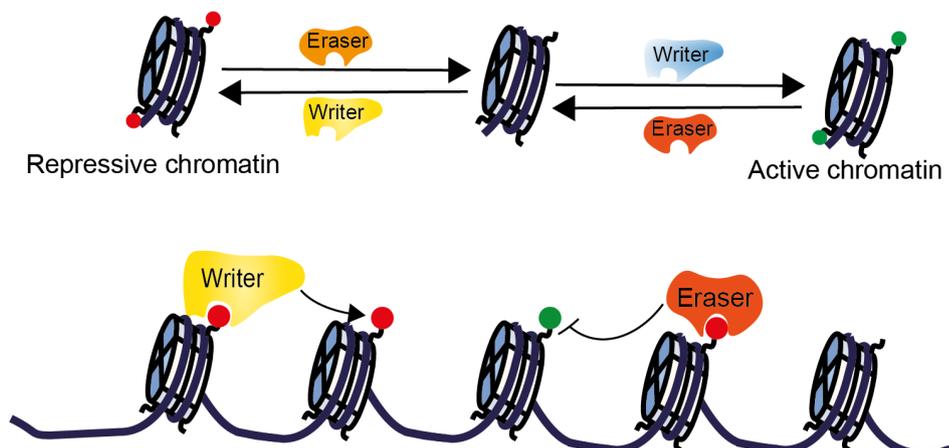


Figure 1-2 Histones as information carriers. Histones are modified on their NH₂-tails by specific writing proteins that can either add repressive (red circles) or activating (green circles) modifications. Histone modifications establish a feedback loop where the modifications stimulate the specific writer to add more of the same modification. Similarly, eraser proteins can recognize a mark and remove the opposing chromatin mark.

In summary, modifications of the histones serve as excellent carriers of information to maintain different expression profiles between different

cells in a multicellular organism. These modifications need to be faithfully inherited through DNA replication. How these modifications are delivered and inherited is central to the work done in this thesis. These questions require excellent model systems to unpick the complexity. In the next section I will describe one such model system that has enhanced our mechanistic understanding of epigenetic memory.

1.2 *FLC* – a model for epigenetic memory

A key developmental change in plants is the switch from vegetative to reproductive growth. A key regulator of this switch in the model plant *Arabidopsis thaliana*, hereafter referred to as *Arabidopsis*, is the developmental regulator *FLOWERING LOCUS C (FLC)*. *FLC* is a MADS box transcription factor that represses the expression of a set of key genes required for floral transition (Whittaker and Dean, 2017). *Arabidopsis* has been used extensively as a reference plant for decades, partly because of the short generation time of the common lab accession. The common lab accession of *Arabidopsis* is a rapid-cycling accession, in which flowering time is mainly regulated by the photoperiod and the autonomous pathway (He and Amasino, 2005). In contrast, most natural *Arabidopsis* accessions require exposure to prolonged cold for flowering. The acceleration of flowering by exposure to prolonged cold is called vernalization and it is central to flowering in many plant species, including some of our most important crops (Michaels et al., 2005). The need for vernalization ensures that plants which germinate in the summer and autumn do not flower during winter but instead over-winter vegetatively before flowering in the following spring. The adaptation of specific *Arabidopsis* accessions to different climates has been associated with changes in the length of cold required for the full promotion of flowering (Shindo et al., 2006).

1.2.1 Vernalization

Accessions of *Arabidopsis* that grow in the northern hemisphere generally adopt a winter annual reproductive strategy, requiring long winters to flower, and thus aligning flowering with favourable conditions in the spring (Shindo et al., 2006). This adaptation to the environment provides an excellent system for dissecting the molecular mechanisms of gene expression switching. Winter annual accessions of *Arabidopsis* are very late flowering compared to rapid cycling accessions grown in inductive conditions. This late flowering (and vernalization requirement) is due to high levels of *FLC*, a consequence of the accession having an active *FRIGIDA (FRI)* allele (Michaels and Amasino, 1999). *FRI* is a coiled-coil protein that through interaction with chromatin modifiers and RNA maturation factors establishes a high transcriptional state at *FLC* during early development by antagonising early *FLC* transcriptional termination (Choi et al., 2011; Geraldo et al., 2009; Schon et al., 2021).

Rapid cycling accessions, including the standard lab accession Col-0, carry mutations of *FRI*. This suggests that the loss of *FRI* has been a common strategy through evolution for acquiring rapid-cycling behaviour and enabling multiple generations per year (Johanson et al., 2000). During the expansion of the *Arabidopsis* international community, most genetic tools, including the collection of loss-of-function mutants, were established in the Col-0 background. However, to study the effect of mutants on vernalization, it is necessary to have high *FLC* starting levels. In order to achieve this, the near isogenic line ColFRI^{SF2} was established by introgression of an active *FRI* gene from the San Feliu-2 accession (Michaels and Amasino, 1999). A detailed understanding of *FLC* regulation during the life cycle of *Arabidopsis* has provided concepts relevant to floral transition in plants, generally helping the development of crops that maintain their yield in a changing climate.

1.2.2 Molecular events during *FLC* repression

Winter annual *Arabidopsis* accessions contain an active *FRI* allele that stimulates a high *FLC* expression state. Very early during embryo development, *FRI* antagonizes the activity of the repressor *FCA* (Schon et al., 2021). In the absence of *FRI*, *FCA* promotes the proximal termination of the *FLC* transcription, resulting in a repressive chromatin state that mediates a low transcription state of *FLC* (Schon et al., 2021). In contrast, when *FRI* is present, it antagonizes this activity of *FCA* and sets a high expression state of *FLC*, which leads to the requirement of vernalization for flowering (Schon et al., 2021).

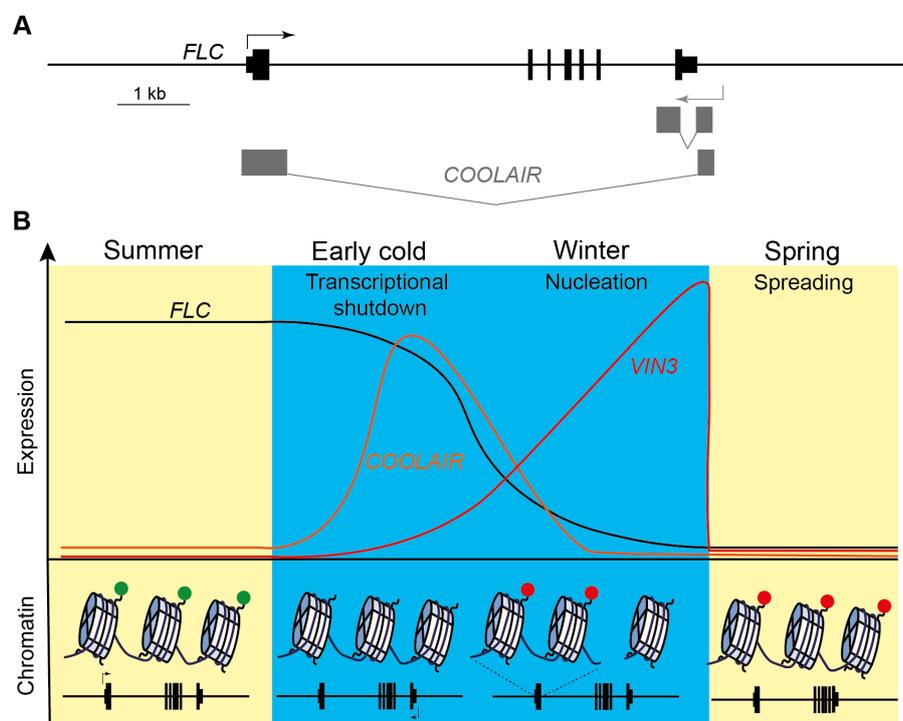


Figure 1-3 Expression and chromatin dynamics during vernalization. A) Schematic of the *FLC* locus with black boxes representing exons. Beneath in grey are drawn the antisense transcripts *COOLAIR*. **B)** *FLC* is expressed in the autumn and the chromatin is covered with active chromatin marks (green). *FLC* is transcriptionally repressed in early cold, coinciding with upregulation of the antisense *COOLAIR* and removal of the active marks. The *VEL* gene *VIN3* is quantitatively upregulated during the cold of winter and is required for establishment of the repressed state at *FLC* where the region around exon 1 becomes enriched in repressive chromatin marks (red). *FLC* is repressed in the following spring through the spreading of repressive marks over the entire locus.

Detailed analysis of vernalization-induced silencing of *FLC* has revealed that the process can be separated into two distinct phases: a slow phase during autumn and early winter and a faster phase when the temperature is

lower during winter (Hepworth et al., 2020) (Fig. 1-3). The initial phase leads to transcriptional shutdown of *FLC* expression, but not epigenetically stable repression. Transcriptional shutdown of *FLC* involves antisense transcripts, collectively named *COOLAIR*, which are upregulated by cold and whose expression is mutually exclusive with nascent sense *FLC* expression (Rosa et al., 2016; Swiezewski et al., 2009). Cold-upregulated antisense transcription leads to a reduction in the levels of the active chromatin mark H3K36me3 over the site of transcription initiation and the early transcriptional elongation zone of *FLC*, correlating with reduced expression of *FLC* sense RNA (Csorba et al., 2014; Yang et al., 2014). In addition, increased *COOLAIR* expression in early cold correlates with altered chromatin conformation of the *FLC* locus. Before cold, the *FLC* locus forms a loop conformation that connects the transcription start site (TSS) and the transcription termination site (TTS) (Crevillén et al., 2013), similar to many transcription units (TU) in yeast (O'Sullivan et al., 2004). The exposure to cold and increase in *COOLAIR* expression leads to disruption of the loop conformation, potentially helping to mediate the transcriptional shutdown of *FLC* (Crevillén et al., 2013). Induction of *COOLAIR* happens individually at each allele, showing that the induction is, at least partially, controlled in cis and not solely by trans-factors (Rosa et al., 2016).

The second, and faster, phase of *FLC* repression is linked to a group of proteins that is required for the epigenetic silencing of *FLC* (Hepworth et al., 2020). One of these proteins is the plant homeodomain (PHD) protein VERNALIZATION-INSENSITIVE 3 (VIN3) (Sung and Amasino, 2004). *VIN3* is specifically expressed in the absence of warm, which ensures that the epigenetic silencing of *FLC* only happens when the plants experience winter and not just a particularly cold week in autumn (Hepworth et al., 2018; Sung and Amasino, 2004). During prolonged cold, the *FLC* chromatin becomes enriched in the repressive chromatin marks H3K9me2 and H3K27me2/me3 (Bastow et al., 2004; Schubert et al., 2006). H3K27me3 initially accumulates at a small region around the first exon-

intron boundary, before spreading over the entire locus as the plants resume faster growth in the warmer condition of the spring (De Lucia et al., 2008; Schubert et al., 2006; Yang et al., 2017). The small region at the 5' end of *FLC* that initially accumulates H3K27me3 during the exposure to cold is known as the “nucleation region” (Angel et al., 2011). As plants come out of cold, the repressive chromatin mark H3K27me3 spreads over the entire *FLC* locus in a process that is connected to DNA replication (Hyun et al., 2013; Jiang and Berger, 2017; Yang et al., 2017). The spreading of H3K27me3 is limited to the *FLC* TU, despite neighbouring genes also having reduced expression in the cold (Finnegan et al., 2004; Schubert et al., 2006). This suggests that transcription or physical constraints prevent the spreading of H3K27me3 into neighbouring genes. To ensure that the requirement for vernalization is maintained in the next generation, the repression of *FLC* is reset during early embryogenesis through a process that requires demethylases, which remove the H3K27me3 mark (Crevillén et al., 2014; Sheldon et al., 2008).

1.3 Polycomb repression

To elucidate the mechanism behind vernalization, a forward screen was performed to identify mutants that were impaired in the vernalization response (*vrn*) (Chandler et al., 1996). One of the *vrn* mutants, *vrn2*, was shown to be a homolog of the *Drosophila* protein Suppressor of zeste 12 (Suz12) (Gendall et al., 2001). Suz12 is a core component of the highly conserved Polycomb Repressive complex 2 (PRC2), which methylates H3K27 and is essential for correct development in *Drosophila* (Müller et al., 2002; Tie et al., 2001). *vrn2* mutants showed normal transcriptional shutdown of *FLC* during vernalization but *FLC* expression reactivated after the plants returned to warm conditions (Gendall et al., 2001). These observations revealed that the memory of cold exposure at *FLC* is linked to the activity of Polycomb proteins and changes in the chromatin environment in a similar way to the memory of gene expression during development in animals (Gendall et al., 2001; Tie et al., 2001).

Polycomb group proteins (PcG) are central to maintaining the correct expression of genes after they have initially been repressed during development (Margueron and Reinberg, 2011). Initial work in *Drosophila* identified many of the core principles and core players in Polycomb repression. During development, HOX genes are tightly regulated in order to ensure the correct expression patterns and normal development (Beuchle et al., 2001). The silencing of the HOX genes occurs in two phases: gap proteins lead to transcriptional shutdown, and stable epigenetic repression then occurs after the levels of Gap proteins decline (Beuchle et al., 2001; Yu et al., 2019). The pattern of HOX gene repression is therefore like the pattern of *FLC* repression during vernalization, reflecting the evolutionary conservation in epigenetic memory mechanisms.

Initial work on *Drosophila* identified three Polycomb complexes, Polycomb Repressive Complex 1 (PRC1), Polycomb Repressive Complex 2 (PRC2) and Pho Repressive Complex (PhoRC) (Geisler and Paro, 2015). Later a fourth complex, Polycomb Repressive Deubiquitinase (PR-DUB) was characterized (Scheuermann et al., 2010). Three of the Polycomb complexes, PRC1, PRC2 and PR-DUB, have enzymatic activity and can catalyse the modification of histone tails. In contrast, the fourth complex PhoRC has no known enzymatic activity but is rather involved in the recruitment of the PcG proteins (Klymenko et al., 2006).

1.3.1 Polycomb Repressive Complex 2

The core of PRC2 consists of four proteins. In addition to Suz12, the core proteins in *Drosophila* are Enhancer of Zeste (Ez), Extra Sex Combs (ESC) and the Nucleosome remodelling factor 55 (NURF55/p55) (Joyce et al., 2000; Tie et al., 2001). The core PRC2 proteins are very well conserved between organisms but have undergone some duplication (Fig. 1-4). Vertebrates contain two homologs of Ez, enhancer of Zeste 1 and 2 (Ezh1/Ezh2), while *Drosophila* contains two homologs of ESC, Esc1 and Esc2 (Margueron and Reinberg, 2011). PRC2 has become more

evolutionarily complex in plants (Hennig and Derkacheva, 2009). For example, *Arabidopsis* contains three homologs of E(z), CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA); three homologs of Su(z)12, EMBRYONIC FLOWER 2 (EMF2), FERTILISATION INDEPENDENT SEED 2 (FIS2) and VRN2; and five homologs of p55, MSI1-5 (Hennig and Derkacheva, 2009), although only MSI1 is found in PRC2 complexes (Mozgova et al., 2015). The conserved role of PRC2 is to catalyse the methylation of H3K27, and through biochemical and more recent structural analysis the specific role of each core component has been elucidated. Su(z)12 serves as the scaffold protein that coordinates the assembly of the core PRC2 complex, mainly through the C-terminal VEFS (VRN2-EMF2-FIS2-Su(z)12) domain, but also through interactions with non-core PRC2 proteins to establish distinct PRC2 subcomplexes through the N-terminus (Chen et al., 2018; Kasinath et al., 2018). Ez is a methyltransferase with a SET-domain and it is the catalytic subunit of PRC2 that catalyses the addition of methyl groups to the tail of histone 3 at lysine 27 (H3K27). ESC binds H3K27me₂/me₃ and allosterically stimulates the methyltransferase activity of PRC2, a function that is essential for the spreading of H3K27me₃ (Margueron et al., 2009; Oksuz et al., 2018). Like ESC, Nurf55 is a WD-40 protein and binds histones, which likely stabilizes the interaction of PRC2 with the chromatin at target sites (Song et al., 2008).

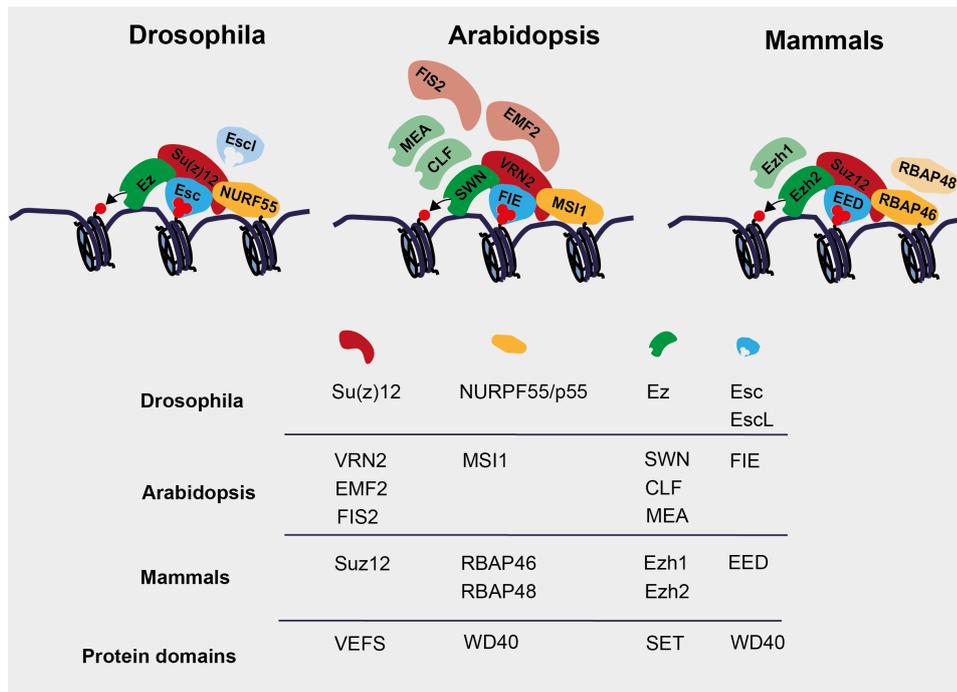


Figure 1-4 Polycomb Repressive Complex 2 (PRC2) core proteins. Overview of core PRC2 proteins in *Drosophila*, *Arabidopsis*, and mammals. PRC2 interacts with the chromatin through contacts to both the DNA and nucleosomes. PRC2 activity is allosterically activated through recognition of trimethylated (red circles) H3K27. PRC2 catalyses the methylation of H3K27. The table shows an overview of the core components in *Drosophila*, *Arabidopsis*, and mammals and the characteristic protein domain in each of the core proteins (adapted from Berry, 2015).

The role of PRC2 is to establish H3K27me₃ over target genes to maintain them in a transcriptionally repressed state. The inability to form H3K27me₃ domains leads to the derepression of target genes (García Reyes, 2013). However, exactly how H3K27me₃ and PRC2 cause transcriptional repression is not completely understood. Early models proposed that the occupancy of PRC2 and PRC1 leads to chromatin compaction, which limits the ability of transcription factors, chromatin remodellers and RNAPII to access the underlying DNA (Chopra et al., 2011; Di Croce and Helin, 2013). How PRC1 can function to achieve chromatin compaction will be described below. In addition to local chromatin compaction, Polycomb repressed loci come together in the nuclear space to form higher-order domains that are physically separated from the active chromatin (Sexton et al., 2012). In addition, to limit the ability of RNAPII to access chromatin, H3K27me₃ in the gene body inhibits the elongation of RNAPII (Aranda et al., 2015). Recent work in

different organisms has focussed on the role of H3K27me3 reader proteins and shown that these associate with deacetylases and RNAPII phosphatases, maintaining a repressed chromatin environment and/or preventing the release of RNAPII into the transcription elongation state (Fan et al., 2021; Wiles et al., 2020; Zhang et al., 2020b). These new mechanistic insights are consistent with earlier studies which showed that despite being repressed by PcG proteins, some genes are occupied by non-productive RNAPII and accumulate short transcripts (50–200 nucleotides long) (Brookes et al., 2012; Kanhere et al., 2010). Collectively, these observations show an interesting coupling between the progression of RNAPII transcription and Polycomb silencing.

Recent work has shown that the separation of nucleation and spreading, as identified at *FLC*, can also be observed in mammalian PcG targets (Oksuz et al., 2018; Yang et al., 2017). This highlights the conserved mechanisms in PRC2 function across organisms. Additional work has further shown that the initial establishment of H3K27me3, in other words, nucleation, depends on a number of proteins that associate with PRC2, in addition to the four core PRC2 proteins (Højfeldt et al., 2019; Laugesen et al., 2019).

1.3.1.1 Accessory proteins make PRC2 specific

Several proteins associate with the core PRC2 to form two distinct PRC2 complexes named PRC2.1 and PRC2.2 (Chen et al., 2018; Grijzenhout et al., 2016). PRC2.1 refers to the complex that, in addition to the core components, contains one of the Polycomb-like proteins (PCL1-3), also named PHF1, MTF2, or PHF19 in mammals (Hauri et al., 2016). Two other proteins, EPOP or the vertebrate-specific PALI, interact with PRC2.1 (Alekseyenko et al., 2014; Hauri et al., 2016). The second PRC2 complex, PRC2.2, contains the accessory proteins JARID2 and AEBP2 (Hauri et al., 2016). In general, the accessory proteins work to increase the activity of PRC2, either through allosteric activation or by increasing the association of the PRC2 complex with chromatin (Yu et al., 2019).

1.3.1.1.1 PRC2.1 – The PCL containing PRC2 complex

The PCL proteins, in particular PCL2 (MTF2), are important for the binding of PRC2 at target regions (Højfeldt et al., 2019; Li et al., 2017; Perino et al., 2018). All three PCL proteins contain a Tudor domain followed by two Plant homeodomains (PHDs), a Winged-Helix (WH) domain and a C-terminal conserved domain (van Mierlo et al., 2019). The Tudor domain recognizes the active histone mark H3K36me2/me3 (Ballaré et al., 2012), which otherwise antagonizes PRC2 activity (Schmitges et al., 2011), and it may play a role at genes that switch from an active transcription state to a PRC2 silenced state, by stabilising PRC2 at H3K36me2/me3 mark genes and facilitating H3K27me3 deposition. In addition, PCL3 (PHF19) directly links the demethylation of H3K36 with the acquisition of H3K27me3 by interacting with the demethylase NO66, as well as potentially mediating the recruitment of another demethylase KDM2B (Ballaré et al., 2012; Brien et al., 2012). While the PCL proteins share a high degree of sequence conservation, and at first glance can appear identical, small differences give them different characteristics. For example, a single amino acid difference in the Tudor domain of MTF2 compared to PHF1 and PHF19 results in significantly lower affinity for H3K36me3 (Gatchalian et al., 2015; Li et al., 2017). This highlights how small changes in the PRC2 accessory proteins can potentially change the functional output of PRC2. Like the Tudor domain, PHDs are best known for their ability to bind to chromatin, particularly H3K4 (Mellor, 2006). However, some of the PHDs in the PCL proteins are involved in protein-protein interaction with both core PRC2 components, as well as with other non-PRC2 proteins, particularly through the second PHD located next to the WH domain (Boulay et al., 2011; Brien et al., 2015; Liu et al., 2018; O’Connell et al., 2001). Recognition of methylated H4R3 has also been reported for PCL1 (Liu et al., 2018). Immediately downstream of the second PHD is a WH domain, which was previously known as the Extended Homology (EH) region, and like other WH domains, the WH domain of the PCL can bind DNA *in vitro* (Choi et al., 2017; Li et al., 2017; Perino et al., 2018). This is to some extent required for MTF2-

mediated binding of PRC2 at target sites (Li et al., 2017). The specificity of the DNA binding mediated by the WH is still debated. Choi et al. reported that the binding of PHF1 to DNA was not specific, while Li et al. reported preferential binding to unmethylated CpG sequences (Choi et al., 2017; Li et al., 2017). In addition, a study of the preferential binding of MTF2 to certain CpG sequences suggested that the helical shape of the DNA is important for recognition by the WH, at least for MTF2 (Perino et al., 2018).

While the functions of the PCL proteins are somewhat understood, the function of the other recently identified vertebrate PRC2.1 accessory proteins EPOP and PALI are less well understood. However, similar to the other PRC2 accessory proteins, they have been reported to stimulate the methyltransferase activity of PRC2 (Conway et al., 2018; Piunti and Shilatifard, 2021).

1.3.1.1.2 PRC2.2 – The JARID2/AEBP2 PRC2 complex

Similar to the role of the PCL in PRC2.1, it has been suggested that AEBP2 increases the residence time of PRC2 at chromatin, by binding to DNA and nucleosomes. DNA binding is mediated by a zinc-finger reported to specify PRC2.2 for recognition of methylated CpG islands (Kim et al., 2009; Wang et al., 2017). This is in contrast to the inhibitory role of CpG methylation in the binding of the PCL proteins. Furthermore, AEBP2 contains a basic motif that is rich in the basic amino acids lysines (K) and arginines (R). This motif, known as the KR motif, is involved in nucleosome binding, and in stimulating the activity of PRC2.2 (Lee et al., 2018). Mutations that disrupt the basic motif lead to less PRC2 binding and lower H3K27me3 methylation *in vivo* (Lee et al., 2018).

The other accessory factor in PRC2.2, JARID2, is reported to be involved in all the same mechanisms as the other accessory factors. Through its N-terminal region, JARID2 recognizes and binds H2Ub, a mark associated with the activity of PRC1 (Cooper et al., 2016). Both AEBP2 and JARID2

interact with ubiquitinated nucleosomes and stimulate the activity of PRC2, which helps to overcome the inhibitory effect of H3K4me3 and H3K36me3 on PRC2 (Kasinath et al., 2021). Through the middle domain, JARID2 contributes to nucleosome binding and through the C-terminal domain it binds DNA (Li et al., 2010). Similar to the theme for some of the other accessory proteins, the DNA binding of JARID2 does not show strong specificity, other than a potential bias for GC-containing sequences (Li et al., 2010). In addition to the common mechanisms of enhancing PRC2 activity, JARID2 also stimulates PRC2 through a specific mechanism that mimics the read-write model introduced above. PRC2 catalyses the di- and tri-methylation of K116 on JARID2, which, like H3K27me3, is bound by the aromatic cage in Esc^{EED} to allosterically activate PRC2 (Sanulli et al., 2015). Interestingly, a recent study showed that PALI, the vertebrate PRC2.1 accessory factor, likewise can be methylated by PRC2 to allosterically activate PRC2 through binding of the PALI K1241me2/3 to Esc^{EED} (Zhang et al., 2020a). This suggests that the PRC2 methylation of accessory proteins is a common mechanism in both PRC2 subcomplexes; acquisition of a methylated lysine residue allosterically activates PRC2 in the absence of methylated histone tails.

While one might expect that PRC2.1 and PRC2.2 would be specific for certain genomic targets, recent studies on mammalian cells have shown that PRC2.1 and PRC2.2 work synergistically and share most target sites. Consistently the accessory proteins are collectively required for genomic targeting of H3K27me3 (Healy et al., 2019; Højfeldt et al., 2019). From the domain architecture and reported functions of the accessory proteins, we see that a common theme seems to be that the accessory proteins provide a range of mechanisms for the association of PRC2 with chromatin, suggesting that PRC2 association with chromatin relies on combinatorial interactions.

1.3.1.2 Arabidopsis PRC2 proteins

While the core Polycomb proteins at first glance appear to be strongly conserved between different species, the plant PRC2 core proteins possess some interesting features. The conserved function of PRC2 is to mediate H3K27me₃ and transcriptional repression. The function of PRC2 is antagonized by active transcription and active histone marks (Berry et al., 2017a; Riising et al., 2014; Yuan et al., 2011). Biochemical studies have shown that the presence of K36me₃ and K4me₃ on the same histone tail inhibits the deposition of H3K27me₃ (Schmitges et al., 2011; Yuan et al., 2011). However, in contrast to other PRC2 complexes, including the Arabidopsis EMF2-PRC2 complex, the H3K27 methylation activity of a VRN2-PRC2 complex is not inhibited by active chromatin marks (Schmitges et al., 2011). This offers the possibility of a PRC2 complex being able to associate with loci marked by active chromatin marks to trigger repression through H3K27 methylation. This observation might also help to explain why the removal of H3K36me₃ and the gain of H3K27me₃ during vernalization can be uncoupled, as observed in lines with low *FLC* antisense transcription (Csorba et al., 2014).

1.3.1.2.1 The vernalization PRC2 complex

As introduced above, in contrast to the situation in other organisms, the Arabidopsis genome contains several copies of each of the PRC2 core components, except for FIE^{Esc}. Through genetics, protein chromatin immunoprecipitation (ChIP) and immunoprecipitation followed by mass spectrometry (IP-MS), the core components essential for the repression of *FLC* during vernalization have been identified. The core components that make up the vernalization-relevant PRC2 complex are VRN2^{Su(z)12}, FIE^{Esc}, MSI1^{NURF55}, and CLF/SWN^{Ezh} (Gendall et al., 2001; Wood et al., 2006; Yang et al., 2017). An interesting observation from recent studies was the functional difference between the Ezh homologs SWN and CLF, as *SWN* is mainly involved in the establishment of H3K27me₃ at the nucleation region, while *CLF* is required for the stable epigenetic repression of *FLC*, which it does by mediating the spreading of H3K27me₃ over the locus

(Yang et al., 2017). The pattern observed at *FLC* is similar to the pattern observed at several other genomic sites, where H3K27me3 is maintained at a small genomic patch in a *clf* mutant (Shu et al., 2019). This suggests that the separation of nucleation and spreading observed at *FLC* are true for other Polycomb targets in Arabidopsis as well.

Consistent with *CLF*'s role in propagating and maintaining H3K27me3 in Arabidopsis, a *clf* mutant has a much stronger phenotype than a *swn* mutant. The *clf* mutant phenotype cannot simply be overcome by overexpression of *SWN*, showing the distinct functions of the Ezh homologs (Chanvivattana et al., 2004). While it has been established that *CLF* is required for the stable repression of *FLC* (Yang et al., 2017), the absolute need for *SWN* in the initial establishment of H3K27me3 at *FLC* has not yet been confirmed. However, the use of IP-MS suggests that the main PRC2 complex involved in the nucleation of PRC2 and H3K27me3 at *FLC* during vernalization contains SWN and not CLF (De Lucia et al., 2008).

To some extent, the functional difference between SWN and CLF is similar to the different roles of the mammalian Ezh homologs Ezh1 and Ezh2 (Margueron et al., 2008). One of the differences between Ezh1 and Ezh2 is methyltransferase activity (Margueron et al., 2008). This had led to the suggestion that Ezh2, which has high levels of methyltransferase, is required for the propagation and maintenance of H3K27me3 when cells are dividing, and H3K27me3 is disrupted by replication. Interestingly, the low catalytic activity of Ezh1-PRC2 can be complemented by non-core PRC2 components (Lee et al., 2018), highlighting the importance of not only the four core PRC2 components but also the proteins that associate with PRC2.

1.3.1.2.2 Arabidopsis accessory proteins

Through approaches similar to those used in the animal Polycomb field, several potential PRC2 accessory proteins have been identified in

Arabidopsis. Most of these proteins have been identified through genetic screens or by IP-MS. Therefore, confirmation is required of whether or not they associate directly with core PRC2 through physical contacts. Despite this lack of confirmation, due to their abundance in the IP-MS of PRC2 components and their often mutual exclusivity with other accessory factors, it seems reasonable to consider these proteins as plant PRC2 accessory proteins that give rise to distinct PRC2 complexes.

Most likely because of their important role in controlling flowering time, the first identified PRC2 accessory proteins were the homologs *VIN3* and *VERNALIZATION5 (VRN5)/VIN3-LIKE 1 (VIL1)* (Greb et al., 2007; Sung and Amasino, 2004; Sung et al., 2006). These were initially identified as having the same phenotype as the *vrn2* mutant and later shown to be co-immunoprecipitated with the core PRC2 components VRN2, FIE, SWN, and MSI1 and their homologue *VERNALIZATION5/VIN3-Like (VEL1)/VIN3-LIKE 2 (VIL2)* (De Lucia et al., 2008; Wood et al., 2006).

The VEL (*VEL1*, *VIN3* and *VRN5*) proteins robustly co-purify with the core PRC2 components CLF, SWN, and MSI1 (Bloomer et al., 2020; Derkacheva et al., 2013; Liang et al., 2015). The Arabidopsis genome contains five members of the VEL protein family (*VEL1-3/VIL2-4*, *VIN3* and *VRN5/VIL1*). With the exception of *VEL3*, they all share the same protein domain architecture with a PHD, a FNIII domain and a C-terminal domain. The latter shared no similarity with other known domains and was therefore named the VEL/VID domain (Mylne et al., 2004; Sung et al., 2006) (Fig. 1-5). Two groups identified this family within 1–2 years of each other, which led to the confusing naming of these proteins as VEL or VIL proteins. For clarity the proteins will hereafter only be referred to by their VEL name, which was also the naming used in the first publication describing this family (Mylne et al., 2004).

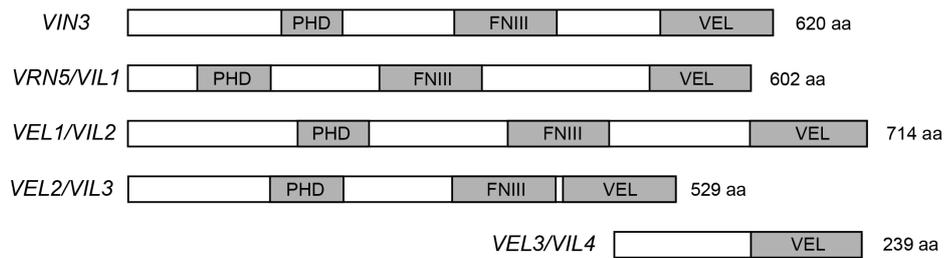


Figure 1-5 The Arabidopsis VEL family. Gene structure of the five Arabidopsis VEL homologs showing the three known domains, the Plant homeodomain (PHD), Fibronectin III (FNIII) domain, and VEL domain.

As introduced earlier, the *VIN3* RNA level increases as a response to prolonged cold (Fig. 1-3), while both *VRN5* and *VEL1* are relatively unchanged during vernalization. *VEL1* increases slightly, while *VEL2* increases during cold, but in contrast to *VIN3* remains relatively highly expressed after vernalization (Greb et al., 2007; Sung et al., 2006).

VEL3 appears to be specifically expressed in the seed (X. Chen unpublished). In contrast to *vrn5* and *vin3*, *vell* and *vel2* do not appear to affect the repression of *FLC* during vernalization, but are required for repression of some of the other members of the *FLC* gene family (*MAF2-MAF5*) (Kim and Sung, 2013).

Another putative PRC2 accessory factor with a role in the repression of *FLC* is *INCURVATA 11 (ICU11)* (Bloomer et al., 2020). Like *VIN3* and *VNR5*, *ICU11* is required for the acceleration of flowering by vernalization (Bloomer et al., 2020), and as expected for a PRC2 accessory protein, *icu11* showed lower levels of H3K27me3 compared to wild-type at some PRC2 targets (Bloomer et al., 2020). In contrast to the VEL proteins, *ICU11* was shown to co-purify with the other Arabidopsis Suz12 homolog EMF2 (Bloomer et al., 2020), and as expected, it is required for H3K27me3 at several other genes in Arabidopsis (Mateo-Bonmatí et al., 2018). Functionally similar to the association of the H3K36 demethylase N066 with the PRC2.1 complex in mammalian cells (Brien et al., 2012), *ICU11*'s main role appears to be the removal of H3K36me3 (Bloomer et al., 2020; Mateo-Bonmatí et al., 2018). This may potentially explain why *ICU11* is found in complex with EMF2, which as introduced above is

inhibited by the presence of H3K36me3, while VRN2 is not (Schmitges et al., 2011). Consistent with the existence of distinct PRC2 complexes in Arabidopsis, ICU11 does not co-purify with the VEL proteins but instead with the other putative accessory proteins – EMF1, LHP1 and the TELOMERE-REPEAT-BINDING 1-3 (Bloomer et al., 2020). At *FLC*, *LHP1* is required for the spreading of H3K27me3 after vernalization (Yang et al., 2017), and consistent with a function in association with a EMF2-PRC2 complex, *lhp1* affects several PRC2 targets similar to *icu11* (Berry et al., 2017b; Mateo-Bonmatí et al., 2018). Performing a similar role to that of animal PRC2 accessory proteins, TRBs are required for the deposition of H3K27me3 at many Arabidopsis Polycomb targets (Zhou et al., 2018).

At least one other distinct PRC2 complex has been identified in Arabidopsis, composed of the core components EMF2, MSI1, FIE and SWN/CLF) and the accessory factors *ALP1* and *ALP2* (*ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1/2*) (Liang et al., 2015; Velanis et al., 2020). In contrast to the other PRC2 accessory proteins, the ALPs are reported to antagonize the function of PRC2, potentially by competing with LHP1 and EMF1 for binding with MSI1 (Velanis et al., 2020). They thus demonstrate similarities with the mammalian protein EZH inhibitory protein (EZHIP)/CATACOMB, which likewise antagonizes the stimulatory effect of accessory factors like AEBP2 and JARID2, potentially through a displacement model (Piunti and Shilatifard, 2021; Ragazzini et al., 2019).

Combined, these studies suggest the existence of at least four (or eight by exchanging SWN and CLF) different PRC2 complexes (Table 1-1). While some potential similarities can be hypothesized between the plant and animal accessory proteins, the molecular function of the majority of the proteins is still not understood.

Table 1-1 Potential PRC2 subcomplexes in Arabidopsis

PRC2 core	Associating factors
VRN2-FIE-MSI1-SWN/CLF	VIN3, VRN5, VEL1
EMF2-FIE-MSI1-SWN/CLF	ICU11, EMF1, LHP1, TRB1-3
EMF2-FIE-MSI1-SWN/CLF	ALP2, ALP1
FIS2-FIE-MSI1-SWN/CLF	

1.3.2 Polycomb Repressive Complex 1

Another Polycomb complex that has been studied in detailed is Polycomb Repressive Complex 1 (PRC1). PRC1 generally occupies the same regions in the genome as PRC2 (Ku et al., 2008). It has E3 ligase activity and facilitates the ubiquitination of H2A (H2AUb) (Aranda et al., 2015). In addition to its enzymatic activity, PRC1 also causes chromatin compaction (Grau et al., 2011). Chromatin compaction by PRC1 happens locally at target genes and through higher-order clustering required for Polycomb mediated silencing, through stabilization of transcriptional repression (Grau et al., 2011; Isono et al., 2013). The *Drosophila* PRC1 complex consists of the core proteins Polycomb (Pc), Polyhomeotic (Ph), Posterior sex comb (Psc), and Sex combs extra (Sce) (Steffen and Ringrose, 2014). In contrast to PRC2, studies on mammalian cells have revealed the existence of many homologs of the *Drosophila* core PRC1 proteins, which give rise to distinct PRC1 complexes (Gao et al., 2012; Geng and Gao, 2020). The association with one of the six different mammalian Psc homologs (PCGF1-6) gives rise to PRC1.1-6. However, PRC1 is often referred to as either canonical PRC1 or variant PRC1, while PRC1.2/4 is referred to as the cPRC1, as its composition is similar to that of the original PRC1 complex identified in *Drosophila* (Gao et al., 2012; Geng and Gao, 2020).

While the existence of a plant PRC2 complex has been accepted for a long time, the existence of an authentic PRC1 complex in *Arabidopsis* has long been a matter of controversy. Recently, studies have shown the existence of a core PRC1 complex containing RING1A/B^{Sce} and BMI1A/B/C^{Psc}

(Calonje, 2014). In addition, LHP1^(Pc) and EMF1^(Psc) are sometimes referred to as being members of the PRC1 complex (Merini and Calonje, 2015). Similar to PRC1 in other organisms, PRC1 associates with a range of different proteins in Arabidopsis, including VAL1 (Calonje, 2014).

The work in this thesis concentrates on the function of the VEL proteins that are considered to be PRC2 accessory factors, as introduced above. Therefore, a detailed introduction of PRC1 and the role of cPRC1 and the vPRC1 complexes is not required to understand the work in this thesis. However, one interesting feature of the animal PRC1 subunit Ph and the PRC1 associating protein Sex-comb-midleg (Scm) is the presence of a sterile alpha motif (SAM). This is a polymerization domain enabling the formation of head-to-tail polymers (Kim et al., 2002). Through its polymerization ability, the SAM domain is involved in the formation of small but microscopically visible foci named PcG bodies (Isono et al., 2013). It is also required for the silencing of Polycomb genes, putatively by establishing a chromatin conformation that excludes RNAPII (Isono et al., 2013; Wani et al., 2016). In addition, the formation of condensates (PcG bodies) enhances the ubiquitination activity towards H2A (Seif et al., 2020). This enhanced activity is presumably achieved through the compartmentalisation of the SAM domain containing PRC1 subunits at the chromatin, which then recruits the enzymatically active components (Frey et al., 2016; Isono et al., 2013; Seif et al., 2020). The following section introduces the interesting concept of how PcG proteins, particularly PRC2, are recruited.

1.4 Recruitment of Polycomb

A common question in Polycomb biology is how PcG proteins find and bind their targets. In *Drosophila*, PcG proteins are highly enriched at specific cis-motifs that are required for the repression of Polycomb genes; these cis-motifs are named Polycomb response elements (PREs) (Müller and Kassis, 2006). The identification of vital cis motifs for Polycomb repression means that the recruitment of PcG proteins most likely also

involves specific DNA binding proteins. The PcG protein Pleiohomeotic (Pho) has been identified as a sequence-specific DNA binding protein important for Polycomb repression (Brown et al., 1998). The potential role of Pho as functioning at the base of the Polycomb recruitment system has stimulated a lot of interest in the link between PRE-Pho and PRC1/PRC2. This has recently led to the discovery of a third Polycomb complex, PhoRC (Frey et al., 2016; Klymenko et al., 2006), which plays a key role in the association of PRC1, and to some extent PRC2, with its genomic targets.

1.4.1 PhoRC

The three other Polycomb complexes (PRC1, PRC2, and PR-Dub) all have known enzymatic activity. In contrast, the fourth Polycomb complex PhoRC has no known activity. PhoRC was initially shown to consist of two proteins: the sequence-specific DNA binding protein Pho and the SAM domain containing protein dSfmc (Klymenko et al., 2006). dSfmc has the ability to transcriptionally silence its target gene but it relies on a DNA binding protein for its targeting to DNA (Klymenko et al., 2006). The SAM domain of dSfmc functions as a bridge to PRC1 by binding the SAM domain containing PRC1 protein Scm. Scm can further interact with the SAM domain containing PRC1 protein Ph (Frey et al., 2016). In addition to linking to PRC1, Scm has also been shown to co-purify with PRC2 in *Drosophila* (Kang et al., 2015). This had led to an attractive model where PhoRC is the first component that initiates Polycomb repression and through SAM-domain-mediated protein-protein interaction recruits both PRC1 and PRC2 (Frey et al., 2016) (Fig. 1-6). In line with this model is the fact that PhoRC precedes PRC2 at the two *giant* PREs, consistent with PhoRC being a nucleation factor in *Drosophila* (Abed et al., 2018).

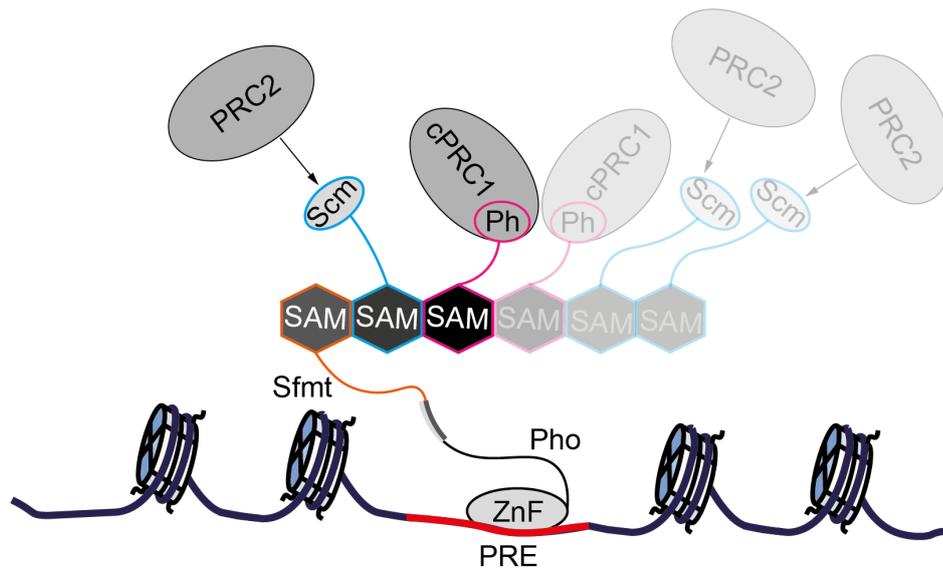


Figure 1-6 Model of PhoRC acting as a platform for PRC1/PRC2 recruitment through head-to-tail polymerization. PhoRC consists of the DNA binding protein Pho and the SAM domain protein Sfmt. Pho binds the DNA at the Polycomb response element (PRE) through its ZnF domain. The SAM domain of Sfmt interacts with the SAM domain of Scm that can associate with PRC2. The SAM domain of Scm can form homopolymers with other Scm proteins or heteropolymers with the SAM domain of the PRC1 protein Ph. The polymerization ability of the Scm-SAM and Ph-SAM enables the potential formation of longer polymers that could mediate the local concentration of PRC1 and/or PRC2.

1.4.2 Recruitment – a Rethink

The PhoRC offers a very simple explanation for the recruitment of Polycomb to their genomic targets, through direct links from a specific DNA sequence to PRC1 and PRC2, via direct protein-protein interactions. This is similar to the classic view of how transcription factors bind to DNA motifs to recruit other factors that modulate the transcriptional output. However, additional studies have shown that the recruitment of PcG complexes does not solely rely on sequence-specific binding through, for example, Pho; it also relies on combinational interactions (Kahn et al., 2014). The lack of direct sequence-specific targeting is even clearer in Polycomb targeting in mammalian cells. Sequence-specific factors, including the mammalian homolog of Pho, YY1, cannot account for most of the Polycomb binding *in vivo* (Klose et al., 2013). In mammalian cells, Polycomb binding occupancy overlaps with regions (1–2 kb) that are rich in CpG DNA, known as CpG islands (CGIs) (Ku et al., 2008). Consistent with CGI having an important role in the recruitment of PcG proteins in

mammalian cells, the integration into mammalian cells of a bacterial sequence similar to CGIs was found to be sufficient for the recruitment of PRC2 (Mendenhall et al., 2010). However, in order to recruit PRC2, the CGI must be deprived of transcription-activating motifs (Mendenhall et al., 2010).

In contrast to the recruitment mechanism of *Drosophila* PhoRC, no mammalian trans factors with high affinity for CGIs have been found. Instead, the binding is mediated directly by the PRC2 complex, in part through the activity of the accessory factors (Laugesen et al., 2019; Yu et al., 2019), as introduced earlier. The lack of strong affinity binders has led to a chromatin sampling model, where PRC2 constantly touches-and-leaves most of the chromatin and where stable binding is particularly prohibited by active transcription (Klose et al., 2013). Instead of high-affinity binders, H3K27me3 is mediated by many low-affinity binders that not only recognize CGIs but also DNA conformation, and the presence of PRC1 and H2Ub (Laugesen et al., 2019; Yu et al., 2019). Therefore, the word ‘recruitment’ in Polycomb biology, except for a few cases, refers to a mechanism by which PRC2’s residence time at the chromatin is increased through a range of multivalent interactions (Laugesen et al., 2019).

1.4.3 Recruitment of PRC2 in Arabidopsis

Similar to the work on animals, several studies have tried to find plant PREs. The *LEAFY COTYLEDON2* (*LEC2*) promoter was shown to contain a short 50 bp sequence that was able to cause repression and H3K27me3 accumulation when inserted into an unrelated promoter sequence, showing similarities with *Drosophila* PREs (Berger et al., 2011). Similarly, the ASSYMETRIC LEAVES (AS) complex mediates the recruitment of PRC2 through direct protein-protein interactions with the core PRC2 component CLF. It is recruited to the promoter of *KNOX* genes, reportedly through sequence-specific recognition (Lodha et al., 2013). Similar to the putative PRE from *LEC2*, the PRE from *KNOX* containing the AS binding

sites is enough to cause repression of an engineered reported construct (Lodha et al., 2013). AS1 and AS2 together can bind DNA, as examined by EMSA (Guo et al., 2008), and link a transcriptional repressor complex with PRC2 silencing.

In general, the association of PRC2 with its targets in the Arabidopsis genome appears to resemble the scenario in Drosophila, where PRC2 is recruited through trans factors with DNA binding domains that bind specific DNA sequences (Huo et al., 2016). However, recruitment of PRC2 in Arabidopsis can apparently be mediated by a range of different trans factors and cis sequences.

At *FLC*, the only sequence-specific DNA binding protein that has been identified to play a role in vernalization is VAL1 (Qüesta et al., 2016). In addition, it has also been shown that VRN1 (Levy et al., 2002), binds DNA. VRN1 is not a sequence-specific binding factor and it associates with all metaphase chromosomes in Arabidopsis (Mylne et al., 2006), preferentially to long poly-A motifs (King et al., 2013; O'Malley et al., 2016). How VRN1 would potentially connect to PRC2 is not known, but interestingly *vrn1* is genetically non-additive with *vin3*, suggesting that they are working through the same pathway to repress *FLC* (Greb et al., 2007). Similar to VRN1, VAL1 is a B3 domain containing protein, but in contrast to VRN1 it possesses sequence specificity and it specifically recognizes the RY-motif (TGCATG) (Qüesta et al., 2016; Yuan et al., 2016). Supporting the importance of VAL1 for the repression of *FLC*, earlier studies showed that a sequence containing the RY-motifs is required for an accurate vernalization response at *FLC* (Sheldon et al., 2002; Yuan et al., 2016). At *FLC*, a direct link from the RY-motif to PRC2 through VAL1 has not been established, but other studies have shown that VAL1 can interact with the PRC2 component SWN as well as with LHP1 (Yuan et al., 2020, 2016). Instead, it has been proposed that at *FLC* the connection of VAL1 to PRC2 is mediated by protein-protein interaction between VAL1 and the apoptosis- and splicing-associated protein (ASAP)

complex and the histone deacetylase HDA19, which in turn interact with VIN3 (Qüesta et al., 2016). As the name suggests, the ASAP complex is involved in RNA processing, but the exact role of RNA in PRC2-mediated silencing has been heavily debated.

