



**HAL**  
open science

# Functional, structural and evolutionary aspects of the auxin transcriptional response

Raquel Martín-Arevalillo

► **To cite this version:**

Raquel Martín-Arevalillo. Functional, structural and evolutionary aspects of the auxin transcriptional response. Structural Biology [q-bio.BM]. Université Grenoble Alpes, 2017. English. NNT : 2017GREAV073 . tel-03029246

**HAL Id: tel-03029246**

**<https://theses.hal.science/tel-03029246>**

Submitted on 28 Nov 2020

**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 LA COMMUNAUTE UNIVERSITE  
GRENOBLE ALPES**

Spécialité : **Biologie Structurale**

Arrêté ministériel : 25 mai 2016

Présentée par

**Raquel MARTÍN-AREVALILLO**

Thèse dirigée par **Renaud DUMAS** préparée au sein du  
**Laboratoire de Physiologie Cellulaire et Végétale** dans l'École  
**Doctorale de Chimie et Sciences du Vivant**

## **Aspects fonctionnels, structuraux et évolutifs de la réponse transcriptionnelle à l'auxine**

Thèse soutenue publiquement le **27 Novembre 2017**  
devant le jury composé de :

**Dr. Catherine BELLINI**

Directrice de recherche, INRA, Rapportrice

**Pr. Miguel BLÁZQUEZ**

Professeur, Université de Valencia, Rapporteur

**Pr. Eva PEBAY-PEYROULA**

Professeure, Université Grenoble Alpes, Présidente

**Dr. Teva VERNOUX**

Directeur de recherche, CNRS, Examinateur

**Dr. François PARCY**

Directeur de recherche, CNRS, Examinateur

**Dr. Renaud DUMAS**

Directeur de recherche, CNRS, Directeur de thèse





*None but those who have experienced them can conceive of the enticements of science. In other studies, you go as far as others have gone before you, and there is nothing more to know; but in a scientific pursuit there is continual food for discovery and wonder.*

*Mary Shelley, Frankenstein*



## Abstract

Auxin is a plant hormone implicated in almost all plant developmental stages, since the embryo formation till flowering, determining the position of the organs in the plant and thus, its whole structure. As for any other hormone, auxin perception is followed by a signal transduction that finishes in a series of changes in a plant cell, including transcriptional changes. This thesis is divided in 3 chapters, each with a focus on the structural, molecular and evolutionary aspects of different proteins involved in the regulation of auxin genes response.

First, we focused our studies on TOPLESS (TPL), a co-repressor implicated, not only in auxin responsive genes repression, but also in many other plant processes due to its interactions with numerous transcriptional repressors in plants. Our determination of the TPL N-terminal structure allowed us to understand that TPL can interact with different partners through the same binding site. Moreover, it revealed that TPL is a tetrameric protein, with the tetramerization interface formed by a newly identified domain, the CRA domain, that is also part of the binding site. The high residues conservation in both tetramerization interface and TPL binding site since m.y.a indicates the importance of TPL role since the origin of plants. This work also shows that the structural similarities between TPL and other co-repressor with similar domains but different function nicely exemplify how evolution plays with common features for creating new functions.

Second, we studied ARF proteins, the transcription factors of the auxin transcriptional response, with a focus on their DNA binding preferences. For this, we used a combination of bioinformatic analyses of DAP-seq ARFs genomic binding, with *in vitro* DNA binding tests and structure modelling. Our results point out that different ARFs can have different preferential binding sites within the genome, with these preferences being determined by the orientation and spacing of the binding motifs. Moreover, our studies suggest that depending on the binding site, ARFs could bind with different conformations using dimerization interfaces not yet discovered. These results can explain how different ARFs co-expressed inside a plant cell can collaborate to the specificity and robustness of auxin transcriptional response by differential bindings to the genome.

Finally, we travelled back in time to position the origin of auxin signalling pathway in the evolution of plants. Here we looked for protein homologues of the auxin signalling pathway in charophyte green algae, the most ancient plants ancestor (450 M years). This search retrieved an ARF and a TPL homologue in the first multicellular charophyte algae (*Chlorokybus atmophyticus*). The biochemical characterization of *C. atmophyticus* ARF indicated that it presented already the same properties of the ARFs from land plants and that it was able to interact with TPL protein, as it is the case for some ARFs. The absence of auxin receptor homologues in these primitive algae indicates however that auxin-dependency appeared with the acquisition of TIR1/AFB-Aux/IAA coreceptor system, after charophytes divergence into land plants.

## Résumé

L'auxine est une hormone végétale impliquée dans presque toutes les étapes du développement des plantes, de la formation de l'embryon jusqu'à la floraison, déterminant la position des organes et donc la structure de la plante. Comme pour les autres hormones, la perception de l'auxine est suivie par une transduction du signal qui produit une série de changements dans les cellules végétales dont des régulations transcriptionnelles. Cette thèse est divisée en 3 chapitres, chacun d'eux étant focalisé sur des aspects structuraux, moléculaires et évolutifs de différentes protéines impliquées dans la régulation des gènes de réponse à l'auxine.

Nous avons tout d'abord centré nos études sur TOPLESS (TPL), un corépresseur qui agit au niveau de la répression des gènes de réponse à l'auxine, mais aussi dans d'autres processus végétaux compte tenu de son interaction avec de nombreux répresseurs transcriptionnels. Nous avons déterminé la structure de la partie N-terminale de TPL et compris comment TPL interagit avec différents partenaires au niveau d'un même site de liaison. Nous avons alors démontré que TPL forme un tétramère à l'aide d'une surface de tétramérisation constituée par un nouveau domaine, le domaine CRA, qui fait aussi partie du site de liaison. Les résidus impliqués dans la tétramérisation et l'interaction avec des partenaires sont très conservés depuis des centaines de millions d'années montrant ainsi l'importance du rôle de TPL depuis l'origine des plantes. Enfin, les similarités de structure entre TPL et d'autres corépresseurs qui possèdent des domaines similaires mais possédant une fonction différente montrent un bel exemple de la manière dont l'évolution joue avec des domaines protéiques pour créer de nouvelles fonctions.

Nous avons ensuite étudié les préférences de liaison à l'ADN des facteurs de transcription de la réponse à l'auxine (ARF). Pour cela nous avons utilisé une combinaison d'analyses bio-informatiques de données de DAP-seq sur la liaison des ARFs sur le génome, des tests d'interaction ADN-protéine *in vitro* et de la modélisation de structures. Nos résultats indiquent que les différents ARFs ont des sites préférentiels de liaison sur le génome et que ces préférences sont déterminées par l'orientation et l'espacement entre motifs de liaison. Enfin, ces études suggèrent qu'en fonction du site de liaison, les ARFs pourraient se lier avec différentes conformations à l'aide de surfaces de dimérisation qui ne sont pas encore décrites. Ces résultats permettent d'expliquer comment différents ARFs coexprimés dans la même cellule peuvent fonctionner ensemble pour contribuer à une réponse transcriptionnelle à l'auxine spécifique et robuste.

Finalement, nous avons remonté le temps pour positionner l'origine de la voie de signalisation de l'auxine chez les plantes. Pour cela, nous avons recherché des homologues des protéines de la voie de signalisation de l'auxine dans des algues vertes charophytes, les ancêtres les plus lointains (450 Millions d'années) des plantes. Nous avons alors trouvé un homologue des ARFs et TPL chez les premières algues multicellulaires (*Chlorokybus atmophyticus*). La caractérisation biochimique de

l'ARF de *C. atmophyticus* indique qu'il partageait déjà les mêmes propriétés que les ARFs des plantes terrestres et était aussi capable d'interagir avec TPL comme certains ARFs. L'absence d'homologues du récepteur de l'auxine chez ces algues primitives indique cependant que la dépendance à l'auxine aurait été acquise plus tard avec l'apparition du système corécepteur TIR1/AFB-Aux/IAA après la divergence des charophytes vers les plantes terrestres.



## Acknowledgements

It might be my name which appears in the cover of this manuscript and it will be me the person that will have to defend these results in front of everybody (trying not to freak out), but all this it is not just my work. It would not have been possible without the help of many people and therefore, I need to say many, many thanks to you all.

First, thank you to the components of my thesis jury, Miguel Blázquez, Catherine Bellini, Eva Pebay-Peyroula and Teva Vernoux for having taken the time to evaluate this work and discuss it with me.

Thank you, Norbert Rolland, Laurent Blanchoin, Eric Maréchal and François Parcy for having welcome me in your laboratory.

François, to you I have to say thanks for many other things. Thank you for having shared with me your experience, knowledge, enthusiasm, ideas and jokes. You have always been available for me for whatever need I had, no matter if it was for discussing results, making “sexier” presentations or trying to help me with my incompetence with computers. I learnt a lot from you both at the professional and the personal levels. Merci beaucoup.

Renaud... I can't believe how lucky I am to have had you as my supervisor. I don't think I can thank you enough for everything you've done for me during these 3 years. You have taught me protein biochemistry and structure. You taught me to be critic with myself and my results at the same time that you gave me freedom for doing things the way I thought they should be done. You were always available for discussing with me. And I guess all this was part of your work, an amazing work. But it is not just that; you have done for me many other things that were not part of your work. When things weren't going good, you've always had a joke to say to make me smile again. When you've seen me stressed, tired, sad or ill you have been worried about me and you have always offered me your help. You have even tried to find a husband for me! And this is the only thing in which you failed (:P) because with everything else you helped me to feel like home. Merci, merci et merci pour tout, Renaud.

Thomasito and Manuelito, thank you for all the help with the experiments. I remember arriving to the lab, not being able to find anything (not even my own office) and you were always willing to help and to teach me new stuffs. You taught me more than science, you also taught me French! Thanks to you two and Gregounette and PH, now I can go as rude as I want when speaking French. I had a lot of fun with you guys 😊.

Euge! Luckily you were there, in the bench next to mine helping me not to forget how to speak my own language. It's been really “con padre” and “con madre” to have our Hispanic side of the lab.

Muchas gracias por haberme dedicado tiempo, paciencia y una sonrisa cuando estaba a punto de subirme por las paredes.

Arnaud and Adrien, A&A, you've turned my last months here into a nightmare of oligos. Just kidding. I've loved to bother you every 5 minutes with all my questions about bioinformatics and I had a lot of fun discussing with you.

Marie and Gaby, thank you for always transmitting so much happiness. You're always in such a good mood and so ready to help everybody that it's been a huge pleasure to share lunches, time and jokes with you. Marie, please, don't give up in our task of hiring a handsome cleaning gentleman. Send pictures!

Sophie, Tiffany and Danielle, thank you for your good mood and efficiency. You helped me a lot not being such a disaster with the administrative issues.

Leonie, unluckily it's been just a few months that we could spend together but still, it's been enough to realize your passion for research. Good luck with your crystals!

Cristel, Gilles, Agnes, Vero, Aditya and Chloe, I didn't have the chance to work with you directly but it's been very nice to meet you all. Thank you for all your help and advices.

Max Nanao and David Cobessi, thank you for the crystallographic work and for inviting me over to the lines. It's been really exciting to see our crystals turning around in the robot.

Teva, Antoine and David, les Lyonnais, thanks for all the amazing work you have developed in Lyon, for the discussions, the ideas and for inviting us over for beers. It has been very nice collaborating with you, both at the professional and personal levels.

Thank you to all my friends here. We've travelled, we've danced, we've eaten and we've drunk. We've had so much laughter and fun in a country that was not ours but that we made ours, somehow. Thank you for helping me to not miss my family and culture that much by turning yourselves into my family and culture.

Thank you to my family, the real one this time, the one in Spain. Gracias por el apoyo incondicional que siempre me brindáis, decida lo que decida y vaya donde vaya. Si estoy aquí es indudablemente por vosotros. Os he echado mucho de menos.

Along the manuscript you'll find conclusions for the results, but here you have my conclusion, the summary of these three years of my life: I have been happy and this is thanks to each and all of you.

Thank you.

Merci.

Gracias.

## Abbreviations

2,4-D	2,4-Dichlorophenoxyacetic Acid
4-Cl-IAA	4-Chloroindole-3-Acetic Acid
ABA	Absciscic Acid
ABC	ATP-Binding Cassette
ABI3	Absciscic Acid Insensitive 3
ABP1	Auxin Binding Protein 1
AG	AGAMOUS
AP2	APETALA 2
ARF	Auxin Responsive Factor
Aux/IAA	Auxin/ Indole-3-Acetic Acid Protein
AUX1	Auxin Transporter Protein 1
AuxRE	Auxin Response Element
BR	Brassinosteroid
BRM	BRAHMA
cDNA	Complementary Deoxyribonucleic Acid
CK	Cytokinin
COI1	Coronatine Insensitive 1
CRA	CT11 RanBPM
CTLH	C-terminal to LisH
DAP-seq	DNA-affinity Purification Sequencing
DBD	DNA-Binding Domain
DD	Dimerization Domain
DNA	Deoxyribonucleic Acid
DNA	MTase DNA Methyltransferases
DR	Direct Repeat
EAR	Ethylene-responsive element binding factor-associated Amphiphilic Repression
EMSA	Electrophoretic Mobility Shift Assay
ER	Endoplasmic reticulum
ER	Everted Repeat

ERF	Ethylene-Responsive Factors
FD	Flanking Domain
FL	Full-Length
GA	Gibberellins
gDNA	genomic Deoxyribonucleic Acid
GH3	Gretchen Hagen 3
Gro	Groucho
GST	Glutathione-S-transferase
HAT	Histone Acetyltransferases
HDA	Histone Deacetylases
HTRF	Homogeneous Time Resolved Fluorescence
IAA	Indole-3-Acetic Acid
IAA-Ala	IAA-Alanine
IAA-Asp	IAA-Aspartate
IAA-Glu	IAA-Glutamate
IAAld	Indole-3-Acetaldehyde
IAA-Leu	IAA-Leucine
IAA-Trp	IAA-Tryptophan
IAM	Indolacetamide
IAN	Indole-3-Acetonitrile
IAOx	Indole-3-Acetaldoxine
IGP	Indole-3-Glycerol Phosphate
IPTG	Isopropyl- $\beta$ -D-1-thyogalactopiranoside
IPyA	Indole-3-Pyruvic Acid
IR	Inverted Repeat
JA	Jasmonic Acid
JAZ	Jasmonate-ZIM Domain
LAX	Auxin Transporter-like Protein
LEC2	Leafy Cotyledon 2
LisH	Lissencephaly
LRR	Leucine Rich Repeat
L-Trp	L-Tryptophan

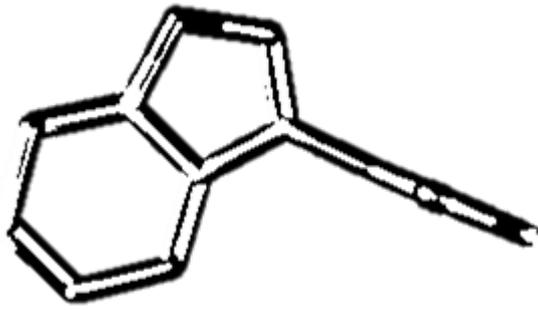
LUG/LUH	LEUNIG/LEUNING-HOMOLOGUE (LUG/LUG)
MBP	Maltose Binding Protein
mRNA	Messenger Ribonucleic Acid
NAA	1-Naphtaleneacetic Acid
NAP	Nuclear Auxin Pathway
NCoR	Nuclear Receptor Co-repressor
NO	Nitric Oxide
PAA	Phenylacetic Acid
PB1	Phox/Bemp1
PWM	Position Weight Matrixes
RAV	Related to ABI3/VP1
REM	Reproductive Meristem
SA	Salicylic Acid
SDM	Site Directed Mutagenesis
SEC	Size Exclusion Chromatography
SEC-MALLS	SEC-Multi-Angle Laser Light Scattery
SEU	SEUSS
SKP2a	S-Phase Kinase-Associated Protein 2a
SL	Strigolactone
SMRT	Silencing Mediator of Retinoic cid and Thyroid hormone receptor
SYD	SPLAYED
TAA1	Tryptophan Aminotransferase 1
TBL1	Transducin Beta-like 1
TF	Transcription Factor
TIR1/AFB	Transport Inhibitor Resistant 1/Auxin Signalling F-box
TPL/TPR	TOPLESS/TOPLESS-related
VAL	VP1/ABI3-like
WUS	WUSCHEL
Y2H	Yeast-2-hybrid
ZAT	Zinc Finger of <i>Arabidopsis</i>



## Contents

<b>General introduction to auxin, the hormone growth .....</b>	<b>19</b>
Hormones in plants! Really? Why?.....	19
A little bit of history... ..	20
Hormones inside a plant cell .....	21
Auxins: the hormone that influences almost everything .....	22
Auxin synthesis .....	23
Auxin inactivation .....	24
Auxin transport.....	25
Auxin signalling .....	27
<b>Objectives.....</b>	<b>35</b>
<b>Chapter I.....</b>	<b>37</b>
<b>Introduction .....</b>	<b>39</b>
Gene expression control in eukaryotic organisms .....	39
Co-repressors.....	40
Co-repression in plants .....	43
TOPLESS co-repressor mechanism .....	45
TPL in the specific context of auxin signalling.....	47
<b>Article 1 .....</b>	<b>51</b>
<b>Supplementary Information .....</b>	<b>57</b>
SI Figures .....	58
SI Tables.....	67
SI Materials and Methods.....	72
SI References.....	78
<b>Complementary results and discussion .....</b>	<b>80</b>
<b>Conclusions .....</b>	<b>83</b>
<b>Chapter II.....</b>	<b>85</b>
<b>Introduction .....</b>	<b>87</b>
DNA-TFs interplay in the control of gene expression .....	87
ARFs, transcription factors of the auxin response.....	88
ARFs DNA binding sites and mechanism.....	92
<b>Results.....</b>	<b>97</b>

<b>Discussion .....</b>	<b>112</b>
<b>Conclusions.....</b>	<b>117</b>
<b>Chapter III.....</b>	<b>119</b>
<b>Introduction.....</b>	<b>121</b>
Charophyte organisms: from water to land.....	121
B3 family .....	122
Auxin signalling: from bryophytes to flowers .....	125
Auxin signalling clues in charophytes? .....	126
<b>Results.....</b>	<b>128</b>
<b>Discussion .....</b>	<b>141</b>
<b>General discussion and conclusions.....</b>	<b>147</b>
<b>Materials &amp; Methods.....</b>	<b>151</b>
<b>Bibliography.....</b>	<b>161</b>
<b>Annexe.....</b>	<b>169</b>



# General introduction to auxin, the hormone growth





## General introduction to auxin, the hormone growth

### Hormones in plants! Really? Why?

Whereas most animals can move around escaping from danger or from undesired environmental conditions, plants, as sessile organisms, are subjected for their whole life to an ever-changing (and often adverse) environment from which they will receive very different stimuli. Diverse light, temperature, atmosphere and soil conditions, pathogens or abiotic stresses are some of them. Therefore, plants need to coordinate all these external cues with their physiological and developmental state. Contrary to animals, plants have continuous sources of stem cells, called meristems. Meristems allow the *de novo* formation of organs all along a plant's life conferring them a high developmental plasticity that helps them coping with external and internal signals (Jaillais and Chory, 2010).

Plant hormones, also called phytohormones, are a structurally and chemically diverse group of plant secondary metabolites with a main role in the integration of all these external and internal signs. Instead of being produced in specific organs, as it happens in animals, phytohormones can be produced and sensed by all or almost all plant cells where they will have a range of physiological effects at specific stages of a plant's life (Dharmasiri et al., 2013; Kumar et al., 2016; Rigal et al., 2014).

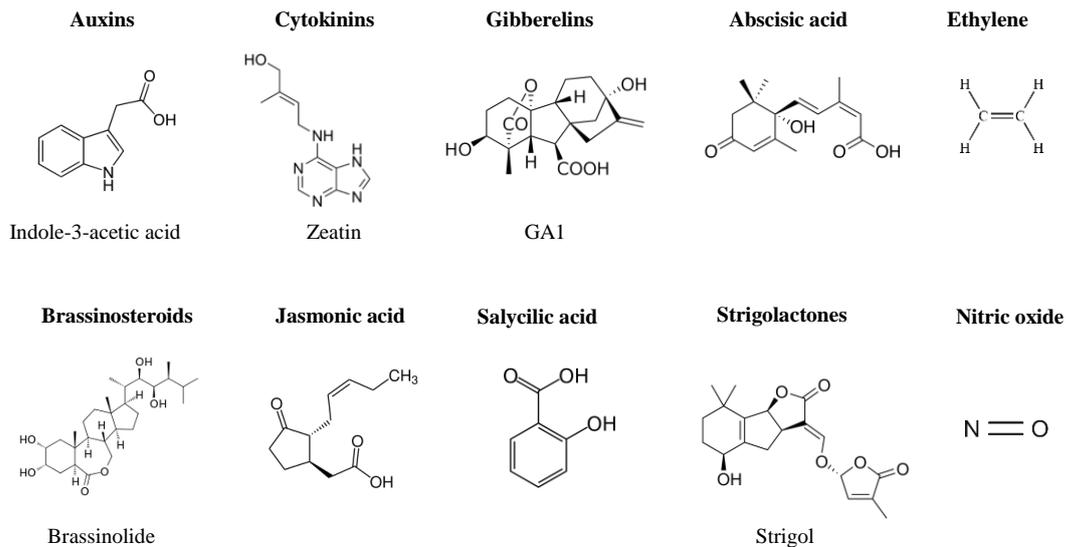
In order to more specifically determine if a molecule corresponds to a plant hormone or not several criteria have been established (Leyser, 1998):

- It must be active at concentrations lower than  $10^{-6}$ M
- It must be synthesized by the plant
- It must be transported some distance (at least one cell diameter)
- It must have some important physiological effect on the plant
- It must act by non-covalent binding to a specific receptor, remain non-covalently bound and non-covalently modified while acting.

According to these criteria, nowadays 10 hormones have been or are being described: auxins (IAA, Indole-3-acetic acid), cytokinins (CKs), gibberellins (GAs), abscisic acid (ABA), ethylene, brassinosteroids (BRs), jasmonic acid (JA), salicylic acid (SA), strigolactones (SLs) and nitric oxide (NO) (Fig 1).

### A little bit of history...

At the end of the 18<sup>th</sup> century a pineapple farmer in the Azores discovered that the fumes he was using to try to kill insects in the greenhouses were promoting an early flowering of pineapple trees. Since then, smoke started being used as a synchronizer of flowering in agriculture without people knowing that the phytohormone ethylene was behind these effects.



**Fig 1. Chemical structure of plant hormones.**

The first evidences of some kind of mobile signal within plants came from the Darwin family. In the end of the 18<sup>th</sup> century, Charles and Francis Darwin discovered that light induced a differential elongation in grass coleoptiles and proposed that this was mediated by a signal transported towards the roots creating an unequal redistribution that regulated plant curvature towards the light (Sauer et al., 2013) (Fig 2).

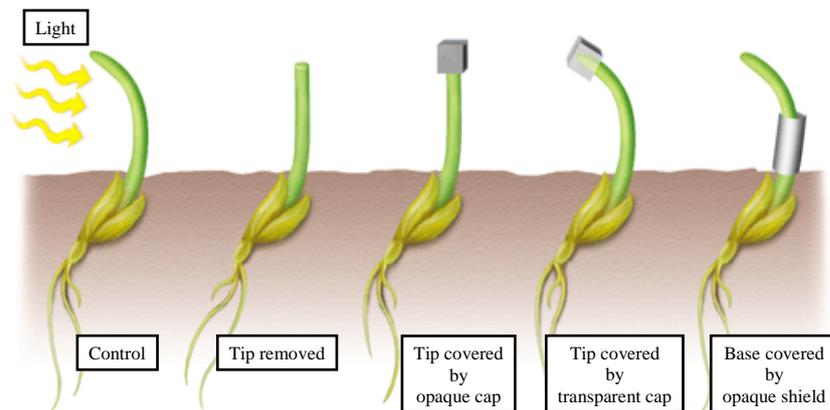
It took around 50 years more (Went, 1926) to discover that the mobile signal causing these effects was auxin, and even a few more for auxin to be isolated from higher plants (1946), being this the first plant hormone to be discovered.

Simultaneously, in 1926 in Japan, the scientist Eiichi Kurosawa was studying the “bakanae” (in Japanese “foolish seedling”) disease in rice that caused an excessive growth of the plants which could no longer stand their own weight -quite foolish, indeed. He discovered that these effects were due to the infection by the fungus *Gibberella fujikuroi*. Years later, the molecule responsible for

this was isolated from fungus extracts and was named gibberellins. Gibberellins were first isolated from plants and thus defined as plant hormones in 1956 (Baca and Elmerich, 2007).

But if there is a winner in the weirdest story, those are cytokinins. Carlos Miller, a postdoc in Folke Karl Skoog's lab (the same Skoog to whom we must thank for Murashige&Skoog medium) was working in the improvement of plant tissues growing medium by adding different extracts. Surprisingly, the additive that seemed to work the best as a growth promoter was autoclaved herring sperm, so he ordered a whole keg of it! Cytokinins were then, the third of the "classical phytohormones" to be discovered, as adenine-derivatives that resulted from DNA degradation (Baca and Elmerich, 2007).

Even if we have been using plant hormones effects on agriculture for decades, it is now that we are starting to understand phytohormones and the molecular mechanisms below their actions mainly thanks to the recent advances in *Arabidopsis thaliana* model plant genetics, that have allowed an extensive characterization of many of the components involved in the production, sensing and response of plants to phytohormones.



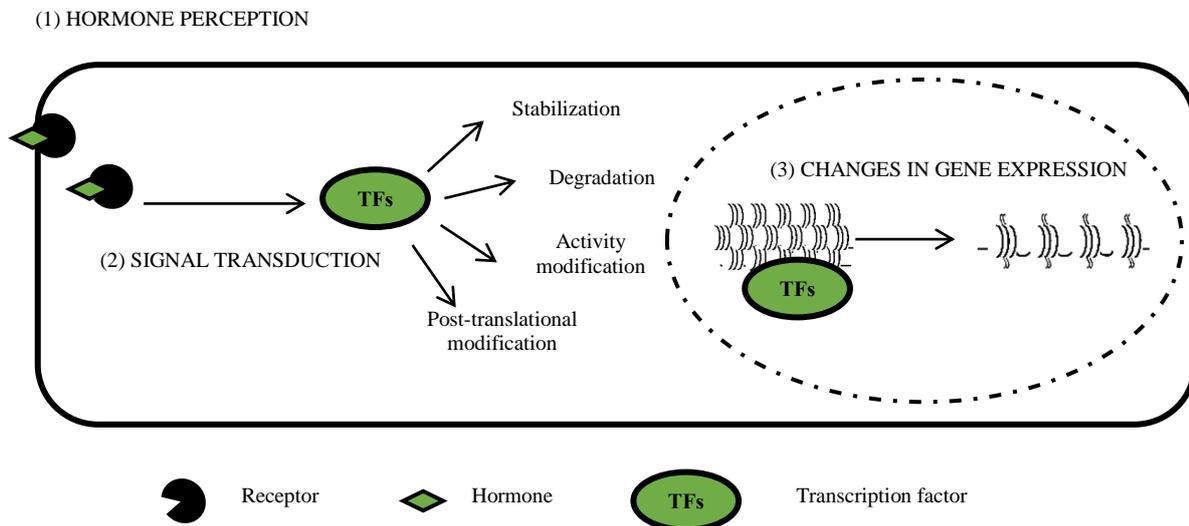
**Fig 2. Darwin's phototropic response experiments.** Grass seedlings grow towards light. When the tip is removed or covered by an opaque cap the shoots grow straight up. When the opaque cap is displaced by a transparent one the curvature still takes places. If the opaque cap is placed on the base of the seedling instead of on the tip there is still a phototropic response. These results suggested the presence of a mobile signal induced by light on the tip that is transported down to the growing region of the shoot.

### Hormones inside a plant cell

The response of a plant cell to a phytohormone will depend on two aspects. First, on the local concentration of a hormone and its availability, which in turn can be modulated through different processes: its synthesis, conjugation, degradation, compartmentalization within the cell and transport within the whole plant.

On the other hand, it will be dependent on the sensitivity of a cell to the available hormone concentration: a hormone signal will trigger a series of changes and responses in a cell that can be transcriptional changes (slow response) or non-transcriptional changes (quick response). The hormonal signal will be detected by receptors either in the plasma membrane or inside the cell that will, normally, provoke changes in transcription factors (stabilization, degradation, activity modification or post-translational modifications) which will induce activation or repression of genes responding to this hormone. Depending on the signalling components present in the cell at that particular moment the response to this hormone concentration can vary (Dharmasiri et al., 2013) (Fig 3).

In most of the cases, part of the signalling process includes a feedback regulation by the hormone itself, which contributes to the hormone homeostasis.



**Fig 3. Hormonal signalling cascade.** Hormone perception by receptors in the membrane or inside the cell (1) triggers changes in transcription factors (2) which in turn will regulate expression of hormone-responsive genes (3).

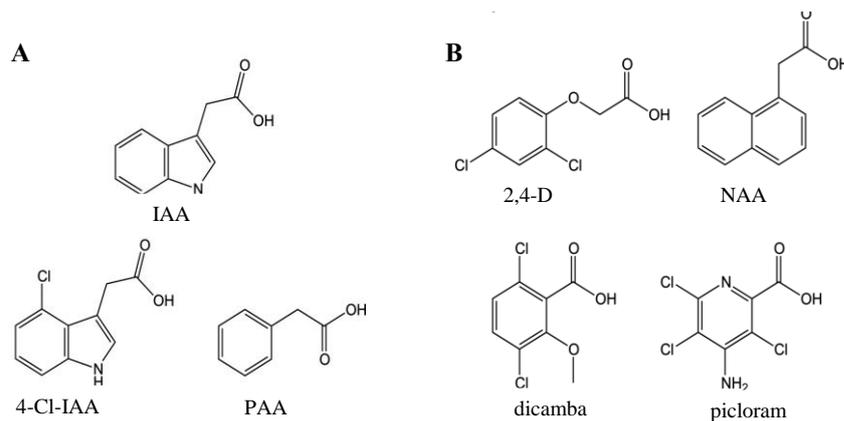
### Auxins: the hormone that influences almost everything

As previously mentioned, auxins were the first hormones to be discovered and since, they have been the most studied ones due to its importance in plant development. Auxin effects are apparent along all the plant and its developmental stages, from embryogenesis, during which it controls apical-basal polarization, to senescence and from tip of the root to tip of the shoot (Dharmasiri et al., 2013; Ruiz Rosquete et al., 2012). It has been involved in hypocotyl elongation, tropic responses and responses to pathogens or abiotic stresses (Sauer et al., 2013; Wang and Estelle, 2014). Furthermore, it induces organogenesis at the meristems, which determines plant architecture. At the

cellular level auxins have been related to cell division, cell expansion and cell differentiation (Paque and Weijers, 2016).

Chemically speaking, auxins are weak organic acids that present an aromatic ring coupled to a side chain containing a terminal carboxyl group (Fig 4) (Ljung, 2013; Sauer et al., 2013).

A more physiological definition describes active auxins as molecules that when exogenously supplied to a plant, induce in it an auxin response. In this sense, several naturally occurring auxin forms have been identified: indole-3-acetic acid (IAA), 4-chloroindole-3-acetic acid (4-Cl-IAA) and phenylacetic acid (PAA) (Fig 4.A). On the other hand, they also exist synthetic active auxins compounds: 2,4-dichlorophenoxyacetic acid (2,4-D), 1-naphtaleneacetic acid (NAA), 3-6-dichloro-2-methoxy-benzoic acid (dicamba) and 4-amino-3,5,6-trichloropicolinic acid (picloram) (Fig4.B) (Korasick et al., 2013). Synthetic auxins are frequently used in agriculture as herbicides or rooting agents and in plant tissue cultures, due to its higher stability and easier absorption by cells (Ljung, 2013).



**Fig 4. Chemical structures of naturally occurring (A) and artificial (B) auxin compounds.** Adapted from (Korasick et al., 2013).

Same as it happens with any other hormone, the physiological auxin response is first subjected to the auxin content inside a cell -in turn dependent on auxin synthesis, conjugation, compartmentalization and transport- and then to the combination of components involved in auxin signalling present at that precise moment inside the cell nucleus (Wang and Estelle, 2014).

### Auxin synthesis

Although many cell types can produce auxin, it is preferentially synthesized in the aerial parts of the plants, especially in young leaves and meristems from which it can be transported to other parts of the plant (Paque and Weijers, 2016; Ruiz Rosquete et al., 2012).

L-Tryptophan (L-Trp)-dependent biosynthesis is thought to be the main way of IAA production in plants. L-Trp is produced from chorismate, the final compound of the shikimate pathway, that takes place in the plastid. Later, in the cytosol L-Trp can be transformed into IAA through different precursors: indole-3-pyruvic acid (IPyA), indolacetamide (IAM), indole-3-acetaldoxime (IAOx), indole-3-acetonitrile (IAN) and indole-3-acetaldehyde (IAAld) (Korasick et al., 2013; Ljung, 2013). Although some of the enzymes that produce these different compounds have been identified, IAA synthesis from IPyA is the only fully described auxin synthetic pathway in plants. It involves a two-step process: first the transformation of L-Trp in IPyA by a tryptophan aminotransferase (TAA1) followed by its conversion into IAA by flavin mono-oxygenases from the YUCCA family (Finet and Jaillais, 2012) (Fig 5).

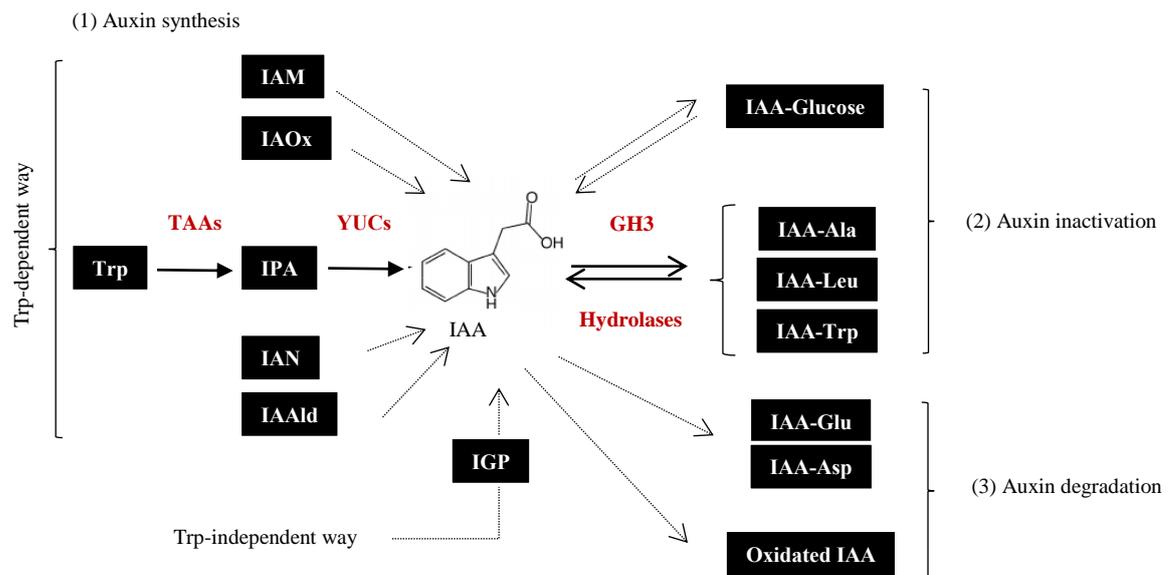
However, it has been observed that *Arabidopsis* mutants impaired in Trp synthesis present comparable levels of IAA to a wild-type plant, being this indicative of the existence of a Trp-independent IAA production. Although little is known about this other synthetic pathway it is thought to derive from another product of the shikimate route, indole-3-glycerol phosphate (IGP) (Korasick et al., 2013) (Fig 5).

Alterations in the different biosynthetic pathways have been observed in response to diverse environmental cues, indicating this that the redundant auxin pathways could collaborate to combinatorial responses for the integration of external signals (Ruiz Rosquete et al., 2012).

### **Auxin inactivation**

IAA levels can also be attenuated by its inactivation, which can happen by conjugation to other molecules (normally reversible) or through IAA catabolism (irreversible).

IAA conjugates are considered as temporary storage forms from which, by hydrolysis, active IAA can be recovered. Its importance has been proven in response to seed imbibition, auxin-mediated symbiosis, nodule organogenesis and auxin toxic levels. Although the specific role of these compounds is not clear yet they are thought to antagonize auxin effects competing with its receptors or transporters (Korasick et al., 2013; Ruiz Rosquete et al., 2012).



**Fig 5. Auxin synthetic and inactivation pathways.** Auxin can be synthesized in a Trp-dependent (from IAM, IAOx, IPA, IAN or IAAlD precursors) or a Trp-independent way (through IGP) (1). Auxin free levels can be regulated by the hormone conjugation to sugars or amino acids (2) or by its degradation (3). Continuous arrows indicate known steps and the enzymes that mediate them (in red). Discontinuous arrows show not yet characterized reactions. Double arrows indicate reversible steps (Adapted from (Ljung, 2013)).

IAA can be ester-linked to sugars or amide-linked to amino acids (low molecular weight conjugates) or to peptides or proteins (high molecular weight conjugates). Although the composition of IAA conjugates varies between different species, IAA amides are the most common forms of conjugation, more specifically IAA-Alanine (IAA-Ala), IAA-Leucine (IAA-Leu), IAA-Tryptophan (IAA-Trp), IAA-Aspartate (IAA-Asp) and IAA-Glutamate (IAA-Glu). These reactions are carried out by the Gretchen Hagen 3 (GH3) protein family and, except for IAA-Asp and IAA-Glu compounds, they can be reversed back by amidohydrolases. *GH3* genes are auxin-early response genes, indicating this another point for the feedback control of available auxin inside a cell (Finet and Jaillais, 2012; Korasick et al., 2013; Ruiz Rosquete et al., 2012) (Fig 5).

On the other hand, IAA and its conjugated forms IAA-Asp and IAA-Glu can be oxidized which leads to the irreversible degradation of the hormone. The molecular mechanisms behind these processes are not well described till now (Ruiz Rosquete et al., 2012) (Fig 5).

### Auxin transport

Same as it happens with nutrients and photosynthetic assimilates, plant hormones can be transported long distances in a fast-non-directional way along the phloem. However, auxin is the only

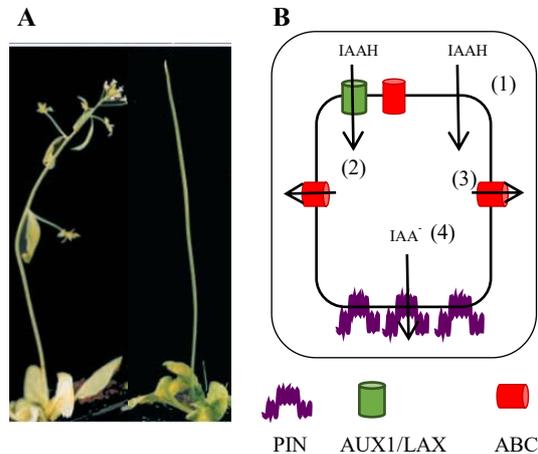
phytohormone known to also be transported short distances from cell to cell. This type of transport is known as polar transport as it happens in a directional way. It has been proven to be fundamental for several plant developmental processes such as apical-basal early embryo determination, hypocotyl elongation, lateral roots formation or flower development, processes all of them based in peaks of minimums and maximums of auxin levels (Adamowski and Friml, 2015). Indeed, alteration of polar auxin transport causes dramatic effects on plant development leading to the formation of naked shoots (Fig 6.A) (Vernoux et al., 2000).

Cell-to-cell auxin transport happens according to the chemiosmotic model: in the apoplast of the plant cell, where the pH is around 5, IAA will predominate in its protonated form (IAAH) that can diffuse across the cell wall, or can enter the cell by active transport (influx carriers). Once inside the cell, given the neutral pH of the cytosol, IAA will be found mostly as IAA<sup>-</sup> that can only exit the cell by efflux carriers (Fig 6.B).

There are four main classes of active auxin transporters in *A. thaliana*: auxin transporter protein 1 (AUX1) and auxin transporter-like protein (LAX) families of proteins constitute auxin influx carriers. On the other hand, auxin efflux carriers are members of the ATP-binding cassette (ABC) transporters family and the pin-formed proteins (PIN). Whereas ABC are uniformly localized in the plasma membrane, PIN proteins are polarly present in one side of the plant cell. This polar localization of PIN efflux carriers is what drives cell-to-cell auxin transport in a polar way (Grones and Friml, 2015) (Fig 6.B).

Out of the 8 PIN proteins found in *A. thaliana*, PIN5, 6 and 8 are localized in the endoplasmic reticulum (ER). This positioning is thought to contribute to auxin homeostasis by auxin compartmentalization. As for the rest of the PIN proteins, they can agglomerate in the basal or apical part of the cell in a dynamic manner: PIN proteins go through continuous endocytosis and recycling processes that will control the abundance of the protein in the plasma membrane and thus the auxin efflux. PIN maintenance in the plasma membrane is highly regulated. Post-translational modifications of PIN, such as phosphorylation or ubiquitination, environmental signals, microtubules orientation and other phytohormones (strigolactones, cytokynins, gibberelins and auxins themselves) will influence PIN presence in the plasma membrane (Adamowski and Friml, 2015; Habets and Offringa, 2014).

PINs are also regulated at the transcriptional level. Different expression patterns exist for each PIN gene within different tissues of the plant. Auxin is one of the signals that contribute to the transcriptional regulation of these proteins through Auxin Responsive Factors (ARFs) transcription factors (Habets and Offringa, 2014).



**Fig 6. Auxin polar transport.** A. Wt *A. thaliana* plant (left) vs *pin1* mutant (right) affected in auxin transport. Adapted from (Vernoux et al., 2000). B. Auxin transport scheme: in the apoplast IAA predominates as IAAH which can enter the cell by diffusion (1) or by active transport driven by influx carriers (2). Inside the cytosol, IAA is found in its deprotonated form, IAA<sup>-</sup>, and it will only exit the cell with the help of efflux carriers: ABC (3); uniformly distributed along the plasma membrane or PIN (4) polarized on one side of the cell.

### Auxin signalling

The auxin response inside the cell starts with the detection of the hormone by auxin receptors. Three possible auxin perception mechanisms have been described till now: the TIR1/AFB (Transport Inhibitor Resistant 1/Auxin Signalling F-box)-Aux/IAA co-receptor mechanism, the SKP2a (S-Phase Kinase-Associated Protein 2a) and ABP1 (Auxin Binding Protein 1).

- SKP2a

SKP2a involvement in auxin response has been recently proposed when auxin binding to this protein was found. SKP2a is a component of the SCF (Skp, Cullin, F-box) complex that participates in the G1/S checkpoint in cell cycle mediating the degradation of E2FC and DPB factors. This relieves repression of cell cycle control genes. It presents a F-box and a Leucine Rich Repeats (LRR) domain, which mediates binding to auxin (Jurado et al., 2010; Peer, 2013; Powers and Strader, 2016). By structure modelling with human SKP2 and further comparison of the modelled SKP2a with TIR1 structure (Fig 10.C), the SKP2a auxin binding pocket was identified. SKP2a mutants in this site were not able to induce E2FC and DPB degradation showing that auxin binding was necessary for SKP2a ubiquitin ligase role. SKP2a mutants were auxin-resistant and overexpression of the protein gave auxin-related phenotypes (Jurado et al., 2010; Powers and Strader, 2016), proving all this SKP2a implication in auxin response. SKP2a auxin perception could partially explain auxin roles in cell division.

- ABP1

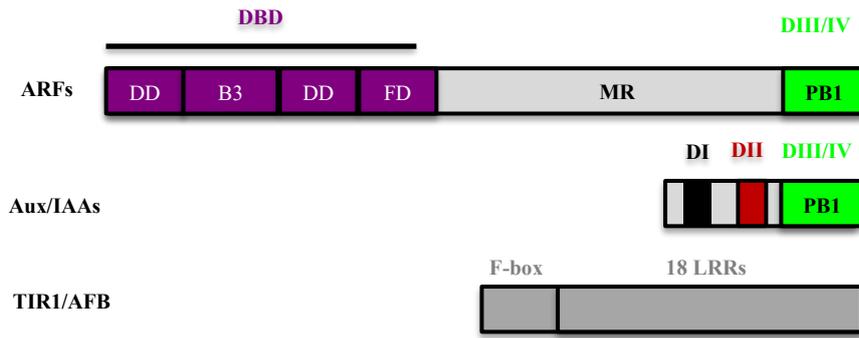
ABP1 auxin receptor role is a controversial matter in the auxin field nowadays. Even if it was the first auxin receptor to be discovered, its study became complicated due to the lack of auxin-related phenotypes associated to these receptors. In 2001, the first ABP1 mutant, *abp1-1*, was generated in *A. thaliana* and this allowed the generation of other mutant lines in downstream elements associated with this signalling cascade. However, recent genetic studies suggest that ABP1 mutations give no phenotype and that the already described ones are the consequence of off-targets mutations (Feng and Kim, 2015).

ABP1 protein is localized in the plasma membrane or in the ER. Although ABP1 has been proven to bind to auxin (Woo et al., 2002), the conditions in the ER are not favourable for the binding of the hormone, so the role of the ABP1 localized in the organelle is still to be studied.

On the other hand, ABP1 located in the plasma membrane has been associated mostly with the auxin non-transcriptional response, among others with the ROP-GTPase pathway that leads to clathrin-mediated endocytosis of PIN proteins (Peer, 2013).

- TIR1/AFB-Aux/IAA

Finally, the TIR1/AFB-Aux/IAA co-receptor mechanism has been extensively studied together with the transcription factors and the transcriptional response downstream auxin perception. All these proteins constitute the Nuclear Auxin Pathway (NAP). NAP reconstitution in a yeast system proved that only three components conform the core auxin response module: the TIR1/AFB and Aux/IAA co-receptors and the Auxin Response Factors, ARFs; and that these are enough for obtaining an auxin basic transcriptional response (Pierre-Jerome et al., 2014). However, the auxin transcriptional response in plants is everything but basic: different auxin inputs and gradients controlled by cell type, developmental stages and environmental cues give rise to a complex set of outputs in auxin response (Paque and Weijers, 2016; Pierre-Jerome et al., 2013). In the next lines, we will see how auxin is able to play its magic always through the same repertoire of proteins: TIR1/AFB, Aux/IAAs and ARFs (Fig 7). Due to the interactions established between them, it is difficult to talk about these proteins individually and thus we will treat them as proteinic modules.



**Fig 7. Domains organization of the protein families involved in the Nuclear Auxin Pathway.** DBD (DNA-binding domain); DD (Dimerization Domain); FD (Flanking Domain); MR (Middle Region); DIII/IV (Domains III/IV) also called PB1 (Phox/Bemp1); DI (Domain I, containing EAR-motifs); DII (Domain II, containing degron motif); LRR (Leucine Rich Repeat).

○ Module ARF-DNA

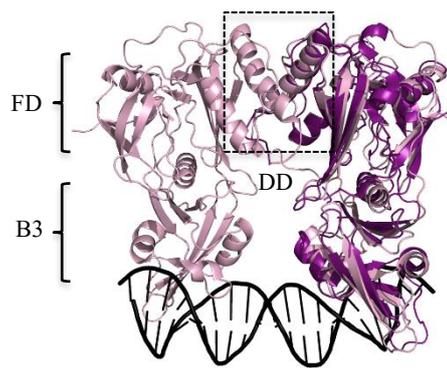
ARF proteins are transcription factors that bind to DNA sequences named Auxin Response Elements (AuxREs) located in the promoters of auxin responsive genes. Only in the model plant *A. thaliana*, 23 ARF proteins have been found. Out of the 23, just 5 of them (ARF5-8 and ARF19) are considered as ARF activators. The rest have been classified as repressors although there is not yet evidence of a repressor role for all of them (Peer, 2013).

ARFs-DNA interaction is mediated by a B3 domain located in their N-terminus. The B3 domain is flanked by regions that mediate dimerization of the protein (Dimerization Domains, DD), thanks to which ARFs bind to double AuxREs as dimers. In AtARF1 and AtARF5 DBD structures, a third domain, Flanking Domain (FD) was identified but its role remains unknown (Fig 7-8) (Boer et al., 2014)<sup>1</sup>.

○ Module ARF-Aux/IAA

On the other hand, in their C-terminal ARF proteins present a Phox/Bem1p (PB1) domain which is shared with Aux/IAAs (corresponding to domains III/IV, DIII/IV) (Fig 7). PB1s are protein-protein interaction domains that have been shown to mediate homo or hetero-oligomerization of ARFs or Aux/IAAs and ARFs-Aux/IAAs, respectively (Farcot et al., 2011; Guilfoyle and Hagen, 2012; Li et al., 2011; Piya et al., 2014; Trigg et al., 2017).

<sup>1</sup> ARFs-DNA interaction and the details of ARFs structure and domains organization will be properly explained in Chapter II of this manuscript.

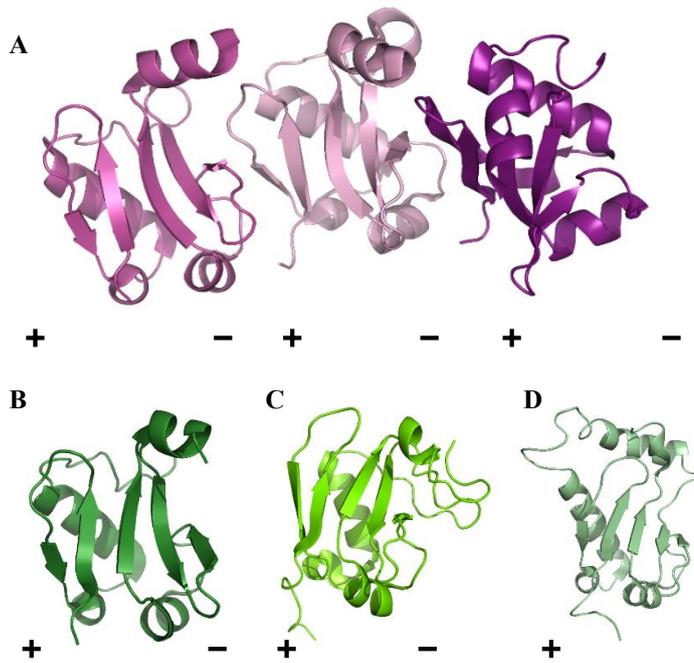


**Fig 8. ARFs DBD structure.** Structure of AtARF1 DBD (light pink, 4LDX) in complex with DNA (black) overlapped with ARF5 DBD structure (purple, 4LDU) showing the high similarity between them. FD, Flanking Domain; DD, Dimerization Domain. DNA-interactions are mediated by two loops within the B3 domain. Adapted from (Boer et al., 2014).

These interactions take place in a front-to-back manner due to the presence of two interaction interfaces in these domains: a negative and a positive interface, that allow chaining a high number of molecules by means of electrostatic interactions (Chandrasekaran et al., 2015; Kim et al., 2015; Korasick et al., 2014; Nanao et al., 2014; Parcy et al., 2016; Wang et al., 2015) (Fig 9). A high-throughput Yeast-2-hybrid (Y2H) experiment proved that most Aux/IAA proteins can homo-oligomerize or hetero-oligomerize mainly with ARF activators. ARF activators were also able to homo-oligomerize but ARFs repressors showed barely any interactions among themselves or with Aux/IAA proteins (Farcot et al., 2011).

The specificity and affinity of Aux/IAA-ARF and ARF-ARF interactions is related to the residues present in the acidic and basic interfaces of their PB1 domains (Parcy et al., 2016), meaning that certain combinations of PB1 domains are more favourable than others (Han et al., 2015). Also, post-translational modifications have been shown to affect these interactions (Dinesh et al., 2016; Hill, 2015; Vert et al., 2008; Wang and Estelle, 2014).

It is necessary to point out that ARF PB1 domains can also interact with other transcription factors such as MYB77, PIF4, bZIP or BZR1/2. These interactions might help the cross-talk between auxin and other signalling pathways (Dinesh et al., 2016).



**Fig 9. PB1 structures.** 4 PB1 domains structure have been determined till now. (A) AtARF5-PB1 oligomer (4CHK,(Nanao et al., 2014)) showing the alternance of positive and negative interfaces that mediate the front-to-back interactions. Each PB1 monomer shown in a different violet intensity. (B) AtARF7-PB1 (4NJ6, (Korasick et al., 2014)). (C) PsIAA4 (*Peasum sativum*) (2M1M, (Dinesh, D.C. et al, 2015)) . (D) AtIAA17 (2MUK, (Han et al., 2015)). Positive and negative interface indicated for each.

○ Module Aux/IAA-TIR1

Apart from the DIII/IV in Aux/IAAs, that mediates interactions with ARFs, in their N-terminal Aux/IAAs present two other domains I and II (DI and DII). DII, also called degron motif, mediates the interaction with TIR1/AFB auxin receptor.

TIR1/AFB are a family of proteins that belong to the E3 ubiquitin ligase systems. They present a F-box domain and a Leucine Rich Repeats domain (LRRs) composed by 18 LRRs (Fig 7). TIR1 was the first plant hormone receptor which structure was solved, helping this to the understanding of auxin perception: auxin binding takes places in an auxin binding pocket located within the structure formed by the LRRs. The hormone does not induce any conformational change in the TIR1/AFB receptors but it acts as a “molecular glue” between TIR1/AFB proteins and DII of Aux/IAAs (Tan et al., 2007) that leads to Aux/IAAs ubiquitination and further proteosomal-mediated degradation (Salehin et al., 2015) (Fig 10).

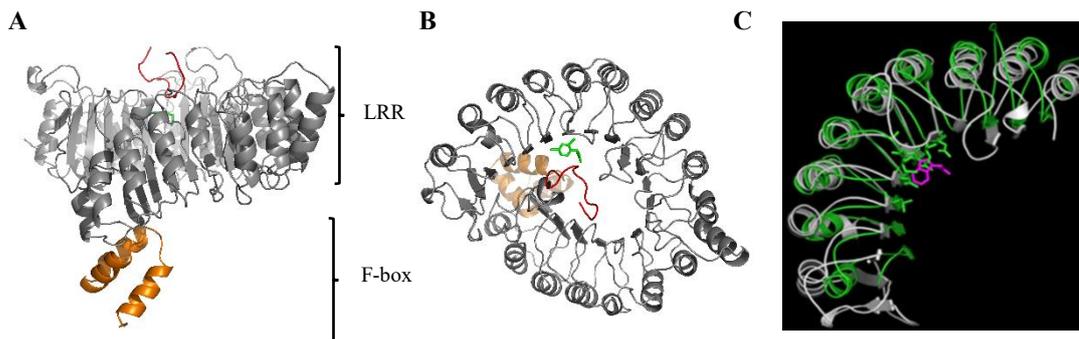
Recent studies have shown that TIR1 can also form oligomers and that the receptor oligomerization could help binding and poly-ubiquitination of Aux/IAAs (Dezfulian et al., 2016).

In *A. thaliana* 6 TIR1/AFB and 29 Aux/IAA proteins have been described. All of them have different patterns of gene expression in different plant tissues (Farcot et al., 2011; Wang and Estelle, 2014). In addition to this, TIR1/AFB-Aux/IAA combinations present

different affinities between themselves and for auxin hormone, which is translated into different degradation rates of Aux/IAA proteins, with half-lives that can range from 6 to 80 min in the presence of the hormone (Dinesh et al., 2016).

Although residues GWPPV within Aux/IAA-DII are the ones necessary for this interaction and further Aux/IAA degradation other parts inside the DII and in between DI and DII have been shown to influence the degradation rate of the protein, contributing this to the diversity of affinities (Calderón Villalobos, L; Lee, S; De Oliveira, C; Ivetac, A; Brandt, W; Armitage, L; Sheard, LB; Tan, X; Parry, G; Mao, H; Zheng, N; Napier, R; Kepinski, 2012).

Finally, post-translational modifications have been described for TIR1/AFB proteins that can modify their stability and interaction with Aux/IAA proteins. Altogether, transcriptional regulation, post-translational modifications and the diversity of TIR/AFB-Aux/IAA complexes contributes to the control and fine-tuning of auxin response (Wang and Estelle, 2014).



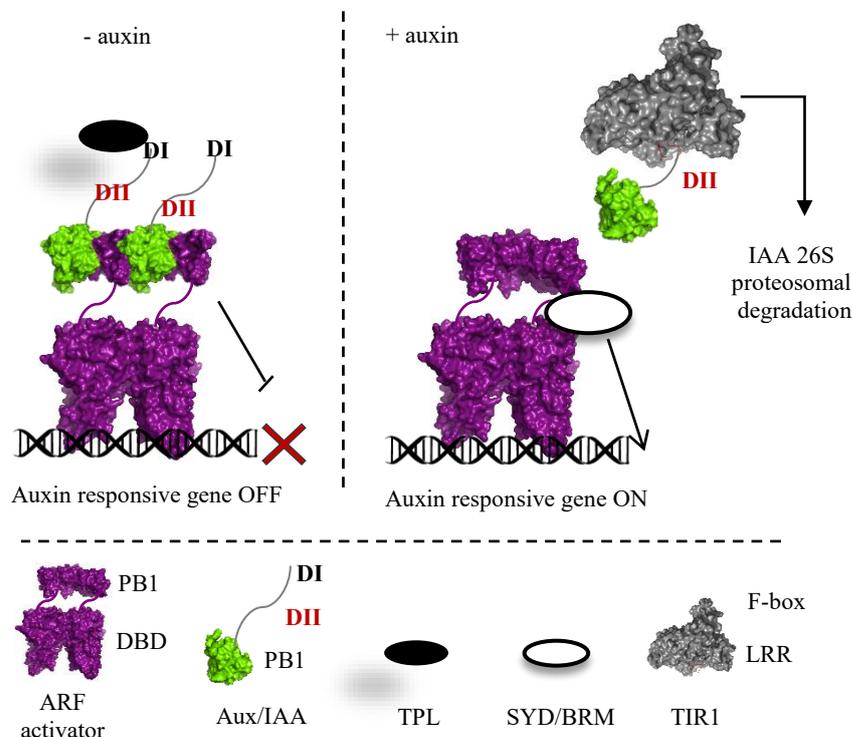
**Fig 10. TIR1 structure.** A-B. AtTIR1 structure in complex with IAA7-DII (red) and auxin (green) (2P1N). (B) Rotated 90° with respect to (A). LRR (grey) and F-box domains (orange) form a mushroom-shaped structure, with the LRR mediating the interaction with both the hormone and the degron in the same binding pocket. The structure was also obtained in complex with ASK1 (ubiquitin ligase system) and Ins6P, identified as a co-factor of unknown function (structure not shown here) (Tan et al., 2007). (C) Structure model of SKP2a (green) aligned to LRR domain of TIR1 that allowed SKP2 binding site for auxin (purple) (Adapted from (Jurado et al., 2010)).

The combinatorial action of all these modules and the diversity of combinations within each, allows the repression or activation of auxin transcriptional response in a fine-tuned and spatial-temporal way. We must add a fourth component to the equation, though: the co-regulators. Although not specific to auxin signalling pathway, and thus not included into the so-called NAP, the co-repressors

TOPLESS (TPL) and the chromatin remodelling ATPases SPLAYED (SYD) and BRAHMA (BRM) collaborate to the repression or activation of the signal<sup>2</sup>.

Putting all the pieces together they create the following, precise, and stylish puzzle (Fig 11): under low auxin concentrations, ARF activators are blocked by Aux/IAA repressor proteins that bind to them through the DIII/IV present in both proteins. At the same time, the Aux/IAA domain I (DI) can recruit TOPLESS co-repressor (TPL) that will in turn recruit chromatin modifier enzymes, forming a repressor complex that will avoid the transcription of auxin responsive genes through chromatin compaction. However, when auxin levels increase in the plant cell Aux/IAA proteins will bind auxin together with TIR1/AFB which leads to Aux/IAA degradation and thus all the co-repressor protein complex will be taken away from the DNA. This leaves ARF activators free to do their transcriptional activator role, for which they will recruit chromatin remodelling proteins (such SYD and BRM) that will help opening chromatin so that transcription of auxin responsive genes can take place (Lavy and Estelle, 2016; Wu et al., 2015b).

This mechanism is relatively well described for ARF activators. However, ARF repressors are much more of a mystery. Most of them are unable to interact with Aux/IAA repressors (Farcot et al., 2011) and therefore they are thought to work in an auxin-independent manner. Three possible mechanisms are proposed for ARF repressors working mode: “kidnapping” of ARF activators by binding to them, competition with ARF activators for AuxREs and/or TOPLESS recruitment (Chandler, 2016).



<sup>2</sup> TPL involvement in auxin signalling will be further discussed in Chapter I of this manuscript

**Fig 11. Nuclear Auxin Pathway.** In the absence of auxin (left), Aux/IAA repressors bind to ARF activators through their PB1 domains and recruit TPL through its DI preventing auxin responsive genes transcription. In the presence of auxin (right) a TIR1/AFB-Aux/IAA auxin complex is formed and triggers the degradation of Aux/IAAs and to the release of ARF activators, now free to recruit SYD and BRM chromatin remodellers, leading to auxin responsive genes expression. The domains or proteins pictured as structures represent real structures already determined for some of these proteins (Adapted from (Nanao et al., 2014)).

