



**HAL**  
open science

# Étude biochimique et structurale du complexe PRC2 et de la protéine ULTRAPETALA1, deux régulateurs antagonistes de la chromatine chez *Arabidopsis thaliana*

Jean-Baptiste Izquierdo

## ► To cite this version:

Jean-Baptiste Izquierdo. Étude biochimique et structurale du complexe PRC2 et de la protéine ULTRAPETALA1, deux régulateurs antagonistes de la chromatine chez *Arabidopsis thaliana*. Sciences du Vivant [q-bio]. Université Grenoble Alpes (UGA), 2023. Français. NNT : 2023GRALV069 . tel-04837705

**HAL Id: tel-04837705**

**<https://hal.science/tel-04837705v1>**

Submitted on 13 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE  
Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : CSV- Chimie et Sciences du Vivant  
Spécialité : Biologie Structurale et Nanobiologie  
Unité de recherche : LPCV - Laboratoire de Physiologie Cellulaire Végétale

**Etude biochimique et structurale du complexe PRC2 et de la protéine ULTRAPETALA1, deux régulateurs antagonistes de la chromatine chez *Arabidopsis thaliana***

**Biochemical and structural study of the plant PRC2 complex and ULTRAPETALA1, two antagonistic regulators of chromatin in *Arabidopsis thaliana***

Présentée par :  
**Jean-Baptiste IZQUIERDO**

Direction de thèse :

**Christel CARLES**  
Université Grenoble Alpes  
**Jan KADLEC**  
Université Grenoble Alpes

Directrice de thèse  
Co-directeur de thèse

Rapporteurs :

**Elisa BERGAMIN**  
CHARGE DE RECHERCHE HDR, CNRS  
**Raphaël MARGUERON**  
DIRECTEUR DE RECHERCHE, INSERM

Thèse soutenue publiquement le 12 octobre 2023, devant le jury composé de :

<b>Elisa BERGAMIN</b> CHARGE DE RECHERCHE HDR, CNRS	Rapporteuse
<b>Raphaël MARGUERON</b> DIRECTEUR DE RECHERCHE, INSERM	Rapporteur
<b>Aline PROBST</b> DIRECTEUR DE RECHERCHE, CNRS	Examinatrice
<b>Winfried WEISSENHORN, Président du jury</b> PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Examinateur
<b>Christel CARLES</b> PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Directrice de thèse
<b>Jan KADLEC</b> CHARGE DE RECHERCHE HDR, INSERM	Co-directeur de thèse





# Table des matières

<b>Table des matières</b>	<b>3</b>
<b>Liste des figures et des tableaux</b>	<b>6</b>
<b>Résumé</b>	<b>9</b>
<b>Summary</b>	<b>11</b>
<b>Liste des abréviations</b>	<b>13</b>
<b>I. Introduction</b>	<b>16</b>
<b>A. Structure et dynamique de la chromatine</b>	<b>17</b>
1. Structure et fonction de la chromatine	17
2. Modification de la chromatine	20
a. Variants d'histones	20
b. Complexe de remodelage de la chromatine dépendant de l'ATP	21
c. Méthylation de l'ADN	21
d. Modification d'histones	22
1) Acétylation des histones	23
2) Méthylation des histones	24
a) Méthylation sur H3K4 et H3K36	24
b) Méthylation sur H3K9 et H3K27	26
3. État chromatiniens	29
<b>B. Les protéines des groupes Polycomb et Trithorax fonctionnent de manière antagoniste</b>	<b>29</b>
1. Composition et fonction du complexe protéique PRC2	31
2. Structures du PRC2	33
3. Régulation de l'activité du PRC2	36
4. Composition des complexes PRC2 canoniques chez Arabidopsis	38
5. Mécanisme de recrutement du PRC2 chez Arabidopsis	42
6. Dépôt de la marque H2AK121ub par le complexe PRC1	43
7. Rôles d'H2AK121ub et H3K27me3 dans la régulation de la transcription chez les plantes	45
8. De nombreux effecteurs influencent la répression des gènes <i>via</i> divers mécanismes	46
9. Les protéines Trithorax, acteurs associés à l'activité de transcription	49
<b>C. Rôles de PcG et TrxG dans le développement des plantes</b>	<b>50</b>
1. Rôle de Polycomb et Trithorax dans le maintien de l'homéostasie des méristèmes	52
2. Rôle de Polycomb et Trithorax dans l'induction de la floraison et l'organogenèse florale	53
<b>D. ULTRAPETALA1 (ULT1), une protéine à domaine SAND</b>	<b>55</b>
1. ULT1 est impliquée dans plusieurs processus développementaux, tous liés aux transitions d'identité cellulaire	56
2. ULT1, une protéine antagoniste au PRC2	59
<b>E. Objectifs de la thèse</b>	<b>62</b>
<b>II. Matériels et Méthodes</b>	<b>64</b>
<b>A. Production d'ULT1</b>	<b>65</b>
1. Clonage des constructions ULT1 <sub>1-237</sub> , ULT1 <sub>1-121</sub> et ULT1 <sub>128-237</sub>	65
2. Production de ULT1 <sub>1-237</sub> , ULT1 <sub>1-121</sub> et ULT1 <sub>128-237</sub> en système bactérien	67
3. Lyse des culots bactériens	67
4. Purification d'ULT1 FL et SAND	67
a. Chromatographie d'affinité sur résine d'amylose	67

b.	Première chromatographie d'exclusion de taille	68
c.	Clivage de l'étiquette par la protéase TEV	68
d.	Purification d'ULT1 FL et SAND sur colonne d'héparine	69
e.	Deuxième chromatographie d'exclusion de taille	69
5.	Purification d'ULT1 C-ter	70
a.	Chromatographie d'affinité sur résine d'amylose	70
b.	Chromatographie d'affinité sur résine de Nickel	70
c.	Chromatographie d'exclusion de taille	70
6.	Identification des régions structurés d'ULT1 par protéolyse limitée par la trypsine	71
<b>B.</b>	<b>Production du complexe PRC2</b>	<b>71</b>
1.	Principe de l'expression de protéines en cellules d'insecte	71
2.	Clonage des différentes constructions de PRC2	73
a.	Clonage des gènes codants pour les sous-unités du complexe PRC2	73
3.	Recombinaison des vecteurs MultiBac par le système Cre-loxP	75
4.	Expression du PRC2 en cellules d'insecte	76
5.	Lyse des cellules d'insectes	77
6.	Purification de PRC2	77
a.	Chromatographie d'affinité sur résine de nickel	77
b.	Chromatographie d'affinité sur résine Strep-Tactin XT (IBA)	77
c.	Chromatographie d'exclusion de taille	78
<b>C.</b>	<b>Caractérisation biochimique et structurale d'ULT1</b>	<b>78</b>
1.	SEC-MALLS	78
2.	Essais de cristallisation des domaines N-Terminal et C-Terminal d'ULT1	79
3.	Phasage et affinement du modèle cristallin	80
4.	Caractérisation de l'interaction entre ULT1 C-Ter et le peptide HsH3.1 <sub>26-44</sub>	81
5.	Pull-down entre le domaine C-Terminal d'ULT1 et AtH3.1 <sub>26-44</sub> ou AtH3.3 <sub>26-44</sub>	82
6.	Mesure de l'interaction entre ULT1 <sub>128-237</sub> et le peptide HsH3.1 <sub>26-44</sub> par ITC	82
<b>D.</b>	<b>Caractérisation biochimique et structurale de l'interaction entre ULT1 et le complexe PRC2<sup>83</sup></b>	
1.	Pull-down sur résine Strep-Tactin XT	83
2.	Résonance Plasmonique de Surface	83
3.	Western-Blot	85
4.	Caractérisation structurale de PRC2-CLF et de l'interaction entre ULT1 et PRC2-SWN	85
a.	Reconstitution du complexe PRC2-SWN/ULT1	85
b.	Pontages chimiques des complexes PRC2-CLF et PRC2-SWN/ULT1 pour la cryo-EM	85
c.	Analyse du complexe PRC2-SWN/ULT1 par XL-MS	86
d.	Cryo-microscopie électronique (Cryo-EM)	87
e.	AlphaFold2	88
<b>III.</b>	<b>Résultats</b>	<b>90</b>
<b>A.</b>	<b>Étude biochimique et structurale de la protéine ULT1</b>	<b>91</b>
1.	Expression d'ULT1 : importance du Zinc	91
2.	Identification des domaines structurés d'ULT1 <sub>1-237</sub> par protéolyse limitée	95
3.	Purification du domaine C-terminal d'ULT1 (C-ter) : ULT1 <sub>128-237</sub>	97
4.	Analyse de l'état oligomérique d'ULT1 <sub>128-237</sub> en solution	98
5.	Purification du domaine N-terminal d'ULT1 (SAND) : ULT1 <sub>1-121</sub>	99
6.	Essai de cristallisation d'ULT1 <sub>128-237</sub> et ULT1 <sub>1-121</sub>	101
7.	Structure cristallographique du domaine C-Terminal d'ULT1	103
8.	ULT1 <sub>128-237</sub> pourrait se lier à l'histone 3 par l'intermédiaire de modifications sur K36	105
9.	AlphaFold dévoile la structure entière d'ULT1	108

<b>B.</b>	<b>Caractérisation biochimique et structurale de l'interaction entre ULT1 et le complexe PRC2</b>	<b>112</b>
1.	Purification du complexe PRC2-CLF _____	115
2.	Tests d'interaction entre PRC2-CLF et ULT1 _____	116
a.	Test d'interaction entre PRC2-CLF et ULT1 par filtration sur gel _____	118
b.	Test d'interaction entre PRC2-CLF et ULT1 par <i>pull-down</i> sur Résine Ni-NTA _____	118
c.	Test d'interaction par Résonance Plasmonique de Surface (SPR) _____	119
d.	Test d'interaction par <i>pull-down</i> sur résine Strep-Tactin XT _____	120
e.	Purification de la co-expression du PRC2-CLF et ULT1 en cellules d'insecte _____	121
3.	Purification du complexe PRC2-SWN _____	123
4.	Tests d'interaction entre PRC2-SWN et ULT1 _____	124
a.	Test d'interaction entre PRC2-SWN et ULT1 par <i>pull-down</i> _____	124
b.	Test d'interaction entre PRC2-SWN et ULT1 par filtration sur gel _____	124
c.	Test d'interaction entre PRC2-SWN et les parties N-Ter et C-ter d'ULT1 par <i>pull-down</i> _____	125
5.	Caractérisation des complexes PRC2-CLF et PRC2-SWN/ULT1 par Cryo-EM, AlphaFold et XL-MS __	127
a.	Modélisation par AlphaFold2 des complexes ternaires PRC2-CLF et PRC2-SWN _____	127
b.	Préparation d'échantillons du complexe PRC2-CLF pour la Cryo-EM _____	128
c.	Préparation d'échantillons du complexe PRC2-SWN/ULT1 pour la Cryo-EM _____	132
d.	Structure du complexe PRC2-SWN/ULT1 par Cryo-EM _____	134
e.	Pontage chimique couplé à la spectrométrie de masse (XL-MS) _____	135
<b>IV.</b>	<b><i>Discussion et Perspectives</i></b> _____	<b>138</b>
<b>A.</b>	<b>Etude biochimique et structurale de la protéine ULT1</b> _____	<b>139</b>
<b>B.</b>	<b>Caractérisation biochimique et structurale de l'interaction entre ULT1 et le PRC2</b> _____	<b>143</b>
<b>V.</b>	<b><i>Conclusion</i></b> _____	<b>148</b>
<b>VI.</b>	<b><i>Références</i></b> _____	<b>151</b>

# Liste des figures et des tableaux

## INTRODUCTION

**p18//Figure I.1** : Organisation multi-échelle et plasticité de la chromatine

**p27//Figure I.2** : Principaux acteurs du dépôt/retrait de groupements méthyl sur certaines lysines de la queue d'histone d'H3.1 chez Arabidopsis

**p28//Figure I.3** : Caractérisation d'éléments génomiques fonctionnels par des combinaisons spécifiques de marques épigénétiques

**p35//Figure I.4** : Structure et domaines du complexe PRC2 chez les mammifères

**p38//Figure I.5** : Structure dimérique du complexe PRC2 formée par un G-quadruplexe (G4) d'ARN

**p41//Figure I.6** : Diversité des complexes PRC2 canoniques chez Arabidopsis

**p43//Figure I.7** : Recrutement du complexe PRC2 vers ses cibles chez Arabidopsis

**p45//Figure I.8** : Interconnexions entre les complexes PRC1 et PRC2 chez Arabidopsis

**p48//Figure I.9** : Coopération entre le PRC2 et des effecteurs chromatiniens impliqués dans la répression des gènes

**p51//Figure I.10** : Cycle du développement chez Arabidopsis

**p57//Figure I.11** : Effets phénotypiques de la perte de fonction d'ULT1

**p60//Figure I.12** : ULT1 agit comme un antagoniste du PRC2

**p63//Figure I.13** : Schéma récapitulatif des objectifs de la thèse

**p33//Tableau I.1** : Sous-unités conservées du PRC2 chez la Drosophile, chez l'Homme, la souris et chez Arabidopsis

**p43//Tableau I.2** : Sous-unités du complexe PRC1 chez la Drosophile, chez l'Homme, et chez Arabidopsis

## MATERIELS ET METHODES

**p76//Figure II.1** : Stratégie d'assemblage des vecteurs d'expression des différentes sous-unités du PRC2-CLF par la méthode Multibac

**p80//Figure II.2** : Représentation de la carte de densité électronique 2Fo-Fc couvrant une partie du domaine C-terminal d'ULT1 et délimitée à  $1,2 \sigma$

**p85//Figure II.3** : Illustration schématique d'une expérience de SPR pour la mesure d'interaction entre deux molécules

**p86//Figure II.4** : Schéma de la molécule de BS3 (ThermoFischer)

**p66//Tableau II.1** : Liste des constructions de ULT1 utilisées dans cette thèse

**p74//Tableau II.2** : Liste des constructions PRC2 utilisées dans cette thèse

**p81//Tableau II.3** : Statistiques de collecte et d'affinement des données

## RESULTATS

**p92//Figure III.1** : Effet du zinc sur l'expression d'ULT1

**p94//Figure III.2** : Purification d'ULT1

**p96//Figure III.3** : Recherche des domaines structurés de MBP-ULT1 et ULT1<sub>1-127</sub> par protéolyse limitée et spectrométrie de masse

**p98//Figure III.4** : Purification d'ULT1<sub>128-237</sub>

**p99//Figure III.5** : Analyse par SEC-MALLS du domaine C-Terminal d'ULT1

**p100//Figure III.6** : Purification d'ULT1<sub>1-121</sub>

**p101//Figure III.7** : Alignement de multiples séquences d'ULT1 chez différentes espèces et d'ULT2 chez Arabidopsis

**p102//Figure III.8** : Cristallisation d'ULT1<sub>128-237</sub>

**p104//Figure III.9** : Structure cristallographique du domaine C-Terminal d'ULT1

**p107//Figure III.10** : ULT1<sub>128-237</sub> pourrait se lier à l'histone 3 par l'intermédiaire de modifications sur K36

**p109//Figure III.11** : Modèle structural d'ULT1 prédit par AlphaFold2

**p110//Figure III.12** : Comparaison des structures des domaines SAND de GMBE1 et ULT1

**p112//Figure III.13** : Hypothèses sur le mode de fonctionnement d'ULT1 lors d'une interaction directe avec le PRC2

**p114//Figure III.14** : Stratégie d'expression du lobe catalytique des complexes PRC2-CLF et PRC2-SWN

**p116//Figure III.15** : Purification du complexe PRC2-CLF ternaire

**p117//Figure III.16** : Test d'interaction entre PRC2 et MBP-ULT1 par filtration sur gel

**p119//Figure III.17** : Test d'interaction entre PRC2-CLF et ULT1 par Pull-Down sur résine Ni-NTA

**p120//Figure III.18** : La Résonance Plasmonique de Surface indique une interaction d'ULT1 avec le nickel

**p121//Figure III.19** : Test d'interaction entre PRC2-CLF et ULT1 par pull-down sur résine Strep-Tactin XT

**p122//Figure III.20** : Purification de la co-expression entre PRC2-CLF et ULT1 sur résine Strep-Tactin XT

**p123//Figure III.21** : Purification du complexe PRC2-SWN ternaire

**p126//Figure III.22** : ULT1 interagit avec PRC2-SWN *in vitro*

**p129//Figure III.23** : Structures de PRC2-CLF et PRC2 SWN modélisées par AlphaFold2

**p130//Figure III.24** : Caractérisation structurale de PRC2-CLF par Cryo-EM

**p131//Figure III.25** : Carte de densité 3D de PRC2-CLF obtenue par Cryo-EM

**p133//Figure III.26** : Préparation et optimisation d'échantillons pour la caractérisation structurale du complexe PRC2-SWN/ULT1

**p134//Figure III.27** : Caractérisation structurale de PRC2-SWN-ULT1 par Cryo-EM

**p136//Figure III.28** : Analyse structurale du complexe PRC2-SWN/ULT1 réticulé, par spectrométrie de masse

## Résumé

Le développement des organismes multicellulaires tels que la plante modèle *Arabidopsis thaliana* dépend de facteurs épigénétiques qui contrôlent la dynamique de la chromatine et régulent la transcription de gènes spécifiques en la réprimant ou l'activant. Parmi les acteurs impliqués dans ces régulations, deux groupes de protéines fonctionnent de manière antagoniste : le groupe Polycomb (PcG) et le groupe Trithorax (TrxG).

Le *Polycomb Repressive Complex 2* (PRC2) joue un rôle essentiel dans le maintien de milliers de gènes à l'état réprimé, par la déposition de marques de méthylation sur la lysine 27 de l'histone 3 (H3K27me3). A l'inverse, les protéines du groupe Trithorax contrecarrent spécifiquement la répression médiée par le groupe Polycomb, sur des cibles géniques communes. Alors que les protéines des groupes PcG et TrxG sont de mieux en mieux identifiées et caractérisées, les mécanismes moléculaires qui permettent de commuter un gène d'un état réprimé par PcG vers un état activé par TrxG sont encore peu connus. Chez *Arabidopsis*, la protéine ULTRAPETALA1 (ULT1), fonctionne comme un antagoniste du complexe PRC2, participant à la commutation de gènes cibles de PRC2 vers un état actif. De plus, des analyses d'interaction *in vivo* (Y2H, BiFC et IP-MS) ont révélé qu'ULT1 interagit avec les méthyltransférases CLF ou SWN des complexes PRC2. Ces observations indiquent qu'ULT1 pourrait directement perturber la fonction de ces complexes. Cependant, l'absence de donnée biochimique et structurale sur ces protéines rend l'étude de leur fonction difficile.

Le projet de thèse vise, par une approche de biologie structurale et de biochimie, à mieux comprendre les mécanismes de leur fonction antagoniste. Pour ceci, j'ai caractérisé *in vitro* la protéine ULT1 et le complexe PRC2 d'*Arabidopsis*, dans le but de mettre en évidence les détails moléculaires de leur interaction.

Dans une première partie concernant la caractérisation biochimique de la protéine ULT1, j'ai pu, après mise au point de ses conditions d'expression et de purification, y identifier deux domaines structurés. La résolution du domaine C-Terminal d'ULT1 par cristallographie aux rayons X a révélé un domaine CW atypique qui pourrait reconnaître des modifications sur H3K36.

Dans une deuxième partie concernant la caractérisation *in vitro* de l'interaction entre ULT1 et les complexes PRC2, j'ai pu révéler une interaction physique entre ULT1 et le complexe contenant la méthyltransférase SWN (complexe PRC2-SWN). J'ai également

démontré que cette interaction était stable et que la liaison d'ULT1 ne dissociait pas le complexe PRC2-SWN. Enfin, une étude par Cryo-EM du complexe composé d'ULT1 et PRC2-SWN n'a pas révélé de densité correspondant à ULT1, indiquant qu'elle pourrait se lier à une région flexible du complexe PRC2-SWN.

En conclusion, cette thèse a permis de développer des outils et méthodes essentiels à l'étude *in vitro* d'ULT1 et des complexes PRC2 chez *Arabidopsis*, et apporte également de nouvelles pistes de réflexion concernant la fonction d'ULT1 et son antagonisme avec le complexe PRC2-SWN. Ces résultats nous permettent d'orienter nos recherches pour une caractérisation plus précise des mécanismes de commutation de la chromatine d'un état réprimé à actif.

## Summary

The development of multicellular organisms such as the model plant *Arabidopsis thaliana* relies on epigenetic factors that govern chromatin dynamics and regulate the transcription of specific genes by either repressing or activating them. Two groups of proteins function antagonistically in these regulations: the Polycomb group (PcG) and the Trithorax group (TrxG).

The *Polycomb Repressive Complex 2* (PRC2) plays a crucial role in maintaining thousands of genes in a repressed state by depositing methylation marks on lysine 27 of histone 3 (H3K27me3). Conversely, proteins from the TrxG specifically counteract repression mediated by the PcG group, particularly on common gene targets. Although the PcG and TrxG proteins are progressively being identified and characterized, the molecular mechanisms that govern the transition of a gene from a PcG-repressed state to a TrxG-activated state remain poorly understood. In *Arabidopsis*, the ULTRAPETALA1 (ULT1) protein acts as an antagonist to the PRC2 complex, contributing to the conversion of an inactive state of PRC2 target genes into active state. Furthermore, *in vivo* interaction analyses (Y2H, BiFC, and IP-MS) have revealed that ULT1 interacts with the CLF and SWN methyltransferases of the PRC2 complexes. These observations indicate that ULT1 might directly disrupt the function of these complexes. However, the lack of biochemical and structural data on these proteins makes the study of their function challenging.

This PhD project aims to gain a better understanding of ULT1 and PRC2 antagonistic function using a structural biology and biochemistry approach. To achieve this, I characterized the ULT1 protein and the PRC2 complex from *Arabidopsis in vitro*, uncovering the molecular details of their interaction.

In the first section, dedicated to the biochemical characterization of the ULT1 protein, I successfully identified two structured domains after optimizing their expression and purification conditions. The resolution of the C-terminal domain of ULT1 by X-ray crystallography revealed an atypical CW domain that may recognize modifications on H3K36.

In the second part dedicated to the *in vitro* characterization of the interaction between ULT1 and PRC2 complexes, I demonstrated a physical interaction between ULT1 and the complex containing the methyltransferase SWN (PRC2-SWN complex). I also showed that this interaction was stable and that ULT1 binding did not dissociate the PRC2-SWN complex. Finally, a Cryo-EM study of the ULT1 and PRC2-SWN complex did not reveal

density corresponding to ULT1, indicating that it might bind to a flexible region of the PRC2-SWN complex.

In conclusion, this PhD work enabled the development of essential tools and methods for *in vitro* study of ULT1 and PRC2 complexes in Arabidopsis, while also bringing new insights into the function of ULT1 and its antagonism with the PRC2-SWN complex. These findings direct our research toward a more precise characterization of the mechanisms that govern transition of chromatin from a repressed to an active state.

## Liste des abréviations

AG : AGAMOUS	HMT : histone méthyltransférases
AIPP3 : ANTISILENCING 1 (ASI1)- IMMUNOPRECIPITATED PROTEIN 3	ISWI : Imitation Switch
AP2 : APETALA2	CHD : Chromodomain Helicase DNA- binding
AP3: APELATA3	Chip-Seq : <i>Chromatin immunoprecipitation</i> – <i>sequencing</i>
ARN Pol II : ARN Polymerase II	COMPASS-Like : COMpLex Proteins ASsociated with SET1
ASH2R : ARABIDOPSIS ASH2 RELATIVE	Cryo-EM : Cryo-microscopie électronique
ASI1 : ANTISILENCING 1	CS : <i>Chromatin State</i>
ATAC-seq : <i>Assay for Transposase- Accessible Chromatin using sequencing</i>	DNase-seq : DNase I hypersensitive sites – sequencing
ATP : Adénosine Triphosphate <sup>7</sup>	EAR : Ethylene-responsive element binding factor-associated Amphiphilic Repression
ATXR2/3/7 : ATX-RELATED 2 , 3 ou 7	EBD : EED BINDING DOMAIN
ATX1-5 : ARABIDOPSIS TRITHORAX 1 à 5	EBS : EARLY BOLTING IN SHORT DAYS
AZF1 : ARABIDOPSIS ZINC-FINGER PROTEIN 1	EED : EMBRYONIC ECTODERM DEVELOPMENT
BAM : Motif de feuilletés $\beta$	EMF1/2 : EMBRYONIC FLOWER 1/2
BAH : Bromo-adjacent homology	EMSA : <i>Electro-Mobility Shift Assay</i>
BBX : B-box	EZH2 : ENHANCER OF ZESTE 2
BiFC : <i>Bimolecular Fluorescence Complementation</i>	FIE : FERTILIZATION INDEPENDENT ENDOSPERM
BPC : BASIC PENTACYSTEINE	FIS2 : FERTILIZATION INDEPENDENT SEED 2
BRM : BRAHMA	FLC : FLOWERING LOCUS C
CAF-1 : CHROMATIN ASSEMBLY FACTOR 1	FLD : FLOWERING LOCUS D
CBP : cAMP-Responsive Element Binding Protein (CREB)-Binding Protein	FT : Facteur de transcription
CLF : CURLY LEAF	FT : FLOWERING LOCUS T
CQ : Centre Quiescent	GNAT : General Control Nondepressible 5- related Acetyltransferase
CREB : cAMP-Responsive Element Binding Protein	HAT : histones acetyltransferases
FLC : FLOWERING LOCUS C	HDAC : histones deacetylases
Pb : Paires de Bases	HD2 : histone deacetylase 2
PTM : <i>Post Translational Modifications</i>	HMT : histone methyltransferases
SWI/SNF : Switch defective/Sucrose Non- Fermentable	

HP1 : HETEROCHROMATIN PROTEIN 1  
 INO80 : Inositol requiring 80  
 IUC11: INCURVATA11  
 ITC : Titration Calorimétrique Isotherme  
 JMJ : JUMONJI  
 KAN1/2 : KANADI1/2  
 LDL : Lysine Specific Demethylase Like  
 LFY: LEAFY  
 LHP1 : LIKE- HETEROCHROMATIN PROTEIN 1  
 lncRNA : Long non-coding RNA  
 LSD : Lysine-specific demethylase  
 MAC : Méristème apical caulinaire  
 MADS : MCM1, AGAMOUS, DEFICIENS, SRF  
 MAR : Méristème apical racinaire  
 MBD : Methyl Binding Domain  
 MBP : Maltose Binding Protein  
 MBT : Malignant Brain Tumor  
 MEA : MEDEA  
 MF : Méristème floraux  
 MI : Méristèmes d'inflorescences  
 MINU1 : MINUSCULE1  
 MINU2 : MINUSCULE2  
 MI : Méristème d'inflorescence  
 MMD1 : MALE MEIOCYTE DEATH1  
 MSI1 : MULTICOPY SUPPRESSOR OF IRA 1  
 MYST : MOZ-YBF2/SAS3-SAS2/TIP60  
 PcG :Protéines du groupe Polycomb  
 PHD : Plant Homeobox Domain  
 PI : PISTILATA  
 PIC : Pré-initiation de la transcription  
 PRC1/2 : Polycomb Repressive Complex 1 et 2  
 PRE : Polycomb Response Element  
 RBL : RBBP5 LIKE  
 RBBP : Retinoblastoma Binding Protein  
 RPD3/HDA1 : Potassium Dependency 3/histone deacetylase 1  
 RPKM : *Read Per Kilobase per Million reads mapped*  
 SAL : boucle d'activation du domaine SET  
 SAM : S-Adenosyl-Méthionine  
 SAND : Sp100, AIRE, NucP41/75, DEAF-1  
 SANT : Swi3, Ada2, N-CoR, TFIIB  
 SAP18 : STRESS ASSOCIATED PROTEIN 18  
 SEP3 : SEPALLATA3  
 SBD : Domaine de liaison de SANT1  
 SEP3 : SEPALATA3  
 SET : Su(var) 3e9, Enhancer-of-zeste, and Trithorax  
 SHL : SHORT LIFE  
 SIR2 : Silent Information Regulator 2)  
 SOC1: SUPPRESSOR OF CONSTANS OVEREXPRESSION  
 SPR : Résonance Plasmonique de Surface  
 SPT : SPATULA  
 SRA : SET and RING finger-associated  
 SRM : Motif de réponse à la stimulation  
 STM : SHOOT MERISTEMLESS  
 SUVH : SU(VAR)3-9 HOMOLOGS  
 SUVR : SU(VAR)3-9-RELATED  
 SUZ12 : Suppressor of Zeste 12  
 SAH : S-adénosylhomocystéine  
 SAM : S-adénosylméthionine  
 SWR : SWI2/SNF2-related 1  
 SWN : SWINGER  
 SYD: SPLAYED  
 TAD : Domaine de transactivation  
 TAFII250 : TATA-binding protein Associated Factor 1

TBP : TATA-binding protein  
TE : Elément transposable  
TPL : TOPLESS  
TRB : TELOMERE-REPEAT BINDING FACTOR  
TrxG : Trithorax  
TSS : Site de démarrage/initiation de la transcription  
TES : Site de terminaison de la transcription  
UIF1 : ULTRAPETALA INTERACTING FACTOR 1  
ULT1/2: ULTRAPETALA1/2  
VAL : VIVIPAROUS1/ABI3-LIKE  
VEFS : VRN2 EMF2 FIS2 et SUZ12  
VRN : VERNALIZATION  
WDR5a : HUMAN WDR5 HOMOLOG A  
WT : phénotype sauvage (wild-type)  
WUS: WUSCHEL

# I. Introduction

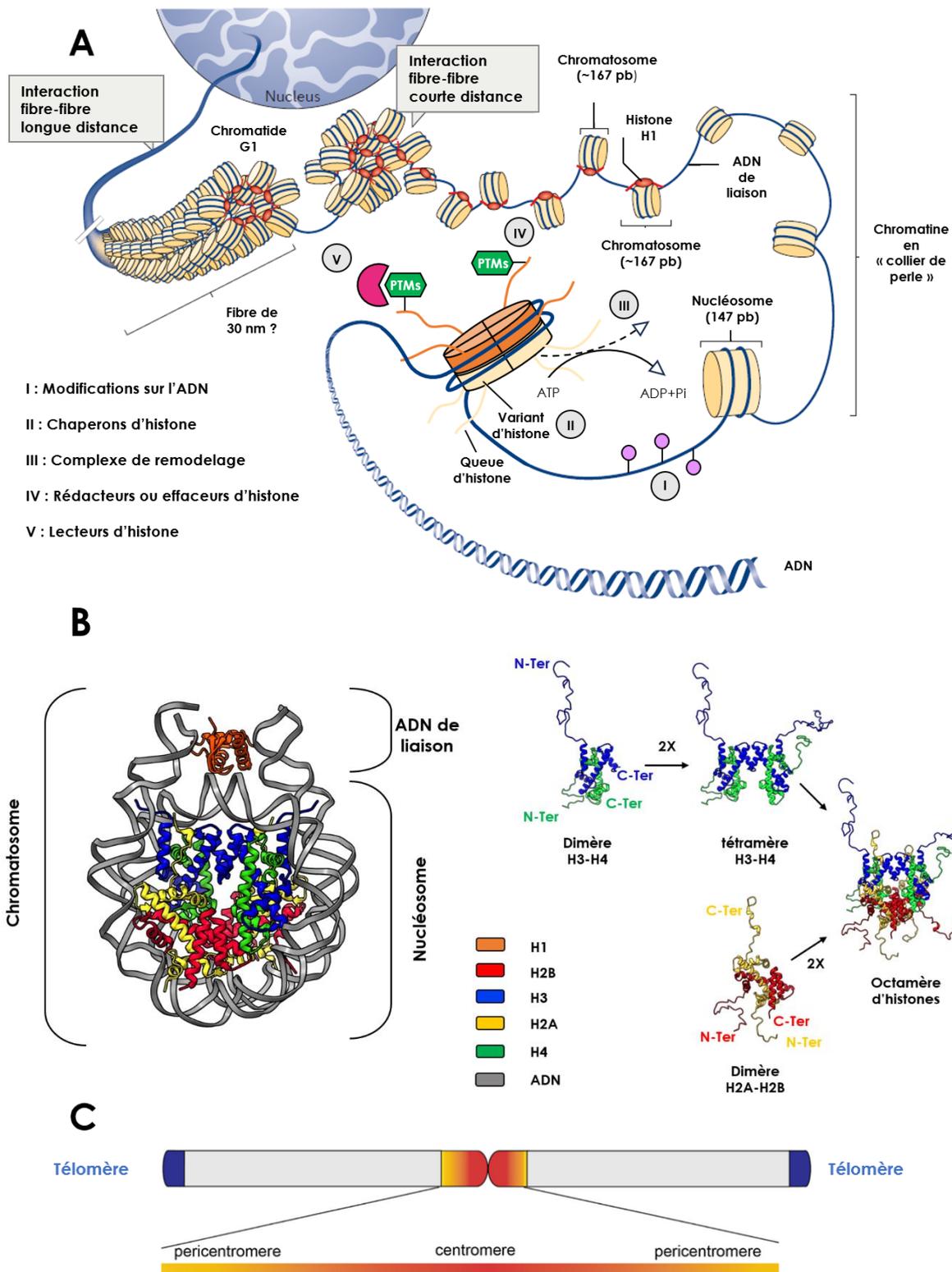
# A. Structure et dynamique de la chromatine

## 1. Structure et fonction de la chromatine

Chaque cellule d'un organisme renferme dans son noyau toutes les informations génétiques nécessaires à son développement et à son fonctionnement. Ces informations sont stockées sous la forme de l'ADN, qui est identique pour toutes les cellules d'un même individu. Cependant, ces informations génétiques ne sont pas exprimées simultanément et lorsque certains gènes sont transcrits, c'est-à-dire actifs, d'autres ne le sont pas et sont alors réprimés. C'est ce programme d'activation et de répression de gènes spécifiques qui va conduire les cellules à acquérir une identité et une fonction propre. Ce processus est très dynamique et varie en fonction de stimuli internes ou externes pour répondre aux changements dans les besoins de l'organisme ou dans leur environnement.

L'expression des gènes est directement liée à des variations dans l'accessibilité des informations stockées dans l'ADN. En effet, bien que ce dernier s'associe avec des protéines pour former une structure compacte que l'on appelle chromatine, certaines régions sont rendues plus ou moins accessibles afin de réguler l'accès à la machinerie de transcription à une position précise d'un gène (locus).

On peut distinguer plusieurs niveaux d'enroulement de l'ADN, correspondant à plusieurs degrés de compaction de la chromatine : Le premier degré de compaction correspond au nucléosome qui représente l'unité de base de la chromatine (**Figure I.1A**). Il est formé d'un octamère de protéines histones caractérisé par un assemblage de deux dimères des histones H2A-H2B et de deux dimères des histones H3-H4 autour duquel est enroulé 147 paires de bases (pb) d'ADN (Luger *et al.*, 1997) (**Figure I.1B**). Une autre protéine, l'histone H1, peut également s'ajouter au nucléosome au niveau de la jonction entre les brins entrant et sortant de l'ADN, et joue un rôle supplémentaire dans l'organisation de la chromatine en régulant sa compaction à des degrés supérieurs (Song *et al.*, 2014; Bednar *et al.*, 2017) (**Figure I.1A et B**). La succession de nucléosomes forme une deuxième structure de la chromatine dite en « collier de perle » d'environ 10 nm de diamètre (Woodcock, Safer and Stanchfield, 1976). Cet arrangement est ensuite condensé en une fibre qui va se replier sur elle-même pour atteindre une structure avec un niveau de compacité maximal lors de la métaphase et que l'on appelle chromosome mitotique (Luger, Dechassa and Tremethick, 2012; Antonin and Neumann, 2016; Fyodorov *et al.*, 2018; Zhou and Bai, 2019).



**Figure 1.1: Organisation multi-échelle et plasticité de la chromatine. (A)** Organisation hiérarchique de la chromatine dans un noyau en interphase. L'ADN est enroulé autour d'un octamère d'histone pour former l'unité de base de la chromatine : le nucléosome. La succession de nucléosomes forme ensuite une structure de la chromatine dite en « collier de perle » d'environ 10 nm de diamètre capable de se condenser en une fibre par l'intermédiaire de l'histone H1. Cette fibre peut ensuite se replier sur elle-même pour atteindre des structures avec un niveau de compacité plus élevé. La chromatine sous la forme d'une fibre de 30 nm n'a pas été observée *in vivo* et son existence reste discutable. Le nucléosome est un module qui peut varier selon la combinaison de plusieurs mécanismes tel que

l'assemblage/désassemblage du nucléosome par des complexes de remodelage, l'incorporation ou l'éviction de variants d'histones, mais aussi l'ajout et la lecture de modifications post-traductionnelles (PTMs pour Post Translational Modifications) sur les histones et sur l'ADN. Tous ces mécanismes contribuent à la plasticité de la chromatine. **(B)** Structure cristallographique du chromatosome contenant le domaine globulaire de la protéine H5 (H1.0 ; orange) chez le poulet et des protéines histones H2A (jaune), H2B (rouge), H3 (bleu) et H4 (vert). PDB : 4QLC. Le nucléosome comprends un double-brin d'ADN enroulé autour d'un octamère d'histone composée d'un tétramère H3-H4 ainsi que deux dimères H2A-H2B. **(C)** Représentation schématique d'un chromosome d'*Arabidopsis thaliana*. L'agrandissement de la partie centrale révèle la région centromérique (rouge) et péri-centromérique (jaune). Les régions télomériques aux extrémités du chromosome sont représentées en bleu. Figure adaptée de Simon et al., 2015; Fyodorov et al., 2018; Yadav, Quivy and Almouzni, 2018.

Dans le noyau, lors de l'interphase, la chromatine n'est pas structurée uniformément, et deux états peuvent être observés avec des degrés de compaction variables : l'euchromatine et l'hétérochromatine (Heitz, 1928). L'euchromatine est associée à une faible densité de nucléosomes et à un faible niveau de condensation de l'ADN, ce qui la rend plus permissive à la liaison des facteurs de transcription (FT) et à la machinerie de transcription (Klemm, Shipony and Greenleaf, 2019) . A l'inverse, l'hétérochromatine représente le plus fort niveau de compaction de la chromatine et est corrélée à une faible activité de transcription (Rosa and Shaw, 2013; Morrison and Thakur, 2021). L'hétérochromatine dite constitutive est l'état le plus condensé de la chromatine. Elle contient des séquences répétées d'ADN ainsi que des éléments transposables et se trouve à la fois aux niveaux des télomères et dans les régions centromériques et péri-centromériques (Law and Jacobsen, 2010; Simon *et al.*, 2015) (**Figure 1.1C**). L'hétérochromatine dite facultative correspond à des régions qui peuvent alterner de manière réversible entre un état de répression de l'activité de transcription et un état actif (Trojer and Reinberg, 2007).

Le changement d'état de la chromatine est un processus dynamique et finement régulé qui implique de nombreux mécanismes. Ces mécanismes incluent l'assemblage/désassemblage du nucléosome par des complexes de remodelage, l'incorporation ou l'éviction de variants d'histones, mais aussi l'ajout de modifications post-traductionnelles (PTMs pour *Post Translational Modifications*) sur les histones et sur l'ADN (Bannister and Kouzarides, 2011; Allis and Jenuwein, 2016; Clapier *et al.*, 2017; Reyes, Marcum and He, 2021; Foroozani, Holder and Deal, 2022) (**Figure 1.1A**). Tous ces éléments vont contribuer à contrôler l'expression des gènes sans altérer la séquence d'ADN.

Ces PTMs peuvent être transmises lors de la mitose mais sont également réversibles. Elles sont régies par une série d'effecteurs tel que des « rédacteurs » qui les déposent, des « lecteurs » qui les interprètent et d' « effaceurs » qui les suppriment. L'étude de ces modifications et de ces mécanismes qui peuvent perpétuer des états alternatifs d'activité

génique dans un contexte d'ADN unique est appelée épigénétique (Cavalli and Heard, 2019).

## 2. Modification de la chromatine

### a. Variants d'histones

L'incorporation dynamique de variants d'histones dans les nucléosomes est un mécanisme important dans la régulation de la structure de la chromatine et de l'accessibilité de l'ADN pour la transcription. Ces variants sont des isoformes protéiques qui diffèrent de quelques résidus dans leur séquence d'acides aminés, ce qui a pour conséquence de changer les propriétés des nucléosomes comme par exemple modifier leurs interactions avec les autres effecteurs chromatinien. A la différence des histones canoniques qui sont incorporées lors de la phase répliative de l'ADN, les variants d'histones peuvent être déposés tout au long du cycle cellulaire (Weber and Henikoff, 2014; Jiang and Berger, 2017; Talbert and Henikoff, 2017).

Chez les plantes, l'histone H3 existe sous trois isoformes principaux : H3.1, H3.3 et CENH3 (Jiang and Berger, 2017; Probst, 2022). Le variant CENH3 est localisé au niveau de la région centromérique où il est crucial pour l'assemblage des kinétochores et la séparation des chromosomes (Lermontova *et al.*, 2011, 2013; McKinley and Cheeseman, 2016). Le variant H3.3 quant à lui, se localise dans des régions enrichies en marque H3K4me3 et corrèle avec l'expression des gènes tandis que l'histone canonique H3.1 est enrichie au niveau de l'hétérochromatine et des gènes/régions réprimés (Ingouff and Berger, 2010; Stroud *et al.*, 2012; Wollmann *et al.*, 2012; Shu *et al.*, 2014; Sequeira-Mendes and Gutierrez, 2016). Récemment, il a été montré que la phénylalanine 41 d'H3.1 est essentielle pour la localisation de ce variant dans les régions silencieuses du génome (Lu *et al.*, 2018).

Trois variants de l'histone H2A sont codés par le génome de la plante modèle *Arabidopsis thaliana* (appelé par la suite : Arabidopsis) : H2A.Z, H2A.X et H2A.W (Talbert *et al.*, 2012; Yelagandula *et al.*, 2014). Le variant H2A.X est localisé dans le corps des gènes fortement exprimés et possède également un rôle essentiel dans la réponse aux dommages de l'ADN (Yelagandula *et al.*, 2014; Lorković *et al.*, 2017). A l'inverse, le variant H2A.W est peu présent dans l'euchromatine et semble participer à la condensation de chromatine à travers son motif SPKK situé dans sa partie C-terminale (Yelagandula *et al.*, 2014). Enfin, le variant H2A.Z peut être à la fois associé aux gènes actifs où il est fortement

enrichi au niveau du site de démarrage de la transcription (TSS), et au corps des gènes réprimés (Coleman-Derr and Zilberman, 2012; Carter *et al.*, 2018). Dans ce dernier cas, la mono-ubiquitylation d'H2A.Z joue un rôle dans la répression de la transcription (Gómez-Zambrano, Merini and Calonje, 2019)

## **b. Complexe de remodelage de la chromatine dépendant de l'ATP**

De larges complexes protéiques, appelés aussi remodeleurs, sont capables d'altérer la structure de la chromatine. Ils utilisent l'énergie de l'hydrolyse de l'ATP (Adénosine Triphosphate) afin de déplacer ou d'évincer les nucléosomes et ainsi libérer l'accès aux protéines de liaison de l'ADN telles que les facteurs de transcription. Les complexes de remodelage peuvent être classés en plusieurs groupes dont les principaux sont : Switch defective/Sucrose Non-Fermentable (SWI/SNF), Imitation Switch (ISWI), Chromodomain Helicase DNA-binding (CHD) et Inositol requiring 80 (INO80) / SWI2/SNF2-related 1 (SWR) (Han *et al.*, 2015).

Chez *Arabidopsis*, la famille SWI/SNF inclue les enzymes BRAHMA (BRM), SPLAYED (SYD), MINUSCULE1 (MINU1) et MINUSCULE2 (MINU2). Elles catalysent le glissement des nucléosomes le long de l'ADN et/ou leur retrait et sont capables de s'associer à de nombreux régulateurs transcriptionnels comme par exemple l'histone désacétylase HDA6 et l'histone déméthylase d'H3K27me3, REF6 (Li *et al.*, 2016; Yang *et al.*, 2020).

CHROMATIN REMODELLING 11 et 17 (CHR11/17) appartiennent à la famille ISWI. Ces enzymes contiennent un domaine HAND-SANT-SLIDE dans leur région C-Terminale, qui permet de réguler l'espacement entre deux nucléosomes adjacents (Clapier *et al.*, 2017).

Les remodeleurs de la famille CHD quant à eux, interviennent dans l'assemblage du nucléosome, la régulation de l'accès à la chromatine et l'incorporation du variant d'histone H3.3 (Konev *et al.*, 2007).

Enfin, la famille INO80/SWR1 est également capable de modifier la composition du nucléosome par l'incorporation ou l'éviction de variants d'histones. C'est le cas du variant d'histone H2A.Z qui est incorporé par le complexe SWR1 mais dont le retrait serait réalisé par le complexe INO80 (Mizuguchi *et al.*, 2004; Papamichos-Chronakis *et al.*, 2011; Luo *et al.*, 2020; Xue *et al.*, 2021).

## **c. Méthylation de l'ADN**

La méthylation de l'ADN participe à la régulation de l'expression des gènes et à la répression des éléments transposables (appelés également transposons). Chez les plantes,

elle est catalysée sur le carbone n°5 des résidus cytosines (5-mC) selon trois différents contextes : CG, CHG et CHH (où H représente les autres bases de l'ADN : A, T ou C) (Zhang *et al.*, 2006a). La méthylation de novo de l'ADN se fait par une voie qui implique de petits ARNs interférents et la méthyltransférase DRM2 bien que le maintien de cette modification puisse s'effectuer, selon le contexte, par d'autres méthyltransférases comme MET1 (contexte CG) , CMT2 (contextes CHG et CHH) et CMT3 (contexte CHG) (Lindroth *et al.*, 2001; Kankel *et al.*, 2003; Zemach *et al.*, 2013; Liu *et al.*, 2014; Stroud *et al.*, 2014). Ces méthylation de l'ADN peuvent être reconnues par le domaine de liaison de l'ADN (MBD pour *Methyl Binding Domain*) et le domaine SET (Su(var) 3e9, Enhancer-of-zeste, and Trithorax) et associé aux doigts de zinc RING (SRA pour *SET and RING finger-associated*) (Grimanelli and Ingouff, 2020).

Chez *Arabidopsis*, l'hétérochromatine péri-centromérique et certaines régions d'euchromatine contenant des transposons ou des séquences répétées sont fortement méthylées et corrélatent avec leur extinction (Zhang *et al.*, 2006b; Cokus *et al.*, 2008). La méthylation de l'ADN est également impliquée dans la régulation de l'expression génique en étant localisée soit dans le promoteur, soit dans le corps du gène, mais leur fonction dans ce contexte reste encore peu comprise (Zhang, Lang and Zhu, 2018).

Concernant la déméthylation de l'ADN, elle s'effectue par excision de la base 5-mC. Chez les plantes elle s'effectue directement par l'intervention d'ADN glycosylases qui vont exciser la base méthylée, alors que chez les mammifères, elle est d'abord initiée par l'oxydation et/ou la désamination de cette base (Zhu, 2009; Wu and Zhang, 2010; Zhang and Zhu, 2012).

#### **d. Modification d'histones**

Les protéines histones sont structurées en deux parties : un domaine globulaire basique autour duquel va s'enrouler l'ADN et une région flexible en position N-Terminale et/ou C-terminale (**Figure I.1A**). Ces deux parties sont soumises à de nombreuses modifications post-traductionnelles telles que la méthylation, la phosphorylation, l'acétylation ou l'ubiquitination. L'ajout de ces modifications (appelées aussi « marques histones ») va changer les propriétés physico-chimiques des histones qui peuvent influencer directement la structure de la chromatine, recruter des protéines associées mais aussi moduler leur fonction. Selon leur position sur la protéine histone, ces modifications, seules ou en combinaison, peuvent induire différents états chromatiniens, corrélatant avec une activation ou répression du locus associé (Margueron, Trojer and Reinberg, 2005;

Kouzarides, 2007; Lee, Smith and Shilatifard, 2010; Bannister and Kouzarides, 2011; Tessarz and Kouzarides, 2014; Zhang *et al.*, 2021; Zhao, Allis and Wang, 2021).

## 1) Acétylation des histones

L'acétylation des histones, est une modification post-traductionnelle caractérisée par le dépôt d'un groupement acétyl ( $\text{CH}_3\text{COO}^-$ ) au niveau du groupement amine des lysines (K). L'ajout de cette modification a pour conséquence une perte de la charge positive des lysines et peut perturber les interactions entre les résidus modifiés et d'autres macromolécules (protéines ou acides nucléiques) (Strahl and Allis, 2000; Yang and Seto, 2007; Bannister and Kouzarides, 2011). L'acétylation des lysines peut donc entraîner un changement d'état de la chromatine où l'ADN est plus accessible, ce qui corrèle généralement avec la transcription des gènes (Shogren-Knaak *et al.*, 2006). Les lysines acétylées peuvent être reconnues par des Bromodomains ou des domaines PHD (Plant Homeobox Domain) et participer aux recrutements de protéines régulatrices (Dhalluin *et al.*, 1999; Zeng *et al.*, 2010). L'ajout de ces modifications peut s'opérer au niveau des lysines K4, K9, K14, K18, K27 et K36 de l'histone H3 et sur les lysines K5, K8, K12, K16 et K20 de l'histone H4 (Shahbazian and Grunstein, 2007). Chez Arabidopsis, elles sont catalysées par 12 enzymes de la famille des histones acétyltransférases (HAT) qui sont classées en quatre groupes : GNAT (General Control Nondepressible 5-related Acetyltransferase), MYST (MOZ-YBF2/SAS3-SAS2/TIP60), CBP (cAMP-Responsive Element Binding Protein (CREB)-Binding Protein) et TAFII250 (TATA-binding protein Associated Factor 1) (Pandey *et al.*, 2002). A l'inverse, le retrait du groupement acétyl est catalysé par des histones désacétylases (HDAC). Elles sont au nombre de 18 chez Arabidopsis et sont divisées en 3 groupes : RPD3/HDA1 (Reduced Potassium Dependency 3/histone deacetylase 1), SIR2 (Silent Information Regulator 2) et HD2 (histone deacetylase 2) (Pandey *et al.*, 2002). Les HAT et HDAC interviennent dans la régulation de nombreux processus essentiels pour la croissance et le développement mais aussi en réponse aux stimuli externes. Par exemple, l'association physique entre l'histone désacétylase HDA19, APETALA2 (AP2) et TOPLESS (TPL) a montré une répression des gènes impliqués dans l'organogenèse florale comme AGAMOUS (AG) (Krogan, Hogan and Long, 2012).

## 2) Méthylation des histones

La méthylation des histones opère au niveau du groupement amine des lysines (K) qui peut être mono-, di- ou triméthylé et au niveau du groupement guanidine des arginines (R) qui peut être mono- ou di- méthylé (Gary and Clarke, 1998; Zhang and Reinberg, 2001). Le degré de méthylation déposé sur les lysines et les arginines change l'encombrement stérique de ces résidus et augmente leur hydrophobicité mais ne modifie pas leur charge (Luo, 2018). Chez *Arabidopsis*, les principaux sites de méthylation des lysines sur l'histone 3 (H3) sont les résidus K4, K9, K23, K27 et K36 (K. Zhang *et al.*, 2007; Trejo-Arellano *et al.*, 2017) (**Figure I.2**). Certaines des méthylations sur ces acides aminés ont été associées à des gènes actifs comme H3K4me3 et H3K36me2/me3 ou réprimés pour les marques H3K9me3/H3K27me3 (Liu *et al.*, 2010; Xiao, Lee and Wagner, 2016). Ces méthylations sont catalysées par des enzymes appelées histone méthyltransférases (HMTs) qui appartiennent généralement à la famille à domaines catalytique SET (Su(var) 3e9, Enhancer-of-zeste, and Trithorax) (Ng *et al.*, 2007).

Plusieurs domaines protéiques peuvent reconnaître et lier ces méthylations. Les méthylations sur H3K4 sont reconnues par les domaines PHD (pour Plant Homéodomain), TUDOR, MBT (pour Malignant Brain Tumor) et CW. Les méthylations sur H3K9 peuvent être reconnues par les Chromodomaines, les domaines PHD, WD40 et TUDOR. Les méthylations sur H3K27 sont quant à eux reconnus par les Chromodomaines et domaines WD40. Enfin, les méthylations sur K36 peuvent être reconnues par les chromodomaines et les domaines PWWP (Yun *et al.*, 2011).

Le retrait de ces marques s'effectue par des histones déméthylases de la famille LSD (Lysine-specific demethylase) ou JMJ (JUMONJI) (Lu *et al.*, 2008; Foroozani, Vandal and Smith, 2021; Hu and Du, 2022).

### a) Méthylation sur H3K4 et H3K36

Les méthylations d'H3K4 sont majoritairement localisées dans les régions géniques avec une accumulation d'H3K4me2 et d'H3K4me3 dans le promoteur et la région 5' des gènes (Zhang *et al.*, 2009; Roudier *et al.*, 2011) (**Figure I.3**). Elles sont déposées par les protéines *ARABIDOPSIS TRITHORAX* 1 à 5 (ATX1-5) et *ATX-RELATED* 2, 3 ou 7 (ATXR2/3/7) (Ornelas-Ayala *et al.*, 2023) (**Figure I.2**). Des différences dans l'activité catalytique ont pu être observées parmi ces protéines notamment chez ATX1 qui triméthyle H3K4 et ATX2 qui le diméthyle (Saleh *et al.*, 2008). Les méthyltransférases d'H3K4

ont été retrouvées en association avec les protéines ASH2R (ARABIDOPSIS ASH2 RELATIVE), RBL (RBBP5 LIKE) et WDR5a (HUMAN WDR5 HOMOLOG A) pour former plusieurs complexes COMPASS-Like impliqués dans l'activité de transcription. C'est le cas du complexe ASH2R-ATX1 COMPASS-like qui participe à l'expression du gène *FLOWERING LOCUS C (FLC)* qui réprime la transition florale et pour lequel les mutants *atx1* et *ash2r* présentent une diminution de l'expression de *FLC* corrélée à une réduction de la marque H3K4me3 (Jiang *et al.*, 2011). Le complexe ATX1-COMPASS-Like est également connu pour participer au recrutement et/ou à la stabilisation de l'ARN Polymerase II (ARN Pol II) et de la TATA-binding protein (TBP), une protéine du complexe de pré-initiation de la transcription (PIC). Ce recrutement/stabilisation se fait de manière indépendante de l'activité catalytique puisqu'un mutant d'ATX1 dans son domaine SET n'impacte pas la présence de ces protéines (TBP et Pol II) dans la région promotrice. En revanche la perte de fonction méthyltransférase du mutant dans le domaine catalytique SET induit une diminution de l'élongation de la transcription. Ces résultats indiquent que le dépôt de la marque H3K4me3 par le complexe ATX1-COMPASS-Like n'est pas requise pour l'initiation de la transcription mais quelle est une marque qui favorise l'élongation (Ding, Avramova and Fromm, 2011; Ding *et al.*, 2012). De plus, d'autres études semblent également montrer que la marque H3K4me3 (mais aussi H3K27ac) est une conséquence de la transcription plutôt qu'une cause et pourrait donc être un bon marqueur de l'activité d'un gène (Martin *et al.*, 2021; Z. Wang *et al.*, 2022).

Concernant la déméthylation d'H3K4, elle est catalysée par les protéines de la famille LSD et du groupe 5 (KDM5) de la famille JMJ. Les enzymes *FLOWERING LOCUS D* et Lysine Specific Demethylase Like 1-3 (LDM1-3) catalysent le retrait de la mono- et diméthylation d'H3K4 tandis que JMJ14 à 18, catalysent en plus le retrait de la triméthylation (**Figure I.2**) (Shi *et al.*, 2004; Tsukada *et al.*, 2006). Les déméthylases de ces deux familles jouent un rôle important dans la régulation de la floraison comme FLD (*FLOWERING LOCUS D*) (Liu *et al.*, 2007) et JMJ18 (Yang *et al.*, 2010) qui interviennent dans la répression de *FLC* (*FLOWERING LOCUS C*) mais participent aussi dans la régulation d'autres processus comme la défense de la plante (JMJ14, (Li *et al.*, 2020)), le stress salin (JMJ15, (Shen *et al.*, 2014)), le stress hydrique (JMJ17, (Huang *et al.*, 2019)) et la sénescence (JMJ16, (P. Liu *et al.*, 2019)).

Les méthylations sur H3K36 sont majoritairement retrouvées sous la forme di- et triméthyles au niveau de la région 5' des gènes pour H3K36me3 et 3' pour H3K36me2 où elles sont associées à des gènes où la transcription est active (Xu *et al.*, 2008; B. Liu *et al.*, 2019; Leng *et al.*, 2020; Shim *et al.*, 2020) (**Figure I.3**). Chez Arabidopsis, la méthylation

spécifique sur H3K36 est principalement catalysée par la protéine SDG8 (ASHH2) bien que les méthyltransférases d'H3K4 : ATXR2/SDG36, ATXR7/SDG25, SDG26 et SDG4 en soient également capable (Xiao, Lee and Wagner, 2016; Ornelas-Ayala *et al.*, 2023) (**Figure I.2**). Il a été montré qu'SDG8 interagit avec le domaine C-terminal de l'ARN Pol II et participe à l'induction de la transcription de gènes *PR1* et *PR2* (PATHOGENESIS RELATED GENE 1/2) en réponses aux pathogènes bactériens (X. Zhang *et al.*, 2020) ce qui indique que l'activité de transcription est également associée aux méthylations sur H3K36. A la différence des animaux, les déméthylases d'H3K36 ne sont pour l'instant pas connues bien que cette fonction ait été proposée pour la protéine INCURVATA11 (ICU11) (Bloomer *et al.*, 2020) (**Figure I.2**).

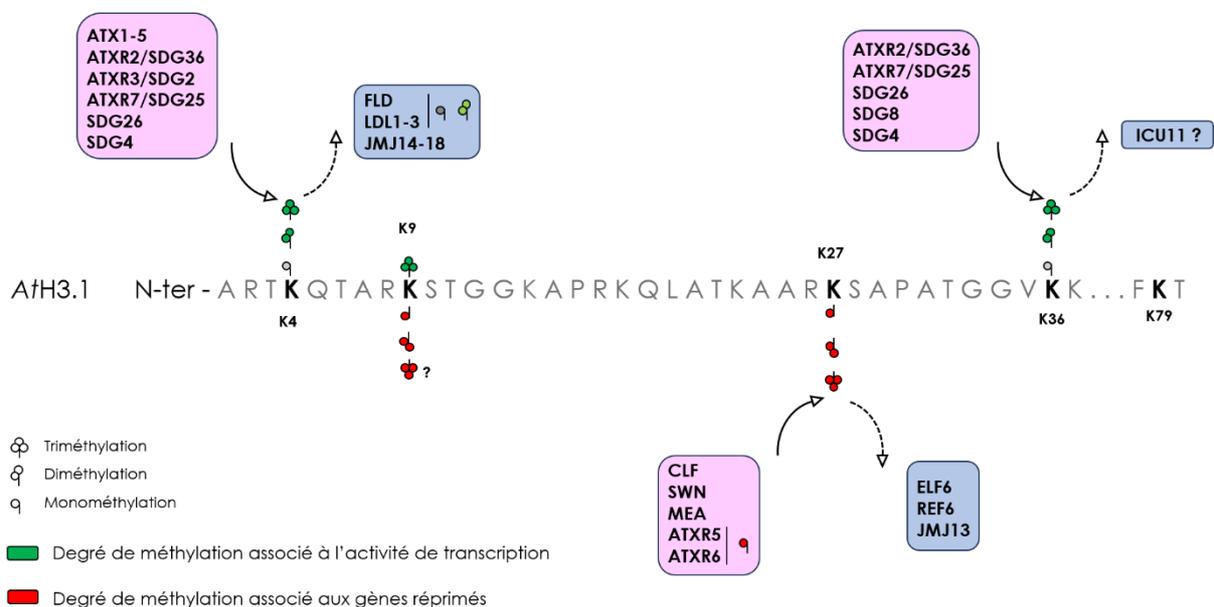
### b) Méthylation sur H3K9 et H3K27

Chez *Arabidopsis*, les marques H3K9me1/2 sont principalement enrichies au niveau de l'hétérochromatine constitutive et sont associées à la répression des gènes (**Figure I.3**) (Zhou *et al.*, 2010). A l'inverse, la marque H3K9me3 est retrouvée au niveau de l'euchromatine et corrèle avec la transcription (Roudier *et al.*, 2011). Cependant, de faibles niveaux d'H3K9me3 ont également été détectés au niveau d'éléments transposables et des séquences répétées, ce qui pourrait indiquer un rôle répresseur spécifique de cette marque sur ces éléments (Veiseth *et al.*, 2011; Sequeira-Mendes and Gutierrez, 2016). Les protéines SU(VAR)3-9 HOMOLOGS 1 à 10 (SUVH1-10) et SU(VAR)3-9-RELATED 1 à 5 (SUVR1-5) ont été identifiées pour catalyser la méthylation sur H3K9. La déméthylation de cette marque est quant à elle catalysée par les protéines JMJ25-29 (Lu *et al.*, 2008; Dutta *et al.*, 2017; Cheng *et al.*, 2022). Certaines enzymes peuvent changer de substrat lorsqu'elles sont modulées par d'autres partenaires protéiques comme JMJ16 qui présente une activité déméthylase sur H3K4 mais qui montre également une activité alternative sur H3K9 lorsqu'elle interagit avec la protéine MALE MEIOCYTE DEATH1 (MMD1) (Wang *et al.*, 2020).

Chez les animaux comme chez les plantes, la mono-, di- et triméthylation sur H3K27 sont effectuées par le Polycomb Repressive Complex 2 (PRC2) (Margueron and Reinberg, 2011; Schmitges *et al.*, 2011; Mozgova and Hennig, 2015) (**Figure I.2**). Toutefois, une particularité a été observée chez *Arabidopsis*, où d'autres protéines, ATXR5 et ATXR6, déposent spécifiquement la marque H3K27me1 (Jacob *et al.*, 2009). La marque H3K27me3 est retrouvée au niveau de gènes maintenus dans un état réprimés tandis que la marque H3K27me1 est enrichie dans les régions péri-centromériques de l'hétérochromatine (X. Zhang, Clarenz, *et al.*, 2007; Jacob *et al.*, 2014) (**Figure I.3**). ATXR5/6

jouent un rôle important dans la réplication de l'ADN, où elles déposent spécifiquement la marque H3K27me1 sur l'histone H3.1 nouvellement incorporée par CAF-1 (CHROMATIN ASSEMBLY FACTOR 1) au niveau de la fourche de réplication (Jacob *et al.*, 2014). De plus, la reconnaissance spécifique d'H3K4 non méthylée par le domaine PHD d'ATXR5 empêche également cette enzyme de déposer H3K27me1 sur les histones porteuses de la marque H3K4me3 (Bergamin *et al.*, 2017). Cette reconnaissance couplée avec d'autres mécanismes de régulation permettrait de protéger les gènes actifs de la répression.

Trois protéines qui appartiennent au groupe 4 (KDM4) de la famille JMJ sont impliquées dans le retrait des méthylations sur H3K27me3 : ELF6, REF6 et JMJ13 (Crevillén, 2020) (**Figure I.2**). Ces enzymes contiennent des motifs en doigt de Zinc dans leur région C-terminale leur permettant de cibler directement des séquences d'ADN spécifiques (Cui *et al.*, 2016; Zheng *et al.*, 2019). C'est le cas de REF6 qui reconnaît spécifiquement la séquence 5'-CTCTGYTY-3' (où Y représente C ou T) (Cui *et al.*, 2016). De plus, ELF6 peut interagir avec l'histone méthyltransférase d'H3K36 SDG8 et pourrait constituer un lien entre l'activité transcriptionnelle médiée par la méthylation d'H3K36 et la dé-répression médiée par la déméthylation d'H3K27 (Yang, Howard and Dean, 2016). Enfin, les déméthylases de la marque H3K27me3, REF6, ELF6 et JMJ13 semblent coopérer pour contrôler les limites spatiales de cette marque afin de limiter son expansion au-delà des gènes ciblés par PRC2 (Yan *et al.*, 2018).



**Figure I.2 : Principaux acteurs du dépôt/retrait de groupements méthyl sur certaines lysines de la queue d'histone d'H3.1 chez Arabidopsis.** Le nom des méthyltransférases sont représentés dans un cadre rose et le dépôt de groupements méthyl est indiqué par une flèche pleine. Le nom des déméthylases sont représentés dans un cadre bleu et le retrait de groupements méthyl est indiqué par une flèche en pointillé. Les groupements méthyl associés à l'activité de transcription sont représenté en vert et ceux associés aux gènes réprimés en rouge. Le groupement dont l'association n'est pas définie est représenté en gris. La fonction de ces acteurs est expliquée tout au long du manuscrit.