1.4.4 RNA in Polycomb recruitment

Over the last couple of decades, the role of RNA in Polycomb association with chromatin has been heavily discussed and debated. Biochemical studies have shown that PRC2 binds RNA with relatively high affinity, although promiscuously (Davidovich et al., 2013). More recent work has reported that RNA is required for PRC2 chromatin occupancy (Long et al., 2020), suggesting that RNA directly or indirectly through RNA binding proteins is required for the strong association of PRC2 with its genomic target sites (Long et al., 2020). Similar to the case for PRC2 association with chromatin mediated by DNA, it has also been suggested that RNA regulates PRC2 activity rather than directly facilitating its recruitment (Almeida et al., 2020). RNA has generally been found to inhibit the catalytic activity of PRC2 (Kaneko et al., 2014), leading to the question of how PRC2 functions in the presence of RNA, and how RNA can be involved in the association of PRC2 with its targets. A recent study suggested that PRC2 associates with RNAs that inhibit its catalytic activity; however, the RNA can make RNA-RNA base pairing with nascent RNA, which changes the conformation of the RNA to a state that is less inhibitory to PRC2 activity (Balas et al., 2021). This may explain how RNA can be generally inhibitory to PRC2 in the nuclear space but also can mediate specific association of PRC2 at target sites.

1.4.4.1 The role of RNA in the association of PRC2 with *FLC*

The hot theme of the role of RNA in Polycomb repression naturally led to the study of RNA in the repression of *FLC*. At *FLC*, several ncRNAs have been described as being involved in the cold-induced repression of *FLC* (Heo and Sung, 2011; Kim and Sung, 2017; Swiezewski et al., 2009). As introduced previously, cold leads to the induction of *COOLAIR*, and while

most studies had studied the role of *COOLAIR* in the transcriptional shutdown of *FLC* (Csorba et al., 2014; Rosa et al., 2016; Swiezewski et al., 2009), recent work has suggested that *COOLAIR* can mediate the association of PRC2 with chromatin through the RNA binding protein FCA (Tian et al., 2019). Similar to the debated roles of ncRNAs in animals, the role of the antisense RNA *COOLAIR* in vernalization has been debated since its discovery. The use of T-DNA mutants that uncouple *COOLAIR* from *FLC* suggested that *COOLAIR* was not required for H3K27me3 deposition; consistent with this was the finding that transgenic lines that are disruptive in *COOLAIR* induction show no defect in H3K27me3 accumulation (Csorba et al., 2014; Helliwell et al., 2011). However, recent elegant studies have shown that in all tested cases, transgenic lines used to study the role of *COOLAIR* failed to completely remove antisense transcription from the *FLC* locus (Zhao et al., 2021).

Another ncRNA reported to be involved in Polycomb repression is the sense ncRNA *COLD AIR* (Heo and Sung, 2011). *COLD AIR* is transcribed from within intron 1 and it is induced as *FLC* sense decreases (Heo and Sung, 2011). Similar to *COOLAIR*, *COLD AIR* is induced by vernalization and appears to be maximally expressed after three weeks of vernalization, coinciding with the decrease in sense *FLC* expression (Heo and Sung, 2011). This behaviour is consistent with the presence of cryptic promoters within the intron 1, a well-known phenomenon in eukaryotes, including plants, known as transcriptional interference (Ard et al., 2017; Nielsen et al., 2019). It has been suggested that *COLD AIR*, like *COOLAIR*, recruits PRC2 to *FLC*, but in contrast to *COOLAIR*, *COLD AIR* has been reported to directly interact with the Ezh homolog CLF (Heo and Sung, 2011). Interestingly, a specific region of *COLD AIR* was reported to be able to bind to PRC2, and it was argued that a stem-and-loop structure might mediate interaction with CLF, and act as a direct link for recruitment of PRC2 to *FLC* (Kim et al., 2017). This somewhat resembles the reported binding of PRC2 to the stem-loop structure within the XIST/RepA

lncRNA (Zhao et al., 2008), involved in PRC2 recruitment to the inactivated X chromosome.

A third ncRNA, *COLDWRAP*, has been described, and similar to *COLDAIR* it is transcribed in the sense orientation relative to *FLC* (Kim and Sung, 2017). In contrast to all other transcripts at the *FLC* locus, *COLDWRAP* is induced throughout vernalization and it only seems to be slightly repressed as the plants return to warm (Kim and Sung, 2017b). Interestingly, *COLDWRAP* terminates exactly at the end of exon 1 of *FLC* sense around the nucleation region, which may indicate that *COLDWRAP* results as a by-product of stalled RNAPII, as a consequence of nucleating PRC2 and/or H3K27me3. This resembles the observation of the presence of short RNAs from mammalian Polycomb targets (Kanhere et al., 2010). Similar to the short mammalian Polycomb RNAs and *COLDAIR*, *COLDWRAP* has been reported to associate with CLF through a part of *COLDWRAP* that is predicted to adopt a stem-loop structure (Kim and Sung, 2017b).

Over the years, one of the main arguments for the lack of function of ncRNAs has been their general low sequence conservation. However, lncRNAs can work through transcription or through the structure the RNA adopts. The antisense *COOLAIR* has been reported to work through both mechanisms. As mentioned earlier, full-length *COOLAIR* transcription is mutually exclusive with full-length *FLC* transcription (Rosa et al., 2016). In addition, despite low sequence conservation, the structure of *COOLAIR* is conserved in other Brassica species (Hawkes et al., 2016) but the sense ncRNAs, *COLDAIR* and *COLDWRAP* have not been identified beyond *Arabidopsis* (Castaings et al., 2014; Li et al., 2016). However, another Polycomb target in plants *AGAMOUS* (*AG*) has likewise been shown to contain several sense ncRNAs, one of which specifically interacts with CLF and has been reported to be required for PRC2 binding at *AG* (Wu et al., 2018). This suggests that the involvement of sense ncRNAs could be a

more general factor for the association of PRC2 with their genomic targets in *Arabidopsis*.

In addition to mediating the recruitment of CLF to *FLC*, the expression of *COLDWRAP* was reported to alter the chromatin conformation at *FLC* (Kim and Sung, 2017b). Likewise, recent genome-wide studies in mammalian cells have reported that PRC2, in addition to mediating the deposition of H3K27me₃, also drives the formation of long-range interaction along the chromosome fibre, an activity that is required for the spreading of H3K27me₃ from nucleation sites (Kraft et al., 2020; Oksuz et al., 2018). Consistent with the reported role of *COLDWRAP*, the formation of long-range interactions mediated by PRC2 is dependent on the RNA binding ability of PRC2 (Kraft et al., 2020).

1.5 Memory in 3D

Until this point, the role of Polycomb proteins has been attributed to modification of the histone code and PRC1-mediated chromatin compaction. However, PcG proteins are also involved in the regulation of the 3D organisation of the chromatin in the nucleus. With the use of high-throughput chromatin conformation capture (Hi-C), a technique that maps 3D contacts genome-wide, many chromatin loops were identified within Polycomb repressed regions in the *Drosophila* genome, and these were associated with gene repression (Eagen et al., 2017). Analysis revealed that anchor regions overlap with the cPRC1 protein Ph and that Ph-mediated chromatin topology requires the polymerization ability of the SAM domain (Eagen et al., 2017; Wani et al., 2016) (Fig. 1-7). It has been suggested that the formation of chromatin loops plays an instructive role in gene silencing as the prevention of loop formation decreases the silencing efficiency (Ogiyama et al., 2018). One function of loops related to Polycomb is presumably to facilitate the spreading of H3K37me₃ from nucleation sites (Oksuz et al., 2018).

Genome-wide studies on *Arabidopsis* have shown that in contrast to animals, the higher-order units in *Arabidopsis* are formed with lengths that correlate with the lengths of gene bodies (Liu et al., 2016), thereby more closely resembling the scenario reported in yeast.

In the early 2000s, the formation of local gene loops at yeast genes was shown to correlate with active transcription (O’Sullivan et al., 2004). The loop forms between the promoter and terminator of a TU, is dependent on active transcription and requires the activity of the RNA 3’ end processing complex (Ansari and Hampsey, 2005). It is suggested that the formation of a local gene loop stimulates fast recycling of RNAPII from the terminator to the promoter through direct protein-protein interaction between the 3’ end processing complex component SSU72 and the initiation complex component TFIIB (Ansari and Hampsey, 2005) (Fig.1-7). While the original data suggested that gene looping was a simple consequence of an initial round of transcription facilitated by initiation and termination factors, a recent study has shown that RNAPII plays an active role through proteins that associate directly with RNAPII (Allepuz-Fuster et al., 2019).

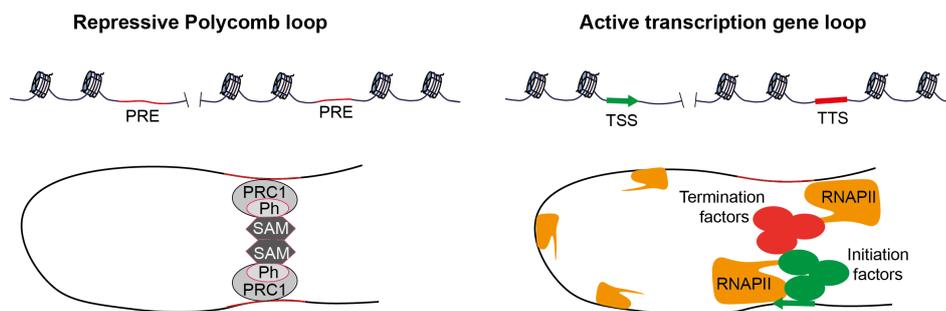


Figure 1-7 Two models for the function of chromatin loops. Repressive loops form through SAM-domain-mediated polymerization of the cPRC1 component Ph. Active gene loops juxtapose transcription start sites (TSS) and transcription termination sites (TTS) through interaction between transcription initiation and termination factors.

The work in this thesis is focused on cellular memory, and while local transcription gene loops have not been reported to play a role in epigenetic memory, they do reportedly provide another form for memory known as transcriptional memory. Transcriptional memory implies that reactivation

of gene expression occurs faster after a second induction compared to the first (Deng and Blobel, 2010). A study in yeast directly linked the presence of a gene loop with faster reinduction after a short intervening period of repression (Tan-Wong et al., 2009).

1.5.1 Arabidopsis chromatin loops

As mentioned above, the first genome-wide study in Arabidopsis identified that most gene loops form locally at TUs. However, newer studies have shown that similar to animals, long-range interactions are formed and these compartmentalize chromatin with the same chromatin marks (Feng et al., 2014; Huang et al., 2021). Consistent with chromatin and/or PcG proteins playing a vital role in the establishment of chromatin contacts is the fact that mutants which affect H3K27me3 also lead to changes in chromatin interaction. For example, mutants that increased levels of H3K27me3 led to the formation of new chromatin loops (Huang et al., 2021). This is consistent with the observation that *FLC* copies come together in the nucleus during vernalization as they acquire H3K27me3 (Rosa et al., 2013).

At *FLC*, *LHP1* is required for the spreading of H3K27me3 and this requires RNA binding (Berry et al., 2017). The spreading of H3K27me3 at other Arabidopsis genes likewise relies on functional *LHP1*. Loss of *lhp1* not only leads to disrupted spreading of H3K27me3, but it also results in distinct changes in chromatin topology compared to WT (Veluchamy et al., 2016). Thus, LHP1 links the spreading of H3K27me3 to the presence of chromatin loops, similar to the mechanism reported in other systems (Kraft et al., 2020; Oksuz et al., 2018).

As described above, gene loops in yeast are thought to facilitate a high transcription state by juxtaposing the promoter with the terminator at TUs, enabling efficient RNA PolII recycling. Several such loops have been identified in plants, and the common trend is that formation of the gene loop correlates with expression of the gene (Jégu et al., 2015; Liu et al.,

2013). A gene loop between the promoter and terminator formed at *FLC* relates to active transcription, and as the plant experiences cold and *FLC* transcription is reduced, the gene loop is disrupted (Crevillén et al., 2013).

In addition to long-range contacts, short-range interactions, e.g., between the TSS region and intragenic regions, play a role in gene regulation locally. An intragenic loop between the promoter and the fourth intron at the sunflower *HaWRKY6* locus correlates with low expression (Gagliardi et al., 2019). This intragenic loop correlates with reduced RNAPII occupancy downstream of the loop, suggesting that the establishment of an intragenic loop can add an extra layer of memory that blocks full-length transcription, thereby maintaining the locus in a low expression state (Gagliardi and Manavella, 2020). Another example is the intragenic loop that forms at *WUSHEL (WUS)*; here the loop encompasses the entire TSS and represses transcription, likely by blocking the recruitment of RNAPII to the TSS (Guo et al., 2018). This is similar to how gene loops have been shown to inhibit the expression of genes in yeast (Nguyen et al., 2014).

Overall, Arabidopsis chromatin organisation resembles that of other organisms with the formation of both long-range interactions that compartmentalize chromatin and short-range interactions that regulate the local transcription state, either through the formation of a chromatin loop that stimulates transcription, or loops that repress transcription.

1.6 Thesis outline

This Introduction chapter has summarized the epigenetic switching and maintenance mechanisms central to growth and development. PRC2 accessory proteins are key regulators in these mechanisms, but how they function is still unclear. As discussed in animal systems, they stimulate the catalytic activity of PRC2, mediate the association of PRC2 with its targets, stabilize a higher oligomeric state of PRC2, and maintain contacts between distant chromosomal regions (Chen et al., 2020; Eagen et al., 2017; Højfeldt et al., 2019). The goal of my thesis was to investigate PRC2

accessory protein's function in exploiting the well-characterized PcG switching mechanism at the Arabidopsis locus *FLC*. The two VEL proteins VIN3 and VRN5 are required for this mechanism and are likely to function as general PRC2 accessory proteins in plants. Despite the importance of their role, very little was known about their molecular function at the start of this work.

The first Results chapter, Chapter II, discusses the investigation into the VEL protein's function at the nucleation region at the 5' end of *FLC*, with a focus on the cold-induced VEL protein VIN3. Through collaborative work it is shown that the N-terminal region of VIN3 contains a more complicated protein domain architecture than previously appreciated. Furthermore, the potential of the nucleic acid binding of VIN3 is investigated. This chapter also describes proteomics and genomics experiments aimed at uncovering the mechanism behind VIN3's specific association with the nucleation region.

In Chapter III, an investigation of the VEL domain is described. The VEL domain is shown to be a polymerization domain with a novel polymerization protein fold. We show that the VEL domain mediates protein-protein interaction between VIN3 and VRN5 and that *in vivo* VIN3 forms visible dynamic, nuclear foci.

In Chapter IV, the perspective shifts to examining the 3D organization of the *FLC* locus. The activity of the antisense promoter is shown to be important for gene loop formation at *FLC*.

Chapter V describes the rationale and methodology for generating transgenic lines in order to test the hypotheses discussed in Chapters II and III.

Finally, Chapter VI is a general discussion summarizing the results obtained in this thesis and considering Polycomb silencing and gene regulation in a wider context.

1.7 Statement of collaborations

This PhD project was part of a larger project trying to understand the core mechanism of the epigenetic switch at *FLC*. The overall project involved several people with different expertise. My PhD project work has greatly benefitted from close collaboration with Drs Marc Fiedler and Elsa Franco-Echevarría at MRC, LMB Cambridge, and Dr Anna Schulten (JIC). To acknowledge their help and make each contribution transparent, the beginning of each chapter contains a statement summarizing any results they contributed. Everything shown and discussed in this thesis is done with the acceptance of the people involved.

Chapter II

2 VIN3 recruitment to the nucleation region

This chapter describes my work on the chromatin-binding properties of VIN3 and how this connects to PRC2 association at the nucleation region of *FLC*. This work is a collaboration with Dr Elsa Franco Echevarria from MRC LMB. During my time at MRC LMB, I made the initial constructs that enabled Dr Echevarria to achieve protein expression and later structural determination of the composite domain of VIN3 from *Phoenix dactylifera*. I also contributed to the successful expression and purification of the Arabidopsis VIN3 composite domain from insect cells. I performed all the *in vitro* EMSA assays after receiving training from members of the Lori Passmore group at MRC LMB. I performed all the Arabidopsis analysis, apart from some of the ChIP-seq analysis, which was performed by Jitender Cheema.

2.1 Introduction

How PcG proteins are recruited to their target sites is a major question in Polycomb regulation. This question has been extensively studied in several different organisms. Early work on *Drosophila* led to a simple model where PcG proteins are recruited through direct protein-protein

interactions with sequence-specific factors (Simon and Kingston, 2009). These sequence-specific factors bind to DNA sequences termed Polycomb Response Elements (PREs) (Kassis and Brown, 2013). PREs contain sequence motifs for several trans-factors such as Pho, GAGA factor and several others (Kassis and Brown, 2013). From these, Pho in particular plays a key role in recruiting PcG complexes (Frey et al., 2016). Computational analysis of *Drosophila* PREs has shown that a single DNA motif is not enough to specify PREs; instead, clusters of several motifs function together to form a functional PRE in the genome (Ringrose et al., 2003). The complexity in *Drosophila* is indicative of the complex relationship between chromatin and PcG recruitment also found in other organisms.

The emerging functional role of cellular condensates reinforces the importance of multivalent interactions in driving biological processes (Sabari et al., 2020). In mammalian cells, PRC2 association with genomic targets relies on multivalent interactions (Margueron and Reinberg, 2011), chiefly at CpG-rich DNA around the transcription start sites TSS (Mendenhall et al., 2010; Tanay et al., 2007). These PRC2-bound CpG regions are also characterized by the co-occupancy of PRC1 and H2AK199ub1, a special DNA conformation and the lack of active transcription (Laugesen et al., 2019). In particular, the lack of active transcription is an important element of PRC2 recruitment in mammalian cells, and transcriptional arrest is sufficient to cause PRC2 recruitment to CGIs genome-wide (Riising et al., 2014). Unlike *Drosophila*, there are no well-characterized proteins with high affinity for DNA motifs in mammals; instead, the association of PRC2 with CGIs is thought to be the result of many low-affinity interactions (Yu et al., 2019). This has led to a model where PcG proteins constantly sample the chromatin, with the difference between target and non-target sites being an increased residence time of PcG proteins at the former (Laugesen et al., 2019).

The hunt for the mechanism of PRC2 recruitment in Arabidopsis has led to the identification of many transcription factors involved in the association of PRC2 with target sites in the Arabidopsis genome. In addition, research has identified genomic features associated with Arabidopsis PRC2 sites that are sufficient to trigger silencing when integrated into reporter constructs; these show similarities with Drosophila PRE sequences (Xiao et al., 2017). One hundred and seventy candidate PREs were identified at 132 high-confidence Arabidopsis Polycomb targets (Xiao et al., 2017). This study identified six enriched motifs, several of which are found at the *FLC* nucleation region (Fig.2-1) (Xiao et al., 2017). Several of these were also identified from genome-wide occupancy of FIE – a core member of Arabidopsis PRC2. The identified motifs include GAGA factor binding sites, also found at Drosophila PREs (Deng et al., 2013; Schuettengruber et al., 2009). These PREs enabled identification of trans-factors putatively involved in recruiting PcG proteins. For example, GAGA motifs are recognized by a family of BASIC PENTACYSTEINE (BPC) proteins, which bind as multimers and cause DNA conformational changes (Kooiker et al., 2005). Interestingly, the importance of linker DNA conformation at PREs in PcG recruitment has also been discussed for animal PRC2 accessory proteins, particularly MTF2 (Perino et al., 2018). In addition, BPC6 has been shown to directly interact with LHP1, revealing a direct way of recruiting PcG proteins (Hecker et al., 2015).

Another well-characterized cis motif – trans factor relationship – involves the Telobox motif and telomere repeat binding factors (TRBs). TRBs bind the Telobox motif and they have been shown to interact with the Ezh2 homologs CLF and SWN (Zhou et al., 2016, 2018). Consistent with an important role in PRC2 recruitment in Arabidopsis, a triple *trb* mutant resembles the mutant phenotype of the strong Polycomb mutants (Zhou et al., 2018).

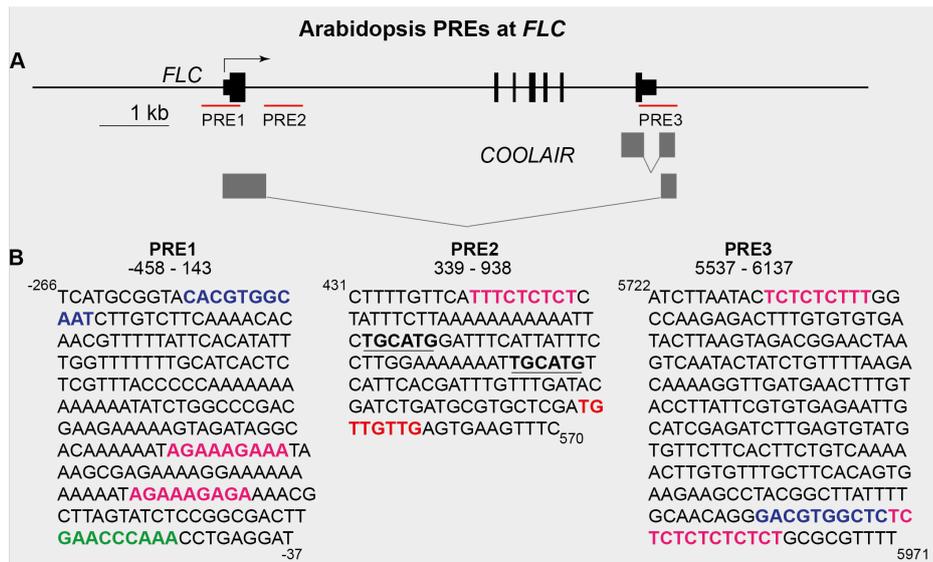


Figure 2-1 Bioinformatically identified PRE cis motifs at Arabidopsis *FLC*. PRE and associated cis motifs at *FLC* (Xiao et al., 2017). **A**) Three 600 bp candidate PREs (underlined) at *FLC*. **B**) Each PRE at *FLC* contains several of the identified PRE associated cis motifs; G-boxes (dark blue), GA repeats (pink), a Telobox (green), and an AC-rich (red). The two RY-motifs in PRE2 are underlined.

As illustrated in Fig.2-1, several of the PRE-associated cis motifs are present at *FLC*. However, for *FLC* repression a different cis motif appears to be the most important. This is the RY motif, identified through genetic approaches and occurring 500 bp downstream of the transcription start site. Two RY motifs are important for repression during vernalization and for memory of the repressed state after cold (Qüesta et al., 2016; Yuan et al., 2016). This region overlaps with the second candidate PRE identified from bioinformatic analysis (Fig.2-1B) (Xiao et al., 2017). The RY motif is recognized by B3 domain proteins including VAL1/VAL2, and cis mutations in the RY motifs have a similar loss of *FLC* repression phenotype as *val1* and *val1/val2* mutants (Qüesta et al., 2016; Yuan et al., 2016). As well as the DNA-binding B3 domain, VAL1 and VAL2 contain an EAR domain involved in protein-protein interactions with transcriptional repressors such as TOPLESS (TPL) and SAP18 (Kagale and Rozwadowski, 2011). VAL1 interacts directly with both PRC1 and PRC2 components (Yang et al., 2013a; Yuan et al., 2020), so it works as a central platform, assembling both transcriptional repressors and for epigenetic silencing (Baile et al., 2021). Similar to the triple *trb* mutant, a strong *val1/val2* double mutant shows a phenotype similar to the strongest

PRC2 mutants. Thus, VAL1/VAL2 is likely to be important for PRC2 recruitment at many targets genome wide (Yuan et al., 2020). However, VAL1/VAL2-RY is independent of both the TRB and BPC recruitment system (Yuan et al., 2020), consistent with RY-VAL1 being the main factor in PRC2 recruitment at *FLC*.

Non-core PRC2 components, also referred to as PRC2 accessory proteins, are becoming increasingly recognized as central to PRC2 recruitment. For example, the PRC2.2 subunit JARID2 binds H2AK119ub1 and mediates the recruitment of PRC2.2 (Blackledge et al., 2014; Cooper et al., 2016). Recent biochemical work has shown that naked DNA rather than nucleosomes dominates the interaction with Polycomb complexes (Wang et al., 2017). Consistent with this, the Polycomb-like proteins (PCLs), which are Polycomb accessory proteins that are part of PRC2.1, enhance the affinity of PRC2.1 for DNA (Choi et al., 2017; Li et al., 2017; Perino et al., 2018). It is suggested that the increased affinity obtained through the accessory proteins increases PRC2 residency time at target sites, holding it long enough to facilitate H3K27 trimethylation. Both JARID2 and the PCL protein MTF2 are required for *de novo* PRC2 recruitment in mammalian cells (Oksuz et al., 2018). Thus, PRC2 accessory proteins appear central to *de novo* recruitment and the maintenance of PRC2 at genomic target sites (Højfeldt et al., 2019; Laugesen et al., 2019).

The Arabidopsis VEL proteins robustly co-purify with core PRC2 components (De Lucia et al., 2008) (Dean lab unpublished), and these proteins are considered to be Arabidopsis PRC2 accessory proteins. The VEL proteins do not co-purify with other PRC2 accessory proteins, supporting the existence of a distinct VEL-PRC2 complex (Bloomer et al., 2020). The VEL proteins contain a PHD domain, implying an involvement in chromatin engagement, potentially by increasing the dwell time of PRC2 at the nucleation site (Kim and Sung, 2017a; Sung and Amasino, 2004).

This chapter describes my investigation of the cold-induced VEL protein VIN3 and its possible role in PRC2 targeting of the nucleation region. The chapter begins with the characterization of the N-terminal part of the VEL proteins; this contains a more complicated protein domain organisation than previously appreciated. This work led to an investigation of a putative nucleic acid binding ability of VIN3, and the identification of a conserved positively-charged patch in a winged-helix/Bromo-like domain that binds non-sequence specifically to nucleic acid. To complement the *in vitro* studies, I searched for specificity factors *in vivo* using proteomics and genomics.

2.2 Results

2.2.1 The VEL proteins contain additional domains

In vertebrates, PRC2 accessory proteins are important for the localization of PRC2 at nucleation regions (Oksuz et al., 2018). Similarly, the nucleation of H3K27me3 at *FLC* is dependent on *VIN3* and exposure to cold (Sung and Amasino, 2004). It is therefore intriguing to speculate that *VIN3* has a crucial role in increasing the association of PRC2 to the nucleation region at *FLC*, similar to mammalian accessory proteins.

To identify how *VIN3* might increase PRC2 association with *FLC*, I initially focused on the biochemical and biophysical properties of *VIN3*. In the initial characterization, *VIN3* was shown to contain a PHD domain (Sung and Amasino, 2004), the presence of which provided an intriguing mechanism of chromatin recognition. Previous studies had focussed on *VIN3* binding to different histone tail modifications and specific binding to H3K9me2 was found, indicating that the *VIN3* PHD domain is different from the majority of PHD domains, which bind trimethylated residues, particularly H3K4me3 (Kim and Sung, 2017a). Since my aim was to understand how VEL-PRC2 associates with the nucleation region, I focussed on the role of the *VIN3* PHD domain.

The atypical PHD region of the VEL proteins was explored by performing *in silico* secondary structure predictions. Using the phyre2 software, the VEL proteins, including VIN3, were predicted to have a well-defined PHD motif, with β -sheets and an α -helix (Fig. 2-2) (Kelley et al., 2015). Interestingly, the amino acid sequence was also predicted to contain regions of α -helices both up and downstream of the PHD domain. The upstream region is specific to VIN3, while the downstream four α -helical bundle is found in all the VEL proteins. We initially predicted that this four α -helical bundle could form a winged-helix-like fold. In addition, the region just upstream of the PHD domain was predicted to fold into a Zinc-finger domain. The shared ZnF, PHD and WH-like domains are in close proximity, suggesting that the individual domains could influence each other. We, therefore, defined this three-domain region as the composite domain (CD). The presence of several chromatin reader domains and a putative WH resembles the structure of mammalian PCLs (Choi et al., 2017; Li et al., 2017).

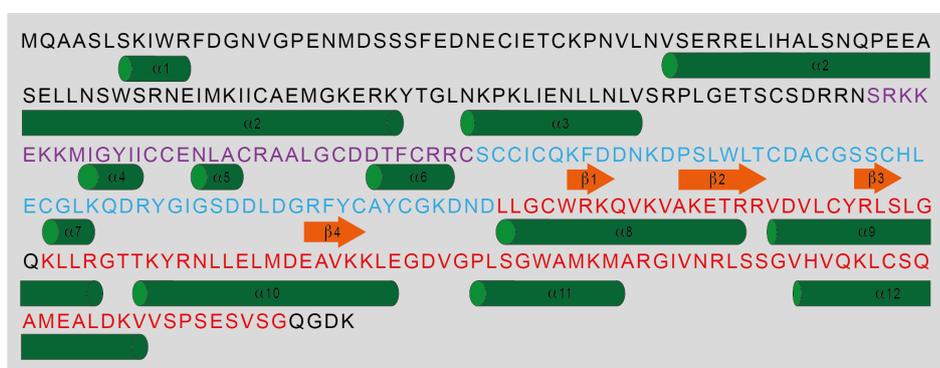


Figure 2-2 Secondary structure prediction of the N-terminal region of VIN3. The amino acid sequence of the N-terminal part of VIN3 is shown with the predicted secondary structure depicted below. α -helices are symbolized with green barrels and β -sheets as orange arrows. The secondary structure was predicted using the Phyre2 web portal (Kelley et al., 2015). The conserved Composite Domain (CD) shared between the VEL proteins contains the ZnF (purple), PHD (light blue), and a WH/Bromo-like domain (red).

To understand the function of the CD in VIN3, we undertook an in-depth analysis of the protein features through *in vitro* analysis. We first tried to express the individual domains in *E. coli*. The ZnF and PHD domains could be successfully expressed in bacteria; however, we were unable to

express the WH/Bromo-like domain. We speculated that the apparent toxicity of the WH/Bromo-like domain in bacteria could be due to its potential nucleic acid binding property. Because of the close arrangement of the three subdomains (Fig. 2-2), it is likely that the correct folding of the individual domains will depend on the neighbouring domains. We, therefore, decided to focus on the functional analysis of the full CD as a unit. Because the WH/Bromo-like domain could not be expressed in bacterial cells, we instead tried to express the VIN3 CD in Sf9 insect cells. The CD of VIN3 could be successfully expressed in Sf9 insect cells and following purification and removal of the tags, it was possible to obtain well-behaving protein samples that enabled me to perform biochemical experiments.

The proteins of Arabidopsis VIN3 CD from Sf9 cells failed to form crystals that were useful for structural analysis (E. Franco-Echevarria unpublished). Structural analysis is required to fully understand domain structure and a common approach in crystallography is to use homologues from other species. Using this approach, we searched for VIN3 homologues in the plant kingdom. Our analysis showed that the composite domain has been very well conserved throughout evolution and is present in most plant species (M. Fiedler personal communication). We selected homologues from four tropical species likely to have a more rigid protein conformation and thereby better suited for crystallography. The four species were *Hevea brasiliensis*, *Amborella trichopoda*, *Theobroma cacao* and *Phoenix dactylifera*, all of which contain a VIN3 CD domain very similar to that in Arabidopsis (Fig. 2-3A).

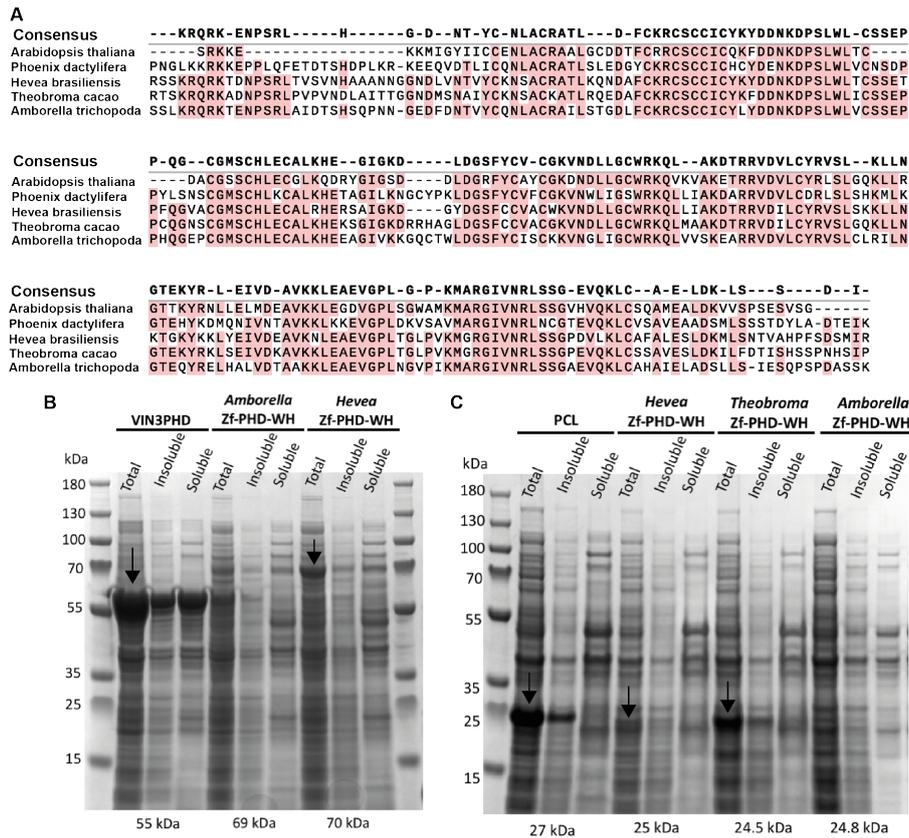


Figure 2-3 Exotic VIN3 CDs. **A)** Amino acid sequence alignment of the VIN3 CD from *Phoenix*, *Hevea*, *Theobroma*, and *Amborella* with VIN3 CD from *Arabidopsis*. Conserved residues are shaded pink. **B-C)** Expression of VIN3 CD from exotic species in pETM41 (**B**) or pEC-Lic (**C**). The expected size of the CD with the respective tags is shown below the SDS-Page gel. Total, insoluble, and soluble proteins were visualized by Coomassie staining.

The VIN3 CD from these four species was expressed, fused to His-MBP-H3K9M or through use of the pEC-Lic vector, which has been used successfully for expression of the PHD-WH domain of the PRC2 accessory protein PHF1 (Choi et al., 2017). In contrast to the *Arabidopsis* VIN3, we were able to express the CD from the exotic species in *E. coli* (Fig.2-3 B-C) (E. Franco-Echevarria unpublished), suggesting that the putative WH/Bromo-like domain from these proteins is not toxic for bacteria. The ability to express the CD in bacteria enabled not only structural determination but also biochemical analysis.

2.2.2 Structural determination of the PhVIN3 composite domain

The bacterially expressed CDs from the exotic species were used in crystallisation assays and one of them, the CD from *Phoenix dactylifera*, hereafter referred to as Phoenix, formed crystals useful for structural determination (E. Franco-Echevarria unpublished). The three other CDs did not form crystals in the initial screens, likely due to differences in the chosen domain boundaries. From the protein crystals, it was possible for Dr Elsa Franco-Echevarria to solve the crystal structure of the Phoenix CD. While a description of the structure of the CD is beyond the work in this thesis, some interesting features were observed in the structure, features that help to frame the experiments conducted in the following sections. Therefore, the key findings are summarized below.

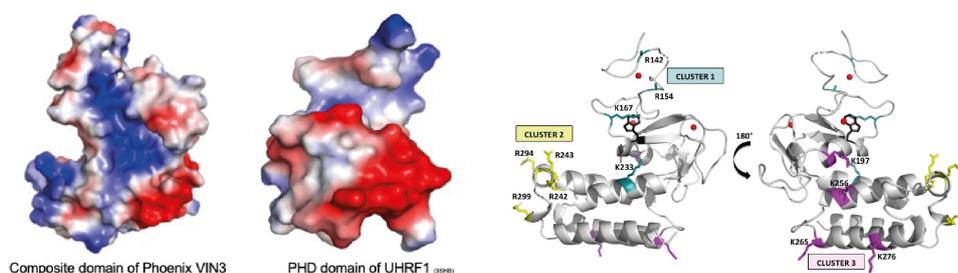


Figure 2-4 The composite domain of VIN3 mainly presents a positive surface. A) Electrostatic map of the composite domain of VIN3 from Phoenix and the ZnF-PHD of the human protein UHRF1. It shows the positive (blue) surface charge of the CD in the putative histone pocket and the more common negative charge (red) in the histone binding pocket of UHRF1. **B)** Ribbon structure of Phoenix VIN3 CD, showing the three conserved positive patches.

Very surprisingly, it was found that the PHD domain has a positive surface charge in the putative histone tail binding pocket (Fig. 2-4A). The majority of PHD domains that bind histone tails have a negative charge in order to accommodate the positively charged histone tail (Bienz, 2006; Mellor, 2006). This observation might explain why our attempts to identify histone tail binding of the CD from VIN3 have so far been unsuccessful, even with the use of NMR, which can identify even low-affinity binding (Dean lab unpublished).

The crystal structure confirmed the secondary structure prediction of a four α -helical bundle downstream of the PHD. Using the TM-align program, which compares protein structures, we determined that the structure of this region shows some similarities with a WH domain and even more similarities with a bromodomain (E. Franco-Echevarria unpublished). As mentioned earlier, PHF1, the mammalian PRC2.1 accessory protein, contains a WH-domain, and despite being mostly recognized as a histone reader domain, several bromodomains have been shown to bind DNA through positively charged patches (Miller et al., 2016; Weaver et al., 2018). Consistent with a role in DNA binding, the atypical-WH/Bromo-like domain of VIN3 contains four well-conserved arginine residues that could be involved in nucleic acid binding (Fig. 2-4B). Together, the presence of a positively charged PHD domain and a WH/Bromo-like domain prompted us to ask if the CD could associate with nucleic acids (Choi et al., 2017; Li et al., 2017).

2.2.3 Composite domain of VIN3 associates with nucleic acid

2.2.3.1 VIN3-CD associates with DNA non-specifically *in vitro*

We investigated whether the AtVIN3 CD purified from insect cells could bind to nucleic acid by electrophoretic mobility shift assays (EMSAs). VIN3 has only been reported to bind *FLC* around the TSS of *FLC*, in the region within the PRE1 region described earlier (Fig. 2-1) (Yang et al., 2017). We therefore initially used a 330bp PCR product encompassing the VIN3 ChIP-peak region as the probe in the EMSA experiments (Fig. 2-5A). Increasing amounts of AtVIN3 CD were able to cause a gel-shift indicative of binding (Fig. 2-5B). The AtVIN3 CD produced a “well-shift”, which is also the case for other EMSA analyses of PcG complexes for nucleic acid binding (Alecki et al., 2020). The “well-shift” rather than discrete bands means that the quantification of the binding observed in these experiments would not be very accurate.

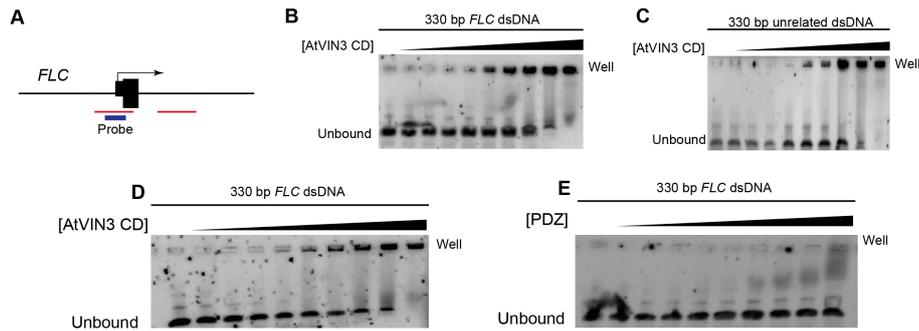


Figure 2-5 AtVIN3 CD associates with dsDNA. A) Schematic drawing of the 5' end of *FLC* with the two PRE regions (red) identified in Xiao et al. (2017) and the 330 bp probe used for the EMSA analysis (purple). B-C) *Arabidopsis thaliana* VIN3 composite domain (AtVIN3 CD) binding to dsDNA from *FLC* (B) and an unrelated gene (C). D) AtVIN3 CD binding to dsDNA from *FLC*. E) No binding of an unrelated protein (PDZ) with similar biochemical properties to AtVIN3 CD. [AtVIN3 CD] = 0.12-31.2 μ M.

DNA fragments containing sequences from within the region covering *FLC* TSS to the RY-motif showed a similar binding to that in the EMSA in Fig.2-5. The same was also observed when an unrelated bacterial sequence of the same length was used as the probe (Fig. 2-5 C). Together, the EMSA revealed non-specific association of the VIN3 CD with DNA. This is similar to previous reports on the DNA-binding ability of the WH of PHF1 (Choi et al., 2017).

Like many other nucleic acid binding proteins, the CD has an overall alkaline isoelectric point (Castello et al., 2012), which means that at the pH used for the EMSA assays, the protein will have an overall negative surface charge. Because of the observed non-specificity, we wondered whether a similar shift would be caused by a non-related protein with a similar pI. We therefore performed the same EMSA analysis with a PDZ domain from an unrelated animal membrane protein with similar pI. Consistent with nucleic acid binding being a property of the VIN3-CD, we did not observe any obvious shift when the 330 bp *FLC* DNA probe was incubated with increasing amounts of the PDZ protein (Fig.2-5 D-E).

2.2.4 The CD associates with nucleic acids generally *in vitro*

2.2.4.1 RNA

In the previous section, we showed that the VIN3 CD is able to associate with DNA non-sequence specifically. In recent years, RNA has been increasingly recognized as important for the association of PRC2 to its target sites (Kretz and Meister, 2014). Similarly, the importance of RNA in recruiting PRC2 to *FLC* has been suggested; several ncRNAs originating from *FLC* have been reported to interact with core PRC2 components (Heo and Sung, 2011; Kim and Sung, 2017b; Tian et al., 2019). Therefore, we tested whether the VIN3-CD could bind to RNA and whether potentially it would have higher affinity for RNA than DNA. COLDAIR, COLDWRAP and different distal COOLAIR forms were *in vitro* transcribed, purified and folded as described previously for COOLAIR (Hawkes et al., 2016). As observed for dsDNA, we saw a “well-shift” of all RNA tested with increasing amounts of VIN3-CD, indicative of association of the protein with the RNAs (Fig. 2-6 A-F). To test whether the binding was dependent on the RNA being folded, we performed the same experiments on RNA that was heated at 95°C for 2 min and then snap-frozen. This did not affect the pattern of retarded migration, suggesting that the association does not require the RNA to be folded. However, as the EMSAs were performed at 4 °C, the RNA would very likely adopt a folded conformation, therefore the assay is not strictly performed on linear RNA. To further test whether the association is independent of structure, we performed EMSA using annealed short RNA oligos. The same EMSA patterns were found, showing that the RNA association is independent of RNA sequence or structure (Fig.2-6G).

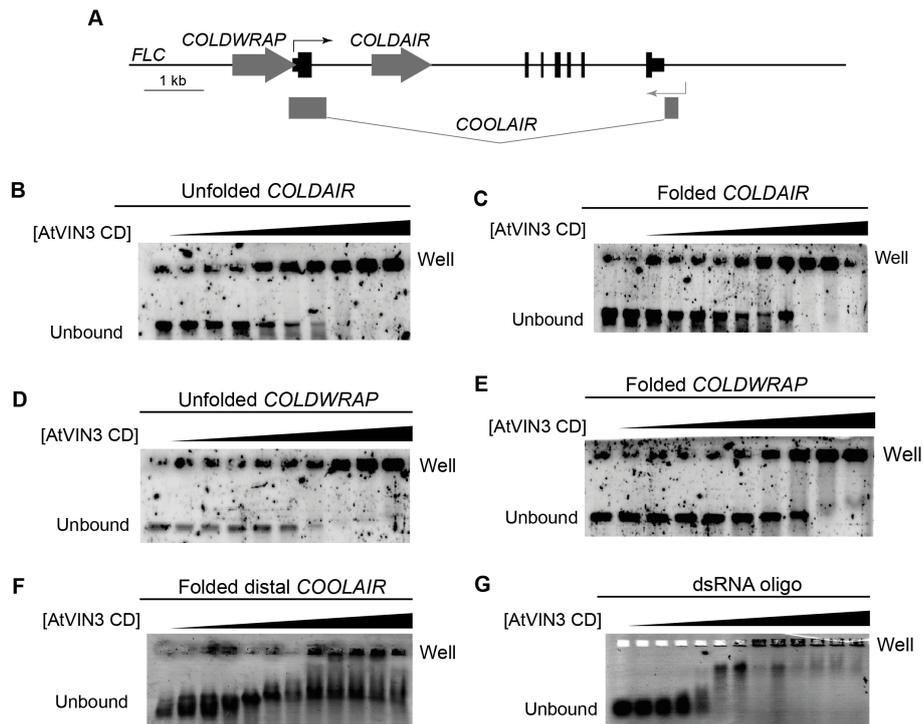


Figure 2-6 AtVIN3 CD binding to RNA. A) Schematic drawing of the *FLC* locus illustrating the three ncRNAs (COLDAIR, COLDWRAP, and distal COOLAIR) used for the EMSAs in B-G. B-C) AtVIN3 CD binding to COLDAIR “unfolded” (B) and folded (C) [AtVIN3 CD] = 0.14-32 mM. D-E) AtVIN3 CD binding to COLDWRAP “unfolded” (D) and folded (E) [AtVIN3 CD] = 0.14-32 mM. F) AtVIN3 CD binding to folded distal COOLAIR class2-I [AtVIN3 CD] = 0.25-10 mM. G) AtVIN3 CD binding to annealed dsRNA oligo (25 bp) [AtVIN3 CD] = 0.25-10 mM.

2.2.4.2 Different nucleic acids

In the sections above we observed that the CD of VIN3 associates with all the tested nucleic acids with similar characteristics. To explore this apparent non-specificity further, we use labelled synthesized oligos to analyse binding to different nucleic acid species and sequences. We first tested whether the GC content affected the binding of the CD. Consistent with non-specific association of the CD to nucleic acid, we did not observe any difference (Fig.2-7 A-B). We then tested whether the VIN3 CD has a preference for DNA:RNA hybrids or single-stranded DNA, the two components of an R-loop. R-loops, three-stranded nucleic acid structures, are formed when an RNA hybridizes with the complementary DNA strand and displaces the second DNA strand. They have been implicated in PRC2 recruitment in mammalian cells (Skourti-Stathaki et al., 2019). Furthermore, both PRC1 and PRC2 have been shown to bind to R-loops *in vitro* using EMSA analysis, similar to the approach presented in this

chapter (Alecki et al., 2020). However, again we observed a band shift pattern similar to that of all the other nucleic acids tested (Fig.2-7 C-D).

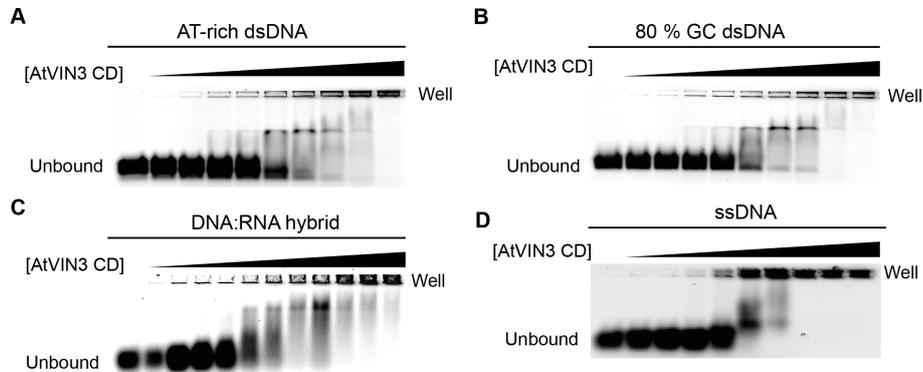


Figure 2-7 AtVIN3 CD binding to nucleic acid. **A-B)** AtVIN3 CD binding to dsDNA with high AT content (A) or GC content (B). [AtVIN3 CD] = 0.25-6 mM. **C)** AtVIN3 CD binding to DNA:RNA duplex (C) [AtVIN3 CD] = 0.25-8 mM. **D)** AtVIN3 CD binding to single stranded DNA (ssDNA) [AtVIN3 CD] = 0.1-31.3 mM.

2.2.5 The atypical WH/Bromo-like domain is required for binding

The *in-silico* prediction and later the crystal structure showed that the last part of the CD folds into a structure that has some similarities to a WH- or Bromo- domain. Both domains have been shown to be able to bind nucleic acid. We therefore hypothesized that the observed property of the VIN3 CD to associate with nucleic acid is caused by this region of the CD. However, because the crystal structure revealed a positive surface charge of the PHD domain, it is possible that the PHD could also be the causative domain, by interacting with the negatively charged nucleic acid. We performed EMSA with the 330 bp *FLC* sequence with protein samples that lacked the WH/Bromo like domain. Consistent with the atypical-WH/Bromo-like domain being responsible for the association, neither the PHD domain alone nor the ZnF-PHD were able to cause the observed association with the DNA probe (Fig.2-8 A-C). However, as was observed in other protein structures, we cannot rule out the possibility that the three subdomains stabilize each other, so although the proteins missing the atypical WH/Bromo-like domain were well folded based on gel filtration and NMR analysis (Dr E. Franco-Echevarria), the atypical WH/Bromo-like domain could have a role in stabilizing the overall fold of the domain.

To address this potential drawback, we designed specific point mutations in the putative nucleic acid binding interface on the CD (Fig. 2-8 A), together with mutations in the two other conserved surface-exposed positively charged patches (Fig.2-4 B). Because the CD of Phoenix is more easily expressed in bacteria, we generated the mutations in the Phoenix CD. Importantly, in EMSA analysis the VIN3 CD from Arabidopsis and Phoenix behave similarly (Fig. 2-8 D-E). Consistent with the arginine residues being involved in the nucleic acid association, the charge swap mutations abolish the EMSA interaction (Fig. 2-8 D-F). However, the same effect was also observed with charge-swap mutations in the other conserved patches, suggesting that all patches are required for the association.

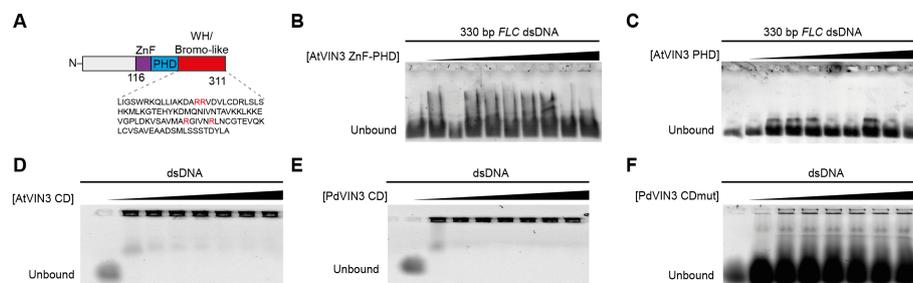


Figure 2-8 The WH/Bromo-like domain is required for nucleic acid binding. A) Drawing of the N-terminal region of AtVIN3 with the three domains of the composite domain (ZnF, PHD and WH/Bromo-like). The sequence of the WH/Bromo-like domain from *Phoenix dactylifera* (Pd) is shown below with the conserved exposed Arginines marked in red. B-C) EMSA on 330 bp dsDNA with AtVIN3 ZnF-PHD (B) and PHD only (C) [AtVIN3 ZnF-PHD/PHD] = 0.1-31.3 μ M. D-F) EMSA on annealed dsDNA oligos with AtVIN3 CD (D) PdVIN3 CD (E) and PdVIN3 CD carrying four Arginine to Aspartic acid mutations of the conserved Arginines shown in (A) [AtVIN3CD/ PdVIN3CD/ PdVIN3 CDmut] = 2-10 μ M.

