



**HAL**  
open science

# Structure, fonction et évolution de LEAFY, facteur de transcription clé du développement floral

Camille Sayou

► **To cite this version:**

Camille Sayou. Structure, fonction et évolution de LEAFY, facteur de transcription clé du développement floral. Sciences agricoles. Université de Grenoble, 2013. Français. NNT : 2013GRENV029 . tel-01222909

**HAL Id: tel-01222909**

**<https://theses.hal.science/tel-01222909>**

Submitted on 31 Oct 2015

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

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

## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Végétale**

Arrêté ministériel : 7 août 2006

Présentée par

**Camille SAYOU**

Thèse dirigée par **Renaud DUMAS**

préparée au sein du **Laboratoire de Physiologie Cellulaire et Végétale**

dans l'**École Doctorale Chimie et Sciences du Vivant**

## Structure, fonction et évolution de LEAFY, facteur de transcription clé du développement floral

Thèse soutenue publiquement le **30 septembre 2013**,  
devant le jury composé de :

**Dr. Marie-Hélène LE DU**

Rapporteur

**Dr. Thierry LAGRANGE**

Rapporteur

**Dr. Peter ROGOWSKY**

Examineur

**Dr. François PARCY**

Examineur

**Dr. Renaud DUMAS**

Directeur de Thèse



## Remerciements

Le travail présenté ici est avant tout un travail d'équipe, je tiens donc à remercier ceux qui y ont participé, de près ou de loin, et sans qui cette belle histoire ne serait pas la même.

Je remercie **Norbert Rolland**, **Laurent Blanchoin** et **Marilyn Vantard** de m'avoir accueillie au sein de leur laboratoire.

Je remercie également **Marie-Hélène Le Du**, **Thierry Lagrange**, **Peter Rogowsky** et **Rob Ruigrok** d'avoir accepté d'évaluer mon travail.

Un IMMENSE Merci à **Renaud** qui m'a accompagné durant ces 3 super années passées au labo. Merci pour ta grande générosité et pour ta disponibilité permanente, tu as toujours été à l'écoute et attentionné. Merci aussi pour ta patience face à mon esprit de contradiction parfois persistant. Ta grande rigueur mélangée à tes idées parfois un peu folle ont rendu chaque petit résultat passionnant. Quand j'ai frappé à la porte de l'équipe, c'était pour voir des plantes, finalement tu as su me transmettre ton enthousiasme pour la vie trépidante des protéines et autres objets moléculaires. On aura beaucoup bossé, mais aussi bien rigolé !

Un merci tout aussi grand à **François** qui, même si cela n'apparaît pas dans les lignes officielles, m'a encadré au même titre que Renaud. J'ai eu beaucoup de chance de vous avoir comme duo de "chefs" et de profiter de vos expertises respectives. Merci de partager toutes tes idées, et ça même depuis l'autre côté de l'Atlantique. Merci pour ton enthousiasme à l'égard des différents projets, et pour les heures passées à discuter de LFY dans le bureau. Merci aussi pour ta confiance et pour l'énergie que tu passes à valoriser notre travail.

Un grand merci à **Marie** qui a été une associée en or ! C'était un plaisir de bénéficier de ta bonne humeur constante, de ta bonne volonté et de ta générosité pendant tout ce temps passé au labo. On aura fait une bonne équipe pour résoudre les mystères de l'évolution !

Un super merci à super **Manou**, notre lab-manager de choc ! On sait bien que sans toi le labo ne tournerait pas aussi bien. Merci d'avoir toujours été volontaire pour me donner des coups de main à chaque fois que j'en avais besoin, et d'avoir pris à cœur toutes les missions que je t'ai confiées.

Merci également à **Max**, notre associé indispensable pour la cristallographie. Tu as passé des dizaines (peut être bien des centaines) de nos cristaux sur ta ligne et tu nous as résolu de superbes structures. Merci pour ta disponibilité et ton enthousiasme à l'idée de faire parler LFY.

Merci aussi à **Marc** qui a passé plusieurs journées à analyser mes échantillons sur le SEC-MALLS et sans qui on n'aurait toujours pas résolu la taille de LFY.

Un grand merci à **Patrick** et **Carlo** qui ont fait partie de mon comité de thèse et m'ont bien aidé à faire avancer le projet dans le bon sens. Merci également à **Chloé** avec qui on a plein de nouveaux projets super excitants, c'est vraiment dommage que tu arrives quand je pars !

**Gaby**, c'est toi qui m'as encadré à mes débuts. On n'avait dit qu'on ferait pousser des plantes, finalement c'est des prots que j'ai cultivées avec Renaud. Heureusement, on a eu l'occasion de se rattraper et tu m'as donné un grand coup de main quand je ne maîtrisais plus tout à fait la situation !

Deux grands mercis à **Edwige & Cécile**, les deux thésardes qui nous ont précédé Marie et Moi. Edwige, tu as été un super exemple, tu m'as transmis ton LFY-virus et j'ai beaucoup d'admiration pour ton travail. Cécile, tu m'avais laissé une super thèse pleine de bons plans pour apprivoiser LFY, les pages commencent d'ailleurs à tomber tellement je les ai tournées ! Même si vous aviez fait du super boulot, vous m'aviez laissé des bases de travail très prometteuses, alors merci d'avoir passé le relais.

Hicham, ou plutôt **Hichaminator** ! Merci pour ton enthousiasme et théories révolutionnaires mais bien fondées sur tous les articles que tu lis. J'ai toujours espéré que la hichamite soit contagieuse pour devenir aussi efficace que toi en biblio. Dommage qu'on n'ait pas eu le temps d'écrire une revue, j'aurais sans aucun doute appris plein de trucs.

**Fanny**, depuis que tu es là, notre bureau est bien plus gai, grâce à ton énergie et ton rire débordant. T'es bien plus qu'impressionnante sur des parois verticales. J'ai d'abord cru que cette force venait de ton alimentation, mais même après que tu m'aies fait déguster de nombreuses fois ton menu fétiche (pain, fromage, C.d.R.), je suis toujours aussi mauvaise en grimpe ! En tout cas merci pour ta patience. Je parierais volontiers sur le 7 avant la fin de ta thèse : une tourte au thon contre une rando-refuge - tu valides ?

Merci à **Cristel, Gilles, Robert & Julia** toujours prêts à répondre à mes questions concernant les plantes et la chromatine ! C'était vraiment chouette d'encadrer les TP avec toi Cristel, dommage que ça s'arrête, on commençait à être super rodées !

Un grand merci à **Marie-MLM** pour ta gentillesse et ton application, que ce soit pour les manips, l'organisation du labo, mais aussi nos fameuses sorties d'équipe. On a bien de la chance que tu sois venue dans notre équipe !

Merci à **Marta**, my favourite 'stagiaire'. You came to learn protein science with Renaud and I but you helped me a lot with the oligomerisation story. I remember the terrible 15 days we spent preparing samples for the SAXS ! Thanks for your help and good luck for you PhD !

Merci à **Sandrine, Mylène, Eugenio, Sophie, Thomas, Aude, Elena, Pierre, Meryl, Béatrice, Jeanne** qui font ou ont fait partie de l'équipe et ont contribué à la bonne humeur qui y règne !

Merci à l'équipe des jeunes du labo, anciens et actuels, et à la promo Irtelis-Grenoble 2010 pour tous les bons moments passé ensemble. Merci aussi à Sylvianne, Sophie et l'équipe des secrétaires, toujours prêtes à prendre en charge nos demandes administratives, ainsi qu'à tous les membres de PCV grâce à qui c'est un plaisir de venir au labo chaque jour.

Merci aussi à l'équipe de Grenoble et celle des drômois sur qui j'ai pu compter pour tout ce qui ne concerne pas cette thèse !

Un merci particulier à Sylvian qui accepte depuis longtemps les contraintes que je nous impose et qui est toujours d'un grand soutien.

Et enfin, merci à ma famille qui m'a encouragé et soutenu, ainsi qu'à ma famille d'adoption à Margès avec qui je passe toujours d'agréables week-ends et de jolies vacances. Une pensée très spéciale à mon papa qui aurait sans doute été le plus fier de me voir soutenir cette thèse.

## Sommaire

<b>INTRODUCTION</b> .....	<b>5</b>
<b>I. La régulation de l'expression des gènes</b> .....	<b>7</b>
1. L'information génétique, un mode d'emploi à décoder .....	7
2. La régulation de l'expression du génome .....	7
3. Notion de réseau de régulation .....	9
<b>II. Les facteurs de transcription</b> .....	<b>10</b>
1. Des acteurs majeurs de la régulation transcriptionnelle .....	10
2. Diversité des modes de liaison ADN-protéine .....	10
<b>III. LEAFY, régulateur clé du développement floral</b> .....	<b>15</b>
1. La floraison, un processus crucial et fascinant .....	15
2. <i>LFY</i> , chef d'orchestre du développement floral .....	16
3. La protéine <i>LFY</i> .....	20
4. <i>LFY</i> au cours de l'évolution.....	24
<b>OBJECTIFS</b> .....	<b>28</b>
<b>CHAPITRE I. Le domaine N-terminal de LEAFY, de la structure à la fonction</b> .	<b>31</b>
<b>I. Introduction</b> .....	<b>33</b>
<b>II. Article 1</b> .....	<b>35</b>
<b>III. Résultats et discussions complémentaires</b> .....	<b>71</b>
1. Production de protéines recombinantes <i>LFY</i> , domaine N-terminal et protéines entières 71	
2. Analyse structurale de <i>LFY</i> .....	80
3. Structure du domaine N-terminal de <i>LFY</i> du <i>Ginkgo biloba</i> , un nouveau domaine SAM qui oligomérisé .....	84
4. Les diverses fonctions des domaines SAM .....	88
5. Rôle de l'oligomérisation pour la fonction de <i>LFY</i> comme régulateur floral.....	92
6. Rôle de l'oligomérisation du domaine SAM pour la liaison à l'ADN <i>in vitro</i> .....	96
7. Oligomérisation et liaison à l'ADN <i>in planta</i> , à l'échelle du génome .....	105

<b>IV. Bilan et perspectives .....</b>	<b>109</b>
1. Bilan .....	109
2. Perspectives .....	110
<b>CHAPITRE II. Évolution du mode de liaison à l'ADN de LEAFY .....</b>	<b>113</b>
<b>I. Introduction .....</b>	<b>115</b>
1. Régulation et évolution .....	115
2. Modifications en <i>cis</i> ou en <i>trans</i> ? .....	115
3. LFY, un bon candidat pour étudier l'évolution des facteurs de transcription .....	117
<b>II. Article 2 .....</b>	<b>120</b>
<b>III. Bilan et perspectives .....</b>	<b>136</b>
<b>MATERIEL ET METHODES .....</b>	<b>143</b>
<b>I. Matériel .....</b>	<b>143</b>
1. Matériel végétal .....	143
2. Matériel bactérien .....	143
<b>II. Méthodes .....</b>	<b>143</b>
1. Biologie moléculaire .....	143
2. Biochimie des protéines .....	145
3. Cristallographie des protéines .....	149
4. Biologie végétale .....	149
5. Caractérisation des interactions ADN-protéine chez <i>A. thaliana</i> par CHIP-seq .....	153
<b>CONCLUSION .....</b>	<b>139</b>
<b>REFERENCES .....</b>	<b>155</b>
<b>ANNEXES .....</b>	<b>165</b>

## Abréviations

<b>ADN</b>	Acide desoxyribonucléique
<b>ADNc</b>	ADN complémentaire, issu de la rétrotranscription d'un ARN messager
<b>AG</b>	<i>AGAMOUS</i> , gène chez <i>A. thaliana</i>
<b>AP1</b>	<i>APETALA1</i> , gène chez <i>A. thaliana</i>
<b>AP3</b>	<i>APETALA3</i> , gène chez <i>A. thaliana</i>
<b>ARN</b>	Acide ribonucléique
<b>ARNm</b>	ARN messager
<b>ATP</b>	Adenosine triphosphate
<b>ATPase</b>	Enzyme utilisant l'énergie de l'hydrolyse de l'ATP
<b>ChIP-Seq, ChIP-chip</b>	Immunoprécipitation de la chromatine (chromatin immunoprecipitation) suivie d'un séquençage massif (Seq) ou de l'hybridation des séquences sur puce (chip)
<b>DNase</b>	Desoxyribonucléase, enzyme de dégradation de l'ADN
<b>EDTA</b>	Acide Ethylène Diamide Tétraaécétique
<b>EMSA</b>	Electrophoretic Mobility Shift Assay, ou gel retard
<b>FT</b>	Facteur de transcription
<b>GFP</b>	Green Fluorescent Protein, protéine fluorescente verte
<b>HTH</b>	Helix-Turn-Helix (hélice-tour-hélice), motif de liaison à l'ADN
<b>IPTG</b>	Isopropyl- $\beta$ -D-1-thiogalactopyranoside
<b>Kd</b>	Constante de dissociation
<b>kDa</b>	kilodalton
<b>LFY</b>	Le gène <i>LEAFY</i> (police italique)
<b>LFY</b>	La protéine LEAFY
<b>LFY-C</b>	Domaine C-terminal de LFY (domaine de liaison à l'ADN)
<b>LFY-N</b>	Domaine N-terminal de LFY
<b>LFY<math>\Delta</math></b>	Protéine LFY tronquée de 40 à 50 acides aminés du côté N-terminal
<b>MADS-box</b>	Boîte MADS, domaine des gènes homéotiques de type MADS, du nom des premiers membres identifiés ( <i>MCM1</i> , <i>AGAMOUS</i> , <i>DEFICIENS</i> , <i>SRF</i> )
<b>MW</b>	Molecular Weight, poids moléculaire
<b>NLY</b>	<i>NEEDLY</i> , paralogue de <i>LFY</i> chez les gymnospermes
<b>p35S</b>	Promoteur constitutif fort en amont de l'ARN 35S du virus de la mosaïque du chou-fleur (CaMV)
<b>PCR</b>	Polymerase Chain Reaction, réaction en chaîne par polymérase
<b>pb</b>	Paire de bases
<b>PDB</b>	Protein Data Bank, base de données des structures de macromolécules biologiques ( <a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a> )
<b>pLFY</b>	Promoteur du gène <i>LFY</i>
<b>PpLFY1 et 2</b>	<i>Physcomitrella patens LFY 1 et 2</i>
<b>PWM</b>	Position Weight Matrix, matrice poids-position associant un poids à chaque nucléotide du site de liaison d'un facteur de transcription
<b>QuMFRA</b>	Quantitative Multiple Fluorescence Relative Affinity
<b>rmsd</b>	Root Mean Square Deviation, mesure de similarité structurale entre deux protéines
<b>RT-PCR</b>	Reverse Transcription (Rétrotranscription) suivie d'une PCR
<b>SAM</b>	Sterile Alpha Motif
<b>SDS-PAGE</b>	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis, gel de polyacrylamide en conditions dénaturantes
<b>SEC</b>	Size Exclusion Chromatography, chromatographie d'exclusion de taille
<b>SEC-MALLS</b>	SEC-Multi Angle Laser Light Scattering, chromatographie d'exclusion de taille couplée à une mesure de masse moléculaire par diffusion de lumière
<b>SELEX</b>	Systematic Evolution of Ligands by Exponential enrichment
<b>SEP3</b>	<i>SEPALATA3</i> , gène chez <i>A. thaliana</i>
<b>TCEP</b>	Tris(2-carboxyethyl)phosphine
<b>TBE</b>	Tampon Tris Borate EDTA
<b>TEV (protéase)</b>	Protéase du Virus Tobacco Etch
<b>TFL1</b>	<i>TERMINAL FLOWER1</i> , gène chez <i>A. thaliana</i>
<b>Tris</b>	Trishydroxyméthylaminométhane
<b>U.A.</b>	Unité Arbitraire
<b>UFO</b>	<i>UNUSUAL FLORAL ORGANS</i> , gène chez <i>A. thaliana</i>
<b>WT</b>	Wild-type, protéine ou plante sauvage





# INTRODUCTION



# I. La régulation de l'expression des gènes

## 1. L'information génétique, un mode d'emploi à décoder

Les caractéristiques biologiques d'un organisme reposent sur son information génétique et sur la manière dont celle-ci est utilisée. La régulation précise de l'expression des gènes est indispensable au bon fonctionnement de la cellule ou de l'organisme. Le développement et l'adaptation aux conditions extérieures en dépendent.

L'information génétique est stockée sous forme d'ADN. Pour être utilisée cette information est transférée de manière sélective, en fonction des besoins de la cellule, d'abord sous forme d'ARN, puis dans certains cas sous forme de protéines. Les protéines, ainsi que l'ARN, sont les molécules effectrices de la cellule.

## 2. La régulation de l'expression du génome

Différents éléments sont impliqués dans la régulation de l'expression du génome (**Figure 1**) (Weake et Workman, 2010 ; Levine et Tjian, 2003 ; Barrera et Ren, 2006 ; Babu *et al.*, 2004). D'une part, au niveau de l'ADN, les gènes sont entourés de séquences régulatrices appelées éléments *cis*. D'autre part, des facteurs dits *trans* reconnaissent les éléments *cis* et permettent l'activation ou la répression des gènes.

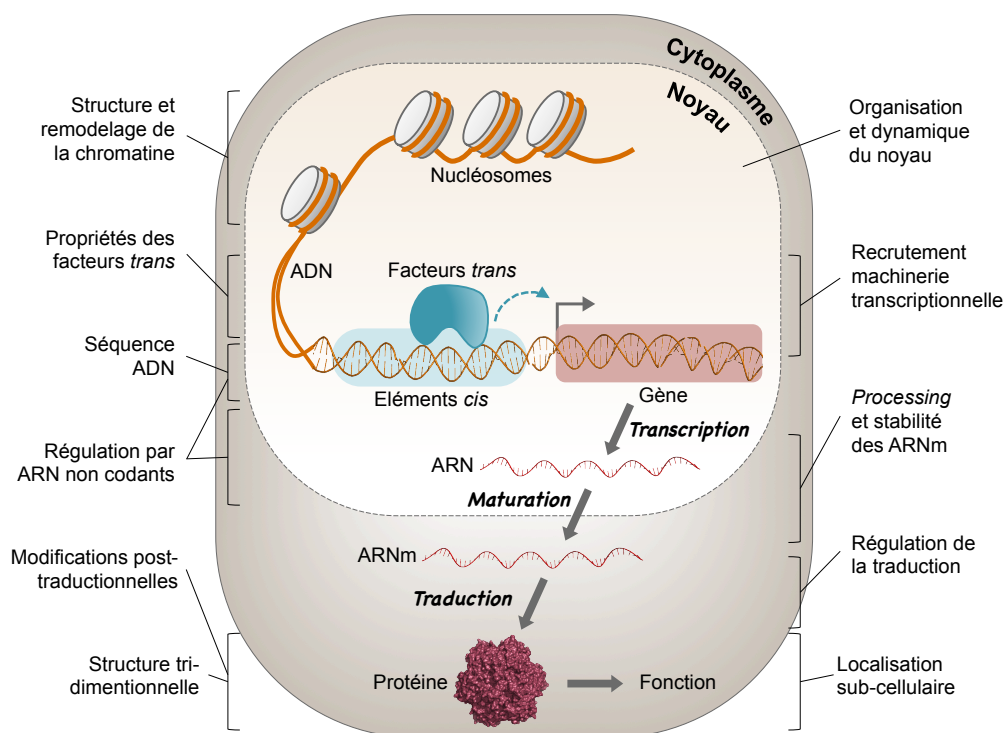
La molécule d'ADN, avec sa structure en double hélice, est organisée par les nucléosomes, composés d'histones, pour former la chromatine. Les facteurs *trans* sont donc influencés par la séquence de l'ADN et par la structure de la chromatine.

Parmi les facteurs qui agissent sur les éléments *cis* on compte de nombreux acteurs. Les facteurs de transcription (FTs) apportent le caractère spécifique à la régulation, car ils reconnaissent uniquement certaines séquences d'ADN. Les co-activateurs qui s'associent aux FTs peuvent apporter un niveau de spécificité supplémentaire, en étant spécifiques de certains tissus par exemple, ou bien ils peuvent servir d'intermédiaire entre les FTs et d'autres éléments régulateurs.