### 3. État chromatinien

L'analyse combinatoire des différentes modifications et de leur localisation a permis de révéler un nombre restreint de possibilités que l'on appelle états chromatinien (CS / *Chromatin State*). Chez *Arabidopsis*, quatre états chromatinien ont été définis à partir de l'analyse, à l'échelle du génome, de la répartition de douze variants ou PTMs (Roudier *et al.*, 2011). L'état 1 (CS1) est principalement enrichi en marques H3K4me3 et H3K36me3 et corrèle avec des régions de l'euchromatine où l'activité de transcription est élevée. Le CS2 correspond aux régions géniques réprimées et fortement enrichies en marque H3K27me3 tandis que le CS3 correspond aux régions de l'hétérochromatine riches en éléments transposables. Enfin, le CS4 ne contient aucun enrichissement d'une marque particulière et semble être associé aux régions inter-géniques. Plus récemment, une autre étude basée sur l'analyse de seize modifications a révélé 9 états chromatinien (Sequeira-Mendes *et al.*, 2014). Parmi ces nouveaux états identifiés, l'état 2 semble intermédiaire au CS1 et CS2 caractérisés par Roudier *et al.*, 2011 et représente des régions fortement enrichies avec les marques H3K4me3 et H3K27me3 situées au niveau des promoteurs des gènes. En 2018, une nouvelle étude combinant de récents jeux de données épigénomiques (*Chromatin immunoprecipitation - sequencing* (ChIP-Seq), *DNase I hypersensitive sites - sequencing* (DNase-seq), *Assay for Transposase-Accessible Chromatin using sequencing* (ATAC-seq)) a révélé la présence de 36 états chromatinien chez *Arabidopsis* (Liu *et al.*, 2018).

Ces résultats indiquent que la régulation de l'expression des gènes par les PTMs n'est pas limitée à la présence de modifications individuelles mais plutôt à la combinaison, l'enrichissement et à la distribution de celles-ci sur le génome.

## B. Les protéines des groupes Polycomb et Trithorax fonctionnent de manière antagoniste

Nous avons vu que de nombreuses marques épigénétiques régulent la structure de la chromatine et ainsi la dynamique d'expression des gènes (activation ou répression). Ces marques elles-mêmes sont très dynamiques, grâce à l'activité d'un grand nombre de régulateurs protéiques qui déposent, retirent ou lisent spécifiquement les modifications portées par la chromatine. Parmi ces régulateurs épigénétiques, il existe deux groupes de protéines qui agissent de manière antagoniste dans l'activité de transcription, et qui jouent un rôle majeur dans la régulation du développement des eucaryotes, avec des

rôles globalement conservés entre animaux et plantes, mais aussi quelques divergences. Il s'agit des groupes Polycomb et Trithorax.

Historiquement, les protéines du groupe Polycomb (PcG) ont été découvertes par des approches de génétique chez la *Drosophile* où elles ont montré un rôle dans le maintien de la répression des gènes homéotiques *Hox* (Lewis, 1947, 1978). Lorsque les protéines de ce groupe sont mutées, elles conduisent à une expression ectopique des gènes *Hox*, ce qui engendre de nombreuses modifications morphologiques (Lewis, 1978). Par opposition, les gènes dont la mutation restaure le phénotype causé par la dérépression des gènes *Hox* ont été classés dans le groupe Trithorax (TrxG) (Ingham, 1983, 1985).

Depuis cette découverte, de nombreuses protéines du groupe PcG et TrxG ont été identifiées chez les mammifères et chez les plantes. Le fait que ces protéines soient conservées chez de nombreuses espèces et causent de sévères effets phénotypiques (souvent embryo-létaux chez les animaux) lorsqu'elles sont mutées, souligne l'importance de leur fonction pour l'établissement des plans développementaux des individus (Faust *et al.*, 1995; O'Carroll *et al.*, 2001; Wang *et al.*, 2007; Kassis, Kennison and Tamkun, 2017).

Les protéines PcG peuvent former deux complexes majeurs : PRC1 et PRC2. Ces complexes interviennent dans la répression de gènes cibles par le dépôt de modifications post-traductionnelles, telles que la tri-méthylation sur H3K27 (PRC2) ou l'ubiquitination de l'histone H2A (PRC1) (Cao *et al.*, 2002; H. Wang *et al.*, 2004). Chez les plantes, les protéines des complexes PcG ont été identifiées, comme chez la *Drosophile* par caractérisation de mutants homéotiques (génétique directe), ou bien par homologie de séquence avec des protéines animales connues (génétique inverse).

Chez les plantes et les animaux, le complexe PRC2 joue un rôle essentiel dans la détermination de l'identité cellulaire et la répression des gènes homéotiques qui sont cruciaux pour le développement des organismes multicellulaires. Cependant, le complexe PRC2 n'est pas uniquement associé aux organismes multicellulaires. En effet, plusieurs homologues des sous-unités de ce complexe ont été identifiés chez des eucaryotes unicellulaires tels que la paramécie *Paramecium tetraurelia* (Frapporti *et al.*, 2019), la levure *Cryptococcus neoformans* (Dumesic *et al.*, 2015), l'algue rouge *Cyanidioschyzon merolae* (Mikulski *et al.*, 2017) mais aussi la diatomée *Phaedodactylum tricornutum* (Veluchamy *et al.*, 2015) ou encore l'algue verte *Chlamydomonas reinhardtii*, un eucaryote unicellulaire de la lignée verte (Viridiplantae) (Shaver *et al.*, 2010;

Vijayanathan, Trejo-Arellano and Mozgová, 2022). La présence de sous-unités du PRC2 dans ces organismes ainsi que la récente découverte d'orthologues d'Enhancer-of-Zeste (E(z) ; voir tableau I.1 p33) dans *Discoba* (Sharaf *et al.*, 2022), une lignée hypothétiquement proche de l'embranchement des Eucaryotes, indiquent donc que ce complexe serait apparu tôt dans l'évolution, avant la diversification des eucaryotes. En revanche certains organismes unicellulaires comme *Schizosaccharomyces pombe* ou *Saccharomyces cerevisiae* sont dépourvus du complexe PRC2, ce qui indique qu'il a été également perdu à certains moment au cours de l'évolution (Shaver *et al.*, 2010).

La fonction ancestrale du complexe PRC2 chez les eucaryotes unicellulaires est encore mal caractérisée mais pourrait impliquer la répression des éléments transposables comme cela été montré pour Ezl1 (Enhancer-of-Zeste-like 1), l'homologue d'EZH2 (Enhancer-of-Zeste 2) chez *Paramecium tetraurelia* (Frapporti *et al.*, 2019).

Les protéines Trithorax quant à elles, sont impliquées de manière générale dans l'activation de la transcription et le maintien de cette activité (Schuettengruber *et al.*, 2011; Kingston and Tamkun, 2014). Au-delà de leur rôle dans l'activité de transcription, elles peuvent également contrecarrer la répression des gènes médiée par PcG à travers le dépôt de modifications post-traductionnelles covalentes sur les histones (voir p36). Les protéines TrxG sont un groupe de protéines plus hétérogènes que le groupe PcG et dont les fonctions sont très diverses. Parmi ces protéines, on trouve des méthyltransférases à domaine SET qui déposent les marques H3K4me2/3 et H3K36me2/3 associées à une activité de transcription élevée, ou encore des complexes de remodelage de la chromatine permettant de rendre l'ADN accessible à la machinerie de transcription. Chez les plantes, les protéines TrxG sont moins bien caractérisées même si certaines ont pu être identifiées par homologie avec des protéines TrxG connues chez d'autres espèces ou bien par caractérisation génétique basée sur la capacité de mutants à supprimer les phénotypes de mutants PcG. (Mozgova and Hennig, 2015; Schuettengruber *et al.*, 2017).

## **1. Composition et fonction du complexe protéique PRC2**

Le PRC2 est un acteur déterminant dans le devenir des cellules, par le contrôle de l'expression des gènes qui régulent leur différenciation. Chez les mammifères par exemple, une déplétion ou une dérégulation de ce complexe a des conséquences létales aux stades précoces du développement et peut également favoriser l'apparition de cancer (Varambally *et al.*, 2002; Boyer *et al.*, 2006). Son rôle essentiel dans le développement en fait donc l'un des complexes les plus étudiés à ce jour chez les mammifères.

Le PRC2 est un complexe de quatre sous-unités qui sont conservées chez de nombreuses espèces (**Tableau I.1**). Chez les mammifères il est composé de :

- Enhancer of Zeste 2 (EZH2) ou de son paralogue EZH1. Ce sont des histones méthyltransférases à domaine SET qui catalysent le dépôt de la marque épigénétique H3K27me3 (Cao *et al.*, 2002; Czermin *et al.*, 2002; Kuzmichev *et al.*, 2002; Müller *et al.*, 2002). Plusieurs différences ont pu être observées entre PRC2-EZH2 et PRC2-EZH1. Par exemple, chez la souris, alors que le niveau d'expression de PRC2-EZH2 est élevé dans les tissus précoces et diminue au cours du développement, l'expression de PRC2-EZH1 reste constante (Margueron *et al.*, 2008). De plus, PRC2-EZH2 semble montrer une activité méthyltransférase intrinsèque plus importante que PRC2-EZH1 (Margueron *et al.*, 2008; Lee *et al.*, 2018). Ce dernier, à la différence de PRC2-EZH2 a une plus forte affinité pour la chromatine et favorise sa compaction (Margueron *et al.*, 2008; Grau *et al.*, 2021a).
- Suppressor of Zeste 12 (SUZ12), dont le rôle est de stabiliser le complexe et de stimuler son activité HMT grâce à son domaine VEFS (**V**RN2 **E**MF2 **F**IS2 **S**UZ12) (Cao and Zhang, 2004; Pasini *et al.*, 2004). SUZ12 joue également un rôle de plateforme d'interaction pour de nombreux co-facteurs.
- Embryonic Ectoderm Development (EED) est une protéine à domaine WD40 qui reconnaît la marque H3K27me3. Des études structurales de ce domaine indiquent que la liaison d'H3K27me3 à EED entraîne un changement conformationnel d'EZH2 ce qui renforce son activité catalytique et permet la propagation d'H3K27me3 vers les nucléosomes voisins (Margueron *et al.*, 2009; Jiao and Liu, 2015). EED peut également reconnaître d'autres acides aminés méthylés, tel que la lysine 116 de JARID2 (Sanulli *et al.*, 2015)
- Retinoblastoma Binding Protein 4 et 7 (RBBP4/7) est une autre protéine à domaine WD40. Elle est connue pour interagir avec le nucléosome mais n'est pas essentielle pour l'activité HMT *in vitro* (Ketel *et al.*, 2005).
- Ce complexe PRC2 à 4 sous-unités est appelé PRC2 canonique mais de nombreuses protéines accessoires peuvent s'y associer pour moduler son activité et former au moins deux complexes : PRC2.2 qui est caractérisé par la présence des protéines JARID2 et AEBP2 et PRC2.1 qui peut inclure un des trois homologues

de la protéine Polycomb-like (PHF1, PH19 et MTF2), et les protéines EPOP ou PAL1/2 (Mierlo *et al.*, 2019).

**Tableau I.1 : Sous-unités conservées du PRC2 chez la Drosophile, chez l'Homme, la souris et chez Arabidopsis**

Drosophile	Homme/Souris	Arabidopsis
E(z)	EZH1 ; EZH2	CLF ; SWN ; MEA
Su(z)12	SUZ12	FIS2 ; EMF2 ; VRN2
Esc	EED	FIE
Nurf55	RBBP4/7	MSI1

## 2. Structures du PRC2

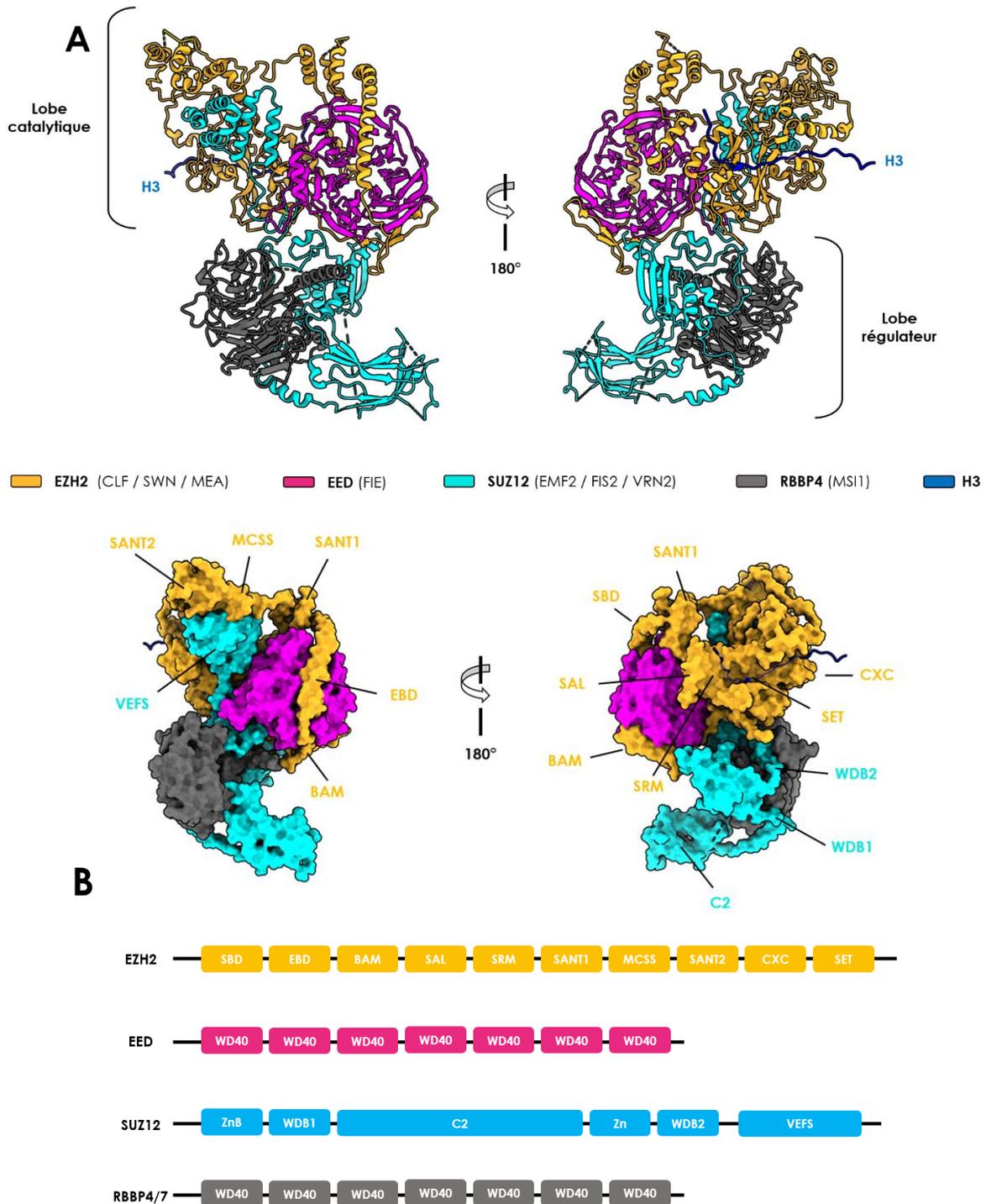
Avec la cristallographie aux rayons X et l'essor de la Cryo-microscopie électronique (Cryo-EM), de nombreuses structures du PRC2 chez différentes espèces (à l'exception des plantes) ont été résolues ces dernières années (Jiao and Liu, 2015; Justin *et al.*, 2016; Chen *et al.*, 2018, 2020; Kasinath *et al.*, 2018, 2021; Poepsel, Kasinath and Nogales, 2018; Finogenova *et al.*, 2020; Grau *et al.*, 2021a). Dans l'ensemble de ces structures, du fait de la conservation des différentes sous-unités, l'assemblage du complexe est similaire et révèle que le PRC2 peut être séparé en deux régions distinctes : le lobe catalytique et le lobe régulateur (ou lobe cible / *targeting lobe* en anglais) (**Figure I.6A**).

Le lobe catalytique comprend l'histone méthyltransférase EZH2/1, EED et le domaine C-terminal de SUZ12. La région N-Terminale d'EZH2/1 est composée d'une longue hélice  $\alpha$  (EBD) et d'un motif de feuillets  $\beta$  (BAM) qui enveloppe et stabilise la protéine EED, tout en laissant l'une de ses faces exposée pour la liaison avec des ligands (**Figure I.4A et B**). Elle est suivie de la boucle d'activation du domaine SET (SAL) et du motif de réponse à la stimulation (SRM) qui contribuent à stimuler l'activité HMT (**Figure I.4A et B**). En effet, le motif SRM se structure en hélice  $\alpha$  lorsqu'un peptide triméthylé se lie à EED, induisant ainsi une ouverture du site de liaison du substrat qui favorise sa liaison et l'activité catalytique. Par la suite, dans EZH2/1, on trouve le domaine SANT1, qui interagit avec une région distale d'EZH2/1 appelée domaine de liaison de SANT1 (SBD) (**Figure I.4A et B**). Enfin, les domaines MCSS (Motif Connecting SANT1 and SANT2), SANT2 d'EZH2/1 et le domaine VEFS de la partie C-terminale de SUZ12 forment la région centrale tandis que les domaines CXC et SET d'EZH2/1 composent la région C-terminale (**Figure I.4A et B**) (Chammas, Mocavini and Di Croce, 2020). Ce lobe est capable d'interagir avec des nucléosomes à la fois du côté C-terminal d'EZH2/1 par l'intermédiaire des domaines CXC et SET, mais aussi

du côté N-Terminal avec le domaine SBD et la protéine EED (Poepsel, Kasinath and Nogales, 2018).

Le lobe régulateur quant à lui, est formé de la région N-Terminale de SUZ12 qui interagit via ses domaines C2, WDB1 et WBD2 avec une autre protéine à domaine WD40, RBBP4/7. L'association de ces deux protéines forme une plateforme d'assemblage pour diverses protéines accessoires tel que AEBP2 (Chen *et al.*, 2018).

Le complexe PRC2-EZH2 avec les 4 sous-unités est fortement flexible et l'ajout de protéines accessoires comme AEBP2 et JARID2 a été nécessaire pour le stabiliser et résoudre sa structure par Cryo-EM (Kasinath *et al.*, 2018). Cependant, d'autres structures résolues par Cryo-EM en présence de protéines accessoires et de nucléosomes n'ont pas révélé de densité pour la région du lobe régulateur, confirmant ainsi sa grande flexibilité (Poepsel, Kasinath and Nogales, 2018; Finogenova *et al.*, 2020). Récemment, une structure d'un dimère de PRC2-EZH1 en complexe avec un nucléosome a révélé une orientation particulière du lobe régulateur, mettant ainsi en lumière son caractère flexible et un rôle dans la promotion de la compaction de la chromatine (Grau *et al.*, 2021a).



**Figure I.4 : Structure et domaines du complexe PRC2 chez les mammifères. (A)** Structure Cryo-EM du complexe PRC2-EZH2 humain (PDB : 6WKR). La méthyltransférase EZH2, la protéine à domaine WD40 EED et le domaine VEFS de SUZ12 forment le lobe catalytique du complexe tandis que la partie N-Terminale de SUZ12 et la protéine à domaine WD40 RBBP4/7 forment le lobe régulateur. Chaque sous-unité est représentée par une couleur différente : EZH2 est en jaune ; FIE en rose ; SUZ12 en bleu et RBBP4/7 en gris. Le ou les homologues de ces protéines chez *Arabidopsis* sont notés entre parenthèses. **(B)** Domaines identifiés pour chaque sous-unité du PRC2

### 3. Régulation de l'activité du PRC2

De nombreuses expériences biochimiques, ont révélé que l'activité catalytique du PRC2 peut être régulée par plusieurs facteurs.

Tout d'abord, elle peut être modulée par des changements allostériques au niveau du site catalytique, induits par la liaison de la marque H3K27me3 à EED (Margueron *et al.*, 2009; Jiao and Liu, 2015). Cependant ce cas n'est valable que pour EZH2, car EZH1 possède une substitution d'acide aminé dans son domaine SRM (C130 pour EZH1 vs H129 dans EZH2) qui influence l'activation allostérique et par conséquent réduit l'activité HMT du complexe correspondant (Lee *et al.*, 2018). L'activité catalytique peut donc être régulée directement par des différences au niveau de la composition du complexe canonique (présence de paralogue).

De plus, une interaction stable du complexe avec la chromatine améliore également son activité méthyltransférase. Cela a été démontré par plusieurs tests *in vitro* avec différents substrats qui ont révélé une meilleure activité catalytique du complexe avec deux nucléosomes liés par 40 pb d'ADN (dinucléosome) plutôt qu'avec un seul (mononucléosome) (Lee *et al.*, 2018). Par exemple, des mutations ponctuelles au niveau du domaine CXC qui déstabilisent l'interaction du PRC2 avec l'ADN ont montré une diminution de son activité méthyltransférase (Finogenova *et al.*, 2020). De même, des protéines accessoires telles qu'AEBP2 peuvent stabiliser l'interaction du PRC2 avec le nucléosome, contribuant ainsi à stimuler son activité (Lee *et al.*, 2018).

L'activité du complexe peut être également régulée par la présence de PTMs sur EZH2 et sur les queues d'histones. Ces PTMs peuvent avoir un effet positif tel que l'autométhylation d'EZH2 qui peut stimuler son activité catalytique en favorisant la conversion d'H3K27me2 en H3K27me3 (Lee *et al.*, 2019; Wang *et al.*, 2019; Sauer *et al.*, 2023), la présence de la marque H3K27me3 sur un nucléosome adjacent qui favorise également son activité (comme mentionné dans le paragraphe précédent), ou bien encore un effet négatif par la présence de marques sur la même histone 3, comme H3K4me3 et H3K36me3. En effet, des expériences *in vitro* ont révélé que ces marques réduisaient l'activité méthyltransférase de PRC2 chez l'Homme, la Drosophile mais aussi chez Arabidopsis (Schmitges *et al.*, 2011; Finogenova *et al.*, 2020; Kasinath *et al.*, 2021). En revanche, l'activité du complexe PRC2-VRN2 chez Arabidopsis (présenté plus tard dans ce manuscrit) ne semble pas être perturbée par les marques H3K4me3 et H3K36me3 (Schmitges *et al.*, 2011).

La phosphorylation de la sérine 21 (S21) d'EZH2 sous une forme non-PRC2 (c'est-à-dire sans les protéines EED, SUZ12 et RBBP4/7) a permis de mettre en évidence un domaine de transactivation (TAD) qui est composé des domaines SRM et SANT1. Lorsque S21 est phosphorylée, le TAD est « déverrouillé » et permet l'activation de gènes rapporteurs via le recrutement de la protéine p300, une acétyltransférase (Jiao *et al.*, 2020). Bien que d'autres études montrent également l'interaction entre EZH2 non-PRC2 et p300 par Co-IP (J. Wang *et al.*, 2022), le manque de tests d'interactions complémentaires sur la nature de cette interaction *in vivo*, requiert de la prudence dans l'interprétation de ce résultat.

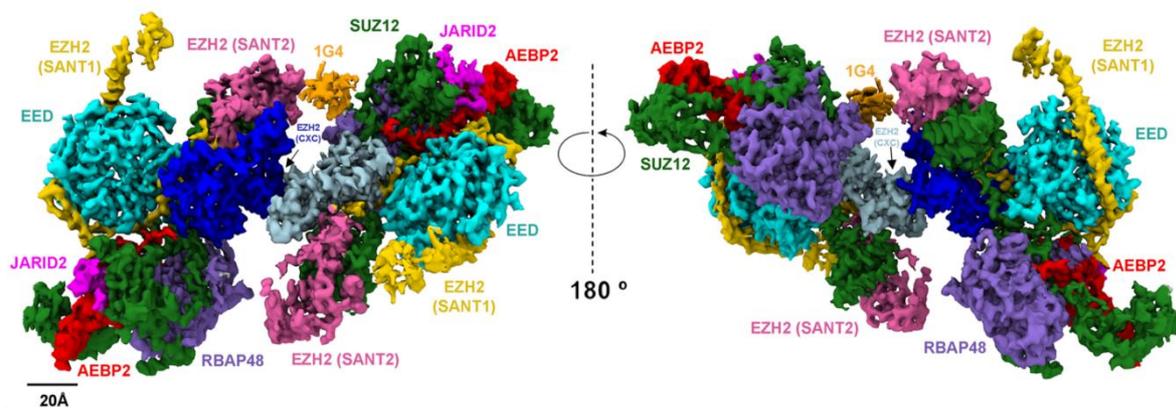
Récemment, des PTMs sur les autres sous-unités du complexe PRC2 et qui influencent son activité ont également été identifiées. Par exemple, la présence de phosphorylations sur FIE (l'homologue de EED chez les Arabidopsis) permet la translocation de cette protéine du cytoplasme vers le noyau. De plus, une mutation de ces résidus pour prévenir la phosphorylation de FIE, modifie le profil global d'H3K27me3, et perturbe l'organogenèse chez la plante (Ye *et al.*, 2022).

Chez l'homme la phosphorylation de la sérine 583 présente dans le domaine VEFS de SUZ12, stabilise le site catalytique de PRC2, augmente son activité enzymatique et la liaison au nucléosome *in vitro* (Gong *et al.*, 2022).

La résolution d'une structure du PRC2 par Cryo-EM en interaction avec un nucléosome porteur de la marque H3K4me3 a révélé deux états du complexe : l'un où la queue d'histone atteint le site catalytique et l'autre où elle est absente, ce qui indique que cette modification pourrait provoquer une instabilité du peptide et un mauvais placement de la lysine K27 dans le site catalytique (Kasinath *et al.*, 2021). Les autres structures du PRC2 résolues par Cryo-EM en complexe avec un mono- ou di-nucléosome ont également révélé le rôle critique du résidu K36 dans l'orientation de la queue d'histone. Il a donc été proposé que la présence de modification (ou de mutations) sur ce résidu pourrait également entraîner une déstabilisation du peptide et expliquer la diminution de l'activité du PRC2 *in vitro*. On peut noter que la réduction de l'activité du complexe par la présence d'H3K36me3 ne perturbe pas la liaison du PRC2 avec le nucléosome (Finogenova *et al.*, 2020).

Enfin, l'interaction du PRC2 avec des ARNs semble jouer un rôle dans le recrutement de ce complexe à ses cibles, mais aussi dans la régulation de son activité. En effet, de nombreuses études ont révélé l'association du PRC2 avec des ARNs *in vivo* et *in vitro* (Kaneko *et al.*, 2013; Cifuentes-Rojas *et al.*, 2014). Cette liaison ne semble pas indiquer de spécificité de séquence mais plutôt une préférence pour les séquences riches en CG

et G-quadruplexes (G4) (Wang *et al.*, 2017). Par exemple, une étude a montré *in vitro* que l'inhibition de l'activité du complexe pouvait être en partie due à la liaison d'ARNs au niveau du site de régulation allostérique du PRC2. L'étude montre également que cette liaison est affectée par la présence de la marque H3K27me3 (Zhang *et al.*, 2019). Dernièrement, un autre mécanisme d'inhibition de l'activité de PRC2 a été révélé. En effet une structure Cryo-EM a dévoilé un état dimérique induit par la présence d'un G4 d'ARN. Ce dimère PRC2 adopte une conformation où les deux sites catalytiques se font face, limitant ainsi l'interaction du complexe avec le nucléosome et l'accessibilité de la queue d'histone (**Figure I.5**) (Song *et al.*, 2023).



**Figure I.5 : Structure dimérique du complexe PRC2 formée par un G-quadruplexe (G4) d'ARN.** Carte de densité Cryo-EM du complexe PRC2 en liaison avec un G4 d'ARN. EZH2 (SANT1) est surligné en jaune, EZH2 (SANT2) en rose, EZH2 (CXC-SET) du protomère 1 en bleu, EZH2 (CXC-SET) du protomère 2 en bleu clair, EED en cyan, RBAP48 en violet, SUZ12 en vert, JARID2 en magenta, AEBP2 en rouge et le G4 d'ARN en orange. Figure issue de la publication de Song *et al.*, 2023.

#### 4. Composition des complexes PRC2 canoniques chez Arabidopsis

Chez Arabidopsis, le PRC2 canonique est également composé de 4 sous-unités. Cependant, due à la multiplicité de certains gènes, plusieurs homologues de ces sous-unités sont codés dans son génome et peuvent se combiner pour former différents complexes (**Figure I.6A**).

- CURLY LEAF (CLF), MEDEA (MEA) et SWINGER (SWN) sont les trois homologues des méthyltransférases EZH1/2 chez les mammifères. Alors que MEA est spécifique de la famille des Brassicaceae et joue un rôle fondamental dans le développement du gamétophyte et des graines, CLF et SWN agissent préférentiellement au cours du développement végétatif et sporophytique (Goodrich *et al.*, 1997; Grossniklaus *et al.*, 1998; Kiyosue *et al.*, 1999; Chanvivattana *et al.*, 2004) (**Figure I.6A**).

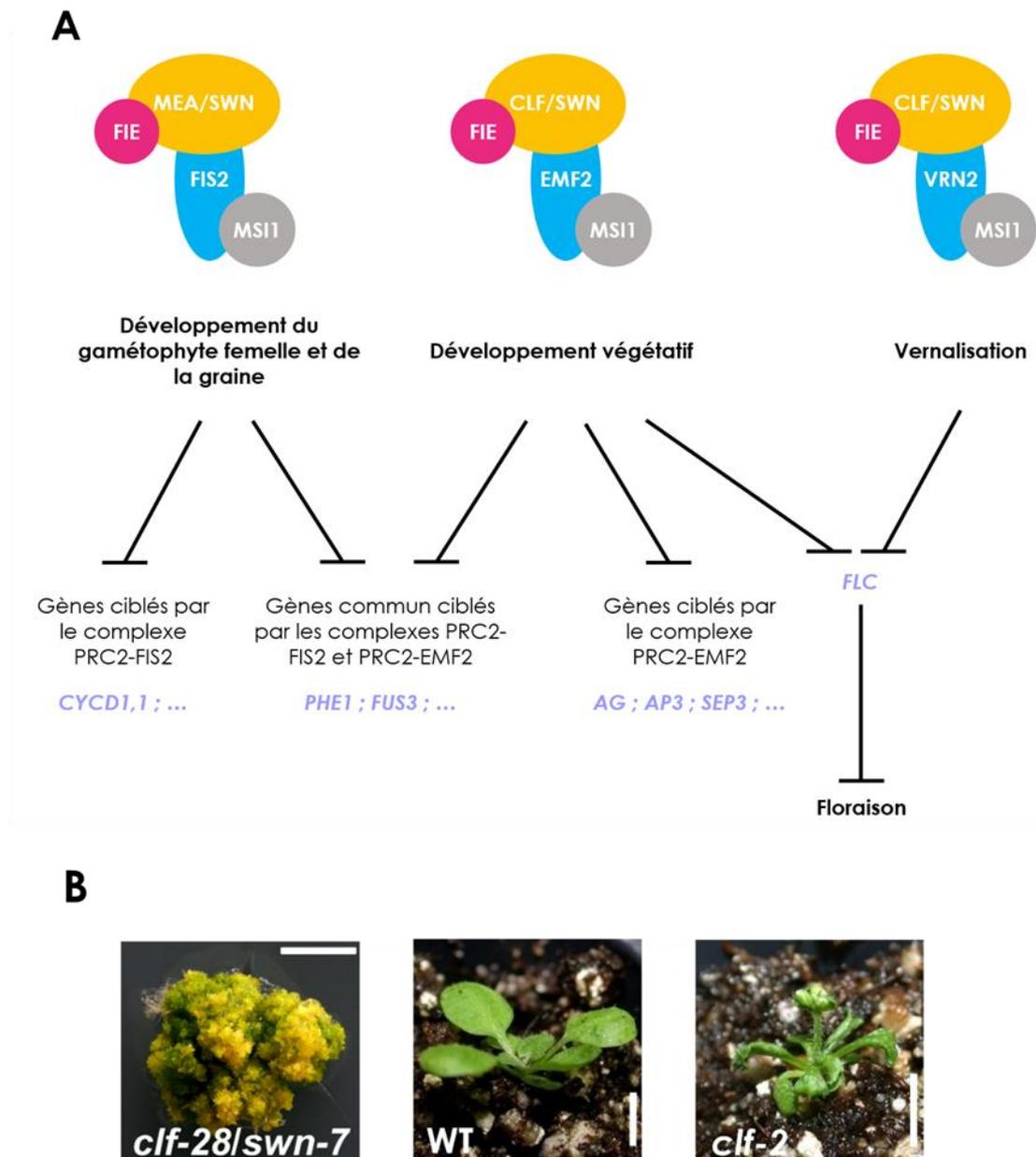
- Les protéines FERTILIZATION INDEPENDENT SEED 2 (FIS2), EMBRYONIC FLOWER 2 (EMF2) ou VERNALIZATION 2 (VRN2) correspondent à l'homologue de SUZ12. Chez *Arabidopsis*, 3 complexes PRC2 ont été nommés d'après ces 3 protéines et permettent la répression des gènes impliqués dans différentes étapes du développement : le complexe PRC2-FIS2 (FIS2, MEA/SWN, FIE, MSI1) intervient pendant le développement du gamétophyte et de l'endosperme, le complexe PRC2-EMF2 (EMF2, CLF/SWN, FIE, MSI1) opère lors du développement végétatif et embryonnaire, et enfin le complexe PRC2-VRN2 (VRN2, CLF/SWN, FIE, MSI1) agit lors de la réponse à la vernalisation pour la transition florale (Luo *et al.*, 1999; Gendall *et al.*, 2001; Yoshida *et al.*, 2001; Chanvivattana *et al.*, 2004) (**Figure I.10A**).
- FERTILIZATION INDEPENDENT ENDOSPERM (FIE) est l'unique homologue du composant EED chez l'homme (Ohad *et al.*, 1999).
- Enfin, la protéine MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) est apparentée aux protéines RBBP4 et 7 chez l'homme (Hennig *et al.*, 2003; Ketel *et al.*, 2005).

Des analyses génétiques et phénotypiques ont permis de mettre en évidence différentes caractéristiques fonctionnelles pour chacune des sous-unités qui composent les complexes PRC2 canoniques chez *Arabidopsis*. Par exemple les phénotypes des mutants perte de fonction de CLF (*clf*), qui exhibent des feuilles recourbées vers le haut et une transition florale rapide, sont beaucoup plus sévères que ceux des mutants perte de fonction de SWN (*swn*) qui ne présentent aucun défaut majeur (**Figure I.6B**). En revanche, les doubles mutants *clf swn* qui correspondent à une perte totale de l'activité catalytique de PRC2 dans la phase sporophytique, entraîne un développement cellulaire non-contrôlé qui se traduit par la formation d'un amas de cellules de plusieurs types (**Figure I.6B**). Ces phénotypes semblent donc révéler à la fois un rôle prépondérant de CLF, ainsi qu'une fonction différente de SWN, mais néanmoins essentielle en absence de CLF (Goodrich *et al.*, 1997; Chanvivattana *et al.*, 2004).

Récemment, la redondance partielle entre CLF et SWN a été illustrée par des expériences de ChIP-seq sur des plantules (Shu *et al.*, 2019). Bien que les résultats correspondants montrent que CLF et SWN sont capables de cibler un très grand nombre de gènes communs (quelques centaines de plus par SWN), leurs effets sur le dépôt de la marque H3K27me3 est différent. En effet, le niveau global d'H3K27me3 est largement réduit dans le mutant *clf* tandis que chez le mutant *swn* ce niveau reste quasi inchangé, en comparaison à des plantes sauvages (*Wild Type* ; WT). Le double mutant *clf swn* quant à lui, présente un niveau d'H3K27me3 quasi inexistant ce qui indique que ces enzymes

pourraient être les seules responsables du dépôt de cette marque à cette étape du développement. Ces résultats semblent ainsi confirmer des fonctions spécifiques à CLF et SWN mais aussi redondantes.

L'importance de la fonction de PRC2 dans le développement des plantes est bien décrite dans la littérature mais sa composition combinatoire en différentes sous-unités, et les fonctions des potentiels variants du complexe PRC2 correspondants, restent encore peu élucidés. Cela est largement dû au faible nombre de caractérisations biochimiques et à l'absence de données structurales sur le PRC2 chez les plantes, qui limitent grandement la compréhension du mécanisme d'action du complexe végétal (Bieluszewski *et al.*, 2021; Vijayanathan, Trejo-Arellano and Mozgová, 2022).



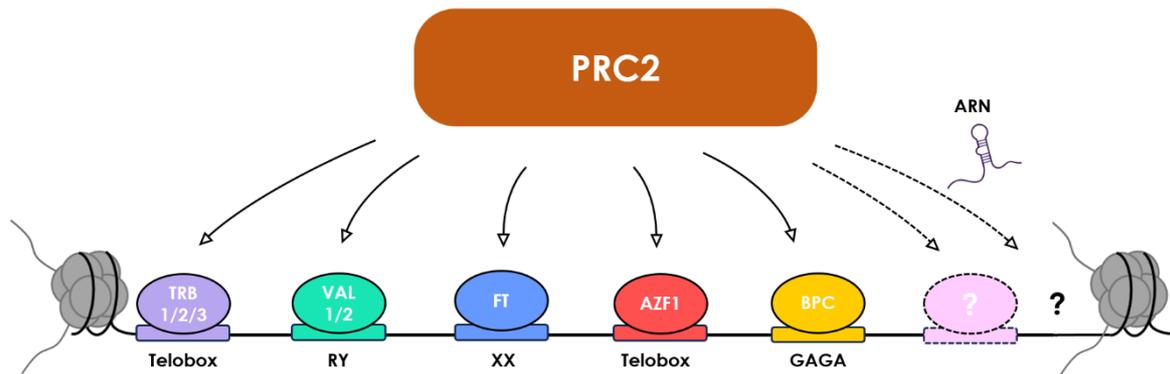
**Figure 1.6 : Diversité des complexes PRC2 canoniques chez Arabidopsis. (A)** Présentation de la composition des trois complexes PRC2 canoniques identifiés chez Arabidopsis. Ces complexes opèrent à différents étages du développement de la plante en réprimant des gènes impliqués dans d'autres étapes. Quelques gènes réprimés par ces différents complexes sont notés en violet. A titre d'exemple, le complexe PRC2-EMF2 agit lors du développement végétatif en réprimant les gènes d'identité florale tels qu'AGAMOUS (AG), SEPALATA3 (SEP3) ou APELATA3 (AP3). **(B)** Phénotype sauvage (WT), du double mutant perte-de-fonction de CLF et SWN (*clf-28/swn-7*) et du simple mutant perte-de-fonction de CLF (*clf*). Le double mutant *clf-28/swn-7* provoque un développement cellulaire non-contrôlé qui se traduit par la formation d'un amas de cellules de plusieurs types tandis que le simple mutant *clf* qui exhibe des feuilles recourbées vers le haut et une transition florale rapide. Les barres blanches sur les photos indiquent une échelle de taille de 0,5 cm. Les images des mutants proviennent des publications de Lafos *et al.*, 2011 (*clf-28/swn-7*) et de Carles and Fletcher, 2009a (*clf-2* et WT).

## 5. Mécanisme de recrutement du PRC2 chez Arabidopsis

Le PRC2 n'a pas la capacité de se lier à l'ADN de manière spécifique et le recrutement de ce complexe au niveau de gènes cibles chez Arabidopsis, semble faire intervenir des régulateurs transcriptionnels qui reconnaissent des séquences *cis*-régulatrices appelées *Polycomb Response Element* (PRE), telles qu'initialement définies chez la Drosophile (Simon *et al.*, 1993; Kassis and Brown, 2013) (**Figure I.7**). Par exemple, il a été montré que plusieurs facteurs pouvaient se lier à des séquences spécifiques d'ADN. C'est le cas d'ARABIDOPSIS ZINC-FINGER PROTEIN 1 (AZF1) qui interagit avec les motifs répétés GAGA, et de BASIC PENTACYSTEINE 1 (BPC1) qui se lie aux Telobox AAACCCTA. Des expériences de complémentation de fluorescence bimoléculaire (*Bimolecular Fluorescence Complementation* ; BiFC) ont révélé qu'AZF1 et BPC1 interagissent *in vivo* avec FIE. De plus, l'insertion artificielle de motifs GA et Telobox au niveau d'un gène rapporteur, est capable d'induire sa répression (Xiao *et al.*, 2017). Par ailleurs, les TELOMERE-REPEAT BINDING FACTOR 1, 2 et 3 (TRB1/2/3) sont également capables d'interagir physiquement avec PRC2 mais aussi de reconnaître le motif Telobox *via* un domaine MYB-Like situé dans leur région N-Terminale (Zhou *et al.*, 2018).

D'autres facteurs comme VIVIPAROUS1/ABI3-LIKE1 (VAL1) et VAL2 contribuent aussi au recrutement de PRC2 à ses gènes cibles. En effet ces facteurs reconnaissent à la fois la séquence d'ADN TGCATG (RY) et interagissent avec PRC2 *in vivo* (double-hybride de levure ; Y2H). Il a été montré que l'absence de VAL1 et de VAL2 induit une réduction de la présence de CLF et de SWN à certains *loci* comme celui du répresseur floral *FLC* (Yuan *et al.*, 2016, 2021).

Par ailleurs, plusieurs ARN non-codants (*Long non-coding RNA* ; lncRNA) provenant des gènes *FLC* (*COLDAIR*, *COLDWRAP*) et *AGAMOUS* (*incRNA4*) semblent également moduler l'activité du PRC2 et participer à son recrutement (Heo and Sung, 2011; Wu *et al.*, 2018). Cependant, bien que des interactions entre CLF et ces ARNs ont pu être mis en évidence, le détail des mécanismes moléculaires qui régissent le recrutement du PRC2 par ces derniers sont encore mal connus. De plus, il est possible que le recrutement du PRC2 au niveau de gènes spécifiques soit plutôt une conséquence d'une inhibition de la transcription, qu'un ciblage direct par ces ARNs. En effet, il a été montré récemment dans des cellules animales que la répression transcriptionnelle du gène *HOXD* induite par l'ARN *HOTAIR* était indépendante du PRC2 (Portoso *et al.*, 2017).



**Figure I.7 : Recrutement du complexe PRC2 vers ses cibles chez Arabidopsis.** Le recrutement du PRC2 chez Arabidopsis semble faire intervenir des régulateurs transcriptionnels et facteurs de transcriptions qui reconnaissent des séquences *cis*-régulatrices appelées *Polycomb Response Element* (PRE), telles qu'initialement définies chez la Drosophile (Kassis and Brown, 2013). Les régulateurs transcriptionnels présentés dans la figure et décrits dans le texte sont capables d'interagir *in vivo* avec le PRC2. En plus d'un rôle dans l'inactivation de l'activité de PRC2 (Beltran *et al.*, 2019; Zhang *et al.*, 2019), des ARNs semblent également jouer un rôle dans le recrutement (direct ou indirect) de ce complexe (Alecki *et al.*, 2020; Long *et al.*, 2020). Chez Arabidopsis l'ARN non-codant COLDAIR, induit lors de la vernalisation, conduit à une réduction de l'occupation du PRC2 et d'H3K27me3 lorsqu'il est muté (Kim, Xi and Sung, 2017). Cependant, le détail des mécanismes moléculaires qui régissent le recrutement du PRC2 par ces ARNs sont encore largement méconnus.

## 6. Dépôt de la marque H2AK121ub par le complexe PRC1

**Tableau I.2 : Sous-unités du complexe PRC1 chez la Drosophile, chez l'Homme, et chez Arabidopsis (adapté de Mozgova and Hennig, 2015).**

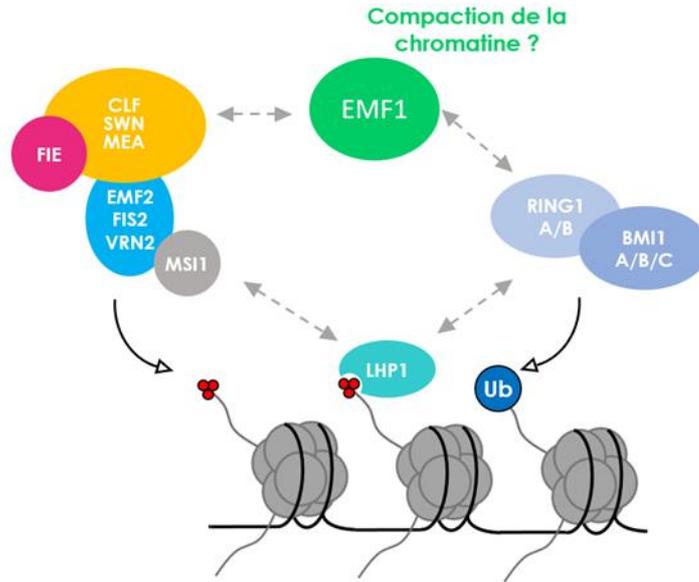
Drosophile	Homme	Arabidopsis
Pc	CBX2	-
	CBX4	
	CBX6	
	CBX7	
	CBX8	
Ph	PH1-3	-
Ph-p		
Ph-d		
Psc-CTR Su(z)2	PCGF1-6	BMI1A/B/C
Sce	RING1A/B	RING1A/B

Le PRC1 est un autre complexe formé par les protéines PcG et impliqué dans la répression des gènes. Chez Arabidopsis les protéines RING1A ou B et les protéines BMI1A,B ou C ont été identifiées comme membres de ce complexe par homologie avec d'autres sous-unités présentes chez les vertébrés et la Drosophile (Bratzel *et al.*, 2010; Wang and Shen, 2018) (**Tableau I.2**). Elles possèdent toutes un domaine E3 ubiquitine ligase dans leur

région N-terminale appelé RING qui est responsable de la monoubiquitination de la lysine 121 d'H2A (H2AK121ub). Chez les animaux, les homologues de RING1A ou B forment un hétérodimère avec PCGF1-6, les homologues des protéines BMI1A/B/C, tandis que chez Arabidopsis, toutes ces protéines interagissent les unes avec les autres (Xu and Shen, 2008; Bratzel *et al.*, 2010; D. Chen *et al.*, 2010). Il n'est donc pas encore clair si les protéines RING1 et BMI1 forment un hétérodimère *in vivo*.

Chez les plantes, d'autres protéines peuvent également s'associer au complexe PRC1 comme par exemple LIKE- HETEROCHROMATIN PROTEIN 1 (LHP1) qui est un homologue de la protéine animale HETEROCHROMATIN PROTEIN 1 (HP1) (Gaudin *et al.*, 2001; Kotake *et al.*, 2003). Cependant, à la différence de cette dernière qui interagit avec l'hétérochromatine et la marque H3k9me2, LHP1 est associée à l'euchromatine et peut reconnaître la marque H3K27me3 via un chromodomaine situé dans sa région N-Terminale (X. Zhang, Germann, *et al.*, 2007; Liu *et al.*, 2022). De plus, LHP1 semble aussi faire partie du complexe PRC2 puisque des expériences de purification d'affinité couplée à la spectrométrie de masse (IP-MS) ont dévoilé une co-purification de LHP1 avec différentes sous-unités du complexe PRC2 comme MSI1 (Derkacheva *et al.*, 2013) (**Figure I.8**). LHP1 participe au maintien et à la diffusion de la marque H3K27me3 mais ne semble pas être nécessaire pour le dépôt de la marque H2AK121ub (Derkacheva *et al.*, 2013; Yang *et al.*, 2017; Zhou *et al.*, 2017) .

Des expériences de *pull-down* ont également identifié EMBRYONIC FLOWER 1 (EMF1) comme membre du complexe PRC1. En effet EMF1 est capable d'interagir *in vitro* avec les sous-unités BMI1A/B ; RING1A/B et a été caractérisée comme un équivalent fonctionnel de la protéine de Drosophile Psc-CTR (Calonje *et al.*, 2008; Bratzel *et al.*, 2010; Beh, Colwell and Francis, 2012). De par la présence localisée d'acides aminés chargés positivement, Psc-CTR et EMF1 sont capables d'interagir avec la chromatine avec une affinité de l'ordre du nano-Molaire (nM) et d'induire la répression de gènes. De plus, Psc-CTR semble jouer un rôle dans la compaction de la chromatine, ce qui indique que cette fonction pourrait également être présente chez EMF1 (Francis, Kingston and Woodcock, 2004). Enfin, comme pour LHP1, EMF1 est co-purifiée avec des protéines du PRC2, ce qui la place aussi comme une protéine accessoire de ce complexe (**Figure I.13**) (Liang *et al.*, 2015; Bloomer *et al.*, 2020).



**Figure 1.8 : Interconnexions entre les complexes PRC1 et PRC2 chez Arabidopsis. Les complexes PRC1 et PRC2 catalysent le dépôt de la monoubiquitination de la lysine 121 sur H2A (H2AK121ub) et de méthylations sur H3K27 respectivement. Les protéines LHP1, qui reconnaît la marque H3K27me3 et EMF1 qui pourrait avoir un rôle dans la compaction de la chromatine, interagissent in vivo avec PRC2 et PRC1 (flèches grises en pointillées) faisant un lien entre ces deux complexes.**

## 7. Rôles d'H2AK121ub et H3K27me3 dans la régulation de la transcription chez les plantes

PRC1 et PRC2 exercent des fonctions essentielles dans le développement des organismes en contrôlant l'expression des gènes par le dépôt d'ubiquitine sur H2A ou de tri-méthylations sur H3K27 respectivement.

Chez les vertébrés, plusieurs études ont révélé que les complexes PRC1 et PRC2 semblent fonctionner en étroites relations. En effet, la marque H2AK119ub déposée par le PRC1 peut recruter le PRC2 *via* sa protéine accessoire JARID2 et à l'inverse, le dépôt de la marque H3K27me3 par le PRC2, peut faciliter le recrutement de PRC1 (Cao *et al.*, 2002; Min, Zhang and Xu, 2003; L. Wang *et al.*, 2004; Blackledge *et al.*, 2014; Cooper *et al.*, 2014; Kalb *et al.*, 2014; Sugishita *et al.*, 2021).

Chez Arabidopsis, la marque H2AK121ub déposée par le PRC1 est majoritairement distribuée directement après le site d'initiation de la transcription et ne chevauche que partiellement la marque H3k27me3, qui elle, semble répartie équitablement le long du corps du gène (Zhou *et al.*, 2017). Cependant bien qu'H2AK121ub et H3K27me3 co-localisent pour certains gènes, d'autres sont uniquement marqués par H2AK121ub ou H3K27me3 ce qui indique que ces marques pourraient jouer des rôles différents et indépendants dans la répression de la transcription (Zhou *et al.*, 2017; Kralemann *et al.*,

2020). L'activité du PRC2 n'est pas requise pour le dépôt de la marque H2AK121ub alors qu'à l'inverse, une perte de fonction du PRC1 induite par la mutation des protéines BMI1A/B/C (*bmi1a/b/c*) induit une réduction de la marque H3K27me3. Par conséquent le PRC1 agirait en amont du PRC2 à certain *loci* (Zhou *et al.*, 2017).

Il a été montré qu'environ 60 % des gènes marqués uniquement par H2AK121ub restent actifs même si le niveau moyen d'expression est plus faible que les gènes qui ne sont pas marqués par cette modification (Zhou *et al.*, 2017). De plus, ce niveau moyen d'expression diminue lorsque la marque H2AK121ub et H3K27me3 co-localisent et devient encore plus faible pour les gènes seulement marqués par H3K27me3 (Zhou *et al.*, 2017; Kralemann *et al.*, 2020). Récemment une étude a montré que la marque H2AK121ub est associée avec un état de la chromatine moins accessible mais permissif à la transcription c'est pourquoi il a été proposé que cette marque joue un rôle modulateur plutôt que répresseur. En revanche, l'accessibilité de la chromatine aux facteurs de transcription est beaucoup plus réduite pour les gènes porteurs des marques H2AK121ub et H3K27me3 et l'est encore plus pour les gènes uniquement marqués par H3K27me3 (Yin *et al.*, 2021).

## **8. De nombreux effecteurs influencent la répression des gènes via divers mécanismes**

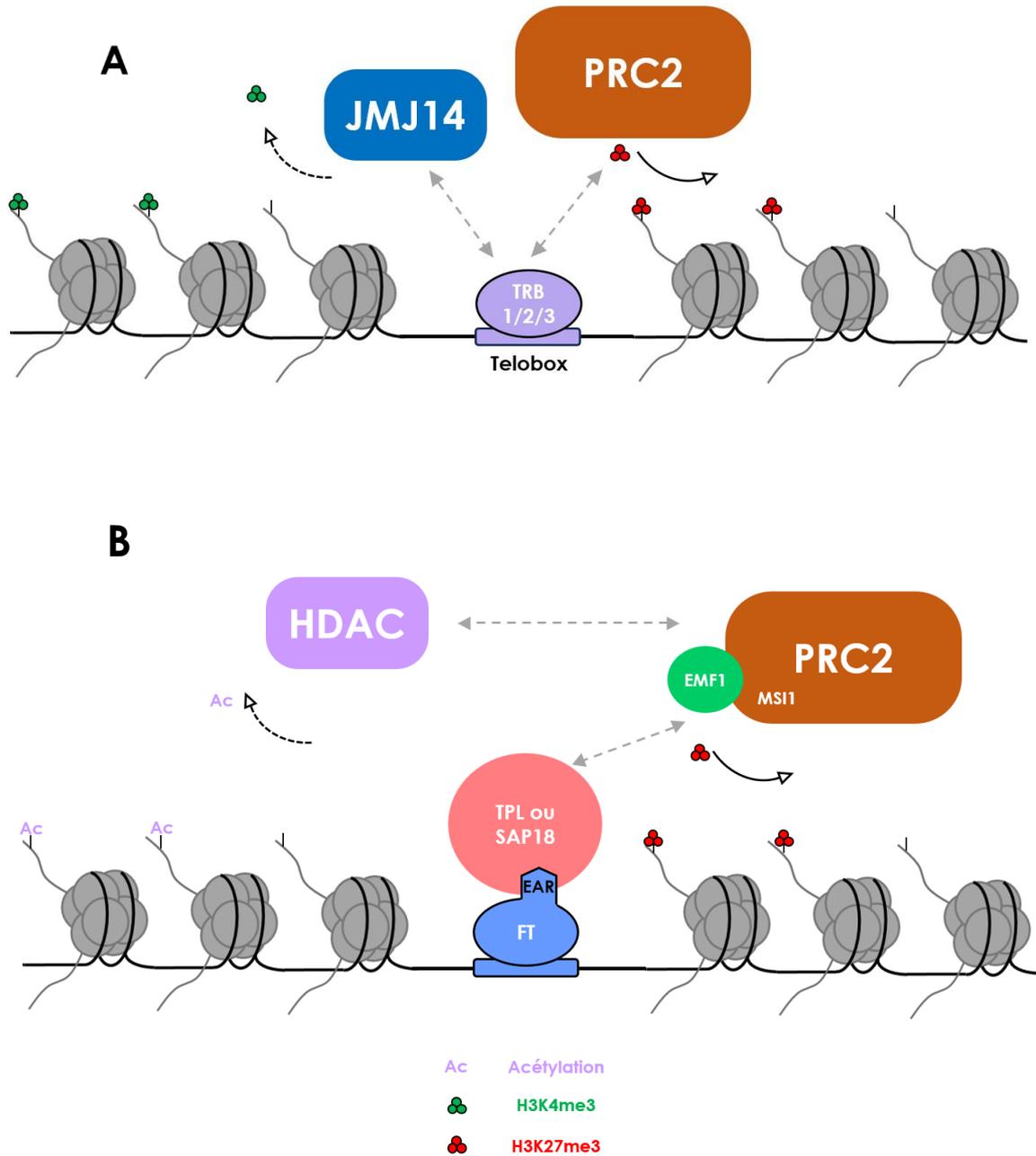
A l'échelle du génome, le dépôt de la marque H3K27me3 et l'occupation par PcG corrélaient avec un état répressif de la chromatine. Cependant, bien qu'il y ait eu de grandes avancées sur la compréhension de la répression de la transcription par ces protéines, les mécanismes moléculaires sous-jacents restent encore largement méconnus (Schuettengruber *et al.*, 2017; Bieluszewski *et al.*, 2021; Blackledge and Klose, 2021; Guo, Zhao and Wang, 2021).

Ces dernières années, de nombreux effecteurs qui reconnaissent et interprètent la marque H3K27me3 et/ou qui se lient avec les protéines du PRC2 ont été identifiés. Parmi ces protéines, SHORT LIFE (SHL) et EARLY BOLTING IN SHORT DAYS (EBS) chez *Arabidopsis*, reconnaissent la marque H3K27me3 via un domaine BAH (Bromo-adjacent homology) et forment un complexe avec EMF1 qui est impliqué dans la répression de nombreux gènes (Li *et al.*, 2018; Qian *et al.*, 2018; Yang *et al.*, 2018). EMF1 partage des caractéristiques fonctionnelles avec la protéine de *Drosophila* Psc-CTR et pourrait restreindre par compaction l'accessibilité de la chromatine à divers effecteurs impliqués dans la régulation transcriptionnelle, comme cela a été montré pour le complexe de remodelage SWI/SNF (Shao *et al.*, 1999).

Récemment, une autre protéine à domaine BAH : ANTISILENCING 1 (AS11)-IMMUNOPRECIPITATED PROTEIN 3 (AIPP3), a été identifiée chez *Arabidopsis* (Y.-Z. Zhang *et al.*, 2020). AIPP3 forme un complexe avec les protéines AIPP2 (un lecteur d'H3K4) et CPL2 qui participent à la répression de la transcription de centaines de gènes par une inhibition de l'ARN Pol II. En effet, CPL2 est une phosphatase de la sérine 5 du domaine C-Terminal de l'ARN Pol II dont la phosphorylation après l'étape d'initiation est nécessaire à la libération de l'enzyme et au début de la transcription (Koiwa *et al.*, 2004). Par conséquent, la perte de cette phosphorylation sur la sérine 5 par CPL2 conduit à une inhibition de la transcription (Y.-Z. Zhang *et al.*, 2020).

Le complexe PRC2 peut s'associer et/ou coopérer avec d'autres effecteurs impliqués dans la répression des gènes par le retrait des marques associées à un état actif de la transcription, comme H3K4me3 et H3K36me2/3 (déposées par le groupe Trithorax) ou les acétylations d'histones (Hollender and Liu, 2008; Lu *et al.*, 2010). Par exemple, la déméthylase JMJ14 et PRC2 semblent coopérer pour réprimer certains gènes. En effet, les TRBs qui sont impliqués dans le recrutement de PRC2 peuvent également interagir avec JMJ14 (**Figure I.9A**). De plus lorsque ces TRBs sont spécifiquement dirigés au niveau du gène *FWA* par l'intermédiaire d'un *Zinc-finger*, une réduction de l'expression du gène est observée. Les protéines TRBs coordonnent donc l'activité de JMJ14 et PRC2 pour combiner le retrait de la marque H3K4me3 et le dépôt de la marque H3K27me3 pour ainsi renforcer la répression des gènes (Wang *et al.*, 2023).

Plusieurs régulateurs de la transcription impliqués dans le recrutement de PRC2 comme BPC1 ou AZF1 contiennent des domaines EAR (Ethylene-responsive element binding factor-associated Amphiphilic Repression) qui peuvent être reconnus par les co-répresseurs STRESS ASSOCIATED PROTEIN 18 (SAP18) ou TOPLESS (TPL). Ces protéines interagissent avec des membres du groupe Polycomb comme EMF1 (TPL) ou MS11 (SAP18) et sont associées à des activités déacétylases (**Figure I.9B**) (Kagale and Rozwadowski, 2011; Mehdi *et al.*, 2016; Bloomer *et al.*, 2020). En effet, le recrutement d'un domaine EAR au niveau d'un gène rapporteur est capable d'induire le dépôt de la marque H3K27me3 et le retrait des acétylations sur H3 (Baile *et al.*, 2021). Par ailleurs, des interactions physiques ont été identifiées entre des sous-unité du PRC2 et des HDAC comme par exemple HDAC6 et MS11 (Xu *et al.*, 2022).



**Figure I.9 : Coopération entre le PRC2 et des effecteurs chromatinien impliqués dans la répression des gènes. (A)** Les TELOMERE-REPEAT BINDING FACTOR 1, 2 et 3 reconnaissent les séquences Telobox et interagissent *in vivo* avec le PRC2 mais aussi avec la déméthylase d'H3K4me3, JMJ14. **(B)** Des facteurs de transcription tels que BPC1 ou AZF1 contiennent des domaines EAR (Ethylene-responsive element binding factor-associated Amphiphilic Repression) qui peuvent être reconnus par les protéines STRESS ASSOCIATED PROTEIN 18 (SAP18) ou TOPLESS (TPL). Ces protéines interagissent avec membres du groupe Polycomb comme EMF1 (TPL) ou MSI1 (SAP18) et sont associées à des activités déacétylases. Figures adaptées de la publication de Baile *et al.*, 2021.

## 9. Les protéines Trithorax, acteurs associés à l'activité de transcription

Tout comme les protéines PcG, les protéines TrxG jouent un rôle fondamental dans le développement des organismes en contrôlant spatialement et temporellement l'expression des gènes. Elles agissent de manière antagoniste aux protéines Polycomb en promouvant et en maintenant un état actif de la transcription.

De nombreuses protéines qui présentent des caractéristiques associées aux protéines Trithorax ont été identifiées ces dernières années (Ornelas-Ayala *et al.*, 2023). Elles forment un groupe plus hétérogène que les protéines PcG et incluent : des histones méthyltransférases (HMTs) qui catalysent le dépôt des marques H3K4me2/3 ou H3K36me2/3 ; des histones déméthylases qui catalysent le retrait des méthylations sur H3K27 ou H3K9 ; mais aussi des protéines issues de complexes de remodelage dépendants de l'ATP ou du complexe COMPASS-Like (COMPLex Proteins ASSociated with SET1).

Alors que plusieurs effecteurs TrxG ont été caractérisés individuellement chez les plantes, très peu d'informations concernant leur association sous forme de complexes est disponible à ce jour. Cependant, il a été montré que certaines de ces protéines pouvaient interagir physiquement afin de réguler positivement l'activité de transcription. Par exemple, les protéines qui constituent le cœur du complexe COMPASS-Like, WDR5a, ASHR2 RBL, peuvent interagir avec des HMTs (Jiang *et al.*, 2011) comme ATX1 (Ding *et al.*, 2012) mais aussi avec des déméthylases comme JMJ24 et JMJ28 (Shang *et al.*, 2021). Récemment, en plus de son activité catalytique, il a été montré que JMJ28 est capable de reconnaître le motif d'ADN 3'-TGCAG-5' via son domaine WRC, et semble participer au recrutement spécifique du complexe COMPASS-Like au niveau de certains loci (Xie *et al.*, 2023). De plus, les protéines qui constituent le cœur du complexe COMPASS-Like peuvent également interagir avec des complexes de remodelage de la chromatine pour réguler le dépôt de la marque H3K4me3 comme c'est le cas pour INO80 (Shang *et al.*, 2021).

Par conséquent, la grande diversité des protéines TrxG et leurs interactions avec le complexe COMPASS-like semblent indiquer l'existence de plusieurs complexes Trithorax dont la fonction pourrait être spécifique à certains gènes ou certains tissus.

Bien que de nombreux facteurs Trithorax aient été identifiés et leur rôle dans l'activité de la transcription caractérisé, les mécanismes moléculaires sous-jacents qui

permettent de passer d'un état répressif (médié par les protéines Polycomb) à actif, sont encore mal connus.