Overall, the EMSA experiments suggested that the CD, most likely through the WH/Bromo-like domain, can associate with nucleic acid. All tested nucleic acids, independent of type (ssDNA, dsDNA, ssRNA, dsRNA and RNA:DNA hybrid) or sequence showed a similar EMSA pattern, indicative of non-specific association. This is similar to observations for PRC2 accessory proteins in other organisms (Choi et al., 2017).

2.2.6 ChIP-seq

Our *in vitro* binding assays did not reveal any specific binding to nucleic acid. To try and gain an understanding of what is required for a VIN3 target gene *in vivo*, we performed VIN3 ChIP-seq in 6 weeks cold treated seedlings of the previously published pVIN3:VIN3-eGFP line. In this paragraph, I will summarize some of the characteristics of the ChIP-seq analysis. Unfortunately, we only performed one replicate of the ChIP-seq so the strength of the data is not high. To date, *FLC* is the only known target of VIN3 in *Arabidopsis*. In order to validate the ChIP-seq method, we first looked at the reads that corresponded to the *FLC* genomic region. Consistent with the results of the ChIP-qPCR, we observed a clear peak of reads corresponding to the TSS region of *FLC*. This observation suggested that the ChIP-seq had worked.

As the ChIP-seq was only performed once, we tried to gain some confidence in the enriched genes by comparing them to published ChIP-seq data of the other PcG proteins, SWN and CLF, as well as VAL1 (Shu et al., 2019; Yuan et al., 2020). This showed a limited number of common targets, including some of the other *FLC* clade members. Interestingly, analysis of the VIN3 peaks revealed a slight tendency of VIN3 to bind at the 5' end of VIN3 target genes. It would be interesting to not only repeat the VIN3 ChIP-seq but also to include RNA-seq analysis in the *vin3* mutant background at different timepoints during vernalization, similar to work carried out with wildtypes.

2.2.7 The VEL proteins interact with other transcription regulators

Our EMSA analysis did not reveal specificity for any type of nucleic acid or sequence. In parallel with the ChIP-seq analysis, we therefore performed immunoprecipitation-mass spectrometry (IP-MS) of VIN3 and VRN5 to determine whether the VEL proteins associated with other proteins. This could help to explain the localization of the VEL-PRC2 complex in the nucleation region at *FLC*. We had previously performed

IP-MS for many of the relevant proteins (De Lucia et al., 2008; Qüesta et al., 2016). In an attempt to identify more transient interactions, we performed crosslinked-mass spectrometry (CLNIP-MS), which has previously been used successfully to identify transient interactions between protein complexes in the autonomous pathway (Fang et al., 2019, 2020). Two repeats of the CLNIP-MS for VRN5-YFP were performed, one of which was processed by the JIC proteomics platform. The second was processed by our collaborator (Tiancong Lu), and VIN3-GFP was also included. Common in all the CLNIP-MS lists were the VEL-PRC2 proteins, with the most abundant peptides coming from VEL1 (Table 2-1). This suggests that VEL1 may act as a bridge between VIN3 and VRN5. VEL1 is also the most abundant interactor for VRN5 in the absence of VIN3 (NV and 6WT7), showing that VRN5-VEL1 is also interacting in the absence of VIN3. The next most abundant interactors were in both cases the other VEL-PRC2 subunits (SWN, MSI1, FIE, VRN2 and VIN3/VRN5). As expected, the IP-MS of VRN5 post-vernalization (6WT7) showed that very few VIN3 peptides associated with VRN5; however, the remaining VEL-PRC2 subunits were robustly identified.

Table 2-1 VIN3 and VNR5 robustly associated with VEL-PRC2

Protein	<i>ColFRI</i>	<i>gVIN3-GFP</i> <i>vin3-4 FRI</i>		<i>gVRN5-YFP</i> <i>vrn5-8 JU223</i>		
	6W0	6W0	6WT7	NV	6W0	6WT7
VIN3	0-0	102	8	0	7-1	0-1
VRN5	0-2	19	0	27	22-22	38-57
VEL1	0-0	43	5	15	27-16	32-60
SWN	0-0	13	1	1	3-13	2-17
MSI1	0-0	11	1	0	10-7	6-14
FIE	0-0	8	1	0	7-6	3-13
VNR2	0-0	8	0	0	2-3	1-8

IP-MS of VIN3-GFP and VRN5-YFP in non-vernalized conditions (NV), after six weeks of cold treatment (6W0) and after six weeks of cold treatment followed by 7 days warm growth (6WT7). The numbers indicate unique peptides from each protein identified by IP-MS. For VRN5, the numbers are from two independent experiments.

In addition to the VEL-PRC2 components, we identified a few other reproducible interactors with VIN3/VRN5. TOPLESS (TPL) was found in both the VIN3 and VRN5 IP-MS data (Table 2-2). This supports the previous finding of a direct interaction between VRN5 and TPL (Causier et al., 2012; Collins et al., 2019). TPL, the Arabidopsis Groucho homologue, is an important corepressor involved in the transcriptional repression of several important developmental genes (Kagale and Rozwadowski, 2011; Long et al., 2006). Corepressors like TPL often work through histone deacetylases (HDACs) to set a low transcriptional chromatin state, including HDA19 (Long et al., 2006). Interestingly, we also identified several HDACs including HDA19 in the list of interactors, which is consistent with previous observations of VIN3 and VRN5 interacting with HDA19 and the ASAP complex (Qüesta et al., 2016).

Table 2-2 VIN3 and VRN5 associated with TOPLESS and members of the TOPLESS interactome

Protein	AGI code	<i>ColFRI</i>	<i>gVIN3-GFP</i> <i>vin3-4 FRI</i>		<i>gVRN5-YFP</i> <i>vrn55-8 JU223</i>	
		6W0	6W0	6WT7	6W0	6WT7
VIN3	AT3G24440	0	102	8	1	1
VRN5	AT5G57380	2	19	0	22	57
TPL	AT1G15750	0	7	4	6	7
TPR2	AT3G16830	0	3	1	2	4
TPR1	AT1G80490	0	0	0	0	1
HDA4	AT5G22650	0	0	0	0	1
HDA19	AT4G38130	0	0	0	0	1
FVE	AT2G19520	0	2	1	1	1

IP-MS of VIN3-GFP and VRN5-YFP in non-vernalized conditions (NV), after six weeks of cold treatment (6W0) and after six weeks of cold treatment followed by 7 days warm growth (6WT7). The numbers indicate unique peptides from each protein identified by IP-MS.

The previous identification of the ASAP complex among the interactors of VIN3 and VRN5 suggests an interesting link between RNAPII transcription, RNA processing and Polycomb repression (Qüesta et al., 2016). With crosslinked IP-MS, I did not identify the ASAP components, but I did identify several proteins that are part of, or associated with, the NineTeen complex (NTC), which interacts with the ASAP complex (Table 2-3) (Monaghan et al., 2009). The NTC is important for the splicing of RNA and it is critical for several steps in the splicing process (Chanarat and Sträßer, 2013). In addition to the well-characterized role in the splicing process, NTC has also been more directly linked to transcription elongation in yeast by facilitating the interactions with the THO/TREX complex (Chanarat et al., 2011). Some subunits of this complex were also found in our proteomics.

Interestingly, many of these splicing factors have also been found in VRN1 proteomics (D. Zhu unpublished), potentially linking the progression of RNAPII, RNA maturation and the chromatin state with PRC2 recruitment. A recent study on mammalian cells likewise identified that the splicing-associated protein RbFox2 can associate with PRC2. It was suggested that

this interaction plays a role in associating PRC2 with its genomic targets (Wei et al., 2016).

Table 2-3 VIN3 associates with splicing factors

Protein	AGI code		<i>ColFRI</i>	<i>gVIN3-GFP</i> <i>vin3-4 FRI</i>		<i>gVRN5-YFP</i> <i>vrn5-8 JU223</i>	
			6W0	6W0	6WT7	6W0	6WT7
VIN3	AT3G24440		0	102	8	1	1
VRN5	AT5G57380		2	19	0	22	57
MAC3A	AT1G04510		0	3	0	2	1
MAC3B	AT2G33340		0	8	1	4	3
CDC5	AT1G09770	NTC/NTR	0	3	0	0	0
SKIP	AT1G77180		0	1	0	0	0
MAC7	AT2G38770		0	2	0	0	0
	AT5G41770		0	2	0	0	0
CLO	AT1G06220	U5	0	5	3	2	6
EMB1507	AT1G20960		0	8	5	3	5
U2A	AT1G09760	U2	0	1	0	1	1
SR34	AT1G02840	SR	0	3	1	2	1
SR34A	AT3G49430		0	0	1	0	1
SCL30A	AT3G13570		0	3	0	0	1
RH2	AT3G19760	EJC	0	9	1	2	2
MAGO	AT1G02140		0	1	0	1	1

The numbers indicate unique peptides from each protein identified by IP-MS. VIN3 associated with proteins of the NineTeen complex (NTC) or those related to the NTR. VIN3 further associates with components of the U2 and U5 sRNP, Serine/Arginine splicing proteins (SR), and Exon-Junction Complex proteins (EJC).

2.2.7.1 VIN3 associates with other PRC1-related proteins

Two other classes of potentially interesting proteins were identified in samples harvested at 6WT0. The first was UBIQUITIN SPECIFIC PROTEASES 12 and 13 (UBP12/UBP13), proteins previously identified as associating with LHP1 and required for *FLC* repression in Col-0 (Derkacheva et al., 2016). Interestingly, recent work suggests that PRC1-deposited H2Ub can function at genes to mediate rapid switching between an active and a repressed transcriptional state. However, that H2Ub must

then be removed for stable PRC2 dependent repression (Kralemann et al., 2020). Recent work in our lab has shown that H2Ub does accumulate at *FLC* during vernalization in a PRC1-dependent manner (Mikulski et al., 2021). Further supporting the link between VIN3 and the removal of H2Ub, H2Ub was shown to accumulate further shortly after cold, when VIN3 is no longer present (Mikulski et al., 2021).

Table 2-4 VIN3 associated with PRC1 and PR-DUB proteins

Protein	<i>ColFRI</i>	<i>gVIN3-GFP</i> <i>vin3-4 FRI</i>		<i>gVRN5-YFP</i> <i>vrn5-8 JU223</i>	
	6W0	6W0	6WT7	6W0	6WT7
VIN3	0	102	8	1	1
VRN5	2	19	0	22	57
UBP13	0	5	0	0	0
UBP12	0	1	0	0	0
AL2	0	1	1	0	1
AL4	0	1	1	0	2
AL6	0	3	0	0	2
AL7	0	1	0	1	1

The numbers indicate unique peptides from each protein identified by IP-MS

The second interesting class was the ALFIN-like family proteins (AL1-7). AL proteins are PHD proteins that associate directly with the core PRC1 components RING1 and BMI1 through multiple interactions (Peng et al., 2018). As discussed earlier in this chapter, PHD domains are mostly known for their ability to bind H3K4 and this is the function of the AL PHDs, which preferentially bind H3K4me3 (Peng et al., 2018). This has led to a model where ALs play a role in initiating the repression of actively transcribed genes marked by H3K4me3. H3K4me3 is recognized by the ALs, which then recruit PRC1 to mediate transcriptional repression and PRC2-mediated H3K27me3 deposition (Molitor et al., 2014). Supporting the interactions observed here, we previously observed other AL proteins interacting with VEL-PRC2 components and the PRC1-interacting protein VAL1 (unpublished). Despite being reported to be involved in other developmental pathways in Arabidopsis, there are no studies on the effects

of ALs on *FLC* expression or flowering time. The role of ALs may have been overlooked due to redundancy between the 7 AL proteins in Arabidopsis.

2.2.8 VAL1 is not required for VIN3 association with FLC

VAL1 is a central player in PcG recruitment at *FLC* and the Arabidopsis genome generally (Qüesta et al., 2016; Yuan et al., 2020, 2016). As a sequence-specific DNA binding protein, it has been suggested that it is a functional homologue of Drosophila Pho, which binds the PRE and recruits PRC1 and PRC2 through direct protein-protein interactions. It is intriguing to speculate that VAL1 also recruits VEL-PRC2 at *FLC*. To test whether *VAL1* is directly or indirectly required for VIN3 chromatin binding at *FLC*, I performed protein-ChIP of VIN3-GFP in *val1-2* after six weeks of vernalization. Surprisingly, I observed no difference in VIN3 occupancy at *FLC* between the wild type and *val1-2* (Fig.2-9). This suggests that VAL1 is not strictly required for VIN3 recruitment to *FLC*, at least when measured after six weeks of cold exposure. Genotypes to assess whether cis mutations affect VIN3 recruitment are being generated but were not available in time for this thesis.

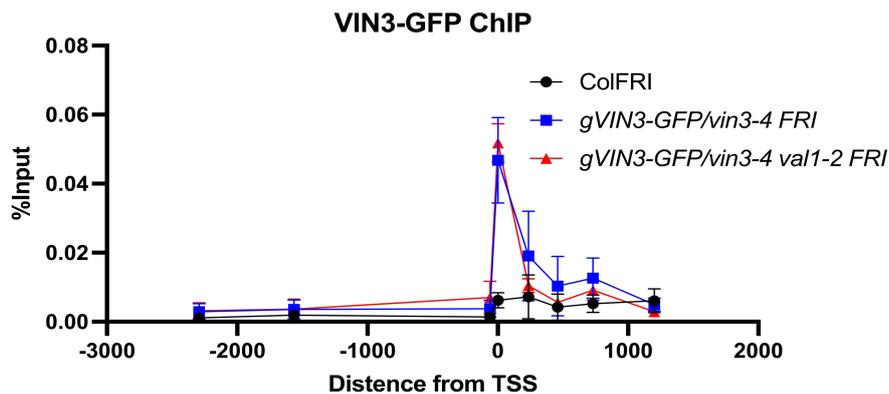


Figure 2-9 VIN3-GFP protein-ChIP in *val1-2 FRI* background

2.3 Discussion

How PcG complexes associate with their target sites is a major question in Polycomb biology. This chapter investigated the putative role of VIN3 in the association of the VEL-PRC2 complex with the nucleation region at

FLC. The results provide new understanding of the *in vivo* function of VIN3, but further work will be required to dissect the complex nature of PRC2 association with *FLC* and the precise function of VIN3.

2.3.1 Several chromatin-interacting domains could enhance the avidity of VIN3 to chromatin

Many chromatin proteins have been characterized as “reader” proteins by the direct pairing of a protein fold and a chromatin feature (DNA sequence, histone tail modifications, nucleosome features etc.) (Ruthenburg et al., 2007; Villaseñor and Baubec, 2021). Other chromatin-interacting proteins consist of multiple reader domains (Jain et al., 2020) that enable multivalent interactions, increasing avidity (functional affinity) and residence time (Ruthenburg et al., 2007). Here, we showed that the latter mechanism seems more relevant for VIN3. The PHD of the VEL proteins is surrounded by a ZnF and a 4- α -helix bundle with some similarities to a Winged-helix- and Bromo-domain. The structure shows that the three domains form a single structural unit, which we name the composite domain. The formation of a structural unit that consists of several chromatin-interacting domains has also been observed in other chromatin proteins (Savitsky et al., 2016). Multivalent interactions can thus be achieved not only through several independent reader domains in one protein, but also by larger folds that can engage with several chromatin features simultaneously.

The ZnF-PHD arrangement of the CD is similar to that found in UHRF1, a protein required for targeting a DNA methyltransferase during replication (Hu et al., 2011). Similar to our observation for the CD (Dr E. Franco-Echevarria unpublished), the ZnF is involved in stabilising the PHD fold and it is not directly involved in histone tail binding (Hu et al., 2011).

2.3.1.1 The atypical VIN3 plant homeo-domain

Prior to our work, the VEL proteins were mainly characterized with respect to their PHD, which had been connected to chromatin association (Sung and Amasino, 2004). The majority of characterized PHDs recognize either an unmethylated or trimethylated H3K4 residue, while a few have also been reported to bind H3K9me3 (Musselman et al., 2009) and H3K36me3 (Shi et al., 2007). It has previously been suggested that the PHD of VIN3 preferentially binds H3K9me2, as well as H3K4me2 and unmodified H3 tails (Kim and Sung, 2013, 2017a). This could be rationalized as making sense functionally because as active transcription decreases, H3K4me3 decreases over the *FLC* promoter and the nucleation region and H3K4me2 slightly increase (Yang et al., 2014). H3K9me2 has similarly been reported at the nucleation region during vernalization, as well as at the 3' end of *FLC* (Bastow et al., 2004; Swiezewski et al., 2007). H3K9 methylation is associated with DNA replication, where positions of replication fork stalling can become marked with H3K9 methylation (Feng et al., 2019). Recently, the recruitment of PRC2 components in mouse embryonic stem cells has been connected to the cell cycle (Asenjo et al., 2020). However, while DNA replication or the progression of the cell cycle have been shown to be required for spreading H3K27me3 over *FLC* following the return to warm (Hyun et al., 2013; Jiang and Berger, 2017; Yang et al., 2017), the involvement of replication in the initial nucleation event seems unlikely due to nucleation being roscovitine independent (Jiang and Berger, 2017; Yang et al., 2017).

With our increased understanding of the CD domain and its behaviour *in vitro*, we tested whether we could repeat the previously published histone binding. However, using ITC assays, no clear binding was detected to any histone tails tested, including H3K4 and H3K9 methylation (Dr E. Franco-Echevarria unpublished).

The finding that the PHD of the VELs is positively charged, in contrast to the majority of known PHDs, suggests that histone tails, which are also positively charged, are not the likely binding partner of the PHD. However, histones do not exist alone. The chromatin is occupied by negatively charged nucleic acid and indeed it has been suggested that RNA neutralizes the positive charge of histone tails (Dueva et al., 2019). Thus, we cannot confidently rule out an interaction between the PHD in the CD and chromatin, simply on the basis of a charge repulsion argument. A PHD with positively charge surfaces is found in the animal autoimmune regulator AIRE (Gaetani et al., 2012). Similar to the case presented for the PHD of the VELs, the positively charged surface is used to explain the finding that this PHD does not bind histone tails (Gaetani et al., 2012). This PHD is instead involved in protein-protein interactions (Yang et al., 2013b), as is the second PHD domain in the animal PRC2 accessory protein, PHF1 (Liu et al., 2018). Interestingly, work in the Dean group has shown that the PHD of the CD of VRN5 is likely to be involved in protein-protein interaction with core PRC2 components (E. Franco-Echevarria unpublished). Further work is therefore required to investigate the role of the PHD in VIN3, and its putative protein interactor(s). In chapter V, the generation of stable transgenic lines is discussed and I hypothesize that IP-MS of a VIN3 Δ PHD line could reveal potential direct interactors.

2.3.1.2 A conserved 4-helix bundle could link VIN3 to nucleic acid

Downstream of the PHD is a conserved region that forms a 4- α -helix bundle with similarities to a WH- and bromo-domain. Intriguingly, a potential WH downstream of the PHD parallels the domain arrangement in the mammalian Polycomb-accessory proteins (PCLs), where it is involved in DNA binding (Choi et al., 2017; Li et al., 2017). Whilst bromodomains are mostly known for their binding to acetylated histone tails, several bromodomains have also been implicated in DNA binding (Weaver et al., 2018). We believe this is the case for the WH/bromo-like domain of the CD, with non-sequence-specific binding to nucleic acids,

similar to that reported for PHF1 (Choi et al., 2017). In our EMSAs, we estimate that the affinity of the VIN3 CD for nucleic acids is in the micromolar range, similar to the WH of PHF1 (Choi et al., 2017; Li et al., 2017). This relatively low affinity does not support a model where the accessory proteins on their own target PRC2 to the nucleation region. Indeed, PRC2 alone shows a high affinity for DNA, only slightly enhanced by PHF1 association (Choi et al., 2017). Rather, this low affinity supports a model where the accessory proteins enhance the methyltransferase activity of PRC2 at the chromatin, either through conformational changes of PRC2 or by increasing residence time at the chromatin (Chen et al., 2020; Laugesen et al., 2019; Højfeldt et al., 2019). Further experiments would focus on whether mutations of the putative nucleic-acid binding interface interfere with VIN3 protein association and/or H3K27me3 accumulation at the nucleation region.

The EMSAs in this chapter were mostly performed with the Arabidopsis VIN3 CD, but for some experiments it was necessary to express the VIN3 CD from *Phoenix dactylifera* because of the toxicity of the WH/bromo-like domain in *E. coli*. Comparison of the two CDs suggests that they adopt the same structure (E. Franco-Echevarria unpublished). Where comparison was possible, the two domains behaved similarly in EMSAs, increasing our confidence that the Phoenix CD is a good tool for characterizing the biochemical features of the CD. Despite these apparent similarities, we wished to investigate whether the CDs were interchangeable *in vivo*. We therefore created stable transgenic lines where the VIN3 CD of Arabidopsis was replaced with the Phoenix CD. Because of the proximity of the FNIII domain to the CD, we also included a line where we swapped both the CD and FNIII (Chapter 5). At the time of writing this thesis, these lines were about to be analysed by other members of the Dean lab.

2.3.2 VIN3 recruitment

At the start of this work, we hypothesized that VIN3 possessed the ability to direct VEL-PRC2 to the nucleation region. While we cannot rule out that the CD recognizes some untested specific sequence, structure, or histone modifications, it seems unlikely that the paradigm established for the mammalian Polycomb PCLs is relevant to VIN3 (Laugesen et al., 2019). This spurred on our whole plant *in vivo* analysis.

2.3.2.1 VAL1 – the Pho equivalent

The ChIP-seq of VIN3 showed that VIN3 has many additional targets in Arabidopsis, consistent with the VELs being general Polycomb accessory proteins and not only involved in vernalization and flowering time control. However, substantial additional work is required to validate and uncover the importance of VIN3 at these multiple sites. When we compared our ChIP-seq with the published protein ChIP-seq of SWN/CLF and VAL1/VAL2, we could identify some overlaps in target genes, potentially genes where the recruitment of PRC2 is like the mechanism at *FLC*. As VAL1 is the only known sequence-specific factor involved in nucleation at *FLC*, it is tempting to speculate that VAL1 is a central component in the nucleation of VEL-PRC2 at *FLC* during vernalization. Interestingly, a recent study reports that VAL1/VAL2 is needed for PRC2 recruitment in Arabidopsis through direct interaction between VAL1 and the core PRC2 component SWN (Yuan et al., 2020). Similarly, another recent study has shown that the tethering of VAL1 is enough to trigger H3K27me3 in an engineered reporter system in Arabidopsis transgenic lines (Baile et al., 2021). In addition, co-immunoprecipitation transfection experiments in mammalian cells show that VIN3 interacts with VAL1 (E. Franco-Echevarria unpublished). VAL1 or core PRC1 components have not been identified in our proteomics of VIN3, but neither has SWN been detected in VAL1 proteomics. This would suggest that the interaction is transient, rather than representing steady state protein complexes *in vivo*. VAL1 interacts with a range of different proteins (Qüesta et al., 2016; Yuan et al., 2020, 2016; Zeng et al., 2020), potentially acting as an assembly platform,

co-ordinating co-transcriptional repression and chromatin silencing (Mikulski et al submitted). A *vall* mutation did not reduce VIN3-GFP occupancy based on ChIP, but this could be explained by redundancy with VAL2, similar to the relationship between Pho and Pho-Like in *Drosophila* (Brown et al., 2003). Consistent with at least partial redundancy between VAL1 (Pho) and VAL2 (Pho-like), the association of the PcG accessory protein Scm is unaffected when Pho is depleted by RNAi in *Drosophila* (Wang et al., 2010).

Analysis of the *vall-2* mutant further supports the idea that VAL1 is not strictly required for VIN3-PRC2 activity at *FLC*. Firstly, H3K27me3 still accumulates in *vall-2*, although to a less extent than WT (Qüesta et al., 2016). Therefore, *FLC* expression does not reactivate in *vall* as it does in the *vrn* mutants (Qüesta et al., 2016; Yang et al., 2017). Similarly, *vall* does not affect the rate of the fast VIN3-dependent shutdown of *FLC* expression during cold exposure (Hepworth et al., 2020). Thirdly, genetic analysis has shown that *vin3* and *vall* are additive (Qüesta et al., 2016), suggesting that *vin3* is not fully dependent on *VAL1*. This is consistent with our protein ChIP results. However, all these observations could be explained by redundancy with *VAL2*. To better understand the role of VAL1/2 in VIN3 nucleation, we aim to test VIN3 nucleation in a line carrying the *cis* mutation that completely blocks H3K27me3 accumulation during vernalization. We will use the transgenic *FLC-C585T/flc-2* line that carries a transgene with the mutation between the two RY motifs (Qüesta et al., 2016).

VAL1/VAL2 may connect co-transcriptional repression and chromatin silencing so it will be interesting to understand why PRC2 mainly associates with *FLC* chromatin over the first exon, i.e., 5' to the VAL1 binding site. Interestingly, using an artificial recruitment system, H3K27me3 was found to accumulate in the transcription unit next to VAL1 binding, suggesting that PRC2 binding may require RNAPII transcription and/or RNA production. This is consistent with the

increasing understanding of the fact that RNA tethers PRC2 to its genomic targets (Alecki et al., 2020; Long et al., 2020; Yu et al., 2019). In mammalian stem cells, RNA has been shown to be essential for the binding of PRC2 to chromatin and the deposition of H3K27me3 (Long et al., 2020). This puts into context the finding that in Arabidopsis, RNAPII stalls genome-wide in the early part of introns and this stalling event is splicing dependent (Kindgren et al., 2020). This is intriguing given our proteomic data showing co-immunoprecipitation of many splicing and transcription elongation factors with VIN3.

2.3.2.2 Paused RNAPII as a signal for VIN3-PRC2 recruitment?

Our CLNIP-MS of VIN3 and VRN5 included many factors involved in RNAPII transcription, particularly components involved in RNA splicing. This included most of the components of the NTC. A possible link between co-transcriptional RNA processing and PRC2 regulation at *FLC* through the autonomous pathway is an active area of study in the Dean lab. The proximal polyadenylation of *COOLAIR* leads to PRC2-mediated repression of *FLC* expression. Recently, proximal termination of the *FLC* sense transcript has been linked to the establishment of a low *FLC* expression state (Schon et al., 2021), thus connecting non-elongating/terminating RNAPII and Polycomb silencing. Interestingly, in other systems, paused RNAPII is also linked to PRC2-silenced genes (Liu et al., 2017). Therefore, the nucleation region could be induced by cold-enhanced RNAPII pausing or stalling around the first splice site on the sense transcript or introns in the *COOLAIR* antisense transcript. A connection between antisense transcriptional read-through and cold-dependent gene regulation has also been established for *SVALKA* in repressing *CBF1* expression (Kindgren et al., 2018).

Of note here is the fact that a PhD thesis from 2019 suggested a similar model and showed by qPCR-based assay that *FLC* and other Polycomb repressed genes in Arabidopsis accumulate proximal nascent RNA that

remains associated with the chromatin (Mermaz, 2019). By using ChIRP-MS, several components of the NTC complex were identified as interacting with the proximal RNA accumulating from *FLC* (Mermaz, 2019). These were found to interact with VIN3 and VRN5 in my experiments. Collectively, this shows that looking at the *FLC* nucleation region from either the protein or chromatin view identifies splicing factors, particularly the NTC complex, strengthening the support for this model.

Interestingly, the splicing-associated mammalian RNA binding protein RBFox2, which interacts with PRC2, preferentially binds chromatin-associated RNAs near gene promoters (Wei et al., 2016). This is consistent with the observation of short RNAs at the 5' end of Polycomb repressed genes in mammalian cells involved in PRC2 association and gene repression (Kanhere et al., 2010). Previous studies on *Drosophila* have suggested a similar mechanistic link between paused RNAPII and PRC1 (Grossniklaus and Paro, 2014). However, from steady-state experiments it is difficult to untangle cause or consequence of stalled RNAPII or Pc silencing.

These observations are relevant for *Arabidopsis* as a recent study has shown how H3K27me3 and unmodified H3K4 recruit a protein complex that contains the RNAPII phosphatase CPL2. *Arabidopsis* CPL2 prevents the accumulation of Ser5-P on the CTD of RNAPII, a modification that is associated with initiating RNAPII (Zhang et al., 2020b). This phenomenon appears to be a common mechanism in Polycomb repression across organisms (Fan et al., 2021; Wiles et al., 2020). These studies nicely link Polycomb repression to the regulation of RNAPII transcription dynamics. It is therefore intriguing to hypothesize a tight feedback mechanism: H3K27me3 not only blocks productive RNAPII transcription, but also non-productive RNAPII transcription enhances PcG association to chromatin, thus increasing H3K27me3 deposition.

2.3.2.3 Open Chromatin mediates PcG association

The genomic region where RNAPII transcription initiates is often associated with more accessible chromatin. The occurrence of the nucleation region around the TSS may indicate that the PRC2 simply associates with open chromatin, where transcription is reduced. This would fit with PRC2 binding with the highest affinity to naked DNA (Wang et al., 2017). Thus, the availability of free linker DNA between nucleosomes may be important for longer PRC2 association with chromatin. Similar conclusions have been made from *in vivo* analysis where regions associated with PRC2 after transcriptional silencing have been shown to possess more accessible DNA (Riising et al., 2014). Furthermore, it has been shown that Su(z)12 binding and nucleosome density are mutually exclusive at PRC2 target sites (Riising et al., 2014). Interestingly, during vernalization the nucleosomes around the VIN3 binding site become stabilized (Finnegan, 2015; Mikulski et al., 2021). It has similarly been observed that VRN1 is involved in stabilizing the nucleosomes around the TSS (Mylne et al., 2006, D. Zhu unpublished). The *vrn1* and *vin3* mutants are non-additive, implying that they work in the same genetic pathway (Greb et al., 2007). This suggests a model where, during early cold, a chromatin state with fixed nucleosomes and potentially a distinct conformation is established, at least partially through the activity of VRN1. This state is then required for VIN3 to bind to *FLC*. To test this, VIN3 protein-ChIP could be performed in a *vrn1* background. A similar hypothesis has been made for PRC2 in other organisms, where increased residence time of PRC2 is achieved in regions with longer nucleosome-depleted regions (Choi et al., 2017).

2.3.3 VEL proteins working upstream of PRC2

In our VIN3 and VRN5 proteomic analysis, we identified the corepressor protein TPL and its homologue TOPLESS-like. From our proteomics, we cannot conclude whether the interactions are direct or indirect. However, the VRN5-TPL interaction has been reported to be direct (Collins et al., 2019). Supporting this interaction, we have previously reported that VIN3

and VRN5 interact with the TPL-interacting protein HDA19 (Qüesta et al., 2016). Similar to these observations, the PcG protein P55(MSI) has been shown to recruit the HDAC Sin3 to mediate H3K27me3 and histone deacetylation genome-wide in the fungus *Magnaporthe oryzae* (Wu et al., 2021). Together, these observations suggest that the VEL proteins, in addition to bringing in PRC2 for epigenetic repression, also play a role in the transcriptional repression of *FLC*. Consistently, *vrn5* has been shown to have reduced activity in the transcriptional shutdown of *FLC* (Hepworth et al., 2020). Thus, a VRN5-TPL-HDA19 complex at *FLC* could maintain transcriptional repression until the locus is stably repressed. Analysis in our lab shows that loss of HDA19 does not influence *FLC* downregulation during vernalization (P. Mikulski unpublished). This could be explained by redundancy between the different HDACs in Arabidopsis and indeed a recent paper suggests that *HDA9* and *HDA19* both work to repress *FLC* (Zeng et al., 2020). It will be interesting to study the role of the VEL and its associated proteins in the transcriptional downregulation of *FLC*, in addition to their better-known role in the epigenetic repression pathway.

2.4 Summary

This chapter studied the mechanism behind the association of VIN3 with the nucleation region at *FLC*. Characterization of the N-terminal region showed that it contains a composite domain which consists of a ZnF, PHD and atypical-WH/bromo-like domain. Further characterization suggested that the CD, likely through the atypical-WH/bromo-like domain, can associate with nucleic acids in a non-sequence specific manner. *In vivo* analysis showed that despite being known for its role in the repression of *FLC* during vernalization, VIN3 potentially has many other targets in the Arabidopsis genome. Proteomic analysis suggested that VIN3 not only associates with PRC2, but also with proteins known to interact with core PRC1 components as well as proteins involved in transcriptional repression. I further showed that the loss of VAL1 is not sufficient to disrupt VIN3 chromatin association, but this analysis now needs to include a time course and be undertaken using the mutation of the RY motif.

Reinforcing that a simple protein-protein interaction model does not account for VIN3 recruitment, I found that VIN3 interactors included many components of the splicing machinery. VIN3 may therefore be associated with nascent promoter-proximal RNA originating from stalled RNAPII at the *FLC* nucleation region around the 5' spliced site of intron 1 (Fig. 2-10). This link to promoter-proximal stalled RNAPII could explain the location of the nucleation region around the Exon1-Intron1 boundary. Additional work is required to understand this intriguing connection between unproductive transcription and the association of PcG proteins.

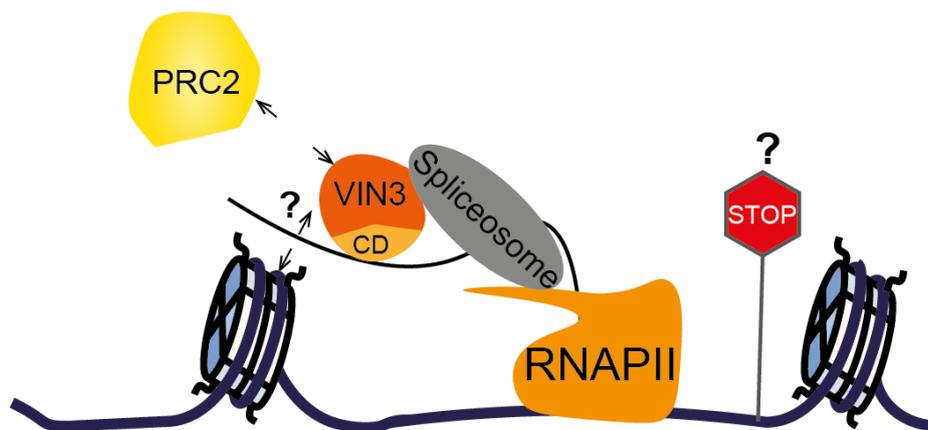


Figure 2-10 Hypothetical model for recruitment of VIN3, and later PRC2, to the nucleation region.

Chapter III

3 A novel polymerization domain involved in the repression of *FLC*

The following chapter discusses the characterization of the polymerization feature of the VEL domain in the VEL proteins. The *in vitro* work was performed in close collaboration with Dr Marc Fiedler, and I spent time during my second year based at MRC, LMB Cambridge. My individual work included the initial gel filtration, yeast-2-hybrid analysis, and most of the cloning involved in the initial VEL mutagenesis that led to the discovery of the polymerization-blocking mutants. The chapter then discusses the role of this domain *in vivo*, firstly in transient assays using *N. benthamiana*, which I did exclusively myself, and then in stable transgenic Arabidopsis plants. I initiated the stable transgenic work and was then joined by post-doc Dr Anna Schulten. We undertook the many transformations and analyses of transgenic lines in close collaboration. I was trained by Dr Silvia Costa in confocal microscopy to obtain the *in vivo* pattern of nuclear GFP proteins.

3.1 Introduction

A central goal in the Dean lab is to understand the key events underlying the epigenetic switch at *FLC* (Song et al., 2012). We hypothesize that interaction between the VEL proteins through the VEL domain is an important step in this switching mechanism (Lövkvist et al., 2021 accepted). Biological aggregation can be the result of biomolecular condensates, which have recently been shown to drive a diverse range of cellular processes (Boeynaems et al., 2018). These condensations contribute to increasing functional affinity – a concept called avidity – thus enabling proteins to be recruited and held at their specific gene targets (Bienz, 2020). In the previous chapter, we found that the composite domain of VIN3 has a low affinity for nucleic acids. Interestingly, the C-terminal region of the VEL proteins contains the novel VEL domain, which is involved in protein-protein interaction between the VEL proteins (Greb et al., 2007). This could therefore provide a mechanism to increase the local concentration of VIN3 and increase avidity for the *FLC* nucleation region. Studies from other Polycomb systems have shown that multimerization of PRC2 with its accessory protein PCL can increase the PRC2 complex's affinity for DNA at PRC2 target sites (Chen et al., 2020). While many biomolecular condensates have been found to form through phase transitions involving intrinsically disordered domains, true structural polymerization has so far only been shown for two structural folds: the DIX/PB1 fold and the SAM fold (Bienz, 2020). The SAM domain is found in a wide variety of proteins, interestingly including Polycomb proteins of other organisms. In *Drosophila*, the SAM domain is found in the Polycomb Repressive Complex 1 proteins Sfm1b, Scm and Ph, and it has been suggested that it forms a bridge between the DNA-binding protein Pho and PRC1, as stated earlier (Fig. 1-6) (Frey et al., 2016).

This chapter focuses on how polymerization, achieved through defined protein-protein interaction of the VEL domain in the VEL proteins, could potentially increase avidity for chromatin binding and thus contribute to

PRC2 targeting at *FLC*. *In vitro* characterization revealed that the VEL domain has intrinsic head-to-tail polymerization properties. Head-to-tail polymers are very difficult to work with *in vitro* because their polymerization makes them highly insoluble. By mutagenising residues required for this polymerization, we could produce *in vitro* recombinant protein and this enabled generation of VEL protein crystals and the structure to be solved by Dr Marc Fiedler. The structure revealed that the VEL domain presents a unique polymerization fold. From the structure we designed further mutations that cleanly prevented polymerization *in vitro*. We introduced them into the full-length Arabidopsis VEL proteins, and into Arabidopsis *vin3* and *vrn5* plants, with the goal of defining the role of VIN3 and VRN5 polymerization in *FLC* repression during vernalization.

3.2 Results

As discussed earlier, the C-terminal part of the VEL family proteins contains a novel domain that is involved in protein-protein interactions between the VEL proteins (Greb et al., 2007; Sung et al., 2006). Over-expression of VEL proteins in Arabidopsis leaves results in the formation of distinct GFP foci (Greb et al., 2007). However, over-expression of proteins often causes foci formation that is not relevant for the mechanism of the protein at endogenous concentrations (Alberti et al., 2019). It is nevertheless intriguing to think that the protein-protein interaction mediated by the VEL domain could help with the recruitment and dwell time of VEL proteins at *FLC*. To investigate the role of the VEL domain, we performed an in-depth characterization of the VEL domain *in vitro* and *in vivo*.

3.2.1 The VEL domain is a four-helical domain distinct from other known polymerization domains

To understand the function of the VEL domain, we performed *in-silico* prediction of its secondary structure. Using Phyre2 we predicted that the VEL domain is likely to consist of four α -helices (Fig. 3-1) (Kelley et al.,

2015). We tried to model the 3D structure of the VEL domain. However, we were unable to build reliable and consistent models using available software because four α -helices regions are present in a wide range of other proteins, folding into a variety of different structures.



Figure 3-1 Alignment of the VEL domain in the Arabidopsis VEL proteins. Amino acid sequence alignment of the VEL domain from Arabidopsis VIN3, VEL1, VEL2, VEL3 and VRN5. The consensus secondary structure predicted with Phyre2 is shown beneath with green barrels indicating α -helices.

The functional role of any protein-protein interaction between the VEL proteins therefore needed to be established experimentally. Inhibition of protein-protein interaction can be achieved through different methods, one of which is deleting the interaction domain. However, such an approach is crude and does strictly not address the function of the protein-protein interaction, as the domain could have other roles in addition to mediating the interaction between the VEL proteins. We ideally wanted to find single residues in the VEL domain that are crucial for the interaction. However, the lack of obvious structural homologues made it impossible to predict the interface. I therefore worked with Marc Fiedler and began an in-depth characterization of the VEL domain by expressing the domain heterologously in *E. coli*. I expressed the C-terminal part of VEL1, VIN3, and VRN5 N-terminal fused to a glutathione S-transferase (GST) tag. The fusion proteins were purified by affinity purification. Following binding of the GST fusion proteins to the matrix, the proteins were eluted by thrombin cleavage, releasing the native VEL protein domain. The eluted proteins were concentrated using Amicon Ultra filters. Purified protein was analysed using analytical Size-exclusion chromatography (SEC) to test the quality and size of the purified protein. By comparing this with the elution volume of proteins with known sizes, we were able to estimate the size of the eluted VEL domain. The size estimation is affected by the overall fold of the protein, as globular and disordered proteins run differently through the column. When the C-terminal part of VRN5 was

assayed by gel filtration it eluted earlier than expected, but crucially after the void volume, indicating that the C-terminal part of VRN5, which includes the VEL domain, forms oligomers *in vitro* (Fig. 3-2A). This observation is consistent with previous transient overexpression assays *in planta*, which showed the formation of large foci indicative of protein aggregation (Greb et al., 2007).

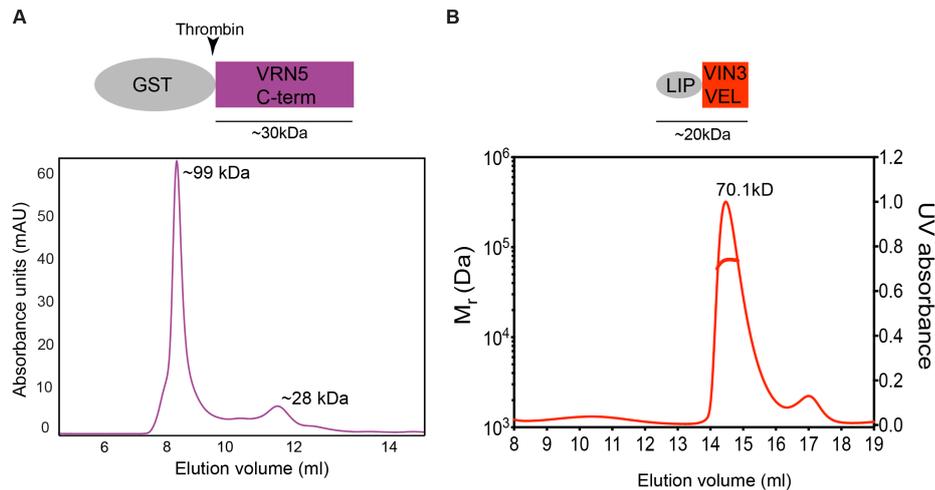


Figure 3-2 Gel filtration of the VEL domain. **A)** Size-exclusion chromatography of affinity purified GST-VRN5_C-term. The construct used is shown on top with the expected size of the purified protein. The sizes on the elution profile indicate estimated sizes based on elution volume. **B)** SEC-MALS of purified Lip-VIN3_VEL (from M. Fiedler). The construct used is shown on top with the expected size of the purified protein.

The preliminary data indicated that the VEL domain, at least that of VRN5, could polymerize/aggregate *in vitro*. This spiked our interest in the characterization of the VEL domain, as only two other structural folds have been characterized as polymerization domains – the DIX/PB1 and the SAM domain (Bienz, 2020). The VEL domain did not seem to have a structure similar to either of these characterized structural folds, meaning that we had the potential to characterize a novel protein fold important for polymer formation.

When studying protein domains *in vitro*, the chosen boundaries of the domain influence the behaviour of the pure protein. For example, the inclusion of unstructured regions is disadvantageous for crystallization because they prevent the ordered packing of protein molecules. The initial constructs used for gel filtration contained the entire C-terminal region of

the VEL proteins, which in addition to the four α -helices also contains a region that was predicted to be unstructured. For the structural analysis of the VEL domain, we cloned a minimal VEL domain fused to a Polyhistidine-tag and a lipoyl (Lip) domain. The Lip domain originates from the pyruvate dehydrogenase complex from *Bacillus stearothermophilus* (Packman et al., 1988). It enhances the solubility of proteins that are prone to aggregation (Lebendiker and Danieli, 2014). It can additionally act as an intramolecular chaperone, helping the folding of the fusion protein (Zou et al., 2008). This is similar to the function of the GST tag in the initial constructs. The new constructs were used like the GST constructs, purified by affinity purification using a Nickel column, and eluted proteins were used for gel filtration. In accordance with our initial observations, the VEL domains of VIN3, VEL1 and VRN5 eluted earlier than expected based on the estimated size of the fusion protein. The gel filtration profile of Lip-VIN3_VEL is shown in Fig. 3-2B. Similar to the initial GST-VRN5_Cterm protein, we observed a sharp peak from proteins that eluted earlier than expected, as well as the presence of a small fraction that eluted later, which likely represent a small monomeric fraction. With these constructs, we observed a sharp peak with a tailing right shoulder, indicative of oligomers of various sizes, similar to that observed for the other known polymerization domain (Fiedler et al., 2011; Kim et al., 2016).

Gel filtration was also used as a purification step, and following further purification and concentration, the protein samples were used in crystallization trials. Initially, we obtained tiny crystals with the LIP-VIN3_VEL constructs under several different conditions. However, probably due to the heterogeneity of polymers, the crystals failed to grow to satisfactory sizes ideal for structural determination, a common problem in the structural determination of polymerization domains. To overcome the problem of polymerization for crystal formation, it was necessary to identify point mutations that would block polymerization. However, the lack of structural models meant that it was difficult to predict the interface.

Thus, we used two different approaches: an unbiased Yeast-2-Hybrid (Y2H) screen and a more targeted approach based on the conservation of residues in the VEL domains.

3.2.1.1 Screens revealed crucial residues for VEL-VEL interaction

Previously we had used Y2H to show that VEL domains from different VEL proteins can interact (Greb et al., 2007). We used this as a foundation to identify residues involved in VEL domain oligomerization. The VEL domain of VIN3 was subjected to error-prone PCR and cloned into the activation domain (AD) containing Y2H plasmid and combined with the WT VEL domain of VIN3, VRN5 and VEL1 fused to the DNA binding domain (BD) (Fig. 3-3A). Yeast clones that were unable to grow on selective media (SD-LWH) indicating a loss of interaction, were isolated, and the mutation(s) in the VEL domain was identified by Sanger-sequencing. The screen was designed to potentially reveal mutations that would specifically block homo or hetero interaction with either VEL1 or VRN5. In addition, by focussing on mutations that only affected one or two of the interactions but not all of them, we would in theory increase the ratio of true interaction mutation over mutation leading to misfolded protein or nonsense mutations. However, since we did not know if the interaction surface(s) is shared between VEL domain homo- and heterointeraction, we also sequenced mutations that blocked all interactions.

A wide range of mutations throughout the VEL domain were identified. Some of the uninteresting ones included likely structural mutations that cause mis-folding of the VEL domain; others introduced an early stop codon. Based on the amino acid mutated we ranked them by their likelihood to affect the polymerization and then tested the top candidates by gel filtration.

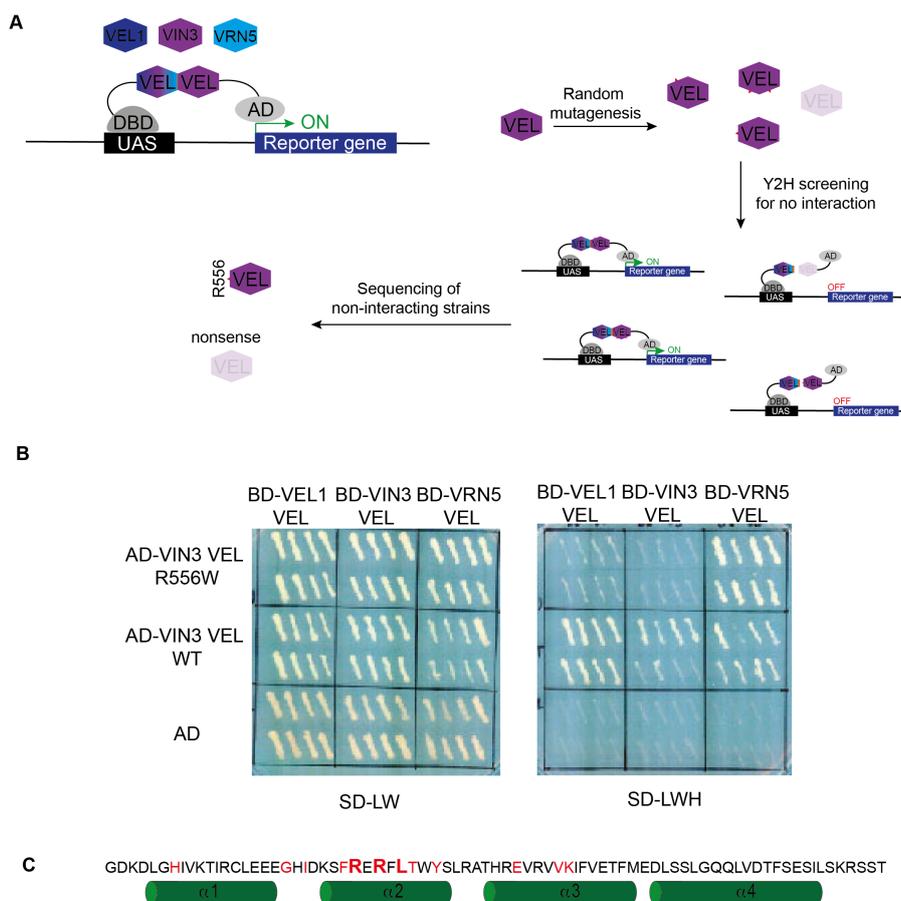


Figure 3-3 Yeast-2-hybrid screen for interaction mutants. **A**) Y2H reporter line for VEL-VEL interaction. The VEL domain from VEL1, VIN3, and VRN5 was fused to the DNA-binding domain (DBD) and tested for interaction with the VEL domain from VIN3, which was fused to the activation domain (AD). The VIN3 VEL domain was random mutagenized and screened for loss of interaction whereafter the mutation(s) was identified by sequencing. **B**) Y2H interaction between VIN3 VEL R556W and the other VEL domains (In collaboration with Dr Yaoxi Li). WT VIN3 VEL and AD-only serves as positive and negative controls respectively. **C**) Amino acid sequence of the VIN3 VEL domain with the secondary structure indicated underneath. Residues identified in the Y2H assay to be able to cause non-homo-interaction are marked in red. The residues involved in VIN3 VEL polymerization in the crystal structure are highlighted in bold.