Les marques épigénétiques sont des modifications chimiques de l'ADN et des protéines associées à l'ADN, en particulier les histones, apportées par des enzymes. Bien que ces marques n'altèrent pas la séquence de l'ADN, elles peuvent être transmises à la génération suivante et ont des rôles majeurs dans la régulation du génome. Les complexes qui jouent sur la structure de la chromatine sont donc très importants pour la régulation transcriptionnelle : ils peuvent apporter des modifications post-traductionnelles sur les

histones, mais aussi les déplacer ou les retirer et contrôler ainsi l'accès à certaines régions (Bell *et al.*, 2011). L'ADN peut aussi être modifié, en particulier méthylé, ce qui influence son état transcriptionnel (Feng *et al.*, 2010). On compte également des ARN non codants, qui peuvent eux aussi être spécifiques de certaines séquences d'ADN et agir sur l'expression des gènes. De nombreuses classes de petits et longs ARN non codants ont été découverts et participent, par des mécanismes divers, à la formation de l'hétérochromatine, à la méthylation de l'ADN, au dépôt de marques sur les histones, et plus généralement à la répression transcriptionnelle (Sabin *et al.*, 2013 ; Castel et Martienssen, 2013).

L'organisation du noyau peut aussi jouer un rôle dans la régulation transcriptionnelle (Lanctot *et al.*, 2007 ; Sexton *et al.*, 2007). En effet, le noyau n'est pas un compartiment homogène, il est divisé en zones où la chromatine est plus ou moins condensée (hétéro et euchromatine) et en régions plus ou moins riches en gènes. Cette distribution, et le repositionnement de la chromatine au sein du noyau, participent à la régulation de l'expression des gènes.



**Figure 1 | Les différents niveaux de régulation du génome.**

Schéma simplifié des multiples paramètres affectant la régulation des gènes dans une cellule (inspiré par Alonso et Wilkins, 2005).

Cette régulation ne se limite pas à la régulation transcriptionnelle, elle implique d'autres processus complexes. Après la synthèse de l'ARN messager, l'expression des gènes est régulée à plusieurs niveaux : au niveau post-transcriptionnel avec la maturation de l'ARN et sa stabilité, au niveau traductionnel avec le passage de l'ARN à la protéine. Une fois la

protéine synthétisée, son activité peut être régulée par des modifications post-traductionnelles.

Ces processus aboutissent à la production de protéines, qui sont un des plus importants « outil du vivant ». Par exemple, les enzymes transforment et fabriquent des métabolites, les protéines structurales permettent le maintien et la mobilité de la cellule, les protéines de signalisation permettent la transmission d'informations, les protéines de transport permettent la communication entre les différents compartiments et avec l'extérieur de la cellule, et des protéines de régulation servent à leur tour à décrypter l'information génétique.

### 3. Notion de réseau de régulation

Même si le nombre de gènes codant pour des protéines varie selon les organismes, il est très élevé. On compte par exemple environ 4000 gènes codant pour des protéines chez la bactérie *Escherichia coli*, 6000 chez la levure *Saccharomyces cerevisiae*, 20 000 chez le ver *Caenorhabditis elegans*, 22 000 chez l'homme et 25 000 chez la plante *Arabidopsis thaliana*. Chaque processus biologique, comme les transitions développementales, la réponse à un stress ou à un changement environnemental, ou tout simplement le fonctionnement basal des cellules nécessitent la régulation coordonnée de nombreux gènes. On parle alors de réseaux de régulation transcriptionnelle (Babu *et al.*, 2004 ; Zhu *et al.*, 2007).

Des changements, même mineurs, du moment et du lieu d'action des facteurs *trans*, ou des modifications des éléments *cis* peuvent donc conduire à des bouleversements phénotypiques importants (Carroll, 2000 ; Peter et Davidson, 2011 ; Prud'homme *et al.*, 2007 ; Wagner et Lynch, 2008 ; Alonso et Wilkins, 2005). Ces éléments régulateurs ont très souvent été recrutés par l'évolution pour créer de la nouveauté.

Pour comprendre un processus biologique, il faut se placer à plusieurs échelles (Fraser *et al.*, 2013). Dans le cas de la régulation de l'information génétique, il s'agit d'identifier les FTs et autres facteurs *trans*-régulateurs, identifier les éléments *cis*-régulateurs et les gènes cibles et comprendre les liens entre tous ces facteurs, aussi bien au niveau moléculaire qu'à grande échelle au niveau du réseau dans sa globalité.

## II. Les facteurs de transcription

### 1. Des acteurs majeurs de la régulation transcriptionnelle

Les FTs sont des éléments clés de la régulation transcriptionnelle. Environ 10% des protéines seraient des FTs chez l'homme et 6 à 8% chez la plante *A. thaliana* (Levine et Tjian, 2003 ; Lang *et al.*, 2010; Riechmann *et al.*, 2000). On distingue les FTs généraux et les FTs spécifiques. Largement impliqués dans la transcription des ARN messagers, les FTs généraux forment des complexes protéiques à proximité du site d'initiation de la transcription et recrutent l'ARN polymérase II (Sikorski et Buratowski, 2009). On s'intéressera plus particulièrement ici aux FTs dits spécifiques. Ce sont eux qui permettent d'activer ou de réprimer certains gènes en fonction des conditions (type cellulaire, stade de développement, signal extérieur). Ils reconnaissent spécifiquement les régions régulatrices de leurs gènes cibles via des contacts ADN-protéine.

Comprendre comment les FTs interprètent le génome, reconnaissent les bonnes séquences d'ADN, accèdent à ces régions dans le contexte du noyau, et régulent leurs cibles passe notamment par l'étude de leur mécanisme d'action au niveau moléculaire.

### 2. Diversité des modes de liaison ADN-protéine

Les FTs sont souvent des protéines multi-domaines, leur point commun étant la présence d'un domaine de liaison à l'ADN (bien que certains FTs n'en aient pas mais fassent partie de complexes qui eux-mêmes lient l'ADN). Ils adoptent une multitude de structures et sont régulés par des mécanismes infiniment variés.

Les motifs de liaison à l'ADN peuvent être classés selon leur structure tridimensionnelle, comme proposé par Rohs et ses collègues dans une revue récente (Rohs *et al.*, 2010). Quelques classes de motifs sont décrites ci-dessous, et des exemples sont donnés pour illustrer la diversité des modes de fonctionnement des FTs.

#### ▪ Les motifs de liaison à l'ADN à hélices $\alpha$

L'hélice  $\alpha$  est la structure secondaire la plus fréquente pour la liaison à l'ADN par les protéines.

Un des motifs le plus représenté est le *helix-turn-helix* (HTH), composé de trois hélices reliées par des boucles (Aravind *et al.*, 2005). La troisième hélice ( $\alpha_3$ ) contacte directement

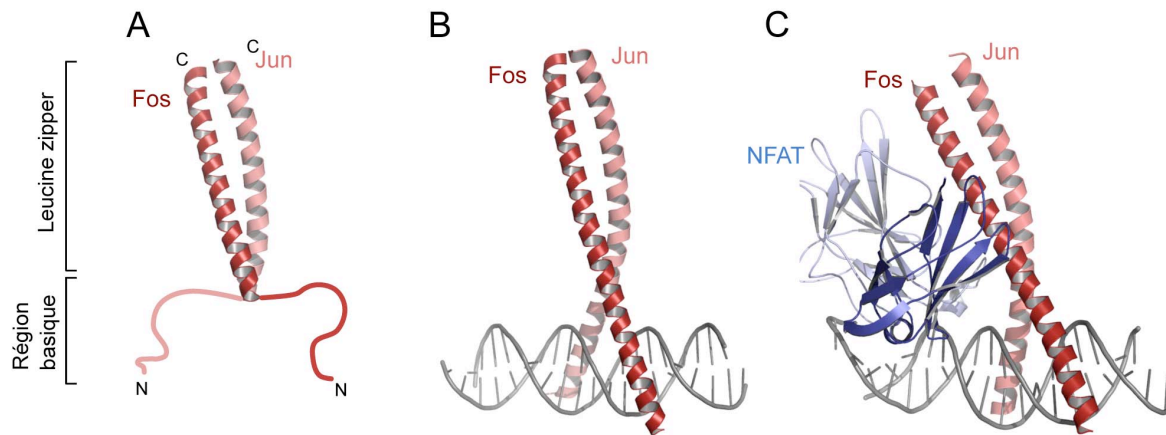
l'ADN par des liaisons hydrogène dans le grand sillon, les autres hélices ont un rôle stabilisateur. Ce motif est inséré dans un domaine plus large, très variable selon les protéines. La protéine LEAFY, qui fait l'objet de cette thèse, lie l'ADN grâce un motif HTH (Hamès *et al.*, 2008) et sera plus largement décrite par la suite.

Deux grandes hélices viennent se fixer de part et d'autre de l'ADN, à la manière d'une pince à linge, pour le motif dit *basic region leucine zipper* (bZIP) (**Figure 2**, exemple1). Le motif *basic helix-loop-helix* (bHLH) est organisé de manière similaire, mais chaque grande hélice est divisée en deux hélices reliées par une boucle. Pour chaque monomère, l'extrémité N-terminale de l'hélice plonge dans le grand sillon de l'ADN, et l'extrémité C-terminale dimérise avec le second monomère par des contacts hydrophobes (via des leucines). La dimérisation détermine l'affinité et la spécificité pour l'ADN. Pour beaucoup de motifs leucine zipper, la région qui lie l'ADN est non-structurée et se replie au contact de l'ADN. L'homo ou hétéro-dimérisation permet de lier soit des séquences palindromiques, soit deux demi-sites de nature différente (Krylov et Vinson, 2001).

#### Exemple 1 : Jun et Fos

Les protéines Jun et Fos sont les sous-unités du complexe régulateur Activator Protein 1 (AP-1), impliqué dans la croissance et la prolifération cellulaires, et la régulation du cycle cellulaire chez les animaux. Grâce à leur domaine bZIP, elles forment des homodimères Jun-Jun ou des hétérodimères Jun-Fos et reconnaissent des séquences palindromiques sur les promoteurs de leurs gènes cibles. Le dimère Jun-Fos adopte sur l'ADN la structure des motifs bZIP décrite précédemment (**Figure 2B**) (Glover et Harrison, 1995). Une étude thermodynamique a montré que la région basique était non structurée en absence d'ADN, et se repliait seulement lors de la liaison (**Figure 2A**) (Seldeen *et al.*, 2008). Nuclear Factor of Activated T cells (NFAT) est un FT multi-domaine importé dans le noyau, après déphosphorylation, en réponse à un stimulus. Des sites de liaison AP-1 sont présents à côté de la plupart des sites NFAT. Les domaines bZIP de Jun et Fos, et deux domaines NHR (NFAT Homology Region) de NFAT forment un complexe sur ces régions (**Figure 2C**). Leur liaison est coopérative et permet d'augmenter l'affinité relativement faible de NFAT et AP-1 individuellement (Chen *et al.*, 1998).

Cet exemple illustre plusieurs mécanismes moléculaires importants pour les FTs : les phénomènes d'homo et hétéro-dimérisation, les changements de conformation induits par les interactions ADN-protéine et protéine-protéine, la coopérativité entre plusieurs facteurs et la régulation par la localisation subcellulaire.



**Figure 2 | Exemple de domaine bZIP (basic region *leucine zipper*) des protéines Jun et Fos.**

**A.** Modèle de la structure du domaine bZIP lorsqu'il n'est pas lié à l'ADN. La partie *leucine zipper* en C-terminal est dimérique, la région basique en N-terminal est non structurée (adapté de Seldeen *et al.*, 2008 ; PDB 1FOS).

**B.** Structure du domaine bZIP sur l'ADN (Glover et Harrison, 1995 ; PDB 1FOS). La région basique est repliée et prolonge l'hélice du *leucine zipper*.

**C.** Structure du domaine bZIP de Jun-Fos en complexe avec la protéine Nuclear Factor of Activated T cells (NFAT) sur l'ADN (Chen *et al.*, 1998 ; PDB 1A02). La liaison de Jun-Fos et NFAT sur ce fragment d'ADN contenant leurs deux sites de liaison est coopérative.

#### ▪ Les motifs de liaison à l'ADN à feuillets $\beta$

Les motifs de liaison à l'ADN majoritairement composés de feuillets  $\beta$  sont peu communs, mais ils sont utilisés par certaines protéines pour lier spécifiquement l'ADN, par exemple les TATA-box Binding Proteins (TBP) qui se fixent dans le petit sillon de l'ADN, ou le célèbre suppresseur de tumeur p53 qui adopte un repliement dit *Immunoglobulin-like* (**Figure 3**, exemple 2).

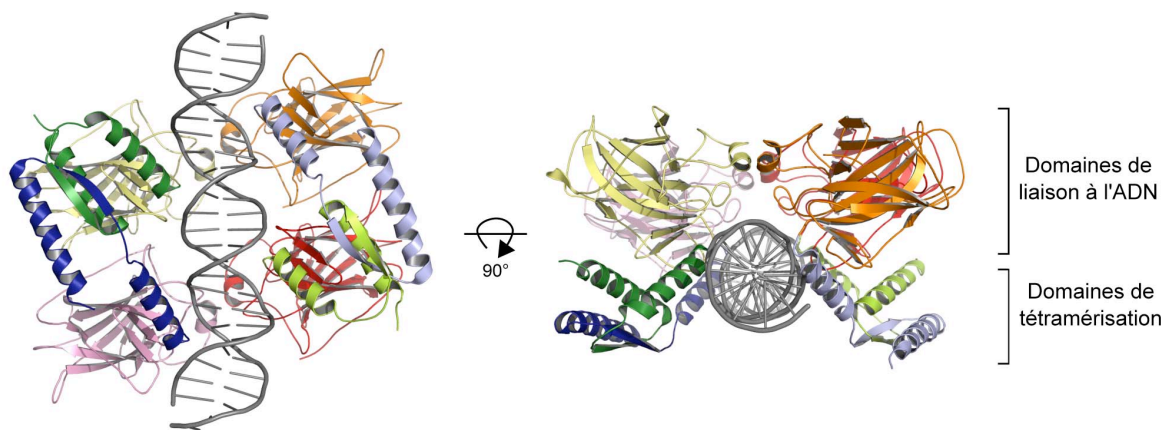
#### Exemple 2 : p53

p53 est un facteur de transcription trouvé chez les animaux et très largement étudié, car impliqué dans de nombreux cancers chez l'homme. Il avait d'abord été qualifié de proto-oncogène, mais c'est en fait un suppresseur de tumeur qui intègre les signaux de stress et induit l'arrêt du cycle cellulaire ou l'apoptose. Il est qualifié de « *master* » régulateur car il intègre de nombreux signaux et orchestre à son tour de multiples processus. La fonction initialement proposée était basée sur l'étude de protéines mutantes, qui sont effectivement présentes dans de très nombreux cancers et ont un effet dominant négatif sur la protéine sauvage (Freed-Pastor et Prives, 2012).

Ce FT est composé de deux domaines structurés : un domaine de liaison à l'ADN et un domaine de tétramérisation en C-terminal. Il possède aussi des domaines non-structurés en N-terminal : un domaine d'activation et un motif riche en prolines. À travers les multiples



structures (haute et basse résolution) de p53 publiées, on décèle la difficulté d'obtenir la structure de la protéine entière, due à sa flexibilité et son caractère multi-domaine. Seuls les deux domaines structurés, comportant d'ailleurs plusieurs mutations et délétions pour stabiliser la protéine, ont été cristallisés (Petty *et al.*, 2011 ; Emamzadah *et al.*, 2011). L'ADN, qui comporte 4 séquences répétées, est lié par deux dimères situés de part et d'autre de la double hélice (**Figure 3**). Les domaines *Immunoglobulin-like* se placent dans le grand sillon de l'ADN et sont reliés par les domaines de tétramérisation. Cette multimérisation a pour effet d'augmenter l'affinité ADN-protéine, par rapport à l'affinité du domaine de liaison à l'ADN isolé. De plus, un changement de conformation de la protéine, au niveau de l'interface avec l'ADN, permet d'assurer la spécificité de la liaison : la protéine peut lier des séquences spécifiques et non spécifiques, mais seule la liaison avec les bonnes séquences persiste (Petty *et al.*, 2011 ; Emamzadah *et al.*, 2011).



**Figure 3 | Exemple de p53, un régulateur multi-domaines oligomérique.**

Structure du domaine de liaison à l'ADN et domaine de tétramérisation liés à l'ADN sous forme de tétramère (Petty *et al.*, 2011 ; PDB 3Q05). Le *linker* entre les deux domaines a été raccourci et le domaine de tétramérisation a été converti en domaine de dimérisation pour permettre la cristallisation. Le domaine de liaison à l'ADN adopte un repliement *Immunoglobulin-like* fait de feuillets  $\beta$ . Les couleurs suivantes correspondent au domaine de liaison à l'ADN/domaine de tétramérisation pour chacun des quatre monomères : jaune/vert foncé ; rose/bleu foncé ; orange/bleu clair ; rouge/vert clair. Une rotation de 90° a été appliquée à l'image de gauche pour obtenir celle de droite.

Cet exemple permet de soulever les questions d'affinité et de spécificité, qui sont des facteurs essentiels pour la liaison ADN-protéine. Dans le cas de p53, qui est un régulateur majeur, ils sont influencés par un domaine d'oligomérisation et par un changement de conformation.

- **Les motifs de liaison à l'ADN combinant hélices  $\alpha$  et feuillets  $\beta$**

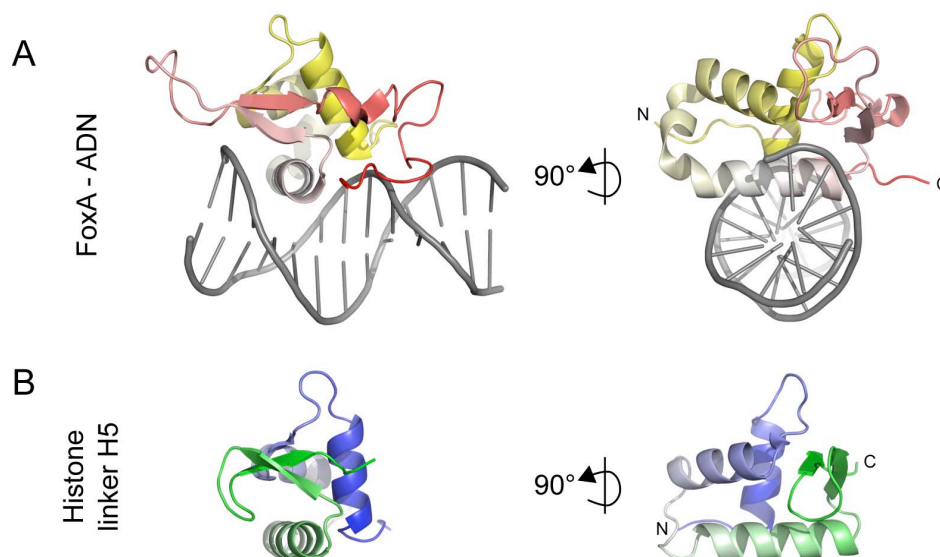
De nombreux domaines de liaison combinent ces deux types de motifs, tel que le *zinc finger* (doigt de zinc), petit motif très répandu chez les FTs. Il est composé d'une hélice, de deux

feuilletés anti-parallèles, et d'un ion  $Zn^{2+}$  coordonné par deux cystéines et deux histidines. On trouve généralement plusieurs *zinc fingers* dans une même protéine, ils se lient dans le grand sillon de l'ADN à 3 paires de bases (pb) d'intervalle.

Le motif *winged helix turn helix* (wHTH), également appelé HNF3/forkhead (HFH) d'après le nom des familles de protéines dans lesquelles il a été découvert, est formé d'un motif HTH suivi de deux à quatre feuilletés  $\beta$  (**Figure 4**, exemple 3). La boucle séparant les deux premiers feuilletés  $\beta$  est généralement assez étendue et forme une « aile » (*wing*) (Gajiwala et Burley, 2000 ; Aravind *et al.*, 2005).

### Exemple 3 : FoxA

FoxA intervient dans le développement embryonnaire chez les animaux, il lie l'ADN via un motif wHTH (**Figure 4A**). C'est un FT dit « pionnier », capable de défier la chromatine condensée. Il peut lier de l'ADN enroulé par des nucléosomes, déplacer des histones *linker*, et décompacter la chromatine (Taube *et al.*, 2010 ; Cirillo *et al.*, 2002). Il a été proposé que ces propriétés de liaison à la chromatine reposent sur la similarité structurale entre son motif wHTH et celui de l'histone *linker* H5 (**Figure 4B**) (Clark *et al.*, 1993 ; Ramakrishnan *et al.*, 1993). La liaison de FoxA à la chromatine a été mise en évidence, mais les mécanismes moléculaires impliqués dans cette liaison restent encore peu compris.