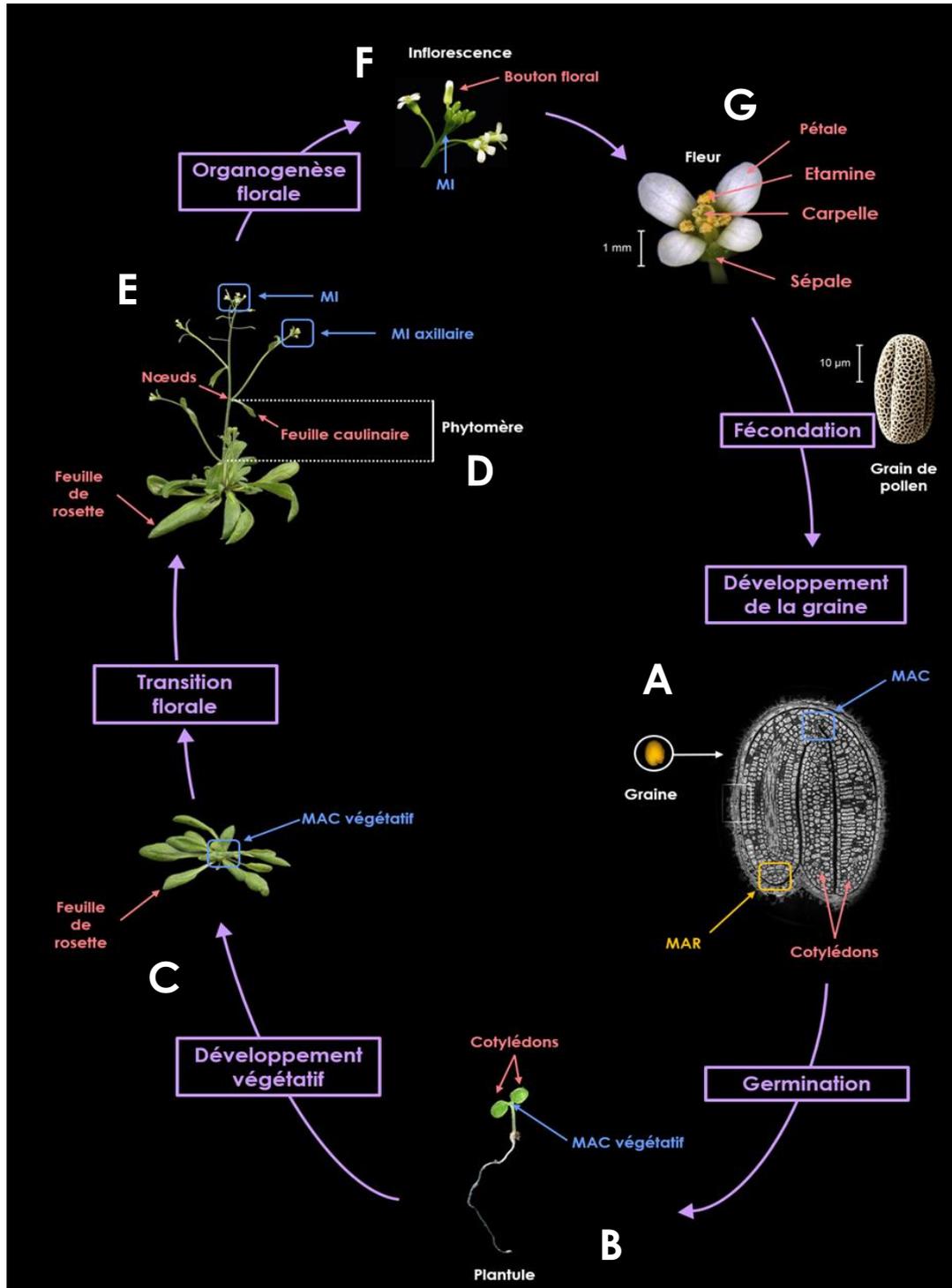
En plus de l'influence des marques H3K4me3 et H3K36me3 sur l'activité du PRC2, chez les mammifères, il a été démontré que le recrutement du complexe de remodelage mSWI/SNF induisait une éviction des complexes Polycomb par compétition avec ces derniers (Kadoch *et al.*, 2017). Par conséquent ce mécanisme pourrait jouer un rôle essentiel dans le basculement d'un état répressif à actif de la chromatine. Chez les plantes cependant, malgré la capacité de BRM et SYD, deux complexes de remodelage de la chromatine de la famille SWI/SNF, d'activer l'expression des gènes, ce mécanisme d'éviction n'a pas encore été identifié (Wu *et al.*, 2012; Li *et al.*, 2015; Xu *et al.*, 2016). En revanche, il a été montré que le facteur de transcription AGAMOUS pouvait évincer le PRC2 en se liant sur le promoteur du gène *KNUCKLES*, conduisant à une activation de ce gène par une perte progressive de la marque H3K27me3 (Sun *et al.*, 2014).

## C. Rôles de PcG et TrxG dans le développement des plantes

Les plantes, en tant qu'organismes sessiles, démontrent une remarquable plasticité phénotypique en ajustant leurs mécanismes de développement en fonction des conditions environnementales auxquelles elles sont confrontées. De plus, à la différence des animaux, elles ont la capacité de former de nouveaux organes en continu tout au long de leur cycle vie. Cette particularité s'explique par la production et la conservation de cellules indifférenciées dans des régions dynamiques appelées méristèmes (Steeves and Sussex, 1989). C'est lors des premières étapes de l'embryogenèse que se mettent progressivement en place au niveau des extrémités apicales de l'embryon, les méristèmes apicaux caulinaire et racinaire (MAC et MAR respectivement) (**Figure I.10 A**). Ils formeront par la suite les parties aériennes de la plante comme les feuilles ou les fleurs (pour le MAC) et souterraines comme les racines (pour le MAR).

Chez *Arabidopsis* et pendant les premières divisions cellulaires, le MAC va apparaître sous la forme d'un dôme entre les deux feuilles primaires (embryonnaires) appelées cotylédons mais ce n'est qu'après l'étape de germination, c'est à dire pendant la phase végétative, que l'organisation cellulaire du MAC sera complètement établie (Fletcher *et al.*, 1999) (**Figure I.10 B**). Lors de la phase végétative, le MAC produit un enchainement de structures appelés phytomères qui sont constitués d'un nœud sur lequel

est insérée une feuille (dite caulinaire) et à l'aisselle de laquelle un méristème axillaire va se développer (**Figure I.10 C et D**). Ces phytomères sont séparés par une tige correspondant à l'entre-nœud.



**Figure I.10 : Cycle du développement chez Arabidopsis. (A)** Développement de la graine. Le méristème apical caulinaire (MAC (encadré bleu)) et apical racinaire (MAR (encadré jaune)) se mettent en place au niveau des apex de l'embryon. **(B)** Après l'étape de germination, une plantule apparaît avec les deux feuilles embryonnaires (cotylédons) situées de part et d'autre du MAC. **(C)** Viens ensuite l'étape de

développement végétatif au cours de laquelle des feuilles de rosettes sont produites ainsi que des phytomères. **(D)** Ces phytomères sont constitués d'un nœud sur lequel est inséré une feuille dite caulinaire et à l'aisselle de laquelle un méristème axillaire va se développer. **(E)** Lors de la phase de transition florale le MAC et les méristèmes axillaires se développent en méristèmes d'inflorescences (MI). **(F)** Les MI vont alors produire sur leurs flancs des méristèmes floraux (ou boutons floraux) qui seront responsables de la formation des différents organes de la fleur. **(G)** Formation des différents organes de la fleur tel que les sépales, les pétales, les carpelles et les étamines. Enfin, la phase reproductive se termine par une étape de pollinisation suite à laquelle une graine sera formée et permettra de débiter un nouveau cycle.

Suite à cette phase végétative, des signaux endogènes tel que l'âge de la plante et exogènes comme l'augmentation de la durée du jour, vont entraîner une transition du développement du méristèmes apical caulinaire et axillaires vers une phase reproductive. Durant cette phase, que l'on appelle transition florale, l'activité de ces méristèmes évolue et un changement de leur identité s'opère où ils vont devenir des méristèmes d'inflorescences (MI) (**Figure I.10 E**). Ces MI vont alors produire sur leurs flancs des méristèmes floraux (MF) (**Figure I.10 F**) qui seront responsables de la formation des différents organes de la fleur : sépales, pétales, étamines et carpelles (**Figure I.10 G**) (Kaufmann, Pajoro and Angenent, 2010). Enfin, la phase reproductive se termine par une étape de pollinisation suite à laquelle une graine sera formée et permettra de débiter un nouveau cycle.

### 1. Rôle de Polycomb et Trithorax dans le maintien de l'homéostasie des méristèmes

Comme évoqué plus haut, la formation des différents organes de la plante dépend de l'activité des méristèmes. Dans le MAC, les cellules souches indifférenciées se divisent en continu et sont maintenues dans un état pluripotent au niveau d'une zone centrale par des signaux provenant de leurs cellules voisines. Au fur et à mesure des divisions cellulaires, certaines cellules filles, liées par leur paroi, vont se déplacer progressivement vers une zone périphérique du méristème où elles vont permettre la production des différents organes latéraux tandis que d'autres vont rester dans la zone centrale pour renouveler le réservoir de cellules souches. Cet équilibre entre la consommation des cellules produites pour la formation des différents organes et la production constante de cellules pluripotentes est finement régulé par l'activité de nombreux réseaux de gènes mais aussi par des hormones végétales.

La protéine WUSCHEL (WUS) est un facteur de transcription essentiel dans l'activité du MAC et le maintien de l'identité des cellules souche car une perte de fonction de cette protéine conduit à disparition progressive des méristèmes et à un arrêt de l'organogenèse (Laux *et al.*, 1996). Dans le MAC, la zone d'expression de WUS est limitée par une boucle de rétro-contrôle qui fait intervenir les protéines de la voie CLAVATA (CLV1/2/3) (Fletcher

*et al.*, 1999; Brand *et al.*, 2000). En effet, bien que WUS soit capable d'induire l'expression de CLV3 par liaison directe à son promoteur, ce dernier produit un peptide diffusible dont l'interaction avec les protéines CLV1 et CLV2 va permettre la répression de WUS, limitant ainsi son domaine d'expression au centre organisateur du méristème (Schoof *et al.*, 2000; Yadav *et al.*, 2011). En parallèle, une autre voie de signalisation qui implique la protéine SHOOT MERISTEMLESS (STM), facteur de transcription de la famille knox, participe également à la prévention de la différenciation des cellules souches dans le MAC (Yanai *et al.*, 2005).

Dans le MAC, Les gènes STM et WUS sont soumis à des régulations épigénétiques par les protéines TrxG et PcG. En effet, il a été montré que les complexes de remodelage de la chromatine SYD et BRM jouent un rôle important dans la promotion de l'expression de WUS (Farrona *et al.*, 2004; Kwon, Chen and Wagner, 2005). De plus, chez Arabidopsis, STM est exprimée exclusivement dans le MAC, et la répression de ce gène dans les ébauches d'organes est associée à la présence de la marque H3K27me3 déposée par PRC2 (Schubert *et al.*, 2006).

## **2. Rôle de Polycomb et Trithorax dans l'induction de la floraison et l'organogenèse florale**

Au cours de son développement, Arabidopsis va transiter d'une phase végétative vers une phase reproductive où le MAC et les méristèmes axillaires vont se convertir en méristèmes d'inflorescences pour former les méristèmes floraux. Cette transition peut se mettre en place par diverses voies de régulation génétique dont l'activation est médiée par des facteurs endogènes tels que l'âge de la plante et les phytohormones de la famille des gibbérellines ou exogènes tels que l'augmentation de la durée du jour (photopériode) et l'exposition prolongée aux températures froides (vernalisation) (Amasino, 2010; Srikanth and Schmid, 2011; Andrés and Coupland, 2012). Par la suite, les signaux induits par ces différentes voies de régulation vont agir sur l'activité des gènes LEAFY (LFY), SUPPRESSOR OF CONSTANS OVEREXPRESSION (SOC1) et FLOWERING LOCUS T (FT), qui sont responsables de l'activation des gènes de l'identité florale comme SEPALATA3 (SEP3), APETALA3 (AP3) et AGAMOUS (AG) (Parcy, 2005; Teper-Bamnolker and Samach, 2005; Wu *et al.*, 2012).

L'un des acteurs majeurs de la répression de la floraison est le gène FLOWERING LOCUS C (FLC) qui code pour un facteur de transcription à domaine MADS (MCM1, AGAMOUS, DEFICIENS, SRF) (Michaels and Amasino, 1999). Durant la phase végétative, l'expression de FLC est active, ce qui va avoir pour conséquence une répression de l'expression de FT et SOC1 (Lee *et al.*, 2000; Michaels *et al.*, 2005). L'activation de FLC est

contrôlée par la protéine FRIGIDA mais aussi par des facteurs TrxG tel que ATX1, ATXR7 et SDG8 qui vont déposer les marques H3K4me3 et H3K36me3 au niveau de ce locus (Xu *et al.*, 2008; Tamada *et al.*, 2009). A l'inverse, l'inhibition de ce facteur de transcription va promouvoir la transition florale et se fait par deux voies différentes : la voie autonome et la vernalisation. Dans le cas de la vernalisation, lorsqu'Arabidopsis est exposée à de longues périodes de froid (6 à 8 semaines), les niveaux d'expression de *FLC* diminuent et les gènes *FT* et *SOC1* qui induisent la floraison sont alors exprimés (Song *et al.*, 2012; Baulcombe and Dean, 2014). La répression de *FLC* par la vernalisation fait intervenir le complexe VRN2-PRC2 (SWN/CLF-VRN-FIE) et les protéines à domaine PHD, VERNALIZATION INSENSITIVE 3 (VIN3) et VERNALIZATION 5 (VRN5). VIN3 et VRN5 sont spécifiquement exprimées lorsque la température diminue et jouent un rôle important dans l'activité de VRN2-PRC2 (Sung and Amasino, 2004; Yang *et al.*, 2017). En effet, VRN5 est co-purifié avec le complexe SWN-VRN2-FIE-MSI1 et les mutants perte de fonction *vin3* et *vrn5* ne présentent pas d'accumulation d'H3K27me3 au niveau de *FLC* (Franco-Echevarria *et al.*, 2023). Une fois que l'expression de *FLC* est réduite par l'action du long ARN non codant (lncRNA) COOLAIR et du PRC1, le complexe PRC2 composé de SWN-VRN2-FIE-MSI1 et des protéines VIN3 et VRN5 va déposer la marque H3K27me3 au niveau d'un site de nucléation proche du promoteur de *FLC* (Yang *et al.*, 2017). Lors du retour à des températures plus élevées, CLF-PRC2 mais aussi LHP1 vont participer à la propagation de cette marque sur l'ensemble du gène *FLC* (Yang *et al.*, 2017). Cette répression sera maintenue au cours des prochaines divisions cellulaires mais sera réinitialisée par la méiose au cours de laquelle la déméthylase d'H3K27me3, ELF6, est impliquée (Crevillén *et al.*, 2014; Costa and Dean, 2019) .

CLF et EMF2 participent également au contrôle des gènes activateurs de la floraison et de l'organogenèse florale, puisque des mutants de ces protéines (*emf2*, *clf*) induisent une baisse de la marque H3K27me3 au niveau des gènes *SOC1*, *FT* et *AG* et présentent un phénotype de floraison précoce (Goodrich *et al.*, 1997; Yoshida *et al.*, 2001; Jiang *et al.*, 2008; Kim, Zhu and Sung, 2010; Shu *et al.*, 2019).

L'activation de l'expression de *LEAFY* (*LFY*) dans les méristèmes floraux, va conduire à l'activation de gènes responsables de l'identité des différents organes de la fleur tels que *APETALA3* (*AP3*) et *AGAMOUS* (*AG*) (Weigel and Meyerowitz, 1993; Pelaz *et al.*, 2000). Cette activation implique les complexes de remodelage de la chromatine SYD et BRM qui s'associent physiquement avec *LFY* et *SEP3* (Wu *et al.*, 2012). En plus du rôle des complexes de remodelage de la chromatine, la protéine TrxG ATX1 participe à l'expression d'autres gènes de l'identité florale tels que *PISTILATA* (*PI*) (Alvarez-Venegas *et al.*, 2003).

Nous avons vu dans ce chapitre que le développement d'*Arabidopsis* fait intervenir de nombreux gènes sous le contrôle des protéines PcG et TrxG. Cependant, alors qu'on sait comment sont établis et maintenus les états réprimés par PRC2, peu d'informations sont disponibles sur les mécanismes de commutation d'un état réprimé par PRC2 vers un état activé. Chez les plantes, le facteur ULT1 (ULTRAPETALA1) a été identifié comme pouvant participer à cette commutation d'état, fonctionnant en antagoniste de PRC2 (Carles and Fletcher, 2009b).

## D. ULTRAPETALA1 (ULT1), une protéine à domaine SAND

La protéine ULT1 est une protéine spécifique du règne végétal et particulièrement à la famille des Angiospermes. Elle est composée de 237 acides aminés et possède une masse moléculaire de 26,7 kilo Daltons (kDa). Avant ma thèse aucune structure d'ULT1 n'était connue mais des alignements de séquences multiples ont révélé qu'ULT1 contenait un domaine SAND (**S**p100, **A**IRE, **N**ucP41/75, **D**EAF-1) dans sa partie N-Terminale et une région riche en cystéines dans sa partie C-terminale (23 dans la séquence totale, 16 dans la région C-Terminale). Chez *Arabidopsis*, il existe également un paralogue d'ULT1 : ULT2 dont la séquence partage 81 % d'identité et 86 % de similarité (Carles *et al.*, 2005).

Le domaine SAND est un domaine de liaison à l'ADN (Bottomley *et al.*, 2001) conservé chez les plantes et les animaux (Carles *et al.*, 2005). Cependant, la fonction de ce domaine est encore mal comprise mais on le retrouve en combinaison avec d'autres domaines dans des protéines impliquées dans l'interaction avec la chromatine ou la régulation de la transcription. C'est le cas de la protéine AIRE (AUTO-IMMUNE REGULATOR) par exemple, qui est capable de se lier à H3K4 via un domaine PHD, d'activer la transcription de gènes cibles, et dont l'activité peut être modulée par la présence d'acétylation dans le domaine SAND (Org *et al.*, 2008; Saare *et al.*, 2012).

Il existe plusieurs structures tridimensionnelles de domaine SAND résolues par RMN et cristallographie (Bottomley *et al.*, 2001; Surdo *et al.*, 2003). Ces structures révèlent un repliement compact et consistent en un feuillet  $\beta$  antiparallèle torsadé et une série d'hélices  $\alpha$  qui exposent des résidus impliqués dans l'interaction avec l'ADN. Cependant, la séquence primaire du domaine SAND est faiblement conservée entre les différentes protéines de cette famille et leur similarité provient de leurs structures secondaires et tertiaires (Carles *et al.*, 2005). Il existe donc des différences structurales entre les domaines SAND, comme par exemple celui de GMEB1/2 dont la structure a révélé la présence d'un

atome de zinc pouvant jouer un rôle stabilisateur et qui est absent chez d'autres protéines comme Sp100b (Surdo *et al.*, 2003).

Une des caractéristiques connues du domaine SAND est qu'il est capable de se lier à l'ADN (Bottomley *et al.*, 2001). Cette propriété a été démontrée *in vitro* par des expériences d'Electro-Mobility Shift Assay (EMSA) et semble indiquer, pour certaines protéines, une spécificité de séquence. Par exemple, pour la protéine GMEB1/2, le domaine SAND montre une affinité pour le motif 5'-ACGT-3' et chez les plantes, l'homologue d'ULT1 chez le riz (OsULT1) a montré une affinité du domaine SAND pour un motif 5'-GAGAG-3' (Surdo *et al.*, 2003; Roy *et al.*, 2019).

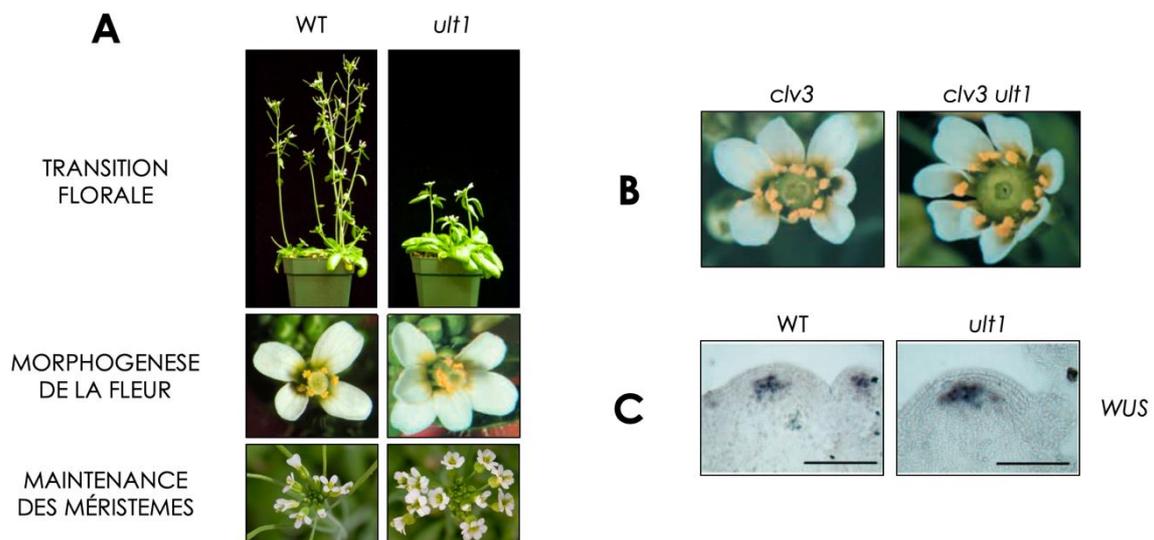
A la suite du domaine SAND, la séquence d'ULT1 révèle une région riche en cystéines dans sa partie C-Terminale (Carles *et al.*, 2005). Cette région partage des similarités avec les domaines B-box (BBX) mais ces domaines restent très mal caractérisés structurellement et fonctionnellement. Chez les eucaryotes les motifs B-Box sont associés à des motifs « doigt-de-zinc » et sont retrouvés dans plusieurs facteurs de transcription (Khanna *et al.*, 2009; Gangappa and Botto, 2014). De plus, il a été proposé que ces motifs B-box étaient impliqués dans des interactions protéine - protéine ou protéine - ARN (Borden, 1998; Torok and Etkin, 2001). ULT1 n'est pas identifié comme une protéine BBX mais la présence de nombreuses cystéines indique à la fois un rôle structural stabilisateur avec la coordination potentielle d'atomes de zinc et un rôle dans des interactions protéine - protéine ou protéine - acides nucléiques d'une manière similaire aux motifs B-box.

### 1. ULT1 est impliquée dans plusieurs processus développementaux, tous liés aux transitions d'identité cellulaire

ULT1 est une protéine impliquée dans de nombreux processus développementaux tels que la transition florale et la morphogenèse de la fleur (**Figure I.11 A**). En effet, à la différence d'une plante sauvage, les phénotypes perte-de-fonction d'ULT1 (*ult1*) révèlent une floraison tardive (jusqu'à 2 semaines supplémentaires) et un élargissement des méristèmes d'inflorescence (MI) et des méristèmes floraux (MF) qui vont conduire respectivement à la formation supplémentaire de fleurs et d'organes floraux comme les sépales et les pétales (Fletcher, 2001; Carles *et al.*, 2004) (**Figure I.11 A**). Il a été montré que cette différence phénotypique dans le mutant *ult1* était due à une augmentation du nombre de cellules souches méristématiques dans les inflorescences, indiquant donc qu'ULT1 joue un rôle dans le maintien de l'homéostasie des méristèmes (Carles *et al.*, 2004).

En effet, ULT1 agit comme un régulateur négatif de l'expression du gène *WUSCHEL* (*WUS*), qui code pour un facteur de transcription important dans le maintien de l'identité

des cellules souches au niveau d'une région spécifique du méristème (appelée centre organisateur) et dont l'activité est également régulée par les protéines de la voie CLAVATA (CLV1 CLV2 CLV3) (Clark *et al.*, 1996; Mayer *et al.*, 1998). Les phénotypes perte-de-fonction des protéines CLV (*clv*) sont similaires à ceux du mutant *ult1* (**Figure I.11 A et B**), conduisant à un élargissement de la zone d'expression de *WUS* (Carles *et al.*, 2004). Cette perte de régulation de *WUS* en absence des fonctions ULT1 (**Figure I.11 C**) et CLV se traduit donc par une accumulation de cellules souches dans les méristèmes et par une production d'organes surnuméraires. Cependant plusieurs observations suggèrent que les protéines CLV et ULT1 régulent la zone d'expression de *WUS* par des voies séparées. Par exemple, dans le mutant *clv*, le domaine d'expression de *WUS* s'étend dans une couche de cellules différente (couche supérieure) tandis que dans le mutant *ult1*, *WUS* s'étend de manière latérale sans modifier la spécificité de sa couche cellulaire (Carles *et al.*, 2004) (**Figure I.11 C**). De plus, le double mutant *clv ult1* montre des effets phénotypiques synergiques sur la taille des méristèmes d'inflorescence et floraux (**Figure I.11 B**). L'ensemble de ces données indiquent donc qu'ULT1 et CLV régulent l'homéostasie des méristèmes par des mécanismes distincts (Fletcher, 2001).



**Figure I.11 : Effets phénotypiques de la perte de fonction d'ULT1.** (A) A la différence des plantes sauvages (Wild Type, WT), Le mutant *ult1* dévoile un retard au niveau de la transition florale, une production surnuméraire d'organes floraux ainsi que davantage de fleurs. (B) Le double mutant *clv3 ult1* présente des effets synergiques sur la production d'organes floraux mais aussi sur la taille du méristème. (C) Dans les méristèmes inflorescenciels du mutant *ult1*, le domaine d'expression de *WUS* est élargi. Les images proviennent des publications de Fletcher, 2001; Carles *et al.*, 2005.

A la différence du méristème apical caulinaire qui produit des cellules souches en continu, les méristèmes floraux ont une croissance finie au cours de laquelle la production

de cellules souches s'arrête pour une production fixe de verticilles floraux correspondant aux couronnes de sépales, pétales, étamines et carpelles. L'arrêt de la prolifération des cellules souches dans les MFs est associé à une répression de *WUS* par les fonctions d'*AGAMOUS* (*AG*) et *ULT1* (Lenhard *et al.*, 2001; Carles and Fletcher, 2009a). Dans les mutants *ult1*, l'expression d'*AG* est retardée et par conséquent l'expression de *WUS* est prolongée dans le MF (Carles *et al.*, 2004). Par la régulation négative de *WUS*, *ULT1* joue donc aussi un rôle clé dans le déterminisme floral, en plus d'être nécessaire au maintien de l'homéostasie des méristèmes.

Le facteur de transcription *ULTRAPETALA INTERACTING FACTOR 1* (*UIF1*) interagit *in vivo* avec *ULT1* (*Y2H*, *BiFC*) et il a été proposé qu'il travaille avec *ULT1* pour réprimer *WUS* dans les MFs (Moreau *et al.*, 2016). En effet, *UIF1* est capable de reconnaître grâce à son domaine *Myb*, des séquences d'ADN spécifiques dans le promoteur de *WUS* (5'-GTAGATTCCT-3') mais aussi dans le premier intron d'*AG* (5'-AAGAATCTTT-3'). De plus, il a également été démontré qu'*UIF1* disposait, par son domaine *EAR*, d'une activité répressive sur la transcription génique.

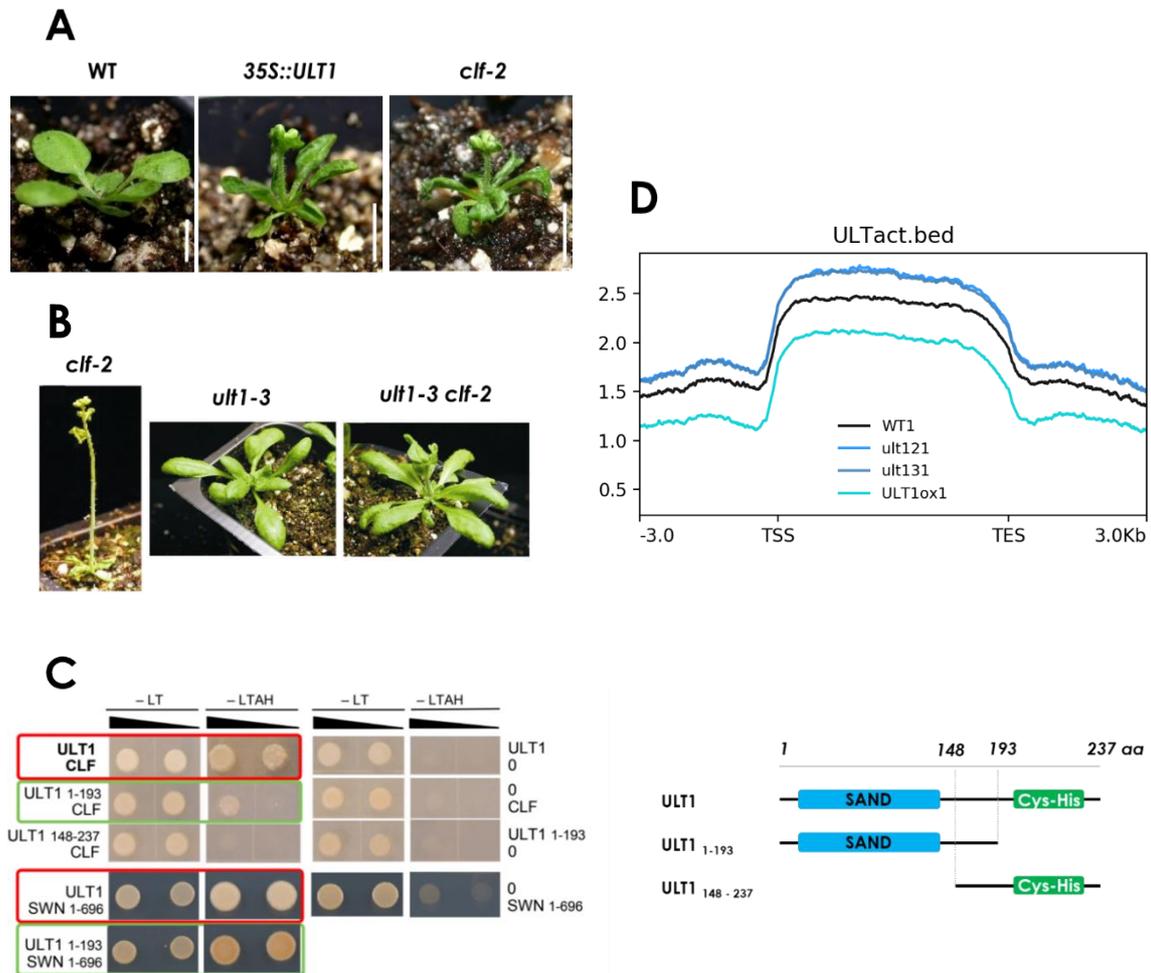
*ULT1* et *ULT2* possèdent une fonction dans la mise en place de la polarité apical-basale du gynécée (pistil) en interagissant physiquement avec les facteurs de transcriptions *KANADI 1* et *2* (*KAN1/2*). Cette coopération agit de manière dose-dépendante pour limiter l'expression du gène *SPATULA* (*SPT*). Cependant, *ULT1/2* et *KAN1/2* agissent de manière antagoniste pour établir l'axe de polarité abaxiale-adaxiale (Pires *et al.*, 2014).

Enfin, *ULT1* participe également à la régulation de l'activité du méristème apical racinaire (*MAR*) et contrôle la production *de novo* de racines (Ornelas-Ayala *et al.*, 2020; Tian *et al.*, 2022). Ces deux fonctions semblent faire intervenir la voie de signalisation par l'auxine, une hormone impliquée dans plusieurs processus développementaux dont la différenciation et l'élongation cellulaire. Dans le *MAR*, des concentrations importantes d'auxine au niveau du Centre Quiescent (*CQ*) permet de maintenir la quiescence des cellules souches de cette région. Cependant un défaut dans la signalisation de l'auxine c'est-à-dire un défaut dans l'activation de certains gènes dépendant de cette hormone, induit une différenciation des cellules souches (Sabatini *et al.*, 1999; Tian *et al.*, 2013). Le mutant *ult1* réduit la signalisation de l'auxine dans le *MAR* ce qui se traduit par une diminution de l'expression de certains gènes (*PIN*) codant pour des transporteurs d'efflux de cette hormone, ainsi que par une différenciation prématurée des cellules souches de la columelle (Ornelas-Ayala *et al.*, 2020). Le mutant *ult1* présente également une accélération de la croissance racinaire *de novo* à partir d'explants de cotylédons mais aussi une augmentation significative de l'expression des gènes *ASA1* et *YUC2* impliqués

dans la voie de biosynthèse de l'auxine (Tian *et al.*, 2022). Ces résultats indiquent donc qu'ULT1 intervient aussi dans la prévention de la surproduction de nouvelles racines, probablement en contrôlant la synthèse de l'auxine.

## 2. ULT1, une protéine antagoniste au PRC2

Une expression ectopique d'*ULT1* sous contrôle du promoteur constitutif fort 35S (*35S::ULT1*), conduit à des phénotypes similaires à ceux d'une surexpression d'AG (*35S::AG*) mais aussi à une perte de fonction de CURLY LEAF (*clf*) (Carles and Fletcher, 2009a) (**Figure I.12 A**). Ces phénotypes présentent une floraison précoce, de petites feuilles retournées vers le haut, ainsi que des transformations dites homéotiques (changement dans le plan d'organisation des différents organes d'un individu) au niveau des organes floraux (Goodrich *et al.*, 1997; Carles and Fletcher, 2009b). Il a été montré que tous les mutants *ult1* étaient capables de restaurer les phénotypes *clf* vers un phénotype sauvage ce qui indique qu'ULT1 agit génétiquement comme un antagoniste de la protéine du groupe Polycomb, CLF (**Figure I.12 B**). Une analyse au niveau moléculaire révèle qu'une surexpression d'*ULT1* ou une perte de fonction de CLF conduit à une expression ectopique des gènes homéotiques floraux AG et *APETALA3* (*AP3*) dans les tissus végétatifs. De plus des expériences de ChIP indiquent qu'ULT1 est capable de s'associer au gène AG au niveau de plusieurs régions régulatrices comme le promoteur et le second intron, mais aussi de modifier le niveau des marques épigénétiques H3K27me3 et H3K4me3 présentes sur ce locus. En effet, dans les plantes *35S::ULT1* et *clf*, la dé-répression d'AG est associée à une augmentation de la marque H3K4me3 et à une diminution de la marque H3K27me3. Par ailleurs, ULT1 semble interagir *in vivo* avec ATX1 (Y2H, BiFC), une méthyltransférase responsable du dépôt de la marque H3K4me3 et impliquée dans l'activation de l'expression d'AG (Alvarez-Venegas *et al.*, 2003). Malgré l'absence d'activité méthyltransférase, ULT1 agit comme un antagoniste de la protéine CLF pour activer l'expression de certains gènes et peut donc, par définition, s'apparenter à une protéine du groupe Trithorax. Récemment des expériences non publiées de notre équipe de Y2H, de BiFC et d'IP-MS (Immunoprécipitation couplée à la spectrométrie de masse) ont mis en évidence *in vivo* une interaction physique entre ULT1 et certaines protéines du groupe Polycomb comme les méthyltransférases CLF et SWN du complexe PRC2 (**Figure I.12 C**). Ces résultats indiquent qu'ULT1 pourrait impacter l'activité du PRC2 par une interaction physique.



**Figure I.12 : ULT1 agit comme un antagoniste du PRC2. (A)** Les plantes qui surexpriment ULT1 (*35S::ULT1*) ont des phénotypes similaires à une perte de fonction de CLF (*clf*). **(B)** La mutation *ult1-3* (lignée *knock-out* par insertion d'un ADN de transfert) restaure le phénotype perte de fonction *clf-2* vers un phénotype sauvage. Les photos proviennent de : Carles and Fletcher, 2009. **(C)** Test d'interaction en double-hybride de levure (Y2H) entre ULT1 et les méthyltransférases du PRC2 CLF et SWN (Emmanuel Thévenon et Caroline Thouly, non publié). La protéine ULT1 entière ou tronquée est exprimée en fusion au domaine activateur du facteur de transcription GAL4 alors que les protéines CLF et SWN entières ou tronquées sont exprimées en fusion avec le domaine de liaison à l'ADN de GAL4. Les souches de levures ayant intégré les deux vecteurs contenant les deux constructions sont sélectionnées sur un milieu dépourvu de leucine, tryptophane (contrôle). Les tests d'interaction sont ensuite réalisés sur un milieu dépourvu de leucine, tryptophane adénine et histidine. Lorsque les deux protéines interagissent *in vivo*, les deux domaines de fusion vont se rapprocher et former un facteur de transcription fonctionnel permettant l'expression de gène rapporteurs de synthèse de l'adénine et de l'histidine. La représentation schématique des différentes constructions d'ULT1 utilisées pour le test de Y2H est présentée dans la partie droite de la figure (C). **(D)** Métagène pour les cibles activées par ULT1. Les courbes représentent le niveau global de la marque H3K27me3 dans des plantules sauvages (WT), *ult1-2* (mutant isolé après mutagenèse par méthane sulfonate d'éthyle), *ult1-3* et *35S::ULT1* (Julia Engelhorn, *et al.*, non publié). Le TSS (Transcription Start Site) représente le premier nucléotide transcrit en ARN tandis que le TES (Transcription End Site) représente le dernier.

Chez le riz, lors d'un stress causé par le froid, l'homologue d'ULT1 (*OsULT1*) est capable de se lier au motif 5'-GAGAG-3' présent dans le promoteur du gène *OsDREB1*. De plus, le stress au froid s'accompagne d'une diminution de la marque H3K27me3 au locus *OsDREB1* (Roy *et al.*, 2019). Le motif GAGAG fait partie des *Polycomb Response*

*Elements* (PRE) reconnus par les protéines de la famille BASIC PENTACYSTEINE (BPC) pour le recrutement de PRC2 (Xiao *et al.*, 2017). Par conséquent, l'une des hypothèses sur le rôle antagoniste entre OsULT1 et le PRC2 serait une compétition pour la reconnaissance de ce motif.

ULT1 fonctionne également comme un antagoniste d'une autre protéine du groupe Polycomb : EMF1. En effet, une réduction de la fonction d'EMF1 dans les boutons floraux (induite par l'expression d'un ARN anti-sens *LFYasEMF1*) conduit à une diminution du niveau de marques répressives H3K27me3 chez certains gènes homéotiques floraux comme *AP3* et *SEPPALATA3 (SEP3)*, alors que le niveau de cette marque est restauré dans le double mutant *LFYasEMF1 ult1* (Pu *et al.*, 2013).

Cependant, des expériences réalisées sur le triple mutant *atx1 emf1 ult1* ont révélé chez des gènes cibles d'EMF1 associés au développement de la graine, une plus forte expression ectopique ainsi qu'un niveau d'H3K27me3 beaucoup moins important que dans les simples mutants *atx1*, *emf1* ou *ult1* (Xu *et al.*, 2018). ULT1 interagit *in vivo* avec ATX1 mais aussi avec EMF1 c'est pourquoi il a été proposé qu'ULT1 forme un complexe avec PRC2 à travers EMF1 pour maintenir la marque H3K27me3 ainsi qu'un état répressif de la chromatine au niveau de certains *loci*. Par conséquent, ces résultats mettent en évidence un rôle activateur mais aussi répresseur d'ULT1 pour certains gènes.

Une analyse par ChIP-seq de la distribution globale de la marque répressive H3K27me3 dans des plantules sauvages, *ult1-2*, *ult1-3* et *35S :: ULT1* a été réalisée au laboratoire (Engelhorn *et al.*, non publié). Ces résultats ont permis d'identifier 916 gènes pour lesquels ULT1 a un effet activateur, c'est-à-dire que les niveaux de la marque H3K27me3 sont augmentés chez les mutants perte de fonction, mais diminués dans lignée d'expression ectopique (**Figure I.12 D**). Ces expériences ont également dévoilé une liste de 1362 gènes pour lesquels ULT1 a un effet répresseur. Cependant, parmi ces 916 et 1362 gènes, la proportion de cibles liées par ULT1 à la chromatine (donc potentielles cibles directes pour la régulation) n'est pour le moment pas élucidée.

Les connaissances actuelles présentent donc ULT1 comme une protéine polyvalente, capable d'interagir avec diverses catégories de protéines comme des facteurs de transcription (UIF1), des protéines du groupe Trithorax (ATX1) et du groupe Polycomb (EMF1, CLF, SWN). L'interaction d'ULT1 avec ces effecteurs chromatiniens semble contrôler l'activation ou la répression de gènes cibles dans plusieurs processus développementaux, en modulant les niveaux de marques épigénétiques antagonistes : H3K27me3 et H3K4me3. Cependant, la diversité de ces interactions suggère que différents

mécanismes d'action doivent exister et rend complexe l'étude moléculaire des fonctions d'ULT1. Il est donc nécessaire de les étudier séparément, en lien avec chacun des interacteurs physiques identifiés. Mes travaux de thèse se sont concentrés sur la fonction d'ULT1 dans le contexte de son interaction avec les protéines du complexe PRC2.

## E. Objectifs de la thèse

Nous avons vu dans les sections précédentes, que le développement des organismes multicellulaires tel que la plante *Arabidopsis thaliana* dépend de facteurs épigénétiques qui vont contrôler la dynamique de la chromatine et réguler la transcription ou la répression de gènes spécifiques.

Parmi les acteurs impliqués dans ces régulations, deux groupes de protéines fonctionnent de manière antagoniste : le groupe Polycomb et le groupe Trithorax.

Au sein du groupe Polycomb, le complexe PRC2 contient une méthyltransférase (CLF ou SWN) qui catalyse la triméthylation de la lysine 27 sur l'histone H3 (H3K27me3), une marque épigénétique associées aux gènes à l'état réprimés. A l'inverse, les protéines du groupe Trithorax contrecarrent spécifiquement la répression des gènes médiées par le groupe Polycomb.

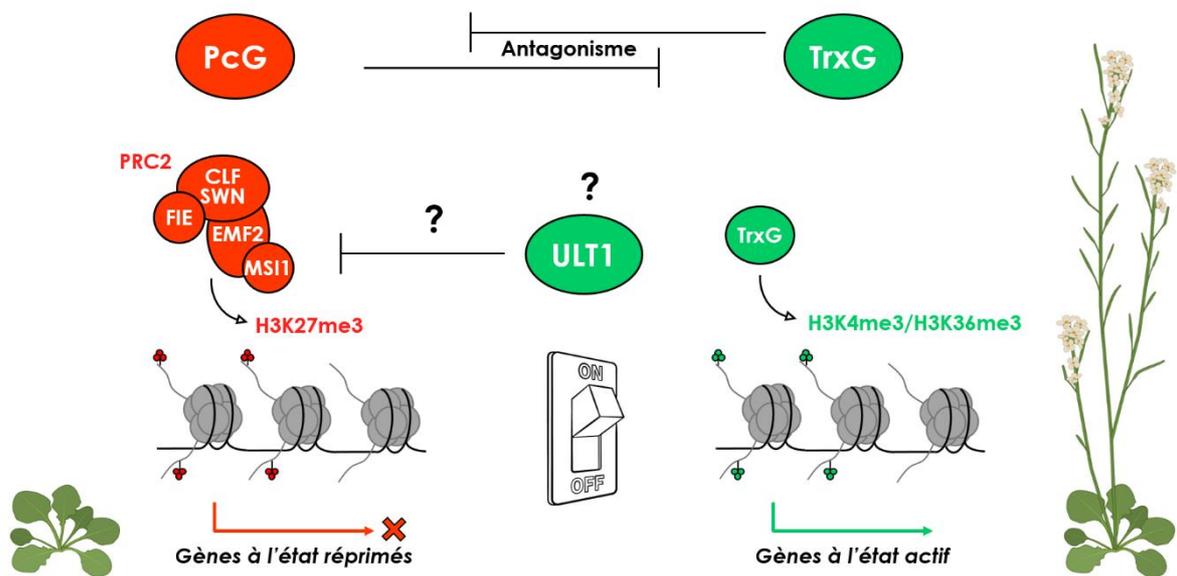
Alors que les protéines de chaque groupe sont de mieux en mieux identifiés et caractérisés, les mécanismes moléculaires qui permettent de commuter un gène d'un état réprimé par le PRC2 vers un état actif sont encore peu connu.

Par des approches de génétiques et de biologie moléculaire, notre laboratoire a identifié la protéine ULT1 comme un antagoniste de la sous unité enzymatique du PRC2, CLF. Alors que de nombreux tests d'interaction *in vivo* indique qu'ULT1 pourrait agir sur l'activité du PRC2 par une liaison directe avec CLF mais aussi SWN, les détails moléculaires de ces interactions ne sont toujours pas caractérisés. De plus connaître l'architecture de ces molécules qui ne sont pas connues à ce jour chez les plantes, pourrait nous permettre de mieux comprendre leur fonction.

**L'objectif de ma thèse est donc d'utiliser une approche par la biologie structurale et la biochimie afin d'étudier les mécanismes moléculaires par lesquels ULT1 contrecarre la fonction de PRC2 et agit comme un commutateur pour activer l'expression des gènes.**

Pour cela, mes deux objectifs principaux sont :

- **(1)** Produire la protéine ULT1 *in vitro*, caractériser ses propriétés biochimiques et résoudre sa structure par cristallographie aux rayons X.
- **(2)** Produire et caractériser le complexe PRC2 d'*Arabidopsis in vitro*. Résoudre sa structure avec ULT1 pour identifier les résidus impliqués dans leur interaction. Evaluer les effets de cette interaction sur les propriétés biophysiques du complexe PRC2.



**Figure I.13 : Schéma récapitulatif des objectifs de la thèse.** Les groupes PcG et TrxG fonctionnent de manière antagoniste et sont associés respectivement à des gènes à l'état réprimés ou actifs. Des approches de génétiques et de biologie moléculaire, la protéine ULT1 a été identifiée comme un antagoniste de la sous unité enzymatique du PRC2, CLF. Alors que de nombreux tests d'interactions *in vivo* indiquent qu'ULT1 pourrait agir sur l'activité du PRC2 par une liaison directe avec CLF mais aussi SWN, les détails moléculaires de ces interactions ne sont toujours pas caractérisés. L'objectif de la thèse est donc d'utiliser une approche par la biologie structurale et la biochimie afin d'étudier les mécanismes moléculaires par lesquels ULT1 contrecarre la fonction de PRC2 et agit comme un commutateur pour activer l'expression des gènes du développement chez *Arabidopsis*.

# **II. Matériels et Méthodes**

Avant mon arrivée au laboratoire, un premier travail de biologie moléculaire et de biochimie avait été initié par Emmanuel Thévenon, ingénieur dans l'équipe de François Parcy et le groupe de Christel Carles. Grâce à lui, plusieurs constructions contenant des séquences codantes d'ULT1 mais aussi de protéines du PRC2 avaient été réalisées et des essais préliminaires de production ont pu être réalisés.

Dans cette thèse, j'ai élaboré de nouvelles constructions pour l'étude structurale d'ULT1 et du complexe PRC2 basées sur d'autres approches comme la protéolyse limitée, l'analyse de structure et les alignements de séquences multiples.

## A. Production d'ULT1

### 1. Clonage des constructions ULT1<sub>1-237</sub>, ULT1<sub>1-121</sub> et ULT1<sub>128-237</sub>

L'ensemble des constructions d'ULT1 ont été exprimées en système bactérien dans le plasmide pETM41, à l'exception d'une construction réalisée pour la co-expression d'ULT1 avec le PRC2-CLF qui a été exprimée en cellules d'insectes. Le pETM41 est un vecteur d'expression qui code pour un gène de résistance à la kanamycine mais aussi pour une étiquette poly-histidines en fusion avec la Maltose Binding Protein (MBP) qui sera placée au niveau N-terminal de l'insert. Cette étiquette inclus également un site de coupure par la protéase TEV situé entre la MBP et la protéine d'intérêt. Le gène d'intérêt est sous le contrôle du promoteur du phage T7 et de l'opérateur lacO qui en présence de d'Isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) va induire l'expression du gène. En effet l'IPTG va capturer le répresseur lacI qui se dissocie de lacO, ce qui va permettre à l'ARN polymérase de pouvoir se fixer sur le promoteur du phage T7 et débiter la transcription.

Les séquences codantes pour les protéines d'intérêts ont été amplifiées par réaction de polymérase en chaîne (PCR) à partir du vecteur exprimant la protéine ULT1<sub>1-237</sub> déjà disponible au laboratoire dans le plasmide pETM40. Une fois amplifié par la polymérase Phusion (*New England Biolab* (NEB)), le produit de la PCR (insert) est déposé sur un gel à 0,8 % d'agarose pour vérifier sa taille puis est purifié grâce au kit NucleoSpin® Gel (Macherey-nagel). L'insert et le plasmide cible pETM41 sont ensuite digérés séparément par les enzymes de restriction NcoI et NotI afin obtenir des extrémités complémentaires nécessaire à leur ligature. Après digestion, le plasmide ainsi que l'insert sont de nouveau déposés sur gel d'agarose et purifié par le même kit que précédemment. Le plasmide et l'insert sont mélangés avec un ratio molaire

(plasmide/insert) de 1 : 4, puis sont fusionnés selon le protocole du mix de ligation commercial de chez Takara (DNA Ligation mix). Après transformation par choc thermique des cellules d'*E.coli* Top 10 avec le produit de ligation, quelques colonies obtenues sur boîte de sélection Luria Bertani (LB) agar + kanamycine (50 µg.mL<sup>-1</sup> final), ont été mises en culture à 37 ° C dans le but d'amplifier le vecteur et de le purifier par le kit NucleoSpin® Plasmid (Macherey-Nagel). La bonne insertion de la séquence dans le vecteur a été vérifiée par digestion enzymatique en utilisant les mêmes enzymes que pour le clonage, puis par séquençage (GATC Biotech/Eurofins). Une fois la séquence du plasmide validée, une transformation dans une souche d'*E.coli* Rosetta 2 (Novagen) a été réalisée. La souche Rosetta 2 contient le plasmide pRARE2, utile pour augmenter le taux de traduction des protéines eucaryotes dans *E. coli* par la synthèse de codons rares et code également pour une résistance au chloramphénicol.

Le **Tableau II.1** résume les différentes constructions d'ULT1 utilisées dans cette thèse.

**Tableau II.1 :** Liste des constructions de ULT1 utilisées dans cette thèse. Dans ce tableau sont représentés : le nom de la construction avec les numéros d'acides aminés définissant le début et la fin d'ULT1 ; le ou les noms alternatifs pour certaines constructions ; le plasmide utilisé (pETM40 ou pETM41) ; l'antibiotique de résistance ; le type d'étiquette de purification et sa position N-Terminale (N-Ter) ; la présence du site de coupure par la protéase TEV ainsi que les enzymes de restriction utilisées pour le clonage. La construction ULT1<sub>1-237</sub> en gris clair représente la construction réalisée par Emmanuel Thévenon avant mon arrivée au laboratoire.

Nom	Noms alternatifs	Plasmide	Étiquette	Résistance	Restriction
ULT1 <sub>1-237</sub>	-	pETM40	N-ter-MBP-TEV	Kanamycine	NcoI/NotI
ULT1 <sub>1-237</sub>	Full Length ou FL	pETM41	N-ter-HIS-MBP-TEV	Kanamycine	NcoI/NotI
ULT1 <sub>1-121</sub>	SAND ou N-Ter	pETM41	N-ter-HIS-MBP-TEV	Kanamycine	NcoI/NotI
ULT1 <sub>1-127</sub>	-	pETM41	N-ter-HIS-MBP-TEV	Kanamycine	NcoI/NotI
ULT1 <sub>13-121</sub>	-	pETM41	N-ter-HIS-MBP-TEV	Kanamycine	NcoI/NotI
ULT1 <sub>128-237</sub>	C-ter	pETM41	N-ter-HIS-MBP-TEV	Kanamycine	NcoI/NotI

## 2. Production de ULT1<sub>1-237</sub>, ULT1<sub>1-121</sub> et ULT1<sub>128-237</sub> en système bactérien

Pour produire la protéine d'intérêt, les cellules bactériennes transformées sont mises en culture à 37°C et sous agitation dans du milieu LB contenant de la kanamycine (50 µg.mL<sup>-1</sup>) et du chloramphénicol (34 µg.mL<sup>-1</sup>). A une densité optique (DO<sub>600 nm</sub>) de 0.8, la température d'incubation est abaissée à 18 – 20 °C et l'expression protéique est induite par ajout d'IPTG (1 mM). 1 mM de sulfate de zinc (ZnSO<sub>4</sub>) est également ajouté dans la culture au début de l'induction pour fournir un apport de zinc nécessaire à la bactérie pour le bon repliement d'ULT1. Après 16 heures d'induction, les cellules sont centrifugées à 5000 g, collectées et stockées à -20° C.

## 3. Lyse des culots bactériens

Les culots bactériens sont décongelés sur glace et repris dans 40 mL de tampon de lyse contenant 25 mM Hepes pH 7,8, 300 mM NaCl, 10 mM DTT et une tablette d'inhibiteur de protéases sans EDTA (Roche). Une fois re-suspendues, les cellules sont lysées par sonication (Vibra-Cell™ 75115) pendant 3 minutes avec des impulsions de 3 secondes à 65% de puissance, intercalées de 6 secondes d'arrêt. L'extrait brut est ensuite centrifugé pendant 30 minutes à 48 000 g et à 4°C pour séparer les fractions soluble et insoluble

## 4. Purification d'ULT1 FL et SAND

### a. Chromatographie d'affinité sur résine d'amylose

L'utilisation de l'étiquette MBP présente plusieurs avantages. Elle permet d'augmenter la solubilité de certaines protéines mais peut servir également comme étiquette de purification. En effet, la MBP a une forte affinité pour l'amylose ce qui est utile pour isoler notre protéine d'intérêt.

Une fois l'étape de lyse réalisée, le surnageant est collecté et déposé sur 5 mL de résine d'amylose (NEB) préalablement équilibrée avec un tampon de lavage contenant 25 mM Hepes pH 7,8, 300 mM NaCl, 10 mM DTT. Une fois chargée, la résine est lavée avec environ 200 mL du même tampon dans le but d'éliminer la plupart des protéines liées non spécifiquement à la colonne. La protéine est ensuite éluée avec 30 mL de tampon d'éluion. Ce tampon correspond au tampon de lavage dans lequel 10 mM de maltose a été ajouté. Le maltose interagit avec la MBP via le même site de fixation que l'amylose avec une plus grande affinité. C'est par cette compétition que la protéine est décrochée

de la résine lors de l'élution. Les échantillons collectés lors de cette étape sont déposés sur SDS-PAGE 15% et analysés avant la prochaine étape de purification.

### **b. Première chromatographie d'exclusion de taille**

La chromatographie d'exclusion de taille permet de séparer en solution les protéines en fonction de leur masse moléculaire et de leur encombrement stérique. Par conséquent, les protéines de masse moléculaire élevée ne passent pas dans les pores de la résine et sont donc éluées en premier contrairement aux protéines de faible masse moléculaire. Cette première chromatographie d'exclusion de taille vise à séparer les protéines ULT1 FL et SAND agrégées de la fraction soluble. L'éviction des agrégats permet d'obtenir une plus grande pureté de la protéine d'intérêt après clivage de l'étiquette par la protéase TEV.

Après l'étape de purification sur résine d'amylose, l'échantillon élué est concentré jusqu'à un volume de 5 mL par un système Amicon Ultra avec seuil de coupure de 30 kDa. La protéine concentrée est ensuite chargée sur une colonne HiLoad Superdex™ 200 16/600 (Cytiva) préalablement équilibrée avec un tampon contenant 25 mM Hepes pH 7,8, 200 mM NaCl, 10 mM DTT et purifiée par le système FPLC *NGC Quest Plus* de chez Biorad. Les fractions collectées sont déposées sur SDS-PAGE 15 % et analysées mais seules les fractions d'intérêt sont conservées pour être clivées par la protéase TEV avant les prochaines étapes de purification.

### **c. Clivage de l'étiquette par la protéase TEV**

Pour des essais de cristallisation ou des tests d'interaction, il peut s'avérer utile de séparer l'étiquette de notre protéine d'intérêt. Dans le cas de ULT1, toutes les constructions contiennent un site de coupure par la protéase TEV, caractérisé par la séquence peptidique suivante : ENLYFQG. La protéase TEV va reconnaître cette séquence et couper entre les acides aminés Q et G.

Les fractions purifiées d'ULT1 sont mélangées avec de la TEV et 10 mM DTT (déjà présent le tampon d'ULT1) puis incubées sur la nuit à 4°C. Pour 1 mg de ULT1, 0.01 mg de TEV est utilisé.

#### **d. Purification d'ULT1 FL et SAND sur colonne d'héparine**

Dans le but de séparer l'étiquette HIS-MBP mais aussi la TEV, une étape supplémentaire de purification est réalisée sur le système FPLC Biorad associé à une colonne d'héparine. L'héparine est un polymère qui fonctionne de la même manière qu'une résine échangeuse de cations par la présence de groupes sulfates anioniques. Elle est donc utilisée pour purifier les protéines se liant aux acides nucléiques, comme c'est le cas pour ULT1.

L'échantillon protéique après clivage par la TEV est dilué par 4 avec un tampon contenant uniquement 25 mM d'Hepes à pH 7.8 et 10 mM DTT. Cela permet de diminuer la concentration en NaCl de l'échantillon à 50 mM et ainsi augmenter la force ionique entre ULT1 et l'héparine. Une fois la colonne équilibrée avec un tampon contenant 25 mM d'Hepes à pH 7.8, 50 mM NaCl et 10 mM DTT, l'échantillon protéique est injecté dans le système. Durant cette étape, l'étiquette HIS-MBP ainsi que la protéase TEV vont passer à travers la colonne et seule ULT1 va être adsorbée sur la résine. Après avoir lavé la colonne avec le même tampon que précédemment, l'échantillon est élué par un gradient de tampon à forte concentration de NaCl (25 mM d'Hepes à pH 7.8, 1 M NaCl et 10 mM DTT). Les fractions éluées sont ensuite déposées sur SDS-PAGE 15 % pour être analysées.

#### **e. Deuxième chromatographie d'exclusion de taille**

Cette dernière étape de purification va permettre de vérifier l'état de la protéine clivée mais aussi d'adapter le tampon final en fonction de l'expérience à réaliser.

Les fractions collectées après la chromatographie d'affinité sur colonne d'héparine sont concentrées jusqu'à un volume de 500 µL par un système Amicon Ultra avec seuil de coupure de 10 kDa. L'échantillon concentré est ensuite chargé sur une colonne Superdex™ 200 increase 10/300 ou Superdex™ 75 10/300 (Cytiva) préalablement équilibrée avec un tampon contenant 25 mM Hepes pH 7.8, 200 mM NaCl, 1 mM TCEP et purifié par le système FPLC NGC Quest Plus de chez Biorad. Ce dernier tampon peut varier légèrement en fonction de l'expérience mais sera précisé dans les paragraphes suivants. Les fractions collectées sont déposés sur SDS-PAGE 15% et analysés avant d'être utilisés pour les expériences de caractérisation.

## 5. Purification d'ULT1 C-ter

### a. Chromatographie d'affinité sur résine d'amylose

Cette étape est réalisée de la même manière que pour ULT1 FL et SAND (voir p 68). La différence dans le cas de ULT1 C-ter est l'ajout de 10 mM d'imidazole dans le tampon de lavage contenant 25 mM Hepes pH 7.8, 300 mM NaCl et 10 mM DTT. Le tampon d'éluion est également légèrement modifié et contient 25 mM Hepes pH 7.8, 200 mM NaCl, 5 mM DTT, 10 mM maltose et 10 mM imidazole.

### b. Chromatographie d'affinité sur résine de Nickel

Après clivage de l'étiquette par la protéase TEV, comme décrit ci-dessus (voir p 69), une deuxième étape de purification par chromatographie d'affinité sur résine de nickel est réalisée. Cette étape va permettre de fixer les protéines fusionnées par une étiquette poly-histidine comme par exemple le produit du clivage HIS-MBP et la TEV. La protéine d'intérêt ULT1 C-ter, quant à elle, va passer à travers la colonne et sera collectée directement.

L'échantillon après clivage par la protéase TEV est déposé sur une résine de Ni-NTA préalablement équilibrée avec un tampon contenant 25 mM Hepes pH 7.8, 200 mM NaCl, 5 mM DTT et 10 mM imidazole. Pour cette étape, l'utilisation de la résine Ni-NTA est indispensable car elle présente une forte tolérance aux agents réducteurs (10 mM DTT maximum) nécessaires pour la purification de ULT1 C-ter. La protéine d'intérêt est collectée directement en sortie de colonne et sera analysée par SDS-PAGE 15 %. L'étiquette HIS-MBP et la TEV sont ensuite élués avec le tampon contenant 25 mM Hepes pH 7.8, 200 mM NaCl, 5 mM DTT et 300 mM imidazole.

La protéine est ensuite purifiée par filtration sur gel, comme décrit ci-dessous.

### c. Chromatographie d'exclusion de taille

Les fractions collectées après la chromatographie d'affinité sur colonne Ni-NTA sont concentrées jusqu'à un volume de 500 µL par un système Amicon Ultra avec seuil de coupure de 10 kDa. L'échantillon est ensuite chargé sur une colonne Superdex™ 200 increase 10/300 ou Superdex™ 75 10/300 (Cytiva) préalablement équilibrée avec un tampon contenant 25 mM Hepes pH 7.8, 200 mM NaCl, 1 mM TCEP et purifié par le système NGC Quest Plus de chez Biorad. Le dernier tampon peut différer en fonction de

l'expérience mais sera précisé dans les paragraphes suivants. Les fractions collectées sont déposées sur SDS-PAGE 15% et analysées avant d'être utilisées pour les expériences de caractérisation.

## **6. Identification des régions structurées d'ULT1 par protéolyse limitée par la trypsine**

La protéolyse limitée est une méthode efficace pour identifier les domaines structurés d'une protéine. Cette méthode se base sur l'utilisation d'une protéase qui va cliver notre protéine d'intérêt au niveau d'acides aminés spécifiques et accessibles. Les acides aminés protégés par l'organisation 3D de notre protéine d'intérêts ne seront donc pas impactés par l'enzyme et les domaines structurés pourront être identifiés par spectrométrie de masse.

Des échantillons de MBP-ULT1 et d'ULT1<sub>1-127</sub> purifiées sont mélangés avec de la trypsine dans un volume final de 60 µL puis incubés à température ambiante. Le ratio ULT1 : trypsine utilisée est de 1 : 500. Après 5, 15, 30 et 60 minutes, 10 µL de la réaction sont prélevés pour être analysés sur SDS-PAGE. Après 60 minutes de digestion, la réaction est stoppée par dénaturation de l'échantillon à 95°C et une fraction est envoyée à la plateforme de spectrométrie de masse de l'IBS pour y être analysée par ESI - MS (*Electrospray Ionisation – Mass Spectrometry*). Comme la trypsine ne clive qu'après les résidus lysine et arginine, les limites de séquence des fragments analysés peuvent être déterminées en comparant la masse moléculaire des fragments théoriques de clivage par la trypsine avec celle des fragments mesurés.

## **B. Production du complexe PRC2**

### **1. Principe de l'expression de protéines en cellules d'insecte**

Le système d'expression en cellules d'insecte est basé sur l'utilisation d'un baculovirus qui après infection de ces dernières va permettre la production de nos protéines d'intérêts.

Dans cette thèse nous avons utilisé la méthode MultiBac, développée par l'équipe de Imre Berger à l'EMBL afin de produire de larges complexes multi-protéiques par la combinaison de plusieurs vecteurs d'expressions (Fitzgerald *et al.*, 2006). Ces vecteurs permettent le clonage de séquences d'intérêts au niveau d'un ou plusieurs sites de

clonage multiple (MCS) qui seront sous le contrôle du promoteur viral de la polyhédrine ou de la protéine p10 lors de la phase tardive du baculovirus.

L'ensemble des vecteurs MultiBac codent également pour un gène de résistance à un antibiotique spécifique comme par exemple la kanamycine, l'ampicilline et le chloramphénicol pour les vecteurs pIDK, pFL et pUCDM, respectivement.

Une des particularités de ces vecteurs est la présence du site *LoxP* dans leur séquence qui permet leur assemblage par l'intermédiaire de la recombinase Cre. On obtient ainsi un seul vecteur codant pour plusieurs protéines ce qui est idéal pour l'expression et l'étude de complexes macromoléculaires.

Parmi les vecteurs MultiBac, on trouve deux catégories : les accepteurs (pFL, pKL) et les donneurs (pUCDM, pIDK). A la différence des donneurs, les accepteurs possèdent dans leur séquence un site de transposition Tn7 nécessaire à l'intégration des vecteurs MultiBac dans le génome du baculovirus (appelé bacmide). Les donneurs quant à eux, possèdent une origine de réplication bactérienne dérivée du plasmide R6K $\gamma$  rendant sa propagation dépendante d'un hôte exprimant le gène *pir* et permettant ainsi de sélectionner (en plus de la sélection par l'antibiotique de résistance), les vecteurs recombinés des vecteurs non-recombinés entre donneurs et accepteurs.

Une fois les vecteurs recombinés et sélectionnés, le vecteur final codant pour l'ensemble des protéines d'intérêt va être inséré par transformation dans une souche d'*E. coli* EMBACY. Cette souche possède un exemplaire du bacmide qui va pouvoir intégrer le nouveau vecteur par recombinaison grâce à une transposase et au site Tn7 présent dans les deux séquences. Le site Tn7 du bacmide est situé dans le gène *lacZ* codant pour une  $\beta$ -galactosidase capable d'hydrolyser le Bluo-Gal (ThermoFisher) présent dans les boîtes de sélection. Lorsque le Bluo-Gal est hydrolysé, un pigment bleu est produit ce qui permet d'identifier les bactéries ayant intégré le bacmide. Les bactéries ayant intégré le bacmide seront donc de couleurs blanches tandis que celles qui ne l'auront pas intégré deviendront bleues.