The Y2H screen was combined with a screen where we systematically mutated all of the residues in the VEL domain. The mutations changed the amino acid to alanine, caused a charge swap, or resulted in opposite amino acid features. All the mutations were made in our standard bacterial His-Lip-VIN3_VEL expression construct. The mutated VEL domains were expressed in *E. coli*, His- purified and gel filtrated in batches. Expression and gel filtration showed that the mutations fell into three main categories, some caused misfolding and eluted with the void volume after gel filtration

or were not expressed in *E. coli*, some showed no effect on polymerization, and a few gave rise to smaller oligomers. Interestingly, both the Y2H assay and the systematic mutation screen identified R556 as an important residue for VIN3 VEL polymerization (Fig. 3-3 B). It is worth noting that in the Y2H assay the R556 mutation (R556W) did block interaction with VEL1 and VIN3 but not VRN5. In addition, the neighbouring R554 residue was also identified using both approaches to be able to prevent interaction (Fig. 3-3 C). From these combined approaches, we identified two residues that gave rise to smaller oligomers R554A and R556A/R556D (numbers based on VIN3 position) (Fig. 3-4) and were, therefore, very likely to be in the putative head or tail interface. Further refinement showed that the R554A/R556D double point mutant gave rise to almost only monomeric protein based on SEC-MALS analysis (Fig. 3-4). Protein samples of the VEL domain containing the R554A/R556D gave much better crystal formation and eventually enabled Dr Marc Fiedler to solve the crystal structure of the VEL domain.

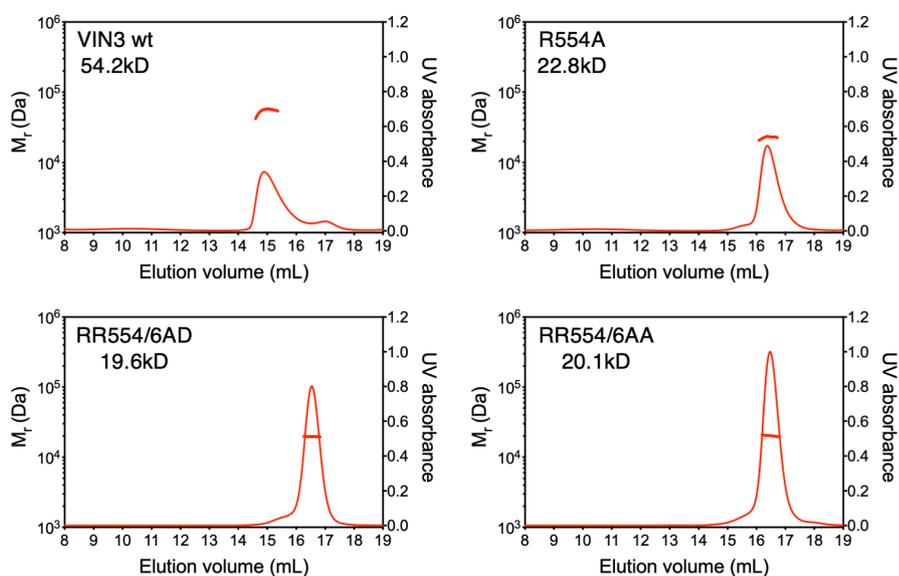


Figure 3-4 SEC-MALS of VIN3 VEL domain mutants. SEC-MALS of purified wt or mutant Lip-VIN3 VEL domain at same concentration (2 mg/ml). Monomeric protein expected to be 20 kDa. (Obtained from M. Fiedler).

3.2.1.2 The VEL domain a novel polymerization domain

This paragraph provides some of the results of the structural determination of the VEL domain done by Dr Marc Fiedler. The structure provides

important insights into the interaction of VEL proteins vital to understand and discuss some of the following plant data and is therefore included here. The crystal structure of the VEL domain revealed that the domain folds into a four helical bundle (Fig. 3-5A) as we had predicted from the sequence. The actual fold revealed a globular fold with antiparallel helices.

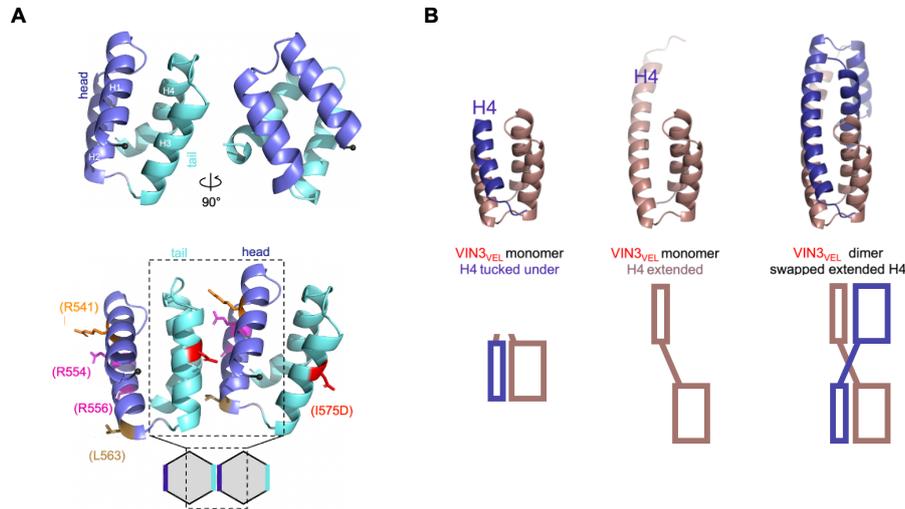


Figure 3-5 VEL domain structure. **A)** Ribbon diagram of the VEL domain monomer showing the head (blue) and tail (cyan) interfaces. The interface in the VEL polymer is shown beneath with the residues involved in head-to-tail interactions shown. **B)** Two conformations of the VIN3 VEL domain, the monomer with helix 4 (H4) tucked under or with H4 extended. The domain swapped dimer of two monomers with H4 extended shown on the right. Schematic drawing of the different conformations is shown below. (Reproduced from Fiedler et al.)

The VEL structure revealed two surfaces involved in the interaction between individual VEL proteins. Because of the organization of the VEL proteins into a head-to-tail polymer, we refer to these interfaces as the head and tail interface. An important finding from the VEL structure is that the head mutant R554 not only participates in VEL-VEL interaction but also in forming the core of the VEL domain. It is, therefore, possible that R554A is also affecting the stability of the VEL domain. The crystal structure revealed that several of the residues identified in the Y2H assay as being able to block the VEL-VEL interaction were indeed part of the polymerization interface (Fig. 3-3 C). This finding demonstrates the usefulness of the Y2H approach in the absence of structural information.

The crystal structure enabled the design of additional head and tail mutations that disrupt polymerization. The residues involved in polymerization are well conserved in the VEL domains. This shows that these residues, and thereby polymerization, are important for VEL domain function. The residues we had previously focussed on (R554/R556) were both located in the head surface (Fig. 3-5). From the structure, the I575 residue in the tail surface was identified and when mutated to aspartic acid, polymerization was blocked to the same extent as with the R554A/R556D mutations (M. Fiedler unpublished). An interesting observation in the crystal structure of the VIN3 VEL domain was that the minimal unit that made up the crystal polymer was a VIN3 VEL dimer, formed through so-called “domain-swapping”. Domain swapping is the phenomenon where two (or more) proteins form multimers through interwinding parts, where the exchanging part can be an entire domain, or as in the case of VIN3, only a part of a domain (Rousseau et al., 2003). An important concept in domain-swapping dimers is that the dimers adopt the same conformation as the individual monomers, with the exception of the hinge that connects to the exchanging parts. This hinge often folds back on itself in the monomer or extends in the domain-swapped conformation (Rousseau et al., 2003). In the case of VIN3, the fourth helix (H4) from one molecule swaps over and interacts with the third helix of the second molecule, and the same happens in the opposite direction for H4 from the second molecule (Fig.3-5 B). As seen in Fig. 3-5B, the dimer follows the definition of domain swapping; the two units in the dimer adopt the same conformation as the monomer. The hinge either allows H4 to fold back on itself or is extended when H4 is involved in domain swapping. This form of dimerization is also observed in the DEP domain of the Wnt signalling component Dishevelled, and it is required for functionality (Gammons et al., 2016). This kind of interaction was only observed for the VEL domain of VIN3 and not VEL1, the other VEL domain for which we obtained the crystal structure (M. Fiedler unpublished). Interestingly, the linker region between H3 and H4 differs in a central residue between VIN3 and VEL1 (Fig. 3-1). VIN3 contains a small leucine residue whereas VEL1 contains

a rigid proline at this position. We speculated that this difference is the causative factor that allows the H4 in the VIN3 VEL domain to extend. We therefore made the L584P mutation in VIN3 to block domain swapping, similar to how a glycine-to-proline mutation in the DEP domain in Dishevelled blocks domain swapping, to confer reduced signalling (Gammons et al., 2016). The additional mutations designed, based on the *in vitro* work to test the importance of VIN3 VEL domain polymerization for *FLC* repression, are discussed in more detail in Chapter V.

3.2.2 The VEL domain is required for VIN3-VRN5 interaction

We had previously shown that VIN3 can co-immunoprecipitate VRN5 when both are heterologously expressed in *N. benthamiana* leaves (Greb et al., 2007). According to Y2H assays, this interaction is dependent on the C-terminal region of the VEL proteins (Greb et al., 2007). However, the initial Y2H assay was based on deleting everything downstream of the FNIII domain, meaning that we could not rule out the possibility that the linker region between the FNIII and VEL domains is involved in the interaction. I therefore used Co-Immunoprecipitation (Co-IP) of transiently expressed VIN3 and VRN5 containing different deletions to map the minimal region required for the interaction (Fig. 3-6A). Consistent with previous observations, VIN3 was able to co-immunoprecipitate VRN5 (Fig. 3-6B). Supporting the idea that the VEL domain is required for interaction between the VEL proteins, deletion of the VEL domain in VIN3 was enough to strongly reduce the Co-IP with HA-VRN5 (Fig. 3-

6B). The same could be observed when the reciprocal experiment was performed using a mutant version of VRN5 as the bait (Fig. 3-6C).

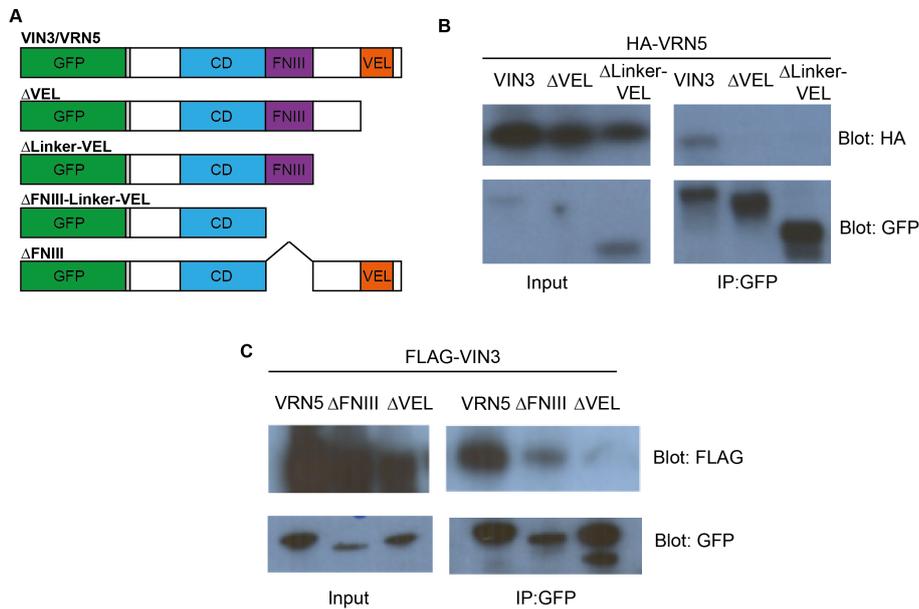


Figure 3-6 The VEL domain is required for interaction between VIN3 and VRN5 in *N. benthamiana*. **A)** Schematic drawing of the constructs used in *N. benthamiana* Co-Immunoprecipitation (Co-IP) experiments. **B)** Co-IP of WT and mutant GFP-VIN3 with HA-VRN5 when co-overexpressed in *N. benthamiana* leaves. **C)** Co-IP of WT and mutant GFP-VRN5 with FLAG-VIN3 when co-overexpressed in *N. benthamiana* leaves.

From our *in vitro* characterization, we predicted that the R554/R556 mutation would block the interaction, mimicking deletion of the VEL domain. Surprisingly, when we made the R554A/R556A double point mutation, we observed no effect on the interaction between VIN3 and VRN5 through Co-IP when the proteins were overexpressed in *N. benthamiana* (Fig. 3-7). Later, we obtained the crystal structure that revealed the other interface between the VEL domains. This made it clear that the head R554A/R556A mutant maintains the availability of the tail interface for interaction, which can explain the observed maintained interaction (Fig. 3-7). This is consistent with the observation that VIN3 homo-interaction is not blocked when only one of the proteins carries the R554/R556 mutations in mammalian Co-IP experiments (Fiedler et al., unpublished). Another explanation could be that the VIN3-VRN5 interface is different from the VIN3-VIN3 one; to support this, we noted that the

VIN3 R556 single mutation also maintains interaction with VRN5 in Y2H but not with VEL1 or VIN3 (Fig. 3-3B).

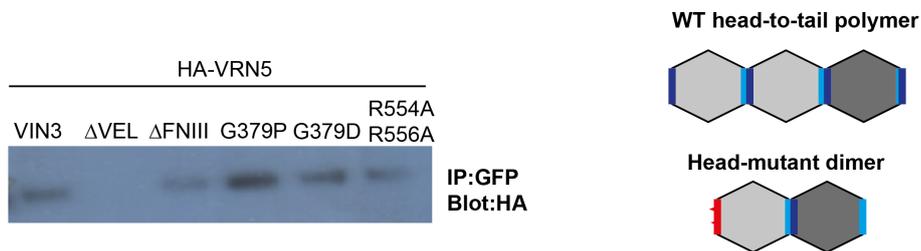


Figure 3-7 Point mutations do not block interaction of VIN3 and VRN5. Co-IP of WT and mutant GFP-VIN3 with HA-VRN5 when co-overexpressed in *N. benthamiana* leaves. Model for the observed maintained interaction in the head mutant R554A/R556A /D, light grey represents the VIN3 VEL domain and dark grey the VEL domain of VRN5.

3.2.2.1 The FNIII domain is not required for VEL-VEL interaction

Chapter II focused on the N-terminal part of the VEL proteins, the CD, whereas the previous sections of this chapter have focused on the C-terminal VEL domain. However, the function of the FNIII domain in the VEL proteins remains elusive. FNIII domains are found in many proteins, where their functions remain to be characterized. However, some FNIII domains have been reported to be involved in protein-protein interaction, including the FNIII domain of a rice VEL protein, which has been reported to interact with the Suz12 homologue EMF2 (Jeong et al., 2016). At the beginning of this project, before we had obtained the structural information of the VEL domain, we hypothesized that the FNIII domain could be involved in the cross-linking of VEL polymers through dimerization, similar to the “Dishevelled paradigm” (Bienz, 2020). The Dishevelled paradigm, which is also true for other proteins, builds on the observations of how Dishevelled condensates forms. These condensates form through head-to-tail polymerization of the DIX domain, DIX polymers are then cross-linked through dimerization of the DEP domain (Bienz, 2020; Gammons et al., 2016; Schwarz-Romond et al., 2007). We therefore initially speculated that the FNIII domain could be a functional equivalent of the DEP domain.

The literature contains several examples of FNIII domains that form dimers (Moore et al., 2017; Schumacher et al., 2013). The initial gel filtration of *in vitro* expressed FNIII domains suggested that FNIII from the VEL proteins could dimerize as well, although the dimerization was not as obvious as the polymerization behaviour of the VEL domain. In *Drosophila*, the DEP domain dimerization of Dishevelled mediates strong interaction, enabling efficient co-immunoprecipitation of Dishevelled proteins, even in the absence of the polymerization DIX domain (Gammons et al., 2016). We therefore investigated whether the FNIII domain in VIN3 contributed to the interaction with VRN5. In our experiments, when both proteins were overexpressed in *N. benthamiana* leaves, the interaction between VIN3 and VRN5 is maintained when the FNIII domain is deleted (Fig. 3-6 C, Fig. 3-7, and Fig. 3-8).]

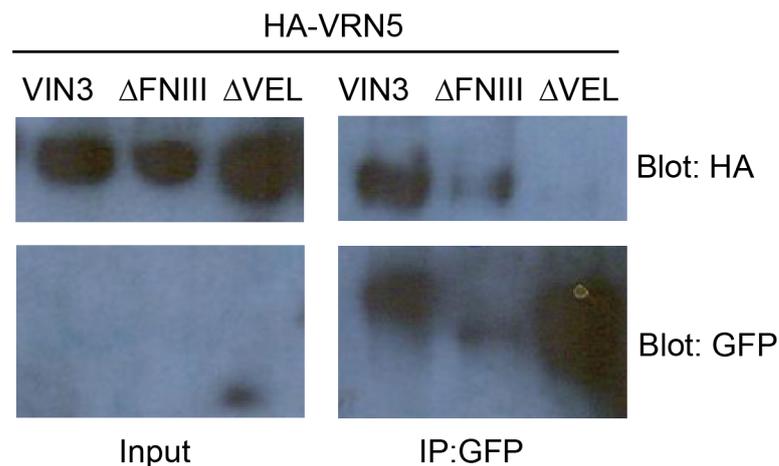


Figure 3-8 VIN3 FNIII domain is not required for VIN3-VRN5 co-immunoprecipitation. Co-IP of WT and mutant GFP-VIN3 with HA-VRN5 when co-overexpressed in *N. benthamiana* leaves

An important observation from the expression of GFP-VIN3_ΔFNIII in *N. benthamiana* was the repeatedly low levels of protein, potentially because of destabilization of VIN3 when the FNIII domain was deleted. The interpretation of the interaction between VIN3_ΔFNIII and VRN5 is therefore complicated by the apparent destabilization of the protein.

As mentioned above, FNIII domains can be involved in dimerization. Furthermore, FNIII domains have been reported to form dimers through

domain swapping (Hu et al., 2007; Teplyakov et al., 2014; Wojcik et al., 2010), like the DEP domain of Dishevelled (Fig. 3-9 A). Interestingly, the putative hinge region in the FNIII domains of the VEL proteins is conserved. The putative hinge region contains the GxxP motif except for VRN5 where the flexible glycine is replaced by a less flexible aspartic acid (Fig. 3-9 B).

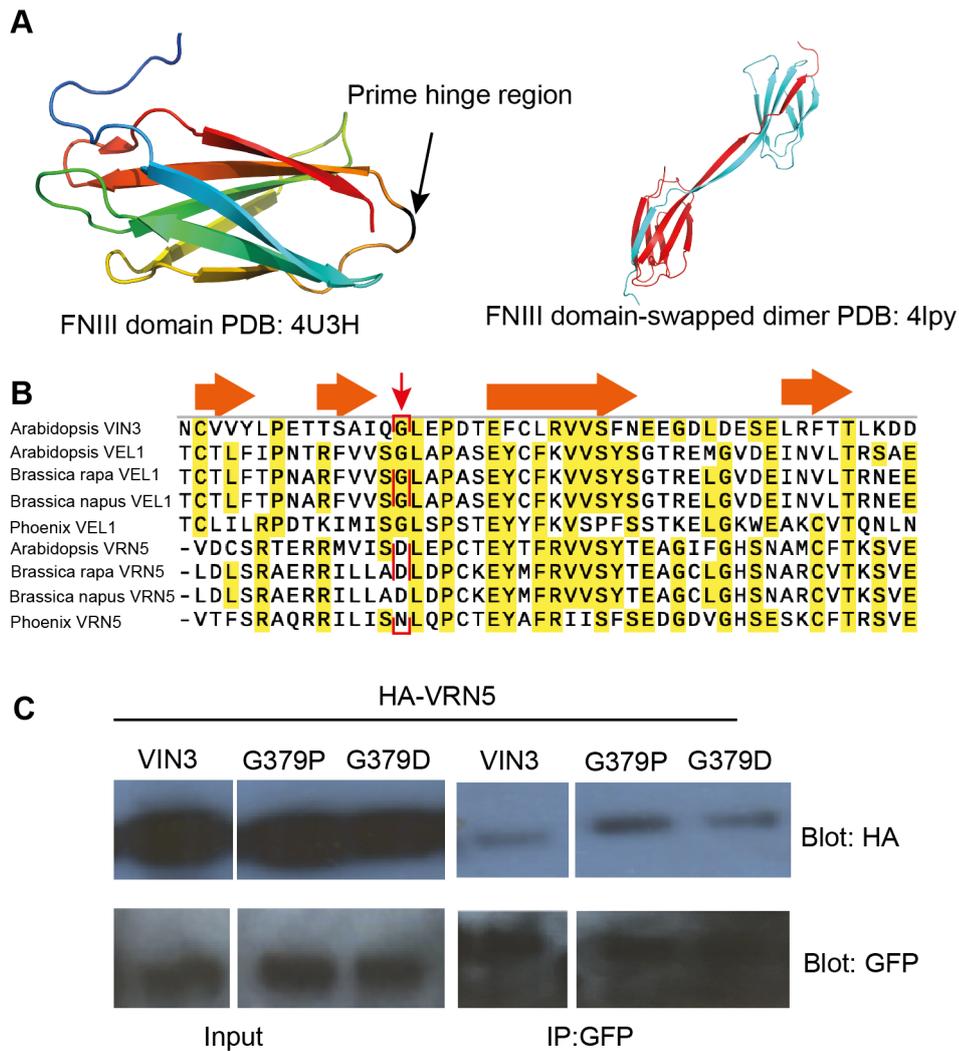


Figure 3-9 FNIII mutations do not affect co-immunoprecipitation of VIN3 with VRN5. **A)** Left: The crystal structure of the monomeric FNIII domain from Porebski et al. (2015) (4U3H), with the prime domain-swap hinge glycine residue marked in black. Right: One of the domain-swapped FNIII dimer conformations from Teplyakov et al. (2014) (4lpy). **B)** Alignment of the last region of the FNIII domain of the VEL proteins from Arabidopsis, Brassica and Phoenix, secondary structure prediction is shown above with arrows symbolising β -sheets. The red arrow indicates the hinge glycine residue, highlighting the conserved difference between VIN3/VEL1 and VRN5. **C)** Co-IP of WT and mutant GFP-VIN3 with HA-VRN5 when co-overexpressed in *N. benthamiana* leaves.

To test whether a flexible hinge in FNIII is important for the interaction, we made constructs where the glycine residue in VIN3 was replaced by either the VRN5 specific aspartic acid or a rigid proline residue. The more rigid proline has been shown to stop domain swapping of the DEP domain in Dishevelled (Gammons et al., 2016). These mutants showed no effect on the hetero-interaction between VIN3 and VRN5 (Fig. 3-9 C). Importantly, the hinge mutations did not affect the protein stability of VIN3, making the interpretation of the Co-IP results more straightforward. In conclusion, multiple analyses showed that the main interaction between VIN3 and VRN5 relies on the VEL domain. However, we cannot rule out the possibility that the FNIII domain has a more important role during, for example, homo-interaction, or that the interaction between FNIII domains is too weak to be revealed by Co-IP analysis, like the DIX-DIX interaction in Dishevelled. Lastly, as is reportedly the case for the rice VRN5 homologue, FNIII could be involved in protein-protein interactions with other proteins (Jeong et al., 2016). Overall, the observations in *N. benthamiana* were consistent with the results obtained through a similar approach using mammalian HEK cells (M. Fiedler unpublished data).

3.2.2.2 The linker region in VIN3 is required for nuclear localization

The deletions of protein domains can potentially affect interactions by changing the localization of proteins. I therefore also checked for nuclear localization of the different deletion constructs using standard stereomicroscopy. Observation of the GFP expression showed that the linker region between the FNIII and VEL domain is required for nuclei

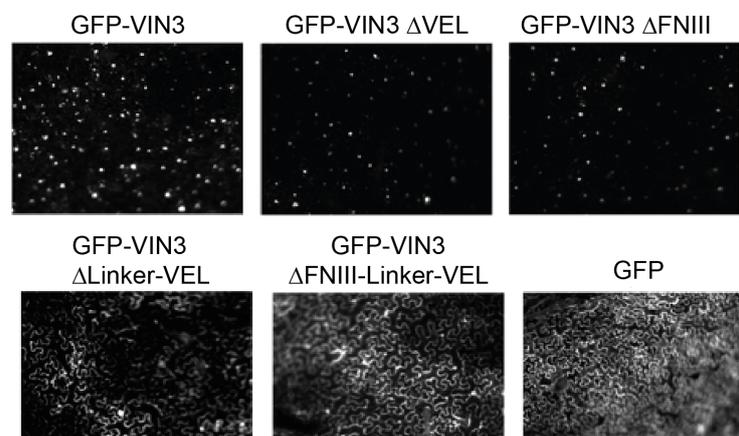


Fig. 3-10 Subcellular localization of WT and mutant GFP-VIN3

localization of VIN3, as both the Linker-VEL and the FNIII-Linker-VEL deletions showed cytoplasmic localization (Fig. 3-10). However, VEL and FNIII-only deletions maintained nuclear localization. This is consistent with the prediction of a nuclear localization signal just upstream of the VEL domain in VIN3. Later analysis showed that the same behaviour can be observed in VIN3 expressed in mammalian cells (E. Franco Echevarria unpublished data). This means that the loss of interaction observed for the Δ VEL construct is not due to the separation of proteins in different cellular compartments.

3.2.3 VIN3 head mutation R556D in Arabidopsis

In the previous sections we focused on the intrinsic property of the VEL domain to polymerize; however, the relevance of VEL domain polymerization for the function of the VEL proteins remained an open question. To address the role of the polymerization of VIN3 in the repression of *FLC* during vernalization, we introduced the polymerization mutant R556D into a *VIN3* transgene in an *Agrobacterium* binary vector and stably transformed Arabidopsis *vin3-1* mutants. Due to the time frame of making stable transgenic lines in a late flowering background, we tested the R556D mutations before we had obtained the structure, as the *in vitro* data looked promising for R556D being able to block polymerization, as discussed above. As described in Chapter V, the mutated VEL proteins were cloned with a C-terminal fluorescent tag. A 10-amino-acid linker was introduced between the end of the VEL protein and the fluorescent tag in order to minimize the putative influence of the fluorescent tag (GFP) on the polymerization of the C-terminal VEL domain. Our *in vitro* work showed that as with the other polymerization domains (Sayou et al., 2016; Schwarz-Romond et al., 2007), VEL polymerization is concentration dependent, and a high concentration of VIN3 VEL R556D can overcome the polymerization deficiency induced by the R556D mutation (M. Fiedler, personal communication). Therefore, to analyse the role of the polymerization of the VEL proteins in the vernalization response *in planta*, we selected transgenic lines that had only one transgene insertion to ensure

that the expression levels of the *VIN3* or *VRN5* transgenes were close to endogenous levels.

We first confirmed the expression and nuclear localization of VIN3_R556D-eGFP with confocal microscopy. Seeds from Basta-selected and PCR-confirmed T1 individual transgenic plants were used. We initially obtained T2 seeds from three single-copy individual transgenic lines of VIN3 WT and six single-copy individual transgenic lines of VIN3_R556D. We monitored the levels of *FLC* spliced and unspliced, as well as the *VIN3* transcript levels in non-vernalized seedlings (NV), after six weeks of cold treatment (6WT0), and after six weeks of cold treatment followed by 10 days of growth in warm conditions. This enabled us to monitor any potential effect on the reactivation of *FLC* expression.

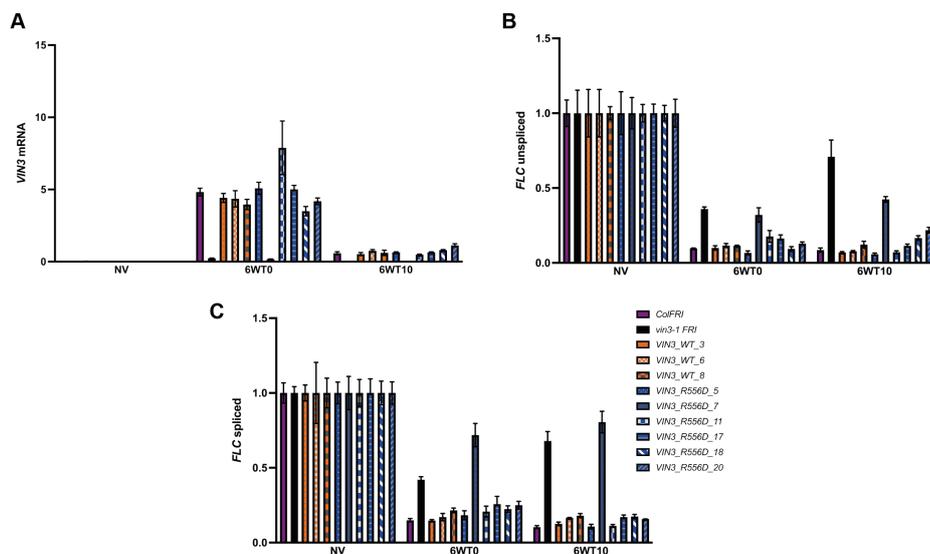


Figure 3-11 FLC vernalization responds in *VIN3*_R556D. A) *VIN3* mRNA levels B) *FLC* unspliced C) *FLC* spliced measured in non-vernalized (NV) seedlings, after six weeks of cold treatment (6WT0), or after six weeks of cold treatment followed by 10 days of warm growth (6WT10). Data were normalized to the geometric mean of UBC and PP2A. For *FLC* unspliced and spliced levels were further normalized to the NV levels. Error bars represent SEM (n = 3).

All transgenic *VIN3* lines, both WT and R556D, showed patterns of *VIN3* RNA expression similar to those of endogenous *VIN3* (Fig. 3-11 A). We detected very little *VIN3* mRNA before the exposure to cold (NV), relatively high levels after vernalization (6WT0), and low levels of

expression post-cold (6WT10). One exception was the VIN3-R556D line 7, which showed no expression of *VIN3* (Fig. 3-11 A), consistent with microscopy analysis where we were unable to observe any GFP signal.

The three individual *VIN3 WT* lines all complemented the *vin3-1* mutant and showed similar *FLC* repression during and after vernalization to the non-transgenic control ColFRI (Fig. 3-11 B-C). The VIN3 R556 mutation did not seem to affect *FLC* repression after 6WT0 or after 6WT10 as all lines showed similar repression of *FLC* to the WT transgenic controls (Fig. 3-11 B-C). Overall, this analysis showed that a single mutant in the head-surface of the VEL domain is not enough to cause a detectable effect on *FLC* repression, despite its *in vitro* polymer deficiency.

3.2.4 VIN3 forms distinct nuclear foci

The ability of the VEL domain to form polymers *in vitro* prompted us to look at the *in vivo* pattern of VIN3-eGFP in the nucleus. After vernalization, VIN3-eGFP appears diffuse with small foci present in the nucleus (Fig. 3-12). Interestingly, initial observations of the polymer mutant R556D and the full VEL deletion revealed a similar pattern. This raises the question of whether the pattern observed for VIN3-eGFP was VIN3 specific or like that of other nuclear proteins. To test this, we performed side-by-side imaging of CLF-GFP, SWN-GFP, VAL1-GFP and HTR5(H3.3)-GFP. In addition, FCA-GFP was included as a positive control, as its pattern in the nucleus shows clear GFP foci (Fang et al., 2019). Importantly, the pattern observed for VIN3-eGFP is not a general feature of nuclear protein, as both SWN and HTR5 show a different pattern. However, the pattern of GFP-CLF and VAL1-GFP is somewhat

similar to VIN3-GFP, suggesting that several chromatin proteins are forming small foci in Arabidopsis root nuclei (Fig. 3-12).

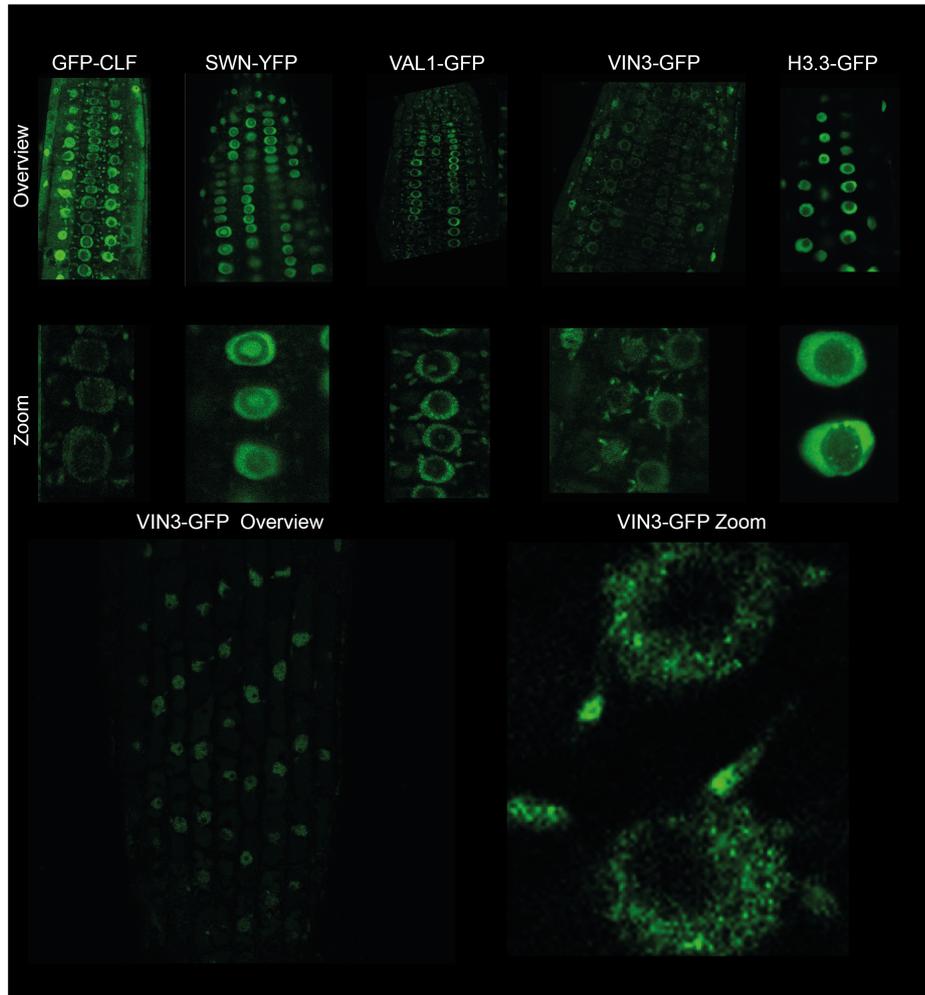


Figure 3-12 VIN3-GFP forms small foci in root nuclei. Overview and zoomed-in images of root nuclei expressing GFP-CLF, SWN-YFP, VAL1-GFP, VIN3-GFP and H3.3-GFP. Two larger images of VIN3-GFP are shown below to highlight the appearance of small VIN3-GFP foci.

3.3 Discussion

In this work, we identified the VEL domain as a novel polymerization domain able to form head-to-tail polymers. Our hypothesis is that the polymerization property is likely to play an important role in the function of the VEL proteins. However, our work so far has not been able to uncover the precise function of polymer formation *in vivo*. Based on paradigms from other multimerization proteins, we can hypothesize potential functions for polymerization. The VEL polymer might form a

protein condensate of VEL-PRC2 molecules, thereby providing a locally high concentration of PRC2 at the nucleation region to facilitate trimethylation of H3K27 (Fig. 3-12). Trimethylation of H3K27 is the slowest methylation step of PRC2, as shown by *in vitro* biochemical measurement (Sneeringer et al., 2010). This is consistent with *in vivo* studies in mammalian cells which showed that the restoration of H3K27me3 after DNA replication or PRC2 inhibition is slow compared to the more rapid restoration of mono- and di-methylation (Alabert et al., 2015; Højfeldt et al., 2018). A conceptually similar mechanism involves polymerization being required for VIN3 to bind at the nucleation region; as I discussed in the previous chapter, VIN3's affinity for nucleic acids is relatively low. A high concentration of VIN3 protein might therefore be required to provide an avidity effect for nucleic acid binding and thus PRC2 localization at the nucleation region. However, as discussed in the previous chapter, some PRC2 components localize to the nucleation region even in warm-grown plants where *VIN3* is not expressed (Shu et al., 2019). Nevertheless, *VIN3* could be required for the initial association of PRC2 with the nucleation region of *FLC* in vernalization-requiring *Arabidopsis* accessions. Ongoing work in the Dean lab aims to test the requirement for *VIN3* in the binding of the PRC2 components to *FLC* to address this model.

Polymerization may be a generally important concept in epigenetic mechanisms. The function of the *Drosophila* Polycomb protein complex PhoRC relies on the polymerization of the SAM domain (Frey et al., 2016). It has been suggested that polymerization of the SAM domain containing proteins Scm and Ph acts as a platform for concentrating PRC1 and PRC2 (Fig. 1-5). This is similar to how we envisage the VEL polymer formed by *VIN3* and *VEL1*. An interesting parallel was recently reported in *in vivo* studies investigating the development of *Drosophila* germ cells. Here the SAM domain protein Scm is required to concentrate PRC2 and PRC1 at PREs, which is partly achieved through the induction of Scm (DeLuca et al., 2020). Applying the PhoRC model for *FLC*, increasing *VIN3* levels in the cold leads to the formation of a VEL polymer, probably together with

VEL1, which could act as a platform for VRN5-PRC2, helping to concentrate PRC2 at the nucleation region (Fig. 3-13). In addition, as mentioned in Chapter II, VRN5 can interact with proteins other than PRC2, such as the corepressor TOPLESS. Similarly, VIN3 and VEL1 could also interact with other proteins through regions other than the VEL domain, for example the FNIII and Composite domains. Therefore, a VEL polymer may act as an assembly platform not only for PRC2, but also for other proteins involved in the repression of *FLC*.

Another role of the SAM domains in animal PcG proteins is to mediate long-range interactions and the clustering of Polycomb targets. The clustering of Polycomb targets requires the polymerization ability of the SAM domain (Isono et al., 2013; Wani et al., 2016). This observation is very interesting as the *FLC* gene copies have also been shown to cluster together during vernalization (Rosa et al., 2013). By combining an *FLC-lacO* transgene with a *LacI-YFP*, it was shown that the *FLC-lacO* transgenes at each insertion site cluster together when the plants are exposed to cold (Rosa et al., 2013). This clustering follows the dynamics of H3K27me3 accumulation and it is dependent on *VRN2* and the VEL protein *VRN5* (Rosa et al., 2013). The clustering was not impaired in *lhp1*, the HP1 homologue in Arabidopsis required for spreading and maintaining H3K27me3 after cold, which otherwise is known to form subnuclear foci (Berry et al., 2017b; Rosa et al., 2013). Further supporting a role for the VEL-PRC2 complex in the clustering of *FLC* copies is the observation that clustering is slightly reduced after continued growth in warm conditions after vernalization. Likewise, the presence of the VEL protein VRN5 at *FLC* decreases slowly as plants are returned to warm after vernalization (Rosa et al., 2013; Yang et al., 2017). However, the deterministic factor that leads to clustering has yet to be established. The clustering observed in the cold could be related to several events that happen during vernalization, in addition to the presence of VEL protein, for example, transcriptional silencing, PRC2 nucleation or H3K27me3 deposition. The experimental uncoupling of VEL polymerization from PRC2, H3K27me3

or transcriptional silencing has currently not been achieved, so detangling the correlations to establish casual roles for clustering will be difficult. However, the observation that *VRN2* and *VRN5* are required for clustering but *LHP1* is not, hint to a role of the VEL-PRC2 complex in clustering, like the SAM domain in PRC1 in *Drosophila* (Fig. 3-12). It will be very interesting to monitor *FLC* clustering in our newly established polymer-deficient VEL protein lines.

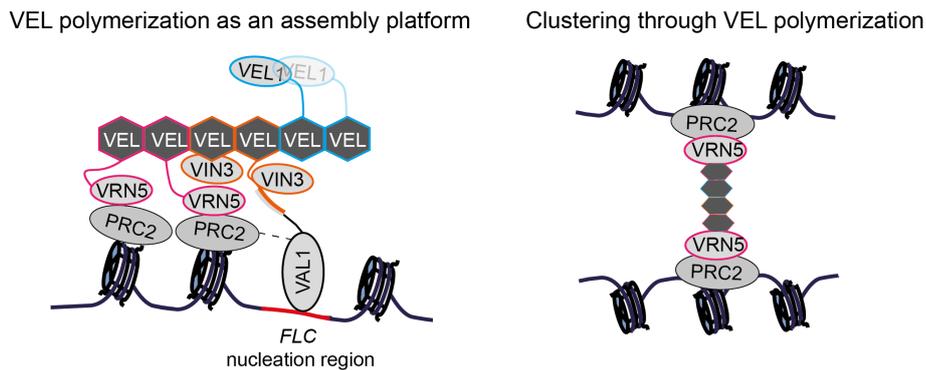


Figure 3-13 Speculative model for the function of VEL polymerization. VEL polymerization could act as a platform and help to concentrate PRC2 and other proteins that potentially interact with the VEL proteins. Another model could be that VEL polymerization helps to bring different VEL-PRC2-bound targets together, for example the two *FLC* loci.

3.3.1 Multimerization of PRC2 through domain swapping – A role for the FNIII domain?

In the previous chapter, we focussed on the binding ability of VIN3 to nucleic acids. Although our analysis showed that VIN3 has only a weak and non-specific affinity for DNA and RNA, it raised the question of whether VIN3 enhances the association of PRC2 to *FLC* during the cold, paralleling the function of the mammalian PRC2 accessory proteins (Choi et al., 2017; Højfeldt et al., 2019; Li et al., 2017). The discovery of the polymerization ability of the VEL domain suggests that an avidity effect could be mediated by VEL polymerization, as a possible mechanism to enhance chromatin binding of VEL-PRC2, as discussed above. A similar mechanism of oligomerization to enhance chromatin binding has been

suggested for a mammalian PRC2 complex (Chen et al., 2020), highlighting a potential functionally conserved mechanism. For the mammalian PRC2 complex, oligomerization is mediated by the dimerization of core PRC2 components and the dimeric structure is stabilized by the accessory protein PHF19 (PCL3) (Chen et al., 2020).

In the structural analysis of the VEL domain of VIN3, it was found that it can form dimers through domain swapping (Fig. 3-5 B). Domain swapping is an interesting property as the domain-swapped conformations are often metastable, due to a high energy barrier between the monomeric and dimeric form (Rousseau et al., 2003). This means that the interconversion between the two forms is slow. The repression of *FLC* during vernalization is a slow process that requires weeks of cold, and it is one where the individual *FLC* alleles switch independently in a digital manner between an ON and an OFF state (Angel et al., 2011; Berry and Dean, 2015; Berry et al., 2015). It is intriguing to speculate that the domain-swapping property of the VIN3 VEL domain is part of the molecular mechanism underlying this switch behaviour of *FLC* repression during vernalization. We hypothesize that a certain concentration of VIN3 must be achieved before the VEL domain undergoes domain swapping, explaining why long periods of cold are required. Switching from a monomer to a domain-swapped dimer is per definition a digital switch, so it could be that VIN3 nucleates at individual alleles and undergoes domain swapping when the threshold concentration is reached. This would explain why the two *FLC* copies have been observed to respond independently (Berry et al., 2015).

In contrast to the domain swapping observed for the VEL domain of VIN3, domain swapping in the mammalian PRC2 dimer occurs through the exchange of protein elements between the two core proteins Suz12 and RBBP4 (RBAP48) (Chen et al., 2020). More precisely, the domain swapping is mediated by the C2 domain of Suz12, which swaps to the other

PRC2 complex and interacts with RBP4 (RBAP48) (Fig. 3-14 A) (Chen et al., 2020).

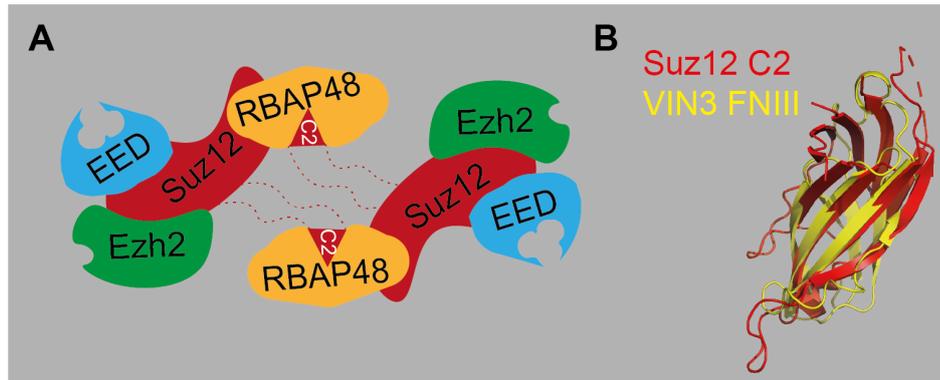


Figure 3-14 PRC2 dimerization through domain swapping. **A)** Domain-swapped dimer of human PRC2. The domain-swapped dimer is formed through the donation of the Suz12 C2 domain (Red) from one proteome to RBAP48 of the other proteome (based on Chen et al., 2020). **B)** The FNIII domain (yellow) has a fold similar to that of the C2 of Suz12 (PDB:5wai).

Interestingly, the C2 region is conserved between Suz12 and EMF2 but not VRN2 (Gendall et al., 2001). The C2 domain is formed by 7–8 β -sheets, whose secondary structure resembles that of the FNIII domain (Fig. 3-14 B). As introduced in the previous section, FNIII domains have been shown to form domain-swapped dimers and our analysis shows strong conservation of the potential hinge residues in the FNIII domains of the VEL proteins. This might indicate that domain swapping is an important feature of the FNIII domain in the VEL proteins. In our cross-linked immunoprecipitation-mass spectrometry (IP-MS) (Chapter II) and in others, native IP-MS, VIN3 and VRN5 only immunoprecipitated with VRN2, not with EMF2 (De Lucia et al., 2008). It is intriguing to speculate that the FNIII domain of the VEL proteins replaces the missing C2 of VRN2 and, together with VEL polymerization, helps to form oligomeric PRC2, which provides PRC2 with increased avidity for longer chromatin engagement. Furthermore, a PRC2 holocomplex containing a FNIII domain could be mutually exclusive with one that contains a C2 domain, because they both interact with MS11^{RBAP48}. This would explain why the VEL proteins appear to only be part of a VRN2-PRC2 complex.

3.3.2 Avidity through polymerization

The discovery of VEL-domain-mediated polymerization prompted us to propose a model where polymerization potentially provides an avidity effect for nucleic acid binding of VIN3. A similar model has been proposed for another important developmental regulator in plants. The *LEAFY* (*LFY*) transcription factor, which is essential for flower development, contains a SAM domain at its N-terminus (Sayou et al., 2016). Similar to what we envisage for the VEL domain, *LFY* proteins that carry a mutation in the polymerization domain show lower activity, although their activity is not completely dependent on oligomerization (Sayou et al., 2016). Experiments similar to those discussed in Chapter II showed that the SAM domain is not involved in DNA binding *in vitro*. However, *in vivo*, *LFY* polymerization mutants showed less binding to some genomic target sites (Sayou et al., 2016). Therefore, the model we propose for VIN3 shares the same paradigm as that proposed for *LFY*, where oligomerization provides avidity, enabling binding to genomic sites.

Our initial analysis of the polymerization mutant R556D in the *VIN3* transgenic lines surprisingly showed no effect on *FLC* repression or flowering time during and after vernalization. This unexpected result requires further investigation but our current hypothesis is that one mutation is not sufficient to block functional interaction *in vivo*. My *in vitro* analysis showed that the R556D mutations block polymerization but maintain the ability to interact with the other VEL proteins (as assayed by Co-IP in both mammalian cells and *N. benthamiana* leaves), likely through the maintained functional tail interface in the VEL domain (Fig. 3-7). This resembles observations for single interface mutants in other polymerization domain-containing proteins. In *LFY*, a single point mutation in the head interface, *lfy-22*, has very little effect (Levin and Meyerowitz, 1995). Only when both interfaces are mutated does *LFY* show reduced activity (Kim et al., 2016). To test whether this is the case for the VEL domain, we used the VIN3 structure to design a new double point mutant, where both interfaces are mutated, named VIN3 R556D/I575D.

The prediction would be that this mutant has slower *FLC* repression dynamics during vernalization.

Another possibility is that VIN3 polymerization is made slightly redundant by the polymerization of the other VEL proteins. *vin3* shows a clear late-flowering phenotype, showing that the VEL proteins are not strictly redundant; however, it is possible that the VEL domain itself is redundant. It could be imagined that in the VIN3 R556D background, VIN3 nucleates at *FLC* as normal, and a polymer is then built by VEL1. However, *VEL1* does not appear to be required for vernalization when analysed in an *fca-9* background (unpublished data), supporting the idea that at *FLC* the VEL proteins are only partly redundant. The analysis of the role of *VEL1* in the vernalization pathway is complicated by the observation that the combination of a *vell* mutant with an active *FRI* allele is lethal and results in plants that die a few days after germination (D. Zhu, unpublished). An additional complication is the cold upregulation of *VEL2* (Greb et al., 2007), as the presence of functional *VEL2* may be enough to compensate for both the *vell* and the *VIN3* R556D mutations. It will be interesting to investigate the effect of *VIN3* R556D in a *vell* and/or *vel2* background, although the lethality of *vell FRI* could complicate this analysis. Similar to this thinking, combining the weak *lfy-22* allele with other mutants revealed the compromised function of the polymer-deficient *lfy-22* allele (Levin and Meyerowitz, 1995).

3.4 Summary

In this chapter, we found that the C-terminal VEL domain of the VEL proteins is a polymerization domain with a novel protein fold. The VEL domain is required for the interaction between VEL proteins. The VEL domain forms concentration-dependent head-to-tail polymers, similar to other well-known polymerization domains. Consistent with the polymerization properties of the VEL domain, *in vivo* GFP-tagged *VIN3* forms small foci indicative of the formation of *VIN3* condensates.

Through *in vitro* characterization, we defined the key residues in the interfaces required for interaction. We made transgenic plants where we aimed to disrupt polymerization to analyse the requirement of VIN3 polymerization for *FLC* repression during vernalization. Unexpectedly, VIN3 lines that only carry mutations in the head interface behave very similarly to wild-type lines. The maintained functionality in R556D lines suggests that dimerization, potentially through domain swapping or partial redundancy between VEL proteins, provides enough VEL protein condensation to mediate the epigenetic switching at *FLC*.

Chapter 4

4 Chromatin conformation at FLC

This chapter discusses the study of chromatin looping at *FLC* in different genetic mutants. I was helped with the optimization of the 3C assay and the discussion of strategies to overcome the experimental variability by Dr Yusheng Zhao and other members of the Dean lab.

4.1 Introduction

While the previous chapters discussed the role of the VEL-PRC2 complex in the establishment and memory of the repressed state at *FLC* during and following vernalization, this chapter focuses on the different 3D conformations of the *FLC* locus and how the 3D conformation could act as another memory element for *FLC* expression.