**Figure 4 | Exemple de FoxA, un facteur pionnier qui ressemble à un histone *linker*.**

**A.** Structure du domaine wHTH de FoxA sur l'ADN, deux vues orthogonales (Clark *et al.*, 1993 ; PDB 1VTN).

**B.** Structure du domaine wHTH de l'histone *linker* H5, deux vues orthogonales (Ramakrishnan *et al.*, 1993 ; PDB 1HST).

FoxA permet d'introduire la notion de facteur pionnier, une propriété nécessaire à l'initiation des phénomènes de régulation transcriptionnelle : lorsque la chromatine est condensée et

défavorable à la fixation des FTs, les facteurs pionniers sont les premiers à accéder aux régions régulatrices et permettent la mise en place d'un environnement favorable à la liaison d'autres facteurs.

### III. LEAFY, régulateur clé du développement floral

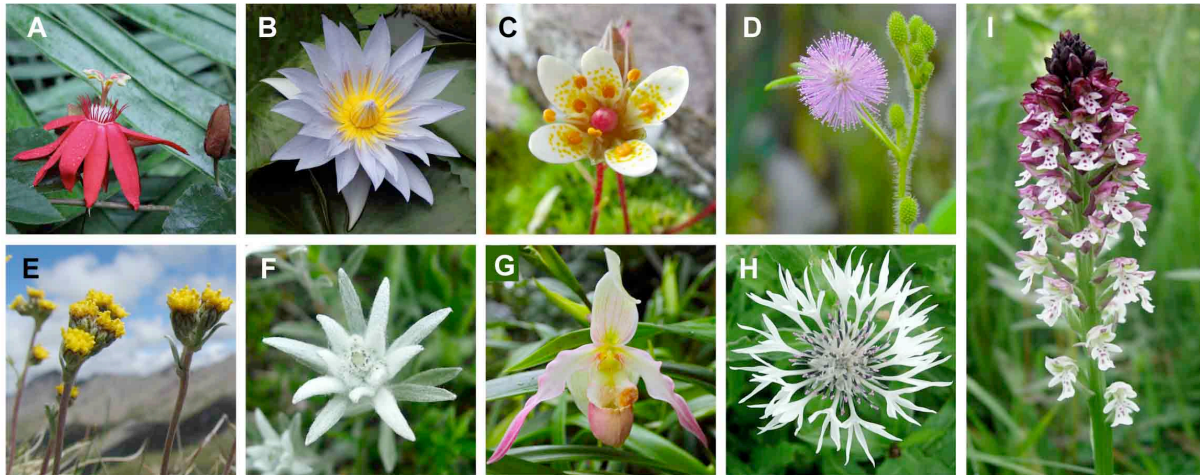
Nous nous intéressons ici au facteur de transcription LEAFY (LFY), impliqué dans des processus développementaux chez les plantes, en particulier la formation des fleurs. Son rôle, ce qu'on sait sur le fonctionnement de la protéine, et sa place dans l'évolution sont expliqués dans cette partie.

#### 1. La floraison, un processus crucial et fascinant

Chez les Angiospermes, la floraison est une étape cruciale car elle permet la reproduction sexuée. Le maintien et la propagation de la plupart des espèces en dépendent. Il est essentiel que le moment de la floraison soit contrôlé pour que la phase reproductrice se déroule dans des conditions environnementales favorables. Les fleurs comptent parmi les créations les plus spectaculaires de l'évolution et ont joué un rôle prépondérant dans le succès évolutif des angiospermes. Elles attirent les pollinisateurs et produisent les fruits renfermant les graines. La diversité morphologique des fleurs est particulièrement remarquable (**Figure 5**), avec une multitude de couleurs, de formes, de tailles et de parfums. Outre leur attrait esthétique, les plantes à fleurs fournissent à l'homme de quoi se nourrir, se vêtir, se soigner et s'abriter.

La transition florale est le passage du stade végétatif au stade reproducteur. Pendant le développement végétatif, le méristème apical, ensemble de cellules indifférenciées, permet la formation de feuilles et de tiges secondaires. Il est converti en méristème d'inflorescence qui produit des méristèmes floraux permettant la formation de fleurs. La transition florale est contrôlée par des facteurs environnementaux et des signaux endogènes (Amasino, 2010).

L'étude de mutants affectés pour la floraison chez la plante modèle *A. thaliana*, mais aussi chez d'autres espèces, a ouvert la voie à la compréhension des mécanismes impliqués dans ce processus (Meyerowitz, 1987; Bowman *et al.*, 1989). Le mutant pour le gène *LFY* est l'un des plus remarquables.



**Figure 5 | La diversité morphologique des fleurs.**

**A.** Passiflore rouge (*Passiflora* sp., *Passifloraceae*). **B.** Lotus bleu (*Nymphaea caerulea*, *Nymphaeaceae*). **C.** Saxifrage faux bryum (*Saxifraga bryoides*, *Saxifragaceae*). **D.** Mimosa pudique (*Mimosa pudica*, *Mimosaceae*) **E.** Génépi des glaciers (*Artemisia glacialis*, *Asteraceae*). **F.** Edelweiss (*Leontopodium nivale* subsp. *Alpinum*, *Asteraceae*). **G.** Sabot de Vénus (*Cypripedium* sp., *Orchidaceae*). **H.** Bleuets blancs (*Centaurea montana* 'alba', *Asteraceae*). **I.** Orchis brûlée (*Neotinea ustulata*, *Orchidaceae*).

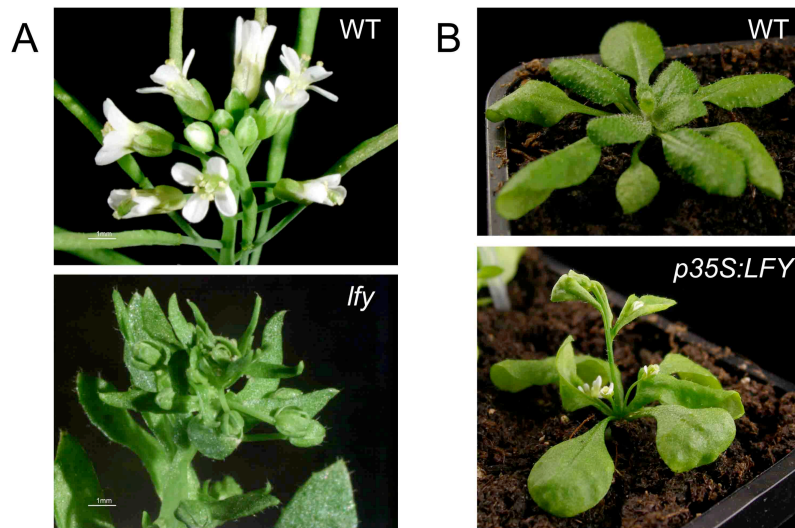
## 2. *LFY*, chef d'orchestre du développement floral

### ▪ *LFY* est nécessaire à la présence de fleurs

D'abord identifié chez le muflier *Antirrhinum majus* (Carpenter et Coen, 1990 ; Coen *et al.*, 1990) puis chez *Arabidopsis* (Huala et Sussex, 1992 ; Schultz et Haughn, 1991 ; Weigel *et al.*, 1992), le mutant *lfy* est caractérisé par l'absence de fleur fertile. Un phénotype semblable est observé chez de nombreuses espèces (*Petunia hybrida*, *Zea mays*, *Oryza sativa*, *Pisum sativum* par exemple) (Moyroud *et al.*, 2010).

Chez *Arabidopsis*, des tiges se forment à la place des premières fleurs chez le mutant *lfy*, et des structures intermédiaires fleurs/tiges remplacent les fleurs tardives (Huala et Sussex, 1992 ; Schultz et Haughn, 1991 ; Weigel *et al.*, 1992) (**Figure 6**). À l'inverse, l'expression constitutive de *LFY* entraîne une floraison précoce et la formation de fleurs ectopiques, directement à la base de la rosette (la rosette désigne l'ensemble des feuilles partant de la base de la plante) (Weigel et Nilsson, 1995) (**Figure 6**). *LFY*, qui code pour un FT (Parcy *et al.*, 1998), est donc nécessaire et suffisant pour transformer les méristèmes végétatifs et les méristèmes d'inflorescence en méristèmes floraux.

On peut décomposer la fonction de *LFY* en deux étapes : il intervient dans l'établissement du méristème floral puis dans la formation des organes floraux (**Figure 7**).



**Figure 6 | LFY est nécessaire et suffisant à la formation de fleurs.**

**A.** Inflorescence d'une plante sauvage (WT) et d'un mutant *lfy* chez *A. thaliana*.

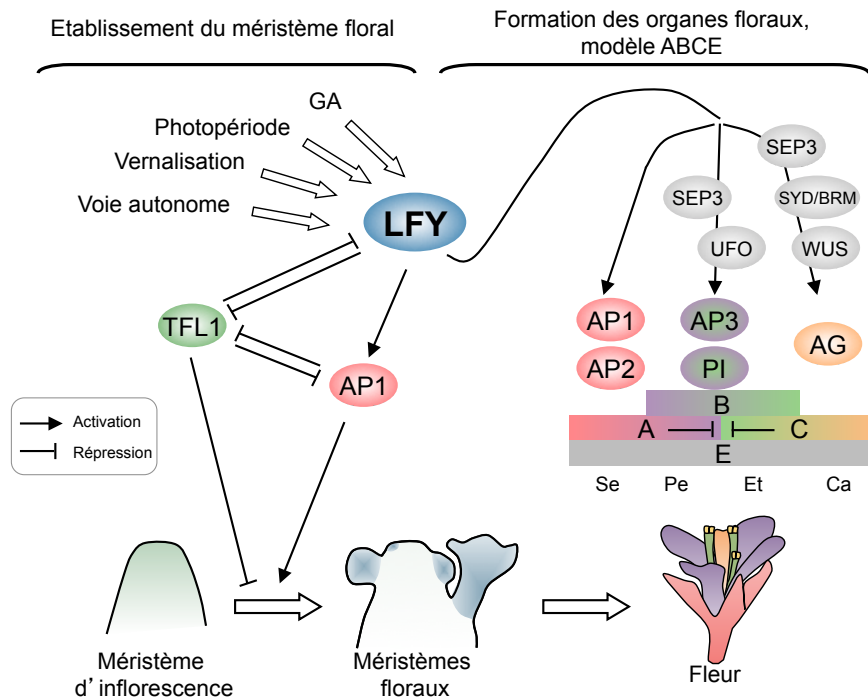
**B.** Plante WT avant montaison de la hampe florale, et plante transgénique exprimant *LFY* de manière constitutive (*p35S:LFY*) au même stade de développement.

#### ▪ L'établissement du méristème floral

*LFY* est exprimé très précocement dans la zone de formation des méristèmes floraux en réponse aux signaux qui influencent la floraison : la lumière (photopériode), la température (vernalisation), l'âge de la plante (voie autonome), les signaux hormonaux (gibbérélines). Comme *LFY* centralise ces différents signaux, il est qualifié « d'intégrateur floral » (Simpson et Dean, 2002 ; Irish, 2010 ; Parcy, 2005) (**Figure 7**). *LFY*, ainsi que le gène *APETALA1* (*AP1*), spécifient l'identité des méristèmes floraux. L'expression d'*AP1* est directement activée par *LFY*, mais est aussi induite de manière indépendante (Parcy *et al.*, 1998 ; Wagner et Meyerowitz, 2002 ; Wagner *et al.*, 2004 ; Ruiz-Garcia *et al.*, 1997). *LFY* et *AP1* sont antagonistes au gène *TERMINAL FLOWER1* (*TFL1*) qui spécifie l'identité végétative et inflorescentielle. *TFL1* est directement réprimé par *LFY*, ainsi que par *AP1* (Liljegren *et al.*, 1999 ; Parcy *et al.*, 2002 ; Ratcliffe *et al.*, 1998 ; Jaeger *et al.*, 2013 ; Kaufmann *et al.*, 2010). Le rôle de *LFY* pour l'établissement du méristème floral est décrit dans de nombreuses revues, dont Irish, 2010 ; Chandler, 2012 ; Wellmer et Riechmann, 2010 ; Moyroud *et al.*, 2009 ; Parcy, 2005 ; Krizek et Fletcher, 2005.

Un rôle pour *LFY* dans l'émergence des méristèmes floraux a été mis en évidence récemment. Ces méristèmes émergent d'un méristème d'inflorescence à des points où la concentration en auxine est maximale. Ce pic d'auxine active l'expression de *LFY* qui régule à son tour la voie de signalisation de l'auxine (Yamaguchi *et al.*, 2013 ; Li *et al.*, 2013). D'autre part, nous avons montré dans l'équipe que *LFY* permettait l'émergence des

méristèmes via l'activation du gène *REGULATOR OF AXILLARY MERISTEMS1 (RAX1)*, connu pour contribuer au développement méristématique (Chahtane *et al.*, 2013, voir l'Annexe 4).



**Figure 7 | LFY, le régulateur clé du développement floral.**

Schéma simplifié du réseau floral de LFY (adapté à partir de Parcy, 2005 et Irish, 2010). *LFY* intègre les signaux de floraison (photopériode, vernalisation, voie autonome, GA : gibbérélines) et induit la formation de méristèmes floraux à partir du méristème d'inflorescence via des interactions avec *AP1* et *TFL1*. Dans un second temps, *LFY*, avec l'aide de corégulateurs (en gris), induit les gènes ABCE qui déterminent l'identité des organes floraux (Se : sépales, Pe : pétales, Et : étamines, Ca : carpelles).

#### ▪ La formation des organes floraux

L'expression de *LFY* persiste ensuite dans le méristème floral, où il régule les gènes impliqués dans la formation des organes floraux (**Figure 7**) : les sépales (qui enveloppent et protègent le bourgeon floral), les pétales (souvent colorés, ils attirent les pollinisateurs), les étamines (organes mâles qui portent le pollen) et les carpelles (organes femelles qui contiennent les ovules). Ces organes sont répartis selon un patron très précis organisé en verticilles (de manière circulaire autour d'un axe).

La mise en place de l'identité des organes floraux a été expliquée par le modèle ABC, ensuite étendu au modèle ABCE (Bowman et Meyerowitz, 1991 ; Coen et Meyerowitz, 1991 ; Theissen, 2001 ; Causier *et al.*, 2010 ; Smaczniak *et al.*, 2012a ; Irish, 2010). L'activité des gènes de classe A conduit à la formation de sépales dans le premier verticille. L'activité

combinée des gènes A et B dans le second verticille permet la formation de pétales. De même, la combinaison des activités B et C conduit à la formation des étamines dans le troisième verticille et les gènes C seuls permettent le développement de carpelles au centre de la fleur (**Figure 7**). Les gènes E sont nécessaires à la formation de tous les organes. Chez *Arabidopsis*, la fonction A est assurée par *AP1* et *AP2*, les gènes B sont *AP3* et *PISTILLATA (PI)*, la fonction C est assurée par *AGAMOUS (AG)* et les gènes *SEPALLATA1-4 (SEP1-4)* assurent la fonction E. A l'exception d'*AP2*, tous les gènes du modèle ABCE appartiennent à la grande famille des gènes à boîte MADS. Les protéines correspondantes sont des FTs capables de s'assembler en complexes tétramériques selon le modèle du « *floral quartet* ». En fonction de leur composition, ces complexes permettent d'orienter la différenciation des organes floraux (Theissen, 2001 ; Theissen et Saedler, 2001 ; Smaczniak *et al.*, 2012b ; Smaczniak *et al.*, 2012a).

*LFY* active directement l'expression d'*AP3*, *PI* et *AG* (Busch *et al.*, 1999 ; Lohmann *et al.*, 2001 ; Lamb *et al.*, 2002). Comme *LFY* est exprimé dans l'ensemble du méristème floral, des co-régulateurs sont nécessaires pour l'activation région-spécifique des gènes ABC. Ainsi, *LFY* interagit avec *SEP3* et *UNUSUAL FLORAL ORGANS (UFO)* pour activer *AP3* (Liu *et al.*, 2009 ; Levin et Meyerowitz, 1995 ; Chae *et al.*, 2008 ; Souer *et al.*, 2008 ; Parcy *et al.*, 1998) et avec *SEP3* et *WUSCHEL (WUS)* pour activer *AG* (Lohmann *et al.*, 2001) (**Figure 7**).

Des études d'expression à grande échelle par puces à ADN (*microarray*) et l'analyse des régions liées par *LFY in planta* par immuno-précipitation de la chromatine et séquence (ChIP-seq) ou hybridation sur puce (ChIP-Chip) ont montré que *LFY* était à la tête d'un vaste réseau et confirmé un grand nombre de ses gènes cibles (Schmid *et al.*, 2003 ; Moyroud *et al.*, 2011 ; Winter *et al.*, 2011).

Enfin, il a été montré que *LFY* était impliqué dans des processus de remodelage chromatinien pour la régulation des gènes B et C, *AP3* et *AG* (Wu *et al.*, 2012). Pendant le développement végétatif, l'activité de ces gènes est réprimée par les complexes *Polycomb* (PRC2 et PRC1). PRC2 dépose des marques répressives sur les histones et PRC1 compacte la chromatine et stabilise l'état réprimé du locus. Au cours du développement floral, *LFY* réprime directement l'expression de *EMBRYONIC FLOWER1 (EMF1)*, qui code pour une sous-unité de PRC1 (Chen *et al.*, 1997 ; Chou *et al.*, 2001 ; Wu *et al.*, 2012 ; Calonje *et al.*, 2008).

Dans les méristèmes floraux, la répression des gènes B et C par les complexes *Polycomb* est contrée par des protéines du groupe trithorax telles que *HOMOLOG OF TRITHORAX1 (ATX1)*, qui transfère des groupements méthyl sur l'histone H3), *PICKLE (PKL)*, remodeleur à

chromo-domaine) et *ULTRAPETALA1* (*ULT1*, protéine à domaine SAND) (Goodrich *et al.*, 1997 ; Alvarez-Venegas *et al.*, 2003 ; Katz *et al.*, 2004 ; Saleh *et al.*, 2007 ; Carles et Fletcher, 2009 ; Carles et Fletcher, 2010). Enfin, il a été montré récemment que LFY recrutait les remodeleurs chromatiniens SPLAYED (SYD) et BRAHMA (BRM) aux loci *AP3* et *AG* (Wu *et al.*, 2012). *SEP3* semble également intervenir dans ce processus. SYD et BRM sont des ATPases de type SWI2/SNF2 qui utilisent l'énergie de l'hydrolyse de l'ATP pour modifier les interactions nucléosome-ADN (Clapier et Cairns, 2009). Le recrutement de SYD et BRM par LFY est probablement direct car ces protéines interagissent ensemble *in vitro* (Wu *et al.*, 2012).

### 3. La protéine LFY

#### ▪ Un facteur de transcription unique

Le gène *LFY* code pour un FT unique dans le règne végétal, il ne ressemble à aucune autre protéine connue. Cette protéine d'environ 400 acides aminés possède deux domaines conservés (**Figure 8**) : le domaine de liaison à l'ADN en position C-terminale (environ 200 acides aminés) et un second domaine de nature peu connue en position N-terminale (environ 75 acides aminés). Le domaine N-terminal est précédé d'environ 40 acides aminés non conservés, et un *linker* peu conservé le relie au domaine de liaison à l'ADN. Un signal de localisation nucléaire se trouve juste avant le domaine de liaison à l'ADN. Tous les allèles *lfy* identifiés chez les plantes portent des mutations dans ces deux domaines (Maizel *et al.*, 2005).

La protéine ne possède pas de pouvoir intrinsèque pour réguler la transcription et sollicite des partenaires pour influencer sur l'expression de ces gènes cibles (Parcy *et al.*, 1998 ; Lohmann *et al.*, 2001).

#### ▪ Structure du domaine de liaison à l'ADN

La structure cristallographique du domaine de liaison à l'ADN (LFY-C) en complexe avec l'ADN a été obtenue en 2008 par l'équipe, en collaboration avec l'équipe de cristallographie de C. Müller (Hamès *et al.*, 2008) (**Figure 9**).