Le génome du baculovirus code également pour le gène de la *Yellow Fluorescent Protein* (YFP) sous contrôle du promoteur de la polyhédrine et qui servira d'indicateur de la phase d'expression du complexe protéique. De manière générale, le maximum de fluorescence de la YFP est obtenu 3 à 4 jours après l'arrêt de la division des cellules d'insectes.

La première étape de préparation du baculovirus commence par la transfection de  $1,5 \cdot 10^6$  cellules Sf21 (issues d'ovaires de *Spodoptera frugiperda*) avec le bacmide par le réactif XtremeGene Transfection Reagent (Roche). 48 à 60 heures après la transfection,

le milieu de culture contenant le virus de première génération appelé  $V_0$  est collecté (environ 6 mL) mais la quantité de particules virales étant faible, une étape d'amplification doit être réalisée. Pour cela, 25 mL de Sf21 à  $0,5 \cdot 10^6$  cellules.mL<sup>-1</sup> vont être infectés par le  $V_0$  puis dilués par deux après 48 h afin d'éviter une trop forte confluence des cellules. Enfin, 4 jours après l'arrêt de la division cellulaire, les 50 mL du milieu de culture sont collectés pour former le  $V_1$  qui servira à l'infection d'un plus grand nombre de cellules lors de la production de notre complexe.

## 2. Clonage des différentes constructions de PRC2

Les plasmides contenant les séquences codantes des protéines du PRC2 (à l'exception de SWN dans cette thèse) ont été obtenus de Justin Goodrich (Université d'Edimbourg), et ont servi de matrices pour la conception de nouveaux vecteurs d'expressions. La séquence codante de SWN, quant à elle, a été obtenue grâce à l'*Arabidopsis Biological Resource Center* (ref : DKLAT4G02020)

### a. Clonage des gènes codants pour les sous-unités du complexe PRC2

Le clonage des gènes codants pour les sous-unités du PRC2 a été effectué comme ci-dessus (voir p 65).

Les séquences complètes de CLF (1-902) et SWN (1-856) ont été clonées avec une étiquette poly-histidine en position N-Terminale entre les sites de restriction RsrII et XbaI du vecteur accepteur pFL.

La séquence complète de FIE (1-369) a été clonée en fusion avec une étiquette STREP N-Terminale (aa WSHPQFEK) entre les sites de restriction RsrII et XbaI du vecteur donneur pUCDM. La souche d'*E. coli* C8 a été utilisée et code le gène *pir* nécessaire à la multiplication des vecteurs donneurs dans le système Multibac.

La séquence codante du domaine VEFS de EMF2 (aa 467-626) a été clonée avec une étiquette FLAG C-terminale (aa DYKDDDDK) entre les sites de restriction XhoI et KpnI du vecteur donneur pIDK. La souche *E. coli* C8 a également été utilisée pour le clonage dans ce vecteur donneur.

Afin de co-exprimer ULT1 sans étiquette avec PRC2, sa séquence a été clonée entre les sites de restriction XmaI et NsiI dans un vecteur pFL contenant déjà la protéine CLF.

L'ensemble des constructions réalisées pour le PRC2 est résumé dans le **Tableau II.2**.

**Tableau II.2** Liste des constructions PRC2 utilisées dans cette thèse. Dans ce tableau sont représentés : le nom du vecteur MultiBac utilisé, l'antibiotique de résistance de chaque vecteur, le nom de la protéine exprimée avec les numéros d'acides aminés définissant le début et la fin de la construction ; le type d'étiquette de purification et sa position (N-Terminale ou C-terminale) ainsi que les enzymes de restriction utilisées pour le clonage. Pour chaque protéine il est indiqué le site de clonage (1 ou 2) et le type de promoteur viral associé (Polyhédrine ou P10).

<b>Vecteur</b>	<b>Résistance</b>	<b>Promoteur Polyhédrine (Site 1)</b>	<b>Promoteur P10 (Site 2)</b>	<b>Étiquette</b>	<b>Restriction site 1</b>	<b>Restriction site 2</b>
pFL	Ampicilline/ Gentamycine	CLF <sub>1-902</sub>	-	N-ter HIS-CLF	RsrII/XbaI	-
pFL	Ampicilline/ Gentamycine	CLF <sub>1-902</sub>	ULT1 <sub>1-237</sub>	N-ter HIS-CLF	RsrII/XbaI	XmaI/NsiI
pFL	Ampicilline/ Gentamycine	SWN <sub>1-856</sub>	-	N-ter HIS-SWN	RsrII/XbaI	-
pUCDM	Chloramphénicol	FIE <sub>1-369</sub>	-	N-ter STREP-FIE	RsrII/XbaI	-
pIDK	Kanamycine	-	EMF2 <sub>467-626</sub> (VEFS)	C-ter EMF2-FLAG	-	XhoI/KpnI

### 3. Recombinaison des vecteurs MultiBac par le système Cre-loxP

Pour permettre la co-expression des différentes sous-unités du PRC2, les vecteurs MultiBac dans lesquels ont été cloné nos protéines d'intérêts doivent être assemblés avant d'être intégré dans le génome du baculovirus. Cet assemblage est réalisé par translocation du site *LoxP* présent sur chaque vecteur grâce à la recombinase Cre. Pour un meilleur contrôle des assemblages, les vecteurs sont assemblés un par un et vérifiés avant chaque nouvelle recombinaison.

Dans un volume final de 10 µL, environ 500 ng de vecteur A (accepteur) et 500 ng de vecteur B (donneur 1) sont mélangés avec le *Cre reaction buffer* (NEB) et la recombinase Cre (NEB). Après 30 minutes à 37°C, une souche *E.Coli* Top 10 va être transformée avec le produit de réaction puis déposée sur boîte de sélection contenant les antibiotiques de résistances des vecteurs A et B. Les colonies obtenues sont mises en culture à 37 °C afin d'amplifier le vecteur recombiné et de le purifier par le kit NucleoSpin® Plasmid. Dans le but de vérifier la bonne recombinaison des vecteurs A et B, la séquence théorique du vecteur recombiné (obtenu par le logiciel : CreRec (EMBL)) va servir à établir un profil virtuel de digestion, en fonction d'enzymes de restriction spécifiques (choisies grâce au site [nbcutter.com](http://nbcutter.com)). Ce profil virtuel va ensuite être comparé au profil de digestion expérimental réalisé sur gel d'agarose. Pour consolider le résultat du profil de digestion, une PCR pour chaque séquence d'intérêt clonée est également réalisée. Une fois ces deux contrôles effectués, une nouvelle étape de recombinaison est réalisée entre le vecteur recombiné et le vecteur C (donneur 2). Le nouveau produit de recombinaison est alors intégré dans une souche *E.coli* Top 10. Les différents complexes réalisés sont les suivants :

#### **Complexes PRC2 ternaires :**

- CLF<sub>1-902</sub> + FIE<sub>1-369</sub> + EMF<sub>2467-626</sub> (VEFS)
- CLF<sub>1-902</sub> + FIE<sub>1-369</sub> + EMF<sub>2467-626</sub> (VEFS) + ULT<sub>1-237</sub>
- SWN<sub>1-856</sub> + FIE<sub>1-369</sub> + EMF<sub>2467-626</sub> (VEFS)



## 5. Lyse des cellules d'insectes

Les culots cellulaires sont décongelés sur glace et repris dans 40 mL de tampon de lyse contenant 25 mM Hepes pH 7,8, 300 mM NaCl, 1 mM TCEP, 2mM MgCl<sub>2</sub>, 10 mM Imidazole et une tablette d'inhibiteur de protéases sans EDTA (Roche). Une fois re-suspendues, les cellules sont lysées par sonication pendant 3 minutes avec des impulsions de 3 secondes à 60% de puissance (Vibra-Cell™ 751 15), intercalées de 6 secondes d'arrêt. L'extrait brut est ensuite centrifugé pendant 30 minutes à 48 000 g et à 4°C.

## 6. Purification de PRC2

### a. Chromatographie d'affinité sur résine de nickel

Une fois l'étape de lyse réalisée, le surnageant est collecté et déposé sur 5 mL de résine Ni-NTA (NEB) préalablement équilibrée avec un tampon de lavage contenant 25 mM Hepes pH 7,8, 300 mM NaCl, 1 mM TCEP, 2 mM MgCl<sub>2</sub>, 10 mM Imidazole. Une fois chargée, la résine est lavée avec environ 50 mL du même tampon dans le but d'éliminer la plupart des protéines liées non spécifiquement à la colonne. Une seconde étape de lavage est réalisée avec un tampon proche du précédent mais modifié à 1M NaCl afin de réduire la présence d'acide nucléique pouvant se lier à notre complexe. Le PRC2 est ensuite élué avec 10 mL de tampon d'éluion et correspond au tampon de lavage dans lequel 300 mM d'imidazole a été ajouté. Les échantillons collectés lors de cette étape sont déposés sur SDS-PAGE 15% et analysés avant la prochaine étape de purification.

### b. Chromatographie d'affinité sur résine Strep-Tactin XT (IBA)

Dans le but d'obtenir notre complexe ternaire avec un ratio protéique de 1 :1 :1 (CLF/SWN : FIE : EMF2) une seconde étape de purification sur chromatographie d'affinité est réalisée. Cette étape a pour but d'éliminer les excès de protéines produites ne participant pas à la formation du PRC2 en purifiant le complexe par l'intermédiaire d'une autre étiquette. La première étape de purification a été réalisé sur résine Ni-NTA grâce à l'étiquette poly-histidine présente sur CLF ou SWN mais pour cette deuxième étape nous avons choisi d'utiliser l'étiquette STREP présente sur FIE. L'étiquette STREP est caractérisée par une séquence polypeptidique de huit acides aminés (WSHPQFEK), spécialement conçue pour se lier avec une forte affinité à la résine Strep-Tactin XT. Cette résine est dérivée de la streptavidine et va lier l'étiquette STREP au niveau du site de fixation de la biotine.

Après purification du PRC2 sur résine Ni-NTA, la solution protéique est déposée sur une résine Strep-Tactin XT préalablement équilibrée avec un tampon de lavage contenant 25 mM Hepes pH 7,8, 200 mM NaCl, 1 mM TCEP et 2 mM MgCl<sub>2</sub>. Une fois l'étape de lavage réalisée, l'échantillon est élué par ajout de 50 mM de biotine dans le tampon puis est analysé par SDS-PAGE.

### c. Chromatographie d'exclusion de taille

Les fractions collectées après la chromatographie d'affinité sur résine Strep-Tactin XT sont concentrées jusqu'à un volume de 500 µL par un système Amicon Ultra avec seuil de coupure de 50 kDa. L'échantillon est ensuite chargé sur une colonne Superdex™ 200 increase 10/300 préalablement équilibrée avec un tampon contenant 25 mM Hepes pH 7,8, 100 mM NaCl, 1 mM TCEP, 2% glycérol et purifié par le système FPLC NGC Quest Plus de chez Biorad. Le dernier tampon peut différer en fonction de l'expérience mais sera précisé dans les paragraphes suivants. Les fractions collectées sont déposés sur SDS-PAGE 15% et analysés avant d'être utilisés pour les expériences de caractérisation.

## C. Caractérisation biochimique et structurale d'ULT1

### 1. SEC-MALLS

Le SEC-MALLS (*Size Exclusion Chromatography Multi-Angles Laser Light Scattering*) est une technique qui permet de calculer la masse moléculaire d'un objet en solution en mesurant la diffusion statique de la lumière à différents angles. L'intensité de la lumière détectée est directement proportionnelle à la masse moléculaire moyenne de notre protéine (ou complexe) d'intérêt ainsi qu'à sa concentration. Le système de mesure de diffusion statique de la lumière à différents angles (MALLS) est associé à un réfractomètre mais aussi à une chromatographie d'exclusion de taille afin de séparer les différents constituants d'un échantillon si ce dernier n'est pas homogène.

Les expériences de SEC-MALLS ont été réalisées sur la plateforme ISBG (*Integrated Structural Biology Grenoble*) avec l'aide de Caroline Mas. Les échantillons ont été concentrés à environ 3 mg.mL<sup>-1</sup> puis injectés sur une colonne Superdex S75 10/300 GL préalablement équilibrée avec le tampon de purification. Le débit utilisé est de 0,5 mL.min<sup>-1</sup>.

## 2. Essais de cristallisation des domaines N-Terminal et C-Terminal d'ULT1

Dans le but de résoudre la structure 3D d'une protéine à l'échelle atomique, l'une des techniques les plus couramment utilisée est la cristallographie au rayon X. Cette technique nécessite la formation d'un arrangement compact, ordonné et périodique de notre macromolécule d'intérêt en forme d'un cristal. Cet arrangement cristallin sert d'amplification du signal de diffraction des rayon X nécessaire à la résolution de la structure de notre objet. La qualité des données de diffractions collectées dépend en partie de la qualité d'organisation de notre macromolécule dans le cristal c'est pourquoi il est important d'avoir un échantillon initial soluble et homogène. La cristallisation consiste à trouver les conditions nécessaires pour que notre macromolécule en solution passe d'une phase liquide à solide. Généralement, il suffit de varier les propriétés physico-chimiques de notre solution comme le pH ou la force ionique afin de réduire la solubilité de notre molécule et induire une nucléation de celle-ci. La concentration en protéine de notre échantillon est également un facteur important de cristallisation.

Les domaines N-Terminal et C-Terminal d'ULT1 ont été purifiés dans un tampon contenant 50 mM Tris pH 7, 100 mM NaCl et 1mM TCEP puis ont été concentré à 5, 10 et 15 mg.mL<sup>-1</sup>. Une fois préparés, ces échantillons ont été envoyés à la plateforme HTX de l'EMBL Grenoble afin de cribler différentes conditions de cristallisation. Ces cribles proviennent de 6 kits commerciaux de 96 conditions chacune : The PEGs I (Qiagen) ; Salt-Grid (Hampton) ; The PACT (MD) ; The JCMG (MD) ; Wizard 1 & 2 (Rigaku) et The Classic Suite (Qiagen).

Des gouttes ont été réalisées selon la technique de diffusion en goutte assise en mélangeant 100 nL de protéines et 100 nL de solution de cristallisation. Une fois réalisées, le boites contenant nos essais de cristallisation, ont été scellée et placée à 4 °C ou 20 °C.

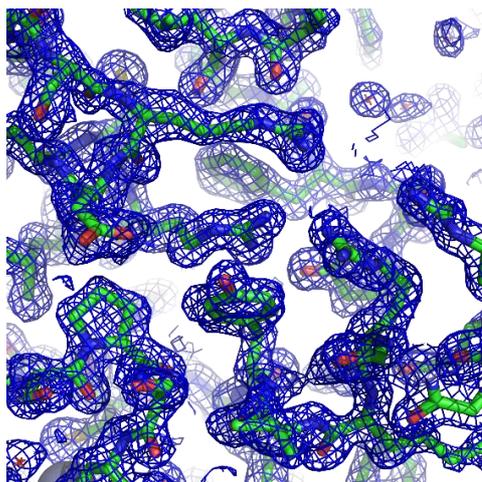
Une fois que la condition de cristallisation d'ULT1<sub>128-237</sub> a été identifiée, les cristaux obtenus ont été reproduits manuellement en gouttes suspendues (1+1 µl) dans des boites 24 puits placées à 4 °C.

Enfin, pour la cryoprotection, ces cristaux ont été plongés dans la solution de cristallisation supplémentée à 30 % de glycérol avant d'être congelés dans l'azote liquide. Ils ont été par la suite testés pour diffraction sur ligne ID23-1 de l'ESRF.

### 3. Phasage et affinement du modèle cristallin

Les données de diffraction ont été recueillies sur la ligne de lumière ID23-1 de l'*European Synchrotron Radiation Facility* (ESRF, Grenoble, France) et traitées à l'aide du logiciel XDS (Kabsch, 2010). Les cristaux d'ULT1<sub>128-237</sub> diffractent jusqu'à une résolution de 1,5 Å et appartiennent au groupe d'espace P2<sub>1</sub> avec une maille de dimensions  $a=49.6$  Å,  $b=56.8$  Å,  $c=84.7$  Å et  $\beta=93.4^\circ$ . La structure a été résolue à l'aide d'un phasage expérimental en utilisant le signal anomal du zinc présent dans la protéine (Zn-SAD, *Single-wavelength anomalous dispersion*). Les données ont été collectées au bord d'absorption de cet atome c'est-à-dire à une longueur d'onde des rayons X de 1,26515 Å. Ensuite, le logiciel SHELXD (Usón and Sheldrick, 1999) a permis de trouver 16 atomes de Zn, correspondant à quatre molécules d'ULT1<sub>128-237</sub> par unité asymétrique contenant chacune quatre atomes de zinc. Les positions de ces atomes ont été utilisées pour le phasage et la construction automatique d'un premier modèle dans SHELXE (Thorn and Sheldrick, 2013). Le modèle final a ensuite été construit manuellement dans COOT (Emsley and Cowtan, 2004) puis affiné par les logiciels REFMAC5 (Murshudov, Vagin and Dodson, 1997) et Phenix jusqu'à un Rfactor final de 25,3 % et un Rfree de 27,79 % (**tableau 4**) avec tous les résidus se trouvant dans les régions autorisées (97,5 % dans les régions favorisées) par le diagramme de Ramachandran, tel qu'il a été analysé par MOLPROBITY (V. B. Chen *et al.*, 2010).

Une partie représentative de la carte de densité électronique 2Fo - Fc couvrant une région du domaine C-Terminal d'ULT1 est présentée dans la **figure II.2**



**Figure II.2 :** Représentation de la carte de densité électronique 2Fo-Fc couvrant une partie du domaine C-terminal d'ULT1 et délimitée à 1,2  $\sigma$ .

**Tableau II.3** : Statistiques de collecte et d'affinement des données

<b>ULT1-Cter</b>	
<b>Data collection</b>	
<b>Space group</b>	$P2_1$
<b>Cell dimensions</b>	
<b><math>a, b, c</math> (Å)</b>	49.6, 56.8, 84.7
<b><math>\alpha \beta \gamma</math> (°)</b>	90, 93.4, 90
<b>Resolution (Å)</b>	84.6-1.6 (1.7-1.6) <sup>a</sup>
<b><math>R_{\text{merge}}</math> (%)</b>	16.5 (114.4)
<b><math>I / \sigma I</math></b>	4.04 (1.3)
<b>CC <math>\frac{1}{2}</math> (%)</b>	98.9 (53.3)
<b>Completeness (%)</b>	94.9 (93.6)
<b>Redundancy</b>	3.4 (2.9)
<b>Refinement</b>	
<b>Resolution (Å)</b>	49.5-1.6
<b>No. reflections</b>	61625
<b><math>R_{\text{work}} / R_{\text{free}}</math></b>	0.253/0.277
<b>B-factors</b>	22.5
<b>R.m.s. deviations</b>	
<b>Bond lengths (Å)</b>	0.007
<b>Bond angles (°)</b>	0.880

<sup>a</sup> Les valeurs entre parenthèses correspondent à la résolution la plus élevée.

#### 4. Caractérisation de l'interaction entre ULT1 C-Ter et le peptide HsH3.1<sub>26-44</sub>

La matrice de peptides (<https://www.activemotif.com/catalog/668/modified-histone-peptide-array>) a été bloquée pendant 1h dans une solution de TBS Tween (TTBS) contenant 5% de lait ou de BSA. Celle-ci a ensuite été lavée 3 x 5 minutes à l'aide de TTBS puis a été incubée pendant 1 H à température ambiante avec 10 mL de HIS-MBP-ULT1<sub>128-237</sub> à 2  $\mu$ M (50 mM Tris pH 7,5, 200 mM NaCl, 10mM DTT). Suite à cette étape, trois lavages de 5 minutes ont de nouveau été effectués avec du TTBS. La révélation de la liaison d'HIS-MBP-ULT1<sub>128-237</sub>, a été effectuée par un anticorps anti-HIS (A7058 SigmaAldrich) directement conjugué à la peroxydase de raifort (HRP) et utilisé à 1/2000 sur la nuit à 4°C. Le signal de liaison de l'anticorps a été réalisée avec un kit ECL de chez Bio Rad (Clarity™ Western ECL Substrate) et visualisée sur un scanner ImageQuant 800 (Amersham GE) en utilisant la fonction de chimiluminescence.

## 5. Pull-down entre le domaine C-Terminal d'ULT1 et AfH3.1<sub>26-44</sub> ou AfH3.3<sub>26-44</sub>

Le pull-down est une technique biochimique pour tester l'interaction entre deux protéines *in vitro*. Dans cette expérience, une des protéines est fixée spécifiquement sur une résine à travers une étiquette, tandis que l'autre ne l'est pas. Dans le cas où il y a une interaction, les deux protéines seront co-éluées ensemble.

Le clonage des séquences codantes des peptides AfH3.1<sub>26-44</sub> et AfH3.3<sub>26-44</sub> a été effectué dans le pETM41 entre les sites de restriction NcoI et NotI et validé par séquençage. Après avoir été purifiée séparément dans un tampon contenant 50 mM Tris pH 7,5, 200 mM NaCl, 10mM DTT), les protéines HIS-MBP-H3.1<sub>26-44</sub> (ou HIS-MBP-H3.3<sub>26-44</sub>) et ULT1<sub>128-237</sub> ont été mélangées puis déposées sur une résine d'amylose. Après plusieurs lavages de la résine avec le tampon de purification, les protéines fixées ont été éluées avec le même tampon supplémenté de 10 mM Maltose. Les fractions collectées ont été analysées par SDS-PAGE.

## 6. Mesure de l'interaction entre ULT1<sub>128-237</sub> et le peptide HsH3.1<sub>26-44</sub> par ITC

La Titration Calorimétrique Isotherme (ITC) permet de mesurer la chaleur dégagée ou absorbée lors de l'interaction entre deux molécules en solution, généralement une macromolécule biologique (protéine, acide nucléique) et un ligand. La mesure de la chaleur dégagée ou absorbée est réalisée à l'aide d'un microcalorimètre, un instrument qui mesure la différence de température entre deux cellules (cellule de référence et cellule contenant l'échantillon protéique) et la compense en injectant de l'énergie pour maintenir une température constante. Au cours de la titration calorimétrique isotherme, le ligand est injecté par petites quantités dans la cellule contenant l'échantillon protéique, et la chaleur dégagée ou absorbée lors de l'interaction est mesurée à chaque injection. La courbe d'interaction obtenue permet de déterminer la stœchiométrie (nombre de sites de liaison entre les deux molécules,  $n$ ), l'enthalpie (chaleur dégagée ou absorbée lors de l'interaction,  $\Delta H^\circ$ ), et la constante d'association (affinité,  $K_d$ ) entre les deux molécules. Par la suite, ces données permettent de calculer la variation d'énergie libre du système ( $\Delta G^\circ$ ), l'entropie ( $\Delta S^\circ$ ) ainsi que la constante de dissociation  $K_d$ .

Les expériences d'ITC ont été effectuées au niveau de la plateforme ISBG avec l'aide de Caroline Mas. La protéine ULT1<sub>128-237</sub> a été purifiée dans un tampon contenant 50 mM Tris pH 7,5, 100 mM NaCl, 1mM TCEP puis a été concentrée à 75  $\mu$ M et placée dans la cellule. En parallèle, le peptide HsH3.1<sub>26-44</sub> avec une triméthylation sur le résidu K36 et une acétylation en N-ter (Genecust) a été dialysé avec le même tampon de purification

qu'ULT1<sub>128-237</sub> et a été concentré à 872  $\mu\text{M}$  avant d'être placé dans la seringue. La titration du peptide avec ULT1 s'est faite par 18 injections de 2,5  $\mu\text{L}$  et ont été réalisées à 25 °C sur un appareil MicroCal ITC200 (Malvern).

## D. Caractérisation biochimique et structurale de l'interaction entre ULT1 et le complexe PRC2

### 1. Pull-down sur résine Strep-Tactin XT

Le complexe PRC2-CLF et la protéine ULT1 sans étiquette ont été purifiés séparément dans un tampon contenant 25 mM Hepes pH 7,8, 100 mM NaCl, 1 mM TCEP 2 % glycérol et 2 mM  $\text{MgCl}_2$ . Les 3 échantillons : ULT1 ; PRC2-CLF et ULT1 + PRC2-CLF ont été préparés et incubés pendant 20 minutes avant d'être déposés sur une résine Strep-Tactin XT. Une fois ces échantillons déposés et la résine lavée avec 10 volumes de tampon de purification, les protéines liées ont été éluées avec le même tampon supplémenté avec 50 mM de biotine. Les fractions collectées ont été analysées par SDS-PAGE.

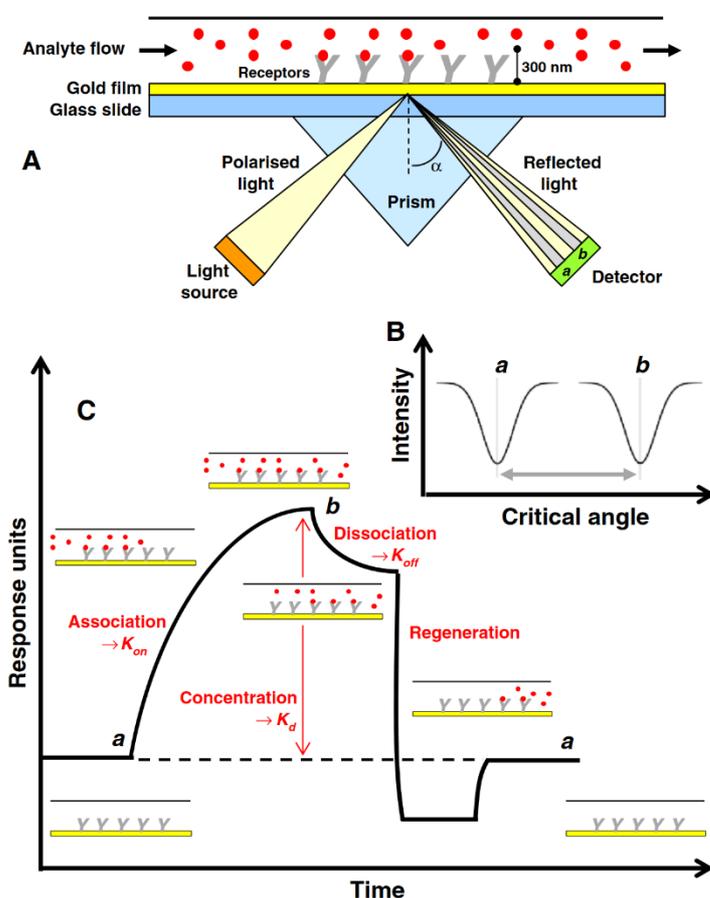
Le même protocole a été utilisé pour tester l'interaction entre PRC2-SWN et ULT1 mais aussi entre PRC2-SWN et les domaines N-terminal et C-Terminal d'ULT1.

### 2. Résonance Plasmonique de Surface

La Résonance Plasmonique de Surface (SPR) est une technique d'analyse optique qui permet de détecter et caractériser des interactions moléculaires en temps réel en mesurant de petites variations d'indice de réfraction à la surface d'un matériau recouvert d'une fine couche de métal tel que l'or ou l'argent. Elle fonctionne en deux étapes : tout d'abord, un ligand est immobilisé sur une surface, ce qui provoque une modification de densité et de l'indice de réfraction. Ensuite, le récepteur est injecté sur la surface par microfluidique, entraînant une seconde modification de densité si une interaction a lieu entre le récepteur et le ligand. Lorsqu'un rayon lumineux monochromatique est dirigé vers la surface sous un angle précis, un phénomène appelé réflexion totale de la lumière provoque l'émission d'une onde de plasmon de surface, amplifiée par la présence de métal sur la surface, d'où le nom de *Surface Plasmon Resonance* (SPR). L'angle de réflexion de cette onde est lié à l'intensité du phénomène de SPR. Lorsque la densité à la surface change, par exemple lors de l'association ou de la dissociation d'un ligand, l'angle de réflexion de l'onde varie légèrement. Cette variation d'angle est mesurée par l'appareil

SPR en utilisant une grandeur arbitraire appelée Unité de Résonance (ou *Response Unit*, RU). Cette modification de l'angle de réflexion est observée lorsque le récepteur est fixé et lorsque le ligand se fixe sur celui-ci. Cette technique permet donc de caractériser les paramètres cinétiques d'une interaction tel que les constantes d'association et de dissociation mais aussi la constante thermodynamique d'association.

Les expériences de SPR ont été effectuées au niveau de la plateforme ISBG avec l'aide d'Etienne Dubiez et de Jean-Baptiste Reiser. Elles ont été réalisées sur un instrument Biacore 2000 (GE Healthcare) dans 200 mM de NaCl, 1 mM de TCEP, 50  $\mu$ M EDTA et 0,05 % Tween 20. Le PRC2-CLF a été immobilisées sur une puce de détection Ni-NTA (Cytiva) et différentes concentrations d'ULT1 ont été injectées à 30  $\mu$ L.min<sup>-1</sup>.



**Figure II.3 :** Illustration schématique d'une expérience de SPR pour la mesure d'interaction entre deux molécules. (A) Principe de fonctionnement de l'expérience de SPR. (B) Changement de l'angle de réflexion de la lumière polarisée provoqué par la liaison d'un récepteur sur le ligand. (C) Courbe représentant la variation d'angle de réflexion en fonction du temps dont les valeurs permettent de calculer la vitesse d'association ( $k_{on}$ ) et de dissociation ( $k_{off}$ ) entre le ligand et le récepteur. La constante de dissociation peut être mesurée par la formule suivante :  $K_d = k_{off}/k_{on}$ . La figure a été extraite d'une publication de Simon Patching, 2014.

### 3. Western-Blot

Le transfert des protéines sur une membrane PVDF pré-activée (lavage rapide à l'éthanol 100% puis rinçage à l'eau) a été réalisé dans un tampon de transfert (20% éthanol, 20% de tampon BioRad TransBlot turbo 5x et 60% eau Milli Q) pendant 7 minutes à 2,5 A (jusqu'à 25 V). Après le transfert, la membrane a été incubée pendant 1h à température ambiante dans une solution de blocage (5% de lait, 1X TBS, 0,1% Tween). Après 6 lavages de 5 min de TTBS (1X TBS, 0,1% Tween), la membrane a été incubée pendant une nuit à 4°C avec l'anticorps primaire dirigé contre ULT1. Après 6 lavages de 5 min (1X TBS, 0,1% Tween), la membrane a été incubée pendant 1 h en présence de l'anticorps secondaire (anticorps de *Goat-Anti-Rabbit*) couplé à la peroxydase (HRP) dilué au 1:10 000, 1X TBS, 0,1% Tween). Après 6 lavages de 5 min pour éliminer l'anticorps secondaire, la révélation de la membrane a été réalisée avec le kit ECL de chez BioRad et visualisée sur le scanner ImageQuant 800 (Amersham GE).

### 4. Caractérisation structurale de PRC2-CLF et de l'interaction entre ULT1 et PRC2-SWN

#### a. Reconstitution du complexe PRC2-SWN/ULT1

La dernière étape de purification de PRC2-SWN et d'ULT1 par chromatographie d'exclusion de tailles (*S200 increase*) s'est déroulée dans un tampon contenant 25 mM Hepes pH 7,8, 100 mM NaCl, 1mM TCEP, 2 mM MgCl<sub>2</sub> et 2 % glycérol. Par la suite, ces protéines, ont été mélangées avec un ratio molaire 1 : 2 puis incubés pendant 20 minutes à température ambiante. Enfin, le mélange est concentré jusqu'à un volume de 500 µL par un système Amicon Ultra avec seuil de coupure de 10 kDa avant d'être de nouveau injecté sur chromatographie d'exclusion de tailles (*S200 increase*). Les fractions éluées du complexe PRC2-SWN / ULT1 sont analysées par SDS-PAGE et sont regroupées pour l'étape de pontage chimique.

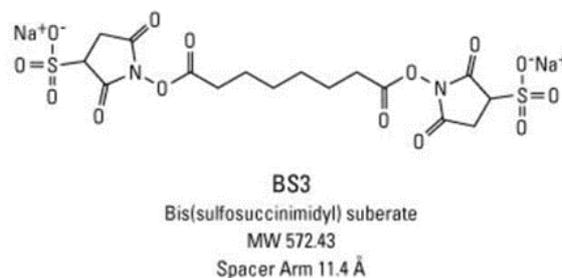
#### b. Pontages chimiques des complexes PRC2-CLF et PRC2-SWN/ULT1 pour la cryo-EM

Le pontage chimique des protéines est un processus par lequel des liaisons covalentes sont formées entre plusieurs acides aminés. Ces liaisons covalentes peuvent être inter-moléculaires, reliant deux chaînes polypeptidiques provenant de différentes sous-unités, ou intramoléculaires, reliant deux acides aminés dans la même chaîne polypeptidique. C'est un moyen efficace de stabiliser leur repliement tridimensionnel en vue de résoudre leur structure.

Le principal agent de pontage chimique utilisé pour stabiliser le complexe PRC2 et l'interaction avec ULT1 a été le subérate de bissulfosuccinimidyle, également connu sous le nom de BS3 (Thermo Fischer). C'est un réactif d'une longueur de 11,4 Å couramment utilisé pour le pontage chimique des protéines. Il s'agit d'un agent de couplage qui possède deux groupes réactifs d'ester de N- hydroxysulfosuccinimide (NHS) et qui permet la formation de liaisons covalentes entre les groupements amines primaires des lysines mais aussi ceux de l'histidine ou celui du résidu en position N-terminale.

Dans un volume final de 60  $\mu\text{L}$ , 0,3  $\text{mg.mL}^{-1}$  de PRC2-CLF ont été réticulé à température ambiante en présence de 2,7 mM de BS3, 4,3  $\mu\text{M}$  *Afh3.1<sub>20-32\_K27M</sub>*, 4,3  $\mu\text{M}$  *Afh3.1<sub>20-32\_K27me3</sub>* et 4,3  $\mu\text{M}$  de S-adenosyl-méthionine (SAM), puis la réaction a été arrêtée après 45 minutes par l'ajout de 60 mM Tris pH8. Dans le but de vérifier que toutes les sous-unités aient bien été réticulées, une fraction de l'échantillon a été déposée sur SDS-PAGE 6 et 15 %.

Pour le complexe PRC2-SWN/ULT1, 0,4  $\text{mg.mL}^{-1}$  ont été réticulé à température ambiante en présence de 5 mM de BS3.



**Figure II.4 : schéma de la molécule de BS3 (ThermoFischer)**

### c. Analyse du complexe PRC2-SWN/ULT1 par XL-MS

Le pontage chimique couplé à la spectrométrie de masse (XL-MS) est une technique qui permet d'identifier les régions spatialement proches d'une protéine ou d'un complexe. Le principe de cette méthode repose sur l'utilisation d'une molécule de pontage chimique capable réagir avec la protéine cible pour former une liaison covalente stable entre deux acides aminés. Une fois que la protéine a été réticulée, celle-ci est soumise à une digestion enzymatique, qui va permettre de cliver la protéine en peptides plus petits. Ces peptides sont ensuite analysés par spectrométrie de masse pour identifier les résidus liés par l'agent de réticulation.

Un échantillon non réticulé du complexe PRC2-SWN/ULT1 dans un tampon contenant 25 mM Hepes pH 7,8, 100 mM NaCl, 2mM MgCl<sub>2</sub> et 1 mM TCEP a été concentré à 1,4 mg.mL<sup>-1</sup> puis analysé à la plateforme de spectrométrie de masse du laboratoire EDyP (Grenoble). Dans ce laboratoire, le complexe a été réticulé par l'urée dibutyrique disuccinimidyle (DSBU), digéré par la trypsine puis analysé par chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS).

#### **d. Cryo-microscopie électronique (Cryo-EM)**

Comme la cristallographie aux rayons X et la Résonance Magnétique Nucléaire (RMN), la Cryo-microscopie électronique (Cryo-EM) est une méthode pour déterminer la structure de molécules biologiques. Ces dernières années elle a connu un développement technologique important, permettant d'atteindre des résolutions quasi atomique des objets étudiés. L'un des avantages de cette technique est de pouvoir résoudre la structure de larges protéines ou complexes dans des conditions proches de leur état natif et qui sont difficilement cristallisables. Pour cela, l'échantillon est déposé sur une grille perforée qui est vitrifiée par congélation rapide et les objets de l'échantillon sont alors immobilisés dans plusieurs orientations. La grille est ensuite placée dans le microscope électronique où elle va être exposée à un faisceau d'électrons. En contact avec les atomes de l'échantillon, les électrons envoyés par le microscope vont alors être dispersés et détectés par une caméra pour former une image en deux dimensions des objets présents (micrographie). Lors du traitement des images, ces projections 2D sont d'abord alignées et classées puis combinés pour obtenir une représentation 3D de l'objet étudié.

Pour PRC2-CLF réticulé, 3,5  $\mu$ L du complexe à 0,3 mg/mL ont été déposés sur une grille UltraAuFoil R 1,2/1,3 300 mesh (Quantifoil) préalablement rendue hydrophile par décharge lumineuse (25 mA, 30 s). L'excès de protéines sur la grille est ensuite retiré et cette dernière est rapidement plongée dans de l'éthane liquide en utilisant un Vitrobot Mark IV (FEI, Thermo Fisher Scientific) afin de vitrifier l'échantillon. Un ensemble de données d'environ 2 414 films de 40 images a été acquis sur un microscope électronique Glacios (Thermo Fisher Scientific) de 200 kV à un grossissement nominal de 36 000 avec une taille de pixel de 1,145 Å.

Les films bruts enregistrés avec une caméra Gatan K2 Summit ont été importés dans CryoSPARC version 4.2.1 pour y être traités. Toutes les classes 2D représentant des caractéristiques protéiques claires ont été sélectionnées et les 162 829 particules de ces classes ont été utilisées pour construire quatre modèles *Ab-initio*, suivis d'un affinement hétérogène. Parmi les quatre volumes obtenus, un seul a été de nouveau affiné, en

utilisant cette fois un affinement non-uniforme (Punjani, Zhang and Fleet, 2020). Une carte 3D avec une résolution allant jusqu'à 4,1 Å a été générée avec seulement 32 676 particules et les figures ont été préparées avec ChimeraX (Pettersen *et al.*, 2021).

Concernant PRC2-SWN/ULT1 réticulé, 3,5 µL de complexe à 1 mg/mL ont été déposés sur les grilles qui ont été congelées selon la même procédure que précédemment. 13 148 images ont été ensuite acquises sur un microscope électronique Krios (Thermo Fisher Scientific) de 300 kV à un grossissement nominal de 105 000 avec une taille de pixel de 0,84 Å.

Les données ont été traitées sous le logiciel CryoSPARC et 1 308 208 particules réparties en 116 classes 2D ont été utilisées pour construire un premier modèle *Ab initio* du complexe. À partir de cette carte et après plusieurs étapes de traitement, 135 012 particules ont finalement été sélectionnées et utilisées pour construire une carte 3D du complexe avec une résolution allant jusqu'à 3,1 Å. La construction du modèle dans la carte 3D est toujours en cours.

#### **e. AlphaFold2**

AlphaFold2 est un système révolutionnaire d'intelligence artificielle qui a été développé par la société DeepMind dans le but de prédire la structure 3D des protéines à partir de leur séquence primaire (Jumper *et al.*, 2021).

Lorsqu'une séquence d'acides aminés est soumise au logiciel AlphaFold2, celui-ci va chercher les séquences homologues et créer un alignement de multiples séquences (MSA). L'information extraite de ce MSA va permettre à AlphaFold2 d'identifier les acides aminés qui ont co-variés et donc qui sont susceptibles d'interagir ensemble. AlphaFold va également construire une matrice de distances deux à deux entre les acides aminés et chercher si les homologues de la séquence d'entrée ont des structures 3D déjà résolues expérimentalement afin d'obtenir des informations structurales. Ensuite, le MSA et la matrice de distance vont être utilisés dans un module d'apprentissage profond où ils vont être confrontés l'un à l'autre et pourront être réinterprétés au fur et à mesure que les données vont s'améliorer. Enfin, une fois que ces données ont été traitées, elles vont être décodées par un autre module afin de construire le modèle 3D de la protéine (Yang *et al.*, 2023).

Pour chaque résidu de la structure 3D obtenue, un score de confiance de 0 à 100 appelé *predicted local-distance difference* (pLDDT) est attribué. Ce score est calculé en comparant les mesures de distances entre les C $\alpha$  des résidus proches dans l'espace de la structure générée par AlphaFold2 et de la structure expérimentale (Mariani *et al.*, 2013).

Pour une valeur supérieure à 90, la confiance du modèle est très élevée. Elle reste élevée entre 70 et 90, mais devient faible entre 70 et 50. Les valeurs inférieures à 50 ne devraient pas être interprétées mais peuvent correspondre à des régions désordonnées de la protéine.

Un autre score est également calculé par AlphaFold2 et correspond pour une paire de résidus ( $x$  et  $y$ ) à l'erreur de position estimée du résidu  $X$  lorsque les résidus  $Y$  de la structure expérimentale et de la structure prédite par AlphaFold2 sont alignés.

AlphaFold est capable de prédire des structures de protéines individuelles mais aussi de complexes. Dans cette thèse, pour générer des modèles d'ULT1, PRC2-CLF, PRC2-SWN et PRC2/ULT1, nous avons utilisé AlphaFold2 sur le serveur public ColabFold mais aussi sur la station de calcul du Commissariat à l'Énergie Atomique (CEA).

# III. Résultats

## A. Étude biochimique et structurale de la protéine ULT1

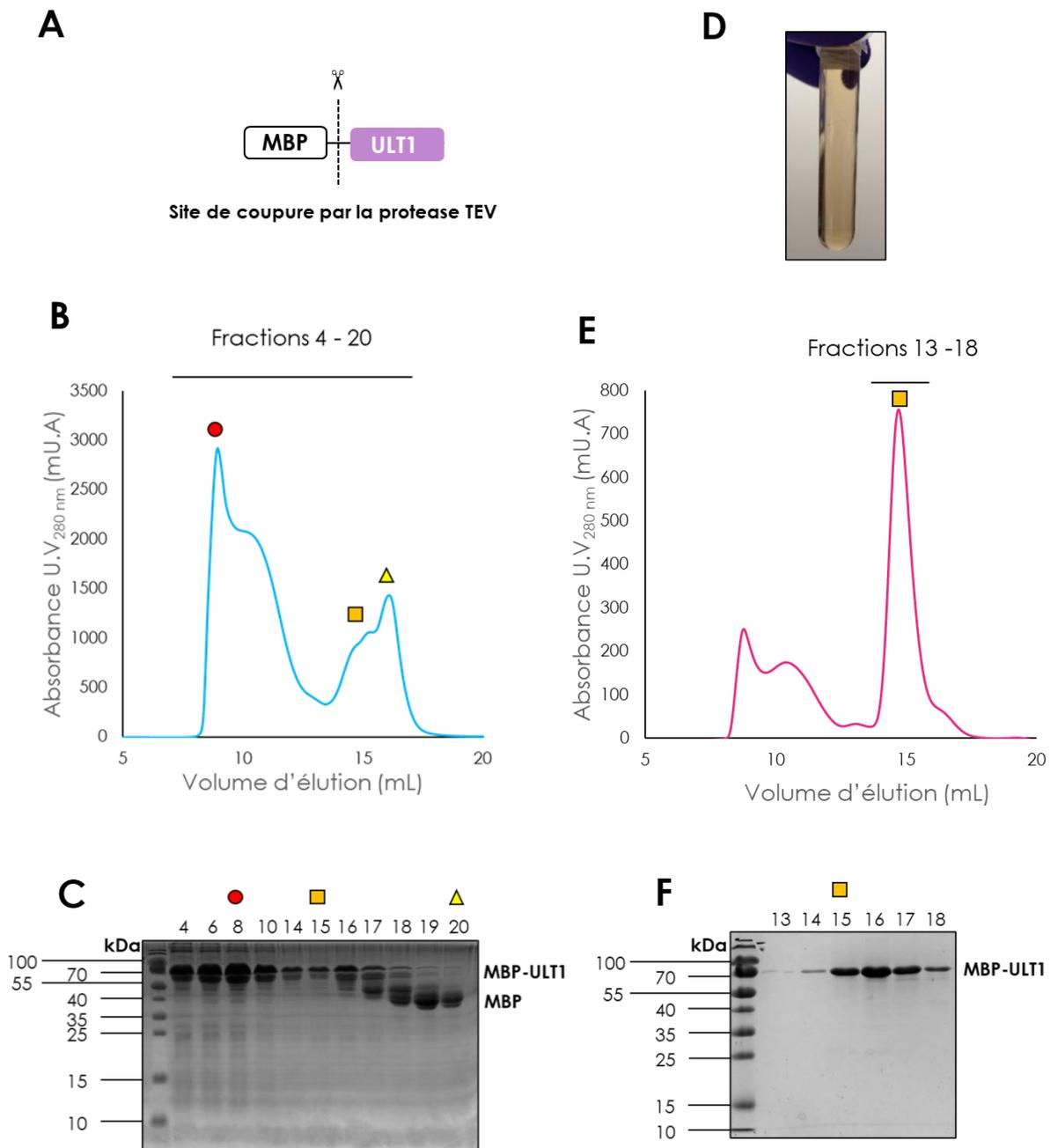
Cette première partie de la thèse traite de la caractérisation biochimique et structurale de la protéine ULTRAPETALA1 chez *Arabidopsis* (AtULT1 ; nommée ULT1 par la suite) dans le but de mieux comprendre sa fonction.

L'expression d'ULT1 recombinante et mono-dispersée a longtemps été un défi au sein du laboratoire et cela a grandement freiné l'étude biochimique de cette protéine. L'une des raisons possibles était la présence d'un grand nombre de cystéines dans sa séquence. Avant mon arrivée au laboratoire, de premiers essais d'analyses biochimiques de la protéine avaient été réalisés par Emmanuel Thévenon. Emmanuel a cloné la version pleine taille d'ULT1 (aa 1-237) en fusion avec la *Maltose Binding Protein* (MBP) en position N-terminale (**Figure III.1 A**) et l'a exprimée en système bactérien. Cependant, une majeure partie de la protéine s'est révélée agrégée lors des étapes finales de purification et les rendements obtenus n'étaient pas suffisants pour une étude structurale. J'ai donc commencé la thèse par chercher à optimiser la production et la purification d'ULT1.

### 1. Expression d'ULT1 : importance du Zinc

Après expression de MBP-ULT1<sub>1-237</sub> dans une souche de *E.coli* BL21 star et purification d'affinité sur résine d'amylose, la protéine a été chargée sur une colonne d'exclusion de taille (*Size Exclusion Chromatography*, SEC) (**Figure III.1 B**). Les fractions collectées ont ensuite été analysées sur gel de polyacrylamide 15 % en condition dénaturante (SDS-PAGE) (**Figure III.1 C**).

MBP-ULT1<sub>1-237</sub> a une taille moléculaire attendue aux alentours de 70 kDa, cependant elle est éluée à des volumes correspondant à des molécules de masse moléculaire élevée (**Figure III.1 C**, Fractions 4 à 10). Une explication possible serait la formation de ponts disulfures entre les différentes molécules d'ULT1 liée à son grand nombre de cystéines (23 au total) malgré la présence d'agents réducteurs dans le tampon (10mM DTT). On remarque également sur l'ensemble des fractions la présence d'une protéine d'une taille comprise entre 55 et 70 kDa qui pourrait correspondre à des produits de dégradation de la fusion MBP-ULT1. Malgré ces observations, une partie de la protéine MBP-ULT1<sub>1-237</sub> semble éluer dans la fraction prévue, c'est à dire correspondant à son poids moléculaire (**Figure III.1 C** Fraction 15).



**Figure III.1 : Effet du zinc sur l'expression d'ULT1 :** (A) Représentation schématique de la construction d'ULT1 utilisée pour son expression. ULT1 est exprimée en fusion avec la *Maltose Binding Protein* en position N-Terminale et contient un site de coupure par la protéase TEV. (B) Profil d'élué obtenu après injection de MBP-ULT1 sur colonne d'exclusion de taille : S200 *increase* (SEC). Le pic d'agrégats ou de multimères est identifié par un rond rouge, la fraction non agrégée (15) de MBP-ULT1 est identifiée par un carré orange et enfin, le triangle jaune correspond au pic d'élué de l'étiquette MBP. (C) Analyse par SDS-PAGE des fractions collectées en (B). (D) Fraction de MBP-ULT1 exprimée sans zinc après purification sur résine d'amylose, dont la couleur de la solution est brune. (E) Profil d'élué obtenu pour MBP-ULT1 exprimée avec du zinc et après injection sur colonne d'exclusion de taille : S200 *increase* (SEC). Le carré orange représente à nouveau la fraction non-agrégée (15) de MBP-ULT1 pour cette expérience. (F) Analyse par SDS-PAGE des fractions collectées en (E).

Après plusieurs ajustements dont le changement de souche d'expression par des *E. coli* Rosetta, j'ai réussi à trouver comment améliorer l'expression de MBP-ULT1<sub>1-237</sub>. J'ai découvert que l'ajout de ZnSO<sub>4</sub> pendant l'expression améliorerait considérablement les

rendements de protéines solubles. En effet, en raison de la présence de nombreuses cystéines, j'ai émis l'hypothèse selon laquelle ULT1 pouvait lier des métaux et ce malgré le fait que les analyses bio-informatiques n'aient pas révélé de motifs de liaisons pour ces derniers. Les deux métaux les plus probables étaient soit le fer (cluster fer-soufre) soit le zinc. J'ai décidé de tester les deux, mais seul l'ajout de zinc a montré une différence majeure dans l'expression d'ULT1. J'ai remarqué un changement de couleur dans la fraction éluée de la protéine après purification sur colonne d'amylose. En effet, cette couleur était initialement brune (**Figure III.1 D**) mais après expression en présence de zinc, cette couleur spécifique a disparu. Une explication possible est que la bactérie a utilisé d'autres éléments disponibles dans le milieu de culture pour compenser le manque de zinc, entraînant ainsi un changement de conformation et une absorbance différente de la protéine. De plus, on remarque que le profil du chromatogramme d'exclusion de taille de MBP-ULT1<sub>1-237</sub> a changé (**Figure III.1 B et E**). En effet, après expression de la protéine avec du ZnSO<sub>4</sub>, on note que la proportion d'agrégat a diminué et que la proportion de MBP-ULT1<sub>1-237</sub> non-agrégée a augmenté (**Figure III.1 E**). On observe également moins de dégradation de MBP-ULT1<sub>1-237</sub>, ce qui est probablement dû à un meilleur repliement de la protéine (**Figure III.1 F**).

Une fois l'expression de MBP-ULT1<sub>1-237</sub> établie, j'ai cherché à améliorer sa purification afin d'obtenir une protéine pure et mono-dispersée (sans étiquette) pour les expériences de caractérisation.

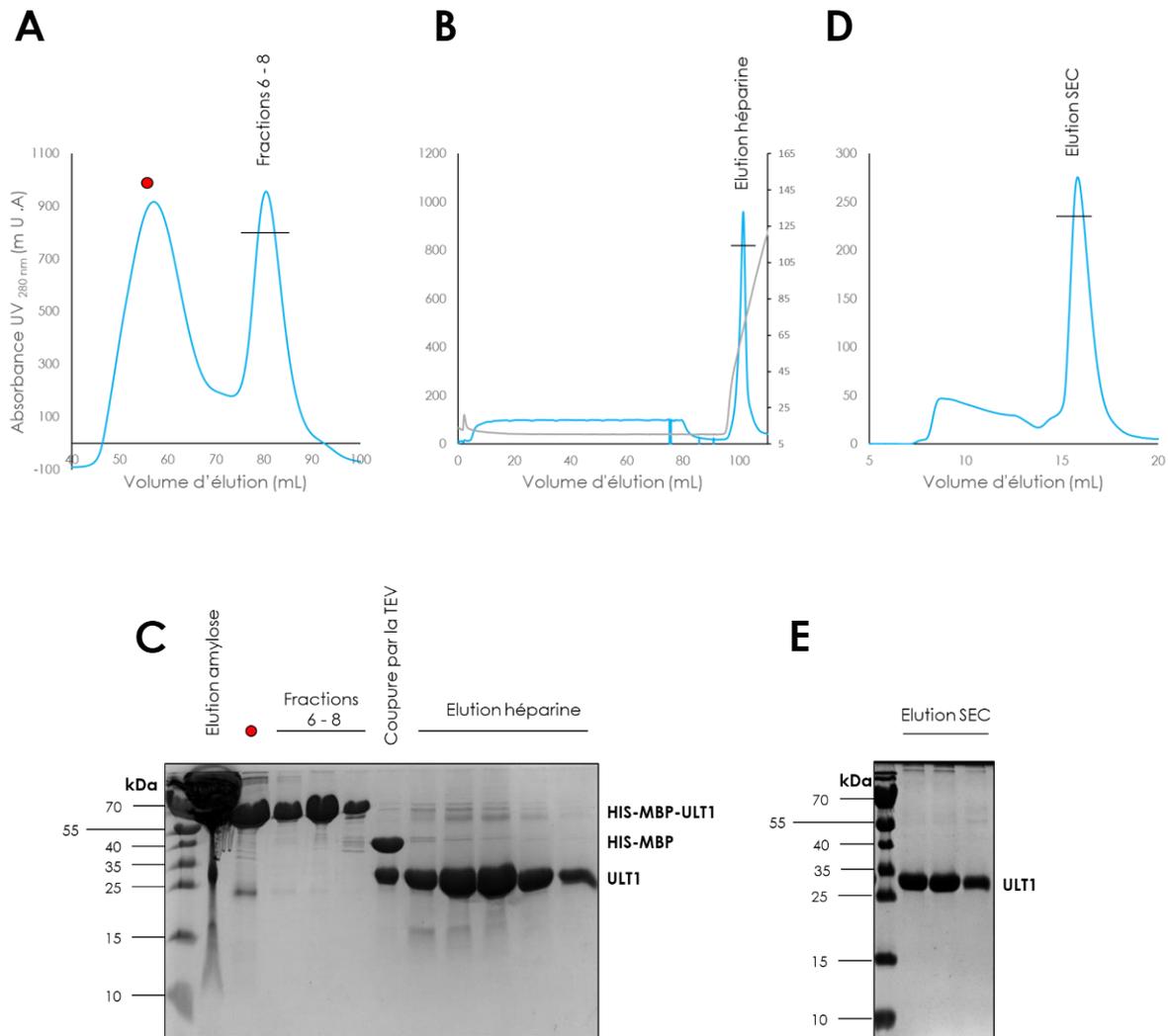
Tout d'abord, j'ai décidé de cloner la séquence ULT1<sub>1-237</sub> dans le pETM41. Ce nouveau vecteur permet d'ajouter une étiquette poly-histidines en position N-Terminale de l'étiquette MBP et peut être utile pour la séparation de cette dernière avec ULT1 si l'étape de purification sur colonne d'héparine ne s'est pas révélée efficace.

Après purification de HIS-MBP-ULT1<sub>1-237</sub> sur résine d'amylose, l'élution obtenue est injectée sur une colonne d'exclusion de taille (**Figure III.2 A**). Le profil d'élution de cette chromatographie est similaire aux résultats de purification obtenus précédemment avec le clone MBP-ULT1<sub>1-237</sub> et révèle deux pics principaux. Un premier pic qui correspond soit à une multimérisation soit à des agrégats d'ULT1 et un deuxième pic élué vers 75 mL qui lui représente la fraction non-agrégée de HIS-MBP-ULT1<sub>1-237</sub>. Ces agrégats ou ces multimères réduisent l'efficacité du clivage de l'étiquette par la protéase TEV c'est pourquoi seule la fraction HIS-MBP-ULT1<sub>1-237</sub> non-agrégée est conservée pour les prochaines étapes de purification.

Pour séparer ULT1<sub>1-237</sub> clivée de l'étiquette HIS-MBP, la meilleure méthode était d'utiliser une colonne d'héparine (**Figure III.2 B**). En effet, l'héparine va mimer les propriétés physico-chimiques de l'ADN et permettre de retenir uniquement ULT1<sub>1-237</sub> probablement

via son domaine SAND. L'analyse par SDS-PAGE des fractions du pic d'éluion obtenues indique la présence majeure d'ULT1<sub>1-237</sub> clivé malgré quelques impuretés (**Figure III.2 C**).

Enfin, après cette étape, la protéine est de nouveau injectée sur une colonne d'exclusion de taille afin d'éliminer les derniers contaminants et de vérifier sa solubilité (**Figure III.2 D**). Le chromatogramme obtenu ainsi que le SDS-PAGE associé, confirment la solubilité et la pureté d'ULT1<sub>1-237</sub> (**Figure III.2 E**).



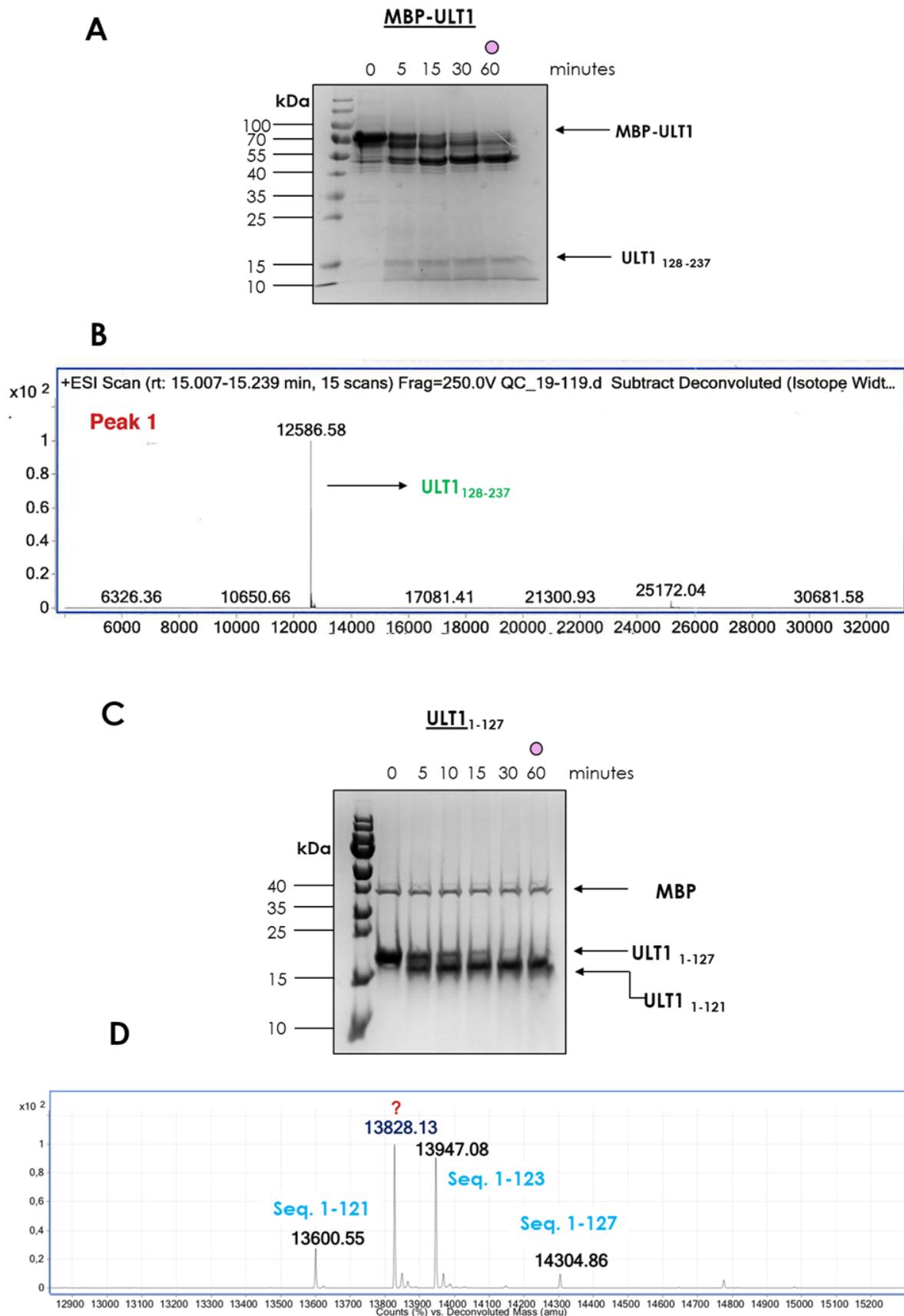
**Figure III.2 : Purification d'ULT1 :** (A) Profil d'éluion obtenu pour HIS-MBP-ULT1 exprimée avec du zinc après injection sur colonne d'exclusion de taille : S200 16/600. Le pic d'agrégats ou de multimères est identifié par un rond rouge. (B) Profil d'éluion obtenu après injection de HIS-MBP-ULT1 clivée par la TEV sur colonne d'héparine. L'éluion de ULT1 se fait par un gradient de NaCl et dont la conductivité est représentée par la courbe grise. (C) Analyse par SDS-PAGE des fractions collectées en (A) et (B). (D) Profil d'éluion obtenu après injection d'ULT1 sur colonne d'exclusion de taille : S200 *increase*. (E) Analyse par SDS-PAGE des fractions collectées en (D).

L'ajout de zinc lors de l'expression d'ULT1 a permis de franchir une étape clé pour la production de la protéine. Après la mise au point de la purification de cette protéine, j'ai donc pu commencer sa caractérisation structurale.

## 2. Identification des domaines structurés d'ULT1<sub>1-237</sub> par protéolyse limitée

Quand j'ai commencé la thèse, aucune structure tridimensionnelle d'ULT1 n'était disponible. L'analyse des alignements de séquences indiquaient la présence d'un domaine SAND à l'extrémité N-terminale et d'une région riche en cystéines à l'extrémité C-terminale. Dans le but d'obtenir des informations sur la fonction d'ULT1, il était important de caractériser sa structure. Comme les domaines semblent être connectés par une région flexible (région non conservée) qui pourrait empêcher sa cristallisation, nous avons voulu en savoir plus sur les domaines structurés d'ULT1. Dans ce but, j'ai réalisé une protéolyse limitée par la trypsine de MBP-ULT1<sub>1-237</sub>. Dès les premières minutes de la réaction, un fragment d'environ 15 kDa apparaît et reste résistant à la digestion enzymatique tout au long de l'expérience (**Figure III.3 A**). Avec l'aide de la plateforme de spectrométrie de masse de l'IBS et des expériences d'ESI-MS (*Electrospray Ionisation – Mass Spectrometry*), nous avons trouvé que ce fragment correspond aux résidus 128 à 237 d'ULT1 (**Figure III.3 B**). J'ai donc cloné la séquence correspondante dans le pETM41 afin de l'exprimer.

Les limites des régions structurées de la partie N-Terminale d'ULT1 étaient moins claires. Cependant après avoir exprimé la partie restante d'ULT1 (résidus 1 à 127) dans le pETM41, j'ai de nouveau réalisé une protéolyse limitée. Cette nouvelle digestion enzymatique couplée à une analyse par spectrométrie de masse a permis d'identifier un domaine structuré entre les résidus 1 à 121 d'ULT1 (**Figure III.3 C et D**). Néanmoins, la région la plus N-Terminale et non conservée d'ULT1 ne comporte pas de lysine ou d'arginine nécessaire pour la protéolyse par la trypsine. Il est donc probable que cette partie de la protéine corresponde à une région désordonnée d'ULT1. Dans le but d'exprimer ce domaine, la séquence correspondant aux acides aminés 1 à 121 d'ULT1 a été clonée dans le pETM41.