In vivo, DNA is tightly wrapped around histone octamers to form nucleosomes, the smallest unit of chromatin (Luger et al., 1997). The nucleosome arrangement of chromatin is popularly referred to as beads on a string, giving the impression of a linear chromatin fibre. However, within the cell, the chromatin fibre adopts a compact 3D arrangement (Szabo et al., 2019). The folding of the chromatin fibre brings different genomic sequences into closer proximity than their genomic position would initially suggest. The organisation of the 3D chromosome fibre adds an extra layer

of epigenetic regulation beyond the histone code. Recently, with the increased understanding of biomolecular condensates and nuclear organization, the 3-D folding of chromatin has received a great deal of attention.

Within the nucleus, the genome is organized at different levels. At the overall level, chromosomes or regions within chromosomes engage in long-range interactions to form distinct compartments, known as the A and B compartments respectively (Lieberman-Aiden et al., 2009). The A compartment contains the gene-rich and actively transcribed euchromatin, while the majority of the B compartment consists of the stably repressed heterochromatin (Lieberman-Aiden et al., 2009). H3K27me3 marked genes can be part of both the A and the B compartment in mammalian cells. At the next level of organization, chromosomes fold into distinct sub-domains named TADs (topologically-associating domains) (Dixon et al., 2012). Metazoan TADs are sub-mega base domains where interaction between chromatin is more frequent within the domain than between domains (Dixon et al., 2012). Despite being present in many eukaryotes, the presence of TADs is less obvious in Arabidopsis. Instead, smaller chromatin loops seem to be more frequent in the Arabidopsis genome, which is similar to the situation in budding yeast (Liu et al., 2016; O'Sullivan et al., 2004). *FLC* was one of the first plant loci to be shown to have chromatin loops – the other being the *bl* locus in maize (Crevillén et al., 2013; Louwers et al., 2009). Consistently, genome-wide studies of contacts in Arabidopsis suggest that *FLC* does not interact with more distant regions (Hövel unpublished). It does, however, seem that chromatin looping is tightly coupled to the regulation of expression and Polycomb repression at *FLC* (Crevillén et al., 2013; Kim and Sung, 2017b; Li et al., 2018).

4.1.1 Chromatin loops involved in *FLC* regulation

4.1.1.1 The active gene loop

Short-range chromatin loops have been identified at *FLC*, and they are associated with both active and repressed states of *FLC* (Crevillén et al., 2013; Kim and Sung, 2017b; Li et al., 2018). In 2013, Crevillén et al. reported a chromatin loop that forms between the 5' and 3' end of *FLC* (Fig. 4-1), similar to many of the first described short-range chromatin loops in *S. cerevisiae* (Crevillén et al., 2013; O'Sullivan et al., 2004). In yeast, the formation of chromatin loops that juxtapose the promoter and terminator has been reported to play a role in transcriptional memory, with the presence of a loop mediating faster reactivation of gene expression (Lainé et al., 2009; Tan-Wong et al., 2009). The formation of a gene loop can also support a high expression state by facilitating quick RNAPII recycling from the terminator to the promoter (Lainé et al., 2009; Lykke-Andersen et al., 2011; O'Sullivan et al., 2004). Observations in yeast tightly link gene loop formation with RNAPII transcription. However, Crevillén et al. initially reported that, while the *FLC* gene loop was disrupted by cold exposure, it was not disrupted through mutations in genes that influence *FLC* transcription levels (Crevillén et al., 2013). Despite this, later studies have suggested that the relative frequency of *FLC* gene loop formation is linked to the expression level of *FLC* (Li et al., 2018), consistent with RNAPII being a key player in formation of gene loops, which are associated with active transcription.

4.1.2 The repressive intragenic loop

Approximately five years after the initial report of the gene loop at *FLC*, Kim and Sung reported a second chromatin loop at *FLC* (Kim and Sung, 2017b). This loop forms between the same 5' end region but links to the 3' end of intron 1 (Kim and Sung, 2017b) (Fig.4-1). The frequency of this smaller intragenic loop was found to be increased by exposure to prolonged cold (Kim and Sung, 2017b). In addition, this loop was more abundant in Col-0 than in ColFRI (Li et al., 2018). These observations linked the formation of the loop to the Polycomb repressed state of *FLC*;

supporting this is the fact that increased loop formation during vernalization was dependent on *VIN3* and *VRN5* (Kim and Sung, 2017b). The boundaries of the intragenic loop are interesting as it has been shown that the 5' end of *FLC* and the end of intron1/start of exon2 are sufficient to cause repression of a *FLC-GUS* reporter construct during vernalization, although the full intron 1 is required for stable repression (Sheldon et al., 2002). The originally described full gene loop has been connected to active transcription, with cold exposure associated with loop disruption and reduced *FLC* expression (Crevillén et al., 2013; Li et al., 2018). However, it can also be considered as a repressive loop, in that it helps to repress the antisense transcription. From this perspective, the gene- and intragenic-loop at *FLC* can be seen to work through the same mechanism. The chromatin loops could block RNAPII recruitment to the promoters of *COOLAIR* and *FLC* respectively, thereby helping to maintain a low expression state. Similar phenomena have been described for the *WUS* and *PID:APOLO* loci in *Arabidopsis*, where the formation of chromatin loops that include the promoters is associated with the repression of these loci (Ariel et al., 2014; Guo et al., 2018).

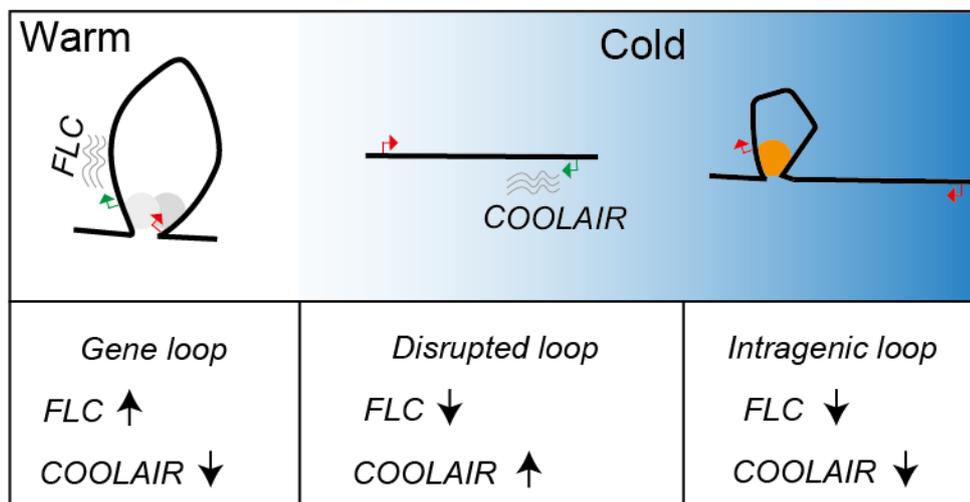


Figure 4-1 Chromatin loop dynamics at *FLC*. Prior to cold, *FLC* exists in a gene loop conformation that forms between the promoter and terminator. The gene loop correlates with high *FLC* sense transcription and low *COOLAIR* transcription. In early cold, the gene loop is disrupted and this correlates with upregulation of the antisense *COOLAIR* and downregulation of *FLC*. With increasing cold, the intragenic loop forms between the promoter and the end of intron 1, correlating with the low expression state of both sense and antisense.

4.1.3 Chromatin conformation capture

In 1993, Cullen and colleagues reported the interaction of the enhancer and promoter regions of the rat *prolactin* gene (Cullen et al., 1993) through a nuclear ligation assay. This signalled the beginning of the era of studying the 3D organisation of nuclear chromatin through ligation-based assays. Almost a decade later, Dekker et al. introduced the Chromatin Conformation Capture (3C) technique to study the interaction frequency between two different genomic locations (Dekker et al., 2002). Since then, many 3C-derived techniques have been reported – 4C, 5C, Hi-C and many others – which allow the study of genomic interactions at a genome-wide level (McCord et al., 2020).

As introduced above, a common theme in Arabidopsis genomic organisation is the existence of local loops limited to the region around or within transcriptional units (Liu et al., 2016). This chapter describes the efforts to describe and understand the behaviour of the 3D structure of the *FLC* locus. 3C is in theory a relatively straightforward approach to study specific genomic interactions, so in this research it was chosen as the experimental approach with which to study *FLC* loop biology. The 3C technique is summarized in the following section.

3C and its derived techniques rely on formaldehyde crosslinking to capture a snapshot of genome organisation (Dekker et al., 2002; Hagège et al., 2007; Louwers et al., 2009). Following nuclear extraction, DNA is digested with a restriction enzyme(s) of choice and then fragments are ligated together. Ligation is performed under dilute conditions to promote intramolecular ligation, although this requirement remains debatable (Sati and Cavalli, 2017). The resulting product is amplified and detected by qPCR assaying two regions of interest (here regions at *FLC*). Primers in tandem orientation are used for the PCR assay to minimize the risk of amplifying fragments resulting from inefficient digestion or cross-linked-independent ligation. The principle of 3C is outlined in Fig. 4-2.

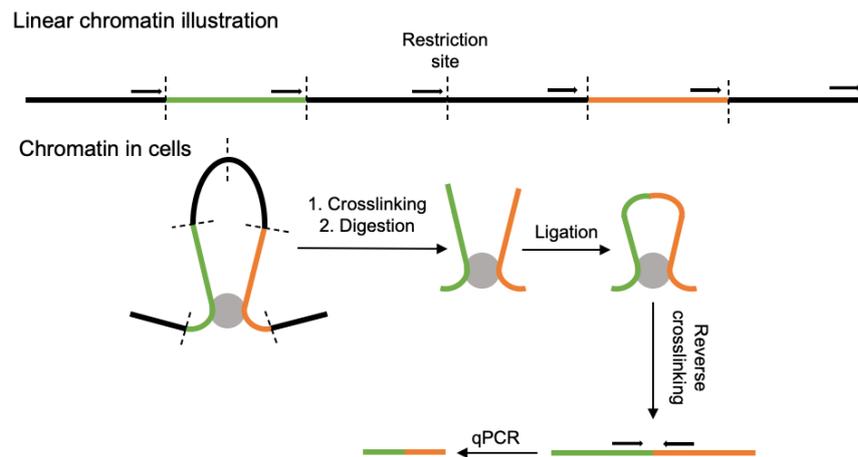


Figure 4-2 Principle of Chromatin Conformation Capture (3C). Chromatin is often illustrated as a linear molecule. However, within the nucleus, the chromatin is folded in 3D. 3C enables the study of regions that are in close proximity in 3D. In 3C, chromatin is cross-linked with formaldehyde and fragmented by restriction enzyme digestion. DNA is ligated under dilute conditions that favour intramolecular ligation. Cross-linking is reversed and DNA purified. The purified DNA is used as template for qPCR, in order to detect interacting fragments.

4.2 Results

4.2.1 At least two chromatin loops form at *FLC*

Previous studies of the 3D conformation at the *FLC* locus have revealed the existence of at least two different loops: a loop that forms between the promoter and terminator of *FLC* (Crevillén et al., 2013) and an internal loop between the promoter and the end of the long first intron (Kim and Sung, 2017b). It is well known that 3C in plants is a challenging technique¹ due to the enzymatic processes of digestion and ligation being performed in suboptimal conditions. This can lead to large variability between different replicates. A rigorous approach therefore needs to be applied to confidently support the existence of chromatin loops (Jamge et al., 2017). We therefore decided to validate the existence of the various *FLC* chromatin loops using different restriction enzymes (Jamge et al., 2017). In the original reports, the loops at *FLC* were detected using the restriction enzymes BglII/BamHI for the promoter-terminator loop and DpnII for the internal loop (Crevillén et al., 2013; Kim and Sung, 2017b).

¹Federico Arial, Maïke Stam and Franziska Turck's lab personal communication

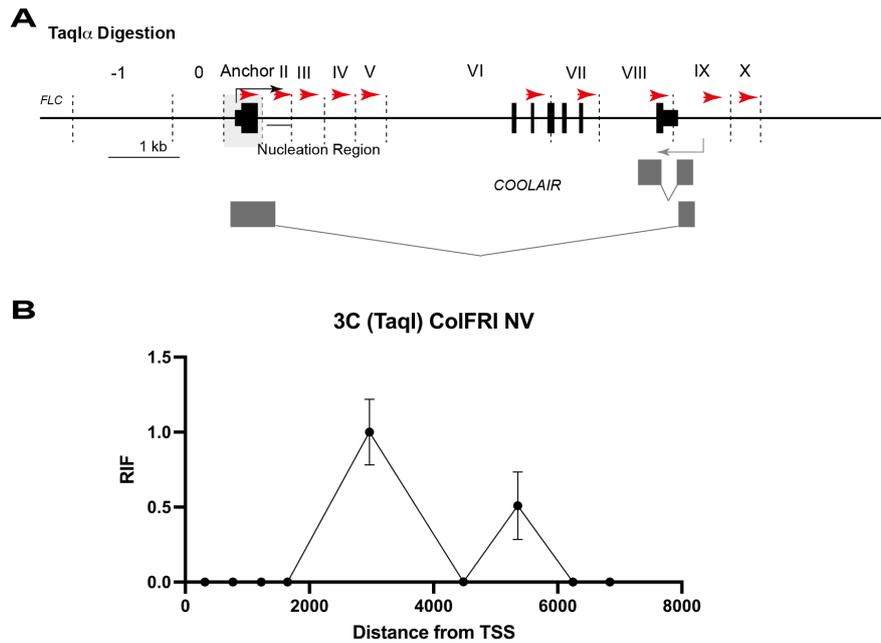


Figure 4-3 Chromatin loop at *FLC* analysed by TaqI α . (A) Diagram of the *FLC* locus showing the TaqI α restriction sites (dashed lines) and primer locations (red arrows). The region that serves as the anchor region is marked in light grey. (B) 3C-qPCR examining chromatin looping at *FLC* in non-vernalizing (NV) seedlings of ColFRI, error bars represent SEM of two biological replicates.

Therefore, I assayed the robustness of the reported *FLC* loops by performing 3C with the restriction enzyme TaqI α . TaqI α is a 4-bp cutter used in several 3C studies in other organisms (Chowdhary et al., 2019). It is also used in the Oxford Biodynamics' EpiSwitchTM platform, a commercial platform that uses chromatin conformations as biomarkers. ColFRI seedlings grown in non-vernalizing conditions were used for the experiment. Following nuclei extraction, TaqI α digestion and ligation, the interaction of the *FLC* TSS region with other regions at the *FLC* locus was investigated with qPCR (Fig. 4-3). Using the fragment around the TSS as the anchor region, we observed specific interactions with primers located in fragment six and eight (Fig. 4-3B). Fragment six encompasses the region previously described as interacting with the 5' TSS region of *FLC* using DpnII (Kim and Sung, 2017b), supporting the existence of this loop conformation at *FLC*. The second specific interaction we observed was between the TSS region and the end of the coding region, consistent with the previously reported gene loop using BglII/BamHI (Crevillén et al., 2013). Overall, the 3C analysis with TaqI α is consistent with the existence

of at least two conformations of the *FLC* locus, one where a loop exists that juxtaposes the *FLC* TSS region with the end of intron 1, and a second that juxtaposes the *FLC* promoter and terminator regions. The 3C analysis with TaqI α was therefore useful for detecting the two loops at *FLC*. However, due to the relatively infrequent occurrence of TaqI α restriction sites at *FLC*, we wanted to continue the 3C to increase the resolution of the 3C analysis. Therefore, 3C was performed using NlaIII, which cuts more frequently than TaqI α , leading to greater fragmentation of the *FLC* chromatin (Fig. 4-4A). NlaIII has previously been used to detect the repressive gene loop at the *WUS* locus in Arabidopsis (Guo et al., 2018).

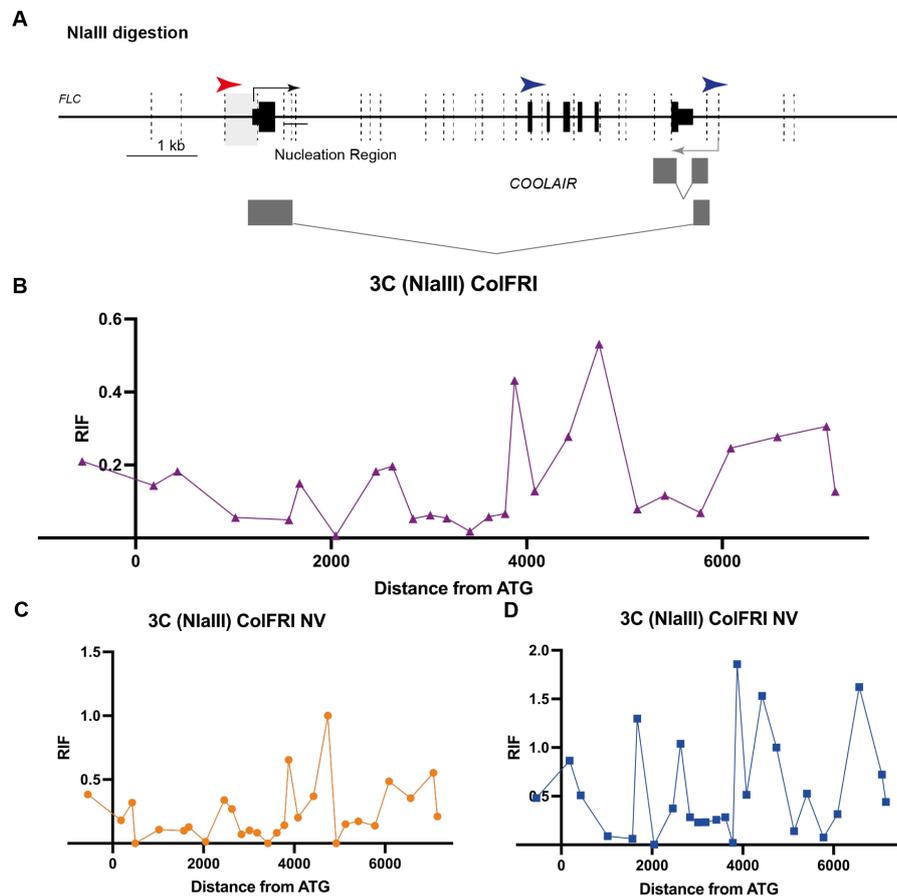


Figure 4-4 Chromatin loop at *FLC* analysed with NlaIII. (A) Diagram of the *FLC* locus showing the NlaIII restriction sites (dashed lines). The region that serves as the anchor region is marked in light grey. Blue arrows indicate regions that most frequently interact with the anchor primer (red). (B) 3C-qPCR examining chromatin looping at *FLC* in non-vernalizing (NV) seedlings of ColFRI with NlaIII. Values are mean relative interaction frequencies (RIF) of two biological replicates. (C-D) RIF of the individual replicates.

Non-vernalized ColFRI seedlings were used for the NlaIII assay to facilitate comparison with the TaqI α assay described above. Consistent with the results using TaqI α , interactions were observed between the *FLC* TSS region and the region around exon 2 and between the *FLC* TSS region and the end of the *FLC* transcription unit region (Fig.4-4B). This further strengthens the support for the existence of these two chromatin loop conformations at *FLC*. With the NlaIII-3C assay, we observed some variation in absolute interaction frequencies; however, the pattern of interaction was similar between the two replicates (Fig.4-4 C and D). In addition to the validation of the known loops, we were able to detect several other interactions. We observed two additional regions with relative high interaction frequency, one around 1500 bp downstream and a second around 2500 bp downstream of the *FLC* TSS. Interestingly, these positions correlate with other known features in the first intron of *FLC*. The 1500 fragment correlates with the vernalization response element (VRE) and the 5' end of the putative sense ncRNA *COLDAIR*. Similarly, the 2500 region correlates with the 3' end of *COLDAIR* (Heo and Sung, 2011) (Fig.4-5). Both regions also overlap with regions with pausing of antisense transcribing RNAPII, as detected by plaNET-seq in cold-treated Col-0 seedlings (Kindgren et al., 2020) (Fig.4-5). This further demonstrates the link between RNAPII transcription and chromatin topology. Interestingly, we also observed a relatively high interaction frequency with the region around the end of exon 6. The relatively high interaction frequencies at the boundaries of the Exon 2 – Exon 6 region suggest a link between productive RNAPII transcription and chromatin loop formation.

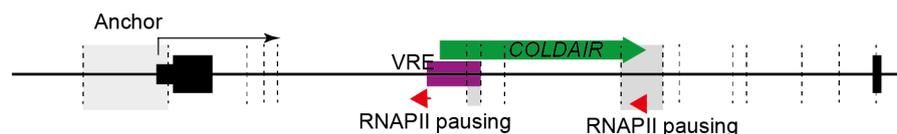


Figure 4-5 Additional interaction regions detected by NlaIII-3C. Schematic representation of the 5' end of *FLC*, showing NlaIII restriction sites (dash lines), the ncRNA *COLDAIR* (green), the VRE region (purple) and the antisense transcribing RNAPII paused sites (red arrows). Interacting regions detected by NlaIII-3C are represented as grey boxes.

Several regions around the TSS also showed relatively high interaction frequencies. However, due to the proximity of these regions to the anchor region, these interactions might be a result of the random collision between those fragments, a phenomenon often observed in 3C assays (Dekker, 2006).

In summary, by performing 3C with different restriction enzymes we have shown a robust ability to detect at least two chromatin loops at *FLC* in non-vernalized ColFRI seedlings. For convenience, we will in the following sections refer to the loop that forms between the TSS region and end of intron 1 as the **internal loop** and the one that forms between the TSS and the end of the transcription unit as the **gene loop**. Additional loops may form at *FLC*; however, as we currently have only observed these with NlaIII-3C, additional 3C assays are required to support the existence of these conformations.

4.2.2 Ectopic expression of *COOLAIR* disrupts gene loop formation

Previous work has shown that the gene loop at *FLC* is disrupted by exposure to cold. Disruption of the gene loop in cold coincides with increased *COOLAIR* transcription and changes in sense transcription dynamics. Because gene looping is often closely linked to transcription, it is intriguing to speculate that the increase in *COOLAIR* transcription leads to disruption of the gene loop. To test whether the *FLC* gene loop is associated with *COOLAIR* transcription or other cold factors, we analysed gene loop formation in a mutant background with increased *COOLAIR* levels even in the absence of cold. Mutations in the NAC domain transcription factor *NTL8* have been shown to cause ectopic expression of both *VIN3* and *COOLAIR* in the warm, resembling the situation during cold (Zhao et al., 2020, 2021). The ectopic expression of *COOLAIR* in the absence of cold makes *ntl8* an ideal system with which to explore *COOLAIR*-associated gene loop changes without cold temperatures.

We performed 3C with BglII/BamHI in the previously described dominant *ntl8* mutant *ntl8-D3* (Zhao et al., 2020, 2021). Similar to the behaviour during cold, the 3C analysis in *ntl8-D3* showed that the gene loop is reduced compared to Col-FRI (Fig.4-6), indicating that antisense transcription is associated with the dynamics of the gene loop.

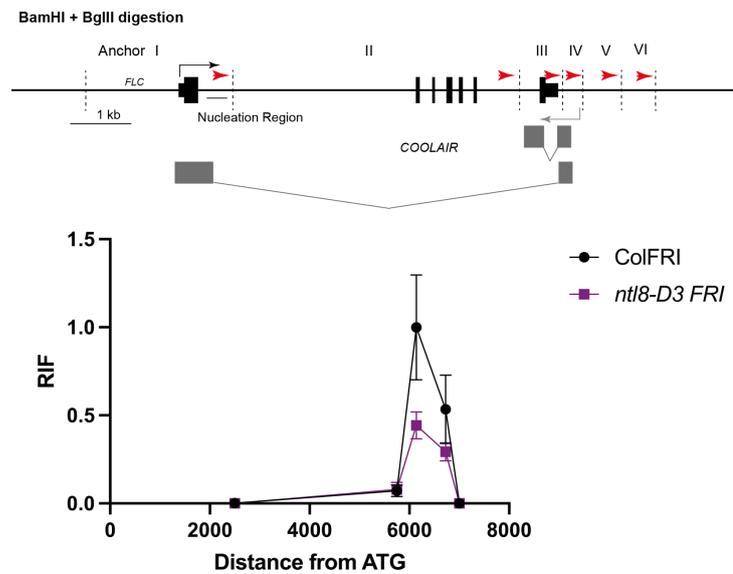


Figure 4-6 Gene loop analysis in ColFRI and *ntl8-D3* FRI. (A) Schematic representation of the *FLC* locus showing the BglII and BamHI restriction sites (dashed lines) and primer locations (red arrowheads). (B) 3C-qPCR analysis of gene looping in non-vernalized seedlings of ColFRI and *ntl8-D3*. Error bars represent the SEM of 8 independent biological replicates.

We cannot distinguish between the *COOLAIR* RNA itself being the causative factor of changes in the gene loop or if these changes are caused by changes in sense/antisense transcription in *ntl8-D3*. We previously reported that full-length transcription of sense and antisense at *FLC* is mutually exclusive (Rosa et al., 2016). As *COOLAIR* is ectopically expressed in *ntl8-D3* before vernalization, this has an impact on full-length sense transcription. It is therefore possible that the observed gene loop disruption in *ntl8-D3* is due to lower *FLC* sense transcription, rather than being directly associated with *COOLAIR*. Consistent with this, it has been suggested that the gene loop is associated with the *FLC* expression state (Li et al., 2018).

4.2.3 Role of *COOLAIR* promoter/3' *FLC* sequences in gene loop formation

In the previous sections, we showed that the gene loop forms between the *FLC* promoter and the 5' end of *COOLAIR* and that the gene loop is disrupted when *COOLAIR* expression increases. One model that explains these data is if a gene loop forms to repress *COOLAIR* expression, paralleling the function of a chromatin loop at the *HMS2:SUT650:BAT2* locus in *S. cerevisiae* (Nguyen et al., 2014). This prompted us to investigate whether the activity of the *COOLAIR* promoter itself is required for formation of the gene loop. We therefore tested whether the ability to cause gene looping is held in the sequence or is related to the function of the *COOLAIR* promoter. 3C with BamHI/BglIII was performed using a transgenic line named Terminator Exchange 2.0 (TEX2.0). TEX2.0 contains an insertion of the NOS terminator sequence that terminates *COOLAIR* transcription without replacing the 3' end of *FLC* (Zhao et al., 2021) (Fig. 4-7). The TEX2.0 construct was introduced into a background (*flclean*) where the *FLC* locus had been deleted by CRISPR/Cas to avoid the complexity of any remaining *FLC* sequences. Previous reports had analysed chimeric *FLC* constructs in *flc-2* mutants, but this genotype still has fragments of the endogenous *FLC* locus in the genome. The chimeric constructs removed *COOLAIR* promoter sequences by replacing *FLC* 3' sequences with 3' sequences from the *NOPALINE SYNTHASE (FLC-NOS-T)* or *MAF2 (FLC-MAF2-T)* genes. This resulted in disruption of the gene loop (Li et al., 2018) (Fig. 4-8).

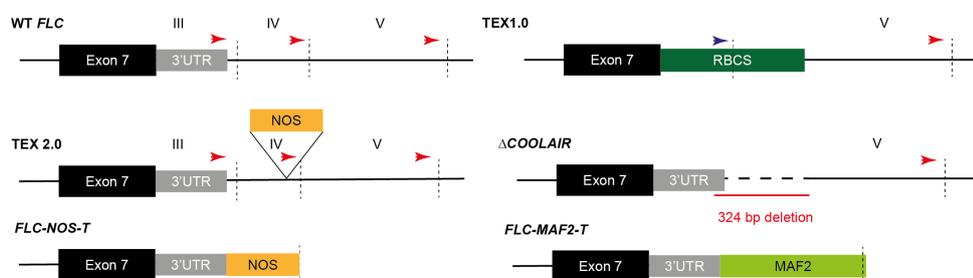


Figure 4-7 Schematic illustration of the 3' end of chimeric *FLC* constructs. Exons are represented by black boxes and the untranslated regions of *FLC* as grey boxes. BamHI and BglIII restriction sites are shown as dashed lines. Red arrowheads and numbers refer to primers and regions in BamHI/BglIII-3C similar to WT. The blue arrowhead represents

TEX1.0 specific primer. RBCS, NOS and MAF2 terminators in TEX1.0, TEX2 and FLC-NOS-T, and in FLC-MAF2-T are indicated by green, orange and light green boxes, respectively. The deleted sequence in Δ COOLAIR is represented by the red line.

Consistent with a role for *COOLAIR* promoter activity in gene loop formation, we observed reduced interaction frequencies in a representative TEX2.0 *flclean* line compared to the non-transgenic ColFRI. We further tried to validate this finding by performing 3C in a line that carries a 293 bp deletion of the *COOLAIR* promoter of the endogenous locus (Luo et al., 2019). Surprisingly, we found that gene looping is not disrupted in this line. This suggests that the organisation of the terminator sequence is important for the gene loop at *FLC*, but that a functional *COOLAIR* promoter is not required. However, so far only the *COOLAIR* dynamics of *FLC* Δ *COOLAIR* in Col-0 have been reported (Luo et al., 2019; Zhao et al., 2021). An in-depth characterization of remaining *COOLAIR* promoter activity in *FLC* Δ *COOLAIR* is missing from the literature. Work in the Dean lab has shown that there is still promoter activity in the *FLC* Δ *COOLAIR* (P. Zhu unpublished).

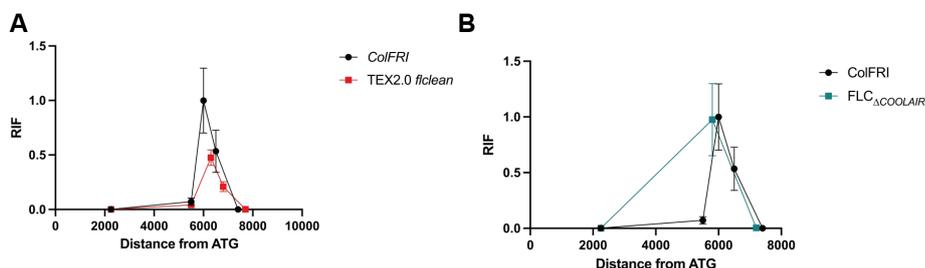


Figure 4-8 Gene loop in *TEX2.0* and *FLC* Δ *COOLAIR*. 3C of the *FLC* locus in *TEX2.0* (A) and *FLC* Δ *COOLAIR* (B) in 10-days-old warm-grown seedlings with BamHI/BglII. Primers as in Fig.4-6A. Relative interaction frequency (RIF) was calculated as described in Material and Methods. The data are the averages of at least six independent biological replicates with error bars showing the SEM. Experiments performed in parallel with Fig. 4-7B.

Overall, the data suggest that changes to the *COOLAIR* promoter/*FLC* terminator sequence can disrupt gene loop formation, but that a 293 bp in that region does not lead to altered gene loop formation. Further studies are therefore required to uncover the complex relationship between the *COOLAIR* promoter/3' end of *FLC* and the formation of the gene loop.

4.2.4 Alternative loop formation through high antisense transcription

In the previous section, I showed that *ntl8-D3* leads to disruption of the gene loop. As described in the introduction, gene loops are frequently associated with RNAPII transcription, with the formation of a gene loop potential facilitating a high transcriptional state (Lykke-Andersen et al., 2011). *ntl8-D3* leads to high antisense transcription at *FLC*, so it was intriguing to speculate that transcription of *COOLAIR* could also lead to the formation of a gene loop, even though coding and non-coding transcription are mechanistically different. We have previously shown that induction of *COOLAIR* in the cold does not lead to detectable loop formation when assayed by BglII and BamHI (Crevillén et al., 2013). In wild-type plants, detectable *COOLAIR* expression is restricted to the pre-vasculature cells in the root (Rosa et al., 2016) and potentially to specific phases of the cell cycle (Zhao et al., 2021) (C. Xu unpublished). *ntl8-D3* upregulates *COOLAIR* expression by expanding the zone of expression – *COOLAIR* is detected in almost all cell types in *ntl8-D3* (Zhao et al., 2021). We therefore reasoned that *ntl8-D3* provides an excellent genotype to test how loop formation is affected by antisense transcription. We performed 3C using BglII/BamHI, similar to the previous analysis, and tested the formation of new interactions. Because of the frequent occurrence of BamHI/BglII restriction sites at the *COOLAIR* TSS region, we assayed interactions using region III, IV and V as anchor regions (Fig. 4-9A), to increase the likelihood of capturing new interactions.

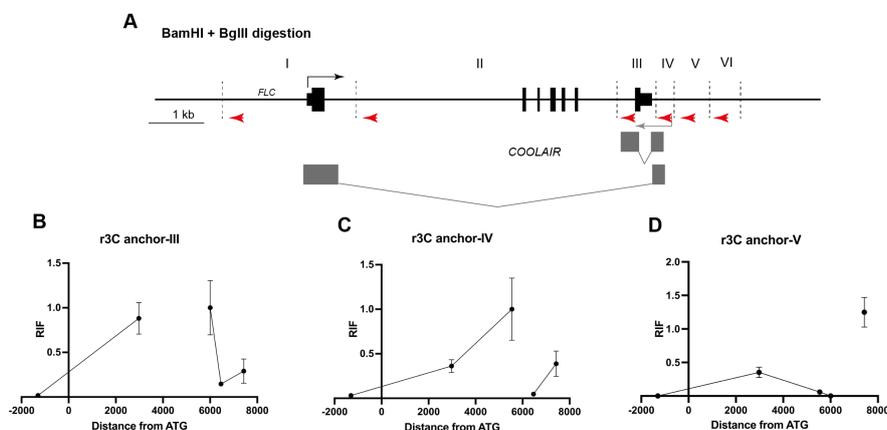


Figure 4-9 Investigation of alternative loops in *ntl8-D3*. (A) Schematic drawing of the *FLC* locus with primers (red arrowheads) and BamHI/BglIII restriction sites (dashed

lines). (B-D) 3C analysis in 10-days-old warm-grown *ntl8-D3* seedlings using regions III (B), IV (C) and V (D) as anchor regions.

Independent of the anchor region used, we observed relatively high interaction frequencies with neighbouring fragments. This was probably caused by the random collision of such fragments, often observed in 3C (Dekker et al., 2002). However, we were unable to detect reliable high frequency interactions with downstream regions in *ntl8-D3*. This is consistent with the previous result that *COOLAIR* transcription does not lead to the formation of alternative loop conformations. As mentioned above, 3C assays are highly dependent on the chosen restriction enzyme(s) so therefore we cannot rule out the existence of alternative loop conformations at *FLC* during *COOLAIR* transcription. However, this would require the use of different restriction enzymes in the assay.

4.3 Internal gene loop dynamics

In the previous sections, the focus was on the gene loop, but as shown in the first section, a second loop, the internal loop that forms between the *FLC* TSS and the end of intron1/start of exon2, also forms at *FLC*. As mentioned earlier, studying 3D organisation with 3C relies on the use of restriction enzymes. We successfully used the combination of BamHI and BglII to study the behaviour of the gene loop. However, we cannot identify other interactions with this enzyme mix. As shown in the first section, the use of more frequent cutters leads to the identification of additional loops. In addition to TaqI α and NlaIII, others have successfully used DpnII to assay the dynamics of the internal loop at *FLC* (Kim and Sung, 2017b; Li et al., 2018). Therefore, we studied the behaviour of the internal loop with 3C using DpnII. Previous studies have linked the gene loop to the high transcription state and the internal loop to the repressed state. To understand whether there is a connection between the expression state and the internal loop, as well as a connection between the gene loop and the internal loop, we performed 3C in the same genotypes as above. *ntl8-D3* reduced *FLC* sense transcription; however, we did not observe an increase in the internal loop. This suggests that the internal loop is not a direct

consequence of a low *FLC* sense transcription state. *ntl8-D3* represents a special case where the locus has low *FLC* sense levels, but the antisense strand is still highly transcribed, meaning that despite having low *FLC* sense, the locus cannot be considered repressed. *TEX1.0*, *TEX2.0* and *FLC_{ΔCOOLAIR}* all have *FLC* levels similar to those of *ColFRI*, and as expected, they do not show any differences in the internal loop.

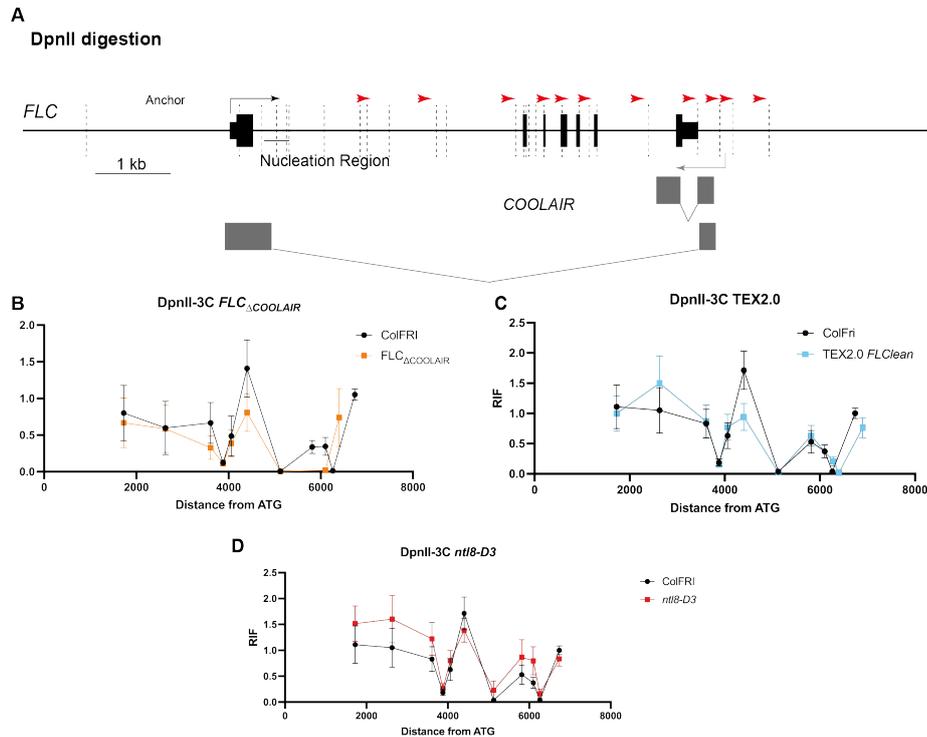


Figure 4-10 Chromatin loops at *FLC* assayed with DpnII. Schematic drawing of the *FLC* locus with primers (red arrowheads) and DpnII restriction sites (dashed lines). (B-D) 3C analysis in 10 days warm grown *FLC_{ΔCOOLAIR}* (B), *TEX2.0* (C) and *ntl8-D3* (D). The data are the averages of at least four independent biological replicates with error bars showing SEM.

4.4 Discussion

Chromatin is, for convenience, often illustrated as a linear fibre; however, within the nucleus, the chromatin is organized in 3D. In animals, the genome is organized into sub megabase TADs. Genome-wide studies have suggested that TADs are not as obvious in *Arabidopsis* as in other higher plants and vertebrates (Liu et al., 2016). Instead, single genes are often organized into functional gene loop units connecting the promoter and terminator. Consistent with a link to transcription, gene loops are often associated with actively transcribed loci. The formation of these local gene

loops at actively transcribed genes is similar to findings reported in the budding yeast *S. cerevisiae*.

4.4.1 RNAPII driving gene loop formation

In line with the genome-wide picture in Arabidopsis, actively transcribed *FLC* forms a gene loop between the promoter and terminator regions. During early cold when *FLC* transcription is reducing, the gene loop is disrupted (Crevillén et al., 2013) and this coincides with increased transcription of *COOLAIR* (Swiezewski et al., 2009). In this chapter, I show that ectopic high *COOLAIR* transcription in *ntl8-D3* also leads to disruption of the gene loop, even in the absence of cold. In both early cold and *ntl8-D3*, the increase in antisense transcription correlates with reduced *FLC* sense levels. Therefore, we cannot determine whether the disruption of the gene loop is caused by reduced *FLC* levels or by increased antisense transcription. Neither can we establish whether disruption of the gene loop derepresses the *COOLAIR* promoter, or whether the activation of *COOLAIR* transcription disrupts the gene loop. It has been suggested that gene loops form through the physical interaction of proteins that associate with the promoter and termination regions. Recently, the interactions of gene-loop-facilitating proteins have been shown to depend on RNAPII (Allepuz-Fuster et al., 2019). The mechanistic insights from *S. cerevisiae* suggest that the RNAPII transcribing *FLC* sense is the deterministic factor for the formation of the gene loop. Consistent with this, genotypes that have higher *FLC* expression than Col-0, like *ndx*, *BAF60* RNAi and ColFRI (Jegu et al., 2014; Li et al., 2018; Sun et al., 2013), show relatively higher interaction frequencies between the promoter and terminator of *FLC*, although *COOLAIR* expression is also higher.

However, I also observed that in a genotype with *COOLAIR* termination through insertion of a NOS terminator (TEX2.0), the gene loop is also disrupted. TEX2.0, together with the other chimeric *FLC* lines (e.g., TEX1.0), is a transgenic line that has been selected for the overall similarity of its *FLC* levels to those of wild-type ColFRI. In contrast to the

observation here for TEX2.0, the lines used by Li et al. (2018), *FLC-NOS-T* and *FLC-MAF2-T*, have reduced *FLC* mRNA levels, potentially explaining the lower gene loop frequency. Strikingly, the disruption of the gene loop in the TEX2.0 line suggests that the formation of a chromatin loop is not simply an outcome of RNAPII transcription, and the presence of a gene loop is also not required to maintain a high expression state.

4.4.2 The *COOLAIR* promoter/3' end of *FLC* is required for gene loop formation

In the section above, I discussed the relationship between RNAPII transcription and the formation of the *FLC* gene loop. This ended with the apparent contradiction that despite the high *FLC* transcription levels in TEX2.0, the gene loop is reduced. This leads to the question of what, in addition to RNAPII transcription, determines the establishment of the gene loop.

Formation of the gene loop at *FLC* has previously been suggested to be a consequence of specific proteins that associate with the 5' end and 3' end of *FLC*; through the physical interaction of these proteins, the gene loop is established (Li et al., 2018). However, the region where these proteins are bound is missing in the *FLC_{ΔCOOLAIR}* line that maintains the gene loop. It is therefore likely that other factors are also involved in the formation of the gene loop or that the proteins that potentially form the gene loop, are cleared as a result of the RNAPII termination of the *FLC* strand. In the TEX2.0 scenario, even the distal primer in fragment VI does not show increased interaction. This suggests that read-through transcription and downstream termination in TEX2.0 are unlikely to explain the reduced gene loop formation between the *FLC* 5' and 3' end. In order to understand any change in RNAPII transcription in the different *flc* lines, the next experimental step would be RNAPII ChIP in these lines.

One possible explanation for these complex observations could be that the CRISPR/Cas *FLC_{ΔCOOLAIR}* line, in contrast with all the other *FLC* modified lines, is located at its normal genomic position. Due to the nature of

introducing transgenes into Arabidopsis through *Agrobacterium*-mediated transformation, the insertion site of the transgenes into the genome is random. Previously it has been suggested that genomic context does not influence the formation of the gene loop at *FLC* (Crevillén et al., 2013). However, these experiments transformed an *FLC:LUC* transgene into the *flc-2* background and the fast neutron *flc-2* allele maintains 4000 bp of the *FLC* locus, including the 3' end (Crevillén et al., 2013). This complex genetic background might complicate the interpretation of the 3C results. Therefore, it would be valuable to perform 3C in lines carrying the WT *FLC* sequence transformed into the *flclean* background, in which the *FLC* genomic region was removed by CRISPER/Cas. However, such a line was not available at the time of writing this thesis.

In yeast, gene loop formation restricts divergent transcription from bi-directional promoters (Tan-Wong et al., 2012). While divergent transcription is prevalent in yeast and metazoans, it is less pronounced in Arabidopsis (Kindgren et al., 2020). While divergent transcription has not been reported for *FLC* or *COOLAIR*, its existence cannot be ruled out due to high RNA turnover. Many divergent transcripts are only detectable in yeast in exosome mutants; the use of Arabidopsis exosome mutants like *hen2* might reveal divergent transcription events and help uncover if the formation of a gene loop at *FLC* is linked to transcription directionality.

4.4.3 Additional chromatin loops form at *FLC*

In the analysis of gene looping at *FLC*, we identified several interactions in addition to the gene and internal loop, suggesting that additional chromatin loops may exist at the *FLC* locus. This raises the question of whether they co-exist at the same allele, whether each loop is mutually exclusive and/or whether they are cell or tissue specific. When performing 3C, the starting material is homogenized tissue of whole seedlings, therefore it is impossible to distinguish between loops originating from different cell types. Studies in other organisms have shown extensive heterogeneity in 3-D nuclear organization between cell types and also

between individual alleles (Finn et al., 2019). Recently, heterogeneity between tissues has also been reported in plants (Yadav et al., 2021). Yadav et al. performed Hi-C on purified nuclei from endosperm and leaf, and this revealed distinct chromatin conformation in those cell types. In ColFRI, many tissues express *FLC* (Sheldon et al., 1999), but it is possible that the gene loop predominantly forms in cell types that highly express *FLC*, and the other loops form in cells with low *FLC* expression.

4.4.4 Internal loop associated with Polycomb repression?

The internal loop forms between the TSS and the end of intron1 and it is unaffected in the tested mutants (*ntl8-D3*, *TEX2.0* and *FLC_{ΔCOOLAIR}*). This suggests that the internal loop is less linked to RNAPII transcription compared to the gene loop. The internal loop has been linked to the low expression state of *FLC* (Li et al., 2018), but despite a low sense expression state in *ntl8-D3*, we do not observe increased levels of the internal loop. In previous reports, the internal loop was linked with the low expression state during vernalization, and in Col-0. Common for *FLC* chromatin in both Col-0 and vernalized plants is the presence of H3K27me3 over the locus. Despite the low *FLC* sense expression in *ntl8-D3*, we do not observe increased H3K27me3 at *FLC* (H. yang unpublished), potentially because of high antisense transcription antagonizing H3K27me3 accumulation. Therefore, our results are consistent with a model in which the internal loop is not just correlating with a low expression state, but also with the presence of H3K27me3 or PcG proteins. In line with this, the apparent increase of the internal loop during vernalization is impaired in *vin3* and *vrn5* mutants (Kim and Sung, 2017b).

The contradiction to this model is the detection of the internal loop in ColFRI in non-vernalized conditions, where *FLC* expression is high. Although this might seem counterintuitive at first, Columbia is not a naturally vernalizing accession of Arabidopsis. The vernalization requirement has been added by the introgression of a functional *FRI* allele.

In Columbia, we can still observe some H3K27me3 at the *FLC* nucleation region even before vernalization, indicative of a small subset of pre-silenced alleles. It is tempting to speculate that this fraction contributes to the observation of the internal loop in ColFRI. The presence of H3K27me3 at *FLC* in non-vernalized seedlings is Columbia specific and is not observed in accessions like Lov-1 and Edi-0 (Qüesta et al., 2020). It would be interesting to test whether the lack of H3K27me3 correlates with the absence of the internal loop in non-vernalized seedlings of these accessions.

4.5 Summary

In this chapter, I used 3C to assay the 3D conformation of the *FLC* locus in different genetic mutants. I assayed how the altered transcriptional state of a locus affects its 3D conformation. 3C is a challenging technique and often raises concerns due to high variability between replicates and a lack of reproducibility². For that reason, I first validated the existence of at least two loop conformations between the TSS region of *FLC* and the end of intron 1 and the end of the gene body, respectively. Through high fragmentation of the locus, it was revealed that additional putative loops may form between the TSS and regions in intron 1, which overlaps with the putative ncRNA COLDAIR. The loop that juxtaposes the sense TSS and TTS was shown to be disrupted in mutants that affect antisense transcription. Ectopic high antisense transcription leads to disruption of the gene loop, resembling the behaviour during early cold. Similarly, insertion of a terminator sequence that prevents antisense transcription from the terminator region also disrupted the gene loop. However, in contrast, the deletion of a fragment of the antisense promoter did not disrupt the gene loop. The role of the antisense transcription unit in gene loop formation therefore requires further investigation. I also conclude that low *FLC* sense expression in *ntl8-D3* does not lead to an increase in the internal loop, suggesting that the previously observed correlation between the internal

² Federico Arial, Maike Stam and Franziska Turck's lab personal communication

loop and low *FLC* expression is dependent on the presence of H3K27me3 and/or PcG proteins.

4.6 Outlook

Our analysis of gene loop dynamics at *FLC* revealed that the gene loop is related to the switching from high sense to relatively high antisense expression. In other words, a switch of chromatin conformation happens at *FLC* as *FLC* sense is transcriptionally repressed by induced antisense expression, for example in the cold or in *NTL8-D3*.

A dynamic transcriptional switch between a sense and antisense transcriptional unit has recently been described for the *HMS2:SUT650* locus in *S. cerevisiae* (Nguyen et al., 2014). Here, transcription of the sense *HMS2* forms a repressive loop that prevents antisense transcription of *SUT650*. Very similar to the switch at *FLC* from sense to antisense in response to cold, a switch is observed at this yeast locus when the source of carbon is changed. Exactly the same as observations at *FLC*, the change in expression is a digital switch causing more cells to express the antisense, rather than an analogue regulation where each antisense already expressing cells is expressing more (Nguyen et al., 2014).

The use of the *FLC-VENUS* and *FLC-mCherry* lines nicely visualizes how the two *FLC* copies switch independently (Berry et al., 2015), and we have previously hypothesized that this means that VIN3-VRN5-PRC2 will nucleate independently at each allele. Similar to *FLC*, X-inactivation is allele specific, and here the transcription of the lncRNA XIST is a key deterministic factor for which allele switches off. A recent study has shown that the trans factors can be present at both alleles, but the repression effect is only achieved at the allele where XIST is expressed (Yu et al., 2021). Many functional similarities therefore seem to exist between XIST and COOLAIR, so it is interestingly to speculate that the allele-specific epigenetic switching is linked to the activation of *COOLAIR* transcription. To test this, we envision creating *FLC-mVenus-TEX2* lines,

and hypothesize that instead of seeing files of ON and OFF cells the switching will be more homogenous because the PcG localization will no longer be allele specific.

Chapter V

5 Transgenic lines

This chapter describes the generation of stable transgenic Arabidopsis lines expressing VIN3 and VRN5 with different domain deletions and point mutations. The lines will be used to elucidate the molecular functions of the VEL proteins and to help understand the role of the VEL proteins in the epigenetic switch at *FLC*.

I generated all the constructs and Agrobacterium strains used for Arabidopsis transformation. Transformation and conformation of lines were done jointly with Dr Anna Schulten and Shuqin Chen. I performed all the microscopy described in this chapter.