## Objectives

This manuscript has a focus on the structural and biochemical mechanisms that lie beyond the proteins involved in the NAP. Along the following chapters, we will try first, to unravel the structural secrets of TPL co-repressor function and its interaction with Aux/IAA repressor proteins. We will immerse ourselves into the details of ARFs-DNA binding mysteries. And last but not least, we will try to answer to a big intriguing question: where did all this come from?

### **Chapter I: Function, structure and evolution of TOPLESS co-repressor**

TOPLESS is a co-repressor protein in plants able to interact with a huge diversity of plant repressors. These interactions are mediated through conserved sequences of amino acids present in all these repressors, the Ethylene-responsive element binding factor-associated Amphiphilic Repression (EAR) motifs.

The first chapter of this manuscript aims at answering:

- how TOPLESS interacts with such a diversity of transcriptional repressors
- how TOPLESS-Aux/IAA interactions happen and how they affect the repression of auxin responsive genes

For this, we followed a “from structure to function” strategy mainly based on *in vitro* studies comprising a wide variety of protein-protein interaction techniques.

### **Chapter II: ARFs-DNA preferences, players of the specificity of auxin transcriptional response**

ARF transcription factors have been classically classified into activators or repressors. Either one or the other they all present DNA binding domains (DBD) that mediate ARFs interaction with Auxin Response Elements, AuxREs. Surprisingly, the residues mediating DNA-interactions are highly conserved among activators and repressors and they have both been shown to bind to the same nucleotide sequences. AuxREs are normally formed by two repeats of ARF DNA binding. Depending on their orientation and spacing different types of AuxREs have been described in literature.

In Chapter II we tried to go in depth into the specificity of the different classes of ARF proteins for the different types of AuxRE repeats. To do this we used an integrated strategy of bioinformatic analysis and *in vitro* studies of protein-DNA interactions.

### **Chapter III: Back to the water looking for ARF ancestors**

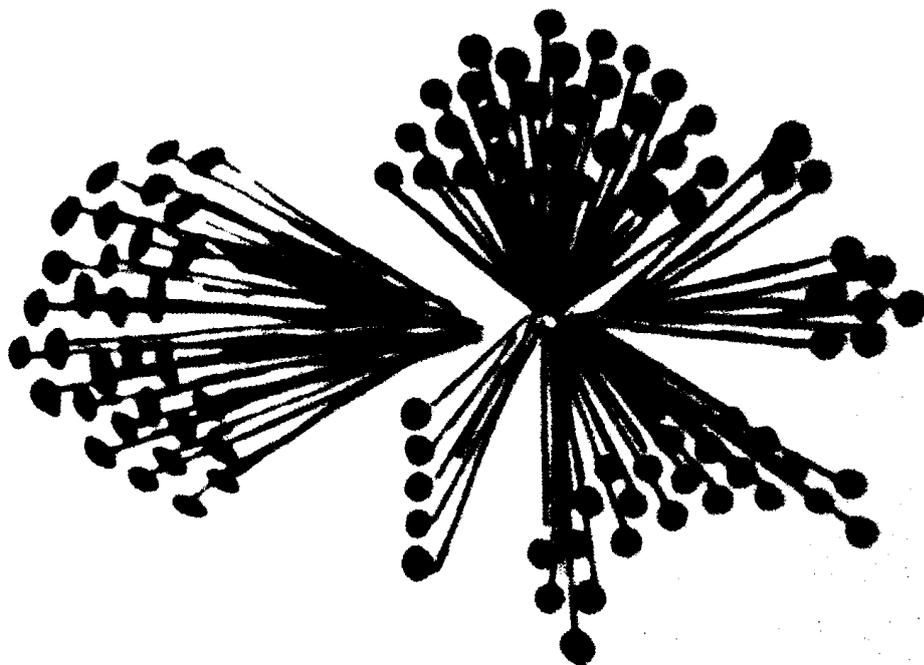
Land plants evolution from algae happened hand in hand with the adaptation to terrestrial stresses and the development of a certain body complexity. Given the importance of auxin hormone in plant growth and architecture, the acquisition of the auxin signalling pathway must have been a breaking point in the transition from water to terrestrial life. However, not much is known about how and when this happened. ARF proteins have been traced back till bryophyte organisms but till now no sign of them has been found in charophyte algae, the proposed ancestors of land plants.

Swimming among the algae, we found by sequences homology and protein structure modelling two possible candidates of ARF ancestors in the charophyte algae *Klebsormidium nitens* and *Chlorokybus atmophyticus*. Both proteins presented a predicted DBD and a PB1 domain with high similarities to those found in plant ARF proteins. Furthermore, as it is the case for some ARF repressors in between both predicted domains we found potential motifs that might recruit TPL.

Along the third and final chapter of this manuscript I will describe the structural and biochemical characterization of these different domains in terms of their DNA binding specificity, oligomerization potential and TPL interactions.

# Chapter I

Function, structure and evolution of  
TOPLESS co-repressor

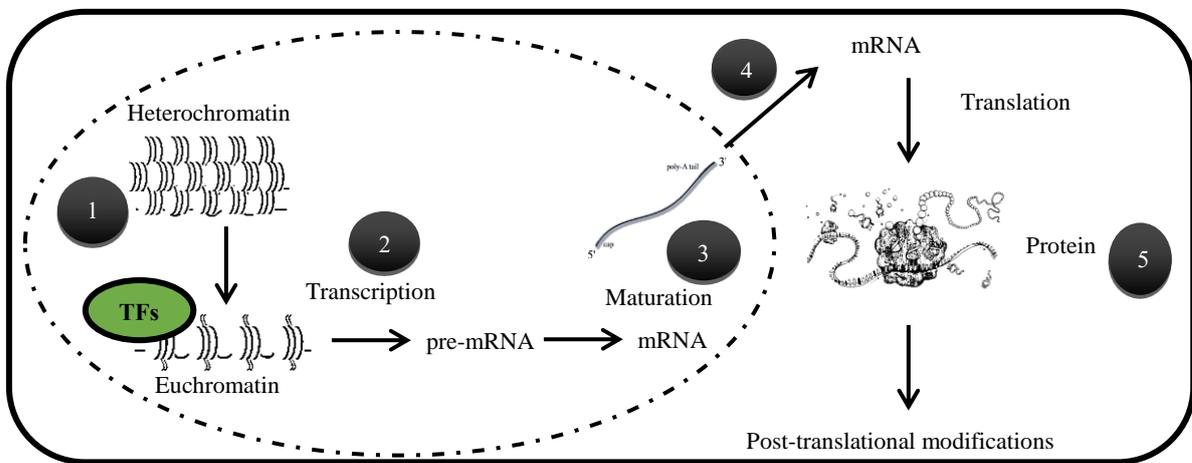




## Introduction

Once upon a time (1958) Francis Crick proposed what till now we know as the “Central Dogma of Biology” according to which, the genetic information is stored inside cells as deoxyribonucleic acid, DNA, that will be transcribed into messenger ribonucleic acid, mRNA, which in turn will be translated into proteins, the elements that will in the end, develop a specific function inside a cell.

Decades after Crick’s proposal we can affirm that the transition from DNA to proteins is not that elemental -dear Watson- with each step of the process being susceptible to different levels of regulation (Fig 12).



**Fig 12. Regulation levels from chromatin to proteins.** (1) Chromatin compaction state can allow or avoid the access of proteins (transcription factors or transcriptional machinery). (2) Euchromatin is a transcriptionally active state in which transcription will be controlled by regulatory sequences in the promoters of genes and the proteins bound to them. (3) Eukaryotic organisms have developed a machinery of processing and splicing of the transcripts that can influence their stability or generate different combinations of exons. (4) Exit of the mature mRNA from the nucleus can control its availability in the cytosol for its translation into proteins. (5) Once the mRNA is translated into proteins the function or availability of these proteins can be also modulated by its tertiary or quaternary structure or by post-translational modifications (Adapted from (Alonso and Wilkins, 2005)).

### Gene expression control in eukaryotic organisms

Gene expression is fundamental in every living organism and thus, it is one of the key regulated points in all this process. Eukaryotic organisms have developed more evolved and complex regulatory systems of gene expression in comparison with prokaryotes.

Bacteria’s DNA is “naked” and “free” inside the cell (Payankaulam et al., 2010). On the contrary, eukaryotes conceal their genetic information inside the nucleus. Thanks to this, transcription

(nuclear) and translation (cytosolic) are physically separated. Therefore, the processing and transport of mRNA from the nucleus to the cytosol adds one level of control with respect to bacteria (Fig 12).

On the other hand, eukaryotic DNA is wrapped around nucleosomes -octamers of histone proteins - that form a structure named chromatin (Liu et al., 2014; Payankaulam et al., 2010) (Fig 12). Chromatin can be in two different forms: euchromatin, that is lightly packed and transcriptionally active and heterochromatin, with a high level of compaction in which transcription cannot happen. Dynamic marks in the DNA or in histone tails influence chromatin packing level. DNA methylation and histones deacetylation are considered as repressive marks whereas histones acetylation is thought to cause relaxation of the chromatin structure. Several families of enzymes have been involved in the deposition or removal of the different marks, including DNA Methyltransferases (DNA MTase), Histone Deacetylases (HDA) or Histone Acetyltransferases (HAT) (Liu et al., 2014; Xiao et al., 2017). Other types of enzymatic histone modifications, such as ubiquitination, methylation or phosphorylation have been described (Liu et al., 2014; Perissi et al., 2010; Xiao et al., 2017). Chromatin state -and all the proteins involved in its dynamics- is consequently another point of control since it constitutes a physical barrier for the access of transcriptional regulatory proteins.

Among these regulatory proteins, transcription factors, either activator or repressors, have the ability to bind to specific DNA sequences in the regulatory regions within genes and from there they will contribute to the activation or repression of these genes.

Transcription factors are present in both prokaryote and eukaryote organisms. In the formers, the regulation of transcription by transcription factors is “direct”: they interact with the basal transcriptional machinery. All over again, eukaryotes went a bit further and they evolved “intermediate proteins”, named coregulators (Payankaulam et al., 2010).

### **Co-repressors**

Co-regulators, either co-activators or co-repressors, are proteins that bridge transcription factors to chromatin modifying enzymes, allowing the formation of a coregulator complex with enzymatic activity that will alter chromatin conformation in specific regions in the DNA (Mottis et al., 2013).

They do not have DNA binding capacity themselves. Instead, they are recruited to specific regions in the genome by transcription factors that do have DNA specificity. One single co-regulator can be recruited to the chromatin environment by many different transcription factors helping to the regulation of a wide set of genes (Watson et al., 2012).

These two facts (co-regulators binding to a diversity of transcription factors and to chromatin remodeller enzymes) bring us to consider co-regulators as “hub” or “scaffolding” proteins that integrate in one single flexible complex diverse protein functions (Watson et al., 2012). Transcriptional repressors present repressor domains that mediate the recruitment of co-repressor proteins to specific regulatory sequences within genes (Kagale et al., 2010a). There, co-repressors will inhibit the expression of these genes by inducing chromatin compaction. Several mechanisms have been described that aid explaining how chromatin closing is triggered. Recruitment of HDA or other chromatin modifying enzymes, direct or indirect -through the Mediator complex- interactions with the transcriptional machinery or blocking of the activation sites of transcription factors are some of them (Agarwal and Mathew, 2015).

On the other hand, co-repressors disposition on chromatin is also thought to be determinant for chromatin compaction. Co-repressors have been classified into long-range and short-range (Payankulam et al., 2010). By “spreading” along chromatin, long-range co-repressors are suggested to help the deposition of repressor marks along long regions of the DNA. On the contrary, short-range co-repressors are locally positioned inducing the formation of DNA loops that antagonize chromatin opening (Sekiya and Zaret, 2007).

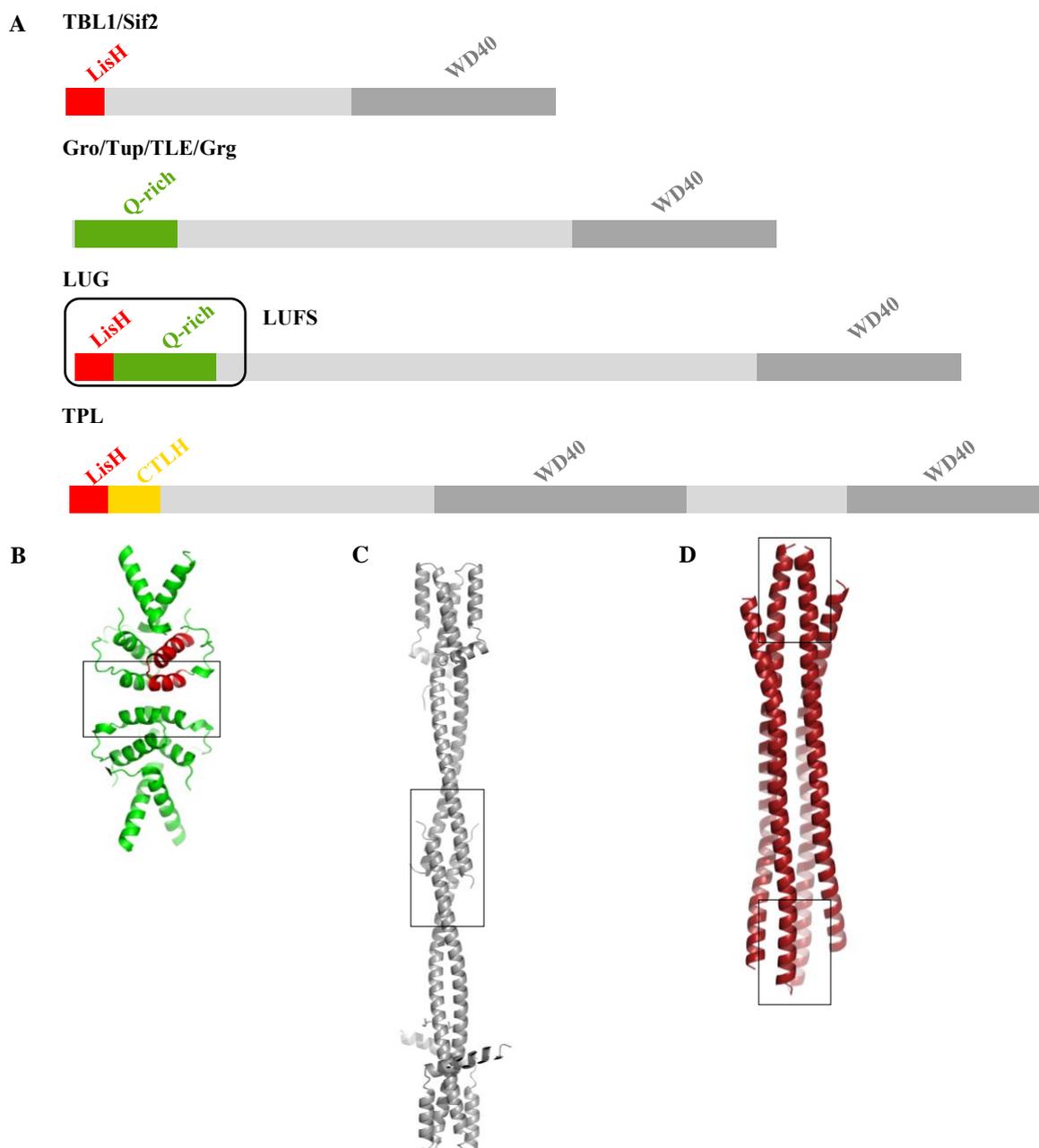
Co-repressor examples have been described in a diversity of eukaryotic organisms. Among the most studied ones are the Gro (Groucho) family in *Drosophila* and the SMRT (Silencing Mediator of Retinoic Acid and Thyroid hormone receptor)/NCoR (Nuclear Receptor Co-repressor) family in humans. Orthologues of these proteins have been found in lower organisms indicating their importance along evolution (Mottis et al., 2013; Watson et al., 2012).

- **SMRT/NCoR family**

SMRT/NCoR co-repressors are associated with nuclear receptors to which they will bind in the absence of ligand. Upon ligand interaction, nuclear receptors will dissociate from co-repressors and will bind coactivators instead. SMRT/NCoR complexes have been implicated in diseases such as cancer or diabetes (Mottis et al., 2013). 10-12 proteins were found to be part of the human SMRT/NCoR complex (Yoon et al., 2003), among them HDA3, GPS2, TBL1 and obviously, SMRT/NCoR. TBL1 “hubs” this complex by establishing interactions with both GPS2 and SMRT/NCoR. TBL1 presents a Lissencephaly (LisH) domain in its N-terminal region and WD40 domains in its C-terminal (Fig 13.A). Structural determination of TBL1<sub>1-90</sub> N-ter revealed that TBL1 is a tetrameric protein with the tetramer established as a dimer of dimers. The LisH domain and a third expanded helix mediate both dimerization and tetramerization and they generate in between a hydrophobic groove that mediates interaction with GPS2 and SMRT/NCoR proteins (Fig 13.B)

(Oberoi et al., 2011). The LisH domain has been proven to be fundamental for TBL1 binding to histones and for repression (Choi et al., 2008; Yoon et al., 2005, 2003).

SET3 complex is thought to be the orthologue of the SMRT/NCoR complex in yeast. Among the proteins integrating SET3, Sif2 is the equivalent to TBL1, presenting a LisH N-terminal domain and WD40 C-terminal repeats (Mottis et al., 2013; Yoon et al., 2005) (Fig13. A). Same as for TBL1, Sif2 LisH domain mediates its tetramerization (Cerna and Wilson, 2005). The function of the WD40 in these proteins remains unknown.



**Fig 13. Structural features of co-repressors.** **A.** Domains distribution of co-repressor proteins from diverse organisms: TBL1 (human)/Sif2 (yeast) of the SMRT/NCOR co-repressor complex, Gro (Drosophila)/Tup1 (Saccharomyces)/TLE (*Homo sapiens*)/Grg (*Mus musculus*), LUG (*A. thaliana*), TPL (*A. thaliana*) (adapted from (Liu and Karmarkar, 2008)). **B.** Structure of the human TBL1 tetramer (2XTC) with LisH domain in red in one monomer (Oberoi et al., 2011). **C.** Structure of the human TLE N-terminal tetramer (4OM3) (Chodaparambil et al., 2014). **D.** Structure of the yeast Tup1p (3VP8) N-terminal tetramer (Matsumura et al., 2012). Tetramerization interfaces framed inside a black rectangle.

- **Groucho family**

Groucho, present in Drosophila, was the first co-repressor protein of this family to be identified. Gro mutation resulted in clumps of extra bristles over the eyes of the poor flies that after this, had an apparently striking resemblance with the famous American humourist Groucho Marx. I bet he did not find this funny (Fig 14.A).

Nowadays, orthologues of Gro have been described in other organisms: Tup1 in yeast, Grg in mouse or TLE in humans. All these proteins present a glutamine (Q)-rich region in their N-ter and a WD40 repeats domain in their C-ter separated by a non-conserved glycine (G) and proline (P)-rich region (Liu and Karmarkar, 2008) (Fig 13.A). Structural studies have proven that Gro family members are tetrameric proteins which tetramerization is mediated by their N-ter. The tetramers are formed as dimers of dimers (Chodaparambil et al., 2014; Matsumura et al., 2012) (Fig 13.C-D) and have been shown to be needed for binding to histones and for repression (Chodaparambil et al., 2014; Flores-Saaib and Courey, 1999; Sekiya and Zaret, 2007). As for the WD40 in the C-terminal regions of Gro proteins, they have been implicated in interactions with transcription factors and in chromatin condensation (Jennings et al., 2006; Sekiya and Zaret, 2007).

### Co-repression in plants

Plants are not less than flies or humans and so, they have also evolved stylish transcriptional repressor mechanisms that involve the action of co-repressor proteins. LEUNIG/LEUNING-HOMOLOGUE (LUG/LUH) and TOPLESS/TOPLESS-related (TPL/TPRs) are the two most studied families of plant co-repressors (Lee and Golz, 2012; Liu and Karmarkar, 2008).

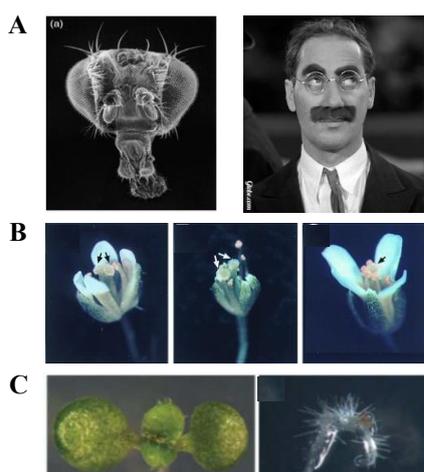
LUG was the first co-repressor to be identified in plants. *lug* mutants presented floral defects associated with an ectopic expression of the floral regulator AGAMOUS (AG) (Conner and Liu, 2000) (Fig 14.B). Further biochemical studies revealed that LUG protein presented a LisH and a Q-rich region in its N-ter and WD40 domains in its C-terminal region (Lee and Golz, 2012; Liu and Karmarkar, 2008). The LisH domain is enclosed into the so-called LUFS domain, responsible for mediating interactions with transcription factors through the intermediate protein SEUSS (SEU) (Sridhar et al., 2004, 2006) (Fig 13.A). LUG repressor capacity has been shown to be dependent on

HDA activity and on components from the Mediator complex, that directly interact with the transcriptional machinery (Gonzalez et al., 2007; Sridhar et al., 2004).

TPL, on the other hand, was discovered thanks to *A. thaliana* *tpl-1* mutants affected in the polarity of the embryo. *tpl-1* mutants were temperature-sensitive, so the mutation originated a diversity of phenotypes being the most striking one the replacement of an apical shoot by an apical root when plants were grown at 29°C (Fig 14.C) (Long et al., 2002). Therefore, these seedlings had no “top”: they were topless. *tpl-1* phenotypes were caused by a N176H substitution in TPL, a protein similar to already known transcriptional co-repressors: TPL had in its N-terminal region a LisH domain followed by a C-terminal to LisH (CTLH) domain and two sets of WD40 repeats in its C-terminus, separated by a non-conserved proline (P)-rich region (Long, 2006) (Fig13. A), a domains organization alike the previously presented Gro family and TBL1.

5 TPL/TPRs genes have been found in *A. thaliana*. Knocking them out individually produces no evident effects in *A. thaliana* plants. Only a quintuple TPL/TPR mutation can phenocopy *tpl-1* striking features, which evidences the redundancy among the different TPL/TPRs (Long, 2006). *tpl-1* mutant is thus defined as a negative dominant.

Expression analysis of markers associated with apical and basal fates within the embryo were found to be altered in *tpl-1* (Long et al., 2002) and several transition repression assays proved that the *TPL-1* protein had a weakened repressor capacity (Ryu et al., 2014; Szemenyei et al., 2008). Despite these studies the *tpl-1* oddity still supposes a puzzle for the plant physiology community.



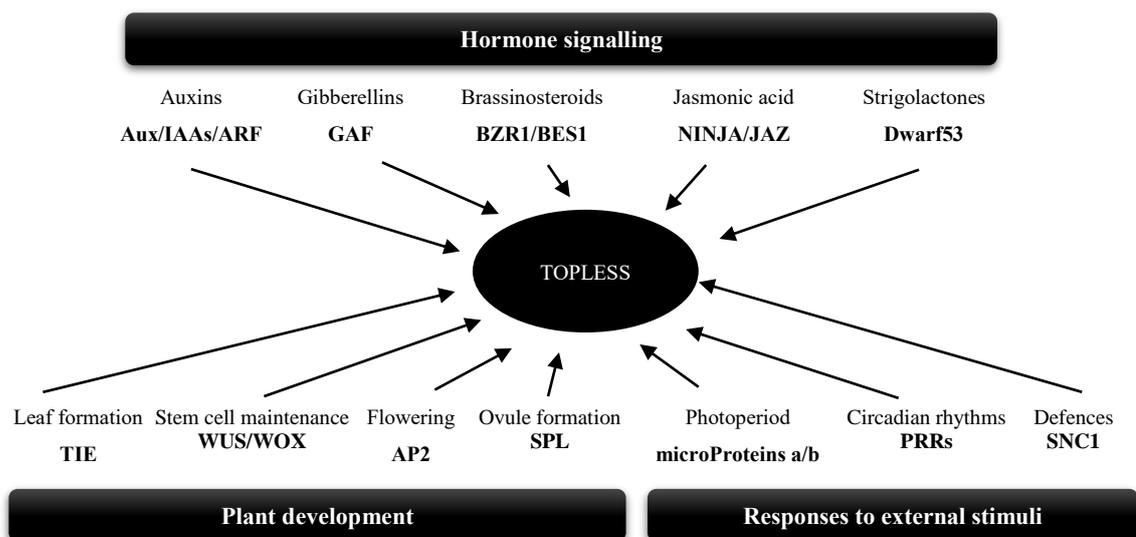
**Fig 14. The monsters of co-repression.** **A.** *groucho* Drosophila mutants (left) exhibit abnormal clumps of bristles over the eyes that researchers found similar to the eyebrows of the famous comedian Groucho Marx (right) (Parkhurst, 1998). **B.** *lug* mutants (left and middle) that show abnormal flower development and organ determination in comparison with the rescued phenotype that looks like a wild type flower (right) (Conner and Liu, 2000). **C.** Wild type *A. thaliana* seedling (left) vs *tpl-1* mutant showing a “transformation” phenotype of an apical shoot into a root (Long, 2006).

### TOPLESS co-repressor mechanism

By definition, TPL, as any co-repressor, is supposed to be able to interact with a diversity of transcriptional repressors. And indeed, it is. Since its discovery and definition as a co-repressor numerous studies have demonstrated the importance of this protein in a variety of plant processes, including hormone signalling (Causier et al., 2012a; Espinosa-Ruiz et al., 2017; Fukazawa et al., 2014; Hao et al., 2014; Jiang et al., 2013; Oh et al., 2014a; Pauwels et al., 2010; Ryu et al., 2014; Szemenyei et al., 2008; Tiwari, 2004; Wang et al., 2013; Wu et al., 2015b), stem cell maintenance (Busch et al., 2010; Ikeda et al., 2009; Kieffer et al., 2006; Pi et al., 2015; Zhang et al., 2014), plant development and defences (Graeff et al., 2016; Krogan et al., 2012a; Long, 2006; Long et al., 2002; Tao et al., 2013a; Wei et al., 2015; Zhu et al., 2010) or circadian rhythms (Wang et al., 2013) due to its interactions with repressors implicated in the cited processes (Fig 15).

TPL interacting partners present repressor domains, named ERF-associated amphiphilic repression (EAR) motifs, that mediate its recruitment.

EAR motifs were first discovered as plant repression domains present in ethylene-responsive factors (ERF) and in certain zinc finger proteins (ZAT). Alignment of the EAR motifs present in these proteins led to the definition of the consensus EAR motif (L/F)DLN(L/F)xP, in which hydrophilic and hydrophobic amino acids alternate. Fusion of EAR motifs to transcriptional activators turned them into repressors, whereas mutations in the leucine residues within the EAR motif abolished the repressor capacity of ERF and ZAT proteins, proving all this the repressor domain function of this conserved amino acidic sequence (Ohta, 2001).



**Fig 15. TPL interactome.** Represented some of the described TPL-interacting partners and the processes they participate in.

Further studies demonstrated that just 5 or 6 residues with a LxLxL consensus sequence were enough, when fused to a DNA-binding domain, to confer a repressor activity to a protein. Although mutations of the leucine residues completely abolished repression, the amino acids in between were also proven to have certain influence in the grade of repression (Hiratsu et al., 2004).

(L/F)DLN(L/F)xP EAR motifs have been found in numerous *A. thaliana* proteins belonging to 18 different families implicated in a variety of processes, some of them with an already described repressor function. Among them, ABI3, WOX, MADS, BZR, Aux/IAA or JAZ proteins (Kagale et al., 2010b).

The analysis of TPL interactome by a huge Y2H screening revealed the presence of other types of repressor motifs, EAR-like motifs, also able to mediate the interaction with TPL. This study allowed to add (K/M/R)LFGV and TLxLFP (also known as WUSCHEL box, WUS box) EAR-like motifs to the already known LxLxL and DLNxxP. Furthermore, it exposed that some proteins can present not just one, but up to 3 EAR motifs of the same or different categories (Causier et al., 2012b). One example of this is the WUS/WOX family, responsible for the maintenance of stem cells in the shoot and root meristems. WUS/WOX present in their C-ter a WUS box and a LxLxL-type EAR motif that have been both implicated in the interaction with TPL and in the repression capacity of the protein (Busch et al., 2010; Ikeda et al., 2009; Pi et al., 2015). Thus, the presence of multiple EAR motifs in transcriptional repressors could have synergetic functions in TPL recruitment and in gene expression inhibition.

In some cases, TPL recruitment is not mediated by the transcriptional repressor factor itself but by intermediate proteins – that also acts as transcriptional repressors but that lack a DBD- that bridge TPL to the transcription factor bound to the DNA. This is the case of Aux/IAA, NINJA, TIE or SPL transcription factors implicated in auxin signalling, jasmonic acid signalling, leaf development and ovule development, respectively. These proteins use different domains to interact with the transcription factor on one side and with TPL on the other (Pauwels et al., 2010; Szemenyei et al., 2008; Tao et al., 2013; Wei et al., 2015).

Recently, microProteins have been added to this list. MicroProteins1a and b (miP1a/b) were shown to bridge the transcription factor CONSTANS to TPL for an appropriate flowering response to photoperiod. The discovery of these proteins as TPL partners also allowed the identification of a new EAR-like motif present in them, (P/F)VL(F/L), which deletion abolished the interaction with TPL (Graeff et al., 2016).

### **TPL in the specific context of auxin signalling**

TPL interactions with several proteins of the Aux/IAA family have been demonstrated by several studies (Causier et al., 2012a, 2012b; Kagale et al., 2010b; Szemenyei et al., 2008). These interactions are known to happen through an EAR-motif type LxLxL present in the Domain I (DI) of Aux/IAAs. Except for Aux/IAA33, all *A. thaliana* Aux/IAA proteins have this motif. Aux/IAA18, 26 and 28 present a small variation of the EAR peptide (LxLxPP) also found in the Aux/IAAs and some ARF proteins from the moss *Physcomitrella patens*. Mutations of LxLxPP motif in these proteins abolished the interaction with TPL (Causier et al., 2012a).

Apart from Aux/IAAs proteins, some ARFs have also been found to interact with TPL (Causier et al., 2012a, 2012b). The TPL interactome analysis in *A. thaliana* revealed that ARF2, ARF9 and ARF18 proteins -all classed as repressors- can interact with TPL and they present a (K/M/R)LFGV EAR-like motif. Apart from this peptide, a LxLxL-type EAR motif was also described in ARF2 (Causier et al., 2012b). Whether ARFs interaction with TPL is necessary for auxin transcriptional repression or not remains an enigma (Chandler, 2016).

Whilst several studies have been done about TPL and its interactions with diverse transcriptional repressors that prove TPL involvement and importance in a variety of plant processes, not much is known about how TPL mediates chromatin compaction.

HDA6 and 19 have been frequently associated with TPL (Liu et al., 2014). Although a direct interaction between TPL and HDA has never been found, their importance in TPL-mediated repression has been demonstrated. *hda* mutants have similar phenotypes to *tpl-1* or other TPL/TPR mutants (Long, 2006; Pi et al., 2015; Zhu et al., 2010) and ChIP-seq results prove that HDA and TPL co-localize in the promoters of the genes they are repressing (Krogan et al., 2012; Pi et al., 2015; Ryu et al., 2014; Wu et al., 2015). Understanding how HDA recruitment by TPL is made and the intermediate/s carrying it out remains to be answered.

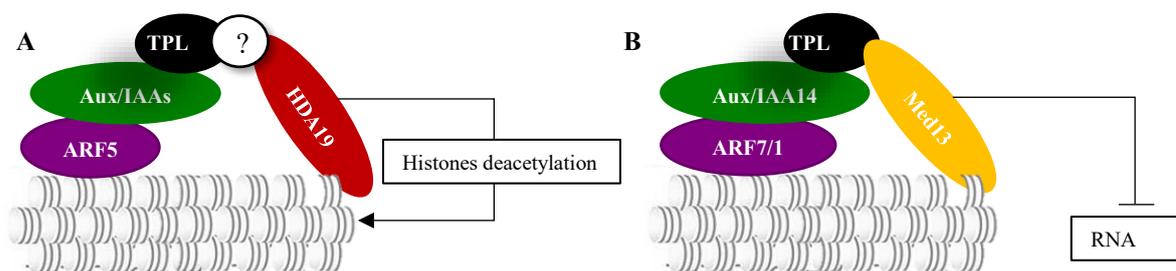
Within auxin signalling, the hormone is thought to trigger a quick chromatin switch that allows the expression of auxin responsive genes. As previously mentioned (Fig 11), under low auxin concentrations auxin transcriptional response is inhibited by Aux/IAA proteins that block ARF activators and that recruit TPL. In agreement with this, TPL and HDA19 have been found to colocalize with the ARF activator, ARF5 (MONOPTEROS, MP), in the promoters of auxin-responsive genes. Auxin treatment led to Aux/IAA degradation and consequent disappearance of the co-repressor complex in these regions of the genome. The removal of the co-repressor complex allows the recruitment of chromatin remodellers by ARF5, that help opening chromatin and lead to auxin-responsive genes transcription. Due to the fast degradation of Aux/IAAs under auxin

presence, this closed-to-opened switch in chromatin is thought to happen rapidly allowing a quick transcriptional response (Wu et al., 2015).

Recently, a similar “quick switch” mechanism has been described involving this time the Mediator complex. The Mediator complex has 4 modules (head, middle, tail and dissociable CDK8 Kinase module, CKM) with several components each. It affects transcription by direct interactions with the RNAPol II that lead to activation or repression of transcription depending on the components of the complex that are intervening. ARF7/19 activators interact with MED25 and 8 (tail and head modules, respectively) whereas Aux/IAA14 was found to be associated with MED13 (CKM module) through TPL, that directly interacts with it. Auxin treatment led to the dissociation of the IAA14-MED13-TPL from auxin-responsive genes whereas ARF7/19-MED25 complex remained stable. CKM module is thought to function as a blocker of the interaction with RNAPol II. Thus, in the absence of the hormone the Aux/IAA-TPL-CKM Mediator module could prevent the access of the transcriptional machinery, whereas auxin presence would induce Aux/IAA degradation and the consequent removal of the whole co-repressor complex in a fast way. RNAPol II could now interact with the tail and head components of the Mediator complex associated to ARF activators inducing the transcription of auxin-responsive genes (Ito et al., 2016).

In summary, these results indicate two possible mechanisms for TPL-mediated transcriptional repression: HDA recruitment by still unknown intermediates and RNAPol II blocking through the Mediator complex. These two mechanisms have also been described for LUG (Gonzalez et al., 2007).

Finally, TPL interactome reveals the possible implication of other chromatin remodellers that could help TPL in its co-repressor role. SDG19, SUVH3 and PKR1, related to histone methyltransferases were found as TPL interactors (Causier et al., 2012b), suggesting that TPL could use different chromatin remodeller partners to exercise chromatin compaction and transcriptional repression.



**Fig 16. TPL co-repressor mechanisms in auxin signalling.** TPL represses transcription of auxin responsive genes by the recruitment of HDA19 that induces deacetylates histone tails leading to chromatin compaction (A, adapted from (Wu et al., 2015b)) or blocking the access of RNA polymerase due to interactions with components of the Mediator complex (B, adapted from (Ito et al., 2016)).



# Structure of the *Arabidopsis* TOPLESS corepressor provides insight into the evolution of transcriptional repression

Raquel Martin-Arevalillo<sup>a</sup>, Max H. Nanao<sup>b,c,1</sup>, Antoine Larrieu<sup>d</sup>, Thomas Vinos-Poyo<sup>a</sup>, David Mast<sup>a,d</sup>, Carlos Galvan-Ampudia<sup>d</sup>, Géraldine Brunoud<sup>d</sup>, Teva Vernoux<sup>d,1</sup>, Renaud Dumas<sup>a,1</sup>, and François Parcy<sup>a</sup>

<sup>a</sup>Laboratoire de Physiologie Cellulaire et Végétale, Université Grenoble Alpes, CNRS, Commissariat à l'Energie Atomique et aux Energies Alternatives/Biosciences and Biotechnology Institute of Grenoble, Institut National de la Recherche Agronomique (INRA), F-38000 Grenoble, France; <sup>b</sup>Structural Biology Group, European Synchrotron Radiation Facility, F-38000 Grenoble, France; <sup>c</sup>European Molecular Biology Laboratory Grenoble, 38042 Grenoble Cedex 9, France; and <sup>d</sup>Laboratoire de Reproduction et Développement des Plantes, Université de Lyon, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, CNRS, INRA, F-69364 Lyon, France

Edited by Mark Estelle, University of California, San Diego, La Jolla, CA, and approved June 13, 2017 (received for review February 22, 2017)

Transcriptional repression involves a class of proteins called corepressors that link transcription factors to chromatin remodeling complexes. In plants such as *Arabidopsis thaliana*, the most prominent corepressor is TOPLESS (TPL), which plays a key role in hormone signaling and development. Here we present the crystallographic structure of the *Arabidopsis* TPL N-terminal region comprising the LisH and CTLH (C-terminal to LisH) domains and a newly identified third region, which corresponds to a CRA domain. Comparing the structure of TPL with the mammalian TBL1, which shares a similar domain structure and performs a parallel corepressor function, revealed that the plant TPLs have evolved a new tetramerization interface and unique and highly conserved surface for interaction with repressors. Using site-directed mutagenesis, we validated those surfaces *in vitro* and *in vivo* and showed that TPL tetramerization and repressor binding are interdependent. Our results illustrate how evolution used a common set of protein domains to create a diversity of corepressors, achieving similar properties with different molecular solutions.

TOPLESS | corepressor | auxin signaling | crystal structure | tetramerization

The regulation of transcription is central to many biological processes. It involves transcription factors (TFs) that bind to regulatory sequences and either activate or repress gene expression. For repression, eukaryotes have developed various strategies involving proteins known as corepressors (1). These proteins usually do not directly interact with DNA, but are addressed to regulatory sequences by sequence-specific TFs. Many corepressors work as hub proteins, as they are able to interact with a wide variety of TFs. In many cases, corepressors recruit chromatin remodeling factors (such as histone deacetylases) to the regulatory sequences of down-regulated genes. In eukaryotes, the most studied corepressors include human SMRT (silencing mediator of retinoid acid and thyroid hormone receptor) and NCoR (nuclear receptor corepressor) complexes (2–5), the yeast proteins Sin3 and Tup1 (6, 7), or the homologous *Drosophila* Groucho (Gro) and mammalian transducin-like enhancer protein (TLE) (8).

Corepressors are also present in plants. In *Arabidopsis thaliana*, the GRO/TUP family of corepressors has about 13 members, including the founding members *LEUNIG* (*LUG*) and *TOPLESS* (*TPL*) (9, 10). Both genes were initially identified genetically, based on developmental defects occurring in *Arabidopsis* mutants bearing mutations in the *TPL* or *LUG* genes. The *tpl-1* dominant mutation causes an abnormal development of the embryo (11), whereas the *lug* mutant shows floral organ defects (12). The TOPLESS protein was later found to be involved in multiple pathways and to interact with numerous transcriptional repressors (13). Notably, TPL interacts with IAA proteins involved in transcriptional repression in auxin signaling (14, 15), as well as with the WUS protein regulating plant stem cell homeostasis (13, 16).

LUG and TPL contain several domains previously characterized in other corepressors. They possess WD40 repeats (Fig. 1A) (9), which are also present in the above-mentioned Tup1, Groucho, TLE, and TBL1 [transducin (beta)-like 1], a member of the SMRT/NCoR complex in *Homo sapiens* (3, 9, 17).

At its N terminus, TPL contains a LIS1 homology (LisH) and a C-terminal to LisH (CTLH) domain. These two domains are often, but not always, found together. LisH domains have been shown to mediate protein dimerization (18, 19). A LisH domain is also present in the TBL1 corepressor; in this case, it confers its capacity both to form tetramers and to interact with hydrophobic peptides from SMRT/NCoR and GPS2 proteins (3). TPL has also been shown to interact with transcriptional regulators through a hydrophobic peptide called the ethylene-responsive element binding factor-associated amphiphilic repression (EAR) motifs domain (13, 20, 21). This interaction was mapped to the LisH CTLH region (14, 22), suggesting TPL and TBL1 might function similarly.

We have solved the crystallographic structure of *Arabidopsis* TPL, which is structurally similar to the previously determined TPR2 structure from rice (23). Our analyses shed light on this essential plant corepressor; we highlight the presence of a previously

## Significance

In most biological processes, genes have to be activated and/or repressed. In plants, the TOPLESS protein is essential for gene repression through its action as a corepressor bridging transcription factor with chromatin remodeling complexes. Here we combine biochemical and structural studies to describe the structure of TOPLESS, how it tetramerizes, and how it interacts with its protein partners. We show that both the tetramerization interface and the binding site for protein partners have been conserved since algae, highlighting the ancestry of TOPLESS function. Comparison of this plant protein with one of its animal counterparts also shows how corepressors can use a common domain differently to achieve similar properties, illustrating the tinkering of evolution in transcriptional repression.

Author contributions: M.H.N., T.V., R.D., and F.P. designed research; R.M.-A., M.H.N., A.L., T.V.-P., D.M., C.G.-A., G.B., and R.D. performed research; R.M.-A., M.H.N., A.L., T.V.-P., D.M., T.V., R.D., and F.P. analyzed data; and R.M.-A., M.H.N., T.V., R.D., and F.P. wrote the paper.

The authors declare no conflict of interest.

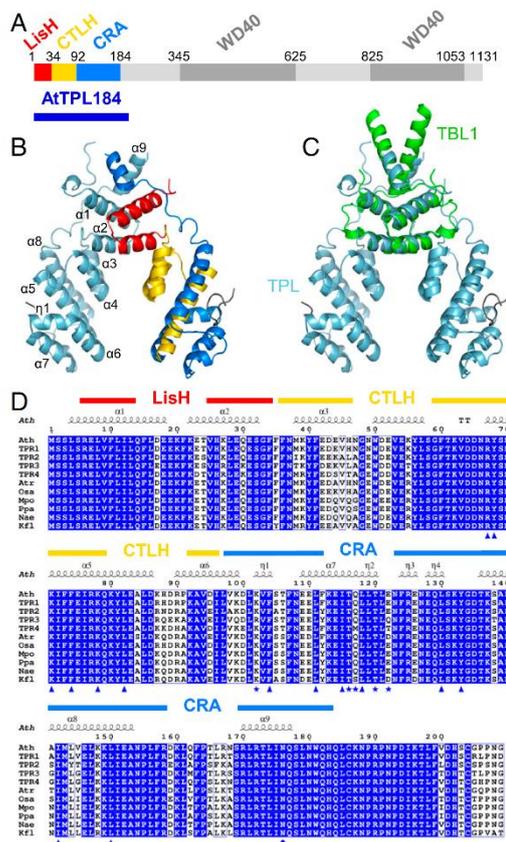
This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID codes 5NQV and 5NQS).

<sup>1</sup>To whom correspondence may be addressed. Email: [teva.vernoux@ens-lyon.fr](mailto:teva.vernoux@ens-lyon.fr), [nanao@esrf.fr](mailto:nanao@esrf.fr), or [renaud.dumas@cea.fr](mailto:renaud.dumas@cea.fr).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1703054114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1703054114/-DCSupplemental).



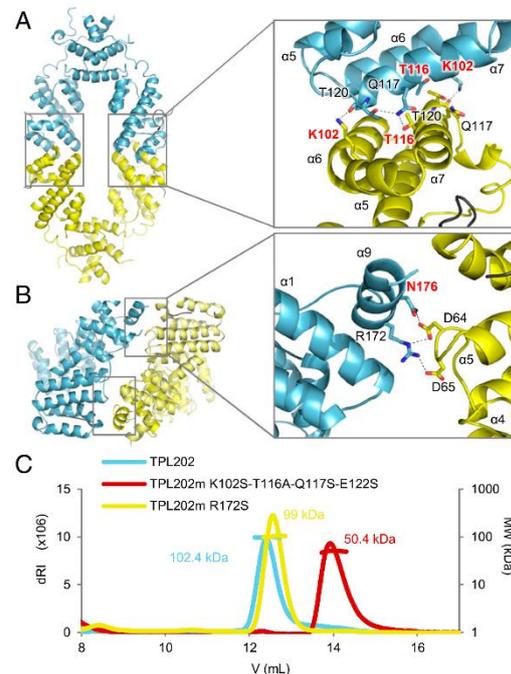
**Fig. 1.** TPL N terminus structure and conservation. (A) AtTPL protein schematic representation. (B) Cartoon representation of AtTPL184 homodimer. LisH, CTLH, and CRA domains are colored in red, gold, and blue in one of the two monomers, respectively. (C) TBL1 (green) and AtTPL184 (blue) structure superimposition. (D) Sequence alignment of the N terminus of TPL and *A. thaliana* (Ath) TPR proteins. Atr, *Amborella trichopoda*; Osa, *Oryza sativa*; Mpo, *Marchantia polymorpha*; Ppa, *Physcomitrella patens*; Nae, *Nothoceros aenigmaticus*; Kfl, *K. flaccidum*. Triangle, amino acids involved in the interaction with the IAA peptide; star, amino acids involved in tetramerization; circle, N176 residue.

unidentified CRA domain, map the tetramerization domain, and reveal the close dependence between peptide binding and tetramerization. We also show that, despite the similarities between TBL1 and TPL, these two tetrameric repressors show remarkably different organizations.

**Results**

**Crystal Structure of *A. thaliana* TPL N-Terminus.** We focused on the highly conserved N-terminal part of TPL that includes the LisH and CTLH domains plus an additional region with no described structural similarities (Fig. 1A). We solved the structure of the first 184 amino acids from *A. thaliana* TPL, AtTPL184 and its selenomethionine derivative at 2.61 and 2.92 Å resolution, respectively (SI Appendix, Table S1). Each monomer presents an  $\alpha$ -helical topology

made by nine  $\alpha$ -helices ( $\alpha$ 1– $\alpha$ 9) connected by loops (Fig. 1B); helices  $\alpha$ 1 and  $\alpha$ 2 form the LisH domain,  $\alpha$ 3– $\alpha$ 5 the CTLH domain, and  $\alpha$ 6– $\alpha$ 9 are a third domain we identified as a CT11-RanBPM (CRA) domain (24) (see detailed explanation on CRA identification in SI Appendix, Fig. S1). Monomers are arranged into dimers, with the long  $\alpha$ 9 from the CRA domain bridging the LisH domain (Fig. 1B). Dimerization is thus mediated by two discontinuous regions (the LisH and  $\alpha$ 9). The configuration adopted by these regions is structurally similar to the LisH-containing domain from the human corepressor TBL1 (3) (Fig. 1C). The AtTPL184 dimers are arranged in tetramers (dimers of dimers), but the crystal packing suggests two possible tetrameric forms (Fig. 2A and B). The AtTPL184 structure has an rmsd of 0.7 Å for 179 C $\alpha$ , using DALI with the recently published OsTPR2 structure from rice (23) (SI Appendix, Fig. S2). TPL/TPR protein sequences are indeed highly conserved over several hundreds of My of evolution (from Charophyte algae to angiosperms; Fig. 1D). DALI searches also established that Smu1 (suppressor of mec-8 and unc-52), a splicing factor and replication regulator, shows a high similarity with TPL (rmsd of 0.8 Å for 35 C $\alpha$  of the LisH domain and of 3.4 Å for 107 C $\alpha$  of the CTLH-CRA domains), consistent with its described resemblance to TPR2 (25). However, the relative positions of the LisH and CTLH-CRA



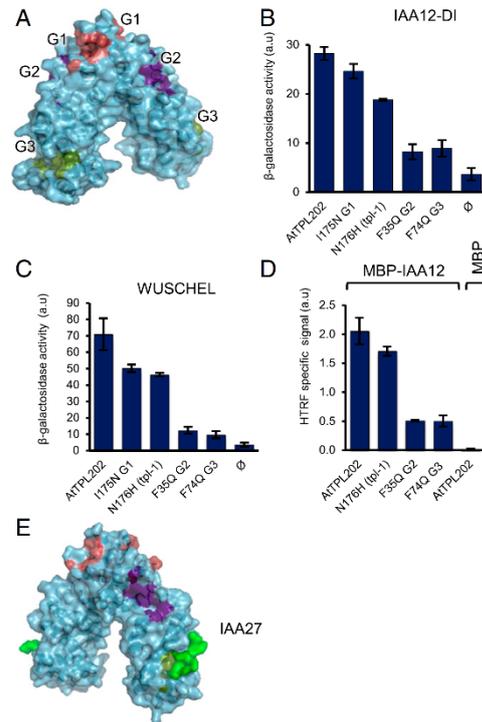
**Fig. 2.** TPL tetramerization. (A) AtTPL184 tetrameric form I (dimers in blue and yellow). (Inset) Close-up view of the tetramerization interface with a rotation of about 90° to better view interacting residues. The interactions between chains C and D (Left) are slightly different from those between A and B (Right), as indicated in SI Appendix, Table S3. Residues mutated in different experiments are shown in red. (B) AtTPL184 tetrameric form II and close-up view of the interface. Main residues shown as sticks; distances <3.4 Å shown as dashes. (C) SEC-MALLS on TPL202 (blue) and TPL202 tetramerization mutants in both tetramerization interfaces I (red) and II (yellow). dRI, differential refractive index; MW, molecular weight.

domains are different, and Smu1 does not tetramerize (*SI Appendix*, Fig. S3).

**Characterization of the Tetramerization Interface of AtTPL.** We tested whether the tetrameric arrangement observed in the crystal is valid in solution, using size exclusion chromatography followed by multiple angle laser light scattering (SEC-MALLS). We determined the size of AtTPL184 and AtTPL202, a more soluble construct comprising TPL's first 202 residues. The measured molecular weights confirmed the tetrameric state of both proteins (*SI Appendix*, Table S2A). Next, we tested which of the two possible tetrameric conformations mentioned earlier (Fig. 2 *A* and *B*) was observed in solution. In the first tetrameric form, also proposed by Ke et al. (23) for OsTPR2, but not experimentally validated, the dimers interact through helices  $\alpha 6$  and 7. This arrangement creates two symmetric tetramerization interfaces of  $600 \text{ \AA}^2$  each (tetramerization interface I) (Fig. 2*A*). The second possible tetramerization interface (tetramerization interface II) involves two symmetric interfaces of  $370 \text{ \AA}^2$  between helix  $\alpha 9$  (with residues R172 and N176) and a small loop between helices  $\alpha 5$  and  $\alpha 6$  (Fig. 2*B*). This surface is less likely. However, the previously identified dominant *tpl-1* point mutation lies within this interface (11). We mutated key residues located in both tetramerization interfaces and determined the molecular weights of resulting proteins using SEC-MALLS (*SI Appendix*, Table S2 *B* and *C*). Mutations of N176 (N176H) or R172 (R172S) from interface II did not affect the complex molecular weight (Fig. 2*C* and *SI Appendix*, Table S2*C*). In contrast, K102S and T116A single mutations in interface I yielded complexes of intermediate molecular weights (3.2–3.5 or 2.6–3.1 monomers, respectively, depending on protein concentrations) (*SI Appendix*, Table S2*B*). This behavior is indicative of a fast dimer–tetramer balance that hampers the separation of the two species in the SEC column. We also assayed the effect of other mutations in interface I. TPL202m/Q117M-E122T and TPL202m/Q117S-E122S were still tetrameric, but the quadruple mutation (TPL202m/K102S-T116A-Q117S-E122S) disrupted the tetramer into a dimer (Fig. 2*C* and *SI Appendix*, Table S2*B*). SEC analyses of the oligomerization state of TPL202 and TPL202m/K102S-T116A-Q117S-E122S at different protein concentrations (from 2.25 to 36  $\mu\text{M}$ ) showed that the oligomeric state of these proteins is independent of the protein concentration, suggesting TPL likely behaves as a tetramer in vivo (*SI Appendix*, Fig. S4). K102 and T116 are thus key residues on interface I involved in TPL tetramerization. The central role of K102 is obvious in both AtTPL184 (Fig. 2*A*) and OsTPR2 structures (*SI Appendix*, Table S3), as it interacts with T120 of the opposite monomer. The role of T116 is less clear, as it forms a hydrogen bond with Q117 only in the AtTPL184 structure and not in OsTPR2's. The K102 and T116 residues are, however, both highly conserved among TPL/TPRs from charophytes to angiosperms, suggesting tetramerization is evolutionarily conserved (Fig. 1*D*).

**TPL Interacts with EAR Motifs Through Hydrophobic Groove 3.** TPL interacts with multiple proteins containing EAR/EAR-like hydrophobic motifs (13). Examination of AtTPL184 monomer structures identified three hydrophobic grooves (G1–G3) as possible candidates to interact with these motifs (Fig. 3*A*). G1 is formed by the LisH domain and the C-terminal  $\alpha 9$  helix of the CRA domain and contains the N176 residue affected by the *tpl-1* mutation (11). In TBL1, mentioned earlier, a similar hydrophobic groove is present and interacts with hydrophobic motifs from the SMRT/NCOR corepressor complex (3). G2 is a small hydrophobic groove positioned between the LisH and the CTLH domains. G3 is formed by a three-helix bundle that includes part of the CTLH domain (helix  $\alpha 5$ ) and helices  $\alpha 7$  and  $\alpha 8$  of the CRA domain.

To determine which groove is involved in the binding of EAR/EAR-like motifs, key hydrophobic amino acids were mutated in each of them, and we assessed the interaction between TPL and two



**Fig. 3.** Hydrophobic groove 3 is implicated in the interaction with LxLxL EAR motifs. (A) Hydrophobic grooves (G1, G2, and G3) identified in TPL N terminus. (B and C) Y2H binding assay between TPL-202 (WT and mutants) with IAA12-DI domain (B) and WUSCHEL protein (C).  $n = 3$  for all experiments. Error bars represent SD,  $\phi$  for empty vector. (D) HTRF binding assays between His-TPL202 (WT and mutants) and MBP-IAA12 protein. HTRF-specific signal is reported.  $n = 3$  for all experiments. Error bars represent SD. (E) Structure of AtTPL184 IAA27 peptide IAA27. a.u., arbitrary units.

of its partners, using yeast two-hybrid (Y2H) and homogeneous time-resolved fluorescence (HTRF) (Fig. 3 *B–D*). For Y2H assays, we used domain I from the AtIAA12 (IAA12-DI) protein, which contains a LxLxL-type EAR motif known to interact with TPL (14, 23), and the WUSCHEL protein (WUS), which contains a LxLxL-type EAR motif and a TLxLFP WUS-box, both involved in WUS repressor function (16, 26). Both assays showed that mutations in the G1 groove (I175N and N176H) only weakly affected the interaction with IAA12 or WUS (Fig. 3 *B–D*), in agreement with previous results showing that the TPL N176H mutant still interacts with IAA12 (14) and with the OsTPR2 structure, where G1 is occluded by a Zinc finger and inaccessible to EAR motifs (23). In contrast, mutations in G2 (F35Q) and G3 (F74Q) considerably weakened the interaction with IAA12 or WUS (Fig. 3 *B–D*), indicating that both G2 and G3 might be important for interacting with EAR/EAR-like motifs. To better characterize these interactions, we cocrystallized AtTPL184 with IAA27 LxLxL-type EAR motif and obtained a high-resolution structure at 1.95  $\text{\AA}$  of the AtTPL184/EAR complex. No conformational changes are induced by the binding of the peptide, localized in G3 (Figs. 3*E* and 4*A*), as observed for OsTPR2 (23). The proline and the three leucine

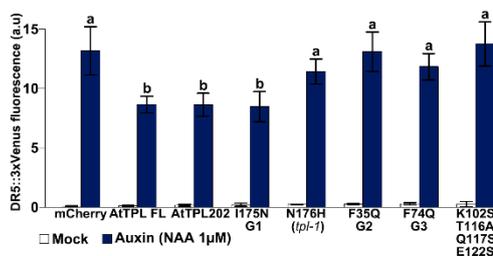


repressive activity in planta. We focused on auxin signaling, as TPL was previously shown to be recruited by AUX/IAA proteins to repress auxin target genes (11, 14, 27). TPL is differentially expressed in tissues (11, 28), and increasing TPL dosage affects both root and shoot development (29), suggesting TPL expression could be limiting in tissues. We thus generated transgenics overexpressing TPL fused to mCherry under the 35S promoter. Surprisingly, none of the lines tested showed reproducible defects in auxin responses, despite a strong TPL-mCherry expression (*SI Appendix, Fig. S7*). Conversely, when tested in protoplasts containing an integrated auxin-inducible DR5::VENUS reporter, overexpression of the same TPL-mCherry fusion had a repressive effect of around 40% on the induction of VENUS fluorescence after treatment with 1  $\mu$ M of the synthetic auxin Naphthalene-1-acetic acid (Fig. 6). Expression of AtTLP202 lacking the two WD40 domains repressed DR5 activity similarly consistent with previous assays in yeast (30, 31). Conversely, the AtTLP202 mutant forms affecting the EAR peptide binding G3 groove (TLP202m/F74Q) and the dimeric mutant TLP202m/K102S-T116A-Q117S-E122S lost their repression activity. A mutant form affected in the G2 groove (TLP202m/F35Q) similarly lost its repressive activity, an observation coherent with the weaker interaction with IAA12 in the Y2H and HTRF assays (Fig. 3). Finally, TLP202m/N176H (*tpl-1*) also shows a reduced repression capacity despite being still able to interact with IAA12 repressors (Fig. 3) (11). These results confirmed the importance of TPL residues involved in peptide binding in the G3 groove or tetramerization in planta and underline the importance of the AtTLP N terminus for transcriptional repression in auxin signaling. They also demonstrate that intact G1 and G2 grooves are required for repression in planta, despite the apparent absence of direct interaction with EAR domains of IAAs at these grooves.

## Discussion

We have obtained the crystallographic structure of *Arabidopsis* TPL N terminus, which allowed us to design a series of experiments to study the oligomerization and peptide binding in vitro and in vivo.

**EAR Peptide Binding and Tetramerization Are Intimately Related.** As previously shown for OsTPR2 (23), we localized the interaction site of the EAR peptide to one of the three hydrophobic grooves present in TPL-N. The higher resolution of our structure allowed us to refine the interactions between EAR peptide and TPL. Our data suggests a different register than previously proposed (with



**Fig. 6.** Importance of TPL G3 groove and tetramerization in auxin signaling. In planta repression assay in Col-0 protoplasts containing the integrated DR5::VENUS reporter gene. A loss of repression activity is observable in TPL mutants impaired in their interaction with repressors, their tetramerization, or carrying the *tpl-1* mutation. DR5 activity is reported ( $n = 200$  protoplasts). Error bars correspond to the 95% confidence interval. Two variables with different letters are significantly different (corrected  $P$  values  $< 0.05$ ; two-sided Kruskal-Wallis test). a.u., arbitrary units.

Martin-Arevalillo et al.

a two-residue shift), which we validated using mutations in the innermost part of the groove.

Based on gel filtration, TPL-N was previously proposed to form a tetramer (23). Using a combination of SEC-MALLS, mutagenesis of two potential tetramerization interfaces, and an in planta assay, we demonstrated that TPL does indeed tetramerize and that the interface proposed for TPR2 (23) is the same as in AtTLP. It remains to be established whether the full-length TPL protein is indeed tetrameric. Our work also shows that tetramerization and peptide binding are intimately linked: The amino acids responsible for EAR binding and tetramerization are in close proximity, and tetramerization cannot be altered without affecting EAR binding. Nevertheless, EAR peptide binding to TPL does not appear to affect its oligomerization.

Tetramerization is a common feature of corepressors, with a few structural commonalities in tetramers arrangement (3, 7, 17, 32) (*SI Appendix, Fig. S8*). Why tetramerization is essential for their function is still elusive. We speculate it could be a way to spread chromatin marks along the DNA; for example, to reproduce chromatin domains after mitosis (33). It could also help to recruit multimeric TFs using weak individual interactions that are compensated for by a high avidity (23, 34). To specifically test the importance of tetramerization, mutations are needed that compromise this property without altering other features (such as complex formation with other proteins). However, our work indicates it might be difficult to study the respective contributions of EAR peptide binding and tetramerization in TPL function. We also show that the two other hydrophobic grooves on TPL-N are important for TPL-N function, as introducing mutations in any of them interferes with the interaction with WUS or IAA12 and with TPL-N repressive activity in vivo. Our data suggest these mutations probably do not affect the interaction with the EAR motif itself (*SI Appendix, Fig. S9*), but TPL-N and rice TPR2-N structures indicate that the G2 groove is accessible and could interact with other protein motifs.