LFY-C a été cristallisé en complexe avec un oligonucléotide contenant un site de liaison issu de la région régulatrice d'*AP1*, et un autre provenant d'*AG*. La séquence de chacun de ces oligonucléotides est pseudo-palindromique, formée de deux demi-sites inversés. LFY-C contacte l'ADN sous forme de dimère, chaque monomère étant lié à un demi-site.



	1	10	20	30	40	50
<i>A. thaliana</i>	MDPEGE	TSGLFRWN	NP . . . . . TR . . . . .	ALVQAPP	VPVPP	ELQQQPVTPTAA
<i>P. hybrida</i>	MDPEAF	SASLFWKDP	RG . . . . . AMPPPNR	. . . . .	LLAVAPP	QPPPLPPPPQPLPPAYSIR
<i>A. trichopoda</i>	MDPDPF	SVTLFKWD	HHHHRPAPP	PNR . . . . .	LLQEATATV	TFFP . . . . . PQTFA
<i>O. sativa</i>	. . . . .	. . . . .	MDPNDAF	SAAHPFR	. . . . .	WDLGPPAPAPVFP
<i>G. biloba</i>	MDPETF	PAAFFKWD	QRSSAAAAPI	HR . . . . .	GLEGQRFV	EFVPTTTNEAL . . . . .
<i>P. radiata</i>	MDPESE	SAAFFKWD	QRPPALAPP	QMORSAGLEA	QRFHDF	GVPNAAAAMA . . . . .
<i>W. mirabilis</i>	MAPESE	SAAFFKWE	QR . . . . .	SSIPLQMQ	ORPGN . . . . .	IFQDFGMPN
<i>C. richardii</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
<i>A. lygodifolia</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
<i>P. patens</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .

	60	70	80	90	100	110	120
<i>A. thaliana</i>	PYGIRFY	TAAKIAEL	GFTASTLV	GKDEE	LEEMMNS	LSHIFRWE	LLVGERY
<i>P. hybrida</i>	AYGIRY	TAAKITEL	GFTVNTLL	DMKDD	ELDDMMNS	LSQIFRWE	LLVGERY
<i>A. trichopoda</i>	QYGVRY	TVAKMVE	GFTVNTLV	NMREE	EIDEMMGS	LSIEFRWE	LLMGERY
<i>O. sativa</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
<i>G. biloba</i>	HYGVRY	TVARISEL	GFTASTLL	AMTER	ELDDMMAA	LAGLFRWD	LLGERF
<i>P. radiata</i>	NYGVRY	TLTKMVD	MGFTVNTLV	NMTEQ	ELDDVIRT	LVDIYRVD	LLVGEKY
<i>W. mirabilis</i>	NYGVRY	TLTKMVD	MGFTVNTLV	NMTEQ	ELDDVIRT	LVDIYRVE	LLVGEKY
<i>C. richardii</i>	DYGVRY	TLTKMVE	MGFTVNTLV	NMTEQ	ELDDVIRT	LVDIYRVE	LLVGEKY
<i>A. lygodifolia</i>	LYGVRY	TLAKMSE	MGFTVNTLV	NMTEQ	ELEDLVMT	VLESCQV	LLVGEKY
<i>P. patens</i>	DFGVQES	TVAVMGQ	MGFTLS	TLINMK	DEVFVI	KSMIEEYHLD	LLMGEQ

	130	140	150	160	170	180	190	200
<i>A. thaliana</i>	RRHLLS	AAGDSG	.THHALD	ALSQED	DWTGL	LSEEPV	QQQD	QTDAA
<i>P. hybrida</i>	RRHIL . . .	SDGG . TN	.VLDALS	QSE . . .	GLSEEP	VQQE	REAA	AGSGGG
<i>A. trichopoda</i>	QHHQLLV	GGGG	EVNHHLD	ALSQE . . .	GLSEEP	VQE .	EKEA	AGSGGG
<i>O. sativa</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
<i>G. biloba</i>	RHHGHQ	SGSTVD	GASQEV	LSDEHDM	MAGSGG	MGDDN	GRRM	VTKGKQ
<i>P. radiata</i>	LD . LFVD	VDGKRR	KADENAL	DTLSQE	GLS .	EEPQGD	NAI	LSQN .
<i>W. mirabilis</i>	MEQLFV	DVDGKR	KIDENAL	DTLSQE	GLSVEE	PPQGD	NAI	LSQNN
<i>C. richardii</i>	LQELFN	DVDGKR	KIDENAL	DTLSQE	GLSVEE	PPHGD	NTIL	LSQNN
<i>A. lygodifolia</i>	RLQLLAT	SSKHK	SDSGM	VVTSLE	GT .	REQRGD	NVMM	FPEA .
<i>P. patens</i>	RDLMP	SKAARK	KIDEP	AALLAP	KEGVA .	KESH	RDDS	IILTD

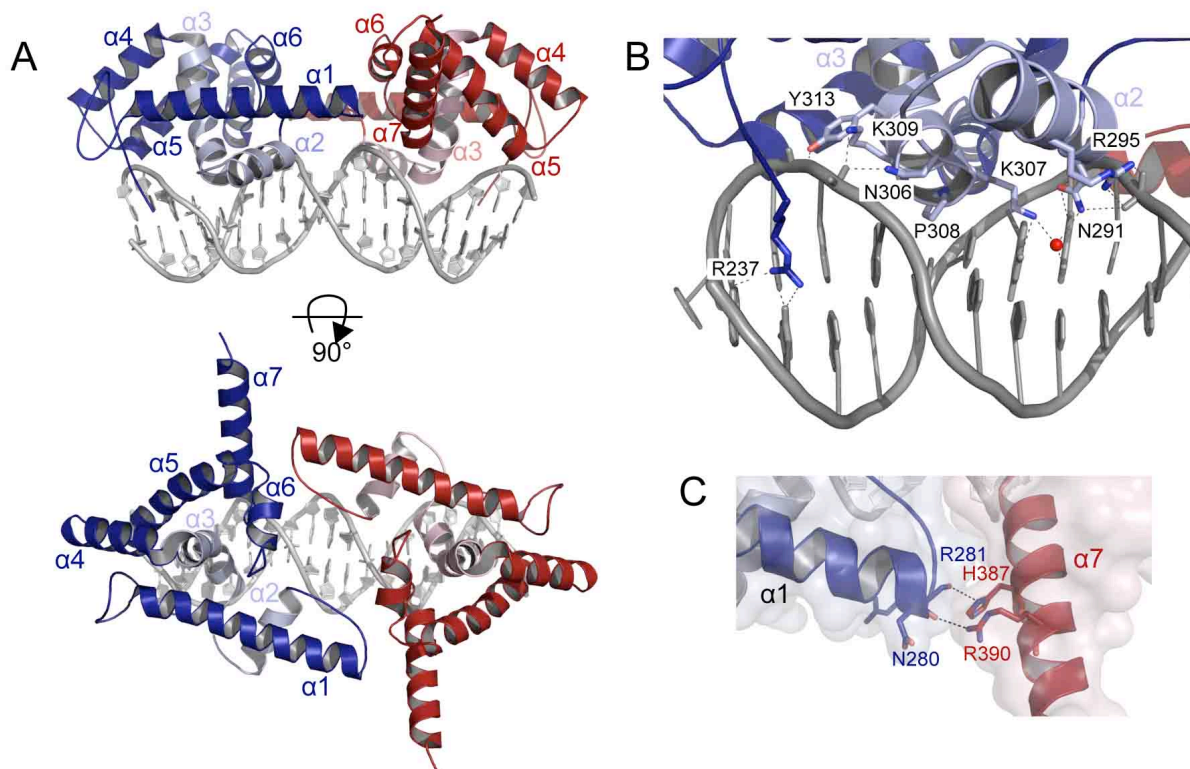
	210	220	230	240	250	260	270
<i>A. thaliana</i>	PMLT	SVETDED	VNEG .	EDDDG	MDNGN	GGSG	GLGTER
<i>P. hybrida</i>	RRGS	SMEDED	TEEGQ	EDN	YINNE	GGGG	ISER
<i>A. trichopoda</i>	. .	ATKGN	VGGDA	E . . . . .	. . . . .	. . . . .	. . . . .
<i>O. sativa</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
<i>G. biloba</i>	QLED	MDCC	DEGGG	SESTE . . . . .	. . . . .	. . . . .	. . . . .
<i>P. radiata</i>	MGLP	DNNY	TNEQ	MKAS .	KKQKR	RSREL	GEDGED
<i>W. mirabilis</i>	IGMP	DTNY	GSEQ	TAC .	KKQKR	RSKDS	GEDGED
<i>C. richardii</i>	AGFP	ENNF	DLHG	KTC .	KKQKR	RLKES	GEEGED
<i>A. lygodifolia</i>	LALPE	PSD	NEG	HKLP	RKKPK	RRLREP	GEDGDD
<i>P. patens</i>	ESQ	SDFG	SDSE	GRKFS	KKQKR	RP	SMET

	280	290	300	310	320	330	340	350
<i>A. thaliana</i>	I AKDR	GEK	CP	KVTN	QVFR	YAKKS	GASY	INPK
<i>P. hybrida</i>	I AKER	GEK	CP	KVTN	QVFR	YAKKA	GASY	INPK
<i>A. trichopoda</i>	L AKER	GEK	CP	KVTN	QVFR	HAKR	AGASH	INPK
<i>O. sativa</i>	MAK	LHG	HK	SP	KVTN	QVFR	YAKKV	GASY
<i>G. biloba</i>	I AKER	GEK	CP	KVTN	QVFR	HAKHS	GAGY	INPK
<i>P. radiata</i>	I AKER	GEK	CP	KVTN	QVFR	HAKHS	GAGY	INPK
<i>W. mirabilis</i>	I AKER	GEK	CP	KVTN	QVFR	HAKYS	GASY	INPK
<i>C. richardii</i>	M ARER	GEK	CP	KVTN	QVFR	HAKL	GASY	INPK
<i>A. lygodifolia</i>	I SKER	GEK	CP	KVTN	QVFR	HAKHN	GCGY	INPK
<i>P. patens</i>	I AKER	GEK	CP	KVTN	QVFR	HAKL	TGAGY	INPK

	360	370	380	390	400	410	420
<i>A. thaliana</i>	VNIACRH .	GWDIDAV	FN	AHPRLS	IWYVPT	KLRLQ	CHLERN
<i>P. hybrida</i>	VAI AARQ .	GWDIDAI	F	NGHPR	LS	IWYVPT	KLRLQ
<i>A. trichopoda</i>	VAMS	RHHHAW	DVDAL	FASHP	SLAIWYVPT	KLRLQ	CHADRS
<i>O. sativa</i>	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
<i>G. biloba</i>	VDIS	ARH .	GFDIDAV	FAAHP	RLAIWYVPT	TR	KLRLQ
<i>P. radiata</i>	VVM	AKEN .	GWDIEG	VFNQ	HEKLR	IWYVPT	KLRLQ
<i>W. mirabilis</i>	VAM	AKDN .	GWDIEG	VFNK	HEKLR	IWYVPT	KLRLQ
<i>C. richardii</i>	VAK	ARN .	GWDIDG	VFNK	HEKLR	IWYVPT	KLRLQ
<i>A. lygodifolia</i>	VD	MARDN .	GWDIEG	VFNK	HEKLR	IWYVPT	KLRLQ
<i>P. patens</i>	VKL	ARQN .	EW	DIDDL	FN	RNDK	LR

### Figure 8 | Alignement de séquences protéiques d'homologues de LFY chez les plantes terrestres.

Les séquences proviennent d'espèces chez les angiospermes (les dicotylédones *A. thaliana* et *Petunia x hybrida*, la dicotylédone basale *Amborella trichopoda*, la monocotylédone *Oriza sativa*), les gymnospermes (*Ginkgo biloba*, *Pinus radiata*, *Welwitschia mirabilis*), les fougères (*Ceratopteris richardii*, *Angiopteris lygodiifolia*) et chez la mousse *Physcomitrella patens* (PpLFY1). La numérotation réfère à la séquence de la protéine d'*A. thaliana*. Le domaine N-terminal et le domaine de liaison à l'ADN sont soulignés en rose et orange respectivement. La structure secondaire du domaine de liaison à l'ADN, déduite de sa structure cristallographique, est indiquée,  $\alpha$  : hélice  $\alpha$ , HTH indique les hélices  $\alpha_2$  et  $\alpha_3$  du motif *helix turn helix*. Les résidus qui contactent le squelette phosphate et les bases de l'ADN sont indiqués par des points verts clairs et verts foncés respectivement. Les résidus impliqués dans la dimérisation du domaine de liaison à l'ADN sont notés par des rectangles violets. Les positions supposées influencer la liaison à l'ADN chez PpLFY1 sont indiquées par des triangles rouges.



### Figure 9 | Structure cristallographique du domaine de liaison à l'ADN de LFY.

**A.** Dimère de LFY-C sur un oligonucléotide comportant un site de liaison pour LFY (Hamès *et al.*, 2008 ; PDB 2VY2). Un monomère est en bleu, l'autre en rouge, les hélices  $\alpha_2$  et  $\alpha_3$  du motif HTH sont plus claires dans chaque monomère. Une rotation de  $90^\circ$  a été appliquée à l'image du haut pour obtenir celle du bas.

**B.** Vue rapprochée de l'interface ADN-protéine pour l'un des monomères. Les chaînes latérales des principaux acides aminés contactant le squelette phosphate ou les bases sont montrées. Les contacts sont représentés par des pointillés.

**C.** Vue rapprochée de l'interface de dimérisation. Les résidus impliqués dans l'interaction sont montrés.

La structure a révélé un repliement original à sept hélices  $\alpha$  ( $\alpha$ 1-7). Les hélices  $\alpha$ 1,  $\alpha$ 2 et  $\alpha$ 3 forment un motif *helix turn helix* (HTH), avec  $\alpha$ 3 qui plonge dans le grand sillon de l'ADN (**Figure 9A**, **Figure 8**). La plupart des contacts avec l'ADN sont apportés par les hélices  $\alpha$ 2 et  $\alpha$ 3 du motif HTH, mais les autres hélices participent aussi, ainsi que la boucle située en N-terminal de  $\alpha$ 1 qui interagit avec le petit sillon. LFY-C contacte à la fois le squelette phosphate de l'ADN et des bases nucléotidiques, assurant ainsi la spécificité de liaison (**Figure 9B**).

Hamès *et al.* ont également montré que LFY-C était monomérique en solution et que la dimérisation avait lieu au contact de l'ADN. Cette dimérisation est coopérative : la fixation d'un premier monomère facilite la liaison du second. Elle est attribuée à l'interaction existant entre les deux monomères via une histidine (H387) et une arginine (R390) sur  $\alpha$ 7 qui contactent la chaîne principale de deux autres résidus sur  $\alpha$ 1 (**Figure 9C**).

#### ▪ Spécificité de LFY pour la séquence d'ADN

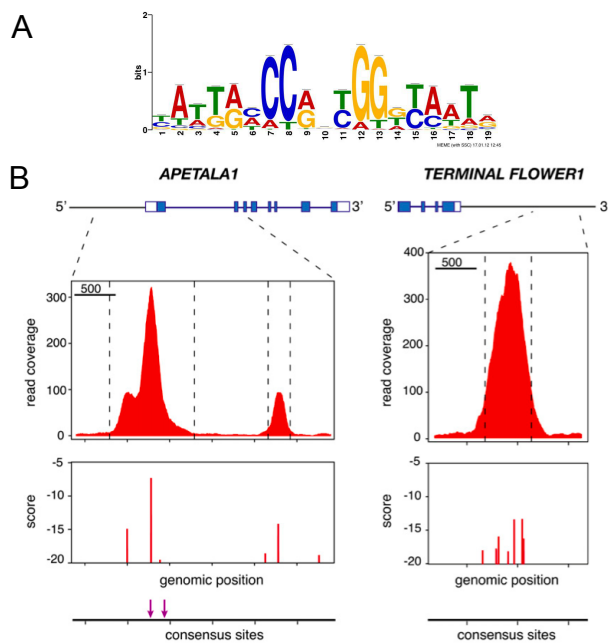
Les données structurales ainsi que l'alignement de 3 sites de liaison de LFY présents dans les gènes *AP1* et *AG* ont d'abord permis de proposer un consensus pour la spécificité de liaison à l'ADN de LFY : (T/A)NNNNCCANTG(T/G)NNNN(T/A) (Busch *et al.*, 1999 ; Hamès *et al.*, 2008).

Depuis, des analyses beaucoup plus exhaustives *in vitro* par *Systematic Evolution of Ligands by EXponential enrichment* (SELEX) ont permis la détermination précise du motif de liaison de LFY (Moyroud *et al.*, 2011). Cette technique, utilisée dans l'équipe, consiste à mettre la protéine en contact avec des oligonucléotides de séquences aléatoires. Les complexes ADN-protéine sont isolés et l'ADN est amplifié par PCR. Cette procédure est réalisée plusieurs fois et permet l'enrichissement en oligonucléotides de bonne affinité pour la protéine. Les oligonucléotides sélectionnés sont séquencés puis alignés, ce qui permet la construction d'une matrice poids-position qui donne la probabilité de présence de chacun des quatre nucléotides à chaque position de la séquence. Cette matrice peut être convertie en logo afin de visualiser le motif de liaison (**Figure 10A**) (Stoltenburg *et al.*, 2007 ; Djordjevic, 2007 ; Stormo, 2000 ; Bulyk, 2003 ; Wasserman et Sandelin, 2004 ; Moyroud *et al.*, 2011).

LFY reconnaît donc un motif pseudo-palindromique de 19 pb (**Figure 10A**) (Moyroud *et al.*, 2011), avec notamment 4 positions dont le contenu en information est particulièrement élevé (2 positions par demi-site, numérotées 7, 8, 12 et 13 sur la **Figure 10A**) et qui coïncident avec des bases contactées de manière spécifique par la protéine dans la structure cristallographique (en particulier par les résidus K307 et N291, **Figure 9B**).

L'alignement des régions liées par la protéine *in vivo* sur le génome (en ChIP-Seq et ChIP-chip) a permis d'obtenir un motif de liaison comparable (Moyroud *et al.*, 2011 ; Winter *et al.*, 2011).

Un modèle décrivant la spécificité de LFY a d'ailleurs été établi dans l'équipe, basé sur la matrice obtenue en SELEX. Ce modèle permet de prédire efficacement les régions liées par LFY sur le génome (Moyroud *et al.*, 2011) (**Figure 10B**).



**Figure 10 | Spécificité de liaison à l'ADN de LFY.**

**A.** Logo représentant le motif de 19 pb reconnu par le domaine de liaison à l'ADN en SELEX. La hauteur d'une base traduit la probabilité qu'elle soit présente à chaque position (Moyroud *et al.*, 2011).

**B.** Profil de liaison de LFY sur le génome et corrélation avec les sites de liaison prédits. Deux régions, issues des loci *AP1* et *TFL1*, sont montrées en exemple. Les pics rouges montrent l'occupation par LFY observée en ChIP-seq (*read coverage*). Les sites de liaison prédits à l'aide du motif montré en A sont représentés par les barres (plus le score est proche de zéro, plus l'affinité prédite est grande), et les sites prédits par l'ancien consensus sont montrés par les flèches en bas (Moyroud *et al.*, 2011).