5.1 Introduction

The VEL proteins (VEL1, VEL2, VIN3, and VRN5) are general Polycomb accessory proteins that not only function in the vernalization pathway that leads to the epigenetic silencing of *FLC*, but also in the regulation of other genes in the Arabidopsis genome (Bordiya et al., 2020; Kim and Sung, 2010). The effect of the VEL proteins has almost entirely been connected to PRC2. However, the mechanism(s) of the VEL proteins themselves has been greatly overlooked. In this work we used the *FLC* paradigm to elucidate the function of VIN3 and VRN5, the two VEL proteins known to be required for the stable silencing of *FLC* (Greb et al., 2007; Sung and Amasino, 2004). A way to elucidate the function of proteins is to make and study more elegant mutations than classical null-mutants. Recently,

the use of point mutations in PRC2 has revealed important features of Polycomb biology in mammalian cells (Long et al., 2020; Youmans et al., 2021). This has highlighted the value of using specifically designed protein mutants to uncover the individual functions of different parts of a protein.

At the beginning of the work for this thesis, we knew that the VEL proteins were part of a family of four proteins with a PHD, FNIII and VEL domain. The PHD was reported to be involved in histone tail binding (Kim and Sung, 2013), the function of the FNIII domain was unknown, and the VEL domain was known to be involved in protein-protein interaction between the VEL proteins (Greb et al., 2007; Sung et al., 2006). Throughout the work done during this thesis period, it became clear that the VEL proteins contain additional interesting features and show functional differences, despite their shared domain architecture. Many of the biochemical properties can be studied with *in vitro* or transient assays. However, so far, the best system for analysing these domains' functional relevance for *FLC* repression requires the generation of stable transgenic lines, where the proteins are expressed at similar levels and in the same cells as the endogenous proteins. The use of genomic constructs with introns and the native promoter and terminator helps with the construction of transgenes with near-endogenous behaviour. This is important to truly elucidate the endogenous function of protein features at *in vivo*- relevant concentrations. In the following, I will summarize the lines I made and the rationale for their design.

5.2 Results and discussion

5.2.1 Characterization of the genetic backgrounds

5.2.1.1 *vin3-1*

The original *vrn* mutants had been identified in Landsberg *erecta* (Chandler et al., 1996) but for comparison with other work I wanted to introduce the *VIN3* transgenes into a Col-0 genotype so I chose *vin3-1 FRI*. *vin3-1* is a fast-neutron allele, which has previously been described (Sung and Amasino, 2004). This background was chosen instead of the more

commonly used T-DNA insertion mutant *vin3-4*. *vin3-4* contains around 20 T-DNA insertions at the *VIN3* locus (Dean lab unpublished information). Therefore, to avoid the potential risk of transcriptional silencing of the T-DNA transgenes, we opted for *vin3-1* as the background for our transgenic lines. *vin3-1 FRI* plants showed a late flowering phenotype similar to *vin3-4* (Hepworth et al., 2020). As the original paper did not report the site or size of the deletion in *vin3-1* very precisely, we characterized the *vin3-1* mutant in more detail. Based on the original paper we amplified the region likely to contain the deletion and sequenced the PCR product. This showed that *vin3-1* carries an 829 bp deletion starting in the 5'UTR region and spanning the first two exons (Fig.5-1A-B). The sequencing revealed that the promoter sequence is maintained and that the remaining *VIN3* sequence contains putative in-frame start codons that could give rise to a VIN3 protein carrying most of the relevant domains. Therefore, we further analysed the expression level of *VIN3* in *vin3-1*. We designed primers for the reverse transcription and qPCR that were located within the last exon (Fig. 5-1A). The aim was to detect any *vin3* transcripts that might arise from the locus in the *vin3-1* mutant and that could result in VIN3 protein being translated from any of the identified start codons. Consistent with the vernalization phenotype, we observed very little induction of *VIN3* transcripts in the *vin3-1* mutant (Fig.5-1C), suggesting that very little productive RNAPII transcription occurs.

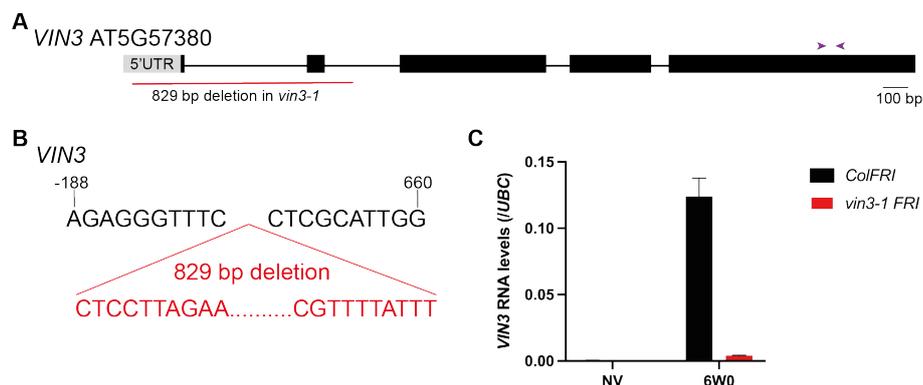


Figure 5-1 Characterization of *vin3-1*. (A) Schematic representation of the *VIN3* genomic sequence, the position of the 829 bp deletion in *vin3-1* is marked underneath the sequence. Black boxes represent exons. (B) Genomic sequence around the *vin3-1* deletion. (C) *VIN3* transcript levels in ColFRI and *vin3-1 FRI* in non-vernalized (NV) seedlings and after 6 weeks of cold treatment (6W0). The locations of primers are marked on the schematic in A. Error bars are the SEM of at least two replicates.

Our characterization confirmed that *vin3-1 FRI* is a suitable background for our analysis of the function of mutated VIN3 proteins.

5.2.1.2 *vrn5-8*

The *VRN5* transgenes were introduced into the *vrn5-8* Col-0 T-DNA line. An active *FRI^{SF2}* allele had been introgressed into the *vrn5-8* line (Dana Zhu, Dean lab) in order to analyse the impact of the *vrn5* mutants in the vernalization pathway. Our new lines were therefore in a different background from the previous complementation *VRN5-YFP* line, which was in a *vrn5-8 FRI (JU223)* background, where an active *FRI* had been introduced as a transgene through transformation (Johanson et al., 2000; Yang et al., 2017). In contrast to *vin3-1*, we had to use a T-DNA background for the *vrn5* transgenics as no EMS-induced *vrn5* allele was available. *vrn5-8 FRI* displayed a late flowering phenotype under all conditions used, consistent with a lack of expression or a loss of function, as previously described (Greb et al., 2007). It is worth noting that the *vrn5-8 FRI* showed a more severe late flowering phenotype than *vin3-1 FRI*. Together our analyses have shown that *vin3-1* and *vrn5-8* are suitable and the best available lines to serve as background for the establishment of stable transgenic lines.

5.2.2 Design of constructs

Earlier chapters discussed how the amount of VIN3 and VRN5 may be an important part of the mechanism for nucleation and the metastable memory. We therefore designed our strategy for the establishment of transgenic lines in such a way to optimize the chance of obtaining lines with similar expression levels and behaviour as the endogenous genes.

We opted for constructs containing the genomic sequence of *VIN3* and *VRN5* with the upstream and downstream regions. For *VIN3*, we used the previously described complementing *pVIN3::VIN3-eGFP* construct as the template for our mutagenesis (Qüesta et al., 2016; Yang et al., 2017; Zhao et al., 2020). The construct consists of the genomic sequence surrounding

the *VIN3* gene. It contains 6318 bp upstream of the ATG and 3100 bp downstream of the stop codon (Fig. 5-2). We opted for this relatively large construct because previous analysis of a very similar *VIN3-luciferase* construct showed a RNA pattern very similar to that of endogenous *VIN3* during vernalization (Zhao et al., 2020). This suggests that the protein concentration could also be similar to that of endogenous *VIN3*. To facilitate the visualization and detection of transgenic proteins, a C-terminal eGFP tag was inserted between the last amino acid and the 3' UTR. To prevent the relatively large eGFP tag interfering with the function of *VIN3*, a 10 amino acid flexible linker (SAAASAAASA) was inserted upstream of the eGFP tag. For *VRN5*, we used a newly made construct (Yaoxi Li, Dean lab) as the template for our mutagenesis. This construct contains 2242 bp upstream and 1260 bp downstream of the coding sequence to maintain endogenous regulation (Fig. 5-2). Like the *VIN3* constructs, *VRN5* was C-terminal fused to sfYFP2 with the same 10 amino acid linker in between. A similar, but not identical, *VRN5*-YFP transgene has previously been shown to rescue the late-flowering phenotype of *vrn5-8 JU223* and *vrn5-1 fca-1* (Greb et al., 2007; Yang et al., 2017).

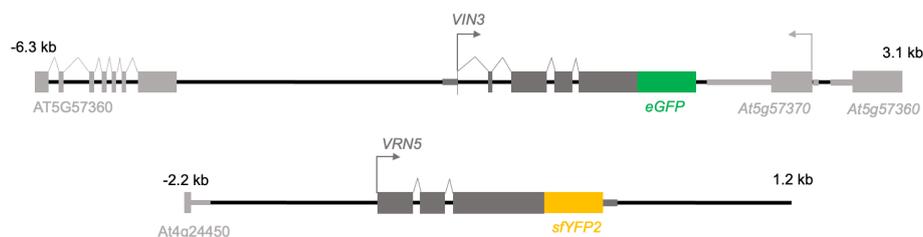


Figure 5-2 Genomic *VIN3* and *VRN5* structure. The region used to generate *VIN3*-eGFP and *VRN5*-sfYFP2 transgenic lines. *VIN3* and *VRN5* exons are represented as dark grey boxes with the splicing pattern represented with lines. Neighbouring gene fragments are shown in light grey.

All constructs were cloned in pENTR and recombined into a destination binary vector based on pSLJ755I6 (Jones et al., 1992). They were transformed into *vin3-1 FRI* and *vrn5-8 FRI*, respectively, by *Agrobacterium*-mediated floral dip. pSLJ755I6 is based on pRK290, a 25 kb plasmid (Jones et al., 1992). The use of this plasmid ensures a higher predominance of single-copy insertions compared to the more commonly

used binary vectors (Muskett et al., 2003). In addition to the use of pSLJ755I6, T1 plants were analysed for transgene copy number (iDnaGENETICS), following selection with BASTA spraying and PCR confirmation.

5.2.3 Mutations to address the functional relevance of polymerization of the VEL domain

In chapter III, I discussed the discovery of the VEL domain as a novel polymerization domain. We showed that the VEL domain is required for the VIN3-mediated repression of *FLC* during cold. However, as discussed in Chapter III, more elegant mutations are required to confidently link this observation to the polymerization property of the VEL domain. We therefore introduced mutations into the interaction surfaces that by gel filtration had been shown to prevent polymerization. However, a single-point mutation, VIN3 R556D, in the head surface of the VIN3 VEL domain was not enough to cause a detectable effect on the downregulation of *FLC*, at least under the conditions used (Chapter III). We therefore designed several other mutations to prevent polymerization and dimerization more effectively. The chosen mutations were based on the residue's location in the two interaction surfaces in the VIN3 VEL polymer that were revealed by the crystal structure (M. Fiedler unpublished). In addition, gel filtration showed that the mutations prevented polymerization and Co-IP revealed that they reduced the homo-interaction of the VIN3 VEL domain (M. Fiedler unpublished). The desired lines related to the VEL domain are summarized in Table 5-1 below, including the reasoning behind the respective mutations.

Table 5-1 Generation of VIN3 VEL domain *in planta* mutations. Description of stable transgenic lines current under generation with the residues/region mutated and the function of the mutations. The information for all the residues in this table has been obtained through our work. See text for more details.

Mutations	Role
VIN3 ΔVEL	Removal of the entire VEL domain, complete disruption of the VEL domain
R556D	Weak head mutant, limits polymerization <i>in vitro</i>
R554A/R556D	Strong head mutant, prevents polymerization <i>in vitro</i>
R556D/I575D	Head and tail mutant, prevents polymerization and dimerization <i>in vitro</i>
R554A/R556D/I575D	Strong head and tail mutant, prevents polymerization and dimerization <i>in vitro</i>
T559E	Putative site of regulation through phosphorylation, the phosphomimic Glutamic acid (E), prevents polymerization <i>in vitro</i>
S562D	Putative site of regulation through phosphorylation, the phosphomimic aspartic acid (D), prevents polymerization <i>in vitro</i>
I575T	The residue that differs between VEL1/VIN3 and VRN5 and may explain the different VEL polymerization abilities.
L584P	Residue likely required for helix 4 extension and domain swapping, observed to be specific for the VIN3 VEL domain. The change to proline would in theory prevent domain swapping.
VEL-DIX Swap	Substitution of the VEL domain with another known polymerization domain,

	the DIX domain, to test for polymerization as the main role of the VEL domain.
VEL-DIX^{mut} Swap	Negative control for the above DIX swap construct. A polymer-dead DIX domain should not complement.

As discussed in Chapter III, the observation that a single mutation in the head (or tail) surface is not enough to cause a phenotype is similar to the lack of phenotype observed in the LFY SAM polymerization domain (Sayou et al., 2016). We therefore designed a series of other mutations that either add to our previous R556D mutant (R554A/R556D) mutant or combine mutations in both the head and tail surface (R556D/I575D and R554A/R556D/I575D). In contrast to the arginine residues in the head surface, the conservation of the isoleucine is less strong (Fig.5-3). Within the Arabidopsis VEL family VIN3, VEL1 and VEL2 contain a hydrophobic isoleucine at this position while VRN5 and VEL3 contain a polar threonine. In contrast to the VEL domain of VEL1 and VIN3, the VEL domain of VRN5 did not form large foci in mammalian cells indicative of the different polymerization features of the VRN5 VEL domain (M. Fielder unpublished). It is tempting to speculate that the difference in this residue might be important for the polymerization behaviour of the VEL domain. Therefore, a single amino acid swap might be enough to change the polymerization behaviour of VIN3 *in vivo* and help to elucidate why the VIN3 VEL domain is required for repression of *FLC*.

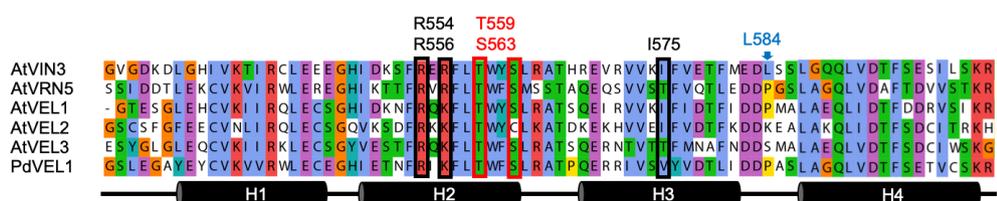


Figure 5-3 Alignment of the VEL domain. ClustalW multiple sequence alignment of the VEL domains from Arabidopsis (At) and from Phoenix dactylifera (Pd). Black boxes represent residues mutated to address polymerization. Red boxes represent potential phosphorylation targets that can block polymerization. The blue arrow indicates the position of the potential hinge residue, L584, involved in VIN3 VEL domain swapping. The barrels below the alignment represent the four α -helices that make up the VEL domain. (Adapted from Fiedler et al., 2021)

At the end of Chapter III, I discussed how the VEL domain could be required for the repression of *FLC* during vernalization. We hypothesized that this was caused by the inability of the VEL domain to polymerize; however, the VEL domain could have additional functions. Therefore, we cannot yet strictly link the missing of the VEL domain to the lack of polymerization. To test whether polymerization is the main role of the VEL domain, we designed constructs where we swapped the VEL domain of VIN3 with another polymerization domain, the DIX domain from Dishevelled (Schwarz-Romond et al., 2007). Dishevelled is a component of the conserved animal Wnt signalling pathway, which is important during development (Bienz, 2020). Similar to the requirement of the VEL domain for functional VIN3 in *FLC* repression, the DIX domain is required for Dishevelled function in Wnt signalling (Schwarz-Romond et al., 2007). As a negative control, we included the polymerization mutant in the DIX domain known as M4. It prevents polymerization and is required for the Dishevelled function (Schwarz-Romond et al., 2007). A similar approach has recently been used to show that the DIX domain from human Dishevelled can substitute for the DIX domain in the Arabidopsis protein SOSEKI (van Dop et al., 2020).

Gel filtration and Co-IP showed that two other mutations in the head surface, T559E and S562D, prevented polymerization and homo-interaction *in vitro* (M. Fielder unpublished). Interestingly, both these mutations introduce a phosphomimic residue. Furthermore, both S562 and T559 are conserved (Fig.5-3) and were predicted to be kinase targets with PhosPhAt (Durek et al., 2010). In line with a model where phosphorylation could be a regulatory element to disassembly the VEL polymer, the introduction of the phosphomimics E/D at these positions disrupted the polymerization *in vitro*. Phosphorylation plays an important role in dissolving protein condensates to facilitate accurate inheritance of the proteins to daughter cells during cell division (Rai et al., 2018). A similar role could be mediated by phosphorylation of the VEL domain. These lines

can therefore act not only as a tool to study the functional importance of polymerization of the VEL domain, but also as a starting point to start elucidating any potential regulation of polymer formation *in vivo*. It is worth noting that, in addition to stopping polymerization, the mutations also lead to lower stability in the VEL domain itself (Marc Fiedler, personal communication). This might also be an *in vivo* regulatory mechanism.

With the structure determined by Dr M. Fiedler, we learned that while the VEL domains from both VEL1 and VIN3 can form polymers, only the VEL domain from VIN3 can undergo domain swapping. The ability to undergo domain swapping is dependent on the ability to extend helix 4 of the VEL domain (M Fiedler, unpublished). Analysis of this region showed that while most VEL proteins have a rigid proline residue in the linker region between helix 3 and 4, Arabidopsis VIN3 has a more flexible leucine (Fig.5-3). The conservation of a non-proline residue in this position is not as strong as for the other features discussed in this thesis. The difference in the helix3-helix4 linker region between VIN3 and the other VEL proteins is an intriguing explanation for the different domain-swapping abilities, although it remains to be confirmed that the proline/leucine difference is the causative factor. However, this is a very likely model as the presence of a rigid proline in the hinge region of the domain-swapping domain in Dishevelled, the DEP domain, is known to stop domain swapping (Gammons et al., 2016). While the different hinge residues may explain the different behaviours of VEL1 and VIN3 polymers *in vitro*, the relevance of domain swapping *in vivo* is an open question. Therefore, to best address the relevance of putative domain swapping *in planta*, we decided to also introduce the putative domain-swapping polymorphism L584P into VIN3 in the plant system. If domain swapping is required, we would expect repression of *FLC* to be compromised in a VIN3 L58P line.

5.2.4 Mutational analysis of the *in planta* function of the composite domain

The N-terminal region of the VEL proteins, except for VEL3, contains domains involved in protein-protein interaction (E. Franco-Echevarria unpublished) and potentially also in chromatin association. By *in silico* prediction and later structural determination, it was shown that the conserved N-terminal region contains three domains arranged in proximity: a Zing-finger (ZnF), a Plant homeodomain (PHD) and a 4 α -helix bundle with some similarities to a winged-helix- (WH) or Bromo-domain. Collectively we named this a composite domain (Chapter II).

As at the beginning of the VEL domain study, we started with deletions of the individual sub-domains of the composite domain as these would be the strongest mutants. We made single deletions of the three sub-domains, as well as the full CD in VIN3 and VRN5. Initially, as discussed in Chapter II, it was thought that these domains were involved in chromatin targeting of the VEL proteins. However, biochemical analysis performed by Dr Elsa Franco-Echevarria showed that the CD of VRN5 is involved in protein-protein interactions between VRN5 and PRC2. It therefore seems plausible to expect that the CD deletions of VRN5 will result in compromised PRC2 interaction. However, the association of PRC2 with *FLC* is likely to rely on multiple factors and is not a stepwise linear process. It is therefore still possible that PRC2 may still associate with *FLC* in the VRN5 CD deletion lines. However, a consequence of the CD deletion could be reduced residence time, which might be revealed by a reduced ChIP signal of PRC2 or increased levels of H3K27me1/me2 at *FLC*, as suggested for the outcome of shorter residence times of PRC2 at mammalian targets (Højfeldt et al., 2019; Laugesen et al., 2019; Youmans et al., 2018). The EMSA analysis suggested that the CD of VIN3 can associate with nucleic acid, meaning that we cannot rule out that the atypical WH/Bromo-like domain in VRN5 is involved in nucleic acid binding or nucleosome binding through the conserved arginines as well.

The determination of the structure of the CD and the sub-domains through crystallisation and NMR by Dr Elsa Franco-Echevarria revealed that the sub-domains of the CD are required for the overall correct folding of the entire CD, as expected from their proximity. This is particularly true for the ZnF, similar to the role of the ZnF in stabilizing the fold of the PHD in UHRF1 (Hu et al., 2011). The interconnection between the sub-domains means that any phenotype observed in, for example, the ZnF deletion line cannot strictly be related to the function of the ZnF domain. However, like the VEL domain deletion, the domain deletions serve as a beginning to explore the functional role of the domains *in vivo*. Similar to the detailed analysis of the polymerization role of the VEL domain, the precise function of different parts of the CD will be elucidated with more defined point mutations.

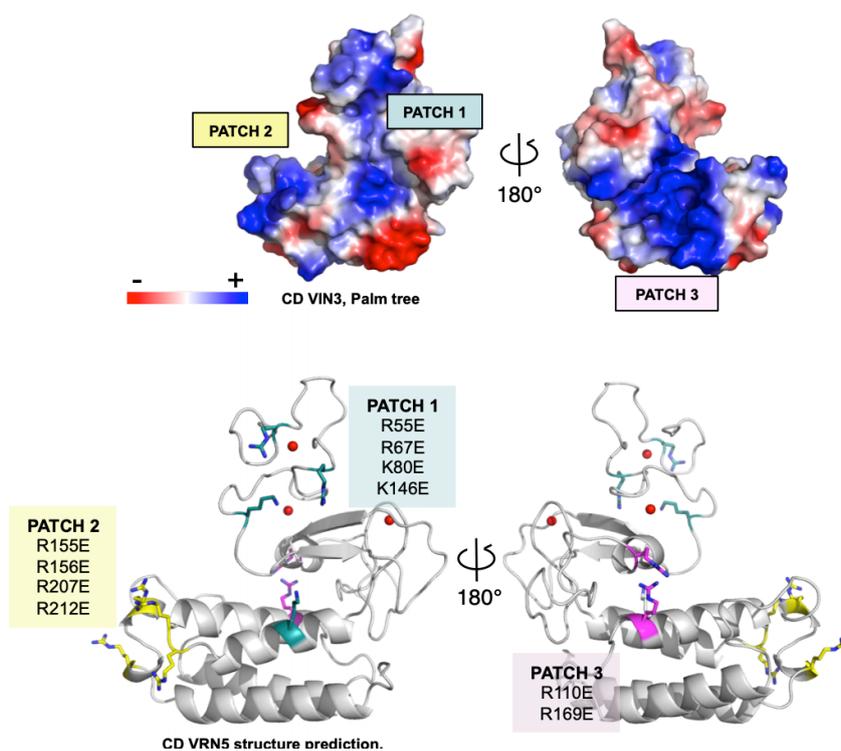


Figure 5-4 Conserved positive patches in the CD in the VEL proteins. Electrostatic map of the CD from *Phoenix dactylifera* showing the location of the three positive charge patches. The residues of *Arabidopsis* VRN5 that make up the three patches are shown in the structure below (from E. Franco-Echevarria)

From the crystal structure of the CD, we found three conserved positive charge patches on its surface (Fig. 5-4). We designed charge-swap mutations where the positively charged lysine or arginine residues were replaced by negatively charged glutamic acid for each of the three patches,

in order to test their function *in vivo* (Fig. 5-4). The mutations are named cluster 1–3 and they are located on the PHD surface (cluster 1), on the atypical WH/Bromo-like (cluster 2) and on the back of the CD (cluster 3) (Fig. 5-4). Similar to the deletions, these mutations also affect PRC2 interaction in Co-IP experiments performed in mammalian cell cultures (E. Franco-Echevarria unpublished). Previously, it had been confirmed that these point mutations do not influence the folding of the CD, as judged by gel filtration (E. Franco-Echevarria). Therefore, any observed effects are unlikely to be caused by a misfolding of the CD or its subdomains, but are rather the true effects of changing the charge of these conserved positively charged patches.

As discussed at the end of Chapter II, solving the crystal structure of the CD required the use of the Phoenix VIN3 CD. In our *in vitro* assays, the Arabidopsis and Phoenix VIN3 CDs behaved similarly (Chapter II). However, to test whether they are truly functionally equivalent domains, we made transgenic lines where we replaced the CD of Arabidopsis VIN3 (AtVIN3) with the Phoenix one. If the composite domains are functionally equivalent, we expect the Phoenix swap to complement the *vin3-1* late flowering phenotype. Because the CD is in proximity to the FNIII, it remains a possibility that the FNIII is associated with or affects the folding of the CD. Therefore, we also included a larger swap, replacing both the CD and FNIII domains in AtVIN3 with the Phoenix domains. A full list of CD-related transgenic lines is provided below.

Table 5-2 Overview of stable transgenic lines generated to address the *in vivo* function of the composite domain of VIN3 (green), VRN5 (yellow) or both (grey)

Lines	Aim
VIN3 Δ ZNF / VRN5 Δ ZnF	Independent domain deletions to address the function of each sub-domain.
VIN3 Δ PHD / VRN5 Δ PHD	
VIN3 Δ WH / VRN5 Δ WH	
VIN3 Δ CD / VRN5 Δ CD	Full CD deletion to address the role of the CD in chromatin association and protein-protein interaction. The VIN3 construct also serves as a control for the swap constructs below.
IN3 Phoenix CD swap	To address the ability of the Phoenix VIN3 CD to be substituted for the Arabidopsis one.
VIN3 Phoenix CD-FNIII swap	Same as above. However, because of the proximity of the CD and FNIII domain, the Phoenix FNII may be required for the correct folding of the Phoenix CD.
VRN5 cluster 1 R55E R67E K80E K146E	Addresses the role of the conserved positively charged patch on the PHD surface in protein-protein interaction and/or chromatin association

<p style="text-align: center;">VRN5 cluster 2 R121E R155E R156E R207E</p>	<p>Addresses the role of the conserved positively charged patch on the atypical WH/Bromo-like surface in protein-protein interaction and/or chromatin association.</p>
<p style="text-align: center;">VRN5 cluster 3 R110E R169E</p>	<p>Addresses the role of the conserved positively charged patch on the back surface of the CD in protein-protein interaction and/or chromatin association.</p>

5.2.5 Mutational analysis of the *in planta* function of the FNIII domain

The final domain in the VEL proteins is the fibronectin type-III domain (FNIII), which is located just after the CD domain. FNIII domains are found in many proteins and interestingly also in the accessory protein of the mammalian H3K9 methyltransferase SETDB1, ATF7IP. Here it has been suggested that it functions as an interaction hub for several repressor proteins, including HDACs and LSD1 (Tsusaka et al., 2020).

In the initial stages of this project, we were interested in the FNIII domain as a potential dimerization and domain swapping domain, as has previously been reported for several other FNIII domains (Gee et al., 2013; Teplyakov et al., 2014). As discussed in Chapter III, a paradigm for the formation of protein condensates is the crosslinking of head-to-tail polymers by a polymerization domain (Bienz, 2020). Our initial hypothesis was that the FNIII domain could act as the cross-linker between VEL head-to-tail polymers through domain swapping. Similar to the work in Chapters II and III, we initially tried to express the FNIII domain for structural analysis. Even though many FNIII domain structures have been solved, we were unable to obtain any crystals of the FNIII domain from the VEL proteins. Because of the novelty of the VEL and CD domains and

their complexity, we focussed the work on those domains. However, our analysis of the FNIII sequence revealed several interesting features, in particular, the conservation of a flexible glycine in the putative hinge region for domain-swapping in VIN3 and VEL1. This was not so in VRN5, where this residue has been replaced by other putative less-flexible residues (Fig. 5-5).

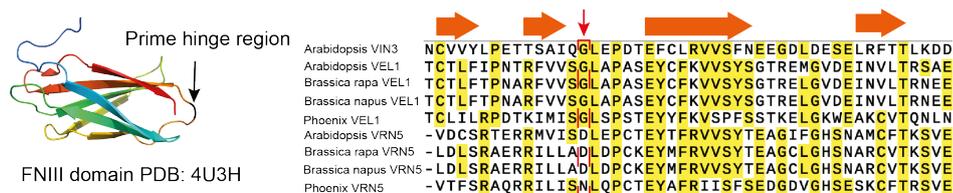


Figure 5-5 FNIII putative hinge region for domain swapping. The crystal structure of the FNIII domain from Porebski et al. (2015) (4U3H), with the prime domain-swapping hinge glycine residue marked in black. Alignment of the last region of the FNIII domain of the VEL proteins from Arabidopsis, Brassica and Phoenix, secondary structure prediction is shown above with arrows symbolising β -sheets. The red arrow indicates the hinge glycine, highlighting the conserved difference between VIN3/VEL1 and VRN5.

It remains a possibility that the FNIII domain participates in protein-protein interaction through domain swapping. Therefore, as with the potential domain-swapping mutants in the VEL domain of VIN3, we designed several mutations that would block the possible domain-swapping ability of the FNIII domain in VIN3. In particular, the replacement of the glycine with a rigid proline residue should prevent domain swapping, similar to the argument for the L584P mutant in the VEL domain, and to how domain swapping was blocked in the DEP domain in Dishevelled (Gammons et al., 2016). As indicated by the alignment in Fig. 5-5, the difference between a flexible glycine in VIN3 and VEL1 and a less flexible aspartic acid or asparagine is well conserved through evolution. The conservation of this difference is so strong that this single residue can be used to distinguish VRN5 from VIN3/VEL1 throughout evolution. This suggests that it has been important to maintain this difference between the different VEL proteins, and it is therefore interesting to study the functional relevance of this residue *in vivo*. The point mutations directly target the putative domain-swapping ability of the FNIII domain. However, as a more unbiased approach to the potentially

different roles of the VIN3 and VRN5 FNIII domains, we made constructs where the FNIII domains were swapped between VIN3 and VRN5. The transgenic lines related to the FNIII domain are summarized in Table 5-3.

Table 5-3 Overview of constructs relevant for addressing the role of the FNIII domain. Constructs in grey were made in both VRN5 and VIN3. VIN3-specific constructs are shown in green.

Lines	Aim
VIN3 Δ FNIII / VRN5 Δ FNIII	To address the role of FNIII <i>in vivo</i>
VIN3 FNIII-swap / VRN5 FNIII-swap	To test whether the FNIII domain is responsible for the different functions of VIN3 and VRN5
VIN3 G379P	A putative strong blocker of domain swapping in VIN3
VIN3 P382G	A putative mild blocker of domain swapping in VIN3
VIN3 G379D / VRN5 D306G	The conserved difference between VRN5 and the other VEL proteins is the presence of Asparagine (N) or as in Arabidopsis Aspartic acid (D) in the putative hinge region. To test the <i>in vivo</i> role, we made single-amino acid swapped lines

In Chapter III the potential destabilization of VIN3 through the deletion of the FNIII domain was discussed based on the observation in *N. benthamiana*. To determine whether a similar observation could be made in stable Arabidopsis lines, two different transgenic lines were imaged by confocal microscopy after 4 weeks of cold treatment. We could observe a

clear nuclear pattern of GFP in line 1 of VIN3- Δ FNIII-eGFP at levels comparable to those we observed in the previously published WT line of VIN3-eGFP (Fig. 5-6). In the second transgenic line, we observed very little GFP signal, most likely mainly a background signal. The different behaviour of the two transgenic lines could be due to different integration sites in the Arabidopsis genome. During the generation of the lines, we noted that where the VIN3- Δ FNIII line 1 showed a normal plant morphology, line 2 showed a more distinct phenotype, potentially due to gene disruption caused by T-DNA integration. Overall, it seems possible to express VIN3 without the FNIII domain in Arabidopsis, although western blotting is required to confirm this, enabling the investigation of the FNIII domain *in vivo*.

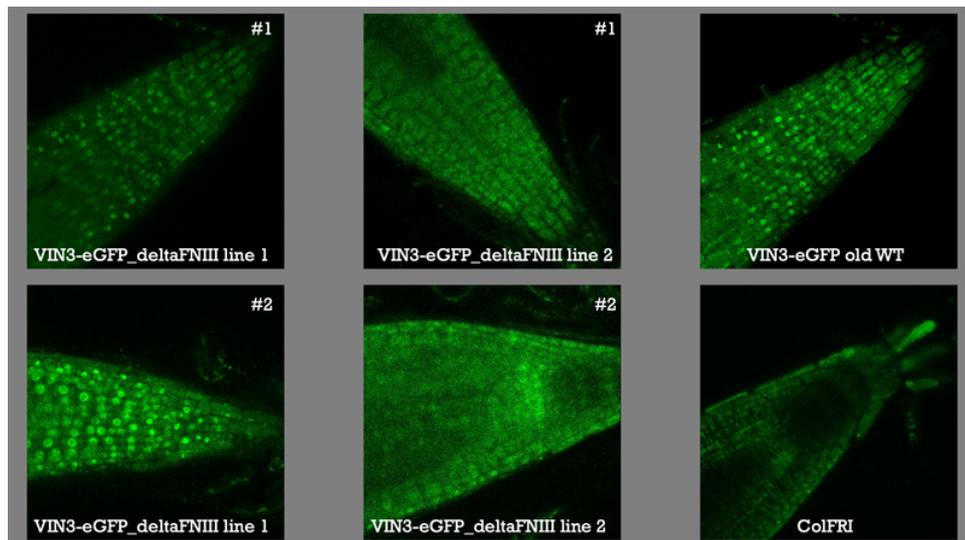


Figure 5-6 Subcellular localization of VIN3- Δ FNIII-eGFP in stable Arabidopsis transgenics. Confocal microscopy images of VIN3-eGFP WT and VIN3- Δ FNIII-eGFP in Arabidopsis root tips.

5.2.6 Analysis of interference of the fluorescence tag

The VEL domains of at least VIN3 and VEL1 can form head-to-tail polymers, as discussed in Chapter III, therefore it is possible that a C-terminal large tag could interfere with polymerization. Such a phenomenon has been observed for the fusion of fluorescence proteins to other polymerization domains (Isono et al., 2013). However, examples also exist where the fusion of a GFP tag close to the polymerization

domains does not affect *in vivo* polymerization (Fiedler et al., 2011). Because of this potential hindrance to polymerization caused by other protein folds, polymerization domains are often separated from other protein domains by an unstructured region, as is the case in the VEL proteins. Because both of the previous *VIN3* and *VRN5* constructs complemented the late flowering phenotype, we assumed that the C-terminal fusion does not interfere with the functions of the nearby VEL domain. However, to test whether the relatively large GFP/sfYFP2 tag next to the VEL domain compromises the *in vivo* function of the VEL proteins, we aimed to create N-terminal-tagged *VIN3* and *VRN5*, as well as versions where we used a smaller tag. At the time of writing this thesis, we have analysed individual sfYFP2-*VRN5* T2 lines by confocal microscopy. Based on the nuclear pattern, the lines do not behave differently from *VRN5*-YFP with regard to the formation of small foci of YFP fluorescence. This supports the notion that the C-terminal fused GFP/YFP tag does not interfere with the function of the VEL domain *in vivo*.

As a precautionary measure and despite the previously published lines where *VIN3* and *VRN5* C-terminal fusions complemented the mutant phenotype (Greb et al., 2007; Qüesta et al., 2016; Yang et al., 2017), we also plan to generate N-terminal fusions.

5.2.7 Overview of transgenic lines

At the time of writing this thesis, many lines have been generated from around 50 different constructs, many of which were described above. The tables below summarize the lines created and how many transformants have been obtained, as well as the number of single-copy lines.

Table 5-4 VIN3 stable transgenic lines. The number of obtained T1 individuals and single-copy lines are given. All lines carry a C-terminal fused eGFP tag.

Line	T1 individuals	Single copy lines
VIN3 WT	9	5
VIN3 ΔZnF	19	8
VIN3 ΔPHD	10	
VIN3 ΔWH	4	
VIN3 ΔCD	12	5
VIN3 ΔFNIII	2	
VIN3 ΔVEL	12	2
VIN3 ΔC-term	8	2
VIN3 Phoenix	9	3
CD swap		
VIN3 Phoenix	10	1
CD-FNIII swap		
VIN3 VRN5	13	4
FNIII swap		
VIN3 G379D	10	3
VIN3 G379P	23	12
VIN3 P382G	15	9
VIN3 R556D	49	18
VIN3 I575D	6	0

Table 5-5 VRN5 stable transgenic lines. The number of obtained T1 individuals and single-copy lines is given. All lines unless specified carry a C-terminal fused sfYFP2 tag, except specified otherwise. **Cluster 2 mutations, ***Cluster 3 mutations

Line	T1 individuals	Single copy lines
VRN5 WT	51	13
VRN5 ΔZnF	9	4
VRN5ΔPHD	14	5
VRN5ΔWH	10	2
VRN5ΔCD	6	
VRN5ΔFNIII	6	
VRN5ΔVEL	96	22
VRN5 VIN3 FNIII	13	2
swap		
VRN5 R121E R155E	5	
R156E R207E**		
VRN5 R110E R169E***	3	1
VRN5 D306G	10	2
VRN5 R544A R546D	26	9
VRN5 T565D	47	3
sfYFP2-VRN5	8	1

As discussed in Chapters II and III, the study of the function of the domains in VIN3 and VRN5 took place in close collaboration with structural biologists and biochemists at MRC LMB. Because of the long time frame required to establish plant transgenic lines, the domain boundaries used in the transgenic lines might differ from the ones later used for *in vitro* characterization. For clarity, Table 5-6 summarizes the boundaries used for the *in vivo* constructs at the protein level.

Table 5-6 Domain boundaries used for the stable transgenic lines. Lines carrying a deletion of the identified functional domains were made in Arabidopsis. The interval identified the region deleted in each line, with the listed residues included in the domain deletions.

Domain	VIN3 boundaries	VRN5 boundaries
ZnF	SRKK-DACG	KKSN-CKRS
PHD	SSCH-KDND	SCCV-KVSQ
WH	CWRK-MTVR	AKEA-AGEL
Composite	SRKK-MTVR	KKSN-AGEL
FNIII	MTVR-AGDQQ	AACR-VEIL
VEL	DKDL-KLWH	SSSI-NGVM
C-term	CLKLWH	N/A

To analyse the role of each domain, we designed lines where we deleted the domain of interest. Removing parts of proteins can affect the nuclear localization or result in unstable proteins. To monitor that, the constructs were stably expressed and maintained nuclear localization. The subcellular localization of the new VIN3-eGFP and VRN5-SYFP2 lines was observed by confocal microscopy, in epidermal root tip cells of T2 plants. When the interaction between VIN3 and VRN5 was tested with transient expression in *N. benthamiana* (Chapter III), localization was also observed. Table 5-7 summarizes the observations in *N. benthamiana* leaves and Arabidopsis roots. Accurate determination of protein levels and the size of the expressed protein requires western blotting. However, because of their low abundance, western blotting of VIN3 and VRN5 requires an enrichment step, so this is not performed until the homozygous T3 generation.

Table 5-7 Nuclear localization of VIN3 and VRN5 in *N. benthamiana* and *A. thaliana*. Subcellular localization of VRN5 and VIN3 mutant lines was observed by stereo (*N. benthamiana*) or confocal microscopy (*A. thaliana*). N/A means that the respective constructs were not transformed into the respective species. A blank box means that microscopy has not been performed at this stage.

Line	<i>N. benthamiana</i>	<i>A. thaliana</i>
VIN3	Nuclear	Nuclear
VIN3 ΔFNIII	Nuclear	Nuclear
VIN3 ΔVEL	Nuclear	Nuclear
VIN3 R556D	N/A	Nuclear
VIN3 Δlinker+VEL	Cytoplasmic	N/A
VIN3 ΔFNIII+linker+VEL	Cytoplasmic	N/A
VIN3-eGFP_G379P	Nuclear	
VIN3-eGFP_G379D	Nuclear	
VIN3eGFP R554AR556A	Nuclear	n/a
sfYFP-VRN5	n/a	Nuclear
VRN5	Nuclear	Nuclear
VRN5 ΔFNIII	n/a	Initially we could not observe fluorescence
VRN5 RRAD	n/a	Nuclear
VRN5 ΔVEL	Nuclear	Nuclear

At this stage, all the imaged lines maintained a nuclear GFP or YFP signal, indicating that the proteins are made and localized like WT. Thus, we should be able to conclude that mutated domains and not destabilization or wrong localization of the proteins is the cause of any phenotype. However, as mentioned, we cannot rule out the possibility that the signal observed originates from degraded protein, until western blotting is complete.

5.3 Summary

Throughout this PhD, a large number of transgenic lines from 50 different constructs have been or are in the process of being generated. The lines generated as part of this PhD will help elucidate the molecular mechanism of the VEL proteins in the epigenetic switching at *FLC*. Molecular

analysis, for example protein-ChIP and proteomics, is likely to reveal the specific contributions of individual domains. In addition, the specific point mutations will enable us to pinpoint the exact role of the different features we have discovered over the last four years, for example, polymerization of the VEL domain, putative domain swapping in the VIN3 VEL and/or FNIII domain, and nucleic acid binding of the CD domain. Deletions and mutations like the ones generated here have greatly increased our understanding of the core PcG proteins in mammalian cells (Højfeldt et al., 2018; Long et al., 2020; Youmans et al., 2021), highlighting the value of introducing and studying such mutations in the *in vivo* context.

The polymerization ability of the VEL domain was clear relatively early in this project, enabling us to design many specific point mutations. However, for the elucidation of the function of the CD, we have so far only a relatively good understanding for the function of the CD in VRN5. This means many more specific point mutations are likely to be generated in the future to separate the potential different functions of the CD. Finally, we still have very little knowledge of the role of the FNIII domain and therefore more point mutations are required to elucidate its function.

These lines should also provide an excellent starting point for studying the role of the VEL proteins in gene regulation globally in Arabidopsis. It is interesting to speculate why only the cold-induced VIN3 seems to carry a VEL domain that can undergo domain swapping. Is this form of polymerization only required for a subset of VEL targets? If so, what is the mechanism behind this special requirement? Similarly, the molecular analysis suggests that the VRN5 VEL domain is not as strong a polymerization domain as the other VEL domains. These findings raise the intriguing question of whether a subset of VRN5 target genes does not require the formation of a VEL polymer.

For the generation of transgenic lines, we focused on the central question of epigenetic switching at *FLC*. We only made the Arabidopsis Phoenix

swap because of the need to use the Phoenix sequence for structural determination. However, additional swaps could be made to address the conservation of the functional features of the VEL proteins throughout the plant kingdom. As discovered in the analysis of the VEL domain, a few amino acid differences can lead to changes in the behaviour of the VEL domain, for example the strength of the polymerization or the ability to undergo domain swapping. It would be valuable to investigate why certain VEL proteins have adopted specific features, as for example hyper-polymerizing VEL domains.

Chapter VI

6 General discussion

The vernalization proteins VIN3 and VRN5 were identified as mutants that were affected in the ability of prolonged cold to accelerate flowering and stably repress *FLC* (Greb et al., 2007; Sung and Amasino, 2004; Sung et al., 2006). Proteomic studies showed that VIN3 and VRN5, together with their homolog VEL1, form a complex with core PRC2 components (De Lucia et al., 2008; Wood et al., 2006). This led to the definition of the VEL proteins as PRC2 accessory proteins, similar to the PCL proteins in animals (Wood et al., 2006). Since their discovery in the 2000s, the molecular mechanism of the VEL proteins has remained mostly elusive. Thus, in my work, through *in planta* and *in vitro* work, I aimed to uncover the molecular role of VEL proteins during vernalization. To fully understand the fundamental molecular mechanism, I teamed up with structural and biochemical experts, whose expertise helped me identify some of the intrinsic properties of the VEL proteins. The project revealed many new functions of the VEL proteins, and these have also been reported for different animal PRC2 accessory proteins. Neither the sequence nor the structure is conserved between the plant and animal accessory proteins, suggesting that they have potentially evolved through convergent evolution to achieve a similar functionality.

6.1 The VEL proteins – part of a plant PhoRC?

One interesting observation from our work is the discovery of the VEL domain as a polymerization domain. To our knowledge, none of the animal PRC2 accessory proteins have domains that can mediate polymerization, although PHF1 is involved in stabilizing a PRC2 dimer (Chen et al., 2020). On the other hand, the PhoRC complex contains proteins with a SAM polymerization domain. This raises the question of whether the VEL proteins should be thought of as part of a plant PhoRC complex rather than being considered PRC2 accessory proteins. PhoRC includes the DNA-binding protein Pho, and Arabidopsis VAL1 can be considered as the functional homologue of Pho. However, in our proteomic analysis we never identified VAL1 when the VEL proteins were used as baits and reciprocally we never identified the VEL proteins when VAL1 was used

as bait, making it difficult to assign the VEL proteins to a plant PhoRC-like complex. Nevertheless, Co-IP performed in mammalian cells showed that VIN3 can co-immunoprecipitate VAL1 (M. Fiedler et al., submitted). There are other interactions where Co-IP gives a positive result, but those proteins are not found in steady state proteomics. For example, as discussed in Chapter II, VAL1 can interact with both core PRC2 components (SWN/CLF) and with the putative PcG accessory protein LHP1 when assayed through Yeast-2-hybrid or Co-IP assays (Yuan et al., 2020, 2016). On the other hand, these were not found in immunoprecipitation followed by mass spectrometry (IP-MS) experiments.

Interestingly, a similar situation occurred during characterization of the *Drosophila* PhoRC; this was not identified by classic proteomics but through IP followed by western blotting or by BioID (Frey et al., 2016; Strübbe et al., 2011). The latter is a technique that allows the detection of weak or transient interactions. Similarly, a mammalian PhoRC equivalent has not been identified *in vivo* through proteomic analysis, but *in vitro*, it has been shown that the Pho homolog YY1 can bind to the SAM domain proteins L3MBTL2, MBTD1, and SFMBT2, albeit with 50-fold lower affinity (Alfieri et al., 2013). It therefore seems plausible to hypothesize that the interaction between VIN3 and VAL1 *in vivo* could be weak and potentially transient.

As mentioned in Chapters II and III, the assumption that trans factor recruitment at the nucleation region is a linear process through direct hierarchical protein-protein interactions is likely to be an oversimplification. One could therefore imagine that the VAL1/2-VIN3 interaction is transiently required after short cold exposure in order to nucleate VIN3, potentially while the transcription levels are only slightly repressed. After VIN3 polymerization, the VEL-domain mediated feedbacks, together with higher avidity for nucleic acid, would be enough to maintain the association of VEL-PRC2 with the nucleation region at

FLC. This would allow VAL1 to interact with other factors required for *FLC* repression, for example LHP1 (Yuan et al., 2016). Combinatorial interactions to mediate PhoRC-PRC1 association with chromatin in *Drosophila* have also been proposed as a way to explain the attenuation of Polycomb binding on actively transcribed loci, despite the constant binding of Pho (Kahn et al., 2014). Likewise, VAL1 has been reported to be associated with *FLC*, also in non-vernalized conditions where the locus is highly transcribed (Yuan et al., 2016). Further work is therefore required to fully dissect the interactions required for nucleation, particularly validation of the VAL1-SWN and VAL1-VIN3 interactions *in planta* at different stages. Co-IP followed by western blotting experiments would support the conjecture that VAL1-VEL proteins are a plant functional equivalent to PhoRC. It would also be interesting to test whether VIN3 or core-PRC2 association at *FLC* is attenuated after short periods of cold or during natural conditions where *VIN3* levels are significantly lower than laboratory conditions (Hepworth et al., 2018). In both circumstances, VIN3-PRC2 association with *FLC* may rely more on VAL1 than after six weeks of constant cold when *VIN3* levels are relatively high. The protein feedbacks mediated by VEL domain polymerization after such a long cold period might also be enough to mediate VIN3-PRC2 association with *FLC* without VAL1 functionality. This could explain why VIN3 and SWN occupancy are unaffected in a *vall* background as shown in Chapter II for VIN3.

6.2 Protein memory

Much of the rationale for part of my project originated from the finding of the persistence/memory of the repressed state at *FLC* in the absence of a broad H3K27me3 domain (Yang et al., 2017). For example, in the spreading mutant *lhp1*, H3K27me3 does not accumulate to a high degree beyond the ~3 nucleosome nucleation region. However, the repressed state of *FLC* is maintained for ~10 cell cycles (Yang et al., 2017). Through mathematical modelling it was shown that if only read-write mechanisms exist for the maintenance of H3K27me3, and assuming random inheritance

of H3K27me3 during cell division, *FLC* should react significantly faster than observed experimentally. This suggests that additional memory elements must exist and it has been suggested that these are achieved through protein feedback (Yang et al., 2017). Additionally, a recent study in budding yeast also observed that the rate of loss of silencing was not dependent on nucleosome numbers (Saxton and Rine, 2019). In a reciprocal study it was likewise shown that despite broad H3K27me3 domains, the PRE is required for maintaining repression through cell division (Laprell et al., 2017). Together these studies point to stable epigenetic repression relying on more than just the inheritance of H3K27me3 histones and read-write mediated feedback.

In our initial analysis, we discovered that the VEL domain of VIN3 is also required for the silencing during cold, suggesting that protein feedback is not only potentially required for memory post-cold but also for the initial establishment of the repressed state. This could potentially be because the initial deposition of H3K27me3 is slow due to lack of H3K27me3 allosteric activation of PRC2. Therefore, the initial nucleation of H3K27me3 requires the formation of a VIN3-mediated VEL oligomer to act as a scaffold and increase the local concentration of PRC2. This would increase overall residence time of PRC2, allowing the initial establishment of H3K27me3. Interestingly, a theoretical study has independently reached the same conclusion that protein feedback through oligomerization is required for the initial nucleation of H3K27me3 (Lövkvist et al., 2021 accepted).

While not strictly similar to protein memory by prions (Harvey et al., 2018), the domain-swapping ability of the VEL domain in VIN3 does suggest that conformational changes in a protein could be an important part of the molecular switching process at *FLC*. Protein memory has been linked to the regulation of *FLC*, as many autonomous pathway components were predicted to contain prion-like domains, including the key components FCA, FPA and LD (Chakrabortee et al., 2016). Resembling

the hypothesis for the vernalization pathway, it was speculated that a conformational change in LD could lead to the formation of larger self-perpetuating protein assemblies with changed protein activity (Chakrabortee et al., 2016). Recent work showed that intrinsically disordered regions in FCA are important for a phase transition to give liquid-like body formation *in vivo*. These condensates are likely to concentrate other 3' end processing factors and enhance 3' end processing (Fang et al., 2019). This is conceptually very similar to the model for the nucleation of H3K27me3 that we propose for the VEL-PRC2 oligomer.

6.3 Avidity as a general concept

The targeting of PRC2 to its target site in the genome has been a long-standing question in Polycomb systems generally. With the exception of strong PREs in *Drosophila*, no direct recruitment model can explain the recruitment of PRC2. Recent advances in structural biology and live-imaging have revealed the existence of multiple interactions between PRC2 and the RNA and chromatin at target sites (Kasinath et al., 2021).