**TPL Structure Sheds Light on Other CTLH-CRA Proteins.** The analysis of the TPL structure revealed the presence of a CRA motif. This motif was first identified in RanBP9 (human Ran binding protein 9) for mediating the interaction with FMRP (fragile X mental retardation protein) in the microtubule organizing center (35). CRA motifs are present in thousands of eukaryotic proteins such as RanBPM in human or E3 ligases proteins (SIE3 in plants or Gid in yeast) (35–37). They are often associated with the CTLH domain, but with variable-length spacers between them. Whether CRA and CTLH motifs were forming a single domain or two separate domains had remained elusive because of the lack of structural data. TPL and TPR2 structures show that CTLH and CRA form two distinct but intertwined domains that can mediate tetramerization and hydrophobic peptide binding. Having unambiguously identified the domain next to the CTLH as a CRA domain now allows the modeling of the wealth of proteins possessing the LisH-CTLH-CRA domains, using OsTPR2/AtTLP or Smu1 as models. We inventoried and modeled with high confidence all such *Arabidopsis* proteins (99.8–100% using Phyre2) (*SI Appendix, Table S5*), revealing that some possess a TPL-like G3 hydrophobic groove, whereas others are more similar to Smu1, which uses another hydrophobic groove to interact with the RED hydrophobic peptide (*SI Appendix, Table S5*). Proteins possessing the three domains (LisH, CTLH, and CRA) are found at the very base of the green lineage (including chlorophyte algae such as *Chlamydomonas reinhardtii*), and also outside the plant kingdom, indicating that those domains have assembled very early in evolution and can now be modeled using TPL/TPR2 structures in many eukaryotic proteins (*SI Appendix, Fig. S1*).

**TBL1 and TPL Use the Same Domain in Different Ways.** TBL1 is a member of the highly studied SMRT/NCOR nuclear receptor complex (3). It functions as a bridge between two members of the complex: GPS2 and SMRT. In TBL1, tetramerization occurs via

PNAS | July 25, 2017 | vol. 114 | no. 30 | 8111

two interfaces in the LisH domain, forming a dimer of dimers. Binding of partner proteins occurs via a hydrophobic groove on the face of LisH that is distal to the dimerization surfaces (*SI Appendix, Fig. S8B*). We therefore anticipated that TPL, which also functions as a tetrameric corepressor hub protein and contains a LisH domain, would tetramerize and bind partner proteins in a similar fashion. However, the crystal structures of TPL reveal that it uses a completely different surface for protein-protein interaction that is within the CTLH-CRA domain. Furthermore, unlike TBL1, which uses the “bottom” face of the LisH domain to form a second dimer interface, dimerization in TPL occurs at the base of the CTLH-CRA domain.

Comparing TPL and TBL1 illustrates how corepressors can differently use a common domain (LisH) to achieve similar properties (tetramerization and binding to hydrophobic peptides), and provides an excellent example of the tinkering of

evolution (38) that, in this case, used a common toolset to achieve transcriptional repression in different organisms.

## Materials and Methods

Material and experimental procedures for vectors construction, protein expression and purification, crystallization, protein structure determination, native molecular mass determination, yeast two hybrid interaction tests, homogeneous-time resolved fluorescence interaction tests, fluorescence anisotropy interaction tests, and repressions assays are described in detail in the *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank R. Miras (Laboratoire Chimie et Biologie des Métaux Grenoble) for help with SEC-MALLS and E. Thevenon and S. Lainé for help. This work was supported by the French National Research Agency programs Auxiflo (ANR-12-BSV6-0005 to T.V. and F.P.) and Gral (ANR-10-LABX-49-01), the Human Frontier Science Program organization (Grant RFG0054-2013 to T.V.), a Université Grenoble Alpes PhD fellowship (R.M.-A.), and the European Community's Seventh Framework Program (FP7) under grant agreement 283570 (BioStruct-X to R.D.).

1. Payankulam S, Li LM, Arnosti DN (2010) Transcriptional repression: Conserved and evolved features. *Curr Biol* 20:R764–R771.
2. Perissi V, Jepsen K, Glass CK, Rosenfeld MG (2010) Deconstructing repression: Evolving models of co-repressor action. *Nat Rev Genet* 11:109–123.
3. Oberoi J, et al. (2011) Structural basis for the assembly of the SMRT/NCOR core transcriptional repression machinery. *Nat Struct Mol Biol* 18:177–184.
4. Watson PJ, Fairall L, Schwabe JWR (2012) Nuclear hormone receptor co-repressors: Structure and function. *Mol Cell Endocrinol* 348:440–449.
5. Mottis A, Mouchiroud L, Auwerx J (2013) Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev* 27:819–835.
6. Grzenda A, Lomberg G, Zhang JS, Urrutia R (2009) Sin3: Master scaffold and transcriptional corepressor. *Biochim Biophys Acta* 1789:443–450.
7. Matsumura H, et al. (2012) Crystal structure of the N-terminal domain of the yeast general corepressor Tup1p and its functional implications. *J Biol Chem* 287:26528–26538.
8. Agarwal M, Kumar P, Mathew SJ (2015) The Groucho/Transducin-like enhancer of split protein family in animal development. *IUBMB Life* 67:472–481.
9. Liu Z, Karmarkar V (2008) Groucho/Tup1 family co-repressors in plant development. *Trends Plant Sci* 13:137–144.
10. Lee JE, Goltz JF (2012) Diverse roles of Groucho/Tup1 co-repressors in plant growth and development. *Plant Signal Behav* 7:86–92.
11. Long JA, Ohno C, Smith ZR, Meyerowitz EM (2006) TOPLESS regulates apical embryonic fate in Arabidopsis. *Science* 312:1520–1523.
12. Sitaraman J, Bul M, Liu Z (2008) LEUNIG\_HOMOLOG and LEUNIG perform partially redundant functions during Arabidopsis embryo and floral development. *Plant Physiol* 147:672–681.
13. Causier B, Ashworth M, Guo W, Davies B (2012) The TOPLESS interactome: A framework for gene repression in Arabidopsis. *Plant Physiol* 158:423–438.
14. Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. *Science* 319:1384–1386.
15. Dinesh DC, Villalobos LIAC, Abel S (2016) Structural biology of nuclear auxin action. *Trends Plant Sci* 21:302–316.
16. Ikeda M, Mitsuda N, Ohme-Takagi M (2009) Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell* 21:3493–3505.
17. Chodaparambil JV, et al. (2014) Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J* 33:719–731.
18. Kim MH, et al. (2004) The structure of the N-terminal domain of the product of the lissencephaly gene Lis1 and its functional implications. *Structure* 12:987–998.
19. Delto CF, et al. (2015) The LisH motif of muskellin is crucial for oligomerization and governs intracellular localization. *Structure* 23:364–373.
20. Kagale S, Links MG, Rozwadowski K (2010) Genome-wide analysis of ethylene-responsive element binding factor-associated amphiphilic repression motif-containing transcriptional regulators in Arabidopsis. *Plant Physiol* 152:1109–1134.
21. Kagale S, Rozwadowski K (2011) EAR motif-mediated transcriptional repression in plants: An underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* 6:141–146.
22. Wang L, Kim J, Somers DE (2013) Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. *Proc Natl Acad Sci USA* 110:761–766.
23. Ke J, et al. (2015) Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. *Sci Adv* 1:e1500107.
24. Kobayashi N, et al. (2007) RanBPM, Muskelin, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8alpha and ARMC8beta are components of the CTLH complex. *Gene* 396:236–247.
25. Ulrich AKC, Schulz JF, Kamprad A, Schütze T, Wahl MC (2016) Structural basis for the functional coupling of the alternative splicing factors Smu1 and RED. *Structure* 24:762–773.
26. Zhang F, et al. (2014) STENOFOLIA recruits TOPLESS to repress ASYMMETRIC LEAVES2 at the leaf margin and promote leaf blade outgrowth in Medicago truncatula. *Plant Cell* 26:650–664.
27. Wu M-F, et al. (2015) Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *eLife* 4:e09269.
28. Busch W, et al. (2010) Transcriptional control of a plant stem cell niche. *Dev Cell* 18:849–861.
29. Espinosa-Ruiz A, et al. (2017) TOPLESS mediates brassinosteroid control of shoot boundaries and root meristem development in Arabidopsis thaliana. *Development* 144:1619–1628.
30. Pierre-Jerome E, Jang SS, Havens KA, Nemhauser JL, Klavins E (2014) Recapitulation of the forward nuclear auxin response pathway in yeast. *Proc Natl Acad Sci USA* 111:9407–9412.
31. Pierre-Jerome E, Moss BL, Lanctot A, Hageman A, Nemhauser JL (2016) Functional analysis of molecular interactions in synthetic auxin response circuits. *Proc Natl Acad Sci USA* 113:11354–11359.
32. Cerna D, Wilson DK (2005) The structure of Sif2p, a WD repeat protein functioning in the SET3 corepressor complex. *J Mol Biol* 351:923–935.
33. Pirrotta V, Gross DS (2005) Epigenetic silencing mechanisms in budding yeast and fruit fly: Different paths, same destinations. *Mol Cell* 18:395–398.
34. Krishnamurthy VM, Estroff LA, Whitesides GM (2006) Multivalency in ligand design. *Fragment-based Approaches in Drug Discovery*, eds Jahnke W and Erlanson DA (Wiley-VCH, Weinheim, Germany).
35. Menon RP, Gibson TJ, Pastore A (2004) The C terminus of fragile X mental retardation protein interacts with the multi-domain Ran-binding protein in the microtubule-organising centre. *J Mol Biol* 343:43–53.
36. Yuan S, et al. (2012) A ubiquitin ligase of symbiosis receptor kinase involved in nodule organogenesis. *Plant Physiol* 160:106–117.
37. Messens R, et al. (2012) Exploring the topology of the Gid complex, the E3 ubiquitin ligase involved in catabolite-induced degradation of gluconeogenic enzymes. *J Biol Chem* 287:25602–25614.
38. Jacob F (1977) Evolution and tinkering. *Science* 196:1161–1166.

## **Supplementary Information**

**SI Appendix**

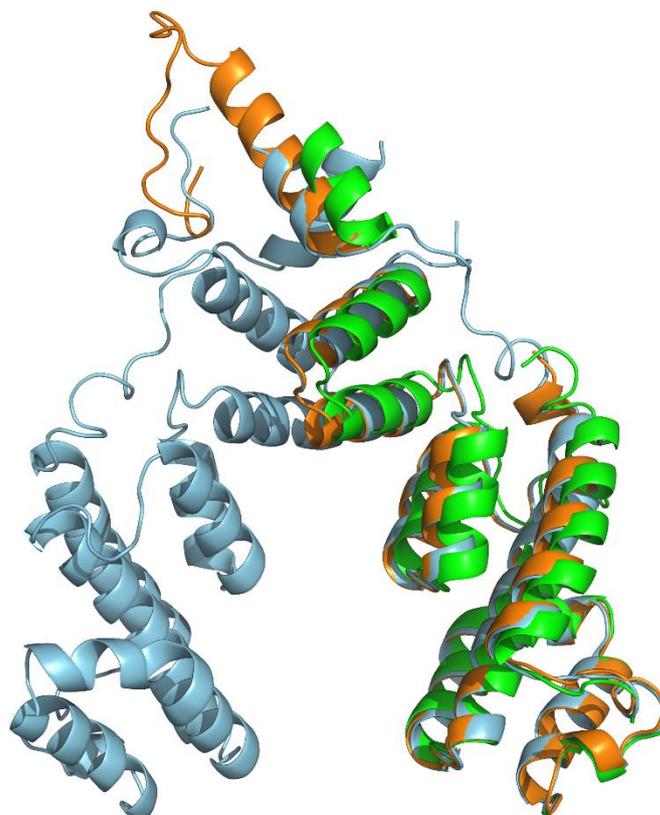
**9 SI figures**

**6 SI tables**

**SI material and methods**

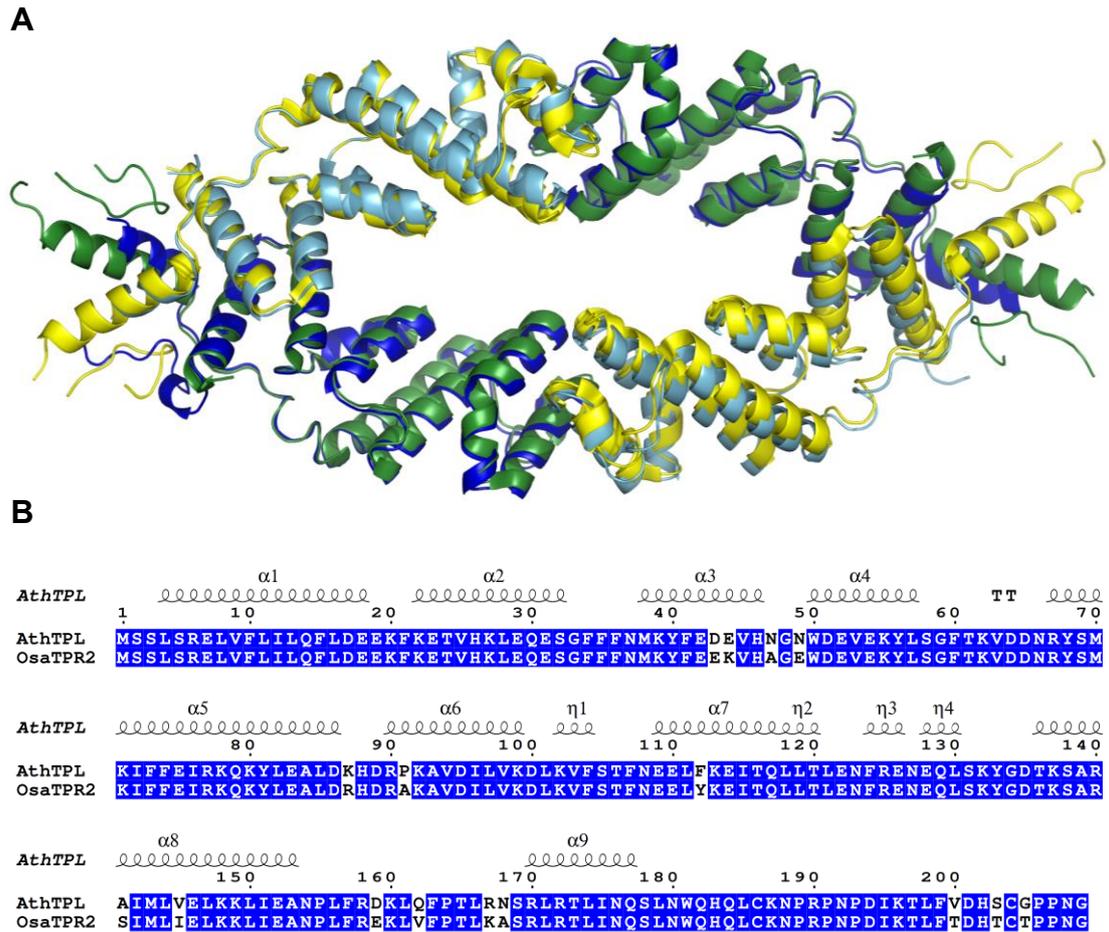
**SI references**

## SI Figures

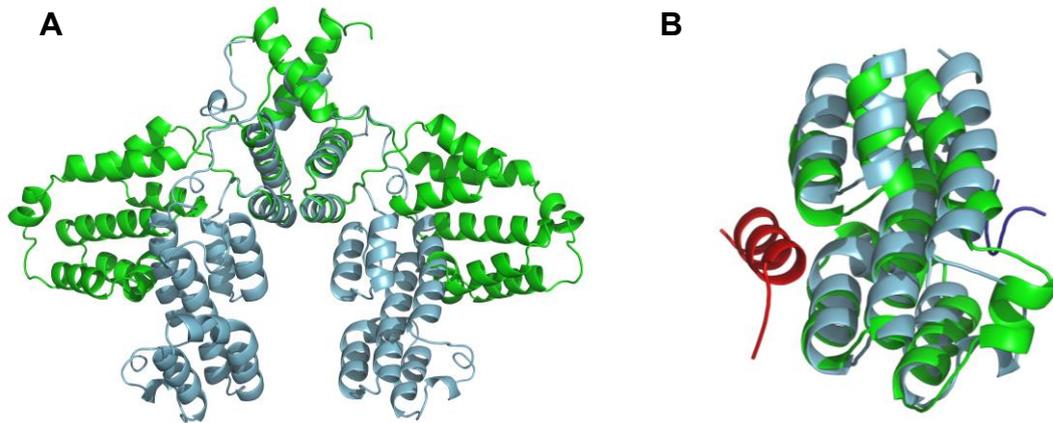


**Figure S1: TPL has a CRA domain.** Superimposition of two proteins possessing the LisH-CTLH-CRA domains (the human RAN-Binding protein 9 [Uniprot Q96S59, green], and a protein from the unicellular algae *C. reinhardtii* [Uniprot A8HQD2, orange] on AtTPL184 structure (blue): Models were generated using PHYRE<sup>2</sup> (1).

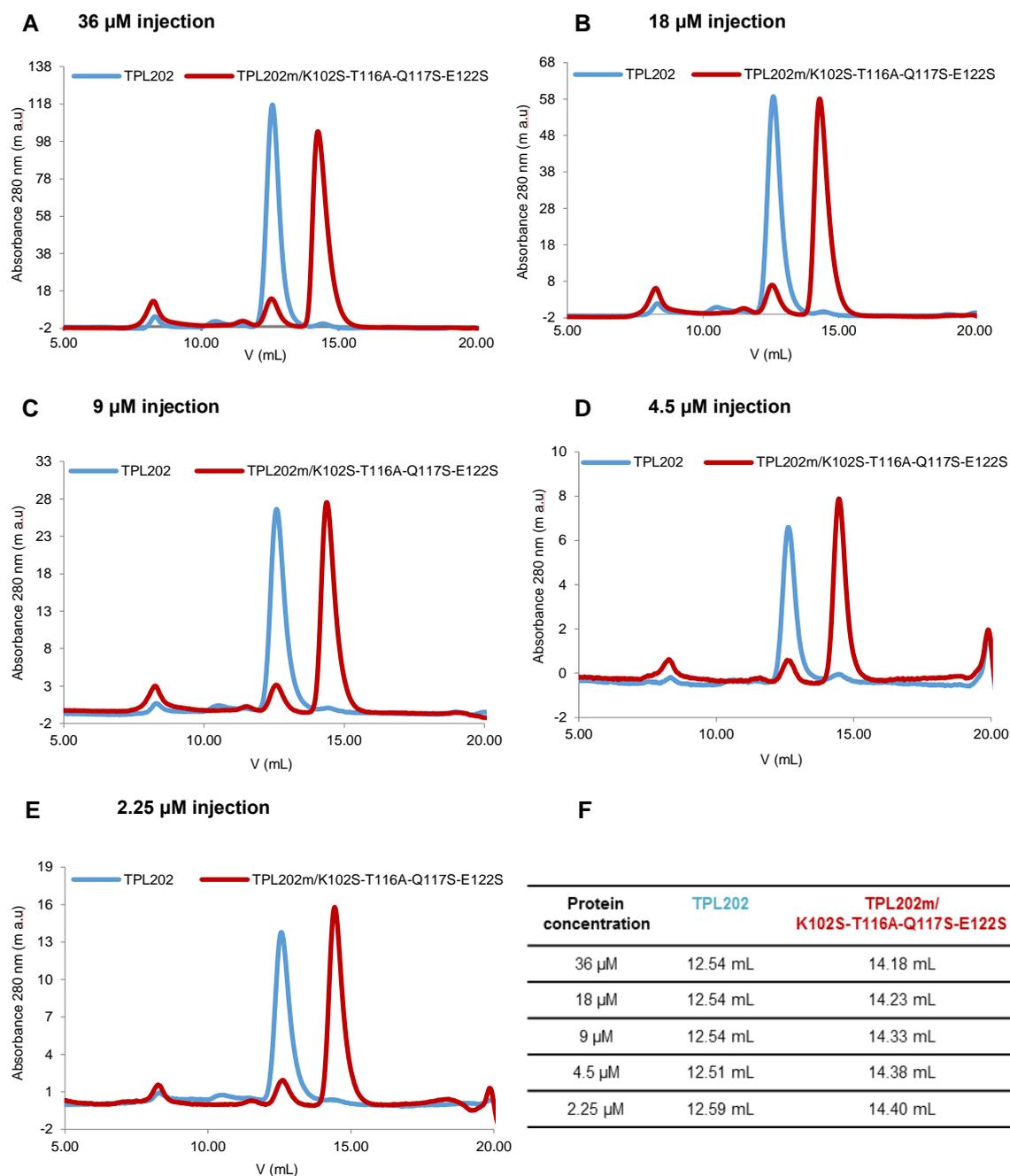
Based on sequence similarity, the CRA domain is detected in numerous proteins containing the LisH and CTLH domains (see Table S5). However, there was no structure available for a CRA domain. When using PHYRE<sup>2</sup>(1), we realized that the CRA domain of all LisH-CTLH-CRA containing proteins was modelled with a high confidence to the C-terminal region of TPL184/TPR2 structures (see Table S5). This analysis revealed that the TPL/TPR2 structures provided the first structure for a CRA domain and that TPL/TPR2 both possessed a CRA domain that had remained unnoticed.



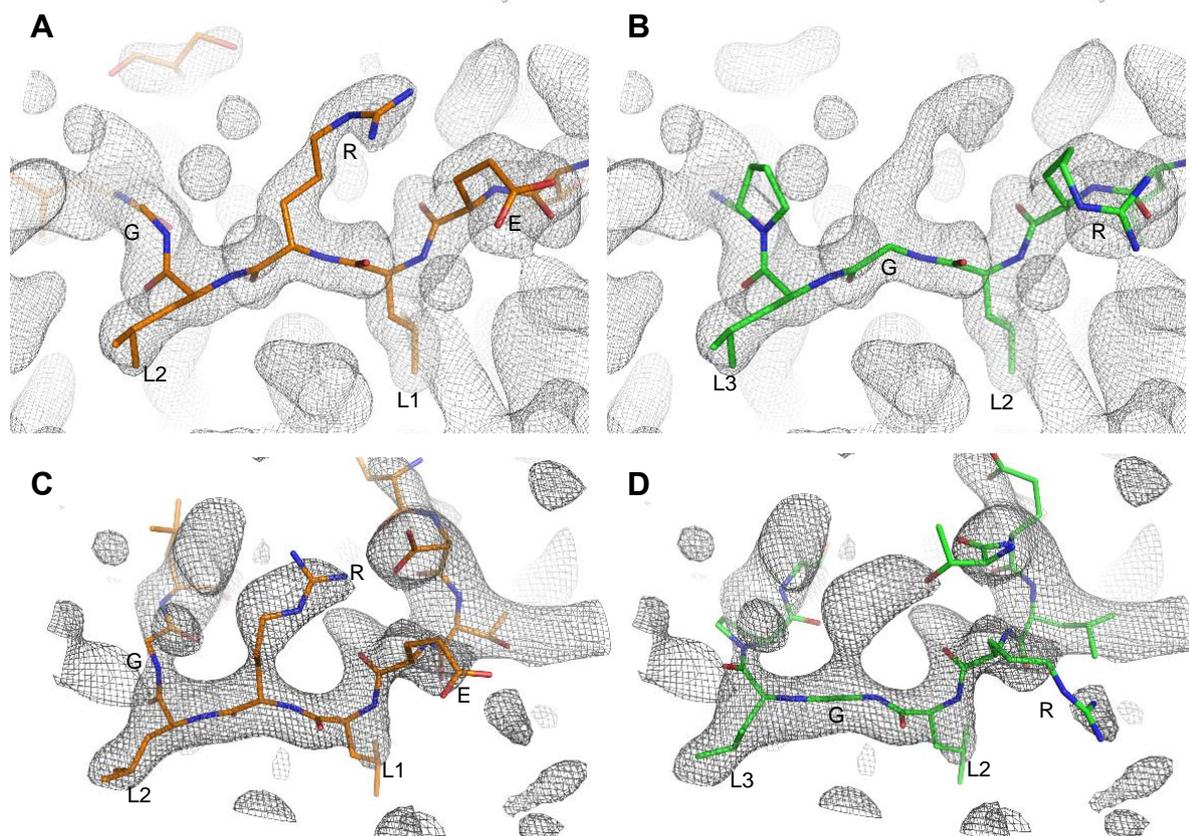
**Figure S2: Structural comparison of AtTPL with OsTPR2.** (A) Superimposition of the tetramers of AtTPL (light and dark blue) and OsTPR2 (yellow and dark green) (5C7F) showing the high similarity between both structures. The AtTPL184 structure has an rmsd of 0.7 Å for 179 Cα using DALI with the OsTPR2 structure. (B) Sequence alignment of the N-terminal part of AtTPL and OsTPR2 using ESPrpt with Multalin (2) and ESPrpt 3.0 (3). These sequences show 92% identity.



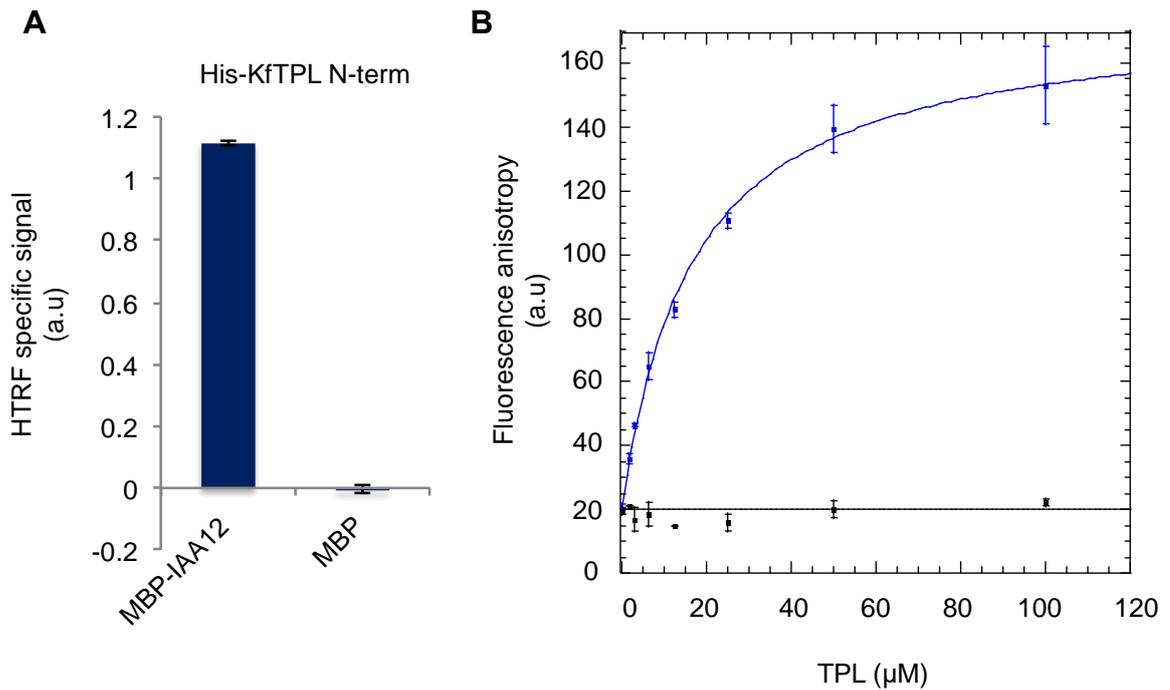
**Figure S3: Structural comparison of AtTPL with Smu1.** (A) Superimposition of the dimer of Smu1 (5EN8) (green) and AtTPL184 (blue) with overlay of the LisH domains. (B) Superimposition of the CRA/CTLH domain of Smu1 interacting with the RED peptide (red) and the CRA/CTLH domain of AtTPL184 interacting with IAA27 (dark blue). The LisH and the CTLH/CRA domains of TPL show a high similarity with Smu1 domains but their relative positions are different between Smu1 and TPL (4).



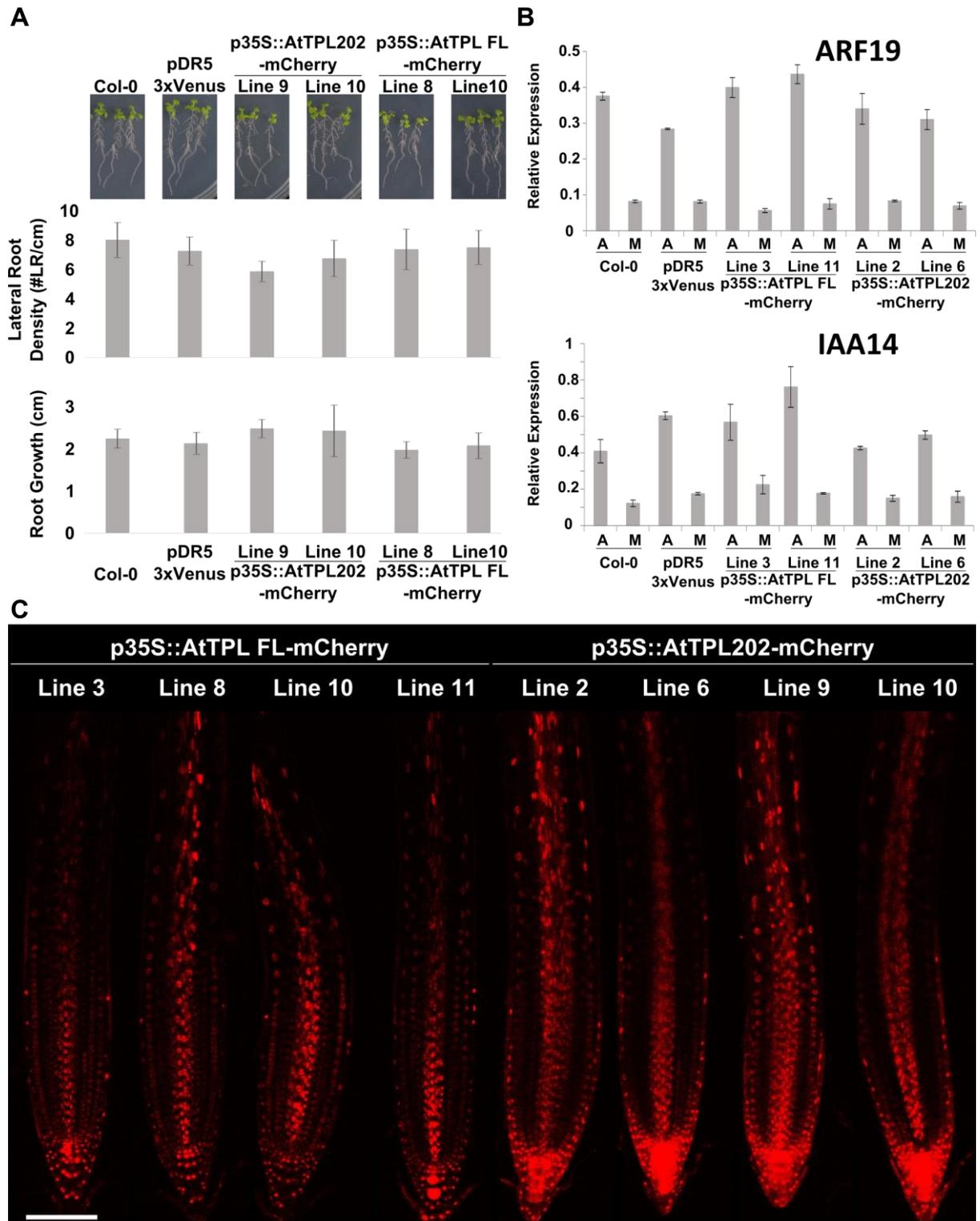
**Figure S4: Analysis of TPL oligomerization state by SEC.** (A-E) Elution profiles of TPL202 (blue) and the dimeric mutant TPL202m/K102S-T116A-Q117S-E122S (red) in a Superdex-S200 increase (GE Healthcare) size-exclusion chromatography column. (F) Elution volumes (mL) of both proteins at different concentrations showing that TPL202 elutes at a tetramer independently of the protein concentration. These experiments have been made at low protein concentrations (2.25  $\mu\text{M}$  - 36  $\mu\text{M}$ ) in comparison with the SEC-MALLS experiments [36  $\mu\text{M}$  (1 mg/mL) – 432  $\mu\text{M}$  (12 mg/mL)].



**Figure S5: Comparing IAA peptide registers between OstPR2 and AtTPL data.** 2Fo-Fc simulated annealed omit electron density map is shown as a grey mesh contoured at one standard deviation above the mean electron density value. In (A) and (B) the electron density for AtTPL184 is shown. The peptide register that we propose is shown in orange sticks in (A). Note the excellent fit in particular of the Arginine residue. In (B), the green sticks represent the peptide register as proposed by Ke *et al.* (5). In (C) and (D), the electron density of the data deposited by Ke *et al.* is shown. The orange model in (C) represents the Ke *et al.* (5) peptide re-modelled using our proposed peptide register whereas the green model in (D) is the peptide as modelled by Ke *et al.*(5).

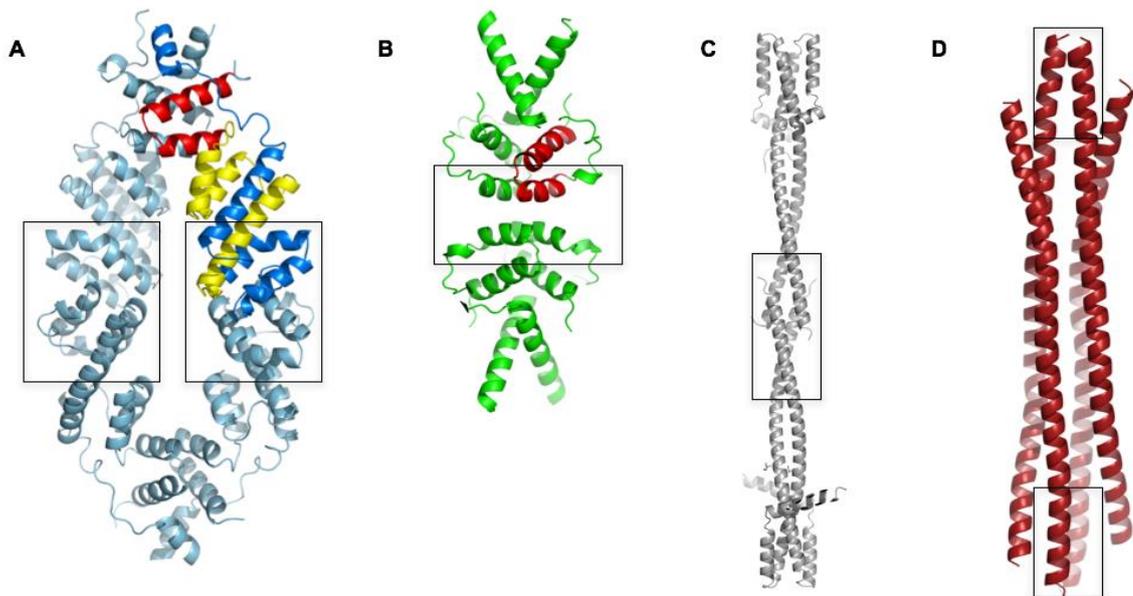


**Figure S6: Evolutionary conservation of TPL interactions.** (A) HTRF interaction assay showing the interaction between the N-terminus of *Klebsormidium flaccidum* TPL and AtIAA12 repressor (n = 3; error bars = SD) (B) Anisotropy interaction assay between KfTPL N-terminus and the FAM-IAA12 EAR peptide (blue) or its mutant version mIAA12 (black) (Table S4). EC<sub>50</sub> for IAA12-EAR = 16.78  $\mu\text{M}$   $\pm$  1.24 (n = 3; error bars = SD).

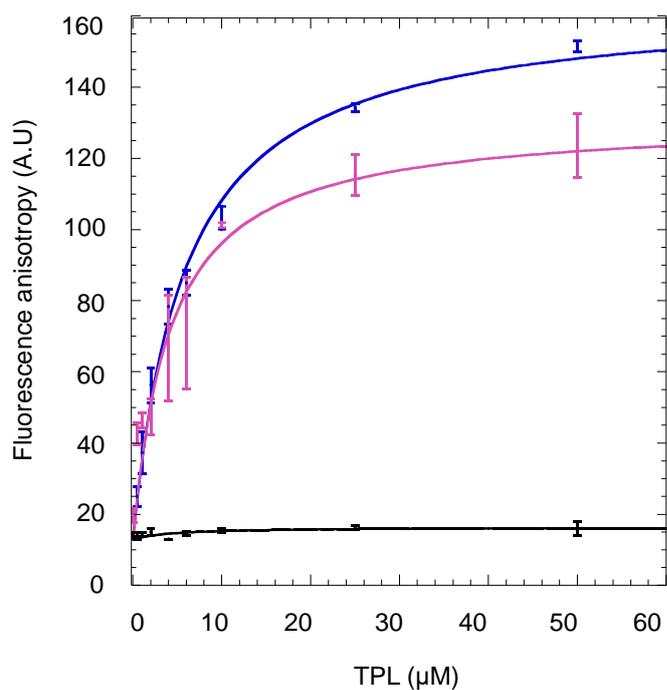


**Figure S7: TOPLESS overexpression lines are not affected in their auxin responsiveness.** (A) Auxin-dependent effects on root growth and lateral root formation was analyzed in 2 independent transgenic arabidopsis lines overexpressing full-length TPL or TPL202. Root growth and lateral root density were determined 5 days after transfer on auxin (NAA 100 nM, n = 10 seedlings). (B) Auxin responses at the transcriptional level

was analyzed in 2 independent overexpression lines of full length TPL or TPL202. *ARF19* and *IAA14* mRNA relative expression were determined 2 hours after transfer on auxin (M: Mock, A: Auxin (NAA 1 $\mu$ M), n = 2 biological replicates). (C) Confocal laser scanning images of root tips of the lines used in the experiments showing relatively identical TPL-mCherry fluorescence levels (scale bar = 100  $\mu$ m). Basta-resistant T2 seedlings were used in all experiments and basta-resistant pDR5::3xVenus were used as controls. Col-0 seedlings grown in the absence of basta were also used as controls.



**Figure S8: Different tetramerization modes of co-repressors.** (A) AtTPL184 tetramer with LisH in red, CTLH in yellow and CRA in marine blue on one of the monomers. (B) Human TBL1 (2XTC) tetramer with LisH in red on one of the monomers (6). (C) Human TLE (4OM3) (7). (D) Yeast TUP1p (3VP8) tetramer (8). Tetramerization interfaces framed with a black rectangle.



**Figure S9: G2 mutant maintains its capacity to interact with EAR motifs.** Fluorescence anisotropy experiment done with FAM-IAA12 EAR peptide or its mutant version mIAA12 (black) (Table S4) showing that TPL202 (blue) and TPL202m/F35Q G2 (pink) similarly interact with LxLxL-type EAR motifs. TPL202 EC<sub>50</sub> = 5.8 μM ± 0.6; TPL202m/F35Q G2 = 4.2 μM ± 1.1 (n = 3; error bars = SD).

## SI Tables

Wavelength, Å	Peak Se*	Apo	IAA peptide
Beamline	ESRF ID29	ESRF ID23-EH2	ESRF ID29
Resolution range, Å	56.76-2.92 (3.03 - 2.92)	23.93-2.61 (2.70 - 2.61)	28.51-1.95 (2.02 - 1.95)
Space group	P 4 <sub>3</sub> 2 <sub>1</sub> 2	P 3 <sub>1</sub> 1 2	P 4 <sub>3</sub> 2 <sub>1</sub> 2
Unit cell axes a,b,c in Å	99.1 99.1 290.0	72.7 72.7 181.1	94.1 94.1 298.0
Total reflections	416016 (32760)	115884 (10928)	709574 (70881)
Unique reflections	32181 (3051)	16927 (1677)	96694 (9366)
Multiplicity	12.9 (10.7)	6.8 (6.5)	7.3 (7.6)
Completeness, %	97 (97)	100 (100)	98 (97)
Mean I/sigma(I)	10.14 (1.45)	12.21 (1.79)	13.11 (1.01)
Wilson B-factor, Å <sup>2</sup>	58.58	57.68	36.99
R-merge	0.2253 (1.718)	0.1059 (0.8868)	0.09131 (2.316)
R-meas	0.2346 (1.802)	0.1144 (0.9624)	0.09837 (2.485)
CC1/2	0.996 (0.571)	0.998 (0.657)	0.999 (0.47)
CC*	0.999 (0.853)	1 (0.89)	1 (0.8)
Reflections used in refinement		16924 (1679)	96685 (9365)
Reflections used for R-free		857 (91)	4830 (441)
R-work, %		20.3 (28.4)	18.5 (36.9)
R-free, %		25.8 (34.7)	21.3 (39.1)
CC(work)		0.955 (0.791)	0.965 (0.740)
CC(free)		0.934 (0.572)	0.939 (0.655)
Number of non-hydrogen atoms		3060	6965
Macromolecules		3029	6359
Ligands			76
Protein residues		357	755
RMS(bonds), Å		0.014	0.014
RMS(angles), °		1.71	1.50
Ramachandran favored (%)		98	98
Ramachandran allowed (%)		1.7	2
Ramachandran outliers (%)		0	0.13
Rotamer outliers (%)		6.5	2
Clashscore		2.14	0.70
Average B-factor Macromolecules		62.22	50.75
Ligands		62.29	50.14
Solvent		55.90	67.12
			55.61

**Table S1:** Data collection and refinement statistics. Outer shell range and statistics are indicated in parentheses.

<b>A</b>	<b>Concentration (mg.mL<sup>-1</sup>)</b>	<b>MW (kDa)</b>	<b>Number of monomers- Quaternary state</b>
AtTPL184	2	95	3.7-tetramer
	1	102.4	3.7-tetramer
AtTPL202	2	103	3.7-tetramer
	4	103	3.7-tetramer
<b>B Tetramerization interface I mutants</b>			
TPL202m/K102S- T116A-Q117S-E122S	2	50.4	1.8-dimer
TPL202m/Q117M- E122T	2	102	3.7-tetramer
TPL202m/Q117S- E122S	2	99.6	3.6-tetramer
	2	72	2.6-equilibrium
	4	78.8	2.9-equilibrium
TPL202m/T116A	6	81.2	3.0-equilibrium
	12	85.3	3.1-equilibrium
	1	87.6	3.2-equilibrium
TPL202m/K102S	2	91.2	3.3-equilibrium
	6	96.9	3.5-equilibrium
<b>C Tetramerization interface II mutants</b>			
TPL202m/N176H	2	103	3.7-tetramer
TPL202m/R172S	2	99	3.6-tetramer
<b>D Groove 3 mutants</b>			
TPL202m/Y68A	2	106	3.8-tetramer
TPL202m/F74Q	2	55	2.0-dimer
TPL202m/L130A	2	55.4	2.0-dimer
TPL202m/Y133A	2	55.4	2.0-dimer
<b>E AtTPL202 + EAR motifs</b>			
+IAA12 EAR motif	2	109.3	3.9-tetramer
+IAA27 EAR motif	2	107.3	3.9-tetramer

**Table S2:** Quaternary structure of AtTPL184 and AtTPL202 (wild type and mutant forms) determined by SEC-MALLS.

	K102/ T120	T120/ K102	Q117/ T116	Q117/ Q117	K113/ Q117	Q117/ K113	E122/ K102
<b>TPL A/B</b>	+	+	+				
<b>TPL C/D</b>	+	+		+	+	+	
<b>5C7F A/C</b>	+	+		+			
<b>5C7F B/D</b>	+	+		+			
<b>4ZHE A/B</b>	+	+					+
<b>4ZHE D/C</b>	+	+		+			
<b>5C6V A/C</b>	+	+		+			
<b>5C6V B/D</b>	+	+		+			
<b>5C7F A/C</b>	+	+		+			
<b>5C7F B/D</b>	+	+		+			

**Table S3:** Interaction (lower than 3.5 Å) observed in tetramer interfaces for AtTPL and OsTPR2 (Ke et al., 2015). TPL: AtTPL/IAA27; 5C7F: OsTPR2/IAA1; 4ZHE: OsTPR2/SeMet; 5C6V: OsTPR2/NINJA; 5C7F: Os-TPR2/IAA10.

EAR peptide	Sequence	
<b>FAM-IAA12</b>	ESELELGLGLSL	Used for anisotropy
<b>FAM-mIAA12</b>	ESEAEAGAGASA	Used for anisotropy
<b>IAA12 long</b>	KSNLPAESELELGLGLSL	Used for SEC-MALLS
<b>IAA27</b>	TELRLGLPGSE	Used for crystallography and SEC-MALLS

**Table S4:** EAR peptides sequences used in this study.

Uniprot	ATG number/name	length	Domain	Function	Model (confidence)	G3/G4
Q94AI7	AT1G15750/TPL	1131	LisH/CTLH/WD40/WD40	Repression	TPR2(100)/Smu1(100)	Y/N
Q0WV90	AT1G80490/TPR1	1120	LisH/CTLH/WD40/WD40	Repression	TPR2(100)/Smu1(100)	Y/N
Q9LRZ0	AT3G16830/TPR2	1131	LisH/CTLH/WD40/WD40	Repression	TPR2(100)/Smu1(100)	Y/N
Q84JM4	AT5G27030/TPR3	1108	LisH/CTLH/WD40/WD40	Repression	TPR2(100)/Smu1(100)	Y/N
Q27GK7	AT3G15880 /TPR4	1135	LisH/CTLH/WD40/WD40	Repression	TPR2(100)/Smu1(100)	Y/N
Q9FNN2	At5g08560	589	LisH/CTLH/WD40		TPR2(100)/Smu1(100)	Y/N
Q9FND4	AT5G43920	524	LisH/CTLH/WD40		TPR2(100)/Smu1(100)	Y/N
Q8W117	AT1G73720/smu1	511	LisH/CTLH/WD40	Splicing	TPR2(99.4)/Smu1(99.8)	N/Y
Q0WVR3	At2g25420/TPR-like	740	LisH/CTLH/LisH/CTLH/WD40		TPR2(99.9)/Smu1(99.8)	Y/N
Q9LXC7	At5g09630	386	LisH/CTLH/CRA/RING		TPR2(99.6)/Smu1(99.7)	Y/N
Q9ZQ45	At2g22690	381	LisH/CTLH/CRA/RING		TPR2(99.6)/Smu1(99.4)	N/N
Q9T075	At4g37880	388	LisH/CTLH/CRA/RING		TPR2(99.6)/Smu1(99.4)	N/N
Q9M2V9	At3g55070	418	LisH/CTLH/CRA/RING		TPR2(99.6)/Smu1(99.3)	N/N
F4HYD7	At1g35470/RanBPM	467	SPRY/LisH/CTLH/CRA		TPR2(99.8)/Smu1(99.8)	Y/N
Q8RX25	At4g09200	397	SPRY/LisH/CTLH/CRA		TPR2(99.8)/Smu1(100)	Y/N
Q8LF14	AT4G09340	429	SPRY/LisH/CTLH/CRA		TPR2(100)/Smu1(99.8)	Y/N
Q8GX44	At1g11110	227	LisH/CTLH/CRA		TPR2(99.9)/Smu1(99.9)	N/N
Q8GYP0	At4g09300	224	LisH/CTLH/CRA		TPR2(99.9)/Smu1(99.9)	Y/Y
A8MQF1	AT1G61150	226	LisH/CTLH/CRA		TPR2(100)/Smu1(100)	Y/N
Q9LNE1	At1g06060	213	LisH/CTLH/CRA		TPR2(99.9)/Smu1(99.9)	N/N
F4K250	At5g66810	750	CTLH/CRA/CTLH/CRA		TPR2(98.9)/Smu1(98.0)	Y/N
Q9SMS2	AT4g09330	111	CRA			
O48847	AT2G32700/LUH	787	LisH/WD40	Repression	LisH: TBL1	
Q9FU2	AT4G32551/LEUNIG	931	LisH/WD40	Repression	LisH: TBL1	
Q9M086	At4g31160/DCAF1	1883	LisH/WD40	Ubiquitination	LisH: TBL1	
Q9FN19	AT5g67320/HOS15	613	LisH/WD40	histone deacetyl.	LisH: TBL1	
Q9LU74	AT5g57120/MUL3	330	LisH/SRP40		LisH: Lis-1	
Q8VYW7	AT5g16210/T21H19	1180	LisH/LisH/Armadillo/Armadillo		LisH: TBL1	
Q9FQ24	AT3G55005/TON 1b	257	LisH	microtubule org.	LisH: TBL1	
Q9FQ25	AT3G55000/TON 1a	260	LisH	microtubule org.	LisH: TBL1	
<b>Examples of proteins from other eukaryotes</b>						
<i>Chlamydomonas reinhardtii</i>						
A8HQD2	CRDRAFT-136822	232	LisH/CTLH/CRA		TPR2(100)/Smu1(100)	Y/N
<i>Homo sapiens</i>						
Q96S59	Hs708182	729	SPRY/LisH/CTLH/CRA	Adapter protein	TPR2(99.6)/Smu1(99.6)	Y/N

**Table S5:** Model of LisH domain containing proteins in *A. thaliana*. Uniprot number, ATG number, length (aa) and predicted domains are indicated. Modeling was performed with PHYRE2 (Kelley et al., 2015). All LisH-CTLH (with an extra WD40) and LisH-CTLH-CRA can be modeled with the N-ter of OsTPR2 or *Caenorhabditis elegans* Smu1 as template with a confidence between 98.0% and 100% (indicated with brackets). These proteins are highly similar to the N-terminus of TPL/TPR2 or Smu1. Most of them contain a hydrophobic groove 3 (G3: Y) found in TPL/TPR2 whereas few contain a hydrophobic groove 4 (G4: Y) specific of Smu1, both potentially involved in protein binding. Some of the modeled proteins contain also a potential tetramer interface (not shown). Proteins predicted to contain only a LisH domain were modeled with the LisH domain of TBL1 or Lis-1 as

template. Two examples of proteins from *C. reinhardtii* and *H. sapiens* shown in Fig. S1 are also indicated.

Name	Strand	Sequence	Experiment
TPL(FL)	Forward	5' CACCATGTCTTCTCTTAGTAGAGAGCTC 3'	Repression assay
	Reverse	5' TCTCTGAGGCTGATCAGATGC 3'	
TPL202 (and mutant)	Forward	5' CACCATGAAACATCACCATCACCATC 3'	Repression assay
	Reverse	5' TTAAGAGTGATCCACAAAAAGATC 3'	
IAA12-DI	Forward	5' CACCATGCGTGGTGTGTGTCAGAATTG 3'	Y2H
	Reverse	5' TCAACTGTTCATCCTGTGTAACCCA 3'	
WUS	Forward	5' CACCATGGAGCCGCCACAGCATC 3'	Y2H
	Reverse	5' CTAGTTCAGACGTAGCTCAAGAGAAG 3'	
AtTPL202	Forward	5' GGCGCCATGGGCTCTTCTCTTAGTAGAGAGCTC 3'	HTRF, Anisotropy, SEC-MALLS
	Reverse	5' CGTGCGGCCGCTTAAGAGTGATCCACAAAAAGATC 3'	
AtTPL202 m/ L130A	Forward	5' AACCTCCGGGAGAATGAACAGGCCTCCAAGTATGGGGACACCAAG 3'	HTRF, Anisotropy, SEC-MALLS
	Reverse	5' CTTGGTGTCCCATACTTGAGGCCTGTTTCATTCTCCCGGAAGTT 3'	
AtTPL202 m/ Y133A	Forward	5' TGCAGACTTGGTGTCCCCAGCCTTGGACAGCTGTTTCATTCTCCCGGAAG 3'	HTRF, Anisotropy, SEC-MALLS
	Reverse	5' CTTCCGGGAGAATGAACAGCTGTCCAAGGCTGGGGACACCAAGTCTGCA 3'	
AtTPL202 m/ T166A	Forward	5' GTTCTCCAATGTCAACAGCTGTGCTATTTTCTTGAAAAGCTCCTCATT 3'	HTRF, Anisotropy, SEC-MALLS
	Reverse	5' AATGAGGAGCTTTTCAAGGAAATAGCACAGCTGTTGACATTGGAGAAC 3'	
AtTPL202 m/K102S	Forward	5'GCTCCTCATTAAAAGTTGAAAACACTGACAGATCTTTCAGTATATCCACA GCCTTGGGAC 3'	HTRF, Anisotropy, SEC-MALLS
	Reverse	5'GTCCAAGGCTGTGGATATACTAGTGAAAGATCTGTCAGTGTTCACCTTTT AATGAGGAGC 3'	
AtTPL202 m/Q117S- E122S	Forward	5'GTTTCATTCTCCCGGAAGTTCGACAATGTCAACAGCGATGTTATTTCTTGAA AAGCTCCTCATTAAAA 3'	SEC-MALLS
	Reverse	5'TTTTAATGAGGAGCTTTTCAAGGAAATAACATCGCTGTTGACATTGTCGAACT TCCGGGAGAATGAAC 3'	
AtTPL202 m/R172S	Forward	5'-CTTTAGAAATTC AAGGCTGTGCGACTTTGATCAACCAGAGCT-3'	SEC-MALLS
	Reverse	5'-AGCTCTGGTTGATCAAAGTCGACAGCCTTGAATTTCTAAGAG-3'	
qCTL1	Forward	5'- AGTGGAGAGGCTGCAGAAGA-3'	qPCR
	Reverse	5'- CTCGGGTAGCACGAGCTTTA-3'	
qARF19	Forward	5'- CACCGATCACGAAAACGATA-3'	qPCR
	Reverse	5'- TGTTCTGCACGCAGTTCAC-3'	
qlAA14	Forward	5'- CAAAGATGGTACTGGATGC-3'	qPCR
	Reverse	5'- GCATGACTCGACAAACATCG-3'	

**Table S6:** DNA probes used for plasmid constructions.

## SI Materials and Methods

### Vectors construction

AtTPL184 cDNA was directly cloned from an AtTPL full-length synthetic coding sequence (Thermo Fisher). AtTPL202 cDNA was obtained by PCR from AtTPL full-length clone using oligonucleotides listed in Table S6. Both constructs were cloned into pETM11 plasmid (EMBL) for the production of N-terminal His-tagged proteins. Mutants were constructed either by site-directed mutagenesis using oligonucleotides referenced in Table S6 or by a restriction-site strategy with synthetic DNA (GeneCust) including the desired mutation.

N-terminal MBP-tagged AtTPL202 wild-type (wt) and mutated constructs were obtained by cloning the corresponding sequences into pETM40 plasmid (EMBL).

IAA12 cDNA was cloned into plasmid pETM33 (EMBL) for the production of full-length N-terminal GST-tagged proteins.

For Yeast-2-Hybrid (Y2H) and repression assays, wt and mutated AtTPL202, IAA12-DI and WUSCHEL coding sequences were amplified with specific primers (Table S6) and cloned into pENTRD-TOPO<sup>®</sup> using CACC strategy. To test in Y2H assay the interactions between AtTPL202 and repressors, AtTPL202 and its mutated versions cDNAs were cloned into bait vector pGBKT7 (Clontech), whereas IAA12-DI and WUSCHEL coding sequences were cloned into prey vector pACT2 (Clontech).

For repression assays, p35S::AtTPL-mCherry expression constructs were obtained by triple LR reactions performed with Gateway<sup>®</sup> recombination technology on the 35S promoter, AtTPL FL, or AtTPL202 wt and mutants cDNAs, and the transfection marker mCherry in the destination vector pDESTR4-R3.

For in planta assays, p35S::AtTPL-mCherry expression constructs were obtained by triple LR reactions performed with Gateway<sup>®</sup> recombination technology on the 35S promoter, AtTPL FL, or AtTPL202 wt and the fluorescent marker mCherry in the destination vector pB7m34 with basta resistance as a selectable marker. Arabidopsis Col-0 plants were transformed with the constructs and basta-resistant T2s were used in all experiments.

## Protein expression and purification

All proteins were expressed in *Escherichia coli* Rosetta 2 strain. Bacteria cultures were grown at 37°C until they achieved an OD<sub>600nm</sub> of 0.6-0.9. Protein expression was induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 400 μM at 18 °C overnight. Bacteria cultures were centrifuged and the pellets were resuspended in the buffers as indicated below, where cells were lysed by sonication.

His-tagged AtTPL184 and AtTPL202 (wt and mutants) bacteria pellets were resuspended in buffer A (CAPS 200 mM pH 10.5, NaCl 500 mM, TCEP 1 mM) with EDTA-free antiprotease (Roche). The soluble fractions recovered after sonication were passed through a Ni-sepharose (GE Healthcare) column previously washed with buffer A and the bound proteins were eluted with buffer A with 300 mM imidazole. A second purification step was carried out on Gel filtration Superdex 200 16/60 (GE Healthcare) equilibrated with buffer A.

Production of selenomethionine (Se-Met) TPL184 for crystallography was done in *E. coli* B834 (DE3) (met-) strain (Novagen) transformed with the AtTPL184 plasmid. Bacteria cultures were grown at 37 °C under agitation up to an OD<sub>600nm</sub> of 0.3. Next, bacteria cultures were recovered by centrifugation and the pellets were washed twice with 100 mL of M9 minimal medium (NaH<sub>4</sub>Cl 37 mM, KH<sub>2</sub>PO<sub>4</sub> 44 mM and Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O 180 mM). Bacteria cultures were re-started after this at 37 °C until an OD<sub>600nm</sub> of 0.6 in 1 L of 2X M9 minimal medium supplemented with MgSO<sub>4</sub> (2 mM), FeSO<sub>4</sub> (25 mg/L), glucose (4 g/L), vitamins (thiamine, pyridoxine, riboflavin and niacinamide at 1 mg/L), and mix of all amino acids (895 mg/L) except methionine. The expression of the Se-Met protein was induced by addition of 400 μM IPTG and Se-Met (40 mg/L). The protein was further purified in buffer B (CAPS 200 mM pH 10.5, NaCl 1 M, TCEP 1 mM) with EDTA-free antiprotease (Roche) as described above for AtTPL184.

MBP-tagged IAA12 bacteria pellets were resuspended in buffer C (Tris-HCl 20 mM pH 8, TCEP 1 mM) with EDTA-free antiprotease (Roche). Protein purification was done from the soluble fraction in an amylose-resin (GE Healthcare) column previously equilibrated with buffer C and from which bound proteins were eluted with maltose 10 mM diluted in buffer C. The same procedure was followed for MBP-tagged AtTPL202 wt and mutant proteins using in this case buffer D (Tris-HCl 20 mM pH 7.4, NaCl 100 mM, EDTA 1 mM, DTT 1 mM) with EDTA-free antiprotease (Roche).

GST-tagged IAA12 bacteria pellets were resuspended in buffer E (Tris-HCl 10 mM pH 8, DTT 10 mM, EDTA 1 mM, Triton 0.1%) with EDTA-free antiprotease (Roche). Soluble fractions were incubated 15 h under slow rotation speed with glutathione–resin previously equilibrated with buffer E at 4°C. Proteins were eluted with glutathione 40 mM pH 7.4 in buffer E.

Proteins dialyses were performed 15 h at 4 °C in their purification buffers and proteins were aliquoted, frozen in liquid nitrogen and stored until use at -80 °C.

## Crystallization

Initial crystallization conditions were identified using the high-throughput crystallization platform at EMBL Grenoble ([embl.fr/htxlab](http://embl.fr/htxlab)). The optimum condition for crystals of native and Se-Met TPL184 was obtained at 20 °C with the hanging drop vapour diffusion method by mixing 1 µL of protein (AtTPL184 70 µM in CAPS 133.3 mM pH 10.5, NaCl 333.3 mM, 0.66 mM TCEP and hexanediol 10% v/v) with 1 µL of reservoir solution (Na<sub>2</sub>HPO<sub>4</sub>-Citrate 0.2 M pH 4.2, 2-propanol 10% v/v and Lithium sulfate 0.3 M). The optimum condition for crystallization of the complex AtTPL184/ IAA27 EAR motif was obtained at 20 °C with the same method by mixing 1 µL of protein (AtTPL184 140 µM, IAA27 EAR motif 1.4 mM, CAPS 20 mM pH 10.5, NaCl 50 mM and TCEP 0.1 mM) with 1 µL of reservoir solution (Di-Ammonium tartrate 1.08 M pH 7 and 2% benzamidine-HCl). Crystals were cryoprotected by plunging them into liquid nitrogen, after incubation in crystallization solution supplemented with 20% glycerol.

## Protein structure determination

Se-Met protein crystallized in space group P43212 with unit cell dimensions of a=b=99.2 Å and c=290.2 Å. A SAD experiment was performed on the ESRF beamline ID29 (9) (Table 1) at the peak of the Se absorption (12.751 keV). All data were reduced in XDS (10) within the Grenades (11) processing pipeline. The heavy atom substructure was determined and iteratively improved in SHELXC/D/E. This substructure was then refined in AUTOSHARP (12) to anomalous phasing power of 3.019 in the inner shell (58.5-12.78Å) and which were greater than 1 to 4.93 Å.  $R_{\text{cullis}}$  in the inner resolution shell was 0.431 and remained below 1 to 3.08 Å. These phases were used to automatically build a model with four protomers in the asymmetric unit: 770 residues of the 848 expected. This model was partially refined and then used as a search model for molecular replacement into a higher resolution native dataset (P3<sub>1</sub>12, a=b=72.681 Å, c=181.147 Å, 2 protomers in the

asymmetric unit, 2.6 Å resolution). Rotational and translational Z-scores of 7.1, 7.6, 12.9 and 47.2. The model was then improved through multiple rounds of manual rebuilding and refinement in COOT (13) and BUSTER (14). Data for the peptide bound structure were collected on the ESRF beamline ID23-EH2 (15). The final apo model was then used for molecular replacement to determine the structure of the IAA-peptide bound structure to 1.95 Å resolution (P43212, a=b=94.1 Å, c=298.0 Å, four protomers in the asymmetric unit). This model was completed in the same manner as the apo structure. Peptides were visible in all four binding grooves, however the electron density was significantly better in subunits A and B compared to C and D. Re-analysis of the 5C7F data was performed as follows: We applied random coordinate shifts to the 5C7F model with Phenix.dynamics to reduce bias (16), then rebuilt the peptide chain in a shifted register with COOT (13) and then refined the structure in BUSTER (14). PHENIX phenix.composite\_omit\_map (17) was run for both sets of data (our own and the data downloaded for 5c7f), omitting the peptide chains and using default parameters. Figures were made in Pymol.