**Figure III.3 : Recherche des domaines structurés de MBP-ULT1 et ULT1<sub>1-127</sub> par protéolyse limitée et spectrométrie de masse. (A)** SDS-PAGE des échantillons obtenus après digestion par la trypsine de MBP-ULT1 en fonction du temps. **(B)** Analyse par spectrométrie de masse de MBP-ULT1 digéré par la trypsine pendant 60 minutes. **(C)** SDS-PAGE des échantillons obtenus après digestion par la trypsine de ULT1<sub>1-127</sub>

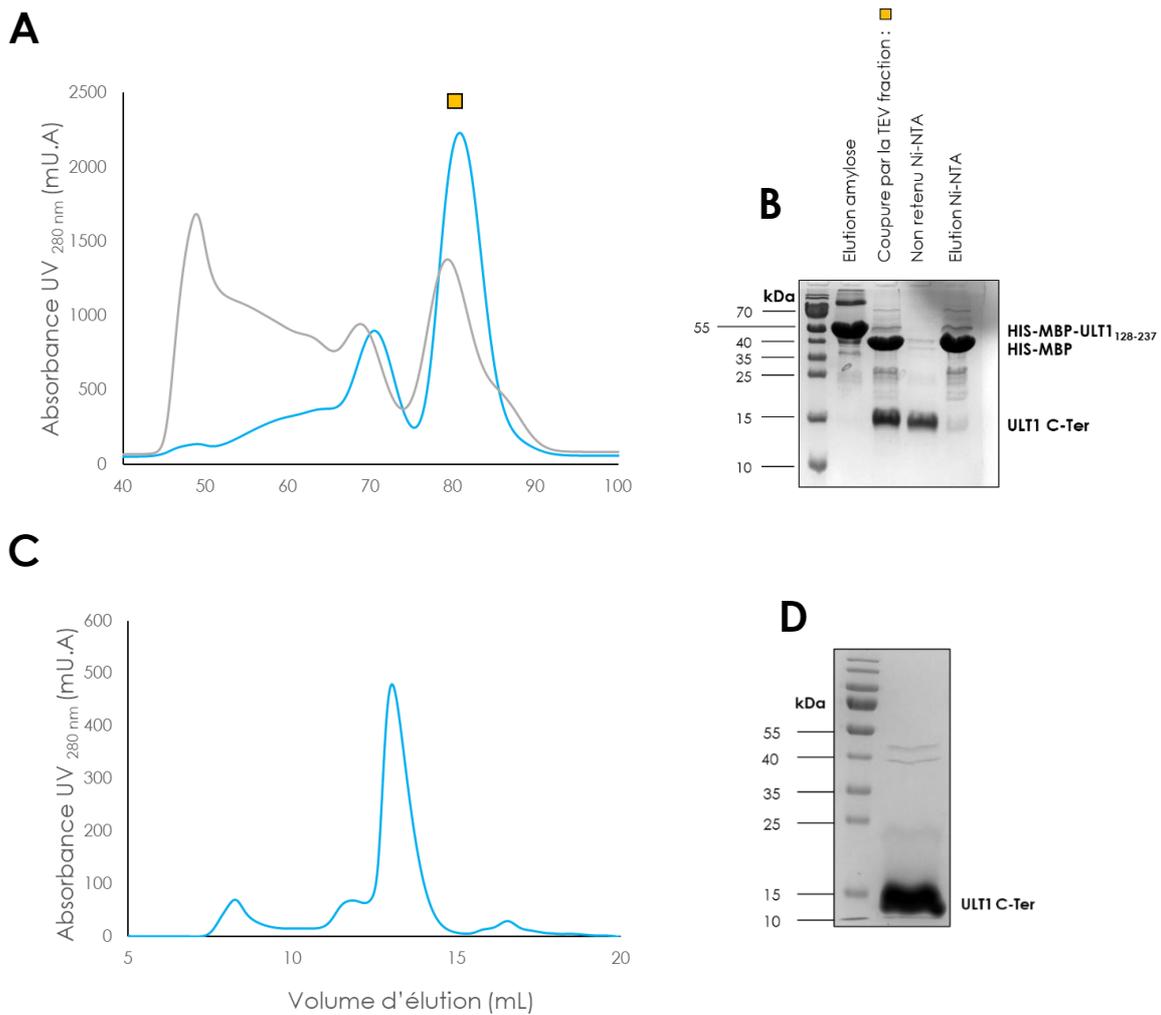
en fonction du temps. **(D)** Analyse par spectrométrie de masse d'ULT1<sub>1-121</sub> après digestion par la trypsine pendant 60 minutes. Les échantillons envoyés à la plateforme sont symbolisés par un point rose.

### 3. Purification du domaine C-terminal d'ULT1 (C-ter) : ULT1<sub>128-237</sub>

ULT1<sub>128-237</sub> a une masse moléculaire de 12,9 kDa et contient 16 cystéines. Pour purifier ULT1<sub>128-237</sub>, plusieurs étapes sont nécessaires mais de la même manière que pour ULT1<sub>1-237</sub>, l'ajout de ZnSO<sub>4</sub> pendant l'expression est essentiel. En effet on peut remarquer une grande différence dans le profil du chromatogramme après injection sur la colonne d'exclusion stérique de HIS-MBP-ULT1<sub>128-237</sub> exprimée avec ou sans zinc (**Figure III.4 A**). Lorsque HIS-MBP-ULT1<sub>128-237</sub> est exprimée avec du zinc, il y a une plus faible proportion d'agrégats ce qui est probablement dû à un meilleur repliement de la protéine. On peut également noter que le volume d'élution d'HIS-MBP-ULT1<sub>128-237</sub> exprimée avec ou sans zinc est différent et peut s'expliquer par une différence de conformation de la protéine. Enfin, au vu de l'intensité de l'absorbance à 280 nm, l'ajout de zinc semble donner de meilleurs rendements comme pour la protéine MBP-ULT1<sub>1-237</sub>. Dans cette expérience, on note également la présence d'un premier pic d'élution vers 70 mL, mais celui-ci s'est avéré correspondre à la formation d'un pont disulfure entre deux molécules de HIS-MBP-ULT1<sub>128-237</sub>. La dégradation ou la quantité insuffisante de DTT dans le tampon de purification sont les deux raisons les plus probables.

Par la suite, la purification d'ULT1<sub>128-237</sub> en condition suffisante de DTT, c'est à dire pour réduire toutes les cystéines libres de cette protéine, a montré une formation d'agrégats et/ou de multimères quasi inexistante et suggère donc une bonne stabilité de ce domaine.

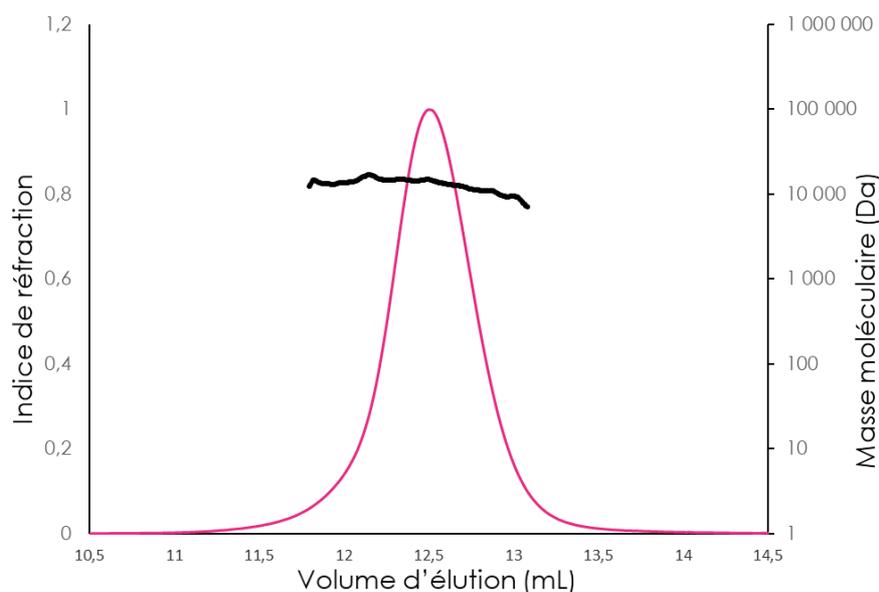
Une fois HIS-MBP-ULT1<sub>128-237</sub> purifiée sur résine d'amylose et l'étiquette clivée, la solution protéique est injectée sur une résine d'affinité Ni-NTA. Après vérification de l'échantillon sur SDS-PAGE (**Figure III.4 B**), la fraction correspondante à ULT1<sub>128-237</sub> est déposée sur chromatographie d'exclusion de taille. Le chromatogramme obtenu révèle un pic d'élution majoritaire vers 12,5 mL (sur colonne S75) et indique qu'ULT1<sub>128-237</sub> est soluble (**Figure III.4 C**). Le gel SDS-PAGE associé révèle également une grande pureté de la protéine (**Figure III.4 D**).



**Figure III.4 : Purification d'ULT1<sub>128-237</sub> :** **(A)** Profils d'éluion obtenus pour HIS-MBP-ULT1<sub>128-237</sub> exprimée avec du zinc (courbe bleue) ou sans zinc (courbe grise) après injection sur colonne d'exclusion de taille : S200 16/600. La fraction soluble de HIS-MBP-ULT1<sub>128-237</sub> est identifiée par un carré orange. **(B)** Analyse par SDS-PAGE de la purification de HIS-MBP-ULT1<sub>128-237</sub> sur résine d'amylose et Ni-NTA. **(C)** Profil d'éluion obtenu après injection d'ULT1<sub>128-237</sub> sur colonne d'exclusion de taille : S75 10/300 GL. **(D)** Analyse par SDS-PAGE de la fraction du pic principal collectée en (C).

#### 4. Analyse de l'état oligomérique d'ULT1<sub>128-237</sub> en solution

Suite à la possibilité de formation d'un dimère d'ULT1<sub>128-237</sub> (**Figure III.4 A**) et avant de réaliser des essais de cristallisation, j'ai analysé l'état oligomérique de ce domaine par SEC-MALLS. Le profil d'éluion obtenu après la filtration sur gel révèle un seul pic dont la masse calculée est de 13,55 +/- 0,38 kDa (**Figure III.5**). La masse théorique d'ULT1<sub>128-237</sub> étant de 12,9 kDa, ce résultat indique qu'en présence d'agents réducteurs, le domaine C-terminal d'ULT1 est monomérique en solution. L'état monomérique de ce domaine en condition réductrice avait déjà été montré pour un homologue d'ULT1 chez le riz (Roy *et al.*, 2019) par filtration sur gel analytique et SDS-PAGE ce qui est en accord avec nos résultats.



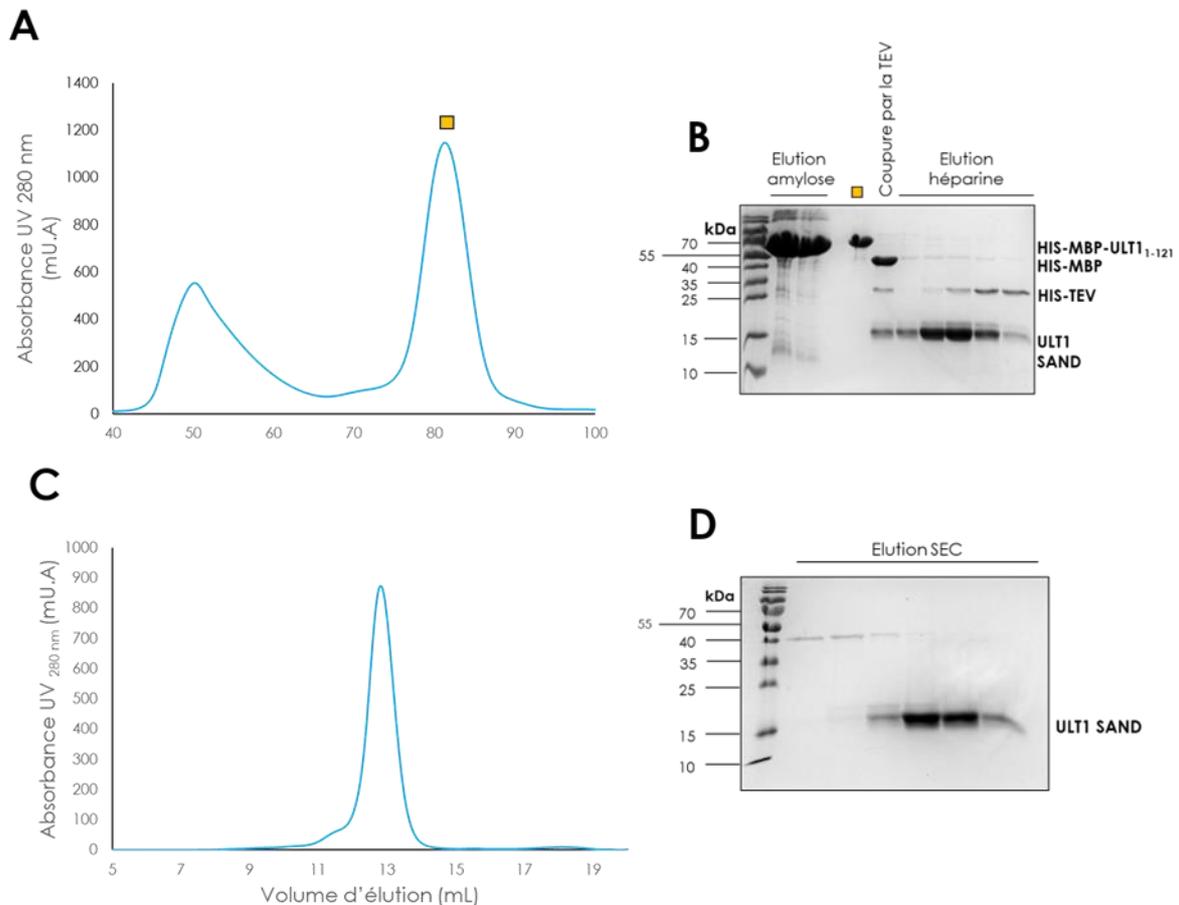
**Figure III.5 : Analyse par SEC-MALLS du domaine C-Terminal d'ULT1** : Chromatogramme d'élution sur une colonne de chromatographie d'exclusion de taille S75 10/300 GL. La courbe représente l'indice de réfraction en fonction du volume d'élution d'ULT1<sub>128-237</sub>. La masse moléculaire calculée (en Dalton) en fonction du volume d'élution est représentée par une courbe noire.

### 5. Purification du domaine N-terminal d'ULT1 (SAND) : ULT1<sub>1-121</sub>

ULT1<sub>1-121</sub> a une masse moléculaire de 13,6 kDa et contient 6 cystéines. Après une première purification de la protéine HIS-MBP-ULT1<sub>1-121</sub> sur résine d'amylose, la solution protéique est injectée sur une colonne d'exclusion de taille afin de vérifier la présence d'agrégats et/ou de multimères (**Figure III.6 A**). Le profil d'élution obtenu est similaire à celui de la protéine HIS-MBP-ULT1<sub>1-237</sub> et révèle la présence d'un premier pic correspondant à des molécules de masse élevée. Ce premier pic n'est pas observé lors de la purification du domaine C-Ter d'ULT1 et pourrait indiquer la formation de multimères au niveau du domaine SAND. Cependant, Il a été montré *in vitro* par filtration sur gel analytique, que la version pleine taille et le domaine SAND d'ULT1 chez le riz, étaient monomérique en présence d'agents réducteurs et que les multimères observés pour ces deux versions de la protéine étaient dû à la formation de pont-disulfures (Roy *et al.*, 2019). Dans notre cas, la présence d'agents réducteurs est assez élevée (10 mM) ce qui suggère que la formation de ce pic est probablement due à une certaine instabilité du domaine SAND pouvant conduire à la formation d'agrégats. Le deuxième pic élué vers 80 mL, quant à lui correspond à la fraction non-agrégée de la protéine HIS-MBP-ULT1<sub>1-121</sub> et va être conservé pour l'étape de clivage de l'étiquette par la TEV.

Une fois le clivage réalisé, la protéine ULT1<sub>1-121</sub> est purifiée sur colonne d'héparine pour séparer l'étiquette HIS-MBP. L'analyse par SDS-PAGE des fractions éluées révèle une

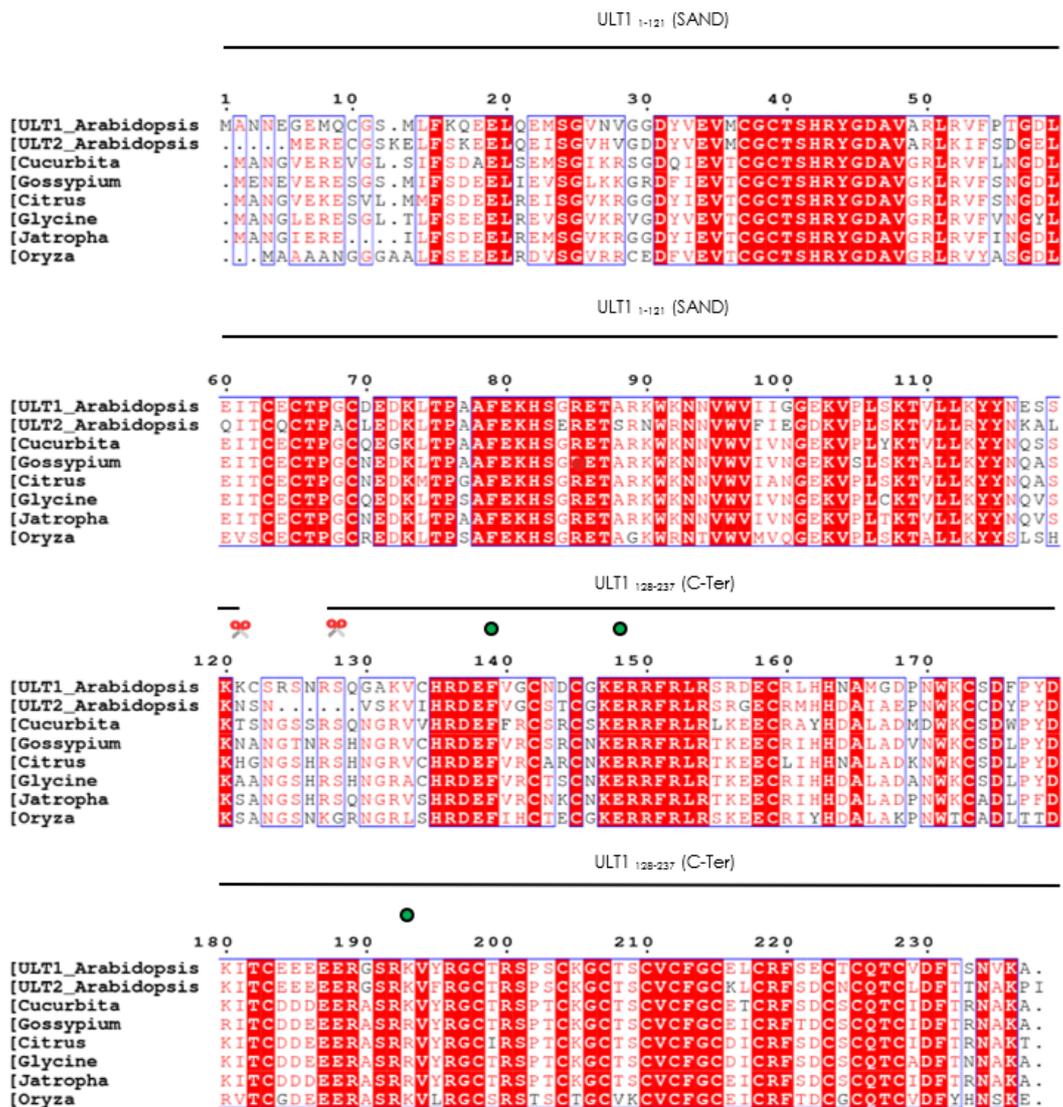
protéine aux alentours de 15 kDa, ce qui correspond à la masse attendue pour le domaine SAND (**Figure III.6 B**). Enfin la dernière étape de chromatographie d'exclusion de taille ainsi que le SDS-PAGE associé dévoile une protéine pure (**Figure III.6 C et D**). On remarque également un contaminant juste au-dessus de la taille d'ULT1<sub>1-121</sub> (plus de 15 kDa), dont l'analyse par spectrométrie de masse a indiqué qu'il pourrait appartenir à l'étiquette HIS-MBP.



**Figure III.6 : Purification d'ULT1<sub>1-121</sub> :** (A) Profils d'éluion obtenus pour HIS-MBP-ULT1<sub>1-121</sub> après injection sur colonne d'exclusion de taille : S200 16/600. La fraction soluble de la protéine HIS-MBP-ULT1<sub>1-121</sub> est identifiée par un carré orange. (B) Analyse par SDS-PAGE de la purification de HIS-MBP-ULT1<sub>1-121</sub> sur résine d'amylose et sur colonne d'héparine après clivage de l'étiquette. (C) Profil d'éluion obtenu après injection d'ULT1<sub>1-121</sub> sur colonne d'exclusion de taille : S75 10/300 GL. (D) Analyse par SDS-PAGE des fractions collectées en (C).

Suite aux résultats obtenus après la mise en œuvre d'un alignement de multiples séquences d'ULT1 (**Figure III.7**), j'ai également testé une autre construction du domaine SAND en clonant la séquence codant pour les résidus 13 à 121. En effet, les résidus 1 à 13 ne sont pas conservés chez les homologues d'ULT1 et indiquent la présence d'une région flexible qui pourrait prévenir la cristallisation de ce domaine. Cependant cette construction semblait plus encline à l'agrégation que la construction comprenant les

acides aminés 1 à 121. J'ai donc décidé d'utiliser la construction ULT1<sub>1-121</sub> pour exprimer le domaine SAND d'ULT1.

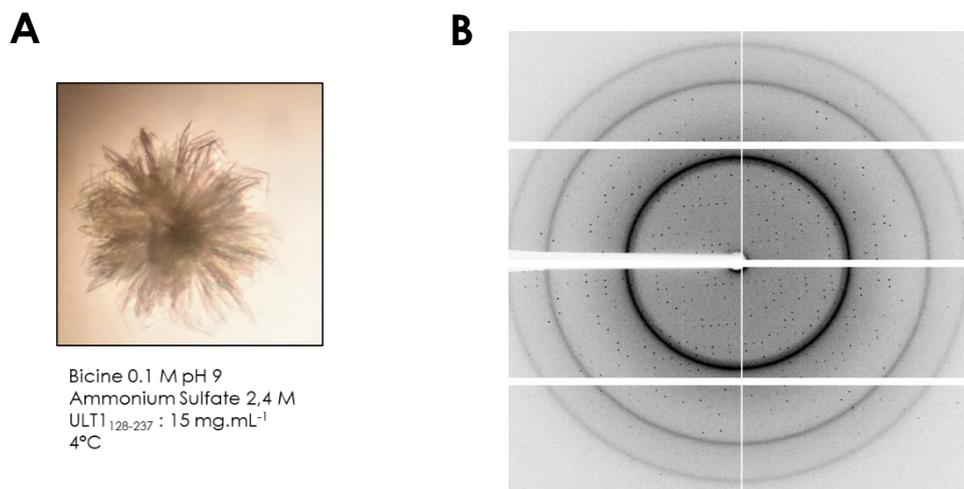


**Figure III.7 Alignement de multiples séquences d'ULT1 chez différentes espèces et d'ULT2 chez Arabidopsis.** Alignements de séquences réalisés par le logiciel ClustalW et colorisés par le logiciel ESPript3. Les lettres rouges représentent les acides aminés équivalents et les résidus en blancs aux acides aminés conservés. Les limites des constructions pour ULT1<sub>1-121</sub> et ULT1<sub>128-237</sub> sont également indiquées par des barres noires au-dessus des séquences. Les ronds verts symbolisent des résidus conservés particuliers et présentés dans la Figure 9C. Les paires de ciseaux au niveau de la lysine 121 (K121) et de l'arginine (127) indiquent les sites de coupures de la trypsine identifiés lors de l'expérience de protéolyse limitée.

## 6. Essai de cristallisation d'ULT1<sub>128-237</sub> et ULT1<sub>1-121</sub>

Après avoir identifié les domaines structurés d'ULT1<sub>128-237</sub> et d'ULT1<sub>1-121</sub> et établi leur protocole de purification, l'objectif était de déterminer la structure de ces domaines par cristallographie aux rayons X.

Dans un premier temps, je me suis concentré sur le domaine C-terminal plutôt que sur le domaine SAND car il est stable sans l'étiquette HIS-MBP et semble très structuré. Après avoir remplacé le DTT par un agent réducteur beaucoup plus stable (le TCEP), j'ai réalisé plusieurs essais de cristallogenèse en utilisant la plateforme de criblage à haut débit de l'EMBL de Grenoble (HTX). Les plaques de cristallisation réalisées ont été placées à 4°C et des cristaux sont apparus après 3 jours, dans une condition contenant un tampon BICINE 0,1 M à pH 9 et du sulfate d'ammonium 2,4 M.



**Figure III.8 : Crystallisation d'ULT1<sub>128-237</sub> :** (A) Cristaux du domaine C-terminal d'ULT1. (B) Cliché de diffraction obtenu sur la ligne de lumière ID23-1 de l'ESRF à Grenoble.

Ces cristaux ont pu être reproduits manuellement en utilisant de plus grosses gouttes (1  $\mu\text{L}$  + 1  $\mu\text{L}$ ). Cependant, comme ils se développent sous forme d'amas de plaques minces, il était assez difficile d'obtenir des cristaux uniques pour collecter des données de diffraction (**Figure III.8 A**). Lorsque nous avons testé la diffraction, en exposant l'arête de l'un des cristaux sur la ligne de lumière ID23-1 de l'ESRF à Grenoble, nous avons pu observer une diffraction jusqu'à une résolution de 1,5 Å (**Figure III.8 B**). De plus, nous savions qu'ULT1 était capable de lier du zinc c'est pourquoi nous avons collecté un ensemble de données à la longueur d'onde correspondant au bord d'absorption de cet atome (1,26515 Å). De ce fait, un jeu de données a pu être traité pour résoudre la structure d'ULT1<sub>128-237</sub> par la technique SAD (*Single-wavelength anomalous dispersion*).

En parallèle, de nombreux efforts ont été mis en place pour cristalliser le domaine SAND d'ULT1 mais sans succès. Lors de la purification de ce domaine je me suis rendu compte de l'instabilité de cette protéine ; il est donc envisageable qu'une partie de la protéine s'agrège et empêche la formation de cristaux. De plus comme l'a suggéré

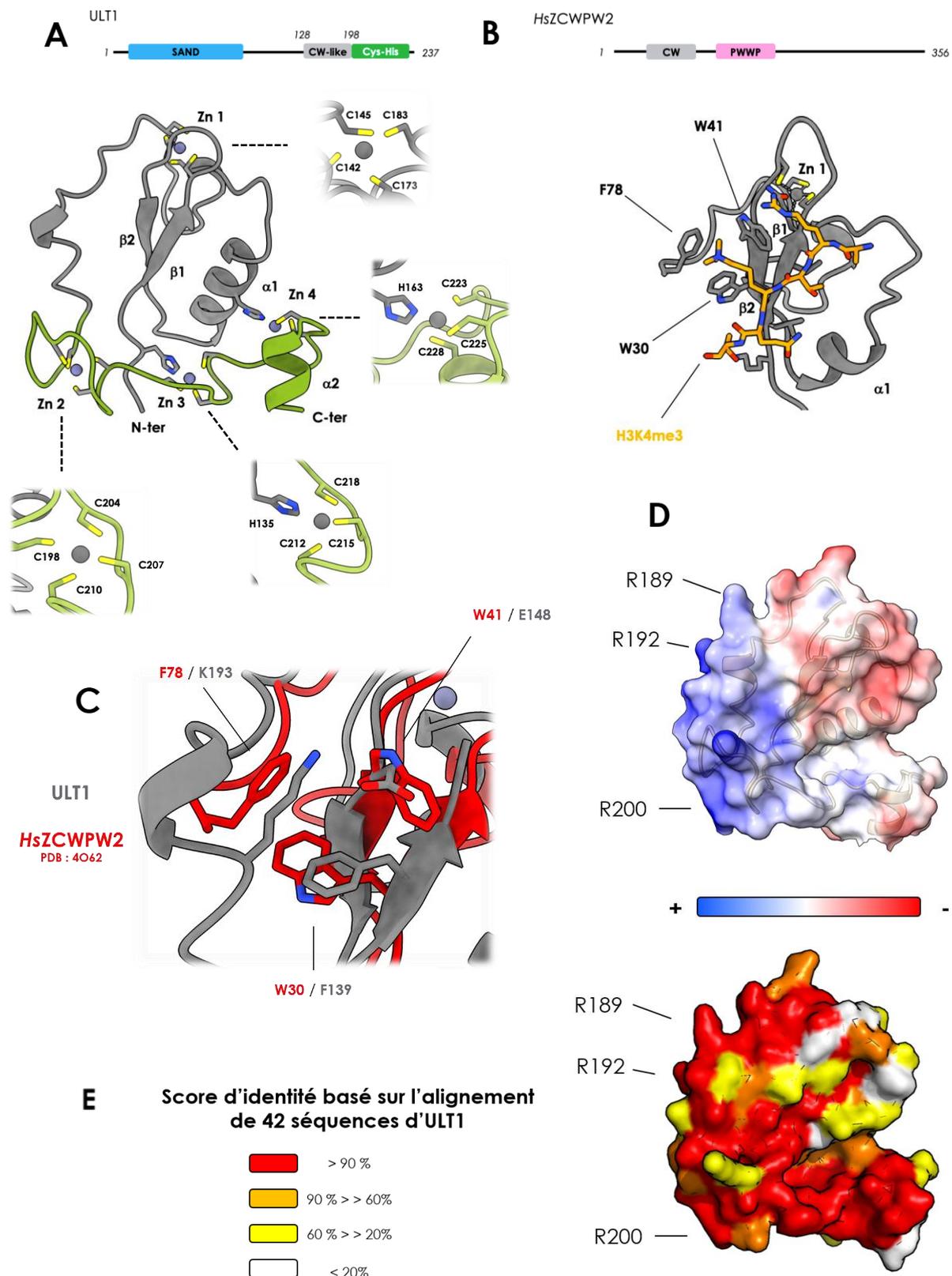
l'alignement de séquences multiples (**Figure III.7**), il est également possible qu'une partie flexible de la protéine en position N-Terminale perturbe la cristallisation. A ce jour, avec la sortie du logiciel de prédiction de structure AlphaFold2 basé sur de l'Intelligence Artificielle (Jumper *et al.*, 2021), nous n'avons pas poursuivi les essais de cristallisation pour ce domaine.

Un autre étudiant de l'équipe de Christel Carles, Vangeli Geshkovski, étudie en parallèle la capacité d'ULT1 à se lier à l'ADN. Il est donc envisageable dans l'avenir d'utiliser des acides nucléiques afin de stabiliser ce domaine et de résoudre sa structure.

## 7. Structure cristallographique du domaine C-Terminal d'ULT1

La structure cristallographique du domaine C-terminal d'ULT1 révèle la présence d'une épingle à cheveux  $\beta$  suivie de trois hélices  $\alpha$  et met en évidence deux sous-domaines stabilisés par la coordination de 4 atomes de zinc (**Figure III.8 A**).

On remarque que le premier sous-domaine qui correspond aux résidus 128 à 197 dévoile un repliement similaire aux domaines CW (**Figure III.9 A**). Il a été démontré que les domaines CW se lient spécifiquement aux marques de méthylation portées par le résidu lysine 4 de l'Histone 3 (H3K4), par l'interaction avec les acides aminés constitutifs d'une cage aromatique et avec un tryptophane conservé (Hoppmann *et al.*, 2011) (**Figure III.9 B**). Les protéines contenant ce domaine sont donc des lecteurs de cette marque. Cependant, le module de liaison aux histones dans la région C-terminale d'ULT1 est atypique (appelé par la suite CW-like). En effet, toutes les caractéristiques de cette cage aromatique ne sont pas présentes chez ULT1. La structure superposée d'ULT1 avec celle du domaine CW de *HsZCWPW2*, révèle la présence d'une lysine (K) et d'un acide glutamique (E) au lieu d'une phénylalanine (F) et d'un tryptophane (W) dans la poche de liaison présumée (**Figure III.9 C**). Un seul résidu hydrophobe est présent dans le domaine CW-Like d'ULT1 ce qui semble le rendre incapable de se lier aux résidus méthylés. Néanmoins, la présence de ce repliement CW suggère qu'ULT1 pourrait interagir avec d'autres formes de l'histone H3.



**Figure III.9 : Structure cristallographique du domaine C-Terminal d'ULT1 : (A)** Structure cristallographique du domaine C-Terminal d'ULT1 résolue à 1.5 Å. La structure dévoile deux sous-domaines stabilisés par 4 atomes de zinc. Le domaine CW-Like est représenté en gris et correspond aux résidus 128 à 197. Le deuxième sous-domaine est représenté en vert et correspond aux résidus 198 à 237. **(B)** Structure cristallographique du domaine CW de la protéine humaine ZCWPW2 (PDB 4O62) en interaction avec un

peptide H3K4me3. **(C)** Structures superposées d'ULT1 (en gris) et HsZCWPW2 (rouge), focalisées sur la cage aromatique. **(D)** Potentiel électrostatique de surface du domaine C-Terminal d'ULT1. Les régions représentées en bleu correspondent à des patchs chargés positivement tandis que les régions en rouge correspondent à des patchs chargés négativement. **(E)** Modèle du domaine C-Ter d'ULT1 montrant la conservation des résidus basée sur l'alignement de 42 séquences. R189 et K192 et R200 sont trois acides aminés chargés positivement et exposés à la surface de la protéine.

En plus du domaine CW-like, la structure cristallographique du domaine C-ter d'ULT1 a révélé un deuxième sous-domaine correspondant aux résidus 198 à 237. Cette partie de la protéine représente une boucle non-structurée mais stabilisée par la coordination de 3 atomes de zinc avec le domaine CW-Like ainsi qu'une hélice  $\alpha$  dans la région la plus C-terminale (**Figure III.9 A**).

L'analyse du potentiel électrostatique de surface pour ce domaine met également en évidence une région chargée positivement entre les résidus 189 et 205 (**Figure III.9 D**). De plus, parmi ces résidus, R189, R192 et R200 sont des acides aminés très conservés entre les différents homologues d'ULT1 chez d'autres espèces (**Figure III.9 E**). La présence localisée de résidus conservés, chargés positivement et exposés à la surface du domaine C-ter, indique qu'il pourrait lui aussi intervenir dans des interactions avec des acides-nucléiques.

## 8. ULT1<sub>128-237</sub> pourrait se lier à l'histone 3 par l'intermédiaire de modifications sur K36

La structure tridimensionnelle du domaine C-ter que nous avons récemment obtenue nous a incités à rechercher si ULT1 était capable de reconnaître des marques d'histones spécifiques. Afin de tester cette hypothèse, j'ai utilisé une matrice où sont greffés différentes combinaisons de peptides, modifiés ou non (*MODified™ Histone peptide array*, Active Motif). Cette matrice contient en duplicata, 384 peptides qui ont été générés à partir des régions N-terminales des quatre histones principales, H2A, H2B, H3 et H4 ainsi que 5 contrôles négatif.

Après avoir incubé cette matrice avec HIS-MBP-ULT1<sub>128-237</sub> la liaison de la protéine avec les différents peptides a été détectée par un anticorps dirigé contre l'étiquette poly-histidines (**Figure III.10 A1**). La matrice a révélé une liaison d'HIS-MBP-ULT1<sub>128-237</sub> pour 5 peptides d'H3 contenant les résidus 26 à 44 et qui diffèrent par la modification présente ou non sur K36. Plus précisément, ULT1 semble se lier à H3K36 mono-, di- et tri-méthylés ainsi qu'à H3K36 acétylé et non modifié. De plus il existe sur la matrice une série de peptides correspondant aux résidus 16 à 35, avec ou sans modification et pour lesquels ULT1 ne semble pas interagir. Cela signifie donc que les résidus 36 à 44 sont importants pour l'interaction.

Le fait qu'ULT1 interagisse avec H3K36 sans spécificité de modification était surprenant, mais en tenant compte de la faible conservation de la cage aromatique révélée par la structure, la première stratégie a été d'étudier la liaison d'ULT1 avec le peptide H3 non-modifié.

Il est important de noter que le peptide greffé sur la matrice correspond à la séquence d'H3.1 Humain et équivaut à une séquence hybride entre H3.1 et H3.3 d'Arabidopsis (**Figure III.10 B**). En effet, entre les résidus 26 à 44 d'H3.1 Humain, on trouve en position 31 une alanine conservée sur H3.1 d'Arabidopsis mais pas sur le variant H3.3. Tandis que la tyrosine 41 d'H3.1 Humain n'est pas présente sur H3.1 d'Arabidopsis mais est conservée sur le variant H3.3. Chez Arabidopsis, l'acide aminé présent au niveau de ces positions est très important, c'est pourquoi j'ai décidé de tester l'interaction entre ULT1 et les peptides d'H3.1 et d'H3.3. Cependant, après plusieurs tests par *pull-down*, aucune interaction entre ULT1<sub>128-237</sub> et les peptides H3 non modifiés d'Arabidopsis n'a été détectée (**Figure III.10 C**).

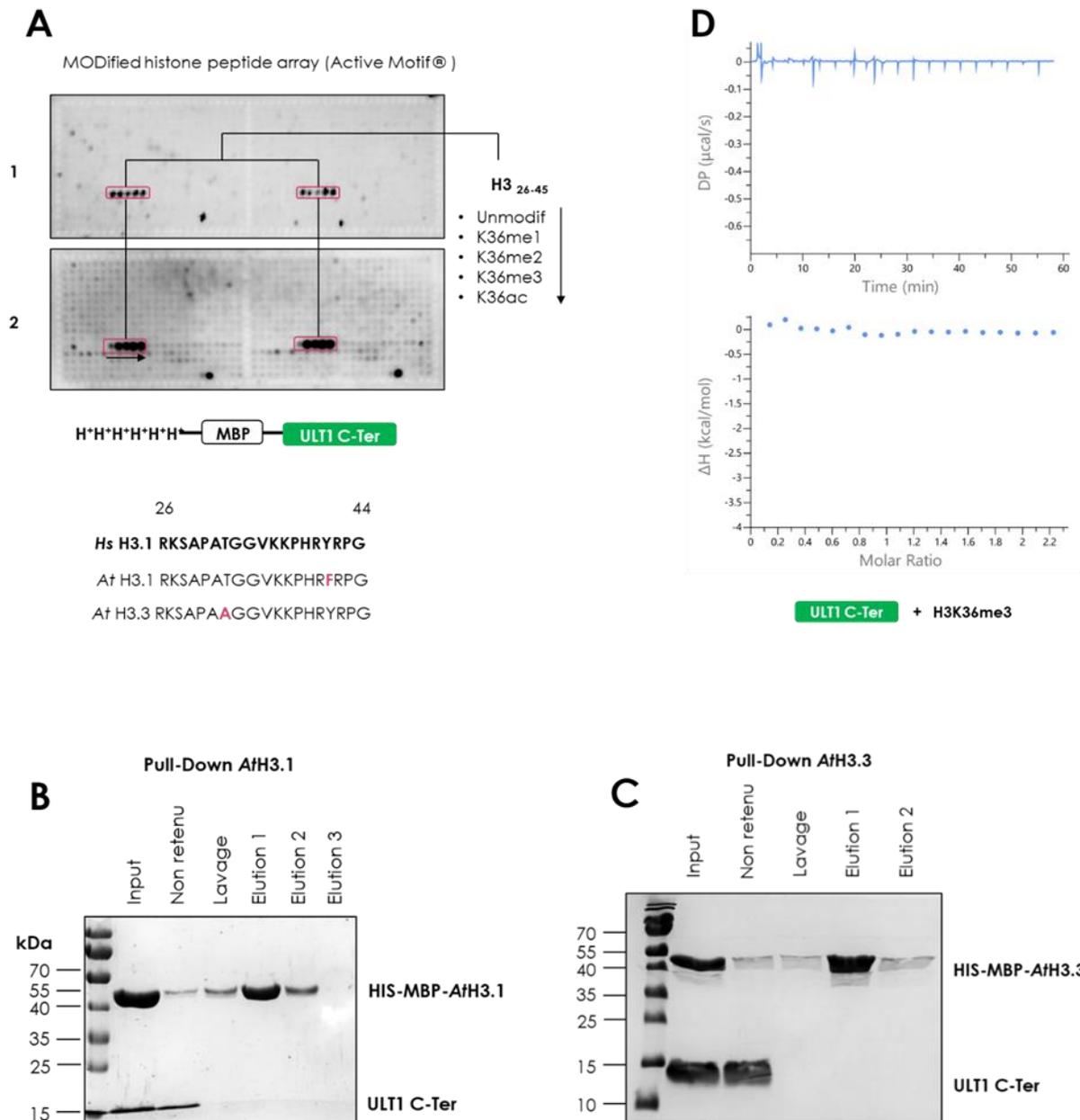
Une seconde expérience a été réalisée pour confirmer nos résultats précédents. Le résultat de l'expérience est similaire à la première et montre qu'ULT1<sub>128-237</sub> se lie toujours aux peptides comprenant les résidus 26 à 44 avec une modification sur la lysine 36 (**Figure III.10 A2**). On note cependant que la zone associée au peptide non-modifié est moins intense que précédemment ce qui suggère une faible affinité et peut expliquer les résultats du *pull-down*. Pour les autres peptides révélés, il est intéressant de noter qu'aucune intensité de signal ne semble être augmentée pour une marque spécifique, ce qui implique l'absence de site de liaison préférentiel pour ULT1.

Un nouveau peptide s'est également révélé sur la matrice et correspond à un peptide biotinylé. Nous avons pu expliquer cette interaction par les nouvelles conditions de l'expérience. L'utilisation du lait comme solution de blocage contient des protéines pouvant se lier de manière non spécifique au peptide et une reconnaissance de cette liaison par l'anticorps utilisé dans l'expérience (anti-HIS) peut expliquer le signal positif observé. Les 4 autres contrôles étaient négatifs.

L'ensemble des résultats obtenus par les deux matrices de peptides semble indiquer que le domaine C-ter d'ULT1 est capable de se lier à H3K36 méthylé ou acétylé et pourrait donc être un lecteur potentiel de marques d'histones.

Enfin, dans le but d'obtenir des informations sur l'affinité et les propriétés thermodynamiques de cette interaction, j'ai tenté de la caractériser par Titration Calorimétrique Isotherme (ITC). Pour cela nous avons utilisé un peptide synthétique correspondant au peptide H3<sub>26-45</sub> Humain, avec tri-méthylation sur K36 (H3K36me<sub>3</sub>). J'ai testé différents ratio ULT1<sub>128-237</sub>/H3K36me<sub>3</sub> ainsi que différents tampons mais aucune interaction n'a été détectée. A titre d'exemple, la **Figure III.10 D** représente un des tests

d'interaction entre ULT1<sub>128-237</sub> (75 μM) et H3k36me3 (872 μM). Il est possible que les conditions expérimentales de l'ITC n'aient pas encore été correctement établies et qu'elles nécessitent une étude plus approfondie. Une hypothèse envisageable serait qu'un certain équilibre entre ULT1<sub>128-237</sub> et le peptide soit nécessaire pour la formation du complexe ce qui expliquerait que l'on observe aucune interaction par ITC. On peut également supposer que la différence de flexibilité du peptide en solution et sur la matrice affecte l'interaction.



**Figure III.10 : ULT1<sub>128-237</sub> pourrait se lier à l'histone 3 par l'intermédiaire de modifications sur K36 : (A)** Liaison du domaine C-Terminal d'ULT1 sur les résidus 26 à 44 d'H3 avec ou sans modification sur K36. Deux matrices ont été utilisées dans des expériences indépendantes et sont notées 1 et 2 respectivement (haut). Comparaison entre *Hs*H3.1<sub>26-44</sub>, *At*H3.1<sub>26-44</sub> et *At*H3.3<sub>26-44</sub>. La séquence représentée en gras est la séquence greffée sur la matrice de peptides. Pour les séquences *At*H3.1 et *At*H3.3, la différence en acide aminé est représentée par un résidu coloré en rouge (bas). **(B et C)** Pull-down réalisés sur résine d'amylose

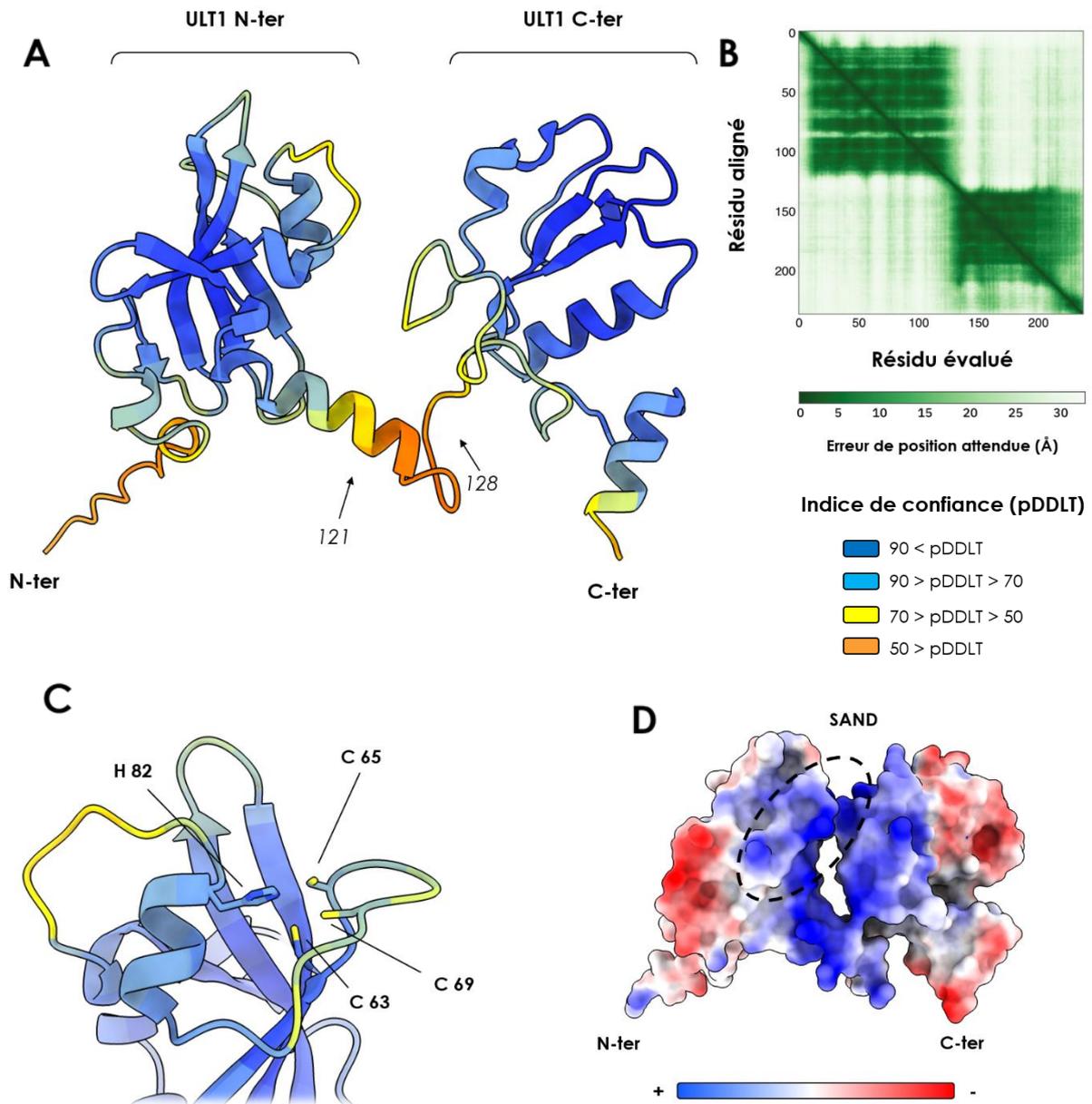
entre le domaine C-Terminal d'ULT1 et A $\theta$ H3.1<sub>26-44</sub> (B) ou A $\theta$ H3.3<sub>26-44</sub> (C) en fusion avec une étiquette HIS-MBP. **(D)** Analyse de l'interaction entre le domaine C-Terminal d'ULT1 et H $\theta$ H3.1<sub>26-44</sub> avec triméthylation sur K36.

## 9. AlphaFold dévoile la structure entière d'ULT1

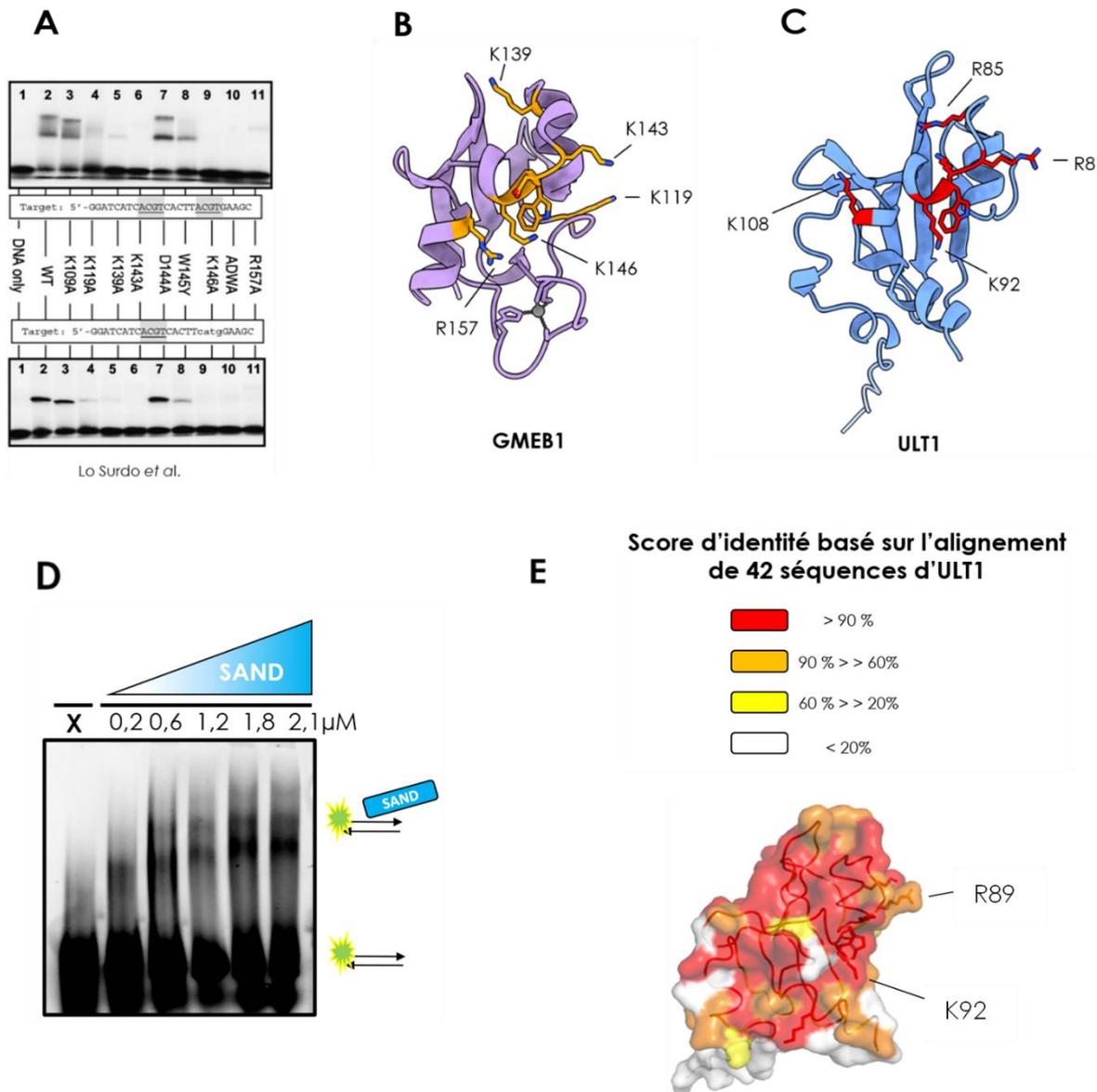
Au moment où je mettais en place des essais de cristallisation pour tenter de résoudre la structure du domaine N-terminal d'ULT1, le logiciel AlphaFold2 de prédiction de structure basée sur de l'intelligence artificielle, est devenu disponible (Jumper *et al.*, 2021). AlphaFold2 est un outil révolutionnaire capable de prédire à partir de la séquence primaire, la structure 3D des protéines ainsi que des complexes. De nombreux modèles, dont celui d'ULT1 ont été prédits avec ce logiciel et sont accessible dans une base de données chez EBI (*European Bioinformatics Institute*, <https://alphafold.ebi.ac.uk/>).

Dans le modèle AlphaFold, le bleu foncé et le bleu clair représentent une prédiction de modélisation respectivement très élevée et confiante. À l'opposé, les parties jaunes et orange sont associées à une confiance faible et très faible du modèle et peuvent être associées à des régions protéiques non structurées. Comme attendu, le modèle d'ULT1 a révélé deux domaines principaux séparés par une région non-structurée (**Figure III.11 A et B**) dont la partie N-terminale d'ULT1 non cristallisée à ce jour. La prédiction montre avec une bonne fiabilité que le domaine SAND se replie en un feuillet  $\beta$  antiparallèle torsadée et d'une série d'hélices  $\alpha$ , de manière similaire aux autres structures à domaine SAND publiés. De plus, on remarque sur le modèle que l'Histidine 82 ainsi que les cystéines 63, 65 et 69 sont regroupées et orientées d'une façon à potentiellement coordonner un ion métallique comme le zinc (**Figure III.11 C**). Le modèle AlphaFold combiné avec une analyse par analogie des structures de domaines SAND déjà publiées, dévoile également la région supposée de liaison à l'ADN, caractérisée par un patch de résidus positifs exposés à la surface du domaine N-Terminal (**Figure III.11 D**). De manière plus intéressante, ce domaine fait face à une autre zone chargée positivement sur la partie C-terminale d'ULT1 et qui a déjà été observée auparavant sur la structure résolue. Cependant, il est peu probable que les deux domaines interagissent ensemble, et une hypothèse serait qu'ULT1 pourrait fonctionner comme une pince pour l'ADN.

Le modèle AlphaFold d'ULT1 correspond également à nos données biochimiques. En effet, l'expérience de protéolyse limitée a clairement identifié une région structurée et non structurée de la protéine. Cependant, comme indiqué précédemment par l'analyse de séquence, le modèle révèle une région flexible au début d'ULT1 qui pourrait expliquer pourquoi je n'ai pas obtenu de cristaux.



**Figure III.11 : Modèle structural d'ULT1 prédit par AlphaFold2 :** (A) Structure d'ULT1 prédite par AlphaFold2 (Jumper *et al.*, 2021). La protéine est colorée en fonction de la valeur de pLDDT (*predicted local distance difference test*). Les valeurs de pLDDT > 70 représentent une prédiction de modélisation confiante tandis que les valeurs < 70 sont associées à une confiance faible. (B) Représentation de l'erreur sur la distance attendue dans la prédiction de la position du résidu X par rapport à la position du résidu Y. Une couleur verte à la position (x, y) indique que la probabilité d'erreur dans la prédiction est faible tandis qu'une couleur blanche traduit une probabilité élevée. (C) Orientation et regroupement d'H82 C65 C69 et C63 dans la structure du domaine SAND prédite par AlphaFold2. L'orientation de ces résidus suggère la coordination d'un ion métallique. (D) Potentiel électrostatique de surface du modèle AlphaFold d'ULT1. Les régions représentées en bleu correspondent à des patches chargés positivement tandis que les régions en rouge correspondent à des patches chargés négativement. La zone supposée de liaison à l'ADN pour le domaine SAND est représentée par une ellipse en pointillé.



**Figure III.12 : Comparaison des structures des domaines SAND de GMBE1 et ULT1 :** (A) EMSA réalisés dans Surdo *et al.*, et révèlent les acides aminés de GMBE1 impliqués dans l'interaction avec l'ADN. (B) Structure cristallographique de GMBE1 (PDB :1OQJ). Les acides aminés impliqués dans l'interaction avec l'ADN sont représentés en orange. (C) Structure AlphaFold du domaine SAND d'ULT1. Les acides aminés impliqués dans l'interaction avec l'ADN sont représentés en rouge. (D) Gel retard montrant le domaine SAND avec de l'ADN double-brin provenant du second intron du gène AGAMOUS. La première ligne ne contient pas de protéine et sert de contrôle négatif. Pour les autres lignes, de gauche à droite, des concentrations croissantes de protéines sont utilisées (V.Geshkovski). (E) Modèle AlphaFold du domaine SAND d'ULT1 montrant la conservation des résidus basée sur l'alignement de 42 séquences. R89 et K92 sont deux acides aminés potentiel pour des mutations

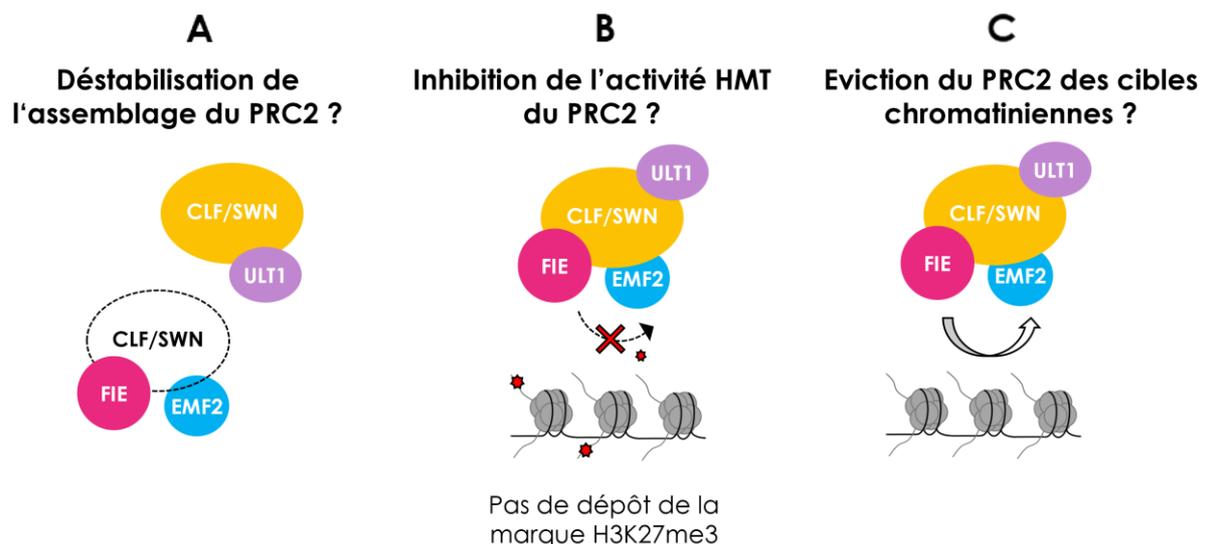
Grâce à la prédiction AlphaFold d'ULT1 j'ai pu réaliser une étude par analogie avec les autres structures à domaine SAND afin d'identifier les résidus impliqués dans l'interaction avec l'ADN. La structure cristallographique de GMEB1 dévoile les acides aminés impliqués dans l'interaction avec l'ADN et ont été validés *in vitro* par des mutations ponctuelles (**Figure III.12 A**). Dans la structure prédite du domaine SAND d'ULT1, on retrouve des acides aminés équivalents à ceux retrouvés dans GMEB1 ce qui suggèrent avec une grande probabilité qu'ils sont impliqués dans l'interaction avec l'ADN (**Figure III.12 B et C**). Dans le but d'étudier la capacité d'ULT1 à interagir avec les acides nucléiques et de savoir comment elle cible des gènes spécifiques, j'ai transmis des aliquots de protéine recombinante au doctorant Vangeli Geshkovski pour qu'il effectue des expériences *in vitro* de gels retards. Ces expériences ont révélé une interaction entre ULT1 et une séquence d'ADN de 100 pb correspondant au second intron d'AGAMOUS. Pour valider expérimentalement les résidus identifiés par la comparaison des structures des domaines SAND publiées et du modèle AlphaFold d'ULT1 mais aussi étudier leur implication dans l'interaction avec les acides nucléiques, des mutants d'ULT1 ont été conçus. En plus de l'analyse par analogie, le choix des mutations a également été supporté par une analyse de la conservation des résidus d'ULT1 chez de nombreuses espèces. Cette analyse montre la forte conservation de R89 et K92 qui sont donc deux résidus à modifier en priorité (**Figure III.12 D**).

Dans ce chapitre j'ai présenté une étude biochimique et structurale d'ULT1. J'ai réussi à mettre au point les conditions d'expression (par l'ajout de zinc) mais aussi de purification de cette protéine ainsi que de ses deux domaines structurés : ULT1<sub>1-121</sub> (SAND) et ULT1<sub>128-237</sub> (C-Ter). La structure du domaine C-Ter a ensuite été résolue par cristallographie au rayon X et a révélé un sous-domaine CW-Like pouvant potentiellement lier l'histone H3 par l'intermédiaire de modification sur K36. De nombreux efforts ont été mis en place pour valider cette interaction par ITC (et Résonance Plasmonique de Surface, SPR ; données non présentées dans ce manuscrit) mais sans succès pour le moment. Le modèle AlphaFold d'ULT1 et la structure cristallographique du domaine C-Ter ont permis d'identifier des résidus potentiellement impliqués dans l'interaction avec des acides nucléiques. Des expériences de gel retard réalisés par Vangeli Geshkovski avec de la protéine recombinante correspondant au domaine SAND, ont confirmé l'interaction de ce domaine avec de l'ADN double-brin. L'ensemble de ces résultats indique donc qu'ULT1 pourrait intervenir au niveau de ses cibles à travers la reconnaissance de modification sur les protéines histones mais aussi par la liaison à l'ADN.

La mise place des conditions d'expression et de purification d'ULT1 permet désormais de tester l'interaction avec PRC2 *in vitro*, qui constitue la seconde partie de cette thèse.

## B. Caractérisation biochimique et structurale de l'interaction entre ULT1 et le complexe PRC2

Plusieurs expériences *in vivo* (Y2H, BiFC, IP-MS) réalisées dans le laboratoire ont indiqué une interaction potentielle entre ULT1 et différents composants du complexe protéique PRC2, notamment les histones méthyltransférases CLF et SWN. Cette liaison directe entre ULT1 et le PRC2 soulève donc de nouvelles hypothèses concernant son mode d'action (**Figure III.13**). En effet, afin de conduire à la dé-répression de certains gènes cibles, ULT1 pourrait par exemple déstabiliser la formation du complexe (**Figure III.13 A**), moduler son activité enzymatique (**Figure III.13 B**), ou bien l'évincer de ses cibles au niveau de la chromatine (**Figure III.13 C**). Dans le but de mieux comprendre la fonction antagoniste d'ULT1, mon objectif était donc de caractériser cette interaction *in vitro* par une approche biochimique et structurale.



**Figure III.13 : Hypothèses sur le mode de fonctionnement d'ULT1 lors d'une interaction directe avec le PRC2. (A)** L'interaction d'ULT1 avec CLF ou SWN pourrait déstabiliser l'assemblage du complexe. **(B)** L'interaction d'ULT1 avec CLF ou SWN pourrait modifier l'activité méthyltransférase du PRC2 et ainsi limiter le dépôt de la marque répressive H3K27me3 sur la chromatine. **(C)** L'interaction d'ULT1 avec CLF ou SWN pourrait évincer le complexe de ses cibles au niveau de la chromatine.

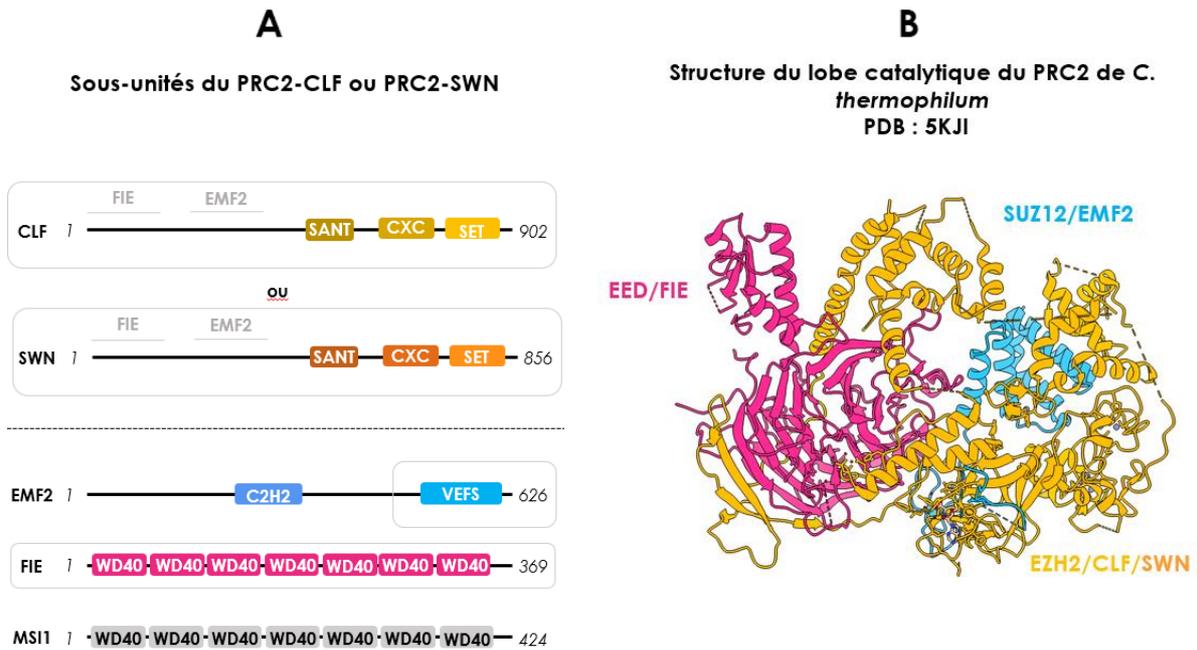
La première étape était donc de mettre au point les conditions d'expression et de purification du PRC2 d'*Arabidopsis thaliana* (AtPRC2). Une fois cette étape réalisée nous pourrions tester l'interaction avec ULT1 *in vitro* et essayer de résoudre la structure du complexe correspondant.