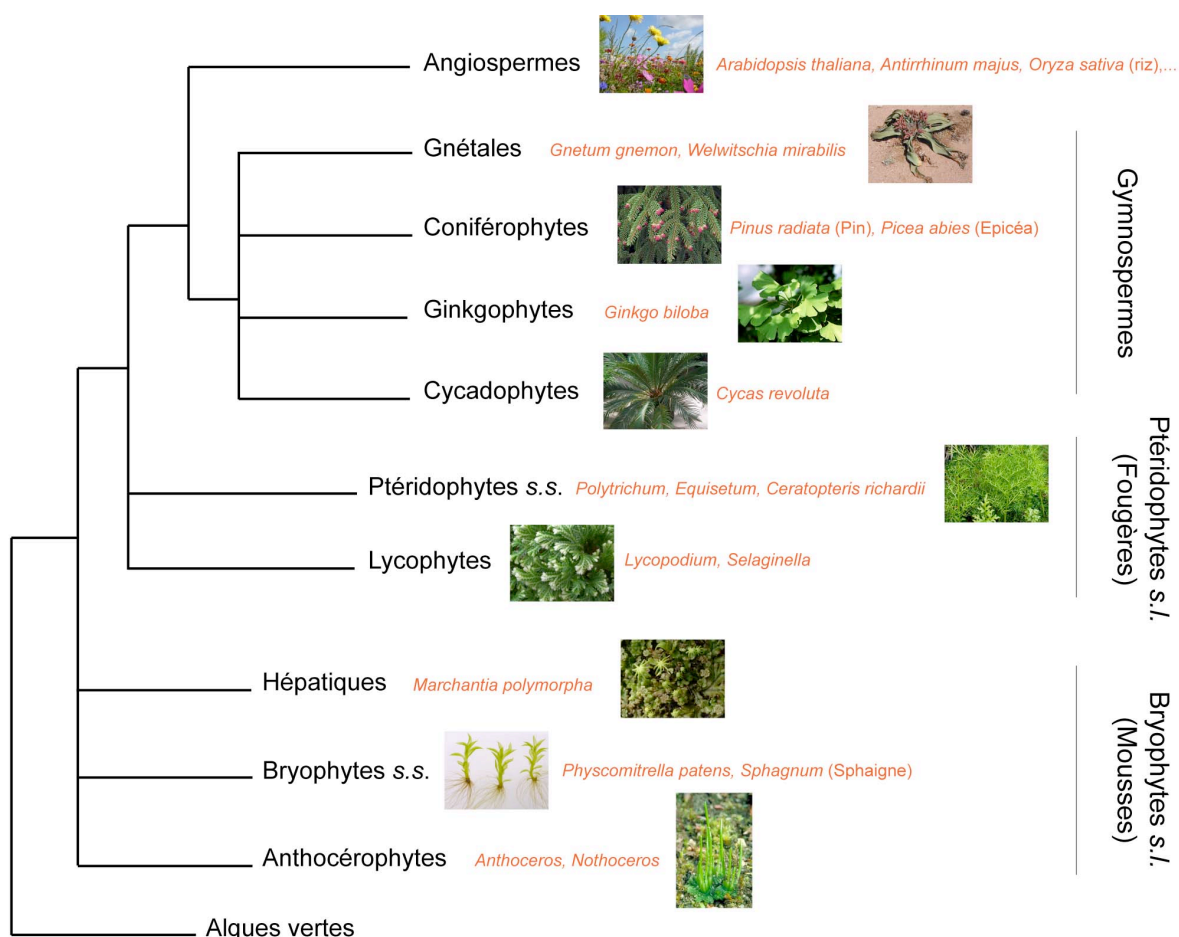
#### 4. LFY au cours de l'évolution

##### ▪ Un gène unique, très conservé

Jusqu'à présent, *LFY* avait été détecté uniquement chez les plantes terrestres (**Article 2**). On classe ces plantes en quatre grands groupes : les angiospermes (plantes à fleurs) ; les gymnospermes, dont font partie les conifères (le sapin par exemple) et le *Ginkgo* ; les fougères ; et les mousses. Les gymnospermes portent des structures reproductrices mâles ou femelles appelées cônes. Cependant ces cônes n'ont pas les caractéristiques d'une fleur : les ovules ne sont pas contenus dans un carpelle mais sont nus, il n'y a pas de structures attirant les insectes et la pollinisation se fait grâce au vent, les organes mâles et femelles sont sur deux cônes séparés, les organes sont organisés en spirale et non en verticilles. Les fougères et les mousses se reproduisent par sporulation. *LFY* a été trouvé chez toutes les espèces analysées dans ces quatre groupes (**Figure 11**).

Contrairement à beaucoup de FTs qui forment des familles multigéniques, *LFY* est généralement présent en une ou deux copies par génome. Les rares cas où plusieurs copies existent sont reliés à des évènements de polyploïdisation du génome (chez le maïs par exemple) et il n'y a aucun signe de divergence par néo-fonctionnalisation. Il y a cependant une exception à cette règle chez les gymnospermes où il existe un paralogue de *LFY*, nommé *NEEDLY (NLY)*, ayant divergé ancestralement (Frohlich, 2003).

La protéine est restée très conservée au cours de l'évolution (**Figure 8**). On compte 43% d'identité et 65% de similarité entre les protéines d'*Arabidopsis* et de la mousse *Physcomitrella patens* : 71% d'identité et 87% de similarité pour le domaine de liaison à l'ADN, 44% d'identité et 79% de similarité pour le domaine N-terminal. Les 15 acides aminés impliqués dans des interactions avec l'ADN d'après la structure du domaine de liaison chez *Arabidopsis* sont parfaitement conservés (**Figure 8**). Cela suggère que les propriétés de *LFY* soient conservées chez les plantes terrestres, et que le réseau contrôlé par *LFY* chez les angiospermes se soit établi plutôt par des modifications dans les éléments *cis*.



**Figure 11 | Arbre phylogénétique simplifié des plantes terrestres.**

Les algues vertes sont choisies comme groupe externe. Les noms d'espèces représentatives de chaque groupe sont inscrits en rouge. s.s. : *sensus stricto* (sens strict), s.l. : *sensus lato* (sens large). D'après Marie Monniaux (inspiré de Frohlich et Chase, 2007).

### ▪ LFY chez les plantes sans fleurs

En dehors des angiospermes, le rôle de LFY est peu connu. Chez les gymnospermes, des données de profil d'expression suggèrent que *LFY* pourrait avoir un rôle dans le développement des organes reproducteurs et contrôler certains gènes MADS (Frohlich, 2003 ; Theissen et Melzer, 2007 ; données de l'équipe en cours de publication).

Chez les fougères, l'expression de *LFY* a été observée dans les tissus végétatifs mais surtout reproducteurs, et ne semble pas coïncider avec l'expression des gènes MADS, suggérant que leur régulation par LFY soit apparue après la divergence des fougères (Himi *et al.*, 2001).

On dispose de données fonctionnelles pour *LFY* chez la mousse *P. patens* (bryophyte modèle). Cette plante se développe sous forme haploïde à partir d'une spore pour former une tige feuillue appelée gamétophyte. Les pieds femelles portent les ovules et les pieds mâles les spermatozoïdes. Après fécondation, le sporophyte, individu diploïde, se développe sur le gamétophyte femelle et donne à son tour des spores (Cove, 2005). *LFY* est présent en deux copies très similaires chez *P. patens*, principalement exprimées à l'apex du gamétophyte et dans le sporophyte. Lorsque les deux copies (*PpLFY1* et *PpLFY2*) sont mutées, le développement du gamétophyte est quasi normal, mais après fécondation, le zygote ne se divise pas et le sporophyte ne se développe donc pas (Tanahashi *et al.*, 2005). *LFY* est donc indispensable au déroulement du cycle de vie chez cette espèce.

### ▪ Évolution des propriétés biochimiques de LFY

La question de l'évolution des propriétés biochimiques de LFY a été adressée par Maizel et ses collaborateurs (Maizel *et al.*, 2005). Des tests fonctionnels avec des homologues de LFY issus des quatre groupes évolutifs ont révélé que contrairement à toutes les protéines testées, PpLFY n'était capable ni de compléter un mutant *lfy* chez *A. thaliana*, ni de réguler les gènes cibles de LFY. *In vitro*, PpLFY ne reconnaît pas les séquences liées par la protéine d'*Arabidopsis*, soit parce que PpLFY ne lie pas l'ADN, soit parce que sa spécificité est différente (Maizel *et al.*, 2005).

Basés sur une comparaison de séquence, certains acides aminés ont été proposés comme candidats pour expliquer cette différence de comportement (Maizel *et al.*, 2005) (**Figure 8**).

Le cas de *P. patens* est très intéressant, il suggère que malgré sa très forte conservation et le fait que le gène ne fasse pas partie d'une famille, les propriétés de liaison à l'ADN de LFY aient subi des modifications au cours de l'évolution.

### ▪ **LFY et les théories sur l'apparition des fleurs**

Les angiospermes sont apparues il y a 130 millions d'années, alors que la flore terrestre était dominée par les gymnospermes. On ne connaît pas d'espèce, actuelle ou fossile, ayant des caractéristiques intermédiaires entre les gymnospermes et les angiospermes. De plus, les angiospermes ont connu une véritable radiation évolutive : elles se sont diversifiées de manière extrêmement rapide. L'origine des fleurs est depuis longtemps une des grandes questions de la biologie évolutive, que Darwin avait qualifié d'« abominable mystère » (Darwin dans une lettre à Hooker en 1879 ; Friedman, 2009 ; Frohlich, 2003 ; Frohlich et Chase, 2007 ; Pennisi, 2009 ; Theissen et Melzer, 2007).

Les théories actuelles relatives à cette apparition s'appuient sur l'étude de fossiles, sur la phylogénie moléculaire, la génétique fonctionnelle, et de plus en plus sur la biologie du développement (Frohlich et Chase, 2007).

Comme *LFY* a un rôle majeur dans le développement floral et a précédé l'invention de la fleur, il a été proposé que des modifications du réseau de régulation contrôlé par *LFY* aient participé à l'apparition de la fleur (Frohlich, 2003 ; Frohlich et Chase, 2007 ; Theissen et Melzer, 2007).

## OBJECTIFS

*LFY* est un régulateur central pour le développement des plantes, en particulier pour la floraison chez les angiospermes. Il centralise les différents signaux qui initient la floraison et dirige à son tour les acteurs de la formation des fleurs.

On dispose de nombreuses données génétiques sur *LFY* et son réseau de régulation chez *Arabidopsis*, la plante modèle. Mais malgré son importance capitale, les mécanismes moléculaires impliqués dans le fonctionnement de ce FT ne sont pas entièrement élucidés.

L'obtention de la structure du domaine de liaison à l'ADN dans l'équipe en 2008, ainsi que la détermination de sa spécificité de liaison à l'ADN, ont grandement contribué à la compréhension du mode d'action de *LFY*.

Cependant, *LFY* a en N-terminal un second domaine conservé dont on ne connaît ni la structure, ni la fonction. La forte conservation de ce domaine au cours de l'évolution suggère qu'il a très certainement un rôle essentiel à jouer pour le fonctionnement de la protéine.

Le travail présenté dans le **Chapitre I** de cette thèse a pour objectif de répondre aux questions suivantes :

- ***Quelle est la fonction du domaine N-terminal ?***
- ***Quelles propriétés confère-t-il à la protéine ?***

Ne disposant que de très peu d'indices sur la nature et fonction du domaine N-terminal, nous avons choisi d'aborder cette étude par une approche structure–fonction. La résolution de la structure tridimensionnelle de ce domaine nous a apporté de précieux indices quant à son rôle. Nous avons ensuite mis en place des approches intégratives permettant de comprendre ses propriétés et leur importance, à la fois *in vitro* et chez la plante.

D'autre part, *LFY* est très conservé chez les plantes terrestres, en particulier les acides aminés impliqués dans l'interaction ADN-protéine, il s'est rarement dupliqué et n'appartient pas à une famille multigénique, et il a des rôles essentiels à la fois chez les angiospermes et chez les mousses, des plantes sans fleurs ayant divergé plus précocement au cours de l'évolution. Ces caractéristiques suggèrent que les possibilités d'évolution de ce FT soient très restreintes. En effet, des modifications de ses propriétés pourraient perturber considérablement le réseau de gènes qu'il contrôle et avoir des conséquences délétères sur le fonctionnement de la plante. Et pourtant, il a été montré que les propriétés de liaison à



l'ADN de LFY chez la mousse *P. patens* différaient de celles observées chez les autres plantes.

Dans le **Chapitre II** de cette thèse, les questions suivantes seront adressées :

- *Un FT peut-il évoluer sans former de famille multigénique ?*
- *Comment expliquer un changement de spécificité au sein d'un domaine de liaison à l'ADN extrêmement conservé ?*

Dans un premier temps, nous avons déterminé la spécificité de liaison de LFY chez différentes espèces issues des grands groupes évolutifs et nous avons mis en évidence plusieurs changements au cours de l'évolution. Par des approches de biochimie et biologie structurale, nous avons apporté des explications, au niveau moléculaire, à ces modifications. Ces résultats ont une portée générale sur l'évolution des FTs et des réseaux de régulation.

Au cours de ma thèse, J'ai aussi participé à deux études publiées. Ma contribution à ces travaux étant mineure, ils ne sont pas intégrés à part entière à la thèse, mais sont annexés à la fin du manuscrit. Ces deux publications visent à préciser la place de LFY dans le réseau floral :

- La transition florale est caractérisée par l'induction du gène *AP1* par *LFY*, et par un autre facteur impliqué dans la réponse à la lumière. Nous avons pu déterminer quels éléments *cis* étaient importants pour la régulation d'*AP1* par ces deux FTs (Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of *APETALA1*, Benlloch *et al.*, 2011, **Annexe 3**).
- LFY contrôle l'identité des méristèmes floraux, mais les mécanismes nécessaires à leur émergence étaient peu connus jusqu'à de récentes études. Nous avons mis en évidence le rôle de LFY dans ce processus, ainsi que sa relation avec un autre gène connu pour son rôle dans le développement des méristèmes (A variant of LEAFY reveals its capacity to stimulate meristem development by inducing *RAX1*, Chahtane *et al.*, 2013, **Annexe 4**).



# CHAPITRE I

Le domaine N-terminal de LEAFY,  
de la structure à la fonction



## I. Introduction

Pour comprendre la fonction de LFY en tant que régulateur clé du développement des plantes, il est important de connaître le fonctionnement de la protéine au niveau moléculaire. Mis à part son domaine de liaison à l'ADN, la protéine LFY comporte un second domaine : le domaine N-terminal. Sa conservation suggère qu'il est important pour la fonction de LFY, mais lorsque nous avons démarré notre étude, il n'avait pas de rôle connu. Il a été proposé qu'il module l'activité du domaine C-terminal (Maizel *et al.*, 2005) ou qu'il interagisse avec des partenaires de LFY (Souer *et al.*, 2008), mais ces hypothèses n'ont pas été clairement démontrées.

Dans le but de comprendre la fonction du domaine N-terminal, nous nous sommes d'abord intéressés à sa structure et à ses propriétés biochimiques.

Suite au travail de Cécile Hamès, qui a fait sa thèse dans l'équipe, la production du domaine de liaison à l'ADN était bien maîtrisée. Mais ni la protéine entière ni le domaine N-terminal n'avaient pu être produits dans des conditions permettant leur analyse.

Après avoir optimisé la production et la purification du domaine N-terminal ainsi que de la protéine entière, nous avons obtenu la structure cristallographique du domaine N-terminal et découvert qu'il présentait une forte homologie structurale avec des domaines appelés Steriles Alpha Motifs (SAM). La structure, ainsi que des analyses biochimiques nous ont permis de montrer que ce domaine SAM était un domaine d'oligomérisation. Nous avons mis en évidence le rôle de cette propriété pour la liaison à l'ADN de la protéine entière d'abord *in vitro*, puis *in vivo* chez *Arabidopsis*.

Les résultats concernant l'étude du domaine N-terminal de LFY sont regroupés dans le chapitre qui suit (**Chapitre I**), sous forme d'un article (**Article 1**) qui sera soumis prochainement pour publication. Des **données complémentaires** sont ensuite exposées et **discutées** plus en détail. Pour terminer, un **bilan** et des **perspectives** sont proposés.

Plusieurs personnes de l'équipe, ainsi que des collaborateurs d'autres laboratoires ont participé à ces travaux. Les compétences apportées par chacun ont permis la mise en place d'une approche intégrative. La production et purification des protéines recombinantes, les analyses biochimiques, et les analyses fonctionnelles chez *Arabidopsis* ont été faites au laboratoire, avec la participation d'Emmanuel Thévenon, technicien CNRS, qui a contribué à la réalisation de plusieurs constructions et analyses biochimiques et Gabrielle Tichtinsky,

maître de conférences à l'Université Joseph Fourier, qui a contribué à la caractérisation fonctionnelle chez les plantes. La collecte des données de diffraction et la résolution de la structure cristallographie ont été faites par Max Nanao, cristallographe et responsable de la ligne ID23-2 au European Synchrotron Radiation Facility (ESRF, Grenoble). Les analyses de Size Exclusion Chromatography - Multi Angle Laser Light Scattering (SEC-MALLS) ont été faites avec l'aide de Marc Jamin, professeur au Unit of Virus Host Cell Interactions (UVHCI, Grenoble). Les expériences d'immunoprécipitation de la chromatine et de séquençage (ChIP-seq) ont été réalisées dans l'équipe de Markus Schmid au Max Planck Institute for Developmental Biology (Tübingen, Allemagne). Les analyses bioinformatiques des données de ChIP-seq ont été réalisées avec l'aide de François Parcy, chercheur CNRS et directeur de notre équipe. Marta Peirats Llobet, stagiaire Erasmus, a également participé à ce projet.

## II. Article 1

### A cryptic SAM domain shapes the DNA binding landscape of the LEAFY transcription factor

#### Authors

Camille Sayou<sup>1234</sup>, Max H. Nanao<sup>56</sup>, Marc Jamin<sup>6</sup>, David Posé Padilla<sup>7</sup>, Felix Ott<sup>7</sup>, Gabrielle Tichtinsky<sup>1234</sup>, Emmanuel Thévenon<sup>1234</sup>, Marta Peirats Llobet<sup>1234a</sup>, Markus Schmid<sup>7</sup>, François Parcy<sup>1234b\*</sup>, Renaud Dumas<sup>1234\*</sup>

<sup>1</sup> CNRS, Laboratoire de Physiologie Cellulaire & Végétale, UMR 5168, F-38054 Grenoble, France

<sup>2</sup> CEA, DSV, iRTSV, LPCV, F-38054 Grenoble, France

<sup>3</sup> Univ. Grenoble Alpes, LPCV, F-38054 Grenoble, France

<sup>4</sup> INRA, LPCV, USC1359, F-38054 Grenoble, France

<sup>5</sup> European Molecular Biology Laboratory, 6 rue Jules Horowitz, BP 181, 38042, Grenoble, France

<sup>6</sup> Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

<sup>7</sup> Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

<sup>a</sup> Present address: Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas- Universidad Politécnica de Valencia, ES-46022 Valencia, Spain

<sup>b</sup> Present address: Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, V5Z 4H4, Canada

\*Correspondence: François Parcy, francois.parcy@cea.fr; Renaud Dumas, renaud.dumas@cea.fr

#### Keywords

LEAFY, transcription factor, Sterile Alpha Motif, oligomerisation, DNA binding, cooperativity, ChIP-seq, flower development

#### Running title

A cryptic SAM domain controls LEAFY DNA binding

## SUMMARY

“Master regulators” are sufficient to induce entire complex developmental pathways. However, the key molecular properties underlying such major switches remain to be uncovered. The LEAFY transcription factor is the master regulator for flower development: it is necessary and sufficient to confer a floral fate to plant stem cells. LEAFY binds its target genes regulatory regions through a dimeric DNA binding domain. Still, how LEAFY gets directed to these regions remains poorly understood. Using a combination of structural biochemistry, *in vivo* characterisation and genome-wide DNA binding analysis, we showed that LEAFY N-terminal conserved domain was a cryptic SAM/Pnt oligomerisation domain, which allows LFY to cooperatively bind regions bearing multiple *cis*-elements, reduces LFY DNA binding promiscuity and gives access to closed chromatin regions. This study illustrates how the DNA binding landscape of a key transcriptional regulator is shaped by a DNA binding domain combined with an oligomerisation domain.

## INTRODUCTION

Master developmental regulators are capable of initiating regulatory networks and activating whole developmental programs. Which molecular properties are key to trigger developmental switches is an open question.

*LEAFY* (*LFY*) is a transcriptional regulator essential for developmental processes in plants (reviewed in Moyroud et al., 2010). It is conserved from algae (streptophyte) to flowering plants (Article 2, submitted).

In the basal land plant *Physcomitrella patens*, *LFY* is essential for the plant life cycle as it is required for cell division during the diploid developmental phase (Tanahashi et al., 2005).