### **Native molecular mass determination**

Molecular masses were determined by Size-Exclusion Chromatography-Multi Angle Light Scattering (SEC-MALLS) on an analytical Superdex-S200 increase (GE Healthcare) connected to an in-line MALLS spectrometer (DAWN HELEOS II, Wyatt Instruments). Analytical size exclusion chromatography was performed at 25°C at a rate of 0.5 mL/min in buffer A for AtTPL184 and AtTPL202 (wt and mutants) and in buffer F (CAPS 20 mM pH 10.5, Tris-HCl 100 mM pH 8.8, NaCl 50 mM, TCEP 0.1 mM) for AtTPL202-EAR motifs complexes (fluorescence anisotropy conditions). The refractive index was measured with in-line refractive index detector (Optirex, Wyatt Instruments) was used to follow the differential refractive index relative to the solvent. Molecular masses calculation was done with the Debye model using ASTRA version 5.3.4.20 (Wyatt Instruments) and a theoretical dn/dc value of 0.185 mL/g.

### **Yeast two hybrid interaction tests**

The vectors pGBKT7 and pACT2 were respectively transformed into yeast strains Y187 and AH109 (Clontech) using standard protocol (18). The analyses were performed after mating of the two yeast strains. Interactions were assessed using β-Gal activity (19). To do so, OD<sub>405nm</sub> were measured during exponential growth phase after incubation at 30°C

in Buffer G (100 mM Phosphate buffer pH 7, 10 mM KCl, 1 mM Mg<sub>2</sub>SO<sub>4</sub>, β-mercaptoethanol 50 mM) with ONPG.

### Homogeneous-Time Resolved Fluorescence (HTRF) interaction tests

His-tagged AtTPL202 wt/mutants and MBP-tagged IAA12 interactions were analysed by HTRF (20) using CisBio Bioassays Anti-His acceptor d2 and Anti-MBP donor Tb antibodies.

HTRF experiments were performed on Greiner 384 Flat bottom wells plate. Three simultaneous replicas were done for each binding mixture. After a 2h-incubation at room temperature in the dark, the binding reactions were excited at 337 nm and emission measurements were taken at 620 nm and 665 nm with a Tecan infinite M1000PRO. HTRF specific signal was calculated as follows:

$$HTRF = \frac{Sample_{665nm}}{Sample_{620nm}} - \frac{Blank_{665nm}}{Blank_{620nm}}$$

For HTRF competition assays, an initial GST-IAA12 (200 nM) – MBP-AtTPL202 (500 nM) complex was formed using CisBio Bioassays Anti-GST donor Tb and Anti-MBP acceptor d2 antibodies. The complex formed was competed by adding increasing amounts of His-tagged AtTPL202 wt and mutant proteins. Three independent replicas were done for each binding mixture and the measurements were done in the same conditions as before.

IC<sub>50</sub> were calculated fitting the competition curves to the following equation:

$$HTRF = \frac{HTRF_{max}}{1 + \frac{[AtTPL]}{IC_{50}}} + HTRF_{min}$$

### Fluorescence anisotropy interaction tests

Fluorescence anisotropy interaction assays were done with wt and mutant IAA12 EAR peptides (GeneCust, Table S4) tagged to a fluorescein amidite fluorophore (FAM) at 100 nM. Increasing concentrations of AtTPL202 and mutants were added in buffer F. Three replicas per binding mixture were performed in Corning 384 Flat Bottom Black Polystyrene plates. Samples were incubated in the dark for 2 hours at room temperature. Anisotropy measurements were done with an excitation wavelength of 470 nm and an emission wavelength of 519 nm with a Tecan infinite M1000PRO.

EC<sub>50</sub> were calculated fitting the binding curves using the following equation:

$$Anisotropy = \frac{Anisotropy_{sat}[AtTPL]}{EC_{50} + [AtTPL]} + Anisotropy_0$$

### Repression assays

To test the repression capacity of AtTPL202 and its mutated versions *in planta*, we used leaf mesophyll protoplasts from *Arabidopsis thaliana* (Col-0) containing the integrated reporter DR5::VENUS. Plants were grown 3 to 5 weeks in short-day condition and protoplasts were isolated and transfected using sandwich tape method and PEG method respectively (21). Protoplasts were incubated overnight in continuous light in presence or absence of 1 μM NAA. Pictures were taken with a confocal microscope Zeiss LSM710 and the fluorescence was measured using Fiji (Image J software). Statistical analysis was tested using non-parametric Kruskal-Wallis test performed with R ([http:// www.r-project.org](http://www.r-project.org)).

### Root growth and lateral root density analyses

Seeds were surface sterilized for 5 minutes in 50 % bleach, 0.1 % triton X-100 then washed three times with sterile ddH<sub>2</sub>O. Seeds were stratified at 4° for 2 days to synchronise germination. 1-week old seedlings of the indicated genotypes were transferred from normal ½ MS plates to medium complemented or not with auxin (NAA 1 μM) for 5 additional days. Root growth after transfer and lateral root density were determined using Fiji (Fiji Is Just ImageJ v1.51e).

### RT qPCR

1-week old seedlings of the indicated genotypes were transferred from normal ½ MS plates to medium complemented or not with auxin (NAA 1 μM) for 2 hours. RNA was extracted from the whole root using the Spectrum Plant Total RNA kit (Sigma). Poly(dT) cDNA was prepared from 500 ng of total RNA with Superscript III reverse transcriptase (Invitrogen) and analyzed on a StepOnePlus apparatus (Life Technologies) with the SYBR® Green PCR Master Mix (Applied Biosystem) according to the manufacturer's instructions. Targets genes were quantified with specific primer pairs designed with the Universal Probe Library Assay Design Center (Roche Applied Science) (Table S6). All reactions were done in quadruplicate and expression levels were normalized to At1G04850 (CTRL1).

## Confocal microscopy

Overexpression lines of TPL-mCherry fusions were imaged on an upright Leica SP5 confocal laser scanning microscope (Leica Microsystems, Germany). Scanner and detectors settings used at the beginning of one experiment were optimized to avoid saturation and to maximize resolution and kept unchanged throughout the experiment. mCherry was excited using a 561 nm laser and fluorescence was collected from 590 to 680 nm (using the AOBS of the SP5).

## SI References

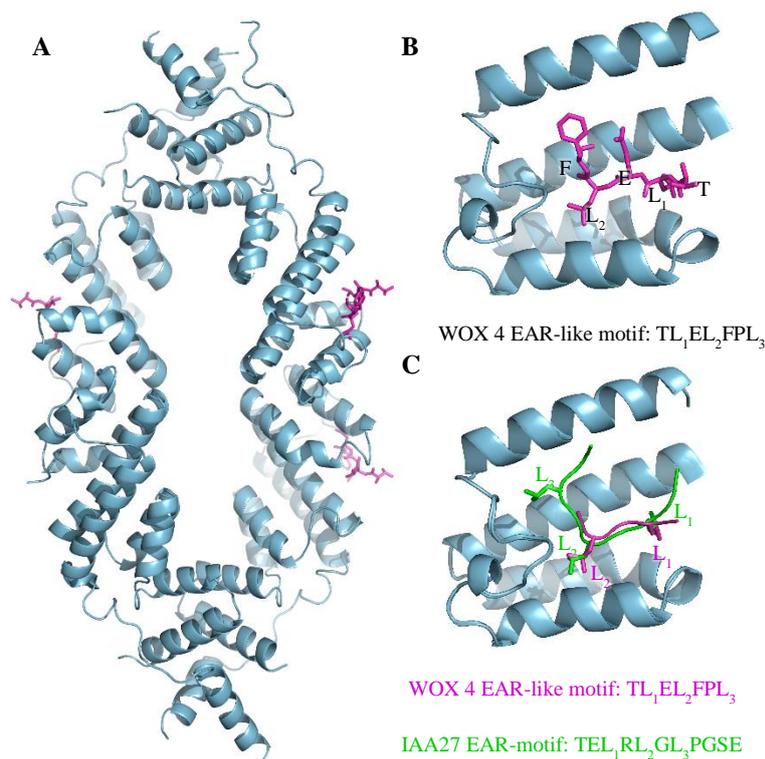
1. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protoc* 10(6):845–858.
2. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881–10890.
3. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42(Web Server issue):W320–4.
4. Ulrich AKC, Schulz JF, Kamprad A, Schütze T, Wahl MC (2016) Structural Basis for the Functional Coupling of the Alternative Splicing Factors Smu1 and RED. *Structure*:1–12.
5. Ke J, et al. (2015) Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. *Sci Adv* 1(6):e1500107.
6. Oberoi J, et al. (2011) Structural basis for the assembly of the SMRT/NCOR core transcriptional repression machinery. *Nat Struct Mol Biol* 18(2):177–184.
7. Chodaparambil JV, et al. (2014) Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J* 33(7):719–731.
8. Matsumura H, et al. (2012) Crystal structure of the N-terminal domain of the yeast general corepressor Tup1p and its functional implications. *J Biol Chem* 287(32):26528–26538.
9. de Sanctis D, et al. (2012) ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering. *J Synchr Rad* 19(Pt 3):455–461.
10. Kabsch W (2010) XDS. *Acta Crystallogr D* 66(2):125–132.
11. Monaco S, et al. (2013) Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. *J Appl Crystallogr* 46(Pt 3):804–810.
12. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–230.
13. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60(Pt 12 Pt 1):2126–2132.
14. Bricogne G, et al. (2011) *BUSTER* (Cambridge, United Kingdom: Global Phasing Ltd.).
15. Flot D, et al. (2010) The ID23-2 structural biology microfocus beamline at the ESRF. *J Synchr Rad* 17(1):107–118.
16. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66(2):213–221.

17. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* 68(Pt 4):352–367.
18. Daniel Gietz R, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Guide to Yeast Genetics and Molecular and Cell Biology - Part B, Methods in Enzymology*. (Elsevier), pp 87–96.
19. Smale ST (2010) Beta-galactosidase assay. *Cold Spring Harbor Protocols* 2010(5):pdb.prot5423–pdb.prot5423.
20. Degorce F (2006) HTRF(®): pioneering technology for high-throughput screening. *Expert Opin Drug Discov* 1(7):753–764.
21. Wu F-H, et al. (2009) Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods* 5(1):16.

## Complementary results and discussion

### Groove 3, more than LxLxL

Although LxLxL-type EAR motifs are the ones most frequently found among TPL interactors, other types of motifs have been also proven to mediate TPL recruitment (Causier et al., 2012a). Co-crystallization was tried for AtTPL184 in complex with DLNxxP, (M/K/R)LFG and TLxLFP-type EAR-like motifs in order to know if different types of peptides could use the same binding mechanism as LxLxL or not. Only AtTPL184-WOX4 peptide (type TLxLFP) co-crystallization gave us some answers. Crystals of the complex revealed that TLxLFP peptides can bind the same groove as LxLxL motifs, here designated as Groove 3 (Fig 18.A-B). Moreover, the position of the leucine residues was overlapped for both peptides when comparing the two complexes structures (Fig 18.C). Unfortunately, the resolution obtained for this structure (3.2Å) was not high enough for a precise orientation of the peptide inside the groove.



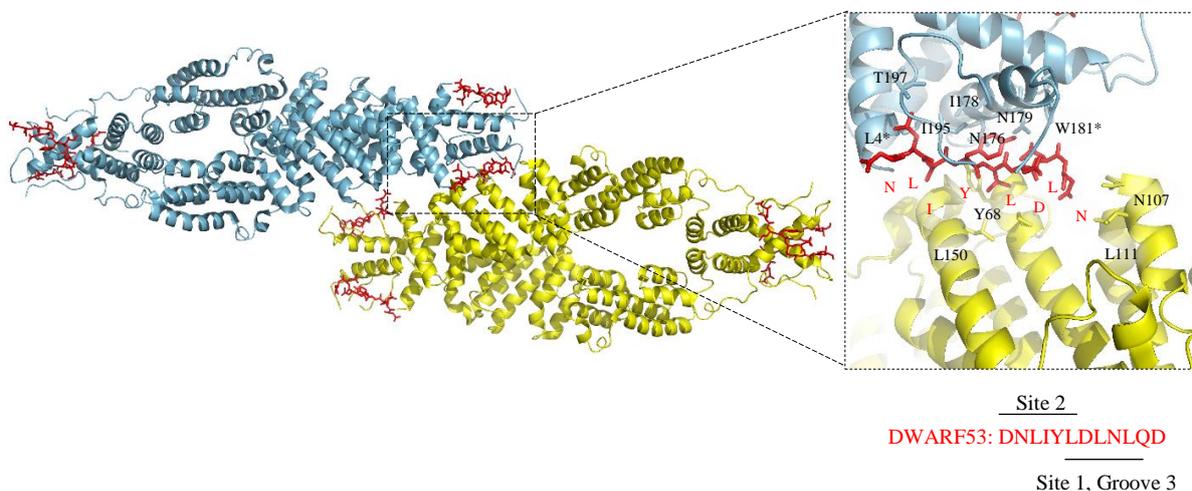
**Fig 17. WUS box EAR-like motifs lodged in the same Groove 3 as LxLxL-type EAR motifs.** **A.** AtTPL184 tetramer (blue) in complex with WOx4 TLxLFP EAR-like motif (magenta). The peptide could only be found in 3 out of the 4 binding sites of the TPL tetramer. **B.** Close-up view of TLELF sequence (WOx4 EAR-like motif) inside AtTPL184 groove 3 in one of the two possible orientations. **C.** WOx4 (magenta) and IAA27 (green) peptides superimposition inside AtTPL184 groove 3. Leucine residues occupy the same positions inside the groove in the two complex structures

### Complementary discussion

Whereas transcription factors interaction with a naked genome seem to be enough for the “simple” and “ancient” prokaryotes, eukaryotes developed more complicated systems that helped solving the

transcriptional complexity of their lives. Co-repressor proteins are an example of this and a very particular one, in fact, due to the contrast that they represent. In opposition to the vast number of TFs that eukaryotes possess, we find only a few co-repressor families, and this few are enough to deal with all transcriptional repressors. The reason for this is their “hub” nature that confers co-repressors (co-regulators in general) the ability to interact with a diversity of transcription factors. The question that inevitably rises to this numeric contrast is how a single protein faces such a diversity of interactions.

Our TPL protein does this job by interacting with conserved sequences of peptides named EAR or EAR-like motifs, for which several categories have been described depending on their amino acidic sequence. The crystallographic and biochemical data presented in this chapter proves that two types of these motifs (LxLxL and TLxLFP) can be lodged by the same hydrophobic groove within TPL N-ter, indicating that this groove greatly contributes to TPL “hub” activity. However, according to recent publications, other regions within TPL N-ter structure might also participate in the interaction with EAR-motifs (Ma et al., 2017). Ma *et al.* suggest that a peptide of the rice DWARF53 strigolactones repressor could bind simultaneously two TPL tetramers through two sites of OsTPR2, one in each tetramer. Site 1 corresponds to our groove 3 (interacting with more external residues) and site 2 is located in the C-terminal of the N-ter (Fig 18).

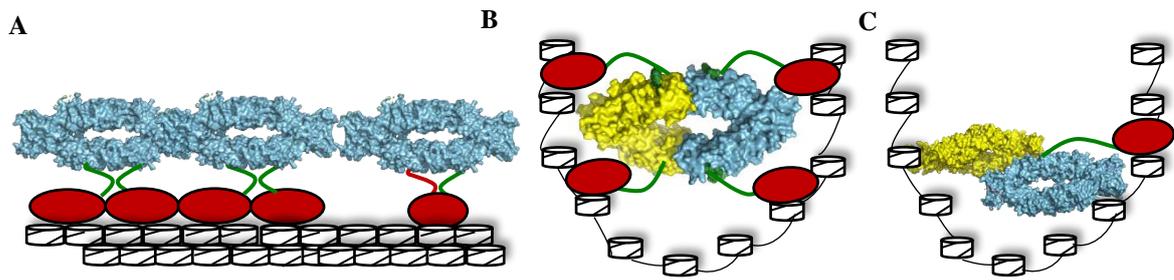


**Fig 18. DWARF53 bipartite EAR motif bound to OsTPR2.** Left, OsTPR2 octamer (one tetramer in blue and the other in yellow) mediated by DWARF53 (red) simultaneous interaction with two OsTPR2 tetramers. Right, close-up view of the two interaction interfaces. DWARF53 interaction with site 2 is mediated by residues of two different chains. Residues of the opposite chain are marked with asterisks. Interaction with site 1 (terminology used by Ma et al. that corresponds to our groove 3) happens in a different manner than for other LxLxL EAR motifs (Ke et al. 2015; Martín Arevalillo et al. 2017) with the interacting residues more towards the outside part of the groove. However, the C-terminal leucine of DWARF53 was not determined in the complex structure, so it could be positioned inner inside the groove (Figure adapted from Ma. et al. 2017. 5J9K pdb file).

Simultaneous binding to two sites was a specific feature of this motif, that presents a “bipartite” nature: the N-ter binds the second site and the C-ter (LxLxL sequence) binds the so-named groove 3. Therefore, TPL N-terminus structure seems to hide the secrets of its promiscuity, with several interacting regions that can bind different EAR-motifs in different manners.

TPL N-terminus is amazingly conserved since m.y.a and it is composed by three domains: LiSH, CTLH and CRA, newly described here. CTLH/CRA domains constitute groove 3. On the other extreme, two WD40 repeats domains form the C-terminal region. LiSH and WD40 domains are shared by other co-repressor proteins, such as TBL1. In TBL1 the LiSH domain mediates dimerization and tetramerization of the protein. OsTPR2 and AtTPL184, as TBL1 and other co-repressors, are also tetramers. However, in TPL, whilst a first dimerization is also mediated by the LiSH it is the CRA domain that forms the second tetramerization interface. Tetramerization has been implicated in other co-repressors in chromatin compaction and binding to histones. We showed that for TPL, tetramerization is necessary for conferring EAR motifs binding site the appropriate conformation for a good affinity binding. However, this does not exclude other functions for TPL tetrameric structure. Having 4 monomers means having 4 binding sites. We can then imagine TPL tetramer lying along chromatin by its simultaneous binding to oligomeric TFs such as Aux/IAAs or ARFs or to TFs with more than one EAR-motif, which would afford an easier spreading of repressive marks on the genome. But we can also picture TPL tetramer interacting with TFs bound on distal regions in the genome, inducing the formation of loops that would contribute to chromatin compaction (Fig 19).

Recent findings make this picture more complicated. Ma *et al.* proposed that first, as it is also the case for other co-repressor proteins, TPL N-ter binds histones. Moreover, they proposed that upon binding of “bipartite” EAR-motifs to two simultaneous sites, TPL could adopt an octameric conformation that increases the binding to nucleosomes (Ma et al., 2017). Although the octameric nature of this complex has not been properly demonstrated by mutations and SEC-MALLS MW determination, these studies lead us to think that higher-order repression complexes can be formed in a very stable manner with TPL oligomers anchored to nucleosomes. Moreover, it implies that the nature of the EAR-motif could be determinant for the intensity and stability of the transcriptional repression.



**Fig 19. Models for different TPL oligomers modes of action.** **A.** Simultaneous binding of tetrameric TPL (blue) to oligomeric proteins or proteins with several EAR motifs could induce TPL spreading along chromatin. **B.** TPL tetramers (blue and yellow) binding to TFs positioned in different regions could help chromatin compaction by loops formation. **C.** (Adapted from Ma. et al. 2017). Certain TFs can induce TPL oligomerization, that is hypothesized to stabilize its binding to nucleosomes.

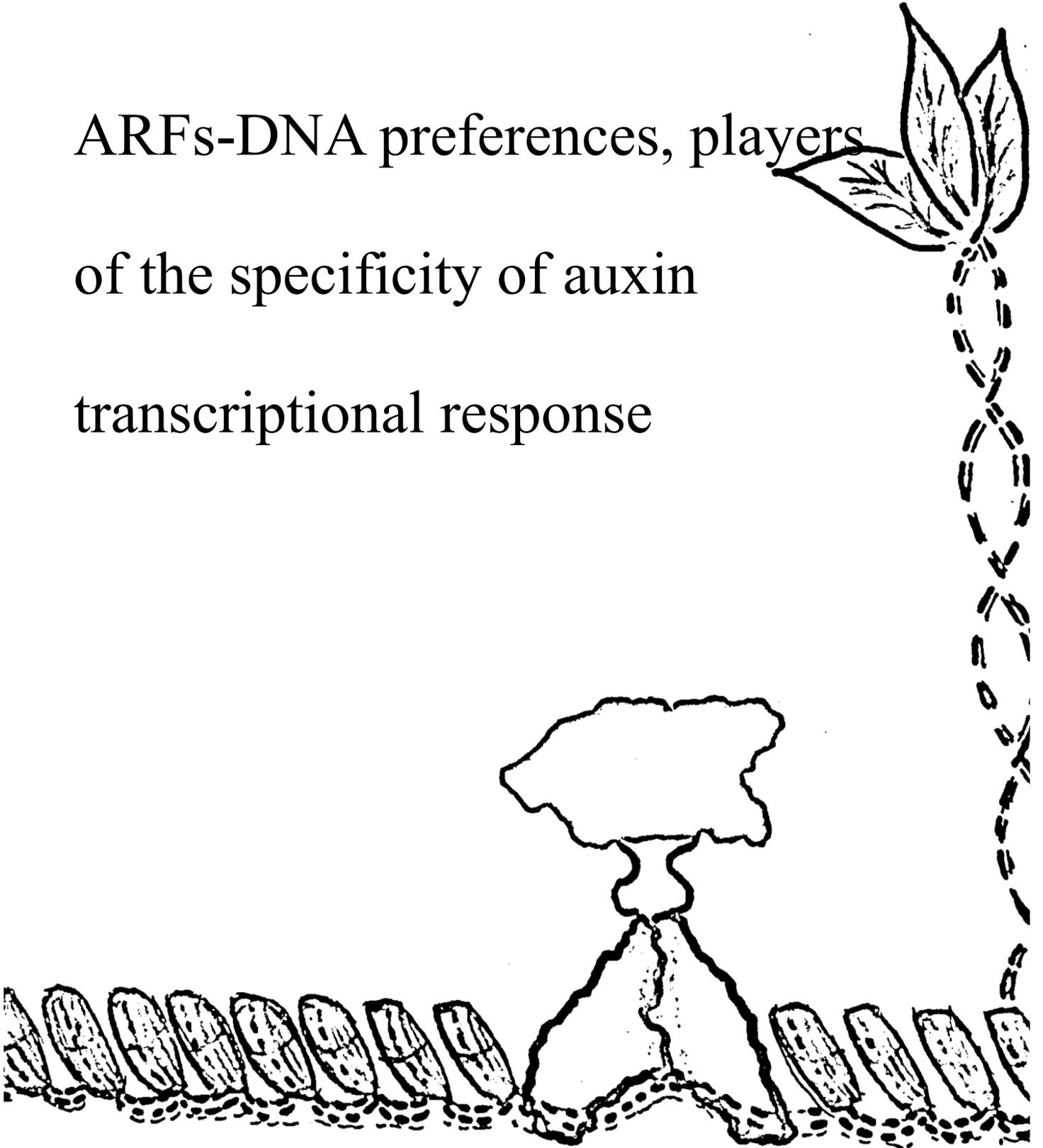
## Conclusions

Eukaryotes invented co-repressors as an extra level of regulation of gene expression and they seemed to do it in a very efficient and simple way. By using a common set of domains or structural properties, co-repressors of very different organisms manage to bind a diversity of transcription factors to trigger repression through a common mechanism, chromatin compaction. Getting into the heart of TPL structure helped us to partially understand how this co-repressor acts a hub and how its oligomeric nature and recruitment mode might contribute to chromatin state. These studies could give us clues for a better understanding of other co-repressors that present structural similarities with TPL.



# Chapter II

ARFs-DNA preferences, players  
of the specificity of auxin  
transcriptional response





## Introduction

Let's think of a plant. Roots, shoot, leaves and sometimes nice flowers. Roots, organs specialized in the uptake of nutrients and water from the soil; leaves, the machinery of photosynthesis; shoots, maintaining the whole structure, sometimes meters and meters high; or flowers, reproductive organs that are formed earlier or later depending on the environmental conditions. And still, no matter where the plant cell comes from, the genetic information inside is the same (or almost the same). So, how does it happen that different cells from different organs behave differently under different stimuli? It all relies in gene regulation, the differential expression of a set of genes under a certain condition. And this, it can be specific for each cell.

Although gene regulation was already addressed in the previous chapter for a mechanism typical of eukaryotes, the co-regulators (co-repressors in our case), in Chapter II we will focus on a system of genetic control also shared by prokaryotes: the interplay DNA-Transcription Factor (TFs).

### DNA-TFs interplay in the control of gene expression

The expression or repression of a certain set of genes is a game with two main players: the DNA regulatory sequences located inside the promoters of the regulated genes, and the transcription factors that can bind to them. A game that is played according to three main rules: DNA accessibility, DNA-TF specificity and TF activator or repressor identity.

In eukaryotes, chromatin state will be determinant for the access of a TF to its DNA binding site. Therefore, transcriptional regulation (activation or repression) is accompanied by chromatin remodelling that will render chromatin opened or closed. In this sense, transcription factors can act by recruiting co-regulator proteins that, as already reviewed in the previous chapter, can alter the state of chromatin. A very special group of TFs, pioneer TFs, can do both jobs: pioneers have not only the capacity to bind DNA, but also to make it more accessible by nucleosomes shifting (Zaret Kenneth S.; Mango, 2017).

Once allowed by chromatin state, the expression of a certain gene will be now controlled by the TFs bound to its promoter. TFs-DNA interaction is mediated by one or more DNA binding domains (DBDs). Additional regions normally accompany DBDs: domains that mediate homo or heterotypic interactions with other transcription factors (dimerization or oligomerization domains) or regions that can recruit co-regulator proteins (Smith and Matthews, 2016). These other domains, although they lack DNA-binding capacity, they can have some influence upon the interaction with the DNA (Sayou et al., 2016).

TF-DNA interactions take place between amino acids within the DBD and nucleotides (Crocker et al., 2016) in both a non-specific and specific manner (Smith and Matthews, 2016). Non-specific interactions occur with the DNA-backbone and provide stability rather than specificity. Specificity is driven by the so-called “base-readout”, base-amino acid interactions mainly with charged and polar amino-acidic residues (Luscombe et al., 2001). TFs can also recognize DNA structural features, such as 3D DNA conformation or flexibility of the DNA, that determine what is known as the “shape readout”.

According to the “base-readout” specificity, for a TF there will be one optimal Binding Site (BS) versus a diversity of less favourable nucleotides combinations that can also be bound by the same TF but with lower affinity. Genome-wide analysis indicates that whereas the most favourable DNA binding sequences are rare inside the genome, clusters or continuum of less favourable BSs are rather the rule (Crocker et al., 2016).

Frequently, eukaryotic TFs act in a cooperative manner to efficiently bind to these lower-affinity sites. They do so by binding as dimers or oligomers that allow simultaneous binding to multiple sites, increasing the length of the binding site and thus the specificity of the binding (Berg and von Hippel, 1987; Marianayagam et al., 2004; Morgunova and Taipale, 2017). The formation of these complexes can also take place in a heterotypic way with TFs involved in other processes. This, apart from conferring specificity to the binding, provides a hub for the integration of different signals and adds a possible extra level of regulation by the formation of complexes which activity can be easily modulated (Marianayagam et al., 2004; Smith and Matthews, 2016).

Under the stimulus “a” inside a cell, the combination of TFs present in it at a specific moment, their specificity for BSs and the accessibility of these BSs within chromatin will, altogether, determine the transcriptional response to “a”. What if “a” stands for “auxin”?

### **ARFs, transcription factors of the auxin response**

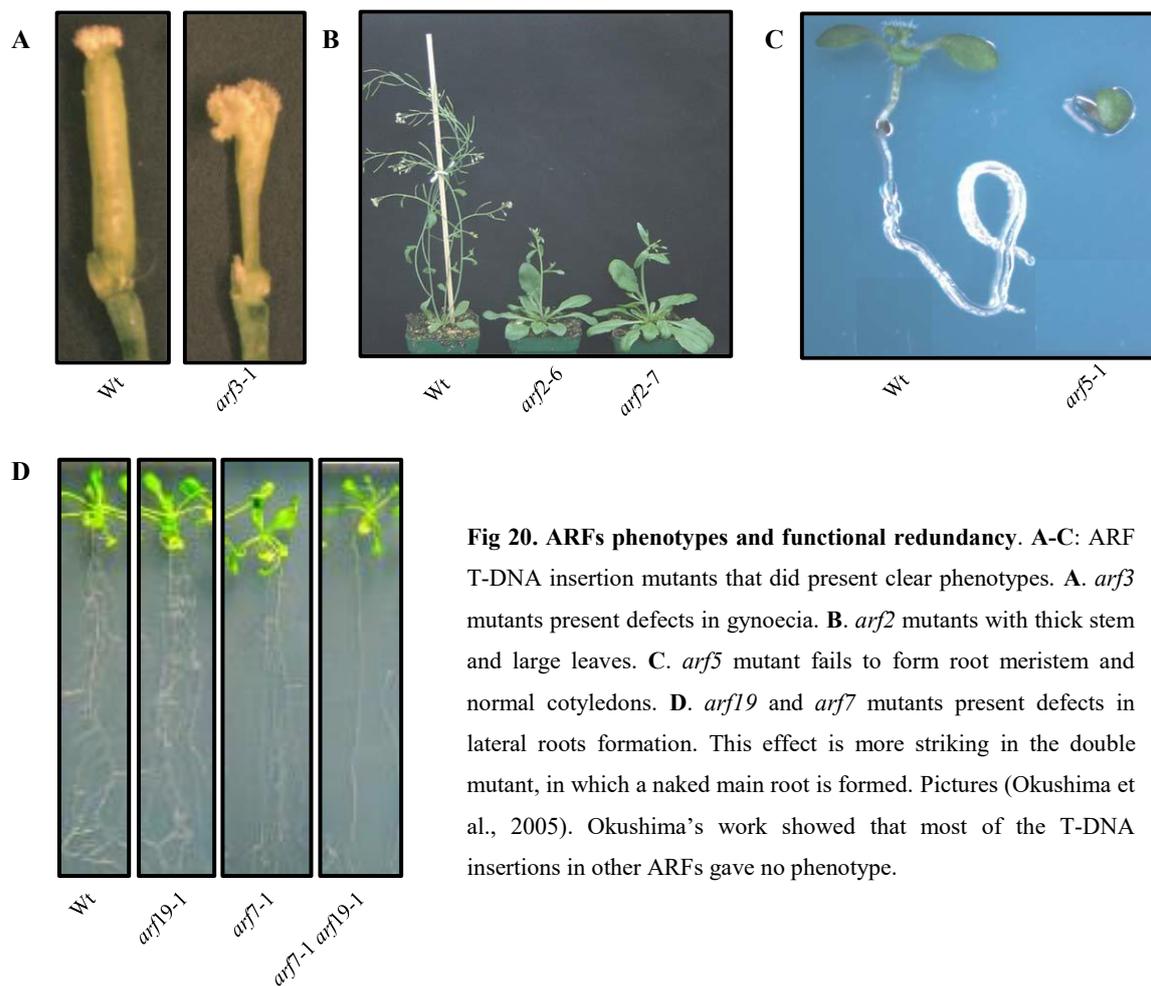
Auxin signal converges towards a family of transcription factors, the ARFs, of which several copies are present in plants with different spatio-temporal expression patterns. The co-expression of different ARFs inside a cell is thought to contribute to the controlled, time and tissue-specific auxin transcriptional response.

As all transcription factors, ARFs present a DBD that mediates their interaction with DNA sequences named Auxin Response Elements, AuxREs, located in the promoters of auxin responsive genes. The DBD is localized in the N-ter of ARF proteins. In the C-terminal region, ARFs present a PB1 domain, also shared by Aux/IAA repressor proteins, that mediates homo and hetero-oligomerization (Chandler,

2016) (Fig7). Both ARFs DBD and PB1 domains can act autonomously: the DBD interacts with AuxREs without the PB1 being necessary for it, and isolated PB1s can mediate oligomerization (Korasick et al., 2014; Nanao et al., 2014; Tiwari et al., 2003; Ulmasov et al., 1997a). For these reasons, ARFs present what is called a “modular domains” structure, with the DBD and PB1 modules that must be fundamental for their function, since they have been extensively conserved along plants evolution (Finet et al., 2013).

Right at the beginning of land plants evolution, we can find the liverwort *Marchantia polymorpha*, a plant that is gaining more and more importance inside the plant physiology field due to the simplicity of its genome. Indeed, *M. polymorpha* presents only 3 copies of ARF genes (Flores-Sandoval et al., 2015) versus the 23 that have been found in the flowering plant *A. thaliana*. This increase in the number of ARF genes is due to a series of gene duplications that on one hand, generated a partial functional redundancy within the ARF family (Okushima et al., 2005) (Fig 20).

However, on the other hand, the duplication events favoured the diversification of ARF biochemical properties, which is thought to contribute to the complexity and precision of the auxin transcriptional



**Fig 20. ARFs phenotypes and functional redundancy.** A-C: ARF T-DNA insertion mutants that did present clear phenotypes. A. *arf3* mutants present defects in gynoecia. B. *arf2* mutants with thick stem and large leaves. C. *arf5* mutant fails to form root meristem and normal cotyledons. D. *arf19* and *arf7* mutants present defects in lateral roots formation. This effect is more striking in the double mutant, in which a naked main root is formed. Pictures (Okushima et al., 2005). Okushima’s work showed that most of the T-DNA insertions in other ARFs gave no phenotype.

response (Roosjen et al., 2017). Diversification that can be found at different levels: functional and phylogenetic.

- **ARFs from a functional point of view**

Functionally, ARFs have been divided in two types: activators or repressors. Out of the 23 ARFs described for *A. thaliana* plant, only 5 have been identified as activators whereas the remaining 18 are considered repressors. Yet, this classification must be carefully taken into account.

First of all, not all ARFs have been functionally tested. The former distribution of the 23 AtARFs into one of the two groups was based on a combination of protoplast transient expression assays and sequences comparison. Analysis of auxin response in protoplasts transformed with AtARFs 1-9 revealed that AtARF5-8 were activators of auxin transcriptional response, whereas AtARF1-4 and AtARF9 were repressors. Moreover, the study of the different domains of AtARF1-9 separately revealed that the region between the DBD and the PB1 of these proteins, named middle region (MR) (Fig 7) confers the activating or repressive activity to each TF. Activators AtARF5-8 MR presented an enrichment of glutamine (Q) residues, whereas the MR of the AtARFs with repressor activity was more enriched in proline (P), serine (S) and threonine (T) amino acids (Tiwari et al., 2003; Ulmasov et al., 1999). Since, ARFs are functionally classed based on the composition of their MRs.

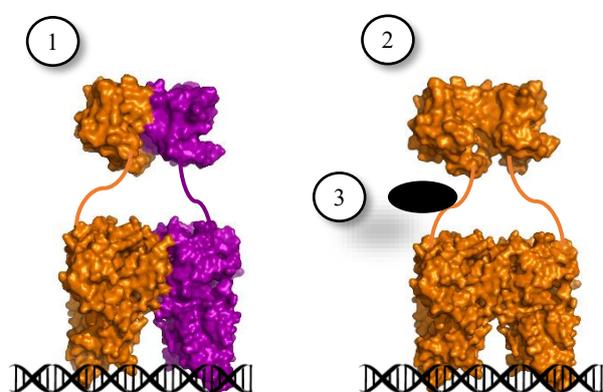
This picture became more complicated with recent evidence showing that some activator ARFs might also work as repressors or even recruit TPL co-repressor (Lokerse and Weijers, 2009; Okushima et al., 2005; Zhao et al., 2010), what suggests that ARFs could function differently in a context-dependent manner.

Finally, although the auxin transcriptional response is well described for ARF activators, there is still no explanation regarding whether and how the potential ARF repressors repress auxin responsive genes. In the absence of the hormone, Aux/IAA and ARF activators form hetero-oligomers through their PB1 domains that block ARF activators. Auxin signal triggers Aux/IAs degradation deblocking ARF activators, now free for the recruitment of co-activator proteins that will allow chromatin opening and the access of the RNA polymerase complex (Fig 11). This model allows us to consider PB1 domains as the “targets” of auxin signal.

However, this idea cannot be applied for ARF repressors. Whereas Aux/IAA and ARF activators PB1 show a very high affinity for each other, which favours the formation of hetero-oligomers (Korasick et al., 2014), most of ARF repressors PB1 do not interact with Aux/IAs or activator ARFs (Farcot et al., 2011; Piya et al., 2014; Trigg et al., 2017). These differences seem to be explicable by the combination of residues in the positive and negative interfaces in the PB1 domains of the different families that leads

to different affinities for one another (Parcy et al., 2016). Although the lack of repressor ARFs-Aux/IAAs interactions led to define these ARFs as auxin independent, they do bind AuxREs and regulate the expression of auxin responsive genes.

Due to this capacity to bind to the same DNA sequence as ARF activators, ARF repressors might act by competition for AuxREs (Lavy et al., 2016). On the other hand, repressor ARFs might block activators, either by interacting through their PB1 domains or by heterodimerizing through their DBDs. Finally, several studies have proven the capacity of some repressor ARFs to bind to TPL (Causier et al., 2012a, 2012b) so chromatin remodelling is proposed as a third mechanism that could explain ARF repression action in an auxin-independent manner (Fig 21) (Chandler, 2016).



**Fig 21. Hypothetical mechanisms for ARF repressors** (represented in orange). (1) “Kidnapping” of ARF activators (purple) by heterodimerization through DBD or PB1 domains. (2) Competition for AuxREs. (3) TPL recruitment, which could be simultaneous to any of the other modes. Adapted from (Chandler, 2016) .

Expression analysis of *ARF* genes indicates that several ARFs, activators and repressors, are co-expressed in different organs of the plant, suggesting that a combination of both is needed for a robust cellular response (Farcot et al., 2011; Rademacher and Mo, 2011). Recent auxin signalling models follow this line: the highly favourable Aux/IAA-ARF activators interactions could provide a stable transcriptional repression of the signal in the absence of auxin. Auxin presence frees ARF activators and the sensitivity of the response would be controlled now by ARF-ARF interactions, with repressor ARFs acting as a “buffers” of the response in the presence of the hormone. This model is supported by the fact that genes coding for ARF repressors are induced by auxin (Farcot et al., 2015; Lavy et al., 2016).

- **ARFs from a phylogenetical point of view**

Given that a functional classification of the ARF proteins might lack some precision, ARFs are starting to be divided into evolutive clades. The most extensive phylogenetic study of ARF proteins included 224 ARF-related protein sequences from different land plants along evolution. As a result, ARFs were classed into 3 clades: A, B and C. Class A includes the 5 AtARF activators, whereas the repressor ARFs fall into clades B and C (Finet et al., 2013).

In some cases, AtARF3 and 4 are dissociated from class B and rearranged into a separate clade, with 4 clades of ARFs instead of 3 (Mun et al., 2012). A, B and C classification will be used from now onwards in this manuscript.

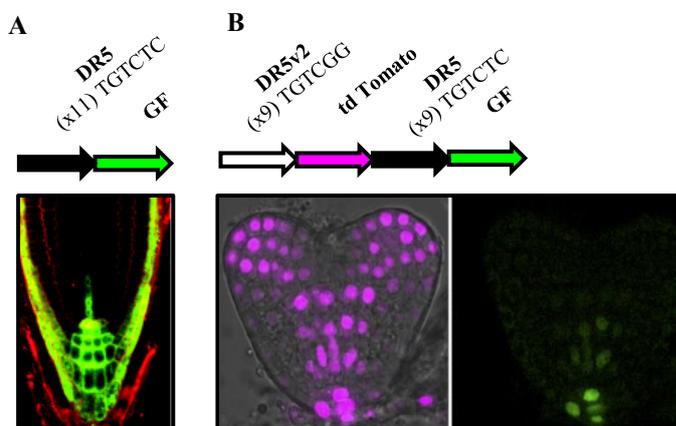
Whilst class-A contains ARFs with Q-rich MRs, classes B and C include ARF members that have suffered C-terminal truncation events (AtARF3, 13, 17 and 23). Despite lacking a PB1 domains, these ARFs, except for ARF23, have been shown to be functional and important for developmental processes (Finet and Jaillais, 2012; Finet et al., 2010; Simonini et al., 2016). Both features, Q-rich regions and C-terminal deletions, seem to have underwent a positive selection pressure (Finet and Jaillais, 2012).

Clade A (or activators) has been extensively studied: structures are available for AtARF5-DBD and for AtARF5 and AtARF7 PB1 domains (Boer et al., 2014; Korasick et al., 2014; Nanao et al., 2014) (Figs 8-9) and their mechanisms of action has less and less secrets for us. Clade B, as repressors, are a bit more of a mystery, but still, inside this class we count with AtARF1, the first ARF protein to be isolated and that provided in 2014 the structural mechanism of interaction between ARFs and palindromic AuxREs (Boer et al., 2014). Last year, another member of this class, AtARF2, together with AtARF5 starred in the DAP-seq (DNA-affinity purification sequencing) opening, a new technology for determining TF binding sites, that contributed to our knowledge about preferential binding of different ARFs (O'Malley R.C. et al., 2016). Class-C is then the great ignored of the ARFs story, (spoiler alert!) at least till the end of these following Chapters II and III.

### **ARFs DNA binding sites and mechanism**

Auxin Response Elements, AuxREs, were first identified in the promoters of *GH3* soybean gene, an early-auxin responsive gene implicated in IAA metabolism (Fig 5). By scraping this promoter into smaller sequences, Liu *et al.* in 1994 isolated the promoter regions that were auxin-inducible, in which the TGTCTC motif was first found as the consensus sequence of the AuxREs (Liu et al., 1994). 3 years later, an artificial AuxRE constructed with 4 tandem copies of TGTCTC tandem repeats (P3 4X) allowed, in a one-hybrid assay, the isolation of the first ARF transcription factor, AtARF1 (Ulmasov et al., 1997a).

Since, numerous researches have been done for trying to unravel the specificity of different ARFs for different AuxREs and how this specificity can lead to activation or repression of auxin responsive genes. Moreover, these findings have permitted the development of several auxin sensors that allow the quantification of the auxin transcriptional response (Fig 22) (Liao et al., 2015; Ottenschläger et al., 2002).

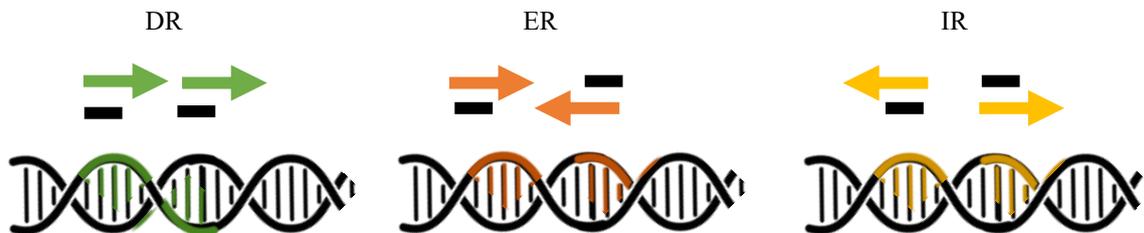


**Fig 22. Reporters of auxin transcriptional response.** **A.** GFP under the transcriptional control of DR5 reporter (TGTCTC repeats) indicating auxin transcriptional response in the root quiescent centre after  $1\mu\text{M}$  IAA treatment (Picture from (Ottenschläger et al., 2002)). **B.** DR5v2 reporter, constructed out of TGTCGG repeats, has a more intense activity (indicated by td Tomato fluorescent protein) than DR5 (represented by GFP signal). Heart-stage embryos are shown in the picture (Liao et al., 2015).

AtARF1 discovery led to the cloning and analysis of many other *A. thaliana* ARF members. The first analysis done with AtARF1-10 proteins proved that all of them bound to the same consensus sequence, TGTCTC. The same research clarified that binding only took place upon double TGTCTC motifs, indicating that a dimeric protein was needed for a cooperative binding (Ulmasov et al., 1999). Mutations within this motif showed that the first four nucleotides, TGTC, were the ones essential for the binding and the auxin response, whereas mutations in positions 5 and 6 had a weaker effect that was different depending on the ARF protein (Ulmasov et al., 1997a, 1999). Another aspect that influenced the ARF1-DNA interaction was the orientation and spacing of the two TGTCTC repeats needed for the dimeric binding. Palindromic TGTCTC repeats, named Everted Repeats (ER) (Fig 23) with a spacing of 7/8 nucleotides, ER7/8, presented a preferential binding by ARF1. Direct repeats (DR) with a 5-nucleotides spacing, DR, could also be bound by this protein. Both types of AuxRE, ER7 and DR5, were auxin-responsive and increasing number of ER7 and DR5 repetitions had a positive influence in their auxin-inducibility (Ulmasov et al., 1997a, 1997b).

In transient expression protoplasts assays, both ER7 and DR5 AuxREs were induced by AtARF5-8 class A and repressed by AtARF1 and 2 class B. However, AtARF3, 4 and 9 were only able to repress DR5-containing promoters and had no effect upon ER7 sequences (Tiwari et al., 2003), indicative of a possible specificity of the response depending on the type of ARF and on the promoter sequence structure. More recent auxin transcriptional response studies in a yeast system sustain this theory: the spacing and orientation of the TGTCTC repeats influenced AtARF5 and AtARF19 activating capacity (Pierre-Jerome et al., 2016).

Latest details of the ARFs-AuxREs interaction came as a result of two different approaches: structural biology and bioinformatic models originated from big data analysis. Combined, they provided a more detailed understanding of the specificity and mechanism of these interactions, that will be described in the next three sections.



**Fig 23. AuxREs configurations.** ARFs binding sites are double sites which can be Direct Repeats (DRs), Everted Repeats (ER) or Inverted Repeats (IR). The spacing in between sites can vary so the double AuxRE can potentially occupy more than one DNA helix turns.

- Structural insights into the ARFs-DNA interaction

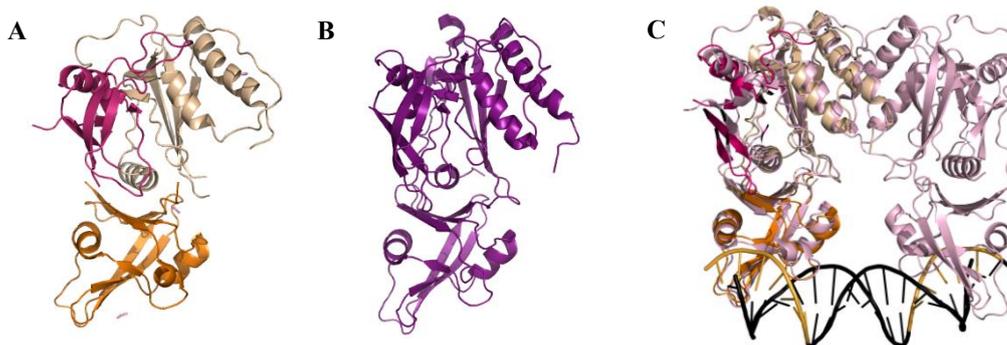
The resolution of the structure of AtARF1 (apoprotein and in complex with ER7 AuxRE) and AtARF5 DBDs was a fundamental contribution for the understanding of the binding mechanism of ARFs to DNA (Boer et al., 2014). Both structures were almost identical: the B3 domain, that mediates interaction with the DNA, is embedded into a dimerization domain (DD), separated in two flanks, that mediates the interaction of two ARF monomers (Fig 5 and 24). B3 and DD are connected by a long flexible loop that, contrary to the B3 or DD, presents a low degree of residues conservation among the AtARF members. A third domain, flanking domain (FD), of yet unknown function is also part of the two ARFs DBD structure.

Consistent with their similar structure and sequence homology, Protein Binding Microarrays (PBM) revealed that both ARFs, considered as monomers, bound the same simple AuxRE. Strikingly, the most enriched consensus was TGTCGG instead of the traditional TGTCTC (Boer et al., 2014; Franco-Zorrilla et al., 2014).

Moreover, AtARF1 was also co-crystallized in complex with an ER7 AuxRE motif. The apoprotein and the complex structures presented barely any differences except for a 25° rotation of the B3 domain upon the protein binding to DNA (Fig 24.C). The distance between the two B3 of the AtARF1 dimer explains the preferential binding of this protein to ER motifs with 7/8 nucleotides spacing. In order to test if the ARFs structure was flexible enough to accommodate differently spaced ERs, the preferential binding of AtARF1 and AtARF5 was tested upon ER5-9 DNA conformations. Whereas AtARF1, in agreement with earlier experiments, bound with a clear preference to 7/8 spaced ERs, AtARF5 was more

permissive and interacted with all tested sequences although the affinity was also higher for ER7/8 (Boer et al., 2014).

After these experiments, ARFs have been considered as “molecular calipers”, able to fit differently spaced ERs depending on the flexibility for the two monomers to open and close. This ability might be dependent on the connecting loops in between B3 and DD, and thus, different for each ARF.



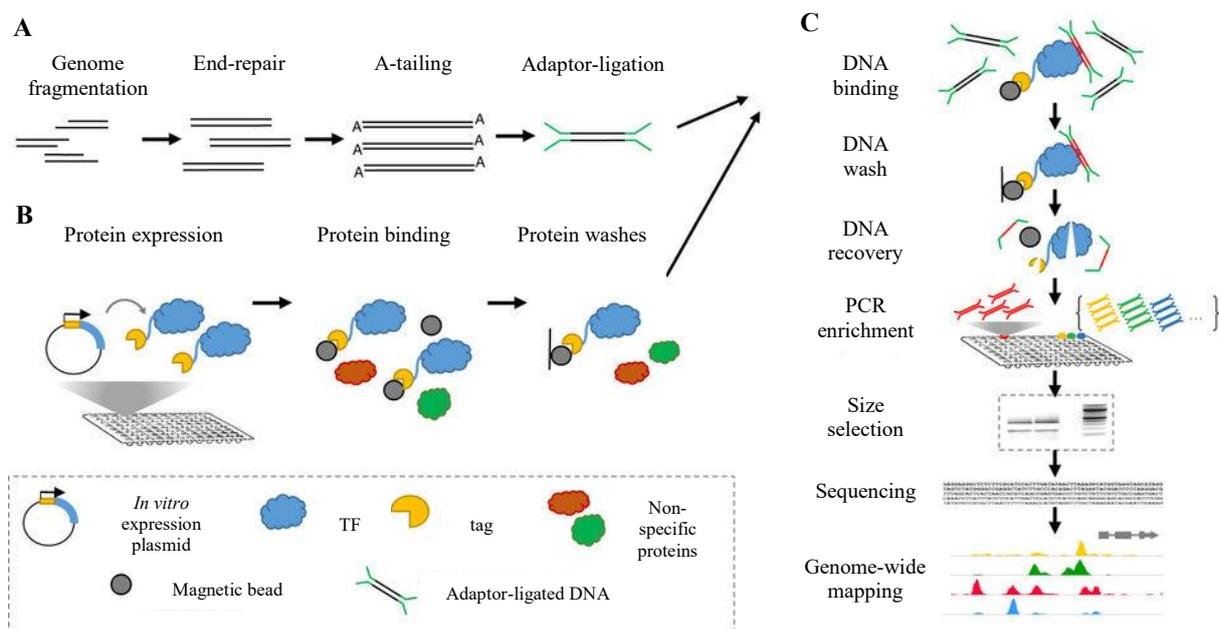
**Fig 24. AtARF1 and AtARF5 DBD structure.** **A.** AtARF1 apoprotein structure (4LDV). Dimerization Domain (DD) and connecting loop in beige, Flanking Domain (FD) in pink and B3-DNA binding domain in orange. **B.** AtARF5 apoprotein structure (4LDU). **C.** AtARF1 apoprotein structure superposed onto AtARF1-DBD (light pink) (4LDX) in complex with AuxRE ER7 (black; TGTCTC repeats in pale orange). A slight rotation of the B3 is observed upon DNA interaction (Boer et al., 2014).

- Bioinformatic approaches for ARFs-AuxREs specificity

Having been determined as the minimal consensus for auxin-inducibility, TGTC motifs have been used for the genome-wide search of auxin-response. This led to the identification of other possible 6-length AuxREs that present variations of the two last bp of TGTCNN motifs, apart from the already known TGTC(TC)/(GG). By crossing the promoter regions with transcriptional profiles resulting from auxin treatment, certain combinations of dinucleotides in positions 5 and 6 were associated with auxin up-regulated genes and others with auxin down-regulated genes (Zemlyanskaya et al., 2016). This indicates a possible specificity of “base readout” depending on the activating or repressive identity of the transcription factor. However, computational approaches based on single AuxREs lack precision due to the known cooperativity of ARFs that bind as dimers (or even oligomers).

Wider analysis of dimeric or oligomeric ARFs specificity binding has been hampered due to the lack of extensive ChIP data. The newly developed DAP-seq technology has broadened the opportunities for mapping the cistrome and epicistrome in the search for TF binding sites (Fig 25).

DAP-seq technology uses the genomic DNA (gDNA) devoid of nucleosomes (cistrome) or the genomic DNA with cytosine-methylated marks (episcistrome). gDNA is sheared in 200bp-fragments that are co-purified with recombinantly produced TFs. The co-purified DNAs are amplified by PCR, sequenced and mapped onto the genome (Bartlett et al., 2017). Due to the *in vitro* nature of this technology, features such as BSs accessibility inside chromatin cannot be taken into account, but at the same time offers the advantage of focusing only on “base readout” and “shape readout” characteristics.



**Fig 25. DAP-seq technology.** **A.** Genomic DNA library preparation: 200bp-DNA fragments are ligated to sequencing adaptors. **B.** Recombinant TFs are expressed and purified with magnetic beads. **C.** DNA fragments and TFs are incubated together. Unspecific interactions are washed away and the specifically-bound DNAs are separated from the proteins by heating treatment. The recovered DNAs are amplified by PCR, followed by selection of the fragments of the right size that are sent for sequencing. DNA sequences are finally mapped onto the genome (Figure taken from (Bartlett et al., 2017)).

AtARF5 and AtARF2 TFs (class A and B or activator and repressor, respectively), were analysed using DAP-seq methodology (O'Malley R C. et al., 2016). TGTC was taken as the model motif for searching ARF binding sites. Their results indicate that, in agreement with the structural and biochemical data, ARF5 could bind a diversity of sequences in terms of spacing and orientation. Enrichment of ARF5 was found in DR4-6 and IR7-8. Moreover, everted repeats (ER) (Fig 23) were also found among the ARF5 bound sequences, with spacings of 0-2 (ER0-2). In all three cases, a second enrichment set was found with a 10bp-interval, that is to say, DR14-17, IR17-19 and ER11-14. Contrary to ARF5, ARF2

was only enriched in IR7-8 motifs. (Please, note that O'Malley *et al.* used an inverted nomenclature of ER and IR).

ARFs being the latest actors of the auxin transduction pathway their study is key for properly understanding the auxin transcriptional response: their preferences, their quarrels and their DNA binding mechanisms. This is why, during this Chapter II we focused on studying the specificity of different ARFs (classes A, B and C) for AuxREs. To carry out these studies, we used a computational approach combined with biochemical and structural studies.

## Results

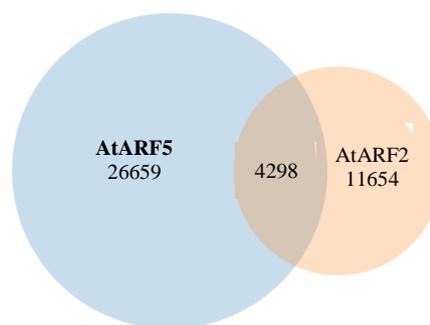
The computational studies consisted on the re-analysis of DAP seq data for AtARF2 and AtARF5 (O'Malley R.C. *et al.*, 2016) using a new model for ARFs DNA binding sites based on Position Weight Matrixes (PWM). These models present an improvement with respect to the TGTC-consensus search tool, used in O'Malley's studies. The resulting predictions from this *in silico* approach were tested by *in vitro* protein-DNA interaction analysis and by structural modelling.

The computational part of Chapter II was developed by Arnaud Stigliani, Adrien Bessy and François Parcy and thus, it will be presented in this manuscript in a more descriptive sense. The presentation of these results will be necessary for a proper understanding of the biochemical and structural modelling data that will be more extensively discussed.

### Results from the computational DAP-seq analysis for AtARF2 and AtARF5 binding sites

The biochemical and structural data obtained in the last 20 years, when the first ARF was isolated, indicated that different types of ARFs could bind similar DNA sequences, with a TGTC consensus, when considered as monomers (Boer *et al.*, 2014; Ulmasov *et al.*, 1997a, 1999). However, the comparison of AtARF2 and AtARF5 DAP-seq data revealed a significant number of regions bound only either by AtARF2 or 5, indicative of a specificity in the binding of these two proteins to DNA (Fig 26). How can these differential binding be explained?

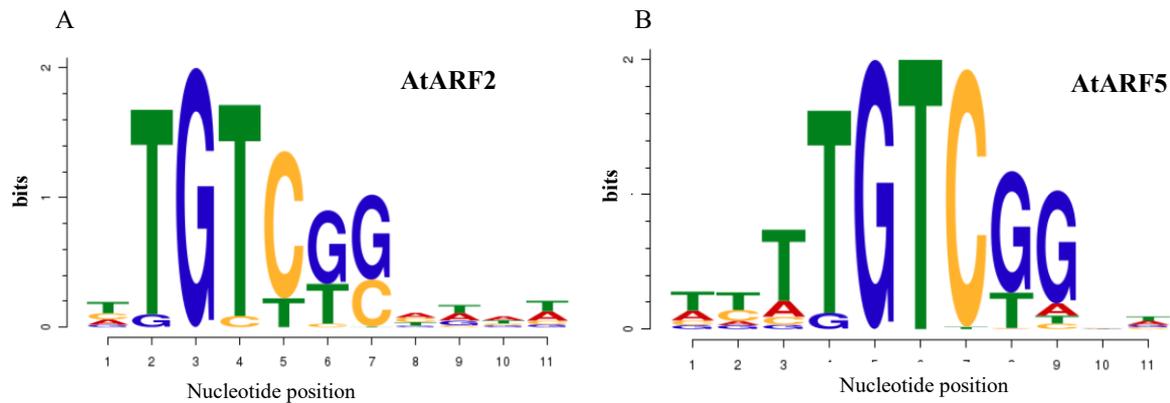
Since the TGTC-based search has proven to be not predictive enough of the auxin response (Keilwagen et al., 2011), the first objective was the establishment of an improved binding model for each factor that could help us to understand these differences. Models for ARFs monomeric binding were developed by Position Weight Matrixes (PWM) that were constructed by alignments of the DAP-seq bound regions for each ARF.



**Fig 26. Venn diagram for regions bound by AtARF2 and AtARF5.** The analysis of regions bound only by AtARF5 or AtARF2 indicates that there is a specificity in the binding of the two proteins, in spite of which there are still some regions in the genome that can be bound by both.

The PWM can be represented as DNA sequences logos in which the size of the letter in each position is informative of the content and frequency of each nucleotide. AtARF2 and AtARF5 logos were very similar. The traditional TGTC sequence is clearly overrepresented but certain variation within these 4 positions seemed to be permitted, especially for AtARF2. Some information also comes from nucleotides flanking the TGTC (Fig 27).