Pour rappel, le complexe PRC2 canonique comprend 4 sous-unités dont une histone méthyltransférase ainsi que 3 sous-unités régulatrices (**Figure III.14 A**). Chez

*Arabidopsis*, il existe plusieurs complexes PRC2 qui diffèrent selon les étapes du cycle de vie de la plante et des transitions développementales (Chanvivattana *et al.*, 2004; Mozgova and Hennig, 2015). Pour étudier l'interaction avec ULT1, nous nous sommes concentrés sur le complexe sporophytique formé par EMF2, FIE, MSI1 et CLF (ou SWN) qui est impliqué dans la répression transcriptionnelle des gènes responsables de l'identité des organes floraux comme AG.

De plus, dans une optique de caractérisation structurale, nous nous sommes posé la question de savoir quelle était la meilleure stratégie à adopter en terme de composition de complexe à produire. En effet il n'existe actuellement aucune structure publiée du PRC2 d'*Arabidopsis thaliana*. Cependant, plusieurs structures du PRC2 humain (*HsPRC2*) et de levure (*CtPRC2*) (**Figure III.14 B**) ont été résolues par cristallographie aux rayons X et par Cryo-EM (Jiao and Liu, 2015; Justin *et al.*, 2016; Kasinath *et al.*, 2018). Ces structures ont permis de mettre en évidence deux régions distinctes que sont le lobe catalytique et le lobe régulateur. Le lobe régulateur est un domaine très dynamique et plusieurs structures ne présentent pas de densité pour cette région (Poepsel, Kasinath and Nogales, 2018; Finogenova *et al.*, 2020). A l'inverse, le lobe catalytique est généralement mieux résolu et contient les sous-unités pouvant interagir avec ULT1 (**Figure I.12** de l'introduction), ce qui nous a confortés dans l'idée de travailler uniquement avec cette partie du PRC2. De plus, le lobe catalytique représente la partie minimale pour obtenir une activité méthyltransférase et constitue donc une stratégie idéale pour l'étude de l'interaction avec ULT1 (Jiao and Liu, 2015).

Notre choix s'est donc porté dans un premier temps sur l'expression dans *E. coli* du complexe PRC2 ternaire, composé de CLF, FIE et de la partie C-Terminale de EMF2 (domaine VEFS). L'expression dans *E. coli* présente de nombreux avantages tels que la facilité de mise en place et la rapidité de production. Cette approche a été encouragée par la publication d'une structure d'*HsPRC2* purifiée et cristallisée après expression dans *E. coli* (Justin *et al.*, 2016). Les constructions pour chaque sous-unité ont été conçues sur la base d'alignements de séquences entre les protéines du PRC2 de plantes et leurs homologues des espèces dont la structure PRC2 a été résolue. Dans cette stratégie j'ai co-exprimé dans deux vecteurs à double cassette, le domaine N-Terminal de CLF (1-246) avec FIE (14-369) et le domaine VEFS d'EMF2 (493-631) avec la partie C-terminale de CLF (522-902). Le choix d'exprimer CLF en deux parties permet de retirer une longue boucle qui aurait pu prévenir la cristallisation du complexe. Malheureusement, l'expression dans *E. coli* ne s'est pas révélée être une solution efficace puisque le complexe n'a pas pu être obtenu dans un état stœchiométrique.



**Figure III.14 : Stratégie d'expression du lobe catalytique des complexes PRC2-CLF et PRC2- SWN. (A)** Représentation schématique des différentes sous-unités du complexe EMF2-PRC2 chez *Arabidopsis thaliana*. Le lobe catalytique du complexe EMF2-PRC2 est composé d'une méthyltransférase qui peut être soit CLF (jaune) soit SWN (orange), de la protéine à domaine WD40 FIE (rose) et du domaine VEFS d'EMF2 (bleu). Ces sous-unités formant le complexe ternaire sont représentées par un cadre gris. Les zones d'interaction de FIE et EMF2 avec CLF ou SWN sont représentées par des barres grises. Les domaines de chaque protéine qui ont été identifiés après analyse de la base de données : Prosite sont représentés par des rectangles de couleurs. **(B)** Structure du lobe catalytique du PRC2 de *Chaetomium thermophilum* résolu par cristallographie ; PDB : 5KJI (Jiao et Liu, Science 2016). Les différentes sous-unités sont représentées selon le même code couleur qu'en (A).

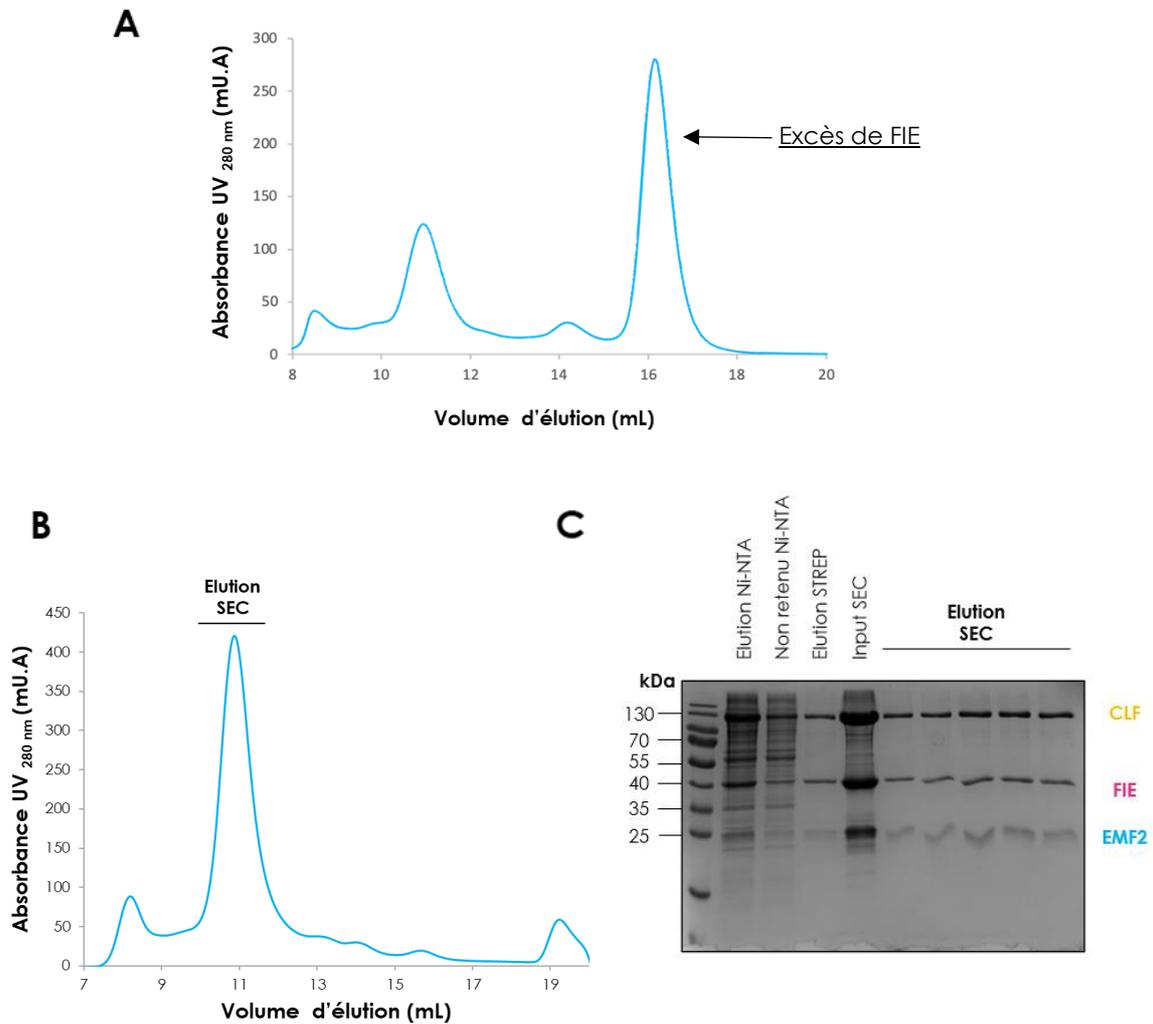
Nous avons donc décidé de changer de stratégie et de poursuivre l'expression de ce complexe dans des cellules d'insecte. Un baculovirus conçu par la méthode Multibac pour l'expression d'*At*PRC2 était déjà disponible au laboratoire. Cependant, les limites des séquences de chaque sous-unité ne correspondaient pas à celles de la structure d'*Hs*PRC2 et les différentes étiquettes de purification présentes dans la construction n'étaient pas adaptées. J'ai donc conçu de nouvelles constructions pour l'expression d'*At*PRC2 dans les cellules d'insecte.

Pour l'expression d'*At*PRC2 ternaire dans *E.coli*, j'ai travaillé avec des protéines tronquées car nous avons décidé d'éliminer les parties non conservées et/ou flexibles de la protéine qui ne sont pas visibles dans la structure cristalline d'*Hs*PRC2. Cependant, ces parties retirées des protéines peuvent être importantes pour l'interaction avec ULT1. Par conséquent, pour l'expression d'*At*PRC2 ternaire dans des cellules d'insecte, nous avons décidé de cloner les protéines CLF et FIE entières. Seule la protéine EMF2 est tronquée pour conserver le domaine VEFS, requis pour obtenir une activité méthyltransférase de PRC2 (Ketel *et al.*, 2005).

Afin de produire le complexe PRC2 ternaire d'*Arabidopsis*, J'ai utilisé le système de co-expression Multibac (voir méthodes).

### 1. Purification du complexe PRC2-CLF

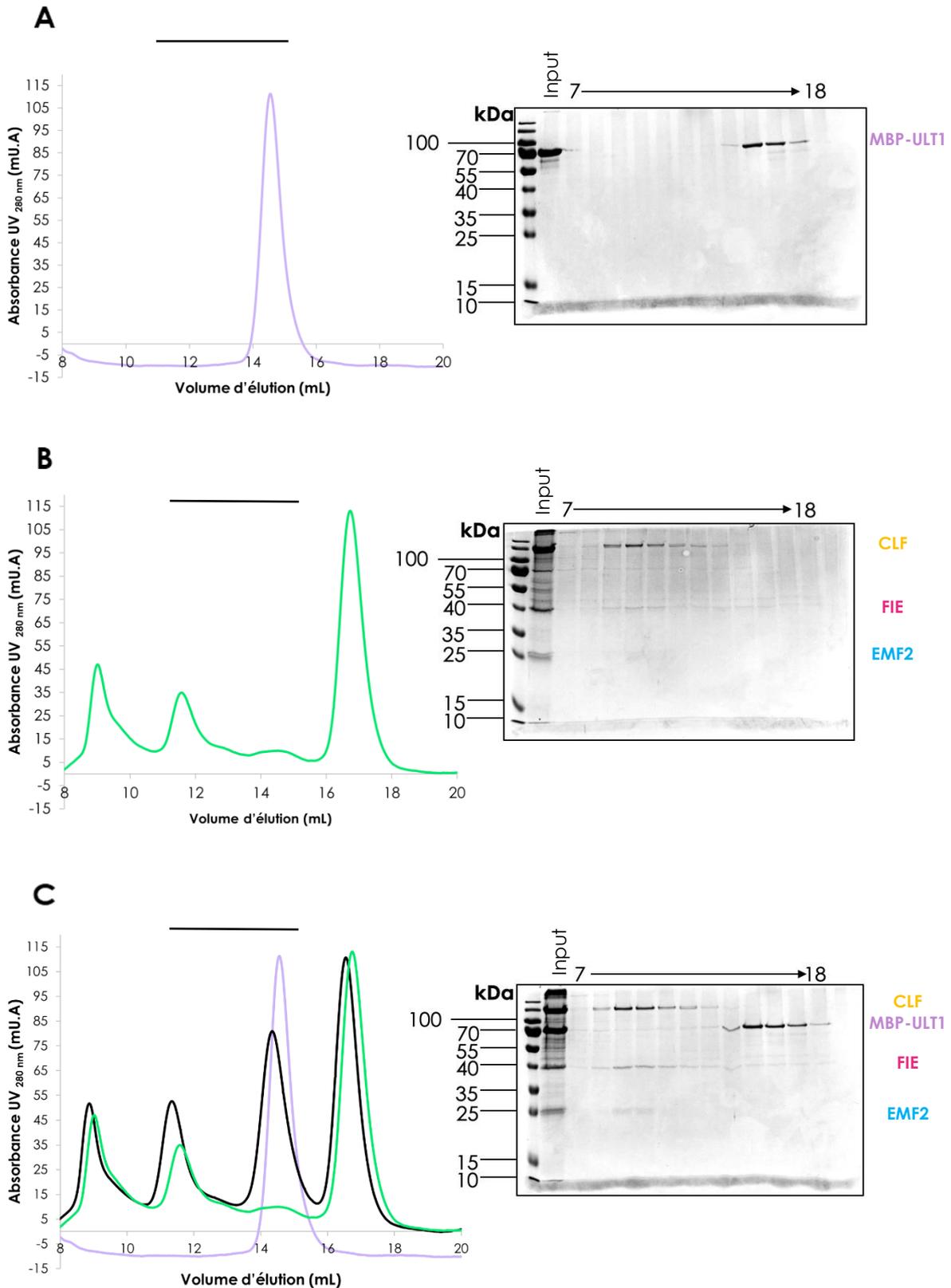
Le complexe PRC2-CLF a été exprimé dans 800 mL cellules d'insecte de type Hi5 pendant 4 jours. Une fois ce temps d'expression réalisé, les cellules sont collectées puis leur lysat est préparé et déposé sur une résine Ni-NTA. Après quelques lavages pour éliminer une majorité de protéines liées non-spécifiquement à la colonne, le complexe est élué avec 300 mM d'imidazole et la fraction collectée est déposée sur SDS-PAGE. Le gel en condition dénaturantes révèle 3 bandes majoritaires qui migrent vers 130 kDa pour CLF, 40 kDa pour FIE et aux alentours de 25 kDa pour EMF2-VEFS. Ensuite, l'échantillon est injecté sur une colonne d'exclusion de taille S200 *increase* où le chromatogramme obtenu a révélé deux pics principaux : un pic minoritaire élué vers 11 mL correspondant au complexe PRC2 et un autre élué vers 16 mL et qui est associé à un excès de FIE (**Figure III.15 A**). L'analyse sur SDS-PAGE de la fraction correspondant au complexe dévoile que les protéines CLF, FIE et EMF2-VEFS sont bien co-éluées ensemble. Cependant, afin d'éliminer les excès de FIE et d'augmenter la pureté de l'échantillon avant l'étape de filtration sur gel, j'ai décidé de rajouter une deuxième étape de purification par affinité. Après avoir purifié le complexe sur résine Ni-NTA en utilisant l'étiquette poly-histidine de CLF, l'échantillon est déposé sur une résine Strep-Tactin XT (Iba) afin de purifier le complexe grâce l'étiquette STREP sur FIE. L'ajout de cette étape a permis d'éliminer les protéines du PRC2 produites en excès mais aussi d'augmenter la pureté de l'échantillon (**Figure III.15 B et C**). La chromatographie d'exclusion de taille après les deux purifications d'affinité, présente finalement un seul pic majoritaire élué vers 11 mL et confirme la monodispersité du complexe PRC2-CLF ternaire (**Figure III.15 B**). Pour 800 mL de cellules Hi5, environ 2 mg de complexe a été purifié.



**Figure III.15 : Purification du complexe PRC2-CLF ternaire.** (A) Profil d'élution obtenu après injection de PRC2-CLF sur une colonne d'exclusion de taille : S200 *increase* (800 mL de culture de cellules). Avant cette étape, le complexe a été purifié uniquement sur une résine Ni-NTA. (B) Profil d'élution obtenu après injection de PRC2-CLF sur une colonne d'exclusion de taille : S200 *increase* (800 mL de culture de cellules). Avant cette étape, le complexe a été purifié sur une résine Ni-NTA puis sur une résine Strep-Tactin XT. (C) Analyse par SDS-PAGE des fractions collectées lors des différentes étapes de purification. Sont représentées : la fraction élue lors de la purification sur résine Ni-NTA ; la fraction non retenue par la résine Ni-NTA ; la fraction élue lors de la purification sur résine Strep-Tactin XT ; la fraction concentrée du complexe avant filtration sur gel ; les fractions collectées du pic d'élution.

## 2. Tests d'interaction entre PRC2-CLF et ULT1

Après avoir réussi à exprimer et purifier PRC2-CLF, l'objectif était de confirmer l'interaction entre le complexe et ULT1 *in vitro*.



**Figure III.16 : Test d'interaction entre PRC2 et MBP-ULT1 par filtration sur gel. (A)** Chromatogramme de MBP-ULT1 après injection de la protéine sur une colonne S200 *increase* et SDS-PAGE associé. **(B)** Chromatogramme de PRC2-CLF après injection du complexe sur une colonne S200 *increase* et SDS-PAGE associé. **(C)** Chromatogramme du mélange MBP-ULT1/PRC2-CLF après injection sur une colonne S200 *increase* et SDS-PAGE associé.

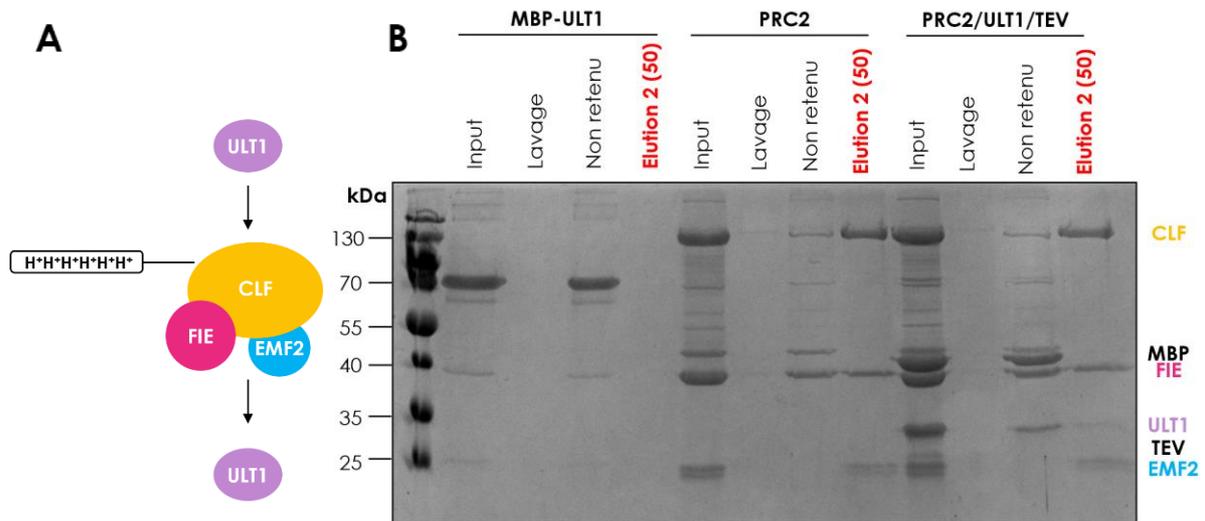
#### **a. Test d'interaction entre PRC2-CLF et ULT1 par filtration sur gel**

Ma première approche a été d'utiliser la chromatographie d'exclusion de taille afin de tester s'il y avait un décalage au niveau du volume d'élution du complexe (**Figure III.16 A, B et C**). Une élution plus rapide du pic correspondant à PRC2 signifierait probablement une augmentation de la masse moléculaire et donc une interaction avec ULT1. Il est important de souligner que cette expérience représente le premier test d'interaction entre PRC2-CLF et ULT1, et qu'elle a été réalisée à un stade où la deuxième étape de purification d'affinité du complexe n'avait pas encore été mise au point. Dans un premier temps, j'ai purifié séparément PRC2-CLF et MBP-ULT1<sub>1-237</sub> puis je les ai mélangés avec un ratio molaire 1 : 2 avant de les injecter sur une colonne d'exclusion de taille (**Figure III.16 C**). Je n'ai observé aucune co-élution significative entre le complexe PRC2-CLF et MBP-ULT1 après analyse des chromatogrammes et des SDS-PAGE associés. Cependant, l'expérience présentait certaines limites. En effet, la filtration sur gel peut être une technique destructive pour les complexes et n'est pas adaptée à la caractérisation des interactions faibles. De plus l'étiquette MBP sur ULT1 pouvait aussi prévenir l'interaction avec PRC2-CLF et donc de rendre difficilement observable la formation du complexe.

#### **b. Test d'interaction entre PRC2-CLF et ULT1 par pull-down sur Résine Ni-NTA**

J'ai ensuite décidé de tester l'interaction entre PRC2-CLF et ULT1 par *pull-down* (**Figure III.17**). Ma stratégie était de purifier les protéines séparément, de les mélanger, de couper l'étiquette MBP d'ULT1 avec la TEV et d'injecter l'ensemble sur une résine Ni-NTA. Dans cette expérience, j'ai choisi de cliver l'étiquette MBP afin de tester l'interaction uniquement entre ULT1 et PRC2-CLF. De plus, le produit du clivage a été mélangé avec PRC2 car l'étape de séparation de la MBP avec ULT1 sur colonne d'héparine n'avait pas encore été mise au point à ce stade d'avancement de la thèse. Après avoir lavé la colonne avec 10 volumes de tampon, l'élution du mélange a été réalisée à l'aide d'un tampon contenant 50 mM d'imidazole. Seul le PRC2 est capable de se lier sur la colonne par l'intermédiaire d'une étiquette poly-histidine sur CLF et d'être élué par le tampon d'élution (**Figure III.17 A et B**). Dans le mélange, l'étiquette MBP et ULT1 clivées sont retrouvées majoritairement dans la fraction non-retenue, ce qui indique qu'elles ne se lient pas ou ne restent pas liées à la colonne (**Figure III.17 B**). Cependant, toujours sur la colonne où le mélange a été déposé, la seconde élution à 50 mM d'imidazole révèle une petite bande de plus de 25 kDa qui pourrait correspondre à ULT1 clivée. Cette bande se trouve principalement dans la fraction non-retenue et ne semble pas être correctement lié par

PRC2 (complexe non stœchiométrique). L'éluition à 50 mM d'imidazole pourrait donc indiquer une faible affinité d'ULT1 pour PRC2-CLF.

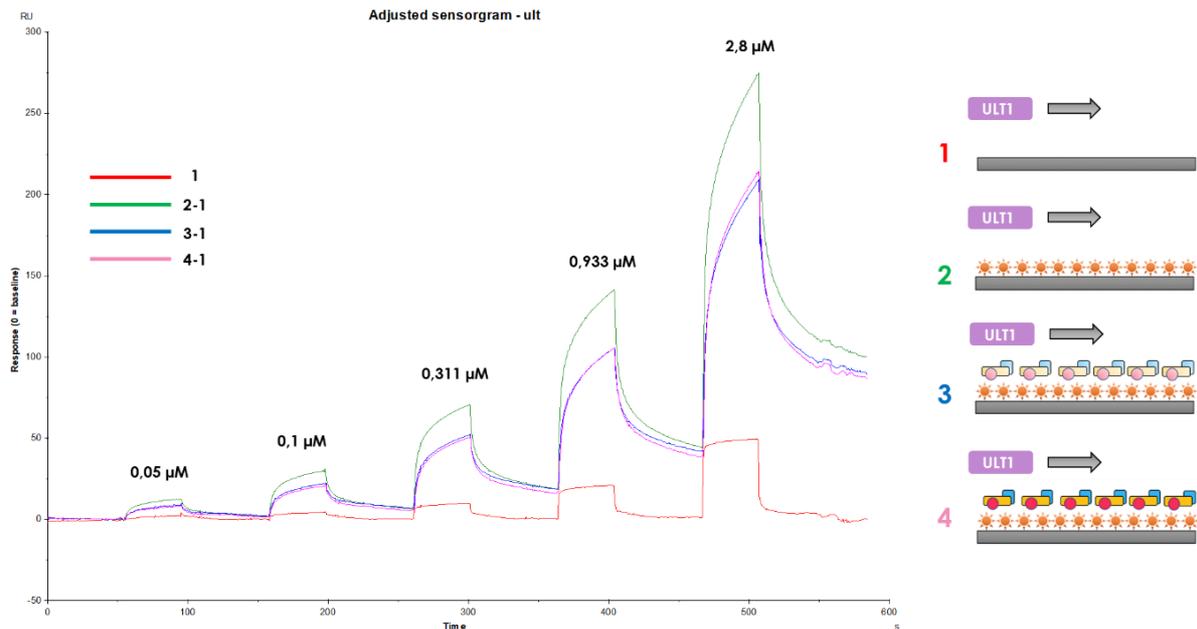


**Figure III.17 : Test d'interaction entre PRC2-CLF et ULT1 par pull-down sur résine Ni-NTA. (A)** Représentation schématique du test d'interaction entre PRC2-CLF et ULT1 par *pull-down*. PRC2-CLF va se fixer à la résine par l'intermédiaire de l'étiquette poly-histidine en position N-Terminale de CLF et ne sera élué qu'en présence de 50 mM d'imidazole. MBP-ULT1 n'a pas d'étiquette poly-histidine. **(B)** Les mélanges MBP-ULT1, PRC2 et MBP-ULT1/PRC2/TEV sont chargés séparément sur le même volume de résine Ni-NTA après une incubation sur la nuit à 4°C. L'*input*, la dernière goutte élue lors de l'étape de lavage, la fraction non-retenue ainsi que la deuxième éluition à 50mM d'imidazole (Elution 2(50)) ont été déposés sur SDS-PAGE. La première éluition à 50 mM d'imidazole correspond au volume de la colonne et rien n'a encore été élué.

### c. Test d'interaction par Résonance Plasmonique de Surface (SPR)

Pour aller, plus loin et tenter de confirmer les résultats du *pull-down*, j'ai essayé une autre méthode en utilisant la Résonance Plasmonique de Surface (SPR) (**Figure III.18**). La SPR est une méthode optique qui est utilisée pour la caractérisation d'interactions ligand/récepteur et qui fournit de nombreuses informations sur la cinétique (association, dissociation et affinité). Dans cette expérience, un ligand est immobilisé sur une puce de détection puis, une fois que le signal détecté est stable, une solution contenant le récepteur est injectée dans le système. Lorsque le récepteur s'associe (ou se dissocie) du ligand, l'indice de réfraction à la surface de la puce change et est converti en *Response Unit* par le détecteur. J'ai choisi le complexe PRC2 comme ligand (que j'ai immobilisé sur une puce Ni-NTA) et ULT1 sans étiquette comme récepteur. Quatre canaux ont été utilisés sur la puce de détection : le premier canal correspond à la surface vide de la puce, le deuxième à la surface de la puce complexée avec le nickel, le troisième à un échantillon de PRC2-CLF décongelé (échantillon stocké à -80°C), et le quatrième au complexe fraîchement purifié. Sur la **Figure III.18** les signaux des canaux 2, 3 et 4 sont normalisés par

rapport au canal 1. De manière surprenante, nous avons observé un fort signal de liaison d'ULT1 à la surface de la puce complexée avec le nickel (courbe verte) et qui peut être due à la capacité d'ULT1 de fixer les ions métalliques. Une dernière observation indique un signal de liaison moins important lorsqu'ULT1 est injecté sur PRC2, mais peut être facilement expliqué par une surface de nickel moins accessible lorsque le complexe est immobilisé. Ainsi, nous ne pouvons pas conclure sur l'interaction entre ULT1 et PRC2 par SPR en utilisant la puce Ni-NTA. Une alternative serait d'utiliser une puce avec une surface de liaison différente comme la Streptavidine (SA) et d'immobiliser PRC2-CLF par le biais de l'étiquette STREP fusionnée à FIE par exemple.

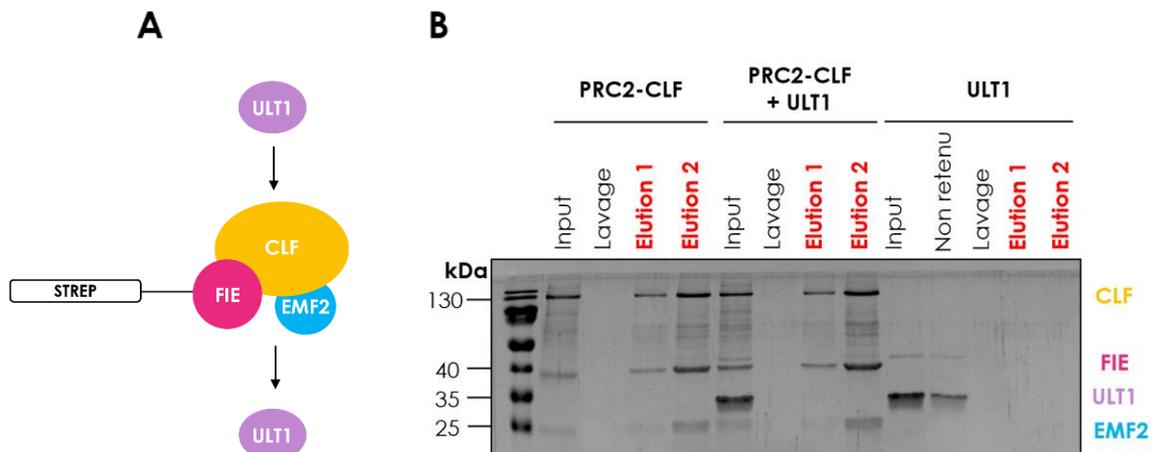


**Figure III.18 : La Résonance Plasmonique de Surface indique une interaction d'ULT1 avec le nickel.** Test d'interaction entre la protéine ULT1 et le complexe PRC2-CLF immobilisé sur puce Ni-NTA. Plusieurs concentrations d'ULT1 sont injectées dans 4 canaux différents. Le canal 1 correspond à la surface vide de la puce, le canal 2 à la surface de la puce complexée avec du nickel, le canal 3 à un échantillon de PRC2 décongelé (échantillon stocké à -80°C) et le canal 4 au complexe fraîchement purifié. Les courbes 2, 3 et 4 ont été normalisées par rapport à la courbe 1.

#### d. Test d'interaction par *pull-down* sur résine Strep-Tactin XT

Après avoir démontré qu'ULT1 avait une affinité pour le nickel, j'ai décidé de tester à nouveau l'interaction par *pull-down* mais cette fois en utilisant la résine Strep-Tactin XT. Le changement de résine d'affinité permet de lier le complexe par l'intermédiaire de l'étiquette STREP de FIE et de s'affranchir de la liaison non spécifique d'ULT1 avec la résine de nickel (**Figure III.19 A**). Après avoir purifié PRC2-CLF et ULT1 sans étiquette, trois échantillons ont été préparés séparément : ULT1 ; PRC2-CLF ; ULT1 + PRC2-CLF. Une fois ces échantillons déposés sur la résine et lavés avec 10 volumes de tampon, la protéine a été éluée à l'aide d'un tampon contenant 50 mM de biotine (**Figure III.19 B**). De manière

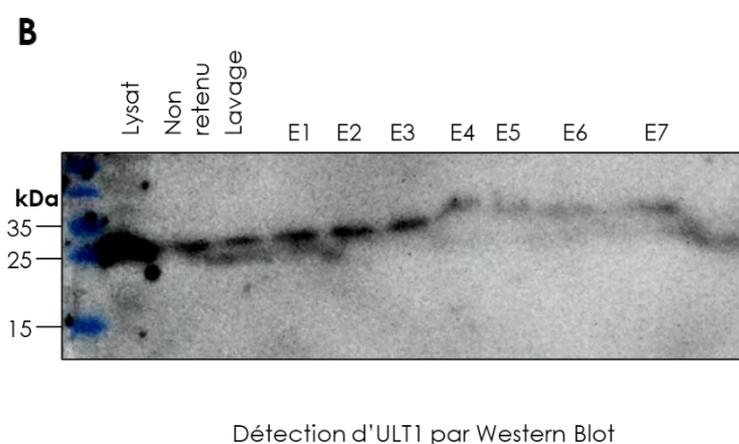
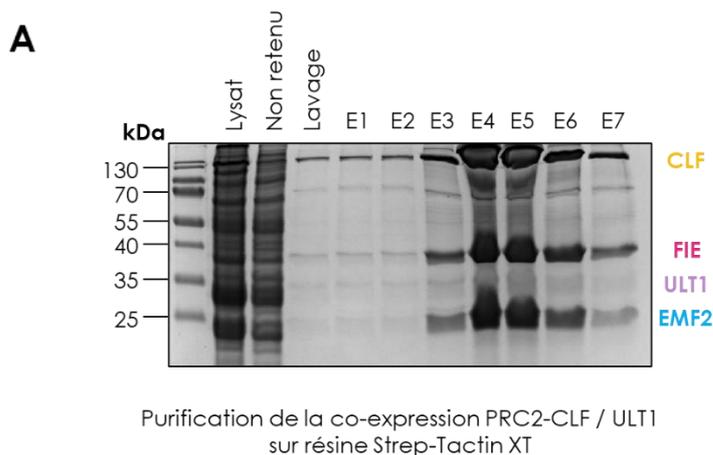
similaire au *pull-down* réalisé sur résine Ni-NTA, la fraction contrôle représentée par ULT1 sans étiquette, n'est pas retenue par la résine Strep-Tactin XT. De plus on remarque que pour la fraction éluee du mélange PRC2-CLF + ULT1, seul le PRC2-CLF est éluee. Ce résultat indique donc que dans ces conditions expérimentales, ULT1 n'interagit pas avec PRC2-CLF.



**Figure III.19 : Test d'interaction entre PRC2-CLF et ULT1 par *pull-down* sur résine Strep-Tactin XT. (A)** Représentation schématique du test d'interaction entre PRC2-CLF et ULT1 par *pull-down*. PRC2-CLF va se fixer à la résine par l'intermédiaire de l'étiquette STREP en position N-Terminale de FIE et ne sera élue qu'en présence de 50 mM de biotine. ULT1 n'a pas d'étiquette. **(B)** Les mélanges ULT1, PRC2-CLF et ULT1/PRC2-CLF sont chargés séparément sur le même volume de résine Strep-Tactin XT après une incubation d'une vingtaine de minutes à température ambiante. L'input, la dernière goutte éluee lors de l'étape de lavage, la fraction non-retenu ainsi que deux éluions à 50mM de biotine ont été déposées sur SDS-PAGE.

### e. Purification de la co-expression du PRC2-CLF et ULT1 en cellules d'insecte

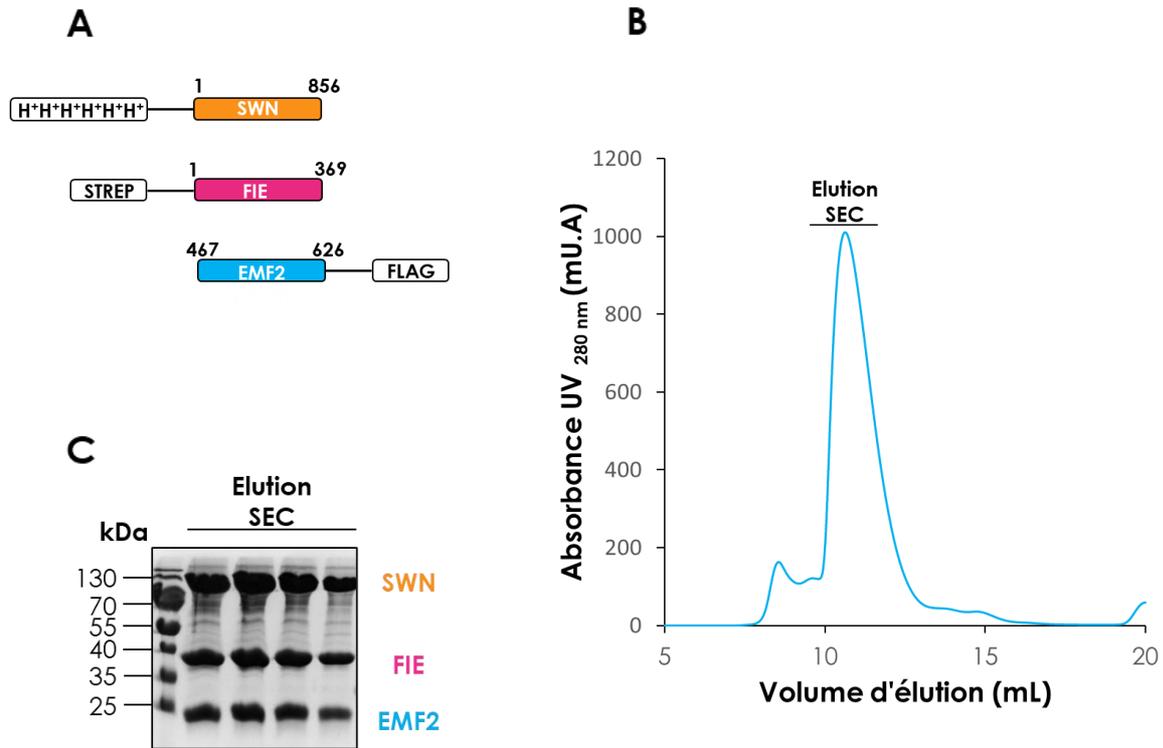
Une approche alternative qui pourrait permettre de caractériser l'interaction entre PRC2-CLF et ULT1 est leur co-expression en cellules d'insecte. Pour cela j'ai inséré la séquence entière d'ULT1 dans le vecteur pFL où était déjà cloné CLF et j'ai généré un nouveau baculovirus permettant la co-expression des 4 protéines. Une fois l'expression réalisée, j'ai lysé les cellules et déposé le surnageant sur une résine Strep-Tactin XT, lavé la colonne avec du tampon et procédé à l'éluion. Les fractions obtenues ont été déposées sur SDS-PAGE (**Figure III.20 A**), révélant la présence d'une bande aux alentours de 25 kDa pouvant correspondre à ULT1. Cette bande est majoritairement non-retenue par la colonne mais semble être également éluee avec le complexe de manière non-stœchiométrique. Un Western Blot nous a permis de confirmer la bonne expression d'ULT1 cependant la détection n'étant pas optimale il est impossible de conclure sur la co-éluion entre PRC2-CLF et ULT1 (**Figure III.20 B**).



**Figure III.20 : Purification de la co-expression entre PRC2-CLF et ULT1 sur résine Strep-Tactin XT. (A)** La fraction lysée, la fraction non-retenu, la dernière goutte éluée lors de l'étape de lavage ainsi que différentes éluions à 50 mM de biotine ont été déposées sur SDS-PAGE. **(B)** Western-Blot dirigé contre ULT1 sur les échantillons collectés lors de l'étape de purification.

Durant cette thèse, j'ai utilisé de nombreuses approches pour tenter de caractériser l'interaction entre ULT1 et PRC2-CLF. Nous avons finalement admis que cette interaction est difficilement caractérisable *in vitro* et pourrait être due à une faible interaction (faible concentration d'ULT1 et de PRC2, interaction transitoire, ...) ou à l'absence de conditions spécifiques (autre partenaire, accessibilité à une séquence spécifique d'ADN, présence de nucléosomes, ...).

### 3. Purification du complexe PRC2-SWN



**Figure III.21 : Purification du complexe PRC2-SWN ternaire. (A)** Représentation schématique des constructions utilisées pour exprimer PRC2-SWN. **(B)** Profil d'élution obtenu après injection de PRC2-SWN sur une colonne d'exclusion de taille : S200 increase (800 mL de cellules). Avant cette étape, le complexe a été purifié sur une résine Ni-NTA puis sur une résine Strep-Tactin XT. **(C)** Analyse par SDS-PAGE des fractions collectées du pic d'élution.

D'après les données dont nous disposons, il était très difficile de caractériser l'interaction entre ULT1 et PRC2-CLF *in vitro*. Cependant, comme les expériences de double hybride de levure et d'IP-MS indiquent également une interaction entre ULT1 et SWN (paralogue de CLF), j'étais curieux de tester cette interaction *in vitro* (Figure I.12 de l'introduction).

Tout d'abord, j'ai cloné et exprimé le complexe ternaire PRC2-SWN pour l'expression en cellules d'insecte, comme je l'ai fait pour PRC2-CLF (Figure III.21 A). La purification de ce nouveau complexe est similaire à celle de PRC2-CLF, mais présente des rendements plus importants, ce qui a permis de tester rapidement l'interaction avec ULT1. La chromatographie d'exclusion de taille après les deux purifications d'affinité, présente un seul pic majoritaire élué vers 11 mL et confirme la solubilité du complexe PRC2-SWN ternaire (Figure III.21 B). Le SDS-PAGE associé révèle la co-élution des trois sous-unités : SWN, FIE et EMF2, ainsi que la pureté du complexe. (Figure III.21 C)

#### 4. Tests d'interaction entre PRC2-SWN et ULT1

##### a. Test d'interaction entre PRC2-SWN et ULT1 par *pull-down*

Après avoir purifié PRC2-SWN et ULT1 indépendamment, j'ai testé leur interaction par *pull-down*, en utilisant l'étiquette STREP fusionnée à FIE (**Figure III.22 A**). Trois échantillons ont été préparés et chargés séparément sur la résine Strep-Tactin XT : PRC2-SWN ; PRC2-SWN + ULT1 ; ULT1. Une fois la résine lavée avec 10 volumes de tampon, j'ai élué les protéines en utilisant un tampon contenant 50 mM de biotine. Tous les échantillons collectés ont été chargés sur un SDS-PAGE afin d'être analysés (**Figure III.22 B**). On observe que l'échantillon contenant PRC2-SWN seul est bien élué par le tampon contenant la Biotine, ce qui signifie que PRC2-SWN est capable de se lier à la résine par l'intermédiaire de l'étiquette STREP de FIE. D'autre part, l'échantillon contenant ULT1 seule se retrouve dans la fraction non-retenue et indique que la protéine n'est pas retenue par la résine. De manière surprenante, l'expérience de *pull-down* a révélé une co-élution entre ULT1 et PRC2-SWN, ce qui démontre pour la première fois leur interaction directe *in vitro*. Ce résultat représente un véritablement changement dans l'orientation de nos recherches.

Le *pull-down* entre ULT1 et PRC2-SWN semble indiquer que la co-élution n'est pas stœchiométrique, mais cela pourrait être dû à un lavage de la résine trop rigoureux.

##### b. Test d'interaction entre PRC2-SWN et ULT1 par filtration sur gel

Pour valider cette interaction, j'ai décidé de purifier à nouveau ULT1 et PRC2-SWN puis de charger le mélange sur une colonne de chromatographie d'exclusion de taille (**Figure III.22 C et D**). Une fois que le mélange avec un excès d'ULT1 (ratio molaire 1 : 2) a été chargé sur la colonne S200 *increase*, j'ai collecté les fractions pour les analyser par SDS-PAGE (**Figure III.22 C**). On observe que le pic correspondant à PRC2-SWN seul est décalé d'environ 1 mL en présence d'ULT1, indiquant ainsi une interaction (**Figure III.22 D**). L'analyse des fractions éluées à partir de ce pic a révélé qu'ULT1 qui est habituellement élué plus tard, est bien co-élué avec le PRC2, validant ainsi l'interaction entre PRC2-SWN et ULT1 (**Figure III.22 C**). De plus, cette interaction semble stœchiométrique et stable puisque le complexe n'est pas déstabilisé. Cette stabilité du complexe indique de manière importante qu'ULT1 ne cause pas le désassemblage du complexe PRC2-SWN, ce qui réduit le nombre de nos hypothèses quant à son mode d'action pour sa fonction antagoniste à PRC2 (**Figure III.22 E**).

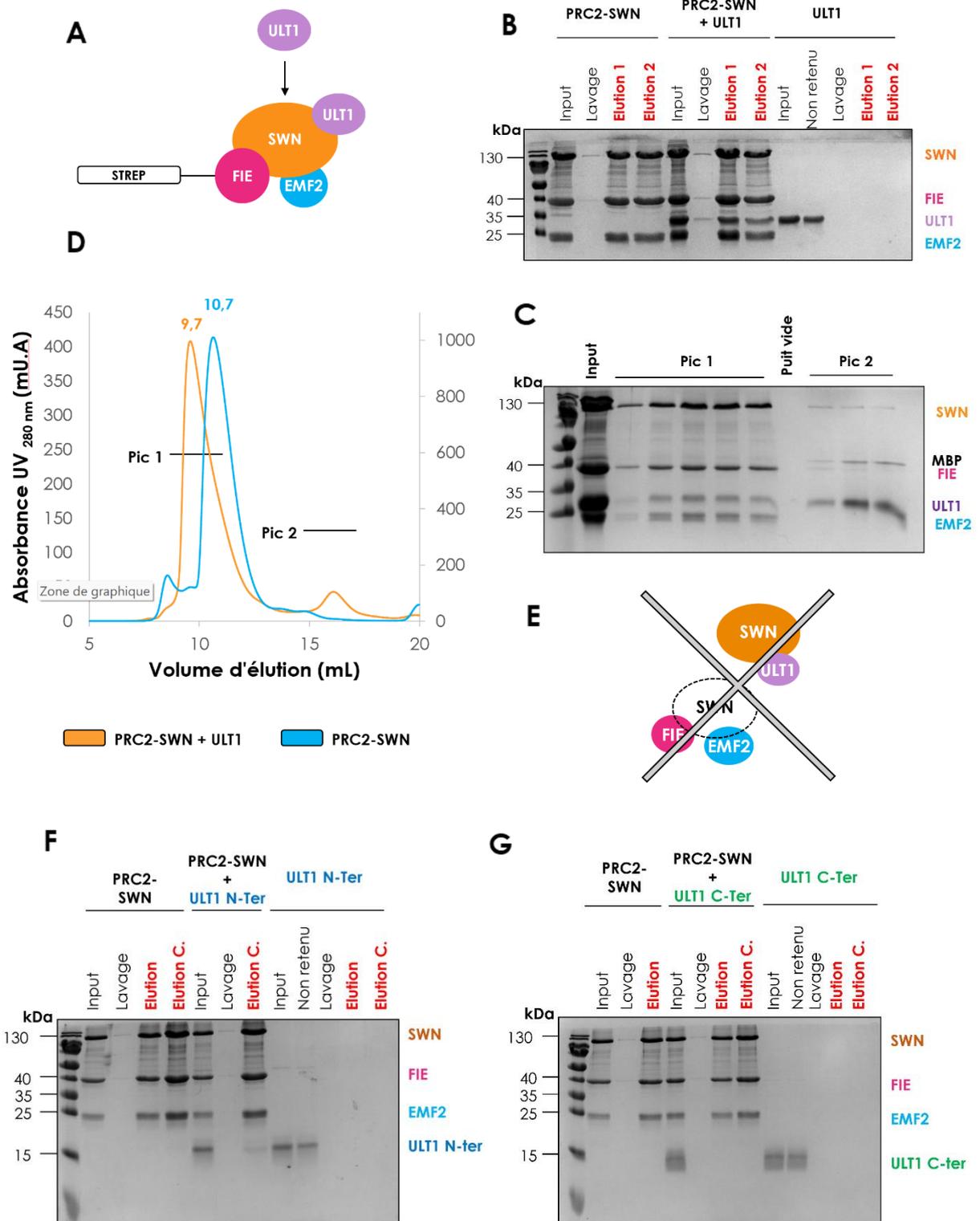
### c. Test d'interaction entre PRC2-SWN et les parties N-Ter et C-ter d'ULT1 par *pull-down*

Pour aller plus loin, j'ai voulu savoir quelle région d'ULT1 interagissait avec PRC2-SWN. Grâce à mes travaux antérieurs, je suis désormais capable de purifier les parties C-terminale et N-terminale de cette protéine et j'ai pu les utiliser pour des tests d'interaction par *pull-down* (**Figure III.22 F et G**).

L'expérience de *pull-down* réalisée avec la partie C-terminale d'ULT1 n'a pas révélé d'interaction avec PRC2-SWN (**Figure III.22 G**). Cependant, alors que la partie N-terminale d'ULT1 seule n'est pas retenue par la colonne, une bande de faible intensité correspondant à la taille attendue semble co-éluer avec le complexe (**Figure III.22 F**). Afin d'augmenter le signal sur le SDS-PAGE, plusieurs fractions éluées ont été concentrées.

Ce faible signal d'interaction obtenu pour la partie N-terminale d'ULT1 met en évidence le domaine par lequel cette protéine pourrait interagir avec PRC2-SWN, et est en accord avec les expériences de Y2H réalisées dans l'équipe, indiquant que les 193 premiers résidus d'ULT1 sont suffisants pour l'interaction avec PRC2 (PRC2-SWN et PRC2-CLF). Cependant, l'absence de stœchiométrie du complexe révélée par le SDS-PAGE semble indiquer une instabilité du domaine SAND ou bien que l'interaction nécessite une partie du domaine C-Terminal afin d'être stabilisée.

Cette découverte nous donne un nouvel aperçu de la fonction de l'ULT1 et de la façon dont cette protéine pourrait contrecarrer l'activité de PRC2. De plus, si la spécificité d'ULT1 pour SWN est confirmée, cela pourrait également révéler de nouvelles informations sur les différences fonctionnelles entre PRC2-CLF et PRC2-SWN.



**Figure III.22 : ULT1 interagit avec PRC2-SWN *in vitro*.** (A) Représentation schématique du test d'interaction entre PRC2-SWN et ULT1 par *pull-down*. PRC2-SWN est fixée à la résine par l'intermédiaire de l'étiquette STREP en position N-Terminale de FIE et ne sera élué qu'en présence de 50 mM de biotine. ULT1 n'a pas d'étiquette. (B) Les mélanges ULT1, PRC2-SWN et ULT1/PRC2-SWN sont chargés séparément sur le même volume de résine Strep-Tactin XT après une incubation d'une vingtaine de minutes à température ambiante. L'input, la dernière goutte élue lors de l'étape de lavage, la fraction non-retenue ainsi que deux éluions à 50mM de biotine ont été déposées sur SDS-PAGE. (C et D) Caractérisation du complexe PRC2-SWN / ULT1 par chromatographie d'exclusion de taille. (C) SDS-PAGE des fractions éluées du complexe PRC2-SWN/ULT1 après chromatographie d'exclusion de taille. (D) Superposition des

chromatogrammes PRC2-SWN (orange) et du complexe PRC2-SWN / ULT1 (bleu). **(E)** Représentation schématique de l'hypothèse selon laquelle ULT1 déstabilise l'assemblage du PRC2. La stabilité du complexe observée lors du *pull-down* et de la filtration sur gel ne favorise pas cette hypothèse. **(F et G)** *pull-down* réalisé entre PRC2-SWN et le domaine N-Terminal (N-Ter) **(F)** ou C-Terminal (C-Ter) **(G)** d'ULT1. Les mélanges ULT1 C -ter ou N-ter, PRC2-SWN et ULT1/PRC2-SWN sont chargés séparément sur le même volume de résine Strep-Tactin XT après une incubation d'une vingtaine de minutes à température ambiante. L'input, la dernière goutte éluée lors de l'étape de lavage, la fraction non-retenue ainsi que la fraction éluée (concentrée ou non) à 50 mM de biotine ont été déposées sur SDS-PAGE. Une étoile indique une bande à la taille attendue pour ULT1 N-Ter.

## 5. Caractérisation des complexes PRC2-CLF et PRC2-SWN/ULT1 par Cryo-EM, AlphaFold et XL-MS

Avant de mettre en évidence l'interaction entre PRC2-SWN et ULT1 *in vitro*, je me suis intéressé à la caractérisation structurale du complexe PRC2-CLF par Cryo-EM. En effet il n'existe actuellement aucune structure publiée des différents complexes PRC2 chez *Arabidopsis* et une telle information pourrait fournir des explications au niveau moléculaire sur leur différences fonctionnelles.

### a. Modélisation par AlphaFold2 des complexes ternaires PRC2-CLF et PRC2-SWN

En parallèle de la mise au point d'un échantillon de PRC2-CLF pour la Cryo-EM, j'ai réalisé des modèles AlphaFold pour les complexes PRC2-CLF et PRC2-SWN. Ces modèles ont pour but de rechercher des différences structurales mais aussi de nous aider à construire les modèles de ces structures à partir des données de Cryo-microscopie électronique. Les modèles générés à partir des séquences clonées de chaque sous-unité du PRC2 sont représentés dans les figures **III.23 A** et **III.23 B** et dévoilent une prédiction de structure avec une confiance élevée. Seules certaines régions des complexes sont prédites avec une confiance faible mais semblent correspondre à des régions non-structurées. Comme le montre la superposition de CLF et de SWN (**Figure III.23 C**), les modèles AlphaFold des complexes PRC2-CLF et PRC2-SWN ne révèlent que très peu de différences structurales. Chez l'humain, la structure du PRC2-EZH1, n'a également pas révélé de différence structurale majeure avec PRC2-EZH2, mais des variations au niveau moléculaire ont mis en évidence des différences fonctionnelles entre ces deux méthyltransférases, notamment au niveau d'une boucle entre le domaine MCSS (Motif Connecting SANT1 and SANT2) et le domaine SANT2 (Grau *et al.*, 2021b). Cependant, pour confirmer la similarité structurale des deux complexes, la résolution de leur structure par Cryo-EM est nécessaire.

Avec la capacité d'AlphaFold de pouvoir modéliser des multimères, des essais ont également été réalisés entre ULT1 et le PRC2 (avec CLF ou SWN) afin de rechercher leur

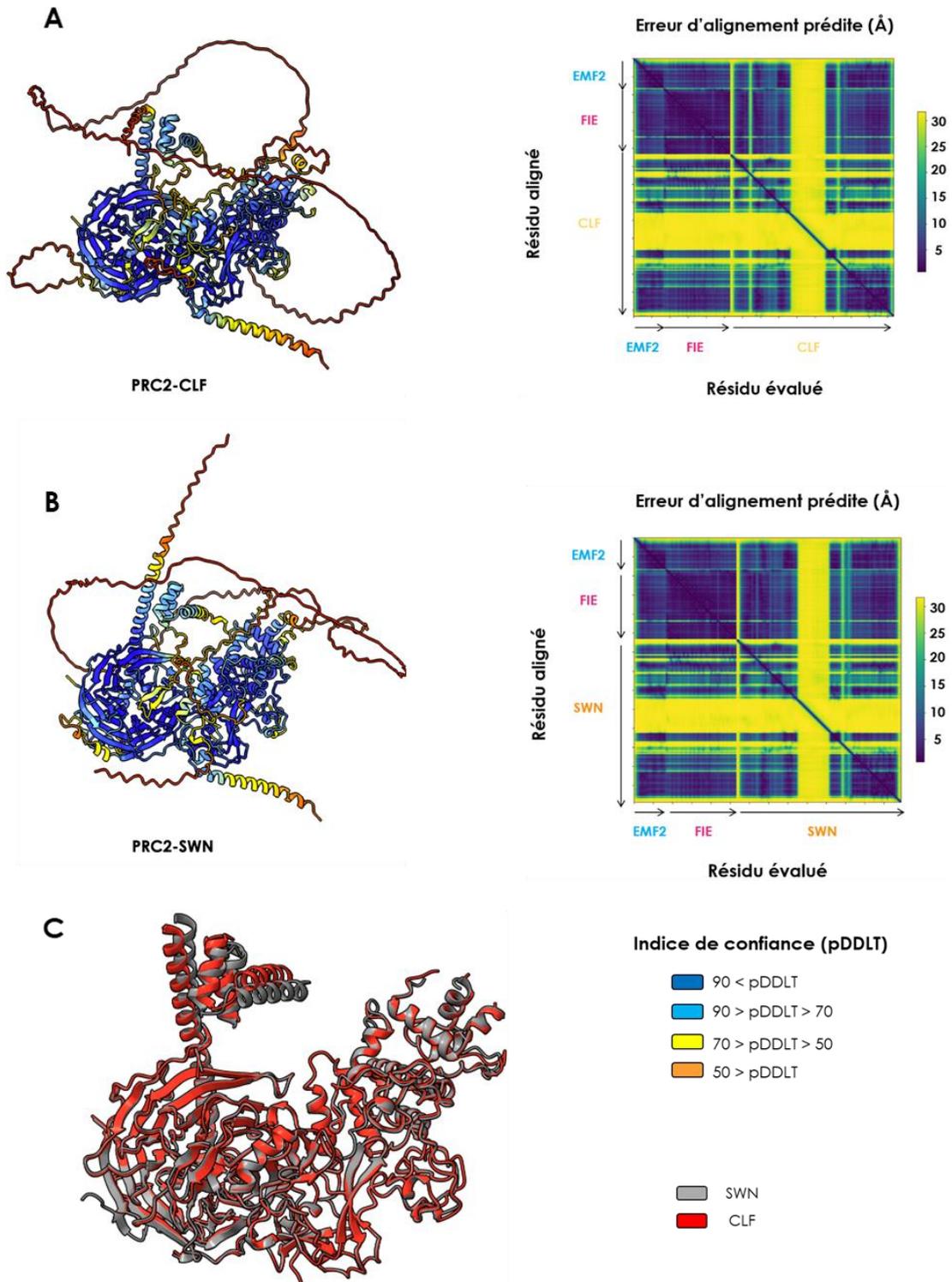
zone d'interaction. Cependant, ces modélisations ont été peu concluantes et la zone d'interaction entre ULT1 et le complexe PRC2 n'a pas pu être identifiée.

### **b. Préparation d'échantillons du complexe PRC2-CLF pour la Cryo-EM**

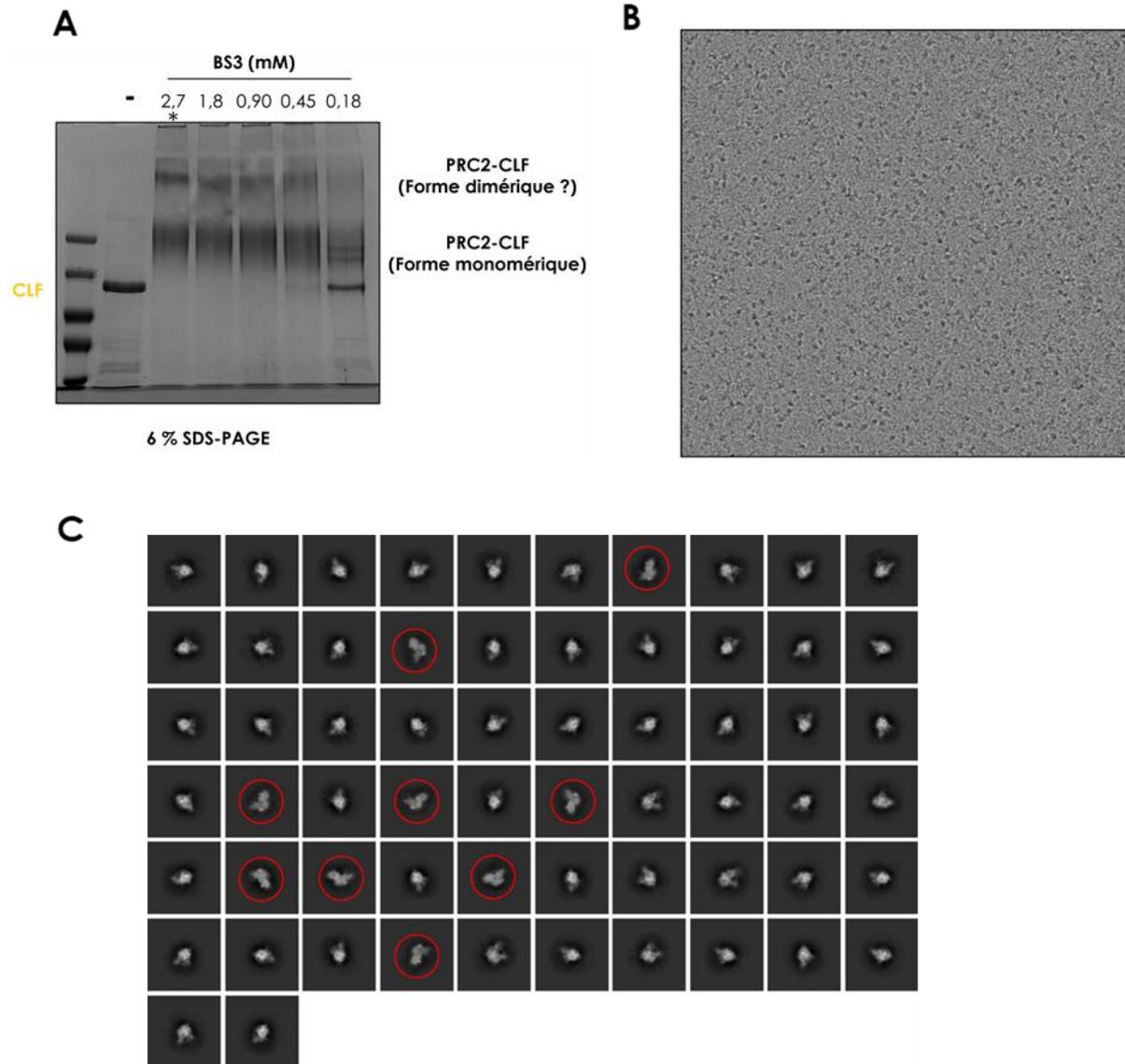
La Cryo-EM est une méthode adaptée pour l'étude structurale de larges complexes multi-protéiques. Cependant, la taille des macromolécules ainsi que l'homogénéité des échantillons sont des conditions importantes pour obtenir une résolution atomique de ces structures. Malgré le nombre de structures publiées pour le complexe PRC2, aucune structure du lobe catalytique n'a été résolue par Cryo-EM. Par conséquent, résoudre la structure du complexe PRC2-CLF ternaire représente un premier défi.

Les premières images collectées et traitées du complexe après l'étape de purification sur gel ont révélé des particules de tailles hétérogènes pouvant provenir d'une dégradation de l'échantillon lors de la congélation des grilles. Les différentes purifications du PRC2 ont montré des dégradations de la protéine CLF (et un peu plus tard de SWN) qui sont caractérisées par la présence de bandes multiples sur le SDS-PAGE. Ces dégradations sont probablement dues à une instabilité de CLF (et SWN) et peuvent fragiliser les interactions entre les différentes sous-unités.

Dans le but de rigidifier l'ensemble des sous-unités du complexe, j'ai mis en place des essais de pontages chimiques. Cette étape est largement utilisée voire s'est avérée absolument requise pour la résolution par Cryo-EM de toutes les structures du complexe PRC2. Sur la base de plusieurs publications, j'ai adapté le protocole de pontage chimique au PRC2 d'*Arabidopsis* en utilisant le substrat de bisulfosuccinimidyle (BS3). Pour une concentration de protéine donnée, j'ai testé différentes concentrations d'agents de réticulation. Après 45 minutes, la réaction a été stoppée et les échantillons ont été chargés sur un SDS-PAGE à 6 % et 15 % afin de vérifier si toutes les sous-unités étaient fixées ensemble. Le SDS-PAGE à 6 % contenant PRC2-CLF réticulé à 0,3 mg.mL<sup>-1</sup> avec différentes concentrations de BS3 est montré sur la **Figure III.24 A** et indique qu'en dessous de 0,9 mM de BS3, la sous-unité CLF n'est pas complètement réticulée. Cette expérience révèle également des bandes de poids moléculaire élevé (agrégats) ainsi que deux formes réticulées du complexe PRC2 : l'une correspondant à la forme monomérique et une autre qui pourrait correspondre à une forme dimérique. Les échantillons obtenus ont ensuite été envoyés à la plateforme de microscopie électronique de l'IBS pour préparer des grilles et collecter de nouvelles données.