Our understanding of how proteins find and assemble at specific sites in the genome has evolved in the past years. Similar to the discussion of the function of the VEL domain in the VEL proteins, many other processes in gene regulation seem to rely on membraneless condensates, which are formed by protein and nucleic acids coming together through a range of multivalent interactions (Sabari et al., 2020). Similarly, both transcriptional shutdown at *FLC* and the 3' end processing of *FLC* rely on nuclear condensates (Zhu unpublished; Fang et al., 2019).

The mechanism of increased local concentration is now a common model for enhanced PcG activity at target sites. Oligomerization of the deubiquitinase Calypso in PR-DUB leads to the increased concentration of PR-DUB and resulting in efficient removal of H2AK119Ub (Foglizzo et al., 2018). Likewise, phase separation of Ph traps PRC1 and increases the ubiquitin ligase activity towards chromatin (Seif et al., 2020).

Polymerization has not been reported for PRC2, but it has been shown to form dimers (Chen et al., 2020) and these engage with chromatin through multivalent interactions.

Overall, this suggests that Polycomb repression, like many other biological processes, relies on a series of interactions rather than a hierarchical recruitment mechanism. This might explain the wild-type VIN3 binding at the nucleation region in a *val1* mutant.

6.4 The VEL proteins beyond flowering time control

An interesting observation from our work was the surprising and previously unappreciated strong conservation of the VEL proteins. In this work we identified VEL proteins from distantly related species to overcome some of the complications of working with Arabidopsis proteins. For example, we expressed the CD as described in Chapter II. We showed that the features of the VEL proteins are conserved; the association of the VIN3 CD with nucleic acid was conserved in the VIN3 CD from Phoenix (Chapter II). Similarly, the polymerization property is widely conserved (Fiedler et al., 2021). This shows that the VEL proteins are likely to play a role in gene regulation beyond vernalization and flowering time control. In Arabidopsis, the VEL proteins have mostly been studied for their role in flowering time regulation at *FLC*. *VRN5* and *VEL1* are also involved in the repression of other *FLC* clade members (Kim and Sung, 2010; Sung et al., 2006). However, recently, *VRN5* has also been implicated in the accurate response to high temperatures (Bordiya et al., 2020). This suggests that, as expected for a general PRC2 accessory factor, the VEL proteins are involved in several gene regulation pathways. This is also suggested by our VIN3 ChIP-seq analysis, which revealed that VIN3 has many other targets in the Arabidopsis genome in addition to *FLC*. Genome-wide studies involving, for example, RNA- and ChIP-seq of the changes in *vel* mutants, as well as genome-wide identification of VEL protein targets, are missing from the Arabidopsis literature. This is due to their relatively weak association to chromatin, making ChIP-seq

experiments difficult. To understand what determines VEL-PRC2 target genes and to try and elucidate why certain genes rely on polymerization, we need to perform these genome-wide analyses. In addition, the tools developed here as part of this thesis could determine whether certain VEL targets are independent of polymerization. Likewise, studies of the animal PRC1 SAM domain containing protein Ph indicated that a fraction of the Ph-regulated genes were independent of the polymerization activity of the SAM domain (Wani et al., 2016).

While studies of genome-wide changes in *Arabidopsis vel* mutants are missing, a recent study profiled the changes in H3K27me3 in a “*vrn5*” mutant in Tomato. In the *crawling elephant (crel)* mutant, almost half of the H3K27me3 covered genes lost the enrichment for H3K27me3 (Shwartz et al., 2021), showing how “VRN5” works as a general PRC2 accessory factor with much broader roles than just flowering time control.

6.5 Conclusions

This project started with the aim of understanding the molecular mechanism of the VEL proteins. While I came a long way in terms of uncovering some of the intrinsic features of the VEL proteins, my studies continuously discovered new complexities in the VEL proteins. One important lesson learned from this study is that despite the lack of sequence conservation, convergent evolution has led to the evolution of similar mechanisms in different organisms. This shows that the lessons learned from the study of Polycomb regulation at *FLC* can be applied to Polycomb systems beyond plants.

Overall, I propose that the VEL proteins work like the SAM domain proteins in animals. Through the VEL domain, the VEL proteins form larger polymers that increase the activity of PRC2 towards chromatin. This is achieved by concentrating PRC2 at the nucleation region and by increasing the residence times of PRC2 at chromatin through the nucleic acid binding property of VIN3.

The stable transgenic lines we are currently developing will help to reveal the core molecular function of the VEL proteins, as well as the contribution of each of the features we have discovered so far.

Chapter VII

7 Materials & Methods

7.1 Materials

7.1.1 Plant material

All plant lines used in this thesis were in the *Arabidopsis thaliana* Columbia background. All mutants – *vin3-4*, *val1-2*, *vin3-1*, *vrn5-8*, *ntl8-D3* and *FLC_ΔCOOLAIR* – have been described before and carry an active *FRI* allele from *FRI^{SF2}* (Table 8-1). The transgenic tagged lines *VIN3-GFP/vin3-4 FRI* and *VRN5-YFP/vrn5-8 JU233* have likewise been described before. All previously described plant material were obtained from other members of the Dean lab or from the Dean lab seed stock. All of the new transgenic lines generated as part of this thesis are described in detail in Chapter V.

Table 7-1 Arabidopsis lines used in this thesis

Plant line	Reference
<i>FRI^{SF2}</i>	Lee et al., 1994
<i>vin3-4</i> (SALK_004766)	Bond et al., 2009
<i>vin3-1</i>	Sung and Amasino, 2004
<i>val1-2</i> (SALK_088606)	Suzuki et al., 2007
<i>vrn5-8</i> (SALK_136506)	Greb et al., 2007
<i>ntl8-D3</i>	Zhao et al., 2020
<i>FLC_ΔCOOLAIR</i>	Luo et al., 2019

<i>VIN3-GFP/vin3-4 FRI</i>	Qüesta et al., 2016
<i>VRN5-YFP/vrn5-8 JU223</i>	Greb et al., 2007; Yang et al., 2017

7.1.2 Bacterial strains

7.1.2.1 Escherichia coli

DH5 α : *F*⁻, *supE44*, *endA1*, *hsdR17* (*r_k*, *m_k*⁺), *thi-1*, λ ⁻, *relA1*, *recA1* Δ *lacU169* (Φ 80 *lacZ* Δ M15) *gyrA96*

Mach1: *F*⁻ ϕ 80(*lacZ*) Δ M15 Δ *lacX74* *hsdR*(*rK-mK*+) Δ *recA1398* *endA1* *tonA*. Mach1 was used for general cloning. Mach1 has a faster doubling time than DH5 α

ccdB survival 2, *F*⁻*mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ 80*lacZ* Δ M15 Δ *lacX74* *recA1* *ara* Δ 139 Δ (*ara-leu*)7697 *galU* *galK* *rpsL* (Str^R) *endA1* *nupG* *fhuA::IS2*

The *ccdB* survival strain is resistant to the gene product of the *ccdB* gene, and it is required for the propagation of gateway vectors, for example the pSLJ-DEST family

BL21 (DE3): *fhuA2* [*lon*] *ompT* *gal* λ *sBamHI*o Δ *EcoRI-B* *int::(lacI::PlacUV5::T7 gene1)* *i21* Δ *nin5* [*dcm*] Δ *hsdS*. BL21 (DE3) was used for protein expression.

BL21 (DE3) pRARE2: BL21 (DE3) carrying pRARE2, which provides phage resistant. BL21(DE3) was used for protein expression.

7.1.2.2 Agrobacterium tumefaciens

C58C1: C58 cured of the TI plasmid pTiC58 carrying pGV2260, Rif^R. C58C1 was used for stable plant transformation with pSJ-based vectors.

GV3101: C58 with the TI plasmid (pMP90 pTiC58DT-DNA), Rif^R, Gent^R. GV3101 was used for transient transformation of *N. benthamiana* leaves with pCAMBIA1300-based constructs.

7.1.3 Yeast strains

AH109: *MATa*, *trp1-901*, *leu2-3, 112*, *ura3-52*, *his3-200*, *gal4Δ*, *gal80Δ*,
LYS2::GAL1_{UAS}-GAL1_{TATA}-HIS3, *GAL2_{UAS}-GAL2_{TATA}-ADE2*, *URA3::*
MEL1_{UAS}-MEL1_{TATA}-lacZ

7.1.4 Plasmids

pSLJ-Dest, binary vector based on pSLJ755I5, containing gateway recombination sites. The plasmid carries the suicide gene *ccdB* and a resistance gene for chloramphenicol. (Previously generated in the Dean lab)

pSLJ755I5, binary vector conferring resistance to Phosphinothricin in transformed plants (Jones et al., 1992)

pSLJ-Dest (Hyg^R), binary vector based on pSLJ6999 containing gateway recombination sites. The plasmid carries the suicide gene *ccdB* and a resistance gene for chloramphenicol.

pSLJ6999, binary vector conferring resistance to hygromycin in transformed plants (Jones et al., 1992)

pBsKSII-FLC15, pBluescript II KS (+/-) (Stratagene) containing a 20 kb region around *FLC* (Crevillén et al., 2013)

pBsKSII FLC-RBSC, pBluescript II KS (+/-) (Stratagene) containing a 15 kb region around *FLC*. The *FLC* terminator was replaced with the RBSC terminator

pCAMBIA1300, conferring Hygromycin resistance in transformed plants

pGADT7-rec, yeast expression plasmid, for the fusing of a gene of interest (GOI) with the *GAL4* activation domain. The plasmid further carries the sequence for a HA-tag, enabling expression to be tested by western

blotting. Plasmid confers resistant to ampicillin in *E. coli* hosts and contains the selectable marker *LEU2* for selection in appropriated yeast strains.

pGBKT7, yeast expression plasmid, for the fusing of a gene of interest (GOI) with the GAL4 DNA binding domain. The plasmid further carries the sequence for a c-Myc-tag, enabling expression to be tested by western blotting. Plasmid confers resistant to kanamycin in *E. coli* hosts and contains the selectable marker *TRP1* for selection in appropriated yeast strains.

7.1.5 Antibodies

Table 7-2 List of used antibodies

Antigen	Antibody	Application	Description
GFP	Abcam ab-290	ChIP	Rabbit Polyclonal
GFP	Roche, 11814460001	Immunoblot	Mixture of two mouse monoclonal
HA	Cell Signalling, HA-Tag, C29F4	Immunoblot	Rabbit monoclonal
FLAG	Sigma ANTI-FLAG M2 (F1804)	Immunoblot	Mouse monoclonal
Mouse IgG	GE healthcare, NXA931V	Immunoblot	Sheep IgG linked to horseradish peroxidase
Rabbit IgG	Ge healthcare, NA934V	Immunoblot	Donkey IgG linked to horseradish peroxidase

7.1.6 Restriction enzymes

Table 7-3 List of enzymes used for 3C

Name	Source	Identifier
BamHI	New England Biolabs (NEB)	R0136M
BglII	NEB	R0144M
DpnII	NEB	R0543M
NlaIII	NEB	R0125L
TaqI α *	NEB	R0149M
T4 Ligase HC	Promega	M1794

*TaqI α is now named TaqI-v2

7.1.7 Standard solutions

10x M9 salts: 422.5 mM Na₂HPO₄, 220.6 KH₂PO₄, 85.5 mM NaCl, and 46.7 mM NH₄Cl.

M9 media: 200 μ l Distilled water agar, 200 μ l 1M MgSO₄, 200 μ l 0.1 M CaCl₂, 200 μ l 10 mg/ml thiamine, 2 ml 20% glucose, 20 ml 10x M9 salts.

10x TE: 0.1 M Tris-HCl, 10 mM EDTA, pH 7.5

10x LiAc: 1 M Lithium acetate, pH 7.5 with acetic acid

7.2 Methods

7.2.1 Plant growth

Seeds were surface sterilized and sown on 1x Murashige and Skoog (MS) media without glucose, stratified for 2–3 days at 4°C and grown for 10 days under long day conditions (16h light, 8h dark at 20°C). For vernalization, the treatment plates were transferred to short day conditions (8h light, 16h darkness at 5°C). Plants harvested 10 days after vernalization were transferred back to long day conditions and harvested from the plates. For longer post-vernalization treatment (more than 10 days), the plants were transferred to soil and grown under long day conditions.

Seedlings grown for confocal microscopy were grown vertically on MS media supplemented with 1% glucose containing 1% Agar.

7.2.2 RNA extraction

Whole seedlings were used for RNA extraction and RNA was extracted using the published Mini hot Phenol procedure as described in Box et al. (2011). Around 100 mg of plant tissue was collected and frozen in liquid nitrogen and grinded to fine powder. 250 μ l acidic phenol was added to the tubes, together with 5 μ l β -mercaptoethanol and 500 μ l homogenisation buffer (100 mM Tris-HCl pH 8.5, 5 mM EDTA pH 8, 100 mM NaCl and 0.5% SDS). The tubes were shaken for 15 min. 250 μ l chloroform was added and the tubes were shaken for a further 15 min. The tubes were centrifugated at maximum speed for 10 min and the aqueous phase was recovered to a fresh tube and mixed with 550 μ l phenol:chloroform 1:1 mixture. The tubes were shaken for 10 min, centrifugated at maximum speed for 10 min, and the aqueous phase was recovered and mixed with 50 μ l 3M sodiumacetate and 400 μ l isopropanol. The mixture was then incubated at -70°C for at least 15 min. The RNA was pelleted with centrifugation at full speed for 30 min at 4°C, the supernatant was removed, and the pellet was dried on tissue paper for 10 min. The pellet was dissolved in 500 μ l DPEC water and the RNA was precipitated with 2M (final concentration) LiCl overnight at 4°C. RNA was pelleted with centrifugation at maximum speed for 30 min at 4°C. and the pellet was washed with 1 ml 80% ethanol twice. The pellet was dried and resuspended in 30 μ l DPEC water and RNA was quantified using a NanoDrop spectrometer.

7.2.3 Dnase digestion

The trace amount of DNA present after RNA isolation was removed with Turbo-DNase (Invitrogen) treatment. DNase treatment was performed in 20 μ l containing 10 μ g of RNA and 2 μ l 10x Turbo DNase buffer and 1 μ l (10u/ μ l) TURBO DNase. Digestion was performed at 37°C for 30 min.

Dnase was inactivated by adding 2 μ l inactivation reagent and incubated at room temperature for 5 min with occasional mixing. The resin was pelleted by centrifugation at 10.000 g for 1.5 min and the supernatant was recovered. RNA concentration after DNase treatment was quantified with NanoDrop.

7.2.4 Complementary DNA (cDNA) synthesis

The Invitrogen superscript cDNA synthesis kit (III or IV) was used according to the manufacturer's instructions. 1.5-2 μ g of total RNA was mixed with 0.5 μ l 10 mM dNTPs and 2 pmol of gene-specific primer or 0.5 μ l 50 μ M oligo(dT) and water to a final volume of 7 μ l and incubated at 65°C for 5 min. The mixture was cooled on ice for 1 min and 1 μ l 5X buffer, 0.5 μ l 0. mM DTT, 0.25 μ l RNaseOUT and 0.25 μ l Reverse transcriptase were added. The reaction was incubated at 50°C (55°C for gene-specific primers) for 30–60 min. The reaction was stopped by incubation at 70°C for 15 min. For superscript IV a similar reaction set-up was used but the reaction was incubated at 50°C for 10 min, followed by 55°C for 2 min. The reaction was stopped by incubation at 80°C for 10 min. As a control for DNA contamination, a similar procedure was performed without adding reverse transcriptase.

7.2.5 Quantitative PCR (qPCR) analysis

The prepared cDNA could be used for qPCR analysis. The cDNA was diluted to 16 ng/ μ l and 5 μ l (total 80 ng cDNA) was used in each reaction. Primers were added at a final concentration of 0.5 μ M (0.05 μ l of 100uM stock) together with 5 μ l Roche Lightcycler 480 SYBRGreen I master mix. The mixture was mixed and spun down, and qPCR was performed on a Lightcycler 480. In most cases a standard program was used: 5 min at 95°C preincubation and 50 cycles of 95°C for 15 sec, 59°C for 20 sec and 72°C for 25 sec.

Expression data are shown as the average of at least three biological replicates and three technical qPCR replicates. Target amplicons were normalized to the geometric mean of PP2A (At1g13320) and UBC (At5g25760)

7.2.6 Cloning

Cloning of constructs for the generation of stable transgenic lines was performed with seamless megaprimer cloning or quick-change mutagenesis. The plasmids containing the wild-type *VRN5* and *VIN3* genomic sequence in a pENTR vector were used as templates. Deletion and point mutations were made by quick-change mutagenesis PCR. For domain swaps and moving of tags, seamless megaprimer cloning was used.

7.2.7 Seamless megaprimer cloning

Seamless megaprimer also named Restriction-free (RF) cloning was used for the majority of the cloning work. Amplicons carrying the insert of interest were amplified with typically ~50 bp primers using Phusion High-fidelity polymerase (NEB). The forward primer contains a ~15 bp overlap with the vector at the 5' end of the desired insertion site, and the reverse primer likewise contains a ~15 bp overlap complementary to the 3' end of the site of insertion. PCR product was checked with agarose electrophoresis and the correct product was gel purified. Amplicons was inserted by a PCR reaction similar to QuickChange PCR. 150 ng of PCR product was used in a second PCR reaction with 50 ng of target plasmids in a 50 µl reaction. An 18 cycle PCR program (1 min annealing at 55°C and 1.5-2 min/kb elongation (temperature depended on polymerase) was used to insert the amplicon. Once completed, the parental methylated plasmid was digested by 1 µl DpnI for 80 min at 37°C. 1-2 µl of the reaction was transformed into electrocompetent *E. coli* cells.

7.2.8 Generation of binary pSLJ clones by LR reaction

The LR clonase Kit was used, following the manufacturers' instructions. 150 ng of entry clone was mixed with 150 ng pSLJ-DEST (prepared with MidiPrep), and TE pH 8 buffer was added to a final volume of 8 μ l. 2 μ l LR clonase II enzyme mix was added, and the reaction was mixed by vortexing and spun down. The reaction was incubated at 25°C for 1–2 hours. To inactivate the enzymes, 1 μ l Proteinase K was added, and the reaction was incubated at 37°C for 10 minutes. 1–2 μ l was transformed into electrocompetent *E. coli* cells. Positive colonies were selected on LB plates containing 10 μ g/ml Tetracycline.

7.2.9 Triparental mating

The binary vector was move into Agrobacterium using triparental mating. Agrobacterium C58 (Rif^R), the helper *E. coli* strain HB101 (pRK2013) (Kan^R) and the donor pSLJ (Tet^R) *E. coli* strain were grown from single colonies. Triparental mating was performed by mixing 800 μ l C58 culture with 200 μ l culture HB101 and 200 μ l construct culture (use LB for negative control). The mixture was mixed by vortexing and it was centrifugated at 6000 rpm for 2 min and spread on LB plates and incubated overnight at 28°C. The plates should have even growth the day after, and 5–6 scrape of cells was resuspended in 1 ml 10 mM MgSO₄. 10⁻² and 10⁻⁴ dilutions were made and 20 μ l of each dilution was spread on LB-Rif-Tet (1 μ g/ml) plates and incubated for 2–3 days at 28°C. Single colonies from the 10⁻⁴ dilution were checked by colony PCR and diagnostic digestion. Positive colonies were streaked out on M9 plates with thiamine and Tet (1 μ g/ml).

7.2.10 Glycerol stocks

Bacterial and yeast strains were preserved in 1.5 ml of 20% (final concentration) glycerol and stored at -70°C.

7.2.11 Floral dip of Arabidopsis

Arabidopsis plants were grown under long day conditions. For *vrn* mutants, the plants were vernalized for 6 weeks before moving to a long day glasshouse. Plants with short inflorescence were used for transformation. Agrobacterium carrying the construct of interest were grown in 10 ml cultures (Rif+tet) and 200 μ l was used to start a 200 ml overnight culture (Tet). Agrobacterium was grown at 28°C with 220 rpm. Cells were harvested with centrifugation for 10 min at 3500 g and resuspended in dipping medium (5% Sucrose, 0.025% Silwet-77). The suspension was transferred to small plastic bags and the aboveground tissues were submerged in the bacterial suspension and gently mixed. Plants were placed in black bags overnight to maintain humidity. Transformation was normally repeated once per week for at least three weeks and the plants were left to set T0 seeds.

7.2.12 Selection of transformed Arabidopsis

The binary vector pSLJ-75515 carries the bialaphos resistance gene, Bar from *Streptomyces hygroscopicus* (De Block et al., 1987; Jones et al., 1992; Thompson et al., 1987). Bar provides resistance to Phosphinothricin (PPT) (De Block et al., 1987). For the selection of the T1 transgenic lines, seeds were sown densely on soil and sprayed with Basta (a Bayer tradename for PPT) 2–4 times at weekly intervals, with the first spray applied approximately 10 days after germination. DNA from surviving plants was isolated and PCR tested for the presence of the transgene. In the following generations (T2, T3), selection was performed on plates with MS media containing PPT.

7.2.13 Copy number analysis

The number of transgenic insertions in new lines was determined by IDna Genetics. Isolated DNA was used in a qPCR-based assay based on a previously described method (Bartlett et al., 2008) adapted for Arabidopsis.

7.2.14 Yeast transformation

Plasmids (pGBKT7 and pGADT7-rec) were co-transformed into the AH109 yeast strain following the manufacturer's protocol. AH109 yeast was grown overnight in YPDA media at 28°C at 220 rpm. Yeast cells were pelleted with centrifugation for 10 min at 4000 rpm. The pellet was washed with 10 ml of LiTE and pelleted with centrifugation for 10 min at 4000 RPM. The pellet was resuspended in 1.5 ml LiTE and pelleted by centrifugation for 5 min at 6000 rpm. The pellet was resuspended according to the number of performed transformations (50 µl LiTE per transformation). 1 µl Plasmid at 100 ng/µl and 8 µl carrier DNA were added to 50 µl yeast cells and vortexed to mix. Carrier DNA was prepared by denaturation of salmon sperm DNA at 99°C for 20 min and cooled on ice immediately before use. 300 µl PEG/LiTE was added to the yeast cells and vortexed to mix. Finally, 10 µl DMSO was added, and the mixture was gently mixed. The yeast was incubated at 30°C for 30 min. The cells were transformed with heat shock at 42°C for 15 min. The cells were chilled on ice for 2 min and pelleted with centrifugation at 6000 rpm for 5 min. All the supernatant was removed, and the pellet was resuspended in 100 µl sterile water and spread on selective media (double dropout media) and grown at 30°C until colonies appeared.

LiTE (10 ml): 1 ml 10x LiAC, 1 ml 10x TE, and 8 ml sterile water

PEG/LiTE (10 ml): 1 ml 10x LiAC, 1 ml 10x TE, and 8 ml 50% PEG3350

7.2.15 Yeast-2-hybrid

To test for interaction between the bait and prey, several single yeast colonies from the initial transformation were picked and dissolved in 100 µl sterile water. They were plated on double dropout (SD-LW), triple dropout (SD-LWH), or quadruple dropout (SD-LWHA) media and left to grow at 30°C for 5–7 days. Growth on triple or quadruple media indicates interaction.

7.2.16 Isolation of plasmid from yeast

A version of the protocol for QIAprep Miniprep was used for isolation of the plasmids. 5 ml of yeast cells were pelleted at 8000 rpm for 3 min and resuspended in 250 µl P1 buffer containing 0.1 mg/ml RNase A. 100 µl acid-washed glass beads were added and grinded in a GenoGrinder at max speed for 2 min. The homogenate was mixed with 250 µl lysis buffer (P2), incubated for 3 min and neutralized with 350 µl neutralisation buffer (N3). The cell debris were pelleted with 10 min centrifugation at maximum speed and the plasmid was recovered by standard miniprep procedure. The plasmid was eluted by adding 30 µl EB buffer (10 mM Tris-HCl pH 8.5) to the column and incubating for at least 10 min (ideally overnight at 4°C). It was used for subsequent transformation into *E. coli*.

7.2.17 Transformation of bacteria

7.2.17.1 Heat-shock transformation of *E. coli*

Competent cells (30–40 µl) were thawed on ice for 10 min, mixed with 1–5 µl DNA and incubated on ice for 30 min. Cells were heat-shocked for 1 min at 42°C and returned to ice for 3 min. 1 ml LB-broth was added, and the cells were incubated at 37°C for 1 hour at 220 rpm. Cells were collected by 15 sec centrifugation and plated on LB plates containing the appropriated antibiotic. They were incubated overnight to single colonies were present.

7.2.17.2 Electroporation of *Agrobacterium*

Competent cells were thawed on ice, mixed with 1 µl plasmid and transferred to a pre-chilled Bio-Rad cuvette. Bacteria were transformed using 2.5 kV, 400 resistance and 25 uF capacitance. 1 ml LB broth was added to the cuvette and the mixture was transferred to an Eppendorf tube and incubated for 3h at 28°C with 220 rpm. Cells were collected by 15 sec centrifugation, plated on LB plates containing the appropriated antibiotic and incubated at 28°C until single colonies were visible, normally 2–3 days. Transformation of agrobacterium was normally highly efficient,

therefore transformed cells were plated on two plates using 1/10 and 9/10 of transformed cells.

7.2.18 Confocal microscopy

Plants were grown vertically on GM plates with 1% agar and 1% glucose. Root tips were mounted in 0.5 x GM media. Analyse of protein localization was performed on a Zeiss LSM780 confocal microscope using the 40x/1.2 water objective. GFP was excited at 488 nm and detected at 491-535 nm. For more detailed imaging, a 63x/1.4 oil objective was used.

7.2.19 Protein expression in *E. coli*

The plasmid carrying the construct of interest was subcloned into BL21 (DE3) cells for protein expression. A single colony was picked and used to inoculate a O/N culture. The O/N culture was used to set up a starter culture, and different volumes were used depending on the aim of the experiment. For gel filtration and crystallography, 100 ml O/N culture was used and 35 ml was diluted in 1 l fresh LB containing the appropriated antibiotic. The starter culture was grown to OD₆₀₀ 0.6-0 at 37°C and moved to a lower temperature of 18–20°C until OD₆₀₀ reached 0.8. Protein expression was induced by 0.4 mM IPTG (isopropyl β-D-1-thiogalactopyranoside). Protein expression was performed for 6h–18h. Cells were harvested by centrifugation at 4500 g for 15 min at 4°C. Cell pellets were flash frozen in liquid nitrogen and stored at -80°C until use.

7.2.20 Protein purification

7.2.20.1 Purification of His-tagged proteins

The cell pellet was dissolved in lysis buffer (25 mM Tris-HCl pH 8, 200 mM NaCl (or KCl), 20 mM Imidazole, 10 ug/ml DNase and EDTA-free protease inhibitor cocktail). The cells were lysed by sonication for 5 min (5 sec on/5 sec off). The cell debris was removed by ultracentrifugation (140.000 g 30 min at 4°C). The soluble fraction was purified by the batch purification method using Ni-NTA resins. Beads were washed several

times with lysis buffer and eluted in lysis buffer containing 500 mM Imidazole.

For crystal screens, the His-tag was removed by O/N cleavage with TEV protease. The purified protein was further purified by gel filtration and by running over an ion exchange column. Finally, the protein was concentrated before being used in crystal trials.

7.2.20.2 Purification of GST-tagged proteins

The cell pellet was dissolved in 5ml/1g lysis buffer (25 mM Tris-HCl pH 7.5, 500 mM NaCl 1% Triton X-100, and EDTA-free protease inhibitor cocktail). The cells were lysed by sonication and debris were removed by centrifugation. GST-tag proteins were purified by column purification, and 1–2 ml of washed glutathione-Sepharose fast-flow resins (GE Healthcare) was added to a disposable chromatography column (Bio-Rad). The cleared soluble fraction was filtered through a 0.45 µm filter and incubated with the resin for 15 min. The resin was washed 5 times with 1 ml of lysis buffer and 10 times with 1 ml low salt wash buffer (25 mM Tris-HCl pH 8.0, 100 mM NaCl). Proteins were eluted by on-column cleavage with Thrombin.

7.2.21 Gel filtration

Gel filtration of purified proteins was performed as described in M. Fielder et al. submitted) using a Superdex 200 10/300 GL column (GE Healthcare).

7.2.22 Electrophoretic mobility shift assay (EMSA)

DNA probes for EMSA analysis were PCR amplified and gel purified, or by annealing of two single-stranded oligos by heating to 95°C and slowly cooling down to 25°C. Oligos were labelled with Cy5 for visualization. RNA probes were made by *in-vitro* transcription as described. For each EMSA experiment, a set of reactions was performed mixing 1 µL 10x loading dye (0.4% w/v orange G, 50% v/v glycerol, 1 mM EDTA), 1 µL 0.5-1 µM probe, 1 µL 10x binding buffer (250 mM MES pH 5.6, 500 mM

NaCl, 100 mM MgCl₂ 10 mM DTT) and 2 μL H₂O with 5 μL serially diluted protein. The reactions were incubated for 1h at 4°C. prior to analysis by 1.6% TBE-Agarose. For non-labelled probes, gels were stained in 0.5x TBE with 1/10.000 SYBR Green II RNA stain for 10 min, followed by 10 min wash in water. Gels were scanned with a Typhoon scanner.

7.2.23 RNA synthesis

The RNA used for the EMSA probes were made as previously described (Hawkes et al., 2016) with minor modifications. The template dsDNA was amplified by PCR with primers that carry a T7 promoter sequence and then gel purified. These templates were used for *in vitro* transcription using the HiScribe™ T7 High Yield RNA Synthesis Kit, following the manufacturer's recommendations. The RNA products were purified by LiCl precipitation. RNA was folded by being initially denatured (2 min at 94°C), snap-cooled on ice and then folded at 22°C for 30 min in 1XHKM buffer (50 mM HEPES-KOH pH 8.0, 100 mM KCl, 6 mM MgCl₂). Unfolded RNA probes were denatured at 94°C. for 2 min and flash frozen in liquid nitrogen.

7.2.24 Crosslinked nuclear immunoprecipitation mass spectrometry

4 g of seedlings were harvested in a 50 ml tube with 37 ml PBS 1% formaldehyde. The seedlings were infiltrated under vacuum for 15 minutes. Formaldehyde was quenched with 2.5 ml 2M Glycine under vacuum for 5 min. The seedlings were washed and frozen in liquid nitrogen. Seedlings were grinded into fine powder and resuspended in 16 ml Lysis buffer (20 mM Tris-HCl pH 7.5, 20 mM KCl, 2 mM EDTA pH 8.0, 2.5 mM MgCl₂, 25% glycerol, 250 mM sucrose, 5 mM DTT and Protease inhibitor), and the mixture was rotated for 10 min at 4°C. The solution was filtered through two layers of Miracloth. The filtrate was spun down for 10 min at 1500g and resuspended in 20 ml NRBT buffer (20 mM Tris pH7.5, 2.5 mM MgCl₂, 25% glycerol, and 0.2% Triton x-100) and centrifugated at 1500 g for 8 min. This wash step was repeated once. The pellet was resuspended in 1 ml NRBT buffer, transferred to a 1 ml tube

and centrifuged for 5 min to remove the buffer completely. The pellet was resuspended in 600 μ l RIPA buffer (PBS 1% NP-40 0.5% Sodiumdeoxycholate 0.1% SDS) and sonicated 3 times for 5 min with 30 sec on/30 sec off cycles. Debris were removed by centrifugation at 13000 rpm for 15 min.

The supernatant was mixed with 15 μ l GFP-trap beads and incubated for 3h with slow rotation. The beads were washed twice for 5 min in Low salt (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, and 20mM Tris-HCl pH 8), High salt (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8) and TE buffer. The complex was eluted by boiling in 1x SDS loading buffer for 15 min.

7.2.25 Preparation of gel slices for Trypsin digestion.

The samples were loaded on 7.5% SDS gel and run less than 0.5 cm into the gel. All parts of the cassette were rinsed under tap water and dried with tissue paper. The gel was stained with InstantBlue, and the lanes were cut with a razorblade and then further cut into smaller pieces. For washing, 1 ml buffer was used for 20 min with strong vortexing, unless mentioned otherwise.

The gel slices were destained with 30% ethanol for 30 min at 65°C. This step was repeated once. The slices were then washed with TEAB/50% ACN, followed by incubation in 10 mM DTT for 30 min at 55°C. The solution was removed, and the slices were incubated in 30 mM IAA (Iodoacetamide) in 50 mM TEAB solution for 30 min at RT in the dark. After incubation, the gel slices were washed with TEAB/50%ACN and then in TEAB. After these wash steps, the gel slices were cut into very small pieces, transferred to Low bind tubes and washed with TEAB/50%ACN for 20 min, followed by wash with 100% acetonitrile. The slices were washed one more time in 100% acetonitrile. The acetonitrile was removed, and the samples were dried in a speed vac for

30 min. The samples were then ready for trypsin digestion and mass spectrometry.

7.2.26 Chromatin Conformation Capture (3C)

Chromatin conformation capture was done as previously described (Crevillén et al., 2013) with some modifications 1g of plant material was harvested and crosslinked in 2% formaldehyde in 1x PBS for a total of 20 min (2+8+10 min). Formaldehyde was quenched with 0.125 mM Glycine for 7 min. The tissue was grinded to fine powder and dissolved in 30 ml Honda buffer. The tissue was dissolved in Honda buffer by slow rotation in the cold room for 15 min before filtering through two layers of miracloth. The nuclei were pelleted with centrifugation at 2500 g for 15 min at 4°C. The pellet was resuspended in 1.8 ml Honda buffer, split in two LoBind Eppendorf tubes and centrifuged for 5 min at 1900 g at 4°C. The supernatant was removed and the remaining Honda buffer was removed by centrifugation for 1 min at 12000 rpm at 4°C. The nuclei were resuspended in 1 ml appropriated 1.2x restriction buffer (3.1 for BamHI/BglIII or DpnII buffer for DpnII) and centrifuged for 5 min at 1900 g at 4°C. Finally, the pellet was dissolved in 500 µl 1.2 x restriction buffer. To inactivate endogenous restriction enzymes, 10 µl (15 µl for DpnII) 10% SDS was added, and the samples were incubated for 30 min at 65°C at 900 rpm. 10% Triton X-100 was added to sequester the SDS and it was incubated for 30 min at 37°C at 900 rpm. Chromatin was digested overnight, for 14–16 h, at 900 rpm, with 600 U of BamHI, BglIII, and TaqIα, 400 U of DpnII, or 200 U of NlaIII. Except for TaqIα, digestion was performed at 37°C; the TaqIα digestion was performed at 60°C. Digestion was stopped with 40 µl 10% SDS and incubated for 10 min at 65°C. The SDS was sequestered with 100 µl 10% Triton x-100 and 300 µl water at 37°C for 10 min. Ligation was performed in 5 ml, the samples were diluted to 5 ml with water, with 50 U T4 ligase (Promega) at 17°C for 7h. Reverse crosslinking was performed overnight at 65°C, followed by proteinase K treatment for 20 min at 45°C. DNA was purified with Phenol:Chloroform:IAA (25:24:1) and precipitated with 1x Isopropanol

1/10x 3M Sodiumacetate and 5 µl Glycogen. Precipitation was performed in 2 ml LoBind tubes at -80°C overnight. The DNA was dissolved in 50 µl water and further purified using the ChIP DNA Clean & Concentration (Zymo research) kit, following the manufacturer's protocol, and then finally resuspended in 500 µl water. The 3C library was then ready for qPCR analysis.

Honda buffer: 0.44 M Sucrose, 1.25% Ficoll, 2.5% Dextran T40, 20 mM Hepes-KOH pH 7.4, 0.5% Triton X-100, 10 mM MgCl₂, 5 mM DTT and EDTA-free complete protease inhibitor cocktail. Aliquots prepared and stored at -20, Triton X-100, DTT and protease inhibitor cocktail were freshly added.

10 x T4 ligase buffer (Promega): 300 mM Tris-HCl pH 7.8, 100 mM MgCl₂, 100 mM DTT and 10 mM ATP

10 x T4 ligase buffer (New England Biolabs): 500 mM Tris-HCl pH 7.5, 100 mM MgCl₂, 100 mM DTT and 10 mM ATP

7.2.26.1 3C quantification

Calculation of relative interaction frequency was performed as previously described (Hagège et al., 2007). In summary, firstly the 3C product for each fragment was normalized to the DNA concentration in the sample to correct for different DNA amounts between samples. The DNA concentration was measured by a loading control amplicon that does not span the restriction sites.

$$LC = 2^{-(\text{Fragment } Cp - LC \text{ } cp)}$$

Secondly, the value was corrected for different primer efficiencies. Primer efficiency was measured by digestion and ligation of pBS-FLC15.

$$CT = \frac{\text{Experimental LC}}{\text{Control template LC}}$$

For convenience, data are plotted as relative interaction frequency, proportional to the fragment of highest interaction frequency.

7.2.27 Protein chromatin immunoprecipitation

Protein ChIP of VIN3-GFP was performed as previously described (Yang et al., 2017). 3g of Arabidopsis seedlings were crosslinked in 1% formaldehyde PBS for 15 min. Formaldehyde was quenched with 0.125 M Glycine for 5 min and the seedlings were washed, dried and frozen in liquid nitrogen. Nucleic was extracted as for 3C with slight modifications. 35 ml Honda buffer was used to resuspend the grinded plant material, and after filtering the solution was centrifuged at 3500 g for 15 min. The pellet was resuspended in 1 ml/g Honda buffer, divided between two 2 ml DNA LoBind tubes and centrifugated at 2500 g for 5 min. This wash step was repeated until the pellet appeared clear. Honda buffer was removed by centrifugation at 12.800 rpm for 3 min and the pellet was resuspended in 600 ul RIPA buffer. Chromatin was sheered by sonication, at 30 sec on/ 30 sec off for 5 min, 5 times at high intensity. The tubes were mixed between each 5 min cycle. Released Chromatin was purified by centrifugation at 12.800 rpm for 10 min, and the supernatant was collected and sonicated twice more with the same settings as the first time. Debris were removed by centrifugation at 12.800 rpm for 10 min.

Immunoprecipitation was performed by combining four tubes (6g starting materials) of chromatin, mixing with 3 µg Anti-GFP antibody (ab-290) and being incubated for 2.5h at 4°C. with slow rotation. After 2.5h, 50 µl (100 ul slurry) Protein A Agarose/Salmon sperm (16-157) was added and the immunoprecipitation was continued for a further 1.5h. The beads were washed twice with Low salt, High salt and TE wash buffer, each buffer for 5 min. The ChIP complex was eluted with 200 µl (1% SDS, 0.1 M NaHCO₃) at 65°C 800 rpm for 15 min twice for a total of 400 µl elution

product. Reverse crosslinking was performed O/N with 16 μ l 5M NaCl at 65°C 600 rpm. For Proteinase K treatment, 16 μ l 1M Tris-HCl pH 6.5, 8 μ l 0.5M EDTA pH 8 and 3 μ l 20 mg/ml proteinase K were added and incubated at 45°C for 1 h. The DNA was purified with phenol:chloroform purification followed by ethanol precipitation.

RIPA buffer: 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS in 1x PBS supplemented with 1x cOmplete EDTA-free protease inhibitor cocktail (1 tablet in 1 ml gives 50x stock solution).

7.2.27.1 ChIP-Seq

Purification of VIN3-GFP associated nucleic acid was performed as above with a few modifications for the purpose of sequencing. VIN3-GFP-antibody complexes were pulled down with fast-flow Protein A-agarose (instead of salmon-sperm-coated Protein A agarose). After purification of nucleic acid, the RNA was removed by RNase treatment. Finally, the nucleic acid was concentrated using the ChIP DNA Clean & Concentration (Zymo research) kit, following the manufacturer's protocol. Concentration was measured by Qubit fluorometric quantification. Immunoprecipitated DNA and input DNA were shipped to BGI Genomics (Hong Kong) for library preparation and sequencing.

7.2.28 Transient expression in *N. benthamiana*

For protein expression in *N. benthamiana*, a single colony of *Agrobacterium* harbouring the constructs of interest was inoculated in LB medium with the relevant antibiotics and grown for 36–48h at 28°C. *Agrobacterium* was pelleted at 3000g for 10 min, resuspended in infiltration buffer and adjusted to an OD₆₀₀ of 0.7. Constructs were mixed in equal OD ratios and incubated at RT for 2–3h. Infiltration was performed using 3–4-week-old plants. Entire leaves were infiltrated, harvested 4–5 days after infiltration and frozen in LN₂. Expression of GFP was confirmed by fluorescence.

7.2.29 Light microscopy

Expression and localization of transient expressed construct in *N. benthamiana* were performed using the Zeiss Lumar V12 stereomicroscope with the GFP filter (excitation 470/40 emission 525/50) Images were processed using ImageJ.

7.2.30 Co-IP

Leaves were ground to fine powder in LN₂, and protein extracted in plant optimized Über buffer (10% glycerol, 100 mM Tris-HCl pH 7.5, 200 mM NaCl, 1% Triton X-100, 10 mM DTT, 2% PVPP, 2 mM Na₃VO₄, 5 mM NaF, and Protease inhibitor cocktail). Plant debris were removed by repeated centrifugation at 3000 g for 10 min and 500 µl was used for immunoprecipitation with 10 µl homemade GBP-beads for 2h. Beads were washed three times with IP buffer and the protein complex was eluted in 50 µl 2X LDS buffer.

7.2.30.1 Immunoblot

Proteins were separated by SDS-PAGE and blotted onto PVDF membrane. Equal loading and transfer were checked by Ponceau staining. Western blotting was performed with appropriate antibodies. Primary antibodies were diluted 1:5000 and secondary 1:50.000 in PBS 0.05% Tween20 and 5% milk powder. Blots were washed with PBS 0.05% Tween20 and developed with ECL Western blotting reagent on film.

7.2.31 Preparation of competent cells

7.2.31.1 Chemically competent *E. coli* cells

Dh5α and Mach1 *E. coli* strains were used for cloning. 50 mL of SOB were inoculated with the desired strain and grown overnight. 1 L of LB was supplemented with 10 ml 1 M MgCl₂, and inoculated with 20 ml preculture (1/50) and grown to OD₆₀₀ 0.6. During the following procedures everything was pre-chilled and kept cold. Cells were chilled on ice for 10 min and collected by centrifugation at 2500g for 10 min at 4°C.

Cells were resuspended in 130 ml CC buffer (10 mM PIPES pH 6.7, 15 mM CaCl₂, 250 mM KCl 55 mM MnCl₂) and incubated 10 min on ice before being collected by centrifugation, resuspended in 50 ml CC buffer and incubated for 10 min on ice. 3.5 ml of DMSO was added and cells were incubated for an additional 10 min on ice before being aliquoted into pre-chilled tubes and flash frozen in liquid nitrogen. Competent cells were stored at -80°C.

7.2.31.2 Electrocompetent *Agrobacterium* cells

A single colony was used to inoculate 10 ml of LB-broth and this was grown for 2 days. 4 ml of starter culture was used to inoculate 400 ml LB-broth and it was grown to OD₆₀₀ 0.5-1. During the following procedures, everything was pre-chilled and kept cold. Cells were cooled on ice and collected by centrifugation (8000 rpm for 9 min at 4°C.) The pellet was resuspended in 400 ml sterile water before being centrifugated again for 9 min at 5000 rpm. Cells were resuspended in 200 ml sterile water and centrifugated at 6000 rpm for 9 min. All supernatant was removed, the pellet resuspended in 10% glycerol and centrifugated at 7000 rpm for 10 min. The pellet was finally resuspended in 2 ml 10% glycerol. 50 ul aliquots were made, flash frozen and stored at -80°C.

8 References

Abed, J.A., Ghotbi, E., Ye, P., Frolov, A., Benes, J., and Jones, R.S. (2018). De novo recruitment of polycomb-group proteins in drosophila embryos. *Dev.* *145*, 1–12.

Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* *29*, 585–590.

Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* *176*, 419–434.

Alecki, C., Chiwara, V., Sanz, L.A., Grau, D., Arias Pérez, O., Boulier, E.L., Armache, K.J., Chédin, F., and Francis, N.J. (2020). RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. *Nat. Commun.* *11*, 1–14.

Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P. V., and Kuroda, M.I. (2014). Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc. Natl. Acad. Sci.* *111*, 2488 LP – 2493.

Alfieri, C., Gambetta, M.C., Matos, R., Glatt, S., Sehr, P., Fraterman, S., Wilm, M., Müller, J., and Müller, C.W. (2013). Structural basis for targeting the chromatin repressor Sfmtb to Polycomb response elements. *Genes Dev.* *27*, 2367–2379.

Allepuz-Fuster, P., O'Brien, M.J., González-Polo, N., Pereira, B., Dhoondia, Z., Ansari, A., and Calvo, O. (2019). RNA polymerase II plays an active role in the formation of gene loops through the Rpb4 subunit. *Nucleic Acids Res.* *47*, 8975–8987.

Almeida, M., Bowness, J.S., and Brockdorff, N. (2020). The many faces of Polycomb regulation by RNA. *Curr. Opin. Genet. Dev.* *61*, 53–61.

Angel, A., Song, J., Dean, C., and Howard, M. (2011). A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* *476*, 105–108.

Ansari, A., and Hampsey, M. (2005). A role for the CPF 3' -end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* *19*, 2969–2978.

Aranda, S., Mas, G., and Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* *1*, 1–16.

Ard, R., Allshire, R.C., and Marquardt, S. (2017). Emerging properties and functional consequences of noncoding transcription. *Genetics* *207*, 357–367.

Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., and Crespi, M. (2014). Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* *55*, 383–396.

Asenjo, H.G., Gallardo, A., López-Onieva, L., Tejada, I., Martorell-Marugán, J., Carmona-Sáez, P., and Landeira, D. (2020). Polycomb regulation is coupled to cell cycle transition in pluripotent stem cells. *Sci. Adv.* *6*, eaay4768.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* *12*, 429–442.

Baile, F., Merini, W., Hidalgo, I., and Calonje, M. (2021). EAR domain-containing transcription factors trigger PRC2-mediated chromatin marking in *Arabidopsis*. *Plant Cell*.

Balas, M.M., Hartwick, E.W., Barrington, C., Roberts, J.T., Wu, S.K., Bettcher, R., Griffin, A.M., Kieft, J.S., and Johnson, A.M. (2021). RNA matchmaking remodels lncRNA structure and promotes PRC2 activity. *BioRxiv* 2020.04.13.040071.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., et al. (2012). Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nat. Struct. Mol. Biol.* *19*, 1257–1265.

Bartlett, J.G., Alves, S.C., Smedley, M., Snape, J.W., and Harwood, W.A. (2008). High-throughput *Agrobacterium*-mediated barley transformation. *Plant Methods* *4*, 1–12.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature* *427*, 164–167.

Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L. (2011). Transcriptional regulation of *Arabidopsis* LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* *23*, 4065–4078.

Berry, S. (2015). Chromatin-based memory of prolonged cold exposure in *Arabidopsis thaliana*. University of East Anglia.

Berry, S., and Dean, C. (2015). Environmental perception and epigenetic memory: mechanistic insight through FLC. *Plant J.* *83*, 133–148.

Berry, S., Hartley, M., Olsson, T.S.G., Dean, C., and Howard, M. (2015). Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance. *Elife* 4, 1–11.

Berry, S., Dean, C., and Howard, M. (2017a). Slow Chromatin Dynamics Allow Polycomb Target Genes to Filter Fluctuations in Transcription Factor Activity. *Cell Syst.* 4, 445-457.e8.

Berry, S., Rosa, S., Howard, M., Bühler, M., and Dean, C. (2017b). Disruption of an RNA-binding hinge region abolishes LHP1-mediated epigenetic repression. *Genes Dev.* 31, 2115–2120.

Beuchle, D., Struhl, G., and Muller, J. (2001). Polycomb group proteins and heritable silencing of *Drosophila* Hox genes. *Development* 128, 993–1004.

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* 31, 35–40.

Bienz, M. (2020). Head-to-Tail Polymerization in the Assembly of Biomolecular Condensates. *Cell* 182, 799–811.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., and Koseki, Y. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 157, 1445–1459.

De Block, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gosselé, V., Movva, N.R., Thompson, C., Van Montagu, M., and Leemans, J. (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *EMBO J.* 6, 2513–2518.

Bloomer, R.H., Hutchison, C.E., Bäurle, I., Walker, J., Fang, X., Perera, P., Velanis, C.N., Gümüs, S., Spanos, C., Rappsilber, J., et al. (2020). The Arabidopsis epigenetic regulator ICU11 as an accessory protein of polycomb repressive complex 2. *Proc. Natl. Acad. Sci. U. S. A.* *117*, 16660–16666.

Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* *28*, 420–435.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. *Science* *330*, 612–616.

Bond, D.M., Wilson, I.W., Dennis, E.S., Pogson, B.J., and Jean Finnegan, E. (2009). VERNALIZATION INSENSITIVE 3 (VIN3) is required for the response of Arabidopsis thaliana seedlings exposed to low oxygen conditions. *Plant J.* *59*, 576–587.

Bordiya, Y., Kim, J., Xi, Y., Kim, D.-H., Pyo, Y., Zhao, B., Zong, W., Ricci, W.A., Zhang, X., and Sung, S. (2020). VIL1, a Polycomb-associated protein, modulates high ambient temperature response via H3K27me3 and H2A.Z in Arabidopsis thaliana. *BioRxiv*.

Boulay, G., Rosnoblet, C., Guérardel, C., Angrand, P.-O., and Leprince, D. (2011). Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. *Biochem. J.* *434*, 333–342.

Box, M.S., Coustham, V., Dean, C., and Mylne, J.S. (2011). Protocol: A simple phenol-based method for 96-well extraction of high quality RNA from Arabidopsis. *Plant Methods* *7*, 7.

Brien, G.L., Gambero, G., O'Connell, D.J., Jerman, E., Turner, S.A., Egan, C.M., Dunne, E.J., Jurgens, M.C., Wynne, K., Piao, L., et al. (2012). Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat. Struct. Mol. Biol.* *19*, 1273–1281.

Brien, G.L., Healy, E., Jerman, E., Conway, E., Fadda, E., O'Donovan, D., Krivtsov, A. V., Rice, A.M., Kearney, C.J., Flaus, A., et al. (2015). A chromatin-independent role of polycomb-like 1 to stabilize p53 and promote cellular quiescence. *Genes Dev.* *29*, 2231–2243.

Brockdorff, N., Bowness, J.S., and Wei, G. (2020). Progress toward understanding chromosome silencing by Xist RNA. *Genes Dev.* *34*, 733–744.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K.J., Carroll, T., Xie, S.Q., Stock, J.K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. *Cell Stem Cell* *10*, 157–170.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J.A. (1998). The *Drosophila* Polycomb Group Gene pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. *Mol. Cell* *1*, 1057–1064.

Brown, J.L., Fritsch, C., Mueller, J., and Kassis, J.A. (2003). The *Drosophila* pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* *130*, 285–294.

Calonje, M. (2014). PRC1 Marks the Difference in Plant PcG Repression. *Mol. Plant* *7*, 459–471.