In *Arabidopsis thaliana*, where *LFY* has been extensively characterised, it acts as a master regulator, integrating environmental and endogenous signals and orchestrating the whole floral network (reviewed in Moyroud et al., 2009; Irish, 2010; Posé et al., 2012; Chandler, 2012). *LFY* triggers the early emergence of meristems on the flanks of the shoot tip (Chahtane et al., 2013; Yamaguchi et al., 2013) and specifies their floral identity, in particular through direct activation of *APETALA1* (*AP1*; Parcy et al., 1998; Wagner et al., 1999) and repression of *TERMINAL FLOWER1* (*TFL1*; Liljegren et al., 1999; Parcy et al., 2002; Ratcliffe et al., 1998; Jaeger et al., 2013). In floral buds, together with organ specific co-regulators (Liu et al., 2009; Chae et al., 2008; Souer et al., 2008; Lohmann et al., 2001), *LFY*



controls genes involved in floral organ identity, such as the B (*APETALA3*, *AP3*) and C (*AGAMOUS*, *AG*) class genes (Weigel and Meyerowitz, 1993; Parcy et al., 1998; Busch et al., 1999; Lohmann et al., 2001; Lamb et al., 2002; Liu et al., 2009). Furthermore, *LFY* is sufficient to trigger early flowering and ectopic flower production when constitutively expressed (Weigel and Nilsson, 1995). Genome-wide studies revealed the DNA binding profile of *LFY* in *Arabidopsis*, confirming its association with a wide range of regulatory regions involved in flower development (Moyroud et al., 2011; Winter et al., 2011).

Chromatin remodelling is crucial for the onset of developmental programs. Nucleosomes, made of histone octamers, are the basic unit of chromatin. When condensed together, they present a barrier for proteins that need to contact the DNA and their rearrangement is a key feature of gene regulation (Bell et al., 2011). Recent findings revealed a role for *LFY* in chromatin remodelling processes.

While plants go through the vegetative phase, two *Polycomb* group (PcG) complexes, Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2), maintain chromatin repression at floral regulators loci such as those of the B and C class genes *AP3* and *AG* (Goodrich et al., 1997). PRC2 deposits repressive marks on histones and PRC1 further compacts the chromatin, thus stabilising the repressed state (Bemer and Grossniklaus, 2012). *LFY* directly down-regulates the expression of *EMBRYONIC FLOWER1* (*EMF1*), a PRC1 subunit (Chen et al., 1997; Chou et al., 2001; Calonje et al., 2008).

During induction of *AP3* and *AG* in flower primordia, polycomb repression is overcome by trithorax group (TrxG) proteins (Goodrich et al., 1997; Alvarez-Venegas et al., 2003; Katz et al., 2004; Saleh et al., 2007; Carles and Fletcher, 2009; Wu et al., 2012). *LFY* was shown to recruit the SWI2/SNF2 chromatin-remodeling ATPase *SPLAYED* (*SYD*) and probably *BRAHMA* (*BRM*) to *AP3* and *AG* regulatory regions (Wu et al., 2012). Using the energy derived from ATP hydrolysis, SWI2/SNF2 chromatin-remodeling ATPases modify the nucleosome-DNA interaction, hence increasing the accessibility of genomic regions to transcription factors and the transcriptional machinery (Clapier and Cairns, 2009). *SYD* and *BRM* recruitment by *LFY* likely occurs through direct interaction, as *LFY* was shown to interact with these proteins *in vitro* (Wu et al., 2012). *SEPALLATA3* (*SEP3*), a *LFY* cofactor, was also suggested to participate to this process (Wu et al., 2012).

The mechanisms highlighted for *LFY* on *AP3* and *AG* loci may represent general processes by which *LFY* could act on the chromatin landscape at its target gene regulatory regions, through genetic and physical interactions with chromatin associated factors. Whether *LFY* is able to access to closed chromatin regions and by which mechanism is not known.

Some of the molecular mechanisms underlying LFY function have been addressed by structural and biochemical studies. The LFY protein binds DNA through its highly conserved C-terminal DNA binding domain (DBD). A structural characterisation revealed that it adopts a unique Helix-Turn-Helix (HTH) fold and dimerises upon DNA binding (Hamès et al., 2008). The DBD determines the sequence preference of LFY for a palindromic 19 bp motif (Hamès et al., 2008; Article 2, submitted). A biophysical model describing LFY DNA binding specificity *in vitro* was built and accurately predicted LFY binding sites on the *Arabidopsis* genome (Moyroud et al., 2011).

In addition to its well-characterised DBD, LFY possesses a second conserved domain in its N-terminal region. This domain was recently shown to mediate LFY self-association in yeast cells and in tobacco leaves, using LFY deletion variants and a combination of mutations, based on a proposed leucine zipper fold (Siriwardana and Lamb, 2012). However, the structure and the role of LFY N-terminal domain remain elusive.

Despite the central role of this transcriptional regulator, the molecular properties required for LFY to trigger such a major developmental switch are not fully understood. LFY DBD allows recognition of DNA, but the molecular mechanisms underlying LFY access and regulation of its target genes remain to be elucidated.

To better understand how LFY works, we first aimed to characterise the poorly understood N-terminal conserved domain. Since this domain has no similarity with other protein, we adopted a structure-function strategy. We obtained the crystal structure of the *Ginkgo biloba* LFY N-terminus and discovered it adopts a Sterile-Alpha-Motif (SAM) fold. We showed that the SAM domain of LFY (LFY-SAM) is an oligomerisation domain and we validated the importance of oligomerisation for flower development in *A. thaliana*. We investigated the effect of oligomerisation on DNA binding. *In vitro* assays pointed out that it prevents promiscuous binding, and promotes cooperative binding of multiple sites. We performed a genome-wide ChIP-Seq analysis in *A. thaliana* and showed that oligomerisation through the SAM domain was also necessary for *in vivo* DNA binding, in particular to regions displaying a high density of binding sites, and where chromatin is packed and hardly accessible for the majority of transcriptional regulators. Interestingly, our data suggest that LFY could be a pioneer transcription factor, acting early during the transcriptional activation process by accessing repressed and closed regions thanks to its SAM oligomerisation domain.

## RESULTS

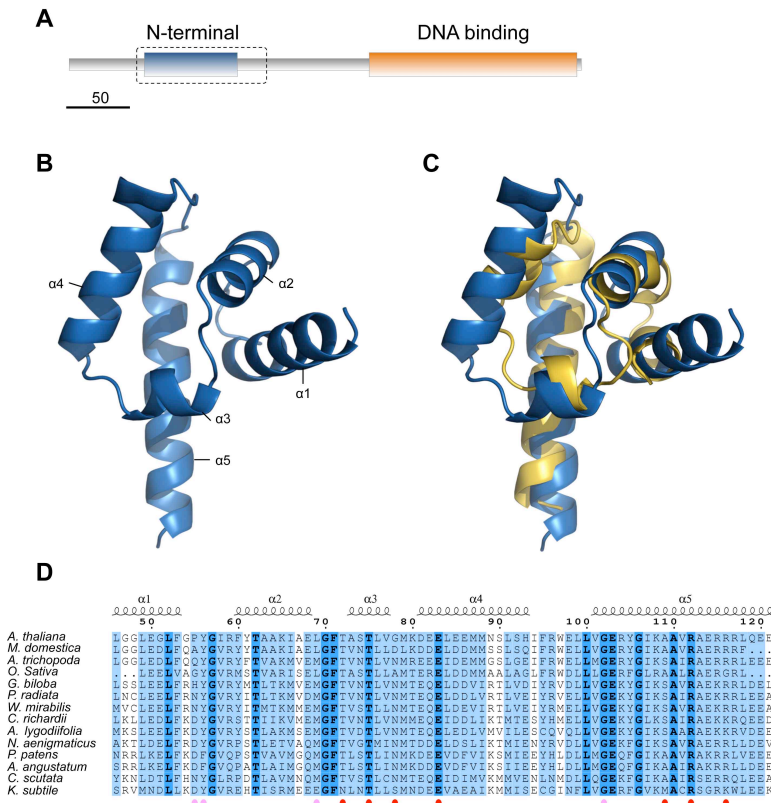
### The crystal structure of *G. biloba* LFY N-terminal conserved domain reveals a cryptic Sterile Alpha Motif

To better understand the function of LFY N-terminal domain and obtain ways to manipulate its properties, we solved its crystal structure. We screened several constructs containing LFY from multiple plant species for crystallisation. We obtained high quality crystals using the N-terminal domain of *G. biloba* LFY (GbLFY-N; from the gymnosperms group) (**Figure 1A**). The X-ray structure was solved with crystals containing selenomethionine, by the single-wavelength anomalous diffraction (SAD) phasing method. It was refined to a  $R_{\text{free}}$  of 22.9 % at 2.3 Å resolution. Details of the structure determination and refinement are given in **Table 1**. The asymmetric unit contained two monomers, both adopting a similar fold (they can be superimposed with a RMSD of 0.3 Å). Each monomer was made of five  $\alpha$ -helices separated by four loops (**Figure 1B**).

Unexpectedly, GbLFY-N shared a strong structural similarity with Sterile Alpha Motifs (SAM), also called Pointed (Pnt) domains (Qiao and Bowie, 2005). GbLFY-N and the SAM domain of p63 $\alpha$  (Sathyamurthy et al., 2011) can be superimposed with a RMSD of 1.9 Å for 62 C $\alpha$  (**Figure 1C**). The SAM domain is a common protein motif found in eukaryotic genomes. It is an interaction domain involved in many cellular functions, including transcriptional regulation (Qiao and Bowie, 2005; Hollenhorst et al., 2011).

Table 1. Data collection and refinement statistics		
	GbLFY-N (Native)	GbLFY-N (Se-Met)
<b>Data collection</b>		
Wavelength (Å)	0.872	0.872
Resolution (Å) range	52.34 - 2.25 (2.33 - 2.25)	100 - 2.61 (2.77 - 2.61)
Space group	P 65	P 65
Cell dimensions (Å, °)	a=b=81.1, c=78.5	a=b=80.6, c=78.5
Number of total reflections	57,705	83,685
Number of unique reflections	14,004	8,879
Multiplicity	4.1(3.9)	9.4 (9.2)
Completeness (%)	99.3 (99.0)	99.6 (97.5)
Mean I/ $\sigma$ (I)	15.67 (2.89)	15.28 (2.54)
R-sym	0.069 (0.514)	0.127 (0.915)
<b>Refinement statistics</b>		
Wilson B-factor	44.1	35.85
R-factor	0.1859 (0.2564)	
R-free	0.2298 (0.2767)	
Number of atoms	1462	
Macromolecule	1335	
Ligands	25	
Water	102	
Protein residues	163	
RMS(bonds)	0.016	
RMS(angles)	1.78	
Ramachandran favored (%)	99	
Ramachandran outliers (%)	0	
Clashscore	13.40	
Average B-factor	46.80	
Macromolecule	46.00	
Solvent	51.80	

Statistics for the highest-resolution shell are shown in parentheses



**Figure 1. LFY N-terminus contains a SAM domain**

(A) Domain representation of LFY protein. The blue box is the conserved N-terminal domain, the orange box is the C-terminal conserved DNA binding domain. The area outlined with black dashes indicates the region used to obtain the crystal structure. The scale bar corresponds to 50 amino acids.

(B) Crystal structure of GbLFY-N monomer.  $\alpha$  indicates the alpha helices

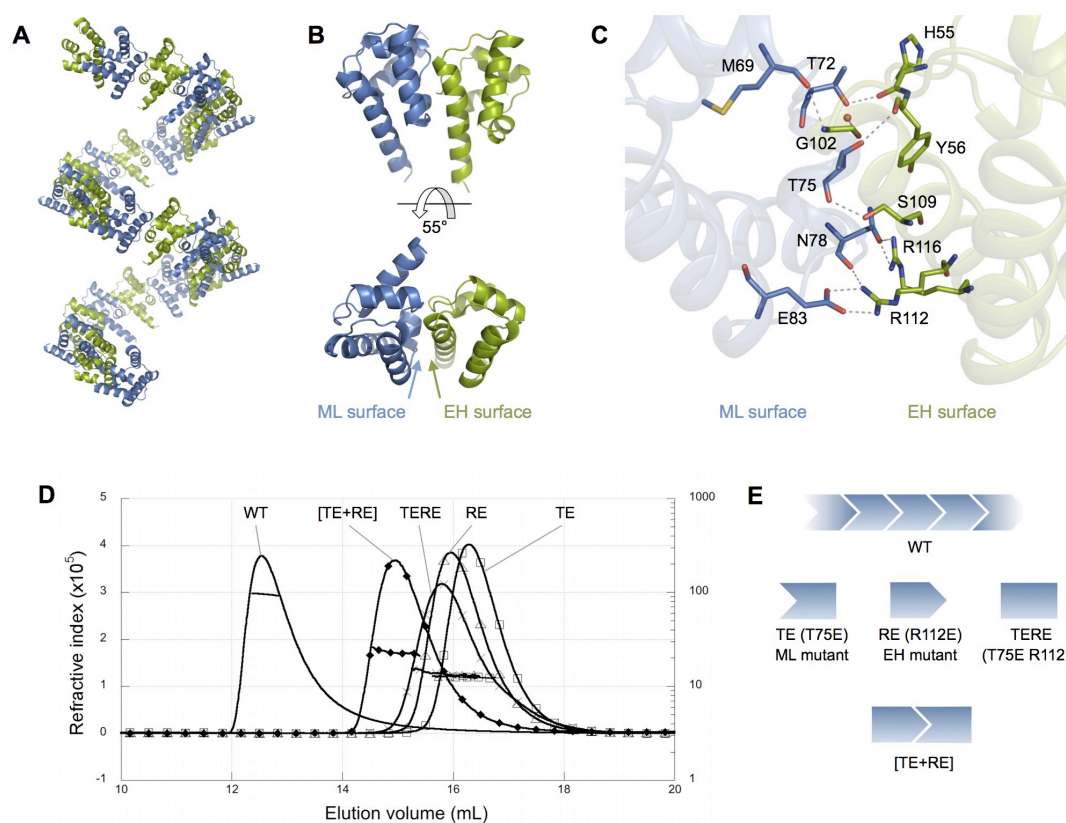
(C) Superimposition of GbLFY-N (blue) and the SAM domain of p63 $\alpha$  (yellow, PDB 2Y9U, Sathyamurthy *et al.*, 2011). The two structures are very similar and can be superimposed with an RMSD of 1.9 Å for 62 C $\alpha$ .

(D) Sequence alignment of LFY-N conserved domain. The residue numbering refers to *A. thaliana* LFY sequence; the secondary structure from GbLFY-SAM crystal structure is indicated above the alignment; red and pink circles indicate residues involved in the interaction between SAM monomers though the lateral and main chain respectively. LFY from various species are shown: *A. thaliana*, *Malus domestica*, *Amborella trichopoda* and *Oryza sativa* belong to angiosperms; *G. biloba*, *Pinus radiata* and *Welwitschia mirabilis* belong to gymnosperms; *Ceratopteris richardii* and *Angiopteris lygodiiifolia* belong to ferns; *Nothoceros aenigmaticus*, *P. patens* (PpLFY1) and *Atrichum angustatum* (AtranFlo1) belong to bryophytes; *Coleochaete scutata* and *Klebsormidium subtilis* belong to charophytes (algae).

**LFY-SAM is an oligomerisation domain**

In the crystal, GbLFY-SAM molecules were arranged in a head-to-tail fashion, each monomer forming two interfaces with its neighbours. Their arrangement formed a helical polymer (**Figure 2A**). Similar polymers had been described for other SAM domains, all with 6 to 8 monomers per turn and a repeat distance ranging from 33 to 53 Å (Kim *et al.*, 2001, 2002, 2005; Qiao *et al.*, 2004; Baron *et al.*, 2006; Harada *et al.*, 2008; Stafford *et al.*, 2011). GbLFY-SAM polymer presented unique characteristics: 12 monomers per turn and a repeat

distance of 78 Å (**Figure 2A**). The two surfaces involved in that arrangement were called the mid-loop (ML) and the end-helix (EH) surfaces (**Figure 2B**), according to their position and in reference to their assignment for the TEL SAM domain (Kim et al., 2001). The ML surface carried negative and polar residues (**Figure 2C**), including T72, T75 and E83 (on loop 2, helix 3 and helix 4, respectively), which interacted with the positively charged EH surface through hydrogen bonds and salt bridges, in particular via R112 and R116 on the helix 5 (residues numbering refers to LFY from *A. thaliana* – AtLFY – sequence and is consistent with the sequence alignment in **Figure 1D**).



### Figure 2. LFY-SAM is an oligomerisation domain

(A) Helical polymer of GbLFY-SAM in the crystal structure. 28 monomers are shown and are alternatively coloured in blue and green.

(B) Side and top views (upper and bottom panels respectively) of the GbLFY-SAM dimer in the asymmetric unit of the crystal. ML and EH interaction surfaces are indicated on the bottom panel.

(C) Detailed view of GbLFY-SAM oligomerisation interface. Residues mediating contacts between two monomers are shown in stick; residues on the ML and EH surfaces are in blue and green respectively.

(D) SEC-MALLS molecular mass analysis of GbLFY-SAM, GbLFY-SAM<sub>TE</sub> (T75E substitution at the ML surface) and GbLFY-SAM<sub>RE</sub> (R112E substitution at the EH surface), GbLFY-SAM<sub>TERE</sub> (T75E, R112E) double mutant and GbLFY-SAM<sub>[TE+RE]</sub> (an equimolar mixture of GbLFY-SAM<sub>TE</sub> and GbLFY-SAM<sub>RE</sub>). 50 µL at 4.4 to 4.7 mg/mL of proteins were used for this experiment. Elution profiles were monitored by excess refractive index (left ordinate axis). The line under each elution pick shows the molecular mass distribution (right ordinate axis). The measured molecular mass for each protein is reported in **Table S1** and is consistent with that of an oligomer for GbLFY-SAM, a monomer for GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub>, GbLFY-SAM<sub>TERE</sub> and a dimer for GbLFY-SAM<sub>[TE+RE]</sub>.

(E) Schematic representation of the oligomeric state of GbLFY-SAM WT and mutants deduced from the SEC-MALLS analysis.

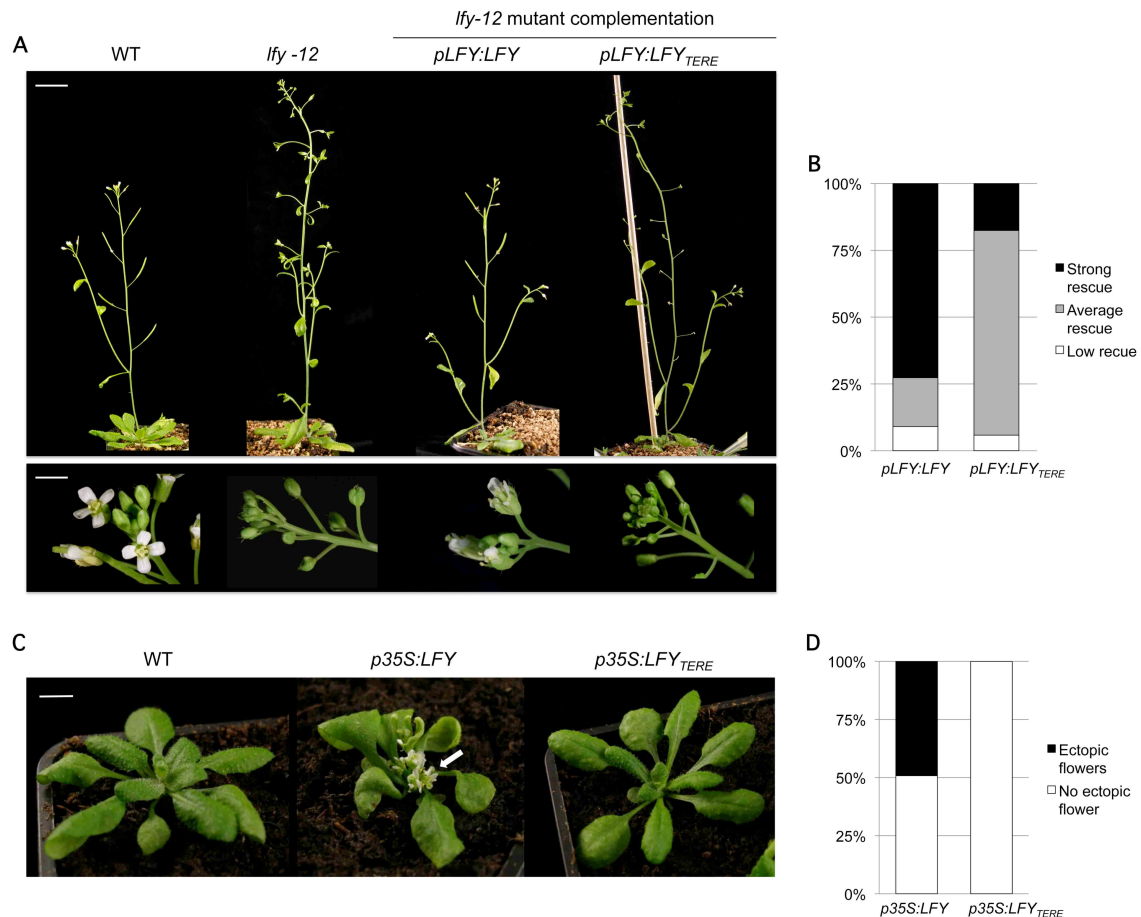
In order to validate this organisation in solution, we substituted highly conserved residues mediating strong side chain contacts (**Figures 2C** and **Figure 1D**), which should disrupt interactions. Hence, we generated two single mutants: GbLFY-SAM<sub>TE</sub> (T75E substitution at the ML surface) and GbLFY-SAM<sub>RE</sub> (R112E substitution at the EH surface). Both mutations were combined to create the GbLFY-SAM<sub>TERE</sub> (T75E, R112E) double mutant, affected on both surfaces. Then, we analysed the oligomerisation state of the wild-type (WT) and mutated SAM domains using Size Exclusion Chromatography - Multi Angle Laser Light Scattering (SEC-MALLS; **Figure 2D**, **Table S1**). The WT GbLFY-SAM formed oligomers in solution (**Figure 2D**, **Table S1**). The number of subunits in the oligomer increased with the protein concentration until a maximum of about 8 molecules per complex (**Figure S1**, **Table S1**). All the different mutations we introduced completely abolished oligomerisation: GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub>, GbLFY-SAM<sub>TERE</sub> proteins were found monomeric (**Figure 2D**, **Table S1**). GbLFY-SAM<sub>TE</sub> and GbLFY-SAM<sub>RE</sub> single mutants dimerised when mixed (forming the GbLFY-SAM<sub>[TE+RE]</sub> complex), demonstrating that the mutations did not alter the overall fold of the protein (**Figure 2D**, **Table S1**) and that simple mutant proteins still interacted with their intact surface. These results showed that the head-to-tail arrangement observed in the crystal structure reflected the state of LFY-SAM in solution, which formed an oligomer of limited size. The point mutations we designed allow controlling the oligomeric state, creating either monomeric or dimeric variants of LFY-SAM (**Figure 2E**).