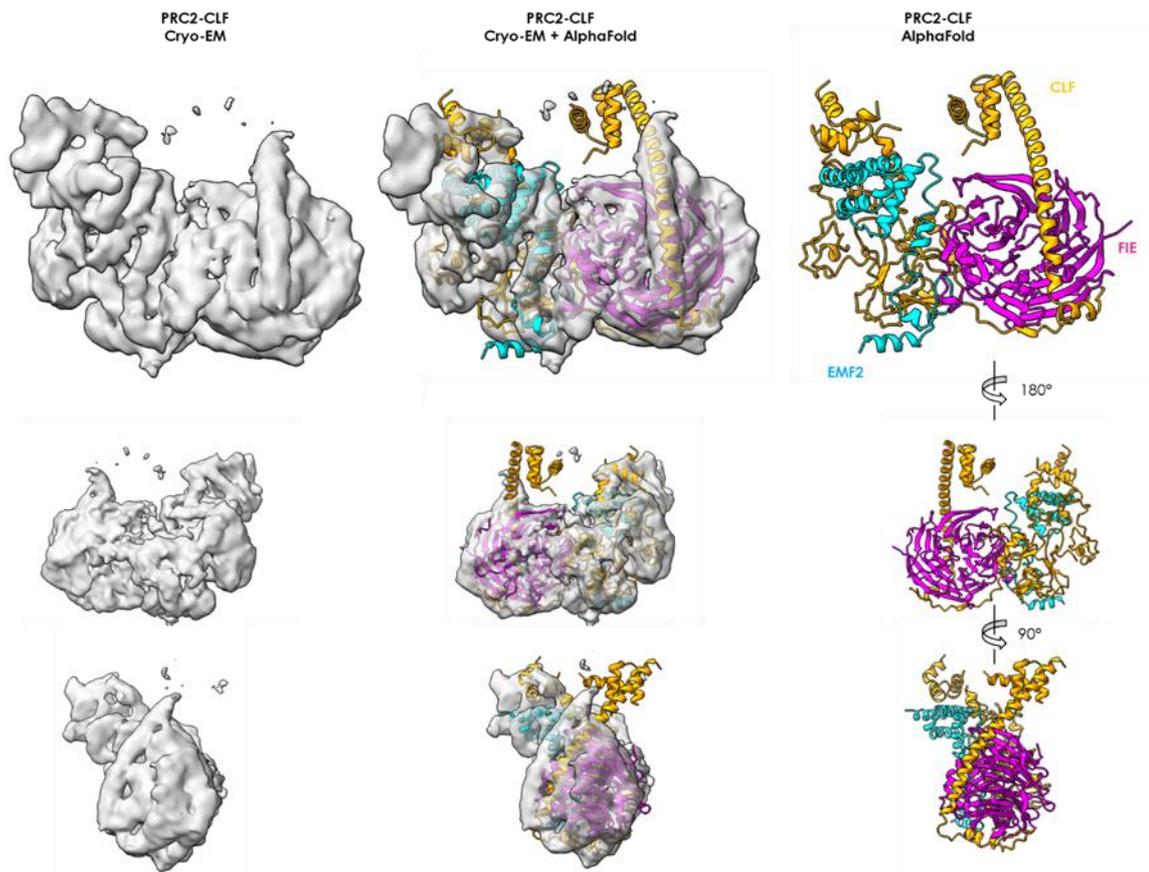
**Figure III.23 : Structures de PRC2-CLF et PRC2 SWN modélisées par AlphaFold2. (A)** Structure ternaire de PRC2-CLF (CLF ; FIE ; EMF2-VEFS). **(B)** Structure ternaire de PRC2-SWN (SWN ; FIE ; EMF2-VEFS). **(C)** Superposition de CLF (rouge) et SWN (gris) modélisés par AlphaFold2. Les résidus dont la valeur de pDDLT est inférieure à 70 ont été masqué.



**Figure III.24 : Caractérisation structurale de PRC2-CLF par Cryo-EM.** (A) SDS-PAGE à 6 % représentant la réticulation de PRC2-CLF à différentes concentrations de BS3 pendant 45 min à température ambiante. Deux populations semblent émerger aux plus fortes concentrations de BS3. Une étoile indique l'échantillon envoyé à la plateforme de microscopie électronique de l'IBS (B) Micrographe obtenu lors de la collecte de données et représentant des particules homogènes. (C) Classification 2D de PRC2-CLF à partir de 162 829 particules. Les cercles rouges indiquent les classes pour lesquelles PRC2-CLF semble complet et devrait être sélectionné pour la reconstruction d'un modèle 3D.

Comme le montre la **Figure III.24 B**, l'échantillon contient de nombreuses particules homogènes et peu d'agrégats. La classification 2D obtenue sur 162 829 particules après traitement des données sur le logiciel CryoSPARC est présentée dans la **Figure III.24 C** et révèle un grand nombre d'orientations différentes du complexe PRC2. J'ai utilisé l'ensemble de ces particules pour construire quatre modèles *Ab-initio*, suivis d'un affinement hétérogène. Parmi les quatre volumes obtenus, un seul semble ressembler au complexe PRC2 et a été de nouveau affiné, en utilisant cette fois un affinement non-uniforme. Une carte 3D avec une résolution allant jusqu'à 4,1 Å a été générée avec

seulement 32 676 particules et est présentée sur la **Figure III.25**. Malgré la faible résolution, due à un faible nombre de particules et à la flexibilité des différents domaines, nous avons pu facilement insérer un modèle AlphaFold de PRC2-CLF dans la carte 3D. Cet ajustement révèle la présence de toutes les sous-unités (CLF, FIE et EMF2), qui semblent être connectées ensemble, de la même manière que ce qui est observé pour le PRC2 humain ou de levure.



**Figure III.25 : Carte de densité 3D de PRC2-CLF obtenue par Cryo-EM.** Modèle 3D de PRC2-CLF à 4,13 Å obtenu à partir de 32 676 particules et après un affinement non-uniforme réalisé sur le logiciel CryoSPARC. Un modèle AlphaFold de PRC2-CLF a été ajusté dans la carte obtenue par Cryo-EM. Le modèle AlphaFold de PRC2-CLF a été coloré par sous-unités. CLF est en jaune, FIE en rose et EMF2 en bleu. Les régions du complexe modélisées avec un score de pDDLT < 70 ont été retirées.

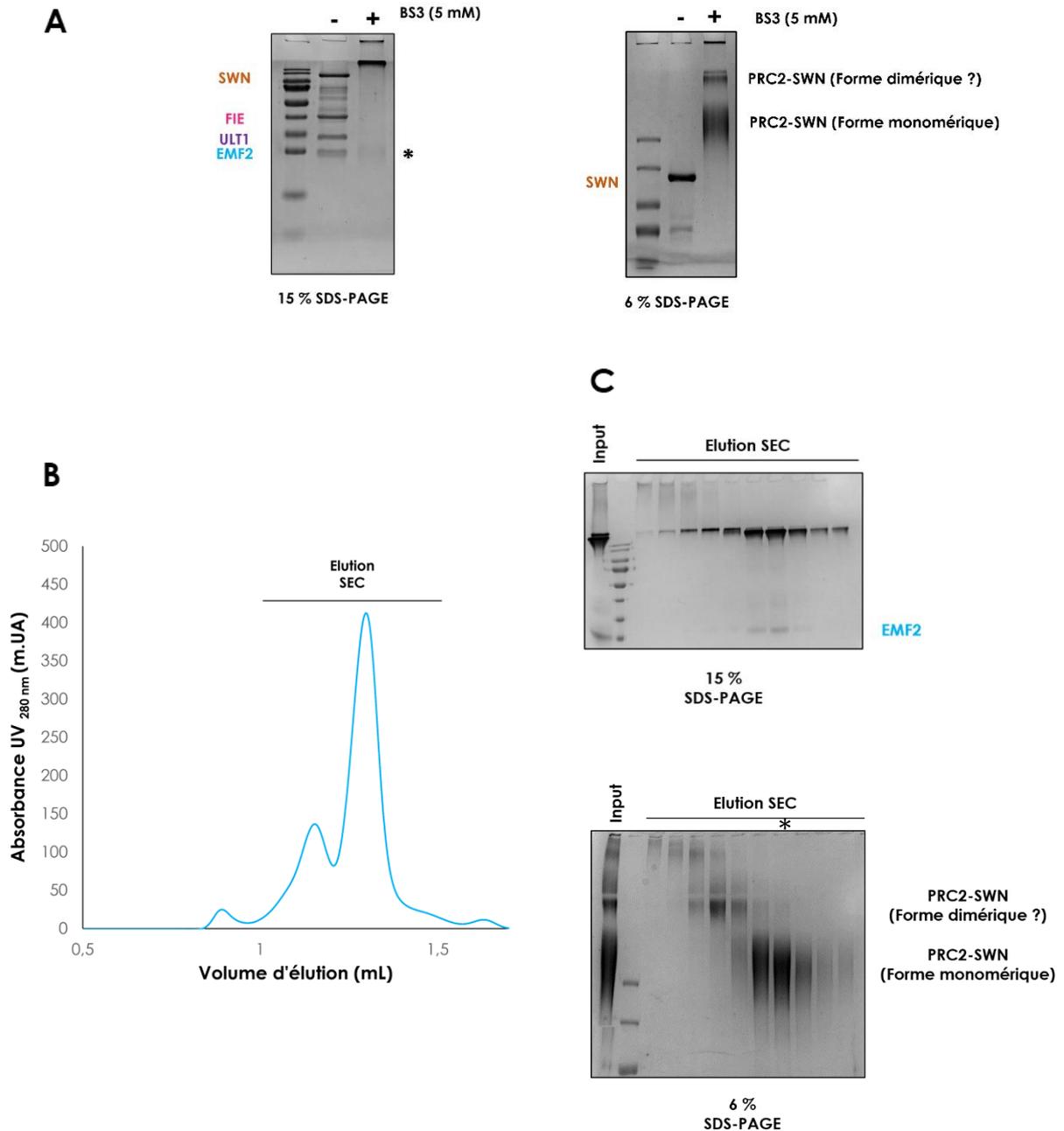
Ces résultats nous permettent désormais d'identifier quelles classes 2D sélectionner afin d'améliorer le traitement des futures données. Le modèle actuel à basse résolution n'a pas encore fourni suffisamment d'informations pour comparer les structures des complexes PRC2 d'Arabidopsis avec celles de l'Homme et de la Levure, c'est pourquoi il est nécessaire de collecter davantage d'images sur le microscope et/ou d'améliorer la concentration et l'homogénéité de l'échantillon.

### c. Préparation d'échantillons du complexe PRC2-SWN/ULT1 pour la Cryo-EM

Maintenant que j'ai démontré l'interaction entre ULT1 et PRC2-SWN *in vitro*, de nombreuses questions doivent être résolues afin de comprendre comment ULT1 contrecarre la fonction de PRC2.

Pour cartographier précisément l'interaction entre ULT1 et PRC2, nous avons choisi d'utiliser la Cryo-EM. Sur la base de mes travaux concernant la préparation des échantillons de PRC2-CLF, j'ai réalisé des tests de réticulation entre ULT1 et PRC2-SWN. Tout d'abord, j'ai décidé de tester une large gamme de concentrations de BS3 pour trouver la bonne condition pour ce nouveau complexe. L'analyse par SDS-PAGE à 15 % et 6 % des complexes réticulés révèlent qu'ULT1 semble être entièrement réticulée à PRC2-SWN, ce qui n'est pas le cas d'EMF2 à 2,7 mM BS3. Par conséquent, j'ai donc décidé d'augmenter la concentration de BS3 jusqu'à 5 mM (**Figure III.26 A**). Dans ces nouvelles conditions une petite partie d'EMF2 ne semble toujours pas être réticulé au complexe PRC2-SWN. Comme SWN semble se dégrader, il pourrait s'agir d'une région impliquée dans la liaison à EMF2, ce qui expliquerait pourquoi cette dernière n'est pas en mesure de se réticuler complètement au complexe. On remarque également la présence de deux populations principales pouvant correspondre comme pour PRC2-CLF à une forme dimérique et monomérique du complexe. La liaison d'ULT1 ne semble pas influencer la formation de ces deux populations.

Après avoir réticulé le complexe, des grilles ont été préparés au sein de la plateforme de microscopie électronique de l'IBS en vue d'une collecte de données sur le Glacios. Cependant, la congélation des grilles ne s'est pas révélée optimale, probablement en raison d'une faible concentration d'échantillon. Avec l'aide d'Hamida Laroussi, Post-Doctorante dans le laboratoire, Nous avons donc cherché à optimiser l'échantillon pour la Cryo-EM. Notre objectif était d'obtenir un échantillon plus concentré et plus homogène. Pour cela nous avons décidé de purifier les échantillons réticulés sur une colonne d'exclusion de taille : Superdex 200 *increase* 3.2/300 avant la formation des grilles. Le chromatogramme obtenu révèle deux pics principaux correspondant aux deux populations principales de l'échantillon et indique également que les fractions n'ont pas (ou peu) été agrégées lors de l'étape de concentration (**Figure III.26 B et C**). Le pic correspondant à la fraction monomérique a été utilisé pour préparer de nouvelles grilles pour la Cryo-EM et les données ont été collectées sur le microscope Krios de l'ESRF.

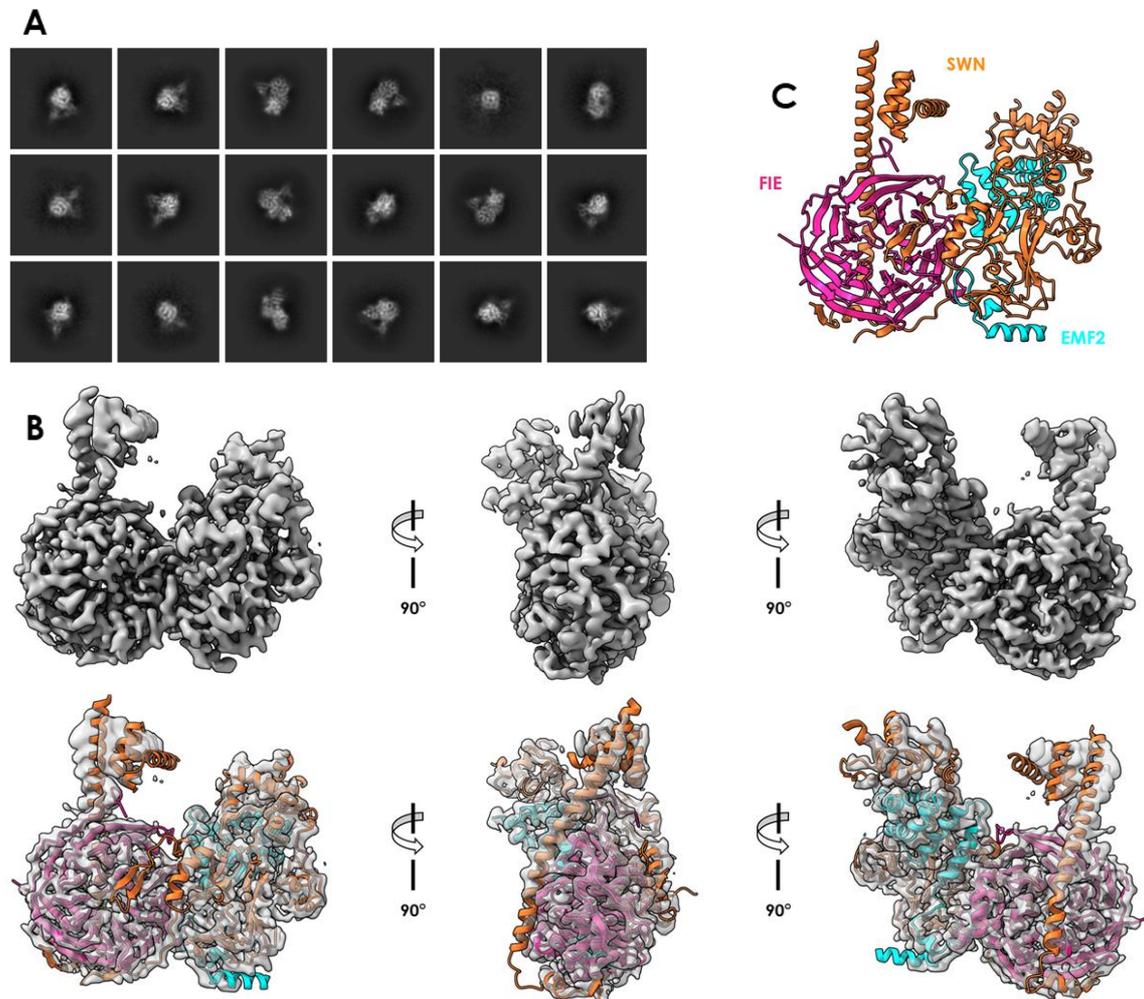


**Figure III.26 Préparation et optimisation d'échantillons pour la caractérisation structurale du complexe PRC2-SWN/ULT1.** (A) SDS-PAGE à 6 et 15 % représentant PRC2-SWN /ULT1 réticulé à 0,4 mg.mL<sup>-1</sup> avec 5 mM de BS3 pendant 45 min à température ambiante. Une étoile indique la présence d'EMF2 non réticulé au complexe PRC2-SWN (B) Profil d'éluion obtenu après injection d'un échantillon concentré de PRC2-SWN/ULT1 réticulé sur une colonne d'exclusion de taille : Superdex 200 increase 3.2/300 (C) SDS-PAGE 6 et 15 % des fractions collectées en (B). Les agrégats et les deux populations du complexe PRC2-SWN réticulé avec ULT1 semblent bien séparés. L'échantillon utilisé pour réaliser des grilles pour la Cryo-EM est symbolisé par une étoile.

Cette étape nous a permis d'augmenter la concentration du complexe et de le rendre plus homogène en séparant les différentes populations présentes dans

l'échantillon. Plus tard, cette procédure pourra également être appliquée à la résolution du complexe PRC2-CLF, et de son éventuelle interaction avec ULT1.

#### d. Structure du complexe PRC2-SWN/ULT1 par Cryo-EM



**Figure III.27 : Caractérisation structurale de PRC2-SWN-ULT1 par Cryo-EM. (A)** Sélection de 18/116 classifications 2D de PRC2-SWN/ULT1 à partir de 1 308 208 particules. **(B)** Carte 3D de PRC2-SWN-ULT1 à 3,11 Å obtenu à partir de 135012 particules. **(C)** Modèle AlphaFold de PRC2-SWN qui a été ajusté dans la carte obtenue par Cryo-EM et coloré par sous-unités : SWN est en orange, FIE en rose et EMF2 en bleu. Les régions du complexe modélisées avec un score de pDDLT < 70 ont été retirées.

Pour cette structure, le traitement des données a été entièrement réalisé par Hamida Laroussi. La classification 2D des particules a révélé une grande diversité des orientations du complexe dont une sélection sont présentés en **Figure III.27 A**. 116 classes 2D correspondant à 1 308 208 particules ont ensuite été sélectionnées pour construire une première carte 3D de PRC2-SWN/ULT1. A partir de cette carte et après plusieurs étapes de traitement, 135 012 particules ont finalement été sélectionnées et utilisées pour construire

une nouvelle carte 3D du complexe avec une résolution allant jusqu'à 3,1 Å (**Figure III.27 B**). De la même manière qu'avec PRC2-CLF un modèle AlphaFold de PRC2-SWN (**Figure III.27 B et C**) a pu être insérée dans la carte de densité et dévoile une organisation similaire. Cependant aucune densité supplémentaire correspondant à ULT1 réticulée au complexe n'a été observée et indique que la liaison de cette protéine pourrait s'effectuer au niveau d'une boucle flexible de SWN.

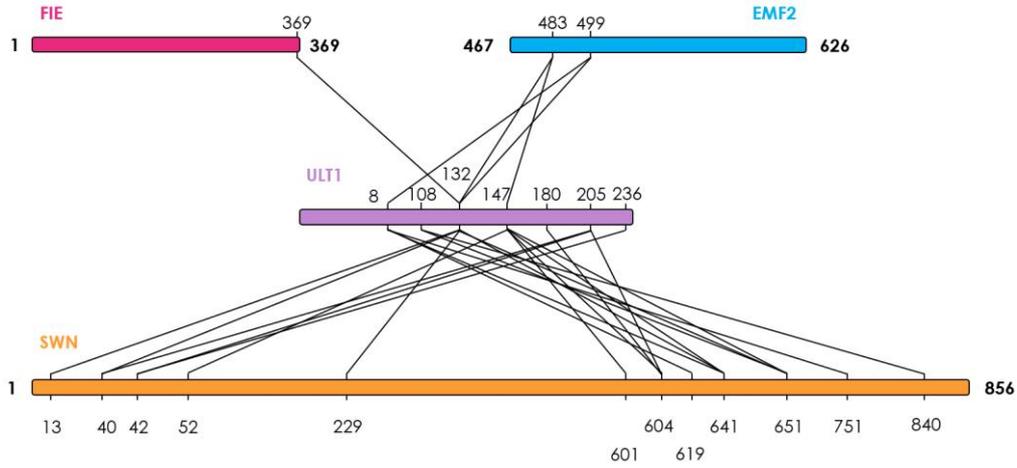
#### **e. Pontage chimique couplé à la spectrométrie de masse (XL-MS)**

Comme nous l'avons vu lors de la résolution de la structure Cryo-EM du complexe PRC2-SWN/ULT1, il est possible qu'ULT1 se lie à une partie flexible de PRC2-SWN. Dans le contexte de ce scénario, nous avons mis en place une collaboration avec l'équipe de Yohann Couté (laboratoire EdYP, Grenoble) afin de réaliser des expériences de spectrométrie de masse sur le complexe réticulé. La réticulation va lier chimiquement différentes régions du complexe et créer des contraintes de distance entre elles, permettant ainsi d'obtenir des informations structurales tridimensionnelles à basse résolution.

Le complexe PRC2-SWN / ULT1 purifié a été traité à l'urée dibutyrique disuccinimidyle (DSBU), un agent de pontage chimique réactif aux amines. Une fois réticulé, le complexe est digéré et les fragments obtenus sont analysés par spectrométrie de masse afin d'identifier les résidus réticulés. De nombreux peptides réticulés ont été identifiés mais seules les lysines impliquées dans l'interaction entre ULT1 et PRC2-SWN sont représentées dans la **Figure III.27 A et B** (réticulations inter-protéines). Les peptides contenant ces lysines ont été validés manuellement pour atteindre une confiance élevée dans les résultats.

Les pontages chimiques observés correspondent à des lysines proches dans l'espace (environ 26,9 Å entre les C $\alpha$  des lysines réticulées). Les données obtenues indiquent que plusieurs lysines dans chaque sous-unité du PRC2-SWN interagissent avec des lysines d'ULT1. On remarque que les lysines 81 et 108 situées dans la partie N-Terminale d'ULT1 sont réticulées avec les lysines 619/641/751/651 et 840 de la partie C-Terminale de SWN, tandis que les lysines 205 et 236 de la partie C-terminale d'ULT1 sont réticulées avec les lysines 42 et 40 situées dans la partie N-Terminale de SWN. Ces résultats indiquent une probable orientation du domaine N-terminal d'ULT1 vers le domaine C-terminal de SWN mais ne permettent pas de conclure sur la région précise d'interaction entre ULT1 et PRC2-SWN.

**A**



**B**

Site 1		Site 2		Nombre d'occurrences	
ULT1	81	641	SWN	26	
		751		9	
		619		6	
	108	499	EMF2	7	
		840	SWN	6	
	132	651		5	
		13	SWN	12	
		40		9	
		751		9	
		641		6	
		229		5	
	147	483		EMF2	7
		499	6		
		369	FIE		8
		180	651		SWN
	601		8		
52	5				
641	5				
604	5				
205	483	EMF2	12		
	604	SWN	14		
236	40	SWN	6		
	42		6		
	604		6		
	42	SWN	6		

**Figure III.28 : Analyse structurale du complexe PRC2-SWN/ULT1 réticulé, par spectrométrie de masse. (A)** Représentation schématique d'ULT1 (violet) et des différentes sous-unités du complexe PRC2-SWN : SWN (orange), FIE (rose) et EMF2 (bleu). Les pontages chimiques validés manuellement entre les lysines des différentes protéines (pontage inter-protéine) sont représentés par des lignes noires. **(B)** Tableau représentant les différents pontages entre ULT1 et les trois sous-unités du complexe PRC2-SWN ainsi que le nombre d'occurrences mesurées.

Dans cette deuxième partie, je me suis intéressé à la caractérisation biochimique et structurale de l'interaction entre ULT1 et le PRC2 d'*Arabidopsis*. D'après les résultats *in vivo* obtenus précédemment au sein du laboratoire, notre choix s'est porté sur l'étude de l'interaction *in vitro* avec le complexe PRC2 contenant la méthyltransférase CLF. Après une première tentative d'expression du complexe PRC2 ternaire (CLF, EMF2-VEFS, FIE) dans *E.coli*, j'ai finalement réussi à le produire dans des cellules d'insecte et à le purifier. J'ai testé l'interaction par différentes approches tel que des *pull-down*, la résonance plasmonique de surface ou la co-expression mais sans succès pour le moment. Nous avons finalement admis que cette interaction est difficilement caractérisable *in vitro* et pourrait être due à une faible affinité ou à l'absence de conditions spécifiques.

Par la suite, j'ai décidé de produire le complexe PRC2 contenant la méthyltransférase SWN et de tester l'interaction avec ULT1. Cette nouvelle approche a permis de révéler pour la première fois *in vitro* l'interaction avec ULT1 et semble mettre en évidence une différence fonctionnelle entre PRC2-CLF et PRC2-SWN. Dans le but d'identifier les résidus impliqués dans l'interaction entre ULT1 et PRC2-SWN, une étude structurale par Cryo-EM et spectrométrie de masse couplée à la réticulation a été mise en place. J'ai trouvé les conditions optimales de réticulation de l'échantillon pour ces deux techniques et une première structure du complexe PRC2-SWN/ULT1 a été obtenue. Cependant, aucune densité correspondant à ULT1 n'a été observée ce qui semble indiquer qu'ULT1 pourrait se lier à une région flexible du complexe.

En parallèle, une étude du complexe par Cryo-EM de PRC2-CLF a été initié dans l'optique de rechercher des différences structurales avec le complexe PRC2-SWN. A ce jour la résolution obtenue pour ce complexe ne permet pas d'identifier des différences structurales avec PRC2-SWN. De plus, les modèles AlphaFold de ces complexes ne semblent pas révéler de différences majeures. Pour déterminer les différences entre PRC2-CLF et PRC2-SWN des études fonctionnelles seront donc nécessaires (affinité avec la chromatine, recherche de partenaires (protéines ou acides nucléiques), activité méthyltransférase, ...).

De nombreux outils moléculaires ont été mis en place dans le laboratoire et nous permettent désormais de focaliser nos recherches sur la caractérisation de l'interaction *in vitro* entre ULT1 et PRC2-SWN.

# **IV. Discussion et Perspectives**

Chez les organismes eucaryotes, les protéines du groupe Polycomb et Trithorax fonctionnent de manière antagoniste pour réguler la dynamique de la chromatine et contrôler la transcription des gènes du développement. Alors que les protéines de chaque groupe sont de mieux en mieux identifiées et caractérisées, les mécanismes moléculaires qui permettent de commuter un gène d'un état réprimé par le PRC2 vers un état actif sont encore peu connus. Chez *Arabidopsis*, notre équipe a identifié la protéine ULTRAPETALA1 (ULT1), et a mis en évidence par des expériences de génétique et de biologie moléculaire, sa fonction antagoniste au complexe PRC2. De plus des tests d'interaction *in vivo* par Y2H et BiFC ont révélé qu'ULT1 pouvait interagir directement avec les méthyltransférases CLF et SWN ce qui indique qu'elle pourrait perturber la fonction de ce complexe. En ce sens, plusieurs hypothèses de travail ont été formulées concernant son mode d'action, dont certaines que j'ai testées pendant ma thèse de doctorat.

Afin de conduire à la dé-répression de certains gènes cibles, ULT1 pourrait par exemple déstabiliser la formation du complexe PRC2, moduler son activité enzymatique, prévenir le recrutement de protéines accessoires ou bien l'évincer de ses cibles au niveau de la chromatine. Cependant, l'absence de donnée biochimique et structurale sur ces protéines chez les plantes rend l'étude de leur fonction difficile à étudier. C'est donc dans cette optique que s'est porté mon projet de thèse. Par une approche de biologie structurale et de biochimie j'ai cherché à caractériser *in vitro* la protéine ULT1 et le complexe PRC2 d'*Arabidopsis*, et à mettre en évidence les détails moléculaires de leur interaction dans le but de mieux comprendre les mécanismes de leur fonction antagoniste.

## **A. Etude biochimique et structurale de la protéine ULT1**

Le premier résultat majeur de ce travail de thèse a été la mise en place des conditions optimales de production et de purification d'ULT1. Cette étape a été déterminante pour les études biochimiques et structurales qui ont suivi. En effet, la présence d'un très grand nombre de cystéines dans la séquence d'ULT1 était problématique pour son expression et les protéines recombinantes utilisées auparavant par notre équipe pour les expériences *in vitro* étaient obtenues en faible quantité et n'étaient probablement pas correctement repliées. Par conséquent, une caractérisation biochimique convaincante d'ULT1 et de ses interactions avec d'autres protéines et acides nucléiques n'avait pas été réalisée au sein du laboratoire.

Les prédictions bio-informatiques identifiaient deux domaines pour ULT1. Ces deux domaines ont été validés par mes expériences biochimiques et correspondent aux résidus 1-121 de la région N-Terminale (dont fait partie le domaine SAND) et aux résidus 128 à 237 de la région C-terminale riche en cystéines.

Par l'analyse du profil de filtration sur gel et la technique de MALLS, j'ai trouvé qu'ULT1<sub>128-237</sub> était monomérique en solution dans des conditions réductrices ce qui est en accord avec les résultats observés *in vitro* pour l'homologue d'ULT1 chez le riz (OsULT1) par Roy *et al.*, 2019.

Des expériences de caractérisation de l'état oligomérique d'ULT1 pleine taille et du domaine SAND ont également été réalisées. Cependant, elles ont été effectuées sur des échantillons dont le protocole de purification n'avait pas encore été correctement établi, ce qui n'a pas permis de mesurer leur poids moléculaire exact. Maintenant que la purification de ces deux constructions est bien maîtrisée dans le laboratoire, ce type d'expérience sera de nouveau réalisé.

En revanche, ces expériences *in vitro* ne permettent pas forcément de prédire l'état d'oligomérisation d'ULT1 *in vivo*. En effet, des expériences de Y2H et de BiFC réalisées dans l'équipe indiquent qu'ULT1 pourrait interagir avec lui-même. Une hypothèse révélée par des gels retards réalisés par Vangeli Geshkovski serait que la protéine ULT1 oligomériserait lorsqu'elle se lie à l'ADN (**Figure III.12.D**).

Ensuite, j'ai cherché à caractériser la structure d'ULT1 par cristallographie aux rayons X. Ayant établi la présence de deux domaines liés de manière flexible au sein de la protéine, j'ai effectué l'analyse structurale pour chacun d'eux, séparément.

Dans un premier temps, j'ai caractérisé la structure de la partie C-terminale d'ULT1 qui a révélé la présence d'un domaine CW et de trois sites de liaison pour le zinc. Les domaines CW sont connus pour se lier aux méthylations d'H3K4 par des acides aminés formant une cage aromatique avec un tryptophane conservé (Hoppmann *et al.*, 2011). Cependant, chez ULT1 d'*Arabidopsis* mais aussi chez des homologues d'autres espèces, à l'exception d'un seul acide aminé (F139), les résidus hydrophobes impliqués dans l'interaction avec les méthylations ne sont pas conservés, ce qui la rend très probablement incapable de se lier aux résidus méthylés. Néanmoins, la présence de ce repliement CW (nommé CW-Like) indique qu'ULT1 pourrait interagir avec d'autres formes de l'histone H3 ; c'est pourquoi nous avons réalisé un test d'interaction entre ce domaine et une matrice de peptides modifiés. Nos résultats ont révélé qu'ULT1 pouvait se lier à un peptide porteur de modifications sur la lysine 36 de l'histone H3 mais cette interaction n'a toujours pas été confirmée *in vitro* malgré de nombreux efforts. Il est possible que les

conditions expérimentales de l'ITC n'aient pas encore été correctement établies et qu'elles nécessitent une étude plus approfondie. Une hypothèse envisageable serait qu'un certain état d'équilibre entre ULT1<sub>128-237</sub> et le peptide soit nécessaire pour la formation du complexe ce qui expliquerait qu'on n'observe aucune interaction par ITC. On peut également supposer que la différence de flexibilité entre le peptide en solution (pour l'ITC) et celui fixé sur la matrice affecte l'interaction. Enfin une dernière hypothèse qui serait en accord avec la non conservation de la cage aromatique du domaine CW-like d'ULT1, serait simplement que l'interaction observée sur la matrice est un artefact.

Cependant, il a été démontré *in vitro* que les modifications sur H3K36 ont une influence sur l'activité catalytique du PRC2 (Schmitges *et al.*, 2011; Finogenova *et al.*, 2020). De plus comme ULT1 fonctionne comme un antagoniste à PRC2, il semble cohérent de continuer à étudier la relation entre les modifications d'H3K36 et ULT1.

Au cours de l'évolution, la fonction de liaison des méthylation d'H3K4 par le domaine CW-Like d'ULT1 pourrait avoir été perdue en raison de la présence d'un domaine CW sur la méthyltransférase de K36 chez Arabidopsis, appelée SDG8 (ou ASHH2). Dans le cas où la liaison d'ULT1 aux méthylation sur H3K36 ne peut être confirmée, une hypothèse serait qu'ULT1 participe au dépôt ce type de marque pour contrecarrer l'activité de PRC2. Enfin, la conservation de ce repliement chez ULT1 pourrait également correspondre à une autre fonction encore méconnue pour ce domaine.

Pour tester la connexion génétique entre les méthyltransférases d'H3K36 comme SDG8, et ULT1 *in vivo*, les phénotypes d'un mutant de SDG8 (*sdg8*) et du double mutant *ult1 sdg8* sont actuellement caractérisés et comparés dans le laboratoire par Vangeli Geshkovski. Il en sera de même prochainement pour SDG26. Également, le niveau de la marque H3K36me<sub>3</sub> a été vérifié par western blot chez le mutant *ult1* et aucun changement n'a été observé ; il se peut que la relation entre ULT1 et les méthyltransférases d'H3K36 soit spécifique de certains tissus ou gènes, rendant ainsi l'effet d'ULT1 modéré à l'échelle globale. Aussi, des expériences de ChIP-seq pour une cartographie génomique de la marque H3K36me<sub>3</sub> dans les différents fonds génétiques ULT1 (*ult1* mutant et lignée de surexpression ectopique *35S::ULT1*) sont prévues. Enfin, dans le but de vérifier la colocalisation entre ULT1 et H3K36me<sub>3</sub> *in vivo*, des expériences d'immuno-localisation couplées à la microscopie à épifluorescence sont en cours d'analyse.

La structure globale du domaine C-terminal d'ULT1 est plutôt unique et ne ressemble à aucune autre structure connue. Étant donné que la liaison des queues d'histones méthylées n'a pas pu être confirmée, son implication dans la liaison des histones reste incertaine. De plus, la structure ne nous a toujours pas permis de comprendre le rôle des différents domaines de liaison au zinc. Cependant, elle a révélé une grande surface chargée positivement qui pourrait être impliquée dans la liaison aux acides nucléiques. En

outre, la cartographie des résidus conservés sur la surface du domaine a également révélé un certain nombre de résidus hautement conservés et positionnés à la surface de la protéine qui sont susceptibles d'être impliqués dans des interactions avec d'autres facteurs protéiques. C'est particulièrement le cas au niveau des domaines de liaison au zinc. Afin de mieux comprendre la fonction de ces surfaces conservées, certains des résidus ont été mutés pour évaluer leur impact *in vivo*.

En ce qui concerne le domaine SAND, des essais de cristallisation ont également été réalisés mais aucun cristal n'a été obtenu à ce jour. Ceci peut être dû à la présence de la région non structurée en position N-Terminale ou bien de l'instabilité de ce domaine. Cependant, grâce à l'arrivée du logiciel de prédiction de structure basé sur l'intelligence artificielle AlphaFold, nous avons pu obtenir des informations structurales sur ce domaine mais aussi sur la protéine ULT1 entière.

La prédiction AlphaFold du domaine C-Terminal est très similaire à notre structure expérimentale, ce qui confirme l'efficacité des prédictions de structure par ce logiciel. Le domaine SAND a également été prédit avec un niveau de confiance très élevé et semble très convaincant. De plus, la prédiction de la forme pleine taille d'ULT1 est parfaitement en accord avec notre caractérisation biochimique qui avait permis d'identifier les limites qui définissent les deux domaines, ainsi que la région flexible qui les sépare. Le modèle du domaine SAND a pu être comparé à d'autres structures de domaines SAND connues, ce qui nous a également permis d'identifier les résidus susceptibles de lier l'ADN. En outre, d'autres résidus conservés en surface ont pu être identifiés.

La structure 3D d'ULT1 est désormais connue. La combinaison de nos résultats biochimiques, de la structure expérimentale du domaine C-Terminal et du modèle AlphaFold, nous a permis d'obtenir des informations détaillées sur chaque domaine mais aussi sur la protéine entière. Alors que la protéine pleine taille n'est probablement pas adaptée aux études structurales en raison de la flexibilité entre les deux domaines, sa structure pourrait être déterminée en présence d'autres partenaires protéiques ou nucléiques.

L'une des principales directions de nos futures recherches sera l'analyse de l'interaction entre ULT1 et l'ADN. Alors que des expériences réalisées dans le laboratoire ont révélé l'interaction entre ULT1 et une séquence d'ADN de 100 pb correspondant au second intron d'AGAMOUS, nous cherchons actuellement à savoir si ULT1 est capable de reconnaître des motifs spécifiques (thèse de doctorat de Vangeli Geshkovski). Ces études permettront de réaliser des tests d'affinité *in vitro* avec des fragments d'ADN plus courts et serviront également à stabiliser le domaine SAND en vue de résoudre sa structure par cristallographie aux rayons X ou RMN. Récemment, les résultats de DAP-seq et l'analyse

bio-informatique des 916 gènes identifiés par ChIP-seq pour lesquels ULT1 a un effet activateur, semblent révéler des motifs riches en GGA ou GAA. Des fragments d'ADN contenant ces motifs, la protéine ULT1 recombinante et les mutations identifiées sur la base de l'analyse structurale seront donc utilisés dans des tests *in vitro* pour déterminer comment ULT1 interagit avec l'ADN.

De plus ils se pourrait que la spécificité d'ULT1 pour l'ADN soit augmentée par l'interaction avec d'autres protéines (Pires *et al.*, 2014; Moreau *et al.*, 2016), c'est pourquoi il est également envisagé de rechercher de nouveaux partenaires. Pour cela, des expériences de ChIP-MS ou de *Proximity Labelling* (Qin *et al.*, 2021) déjà mises au point au laboratoire sont mises en œuvre et les nouveaux interacteurs identifiés seront testés *in vitro* avec la protéine ULT1 recombinante. Pour le moment, le candidat principal pour une interaction directe avec ULT1 est le complexe PRC2.

## **B. Caractérisation biochimique et structurale de l'interaction entre ULT1 et le PRC2**

Le point de départ pour l'analyse de l'interaction entre ULT1 et le complexe PRC2 a été la mise au point des conditions d'expression et de purification du lobe catalytique de ce complexe, composé des protéines CLF, FIE et du domaine VEFS d'EMF2, dans des cellules d'insectes. Le choix de ces 3 sous-unités s'inspirait de la première caractérisation tridimensionnelle du complexe PRC2 par cristallographie chez *Chaetomium thermophilum* (Jiao and Liu, 2015).

Les différents tests d'interaction réalisés par *pull-down*, filtration sur gel analytique, co-expression et SPR, n'ont pas permis de mettre en évidence une interaction directe entre ULT1 et le complexe PRC2 contenant la méthyltransférase CLF. Parce qu'une interaction entre ULT1 et SWN avait également été montrée *in vivo* par Y2H et par IP-MS, j'ai alors décidé d'entreprendre des expériences *in vitro* visant à tester l'interaction entre ULT1 et le complexe PRC2-SWN. Après avoir mis au point l'expression du complexe PRC2-SWN ternaire de la même manière que pour PRC2-CLF, j'ai pu tester *in vitro* son interaction avec ULT1. Ces expériences ont révélé pour la première fois une interaction physique entre PRC2-SWN et ULT1 mais pourrait aussi mettre en évidence une différence fonctionnelle avec PRC2-CLF. Le chromatogramme de filtration sur gel ainsi que le gels SDS associé ont permis de révéler que le complexe formé par PRC2-SWN et ULT1 est stœchiométrique et stable, ce qui indique qu'ULT1 ne dissocie pas le complexe PRC2-SWN. J'ai également réalisé des tests d'interaction *in vitro* entre PRC2-SWN et les domaines N-Terminal et C-terminal d'ULT1, qui semblent indiquer qu'ULT1 interagirait avec SWN *via* son domaine N-

Terminal ; ceci est en accord avec les expériences de Y2H réalisées précédemment par l'équipe (cf Figure 1.19, page 61). Cependant, le complexe formé *in vitro* entre la partie N-Terminale d'ULT1 et le PRC2-SWN n'est pas stœchiométrique, ce qui pourrait indiquer que le domaine C-terminal est lui aussi impliqué pour l'interaction avec SWN.

Afin de mieux caractériser l'interaction, l'affinité de liaison doit être mesurée. Mes essais préliminaires réalisés par ITC ont été compromis en raison de la précipitation de PRC2-SWN dans la cellule de l'appareil pendant l'expérience. Ces mesures seront répétées prochainement. Dans le cas où le complexe continue de précipiter, une alternative sera de tester cette interaction par SPR. Dans ce contexte, il sera également important de mesurer l'interaction avec le complexe PRC2-CLF. En effet, alors qu'aucune interaction n'a été mise en évidence par les expériences de *pull-down* analysées sur gel SDS-PAGE, si la liaison au complexe PRC2-CLF est d'une affinité plus faible, elle pourrait éventuellement être révélée par ITC.

Dans le but de comprendre les détails moléculaires de l'interaction entre ULT1 et complexe PRC2-SWN, mais aussi pour tenter d'évaluer si ULT1 pourrait interagir plus efficacement avec le complexe PRC2-SWN qu'avec le complexe PRC2-CLF, nous avons décidé de caractériser la structure de ces complexes PRC2 par Cryo-EM et spectrométrie de masse couplée à la réticulation (XL-MS). Pour ces deux techniques, j'ai dû mettre au point les conditions optimales de réticulation du complexe. Lors des essais de réticulation entre PRC2-SWN et ULT1 mais aussi lors des essais de réticulation du complexe PRC2-CLF, j'ai observé la présence de deux populations majeures sur SDS-PAGE. Une forme qui correspondrait, d'après le poids moléculaire, à un monomère du complexe et une forme qui correspondrait à un dimère. Cette forme dimérique est surprenante puisqu'elle n'a pas été observée sur les chromatogrammes lors de l'étape de filtration sur gel, ce qui indique que cette forme pourrait être transitoire et dynamique ? en solution. Il est connu que le complexe PRC2 peut former des dimères en solution via le domaine N-Terminale de SUZ12 mais aussi par la mise en présence d'ARN (Davidovich *et al.*, 2014; Chen *et al.*, 2020; Grau *et al.*, 2021b; Song *et al.*, 2023). Cependant ces conditions ne sont pas présentes dans notre système, ce qui pourrait mettre en évidence une autre zone de dimérisation transitoire en solution. La structure cristallographique de la région N-Terminale d'EZH2 en complexe avec FIE a récemment révélé une organisation trimérique avec des zones de contact au niveau des domaines SBD et SANT1 (Jiao *et al.*, 2020). Nous avons pu séparer ces deux populations par filtration sur gel après réticulation du complexe ; il serait donc

possible d'identifier la stœchiométrie de ce complexe réticulé par spectrométrie de masse native ou SEC-MALLS afin de vérifier s'il s'agit bien d'un dimère de PRC2. Il est également important de préciser que la présence d'ULT1 ne semble pas avoir d'impact la présence de ces deux populations.

Les données de Cryo-EM pour la forme monomérique du complexe PRC2-SWN réticulé avec ULT1 ont permis d'obtenir une carte de densité 3D à une résolution de 3,11 Å. Cependant, aucune densité correspondant à ULT1 n'a pu être observée à ce jour. De plus, les résultats d'XL-MS ne permettent pas non plus d'identifier clairement la zone d'interaction avec ULT1. Ces résultats indiquent donc qu'ULT1 pourrait être liée à une région flexible du complexe PRC2 comme par exemple la longue boucle désordonnée entre les domaines SANT1 et SANT2, révélée par le modèle AlphaFold. Si c'est le cas, il sera difficile de résoudre la structure complète de ce complexe. Une possibilité serait de fragmenter SWN et de tester *in vitro* l'interaction avec ULT1. Enfin, des alignements de séquences multiples entre CLF et SWN chez d'autres espèces pourraient mettre en évidence des motifs spécifiques à chacune de ces protéines. Cependant, le nombre de séquences vérifiées et correctement attribuées à CLF ou à SWN est très limitée dans les bases de données.

Nous allons tenter d'identifier des fragments de SWN qui pourraient se lier à ULT1 et qui pourraient être utilisés pour une caractérisation structurale par cristallographie aux rayons X. Enfin, comme décrit ci-dessus, nous avons également préparé de multiples mutations de résidus conservés à la surface d'ULT1 et nous testerons leur impact sur la liaison au PRC2-SWN par *des pull-down*. Si les résidus d'ULT1 impliqués dans la liaison au complexe PRC2 peuvent être identifiés, l'impact de leur mutation et donc de l'inhibition spécifique de l'interaction seront évalués *in vivo* afin d'obtenir de nouvelles informations sur le rôle de l'interaction PRC2-ULT1 dans la régulation épigénétique chez les plantes.

Malgré l'absence de donnée structurale sur l'interaction entre ULT1 et PRC2-SWN, pour comprendre la fonction antagoniste d'ULT1, j'ai cherché à mettre au point un test d'activité méthyltransférase. En effet, l'une des hypothèses proposées serait que l'interaction d'ULT1 avec PRC2-SWN modifie l'activité méthyltransférase de ce dernier. Pour le vérifier, j'ai utilisé un test commercial, le SAMfluoro de G-bioscience, qui avait déjà été utilisé dans le laboratoire. Ce test est basé sur la conversion du produit de la réaction de méthyltransférase, la S-adénosylhomocystéine (SAH), en une molécule fluorescente, la résorufine. Cependant, mes expériences ont révélé que le test commercial n'était pas adapté pour l'étude de l'activité de notre complexe. L'une des raisons principales est l'utilisation du complexe ternaire de PRC2. En effet, bien qu'une activité méthyltransférase ait été démontrée pour ce complexe minimal (Jiao and Liu, 2015), le signal de

fluorescence détecté par le test sur des peptides de queues d'histones était très faible et peu reproductible. La faible activité du PRC2-CLF ternaire n'est pas surprenante car certains essais avec le PRC2 humain présentés dans la littérature indiquent une activité détectée sur le complexe complet (à 4 sous-unités), qui est améliorée lorsqu'il se lie au nucléosome ou lorsque des protéines accessoires sont présentes (Finogenova *et al.*, 2020; Kasinath *et al.*, 2021). Un meilleur test d'activité nécessitera donc le complexe PRC2 quaternaire composé de CLF ou SWN, EMF2, FIE et MSI1 mais aussi d'un nucléosome comme substrat, dont la production est établie dans l'équipe de Carlo Petosa à l'IBS. Pour initier ces futures expériences, j'ai déjà réalisé le clonage de MSI1 mais seul le clonage de la forme pleine taille d'EMF2 n'est pas encore disponible dans le laboratoire.

Chez *Arabidopsis* mais aussi chez les plantes en général, la structure des complexes PRC2 ne sont pas connues et pourraient présenter des différences structurales à la fois avec leurs homologues chez les mammifères mais aussi entre elles de par la diversité des homologues de chaque sous-unité. Bien que les modèles AlphaFold de PRC2-CLF et de PRC2-SWN ne semblent pas dévoiler de différences structurales majeures entre elles, il nous semble important de le vérifier expérimentalement. Nous avons donc cherché à résoudre la structure des complexes PRC2-CLF et PRC2-SWN par cryo-EM et nos données semblent montrer une organisation des sous-unités similaires aux structures déjà résolues. Cependant, dans le cas de CLF, la résolution obtenue n'était pas optimale pour observer ces différences, probablement en partie dû à un faible nombre de particules sélectionnées. De plus, l'échantillon utilisé pour la Cryo-EM n'était pas aussi homogène que celui de PRC2-SWN/ULT1 car la dernière étape de filtration sur gel pour séparer les deux formes du complexe et retirer le glycérol présent dans l'échantillon n'a pas été effectuée. Afin d'obtenir une meilleure structure du complexe PRC2-CLF, nous envisageons de rajouter cette dernière étape de filtration sur gel mais aussi de collecter plus d'images (sur le microscope Krios) dans le but de sélectionner plus de particules. Dans le cas des données structurales obtenues pour le complexe PRC2-SWN/ULT1, malgré une haute résolution globale de la carte de densité, le site catalytique semble moins bien défini et indique qu'il pourrait être très dynamique. Pour le stabiliser, j'avais rajouté lors de l'étape de réticulation, le peptide de stimulation allostérique (A $\theta$ H3.1<sub>20-32\_K27me3</sub>), le peptide substrat muté (A $\theta$ H3.1<sub>20-32\_K27M</sub>) et la S-adénosylméthionine (SAM). Cependant, sur la structure obtenue, la SAM et le peptide substrat ne sont pas présents et ont probablement été perdus lors de la dernière étape de filtration sur gel. Il est donc envisageable de rajouter ces deux éléments dans l'échantillon juste avant de congeler des grilles pour la microscopie.

Afin de progresser dans la caractérisation structurale des complexes PRC2, de nouvelles données de Cryo-EM sur les complexes PRC2-SWN et PRC2-CLF en présence des peptides et de la SAM seront collectées sur le Krios.

Comme semblent l'indiquer nos données de Cryo-EM et AlphaFold, il n'existe que peu de différences structurales entre les complexes PRC2-CLF et PRC2-SWN. Leur différence fonctionnelle pourrait donc s'expliquer par des divergences au niveau de certains acides aminés spécifiques ou bien par le recrutement de protéines accessoires différentes comme ULT1. Plusieurs publications chez les modèles animaux ont mis en évidence des différences d'acides aminés qui peuvent expliquer la différence fonctionnelle entre EZH2 et EZH1 (Lee *et al.*, 2018; Grau *et al.*, 2021b). Les différences spécifiques entre CLF et SWN révélées par les modèles AlphaFold et nos structures Cryo-EM seront identifiées et validées *in vitro* et *in vivo*.

Notre capacité à produire, au sein du laboratoire, les complexes PRC2-SWN et PRC2-CLF ternaires et bientôt quaternaires avec l'ajout de MSI1 et du domaine N-Ter d'EMF2, nous permettent également d'envisager l'étude biochimique des protéines accessoires de ce complexe chez les plantes. En effet, de nombreuses protéines pouvant interagir avec le complexe PRC2 ont été identifiées chez les plantes, mais le faible nombre de caractérisations biochimiques et l'absence de données structurales de ces complexes limitent grandement la compréhension des mécanismes d'action du PRC2 chez les plantes (Bieluszewski *et al.*, 2021; Godwin and Farrona, 2022; Vijayanathan, Trejo-Arellano and Mozgová, 2022).

# V. Conclusion

Par une approche de biologie structurale et de biochimie, j'ai cherché à caractériser *in vitro* la protéine ULT1 et le complexe PRC2 d'Arabidopsis, et à mettre en évidence les détails moléculaires de leur interaction dans le but de mieux comprendre les mécanismes de leur fonction antagoniste.

Ce travail de thèse a permis de mettre en place des outils essentiels pour l'étude *in vitro* d'ULT1 et des complexes PRC2 chez Arabidopsis mais aussi d'obtenir les premières informations structurales sur ces protéines. En effet nous avons résolu la structure du domaine C-Terminal d'ULT1 par cristallographie aux rayons X et mis en évidence un domaine CW atypique qui pourrait reconnaître des modifications sur H3K36. Cependant, ce dernier résultat, obtenu à partir d'une puce portant des peptides histone modifiés n'a pas été confirmé par des approches *in vitro* alternatives de tests d'interaction. Par ailleurs, bien que le domaine SAND n'ait pas pu être cristallisé, nous avons obtenu des informations structurales importantes grâce aux modèles AlphaFold. Ces modèles ont permis d'identifier des résidus certainement essentiels pour l'interaction d'ULT1 avec les acides nucléiques et apporteront de nouvelles informations sur le rôle de cette interaction.

Cette thèse avait également pour but de caractériser *in vitro* l'interaction entre ULT1 et les complexes PRC2 d'Arabidopsis contenant les méthyltransférases CLF ou SWN. Mes travaux n'ont pas permis de confirmer l'interaction d'ULT1 avec PRC2-CLF *in vitro* mais en ont révélé une avec PRC2-SWN. J'ai également démontré que cette interaction était stable et que la liaison d'ULT1 ne dissociait pas le complexe PRC2-SWN *in vitro*. Nos expériences *in vitro* et *in vivo* semblent aussi indiquer que la région N-Terminale d'ULT1 est en interaction physique avec PRC2-SWN. Nous n'avons à ce jour pas réussi à caractériser la structure du complexe PRC2-SWN avec ULT1 par Cryo-EM et à identifier les résidus clés de cette interaction. Seule la structure du complexe PRC2-SWN a pu être révélée et indique qu'ULT1 pourrait se lier à une région flexible du complexe.

Alors que la structure des complexes PRC2 chez les plantes ne sont toujours pas connues et pourraient présenter des différences structurales, mes travaux sur la production et la purification des complexes PRC2-CLF et PRC2-SWN ternaires chez Arabidopsis ont permis d'initier leur étude structurale par Cryo-EM. Ces données préliminaires serviront par la suite à identifier les différences entre les complexes PRC2-CLF et PRC2-SWN afin de mieux comprendre leurs différences fonctionnelles.

En conclusion, mes données apportent de nouvelles pistes de réflexion concernant la fonction d'ULT1 et permettent d'orienter nos recherches vers une caractérisation plus précise de l'interaction de cette protéine avec le complexe PRC2-SWN. De plus, ce travail

de biochimie pourra également servir de base pour l'étude de protéines accessoires impliquées dans la régulation de la fonction des complexes PRC2 chez les plantes.

# **VI. Références**

- Alecki, C. *et al.* (2020) 'RNA-DNA strand exchange by the *Drosophila* Polycomb complex PRC2', *Nature Communications*, 11(1), p. 1781. Available at: <https://doi.org/10.1038/s41467-020-15609-x>.
- Allis, C.D. and Jenuwein, T. (2016) 'The molecular hallmarks of epigenetic control', *Nature Reviews. Genetics*, 17(8), pp. 487–500. Available at: <https://doi.org/10.1038/nrg.2016.59>.
- Alvarez-Venegas, R. *et al.* (2003) 'ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes', *Current biology: CB*, 13(8), pp. 627–637. Available at: [https://doi.org/10.1016/s0960-9822\(03\)00243-4](https://doi.org/10.1016/s0960-9822(03)00243-4).
- Amasino, R. (2010) 'Seasonal and developmental timing of flowering', *The Plant Journal*, 61(6), pp. 1001–1013. Available at: <https://doi.org/10.1111/j.1365-313X.2010.04148.x>.
- Andrés, F. and Coupland, G. (2012) 'The genetic basis of flowering responses to seasonal cues', *Nature Reviews Genetics*, 13(9), pp. 627–639. Available at: <https://doi.org/10.1038/nrg3291>.
- Antonin, W. and Neumann, H. (2016) 'Chromosome condensation and decondensation during mitosis', *Current Opinion in Cell Biology*, 40, pp. 15–22. Available at: <https://doi.org/10.1016/j.ceb.2016.01.013>.
- Baile, F. *et al.* (2021) 'EAR domain-containing transcription factors trigger PRC2-mediated chromatin marking in Arabidopsis', *The Plant Cell*, 33(8), pp. 2701–2715. Available at: <https://doi.org/10.1093/plcell/koab139>.
- Bannister, A.J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Research*, 21(3), pp. 381–395. Available at: <https://doi.org/10.1038/cr.2011.22>.
- Baulcombe, D.C. and Dean, C. (2014) 'Epigenetic regulation in plant responses to the environment', *Cold Spring Harbor Perspectives in Biology*, 6(9), p. a019471. Available at: <https://doi.org/10.1101/cshperspect.a019471>.
- Bednar, J. *et al.* (2017) 'Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1', *Molecular Cell*, 66(3), pp. 384–397.e8. Available at: <https://doi.org/10.1016/j.molcel.2017.04.012>.
- Beh, L.Y., Colwell, L.J. and Francis, N.J. (2012) 'A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence', *Proceedings of the National Academy of Sciences of the United States of America*, 109(18), pp. E1063–1071. Available at: <https://doi.org/10.1073/pnas.1118678109>.
- Beltran, M. *et al.* (2019) 'G-tract RNA removes Polycomb repressive complex 2 from genes', *Nature Structural & Molecular Biology*, 26(10), pp. 899–909. Available at: <https://doi.org/10.1038/s41594-019-0293-z>.
- Bergamin, E. *et al.* (2017) 'Molecular basis for the methylation specificity of ATXR5 for histone H3', *Nucleic Acids Research*, 45(11), pp. 6375–6387. Available at: <https://doi.org/10.1093/nar/gkx224>.
- Bieluszewski, T. *et al.* (2021) 'PRC2 activity, recruitment, and silencing: a comparative perspective', *Trends in Plant Science*, 26(11), pp. 1186–1198. Available at: <https://doi.org/10.1016/j.tplants.2021.06.006>.

- Blackledge, N.P. *et al.* (2014) 'Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation', *Cell*, 157(6), pp. 1445–1459. Available at: <https://doi.org/10.1016/j.cell.2014.05.004>.
- Blackledge, N.P. and Klose, R.J. (2021) 'The molecular principles of gene regulation by Polycomb repressive complexes', *Nature Reviews. Molecular Cell Biology*, 22(12), pp. 815–833. Available at: <https://doi.org/10.1038/s41580-021-00398-y>.
- Bloomer, R.H. *et al.* (2020) 'The Arabidopsis epigenetic regulator ICU11 as an accessory protein of Polycomb Repressive Complex 2', *Proceedings of the National Academy of Sciences of the United States of America*, 117(28), pp. 16660–16666. Available at: <https://doi.org/10.1073/pnas.1920621117>.
- Borden, K.L. (1998) 'RING fingers and B-boxes: zinc-binding protein-protein interaction domains', *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 76(2–3), pp. 351–358. Available at: <https://doi.org/10.1139/bcb-76-2-3-351>.
- Bottomley, M.J. *et al.* (2001) 'The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation', *Nature Structural Biology*, 8(7), pp. 626–633. Available at: <https://doi.org/10.1038/89675>.
- Boyer, L.A. *et al.* (2006) 'Polycomb complexes repress developmental regulators in murine embryonic stem cells', *Nature*, 441(7091), pp. 349–353. Available at: <https://doi.org/10.1038/nature04733>.
- Brand, U. *et al.* (2000) 'Dependence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity', *Science*, 289(5479), pp. 617–619. Available at: <https://doi.org/10.1126/science.289.5479.617>.
- Bratzel, F. *et al.* (2010) 'Keeping Cell Identity in Arabidopsis Requires PRC1 RING-Finger Homologs that Catalyze H2A Monoubiquitination', *Current Biology*, 20(20), pp. 1853–1859. Available at: <https://doi.org/10.1016/j.cub.2010.09.046>.
- Calonje, M. *et al.* (2008) 'EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis', *The Plant Cell*, 20(2), pp. 277–291. Available at: <https://doi.org/10.1105/tpc.106.049957>.
- Cao, R. *et al.* (2002) 'Role of histone H3 lysine 27 methylation in Polycomb-group silencing', *Science (New York, N.Y.)*, 298(5595), pp. 1039–1043. Available at: <https://doi.org/10.1126/science.1076997>.
- Cao, R. and Zhang, Y. (2004) 'SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex', *Molecular Cell*, 15(1), pp. 57–67. Available at: <https://doi.org/10.1016/j.molcel.2004.06.020>.
- Carles, C.C. *et al.* (2004) 'The ULTRAPETALA1 gene functions early in Arabidopsis development to restrict shoot apical meristem activity and acts through WUSCHEL to regulate floral meristem determinacy', *Genetics*, 167(4), pp. 1893–1903. Available at: <https://doi.org/10.1534/genetics.104.028787>.
- Carles, C.C. *et al.* (2005) 'ULTRAPETALA1 encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in Arabidopsis', *Development (Cambridge, England)*, 132(5), pp. 897–911. Available at: <https://doi.org/10.1242/dev.01642>.

- Carles, C.C. and Fletcher, J.C. (2009a) 'The SAND domain protein ULTRAPETALA1 acts as a trithorax group factor to regulate cell fate in plants', *Genes & Development*, 23(23), pp. 2723–2728. Available at: <https://doi.org/10.1101/gad.1812609>.
- Carles, C.C. and Fletcher, J.C. (2009b) 'The SAND domain protein ULTRAPETALA1 acts as a trithorax group factor to regulate cell fate in plants', *Genes & Development*, 23(23), pp. 2723–2728. Available at: <https://doi.org/10.1101/gad.1812609>.
- Carter, B. *et al.* (2018) 'The Chromatin Remodelers PKL and PIE1 Act in an Epigenetic Pathway That Determines H3K27me3 Homeostasis in Arabidopsis', *The Plant Cell*, 30(6), pp. 1337–1352. Available at: <https://doi.org/10.1105/tpc.17.00867>.
- Cavalli, G. and Heard, E. (2019) 'Advances in epigenetics link genetics to the environment and disease', *Nature*, 571(7766), pp. 489–499. Available at: <https://doi.org/10.1038/s41586-019-1411-0>.
- Chammas, P., Mocavini, I. and Di Croce, L. (2020) 'Engaging chromatin: PRC2 structure meets function', *British Journal of Cancer*, 122(3), pp. 315–328. Available at: <https://doi.org/10.1038/s41416-019-0615-2>.
- Chanvivattana, Y. *et al.* (2004) 'Interaction of Polycomb-group proteins controlling flowering in Arabidopsis', *Development*, 131(21), pp. 5263–5276. Available at: <https://doi.org/10.1242/dev.01400>.
- Chen, D. *et al.* (2010) 'The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth', *Cell Research*, 20(12), pp. 1332–1344. Available at: <https://doi.org/10.1038/cr.2010.151>.
- Chen, S. *et al.* (2018) 'Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding', *Molecular Cell*, 69(5), pp. 840–852.e5. Available at: <https://doi.org/10.1016/j.molcel.2018.01.039>.
- Chen, S. *et al.* (2020) 'A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin', *Molecular Cell*, 77(6), pp. 1265–1278.e7. Available at: <https://doi.org/10.1016/j.molcel.2019.12.019>.
- Chen, V.B. *et al.* (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D: Biological Crystallography*, 66(Pt 1), pp. 12–21. Available at: <https://doi.org/10.1107/S0907444909042073>.
- Cheng, J. *et al.* (2022) 'H3K9 demethylases IBM1 and JMJ27 are required for male meiosis in Arabidopsis thaliana', *New Phytologist*, 235(6), pp. 2252–2269. Available at: <https://doi.org/10.1111/nph.18286>.
- Cifuentes-Rojas, C. *et al.* (2014) 'Regulatory interactions between RNA and polycomb repressive complex 2', *Molecular Cell*, 55(2), pp. 171–185. Available at: <https://doi.org/10.1016/j.molcel.2014.05.009>.
- Clapier, C.R. *et al.* (2017) 'Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes', *Nature Reviews Molecular Cell Biology*, 18(7), pp. 407–422. Available at: <https://doi.org/10.1038/nrm.2017.26>.
- Clark, S.E. *et al.* (1996) 'The CLAVATA and SHOOT MERISTEMLESS loci competitively regulate meristem activity in Arabidopsis', *Development (Cambridge, England)*, 122(5), pp. 1567–1575. Available at: <https://doi.org/10.1242/dev.122.5.1567>.

- Cokus, S.J. *et al.* (2008) 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature*, 452(7184), pp. 215–219. Available at: <https://doi.org/10.1038/nature06745>.
- Coleman-Derr, D. and Zilberman, D. (2012) 'Deposition of Histone Variant H2A.Z within Gene Bodies Regulates Responsive Genes', *PLoS Genetics*, 8(10), p. e1002988. Available at: <https://doi.org/10.1371/journal.pgen.1002988>.
- Cooper, S. *et al.* (2014) 'Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment', *Cell Reports*, 7(5), pp. 1456–1470. Available at: <https://doi.org/10.1016/j.celrep.2014.04.012>.
- Costa, S. and Dean, C. (2019) 'Storing memories: the distinct phases of Polycomb-mediated silencing of Arabidopsis FLC', *Biochemical Society Transactions*, 47(4), pp. 1187–1196. Available at: <https://doi.org/10.1042/BST20190255>.
- Crevillén, P. *et al.* (2014) 'Epigenetic reprogramming that prevents transgenerational inheritance of the vernalized state', *Nature*, 515(7528), pp. 587–590. Available at: <https://doi.org/10.1038/nature13722>.
- Crevillén, P. (2020) 'Histone Demethylases as Counterbalance to H3K27me3 Silencing in Plants', *iScience*, 23(11), p. 101715. Available at: <https://doi.org/10.1016/j.isci.2020.101715>.
- Cui, Xia *et al.* (2016) 'REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in Arabidopsis', *Nature Genetics*, 48(6), pp. 694–699. Available at: <https://doi.org/10.1038/ng.3556>.
- Czermin, B. *et al.* (2002) 'Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites', *Cell*, 111(2), pp. 185–196. Available at: [https://doi.org/10.1016/s0092-8674\(02\)00975-3](https://doi.org/10.1016/s0092-8674(02)00975-3).
- Davidovich, C. *et al.* (2014) 'A dimeric state for PRC2', *Nucleic Acids Research*, 42(14), pp. 9236–9248. Available at: <https://doi.org/10.1093/nar/gku540>.
- Derkacheva, M. *et al.* (2013) 'Arabidopsis MSI1 connects LHP1 to PRC2 complexes', *The EMBO Journal*, 32(14), pp. 2073–2085. Available at: <https://doi.org/10.1038/emboj.2013.145>.
- Dhalluin, C. *et al.* (1999) 'Structure and ligand of a histone acetyltransferase bromodomain', *Nature*, 399(6735), pp. 491–496. Available at: <https://doi.org/10.1038/20974>.
- Ding, Y. *et al.* (2012) 'ATX1-Generated H3K4me3 Is Required for Efficient Elongation of Transcription, Not Initiation, at ATX1-Regulated Genes', *PLoS Genetics*. Edited by G.P. Copenhagen, 8(12), p. e1003111. Available at: <https://doi.org/10.1371/journal.pgen.1003111>.
- Ding, Y., Avramova, Z. and Fromm, M. (2011) 'Two Distinct Roles of ARABIDOPSIS HOMOLOG OF TRITHORAX1 (ATX1) at Promoters and within Transcribed Regions of ATX1-Regulated Genes', *The Plant Cell*, 23(1), pp. 350–363. Available at: <https://doi.org/10.1105/tpc.110.080150>.
- Dumesic, P.A. *et al.* (2015) 'Product binding enforces the genomic specificity of a yeast polycomb repressive complex', *Cell*, 160(1–2), pp. 204–218. Available at: <https://doi.org/10.1016/j.cell.2014.11.039>.