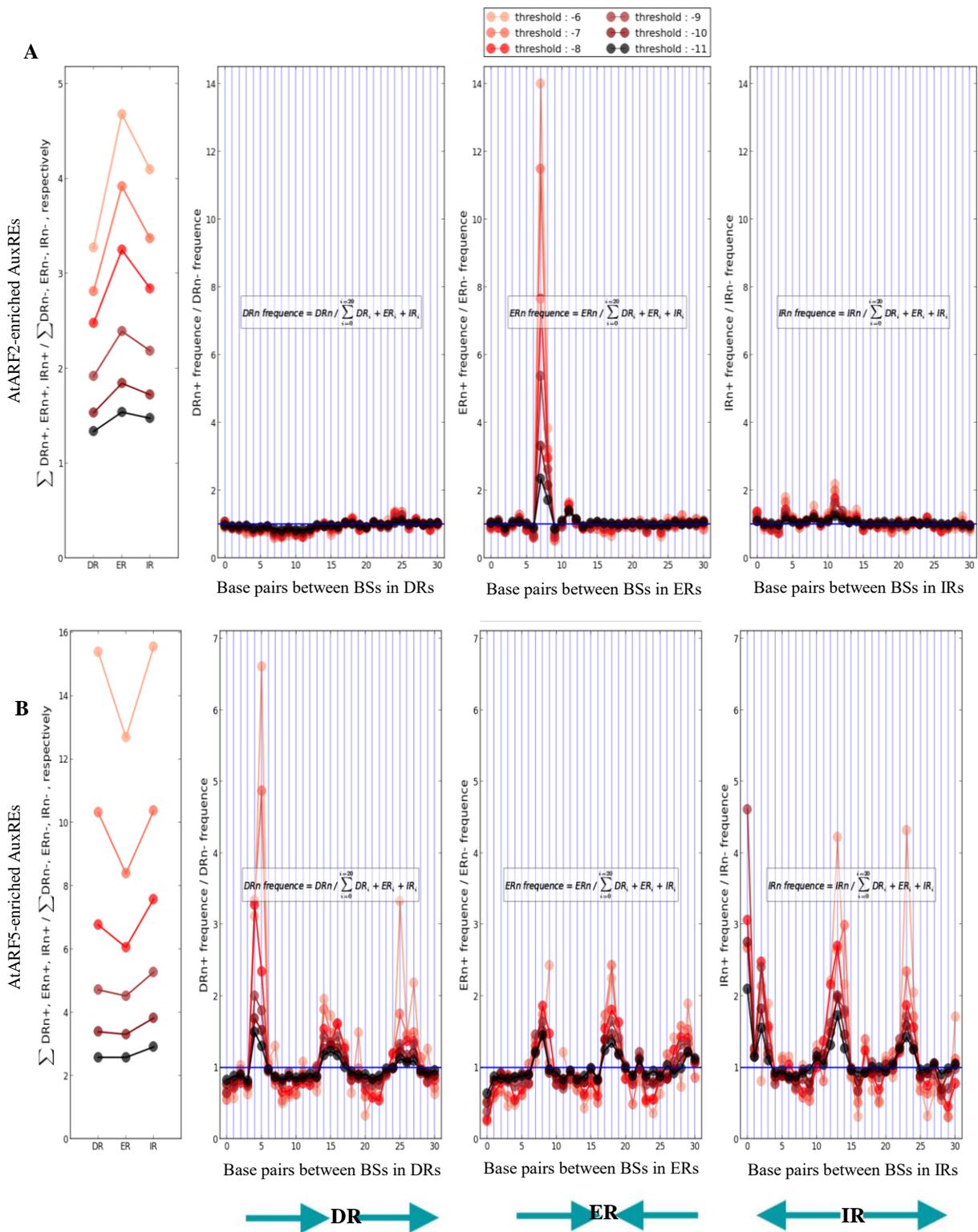
These slight differences in the models for AtARF2 and AtARF5 monomeric binding could be one possible explanation for the genomic regions bound in an exclusive manner by one of the two proteins as shown by the Venn diagram. However, testing AtARF5 PWM on AtARF2 DAP-seq bound regions only slightly decreased the performance of the model. This indicates that, although there is some specificity in the PWM monomeric models, other parameters might have more influence than the nucleotide sequence itself.



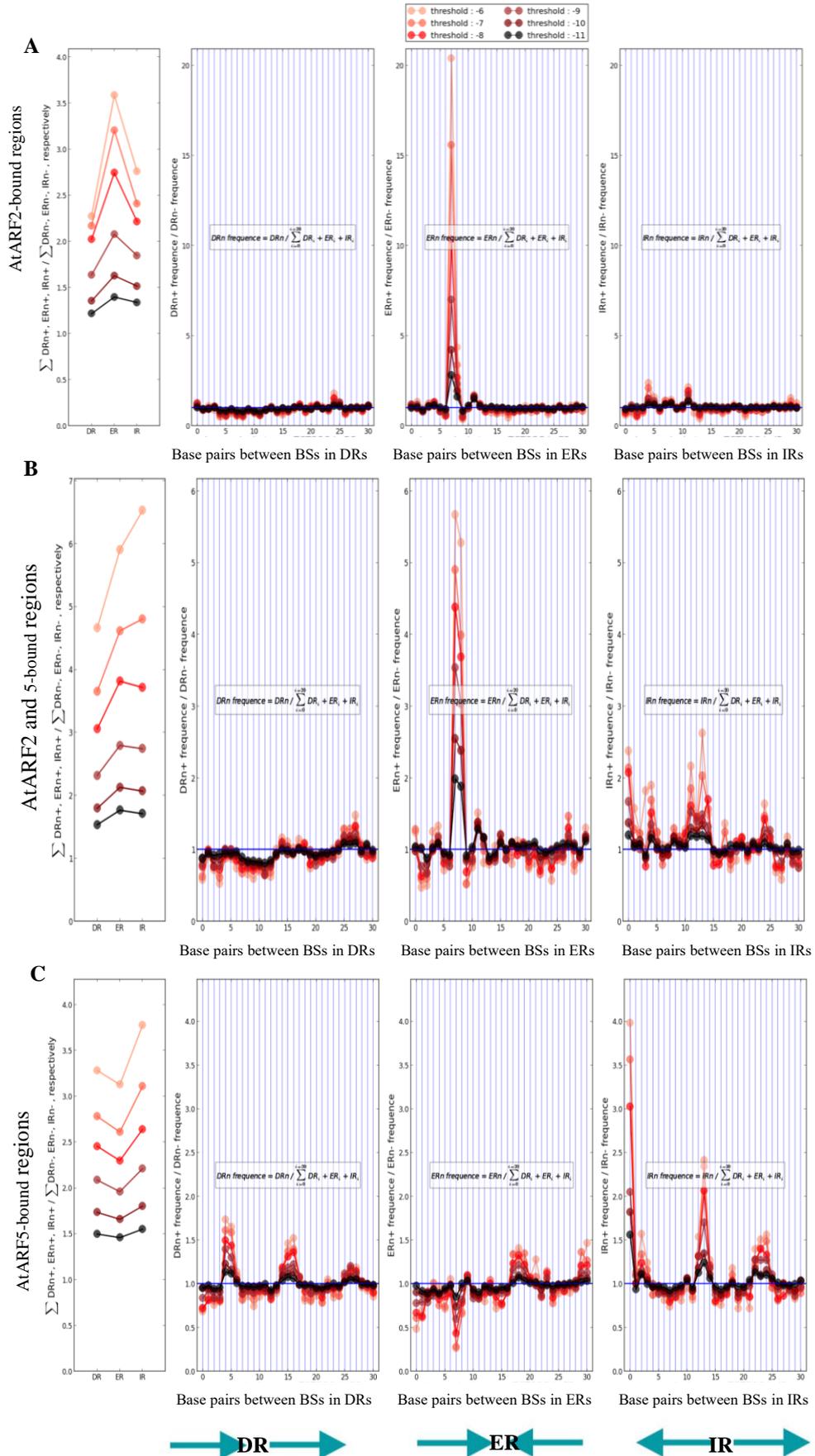
**Fig 27. DNA logos for AtARF2 (A) and AtARF5 (B).** Although the TGTC consensus is the preferential binding site, modifications of it appear among AtARF2 and AtARF5 binding preferences are slightly different with flanking nucleotides being informative as well.

Since ARF proteins are known to bind as dimers to double AuxRE sites, with the orientation and spacing of the two sites being determinant for the binding and auxin-responsiveness (Boer et al., 2014; Pierrejrome et al., 2016; Tiwari et al., 2003; Ulmasov et al., 1997a, 1997b), we next tested DR, ER and IR motifs enrichment for AtARF2 and AtARF5. Whereas AtARF2 presented a clear enrichment on ER7/8 sites, AtARF5 was enriched in all three types of AuxREs, admitting several spacings. The predominant binding sites for AtARF5 were DR4-6, ER7-9 and IR0-3, in agreement with O'Malley's analyses (considering the ER-IR inversion). Moreover, an enrichment of the same sequences but spaced with a periodicity of plus 10bp more was also found (Fig 28).

The Venn diagram was then reanalysed with respect to ARFs binding preferences as dimers scanning the regions bound only either by AtARF2, AtARF5 or both. Whereas the regions bound exclusively by AtARF2 or by both proteins presented mostly ER7/8 AuxREs, the regions bound only by AtARF5 were depleted in ER7/8 binding sites and enriched in the rest of motifs previously found for AtARF5 (Fig 29).



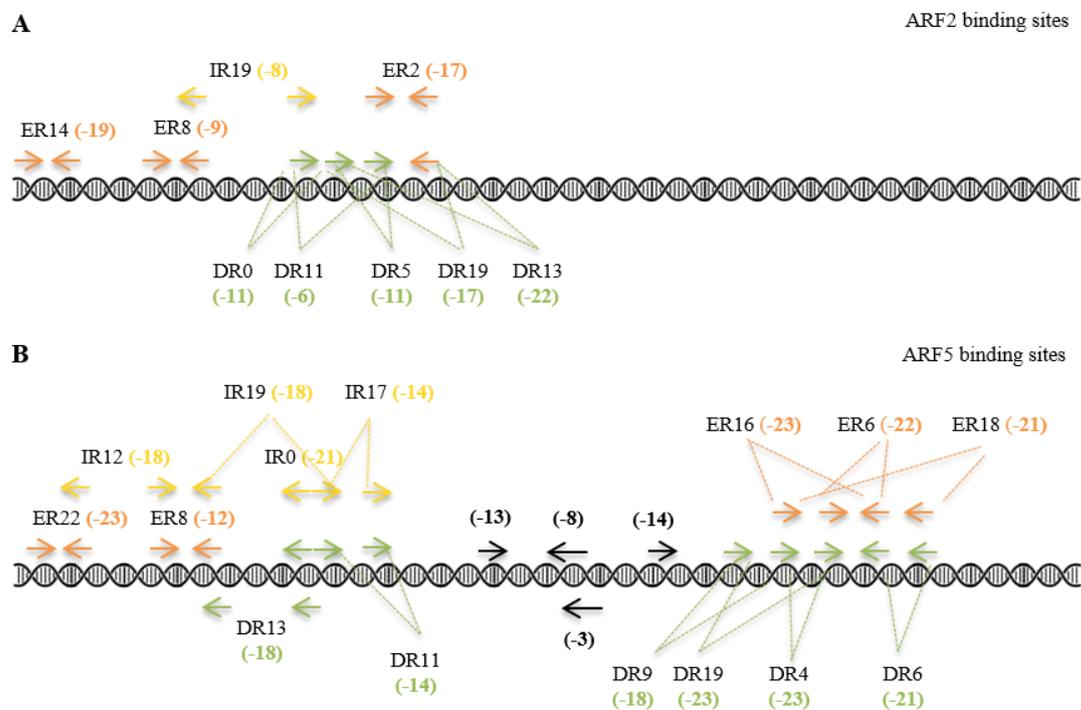
**Fig 28. Double binding sites enrichment for AtARF2 and AtARF5. A.** AtARF2 preferentially bound ER7/8. **B.** AtARF5 was enriched in DR4-5;14-15; 24-26; ER7-8; 16-18; 29-30; IR0-3; 11-14;22-24. Note the periodicity of 10bp-spacing in all types of AuxREs conformations. Threshold represents the cut-off value to qualify a sequence as binding site with the PWM model. Because it cannot be chosen rigorously, we used different values.



**Fig 29. Sites enrichment for regions corresponding to Venn diagram.** **A.** Regions bound exclusively by AtARF2. A ER7/8 enrichment is observed. **B.** Regions bound by both proteins. ER7/8 and several IR spacings stand out. **C.** Regions bound exclusively by AtARF5. DR and IR spacings enriched in AtARF5 complete analyses were also present in regions bound only by AtARF5 but there is a depletion of ER7/8. The square in the left represents the enrichment of DR, ER and IR in the positive set with respect to the negative set.

### *In vitro* test of PWM matrices through natural promoters

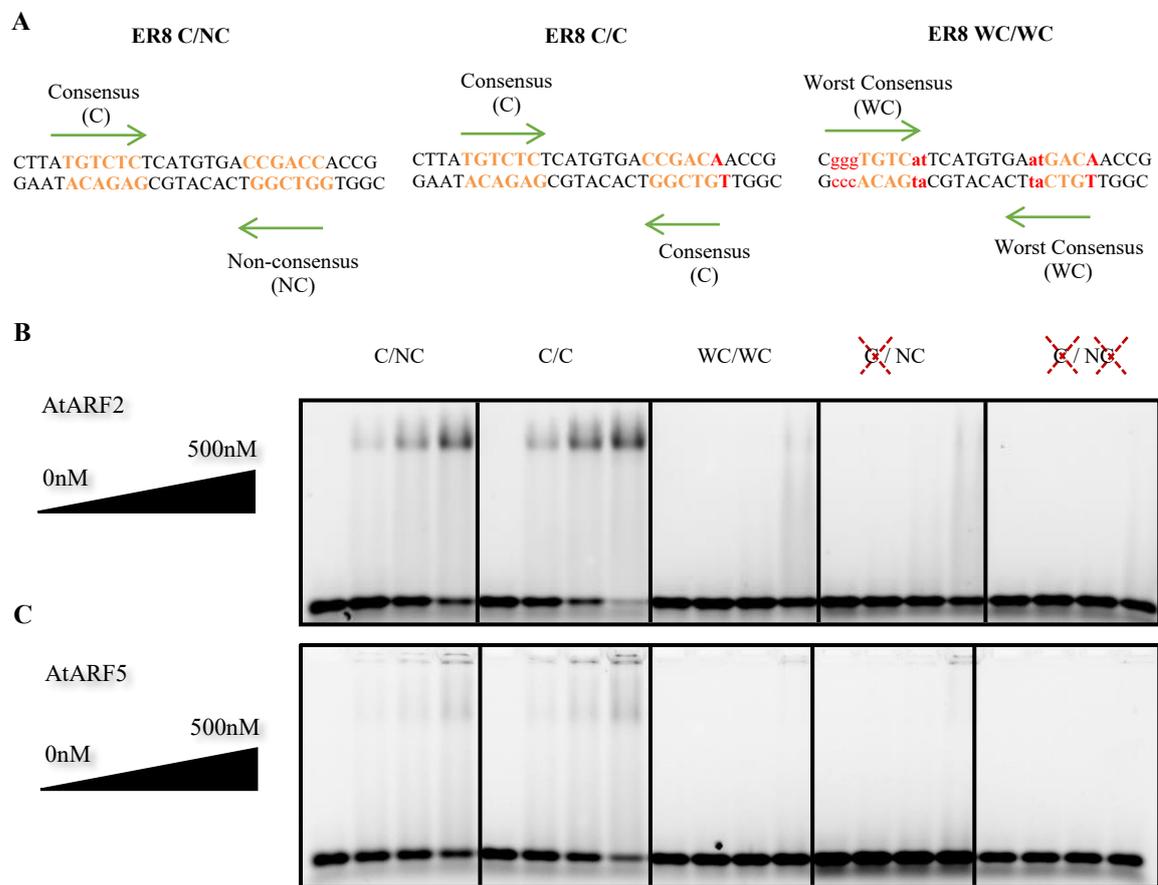
The PWM obtained for AtARF2 and AtARF5 from DAP-seq data indicated that the TGTC consensus is not necessarily required for binding by ARFs (Fig 27). Although AtARF2 seemed to admit more variations within these four nucleotides, a GGTC non-consensus sequence could, for example, be compatible with AtARF2 and AtARF5 according to the logo models. To test the power of our new model to detect ARF BSs, we took as model the natural promoter of *IAA19*, an auxin responsive gene which has been recently used as a tool for the analysis of auxin transcriptional response in yeast (Pierre-Jerome et al., 2016). Within the DAP-seq peak for *IAA19* promoter several AuxREs were found for AtARF2 and AtARF5. Among them, the predicted binding site with the



**Fig 30. Predicted binding sites within *IAA19* promoter.** **A.** ARF2 binding sites. **B.** ARF5 binding sites. Green arrows: DRs. Orange arrows: ERs. Yellow arrows: IRs. Black arrows: single sites. Scores for each site indicated between brackets.

highest score was an ER8 motif with a non-consensus GGTC sequence on the reverse strand (Fig 30; 31.A).

Using ER8-*IAA19* AuxRE (ER8 C/NC) as starting point, we constructed two variants: ER8-*IAA19* (C/C) in which we restored a consensus from the original non-consensus sequence and ER8-*IAA19* (WC/WC for worst consensus) that maintained both TGTC consensus sequences but with poorly represented nucleotides on the flanks (based on the PWM) (Fig 30.A). As controls, we used single ER8 sites and no-site sequences in which one or both sites were mutated. We found that AtARF2 and AtARF5 could bind ER8 C/NC in a similar way to ER8 C/C, whereas the binding was lost with ER8 WC/WC, proving that the TGTC-consensus is neither absolutely necessary nor sufficient for the binding, with flanking nucleotides also playing an important role for a good affinity of both



**Fig 31. *In vitro* new matrix validation through *IAA19* promoter.** **A.** ER8 predicted site inside *IAA19* promoter (left, C/NC) and the subsequent modifications introduced in it: non-consensus sequence substituted by an optimal TGTC consensus (centre, C/C) and both sites substituted by a bad-score TGTC-consensus (right, WC/WC)). DNA sequences with a single site (TttCTC mutation in consensus) and no site (TttCTC mutation in both sites) were also used. **B.** AtARF2 binding to ER8-*IAA19* original and modified sequences at increasing protein concentrations (0, 125, 250 and 500nM). **C.** AtARF5 binding to ER8-*IAA19* original and modified sequences at increasing protein concentrations (0, 125, 250 and 500nM).

proteins. None of the proteins could bind to single sites, in agreement with the need of double sites for a cooperative binding of ARF proteins as dimers (Fig 31.B-C).

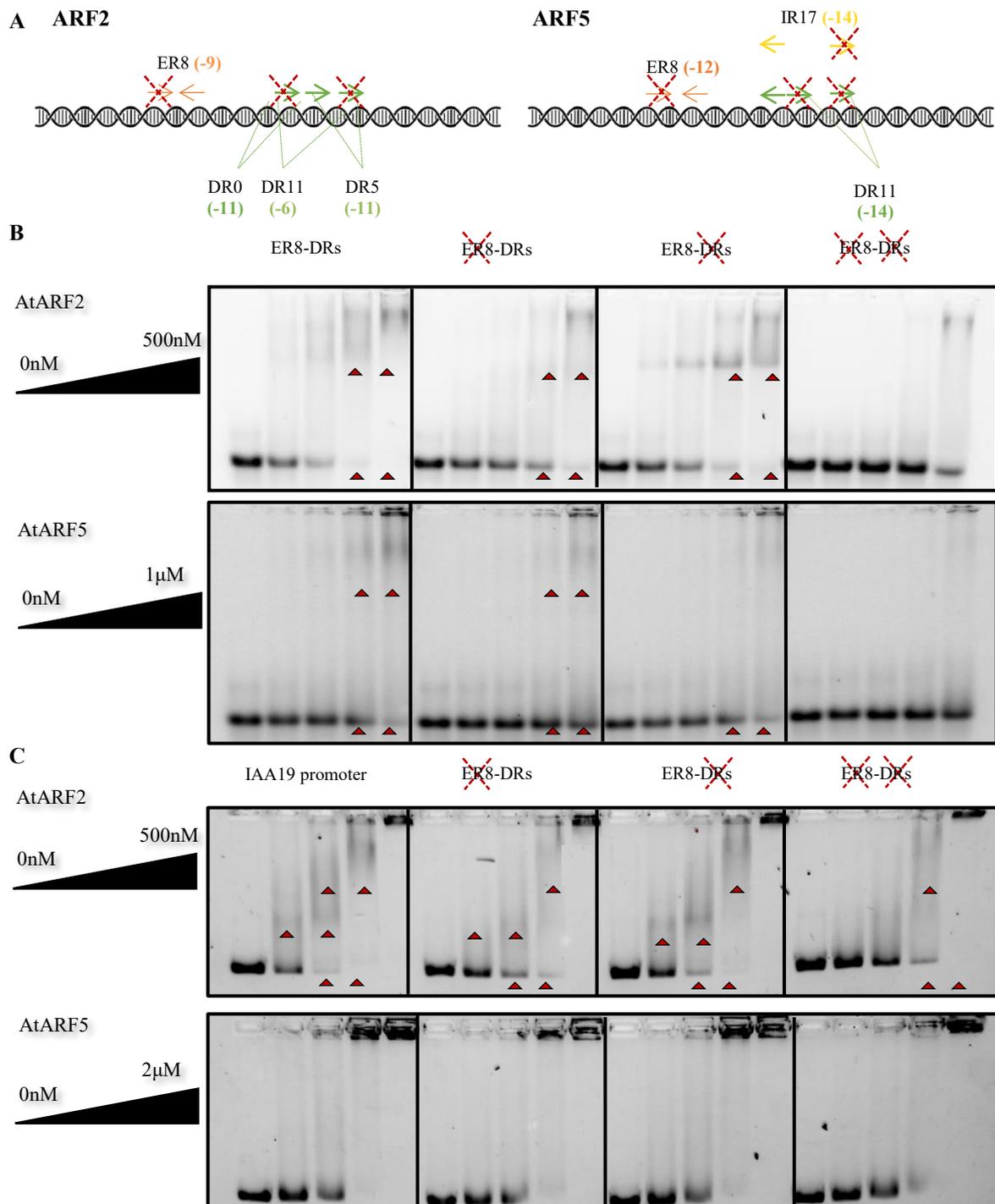
### ***In vitro* validation of binding preferences through natural promoters**

*IAA19* promoter contained other ARF binding sites with lower scores. AtARF5 presented two clusters of predicted sites (one in 5' and one in 3') with a mix of DR, ER and IR motifs inside each cluster (Fig 30.B). For AtARF2, on the other hand, AuxREs were only located in the 5' region, in which, apart from the already studied ER8, a series of DRs with high scores was found (Fig 32.A). Due to the combination of ER8 and DRs, we chose to work first with this region in order to study the binding preferences of our proteins to different types of AuxREs. For this, we isolated an 80bp-fragment (ER8-DRs) that contained the ER8 motif and differently spaced DRs for AtARF2 and AtARF5, and we mutated both set of motifs in a mutually exclusive manner (ER8 mutated-DRs and ER8-DRs mutated) (Fig 32.A). None of the mutations originated striking affinity changes with respect to the wild type. Therefore, we analysed the effect of the different mutations based on two features: the decrease of the free DNA band intensity and the formation of complexes as increasing protein concentration (indicated by red arrows in the figure). Following these criteria, we observed that mutation of the ER8 motif alone had a slightly stronger effect than DRs mutation for AtARF2 (Fig 32.A), allowing us to conclude that ER8 motif plays a more important role in the binding of ARF2 than DRs, in agreement with the predicted preferences from the computational DAP-seq data treatment. The EMSAs done with AtARF5 did not show clear differences not being then, conclusive enough.

We repeated these analyses using the full promoter (350bp) in which we reproduced the same mutations: ER8 mutated-DRs, ER8-DRs mutated and ER8 mutated-DRs mutated. In comparison with smaller DNA sequences, binding to *IAA19* full promoter gave no clear bands but smeared DNA shifts or traces around the wells at higher protein concentrations, indicative of the formation of big DNA-protein complexes unable to penetrate the gel. We used the same criteria as before for evaluating the effect of each mutation.

For AtARF2, mutating ER8 had the same effect as mutation of both ER8 and DRs, whereas the binding to DRs mutated sequence was similar to the one observed for the original *IAA19*. In line with the previous results for ER8-DRs 80bp-fragment, the EMSAs done with the full promoter indicate that AtARF2 shows preference for ER8 motif rather than DRs. On the contrary, AtARF5 showed no clear differences with the different mutated sequences (Fig 32.C).

DNA preferences apart, it is worth noting that both proteins shifted *IAA19* promoter in a different manner: AtARF2 protein seems to progressively form complexes of increasing sizes that turn into a very big complex unable to enter the gel at higher protein concentrations, but AtARF5 does not form these intermediate complexes. On the contrary, DNA is shifted all in once into big DNA-protein complexes that do not penetrate the gel (Fig 32.C).



**Fig 32. *In vitro* test of DR/ER preferences using *IAA19* promoter.** **A.** 80bp fragment from *IAA19* promoter (ER8-DRs) (position 56-138) representing the mutations introduced to annul ER8 site and/or multiple DRs. **B.** AtARF2 and AtARF5 binding to ER8-DRs fragment, and to its mutated versions: ER8mutated-DRs, ER8-DRsmutated and ER8mutated-DRsmutated. Tested increasing protein concentrations: AtARF2 at 0, 62.5, 125, 250 and 500nM and AtARF5 at 125, 250, 500 and 1000nM. **C.** AtARF2 and AtARF5 binding to *IAA19* promoter and its mutated versions, containing the same mutations as the ones introduced in ER8-DRs fragment. Tested increasing protein concentrations: AtARF2 at 0, 62.5, 125, 250 and 500nM and AtARF5 at 250, 500, 1000 and 2000nM. Red arrows point at differences among the effect of the different mutations.

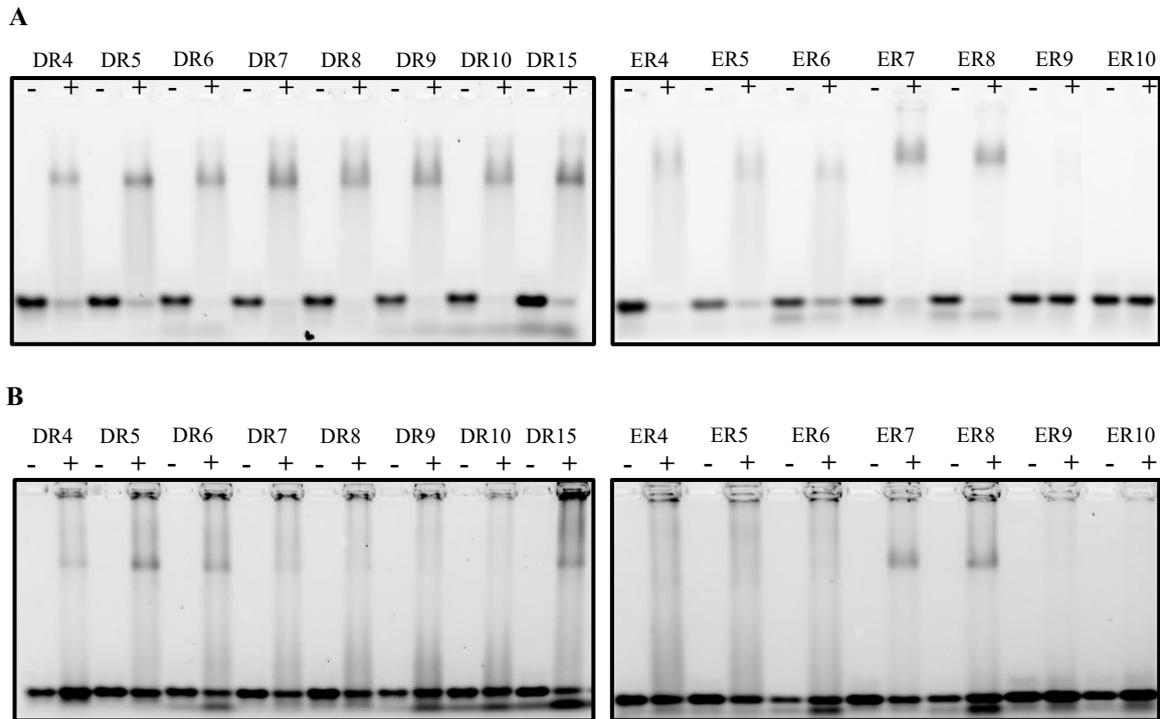
### Classes A, B and C binding preferences outside the *IAA19* nest

For a wider representation of AuxREs in terms of orientation and spacing, next we further tested the predicted binding preferences for AtARF2 and AtARF5 using a set of DRs and ERs spaced 4-10 nucleotides. An additional DR15 was used due to its high enrichment among AtARF5 bound sequences. These AuxREs were designed using 36-47 bp-oligonucleotides (depending on the spacing) with optimal binding sites containing the TGTC consensus (Table 8, Materials&Methods).

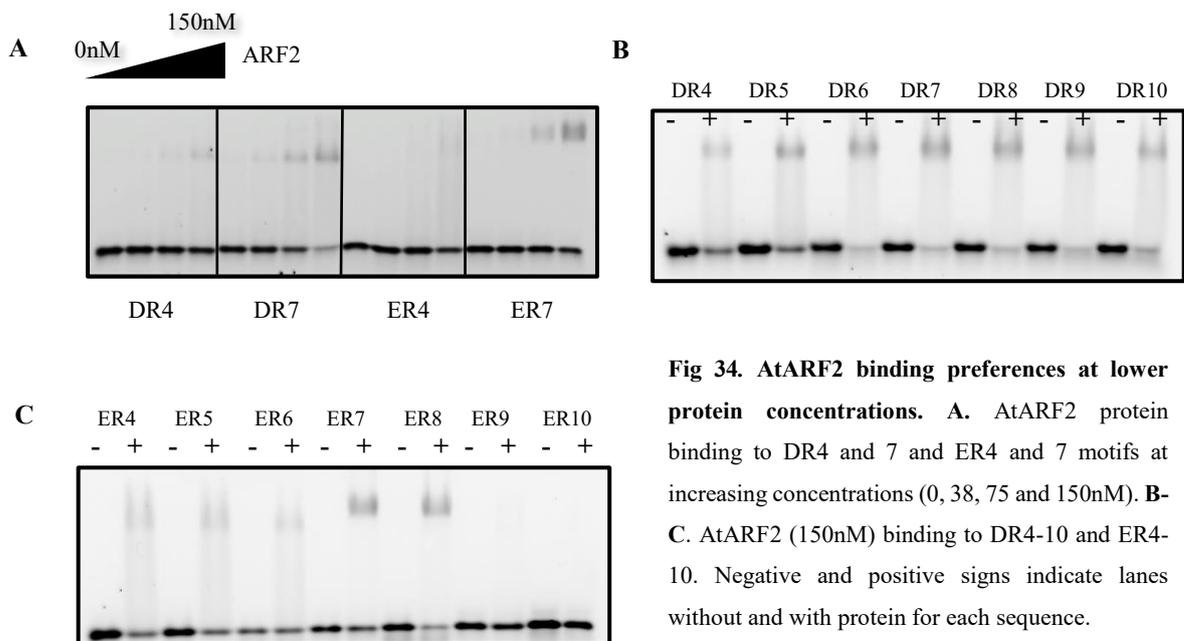
In agreement with the bioinformatic analysis, AtARF2 and AtARF5 bound with a clear preference to ER7/8 motifs among the ER set, although smeared shifts could also be observed for ER4-6 DNA sequences (Fig 28; 33.A-B). AtARF5 binding to DR motifs also fit the DAP-seq analysis predictions, with a preferential binding to DR4-6 and DR15 sequences (Fig 28; 33.B). Although AtARF2 was not enriched in DR sequences in the bioinformatic analysis, EMSAs revealed that this protein bind DRs independently of the spacing (Fig 28; 33.A).

We wondered if AtARF2 not fitting the DR binding predictions could be due to too high protein concentrations, so we tried to optimise the binding conditions by testing first, a lower concentrations range for AtARF2. AtARF2 did not bind the same to differently spaced DRs at lower concentrations (Fig 34.A-B), showing a slightly better interaction for DR6-9 (Fig 34.B). However, when comparing AtARF2-DR7 and AtARF2-ER7 EMSAs (Fig 34.A) we could not observe clear differences, indicating that at least in the context of individual DRs, AtARF2 can bind this type of motifs with the same intensity as ER7/8 (Fig 34. A-C).

Finally, we wondered what would happen with other types of ARF proteins. AtARF2 and AtARF5 belong to classes B and A, respectively. Only AtARF16 genomic binding data exist for class-C ARFs (O'Malley R.C. et al., 2016). However, when analysing these data, they were not appropriate enough for retrieving a good PWM binding model. Instead, we decided to test AtARF10 *in vitro* binding to the already tested DR and ER set as a first analysis for a class-C ARFs.

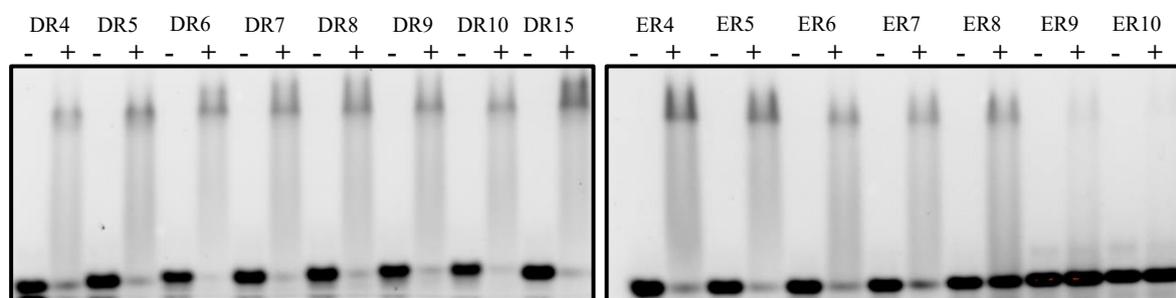


**Fig 33. *In vitro* validation of DR/ER spacing preferences.** DR4-10 and DR15, ER4-10 tested in their binding to AtARF2 (250nM) (A) and AtARF5 (500nM) (B). Negative and positive signs indicate lanes without and with protein for each sequence.



**Fig 34. AtARF2 binding preferences at lower protein concentrations.** A. AtARF2 protein binding to DR4 and 7 and ER4 and 7 motifs at increasing concentrations (0, 38, 75 and 150nM). B-C. AtARF2 (150nM) binding to DR4-10 and ER4-10. Negative and positive signs indicate lanes without and with protein for each sequence.

The DBD of AtARF10 presents a 36-amino acids insertion (Ulmasov et al., 1999) in one of the helices from the dimerization interface<sup>3</sup>, which might impede dimerization or mediate a different dimerization mode, with potential different DNA binding preferences. Indeed, AtARF10 behaved similarly to AtARF2 when tested against DR4-10 plus DR15 (Fig 35.A), being able to equally bind all DRs, but it showed a different ER binding pattern than the one observed for AtARF2 and 5's. AtARF10 presented the strongest binding to ER4-5, being also able to interact with ER6-8. No interaction was observed with ER9-10 (Fig 35.B). AtARF10 could not bind single sites (not shown), proof that, as other ARFs, this protein only binds as a dimer to AuxRE repetitions.



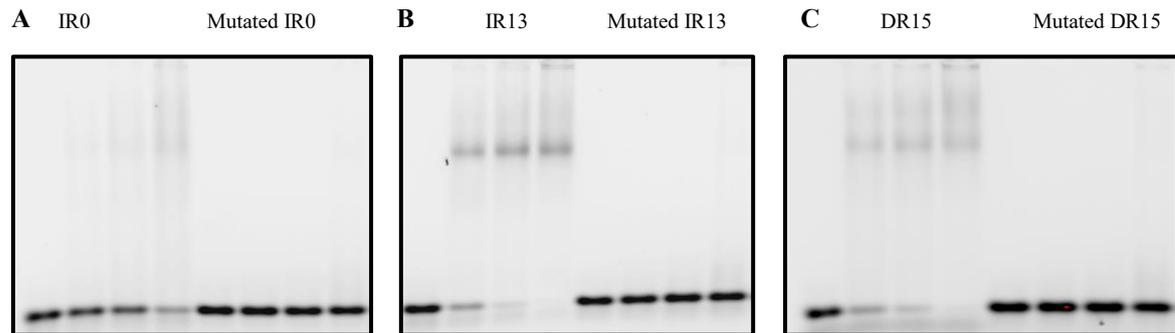
**Fig 35. AtARF10 DR/ER preferences.** DR4-10 and DR15 (left) and ER4-10 (right) tested in their binding to AtARF10 (250nM). Negative and positive signs indicate lanes without and with protein for each sequence.

### Binding to different motifs originates different complexes

Our DAP-seq data analysis indicated that AtARF5 is much more flexible than AtARF2 being enriched in DR, ER and IRs sequences with different spacings, whereas AtARF2 was almost uniquely enriched in ER7/8 sites. Moreover, we observed that for all three AuxRE configurations, DR, ER or IR, there was a periodicity of +10bp in the enriched spacings.

EMSAs were done in order to test AtARF5 binding to the most represented sequences IR0, IR13 and DR15. The protein could bind in a specific manner to IR13 and DR15 at low concentrations, whereas IR0 binding showed a weaker affinity (Fig 36).

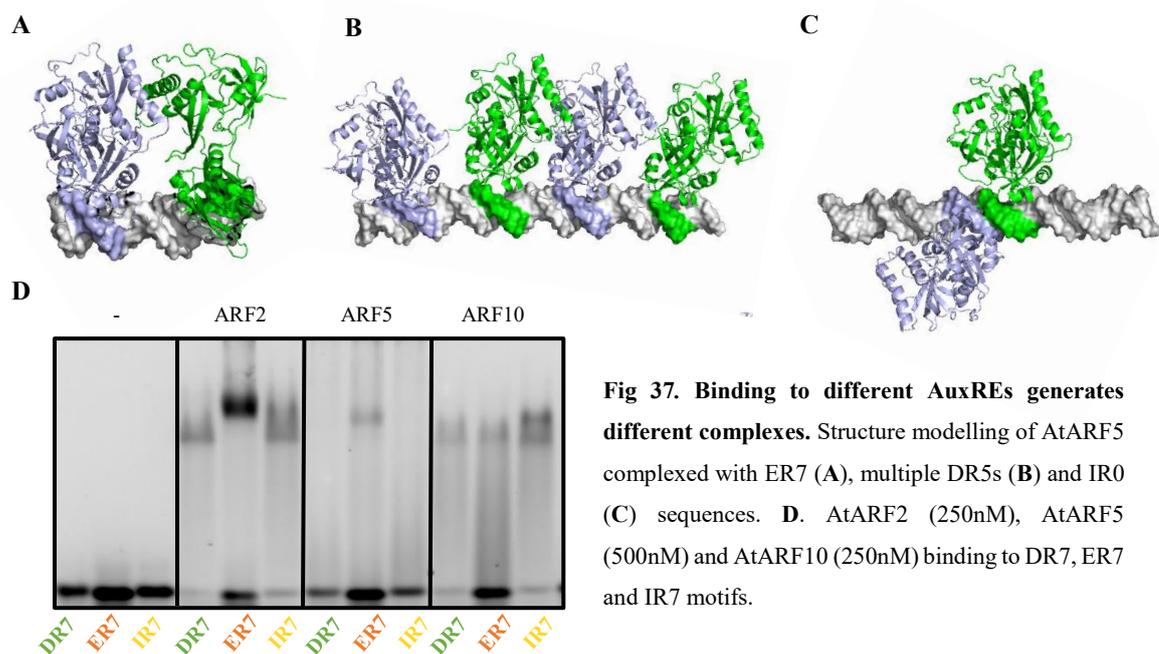
<sup>3</sup> More detailed information about class-C ARFs DBD will be given in Chapter III.



**Fig 36.** *In vitro* test of AtARF5 binding to the new most enriched sequences: IR0 (A), IR13 (B) and DR15 (C). For each AuxRE motif a mutated version of it was used as control. Tested AtARF5 increasing protein concentrations (0, 125, 250 and 500nM).

Despite being able to bind to DR and IR sequences, as proved by DAP-seq data and EMSA tests, the orientation of the ARF BSs in these sites is not compatible with the dimerization mode of the only available structure: AtARF1 bound to ER7 (Fig 37.A). Modelling of AtARF5 binding to DR or IR indicates different binding modes and dimers formation (Fig 37.B-C).

Indeed, when testing all three ARFs, AtARF2, AtARF5 and AtARF10 binding to DR, ER and IR sequences on the same gel we could observe a different migration depending on the type of motif, mostly evident for AtARF2 (Fig 37.D). None of these proteins could bind monomeric sites (not shown) which rules out the option of a monomeric DNA-protein interaction as an explanation for



**Fig 37.** Binding to different AuxREs generates different complexes. Structure modelling of AtARF5 complexed with ER7 (A), multiple DR5s (B) and IR0 (C) sequences. D. AtARF2 (250nM), AtARF5 (500nM) and AtARF10 (250nM) binding to DR7, ER7 and IR7 motifs.

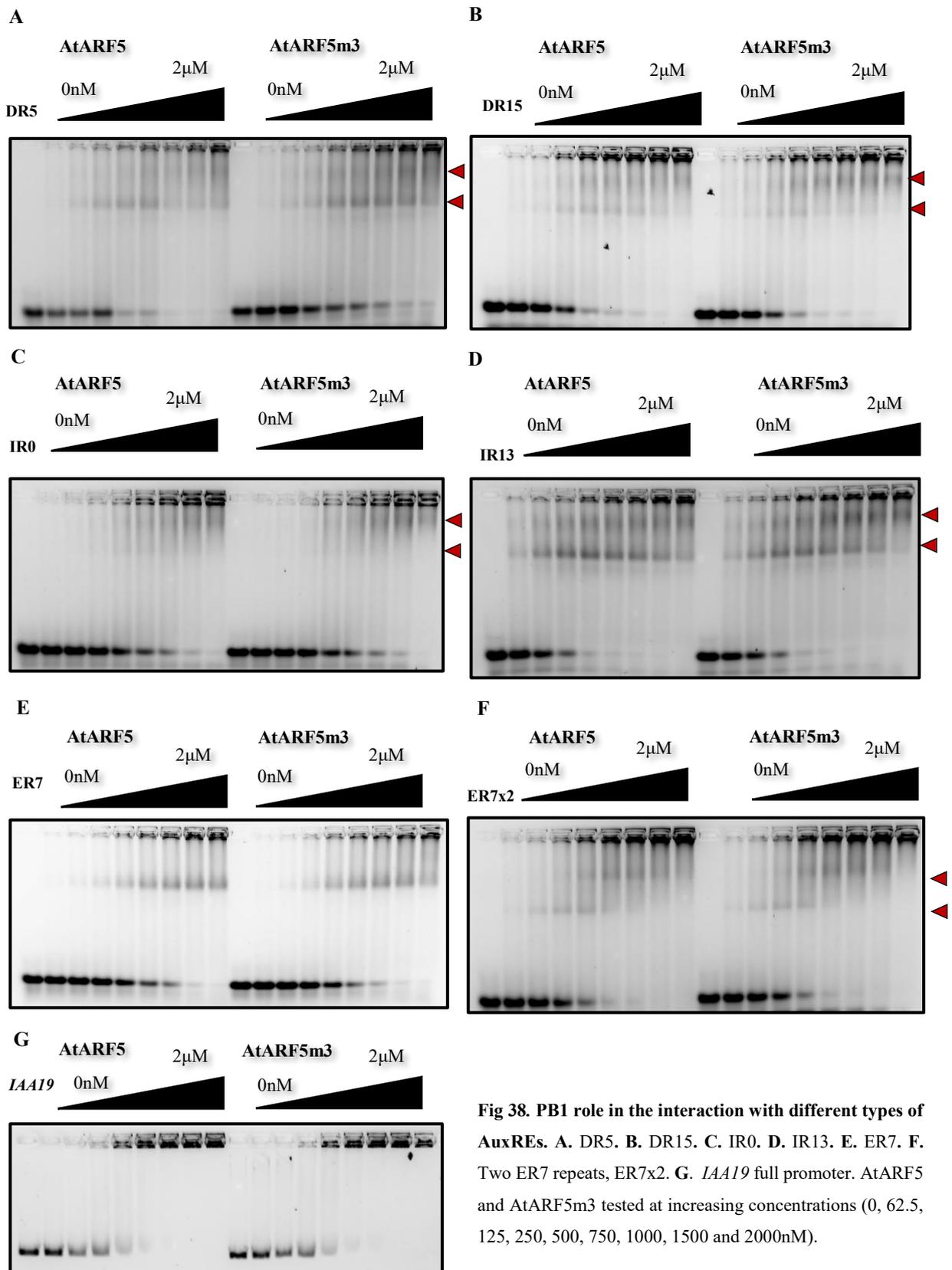
the different complexes formed. Different dimerization interfaces must be thus, mediate the binding as dimers of ARF proteins to DR and IR motifs.

Apart from the DD inside the DBD, the C-terminal PB1 domain is the only other known homodimerization surface described till now for ARF proteins. Although for some ARFs PB1 domain deletions or mutations affect the stability of the DNA binding and auxin response (Pierre-Jerome et al., 2016; Ulmasov et al., 1999) in other cases, the binding to DNA was independent of the PB1 presence or absence (Ulmasov et al., 1999). To understand the potential role of the PB1 domain in the binding to DR and IR sequences, we tested AtARF5 binding in comparison with AtARF5m3 (K797S-D857S in PB1 positive and negative interface), a mutant impaired in oligomerization due to mutations in its PB1 positive and negative interfaces (Nanao et al., 2014). AtARF5 and AtARF5m3 behaved the same in the binding to DR<sup>5</sup> and IR0 motifs (Fig 38.A, C), indicating that the PB1 does not influence the AtARF5 dimerization mode that mediates DR and IR interactions. Both proteins were also equally able to bind more spaced motifs, DR15 and IR13, making us discard the option of AtARF5 PB1 having a role in facilitating the interaction with long-spaced sites (Fig 38.B, D). All 4 DNA sequences seemed to give two different complexes for both AtARF5 wt and the mutant.

Since the PB1 domain does not seem to play a role for binding to DR or IR, we wondered if it could influence the binding to multiple double sites. Thus, we tested AtARF5 and AtARF5m3 binding to ER7 (for comparison with ER7x2), ER7x2, formed by two repetitions of ER7 and *IAA19* promoter. Again, no differences were observed between the two proteins in the binding to none of the DNA sequences (Fig 38.E-G). AtARF5 wt and mutant binding to ER7 resulted in the formation of a single complex, whereas ER7x2 binding pattern indicated the sequential formation of two complexes, most likely a dimer and a tetramer, with the tetramer being the only resulting form at high protein concentrations (Fig 38. E-F). *IAA19* DNA was bound in a cooperative way by both wt and AtARF5m3 proteins, making us discard the option of the PB1 domain having a role in ARF5 cooperative binding to multiple low-affinity sequences (Fig 38.G)

---

<sup>4</sup> By looking at free DNA for DR5 it can be interpreted that there are slight differences between AtARF5 and AtARF5m3. Replicas of this experiment supported that both proteins behaved the same, as indicated in the text.



## Discussion

ARF transcription factors, implicated in the auxin response, translate the auxinic chemical signal into changes in gene expression that end up in an appropriate auxin response. In higher plants, *ARF* genes are numerous. Traditionally divided into activators and repressors, ARFs, as every transcription factor, possess a DBD that mediates interactions with DNA. Despite the high sequence homology among DBDs of different ARFs, several studies point out that different ARFs could prefer different AuxRE motifs. The details of these preferences and the potential structural or biochemical determinants of them are still to be discovered. In the second Chapter of this manuscript we brought AtARF2, a repressor ARF, face to face to the famous activator, AtARF5, in the search of differential binding preferences among them, through a combination of DAP-seq massive data analysis, *in vitro* DNA binding tests and structural modelling.

### TGTC, a search tool to be reconsidered

Although “TGTC” motifs have been traditionally used as THE tool for the search of ARF binding sites and auxin responsive genes, they have been proven not to be the optimal prediction instrument of auxin transcriptional response (Keilwagen et al., 2011). Taking advantage of the available AtARF2 and AtARF5 genomic DAP seq data (O’Malley R.C. et al., 2016), we constructed PWMs for a better prediction of these transcription factors binding sites. The PWM for AtARF2 and AtARF5 revealed some differences in the preferential binding of these two proteins as monomers and showed that variations within the TGTC could be accepted. To confirm our new ARFs binding models we used the ER8 site located inside *IAA19* promoter, constituted by a TGTC sequence in the sense strand and a GGTC non-consensus sequence in the antisense strand. Our EMSA assays done with ER8-*IAA19* motif and alterations of it confirmed that ARFs can bind sequences other than the TGTC consensus, and that nucleotides flanking the binding site importantly contribute to a good affinity binding. Therefore, AtARF2 and AtARF5 PWM developed in this study provide a better prediction tool for the understanding of auxin transcriptional response.

### Different ARFs, different ways

Although AtARF2 and AtARF5 monomeric binding presented some particularities, DAP-seq data treatment revealed that both proteins behave differently mostly when bound as dimers. DAP-seq analysis leads us to define AtARF5 as a rather flexible protein, able to fit the three types of double AuxREs with several spacings: DR4-6, ER7-8, IR0-3, re-enriched with a periodicity in the spacing of 10bp (DR4-6, DR14-16, DR25-26; ER7-8, ER17-18, ER28; IR0-3, IR12-13, IR22-23). On the contrary, AtARF2 was more discriminating, being only enriched on ER7/8 motifs, results that

altogether, coincide with O'Malley's *et al.* predictions except for their confusion between ER and IR.

In agreement with our DAP-seq analysis of enriched motifs for AtARF2, we observed that DR mutations decreased AtARF2 binding to a lesser extent than ER8 mutations in EMSAs done with an 80bp-fragment within *IAA19* promoter. Same goes for AtARF2 binding to the full *IAA19* (350bp) promoter, in which DR mutations barely affected the interaction. It is worth noting that none of the mutations caused striking effects, indicating that the ARFs binding happens with multiple sites rather than a single high affinity site leading the binding of the protein.

AtARF5 behaved differently when binding to *IAA19* full promoter. First of all, instead of gradually shifting DNA into different size complexes, as it was observed for AtARF2, AtARF5-*IAA19* promoter suddenly formed large complexes at high protein concentration, suggesting a possible cooperativity in the binding of this protein that was not observed for AtARF2. Moreover, AtARF5 binding to *IAA19* promoter was unaffected by any of the mutations introduced in the ER and DR motifs. For this protein, an extra cluster of low-score sites was predicted in the 3' of *IAA19* promoter, that were not predicted for AtARF2. Altogether, this suggests that in the genomic context, clusters of low affinity binding sites might play a fundamental role for AtARF5 binding, that could manage to interact with low-score sites with high affinity thanks to a cooperativity mechanism.

Whilst working with long, natural, genomic sequences seemed to fit the DAP-seq analysis, AtARF2 and AtARF5 behaved differently when binding to unique, individual sites. EMSAs done with designed AuxREs containing the TGTC consensus indicated that for ERs, a 7/8 spacing was preferred by both AtARF5 and AtARF2 proteins, in line with DAP-seq ER enriched sequences. AtARF5 binding to DRs was also consistent with DR4-5 and DR14-15 DAP-seq enriched motifs. However, in opposition to the bioinformatic studies, AtARF2 could bind DRs with a similar affinity to ER7 no matter the spacing. This can be due to several reasons, being the first one -plane and simple - differences among the production and purification methods used in O'Malley's study and ours. Second, the individual DRs used in EMSA analysis are optimal binding sites, not necessarily representative of the DRs present in the genome. Third, DAP-seq analysis indicates that AtARF2 does bind DR motifs, although the enrichment is much lower than for ERs and no spacing is overrepresented (Fig 28, left square). Finally, the EMSAs done with natural fragments suggest that binding to one site can be dependent on the sites around. Therefore, AtARF2 EMSAs not fitting DAP-seq inferences could be just a consequence of TFs-DNA game being genome-context dependent: AtARF2 might perfectly bind individual DR motifs diversely spaced, but inside the immensity of the whole genome, where low-score DRs are diluted among other binding sites, AtARF2 will preferentially bind ER7/8.

DAP-seq and EMSAs both employ recombinant transcription factors and naked DNA that are bound to each other *in vitro*. Our data suggest that despite their technical similarities, different results can come out of the two methods, most likely due to the genomic nature of the DAP seq analysis. This suggests that using labelled natural fragments in EMSAs or other DNA-protein interaction methods could provide more relevant results.

### **ARFs, molecular calipers with different arrangements**

Several studies have proven the capacity of ARF proteins for binding ERs and DRs with different spacing (Boer et al., 2014; Pierre-Jerome et al., 2016; Ulmasov et al., 1997a). For this, ARFs have been dubbed as “molecular calipers” (Boer et al., 2014). The only available structure of an ARF in complex with DNA is that of AtARF1-DBD binding to an ER7 motif. The position of the two monomers perfectly fits the orientation and spacing of the ER7 sequence, with only a slight rotation of the B3 domain taking place upon DNA contact in comparison with the apoprotein structure (Boer et al., 2014). However, it is hard to imagine the determined AtARF1 dimer structure upon DR and IR motifs due to the disposition of the two binding sites. We modelled AtARF5-DBD structure in complex with ER7, DR5 and IR0 motifs, that came out among the most represented sequences in our DAP-seq analysis. Our models suggest that AtARF5 binding as dimer to DR and IRs must happen through different dimerization interfaces than the one mediating the binding to ER motifs. Different complexes were also observed in EMSAs done with DR7, ER7 and IR7, in support of this theory. The adoption of different conformations could mediate AtARF5 binding to clusters of multiple sites composed by a combination of DR, ER and IRs as the ones predicted in *IAA19* promoter. Moreover, the AtARF5-DR complex indicates that for DRs, the disposition of the monomers upon the DNA could be compatible with AtARF5 spreading along chained DRs in the genome, in opposition with the binding to ERs, in which the dimerization mode limits the complex to the formation of a dimer at least when referring just to the DBD.

Not only AtARF5 can bind all three described types of AuxREs but also diverse spacings. Although a first analysis indicated that 4-6 bp spacings were preferred for DR conformations, 7-8 for ERs and 0-3 for IRs, widening the AuxREs search window revealed a periodicity of plus 10bp for the three types of motifs. 10bp corresponds to a full DNA helix turn. This implies that for a double AuxRE (DR, ER or IR) the second binding site will be pointing towards the same side of the DNA double strand with a spacing of “n” nucleotides than with a spacing of “n+10”, “n+20”, etc (Fig 39.A). Therefore, AtARF5 dimers should be able to “open” and “close” to a certain extent that admits all the enriched spacings (Fig 39.B-C). However, the fact that no intermediate spacings between “n” and “n+10” are bound indicates that dimers structure is not compatible with turns around the DNA (Fig 39.D). Boer *et al.* suggested, based on their structural data and ARFs

sequences comparison, that ARFs caliper capacity was due to the potential of their B3 domain to move around, which in turn was dependent on the loop that connects the B3 with the DD. This loop is long, flexible and, in opposition to the B3 or DD, presents a low residues conservation among the different ARFs, implying that depending on the length and amino acidic composition of the loop for each ARF, the rigidity of its structure and thus, its DNA preferences, might change.

The enriched spacings can also be explained by an allosteric binding to DNA. Recent studies suggest that the binding of a TF to DNA major groove can favour a second protein binding by widening of the next major grooves, that is to say, at a distance of 10 bp, 20bp, etc from the first binding site (Kim et al., 2013). In the most enriched configurations for AtARF5, DR5, ER7 and IR3, the spacings correspond to a distance of 10bp between two ARFs binding as dimers. Therefore, binding of a first ARF monomer could induce DNA structural changes leading to the binding of a second monomer in the next major groove and thus, the formation of the dimeric complexes. Moreover, since the allosteric binding can take place also within a 20bp distance, this mechanism could also favour the DR15, ER17 and IR13 configurations.

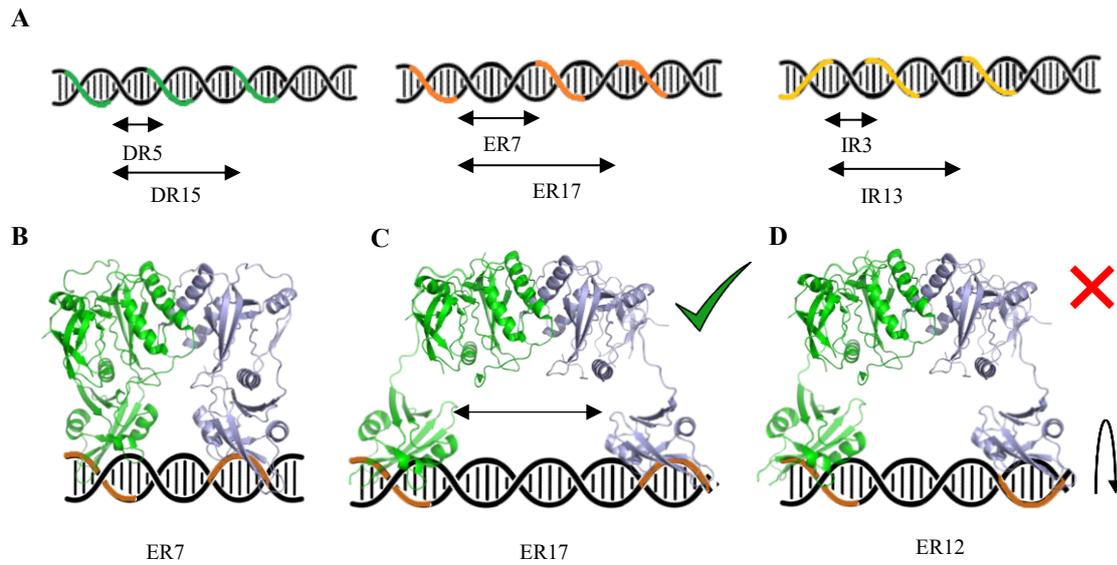
Finally, other ARFs could present a different behaviour still. Class-C ARFs (AtARF10, 16 and 17) present an insertion inside one of the helices of the DD that could induce a different dimer conformation with different DNA specificities. We tested AtARF10 in EMSA assays and we observed that, indeed, the binding was stronger with ER4-5 than for ER7-8 as it was the case for AtARF2 and AtARF5. It would be interesting to test if these binding preferences are generalized to all C-ARFs.

### **ARFs-DNA preferences and configurations, auxin tool for a variety of transcriptional responses?**

Although the complexity of the auxin response is attributed to 3 families of proteins, TIR1/AFB, Aux/IAA and ARFs, once the hormone is perceived by the receptors, it triggers the degradation of Aux/IAs and so, the “only thing left” to finish auxin job are the ARFs. ARFs and the regions in the genome where they are bound must be then, the ultimate determinants of the specificity of the auxin response. Here, we proved that although most regions in the genome are exclusively bound by AtARF2 or by AtARF5, there is also a percentage of regions that can be bound by both proteins. These regions are mostly enriched in ER7/8 motifs, preferential binding sites for AtARF2 and AtARF5 but that can also be bound and respond to auxin when bound by other ARFs (Boer et al., 2014; Pierre-Jerome et al., 2016; Ulmasov et al., 1997a, 1999).

On the other hand, we know that simultaneous expression of different ARFs (activators and repressors) takes place and is needed for an appropriate auxin response inside a cell (Farcot et al.,

2011; Rademacher and Mo, 2011). Therefore, the auxin chemical signal could be translated into a complex and robust transcriptomic response by the combination of ARFs synthesized inside a specific cell, where some of them might compete for ER7/8 sites upon certain genomic regions, bind simultaneously to them forming heterodimers or bind to their preferential sites inside the genome.



**Fig 39. 10bp periodicity in ARFs binding.** **A.** Position of ARF5 binding sites for the highly enriched sequences, DR5, ER7 and IR3 and the corresponding +10bp spacings, DR15, ER17 and IR13 the second binding site is orientated in the same way when spacing it 10bp more. **B-D.** AtARF1 DBD dimers upon differently spaced ERs showing how the dimer could be compatible with +10bp periodicity in the spacing by a more open disposition of the B3 domains, whereas intermediate spacings would need a turn of the B3 domain around the DNA helix.

Moreover, we hypothesize that different ARFs have different binding modes at the full genome scale that could influence auxin transcriptional response. First, more sites, although with worse scores were predicted for AtARF5 inside *IAA19* natural promoter than for AtARF2. Second, for this specific promoter, both proteins did not bind in the same way, with AtARF5 binding to *IAA19* promoter suggesting a cooperative mode to low-affinity sites. In the presence of the hormone, once AtARF5 is free from Aux/IAs, it might expand along longer regions in the genome amplifying the auxin transcriptional response. Cooperativity is a common mechanism for transcription factors binding that increases the specificity of the binding (Morgunova and Taipale, 2017). For several transcription factors, DNA binding cooperativity happens through their oligomerization domains (Katsani et al., 1999; Sayou et al., 2016). This did not seem to be the case for AtARF5 since a mutant in its PB1 domain was equally able to bind multiple site sequences and *IAA19* promoter. However,

studies carried out in a yeast system have proven the importance of this domain in activator ARFs for auxin transcriptional response (Pierre-Jerome et al., 2016). DAP-seq experiments done with AtARF5 deleted or mutated in their PB1 domains could help clarifying this matter since they would provide information at the genomic scale.

Finally, another level of the specificity of the auxin response could come from interactions with other partners or posttranslational modifications (Hill, 2015; Mironova et al., 2014; Oh et al., 2014b). Structural models revealed that ARFs binding to different AuxREs orientations happens through different dimeric conformations. ARFs are known to bind TFs involved in other signalling pathways, for the integration of several signals (Cho et al., 2014; Mironova et al., 2014; Oh et al., 2014b; Vert et al., 2008) and to co-regulators (TPL, LUG for repressors and SYD/BRM for activators) (Causier et al., 2012a, 2012c; Lokerse and Weijers, 2009; Wu et al., 2015a) for the alteration of the chromatin structure. ARFs different dimeric configurations upon different types of AuxREs could allow or avoid the binding to these other proteins, contributing to a more specific auxin response.