Aligning sequence of LFY from various species throughout the plant kingdom revealed that the key interaction residues are well conserved (**Figures 1D** and **S2**), suggesting that the oligomerisation property of GbLFY-SAM is a general characteristic of LFY-SAM. We supported this hypothesis by analysing the SAM domain of AtLFY (AtLFY-SAM). We found that its molecular mass in solution increased with the protein concentration (**Figure S2**), similar to what we observed for GbLFY-SAM (**Figure S1**) confirming that the oligomerisation capacity also exists in angiosperms (flowering plants).

### **LFY-SAM oligomerisation is required to build flowers in *Arabidopsis***

The conservation of LFY-SAM sequence throughout evolution strongly suggests that oligomerisation is essential for LFY function. Several mutations in this domain that do compromise LFY function *in planta* have been described. However, these are poorly informative regarding oligomerisation: examining our crystallographic structure revealed that most of these mutations affect buried hydrophobic residues (Ikeda-Kawakatsu et al., 2012; Siriwardana and Lamb, 2012) and are likely disturbing the folding of the protein. Only the *A. thaliana* very weak *lfy-22* allele (Levin and Meyerowitz, 1995) carries a mutation (G70D) localised at the SAM ML surface and possibly weakens oligomerisation.

In order to specifically assess the importance of oligomerisation, we introduced the TERE mutations in the AtLFY protein, and characterised its ability to perform the LFY function in *A. thaliana* plants.



### Figure 3. LFY-SAM oligomerisation is required for LFY floral function in *Arabidopsis*

(A-B) Functional complementation assay.

(A) Wild-type, *Ify-12* mutant and *Ify-12* expressing *pLFY:LFY* or *pLFY:LFY\_TERE* grown for 40 days under long day conditions. The top panel shows the entire plants with the scale bar corresponding to 2 cm; the bottom panel is a close-up view of the inflorescences with the scale bar corresponding to 2 mm.

(B) Phenotypic classes for *pLFY:LFY* and *pLFY:LFY\_TERE* plants. The number of T1 plants observed is indicated in **Table S2**.

(C-D) Overexpression assay.

(C) Wild-type and transgenic plants expressing *p35S:LFY* or *p35S:LFY\_TERE* grown for 30 days under long day conditions. The white arrows points out ectopic rosette flowers on the *p35S:LFY* plant. The scale bar corresponds to 1 cm.

(D) Phenotypic classes for *p35S:LFY* and *p35S:LFY\_TERE* plants. The number of T1 plants observed is indicated in **Table S3**.

First we assessed AtLFY<sub>TERE</sub> ability to complement the null *Ify-12* mutant (Q32stop) whose flowers are replaced by shoots or sterile shoot/floral intermediates never bearing any petal or stamen (**Figure 3A**) (Weigel et al., 1992; Maizel and Weigel, 2004). When expressed under its own promoter (*pLFY:LFY*), the WT AtLFY protein complements well the *Ify-12* mutant (Maizel et al., 2005; Chahtane et al., 2013). 8 of the 11 *Ify-12* mutants primary transformants

(T1) bearing the *pLFY:LFY* construct showed a high level of complementation, with fertile flowers displaying correctly arranged floral organs; 2 plants showed an intermediate rescue, with structures lacking proper floral organs and mostly sterile and one plant showed poor rescue, strongly resembling a *lfy-12* mutant. In contrast, AtLFY<sub>TERE</sub> was poorly efficient at rescuing the *lfy-12* mutant phenotype: 13 of 17 T1 *pLFY:LFY<sub>TERE</sub>* plants fell in the intermediate complementation class and only 3 plants complemented well the *lfy-12* mutant (**Table S2, Figures 3A and 3B**). These experiments indicated that the capacity of LFY to control flower formation was reduced, although not abolished, when oligomerisation through the SAM domain was compromised.

To further assess the importance of oligomerisation for the floral transition, we drove AtLFY expression from the *p35S* constitutive promoter and monitored its effect at an early developmental stage, when other pathways contributing to flower development have a minimal impact. When expressed precociously, AtLFY is known to induce early flowering, trigger the formation of ectopic flowers from the axils of rosette leaves and an early termination of the shoots (Weigel and Nilsson, 1995). Among the 61 *p35S:LFY* T1 plants we generated, half showed this characteristic phenotype. In contrast, it was never observed in the 51 *p35S:LFY<sub>TERE</sub>* T1 plants (**Table S3, Figures 3C and 3D**). We verified that both types of plants expressed AtLFY and AtLFY<sub>TERE</sub> proteins at similar level by western-blot (**Figure S3**). This result shows that the SAM domain oligomerisation is critical for LFY floral switch function.

Together, these results indicated that the capacity of LFY to control flower formation depends on its SAM domain oligomerisation.

Next, we wished to understand why oligomerisation is important for LFY function at the molecular level. As a transcription factor, LFY specifically binds to *cis*-elements present in its target genes regulatory regions and subsequently participates to their transcriptional regulation. Hence, we investigated the role of oligomerisation for DNA binding, both *in vitro* and *in vivo* at the genome wide level.

### **LFY-SAM is not essential for LFY to bind DNA as a dimer**

We took advantage on our aptitude to tightly control the oligomerisation state of the isolated SAM domain (**Figure 2E**) to assess the oligomerisation requirement for LFY binding to DNA. Previous studies on proteins containing oligomerisation SAM domains have been hampered by the difficulty of producing recombinant soluble proteins: the SAM domain often triggers protein aggregation and precipitation (Green et al., 2010). By optimising our purification protocol, we successfully purified the near full-length GbLFY protein (GbLFY $\Delta$ ). As previously



described for AtLFY (Chahtane et al., 2013; Article 2, submitted), deletion of the first non-conserved residues improved the protein production.

As GbLFY and AtLFY have the same DNA binding specificity (Article 2, submitted), we tested GbLFY $\Delta$  binding to a DNA fragment carrying a single LFY binding site (*AP1* probe) from the promoter of the *Arabidopsis APETALA1* gene, an established AtLFY target. Using Electrophoretic Mobility Shift Assay (EMSA), we showed that GbLFY $\Delta$ , GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  and GbLFY $\Delta_{TERE}$  formed a major complex on *AP1* DNA with a similar affinity (**Figure 4A** and **B**). This major complex corresponded to a protein dimer bound to DNA, as measured by SEC-MALLS (**Figure 4C**). Thus, we deduced that the SAM domain is not essential to bind DNA as a dimer, consistent with the fact that LFY-DBD has the endogenous capacity to dimerise on DNA (Hamès et al., 2008).

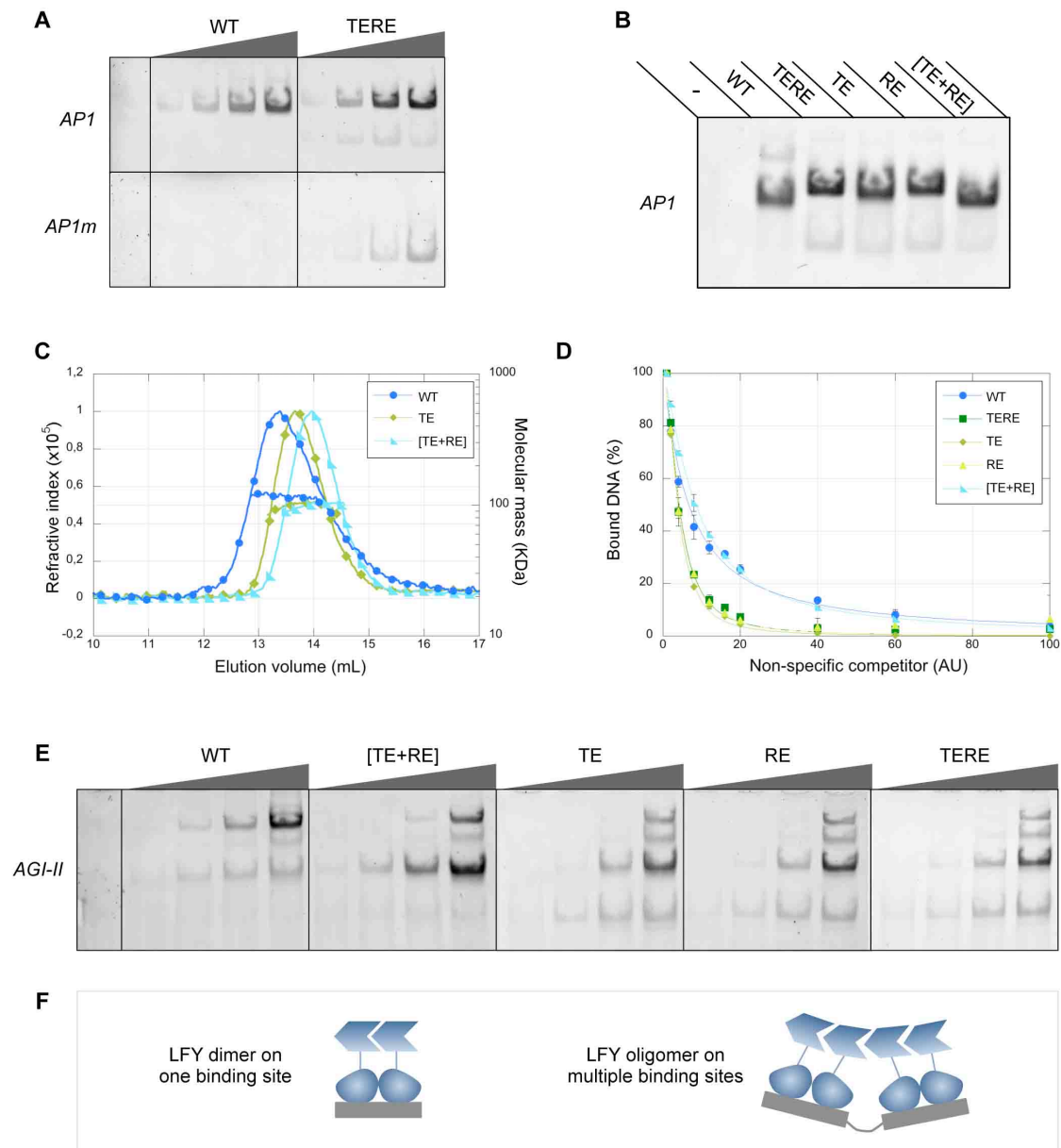
However GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  and GbLFY $\Delta_{TERE}$  major complexes migrated slightly higher than GbLFY $\Delta$  in EMSA (**Figure 4B**), which was likely due to a conformational difference, since a mix of GbLFY $\Delta_{TE}$  and GbLFY $\Delta_{RE}$  (hereafter called GbLFY $\Delta_{[TE+RE]}$ ) migrated as GbLFY $\Delta$  (**Figure 4B**). This result suggested that, as shown for the isolated SAM, the SAM domains interacted within a dimer of the full-length LFY protein when at least one ML and one EH surfaces were functional (GbLFY $\Delta$  and GbLFY $\Delta_{[TE+RE]}$ ), leading to a more compact and fast migrating complex compared to the mutants. A similar result was obtained for the *Arabidopsis* LFY protein (**Figure S4A**). So both LFY-DBD and LFY-SAM participate to the dimer formation on a single DNA binding site (**Figure 4F**).

### LFY-SAM increases selective DNA binding

In addition to the main dimeric complex, GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  and GbLFY $\Delta_{TERE}$  mutants showed a lower complex on the *AP1* probe in EMSA (**Figure 4A**), which likely corresponded to a monomer on DNA. This complex was particularly obvious on the *AP1m* probe that carries an altered version of the *AP1* binding site. Whereas GbLFY $\Delta$  did not recognise the *AP1m* DNA, the mutants showed non-specific binding (**Figure 4A**). GbLFY $\Delta_{TERE}$  was also able to bind a half binding site (one half of the palindromic site, or the other one, were altered on *AP1m1* and *AP1m2* probes, respectively) as a monomer and even as a dimer, whereas GbLFY $\Delta$  showed very weak binding (**Figure S4B**). This suggested that selective DNA binding is increased by the SAM oligomerisation.

To further test this hypothesis, we did a competition assay between the *AP1* specific probe and non-specific unlabelled competitor (fish genomic DNA). GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  and GbLFY $\Delta_{TERE}$  dissociated from *AP1* at lower competitor concentrations than GbLFY $\Delta$  and GbLFY $\Delta_{[TE+RE]}$  (**Figure 4D** and **Figure S4C**), confirming that dimerisation through a functional SAM domain promotes selective binding.

We concluded that even though the SAM domain oligomerisation was not essential to bind DNA as a dimer, it was nevertheless involved in LFY dimerisation and increased DNA binding selectivity.



**Figure 4. Effects of LFY-SAM oligomerisation on *in vitro* DNA binding**

(A) Electrophoretic mobility shift assay (EMSA) with 10 nM AP1 or AP1m DNA and 0, 50, 100, 250, 500 nM GbLFY $\Delta$  or GbLFY $\Delta_{\text{TERE}}$  protein. Only the protein-DNA complexes are shown.

(B) EMSA with 10 nM AP1 DNA and 500 nM GbLFY $\Delta$ , GbLFY $\Delta_{\text{[TE+RE]}}$ , GbLFY $\Delta_{\text{TE}}$ , GbLFY $\Delta_{\text{RE}}$  or GbLFY $\Delta_{\text{TERE}}$  protein. Only the protein-DNA complexes are shown.

(C) SEC-MALLS molecular mass analysis of GbLFY $\Delta$ , GbLFY $\Delta_{\text{TE}}$  and GbLFY $\Delta_{\text{[TE+RE]}}$  in complex with AP1 DNA. 50  $\mu$ L containing 2.6 mg/mL (60  $\mu$ M) protein and 15  $\mu$ M DNA were used for this experiment. Elution profiles were monitored by excess refractive index (left ordinate axis). The line under each elution pick shows the molecular mass distribution (right ordinate axis). The measured molecular mass for each complex is reported in **Table S4** and is consistent with that of a protein dimer on DNA.

(D) Percentage of DNA bound to LFY as a function of non-specific unlabelled competitor concentration. Quantifications were based on EMSA shown in **Figure S4C**, made with 10 nM AP1

DNA, 500 nM GbLFY $\Delta$ , GbLFY $\Delta_{[TE+RE]}$ , GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  or GbLFY $\Delta_{TERE}$  protein and increasing concentration of competitor.

(E) EMSA with 10 nM *AGI-II* DNA and 0, 50, 100, 250, 500 nM GbLFY $\Delta$ , GbLFY $\Delta_{[TE+RE]}$ , GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  or GbLFY $\Delta_{TERE}$  protein. Only the protein-DNA complexes are shown.

(F) Schematic representation of LFY bound as a dimer on a single binding site (left) and as an oligomer on multiple binding sites (right; here a tetramer on two binding sites is represented) deduced from the EMSA and SEC-MALLS analysis. DNA is represented in grey, LFY is shown in blue with the DNA binding domain in direct contact with DNA and connected to the SAM by a linker (shown as a line).

### LFY-SAM oligomerisation allows cooperative binding on multiple LFY binding sites

Many roles have been proposed for SAM proteins, including cooperative DNA binding on large regions (Green et al., 2010).

To test if LFY-SAM could affect binding on larger DNA fragments, we studied GbLFY $\Delta$  binding on a probe from the *AGAMOUS* (*AG*) gene regulatory region and that carries two nearby LFY binding sites (*AGI-II* probe) (**Figure 4E**). We observed two main complexes for GbLFY $\Delta$  protein, likely reflecting binding of one and two dimers on *AGI-II*. We wondered if this binding could involve a positive cooperative interaction between the two dimers bound on each binding site. To test whether such cooperativity existed and depended on the SAM domain oligomerisation, we compared the binding of GbLFY $\Delta$  to that of a GbLFY $\Delta_{[TE+RE]}$ , which mimics a WT dimer but is unable to further oligomerise (**Figure 2E**). We observed that GbLFY $\Delta_{[TE+RE]}$  was not as efficient as GbLFY $\Delta$  at forming the tetrameric complex, which was always present in smaller amount than the dimeric one (**Figure 4E**). We noticed that GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  and GbLFY $\Delta_{TERE}$  were also impaired for tetramer formation. AtLFY protein behaved similarly (**Figure S4A**). We concluded that LFY cooperatively binds on multiple sites, via oligomerisation of its SAM domain (**Figure 4F**).

### LFY genome-wide DNA binding relies on the SAM domain oligomerisation

To test how oligomerisation affects DNA binding *in vivo* at genome wide level, we performed a ChIP-Seq experiment on *Arabidopsis* seedlings constitutively expressing either AtLFY or AtLFY $_{TERE}$ .

ChIP-seq was performed in duplicate on 2 week-old *p35S:LFY*, *p35S:LFY $_{TERE}$*  and non-transgenic WT plants (Col-0), which do not express the endogenous LFY at this stage. After normalisation, genomic binding was compared between genotypes. We focused on the 2000 regions bound with the highest confidence in *p35S:LFY* as compared to Col-0 plants and analysed their binding by AtLFY $_{TERE}$ . We observed that, for all 2000 regions, the binding intensity was reduced in *p35S:LFY $_{TERE}$*  as compared to *p35S:LFY* but that the intensity of the reduction was very variable, from 2-3 fold to more than 300 fold for the most affected regions

(Figure 5A, Figure S5A). We also performed the converse analysis by selecting the highest confidence 2000 bound regions in *p35S:LFY<sub>TERE</sub>* and comparing their binding in *p35S:LFY*. Except for a few regions where the binding is slightly higher in *p35S:LFY<sub>TERE</sub>*, most regions were better bound in *p35S:LFY* (Figure S5A). The coverage reduction for AtLFY<sub>TERE</sub> compared with AtLFY will thereafter be named the TERE-reduction. Binding of AtLFY and AtLFY<sub>TERE</sub> on two genomic regions with low and high TERE-reductions are shown as examples on Figure 5B (more examples are found in Figure S5B). The capacity of LFY SAM domain to oligomerise is thus central for LFY DNA binding *in vivo* but the importance of this property varies between regions.