Castaings, L., Bergonzi, S., Albani, M.C., Kemi, U., Savolainen, O., and Coupland, G. (2014). Evolutionary conservation of cold-induced antisense RNAs of FLOWERING LOCUS C in *Arabidopsis thaliana* perennial relatives. *Nat. Commun.* *5*, 4457.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* *149*, 1393–1406.

Causier, B., Ashworth, M., Guo, W., and Davies, B. (2012). The TOPLESS Interactome: A Framework for Gene Repression in *Arabidopsis*. *Plant Physiol.* *158*, 423–438.

Chakrabortee, S., Kayatekin, C., Newby, G.A., Mendillo, M.L., Lancaster, A., and Lindquist, S. (2016). Luminidependens (LD) is an *Arabidopsis* protein with prion behavior. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 6065–6070.

Chanarat, S., and Sträßer, K. (2013). Splicing and beyond: The many faces of the Prp19 complex. *Biochim. Biophys. Acta - Mol. Cell Res.* *1833*, 2126–2134.

Chanarat, S., Seizl, M., and Sträßer, K. (2011). The prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev.* *25*, 1147–1158.

Chandler, J., Wilson, A., and Dean, C. (1996). *Arabidopsis* mutants showing an altered response to vernalization. *Plant J.* *10*, 637–644.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* *131*, 5263–5276.

Chen, S., Jiao, L., Shubbar, M., Yang, X., and Liu, X. (2018). Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol. Cell* 69, 840-852.e5.

Chen, S., Jiao, L., Liu, X., Yang, X., and Liu, X. (2020). A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin. *Mol. Cell* 77, 1265-1278.e7.

Choi, J., Bachmann, A.L., Tauscher, K., Benda, C., Fierz, B., and Müller, J. (2017). DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nat. Struct. Mol. Biol.* 24, 1039–1047.

Choi, K., Kim, J., Hwang, H.-J., Kim, S., Park, C., Kim, S.Y., and Lee, I. (2011). The FRIGIDA Complex Activates Transcription of FLC , a Strong Flowering Repressor in Arabidopsis , by Recruiting Chromatin Modification Factors. *Plant Cell* 23, 289–303.

Chopra, V.S., Hendrix, D.A., Core, L.J., Tsui, C., Lis, J.T., and Levine, M. (2011). The Polycomb Group Mutant *esc* Leads to Augmented Levels of Paused Pol II in the Drosophila Embryo. *Mol. Cell* 42, 837–844.

Chowdhary, S., Kainth, A.S., Pincus, D., and Gross, D.S. (2019). Heat Shock Factor 1 Drives Intergenic Association of Its Target Gene Loci upon Heat Shock. *Cell Rep.* 26, 18-28.e5.

Clarke, B.J. (2004). A Brief History of Pentecostalism. *Cold Spring Harb. Lab. Press* 6, 12–14.

Collins, J., O’Grady, K., Chen, S., and Gurley, W. (2019). The C-terminal WD40 repeats on the TOPLESS co-repressor function as a protein–protein interaction surface. *Plant Mol. Biol.* 100, 47–58.

Conway, E., Jerman, E., Healy, E., Ito, S., Holoch, D., Oliviero, G., Deevy, O., Glancy, E., Fitzpatrick, D.J., Mucha, M., et al. (2018). A Family of Vertebrate-Specific Polycombs Encoded by the LCOR/LCORL Genes Balance PRC2 Subtype Activities. *Mol. Cell* *70*, 408-421.e8.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., and Agger, K. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* *7*, 1–8.

Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J.* *32*, 140–148.

Crevillén, P., Yang, H., Cui, X., Greeff, C., Trick, M., Qiu, Q., Cao, X., and Dean, C. (2014). Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state. *Nature* *515*, 587–590.

Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* *20*, 1147.

Csorba, T., Questa, J.I., Sun, Q., and Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 16160–16165.

Cullen, K.E., Kladde, M.P., and Seyfred, M.A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. *Science* (80-.). *261*, 203–206.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. *J. Mol. Biol.* *319*, 1097–1113.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* *20*, 1250–1257.

Dekker, J. (2006). The three “C” s of chromosome conformation capture: Controls, controls, controls. *Nat. Methods* *3*, 17–21.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. *Science* (80-.). *295*, 1306–1311.

DeLuca, S.Z., Ghildiyal, M., Niu, W., Pang, L.Y., and Spradling, A.C. (2020). Differentiating *Drosophila* female germ cells initiate Polycomb silencing by altering PRC2 sampling. *Elife* 1–33.

Deng, W., and Blobel, G.A. (2010). Do chromatin loops provide epigenetic gene expression states? *Curr. Opin. Genet. Dev.* *20*, 548–554.

Deng, W., Buzas, D.M., Ying, H., Robertson, M., Taylor, J., Peacock, W.J., Dennis, E.S., and Helliwell, C. (2013). Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics* *14*, 593.

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., Bischof, S., Grussem, W., and Hennig, L. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* *32*, 2073–2085.

Derkacheva, M., Liu, S., Figueiredo, D.D., Gentry, M., Mozgova, I., Nanni, P., Tang, M., Mannervik, M., Köhler, C., and Hennig, L. (2016). H2A deubiquitinases UBP12/13 are part of the Arabidopsis polycomb group protein system. *Nat. Plants* *2*, 16126.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.

van Dop, M., Fiedler, M., Mutte, S., de Keijzer, J., Olijslager, L., Albrecht, C., Liao, C.-Y., Janson, M.E., Bienz, M., and Weijers, D. (2020). DIX Domain Polymerization Drives Assembly of Plant Cell Polarity Complexes. *Cell* 180, 427-439.e12.

Dueva, R., Akopyan, K., Pederiva, C., Trevisan, D., Dhanjal, S., Lindqvist, A., and Farnebo, M. (2019). Neutralization of the Positive Charges on Histone Tails by RNA Promotes an Open Chromatin Structure. *Cell Chem. Biol.* 26, 1436-1449.e5.

Durek, P., Schmidt, R., Heazlewood, J.L., Jones, A., MacLean, D., Nagel, A., Kersten, B., and Schulze, W.X. (2010). PhosPhAt: the Arabidopsis thaliana phosphorylation site database. An update. *Nucleic Acids Res.* 38, D828–D834.

Eagen, K.P., Aiden, E.L., and Kornberg, R.D. (2017). Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci. U. S. A.* 114, 8764–8769.

Fan, H., Guo, Y., Tsai, Y.-H., Storey, A.J., Kim, A., Gong, W., Edmondson, R.D., Mackintosh, S.G., Li, H., Byrum, S.D., et al. (2021). A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. *BioRxiv* 2021.03.11.435004.

Fang, X., Wang, L., Ishikawa, R., Li, Y., Fiedler, M., Liu, F., Calder, G., Rowan, B., Weigel, D., Li, P., et al. (2019). Arabidopsis FLL2 promotes liquid–liquid phase separation of polyadenylation complexes. *Nature* 569, 265–269.

Fang, X., Wu, Z., Raitskin, O., Webb, K., Voigt, P., Lu, T., Howard, M., and Dean, C. (2020). The 3' processing of antisense RNAs physically links to chromatin-based transcriptional control. *Proc. Natl. Acad. Sci. U. S. A.* *117*, 15316–15321.

Feng, G., Yuan, Y., Li, Z., Wang, L., Zhang, B., Luo, J., Ji, J., and Kong, D. (2019). Replication fork stalling elicits chromatin compaction for the stability of stalling replication forks. *Proc. Natl. Acad. Sci. U. S. A.* *116*, 14563–14572.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Mol. Cell* *55*, 694–707.

Fiedler, M., Mendoza-Topaz, C., Rutherford, T.J., Mieszczynek, J., and Bienz, M. (2011). Dishevelled interacts with the DIX domain polymerization interface of Axin to interfere with its function in down-regulating β -catenin. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 1937–1942.

Fiedler, M., Franco-Echevarria, E., Rutherford, T.J., Ahsen, B., Yeates, A., Nielsen, M., Schulten, A., Dean, C., and Bienz, M. (2021). Ancient polymerization fold for cold-induced Polycomb silencing in flowering control.

Finn, E.H., Pegoraro, G., Brandão, H.B., Valton, A.L., Oomen, M.E., Dekker, J., Mirny, L., and Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* *176*, 1502–1515.e10.

Finnegan, E.J. (2015). Time-dependent stabilization of the +1 nucleosome is an early step in the transition to stable cold-induced repression of FLC. *Plant J.* *84*, 875–885.

Finnegan, E.J., Sheldon, C.C., Jardinaud, F., Peacock, W.J., and Dennis, E.S. (2004). A Cluster of Arabidopsis Genes with a Coordinate Response to an Environmental Stimulus. *Curr. Biol.* *14*, 911–916.

Foglizzo, M., Middleton, A.J., Burgess, A.E., Crowther, J.M., Dobson, R.C.J., Murphy, J.M., Day, C.L., and Mace, P.D. (2018). A bidentate Polycomb Repressive-Deubiquitinase complex is required for efficient activity on nucleosomes. *Nat. Commun.* *9*.

Frey, F., Sheahan, T., Finkl, K., Stoehr, G., Mann, M., Benda, C., and Müller, J. (2016). Molecular basis of PRC1 targeting to polycomb response elements by PhoRC. *Genes Dev.* *30*, 1116–1127.

Gaetani, M., Matafora, V., Saare, M., Spiliotopoulos, D., Mollica, L., Quilici, G., Chignola, F., Mannella, V., Zucchelli, C., Peterson, P., et al. (2012). AIRE-PHD fingers are structural hubs to maintain the integrity of chromatin-associated interactome. *Nucleic Acids Res.* *40*, 11756–11768.

Gagliardi, D., and Manavella, P.A. (2020). Short-range regulatory chromatin loops in plants. *New Phytol.* *228*, 466–471.

Gagliardi, D., Cambiagno, D.A., Arce, A.L., Tomassi, A.H., Giacomelli, J.I., Ariel, F.D., and Manavella, P.A. (2019). Dynamic regulation of chromatin topology and transcription by inverted repeat-derived small RNAs in sunflower. *Proc. Natl. Acad. Sci. U. S. A.* *116*, 17578–17583.

Gammons, M. V., Renko, M., Johnson, C.M., Rutherford, T.J., and Bienz, M. (2016). Wnt Signalingosome Assembly by DEP Domain Swapping of Dishevelled. *Mol. Cell* *64*, 92–104.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* *45*, 344–356.

García Reyes, L.E. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. *J. Chem. Inf. Model.* *53*, 1689–1699.

Gatchalian, J., Kingsley, M.C., Moslet, S.D., Rosas Ospina, R.D., and Kutateladze, T.G. (2015). An aromatic cage is required but not sufficient for binding of Tudor domains of the Polycomb-like protein family to H3K36me3. *Epigenetics* *10*, 467–473.

Gee, E.P.S., Yüksel, D., Stultz, C.M., and Ingber, D.E. (2013). SLLISWD sequence in the 10FNIII domain initiates fibronectin fibrillogenesis. *J. Biol. Chem.* *288*, 21329–21340.

Geisler, S.J., and Paro, R. (2015). Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. *Development* *142*, 2876–2887.

Gendall, A.R., Levy, Y.Y., Wilson, A., and Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell* *107*, 525–535.

Geng, Z., and Gao, Z. (2020). Mammalian PRC1 Complexes: Compositional Complexity and Diverse Molecular Mechanisms. *Int. J. Mol. Sci.* *21*, 8594.

Geraldo, N., Bäurle, I., Kidou, S.-I., Hu, X., and Dean, C. (2009). FRIGIDA delays flowering in Arabidopsis via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex. *Plant Physiol.* *150*, 1611–1618.

Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* *25*, 2210–2221.

Greb, T., Mylne, J.S., Crevillen, P., Geraldo, N., An, H., Gendall, A.R., and Dean, C. (2007). The PHD Finger Protein VRN5 Functions in the Epigenetic Silencing of Arabidopsis FLC. *Curr. Biol.* *17*, 73–78.

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. *Development* *143*, 2716–2723.

Grossniklaus, U., and Paro, R. Transcriptional Silencing by Polycomb Group Proteins Cellular Memory. 1–26.

Guo, L., Cao, X., Liu, Y., Li, J., Li, Y., Li, D., Zhang, K., Gao, C., Dong, A., and Liu, X. (2018). A chromatin loop represses WUSCHEL expression in Arabidopsis. *Plant J.* *94*, 1083–1097.

Guo, M., Thomas, J., Collins, G., and Timmermans, M.C.P. (2008). Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of Arabidopsis. *Plant Cell* *20*, 48–58.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* 2, 1722–1733.

Harvey, Z.H., Chen, Y., and Jarosz, D.F. (2018). Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Mol. Cell* 69, 195–202.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. *Cell Rep.* 17, 583–595.

Hawkes, E.J., Hennelly, S.P., Novikova, I. V., Irwin, J.A., Dean, C., and Sanbonmatsu, K.Y. (2016). COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. *Cell Rep.* 16, 3087–3096.

He, Y., and Amasino, R.M. (2005). Role of chromatin modification in flowering-time control. *Trends Plant Sci.* 10, 30–35.

Healy, E., Mucha, M., Glancy, E., Fitzpatrick, D.J., Conway, E., Neikes, H.K., Monger, C., Van Mierlo, G., Baltissen, M.P., Koseki, Y., et al. (2019). PRC2.1 and PRC2.2 Synergize to Coordinate H3K27 Trimethylation. *Mol. Cell* 76, 437-452.e6.

Hecker, A., Brand, L.H., Peter, S., Simoncello, N., Kilian, J., Harter, K., Gaudin, V., and Wanke, D. (2015). The Arabidopsis GAGA-Binding Factor BASIC PENTACYSTEINE6 Recruits the POLYCOMB-REPRESSIVE COMPLEX1 Component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA Motifs. *Plant Physiol.* 168, 1013–1024.

Helliwell, C.A., Robertson, M., Finnegan, E.J., Buzas, D.M., and Dennis, E.S. (2011). Vernalization-repression of Arabidopsis FLC requires promoter sequences but not antisense transcripts. *PLoS One* 6.

Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* 25, 414–423.

Heo, J.B., and Sung, S. (2011). Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. *Science* (80-.). 331, 76–79.

Hepworth, J., Antoniou-Kourouniotti, R.L., Bloomer, R.H., Selga, C., Berggren, K., Cox, D., Collier Harris, B.R., Irwin, J.A., Holm, S., Säll, T., et al. (2018). Absence of warmth permits epigenetic memory of winter in Arabidopsis. *Nat. Commun.* 9, 3–10.

Hepworth, J., Antoniou-Kourouniotti, R.L., Berggren, K., Selga, C., Tudor, E.H., Yates, B., Cox, D., Harris, B.R.C., Irwin, J.A., Howard, M., et al. (2020). Natural variation in autumn expression is the major adaptive determinant distinguishing arabidopsis flc haplotypes. *Elife* 9, 1–30.

Højfeldt, J.W., Laugesen, A., Willumsen, B.M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O.N., and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat. Struct. Mol. Biol.* 25, 225–232.

Højfeldt, J.W., Hedehus, L., Laugesen, A., Tatar, T., Wiehle, L., and Helin, K. (2019). Non-core Subunits of the PRC2 Complex Are Collectively Required for Its Target-Site Specificity. *Mol. Cell* 76, 423-436.e3.

Hu, L., Li, Z., Wang, P., Lin, Y., and Xu, Y. (2011). Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res.* *21*, 1374–1378.

Hu, X., Wang, H., Ke, H., and Kuhlman, B. (2007). High-resolution design of a protein loop. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 17668–17673.

Huang, Y., Sicar, S., Ramirez-Prado, J.S., Manza-Mianza, D., Antunez-Sanchez, J., Brik-Chaouche, R., Rodriguez-Granados, N.Y., An, J., Bergounioux, C., Mahfouz, M.M., et al. (2021). Polycomb-dependent differential chromatin compartmentalization determines gene coregulation in Arabidopsis. *Genome Res.*

Huo, Y., Yan, Z., Zhang, B., and Wang, X. (2016). Recruitment of Polycomb Repressive Complex 2 is Essential to Suppress the Target Chromatin in Arabidopsis. *CRC. Crit. Rev. Plant Sci.* *35*, 131–145.

Hyun, Y., Yun, H., Park, K., Ohr, H., Lee, O., Kim, D.-H., Sung, S., and Choi, Y. (2013). The catalytic subunit of Arabidopsis DNA polymerase α ensures stable maintenance of histone modification. *Development* *140*, 156–166.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* *26*, 565–577.

Jain, K., Fraser, C.S., Marunde, M.R., Parker, M.M., Sagum, C., Burg, J.M., Hall, N., Popova, I.K., Rodriguez, K.L., Vaidya, A., et al. (2020). Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. *Epigenetics and Chromatin* *13*, 1–11.

Jamge, S., Stam, M., Angenent, G.C., and Immink, R.G.H. (2017). A cautionary note on the use of chromosome conformation capture in plants. *Plant Methods* 13, 101.

Jegu, T., Latrasse, D., Delarue, M., Hirt, H., Domenichini, S., Ariel, F., Crespi, M., Bergounioux, C., Raynaud, C., and Benhamed, M. (2014). The BAF60 Subunit of the SWI/SNF Chromatin-Remodeling Complex Directly Controls the Formation of a Gene Loop at FLOWERING LOCUS C in Arabidopsis. *Plant Cell* 26, 538–551.

Jégu, T., Domenichini, S., Blein, T., Ariel, F., Christ, A., Kim, S.-K., Crespi, M., Boutet-Mercey, S., Mouille, G., Bourge, M., et al. (2015). A SWI/SNF Chromatin Remodelling Protein Controls Cytokinin Production through the Regulation of Chromatin Architecture. *PLoS One* 10, e0138276.

Jeong, H.J., Yang, J., Cho, L.H., and An, G. (2016). OsVIL1 controls flowering time in rice by suppressing OsLF under short days and by inducing Ghd7 under long days. *Plant Cell Rep.* 35, 905–920.

Jiang, D., and Berger, F. (2017). DNA replication-coupled histone modification maintains Polycomb gene silencing in plants. *Science* (80-.). 4965, 1146–1149.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science* (80-.). 290, 344–347.

Jones, J.D.G., Shlumukov, L., Carland, F., English, J., Scofield, S.R., Bishop, G.J., and Harrison, K. (1992). Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. *Transgenic Res.* 1, 285–297.

Joyce, N., M., H.C., Kelly, M., and A., S.J. (2000). A *Drosophila* ESC-E(Z) Protein Complex Is Distinct from Other Polycomb Group Complexes and Contains Covalently Modified ESC. *Mol. Cell. Biol.* *20*, 3069–3078.

Kagale, S., and Rozwadowski, K. (2011). EAR motif-mediated transcriptional repression in plants: an underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* *6*, 141–146.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial Interactions Are Required for the Efficient Recruitment of Pho Repressive Complex (PhoRC) to Polycomb Response Elements. *PLoS Genet.* *10*.

Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev.* *28*, 1983–1988.

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. (2015). Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in *Drosophila*. *Genes Dev.* *29*, 1136–1150.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* *38*, 675–688.

Kasinath, V., Faini, M., Poepsel, S., Reif, D., Feng, X.A., Stjepanovic, G., Aebbersold, R., and Nogales, E. (2018). Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science* *359*, 940–944.

Kasinath, V., Beck, C., Sauer, P., Poepsel, S., Kosmatka, J., Faini, M., Toso, D., Aebersold, R., and Nogales, E. (2021). JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science* (80-.). *371*, eabc3393.

Kassis, J.A., and Brown, J.L. (2013). Polycomb group response elements in *Drosophila* and vertebrates. *Adv. Genet.* *81*, 83–118.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* *10*, 845–858.

Kim, D.-H., and Sung, S. (2013). Coordination of the vernalization response through a VIN3 and FLC gene family regulatory network in *Arabidopsis*. *Plant Cell* *25*, 454–469.

Kim, D.-H.H., and Sung, S. (2010). The Plant Homeo Domain finger protein, VIN3-LIKE 2, is necessary for photoperiod-mediated epigenetic regulation of the floral repressor, MAF5. *107*, 17029–17034.

Kim, D.-H.H., and Sung, S. (2017a). The binding specificity of the PHD-Finger domain of VIN3 moderates vernalization response. *Plant Physiol.* *173*, 1258–1268.

Kim, D.H., and Sung, S. (2017b). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. *Dev. Cell* *40*, 302–312.e4.

Kim, C.A., Gingery, M., Pilpa, R.M., and Bowie, J.U. (2002). The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.* *9*, 453–457.

Kim, D.-H., Xi, Y., and Sung, S. (2017). Modular function of long noncoding RNA, COLDAIR, in the vernalization response. *PLoS Genet.* *13*, e1006939–e1006939.

Kim, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* *37*, 2940–2950.

Kim, J., Kwon, J., Kim, M., Do, J., Lee, D., and Han, H. (2016). Low-dielectric-constant polyimide aerogel composite films with low water uptake. *Polym. J.* *48*, 829–834.

Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKKA governs plant cold acclimation. *Nat. Commun.* *9*, 4561.

Kindgren, P., Ivanov, M., and Marquardt, S. (2020). Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in Arabidopsis. *Nucleic Acids Res.* *48*, 2332–2347.

King, G.J., Chanson, A.H., McCallum, E.J., Ohme-Takagi, M., Byriel, K., Hill, J.M., Martin, J.L., and Mylne, J.S. (2013). The Arabidopsis B3 domain protein VERNALIZATION1 (VRN1) is involved in processes essential for development, with structural and mutational studies revealing its DNA-binding surface. *J. Biol. Chem.* *288*, 3198–3207.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet.* *9*, e1003717–e1003717.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* *20*, 1110–1122.

Kooiker, M., Airoidi, C.A., Losa, A., Manzotti, P.S., Finzi, L., Kater, M.M., and Colombo, L. (2005). BASIC PENTACYSTEINE1, a GA Binding Protein That Induces Conformational Changes in the Regulatory Region of the Homeotic Arabidopsis Gene SEEDSTICK. *Plant Cell* 17, 722–729.

Kraft, K., Yost, K.E., Murphy, S., Magg, A., Long, Y., Corces, M.R., Granja, J.M., Mundlos, S., Cech, T.R., Boettiger, A., et al. (2020). Polycomb-mediated Genome Architecture Enables Long-range Spreading of H3K27 methylation. *BioRxiv* 2020.07.27.223438.

Kralemann, L.E.M., Liu, S., Trejo-Arellano, M.S., Muñoz-Viana, R., Köhler, C., and Hennig, L. (2020). Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in Arabidopsis. *Genome Biol.* 21, 1–19.

Kretz, M., and Meister, G. (2014). RNA Binding of PRC2: Promiscuous or Well Ordered? *Mol. Cell* 55, 157–158.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4, e1000242–e1000242.

Lainé, J.-P., Singh, B.N., Krishnamurthy, S., and Hampsey, M. (2009). A physiological role for gene loops in yeast. *Genes Dev.* 23, 2604–2609.

Laprell, F., Finkl, K., and Müller, J. (2017). Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* (80-.). 356, 85–88.

Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* 74, 8–18.

Lebendiker, M., and Danieli, T. (2014). Production of prone-to-aggregate proteins. *FEBS Lett.* 588, 236–246.

Lee, C.-H., Holder, M., Grau, D., Saldaña-Meyer, R., Yu, J.-R., Ganai, R.A., Zhang, J., Wang, M., LeRoy, G., and Dobenecker, M.-W. (2018). Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2. *Mol. Cell* 70, 435–448.

Lee, I., Michaels, S.D., Masshardt, A.S., and Amasino, R.M. (1994). The late-flowering phenotype of FRIGIDA and mutations in LUMINIDEPENDENS is suppressed in the Landsberg erecta strain of Arabidopsis. *Plant J.* 6, 903–909.

Levin, J.Z., and Meyerowitz, E.M. (1995). UFO: an Arabidopsis gene involved in both floral meristem and floral organ development. *Plant Cell* 7, 529–548.

Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R., and Dean, C. (2002). Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science* (80-.). 297, 243–246.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.

Li, H., Liefke, R., Jiang, J., Kurland, J.V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D.J., Bulyk, M.L., et al. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* 549, 287–291.

Li, X., Zhang, S., Bai, J., and He, Y. (2016). Tuning growth cycles of Brassica crops via natural antisense transcripts of BrFLC. *Plant Biotechnol. J.* *14*, 905–914.

Li, Z., Jiang, D., and He, Y. (2018). FRIGIDA establishes a local chromosomal environment for FLOWERING LOCUS C mRNA production. *Nat. Plants* *4*, 836–846.

Liang, S.C., Hartwig, B., Perera, P., Mora-García, S., de Leau, E., Thornton, H., de Alves, F.L., Rapsilber, J., Yang, S., James, G.V., et al. (2015). Kicking against the PRCs – A Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. *PLOS Genet.* *11*, e1005660.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* *326*, 289 – 293.

Liu, C., Teo, Z.W.N., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z., and Yu, H. (2013). A Conserved Genetic Pathway Determines Inflorescence Architecture in Arabidopsis and Rice. *Dev. Cell* *24*, 612–622.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M., and Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res.* *26*, 1057–1068.

Liu, J., Wu, X., Zhang, H., Pfeifer, G.P., and Lu, Q. (2017). Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. *Cell Rep.* *20*, 1307–1318.

Liu, R., Gao, J., Yang, Y., Qiu, R., Zheng, Y., Huang, W., Zeng, Y., Hou, Y., Wang, S., Leng, S., et al. (2018). PHD finger protein 1 (PHF1) is a novel reader for histone H4R3 symmetric dimethylation and coordinates with PRMT5-WDR77/CRL4B complex to promote tumorigenesis. *Nucleic Acids Res.* *46*, 6608–6626.

Liu, Z., Li, F., Ruan, K., Zhang, J., Mei, Y., Wu, J., and Shi, Y. (2014). Structural and functional insights into the human Börjeson-Forssman-Lehmann syndrome-associated protein PHF6. *J. Biol. Chem.* *289*, 10069–10083.

Lodha, M., Marco, C.F., and Timmermans, M.C.P. (2013). The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. *Genes Dev.* *27*, 596–601.

Long, J.A., Ohno, C., Smith, Z.R., and Meyerowitz, E.M. (2006). TOPLESS Regulates Apical Embryonic Fate in Arabidopsis. *Science* *312*, 1520–1523.

Long, Y., Hwang, T., Gooding, A.R., Goodrich, K.J., Rinn, J.L., and Cech, T.R. (2020). RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. *Nat. Genet.* *52*, 931–938.

Louwers, M., Bader, R., Haring, M., van Driel, R., de Laat, W., and Stam, M. (2009). Tissue- and expression level-specific chromatin looping at maize b1 epialleles. *Plant Cell* *21*, 832–842.

Lövkvist, C., Mikulski, P., Reeck, S., Hartley, M., Dean, C., and Howard, M. Hybrid protein oligomer-histone modification mechanism for PRC2-based epigenetic switching and memory. *Elife*.

De Lucia, F., Crevillen, P., Jones, A.M.E., Greb, T., and Dean, C. (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 16831–16836.

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

Luo, X., Chen, T., Zeng, X., He, D., and He, Y. (2019). Feedback Regulation of FLC by FLOWERING LOCUS T (FT) and FD through a 5' FLC Promoter Region in Arabidopsis. *Mol. Plant* *12*, 285–288.

Lykke-Andersen, S., Mapendano, C.K., and Jensen, T.H. (2011). An ending is a new beginning: Transcription termination supports re-initiation. *Cell Cycle* *10*, 863–865.

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

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., Dynlacht, B.D., and Reinberg, D. (2008). Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol. Cell* *32*, 503–518.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury III, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* *461*, 762–767.

Mateo-Bonmatí, E., Esteve-Bruna, D., Juan-Vicente, L., Nadi, R., Candela, H., Lozano, F.M., Ponce, M.R., Pérez-Pérez, J.M., and Micol, J.L. (2018). *INCURVATA11* and *CUPULIFORMIS2* Are Redundant Genes That Encode Epigenetic Machinery Components in *Arabidopsis*. *Plant Cell* *30*, 1596–1616.

McCord, R.P., Kaplan, N., and Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol. Cell* *77*, 688–708.

Mellor, J. (2006). It Takes a PHD to Read the Histone Code. *Cell* *126*, 22–24.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* *6*, e1001244.

Merini, W., and Calonje, M. (2015). PRC1 is taking the lead in PcG repression. *Plant J.* *83*, 110–120.

Mermaz, B. (2019). Dynamics of the histone modification H3K27me3 on plant development. Imperial College London.

Michaels, S.D., and Amasino, R.M. (1999). *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* *11*, 949–956.

Michaels, S.D., Himmelblau, E., Sang, Y.K., Schomburg, F.M., and Amasino, R.M. (2005). Integration of flowering signals in Winter-annual *Arabidopsis*. *Plant Physiol.* *137*, 149–156.

van Mierlo, G., Veenstra, G.J.C., Vermeulen, M., and Marks, H. (2019). The Complexity of PRC2 Subcomplexes. *Trends Cell Biol.* 29, 660–671.

Mikulski, P., Wolf, P., Lu, T., Zhu, D., and Dean, C. (2021). Transcriptional and chromatin regulation at Arabidopsis FLC in advance of long-term PRC2 silencing. Unpublished.

Miller, T.C.R., Simon, B., Rybin, V., Grötsch, H., Curtet, S., Khochbin, S., Carlomagno, T., and Müller, C.W. (2016). A bromodomain-DNA interaction facilitates acetylation-dependent bivalent nucleosome recognition by the BET protein BRDT. *Nat. Commun.* 7.

Molitor, A.M., Bu, Z., Yu, Y., and Shen, W.-H. (2014). Arabidopsis ALPHD-PRC1 Complexes Promote Seed Germination through H3K4me3-to-H3K27me3 Chromatin State Switch in Repression of Seed Developmental Genes. *PLOS Genet.* 10, e1004091.

Monaghan, J., Xu, F., Gao, M., Zhao, Q., Palma, K., Long, C., Chen, S., Zhang, Y., and Li, X. (2009). Two Prp19-Like U-Box Proteins in the MOS4-Associated Complex Play Redundant Roles in Plant Innate Immunity. *PLOS Pathog.* 5, e1000526.

Moore, J.O., Lemmon, M.A., and Ferguson, K.M. (2017). Dimerization of Tie2 mediated by its membrane-proximal FNIII domains. *Proc. Natl. Acad. Sci. U. S. A.* 114, 4382–4387.

Mozgova, I., Köhler, C., and Hennig, L. (2015). Keeping the gate closed: functions of the polycomb repressive complex PRC2 in development. *Plant J.* 83, 121–132.

Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.* 16, 476–484.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* *111*, 197–208.

Muskett, P.R., Clissold, L., Marocco, A., Springer, P.S., Martienssen, R., and Dean, C. (2003). A resource of mapped dissociation launch pads for targeted insertional mutagenesis in the *Arabidopsis* genome. *Plant Physiol.* *132*, 506–516.

Musselman, C.A., Mansfield, R.E., Garske, A.L., Davrazou, F., Kwan, A.H., Oliver, S.S., O'Leary, H., Denu, J.M., Mackay, J.P., and Kutateladze, T.G. (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem. J.* *423*, 179–187.

Musselman, C.A., Lalonde, M.E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* *19*, 1218–1227.

Mylne, J., Greb, T., Lister, C., and Dean, C. (2004). Epigenetic regulation in the control of flowering. *Cold Spring Harb Symp Quant Biol* *69*, 457–464.

Mylne, J.S., Barrett, L., Tessadori, F., Mesnage, S., Johnson, L., Bernatavichute, Y. V, Jacobsen, S.E., Fransz, P., and Dean, C. (2006). LHP1, the *Arabidopsis* homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 5012–5017.

Nguyen, T., Fischl, H., Howe, F.S., Woloszczuk, R., Barros, A.S., Xu, Z., Brown, D., Murray, S.C., Haenni, S., Halstead, J.M., et al. (2014). Transcription mediated insulation and interference direct gene cluster expression switches. *Elife* *3*, 1–21.

Nielsen, M., Ard, R., Leng, X., Ivanov, M., Kindgren, P., Pelechano, V., and Marquardt, S. (2019). Transcription-driven chromatin repression of intragenic transcription start sites. *PLoS Genet.* *15*, e1007969.

O'Connell, S., Wang, L., Robert, S., Jones, C.A., Saint, R., and Jones, R.S. (2001). Polycomb-like PHD fingers mediate conserved interaction with enhancer of zeste protein. *J. Biol. Chem.* *276*, 43065–43073.

O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* *165*, 1280–1292.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* *36*, 1014–1018.

Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.M., and Cavalli, G. (2018). Polycomb-dependent chromatin looping contributes to gene silencing during *Drosophila* development. *Mol. Cell* *71*, 73–88.e5.

Oksuz, O., Narendra, V., Lee, C.-H.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.R.P., Garcia, B.A., et al. (2018). Capturing the onset of PRC2-mediated repressive domain formation. *Mol. Cell* *70*, 1149–1162.

Packman, L.C., Borges, A., and Perham, R.N. (1988). Amino acid sequence analysis of the lipoyl and peripheral subunit-binding domains in the lipoate acetyltransferase component of the pyruvate dehydrogenase complex from *Bacillus stearothermophilus*. *Biochem. J.* *252*, 79–86.

Patel, D.J., and Wang, Z. (2013). Readout of epigenetic modifications. *Annu. Rev. Biochem.* *82*, 81–118.

Peng, L., Wang, L., Zhang, Y., Dong, A., Shen, W.-H., and Huang, Y. (2018). Structural Analysis of the Arabidopsis AL2-PAL and PRC1 Complex Provides Mechanistic Insight into Active-to-Repressive Chromatin State Switch. *J. Mol. Biol.* *430*, 4245–4259.

Perino, M., van Mierlo, G., Karemaker, I.D., van Genesen, S., Vermeulen, M., Marks, H., van Heeringen, S.J., and Veenstra, G.J.C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nat. Genet.* *50*, 1002–1010.

Perišić, O., Collepardo-Guevara, R., and Schlick, T. (2010). Modeling studies of chromatin fiber structure as a function of DNA linker length. *J. Mol. Biol.* *403*, 777–802.

Piunti, A., and Shilatifard, A. (2021). The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* *22*, 326–345.

Porebski, B.T., Nickson, A.A., Hoke, D.E., Hunter, M.R., Zhu, L., McGowan, S., Webb, G.I., and Buckle, A.M. (2015). Structural and dynamic properties that govern the stability of an engineered fibronectin type III domain. *Protein Eng. Des. Sel.* *28*, 67–78.

Qüesta, J.I., Song, J., Geraldo, N., An, H., and Dean, C. (2016). Arabidopsis transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science* *353*, 485–488.

Qüesta, J.I., Antoniou-Kourouniotti, R.L., Rosa, S., Li, P., Duncan, S., Whittaker, C., Howard, M., and Dean, C. (2020). Noncoding SNPs influence a distinct phase of Polycomb silencing to destabilize long-term epigenetic memory at Arabidopsis FLC. *Genes Dev.* 34, 446–461.

Ragazzini, R., Pérez-Palacios, R., Baymaz, I.H., Diop, S., Ancelin, K., Zielinski, D., Michaud, A., Givelet, M., Borsos, M., Aflaki, S., et al. (2019). EZHIP constrains Polycomb Repressive Complex 2 activity in germ cells. *Nat. Commun.* 10, 3858.

Rai, A.K., Chen, J.-X., Selbach, M., and Pelkmans, L. (2018). Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* 559, 211–216.

Ramirez-Prado, J.S., Abulfaraj, A.A., Rayapuram, N., Benhamed, M., and Hirt, H. (2018). Plant Immunity: From Signaling to Epigenetic Control of Defense. *Trends Plant Sci.* 23, 833–844.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. *Mol. Cell* 55, 347–360.

Ringrose, L., Rehmsmeier, M., Dura, J.M., and Paro, R. (2003). Genome-wide prediction of polycomb/trithorax response elements in *Drosophila melanogaster*. *Dev. Cell* 5, 759–771.

Rosa, S., De Lucia, F., Mylne, J.S., Zhu, D., Ohmido, N., Pendle, A., Kato, N., Shaw, P., and Dean, C. (2013). Physical clustering of FLC alleles during polycomb-mediated epigenetic silencing in vernalization. *Genes Dev.* 27, 1845–1850.

Rosa, S., Duncan, S., and Dean, C. (2016). Mutually exclusive sense-antisense transcription at FLC facilitates environmentally induced gene repression. *Nat. Commun.* *7*, 1–7.

Rousseau, F., Schymkowitz, J.W.H., and Itzhaki, L.S. (2003). The unfolding story of three-dimensional domain swapping. *Structure* *11*, 243–251.

Ruthenburg, A.J., Li, H., Patel, D.J., and David Allis, C. (2007). Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* *8*, 983–994.

Sabari, B.R., Dall’Agnese, A., and Young, R.A. (2020). Biomolecular Condensates in the Nucleus. *Trends Biochem. Sci.* *45*, 961–977.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol. Cell* *57*, 769–783.

Sati, S., and Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* *126*, 33–44.

Savitsky, P., Krojer, T., Fujisawa, T., Lambert, J.-P., Picaud, S., Wang, C.-Y., Shanle, E.K., Krajewski, K., Friedrichsen, H., Kanapin, A., et al. (2016). Multivalent Histone and DNA Engagement by a PHD/BRD/PWWP Triple Reader Cassette Recruits ZMYND8 to K14ac-Rich Chromatin. *Cell Rep.* *17*, 2724–2737.

Saxton, D.S., and Rine, J. (2019). Epigenetic memory independent of symmetric histone inheritance. *Elife* *8*, e51421.

Sayou, C., Nanao, M.H., Jamin, M., Posé, D., Thévenon, E., Grégoire, L., Tichtinsky, G., Denay, G., Ott, F., Peirats Llobet, M., et al. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nat. Commun.* 7, 11222.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* 465, 243–247.

Schlissel, G., and Rine, J. (2019). The nucleosome core particle remembers its position through DNA replication and RNA transcription. *Proc. Natl. Acad. Sci.* 116, 20605–20611.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., and Zhou, S. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* 42, 330–341.

Schon, M., Baxter, C., Xu, C., Enugutti, B., Nodine, M.D., and Dean, C. (2021). Antagonistic activities of cotranscriptional regulators within an early developmental window set FLC expression level. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–3.

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. *EMBO J.* 25, 4638–4649.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in *Drosophila* Embryos. *PLOS Biol.* 7, e1000013.

Schumacher, M.A., Chinnam, N., Ohashi, T., Shah, R.S., and Erickson, H.P. (2013). The structure of Irisin reveals a novel intersubunit β -sheet fibronectin type III (FNIII) dimer: Implications for receptor activation. *J. Biol. Chem.* *288*, 33738–33744.

Schwarz-Romond, T., Fiedler, M., Shibata, N., Butler, P.J.G., Kikuchi, A., Higuchi, Y., and Bienz, M. (2007). The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. *Nat. Struct. Mol. Biol.* *14*, 484–492.

Seif, E., Kang, J.J., Sasseville, C., Senkovich, O., Kaltashov, A., Boulier, E.L., Kapur, I., Kim, C.A., and Francis, N.J. (2020). Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nat. Commun.* *11*, 1–19.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome. *Cell* *148*, 458–472.

Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J., and Dennis, E.S. (1999). The FLF MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* *11*, 445–458.

Sheldon, C.C., Conn, A.B., Dennis, E.S., and Peacock, W.J. (2002). Different regulatory regions are required for the vernalization-induced repression of FLOWERING LOCUS C and for the epigenetic maintenance of repression. *Plant Cell* *14*, 2527–2537.

Sheldon, C.C., Hills, M.J., Lister, C., Dean, C., Dennis, E.S., and Peacock, W.J. (2008). Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 2214–2219.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H.A., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G.E., Fingerman, I.M., Briggs, S.D., et al. (2007). Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J. Biol. Chem.* *282*, 2450–2455.

Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of FLC contributes to natural variation in *Arabidopsis* vernalization response. *Genes Dev.* *20*, 3079–3083.

Shu, J., Chen, C., Thapa, R.K., Bian, S., Nguyen, V., Yu, K., Yuan, Z.C., Liu, J., Kohalmi, S.E., Li, C., et al. (2019). Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings. *Plant Direct* *3*.

Shwartz, I., Yahav, C., Kovetz, N., Israeli, A., Bar, M., and Levy, M. (2021). The VIL gene CRAWLING ELEPHANT controls maturation and differentiation in tomato via polycomb silencing. *Biorxiv*

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* *10*, 697–708.

Skourti-Stathaki, K., Triglia, E.T., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019). R-loops enhance polycomb repression at a subset of developmental regulator genes. *Mol. Cell* *73*, 930–945.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl. Acad. Sci.* *107*, 20980–20985.

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

Song, J., Angel, A., Howard, M., and Dean, C. (2012). Vernalization – a cold-induced epigenetic switch. *J. Cell Sci.* *125*, 3723–3731.

Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* *15*, 340–356.

Stewart-Morgan, K.R., Petryk, N., and Groth, A. (2020). Chromatin replication and epigenetic cell memory. *Nat. Cell Biol.* *22*, 361–371.

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

Strübbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C., and Paro, R. (2011). Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. *Proc. Natl. Acad. Sci.* *108*, 5572 – 5577.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the arabidopsis FLC locus. *Science* *340*, 619–621.

Sung, S., and Amasino, R.M. (2004). Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* 427, 159–164.

Sung, S., Schmitz, R.J., and Amasino, R.M. (2006). A PHD finger protein involved in both the vernalization and photoperiod pathways in *Arabidopsis*. *Genes Dev.* 20, 3244–3248.

Suzuki, M., Wang, H.H.-Y., and McCarty, D.R. (2007). Repression of the LEAFY COTYLEDON 1/B3 Regulatory Network in Plant Embryo Development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 Genes. *Plant Physiol.* 143, 902–911.

Swiezewski, S., Crevillen, P., Liu, F., Ecker, J.R., Jerzmanowski, A., and Dean, C. (2007). Small RNA-mediated chromatin silencing directed to the 3' region of the *Arabidopsis* gene encoding the developmental regulator, FLC. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3633–3638.

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* 462, 799–802.

Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci. Adv.* 5, eaaw1668.

Tan-Wong, S.M., Wijayatilake, H.D., and Proudfoot, N.J. (2009). Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* 23, 2610–2624.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. *Science* 338, 671–675.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 5521–5526.

Teplyakov, A., Obmolova, G., Malia, T.J., Luo, J., Jacobs, S.A., Chan, W., Domingo, D., Baker, A., O'Neil, K.T., and Gilliland, G.L. (2014). C-terminal β -strand swapping in a consensus-derived fibronectin Type III scaffold. *Proteins Struct. Funct. Bioinforma.* *82*, 1359–1369.

Thompson, C.J., Movva, N.R., Tizard, R., Cramer, R., Davies, J.E., Lauwereys, M., and Botterman, J. (1987). Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygrosopicus*. *EMBO J.* *6*, 2519–2523.

Tian, Y., Zheng, H., Zhang, F., Wang, S., Ji, X., Xu, C., He, Y., and Ding, Y. (2019). PRC2 recruitment and H3K27me3 deposition at FLC require FCA binding of COOLAIR. *Sci. Adv.* *5*, 1–13.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding p55 and the histone deacetylase RPD3. *Development* *128*, 275–286.

Trojer, P., and Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Mol. Cell* *28*, 1–13.

Tsusaka, T., Fukuda, K., Shimura, C., Kato, M., and Shinkai, Y. (2020). The fibronectin type-III (FNIII) domain of ATF7IP contributes to efficient transcriptional silencing mediated by the SETDB1 complex. *Epigenetics Chromatin* *13*, 52.

Velanis, C.N., Perera, P., Thomson, B., de Leau, E., Liang, S.C., Hartwig, B., Förderer, A., Thornton, H., Arede, P., Chen, J., et al. (2020). The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2). *PLoS Genet.* *16*, 1–30.

Veluchamy, A., Jégu, T., Ariel, F., Latrasse, D., Mariappan, K.G., Kim, S.K., Crespi, M., Hirt, H., Bergounioux, C., Raynaud, C., et al. (2016). LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One* *11*, 1–25.

Villaseñor, R., and Baubec, T. (2021). Regulatory mechanisms governing chromatin organization and function. *Curr. Opin. Cell Biol.* *70*, 10–17.

Wang, L., Jähren, N., Miller, E.L., Ketel, C.S., Mallin, D.R., and Simon, J.A. (2010). Comparative analysis of chromatin binding by Sex Comb on Midleg (SCM) and other polycomb group repressors at a Drosophila Hox gene. *Mol. Cell. Biol.* *30*, 2584–2593.

Wang, X., Paucek, R.D., Gooding, A.R., Brown, Z.Z., Ge, E.J., Muir, T.W., and Cech, T.R. (2017). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat. Struct. Mol. Biol.* *24*, 1028–1038.

Wani, A.H., Boettiger, A.N., Schorderet, P., Ergun, A., Münger, C., Sadreyev, R.I., Zhuang, X., Kingston, R.E., Francis, N.J., Munger, C., et al. (2016). Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* *7*, 10291.

Weaver, T.M., Morrison, E.A., and Musselman, C.A. (2018). Reading more than Histones: The prevalence of nucleic acid binding among reader domains. *Molecules* *23*, 1–25.

Wei, C., Xiao, R., Chen, L., Cui, H., Zhou, Y., Xue, Y., Hu, J., Zhou, B., Tsutsui, T., Qiu, J., et al. (2016). RBFox2 Binds Nascent RNA to Globally Regulate Polycomb Complex 2 Targeting in Mammalian Genomes. *Mol. Cell* 62, 875–889.

Whittaker, C., and Dean, C. (2017). The FLC locus: A platform for discoveries in epigenetics and adaptation. *Annu. Rev. Cell Dev. Biol.* 33, 555–575.

Wiles, E.T., McNaught, K.J., Kaur, G., Selker, J.M.L., Ormsby, T., Aravind, L., and Selker, E.U. (2020). Evolutionarily ancient BAH-PHD protein mediates Polycomb silencing. *Proc. Natl. Acad. Sci. U. S. A.* 117. Wojcik, J., Hantschel, O., Grebien, F., Kaupe, I., Bennett, K.L., Barkinge, J., Jones, R.B., Koide, A., Superti-Furga, G., and Koide, S. (2010). A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat. Struct. Mol. Biol.* 17, 519–527.

Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., and Helliwell, C.A. (2006). The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14631–14636.

Wu, H., Deng, S., Xu, H., Mao, H., Liu, J., Niu, Q., Wang, H., and Chua, N. (2018). A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves. *New Phytol.* 219, 1480–1491.

Wu, Z., Qiu, J., Shi, H., Lin, C., Yue, J., Liu, Z., Xie, W., Kou, Y., and Tao, Z. (2021). Polycomb repressive complex 2 coordinates with Sin3 histone deacetylase complex to epigenetically reprogram genome-wide expression of effectors and regulate pathogenicity in *Magnaporthe oryzae*. *BioRxiv*.

Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J.D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., et al. (2017). Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nat. Genet.* *49*, 1546–1552.

Yadav, V.K., Santos-González, J., and Köhler, C. (2021). INT-Hi-C reveals distinct chromatin architecture in endosperm and leaf tissues of *Arabidopsis*. *Nucleic Acids Res.* 1–15.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013a). VAL- and AtBMI1-Mediated H2Aub Initiate the Switch from Embryonic to Postgerminative Growth in *Arabidopsis*. *Curr. Biol.* *23*, 1324–1329.

Yang, H., Howard, M., and Dean, C. (2014). Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at *Arabidopsis* FLC. *Curr. Biol.* *24*, 1793–1797.

Yang, H., Berry, S., Olsson, T.S.G., Hartley, M., Howard, M., and Dean, C. (2017). Distinct phases of Polycomb silencing to hold epigenetic memory of cold in *Arabidopsis*. *Science* (80-.). *357*, 1142–1145.

Yang, S., Bansal, K., Lopes, J., Benoist, C., and Mathis, D. (2013b). Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 1833–1838.

Youmans, D.T., Schmidt, J.C., and Cech, T.R. (2018). Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. *Genes Dev.* *32*, 794–805.

Youmans, D.T., Gooding, A.R., Dowell, R.D., and Cech, T.R. (2021). Competition between PRC2.1 and 2.2 subcomplexes regulates PRC2 chromatin occupancy in human stem cells. *Mol. Cell* *81*, 488-501.e9.

Yu, B., Qi, Y., Li, R., Shi, Q., Satpathy, A.T., and Chang, H.Y. (2021). B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells. *Cell* *184*, 1790-1803.e17.

Yu, J.R., Lee, C.H., Oksuz, O., Stafford, J.M., and Reinberg, D. (2019). PRC2 is high maintenance. *Genes Dev.* *33*, 903–935.

Yuan, L., Song, X., Zhang, L., Yu, Y., Liang, Z., Lei, Y., Ruan, J., Tan, B., Liu, J., and Li, C. (2020). The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in Arabidopsis. *Nucleic Acids Res.* *49*, 98–113.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* *286*, 7983–7989.

Yuan, W., Luo, X., Li, Z., Yang, W., Wang, Y., Liu, R., Du, J., and He, Y. (2016). A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in Arabidopsis. *Nat. Genet.* *48*, 1527–1534.

Zeng, X., Gao, Z., Jiang, C., Yang, Y., Liu, R., and He, Y. (2020). HISTONE DEACETYLASE 9 Functions with Polycomb Silencing to Repress FLOWERING LOCUS C Expression1. *Plant Physiol.* *182*, 555–565.

Zhang, Q., Agius, S.C., Flanigan, S.F., Levina, V., Owen, B.M., and Davidovich, C. (2020a). Convergent evolution between PALI1 and JARID2 for the allosteric activation of PRC2. *BioRxiv* 2020.05.28.122556.

Zhang, Y., Yuan, J., Zhang, L., Chen, C., Wang, Y., Zhang, G., Peng, L., Xie, S., Jiang, J., Zhu, J., et al. (2020b). transcriptional repression through the BAH-PHD-CPL2 complex in Arabidopsis. *Nat. Commun.*

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.

Zhao, Y., Antoniou-Kourounioti, R.L., Calder, G., Dean, C., and Howard, M. (2020). Temperature-dependent growth contributes to long-term cold sensing. *Nature* 583, 825–829.

Zhao, Y., Zhu, P., Hepworth, J., Bloomer, R., Antoniou-kourounioti, R.L., Doughty, J., Heckmann, A., Xu, C., Yang, H., and Dean, C. (2021). Natural temperature fluctuations promote COOLAIR regulation of FLC. 1–11.

Zhou, Y., Hartwig, B., James, G.V., Schneeberger, K., and Turck, F. (2016). Complementary Activities of TELOMERE REPEAT BINDING Proteins and Polycomb Group Complexes in Transcriptional Regulation of Target Genes. *Plant Cell* 28, 87–101.

Zhou, Y., Wang, Y., Krause, K., Yang, T., Dongus, J.A., Zhang, Y., and Turck, F. (2018). Telobox motifs recruit CLF/SWN–PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nat. Genet.* 50, 638–644.

Zou, Z., Cao, L., Zhou, P., Su, Y., Sun, Y., and Li, W. (2008). Hyperacidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *J. Biotechnol.* *135*, 333–339.