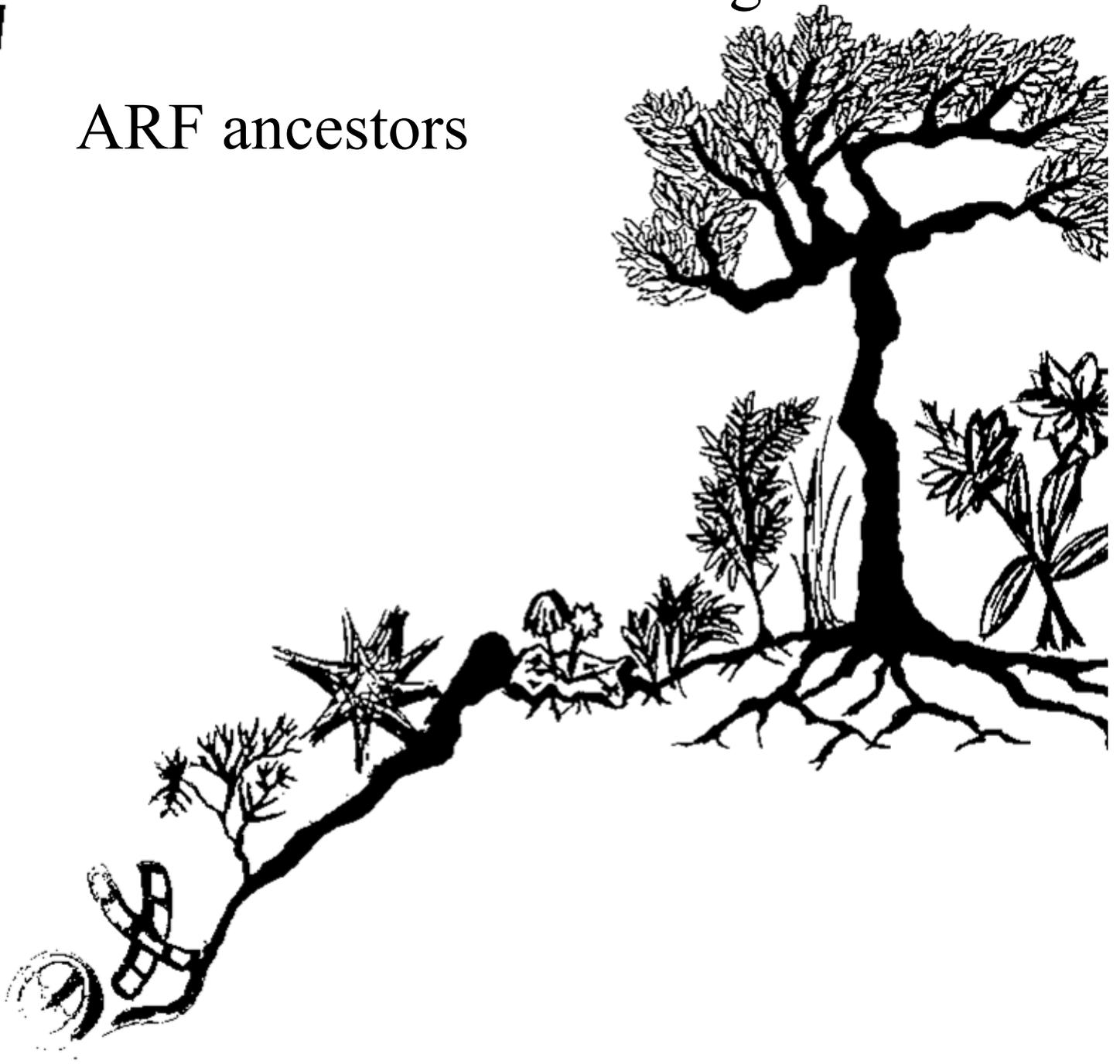
## Conclusions

How a single molecule, such as auxin, diverges its pathway in such a diversity of transcriptional responses is a brain teaser that could find its answer there where the auxin journey ends: the ARFs. Despite their multiple structural and sequence similarities, different ARFs possess different DNA binding capacities and therefore, understanding where and how they bind to regions in the genome can help unravelling the secrets of auxin transcriptional response. DAP-seq technology has provided us with an invaluable method to understand this subject, and thanks to it we have dug here into the DNA likes and dislikes of two different ARFs. Our results, combined with auxin transcriptional profiles, could provide a better knowledge of how different types of ARF organise themselves when co-expressed inside a cell to give an appropriate and robust auxin transcriptional response.



# Chapter III

Back to the water looking for  
ARF ancestors

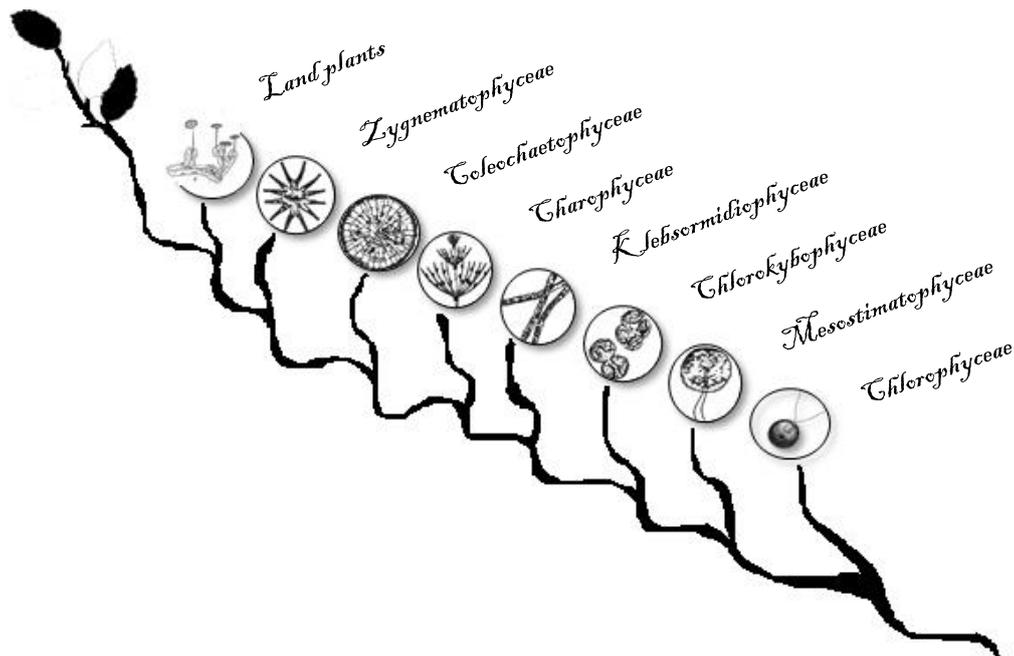




## Introduction

The transition of plants from aquatic to terrestrial life was one of the most striking events in the history of our planet. Land plants altered the atmospheric and soil conditions of the Earth leading to the appearance of other life forms and later on, to the agricultural activity, determinant in human history.

It all started around 450 m.y.a...



**Fig 40. Once upon a time, the land plants.** Tree representing the evolution of land plants from charophyte algae and the 6 successive evolutionary clades described within charophytes, that diverged from Chlorophyte algae: Mesostigmatophyceae, Chlorokybophyceae, Klebsormidiophyceae, Charophyceae (Charales), Coleochaetophyceae and Zygnetatophyceae.

### Charophyte organisms: from water to land

Green algae are defined as photosynthetic eukaryotes with plastids wrapped inside a double membrane and that contain chlorophyll a and b. They are a very diverse group of organisms. From swimming to non-motile, microscopic to 1 meter-length, unicellular, filamentous or branched, green algae are ubiquitous organisms in aquatic and terrestrial environments (Leliaert et al., 2012; Lewis and McCourt, 2004).

Within the green algae two evolutionary lineages have been described: chlorophytes and charophytes. Whereas the formers are distantly related to land plants, charophytes organisms are proposed as the their direct ancestors (Fig 40) (Delwiche and Cooper, 2015).

Charophytes inhabit mainly freshwater and moist terrestrial environments (Delwiche and Cooper, 2015; Leliaert et al., 2012). 6 classes of charophytes have been described till now and divided in two groups: the “early divergent” and the “late divergent” charophytes (Domozych et al., 2016). The clades *Mesostigmatophyceae*, *Chlorokyboceae* and *Klebsormidiophyceae* conform the “early divergent” charophytes. They present relatively simple body structures, although multicellularity is already observed in *Chlorokybus* and *Klebsormidium* genera. On the contrary, the “late divergent” charophytes present much more complex body shapes and sexual reproduction (Fig 41). They are constituted by the classes Charophyceae (also called Charales), Coleochaetophyceae and Zygnematomphyceae (Domozych et al., 2016; Lewis and McCourt, 2004). Due to their morphological and developmental similarities with land plants, Charales were always considered as their immediate ancestor (Delaux et al., 2012), but recent molecular studies propose Coleochaetales and Zygnematales as the sister to terrestrial plants (Bowman, 2013; Timme and Delwiche, 2010; Wodniok et al., 2011) .

Water to land transition was initiated when a charophyte lineage underwent an evolutionary change that allowed it to start handling the harsh terrestrial conditions, such as limited water and carbon supplies, light stresses, strong temperature changes and attacks from new microbes (Delaux et al., 2012; Delwiche and Cooper, 2015). Small but tough, charophytes could go through all these dramatic changes by the acquisition of certain adaptational cues, among them, the appearance of cuticles, rhizoids and the development of multicellular structures that provided advantages such as cell-type specialization and later on, organ diversification (Fig 41) (Bowman, 2013; Delaux et al., 2012; Umen, 2014).

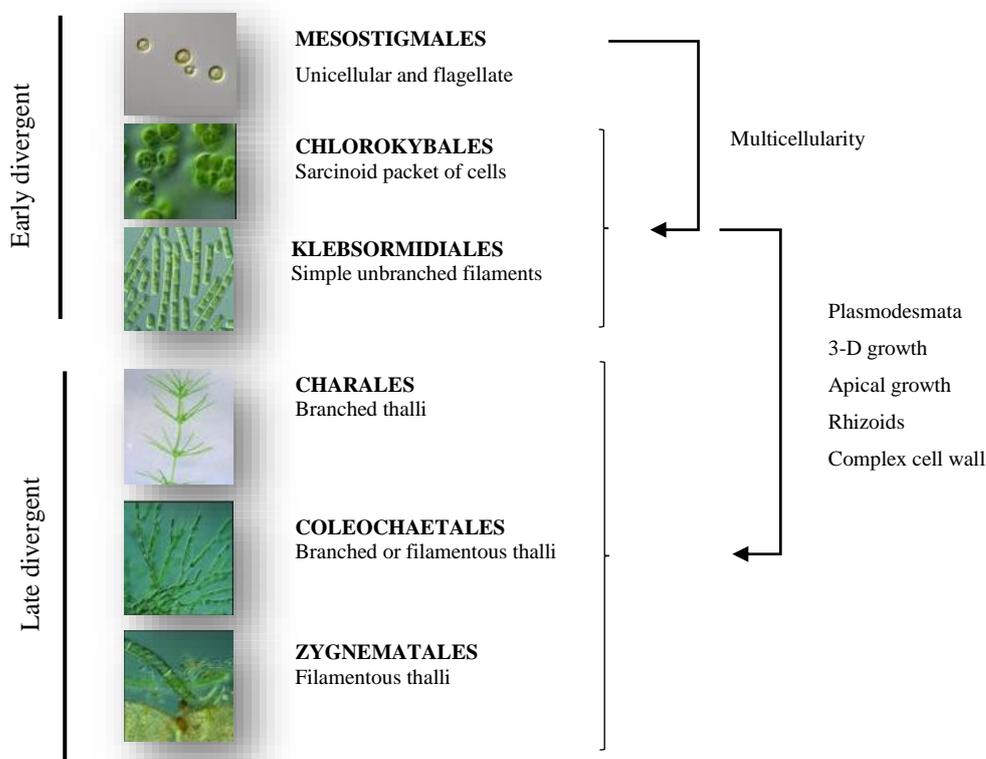
Multicellularity started in its simplest forms in the “early divergent” charophytes, with sarcinoid packets of cells enclosed inside a matrix in the case of *Chlorokybus* or multicellular non-branched filaments in the case of *Klebsormidium*. “Late divergent” charophytes evolved more and more complex structures and developmental patterns that resemble those found in land plants (Umen, 2014) (Fig 41).

### **B3 family**

The physiological and morphological changes that took place during water to land transition were accompanied by an expansion of several transcription factor families with potential importance in the recently acquired adaptations (Lehti-Shiu et al., 2017; Leliaert et al., 2012).

Some of the transcription factors found in plants are shared with other eukaryote organisms whereas others are specific to plants. Among the latter, the B3-family of transcription factors, not only present in land plants but also in green algae, is one of the most represented ones (Lehti-Shiu et al., 2017). B3-containing proteins have been traced back till chlorophyte organisms, in which only one member belonging to the ABI3 family has been identified in *Chlamydomonas reinhardtii* and *Volvox carterii* (Romanel et al., 2009; Wang et al., 2012). Gene duplication events from *ABI3* genes are supposed to

have given rise to the RAV, ARF and REM B3-families that appeared later in evolution and that must have had a fundamental role in the adaptation to land due their current implication in developmental and hormonal responses (Romanel et al., 2009; Swaminathan et al., 2008).



**Fig 41. Charophyte organisms and the acquisition of new traits towards the conquer of land.** Multicellularity in different forms was already present in the “early divergent” charophytes whereas more complex features such as rhizoids or apical growth appeared with the “late divergent” charophytes (Adapted from (Harrison, 2016; Umen, 2014a)).

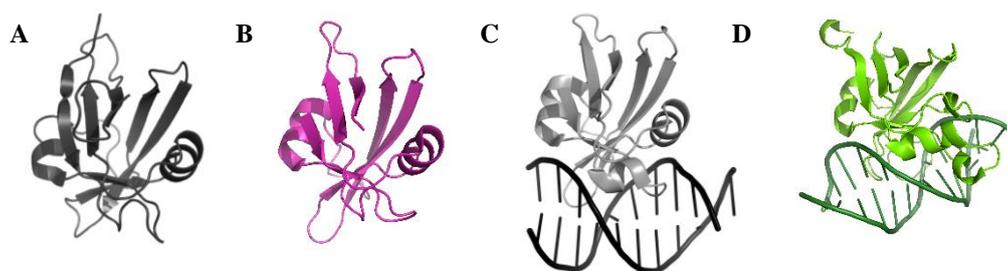
B3 domains, first identified in the maize gene VIVIPAROUS1 (VP1) as the third domain of this protein (being the first two B1 and B2), are composed of around 110 amino acids (Swaminathan et al., 2008). 4 families of B3-containing transcription factors exist in land plants (Romanel et al., 2009; Yamasaki et al., 2013):

- LAV (LEC2 [LEAFY COTYLEDON 2]/ABI3 [ABSCISIC ACID INSENSITIVE 3] – VAL [VP1/ABI3-LIKE]) family composed by two subfamilies: ABI3 and VAL. ABI3 proteins are implicated in ABA responses and seed maturation, whereas VAL proteins are transcriptional repressors involved in vegetative processes (Wang et al., 2012). The repressor activity of VAL proteins is due to the presence of a DLNxxP EAR motif (Tsukagoshi et al., 2007).
- RAV (RELATED to ABI3/VP1) family: it counts with 13 members in *A. thaliana*. Among them, 6 present an additional AP2/ERF (APETALA2/Ethylene Response Factor) DNA binding

domain. Proteins of the RAV family participate in cold stress responses and flower initiation (Fu et al., 2014; Matías-Hernández et al., 2014; Swaminathan et al., 2008; Wang et al., 2012). Some members of this family also present EAR-like motifs (RLFG-type) that interact with TPL (Causier et al., 2012c).

- ARF (Auxin Response Factor) family: old acquaintance in this manuscript already, ARF family members are known for their implication in the regulation of auxin transcriptional response.
- REM (REPRODUCTIVE MERISTEM) family: the most diversified and still the less known B3-family members. They present up to 6 B3 domains (Yamasaki et al., 2013). They have been implicated in the control of flowering time and in the development of sexual organs (Wang et al., 2012).

Several structural studies have allowed to decipher the DNA binding mechanism of transcription factors of the B3-families (Boer et al., 2014; Golovenko et al., 2014; Yamasaki, 2004). Members of the different families share a common structural skeleton with 7  $\beta$ -strands connected in an antiparallel manner and 2  $\alpha$ -helices located in between the  $\beta$ -strands (Yamasaki et al., 2008). The interaction with the DNA takes place through two loops that connect the  $\beta$ -strands and that go into the DNA grooves (Fig 42). The structural similarities between plant B3 domains and the DNA binding domains of the bacterial endonucleases EcoRII and BfiI-C led to the hypothesis that B3-plant specific transcription factors derived from these proteins either by inheritance or more likely by horizontal transfer (Golovenko et al., 2014; Romanel et al., 2009; Swaminathan et al., 2008).



**Fig 42. B3-domain structures.** **A.** AtRAV1-B3 (1WID, (Yamasaki, 2004)). **B.** AtARF5-B3 (4LDU, (Boer et al., 2014)). **C.** AtARF1-B3 in complex with AuxRE binding site (4LDX, (Boer et al., 2014)). **D.** Bfi-I endonuclease from *Bacillus firmus* in complex with DNA (3ZI5, (Golovenko et al., 2014)).

Whereas the residues involved in the establishment of the core B3 structure are highly conserved in the different B3 families (Yamasaki, 2004), the amino acids present in the DNA-interacting loops are specific for each B3-type (Table 1). As a consequence, binding to DNA is sequence-specific for each B3-family (Golovenko et al., 2014; O'Malley R.C. et al., 2016; Swaminathan et al., 2008). AP2/B3 members of the RAV family present the particularity of binding to bipartite DNA sequences composed

by AP2 and B3-RAV binding sites with a preferential spacing of 5 nucleotides in between (Kagaya et al., 1999). Similarly, ARF proteins bind as dimers to double DNA AuxRE sequences, as extensively explained in the previous chapter (Boer et al., 2014).

**Table 1. DNA-binding specificities and residues responsible for them in each B3-family.**

B3-type	Residues mediating specific DNA binding	DNA binding sequence	Members in <i>A. thaliana</i>
ABI3	WPNNKSR	5' CATGCA 3'	6
RAV	WN/RSSQS	5' CACCTG 3'	13
ARF	RGQPK/RR	5' TGTCTC 3'	23
REM	3/ 4 residues deletion	polyA	45

### Auxin signalling: from bryophytes to flowers

As previously mentioned, one of the main innovations in the conquest of land by plants was a higher complexity in body plans with specialized cells and organs. Due to the relevance of auxin in almost every step of land plants development, the acquisition of the auxin signalling machinery must have been a milestone in the transition and adaptation to the terrestrial life. However, where to position this event in the evolution of land plants remains an issue due to the lack of charophytes genomic data.

When looking into bryophyte organisms, right at the beginning of land plants evolution, we can already find the first evidences of the Nuclear Auxin Pathway (NAP), main responsible of the auxin transcriptional response. One example of it is the liverwort *M. polymorpha*. With a very simple body structure, composed by a leaf-like thallus with dorsi-ventral polarity from which sexual organs grow towards the aerial part and rhizoids towards the ground, it is considered as one of the earliest extant land plants. Due to the simplicity and low redundancy of its genome, *Marchantia* is emerging as a new plant model organism. Recent studies done upon this plant have shown that the full NAP is already present in *Marchantia* with only one copy of TIR1, Aux/IAA and TPL (MpTIR, MpAux/IAA and MpTPL). Regarding the transcription factors, 3 ARFs were identified in *M. polymorpha*, each of them belonging to the three evolutionary clades, A, B and C (MpARF1, 2 and 3 respectively). Moreover, these proteins were shown to function in the same way as the TIR1/Aux/IAA/ARF transcriptional response (Flores-Sandoval et al., 2015; Kato et al., 2015).

Even if its body structure is not more complex that *Marchantia*'s, the moss *P. patens* presents already several copies of some of the components of the auxin signalling pathway (3 Aux/IAA and 15 ARFs,

distributed into the 3 clades). The proteins belonging to the different families within this pathway were also proven to present the same molecular mechanism in *P. patens* as the one described for the *A. thaliana* TIR1/Aux/IAA/ARF transcriptional response (Causier et al., 2012a; Lavy et al., 2016; Prigge et al., 2010).

After gymnosperms divergence, gene duplication events led to a huge increase in the number of genes and as a consequence, to certain functional redundancy within each protein family (Chandler, 2016; Finet et al., 2013; Lau et al., 2009). More than 20 *ARF* genes and 30 *Aux/IAAs* are present in flowering plants such as *A. thaliana*, *Zea mays* or *Oryza sativa* (Chandler, 2016; Paul et al., 2016; Wang et al., 2010).

Altogether, these evolutionary studies indicate that the first land plant already presented the full NAP machinery but in its simplest form with just one copy of TIR1/AFB receptor, of Aux/IAA repressor, of TPL co-repressor and an activator and a repressor ARF (Finet and Jaillais, 2012; Finet et al., 2013; Lau et al., 2009). But, when trying to go back even further in the history of plants, till charophyte organisms... Well, there we seem to lose the track.

### **Auxin signalling clues in charophytes?**

Although auxin is a plant hormone it is not just synthesized by plants. Auxin compounds have been detected in other organisms such as animals, fungi or algae. In the case of some algae, auxin is not only synthesized but its application has physiological effects that reminds us to auxin responses in plants (Lau et al., 2009). In some charophyte algae, auxin application affects cytoskeleton responses, cell division or rhizoids formation (Finet and Jaillais, 2012; Lau et al., 2009). Moreover, recent studies done upon *K. nitens* have shown that in this charophyte, auxin not only affected its cell division and elongation, but it also triggered a transcriptional response (Ohtaka et al., 2017). This transcriptional response to auxin involved 576 differentially expressed genes in *K. nitens*, among them, genes coding for Lateral Organ Boundaries Domain (LBD) transcription factors. In *A. thaliana* these genes are also auxin-induced, being AtARF7/19 responsible for their transcriptional activation that is driven through the TIR1/AFB-AUX/IAA-ARF nuclear pathway (Lau et al., 2009; Ohtaka et al., 2017). However, the search of NAP components in *K. nitens* did not reveal the presence of any components of the families TIR1/AFB, Aux/IAA and ARF involved in the common auxin transcriptional response, indicating the possible existence of an alternative and primitive pathway for auxin transcriptional changes in charophyte organisms (Hori et al., 2014; Ohtaka et al., 2017).

In spite of its questionable identity as auxin receptor, ABP1 is one of the favourite candidates for this auxin alternative response since homologues for this protein that conserve the auxin-binding pocket

have been traced back till chlorophyte and charophyte algae (Hori et al., 2014; Lau et al., 2009; Lu, 2015; Ohtaka et al., 2017). The ABP1 homologues lack the ER-retention motif present in AtABP1, indicating that algae ABP1 homologues are only located in the plasma membrane (Tromas et al., 2010). Some of the ABP1-triggered responses in *A. thaliana* are related with cell division and elongation, morphological features that are affected by auxin treatment in charophytes (Ohtaka et al., 2017; Tromas et al., 2010).

More extensive searches for NAP components in other charophytes have retrieved similar results: although genes involved in the synthesis and polar transport of auxin seem to have homologues in charophytes (Finet and Jaillais, 2012; Hori et al., 2014; Yue et al., 2014) only partial Aux/IAA and ARFs homologues have been described in the charophyte lineage (De Smet et al., 2011).

The only evidences of possible ancestors of the NAP in charophyte organisms come from Wang *et al.* in 2015 (Wang et al., 2015). Using transcriptomic data from *K. flaccidum* (now reclassified as *K. nitens* (Ohtaka et al., 2017)) (Klebsormidiophyceae), *Nitella hyalina* and *Nitella mirabilis* (Charales) and *Spyrogira pratensis* and *Penium margaritaceum* (Zygnematophyceae), the authors looked for proteins involved in the different phytohormones synthesis and signalling pathways. TIR1 homologues were found in all these organisms. However, they are proposed as a common ancestor for TIR1 and COI (Coronatine Insensitive), the JA receptor that presents a similar structure and mechanism to TIR1: JA acts as a molecular glue between COI and JA-signalling repressors (JASMONATE-ZIM DOMAIN, JAZ) that follow proteosomal degradation upon the binding of the hormone to the co-receptor complex.

Furthermore, Wang's search provided one protein from *K. nitens* (kfl00094\_0070) that possesses predicted B3 and PB1 (DIII/IV) domains, as in ARF transcription factors. The authors propose the presence of two other domains in this protein, DI and II, corresponding to a LxLxL EAR-motif and a degron motif, that together with PB1 domain (DIII/IV) conform Aux/IAA proteins. Thus, kfl00094\_0070 protein is until now, the only supposed ancestor proposed for both Aux/IAA and ARF transcription factors. Till now...

## Results

### Charophytes: already auxin fans

The lack of genomic and/or transcriptomic charophyte data makes it difficult to settle the origin and evolution of plant molecular pathways. We tried to overcome this issue taking advantage of the One KP database that contains transcriptomic data from more than 1000 species belonging to the *Viridiplantae* clade (Matasci et al., 2014) and of the Marchantia.info database.

We searched into these databases for ARF ancestors in charophyte organisms. For this, we chose to use as query amino acidic sequences from the liverwort *M. polymorpha* due to its position in the beginnings of land plants evolution. Tblastn searches done with the B3 domain of MpARF1 protein (Annexe-sequences) retrieved numerous transcripts from charophytes belonging to the 6 different classes within this lineage. We translated these transcripts (ExPASy translate tool (Gasteiger et al., 2003)) and we classed the resulting proteins into the different B3-types. We focused on ABI3, RAV and ARF B3-families, for being the most studied ones regarding their biochemical and functional properties, in opposition to the REM family.

Classification into ABI3, RAV and ARF B3-families was done depending on the residues present in the hypothetic DNA-interacting loops (Swaminathan et al., 2008) (Table 1). Table 2 shows that members of all B3 families are present in both “early divergent” and “late divergent” charophytes. Moreover, by sequences homology, we identified additional domains in the predicted B3-RAV and B3-ARF charophyte’s containing proteins<sup>5</sup>.

As for some *A. thaliana* B3-RAV proteins, most of the charophyte B3-RAV proteins also presented an AP2 domain. More surprisingly, the predicted AP2/B3-RAV from charophytes also contained a C-terminal PB1 domain (DIII/IV) present in current ARFs and Aux/IAA proteins but not in *A. thaliana* AP2/B3-RAV proteins.

---

<sup>5</sup> Details about organisms, accession numbers and the corresponding databases for the B3-domains search are indicated in Table 1-Annexe

**Table 2. B3 family members in Green Algae.** Possible homologues marked with a tick. The organisms used for the search within each class and the accession numbers for the homologues found in them are indicated in Table1-Annexe.

		ABI	RAV			ARF	
			AP2	B3	PB1	B3	PB1
	Chlorophytes	✓					
	Charophytes						
	Mesostigmatophyceae		✓	✓	✓		
	Chlorokyboceae	✓	✓	✓	✓	✓	✓
	Klebsormidiophyceae	✓	✓	✓	✓		
	Charophyceae	✓		✓		✓	✓
	Coleochaetophyceae		✓	✓	✓	✓	✓
	Zygnematophyceae	✓	✓	✓	✓	✓	✓

It is worth noticing that among the AP2/B3-RAV homologues with an extra PB1 domain, we reencountered kfl00094\_0070 protein from *K. nitens* previously described as an ARF-Aux/IAA ancestor due to the presence of B3 and PB1 domains (Table 1, annexe) (Wang et al., 2015).

In the B3-ARF containing proteins an additional PB1 domain was also found and thus, these proteins were considered as possible ARF ancestors, which we traced back till the charophyte algae *Chlorokybus atmophyticus* (AZZW-2021616, Table 1-Annexe). Despite its classification later in evolution, Klebsormidiophyceae class showed no presence of possible B3-ARF containing proteins, whereas we could find B3-ARF/PB1 proteins in all “late divergent” charophyte classes.

### B3-family evolution

We focused on the characterization of the two most ancient B3/PB1-containing proteins: the previously identified as a possible ARF and Aux/IAA ancestor in *K. nitens*, kfl00094\_0070 (Wang et al., 2015), and the newly described protein from *C. atmophyticus*, AZZW-2021616. Amino acidic sequences analyses of the residues belonging to the DNA-interacting loops revealed that the B3 domains present in these proteins belong to the RAV and ARF family respectively. To confirm this classification, we carried out a phylogeny test including 89 B3-domain sequences from ABI3, RAV and ARF families belonging to chlorophytes, charophytes and land plants organisms (Annexe-sequences).

Phylogenetic trees were constructed with MUSCLE alignment algorithm and different phylogenetic inference methods (Maximum Likelihood, Neighbour Joining and Minimum Evolution). In all the resulting trees, ABI3, RAV and ARF families were separated into different clades. Within the ARF clade, B3-domains were classed into 3 subfamilies corresponding to classes A, B and C (Finet et al., 2013) (Fig 43; Annexe).

*K. nitens* kfl00094\_0070 was always positioned into clade RAV. The different phylogenetic inference methods classed *C. atrophyticus* AZZW-2021616 either as a class-C ARF or in between class-A/B and C (Annexe). However, no matter the method used, AZZW-2021616 was always classed inside ARF clade, confirming this our previous result (Fig 43).

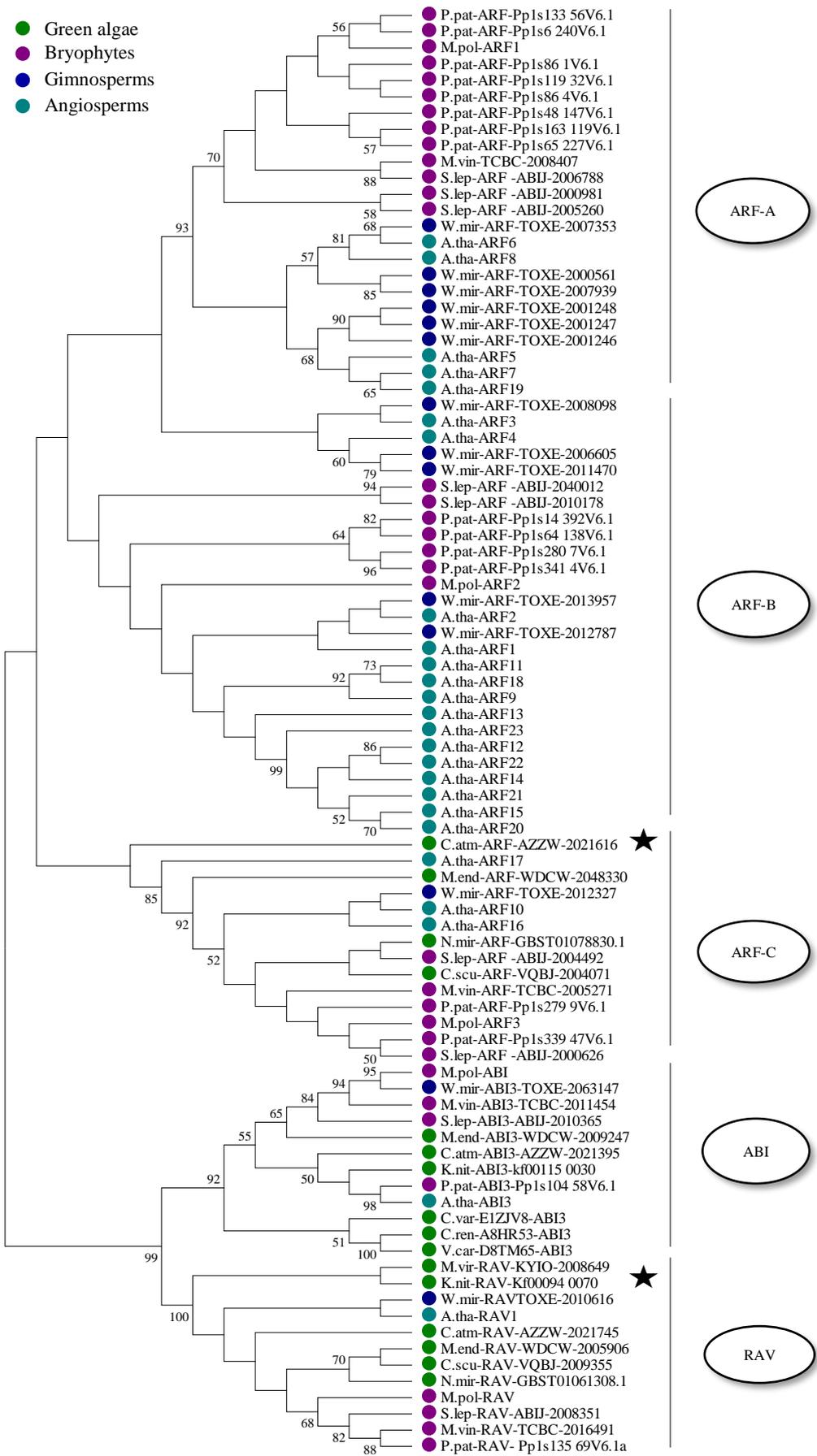
### **Charophycean B3-domains: structural and biochemical characterization**

To determine the transcription factor nature of *K. nitens* kfl00094\_0070 and *C. atrophyticus* AZZW-2021616 (from now on, KnRAV and CaARF respectively) we analysed their potential capacity to specifically bind DNA.

KnRAV-AP2 and B3 predicted domains presented a high conservation of residues implicated in the interaction with DNA sequences (Fig 44.A and C). The structures of these domains were modelled with Phyre 2 (Kelley et al., 2015): AP2 predicted domain was modelled with 100% confidence with AtERF11 (Allen et al., 1998) whereas B3-RAV domain models were based on AtRAV1, AtARF1 and AtARF5 proteins (Boer et al., 2014; Yamasaki, 2004)). The comparison of modelled and real structures returned a very precise superposition of both and revealed that the conserved residues mediating DNA-interactions were equally positioned in the models and in AP2 and B3-AtRAV1 structures (Fig 44.B and D).

To characterize KnRAV DNA-binding capacity and specificity, we isolated the N-terminus of KnRAV containing both the AP2 and B3 domains and we tested by EMSA its capacity to bind different DNA sequences. We tested sequences of DNA-binding sites for the 3 B3-families (ABI3, RAV and ARF, Table 1) together with a bipartite AP2-B3 site with a 5 nucleotide-spacing in between AP2 and B3 binding sites (Kagaya et al., 1999) (Table 8).

KnRAV-AP2/B3 was uniquely able to bind the bipartite site containing both AP2 and B3-RAV DNA-binding sequences. These results indicate that KnRAV can bind to DNA and that this binding presents the same specificity as AtRAV1 transcription factors (Kagaya et al., 1999; Yamasaki, 2004), confirming the classification of KnRAV as a true AP2/B3-RAV type transcription factor.

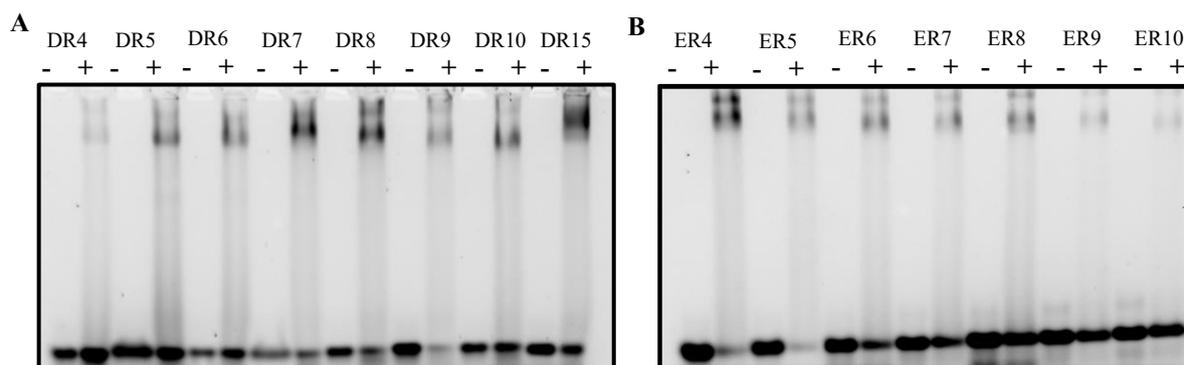








EMSA showed that CaARF-FL could bind all DRs tested with a range of spacings from 4-10 (plus DR15), behaviour shared by AtARF2 and AtARF10 in the previous chapter. As for ERs, CaARF-FL bound ER4-8, with a clear preference for ER4. This preferential binding is more characteristic of AtARF10, whereas AtARF2 and AtARF5 preferred 7/8-spaced ERs (Fig 47; Fig 33-35-Chapter II). Although these particularities in the binding to different AuxREs are consistent with our classification of CaARF into ARF class-C, it is worth noting that CaARF-FL, when binding to DNA (either DR or ER) tends to form two different species in the gel, whereas this was not observed for AtARF10.



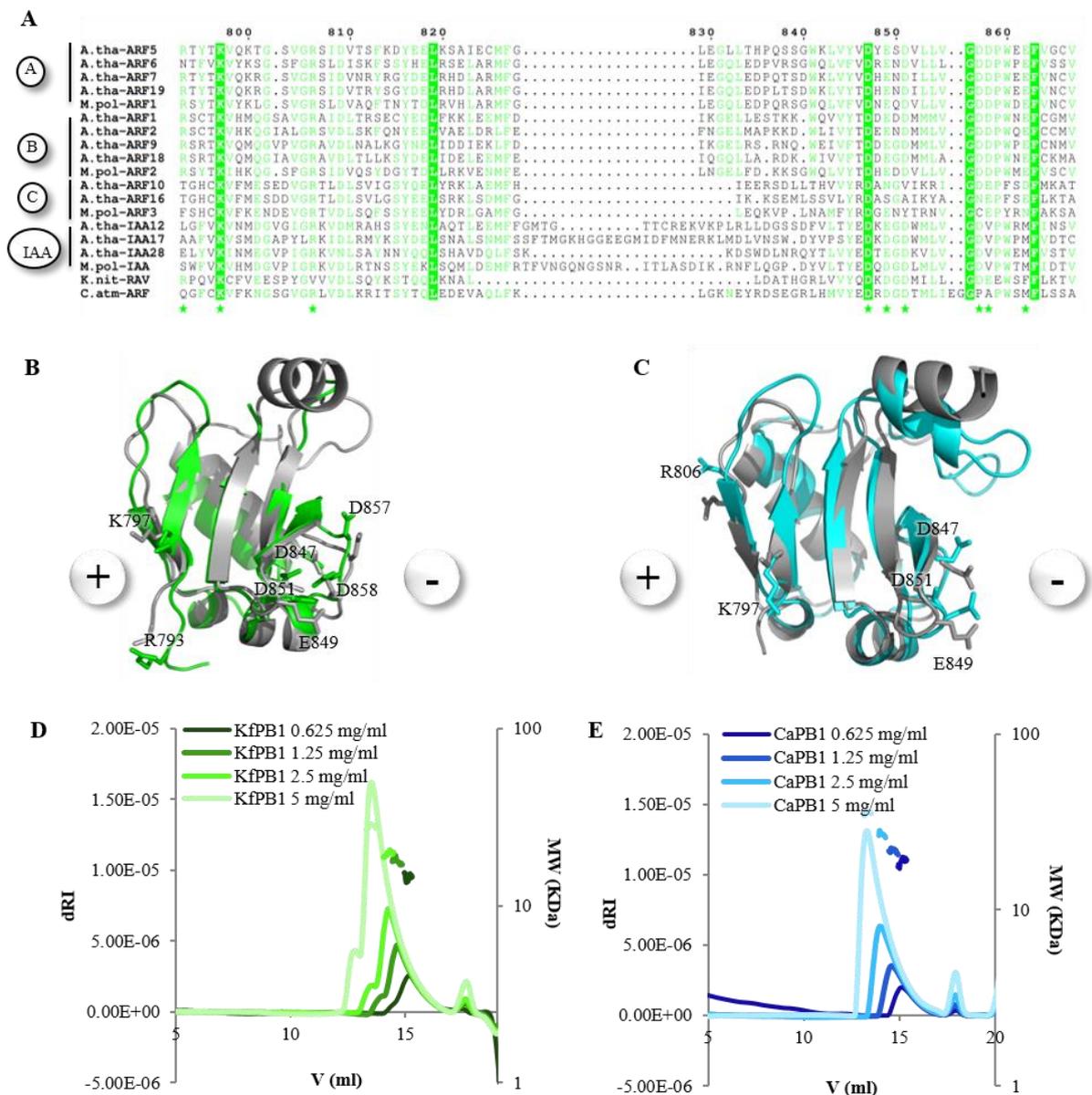
**Fig 47. EMSA shows similarities in the binding preferences between CaARF and AtARF10 class-C. A. DR AuxREs with spacings 4-10 and DR15. B. ER AuxREs with spacings 4-10.**

### Charophycean PB1 domains: structural and biochemical characterization

To characterize the predicted PB1 domains from KnRAV and CaARF we first analysed the conservation of residues implicated in the PB1 oligomerization. These domains are known to form oligomers in a head-to-tail manner thanks to the presence of two interaction interfaces, positive and negative. Different types of ARFs and Aux/IAAs are known to show differences in the residues and electrostatic potential of these interfaces (Parcy et al., 2016).

Multiple sequence alignments combined with structure modelling (Phyre2) allowed us to observe that predicted PB1 domains from both organisms were indeed modelled as PB1 domains from different ARFs or Aux/IAA proteins with 100% confidence (Dinesh, D.C. et al., 2015; Han et al., 2015; Korasick et al., 2014; Nanao et al., 2014). The corresponding modelled structures presented positive and negative interfaces (Fig 48.A-C). Two clusters of negatively charged amino acids are very conserved in KnRAV-PB1, whereas CaARF-PB1 possesses only the first cluster of the negatively charged residues. This feature is not shared by AtARF-C class PB1 domains, in which the negative interface is mainly constituted by the second cluster, or by MpARF3-PB1 that presents only some of the negatively charged residues in both clusters.

The oligomerization potential of these proteins was confirmed by SEC-MALLS. For this, we isolated KnRAV-PB1 and CaARF-PB1 domains and we calculated their MW under increasing protein concentrations (Fig 48.D-E; Table 3). In both cases, the peaks presented an asymmetrical profile of elution. Asymmetric peaks are characteristic of oligomeric proteins due to the mix of species of different molecular sizes that coelute the SEC column. MW estimations confirmed that both PB1 domains were oligomeric with the size of the complex increasing with the protein concentration. KnRAV-PB1 presented a two-peaks profile at high protein concentrations, which might indicate that two different types of oligomers are formed.



**Fig 48. KnRAV-PB1 and CaARF-PB1 oligomerization.** A. KnRAV and CaARF PB1 domains aligned to PB1-ARF and Aux/IAA domains from *A. thaliana* and *M. polymorpha*. Green stars indicate residues involved in head-to-tail interactions referred to AtARF5-PB1 (Nanao et al., 2014). B-C. AtARF5-PB1 structure (grey) (4CHK, (Nanao et al., 2014)) superposed with KnRAV-PB1 (green) and CaARF-PB1 (cyan) models. D-E. SEC-MALLS calculations for KnRAV-PB1 and CaARF-PB1 molecular weights at increasing protein concentrations.

**Table 3. SEC-MALLS-calculated MW for KnRAV-PB1 and CaARF-PB1**

<b>KnPB1</b> (mg/ml; $\mu$ M)	Main peak MW (KDa)// Oligomeric state	Small peak MW (KDa)// Oligomeric state	<b>CaPB1</b> (mg/ml; $\mu$ M)	Main peak MW (KDa)// Oligomeric state
<b>0.625 mg/ml;</b> <b>69<math>\mu</math>M</b>	14.2 $\pm$ 10.8% 1.57 monomers		<b>0.625 mg/ml;</b> <b>52<math>\mu</math>M</b>	18.8 $\pm$ 5.8% 1.56 monomers
<b>1.25 mg/ml;</b> <b>138<math>\mu</math>M</b>	17.8 $\pm$ 2.6% 1.97 monomers	35.1 $\pm$ 5.9% 3.9 monomers	<b>1.25 mg/ml;</b> <b>104<math>\mu</math>M</b>	21.6 $\pm$ 5.3% 1.8 monomers
<b>2.5 mg/ml;</b> <b>277<math>\mu</math>M</b>	19.7 $\pm$ 1.2% 2.19 monomers	33.1 $\pm$ 2.4% 3.68 monomers	<b>2.5 mg/ml;</b> <b>208<math>\mu</math>M</b>	26.9 $\pm$ 1% 2.24 monomers
<b>5 mg/ml;</b> <b>555<math>\mu</math>M</b>	28 $\pm$ 10.8% 3.11 monomers	49 $\pm$ 1% 5.44 monomers	<b>5 mg/ml;</b> <b>416<math>\mu</math>M</b>	35.8 $\pm$ 0.4% 2.98 monomers

### Charophycean: ARFs and something more?

Altogether, the combination of phylogenetical, structural and biochemical studies allowed us to conclude that, although the previously proposed ARF ancestor, kfl00094\_0070, in *K. nitens* is actually an ancestor of the B3-RAV family of transcription factors, ARF proteins were already present in “early divergent” charophytes such as *C. atmophyticus* with similar biochemical properties to those found in plant ARFs. The presence of possible ARF ancestors in charophycean organisms made us wonder if any of the other auxin-signalling components, Aux/IAA, TIR1/AFB or TPL, could also be found in this lineage.

As previously commented, Wang *et al.* 2015 studies proposed kfl00094\_0070 (here KnRAV) as an ancestor of both ARF and Aux/IAA proteins. According to these studies, kfl00094\_0070 presents all the characteristic domains of Aux/IAAs: DI (LxLxL EAR motif), DII (degron, GWPP-motif) and a DIII/IV or PB1 domain. Thus, we reanalysed this sequence and the rest of the RAV-PB1 containing proteins in the other charophycean clades in the search for potential DI and DII motifs. Contrary to Wang *et al.* studies, we could not find any sign of degron motif in any of these proteins (the presence of EAR motifs is discussed later).

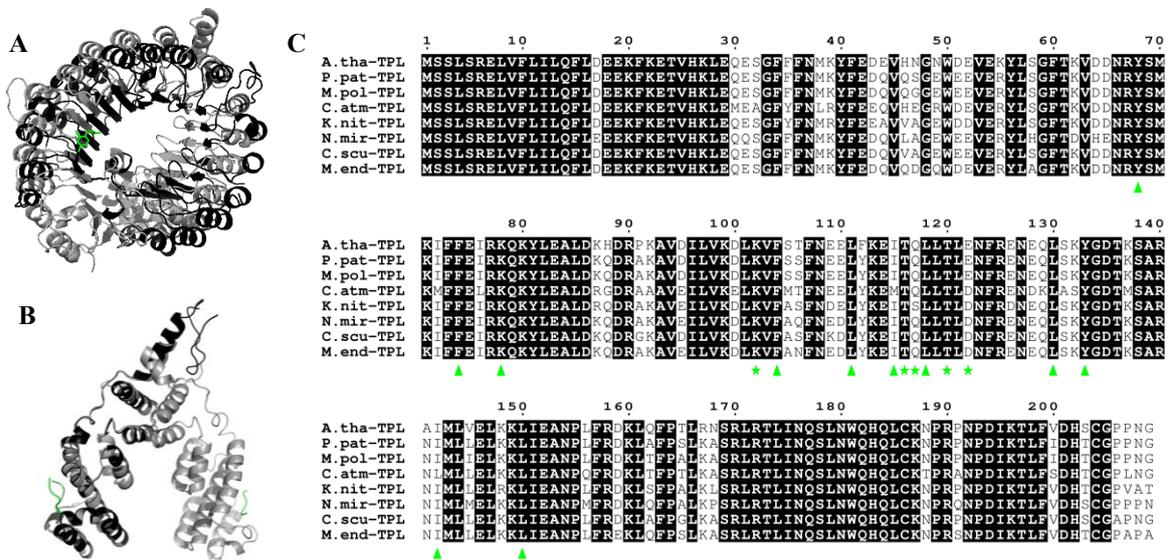
We carried out then a more exhaustive search for Aux/IAA homologues using as query *M. polymorpha* Aux/IAA. This search resulted in homologues that contained only a predicted PB1 (DIII/IV) domain (not shown), but neither DI or DII could be found in these homologues. We concluded then that Aux/IAA ancestors do not exist within the charophycean lineage (Table 4).

We proceeded similarly for TIR1/AFB and TPL homologues. Homologues for MpTIR1 auxin receptor were obtained for the 6 charophyte classes (Table 4 and Table 2-Annexe). Structural models were done for the 6 possible TIR1 ancestors with Phyre2 software (Kelley et al., 2015). Except for *K. nitens* TIR1 candidate, which was only partially modelled, all other homologues gave structures modelled with 100% confidence as AtTIR1 auxin receptor and AtCOI jasmonic acid receptor (Sheard et al., 2010; Tan et al., 2007). These two hormonal receptors present a very similar structure and mechanism but differences among these receptors can be found in the residues that mediate the interaction with the hormones. Therefore, to identify TIR1 charophycean homologues as possible TIR1 ancestors we analysed the conservation of these residues in the obtained structural models. We also determined the conservation of the residues that mediate the interaction with Aux/IAA degron motif (Tan et al., 2007). In none of the TIR1 homologues in charophyte organisms these residues were conserved (Fig 1-Annexe). Moreover, the modelled structures presented some differences in the positioning of the Beta-strands that mediate the interaction with auxin. To exclude the possibility that we could be dealing with COI1 ancestors instead of TIR1 we also studied residues comparison between charophyte TIR1 homologues and AtCOI1 protein (Fig 1-Annexe). Same as it happened with TIR1, the residues implicated in the interaction with JA-Ile conjugate and with JAZ repressors were not present in charophyte proteins. We concluded that the charophytes TIR1, even if they could be considered as ancestral forms of TIR1/AFB auxin receptors, they lacked the ability to bind the hormone and Aux/IAA proteins (Fig 49.A).

Homologues of TPL co-repressor protein were also found in all charophyte classes except for Mesostigmatophyceae (Table 4 and Table 2-Annexe). Protein alignments revealed a high conservation of the N-terminus of TPL ancestors (including the residues involved in the interactions with EAR-motifs and in tetramerization) and were structurally modelled as OsTPR2 or AtTPL184 with 100% confidence (Ke et al., 2015; Martin-Arevalillo et al., 2017) (Fig 49.B and C). Two sets of WD40 domains were predicted in the C-terminus of TPL homologues for *C. atmophyticus*, *K. nitens* and *C. scutata*. *N. mirabilis* TPL presented just one WD40 repeats domain and *M. endlicheranium* sequence was incomplete.

**Table 4. Auxin signalling components in Green Algae.** Possible homologues marked with a tick. The star in TIR1/AFB column indicates that for this family homologues were found but the structure models suggest that these proteins lack the auxin binding pocket. The organisms studied for each class and the accession numbers are indicated in Table 2-Annexe.

	ARF	AUX/IAA	TIR1/AFB	TPL
Chlorophytes				
Charophytes				
Mesostigmatophyceae			★	
Chlorokyboceae	✓		★	✓
Klebsormidiophyceae				✓
Charophyceae	✓		★	✓
Coleochaetophyceae	✓		★	✓
Zygnematophyceae	✓		★	✓



**Fig 49. Auxin signalling components in charophytes.** **A.** *C. atmophyticus* predicted TIR1, CaTIR1, (AZZW-2021242) structure model (black) superposed with AtTIR1 structure (grey) (2P1M, (Tan et al., 2007)) in complex with auxin (green). **B.** *C. atmophyticus* predicted TPL, CaTPL (AZZW-2021890), N-terminus structure model (black) superposed with AtTPL184 structure (5NQV, (Martin-Arevalillo et al., 2017)) in complex with Aux/IAA27 EAR motif. **C.** Protein alignment with the N-ter from charophycean predicted TPL. A. tha (*A. thaliana*); P. pat (*P. patens*); M. pol (*M. polymorpha*); C. atm (*C. atmophyticus*); K. nit (*K. nitens*); N. mir (*N. mirabilis*); C. scu (*C. scutata*); M. end (*M. endlicheranium*). Green triangles: residues responsible for interaction with LxLxL-type EAR motifs. Green stars : residues involved in TPL tetramerization.

### **KnRAV and CaARF interact with TPL**

Although the region in between DBD and PB1 domains for KnRAV and CaARF presented no predicted structure, we identified in these regions possible TPL-recruitment motifs. KnRAV presented a (M/K/R)LFG-type EAR motif and a sequence of amino acids, LGLRIGPP, that resembles the EAR-like motif found in *P. patens* Aux/IAA (LxLxPP) (Causier et al., 2012a). As for CaARF, only a (M/K/R)LFG-type EAR motif was found (Fig 50.A). Since TPL is known to interact with some members of AP2/B3-RAV and ARF families we wondered if that could also be the case for the ancestors of these proteins.

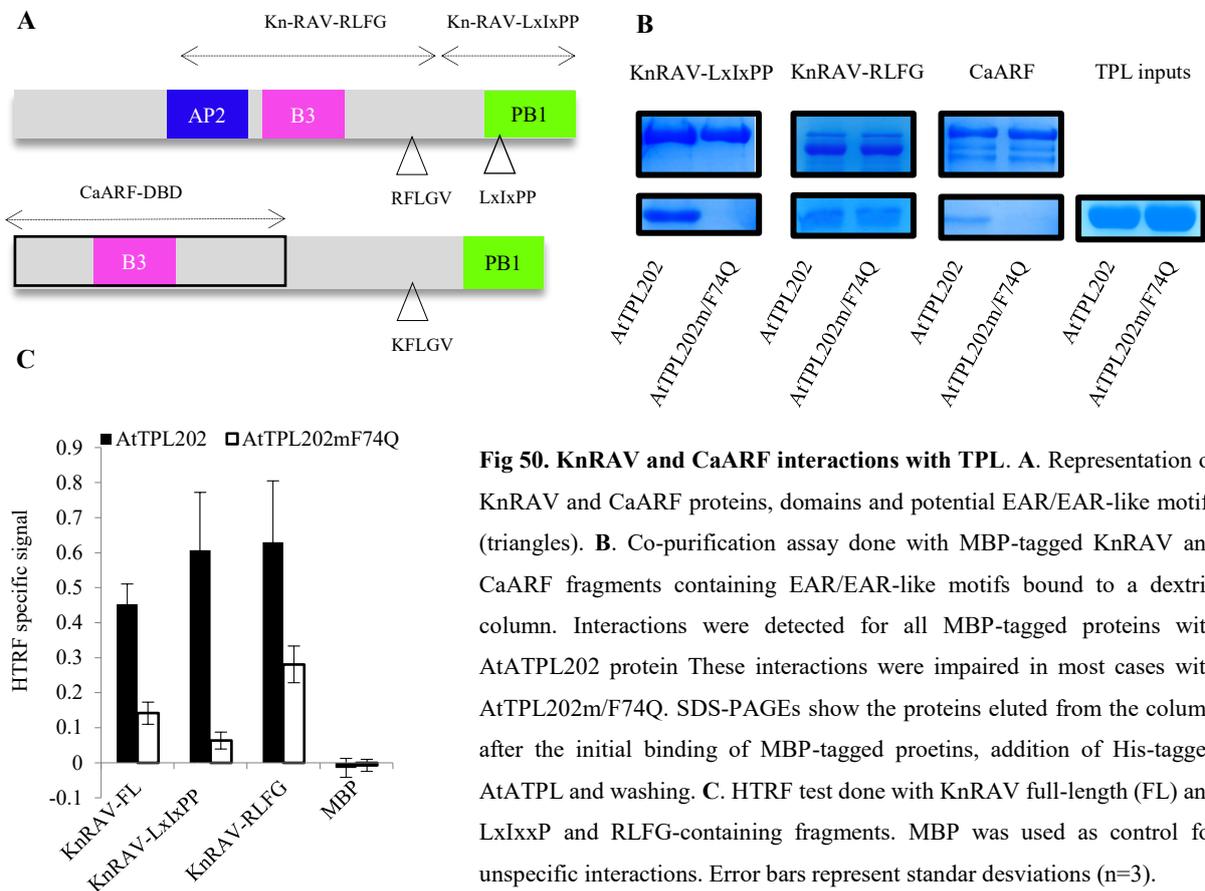
We decided to work with the N-terminal part of TPL protein since till now, is the only one for which interactions with different types of EAR-motifs have been characterised (Ke et al., 2015; Ma et al., 2017; Martin-Arevalillo et al., 2017). Due to the high conservation of residues within TPL N-terminus from charophyte organisms till extant land plants (Results-Chapter I), we used AtTPL (constructs AtTPL202 and AtTPL202/mF74Q [as negative control], Chapter I) instead of TPL proteins for both *K. nitens* and *C. atmophyticus*.

KnRAV and CaARF interactions with AtTPL N-ter interactions were analysed by co-purification and HTRF protein-protein interaction techniques (details in Materials and Methods). In the case of KnRAV, two regions containing RLFG and LxIxPP-motifs in an exclusive manner were used (Fig 50.A) in order to determine which of the two motifs (if any) mediates the interaction with TPL. For CaARF, only a RLFG motif was found as possible TPL-recruiting site and so, we used the full-length protein for the interaction tests.

For co-purification interaction tests, charophyte proteins were first fixed to a dextrin-sepharose column thanks to their MBP-tags. AtTPL202 and AtTPL202m/F74Q were copurified through the columns with the attached KnRAV fragments and CaARF-FL protein. As a control for unspecific interactions MBP protein was used (not shown). TPL interactions were detected for all three proteins: KnRAV-LxIxPP, KnRAV-RLFG and CaARF proteins, with KnRAV-LxIxPP showing the strongest binding. F74Q TPL mutant could not interact with KnRAV-LxIxPP and CaARF proteins, indicating that groove 3 is implicated in these bindings. However, KnRAV-RLFG could interact with both AtTPL202 and the mutant in a similar manner (Fig 50.B).

Confirmation of these interactions was done with HTRF protein-protein interaction technique. This method could not be applied to test interactions between CaARF and AtTPL since CaARF protein by itself gave HTRF positive signal due to the presence of an internal His-tag. KnRAV full-length (KnRAV-FL) and KnRAV-LxIxPP could interact with AtTPL202. These interactions were strongly reduced with F74Q mutation. In agreement with co-purification tests, KnRAV-RLFG was also able to

interact with AtTPL202 although in this case, the interaction with AtTPL202m/F74Q was partially reduced, in opposition to co-purification tests.



**Fig 50. KnRAV and CaARF interactions with TPL.** **A.** Representation of KnRAV and CaARF proteins, domains and potential EAR/EAR-like motifs (triangles). **B.** Co-purification assay done with MBP-tagged KnRAV and CaARF fragments containing EAR/EAR-like motifs bound to a dextrin column. Interactions were detected for all MBP-tagged proteins with AtATPL202 protein. These interactions were impaired in most cases with AtTPL202m/F74Q. SDS-PAGEs show the proteins eluted from the column after the initial binding of MBP-tagged proteins, addition of His-tagged AtATPL and washing. **C.** HTRF test done with KnRAV full-length (FL) and LxIxP and RLFG-containing fragments. MBP was used as control for unspecific interactions. Error bars represent standard deviations (n=3).

## Discussion

Plants water to land transition is one of the most important events that ever occurred in our planet. It changed the atmosphere composition, the soil conditions and it led to the appearance of new organisms and of agriculture: it determined our history, and yet, how it happened remains an opened question. In the last chapter of this manuscript we tried to place one piece into the puzzle of plants evolution, the auxin piece.

### B3: a domain in expansion

Charophytes adaptation to terrestrial conditions was accompanied by an expansion of numerous protein families, especially those implicated in hormone signalling processes (Leliaert et al., 2012). Here we focused on the study of B3 transcription factors, a domain specific of the Viridiplantae clade that is

divided into 4 protein families in land plants: ABI3, RAV, ARF and REM<sup>6</sup>. By sequence homology with the B3 domain from MpARF1 we searched for B3-containing proteins within Green Algae. Whereas in Chlorophyte organisms we only found members of ABI3 family, charophytes already presented predicted B3 belonging to ABI3, RAV and ARF families, which indicates an expansion of the B3 families during or right after the division of these two clades. Our phylogenetic analysis indicates that RAV family appeared earlier, whereas B3-ARF type is a more recent invention.

B3-ARF domains are not found alone; structural data revealed that in ARF proteins B3 domains are embedded within a dimerization domain (DD) with no other structural resemblance till now (Boer et al., 2014). Thus, after the divergence of the B3-ARF family from B3-ABI3/RAV, an insertion of this domain inside the current ARF-DD must have taken place conferring to ARF TF a preference for palindromic sites that can only be bound by dimers (Boer et al., 2014). Already in bryophytes, ARFs had diversified into three evolutive clades: A (activators), B and C (repressors), with just one member of each being enough for a full auxin transcriptional response in the liverwort *M. polymorpha* (Finet et al., 2013; Flores-Sandoval et al., 2015; Kato et al., 2015). But variety is the spice of life and so, numerous gene duplications have led to the appearance of subfamilies within these families, to pairs of ARFs with partial functional redundancy and to truncated ARFs that lack the PB1 domain in high land species (Chandler, 2016; Okushima et al., 2005). What and when was the origin of all this?