- Dutta, A. *et al.* (2017) 'JMJ27, an Arabidopsis H3K9 histone demethylase, modulates defense against *Pseudomonas syringae* and flowering time', *The Plant Journal*, 91(6), pp. 1015–1028. Available at: <https://doi.org/10.1111/tpj.13623>.
- Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', *Acta Crystallographica. Section D, Biological Crystallography*, 60(Pt 12 Pt 1), pp. 2126–2132. Available at: <https://doi.org/10.1107/S0907444904019158>.
- Farrona, S. *et al.* (2004) 'The Arabidopsis thaliana SNF2 homolog AtBRM controls shoot development and flowering', *Development (Cambridge, England)*, 131(20), pp. 4965–4975. Available at: <https://doi.org/10.1242/dev.01363>.
- Faust, C. *et al.* (1995) 'The eed mutation disrupts anterior mesoderm production in mice', *Development (Cambridge, England)*, 121(2), pp. 273–285. Available at: <https://doi.org/10.1242/dev.121.2.273>.
- Finogenova, K. *et al.* (2020) 'Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3', *eLife*, 9, p. e61964. Available at: <https://doi.org/10.7554/eLife.61964>.
- Fitzgerald, D.J. *et al.* (2006) 'Protein complex expression by using multigene baculoviral vectors', *Nature Methods*, 3(12), pp. 1021–1032. Available at: <https://doi.org/10.1038/nmeth983>.
- Fletcher, J.C. *et al.* (1999) 'Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot Meristems', *Science*, 283(5409), pp. 1911–1914. Available at: <https://doi.org/10.1126/science.283.5409.1911>.
- Fletcher, J.C. (2001) 'The ULTRAPETALA gene controls shoot and floral meristem size in Arabidopsis', *Development (Cambridge, England)*, 128(8), pp. 1323–1333. Available at: <https://doi.org/10.1242/dev.128.8.1323>.
- Foroozani, M., Holder, D.H. and Deal, R.B. (2022) 'Histone variants in the specialization of plant chromatin', *Annual review of plant biology*, 73, pp. 149–172. Available at: <https://doi.org/10.1146/annurev-arplant-070221-050044>.
- Foroozani, M., Vandal, M.P. and Smith, A.P. (2021) 'H3K4 trimethylation dynamics impact diverse developmental and environmental responses in plants', *Planta*, 253(1), p. 4. Available at: <https://doi.org/10.1007/s00425-020-03520-0>.
- Francis, N.J., Kingston, R.E. and Woodcock, C.L. (2004) 'Chromatin Compaction by a Polycomb Group Protein Complex', *Science*, 306(5701), pp. 1574–1577. Available at: <https://doi.org/10.1126/science.1100576>.
- Franco-Echevarria, E. *et al.* (2023) *Distinct accessory roles of Arabidopsis VEL proteins in Polycomb silencing.* preprint. Plant Biology. Available at: <https://doi.org/10.1101/2023.05.22.541744>.
- Frapporti, A. *et al.* (2019) 'The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium', *Nature Communications*, 10(1), p. 2710. Available at: <https://doi.org/10.1038/s41467-019-10648-5>.
- Fyodorov, D.V. *et al.* (2018) 'Emerging roles of linker histones in regulating chromatin structure and function', *Nature Reviews. Molecular Cell Biology*, 19(3), pp. 192–206. Available at: <https://doi.org/10.1038/nrm.2017.94>.

Gangappa, S.N. and Botto, J.F. (2014) 'The BBX family of plant transcription factors', *Trends in Plant Science*, 19(7), pp. 460–470. Available at: <https://doi.org/10.1016/j.tplants.2014.01.010>.

Gary, J.D. and Clarke, S. (1998) 'RNA and Protein Interactions Modulated by Protein Arginine Methylation', in K. Moldave (ed.) *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, pp. 65–131. Available at: [https://doi.org/10.1016/S0079-6603\(08\)60825-9](https://doi.org/10.1016/S0079-6603(08)60825-9).

Gaudin, V. *et al.* (2001) 'Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in Arabidopsis', *Development (Cambridge, England)*, 128(23), pp. 4847–4858. Available at: <https://doi.org/10.1242/dev.128.23.4847>.

Gendall, A.R. *et al.* (2001) 'The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis', *Cell*, 107(4), pp. 525–535. Available at: [https://doi.org/10.1016/S0092-8674\(01\)00573-6](https://doi.org/10.1016/S0092-8674(01)00573-6).

Godwin, J. and Farrona, S. (2022) 'The Importance of Networking: Plant Polycomb Repressive Complex 2 and Its Interactors', *Epigenomes*, 6(1), p. 8. Available at: <https://doi.org/10.3390/epigenomes6010008>.

Gómez-Zambrano, Á., Merini, W. and Calonje, M. (2019) 'The repressive role of Arabidopsis H2A.Z in transcriptional regulation depends on AtBMI1 activity', *Nature Communications*, 10(1), p. 2828. Available at: <https://doi.org/10.1038/s41467-019-10773-1>.

Gong, L. *et al.* (2022) 'CK2-mediated phosphorylation of SUZ12 promotes PRC2 function by stabilizing enzyme active site', *Nature Communications*, 13(1), p. 6781. Available at: <https://doi.org/10.1038/s41467-022-34431-1>.

Goodrich, J. *et al.* (1997) 'A Polycomb-group gene regulates homeotic gene expression in Arabidopsis', *Nature*, 386(6620), pp. 44–51. Available at: <https://doi.org/10.1038/386044a0>.

Grau, D. *et al.* (2021a) 'Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction', *Nature Communications*, 12, p. 714. Available at: <https://doi.org/10.1038/s41467-020-20775-z>.

Grau, D. *et al.* (2021b) 'Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction', *Nature Communications*, 12(1), p. 714. Available at: <https://doi.org/10.1038/s41467-020-20775-z>.

Grimanelli, D. and Ingouff, M. (2020) 'DNA Methylation Readers in Plants', *Journal of Molecular Biology*, 432(6), pp. 1706–1717. Available at: <https://doi.org/10.1016/j.jmb.2019.12.043>.

Grossniklaus, U. *et al.* (1998) 'Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis', *Science (New York, N.Y.)*, 280(5362), pp. 446–450. Available at: <https://doi.org/10.1126/science.280.5362.446>.

Guo, Y., Zhao, S. and Wang, G.G. (2021) 'Polycomb Gene Silencing Mechanisms: PRC2 Chromatin Targeting, H3K27me3 "Readout", and Phase Separation-Based Compaction', *Trends in Genetics*, 37(6), pp. 547–565. Available at: <https://doi.org/10.1016/j.tig.2020.12.006>.

Han, S.-K. *et al.* (2015) 'Roles and activities of chromatin remodeling ATPases in plants', *The Plant Journal*, 83(1), pp. 62–77. Available at: <https://doi.org/10.1111/tpj.12877>.

- Heitz, E. (1928) 'Das heterochromatin der Moose', (69), pp. 762–818.
- Hennig, L. *et al.* (2003) 'Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development', *Development*, 130(12), pp. 2555–2565. Available at: <https://doi.org/10.1242/dev.00470>.
- Heo, J.B. and Sung, S. (2011) 'Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA', *Science*, 331(6013), pp. 76–79. Available at: <https://doi.org/10.1126/science.1197349>.
- Hollender, C. and Liu, Z. (2008) 'Histone deacetylase genes in Arabidopsis development', *Journal of Integrative Plant Biology*, 50(7), pp. 875–885. Available at: <https://doi.org/10.1111/j.1744-7909.2008.00704.x>.
- Hoppmann, V. *et al.* (2011) 'The CW domain, a new histone recognition module in chromatin proteins', *The EMBO journal*, 30(10), pp. 1939–1952. Available at: <https://doi.org/10.1038/emboj.2011.108>.
- Hu, H. and Du, J. (2022) 'Structure and mechanism of histone methylation dynamics in Arabidopsis', *Current Opinion in Plant Biology*, 67, p. 102211. Available at: <https://doi.org/10.1016/j.pbi.2022.102211>.
- Huang, S. *et al.* (2019) 'Arabidopsis histone H3K4 demethylase JM17 functions in dehydration stress response', *The New Phytologist*, 223(3), pp. 1372–1387. Available at: <https://doi.org/10.1111/nph.15874>.
- Ingham, P.W. (1983) 'Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes', *Nature*, 306(5943), pp. 591–593. Available at: <https://doi.org/10.1038/306591a0>.
- Ingham, P.W. (1985) 'Genetic Control of the Spatial Pattern of Selector Gene Expression in Drosophila', *Cold Spring Harbor Symposia on Quantitative Biology*, 50, pp. 201–208. Available at: <https://doi.org/10.1101/SQB.1985.050.01.026>.
- Ingouff, M. and Berger, F. (2010) 'Histone3 variants in plants', *Chromosoma*, 119(1), pp. 27–33. Available at: <https://doi.org/10.1007/s00412-009-0237-1>.
- Jacob, Y. *et al.* (2009) 'ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing', *Nature Structural & Molecular Biology*, 16(7), pp. 763–768. Available at: <https://doi.org/10.1038/nsmb.1611>.
- Jacob, Y. *et al.* (2014) 'Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication', *Science (New York, N.Y.)*, 343(6176), pp. 1249–1253. Available at: <https://doi.org/10.1126/science.1248357>.
- Jiang, D. *et al.* (2008) 'Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components', *PLoS One*, 3(10), p. e3404. Available at: <https://doi.org/10.1371/journal.pone.0003404>.
- Jiang, D. *et al.* (2011) 'Arabidopsis COMPASS-Like Complexes Mediate Histone H3 Lysine-4 Trimethylation to Control Floral Transition and Plant Development', *PLoS Genetics*. Edited by G.P. Copenhaver, 7(3), p. e1001330. Available at: <https://doi.org/10.1371/journal.pgen.1001330>.

- Jiang, D. and Berger, F. (2017) 'Histone variants in plant transcriptional regulation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1860(1), pp. 123–130. Available at: <https://doi.org/10.1016/j.bbagr.2016.07.002>.
- Jiao, L. *et al.* (2020) 'A partially disordered region connects gene repression and activation functions of EZH2', *Proceedings of the National Academy of Sciences of the United States of America*, 117(29), pp. 16992–17002. Available at: <https://doi.org/10.1073/pnas.1914866117>.
- Jiao, L. and Liu, X. (2015) 'Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2', *Science (New York, N.Y.)*, 350(6258), p. aac4383. Available at: <https://doi.org/10.1126/science.aac4383>.
- Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: <https://doi.org/10.1038/s41586-021-03819-2>.
- Justin, N. *et al.* (2016) 'Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2', *Nature Communications*, 7, p. 11316. Available at: <https://doi.org/10.1038/ncomms11316>.
- Kabsch, W. (2010) 'Integration, scaling, space-group assignment and post-refinement', *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2), pp. 133–144. Available at: <https://doi.org/10.1107/S0907444909047374>.
- Kadoch, C. *et al.* (2017) 'Dynamics of BAF–Polycomb complex opposition on heterochromatin in normal and oncogenic states', *Nature Genetics*, 49(2), pp. 213–222. Available at: <https://doi.org/10.1038/ng.3734>.
- Kagale, S. and Rozwadowski, K. (2011) 'EAR motif-mediated transcriptional repression in plants', *Epigenetics*, 6(2), pp. 141–146. Available at: <https://doi.org/10.4161/epi.6.2.13627>.
- Kalb, R. *et al.* (2014) 'Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression', *Nature Structural & Molecular Biology*, 21(6), pp. 569–571. Available at: <https://doi.org/10.1038/nsmb.2833>.
- Kaneko, S. *et al.* (2013) 'PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells', *Nature Structural & Molecular Biology*, 20(11), pp. 1258–1264. Available at: <https://doi.org/10.1038/nsmb.2700>.
- Kankel, M.W. *et al.* (2003) 'Arabidopsis MET1 Cytosine Methyltransferase Mutants', *Genetics*, 163(3), pp. 1109–1122. Available at: <https://doi.org/10.1093/genetics/163.3.1109>.
- Kasinath, V. *et al.* (2018) 'Structures of human PRC2 with its cofactors AEBP2 and JARID2', *Science (New York, N.Y.)*, 359(6378), pp. 940–944. Available at: <https://doi.org/10.1126/science.aar5700>.
- Kasinath, V. *et al.* (2021) 'JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications', *Science (New York, N.Y.)*, 371(6527), p. eabc3393. Available at: <https://doi.org/10.1126/science.abc3393>.
- Kassis, J.A. and Brown, J.L. (2013) 'Polycomb group response elements in Drosophila and vertebrates', *Advances in Genetics*, 81, pp. 83–118. Available at: <https://doi.org/10.1016/B978-0-12-407677-8.00003-8>.

- Kassis, J.A., Kennison, J.A. and Tamkun, J.W. (2017) 'Polycomb and Trithorax Group Genes in *Drosophila*', *Genetics*, 206(4), pp. 1699–1725. Available at: <https://doi.org/10.1534/genetics.115.185116>.
- Kaufmann, K., Pajoro, A. and Angenent, G.C. (2010) 'Regulation of transcription in plants: mechanisms controlling developmental switches', *Nature Reviews Genetics*, 11(12), pp. 830–842. Available at: <https://doi.org/10.1038/nrg2885>.
- Ketel, C.S. *et al.* (2005) 'Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes', *Molecular and Cellular Biology*, 25(16), pp. 6857–6868. Available at: <https://doi.org/10.1128/MCB.25.16.6857-6868.2005>.
- Khanna, R. *et al.* (2009) 'The Arabidopsis B-Box Zinc Finger Family', *The Plant Cell*, 21(11), pp. 3416–3420. Available at: <https://doi.org/10.1105/tpc.109.069088>.
- Kim, D.-H., Xi, Y. and Sung, S. (2017) 'Modular function of long noncoding RNA, COLDAIR, in the vernalization response', *PLoS genetics*, 13(7), p. e1006939. Available at: <https://doi.org/10.1371/journal.pgen.1006939>.
- Kim, S.Y., Zhu, T. and Sung, Z.R. (2010) 'Epigenetic Regulation of Gene Programs by EMF1 and EMF2 in Arabidopsis', *Plant Physiology*, 152(2), pp. 516–528. Available at: <https://doi.org/10.1104/pp.109.143495>.
- Kingston, R.E. and Tamkun, J.W. (2014) 'Transcriptional Regulation by Trithorax-Group Proteins', *Cold Spring Harbor Perspectives in Biology*, 6(10), p. a019349. Available at: <https://doi.org/10.1101/cshperspect.a019349>.
- Kiyosue, T. *et al.* (1999) 'Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp. 4186–4191. Available at: <https://doi.org/10.1073/pnas.96.7.4186>.
- Klemm, S.L., Shipony, Z. and Greenleaf, W.J. (2019) 'Chromatin accessibility and the regulatory epigenome', *Nature Reviews. Genetics*, 20(4), pp. 207–220. Available at: <https://doi.org/10.1038/s41576-018-0089-8>.
- Koiwa, H. *et al.* (2004) 'Arabidopsis C-terminal domain phosphatase-like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases', *Proceedings of the National Academy of Sciences of the United States of America*, 101(40), pp. 14539–14544. Available at: <https://doi.org/10.1073/pnas.0403174101>.
- Konev, A.Y. *et al.* (2007) 'CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo', *Science (New York, N.Y.)*, 317(5841), pp. 1087–1090. Available at: <https://doi.org/10.1126/science.1145339>.
- Kotake, T. *et al.* (2003) 'Arabidopsis TERMINAL FLOWER 2 gene encodes a heterochromatin protein 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes', *Plant & Cell Physiology*, 44(6), pp. 555–564. Available at: <https://doi.org/10.1093/pcp/pcg091>.
- Kouzarides, T. (2007) 'Chromatin Modifications and Their Function', *Cell*, 128(4), pp. 693–705. Available at: <https://doi.org/10.1016/j.cell.2007.02.005>.
- Kralemann, L.E.M. *et al.* (2020) 'Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in Arabidopsis', *Genome Biology*, 21(1), p. 144. Available at: <https://doi.org/10.1186/s13059-020-02062-8>.

- Krogan, N.T., Hogan, K. and Long, J.A. (2012) 'APETALA2 negatively regulates multiple floral organ identity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19', *Development*, 139(22), pp. 4180–4190. Available at: <https://doi.org/10.1242/dev.085407>.
- Kuzmichev, A. *et al.* (2002) 'Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein', *Genes & Development*, 16(22), pp. 2893–2905. Available at: <https://doi.org/10.1101/gad.1035902>.
- Kwon, C.S., Chen, C. and Wagner, D. (2005) 'WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in Arabidopsis', *Genes & Development*, 19(8), pp. 992–1003. Available at: <https://doi.org/10.1101/gad.1276305>.
- Lafos, M. *et al.* (2011) 'Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation', *PLoS genetics*, 7(4), p. e1002040. Available at: <https://doi.org/10.1371/journal.pgen.1002040>.
- Laux, T. *et al.* (1996) 'The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis', *Development (Cambridge, England)*, 122(1), pp. 87–96. Available at: <https://doi.org/10.1242/dev.122.1.87>.
- Law, J.A. and Jacobsen, S.E. (2010) 'Establishing, maintaining and modifying DNA methylation patterns in plants and animals', *Nature Reviews Genetics*, 11(3), pp. 204–220. Available at: <https://doi.org/10.1038/nrg2719>.
- Lee, C.-H. *et al.* (2018) 'Distinct stimulatory mechanisms regulate the catalytic activity of Polycomb Repressive Complex 2 (PRC2)', *Molecular cell*, 70(3), pp. 435–448.e5. Available at: <https://doi.org/10.1016/j.molcel.2018.03.019>.
- Lee, C.-H. *et al.* (2019) 'Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma', *Genes & Development*, 33(19–20), pp. 1428–1440. Available at: <https://doi.org/10.1101/gad.328773.119>.
- Lee, H. *et al.* (2000) 'The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis', *Genes & Development*, 14(18), pp. 2366–2376. Available at: <https://doi.org/10.1101/gad.813600>.
- Lee, J.-S., Smith, E. and Shilatifard, A. (2010) 'The Language of Histone Crosstalk', *Cell*, 142(5), pp. 682–685. Available at: <https://doi.org/10.1016/j.cell.2010.08.011>.
- Leng, X. *et al.* (2020) 'A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription', *Trends in Plant Science*, 25(8), pp. 744–764. Available at: <https://doi.org/10.1016/j.tplants.2020.03.005>.
- Lenhard, M. *et al.* (2001) 'Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between WUSCHEL and AGAMOUS', *Cell*, 105(6), pp. 805–814. Available at: [https://doi.org/10.1016/s0092-8674\(01\)00390-7](https://doi.org/10.1016/s0092-8674(01)00390-7).
- Lermontova, I. *et al.* (2011) 'Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation', *The Plant Journal*, 68(1), pp. 40–50. Available at: <https://doi.org/10.1111/j.1365-3113.2011.04664.x>.
- Lermontova, I. *et al.* (2013) 'Arabidopsis KINETOCHORE NULL2 Is an Upstream Component for Centromeric Histone H3 Variant cenH3 Deposition at Centromeres', *The Plant Cell*, 25(9), pp. 3389–3404. Available at: <https://doi.org/10.1105/tpc.113.114736>.

- Lewis, E.B. (1978) 'A gene complex controlling segmentation in *Drosophila*', *Nature*, 276(5688), pp. 565–570. Available at: <https://doi.org/10.1038/276565a0>.
- Lewis, P. (1947) 'New mutant report', *Drosoph. Inf. Serv.*, 21, p. 69.
- Li, C. *et al.* (2015) 'The Arabidopsis SWI2/SNF2 Chromatin Remodeler BRAHMA Regulates Polycomb Function during Vegetative Development and Directly Activates the Flowering Repressor Gene SVP', *PLOS Genetics*, 11(1), p. e1004944. Available at: <https://doi.org/10.1371/journal.pgen.1004944>.
- Li, C. *et al.* (2016) 'Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in Arabidopsis', *Nature Genetics*, 48(6), pp. 687–693. Available at: <https://doi.org/10.1038/ng.3555>.
- Li, D. *et al.* (2020) 'JMJ14 encoded H3K4 demethylase modulates immune responses by regulating defence gene expression and pipecolic acid levels', *The New Phytologist*, 225(5), pp. 2108–2121. Available at: <https://doi.org/10.1111/nph.16270>.
- Li, Z. *et al.* (2018) 'Polycomb-mediated gene silencing by the BAH–EMF1 complex in plants', *Nature Genetics*, 50(9), pp. 1254–1261. Available at: <https://doi.org/10.1038/s41588-018-0190-0>.
- Liang, S.C. *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 Genetics*, 11(12), p. e1005660. Available at: <https://doi.org/10.1371/journal.pgen.1005660>.
- Lindroth, A.M. *et al.* (2001) 'Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation', *Science (New York, N.Y.)*, 292(5524), pp. 2077–2080. Available at: <https://doi.org/10.1126/science.1059745>.
- Liu, B. *et al.* (2019) 'The transcription factor OsSUF4 interacts with SDG725 in promoting H3K36me3 establishment', *Nature Communications*, 10(1), p. 2999. Available at: <https://doi.org/10.1038/s41467-019-10850-5>.
- Liu, C. *et al.* (2010) 'Histone Methylation in Higher Plants', *Annual Review of Plant Biology*, 61(1), pp. 395–420. Available at: <https://doi.org/10.1146/annurev.arplant.043008.091939>.
- Liu, F. *et al.* (2007) 'The Arabidopsis RNA-Binding Protein FCA Requires a Lysine-Specific Demethylase 1 Homolog to Downregulate FLC', *Molecular Cell*, 28(3), pp. 398–407. Available at: <https://doi.org/10.1016/j.molcel.2007.10.018>.
- Liu, P. *et al.* (2019) 'The Histone H3K4 Demethylase JMJ16 Represses Leaf Senescence in Arabidopsis', *The Plant Cell*, 31(2), pp. 430–443. Available at: <https://doi.org/10.1105/tpc.18.00693>.
- Liu, Y. *et al.* (2018) 'PCSD: a plant chromatin state database', *Nucleic Acids Research*, 46(D1), pp. D1157–D1167. Available at: <https://doi.org/10.1093/nar/gkx919>.
- Liu, Y. *et al.* (2022) 'Structural basis for the recognition of methylated histone H3 by the Arabidopsis LHP1 chromodomain', *The Journal of Biological Chemistry*, 298(3), p. 101623. Available at: <https://doi.org/10.1016/j.jbc.2022.101623>.
- Liu, Z.-W. *et al.* (2014) 'The SET Domain Proteins SUVH2 and SUVH9 Are Required for Pol V Occupancy at RNA-Directed DNA Methylation Loci', *PLOS Genetics*, 10(1), p. e1003948. Available at: <https://doi.org/10.1371/journal.pgen.1003948>.

- Long, Y. *et al.* (2020) 'RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells', *Nature Genetics*, 52(9), pp. 931–938. Available at: <https://doi.org/10.1038/s41588-020-0662-x>.
- Lorković, Z.J. *et al.* (2017) 'Compartmentalization of DNA Damage Response between Heterochromatin and Euchromatin Is Mediated by Distinct H2A Histone Variants', *Current Biology*, 27(8), pp. 1192–1199. Available at: <https://doi.org/10.1016/j.cub.2017.03.002>.
- Lu, F. *et al.* (2008) 'Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice', *Journal of Integrative Plant Biology*, 50(7), pp. 886–896. Available at: <https://doi.org/10.1111/j.1744-7909.2008.00692.x>.
- Lu, F. *et al.* (2010) 'JM14 is an H3K4 demethylase regulating flowering time in Arabidopsis', *Cell Research*, 20(3), pp. 387–390. Available at: <https://doi.org/10.1038/cr.2010.27>.
- Lu, L. *et al.* (2018) 'The plant-specific histone residue Phe41 is important for genome-wide H3.1 distribution', *Nature Communications*, 9, p. 630. Available at: <https://doi.org/10.1038/s41467-018-02976-9>.
- Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 Å resolution', *Nature*, 389(6648), pp. 251–260. Available at: <https://doi.org/10.1038/38444>.
- Luger, K., Dechassa, M.L. and Tremethick, D.J. (2012) 'New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?', *Nature Reviews Molecular Cell Biology*, 13(7), pp. 436–447. Available at: <https://doi.org/10.1038/nrm3382>.
- Luo, M. *et al.* (1999) 'Genes controlling fertilization-independent seed development in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences*, 96(1), pp. 296–301. Available at: <https://doi.org/10.1073/pnas.96.1.296>.
- Luo, M. (2018) 'Chemical and Biochemical Perspectives of Protein Lysine Methylation', *Chemical reviews*, 118(14), pp. 6656–6705. Available at: <https://doi.org/10.1021/acs.chemrev.8b00008>.
- Luo, Y.-X. *et al.* (2020) 'A plant-specific SWR1 chromatin-remodeling complex couples histone H2A.Z deposition with nucleosome sliding', *The EMBO journal*, 39(7), p. e102008. Available at: <https://doi.org/10.15252/embj.2019102008>.
- Margueron, R. *et al.* (2008) 'Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms', *Molecular cell*, 32(4), pp. 503–518. Available at: <https://doi.org/10.1016/j.molcel.2008.11.004>.
- Margueron, R. *et al.* (2009) 'Role of the polycomb protein EED in the propagation of repressive histone marks', *Nature*, 461(7265), pp. 762–767. Available at: <https://doi.org/10.1038/nature08398>.
- Margueron, R. and Reinberg, D. (2011) 'The Polycomb complex PRC2 and its mark in life', *Nature*, 469(7330), pp. 343–349. Available at: <https://doi.org/10.1038/nature09784>.
- Margueron, R., Trojer, P. and Reinberg, D. (2005) 'The key to development: interpreting the histone code?', *Current Opinion in Genetics & Development*, 15(2), pp. 163–176. Available at: <https://doi.org/10.1016/j.gde.2005.01.005>.
- Mariani, V. *et al.* (2013) 'IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests', *Bioinformatics*, 29(21), pp. 2722–2728. Available at: <https://doi.org/10.1093/bioinformatics/btt473>.

Martin, B.J.E. *et al.* (2021) 'Transcription shapes genome-wide histone acetylation patterns', *Nature Communications*, 12(1), p. 210. Available at: <https://doi.org/10.1038/s41467-020-20543-z>.

Mayer, K.F. *et al.* (1998) 'Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem', *Cell*, 95(6), pp. 805–815. Available at: [https://doi.org/10.1016/s0092-8674\(00\)81703-1](https://doi.org/10.1016/s0092-8674(00)81703-1).

McKinley, K.L. and Cheeseman, I.M. (2016) 'The molecular basis for centromere identity and function', *Nature reviews. Molecular cell biology*, 17(1), pp. 16–29. Available at: <https://doi.org/10.1038/nrm.2015.5>.

Mehdi, S. *et al.* (2016) 'The WD40 Domain Protein MSI1 Functions in a Histone Deacetylase Complex to Fine-Tune Abscisic Acid Signaling', *The Plant Cell*, 28(1), pp. 42–54. Available at: <https://doi.org/10.1105/tpc.15.00763>.

Michaels, S.D. *et al.* (2005) 'Integration of flowering signals in winter-annual Arabidopsis', *Plant Physiology*, 137(1), pp. 149–156. Available at: <https://doi.org/10.1104/pp.104.052811>.

Michaels, S.D. and Amasino, R.M. (1999) 'FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering', *The Plant Cell*, 11(5), pp. 949–956. Available at: <https://doi.org/10.1105/tpc.11.5.949>.

Mierlo, G. van *et al.* (2019) 'The Complexity of PRC2 Subcomplexes', *Trends in Cell Biology*, 29(8), pp. 660–671. Available at: <https://doi.org/10.1016/j.tcb.2019.05.004>.

Mikulski, P. *et al.* (2017) 'Characterization of the Polycomb-Group Mark H3K27me3 in Unicellular Algae', *Frontiers in Plant Science*, 8. Available at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.00607> (Accessed: 6 January 2024).

Min, J., Zhang, Y. and Xu, R.-M. (2003) 'Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27', *Genes & Development*, 17(15), pp. 1823–1828. Available at: <https://doi.org/10.1101/gad.269603>.

Mizuguchi, G. *et al.* (2004) 'ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex', *Science (New York, N.Y.)*, 303(5656), pp. 343–348. Available at: <https://doi.org/10.1126/science.1090701>.

Moreau, F. *et al.* (2016) 'The Myb-domain protein ULTRAPETALA1 INTERACTING FACTOR 1 controls floral meristem activities in Arabidopsis', *Development (Cambridge, England)*, 143(7), pp. 1108–1119. Available at: <https://doi.org/10.1242/dev.127365>.

Morrison, O. and Thakur, J. (2021) 'Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin', *International Journal of Molecular Sciences*, 22(13), p. 6922. Available at: <https://doi.org/10.3390/ijms22136922>.

Mozgova, I. and Hennig, L. (2015) 'The polycomb group protein regulatory network', *Annual Review of Plant Biology*, 66, pp. 269–296. Available at: <https://doi.org/10.1146/annurev-arplant-043014-115627>.

Müller, J. *et al.* (2002) 'Histone methyltransferase activity of a Drosophila Polycomb group repressor complex', *Cell*, 111(2), pp. 197–208. Available at: [https://doi.org/10.1016/s0092-8674\(02\)00976-5](https://doi.org/10.1016/s0092-8674(02)00976-5).

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) 'Refinement of macromolecular structures by the maximum-likelihood method', *Acta Crystallographica. Section D*,

*Biological Crystallography*, 53(Pt 3), pp. 240–255. Available at: <https://doi.org/10.1107/S0907444996012255>.

Ng, D.W.-K. *et al.* (2007) 'Plant SET domain-containing proteins: structure, function and regulation', *Biochimica Et Biophysica Acta*, 1769(5–6), pp. 316–329. Available at: <https://doi.org/10.1016/j.bbaexp.2007.04.003>.

O'Carroll, D. *et al.* (2001) 'The polycomb-group gene Ezh2 is required for early mouse development', *Molecular and Cellular Biology*, 21(13), pp. 4330–4336. Available at: <https://doi.org/10.1128/MCB.21.13.4330-4336.2001>.

Ohad, N. *et al.* (1999) 'Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization.', *The Plant Cell*, 11(3), p. 407. Available at: <https://doi.org/10.1105/tpc.11.3.407>.

Org, T. *et al.* (2008) 'The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression', *EMBO reports*, 9(4), pp. 370–376. Available at: <https://doi.org/10.1038/sj.embor.2008.11>.

Ornelas-Ayala, D. *et al.* (2020) 'ULTRAPETALA1 maintains Arabidopsis root stem cell niche independently of ARABIDOPSIS TRITHORAX1', *The New Phytologist*, 225(3), pp. 1261–1272. Available at: <https://doi.org/10.1111/nph.16213>.

Ornelas-Ayala, D. *et al.* (2023) 'A Green Light to Switch on Genes: Revisiting Trithorax on Plants', *Plants*, 12(1), p. 75. Available at: <https://doi.org/10.3390/plants12010075>.

Pandey, R. *et al.* (2002) 'Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes', *Nucleic Acids Research*, 30(23), pp. 5036–5055. Available at: <https://doi.org/10.1093/nar/gkf660>.

Papamichos-Chronakis, M. *et al.* (2011) 'Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity', *Cell*, 144(2), pp. 200–213. Available at: <https://doi.org/10.1016/j.cell.2010.12.021>.

Parcy, F. (2005) 'Flowering: a time for integration', *The International Journal of Developmental Biology*, 49(5–6), pp. 585–593. Available at: <https://doi.org/10.1387/ijdb.041930fp>.

Pasini, D. *et al.* (2004) 'Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity', *The EMBO Journal*, 23(20), pp. 4061–4071. Available at: <https://doi.org/10.1038/sj.emboj.7600402>.

Patching, S.G. (2014) 'Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(1, Part A), pp. 43–55. Available at: <https://doi.org/10.1016/j.bbamem.2013.04.028>.

Pelaz, S. *et al.* (2000) 'B and C floral organ identity functions require SEPALLATA MADS-box genes', *Nature*, 405(6783), pp. 200–203. Available at: <https://doi.org/10.1038/35012103>.

Pettersen, E.F. *et al.* (2021) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science: A Publication of the Protein Society*, 30(1), pp. 70–82. Available at: <https://doi.org/10.1002/pro.3943>.

- Pires, H.R. *et al.* (2014) 'ULTRAPETALA *trxG* genes interact with KANADI transcription factor genes to regulate *Arabidopsis* gynoecium patterning', *The Plant Cell*, 26(11), pp. 4345–4361. Available at: <https://doi.org/10.1105/tpc.114.131250>.
- Poepsel, S., Kasinath, V. and Nogales, E. (2018) 'Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes', *Nature Structural & Molecular Biology*, 25(2), pp. 154–162. Available at: <https://doi.org/10.1038/s41594-018-0023-y>.
- Portoso, M. *et al.* (2017) 'PRC2 is dispensable for HOTAIR-mediated transcriptional repression', *The EMBO journal*, 36(8), pp. 981–994. Available at: <https://doi.org/10.15252/emboj.201695335>.
- Probst, A.V. (2022) 'Deposition and eviction of histone variants define functional chromatin states in plants', *Current Opinion in Plant Biology*, 69, p. 102266. Available at: <https://doi.org/10.1016/j.pbi.2022.102266>.
- Pu, L. *et al.* (2013) 'EMBRYONIC FLOWER1 and ULTRAPETALA1 Act Antagonistically on *Arabidopsis* Development and Stress Response', *Plant Physiology*, 162(2), pp. 812–830. Available at: <https://doi.org/10.1104/pp.112.213223>.
- Punjani, A., Zhang, H. and Fleet, D.J. (2020) 'Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction', *Nature Methods*, 17(12), pp. 1214–1221. Available at: <https://doi.org/10.1038/s41592-020-00990-8>.
- Qian, S. *et al.* (2018) 'Dual recognition of H3K4me3 and H3K27me3 by a plant histone reader SHL', *Nature Communications*, 9(1), p. 2425. Available at: <https://doi.org/10.1038/s41467-018-04836-y>.
- Qin, W. *et al.* (2021) 'Deciphering molecular interactions by proximity labeling', *Nature Methods*, 18(2), pp. 133–143. Available at: <https://doi.org/10.1038/s41592-020-01010-5>.
- Reyes, A.A., Marcum, R.D. and He, Y. (2021) 'Structure and Function of Chromatin Remodelers', *Journal of Molecular Biology*, 433(14), p. 166929. Available at: <https://doi.org/10.1016/j.jmb.2021.166929>.
- Rosa, S. and Shaw, P. (2013) 'Insights into Chromatin Structure and Dynamics in Plants', *Biology*, 2(4), pp. 1378–1410. Available at: <https://doi.org/10.3390/biology2041378>.
- Roudier, F. *et al.* (2011) 'Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*', *The EMBO journal*, 30(10), pp. 1928–1938. Available at: <https://doi.org/10.1038/emboj.2011.103>.
- Roy, D. *et al.* (2019) 'Rice Trithorax factor ULTRAPETALA 1 (OsULT1) specifically binds to "GAGAG" sequence motif present in Polycomb response elements', *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms*, 1862(5), pp. 582–597. Available at: <https://doi.org/10.1016/j.bbagr.2019.02.001>.
- Saare, M. *et al.* (2012) 'Autoimmune regulator is acetylated by transcription coactivator CBP/p300', *Experimental Cell Research*, 318(14), pp. 1767–1778. Available at: <https://doi.org/10.1016/j.yexcr.2012.04.013>.
- Sabatini, S. *et al.* (1999) 'An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root', *Cell*, 99(5), pp. 463–472. Available at: [https://doi.org/10.1016/s0092-8674\(00\)81535-4](https://doi.org/10.1016/s0092-8674(00)81535-4).

- Saleh, A. *et al.* (2008) 'The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions', *The Plant Cell*, 20(3), pp. 568–579. Available at: <https://doi.org/10.1105/tpc.107.056614>.
- Sanulli, S. *et al.* (2015) 'Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation', *Molecular Cell*, 57(5), pp. 769–783. Available at: <https://doi.org/10.1016/j.molcel.2014.12.020>.
- Sauer, P.V. *et al.* (2023) 'Activation of automethylated PRC2 by dimerization on chromatin', *bioRxiv: The Preprint Server for Biology*, p. 2023.10.12.562141. Available at: <https://doi.org/10.1101/2023.10.12.562141>.
- Schmitges, F.W. *et al.* (2011) 'Histone methylation by PRC2 is inhibited by active chromatin marks', *Molecular Cell*, 42(3), pp. 330–341. Available at: <https://doi.org/10.1016/j.molcel.2011.03.025>.
- Schoof, H. *et al.* (2000) 'The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes', *Cell*, 100(6), pp. 635–644. Available at: [https://doi.org/10.1016/S0092-8674\(00\)80700-X](https://doi.org/10.1016/S0092-8674(00)80700-X).
- Schubert, D. *et al.* (2006) 'Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27', *The EMBO journal*, 25(19), pp. 4638–4649. Available at: <https://doi.org/10.1038/sj.emboj.7601311>.
- Schuettengruber, B. *et al.* (2011) 'Trithorax group proteins: switching genes on and keeping them active', *Nature Reviews Molecular Cell Biology*, 12(12), pp. 799–814. Available at: <https://doi.org/10.1038/nrm3230>.
- Schuettengruber, B. *et al.* (2017) 'Genome Regulation by Polycomb and Trithorax: 70 Years and Counting', *Cell*, 171(1), pp. 34–57. Available at: <https://doi.org/10.1016/j.cell.2017.08.002>.
- Sequeira-Mendes, J. *et al.* (2014) 'The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States', *The Plant Cell*, 26(6), pp. 2351–2366. Available at: <https://doi.org/10.1105/tpc.114.124578>.
- Sequeira-Mendes, J. and Gutierrez, C. (2016) 'Genome architecture: from linear organisation of chromatin to the 3D assembly in the nucleus', *Chromosoma*, 125(3), pp. 455–469. Available at: <https://doi.org/10.1007/s00412-015-0538-5>.
- Shahbazian, M.D. and Grunstein, M. (2007) 'Functions of site-specific histone acetylation and deacetylation', *Annual Review of Biochemistry*, 76, pp. 75–100. Available at: <https://doi.org/10.1146/annurev.biochem.76.052705.162114>.
- Shang, J.-Y. *et al.* (2021) 'COMPASS functions as a module of the INO80 chromatin remodeling complex to mediate histone H3K4 methylation in Arabidopsis', *The Plant Cell*, 33(10), pp. 3250–3271. Available at: <https://doi.org/10.1093/plcell/koab187>.
- Shao, Z. *et al.* (1999) 'Stabilization of Chromatin Structure by PRC1, a Polycomb Complex', *Cell*, 98(1), pp. 37–46. Available at: [https://doi.org/10.1016/S0092-8674\(00\)80604-2](https://doi.org/10.1016/S0092-8674(00)80604-2).
- Sharaf, A. *et al.* (2022) 'Phylogenetic profiling resolves early emergence of PRC2 and illuminates its functional core', *Life Science Alliance*, 5(7), p. e202101271. Available at: <https://doi.org/10.26508/lsa.202101271>.

- Shaver, S. *et al.* (2010) 'Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*', *Epigenetics*, 5(4), pp. 301–312. Available at: <https://doi.org/10.4161/epi.5.4.11608>.
- Shen, Y. *et al.* (2014) 'Over-expression of histone H3K4 demethylase gene JMJ15 enhances salt tolerance in *Arabidopsis*', *Frontiers in Plant Science*, 5, p. 290. Available at: <https://doi.org/10.3389/fpls.2014.00290>.
- Shi, Yujiang *et al.* (2004) 'Histone demethylation mediated by the nuclear amine oxidase homolog LSD1', *Cell*, 119(7), pp. 941–953. Available at: <https://doi.org/10.1016/j.cell.2004.12.012>.
- Shim, S. *et al.* (2020) 'H3K36me2 is highly correlated with m6A modifications in plants', *Journal of Integrative Plant Biology*, 62(10), pp. 1455–1460. Available at: <https://doi.org/10.1111/jipb.12917>.
- Shogren-Knaak, M. *et al.* (2006) 'Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions', *Science*, 311(5762), pp. 844–847. Available at: <https://doi.org/10.1126/science.1124000>.
- Shu, H. *et al.* (2014) 'Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes', *Genome Biology*, 15(4), p. R62. Available at: <https://doi.org/10.1186/gb-2014-15-4-r62>.
- Shu, J. *et al.* (2019) 'Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings', *Plant Direct*, 3(1), p. e00100. Available at: <https://doi.org/10.1002/pld3.100>.
- Simon, J. *et al.* (1993) 'Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products', *Developmental Biology*, 158(1), pp. 131–144. Available at: <https://doi.org/10.1006/dbio.1993.1174>.
- Simon, L. *et al.* (2015) 'Structure and Function of Centromeric and Pericentromeric Heterochromatin in *Arabidopsis thaliana*', *Frontiers in Plant Science*, 6. Available at: <https://doi.org/10.3389/fpls.2015.01049>.
- Song, F. *et al.* (2014) 'Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units', *Science (New York, N.Y.)*, 344(6182), pp. 376–380. Available at: <https://doi.org/10.1126/science.1251413>.
- Song, J. *et al.* (2012) 'Vernalization - a cold-induced epigenetic switch', *Journal of Cell Science*, 125(Pt 16), pp. 3723–3731. Available at: <https://doi.org/10.1242/jcs.084764>.
- Song, J. *et al.* (2023) *Structural basis for inactivation of PRC2 by G-quadruplex RNA*. preprint. *Biochemistry*. Available at: <https://doi.org/10.1101/2023.02.06.527314>.
- Srikanth, A. and Schmid, M. (2011) 'Regulation of flowering time: all roads lead to Rome', *Cellular and Molecular Life Sciences*, 68(12), pp. 2013–2037. Available at: <https://doi.org/10.1007/s00018-011-0673-y>.
- Steeves, T.A. and Sussex, I.M. (1989) *Patterns in Plant Development*. 2nd edn. Cambridge: Cambridge University Press. Available at: <https://doi.org/10.1017/CBO9780511626227>.
- Strahl, B.D. and Allis, C.D. (2000) 'The language of covalent histone modifications', *Nature*, 403(6765), pp. 41–45. Available at: <https://doi.org/10.1038/47412>.

- Stroud, H. *et al.* (2012) 'Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*', *Proceedings of the National Academy of Sciences of the United States of America*, 109(14), pp. 5370–5375. Available at: <https://doi.org/10.1073/pnas.1203145109>.
- Stroud, H. *et al.* (2014) 'Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*', *Nature Structural & Molecular Biology*, 21(1), pp. 64–72. Available at: <https://doi.org/10.1038/nsmb.2735>.
- Sugishita, H. *et al.* (2021) 'Variant PCGF1-PRC1 links PRC2 recruitment with differentiation-associated transcriptional inactivation at target genes', *Nature Communications*, 12(1), p. 5341. Available at: <https://doi.org/10.1038/s41467-021-24894-z>.
- Sun, B. *et al.* (2014) 'Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells', *Science (New York, N.Y.)*, 343(6170), p. 1248559. Available at: <https://doi.org/10.1126/science.1248559>.
- Sung, S. and Amasino, R.M. (2004) 'Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3', *Nature*, 427(6970), pp. 159–164. Available at: <https://doi.org/10.1038/nature02195>.
- Surdo, P.L. *et al.* (2003) 'Crystal structure and nuclear magnetic resonance analyses of the SAND domain from glucocorticoid modulatory element binding protein-1 reveals deoxyribonucleic acid and zinc binding regions', *Molecular Endocrinology (Baltimore, Md.)*, 17(7), pp. 1283–1295. Available at: <https://doi.org/10.1210/me.2002-0409>.
- Talbert, P.B. *et al.* (2012) 'A unified phylogeny-based nomenclature for histone variants', *Epigenetics & Chromatin*, 5(1), p. 7. Available at: <https://doi.org/10.1186/1756-8935-5-7>.
- Talbert, P.B. and Henikoff, S. (2017) 'Histone variants on the move: substrates for chromatin dynamics', *Nature Reviews Molecular Cell Biology*, 18(2), pp. 115–126. Available at: <https://doi.org/10.1038/nrm.2016.148>.
- Tamada, Y. *et al.* (2009) 'ARABIDOPSIS TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C', *The Plant Cell*, 21(10), pp. 3257–3269. Available at: <https://doi.org/10.1105/tpc.109.070060>.
- Teper-Bamnolker, P. and Samach, A. (2005) 'The flowering integrator FT regulates SEPALLATA3 and FRUITFULL accumulation in *Arabidopsis* leaves', *The Plant Cell*, 17(10), pp. 2661–2675. Available at: <https://doi.org/10.1105/tpc.105.035766>.
- Tessarz, P. and Kouzarides, T. (2014) 'Histone core modifications regulating nucleosome structure and dynamics', *Nature Reviews. Molecular Cell Biology*, 15(11), pp. 703–708. Available at: <https://doi.org/10.1038/nrm3890>.
- Thorn, A. and Sheldrick, G.M. (2013) 'Extending molecular-replacement solutions with SHELXE', *Acta Crystallographica. Section D, Biological Crystallography*, 69(Pt 11), pp. 2251–2256. Available at: <https://doi.org/10.1107/S0907444913027534>.
- Tian, H. *et al.* (2013) 'Auxin gradient is crucial for the maintenance of root distal stem cell identity in *Arabidopsis*', *Plant Signaling & Behavior*, 8(12), p. e26429. Available at: <https://doi.org/10.4161/psb.26429>.
- Tian, J. *et al.* (2022) 'The epigenetic regulator ULTRAPETALA1 suppresses de novo root regeneration from *Arabidopsis* leaf explants', *Plant Signaling & Behavior*, 17(1), p. 2031784. Available at: <https://doi.org/10.1080/15592324.2022.2031784>.

- Torok, M. and Etkin, L.D. (2001) 'Two B or not two B? Overview of the rapidly expanding B-box family of proteins', *Differentiation; Research in Biological Diversity*, 67(3), pp. 63–71. Available at: <https://doi.org/10.1046/j.1432-0436.2001.067003063.x>.
- Trejo-Arellano, M.S. *et al.* (2017) 'H3K23me1 is an evolutionarily conserved histone modification associated with CG DNA methylation in Arabidopsis', *The Plant Journal*, 90(2), pp. 293–303. Available at: <https://doi.org/10.1111/tpj.13489>.
- Trojer, P. and Reinberg, D. (2007) 'Facultative Heterochromatin: Is There a Distinctive Molecular Signature?', *Molecular Cell*, 28(1), pp. 1–13. Available at: <https://doi.org/10.1016/j.molcel.2007.09.011>.
- Tsukada, Y. *et al.* (2006) 'Histone demethylation by a family of JmjC domain-containing proteins', *Nature*, 439(7078), pp. 811–816. Available at: <https://doi.org/10.1038/nature04433>.
- Usón, I. and Sheldrick, G.M. (1999) 'Advances in direct methods for protein crystallography', *Current Opinion in Structural Biology*, 9(5), pp. 643–648. Available at: [https://doi.org/10.1016/S0959-440X\(99\)00020-2](https://doi.org/10.1016/S0959-440X(99)00020-2).
- Varambally, S. *et al.* (2002) 'The polycomb group protein EZH2 is involved in progression of prostate cancer', *Nature*, 419(6907), pp. 624–629. Available at: <https://doi.org/10.1038/nature01075>.
- Veiseth, S.V. *et al.* (2011) 'The SUV4 Histone Lysine Methyltransferase Binds Ubiquitin and Converts H3K9me1 to H3K9me3 on Transposon Chromatin in Arabidopsis', *PLOS Genetics*, 7(3), p. e1001325. Available at: <https://doi.org/10.1371/journal.pgen.1001325>.
- Veluchamy, A. *et al.* (2015) 'An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*', *Genome Biology*, 16(1), p. 102. Available at: <https://doi.org/10.1186/s13059-015-0671-8>.
- Vijayanathan, M., Trejo-Arellano, M.G. and Mozgová, I. (2022) 'Polycomb Repressive Complex 2 in Eukaryotes—An Evolutionary Perspective', *Epigenomes*, 6(1), p. 3. Available at: <https://doi.org/10.3390/epigenomes6010003>.
- Wang, H. *et al.* (2004) 'Role of histone H2A ubiquitination in Polycomb silencing', *Nature*, 431(7010), pp. 873–878. Available at: <https://doi.org/10.1038/nature02985>.
- Wang, J. *et al.* (2020) 'Cell-type-dependent histone demethylase specificity promotes meiotic chromosome condensation in Arabidopsis', *Nature Plants*, 6(7), pp. 823–837. Available at: <https://doi.org/10.1038/s41477-020-0697-0>.
- Wang, J. *et al.* (2022) 'EZH2 noncanonically binds cMyc and p300 through a cryptic transactivation domain to mediate gene activation and promote oncogenesis', *Nature Cell Biology*, 24(3), pp. 384–399. Available at: <https://doi.org/10.1038/s41556-022-00850-x>.
- Wang, L. *et al.* (2004) 'Hierarchical Recruitment of Polycomb Group Silencing Complexes', *Molecular Cell*, 14(5), pp. 637–646. Available at: <https://doi.org/10.1016/j.molcel.2004.05.009>.
- Wang, M. *et al.* (2023) 'Arabidopsis TRB proteins function in H3K4me3 demethylation by recruiting JM14', *Nature Communications*, 14(1), p. 1736. Available at: <https://doi.org/10.1038/s41467-023-37263-9>.

Wang, Q. and Shen, W.-H. (2018) 'Chromatin modulation and gene regulation in plants: insight about PRC1 function', *Biochemical Society Transactions*, 46(4), pp. 957–966. Available at: <https://doi.org/10.1042/BST20170576>.

Wang, S. *et al.* (2007) 'Polycomblike-2-deficient mice exhibit normal left-right asymmetry', *Developmental dynamics*, 236(3), pp. 853–861. Available at: <https://doi.org/10.1002/dvdy.21070>.

Wang, X. *et al.* (2017) 'Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines', *Molecular Cell*, 65(6), pp. 1056–1067.e5. Available at: <https://doi.org/10.1016/j.molcel.2017.02.003>.

Wang, X. *et al.* (2019) 'Regulation of histone methylation by automethylation of PRC2', *Genes & Development*, 33(19–20), pp. 1416–1427. Available at: <https://doi.org/10.1101/gad.328849.119>.

Wang, Z. *et al.* (2022) 'Prediction of histone post-translational modification patterns based on nascent transcription data', *Nature Genetics*, 54(3), pp. 295–305. Available at: <https://doi.org/10.1038/s41588-022-01026-x>.

Weber, C.M. and Henikoff, S. (2014) 'Histone variants: dynamic punctuation in transcription', *Genes & Development*, 28(7), pp. 672–682. Available at: <https://doi.org/10.1101/gad.238873.114>.

Weigel, D. and Meyerowitz, E.M. (1993) 'Activation of Floral Homeotic Genes in Arabidopsis', *Science*, 261(5129), pp. 1723–1726. Available at: <https://doi.org/10.1126/science.261.5129.1723>.

Wollmann, H. *et al.* (2012) 'Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome', *PLoS genetics*, 8(5), p. e1002658. Available at: <https://doi.org/10.1371/journal.pgen.1002658>.

Woodcock, C.L., Safer, J.P. and Stanchfield, J.E. (1976) 'Structural repeating units in chromatin. I. Evidence for their general occurrence', *Experimental Cell Research*, 97, pp. 101–110. Available at: [https://doi.org/10.1016/0014-4827\(76\)90659-5](https://doi.org/10.1016/0014-4827(76)90659-5).

Wu, H.-W. *et al.* (2018) 'A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves', *New Phytologist*, 219(4), pp. 1480–1491. Available at: <https://doi.org/10.1111/nph.15231>.

Wu, M.-F. *et al.* (2012) 'SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors', *Proceedings of the National Academy of Sciences*, 109(9), pp. 3576–3581. Available at: <https://doi.org/10.1073/pnas.1113409109>.

Wu, S.C. and Zhang, Y. (2010) 'Active DNA demethylation: many roads lead to Rome', *Nature Reviews Molecular Cell Biology*, 11(9), pp. 607–620. Available at: <https://doi.org/10.1038/nrm2950>.

Xiao, J. *et al.* (2017) 'Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis', *Nature Genetics*, 49(10), pp. 1546–1552. Available at: <https://doi.org/10.1038/ng.3937>.

Xiao, J., Lee, U.-S. and Wagner, D. (2016) 'Tug of war: adding and removing histone lysine methylation in Arabidopsis', *Current Opinion in Plant Biology*, 34, pp. 41–53. Available at: <https://doi.org/10.1016/j.pbi.2016.08.002>.

- Xie, S.-S. *et al.* (2023) 'JMJ28 guides sequence-specific targeting of ATX1/2-containing COMPASS-like complex in Arabidopsis', *Cell Reports*, 42(3). Available at: <https://doi.org/10.1016/j.celrep.2023.112163>.
- Xu, F. *et al.* (2018) 'Trithorax Group Proteins Act Together with a Polycomb Group Protein to Maintain Chromatin Integrity for Epigenetic Silencing during Seed Germination in Arabidopsis', *Molecular Plant*, 11(5), pp. 659–677. Available at: <https://doi.org/10.1016/j.molp.2018.01.010>.
- Xu, L. *et al.* (2008) 'Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana', *Molecular and Cellular Biology*, 28(4), pp. 1348–1360. Available at: <https://doi.org/10.1128/MCB.01607-07>.
- Xu, L. and Shen, W.-H. (2008) 'Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in Arabidopsis', *Current Biology*, 18(24), pp. 1966–1971. Available at: <https://doi.org/10.1016/j.cub.2008.11.019>.
- Xu, Y. *et al.* (2016) 'Regulation of Vegetative Phase Change by SWI2/SNF2 Chromatin Remodeling ATPase BRAHMA', *Plant Physiology*, 172(4), pp. 2416–2428. Available at: <https://doi.org/10.1104/pp.16.01588>.
- Xu, Y. *et al.* (2022) 'MSI1 and HDA6 function interdependently to control flowering time via chromatin modifications', *The Plant Journal: For Cell and Molecular Biology*, 109(4), pp. 831–843. Available at: <https://doi.org/10.1111/tpj.15596>.
- Xue, M. *et al.* (2021) 'The INO80 chromatin remodeling complex promotes thermomorphogenesis by connecting H2A.Z eviction and active transcription in Arabidopsis', *Molecular Plant*, 14(11), pp. 1799–1813. Available at: <https://doi.org/10.1016/j.molp.2021.07.001>.
- Yadav, R.K. *et al.* (2011) 'WUSCHEL protein movement mediates stem cell homeostasis in the Arabidopsis shoot apex', *Genes & Development*, 25(19), pp. 2025–2030. Available at: <https://doi.org/10.1101/gad.17258511>.
- Yadav, T., Quivy, J.-P. and Almouzni, G. (2018) 'Chromatin plasticity: A versatile landscape that underlies cell fate and identity', *Science*, 361(6409), pp. 1332–1336. Available at: <https://doi.org/10.1126/science.aat8950>.
- Yan, W. *et al.* (2018) 'Dynamic and spatial restriction of Polycomb activity by plant histone demethylases', *Nature Plants*, 4(9), pp. 681–689. Available at: <https://doi.org/10.1038/s41477-018-0219-5>.
- Yanai, O. *et al.* (2005) 'Arabidopsis KNOX1 Proteins Activate Cytokinin Biosynthesis', *Current Biology*, 15(17), pp. 1566–1571. Available at: <https://doi.org/10.1016/j.cub.2005.07.060>.
- Yang, H. *et al.* (2017) 'Distinct phases of Polycomb silencing to hold epigenetic memory of cold in Arabidopsis', *Science (New York, N.Y.)*, 357(6356), pp. 1142–1145. Available at: <https://doi.org/10.1126/science.aan1121>.
- Yang, H., Howard, M. and Dean, C. (2016) 'Physical coupling of activation and derepression activities to maintain an active transcriptional state at FLC', *Proceedings of the National Academy of Sciences*, 113(33), pp. 9369–9374. Available at: <https://doi.org/10.1073/pnas.1605733113>.

- Yang, J. *et al.* (2020) 'SWI3B and HDA6 interact and are required for transposon silencing in Arabidopsis', *The Plant Journal: For Cell and Molecular Biology*, 102(4), pp. 809–822. Available at: <https://doi.org/10.1111/tpj.14666>.
- Yang, W. *et al.* (2010) 'A plant-specific histone H3 lysine 4 demethylase represses the floral transition in Arabidopsis', *The Plant Journal: For Cell and Molecular Biology*, 62(4), pp. 663–673. Available at: <https://doi.org/10.1111/j.1365-313X.2010.04182.x>.
- Yang, X.-J. and Seto, E. (2007) 'HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention', *Oncogene*, 26(37), pp. 5310–5318. Available at: <https://doi.org/10.1038/sj.onc.1210599>.
- Yang, Z. *et al.* (2018) 'EBS is a bivalent histone reader that regulates floral phase transition in Arabidopsis', *Nature Genetics*, 50(9), pp. 1247–1253. Available at: <https://doi.org/10.1038/s41588-018-0187-8>.
- Yang, Z. *et al.* (2023) 'AlphaFold2 and its applications in the fields of biology and medicine', *Signal Transduction and Targeted Therapy*, 8(1), p. 115. Available at: <https://doi.org/10.1038/s41392-023-01381-z>.
- Ye, R. *et al.* (2022) 'Glucose-driven TOR-FIE-PRC2 signalling controls plant development', *Nature*, 609(7929), pp. 986–993. Available at: <https://doi.org/10.1038/s41586-022-05171-5>.
- Yelagandula, R. *et al.* (2014) 'The Histone Variant H2A.W Defines Heterochromatin and Promotes Chromatin Condensation in Arabidopsis', *Cell*, 158(1), pp. 98–109. Available at: <https://doi.org/10.1016/j.cell.2014.06.006>.
- Yin, X. *et al.* (2021) 'H2AK121ub in Arabidopsis associates with a less accessible chromatin state at transcriptional regulation hotspots', *Nature Communications*, 12(1), p. 315. Available at: <https://doi.org/10.1038/s41467-020-20614-1>.
- Yoshida, N. *et al.* (2001) 'EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis', *The Plant Cell*, 13(11), pp. 2471–2481. Available at: <https://doi.org/10.1105/tpc.010227>.
- Yuan, L. *et al.* (2021) 'The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in Arabidopsis', *Nucleic Acids Research*, 49(1), pp. 98–113. Available at: <https://doi.org/10.1093/nar/gkaa1129>.
- Yuan, W. *et al.* (2016) 'A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in Arabidopsis', *Nature Genetics*, 48(12), pp. 1527–1534. Available at: <https://doi.org/10.1038/ng.3712>.
- Yun, M. *et al.* (2011) 'Readers of histone modifications', *Cell Research*, 21(4), pp. 564–578. Available at: <https://doi.org/10.1038/cr.2011.42>.
- Zemach, A. *et al.* (2013) 'The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin', *Cell*, 153(1), pp. 193–205. Available at: <https://doi.org/10.1016/j.cell.2013.02.033>.
- Zeng, L. *et al.* (2010) 'Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b', *Nature*, 466(7303), pp. 258–262. Available at: <https://doi.org/10.1038/nature09139>.

- Zhang, H., Lang, Z. and Zhu, J.-K. (2018) 'Dynamics and function of DNA methylation in plants', *Nature Reviews Molecular Cell Biology*, 19(8), pp. 489–506. Available at: <https://doi.org/10.1038/s41580-018-0016-z>.
- Zhang, H. and Zhu, J.-K. (2012) 'Active DNA Demethylation in Plants and Animals', *Cold Spring Harbor Symposia on Quantitative Biology*, 77, pp. 161–173. Available at: <https://doi.org/10.1101/sqb.2012.77.014936>.
- Zhang, K. *et al.* (2007) 'Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*', *PloS One*, 2(11), p. e1210. Available at: <https://doi.org/10.1371/journal.pone.0001210>.
- Zhang, Q. *et al.* (2019) 'RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2', *Nature structural & molecular biology*, 26(3), pp. 237–247. Available at: <https://doi.org/10.1038/s41594-019-0197-y>.
- Zhang, X. *et al.* (2006a) 'Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*', *Cell*, 126(6), pp. 1189–1201. Available at: <https://doi.org/10.1016/j.cell.2006.08.003>.
- Zhang, X. *et al.* (2006b) 'Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*', *Cell*, 126(6), pp. 1189–1201. Available at: <https://doi.org/10.1016/j.cell.2006.08.003>.
- Zhang, X., Germann, S., *et al.* (2007) 'The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation', *Nature Structural & Molecular Biology*, 14(9), pp. 869–871. Available at: <https://doi.org/10.1038/nsmb1283>.
- Zhang, X., Clarenz, O., *et al.* (2007) 'Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in *Arabidopsis*', *PLOS Biology*, 5(5), p. e129. Available at: <https://doi.org/10.1371/journal.pbio.0050129>.
- Zhang, X. *et al.* (2009) 'Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*', *Genome Biology*, 10(6), p. R62. Available at: <https://doi.org/10.1186/gb-2009-10-6-r62>.
- Zhang, X. *et al.* (2020) 'Arabidopsis SDG8 Potentiates the Sustainable Transcriptional Induction of the Pathogenesis-Related Genes PR1 and PR2 During Plant Defense Response', *Frontiers in Plant Science*, 11. Available at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.00277> (Accessed: 8 April 2023).
- Zhang, Y. *et al.* (2021) 'Overview of Histone Modification', *Advances in Experimental Medicine and Biology*, 1283, pp. 1–16. Available at: [https://doi.org/10.1007/978-981-15-8104-5\\_1](https://doi.org/10.1007/978-981-15-8104-5_1).
- Zhang, Y. and Reinberg, D. (2001) 'Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails', *Genes & Development*, 15(18), pp. 2343–2360. Available at: <https://doi.org/10.1101/gad.927301>.
- Zhang, Y.-Z. *et al.* (2020) 'Coupling of H3K27me3 recognition with transcriptional repression through the BAH-PHD-CPL2 complex in *Arabidopsis*', *Nature Communications*, 11(1), p. 6212. Available at: <https://doi.org/10.1038/s41467-020-20089-0>.
- Zhao, S., Allis, C.D. and Wang, G.G. (2021) 'The language of chromatin modification in human cancers', *Nature Reviews Cancer*, 21(7), pp. 413–430. Available at: <https://doi.org/10.1038/s41568-021-00357-x>.

- Zheng, S. *et al.* (2019) 'The Arabidopsis H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor', *Nature Communications*, 10(1), p. 1303. Available at: <https://doi.org/10.1038/s41467-019-09310-x>.
- Zhou, B.-R. and Bai, Y. (2019) 'Chromatin structures condensed by linker histones', *Essays in Biochemistry*, 63(1), pp. 75–87. Available at: <https://doi.org/10.1042/EBC20180056>.
- Zhou, J. *et al.* (2010) 'Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression', *Plant Molecular Biology*, 72(6), pp. 585–595. Available at: <https://doi.org/10.1007/s11103-009-9594-7>.
- Zhou, Y. *et al.* (2017) 'H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity', *Genome Biology*, 18(1), p. 69. Available at: <https://doi.org/10.1186/s13059-017-1197-z>.
- Zhou, Y. *et al.* (2018) 'Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis', *Nature Genetics*, 50(5), pp. 638–644. Available at: <https://doi.org/10.1038/s41588-018-0109-9>.
- Zhu, J.-K. (2009) 'Active DNA demethylation mediated by DNA glycosylases', *Annual Review of Genetics*, 43, pp. 143–166. Available at: <https://doi.org/10.1146/annurev-genet-102108-134205>.