Our search for B3 homologues retrieved several B3/PB1-containing proteins in charophytes, that due to the presence of both domains could be good candidates of ARF ancestors in green algae. We carried out a structural and biochemical characterization of two of these proteins: kfl00094\_0070 in *K. nitens* (previously proposed as an ARF and Aux/IAA ancestor (Wang et al., 2015)) and *C. atrophyticus*-AZZW-2021616, the most ancient B3-ARF member that we found.

#### ***Klebsormidium nitens* possesses ancestral RAV transcription factors**

kfl00094\_0070 had been previously identified as an ARF ancestor due to the presence of a predicted N-terminal B3 and C-terminal PB1 domains (Wang et al., 2015). However, we determined that the B3 domain present in this protein belongs to the B3-RAV family. Moreover, as for some members of the RAV family in *A. thaliana*, kfl00094\_0070 (KnRAV) presented an additional AP2 domain upstream the B3-RAV. EMSAs done with the N-terminal part of KnRAV, that contained both AP2 and B3-RAV domains, revealed that KnRAV presented the same DNA specificity as AtRAV

---

<sup>6</sup> REM family was annihilated in this study due to the lack of DNA specificity, although recent studies suggest that B3-REM domains bind specifically to poli(A) regions (O'Malley R.C. et al., 2016).

transcription factors, binding bipartite sequences containing sites for AP2 and B3-RAV families (Kagaya et al., 1999; Yamasaki, 2004).

The presence of a predicted PB1 domain with oligomeric capacity in this protein is not shared with any member of land plants RAV family. AP2/B3-RAV/PB1-containing proteins were found in almost all charophytes (except for *N. mirabilis*, within the Charophyceae), including the Zygnematophyceae, considered sisters to land plants. Thus, ancestor RAV transcription factor must have lost the PB1 domain during or after the divergence to land plants. The presence of a sequence with resemblance to EAR-motifs found in *P. patens* Aux/IAA (LxLxPP) (Causier et al., 2012a; Paponov et al., 2009) right upstream KnRAV-PB1 domain suggests that if not ancestor to the ARFs, charophyte RAV PB1-containing proteins could have given rise to Aux/IAA proteins. However, we did not find any evidence of phylogenetic relation between both families of transcriptional regulators (not shown).

Apart from the EAR-like motif (LxIxPP), localized between AP2/B3-RAV and PB1 domains, we found another EAR-like motif (RLFG-type) in this middle region. AtRAV transcription factors act as transcriptional repressors involved in responses to cold and in flower initiation (Fu et al., 2014; Matías-Hernández et al., 2014). Since several members of the AtRAV family have been shown to interact with TPL they are proposed to exercise transcriptional repression through the recruitment of co-repressor complexes (Causier et al., 2012; Matías-Hernández et al., 2014). These interactions happen through (M/K/R)LFG-type EAR motifs (Causier et al., 2012). Protein-protein interaction tests done with KnRAV showed that two different regions, containing either the LxIxPP (KnRAV-LxIxPP) or the RLFG (KnRAV-RLFG) sequence, interact with AtTPL N-ter (AtTPL202). Both KnRAV fragments showed a reduced interaction with AtTPL202m/F74Q, that carries a mutation in TPL binding site for LxLxL-type EAR motifs (Ke et al., 2015; Martin-Arevalillo et al., 2017). Although these effects were much stronger for KnRAV- LxIxPP than for KnRAV- RLFG, these data support that different types of EAR-motif within the same protein could be simultaneously bound to TPL N-ter using the same binding site, as shown in Chapter I for LxLxL and WUS box EAR motifs. Further interaction tests with KnRAV mutants in both LxIxPP and RLFG motifs should be done for full confirmation of these interactions.

The shared biochemical characteristics between KnRAV and AtRAV proteins indicate that in charophytes organisms the RAV family was already established with similar roles and mechanisms as RAV transcriptional repressors in land plants. The current implications of RAV proteins in abiotic stress responses, such as cold or drought (Fu et al., 2014; Matías-Hernández et al., 2014), indicate that the acquisition of this transcription factor family might have had a fundamental role in the adaptation to terrestrial stresses. The presence of the PB1 domain in ancestral charophycean RAVs remains a mystery: is oligomerization needed for KnRAV protein function or is this PB1

simply a remaining vestige with no current use for KnRAV? Only functional studies with deletions or mutations of these KnRAV-PB1 could give a sure answer to this question.

### ***Chlorokybus atmophyticus*: home for the first ARF**

*C. atmophyticus*-AZZW-2021616 (CaARF) was classed as an ARF according to our sequence conservation and phylogenetic studies. Sequence analogies, structural modelling and DNA specificity helped us to further classify this protein into class-C ARFs.

CaARF-DBD bound to palindromic AuxREs DR5 and ER7 in a specific manner, whereas we observed no binding to single-site AuxREs, which indicates that CaARF, as for some AtARFs (Boer et al., 2014), binds DNA as a dimer. However, no evidences exist for ARF-C DBD dimerization. In fact, our structure models for MpARF3 and AtARF10, both class-C ARFs, suggests that the insertion inside their DBD is positioned in one of the helices implicated in the dimerization interface established upon binding to ER motifs. Moreover, Y2H or *in planta* protein-protein interaction tests showed that MpARF3 was not able to dimerize (Kato et al., 2015). One possibility is that class-C ARFs dimerize in a different manner to class A and B ARFs (Boer et al., 2014) or only upon binding to DNA. Structure determination of class-C ARF-DBD remains an “unfinished business” in the auxin field. Structural studies of CaARF-DBD could help understanding class-C ARF mechanism and its evolution (ongoing). Moreover, they could provide information about the DNA binding preferences of these ARFs that, as reviewed in the previous chapter, are necessary for the establishment of models that help predicting auxin transcriptional response.

On the other hand, as land plant ARFs, CaARF presented a predicted C-terminal PB1 domain. In the context of auxin signalling, PB1 domains mediate homotypic and heterotypic interactions between ARFs and Aux/IAA proteins. Auxin interactomes indicate that these interactions happen mainly between class A and Aux/IAA proteins, whereas class B and C are rarely found to interact among themselves or with any of the other two families (Farcot et al., 2011). Our CaARF-PB1 structure model predicts the presence of a positive and a negative interface and SEC-MALLS experiments confirmed the homo-oligomeric potential of this domain. On the other hand, the presence of two proteins with predicted PB1 domains in *C. atmophyticus*, CaARF and CaRAV (AZZW-2021745, Table 1-Annexe) (not characterized in this manuscript), raises the possibility of a heterotypic interaction between these two proteins mediated by their PB1 domains. Mutations in CaARF-PB1 interfaces and structure determination (ongoing) could help confirming PB1 homo and hetero-oligomerization potential.

### Charophytes: setting the scenario for flowers

Most of the auxin-triggered transcriptional changes in land plants are attributed to a set of 4 families of proteins: TIR1/AFB receptors, Aux/IAA repressors, TPL co-repressors and ARF transcription factors (activators and repressors) which are enough to carry out an auxin transcriptional response (Pierre-Jerome et al., 2016; Pierre-Jerome et al., 2014). ARF activators, due to their capacity to interact with Aux/IAA proteins, are proposed to work in an auxin-dependent manner: the hormone triggers the degradation of Aux/IAA proteins which deblocks ARF activators. Regarding ARF repressors... “that’s a different kettle of fish”. If most of them are not able to interact with Aux/IAA (Farcot et al., 2011) how could they trigger an auxin transcriptional response? Several mechanisms are proposed: competition for DNA-binding sites with ARF activators, ARF activators squelching or TPL recruitment (Chandler, 2016). Supporting the latter, several repressor ARFs have been shown to interact with TPL in *A. thaliana* and *P. patens* (Causier et al., 2012a, 2012b) although the repercussions of these interactions have not been studied yet.

Here we showed that CaARF is able to interact with AtTPL N-terminal part. In the middle region in between CaARF-DBD and CaARF-PB1 we found a (M/K/R)LFG-EAR motif that could be responsible for this interaction, although mutation of this motif is needed for final confirmation. AtTPL202m/F74Q lack of interaction with CaARF suggests that the binding to the protein happens through the same groove that had been previously characterized as one of the binding site for EAR motifs (Ke et al., 2015; Martin-Arevalillo et al., 2017). (M/K/R)LFG-EAR motifs are found in AtARF TPL-interacting partners ARF2, 9 and 18, all class-B ARFs (Causier et al., 2012b) whereas the ARFs that were able to interact with TPL in *P. patens* belonged to class-C and presented LxLxL and LxLxPP sequences that mediated the interaction (Causier et al., 2012a).

Although we did not test CaARF interaction with TPL from *C. atmophyticus* we could find homologues for this co-repressor in all classes of charophytes, except for *M. viride*. Comparison of the N-terminal part of these proteins revealed a high conservation of residues indicating that TPL, its structure and hub protein properties the way we know them, were already present in “early divergent” charophytes.

The search for the two other protein families implicated in auxin response was not that successful. Only homologues for Aux/IAA-PB1 domains were found, but these proteins did not present DI and II, containing the EAR-motif and the degron motif and consequently, we did not consider them as precursors of the Aux/IAA family.

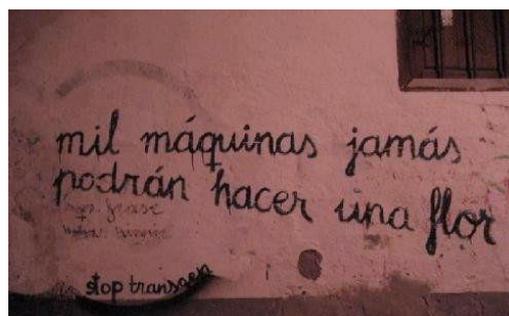
As for TIR1/AFB auxin receptors, homologues were found in *C. atmophyticus* and in all classes of “late divergent” charophytes. The modelled structure of TIR1 homologue in *C. atmophyticus* revealed the presence of both F-box and LRR domains, however, the residues responsible for the

interaction with auxin and the DII-degron motif from Aux/IAA proteins (Tan et al., 2007) were not conserved in the charophyte proteins.

Since TIR1/AFB and Aux/IAA families are absent in charophyte algae we concluded that the NAP was established after the divergence of land plants. However, charophytes already had the basic skeleton and were paving the way towards body complexity. *C. atmophyticus* ARF ancestor had already everything that defines an ARF: a DBD with capacity to bind AuxREs and a PB1 domain able to oligomerise, but settled in a context where no auxin response was possible due to the lack of TIR1/AFB-Aux/IAA co-receptor complex. Therefore, CaARF could have acted as a transcriptional repressor with TPL recruitment potential and no implication in auxin response in *C. atmophyticus*. After gene duplication, adaptations of its PB1 and the appearance of the full TIR1/AFB-Aux/IAA system, ARF ancestors were likely co-opted in auxin transduction.

This system presents two main advantages: having a repressor that systematically blocks the pathway provides stability and its fast degradation through the proteosomal system allows a quick response to auxin, with transcriptional changes observable minutes after the treatment (Abel and Theologis, 1996). This idea can be easily extrapolated to jasmonic acid hormone that plays with the same rules as auxin: COI receptor also belongs to the SCF ubiquitin-ligase complex and it drives repressors degradation under the hormonal signal. Given the structural and mechanistic similarities between these receptors one might hypothesized that they coevolved from a common ancestor, as proposed by Wang *et al.* 2015.

Charophytes' incomplete auxin signalling might have been enough for their simplicity and lack of differentiation of their structures but we know that the full NAP is needed, even in its simplest form to confer *Marchantia* its curious body structure. Multiple copies within each protein family can be found in higher plants. Their different spatio-temporal expression, their interactions with different affinities and their post-translational modifications, all have worked out together for the achievement of one the the highest levels of complexity in plants: the flower. So complex that even some walls in Spain seem to declare that “not even 1000 machines will ever be able to build a flower”. Maybe this thought is a bit romantic, but at least, till that moment arrives there is way to go in the field of auxin research and auxin evolution.



## General discussion and conclusions

### Tradition and innovation in the pathway of auxin

The goal of my work was to get a better understanding of the functional, structural and evolutive properties of proteins implicated in the auxin transcriptional response. One of the big questions with respect to auxin is how a single molecule triggers such a wide diversity of responses in different places and moments within a plant's life. TIR1/AFB receptor, Aux/IAA repressor and ARF TFs families constitute the NAP, in charge of translating the chemical auxin signal into transcriptional changes inside a plant cell. In higher plants, numerous copies of genes exist within each of these families. Their differential patterns of expression, post-translational modifications and the interactome established among them account for the complexity, diversity and robustness of the auxin transcriptional response in land plants.

However, things did not seem so complicated 450 m.y.a. When we looked for homologues of the components of the NAP in charophytes, direct ancestors of land plants, we found one single ARF ancestor candidate in the multicellular “early divergent” charophyte *C. atmophyticus*. The biochemical characterization of CaARF revealed that it corresponds to a class-C ARF, considered as a repressor clade of the ARFs present in land plants. Supporting its repressor identity, we proved that CaARF could interact with TPL. The search of homologues for TIR1/AFB and Aux/IAA repressors indicated that in charophytes the perception auxin system was not present yet. These facts led us to the conclusion that charophyte organisms presented a “preliminary skeleton” of the auxin transcriptional response: an ARF able to interact with TPL but unable to respond to auxin due to the absence of Aux/IAA homologues. Plants ARFs-B/C (repressors) maintained these features: some can recruit TPL (Causier et al., 2012a, 2012b; Lavy et al., 2016) and the great majority of them cannot interact with Aux/IAA proteins (Farcot et al., 2011; Piya et al., 2014; Trigg et al., 2017). Therefore, this “preliminary skeleton” has been maintained in plants as a sort of auxin-independent transcriptional response mechanism.

Apart from classes B/C, a third class of ARFs, class A, exists in land plants. This class contains the commonly named “activator” ARFs. Due to their ability to interact with Aux/IAA repressors, they trigger transcriptional control in an auxin-dependent manner. Therefore, our first ARF, *CaARF* (class C) went through gene duplication events during or after the divergence of charophytes into land plants, that led to the apparition of class-A and B ARFs.

Auxin-dependency within the NAP appears on the scene for the first time in land plants, with the acquisition of TIR1/AFB-Aux/IAA perception system that relies on the degradation of Aux/IAA repressors, proteins with no DNA-binding capacity. This suggests that auxin-dependent

transcriptional regulation was added to an already existing mechanism of transcriptional control formed by ARFs and TPL, in support of our evolutionary studies done in charophyte organisms (Leyser, 2017). TIR1/AFB-Aux/IAA system relies on proteosomal degradation and therefore, constitutes a quick-change mechanism that has been “copied-pasted” by JA or strigolactones signalling pathways.

*M. polymorpha*, right at the beginning of land plants evolution, presents the simplest NAP with just one copy of TIR1/AFB, Aux/IAA and classes A/B/C-ARFs. This is enough for conferring this plant a relative body complexity: polarized thallus, rhizoids and a/sexual reproductive organs. However, the simplicity of *Marchantia*'s system is rather an “exception to the rule”. An impressive diversification of the NAP protein families -mostly Aux/IAAs and ARFs- is observed in other non-vascular plants such as *P. patens* or *S. moellendorffii*. In the gymnosperms to angiosperms transition these genes diversification becomes even more important. Classes A/B/C-ARF and Aux/IAAs duplication events led to a “multi-coloured” auxin landscape made of subtle differences, in which each gene copy can have their own properties.

A nice example of this is the diversity in DNA-binding preferences that we presented in Chapter II. All ARFs possess a domain specific to plants, named B3 domain, that mediates interaction with AuxREs in a monomeric sense. However, structural data proved that ARFs bind to AuxREs repeats as dimers. Whereas the B3 domain is really well conserved, variations can be found in the domain that mediates ARFs dimerization (DD) and in the loops that connect DD with B3. Here we show that ARFs dimeric binding to DNA can happen through different ways and preferential binding sites for classes A/B/C. These properties could be maintained inside each ARF class but it could also happen that within each class, different ARFs can still present more particularities. Our DAP-seq and DNA binding tests showed that whereas some of the DNA motifs are specifically bound by an ARF, there are others that could be bound by several ARFs. This mix between “shared” and “specific” binding sites suggests the existence of mechanisms such as competition or simultaneous binding of different ARFs to a same region, that can increase even more the diversity of the auxin response.

Inside a cell, where the only thing that remains the same is the DNA sequence, the different ARFs and Aux/IAAs combinations and the diversity in their structural, biochemical and mechanical properties, will decide where, when, how and which auxin-responsive genes will be expressed.

Gene duplication and even whole genome duplication events that lead to massive diversifications of protein families are very common among plants. However, plants do not do this in a random manner. When something works and it is useful and efficient in a wide range of physiological processes, they maintain their traditions and they keep it the way it is. This is the case of our

TOPLESS protein, TPL. As it happened with the ARFs, we also found TPL homologues in charophyte organisms and several copies of this protein exist in land plants. However, TPL/TPRs are extremely conserved along evolution and through all the duplication events. Due to this high conservation, we can affirm that the structural and biochemical properties of TPL, presented in Chapter I, were the same million years ago when plants had not yet colonized the Earth. Moreover, not only TPL mechanisms are conserved among plants, but as shown here, some features are also shared by co-repressors of different organisms.

In a general manner, using auxin pathway as an example, this work illustrates the evolutionary guidelines of transcriptional repression. Upstream a signalling cascade we find co-repressors, with very few families existing in each organism. As “hub proteins” involved in many processes, they need to be maintained the same since changes in them would lead to alterations in all these processes. Therefore, at the beginning of a signalling cascade, there is no space for change. However, when moving downstream we find TFs, proteins specific of one single process. Thus, they can diversify with their diversification leading to modifications with less impact that can turn into a more varied and robust response. The antiquity of co-repressors, such as TPL, versus the more recent innovations of certain transcriptional repressors, such as Aux/IAA, suggests that the latter adapt to the former for the sake of transcriptional control.



## Materials & Methods

The information presented in this section is referred to the materials and methods used in Chapters II and III. For Chapter I this information is included in the supplementary materials part of the paper. Techniques used in Chapters II and III that have already been explained in the supplementary materials will be referred back to Chapter I unless extra details are necessary.

### Materials

#### Bacterial material

Clonings were done in *E. coli* DH5 $\alpha$  strain (Invitrogen). Recombinant proteins expression was done in *E. coli* Rosetta2 strain (Novagen).

### Methods

#### Molecular Biology

Bacterial DNA was extracted with Minipreps NucleoSpin Plasmid kit (Machery-Nagel).

- Construction of plasmids used for recombinant protein expression in *E. coli*

cDNA sequences coding for the proteins of interest were cloned into one of the following plasmids: pETM11, for 6xHis tagged proteins; pETM33, for 6xHis-Glutathione S-transferase (GST) tagged proteins; and pETM40 for that allows the addition of a Maltose Binding Protein (MBP) tagged proteins. All tags were fused in N-terminal.

cDNA sequences coding for full-length proteins were synthetic coding sequences (GeneArt or ThermoFisher, Table 5). cDNAs coding for separated domains were amplified by PCR from the corresponding synthetic DNA vectors using Phusion polymerase (ThermoFisher). Amplifications were done with oligonucleotides including cloning restriction sites for pETM11/33/40 plasmids (Table 5). The amplification products were purified from agarose gels with NucleoSpin Gel and PCR clean-up kit (Machery-Nagel), digested with the corresponding restriction enzymes and cloned into pETM11/33 and 40 plasmids. Cloning were verified by sequencing (Eurofins) using primers located in the plasmids (T7 promoter and terminator) or specifically designed primers for very long inserts.

- Site Directed Mutagenesis (Used for TPL mutants)

Site Directed Mutagenesis (SDM) was done by amplification of entire plasmids with oligonucleotides containing the site-specific mutation. Oligonucleotides were designed using

QuickChange Primer Design (<http://www.genomics.agilent.com/primerDesignProgram.jsp>) (Table 6).

PCR reactions contained 5-50ng of the initial vector, 125ng of each of oligonucleotide, 0.2μM of each dNTPs and Phusion polymerase in its buffer (Thermo Scientific).

PCR amplification program consisted of an initial denaturation step (30 sec-95°C) followed by 20 cycles of denaturation (30sec-95°C), annealing (1min- $T_m$  with  $T_m$  or annealing temperature were calculated and optimized for each pair of oligonucleotides) and elongation (1min/kb-72°C).

After DpnI enzyme treatment (1h at 37°C), that digests the initial methylated bacterial plasmid, *E. coli* DH5α bacteria were transformed with the amplification product. Plasmids extracted from the transformed colonies were sent for sequencing (Eurofins) using primers located in the plasmids (T7 promoter and terminator) to verify the presence of the desired mutation.

**Table 5. Cloning methods used for the construction of the different protein expression plasmids.**

<b><i>A. thaliana</i> constructs</b>		
AtTPL FL	Synthetic DNA (Thermofisher)	
AtTPL184	Cloned from AtTPL FL synthetic DNA	
AtTPL202	PCR amplification from AtTPL FL synthetic	5' GGCGCCATGGGCTCTTCTCTTAGTAGAGAGCTC 3' 5' CGTGCGGCCGCTTAAGAGTGATCCACAAAAAGAGTC 3'
<b><i>K. nitens</i> constructs</b>		
KnRAV-FL	Synthetic DNA (Thermofisher)	
KnRAV-AP2-B3	PCR amplification	5' TATACCATGGCAGAACGCGGTCTGGAACAGGGCCTG 3' 5' TAATCTCGAGTTAACGTGCATAACCAATATACAGCTG 3'
KnRAV-RLFG	PCR amplification	5' TATACCATGGCAGAACGCGGTCTGGAACAGGGCCTG 3' 5' GCATCTCGAGTTATTTCGGTGTGAGAACACCCGGTGC 3'
KnRAV-LxIxPP	PCR amplification	5' TATACCATGGCACAGCTGCTGCGTACCGGTAGCCTG 3' 5' GTGCTCGAGTTATTTCGGATGAAGATACGGCT 3'
KnRAV-PB1	PCR amplification	5' TATACCATGGCAGAACAGCGTCCGAGGTTAAATGT 3' 5' GTGCTCGAGTTATTTCGGATGAAGATACGGCT 3'
<b><i>C. atmophyticus</i> constructs</b>		
CaARF-FL	Synthetic DNA (GeneArt)	
CaARF-DBD	Cloned from CaARF FL synthetic	
CaARF-PB1	Cloned from CaARF FL synthetic	

**Table 6. Oligonucleotides used for SDM for AtTPL mutants.**

<b>SDM oligonucleotides</b>	
AtTPL202m/L130A	5' AACTTCCGGGAGAATGAACAGGCCTCCAAGTATGGGGACACCAAG 3' 5' CTGGTGTCCCCATACTTGGAGGCCTGTTCATTCTCCCGAAGTT 3'
AtTPL202m/Y133A	5' TGCAGACTTGGTGTCCCCAGCCTTGGACAGCTGTTCATTCTCCCGAAG 3' 5' CTTCCGGGAGAATGAACAGCTGTCCAAGGCTGGGGACACCAAGTCTGCA 3'
AtTPL202m/T116A	5' GTTCTCCAATGTCAACAGCTGTGCTATTTCTTGAAAAGCTCCTCATT 3' 5' AATGAGGAGCTTTTCAAGGAAATAGCACAGCTGTTGACATTGGAGAAC 3'
AtTPL202m/K102S	5' GCTCCTCATTA AAAAGTTGAAAACACTGACAGATCTTCACTAGTATATCCACAGCCTTGGGAC 3' 5' GTCCCAAGGCTGTGGATATACTAGTGAAAGATCTGTCAGTGTTTTCACTTTTAATGAGGAGC 3'
AtTPL202m/Q117S-E122S	5' GTTCATTCTCCCGAAGTTCGACAATGTCAACAGCGATGTTATTCTTGAAAAGCTCCTCATTA AAA 3' 5' TTTTAATGAGGAGCTTTTCAAGGAAATAACATCGCTGTTGACATTGTCGAACTCCGGGAGAATGAAC 3'
AtTPL202m/R172S	5' CTCTTAGAAATTCAAGGCTGTCGACTTTGATCAACCAGAGCT 3' 5' AGCTCTGGTTGATCAAAGTCGACAGCCTTGAATTTCTAAGAG 3'

### Protein biochemistry

- Expression and purification of recombinant proteins

All proteins were expressed in *E. coli* Rosetta2 strain. Bacteria cultures were grown in liquid LB medium till an O.D<sub>600nm</sub> of 0.6-0.9. Protein expression was induced with isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) at a final concentration of 400 $\mu$ M. Protein production was done overnight at 18°C. Bacteria cultures were centrifuged and the resulting bacteria pellets were resuspended in lysis buffer in which cells were lysed by sonication. Lysis buffers for each protein are summarized in Table 7.

Soluble fractions were recovered after centrifugation of the lysed cultures. Proteins were purified passing the soluble fractions through sepharose resins, previously equilibrated with the corresponding lysis buffer, to which the different tags fused to the proteins were bound. Bound proteins were eluted with high-affinity molecules for each of the resins. MBP-tagged proteins were purified in amylose resin and eluted with Maltose 10mM. His-tagged and MBP-his-tagged proteins were purified in nickel-sepharose resin and eluted with imidazole 300mM after a previous wash step with imidazole 30mM.

Proteins used for crystallography were submitted to a second purification in Gel filtration Superdex 200 16/60 (GE Healthcare) equilibrated with the corresponding lysis buffer.

**Table 7. Proteins used in Chapter II and III. Indicated tags to which they are fused and the purification buffers.**

<b><i>A. thaliana</i> proteins</b>		
AtARF2/10	MBP-His-tagged	Tris 20mM pH8; NaCl 500mM; TCEP 1mM
AtARF5/m3	MBP-His-tagged	Tris 20mM pH8; NaCl 500mM; EDTA 0.5mM; PMSF 0.5mM; TCEP 1mM; triton 0.2%.
<b><i>K. nitens</i> proteins</b>		
KfRAV-AP2-B3	MBP-tagged	Tris 20mM pH 8; NaCl 200mM; DTT 1mM; EDTA 2mM
KfRAV-RLFG	MBP-tagged	Tris 20mM pH 8; NaCl 200mM; DTT 1mM; EDTA 2mM
KfRAV-LxIxPP	MBP-tagged	Tris 20mM pH 8; NaCl 200mM; DTT 1mM; EDTA 2mM
KfRAV-PB1	His-tagged	CAPS 100mM pH 9.6; TCEP 1mM
<b><i>C. atrophyticus</i> proteins</b>		
CaARF-FL	MBP-tagged	Tris 20mM pH 8; NaCl 200mM; DTT 1mM; EDTA 2mM
CaARF-DBD	MBP-tagged	Tris 20mM pH 8; NaCl 500mM; DTT 1mM; EDTA 2mM
CaARF-PB1	His-tagged	Tris 20mM pH8; TCEP 1mM

- Control and quantification of proteins purification
  - a. SDS-page polyacrylamide gels for protein purification checking

The different fractions resulting from proteins expression and purification (non-induced, induced, total fraction, soluble fraction, flow through, wash and elution fractions) were analysed by SDS-page polyacrylamide gel electrophoresis to reveal the presence or absence and the purity of the protein of interest in each fraction.

10%-polyacrylamide gels were used for the detection of proteins with a molecular weight (MW) ranging from 80-120KDa; 12%-polyacrylamide gels were used for the detection of proteins with MW between 40-80KDa; 15%-polyacrylamide gels were used for the detection of proteins with MW between 10-40KDa; for proteins smaller than 10KDa 18%-polyacrylamide gels were used.

Proteins were denaturated by mixing them with a denaturing buffer (50mM Tris-HCl pH6.8; 10% p/v glycerol; 1% p/v SDS; 0.0025% p/v bromophenol blue; 0.4% dithiothreitol) followed by 5 min at 95°C. Denaturated proteins were loaded in SDS-page polyacrylamide gels together with a molecular weight marker (Precision Plus Proteins Standar Dual Color, Bio-Rad). Proteins migration was carried out in LAEMLI buffer (25mM Tris-HCl; 192mM glycine; 0.1% p/v SDS) till bromophenol blue leaves the gel. Proteins were revealed in the gels by colouring them with a Coomassie blue solution (0.25%

p/v bleu R 250; 45% ethanol; 10% acetic acid) and then discolouring them with a discolouring solution (10% ethanol; 10% acetic acid).

Stain-free SDS-page polyacrylamide gels (Biorad) were used for the detection of Tryptophan-containing proteins present a very low concentrations and so, not detectable by colorimetric revealing with Coomassie blue. Proteins were visualized after UV-light gel activation (Biorad ChemiDoc MP Imaging System).

#### b. Protein quantification by 280nm-absorbance measurement

After being purified proteins were quantified by Nanodrop 2000 (ThermoFisher) measurements of their absorbance at 280nm. Concentrations are calculated by Beer-Lambert law, according to which the absorbance of a sample at a given wavelength,  $A_\lambda$ , is proportional to the concentration of the sample, C:

$$A_\lambda = \varepsilon \cdot l \cdot C,$$

where  $\varepsilon$  is the molar extinction coefficient and  $l$  is the optical distance. Molecular weights and  $\varepsilon$  were calculated for each protein using ProtParam tool (Expasy) and introduced in the Nanodrop measurement software for a more precise calculation of the protein concentration.

#### c. Protein quantification on SDS-page polyacrylamide gels

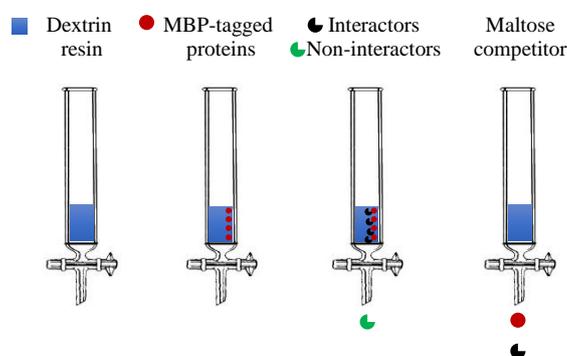
AtARF proteins, when produced in bacteria, are highly degraded. For a precise quantification of the amount of full-length protein AtARF samples at different theoretical calculations were loaded on SDS-page polyacrylamide gels 10%. Bands corresponding with sizes corresponding to the full-length proteins were quantified upon the gel using as reference AtARF2 full-length protein band and protein concentrations were recalculated based on the bands intensities. These calculations were done with ImageLab software.

- Protein-protein interactions characterization

##### a. Co-purification assays

KnRAV-LxIxPP and CaARF co-purifications with AtTPL202 were done by fixing 100 $\mu$ g of MBP-tagged KnRAV-LxIxPP and MBP-tagged CaARF to a dextrin sepharose column previously washed with Milli-Q water and equilibrated with Tris20mM pH8; NaCl

200mM; TCEP 1mM. For KnRAV-RLFG interactions, the buffer used was Arginine/Glutamate (R/E) 50mM. His-tagged TPL proteins (60µg) were passed through the column with the bound MBP-proteins in a second step. For KnRAV-LxIxPP and CaARF interactions, TPL was diluted in Tris 100mM pH8 (final interaction buffer CAPS 20mM pH 10.5; Tris 90mM pH8; NaCl 50mM; TCEP 1mM). For KnRAV-RLFG interactions, TPL was diluted in R/E 50mM (final interaction buffer CAPS 20mM pH 10.5; R/E 45mM; NaCl 50mM; TCEP 1mM). After addition of TPL both proteins were incubated in contact with the resin for 15 minutes. Nonspecific interactions were removed by a washing step with the equilibration buffer. Protein complexes were eluted with elutions 1-4 done with 50 µl of maltose 10mM. MBP was used as control for specific interactions (Fig 51).

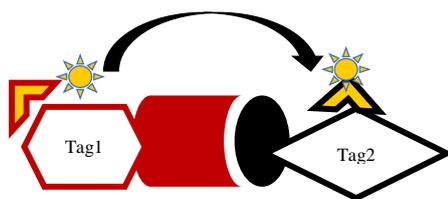


**Fig 51. Co-purification method for protein-protein interactions.** MBP-tagged proteins are bound to the resin. Potential interactors are then added to the column. Proteins that do not interact with the MBP-tagged proteins will flow through the column whereas interactors will remain bound. Complexes are finally eluted with Maltose 10mM competitor.

#### b. Homogeneous Time Resolved Fluorescence (HTRF)

HTRF methodology determines interactions based on a FRET signal. Potential interactors need to present different tags. Two antibodies against the two different tags, one that carries a donor fluorophore and the other that carries an acceptor fluorophore, are added to the reaction. If the two potential interactors do interact, donor and acceptor will get close to each other emitting a FRET signal indicative of the protein-protein interaction at 665nm after a 620nm-excitation (Fig 52).

His-tagged AtTPL202 (625nM) and MBP-tagged KnRAV proteins (100nM) interactions were analysed by HTRF (ref) using CisBio Bioassays Anti-His acceptor d2 and Anti-MBP donor Tb antibodies. Three simultaneous replicas were done for each binding mixture.



**Fig 52. HTRF method for protein-protein interactions.**

Potential interactors must be fused to two different tags. Antibodies against the two tags, labelled with a donor and an acceptor fluorophore each are added to the mix. A FRET signal will happen when the two antibodies come close to each other upon the two proteins interacting.

- DNA-protein interactions characterization by EMSA

- a. Oligonucleotides

DR, ER, and IR binding sequences were artificially designed with TGTC consensus for each site and the appropriate spacing. ER8-DRs DNA sequences were extracted from IAA19 natural promoter sequence. The oligonucleotides corresponding to the sense and antisense strands were synthesized by Eurofins and sent as lyophilized material. Oligonucleotides for the sense strand were designed with a hanging G in 5' that allows the labelling of the DNA. EMSA oligonucleotides are presented in Table 8.

- Annealing

Lyophilized oligonucleotides were resuspended in autoclaved Milli-Q water to a final concentration of 100  $\mu$ M. Annealing of both strands was done by mixing 5  $\mu$ l of each oligonucleotide in Tris 50mM; NaCl 150mM (annealing buffer). The mixes were incubated at 98°C for 5 minutes and then they were let to progressively cool down to room temperature overnight. Annealed oligos are at a final concentration of 10  $\mu$ M.

- Labelling

Annealed oligonucleotides were labelled with Cyanine-5 (Cy5) fluorophore in the 5'-G by a Klenow end filling reaction that adds a Cy5-dCTP to complement the overhanging guanidine. Annealed oligos, at a final concentration of 200nM were incubated at 37°C for 1 hour with Cy5-dCTP (0.4  $\mu$ M) and klenow enzyme in NEB2 buffer (New England Biolabs). Enzyme was inactivated by a 10-minutes incubation at 65°C. Oligonucleotides were conserved at 4° in darkness.

- b. IAA19 sequences

IAA19 promoter sequences were synthetic DNA sequencess (Genewiz) cloned inside pUC57 plasmid. IAA19 sequences were amplified by PCR using labelled oligonucleotides complementary to 5' and 3' extremes of IAA19 (Tm 50°C). IAA19 wt and mutant sequences are presented at the end of this section.

## c. EMSAs

Electrophoretic Mobility Shift Assays, EMSAs, were done on agarose 2% native gels prepared with TBE buffer 0.5X. Gels were pre-run in TBE buffer 0.5X at 90V for 90 minutes at 4°C. Protein-DNA mixes contained competitor Fish DNA (final concentration 0.045mg/ml) and labelled DNA (final concentration 20nM) in the interaction buffer 25mM HEPES pH7,4; 1mM EDTA; 2mM MgCl<sub>2</sub>; 100mM KCl; 10% glycerol; 1mM DTT; 0.5mM PMSF; 0,1% Triton. Mixes were incubated in darkness for 1 hour at 4°C and next loaded in the gels. Gels were run for 1 hour at 90V at 4°C in TBE 0.5X DNA-protein and bindings were visualized on the gels with Cy5-exposition filter (Biorad ChemiDoc MP Imaging System).

**Table 8. Oligonucleotides used for EMSA tests. Only the sense strand is indicated.**

Oligonucleotide	DNA sequence (5'→3')
DR4	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR5	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR6	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR7	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR8	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR9	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR10	GATACACGCAATGTCGGCCTTTGTCGGTTCCACTCA
DR15	GATGTTTTGTCGGCTTCAAACCTTGATTTGTCGGCTCAT
DR15m	GATGTTTTaaCGGCTTCAAACCTTGATTTaaCGGCTCAT
ER4	GATACACGCTTGTGGCAAGCCGACAACCACTCA
ER5	GATACACGCTTGTGGCAAGCCGACAACCACTCA
ER6	GATACACGCTTGTGGCAAGCCGACAACCACTCA
ER7	GATACACGCTTGTGGCAAGCCGACAACCACTCA
ER8	GATACACGCTTGTGGCAAGCCGACAACCACTCA
ER9	GATACACGCTTGTGGCAAGCCGACAACCACTCA
ER10	GATACACGCTTGTGGCAAGCCGACAACCACTCA
IR0	GATGCAGTCATGTGCGGACATGTCGGCATGTGCTCACAT
IR0m	GATGCAGTCATGTGCGGhAtaaCGGCATGTGCTCACAT
IR13	GATGCAGCCGACAAAACACATGATTTGTCGGCTCACAT
IR13m	GATGCAGCCGhA AAAACACATGATTTaaCGGCTCACAT
IR7	GATGCAGTCAACCGACAAAAATTTGTCGGGTGCTCACAT
ER7x2	GATACACGAAATGTCGGTGAATCACCGACATTTTGAATGTCGGTGAATCACCGACATTCGTGTATC
ER8 C/NC	GCAAACCTATGTCCTCTCATGTGACCGACACCGCATC
ER8 C/C	GCAAACCTATGTCCTCTCATGTGACCGACaACCGCATC
ER8 WC/WC	GCAAACgggTGTCaTTCATGTGaatGACaACCGCATC
ER8 Single Site	GCAAACCTATGTCCTCTCATGTGACCGhCACCGCATC
ER8 No site	GCAAACCTATaaCTCTCATGTGACCGhCACCGCATC
ER8-DRs	GACCAAACCTATGTCCTCTCATGTGACCGACACCGCATCCTCAGTTGACCTGTCCTGCCCCACTTTGTCCTCCCCACACAAA
ER8m-DRs	GACCAAACCTATaTCTCTCATGTGACCGaaCACCGCATCCTCAGTTGACCTaTCTCTGCCCCACTTTaTCTCCCCACACAAA
ER8-DRsm	GACCAAACCTATGTCCTCTCATGTGACCGACACCGCATCCTCAGTTGACCTaTCTCTGCCCCACTTTaTCTCCCCACACAAA
ER8m-DRsm	GACCAAACCTATaTCTCTCATGTGACCGaaCACCGCATCCTCAGTTGACCTaTCTCTGCCCCACTTTaTCTCCCCACACAAA

### Phylogenetic and structural analysis

Protein sequences alignments were done with Multialin (<http://multalin.toulouse.inra.fr/multalin/>) and ESPrit (<http://espruit.ibcp.fr/ESPruipt/ESPruipt/>) online tools.

3D protein structure visualization and 3D structures alignments were done with Pymol software.

Protein structure modelling was done with Phyre2 online tool (ref).

Phylogenetic trees were built with MEGA software.

## IAA19 sequences

### IAA19wt

CTTAACCACCTTGTAATGCCGGTCTTTATTAATTTTGAGAAAAAATAAAAAACAGCACCAAACCTTATGTCTCTCATGTGA  
CCGACCACCGCATCCTCAGTTGACCTGTCTCTGCCCCACTTTGTCTCCCACACAACTGAATAACAAGAAGAAGAAG  
ACTTTAGATATCAAATGACTCCACGTGTCGATATTGGATTGGTTTTTCATTGGTTGTATCGTGTGGACCAACGAAGCAACA  
TATAAAAAGCACGACGCGGTGCCATTACTACAATAAGAGAAGTGTAGGAGAAGAAAGTTCTCATTTTCATAATTGTATC  
AAATTGTGAGAGGAAAAAAGAAGTTCAGAAIAA19ER8m

CTTAACCACCTTGTAATGCCGGTCTTTATTAATTTTGAGAAAAAATAAAAAACAGCACCAAACCTTATATCTCTCATGTGA  
CCGAACACCGCATCCTCAGTTGACCTGTCTCTGCCCCACTTTGTCTCCCACACAACTGAATAACAAGAAGAAGAAG  
ACTTTAGATATCAAATGACTCCACGTGTCGATATTGGATTGGTTTTTCATTGGTTGTATCGTGTGGACCAACGAAGCAACA  
TATAAAAAGCACGACGCGGTGCCATTACTACAATAAGAGAAGTGTAGGAGAAGAAAGTTCTCATTTTCATAATTGTATC  
AAATTGTGAGAGGAAAAAAGAAGTTCAGAA

### IAA19DR11m

CTTAACCACCTTGTAATGCCGGTCTTTATTAATTTTGAGAAAAAATAAAAAACAGCACCAAACCTTATGTCTCTCATGTGA  
CCGACCACCGCATCCTCAGTTGACCTATCTCTGCCCCACTTTATCTCCCACACAACTGAATAACAAGAAGAAGAAG  
ACTTTAGATATCAAATGACTCCACGTGTCGATATTGGATTGGTTTTTCATTGGTTGTATCGTGTGGACCAACGAAGCAACA  
TATAAAAAGCACGACGCGGTGCCATTACTACAATAAGAGAAGTGTAGGAGAAGAAAGTTCTCATTTTCATAATTGTATC  
AAATTGTGAGAGGAAAAAAGAAGTTCAGAA

### IAA19ER8m-DR11m

CTTAACCACCTTGTAATGCCGGTCTTTATTAATTTTGAGAAAAAATAAAAAACAGCACCAAACCTTATATCTCTCATGTGA  
CCGAACACCGCACTCAGTTGACCTATCTCTGCCCCACTTTATCTCCCACACAACTGAATAACAAGAAGAAGAAGAC  
TTTAGATATCAAATGACTCCACGTGTCGATATTGGATTGGTTTTTCATTGGTTGTATCGTGTGGACCAACGAAGCAACATA  
TAAAAAGCACGACGCGGTGCCATTACTACAATAAGAGAAGTGTAGGAGAAGAAAGTTCTCATTTTCATAATTGTATCAA  
ATTGTGAGAGGAAAAAAGAAGTTCAGAA



## Bibliography

- Abel, S., and Theologis, A. (1996). Early genes and auxin action. *Plant Physiol.* *111*, 9–17.
- Adamowski, M., and Friml, J. (2015). PIN-Dependent Auxin Transport: Action, Regulation, and Evolution. *Plant Cell Online* *27*, 20–32.
- Agarwal, M., and Mathew, S.J. (2015). Critical Review The Groucho / Transducin-Like Enhancer of Split Protein Family in Animal Development. *Int. Union Biochem. Mol. Biol.* *67*, 472–481.
- Allen, M.D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a  $\beta$ -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* *17*, 5484–5496.
- Alonso, C.R., and Wilkins, A.S. (2005). The molecular elements that underlie developmental evolution. *Nat. Rev. Genet.* *6*, 709–715.
- Baca, B.E., and Elmerich, C. (2007). Microbial Production of Plant Hormones. *Plant Soil* *133*, 1–8.
- Bartlett, A., O'Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* *12*, 1659–1672.
- Berg, O.G., and von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins. *J. Mol. Biol.* *193*, 723–743.
- Boer, D.R., Freire-Rios, A., Van Den Berg, W.A.M., Saaki, T., Manfield, I.W., Kepinski, S., López-Vidriero, I., Franco-Zorrilla, J.M., De Vries, S.C., Solano, R., et al. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell* *156*, 577–589.
- Bowman, J.L. (2013). Walkabout on the long branches of plant evolution. *Curr. Opin. Plant Biol.* *16*, 70–77.
- Busch, W., Miotk, A., Ariel, F.D., Zhao, Z., Forner, J., Daum, G., Suzaki, T., Schuster, C., Schultheiss, S.J., Leibfried, A., et al. (2010b). Transcriptional control of a plant stem cell niche. *Dev. Cell* *18*, 849–861.
- Calderón Villalobos, L., Lee, S., De Oliveira, C., Ivetac, A., Brandt, W., Armitage, L., Sheard, L.B., Tan, X., Parry, G., Mao, H., Zheng, N., Napier, R., and Kepinski, S.E.M. (2012). A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nat. Chem. Biol.* *8*, 477–485.
- Causier, B., Lloyd, J., Stevens, L., and Davies, B. (2012a). TOPLESS co-repressor interactions and their evolutionary conservation in plants. *Plant Signal. Behav.* *7*, 325–328.
- Causier, B., Ashworth, M., Guo, W., and Davies, B. (2012b). The TOPLESS Interactome: A Framework for Gene Repression in Arabidopsis. *Plant Physiol.* *158*, 423–438.
- Cerna, D., and Wilson, D.K. (2005). The structure of Sif2p, a WD repeat protein functioning in the SET3 corepressor complex. *J. Mol. Biol.* *351*, 923–935.
- Chandler, J.W. (2016). Auxin response factors. *Plant. Cell Environ.* *39*, 1014–1028.
- Cho, H., Ryu, H., Rho, S., Hill, K., Smith, S., Audenaert, D., Park, J., Han, S., Beeckman, T., Bennett, M.J., et al. (2014). A secreted peptide acts on BIN2-mediated phosphorylation of ARFs to potentiate auxin response during lateral root development. *Nat. Cell Biol.* *16*, 66–76.
- Chodaparambil, J. V, Pate, K.T., Hepler, M.R.D., Tsai, B.P., Muthurajan, U.M., Luger, K., Waterman, M.L., and Weis, W.I. (2014a). Molecular functions of the TLE tetramerization domain in Wnt target gene repression. *EMBO J.* *33*, 719–731.
- Choi, H.-K., Choi, K.-C., Kang, H.-B., Kim, H.-C., Lee, Y., Haam, S., Park, H.-G., and Yoon, H.-G. (2008). Function of multiple Lis-Homology domain/WD-40 repeat-containing proteins in feed-forward transcriptional repression by silencing mediator for retinoic and thyroid receptor/nuclear receptor corepressor complexes. *Mol. Endocrinol.* *22*, 1093–1104.
- Conner, J., and Liu, Z. (2000). LEUNIG, a putative transcriptional corepressor that regulates AGAMOUS expression during flower development. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 12902–12907.
- Crocker, J., Preger-Ben Noon, E., and Stern, D.L. (2016). The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. *Curr. Top. Dev. Biol.* *117*, 455–69.
- Delaux, P.-M., Nanda, A.K., Mathé, C., Sejalón-Delmas, N., and Dunand, C. (2012). Molecular and biochemical aspects of plant terrestrialization. *Perspect. Plant Ecol. Evol. Syst.* *14*, 49–59.
- Delwiche, C.F., and Cooper, E.D. (2015). The evolutionary origin of a terrestrial flora. *Curr. Biol.* *25*, R899–R910.
- Dezfulian, M.H., Jalili, E., Roberto, D.K.A., Moss, B.L., Khoo, K., Nemhauser, J.L., and Crosby, W.L. (2016). Oligomerization of SCF TIR1 Is Essential for Aux / IAA Degradation and Auxin Signaling in Arabidopsis. *PLoS Genet.* *12*, 1–18.
- Dharmasiri, S., Jayaweera, T., and Dharmasiri, N. (2013). Plant Hormone Signalling: Current Perspectives on Perception and Mechanisms of Action. *Ceylon J. Sci. (Biological Sci.)* *42*, 1–17.
- Dinesh, D.C., Villalobos, L.I. a C., and Abel, S. (2016). Structural Biology of Nuclear Auxin Action. *Trends Plant Sci.* *21*, 302–316.
- Dinesh, D.C., Kovermann, M., Gopalswamy, M., Hellmuth, A., Calderón Villalobos, L., Lilie, H., Balbach, J., and Abel, S.. (2015).

- Solution structure of the PsIAA4 oligomerization domain reveals interaction modes for transcription factors in early auxin response. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 6230–6235.
- Domozych, D.S., Popper, Z.A., and Sørensen, I. (2016). Charophytes: Evolutionary Giants and Emerging Model Organisms. *Front. Plant Sci.* *7*, 1470.
- Espinosa-Ruiz, A., Martínez, C., de Lucas, M., Fàbregas, N., Bosch, N., Caño-Delgado, A.I., and Prat, S. (2017). TOPLESS mediates brassinosteroid control of shoot boundaries and root meristem development in *Arabidopsis thaliana*. *Development* *144*, 1619–1628.
- Farcot, E., Armitage, L., Picard, F., Oliva, M., Das, P., Larrieu, A., Wells, D., Gue, Y., Guyomarc, S., Cellier, C., et al. (2011). The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol. Syst. Biol.* *5*.
- Farcot, E., Lavedrine, C., and Vernoux, T. (2015). A Modular Analysis of the Auxin Signalling Network. *PLoS One* *10*, 1–26.
- Feng, M., and Kim, J.-Y. (2015). Revisiting Apoplastic Auxin Signaling Mediated by AUXIN BINDING PROTEIN 1. *Mol. Cells* *38*, 829–835.
- Finet, C., and Jaillais, Y. (2012). AUXOLOGY: When auxin meets plant evo-devo. *Dev. Biol.* *369*, 19–31.
- Finet, C., Fourquien, C., Vinauger, M., Berne-dedieu, A., Chambrier, P., Paindavoine, S., and Scutt, C.P. (2010). Parallel structural evolution of auxin response factors in the angiosperms. *Plant J.* *63*, 952–959.
- Finet, C., Berne-Dedieu, A., Scutt, C.P., and Marletaz, F. (2013). Evolution of the ARF gene family in land plants: Old domains, new tricks. *Mol. Biol. Evol.* *30*, 45–56.
- Flores-Saaib, R.D., and Courey, A. J. (1999). Analysis of Groucho-histone interactions suggests mechanistic similarities between Groucho-and Tup1-mediated repression. *Nucleic Acids Res.* *28*, 4189.
- Flores-Sandoval, E., Eklund, D.M., and Bowman, J.L. (2015). A Simple Auxin Transcriptional Response System Regulates Multiple Morphogenetic Processes in the Liverwort *Marchantia polymorpha*. *PLoS Genet.* *11*, 1–26.
- Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 2367–2372.
- Fu, M., Kang, H.K., Son, S.H., Kim, S.K., and Nam, K.H. (2014). A subset of arabidopsis RAV transcription factors modulates drought and salt stress responses independent of ABA. *Plant Cell Physiol.* *55*, 1892–1904.
- Fukazawa, J., Teramura, H., Murakoshi, S., Nasuno, K., Nishida, N., Ito, T., Yoshida, M., Kamiya, Y., Yamaguchi, S., and Takahashi, Y. (2014). DELLAs function as coactivators of GAI-ASSOCIATED FACTOR1 in regulation of gibberellin homeostasis and signaling in *Arabidopsis*. *Plant Cell* *26*, 2920–2938.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* *31*, 3784–3788.
- Golovenko, D., Manakova, E., Zakrys, L., Zaremba, M., Sasnauskas, G., Gražulis, S., and Siksnyš, V. (2014). Structural insight into the specificity of the B3 DNA-binding domains provided by the co-crystal structure of the C-terminal fragment of BfiI restriction enzyme. *Nucleic Acids Res.* *42*, 4113–4122.
- Gonzalez, D., Bowen, A.J., Carroll, T.S., and Conlan, R.S. (2007). The transcription corepressor LEUNIG interacts with the histone deacetylase HDA19 and mediator components MED14 (SWP) and CDK8 (HEN3) to repress transcription. *Mol. Cell Biol.* *27*, 5306–5315.
- Graeff, M., Straub, D., Eguen, T., Dolde, U., Rodrigues, V., Brandt, R., and Wenkel, S. (2016). MicroProtein-Mediated Recruitment of CONSTANS into a TOPLESS Trimeric Complex Represses Flowering in *Arabidopsis*. *PLoS Genet.* *12*, 1–22.
- Grones, P., and Friml, J. (2015). Auxin transporters and binding proteins at a glance. *J. Cell Sci.* *128*, 1–7.
- Guilfoyle, T.J., and Hagen, G. (2012). Plant Science Getting a grasp on domain III / IV responsible for Auxin Response Factor – IAA protein interactions. *Plant Sci.* *190*, 82–88.
- Habets, M.E.J., and Offringa, R. (2014). PIN-driven polar auxin transport in plant developmental plasticity: A key target for environmental and endogenous signals. *New Phytol.* *203*, 362–377.
- Han, M., Park, Y., Kim, I., Kim, E., Yu, T., Rhee, S., and Suh, J. (2015). Structural basis for the auxin-induced transcriptional regulation by Aux / IAA17. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 18613–18618.
- Hao, Y., Wang, X., Li, X., Bassa, C., Mila, I., Audran, C., Maza, E., Li, Z., Bouzayen, M., Van Der Rest, B., et al. (2014). Genome-wide identification, phylogenetic analysis, expression profiling, and protein-protein interaction properties of TOPLESS gene family members in tomato. *J. Exp. Bot.* *65*, 1013–1023.
- Harrison, J. (2016). Development and genetics in the evolution of land plant body plans Development and genetics in the evolution of land plant body plans. *New Phytol.* *372*, 20–25.
- Hill, K. (2015). Post-translational modifications of hormone-responsive transcription factors : the next level of regulation. *J. Exp. Bot.* *66*, 4933–4945.
- Hiratsu, K., Mitsuda, N., Matsui, K., and Ohme-Takagi, M. (2004). Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* *321*, 172–178.

- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* *5*, 3978.
- Ikeda, M., Mitsuda, N., and Ohme-Takagi, M. (2009). Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. *Plant Cell* *21*, 3493–3505.
- Ito, J., Fukaki, H., Onoda, M., Li, L., Li, C., Tasaka, M., and Furutani, M. (2016). Auxin-dependent compositional change in Mediator in ARF7- and ARF19-mediated transcription. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 6562–6567.
- Jaillais, Y., and Chory, J. (2010). Unraveling the paradoxes of plant hormone signaling integration. *Nat. Struct. Mol. Biol.* *17*, 642–645.
- Jennings, B.H., Pickles, L.M., Wainwright, S.M., Roe, S.M., Pearl, L.H., and Ish-Horowitz, D. (2006). Molecular Recognition of Transcriptional Repressor Motifs by the WD Domain of the Groucho/TLE Corepressor. *Mol. Cell* *22*, 645–655.
- Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., Meng, X., Liu, G., Yu, H., Yuan, Y., et al. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* *504*, 401–405.
- Jurado, S., Abraham, Z., Manzano, C., López-Torrejón, G., Pacios, L.F., and Del Pozo, J.C. (2010). The *Arabidopsis* Cell Cycle F-Box Protein SKP2A Binds to Auxin. *Plant Cell* *22*, 3891–3904.
- Kagale, S., Links, M.G., and Rozwadowski, K. (2010). Genome-Wide Analysis of Ethylene-Responsive Element Binding Factor-Associated Amphiphilic Repression Motif-Containing Transcriptional Regulators in Arabidopsis. *Plant Physiol.* *152*, 1109–1134.
- Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* *27*, 470–478.
- Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J.L., Nishihama, R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. *PLoS Genet.* *11*, 1–26.
- Katsani, K.R., Hajibagheri, M.A.N., and Verrijzer, C.P. (1999). Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J.* *18*, 698–708.
- Ke, J., Ma, H., Gu, X., Thelen, A., Brunzelle, J.S., and Li, J. (2015). Structural basis for recognition of diverse transcriptional repressors by the TOPLESS family of corepressors. *Sci. Adv.* *1*, 1–12.
- Keilwagen, J., Grau, J., Paponov, I.A., Posch, S., Strickert, M., and Grosse, I. (2011). De-Novo Discovery of Differentially Abundant Transcription Factor Binding Sites Including Their Positional Preference. *PLoS Comput. Biol.* *7*, e1001070.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* *10*, 845–858.
- Kieffer, M., Stern, Y., Cook, H., Clerici, E., Maulbetsch, C., Laux, T., and Davies, B. (2006). Analysis of the transcription factor WUSCHEL and its functional homologue in Antirrhinum reveals a potential mechanism for their roles in meristem maintenance. *Plant Cell* *18*, 560–573.
- Kim, S., Brostromer, E., Xing, D., Jin, J., Chong, S., Ge, H., Wang, S., Gu, C., Yang, L., Gao, Y.Q., et al. (2013). Probing Allostery Through DNA. *Science* *339*, 816–819.
- Korasick, D.A., Enders, T.A., and Strader, L.C. (2013). Auxin biosynthesis and storage forms. *J. Exp. Bot.* *64*, 2541–2555.
- Korasick, D.A., Westfall, C.S., Goo, S., Nanao, M.H., Dumas, R., and Hagen, G. (2014). Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 5427–5432.
- 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*, 4180–4190.
- Kumar, V., Sah, S.K., Khare, T., Shriram, V., and Wani, S.H. (2016). Engineering phytohormones for abiotic stress tolerance in crop plants. *Plant Horm. under Challenging Environ. Factors* *4*, 247–266.
- Lau, S., Shao, N., Bock, R., Jürgens, G., and De Smet, I. (2009). Auxin signaling in algal lineages: fact or myth? *Trends Plant Sci.* *14*, 1360–1385.
- Lavy, M., and Estelle, M. (2016). Mechanisms of auxin signaling. *Development* *143*, 3226–3229.
- Lavy, M., Prigge, M.J., Tao, S., Shain, S., Kuo, A., Kirchsteiger, K., and Estelle, M. (2016). Constitutive auxin response in *Physcomitrella* reveals complex interactions between Aux/IAA and ARF proteins. *eLife* *5*, e13325.
- Lee, J.E., and Golz, J.F. (2012). Diverse roles of Groucho/Tup1 co-repressors in plant growth and development. *Plant Signal. Behav.* *7*, 86–92.
- Lehti-Shiu, M.D., Panchy, N., Wang, P., Uygun, S., and Shiu, S.H. (2017). Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. *Biochim. Biophys. Acta - Gene Regul. Mech.* *1860*, 3–20.
- Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., and De Clerck, O. (2012). Phylogeny and Molecular Evolution of the Green Algae. *CRC Crit. Rev. Plant Sci.* *31*, 1–46.
- Lewis, L.A., and McCourt, R.M. (2004). Green algae and the origin of land plants. *Am. J. Bot.* *91*, 1535–1556.

- Leyser, H.M.O. (1998). Plant hormones. *Curr. Biol.* 8, R5–R7.
- Leyser, O. (2017). Auxin Signaling. *Plant Physiol.* Published August 2017.
- Li, J., Bush, J., Xiong, Y., Li, L., and McCormack, M. (2011). Large-Scale Protein-Protein Interaction Analysis in Arabidopsis Mesophyll Protoplasts by Split Firefly Luciferase Complementation. *PLoS One* 6.
- Liao, C., Smet, W., Brunoud, G., Yoshida, S., Vernoux, T., and Weijers, D. (2015). Reporters for sensitive and quantitative measurement of auxin response. *Nat. Methods* 12, 207–210.
- Liu, Z., and Karmarkar, V. (2008). Groucho/Tup1 family co-repressors in plant development. *Trends Plant Sci.* 13, 137–144.
- Liu, X., Yang, S., Zhao, M., Luo, M., Yu, C., Chen, C., and Tai, R. (2014). Transcriptional Repression by Histone Deacetylases in Plants. *Mol. Plant* 7, 764–772.
- Liu, Z.B., Ulmasov, T., Shi, X., Hagen, G., and Guilfoyle, T.J. (1994). Soybean GH3 promoter contains multiple auxin-inducible elements. *Plant Cell* 6, 645–657.
- Ljung, K. (2013). Auxin metabolism and homeostasis during plant development. *Development* 140, 943–950.
- Lokerse, A.S., and Weijers, D. (2009). Auxin enters the matrix — assembly of response machineries for specific outputs. *Curr. Opin. Plant Biol.* 12, 520–526.
- Long, J.A., Ohno, C., Smith, Z.R., and Meyerowitz, E.M. (2006). TOPLESS Regulates Apical Embryonic Fate in Arabidopsis. *Science* 312, 1520–1523.
- Long, J. A., Woody, S., Poethig, S., Meyerowitz, E.M., and Barton, M.K. (2002). Transformation of shoots into roots in Arabidopsis embryos mutant at the TOPLESS locus. *Development* 129, 2797–2806.
- Lu, Y., and Xu J. (2015). Phytohormones in microalgae : a new opportunity for microalgal biotechnology ? *Trends Plant Sci.* 20, 273–282.
- Luscombe, N.M., Laskowski, R. a, and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* 29, 2860–2874.
- Ma, H., Duan, J., Ke, J., He, Y., Gu, X., Xu, T.-H., Yu, H., Wang, Y., Brunzelle, J.S., Jiang, Y., et al. (2017). A D53 repression motif induces oligomerization of TOPLESS corepressors and promotes assembly of a corepressor-nucleosome complex. *Sci. Adv.* 3, 1–16.
- Marianayagam, N.J., Sunde, M., and Matthews, J.M. (2004). The power of two: Protein dimerization in biology. *Trends Biochem. Sci.* 29, 618–625.
- Martin-Arevalillo, R., Nanao, M.H., Larrieu, A., Vinos-Poyo, T., Mast, D., Galvan-Ampudia, C., Brunoud, G., Vernoux, T., Dumas, R., and Parcy, F. (2017). Structure of the Arabidopsis TOPLESS corepressor provides insight into the evolution of transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* 114, 8107–8112.
- Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., et al. (2014). Data access for the 1,000 Plants (1KP) project. *GIGA Sci.* 3, 1–10.
- Matías-Hernández, L., Aguilar-Jaramillo, A.E., Marín-González, E., Suárez-López, P., and Pelaz, S. (2014). RAV genes: Regulation of floral induction and beyond. *Ann. Bot.* 114, 1459–1470.
- Matsumura, H., Kusaka, N., Nakamura, T., Tanaka, N., Sagegami, K., Uegaki, K., Inoue, T., and Mukai, Y. (2012). Crystal structure of the N-terminal domain of the yeast general corepressor Tup1p and its functional implications. *J. Biol. Chem.* 287, 26528–26538.
- Mironova, V. V., Omelyanchuk, N.A., Wiebe, D.S., and Levitsky, V.G. (2014). Computational analysis of auxin responsive elements in the Arabidopsis thaliana L . genome. *BMC Genomics* 15, S4.
- Morgunova, E., and Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* 47, 1–8.
- Mottis, A., Mouchiroud, L., and Auwerx, J. (2013). Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev.* 27, 819–835.
- Mun, J.-H., Yu, H.-J., Shin, J.Y., Oh, M., Hwang, H.-J., and Chung, H. (2012). Auxin response factor gene family in Brassica rapa: genomic organization, divergence, expression, and evolution. *Mol. Genet. Genomics* 287, 765–784.
- Nanao, M.H., Vinos-Poyo, T., Brunoud, G., Thévenon, E., Mazzoleni, M., Mast, D., Lainé, S., Wang, S., Hagen, G., Li, H., et al. (2014). Structural basis for oligomerization of auxin transcriptional regulators. *Nat. Commun.* 5, 3617.
- O'Malley R.C., Carol Huang, S., Song, L., Galli, M., Gallavotti, A., Ecker Correspondence, J.R., O, R.C., Lewsey, M.G., Bartlett, A., Nery, J.R., et al. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape In Brief A new method for pinpointing transcription factor binding sites in the Arabidopsis genome and their responsiveness to DNA methylation demonstrates the impact of tissue- specific. *Cell* 165, 1280–1292.
- Oberoi, J., Fairall, L., Watson, P.J., Yang, J.-C., Czimmerer, Z., Kampmann, T., Goult, B.T., Greenwood, J. a, Gooch, J.T., Kallenberger, B.C., et al. (2011). Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. *Nat. Struct. Mol. Biol.* 18, 177–184.

- Oh, E., Zhu, J.-Y., Ryu, H., Hwang, I., and Wang, Z.-Y. (2014a). TOPLESS mediates brassinosteroid-induced transcriptional repression through interaction with BZR1. *Nat. Commun.* *5*, 4140.
- Oh, E., Zhu, J., Bai, M., Arenhart, R.A., Sun, Y., and Wang, Z. (2014b). Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. *eLife* *3*, e03031.
- Ohta, M. (2001). Repression Domains of Class II ERF Transcriptional Repressors Share an Essential Motif for Active Repression. *Plant Cell Online* *13*, 1959–1968.
- Ohtaka, K., Hori, K., Kanno, Y., Seo, M., and Ohta, H. (2017). Primitive Auxin Response without TIR1 and Aux/IAA in the Charophyte Alga *Klebsormidium nitens*. *Plant Physiol.* pp.00274.2017.
- Okushima, Y., Overvoorde, P.J., Arima, K., Alonso, J.M., Chan, A., Chang, C., Ecker, J.R., Hughes, B., Lui, A., Nguyen, D., et al. (2005). Functional Genomic Analysis of the AUXIN RESPONSE FACTOR Gene Family Members in Arabidopsis thaliana. *Plant Cell* *17*, 444–463.
- Ottenschläger, I., Wolff, P., Wolverton, C., Bhalerao, R.P., Sandberg, G., Ishikawa, H., Evans, M., and Palme, K. (2003). Gravity-regulated differential auxin transport from columella to lateral root cap cells. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 2987–2991.
- Paponov, I. a, Teale, W., Lang, D., Paponov, M., Reski, R., Rensing, S. a, and Palme, K. (2009). The evolution of nuclear auxin signalling. *BMC Evol. Biol.* *9*, 126.
- Paque, S., and Weijers, D. (2016). Q&A: Auxin: the plant molecule that influences almost anything. *BMC Biol.* *14*, 67.
- Parcy, F., Vernoux, T., and Dumas, R. (2016). A Glimpse beyond Structures in Auxin-Dependent Transcription. *Trends Plant Sci.* *21*, 574–583.
- Parkhurst, S.M. (1998). Groucho: making its Marx as a transcriptional co-repressor. *Trends Genet.* *14*, 130–132.
- Paul, P., Dhandapani, V., Rameneni, J.J., Li, X., Sivanandhan, G., Choi, S.R., Pang, W., Im, S., and Lim, Y.P. (2016). Genome-Wide Analysis and Characterization of Aux/IAA Family Genes in Brassica rapa. *PLoS One* *11*.
- Pauwels, L., Barbero, G.F., Geerinck, J., Tilleman, S., Grunewald, W., Pérez, A.C., Chico, J.M., Bossche, R. Vanden, Sewell, J., Gil, E., et al. (2010). NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature* *464*, 788–791.
- Payankulam, S., Li, L.M., and Arnosti, D.N. (2010). Transcriptional repression: Conserved and evolved features. *Curr. Biol.* *20*, 764–771.
- Peer, W.A. (2013). From perception to attenuation: Auxin signalling and responses. *Curr. Opin. Plant Biol.* *16*, 561–568.
- Perissi, V., Jepsen, K., Glass, C.K., and Rosenfeld, M.G. (2010). Deconstructing repression: evolving models of co-repressor action. *Nat. Rev. Genet.* *11*, 109–123.
- Pi, L., Aichinger, E., van der Graaff, E., Llavata-Peris, C.I., Weijers, D., Hennig, L., Groot, E., and Laux, T. (2015). Organizer-Derived WOX5 Signal Maintains Root Columella Stem Cells through Chromatin-Mediated Repression of CDF4 Expression. *Dev. Cell* *33*, 576–588.
- Pierre-jerome, E., Moss, B.L., Lanctot, A., Hageman, A., and Nemhauser, J.L. (2016). Functional analysis of molecular interactions in synthetic auxin response circuits. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 11354–11359.
- Pierre-Jerome, E., Moss, B.L., and Nemhauser, J.L. (2013). Tuning the auxin transcriptional response. *J. Exp. Bot.* *64*, 2557–2563.
- Pierre-Jerome, E., Jang, S.S., Havens, K. a, Nemhauser, J.L., and Klavins, E. (2014). Recapitulation of the forward nuclear auxin response pathway in yeast. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 9407–9412.
- Piya, S., Shrestha, S.K., Binder, B., Jr, C.N.S., and Hewezi, T. (2014). Protein-protein interaction and gene co-expression maps of ARFs and Aux / IAAs in Arabidopsis. *Frontiers in Plant Sci.* *5*, 744.
- Powers, S.K., and Strader, L.C. (2016). Up in the air: Untethered Factors of Auxin Response. *F1000Research* *5*.
- Prigge, M.J., Lavy, M., Ashton, N.W., and Estelle, M. (2010). Physcomitrella patens auxin-resistant mutants affect conserved elements of an auxin-signaling pathway. *Curr. Biol.* *20*, 1907–1912.
- Rademacher, E.H., and Mo, B. (2011). A cellular expression map of the Arabidopsis AUXIN RESPONSE FACTOR gene family. *Plant J.* *68*, 597–606.
- Rigal, A., Ma, Q., and Robert, S. (2014). Unraveling plant hormone signaling through the use of small molecules. *Front Plant Sci* *5*, 373.
- Romanel, E.A.C., Schrago, C.G., Couñago, R.M., Russo, C.A.M., and Alves-Ferreira, M. (2009). Evolution of the B3 DNA binding superfamily: New insights into REM family gene diversification. *PLoS One* *4*.
- Roosjen, M., Paque, S., and Weijers, D. (2017). Auxin Response Factors – output control in auxin biology. *bioRxiv* 129122.
- Ruiz Rosquete, M., Barbez, E., and Kleine-Vehn, J. (2012). Cellular auxin homeostasis: Gatekeeping is housekeeping. *Mol. Plant* *5*, 772–786.
- Ryu, H., Cho, H., Bae, W., and Hwang, I. (2014). Control of early seedling development by BES1/TPL/HDA19-mediated epigenetic regulation of ABI3. *Nat. Commun.* *5*, 4138.

- Salehin, M., Bagchi, R., and Estelle, M. (2015). SCF TIR1 / AFB -Based Auxin Perception : Mechanism and Role in Plant Growth and Development. *27*, 9–19.
- Sauer, M., Robert, S., and Kleine-Vehn, J. (2013). Auxin: Simply complicated. *J. Exp. Bot.* *64*, 2565–2577.
- Sayou, C., Nanao, M.H., Jamin, M., Pose, D., Thevenon, E., Gregoire, L., Tichtinsky, G., Grégoire, D., Ott, F., Peirats Llobet, M., et al. (2016). genomic binding landscape of the LEAFY transcription factor. *Nat. Commun.* *7*, 11222.
- Sekiya, T., and Zaret, K.S. (2007). Repression by Groucho/TLE/Grg Proteins: Genomic Site Recruitment Generates Compacted Chromatin In Vitro and Impairs Activator Binding In Vivo. *Mol. Cell* *28*, 291–303.
- Sheard, L.B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T.R., Kobayashi, Y., Hsu, F.-F., Sharon, M., Browse, J., et al. (2010). Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* *468*, 400–405.
- Simonini, S., Deb, J., Moubayidin, L., Stephenson, P., Valluru, M., Freire-rios, A., Sorefan, K., Weijers, D., and Østergaard, L. (2016). A noncanonical auxin-sensing mechanism is required for organ morphogenesis in Arabidopsis. *30*, 2286–2296.
- De Smet, I., Voss, U., Lau, S., Wilson, M., Shao, N., Timme, R.E., Swarup, R., Kerr, I., Hodgman, C., Bock, R., et al. (2011). Unraveling the evolution of auxin signaling. *Plant Physiol.* *155*, 209–221.
- Smith, N.C., and Matthews, J.M. (2016). Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Curr. Opin. Struct. Biol.* *38*, 68–74.
- Sridhar, V. V, Surendrarao, A., Gonzalez, D., Conlan, R.S., and Liu, Z. (2004). Transcriptional repression of target genes by LEUNIG and SEUSS, two interacting regulatory proteins for Arabidopsis flower development. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 11494–11499.
- Sridhar, V. V, Surendrarao, A., and Liu, Z. (2006). APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. *Development* *133*, 3159–3166.
- Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. *Trends Plant Sci.* *13*, 647–655.
- Szemenyei, H., Hannon, M., and Long, J. A. (2008). TOPLESS Mediates Auxin-Dependent Transcriptional Repression During Arabidopsis Embryogenesis. *Science* *319*, 1384–1386.
- Tan, X., Calderon-Villalobos, L.I.A., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M., and Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* *446*, 640–645.
- Tao, Q., Guo, D., Wei, B., Zhang, F., Pang, C., Jiang, H., Zhang, J., Wei, T., Gu, H., Qu, L.J., et al. (2013a). The TIE1 transcriptional repressor links TCP transcription factors with TOPLESS/TOPLESS-RELATED corepressors and modulates leaf development in Arabidopsis. *Plant Cell* *25*, 421–437.
- Timme, R.E., and Delwiche, C.F. (2010). Uncovering the evolutionary origin of plant molecular processes: comparison of Coleochaete (Coleochaetales) and Spirogyra (Zygnematales) transcriptomes. *BMC Plant Biol* *10*, 96.
- Tiwari, S.B. (2004). Aux/IAA Proteins Contain a Potent Transcriptional Repression Domain. *Plant Cell* *16*, 533–543.
- Tiwari, S.B., Hagen, G., and Guilfoyle, T. (2003). The Roles of Auxin Response Factor Domains in Auxin-Responsive Transcription. *Plant Cell* *15*, 533–543.
- Trigg, S.A., Garza, R.M., Macwilliams, A., Nery, J.R., Bartlett, A., Castanon, R., Goubil, A., Feeney, J., Malley, R.O., Huang, S.C., et al. (2017). CrY2H-seq : a massively multiplexed assay for deep- coverage interactome mapping. *Nat. Methods* *14*, 819–829.
- Tomas, A., Paponov, I., and Perrot-Rechenmann, C. (2010). Auxin binding protein 1: Functional and evolutionary aspects. *Trends Plant Sci.* *15*, 436–446.
- Tsakagoshi, H., Morikami, A., and Nakamura, K. (2007). Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 2543–2547.
- Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1997a). ARF1 , a Transcription Factor That Binds to Auxin Response Elements. *Science* *276*, 1865–1868.
- Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T.J. (1997b). Aux/IAA Proteins Repress Expression of Reporter Genes Containing Natural and Highly Active Synthetic Auxin Response Elements. *Plant Cell* *9*, 1963–1971.
- Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1999). Dimerization and DNA binding of auxin response factors. *Plant J.* *19*, 309–319.
- Umen, J.G. (2014b). Green Algae and the Origins of Multicellularity in the Plant Kingdom. *Cold Spring Harb. Perspect. Biol.* *6*, 1–29.
- Vernoux, T., Kronenberger, J., Grandjean, O., Laufs, P., and Traas, J. (2000). PIN-FORMED 1 regulates cell fate at the periphery of the shoot apical meristem. *Development* *127*, 5157–5165.
- Vert, G., Walcher, C.L., Chory, J., and Nemhauser, J.L. (2008). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 9829–9834.
- Wang, R., and Estelle, M. (2014). Diversity and specificity: Auxin perception and signaling through the TIR1/AFB pathway. *Curr. Opin. Plant Biol.* *21*, 51–58.
- Wang, C., Liu, Y., Li, S.-S., and Han, G.-Z. (2015). Insights into the Origin and Evolution of the Plant Hormone Signaling Machinery.

Plant Physiol. *167*, 872–886.

Wang, L., Kim, J., and Somers, D.E. (2013). Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. *Proc. Natl. Acad. Sci.* *110*, 761–766.

Wang, Y., Deng, D., Bian, Y., Lv, Y., and Xie, Q. (2010). Genome-wide analysis of primary auxin-responsive Aux/IAA gene family in maize (*Zea mays* L.). *Mol. Biol. Rep.* *37*, 3991–4001.

Wang, Y., Deng, D., Zhang, R., Wang, S., Bian, Y., and Yin, Z. (2012). Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. *Mol. Biol. Rep.* *39*, 6267–6282.

Watson, P.J., Fairall, L., and Schwabe, J.W.R. (2012). Molecular and Cellular Endocrinology Nuclear hormone receptor co-repressors: Structure and function. *Mol. Cell. Endocrinol.* *348*, 440–449.

Wei, B., Zhang, J., Pang, C., Yu, H., Guo, D., Jiang, H., Ding, M., Chen, Z., Tao, Q., Gu, H., et al. (2015). The molecular mechanism of sporocytelless/nozzle in controlling Arabidopsis ovule development. *Cell Res.* *25*, 121–134.

Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A.J., Philippe, H., Melkonian, M., and Becker, B. (2011). Origin of land plants: do conjugating green algae hold the key? *BMC Evol. Biol.* *11*, 104.

Woo, E.J., Marshall, J., Baully, J., Chen, J.G., Venis, M., Napier, R.M., and Pickersgill, R.W. (2002). Crystal structure of auxin-binding protein 1 in complex with auxin. *EMBO J.* *21*, 2877–2885.

Wu, M.F., Yamaguchi, N., Xiao, J., Bargmann, B., Estelle, M., Sang, Y., and Wagner, D. (2015b). Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *eLife* *4*, e09269.

Xiao, J., Jin, R., and Wagner, D. (2017). Developmental transitions: integrating environmental cues with hormonal signaling in the chromatin landscape in plants. *Genome Biol.* *18*, 88.

Yamasaki, K. (2004). Solution Structure of the B3 DNA Binding Domain of the Arabidopsis Cold-Responsive Transcription Factor RAV1. *Plant Cell Online* *16*, 3448–3459.

Yamasaki, K., Kigawa, T., Inoue, M., Watanabe, S., Tateno, M., Seki, M., Shinozaki, K., and Yokoyama, S. (2008). Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol. Biochem.* *46*, 394–401.

Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. *Trends Plant Sci.* *18*, 267–276.

Yoon, H.-G., Choi, Y., Cole, P. a, and Wong, J. (2005). Reading and function of a histone code involved in targeting corepressor complexes for repression. *Mol. Cell. Biol.* *25*, 324–335.

Yoon, H.G., Chan, D.W., Huang, Z.Q., Li, J., Fondell, J.D., Qin, J., and Wong, J. (2003). Purification and functional characterization of the human N-CoR complex: The roles of HDAC3, TBL1 and TBLR1. *EMBO J.* *22*, 1336–1346.

Yue, J., Hu, X., and Huang, J. (2014). Origin of plant auxin biosynthesis. *Trends Plant Sci.* *19*, 764–770.

Zaret Kenneth S.; Mango, S. (2017). Pioneer Transcription Factors, Chromatin Dynamics, and Cell Fate Control. *Curr. Opin. Genet. Dev.* *37*, 76–81.

Zemlyanskaya, E. V., Wiebe, D.S., Omelyanchuk, N.A., Levitsky, V.G., and Mironova, V. V. (2016). Meta-analysis of transcriptome data identified TGTCNN motif variants associated with the response to plant hormone auxin in *Arabidopsis thaliana* L. *J. Bioinform. Comput. Biol.* *14*, 1–16.

Zhang, F., Wang, Y., Li, G., Tang, Y., Kramer, E.M., and Tadege, M. (2014). STENOFOLIA recruits TOPLESS to repress ASYMMETRIC LEAVES2 at the leaf margin and promote leaf blade outgrowth in *Medicago truncatula*. *Plant Cell* *26*, 650–664.

Zhao, Z., Andersen, S.U., Ljung, K., Dolezal, K., Miotk, A., Schultheiss, S.J., and Lohmann, J.U. (2010). Hormonal control of the shoot stem-cell niche. *Nature* *465*, 1089–1092.

Zhu, Z., Xu, F., Zhang, Y., Cheng, Y.T., Wiermer, M., Li, X., and Zhang, Y. (2010). Arabidopsis resistance protein SNC1 activates immune responses through association with a transcriptional corepressor. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 13960–13965.

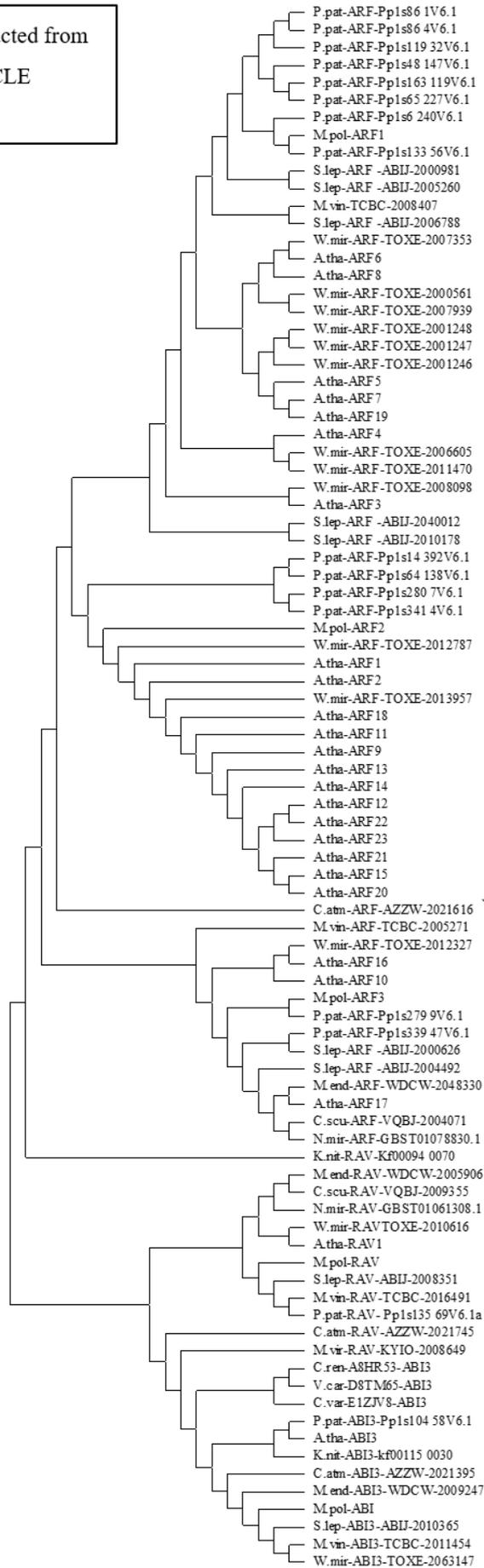


## Annexe

**Table 1. B3-containing proteins from charophyte organisms.** Indicated the lineage, class, genera and specie of the organisms analysed. Accession numbers for transcripts and the databases used for each search are indicated. Proteins were obtained by transcripts translation except for *K.nitens* proteins that were available in the PlantTFDB and chlorophyte proteins (Uniprot). The domains present in each of the predicted or described proteins are indicated with a tick sign. In green, accession numbers of the transcripts coding for the proteins that were extensively studied in this Chapter.

		Accession number	ABI3	RAV			ARF		Data base
				AP2	B3	PB1	B3	PB1	
Chlorophytes		<i>Chlamydomonas reinhardtii</i>	A8HR53	✓					Uniprot
		<i>Volvox carteri</i>	D8TM65	✓					Uniprot
		<i>Chlorella variabilis</i>	E1ZJV8	✓					Uniprot
Charophytes	Mesostigmatophyceae	<i>Mesostigma viride</i>	KYIO-2008649		✓	✓	✓		OneKp
	Chlorokybophyceae	<i>Chlorokybus atmophyticus</i>	AZZW-2021395	✓					OneKp
			AZZW-2021745		✓	✓	✓		
			<b>AZZW-2021616</b>				✓	✓	OneKp
	Klebsormidiophyceae	<i>Klebsormidium subtile</i>	FQLP-2005636	✓					OneKp
			FQLP-2031660	✓					OneKp
			FQLP-2009864		✓	✓	✓		
		<i>Klebsormidium nitens</i>	kfi00115_0030	✓					PlantTFDB
			kfi00073_0220	✓					PlantTFDB
			kfi00219_0030	✓					PlantTFDB
			kfi00085_0040	✓					PlantTFDB
			kfi00822_0050	✓					PlantTFDB
			<b>kfi00094_0070</b>		✓	✓	✓		PlantTFDB
	Charophyceae	<i>Nitella mirabilis</i>	GBST01081320.1	✓					Marchantia.info
			GBST01061308.1			✓			Marchantia.info
			GBST01078830.1					✓	✓
	Coleochaetophyceae	<i>Coleochaete scutata</i>	VQBJ-2009355		✓	✓	✓		OneKp
			VQBJ-2004071					✓	✓
	Zygnematophyceae	<i>Mougeotia</i>	ZRMT-2007919		✓	✓	✓		OneKp
		<i>Mesotaenium endlicherianum</i>	WDCW-2009247	✓					OneKp
			WDCW-2005906		✓	✓			OneKp
			WDCW-2048330				✓	✓	OneKp

Maximum Likelihood tree constructed from  
B3-sequences aligned with MUSCLE  
(MEGA software)



A

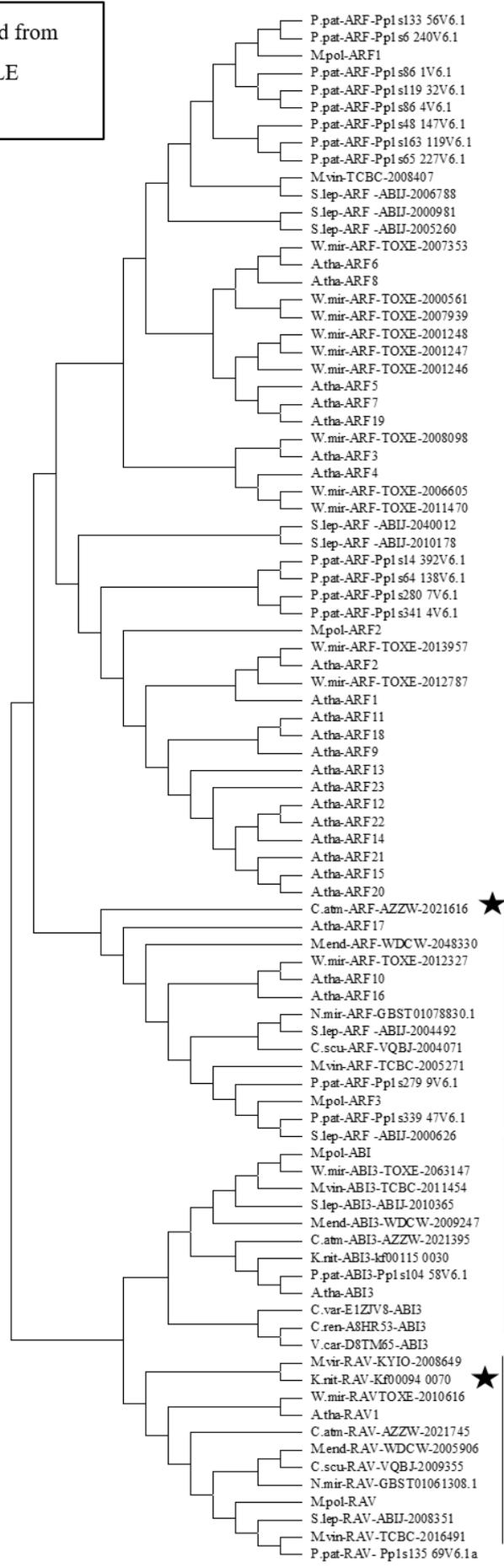
B

C

RAV

ABI

Minimum evolution tree constructed from  
B3-sequences aligned with MUSCLE  
(MEGA software)



A

B

C

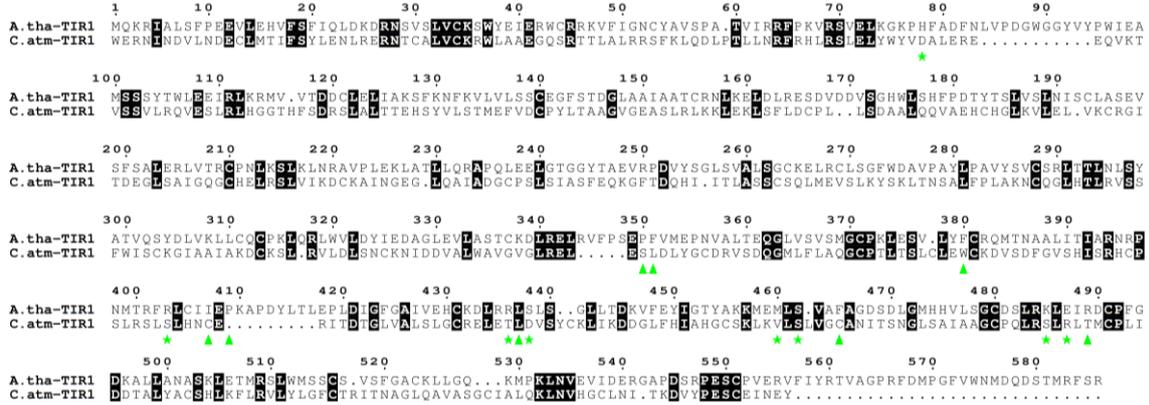
ABI

RAV

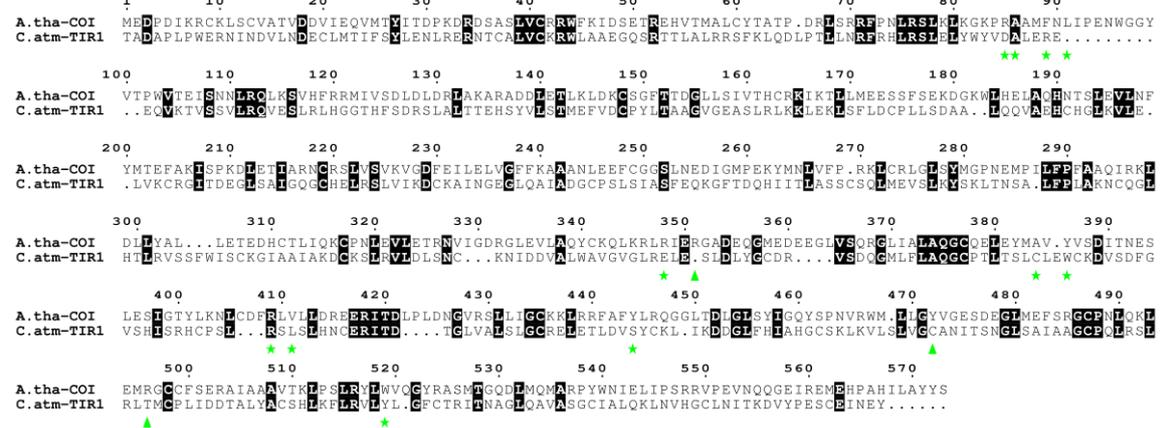
**Table 2. Auxin signalling components in charophyte organisms.** Indicated the lineage, class, genera and specie of the organisms analysed. Accession numbers for transcripts and the databases used for each search are indicated.

	ARF	AUX/IAA	TIR1/AFB (?)	TPL	Data base
Mesostigmatophyceae	<i>Mesostigma viride</i>		KYIO-2005024		OneKp
Chlorokybophyceae	<i>Chlorokybus atmophyticus</i>	AZZW-2021616	AZZW-2021242	AZZW-2021890	OneKp
Klebsiomiophyceae	<i>Klebsormidium nitens</i>			Kfi00881_0020	plantmorphogenesis
Charophyceae	<i>Nitella mirabilis</i>	GBST01078830.1	GBST01045649.1	GBST01062325.1	Marchantia.info
Coleochaetophyceae	<i>Coleochaete scutata</i>	VQBJ-2004071	VQBJ-2011792	VQBJ-2012007	OneKp
Zygnematophyceae	<i>Mesotanium endlicherianum</i>	WDCW-2048330	WDCW-2048136	WDCW-2045609	OneKp

**A**



**B**



**Fig 1. CaTIR1 homologue residues comparison.** **A.** CaTIR1 homologue aligned to AtTIR1 protein sequence. Green stars indicate residues implicated in the interaction with IAA. Green triangles indicate residues implicated in the interaction with Aux/IAA protein. **B.** CaTIR1 homologue aligned to AtCOI1 protein sequence. Green stars indicate residues implicated in the interaction with JA-Ile conjugate. Green triangles indicate residues implicated in the interaction with JAZ proteins. In spite of being modelled as AtTIR1 or AtCOI1 proteins structures, CaTIR1 homologue has very low homology in the amino acidic sequence with any of the two proteins. The residues implicated in the interactions with the phytohormones or with Aux/IAA or JAZ proteins are not conserved. Thus, CaTIR1 homologue could not present the

same mechanism of hormone perception and signal transduction as current TIR1 and COI1 proteins. Same studies were done for the rest of the TIR1 homologues found in charophytes, that led to the same result.

## B3 phylogeny sequences

>C.ren-A8HR53-ABI3

FEKSLTASDVS GGGRV VVPKSIAEQYFPRLEAPSGVTISAADLEGRAYTFKWRFWVNNSSRMYLLEGAGELHR  
NYGLEVDVMVFAQKQDGSLLVAGRCANK

>V.car-D8TM65-ABI3

FEKALTASDVS GGGRV VVPKSIAEQYFPKLEQPSGVTISATDLDGRSYTFKWRFWVNNSSRMYLLEGAGELHR  
NYGLEVDVMVFAQKADGSLMVAGRAASKG

>C.var-E1ZJV8-ABI3

FEKVLTS SDVNGTGR LVIPKSQAEAHFPFLEQQQGMVMSLTDTEGNQHSFRFRFWVNNQSRMYLLENTIEVQ  
AQYKMOVAGDVLVFAKLPDGTYAICGRKGTKD

>M.vir-RAV-KYIO-2008649

FEKIVTPSDVGKLNRLVIPKHAERFFPLDLKEAPVPAMTLEFTDDVGKTWSFRYNFWASSQSYVLTRGWSRF  
VKEKKLYPGDVMIFRRNPTSKQLYIGFKP

>C.atm-ABI3-AZZW-2021395

FEKQLTS SDTGKLGRIVLPKAHAEQHLPRIDTPDGRALQVTDNGRSWAPRFRFWPNNSSRMYLLEGITEILHS  
LHLQTGDSVTFNKDVVTGKLVIGAR

>C.atm-RAV-AZZW-2021745

FEKAVTPSDVGKLNRLVIPKLHAERCFPLDLSVETPAKMLAFLDENNKTNWFRYSYWNSSQSYVLTKGWSRF  
VKEKQLTAGDII VFKRTPDNKMHISYKR

>C.atm-ARF-AZZW-2021616

IAKVLTPSDVSTCGGFSVPRTIADTCLPPLDYSEHPPAQAITARDVHNEEWPFRIHYRGTPKRHLFTTGWGAF  
TAKKLVAGDAIIFVRMRDGSLLRGLGIR

>K.nit-ABI3-kf00115\_0030

FEKALTISDTNPLGRIVMPKVQAEHGLPRMAGKESCTLTVTDADGKTWQLRYSVWLNRRSRMYVLEHAGDF  
LQSRGLAPGDLLAFYRADDRRLIILDKK

>K.nit-RAV-Kf00094\_0070

FEKALTSPSDVGKLNRLVFPKHYAERFFPLDLDQVSVGQTLQFEDERGKFWFRYSYWNSSQSYVLTKGWSRF  
VKEKGLLPGDNVVFEKGHTGQLYIGYAR

>M.end-ABI3-WDCW-2009247

FEKMLSQSDAGR VGR LVVPKAAEAHLPNLHQPEGVPLNVVDMAGRAWHFQYRFWPNNSSRMYVLEGMT  
QCIQTLKLQAGDTLAFGRLDGGSQLVVDCKRA

>M.end-RAV-WDCW-2005906

FEKAVTPSDVGKLNRLVIPKQHAERCFPLDLTSSTPAQTLSEFEDES GKHWRFRYSYWNSSQSYVLTKGWSRFV  
KEKLLVPGDIVFFNRGPAGELYIGFR

>M.end-ARF-WDCW-2048330

FAKTLTQSDANNGGFSVPRYCAETIFPPLDYAVEPPS QSLVARDVHGELWKFRHIYRGTPRRHLLTTGWSTF  
VNAKRLAPGDVVFLRSASGELCVGVRR

>C.scu-ARF-VQBJ-2004071

FAKTLTQSDANNGGFSVPRYCAETIFPALDYTNEPPLQTLTATDIHGKVVHFRHIYRGTPRRHLLTTGWSTFV  
NAKKLVAGDAIVFLRSGDEQLCVGVRR

>C.scu-RAV-VQBJ-2009355

FEKAVTPSDVGKLNRLVIPKQHAERCFPLDLSSGSVAQTLHFEDEGGKHWRFRYSYWNSSQSYVLTKGWSRF  
VKEKLLVPGDIVYFDQGTSQELYI

>N.mir-ARF-GBST01078830.1

FAKTLTQSDANNGGFSVPRYCAETIFPLDYSLDPPVQTLVAKDVHGETWKFRHIYRGTPRRHLLTTGWSTF  
VNAKKLVAGDAIVFLRSVSGDLCVGVRR

>N.mir-RAV-GBST01061308.1

FEKAVTPSDVGKLNRLVIPKQHAEKWFPLEVGSGSPAAQTLSEFEDEAGKHWRFRYSYWNSSQSYVLTKGWSR  
FVKEKLLVPGDIVFFHRGQAAELYIGFRR

>M.vin-ABI3-TCBC-2011454

FEKVLSA SDAGRIGRLVLPKACAEAYFPQISQPEGVPMKIQDVTGKVVWGFQFRFWPNNNSRMYVLEG  
VTPCIQLQLQAGDVTFSRMEPGGRLVMGYRR

>M.vin-RAV-TCBC-2016491

FDKAVTPSDVGKLNRLVIPKQHAERCFPLDLSTNAPGQTLAFEDAAGKQWRFRYSYWNSSQSYVLTKGWSRF  
VKEKLDAGDIVSFERGGNQELYIDFRR

>M.vin-ARF-TCBC-2005271

FAKTLTQSDANNGGFSVPRYCAETIFPLDYSVDPPVQTVLAKDVHGETWKFRHIYRGTPRRHLLTTGWSAF  
VNQKLVAGDAIVFLRSAGGEAPGDLCVGVRR

>M.vin-TCBC-2008407

FCKTLTASDTSTH GGFSIPRKSAAEKVFPPLDYHQAPPAQELVARDLHDQEWFRHIYRGQPRRHLLTTGWSVF  
VSAKRLQAGDAVLFIRDEKQQLLIGIRR

>M.pol-ABI

FEKMLSASDAGRIGRLVLPKACAEAYFPAISQPEGLPMKIQDITGKEWLFQFRFWPNNNSRMYVLEGVTPCIQS  
MQLLAGDVTFSRLDPEGKLVVMGYRR

>M.pol-RAV

FEKAVTPSDVGKLNRLVIPKQHAERCFPLDLALNAPCQTLSEFEDVSGKHWRFRYSYWNSSQSYVFTKGWSRF  
VKEKLEAGDVTFSFERGPNQELYIDFRR

>M.pol-ARF1

FCKTLTASDTSTHGGFIPRRAAEKVFPPLDYSQPPAHPAQELVARDLHDQEWFRHIYRGQPRRHLLTTGWS  
VFSKRLQAGDSVLFIRDDKGQLLIGIRR

>M.pol-ARF2

FTKTLTVSDTSTHGGFSVPRRAADDCLPKLDMSLNPPNQELVAKDLHGNEWRFRHIFRGQPKRHLLTTGWSV  
FVSQKRLVAGDAVFLRGENGQLRVGVRR

>M.pol-ARF3

FAKTLTQSDANNGGFSVPRYCAETIFPLDYSIDPPVQTVLAKDVHGERWKFRHIYRGTPRRHLLTTGWSTF  
VNQKLVAGDAIVFLRTASGELCVGVRR

>P.pat-ABI3-Pp1s104\_58V6.1

LRKELTVTDVDELGRIVLPRKDAEYQLPRLEAKEGKLLTMEDYNSINKWTLRYKWWPNNKSRMYILENTAYF  
VKYYNLREKDEIIVYKDAQEKL VIRGKK

>P.pat-RAV- Pp1s135\_69V6.1a

FDKAVTPSDVGKLNRLVIPKQHAERCF  
PLDLSANSPGQTLSEFEDVSGKHWRFRYSYWNSSQSYVLTKGWSRFVKEKLDAGDIVSFERGRNHELIDFRR

>P.pat-ARF-Pp1s119\_32V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPTLDYQPPAQELVARDLHDQDWHFRHIYRGQPRRHLLTTGWSIFI  
SAKRLQAGDAVLFIRDDKGQLLIGIRR

>P.pat-ARF-Pp1s133\_56V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPLDFTRVPPAQELVARDLHDQEWFRHIYRGQPRRHLLTTGWSVF  
VSAKRLQAGDSVLFIRDDKGNLLIGIRR

>P.pat-ARF-Pp1s14\_392V6.1

FCKTLTASDTSTHGGFSVPRRAAEDCLPLLDHSMNPPCQELVAKDLHGKEWNFRHIYRGHPRRHLLTTGWSV  
FVSQKRLVAGDTVIFLRGENGQLRVGVRR

>P.pat-ARF-Pp1s163\_119V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPLDYSQTPPAQELKARDLHDQEWFRHIYRGQPRRHLLTTGWSVF  
VSAKRLQAGDAVLFIRDDKGQLLIGIRR

>P.pat-ARF-Pp1s279\_9V6.1

FAKTLTQSDANNGGFSIPRYCAETIFPLDYCIDPPVQTVLAKDVHGEVWVKFRHIYRGTPRRHLLTTGWSTFV  
NQKRLVAGDAIVFLRIASGELCVGVRR

>P.pat-ARF-Pp1s280\_7V6.1

FCKTLTASDTSTHGGFSVPRRSAEDCLPLLDYMNPPCQELVAKDLHGQEWKFRHIYRGYPRRHLLTTGWSTF  
VSAKRLVAGDTVIFLRGENGQLRVGVRR

>P.pat-ARF-Pp1s339\_47V6.1

FAKTLTQSDANNGGFSVPRYCAETIFPLDYSSDPPVQTVLAKDVHGDVWVKFRHIYRGTPRRHLLTTGWSTF  
VNQKRLVAGDAIVFLRSASGELCVGVRR

>P.pat-ARF-Pp1s341\_4V6.1

FCKTLTASDTSTHGGFSVPRRSAEDCLPSLDYANPPCQELVAKDLHGHEWKFRHIYRGHPRRHLLTTGWSAF  
VSAKRLVAGDTVIFLRGENGQLRVGVRR

>P.pat-ARF-Pp1s48\_147V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPLDYAQIPPAQELKARDLHDQEWFRHIYRGQPRRHLLTTGWSVF  
SAKRLQAGDAVLFTRDNKGQLLIGIRR

>P.pat-ARF-Pp1s64\_138V6.1

FCKTLTASDTSTHGGFSVPRRAAEECLPLLDHNMVPPCQELVAKDLHGKDWFRHIYRGHPRRHLLTTGWSV  
FVSQKRLVAGDTVIFLRGENGQLRVGVRR

>P.pat-ARF-Pp1s65\_227V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPLDYQTPPAQELKARDLHDQEWFRHIYRGQPRRHLLTTGWSVF  
VSAKRLQAGDAVLFIRDDKGQLLIGIRR

>P.pat-ARF-Pp1s6\_240V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPLDFTKSPPAQELVARDLHDQDWHFRHIYRGQPRRHLLTTGWSVF  
VSIKRLQAGDSVLFIRDDKDHLLLIGIRR

>P.pat-ARF-Pp1s86\_1V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPTLDYNQPPAQELVARDLHDQDWHFRHIYRGQPRRHLLTTGWSVF  
VSAKRLQAGDAVLFIRDDKGQLLIGIRR

>P.pat-ARF-Pp1s86\_4V6.1

FCKTLTASDTSTHGGFSIPRRAAEKVFPTLDWHFRHIYRGQPRRHLLTTGWSVFVSAKRLQAGDAVLFIRDDK  
GQLLIGIRR

>S.lep-ABI3-ABIJ-2010365

FEKTLTVSDAGRIGRLVLPKACAEAFFPPISSPEGIPIKMSDTKGQEWQFQFRFWPNNNSRMYVLEGVTPCVKA  
MQLQAGDVVTFSTRTPGGKMVMGYRR

>S.lep-RAV-ABIJ-2008351

FDKAVTPSDVGKLNRLVIPKQHAERCPLDLSANEKGLLLSFEDITGKVWVFRYSYWN  
SSQSYVLTKGWSRFVKEKLDAGDIVTFERPGQELYISWKRR

>S.lep-ARF -ABIJ-2000626

FAKTLTQSDANNGGFSVPRYCAET  
IFPRLDYSIDPPVQTVLAKDVHGEVWKFRIYRGTPRRHLLTTGWSTFVNQKLVAGDAI  
VFLRSASGELCVGVRR

>S.lep-ARF -ABIJ-2004492

FAKTLTQSDANNGGFSVPRYCAETIFPRL  
DYSLDPPVQKVLAKDVHGEIWKFRHIYRGTPRRHLLTTGWSTFVNHKKLVAGDAIVFLRSTSGELCVGVRR

>S.lep-ARF -ABIJ-2040012

FCKNLTSSDSTH GGFSVPRRAAEECLPPLDYQQSPPAQELVAKDLHGVEWKFRIYRGQPRRHLLTTGWSVF  
VSQKLVAGDAVFLRGGDNGELRIGVRR

>S.lep-ARF -ABIJ-2010178

FCKNLTSSDSTHGGFSVPRRAAEECFPRLDYQQTTPAQELIA  
KDLHGTEWKFRIYRGQPRRHLLTTGWSVFVSQKLVAGDAVLFVVRGDNGELRIGIR

>S.lep-ARF -ABIJ-2006788

FCKTLTASDSTHGGFSIPRRAAEKVFPPLDFTQPPAQEIVARDLHDTEWRFRIYRGQP  
RRHLLTTGWSVFSKRLQGTGDAVLFIRDEKQQLLIGIR

>S.lep-ARF -ABIJ-2000981

FSKTLTASDSTHGGFSIPRRAAEKVFPPLDFTKTPPAQELVARDLHNNEWHFRHIYRGQPRRHLLTTGWSV  
FVSAKRLQAGDTVLFMRDEQGHMLGIR

>S.lep-ARF -ABIJ-2005260

FCKTLTASDSTHGGFSIPRRAAEKVFPPLDFAKQPPAQELVAKDLHNQWTFRIYRGQPRRHLLTTGWS  
VFSKRLQAGDTVLFIFFTSKCFGEKRW

>W.mir-ABI3-TOXE-2063147

FEKILSASDAGRIGRLVLPKACAEAYFPPIQPEGLPLKIQDAKGEWVFQFRFWPNNNSRMVLEGVTP  
CIQSMQLQAGDVTFSRLGPEEKLVMGFRK

>W.mir-RAVTOXE-2010616

FEKAVTPSDVGKLNRLVIPKQHAEKHFPLDGPSTGKGVLLSLEDSSGGKTWVFRYSYWNSSQSYVLTKGWSRF  
VKDKRLVPGDVVSFHRSPSRHLFISFR

>W.mir-ARF-TOXE-2012327

FAKTLTQSDANNGGFSVPRYCAETIFPRLDYSVDPPVQNVLAKDVHGTV  
WKFRIYRGTPRRHLLTTGWSNFVNHKKLVAGDSIVFLRSHNGDLCVGVRR

>W.mir-ARF-TOXE-2006605

FCKTLTASDSTH GGFSVPRRAAEDCFPPLDYSQQRPSQELIAKDLHGQVWKFRIYRGQPRRHLLTTGWSVF  
VSHKGLVSGDAVFLRDENGQLRLGIR

>W.mir-ARF-TOXE-2000561

FCKTLTASDSTHGGFSIPRRAAEKVFPPLDFSQTPPVQELVARDLHDHEWKFRIYRGQPKRHLLTTGWSVF  
SAKRLVAGDSVLFIRNDKGQQLLIGIR

>W.mir-ARF-TOXE-2008098

FCKTLTASDSTHGGFSVPRRAAEDCFPPLDYIQRPSQELIAKDLHGKEWKF  
HIFRGQPRRHLLTTGWSAFVSFKKLTAGDAVFLRDDNGELRLGIR

>W.mir-ARF-TOXE-2007939

FCKTLTASDTSTHGGFSIPRRAAEKVFPPLDFTQQPPAQELVARDLHDTEWKFRHIYRGQPKRHLLTTGWSVF  
VSAKRLVAGDSVLFIRNDGQLLLGIRR

>W.mir-ARF-TOXE-2011470

FCKTLTASDTSTHGGFSVPRRAAEDCFPLDYIQQRPSQELIAKDLHGVEWKFRHIYRGQPRRHLLTTGWSVVF  
SHKGLVSGDAVLFIRGENGELRLGIRR

>W.mir-ARF-TOXE-2012787

FCKTLTASDTSTHGGFSVLKRHADECLPPLDMS  
LQPPSQDLVAKDLHGVEWFRHIYRGQPRRHLLTTGWSVVFVSQKRLVAGDAFIFLRGENG ELRVGVRR

>W.mir-ARF-TOXE-2001246

FCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYSAQPPAQELVARDLHDKVWTFRHIYRGQPKRHLLTTGWSVF  
VSSKRLVAGDSVLFIRDEKSQLLLGIRR

>W.mir-ARF-TOXE-2001248

FCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYSAQPPAQELVARDLHDKVWTFRHIYRGQPKRHLLTTGWSVF  
VSSKRLVAGDSVLFIRDEKSQLLLGIRR

>W.mir-ARF-TOXE-2001247

FCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYSAQPPAQELVARDLHDKVWTFRHIYRGQPKRHLLTTGWSVF  
VSSKRLVAGDSVLFIRDEKSQLLLGIRR

>W.mir-ARF-TOXE-2007353

FCKTLTASDTSTHGGFSVPRRAAEKVFPPLDFTQTPPAQELIARDLHDNEWKFRHIFRGQPKRHLLTTGWSVVF  
SAKRLVAGDSVLFIWNEKGQLLLGIRR

>W.mir-ARF-TOXE-2013957

FCKTLTASDTSTHGGFSVLKHADECLPALDMNQPPIQDLVARDLHGQRWTF  
RHIFRGQPKRHLLTTGWSVVFVSSKRLVAGDAFIFLRGENGDLRVGVRR

>A.tha-ABI3

LQKVLKQSDVGNLGRIVLPKKEAETHLPELEARDGISELAMEDIGTSRVWNMRYRFWPNNKSRMYLLENTGDF  
VKTNGLQEGDFIVIYSDVKLIRGVKV

>A.tha-RAV1

FEKAVTPSDVGKLNRLVIPKHAEKHFPLPSSNVSVKGVLLNFEDVNGKVWRFVRYSYWNSSQSYVLTGWSR  
FVKEKNLRAGDVVSFSRSNGDQQLYIGWKS

>A.tha-ARF1

FCKTLTASDTSTHGGFSVLRRHADDCLPPLDMSQPPWQELVATDLHNSEWHFRHIFRGQPRRHLLTTGWSV  
FVSSKRLVAGDAFIFLRGENEELRVGVRR

>A.tha-ARF2

FCKTLTASDTSTHGGFSVLRRHDECLPPLDMSRQPPTQELVAKDLHANEWFRHIFRGQPRRHLLQSGWSVF  
VSSKRLVAGDAFIFLRGENGELRVGVRR

>A.tha-ARF3

FCKTLTASDTSTHGGFSVPRRAAEDCFPLDYQQRPSQELLARDLHGLEWFRHIYRGQPRRHLLTTGWSAF  
VNKKLVSGDAVLFIRGDDGKRLRVRR

>A.tha-ARF4

FCKTLTASDTSTHGGFSVPRRAAEDCFAPLDYKQQRPSQELIAKDLHGVEWKFRHIYRGQPRRHLLTTGWSIF  
VSQKNLVSGDAVLFIRDEGGELRLGIRR

>A.tha-ARF5

FCKTLTASDTSTHGGFSVPRRAAEKLFPPLDYSAQPPTQELVVRDLHENTWTFRHIYRGQPKRHLLTTGWSLF  
VGSKRLRAGDSVLFIRDEKSQMLVGVRR

>A.tha-ARF6

FCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYSQQPPAQELMARDLHDNEWKFRHIFRGQPKRHLLTTGWSVF  
VSARKLVAGDSVLFIWNDKNQLLLGIRR

>A.tha-ARF7

FFCKTLTASDTSTHGGFSVPRRAAEKIFPALDFSMQPPCQELVAKDIHDNTWTFRHIYRGQPKRHLLTTGWSVF  
VSTKRLFAGDSVLFIRDGKAQLLLGIRR

>A.tha-ARF8

FCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYTLQPPAQELIARDLHDVEWKFRHIFRGQPKRHLLTTGWSVF  
VSARKLVAGDSVIFIRNEKNQLFLGIRR

>A.tha-ARF9

FSKVLTAASDTSTHGGFSVLRKHATECLPPLDMTQQTPTQELVAEDVHGQWKFHIFRGQPRRHLLTTGWSTF  
VTSKRLVAGDTFVFLRGENGELRVGVRR

>A.tha-ARF10

FAKTLTQSDANNGGFSVPRYCAETIFPRLDYSAEPPVQTVIKDIHGETWKFRHIYRGTPRRHLLTTGWSTFV  
NQKKLIAGDSIVFLRSESGDLCVGIRR

>A.tha-ARF11

FVKILTASDTSTHGGFSVLRKHATECLPSLDMTQPTPTQELVARDLHGQWRFKHIFRGQPRRHLLTTGWSTF  
VTSKRLVAGDAFVFLRGETGDLRVGVRR

>A.tha-ARF12

FTKVLTAASDTSAHGGFFVPPKHAIECLPSLDMSQPLPAQELLAIDLHGNQWRFNHNYRGTPQRHLLTTGWNAF  
TTSKKLVAAGDVIVFVRGETGELRVGIRR

>A.tha-ARF13

FSKILTASDVLSGGLIIPKQYAIECFPPLDMSQPSTQNLVAKDLYGQEWFSKHFVFRGTPQRHMFTSGGGWSV  
FATTKRLIVGDIFVLLRGENGELRFGIRR

>A.tha-ARF14

FTKVLTAASDTSVHGGFSVPPKHAIECLPPLDMSQPLPTQEILIDLHGNQWRFRHIYRGTAQRHLLTIGWNAFT  
TSKKLVEGDVIVFVRGETGELRVGIRR

>A.tha-ARF15

FTKVLTAASDISANGVFSVPPKHAIECLPPLDMSQPLPAQELLAIDLHGNQWSFRHSYRGTPQRHLLTTGWNEFT  
TSKKLKGDVIVFVRGETGELRVGIRR

>A.tha-ARF16

FAKTLTQSDANNGGFSVPRYCAETIFPRLDYNAEPPVQTLAKDVHGDVWKFRHIYRGTPRRHLLTTGWSNF  
VNQKKLVAGDSIVFMRAENGDLVCVGIRR

>A.tha-ARF17

FAKILTPSDANNGGFSVPRFCADSVFPLNFQIDPPVQKLYVTDIHGAVWDFRHIYRGTPRRHLLTTGWSKFV  
NSKKLIAGDSVFMRKSADEMFIVGIRR

>A.tha-ARF18

FVKILTASDTSTHGGFSVLRKHATECLPSLDMTQATPTQELVTRDLHGFEWRFKHIFRGQPRRHLLTTGWSTFV  
SSKRLVAGDAFVFLRGENGDLRVGVRR

>A.tha-ARF19

FCKTLTASDTSTHGGFSVPRRAAEKIFPPLDYSQQPPAQEIVAKDLHDTTWTFRHIYRGQPKRHLLTTGWSVFV  
STKRLFAGDSVLFVREKSQMLGIRR

>A.tha-ARF20

FTKVLTASDTSAYGGFFVPPKKHAIIECLPPLPLPAQELLAKDLHGNQWRFRHSYRGTPQRHSLTTGWNEFTTSK  
KLVKGDVIVFVRGETGELRVGIRR

>A.tha-ARF21

FTKVLTASDTSAYGGFSVPPKKHAIIECLPPLDMSQPLPAQEILAILHDNQRFRHNYRGTPQRHSLTTGWNEFI  
TSKLVKGDVIVFVRGETGELRVGIRR

>A.tha-ARF22

FTKVLTASDTSGGFFVPPKKHAIIECLPPLDMSQPLPTQELLATDLHGNQWRFNHNYRGTPQRHLLTTGWNAFTT  
SKKLVAGDVIVFVRGETGELRVGIRR

>A.tha-ARF23

FTKVLTASDTSAQGEFSVPPCKHAIIECLPPLDMSQPIPAQELIAIDLHGNQWRFKHSYRVPRGDTTGWNAFTTSK  
KLVVGDVIVFARGETGELRVGIR









## Abstract

Auxin is a plant hormone implicated in almost all plant developmental stages, since the embryo formation till flowering, determining the position of the organs in the plant and thus, its whole structure. As for any other hormone, auxin perception is followed by a signal transduction that finishes in a series of changes in a plant cell, including transcriptional changes. This thesis is divided in 3 chapters, each with a focus on the structural, molecular and evolutionary aspects of different proteins involved in the regulation of auxin genes response.

First, we focused our studies on TOPLESS (TPL), a co-repressor implicated, not only in auxin responsive genes repression, but also in many other plant processes due to its interactions with numerous transcriptional repressors in plants. Our determination of the TPL N-terminal structure allowed us to understand that TPL can interact with different partners through the same binding site. Moreover, it revealed that TPL is a tetrameric protein, with the tetramerization interface formed by a newly identify domain, the CRA domain, that is also part of the binding site. The high residues conservation in both tetramerization interface and TPL binding site since m.y.a indicates the importance of TPL role since the origin of plants. This work also shows that the structural similarities between TPL and other co-repressor with similar domains but different function nicely exemplify how evolution plays with common features for creating new functions.

Second, we studied ARF proteins, the transcription factors of the auxin transcriptional response, with a focus on their DNA binding preferences. For this, we used a combination of bioinformatic analyses of DAP-seq ARFs genomic binding, with *in vitro* DNA binding tests and structure modelling. Our results point out that different ARFs can have different preferential binding sites within the genome, with these preferences being determined by the orientation and spacing of the binding motifs. Moreover, our studies suggest that depending on the binding site, ARFs could bind with different conformations using dimerization interfaces not yet discovered. These results can explain how different ARFs co-expressed inside a plant cell can collaborate to the specificity and robustness of auxin transcriptional response by differential bindings to the genome.

Finally, we travelled back in time to position the origin of auxin signalling pathway in the evolution of plants. Here we looked for protein homologues of the auxin signalling pathway in charophyte green algae, the most ancient plants ancestor (450 M years). This search retrieved an ARF and a TPL homologue in the first multicellular charophyte algae (*Chlorokybus atmophyticus*). The biochemical characterization of *C. atmophyticus* ARF indicated that it presented already the same properties of the ARFs from land plants and that it was able to interact with TPL protein, as it is the case for some ARFs. The absence of auxin receptor homologues in these primitive algae indicates however that auxin-dependency appeared with the acquisition of TIR1/AFB-Aux/IAA coreceptor system, after charophytes divergence into land plants.

## Résumé

L'auxine est une hormone végétale impliquée dans presque toutes les étapes du développement des plantes, de la formation de l'embryon jusqu'à la floraison, déterminant la position des organes et donc la structure de la plante. Comme pour les autres hormones, la perception de l'auxine est suivie par une transduction du signal qui produit une série de changements dans les cellules végétales dont des régulations transcriptionnelles. Cette thèse est divisée en 3 chapitres, chacun d'eux étant focalisé sur des aspects structuraux, moléculaires et évolutifs de différentes protéines impliquées dans la régulation des gènes de réponse à l'auxine.

Nous avons tout d'abord centré nos études sur TOPLESS (TPL), un corépresseur qui agit au niveau de la répression des gènes de réponse à l'auxine, mais aussi dans d'autres processus végétaux compte tenu de son interaction avec de nombreux répresseurs transcriptionnels. Nous avons déterminé la structure de la partie N-terminale de TPL et compris comment TPL interagit avec différents partenaires au niveau d'un même site de liaison. Nous avons alors démontré que TPL forme un tétramère à l'aide d'une surface de tétramérisation constituée par un nouveau domaine, le domaine CRA, qui fait aussi partie du site de liaison. Les résidus impliqués dans la tétramérisation et l'interaction avec des partenaires sont très conservés depuis des centaines de millions d'années montrant ainsi l'importance du rôle de TPL depuis l'origine des plantes. Enfin, les similarités de structure entre TPL et d'autres corépresseurs qui possèdent des domaines similaires mais possédant une fonction différente montrent un bel exemple de la manière dont l'évolution joue avec des domaines protéiques pour créer de nouvelles fonctions.

Nous avons ensuite étudié les préférences de liaison à l'ADN des facteurs de transcription de la réponse à l'auxine (ARF). Pour cela nous avons utilisé une combinaison d'analyses bio-informatiques de données de DAP-seq sur la liaison des ARFs sur le génome, des tests d'interaction ADN-protéine *in vitro* et de la modélisation de structures. Nos résultats indiquent que les différents ARFs ont des sites préférentiels de liaison sur le génome et que ces préférences sont déterminées par l'orientation et l'espacement entre motifs de liaison. Enfin, ces études suggèrent qu'en fonction du site de liaison, les ARFs pourraient se lier avec différentes conformations à l'aide de surfaces de dimérisation qui ne sont pas encore décrites. Ces résultats permettent d'expliquer comment différents ARFs coexprimés dans la même cellule peuvent fonctionner ensemble pour contribuer à une réponse transcriptionnelle à l'auxine spécifique et robuste.

Finalement, nous avons remonté le temps pour positionner l'origine de la voie de signalisation de l'auxine chez les plantes. Pour cela, nous avons recherché des homologues des protéines de la voie de signalisation de l'auxine dans des algues vertes charophytes, les ancêtres les plus lointains (450 Millions d'années) des plantes. Nous avons alors trouvé un homologue des ARFs et TPL chez les premières algues multicellulaires (*Chlorokybus atmophyticus*). La caractérisation biochimique de l'ARF de *C. atmophyticus* indique qu'il partageait déjà les mêmes propriétés que les ARFs des plantes terrestres et était aussi capable d'interagir avec TPL comme certains ARFs. L'absence d'homologues du récepteur de l'auxine chez ces algues primitives indique cependant que la dépendance à l'auxine aurait été acquise plus tard avec l'apparition du système corécepteur TIR1/AFB-Aux/IAA après la divergence des charophytes vers les plantes terrestres.