Additionally, we found that the binding site logos derived from AtLFY and AtLFY<sub>TERE</sub> bound regions were very similar, confirming that there was no major difference between both proteins for individual binding site recognition (Figure S5C), as suggested earlier by SELEX experiments done on LFY proteins containing or not the N-terminal domain (Moyroud et al., 2011; Chahtane et al., 2013; Article 2, submitted).

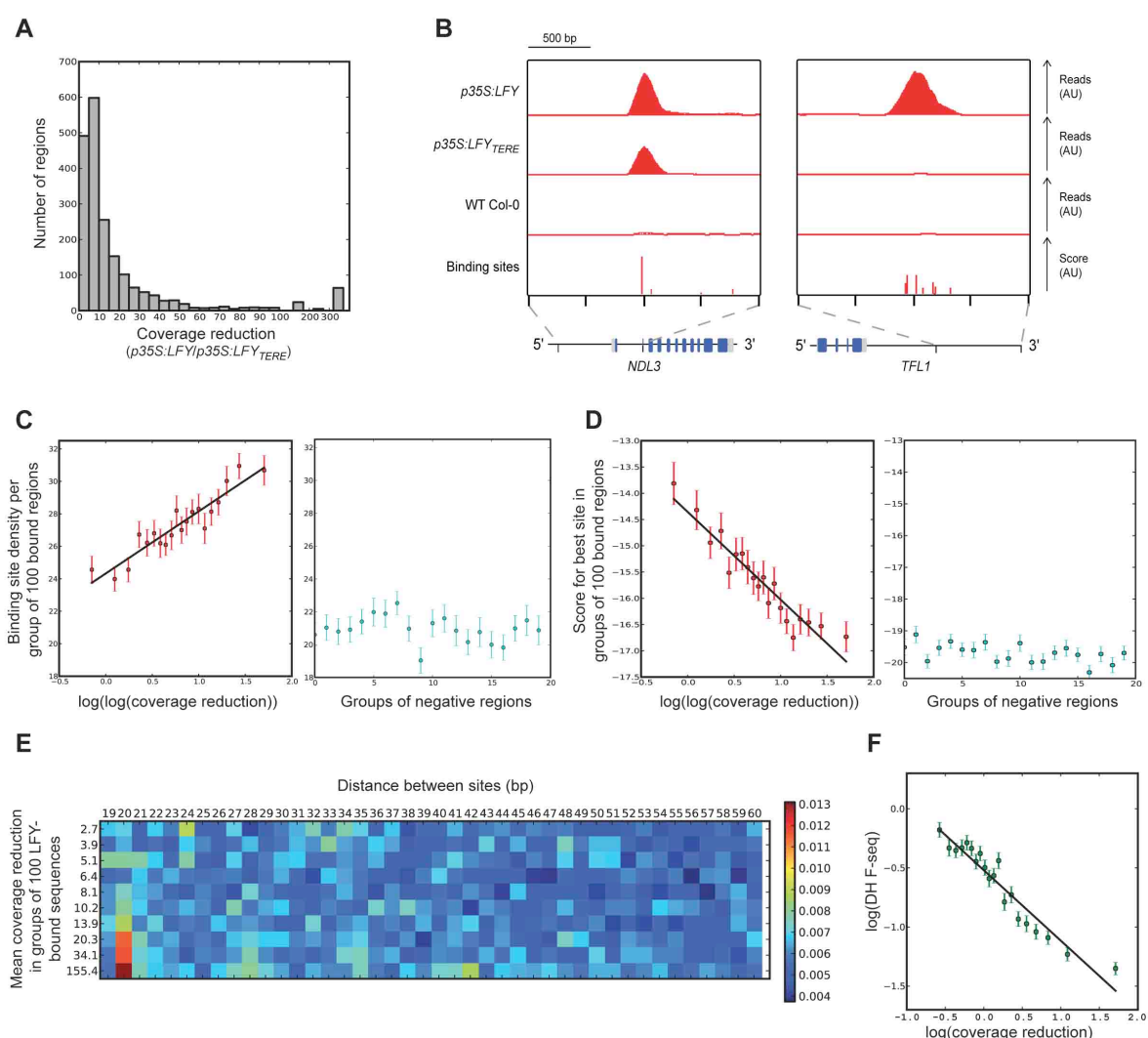


Figure 5. Effects of LFY-SAM oligomerisation on *in planta* DNA binding at the genome wide level from ChIP-seq analysis

(A) Histogram showing the distribution of coverage reductions of  $p35S:LFY_{TERE}$  for the 2000 best-bound regions in  $p35S:LFY$ . Note that the box size on the x-axis is 5 for reduction fold values between 0 and 100 whereas it is 100 for reduction fold values higher than 100. All reduction fold values higher than 300 have been gathered in the last box.

(B) Example of binding of AtLFY and AtLFY<sub>TERE</sub> on two genomic regions. The left panel shows a region with low coverage reduction for AtLFY<sub>TERE</sub> compared with AtLFY (*N-MYC DOWNREGULATED-LIKE 3*, *NDL3*, AT2G19620) and the right panel shows a region where AtLFY<sub>TERE</sub> binding is highly reduced (*TERMINAL FLOWER1*, *TFL1*, AT5G03840) as compared to AtLFY. ChIP-seq read coverage combined from both strands is shown for  $p35S:LFY$ ,  $p35S:LFY_{TERE}$  and WT Col-0 (which shows the background signal). The lower panel shows the LFY binding sites (computed with the model from Moyroud *et al.*, 2011). Noncoding and coding sequences in exons are shown at the bottom as grey and blue boxes respectively. AU: Arbitrary Unit.

(C-D) The 2000 regions bound with the highest confidence in  $p35S:LFY$  were pooled in 20 groups of 100 according to their coverage reduction in  $p35S:LFY_{TERE}$  compared with  $p35S:LFY$  (left panel). A negative set presenting the same size distribution and GC content was also analysed (right panel). (C) Correlation between the mean coverage reduction and the mean number of sites per kb (site density). (D) Correlation between the coverage mean reduction and the mean score for the best LFY binding site per region.

(E) Heat map representing the distribution of distances between LFY binding sites according to the coverage reduction in  $p35S:LFY_{TERE}$  compared to  $p35S:LFY$ . All binding sites with a score higher than -25 were taken into account. The distance is the base pair (bp) number between the first base of each site. The 2000 regions bound with the highest confidence in  $p35S:LFY$  were pooled in 10 groups of 200 regions. In each group, the mean coverage reduction in  $p35S:LFY_{TERE}$  compared to  $p35S:LFY$  was calculated and is shown on the y-axis.

(F) Correlation between the coverage reduction and the chromatin accessibility. The 2000 regions bound with the highest confidence in  $p35S:LFY$  were pooled in 20 groups of 100 according to their coverage reduction in  $p35S:LFY_{TERE}$ ; the F-seq signal derived from DNaseI hypersensitivity data (DH F-seq; Zhang *et al.*, 2012) depicts the chromatin accessibility (low values correspond to closed regions).

### **LFY binding to genomic regions with multiple binding sites highly requires LFY-SAM oligomerisation**

We next tried to understand for which type of regions the SAM oligomerisation was most required.

As our biochemical characterisation suggested that this domain allowed oligomerisation of the LFY protein on multiple binding sites, we examined the LFY binding sites arrangement and tested what features are shared by the regions which most require the SAM domain to be bound. The 2000 regions bound with the highest confidence in  $p35S:LFY$  were partitioned into 20 equal sized bins according to their TERE-reduction. A negative set of non-bound regions was generated for comparison, which shares the same size distributions and GC content as the 2000 bound regions.

We analysed the LFY binding sites, using the previously validated Position Weight Matrix (PWM) that yields scores for any 19 bp sequence, 0 being is the best possible score (binding site of highest affinity) and more negative scores indicating lower affinities (Moyroud *et al.*, 2011). We first examined the bound regions to determine the relevant score threshold to be

used (see methods): we found that a score of -25 was the lower score for specific binding and -14 corresponded to the lowest score for high affinity binding sites.

Comparisons of the most extreme partitions suggested that the regions whose binding was highly SAM dependent possessed, on average, less high affinity binding sites but a higher density of low affinity sites than regions that were less SAM dependent (**Figure S5D**). When analysing LFY binding sites in all 20 partitions, we found a positive correlation between the TERE-reduction and the site density. Whereas a negative set of non-bound regions showed approximately a density of 21 sites/kb, the density increased from 24 sites/kb in the most SAM independent bound regions to 31 sites/kb in the most SAM dependent ones (**Figure 5C**). LFY-SAM oligomerisation is thus more important for binding to regions with a high density of binding sites.

We also analysed the score of the binding sites present in bound regions. For each region, we identified the best LFY binding site and then computed the mean best site score for each partition. In the negative set, the mean best binding site score was between -19 and -20. For most SAM independent regions, this mean binding site score was around -14 and it dropped to about -17 in the SAM dependent regions (**Figure 5D**) (a difference of 3 in score is equivalent to a 20 fold difference in binding site dissociation constant; Moyroud et al., 2011). This correlation shows that the SAM domain is more important to bind to regions bearing binding site of lower affinity than to regions with high affinity binding site.

Together, these analysis of LFY binding sites arrangement in bound regions indicated that oligomerisation through LFY SAM domain mediated binding on regions with a tendency to lack high affinity binding sites but display multiple ones of medium affinity.

As binding to regions with multiple binding sites might depend on distances between the sites, we computed the distribution of these distances in groups of regions with different average TERE-reductions (**Figure 5E**). We considered only LFY binding sites with scores higher than -25, which corresponded to specific binding (**Figure S5D**) and distances ranging from 19 bp (one binding site length) to 60 bp. We found that a spacing of 20 bp was overrepresented in regions with high TERE-reductions. This result shows that, depending on the SAM domain oligomerisation, binding on regions with multiple binding sites preferentially happened when the sites were spaced by 20 bp.

### **LFY-SAM oligomerisation allows access to closed chromatin regions**

As LFY had been proposed to play a role during chromatin remodelling (Wu et al., 2012), we wondered whether the SAM domain could play a role in this process. We took advantage of the fact that DNase I hypersensitivity (DH) was mapped in 2 week-old seedlings (Zhang et

al., 2012), the same developmental stage as for the samples we used for our ChIP-seq. The DH of a genomic region reflects its chromatin accessibility: typically, regions that are sensitive to cleavage by DNase I are nucleosome free.

F-seq values were used to quantitatively characterise the DH level (Boyle et al., 2008): high DH F-seq value indicates an elevated sensitivity to DNase I digestion and thus a high chromatin accessibility. We analysed the DH F-seq, from published data (Zhang et al., 2012), in the 2000 regions bound in *p35S:LFY* pooled in 20 groups according to their TERE-reduction (same region grouping as described above). The DH F-seq strikingly correlated with the TERE-reduction (**Figure 5F**). Indeed, regions with low TERE-reductions showed high DH F-seq values, meaning relatively high chromatin accessibility. On the other hand, the DH F-seq was lower, indicating relatively poor chromatin accessibility, for regions having high TERE-reductions.

The DH can also be considered as a binary parameter (the region can be open or closed) (Zhang et al., 2012). As a complementary analysis, we computed the percentage of closed regions in the 20 groups of regions (**Figure S5E**). This percentage rose from about 10% in low TERE-reduction regions to more than 65% in the most reduced regions, supporting the idea that LFY binding on closed regions is highly dependent on the oligomerisation through the SAM domain.

Altogether, these results showed that LFY binding to closed chromatin regions heavily depends on the presence of the functional SAM domain.

## DISCUSSION

### **The conserved LFY N-terminal domain is a SAM domain which oligomerisation is required for LFY floral function**

In this study, we obtained the crystal structure of LFY conserved N-terminal domain and revealed that it adopts a Sterile Alpha Motif (SAM) fold. We showed that LFY-SAM is an oligomerisation domain and by introducing point mutations, we demonstrated the significance of this property for the function of LFY as a floral master regulator.

LFY-SAM presents several unique features. First, it has no sequence similitude with other SAM domains, making the structural homology hardly predictable. Second, LFY-SAM helical polymer present in the crystal has the largest helical pitch and the widest diameter (with 12 monomers per helix turn instead of 6-8) observed for a SAM polymer. Third, other SAM polymers are mostly held by hydrophobic interactions whereas LFY-SAM oligomerises

through ionic and polar interactions (Kim et al., 2001, 2002, 2005; Qiao et al., 2004; Baron et al., 2006; Harada et al., 2008; Stafford et al., 2011).

Remarkably, LFY-SAM presents a head to tail arrangement, whereas the C-terminal DBD shows a 2-fold symmetry (Hamès et al., 2008). The linker present between the two domains is likely flexible and would permit this organisation. Flexibility was also proposed to conciliate different symmetries for the TEL protein, a human ETS transcriptional regulator, which has a head to tail SAM domain and a DBD able to bind direct as well as inverted binding site repeats (Green et al., 2010).

SAM-containing proteins are involved in a wide range of cellular functions, including transcriptional regulation, signalling, RNA binding, *in vitro* lipid binding and control of cellular localisation (Qiao and Bowie, 2005; Harada et al., 2008). In animals, many transcriptional regulators from the ETS family (Hollenhorst et al., 2011) and from the *Polycomb* group (Kim et al., 2005) carry SAM domains. Within the ETS family, the SAM domain plays versatile roles, as docking module for kinases, polymerisation domain and it also mediates heteromeric interaction between different SAM-containing proteins. These mechanisms are important for the regulation of these proteins activity. For example, in the case of the *Drosophila* ETS transcriptional repressor Yan, the disassembly of the Yan polymer through interaction with Mae, another SAM protein, regulates the transition between its active DNA-bound form and its inactive phosphorylated form (Qiao et al., 2004, 2006). LFY is the first SAM protein characterised in plants, but others are predicted in the *Arabidopsis* genome and it would be worth investigating whether they act as LFY partners.

Oligomerisation of LFY through its SAM-domain is very likely to be a conserved property. Both LFY from *Ginkgo* (gymnosperm) and *Arabidopsis* (angiosperm) oligomerise via their head to tail interface, which is largely conserved from algae to flowering plants. Moreover, chimeric proteins bearing AtLFY-DBD combined to PpLFY-SAM or CrLFY-SAM (from the moss *P. patens* and the fern *Ceratopteris richardii* respectively), complemented well a *lfy* mutant (Maizel et al., 2005), demonstrating the conservation of the SAM domain biochemical properties.

Our data showed that oligomerisation is a key property to LFY floral function in *Arabidopsis*. This is particularly true for LFY switch function as LFY<sub>TERE</sub> poorly rescued a *lfy* mutant and its constitutive expression failed to trigger ectopic floral activation as LFY normally does. LFY variants bearing mutations in the N-terminal region were assessed for dimerisation in a recent study (Siriwardana and Lamb, 2012) assuming that LFY N-terminus was folded as a



leucine zipper motif. Our study strongly suggests that these mutations destabilise LFY folding rather than specifically impairing its oligomerisation.

Oligomerisation was shown to control subcellular localisation of other SAM proteins (Harada et al., 2008; Zhang et al., 2010). This is unlikely the case for LFY as deletions or mutations in the SAM domain did not affect the subcellular localisation nor the trafficking between cells of LFY-GFP fusions *in planta* (Wu et al., 2003; Siriwardana and Lamb, 2012).

### **Oligomerisation through the SAM domain shapes the DNA binding landscape of LFY**

*In vitro* assays pointed out that LFY-SAM plays a role for DNA binding, by increasing the protein selectivity towards DNA thus preventing promiscuous binding, and by promoting cooperative binding between *cis*-elements. Our genome wide DNA binding analysis showed that the ability to bind regions with high binding sites density was highly dependent on the SAM oligomerisation, suggesting that cooperativity was a primary role of LFY SAM domain *in vivo*.

It was known that contacts between LFY-DBD monomers allow LFY dimerisation on a binding site (Hamès et al., 2008) and our biochemical characterisation confirmed that the SAM domain is not required for this. Consistent with this observation, ChIP-seq analysis suggested that LFY oligomerisation is less important for binding to regions with one or few binding sites.

However, we do not exclude that the SAM domain could mediate oligomerisation in the absence of DNA. In EMSA, we observed that the monomeric LFY-DNA complex was less visible when the SAM domain is functional (**Figure 4**), either because it stabilises the dimer on DNA, or because dimer is already formed in solution. Using Bimolecular fluorescence complementation (BiFC), LFY was shown to associate in the cytoplasm of plant cells, where it should not interact with DNA (Siriwardana and Lamb 2012), supporting the later hypothesis.

We found that the presence of a functional SAM domain had an unexpected effect *in vitro*: it prevents promiscuous LFY DNA binding without changing the specificity as shown in SELEX assays (Chahtane et al., 2013; Moyroud et al., 2011) or by comparing ChIP-Seq derived logos for LFY and LFY<sub>TERE</sub>. We propose that the SAM interaction in a dimer makes the DBD more constrained and would thereby prevent the monomeric DNA binding to half sites that are much more frequent in a genome than full-length sites.

We highlighted a key role for LFY SAM oligomerisation domain in cooperative DNA binding. As shown *in vitro* by EMSA (**Figure 4**), only the WT LFY protein was able to efficiently

occupy both binding sites on a double-site probe, as compared to a dimeric variant (LFY<sub>[TE+RE]</sub>). Thus, the SAM domain allows LFY to oligomerise further than a dimer, and its head to tail arrangement makes perfect sense here. This property also appeared as crucial for LFY binding on the genome, since LFY<sub>TERE</sub> binding was dramatically reduced on regions displaying a high density of binding sites.

Cooperative DNA binding was suggested for other SAM proteins. *In vitro*, TEL self-association facilitates cooperative binding to DNA (Green et al., 2010). This property was proposed to allow spreading along chromatin to form large repressive domains. This hypothesis was tested for Yan, the *Drosophila* homolog of TEL, by comparing genome-wide chromatin occupancy of endogenous WT Yan to that of a monomeric variant (Webber et al., 2013). Surprisingly, SAM-mediated oligomerisation did not appear as the primary determinant for Yan spreading over extended chromatin regions. Still, oligomerisation was proven to be essential for Yan function, suggesting that it could be involved in transcriptional regulation downstream DNA binding. Whether other SAM proteins cooperatively bind large genomic regions thanks to oligomerisation remains to be clarified.

Here we pointed out interaction between binding sites located on the same DNA molecule or on the same genomic region. Still, oligomerisation could also mediate DNA looping between distant regions. This issue may be addressed *in vitro* by imaging techniques (Gietl and Grohmann, 2013) and *in vivo* by chromosome conformation capture methods (Yadon and Tsukiyama, 2013).

Our genome-wide study, combined with recently published maps of DNase I hypersensitive in *Arabidopsis* (Zhang et al., 2012), revealed that our LFY variant impaired in oligomerisation (LFY<sub>TERE</sub>) was also impaired in binding regions where chromatin is thickly packed. Hence, oligomerisation through the SAM domain allows LFY to access these closed regions.

Regulatory regions embedded in closed chromatin are usually poorly accessed by transcription factors. Still, to initiate novel regulatory programs, these regions must be accessed by some factors that bind, modify the chromatin state and affect the transcriptional activity. Whether pre-existing chromatin conformation allows transcription factors binding or whether transcription factors binding affect the local chromatin environment is a key question. Most transcription factors need a favourable chromatin structure to bind. However, a handful of factors, termed “pioneer transcription factors”, can be the first to engage in packed chromatin (Zaret and Carroll, 2011; Guertin and Lis, 2012). Such initial binding can modify the nucleosome organisation, either directly, as shown for the FoxA transcription factor (Cirillo et al., 2002), or indirectly by recruiting remodelling factors (Zaret and Carroll, 2011; Guertin and Lis, 2012).

LFY was shown to recruit the SWI2/SNF2 chromatin-remodeling ATPases SYD and BRM to some regulatory regions (Wu et al., 2012) and could thus indirectly drive sequence specific chromatin remodelling.

Therefore LFY appears to be a good candidate to be a novel, plant specific pioneer transcription factor, able to defy packed chromatin thanks to its oligomerisation ability and to modify chromatin through interaction with remodelling factors. Such a property would be central to LFY switch function and remains to be investigated.

#### **SUPPLEMENTAL INFORMATION**

This section, below the references, includes Table S1-S4 and Figure S1-S5

#### **SUPPLEMENTAL EXPERIMENTAL PROCEDURES**

This section, at the end of the paper, includes Supplementary Table 1-3 and Supplementary Figure 1, which refer to the methods used in the study

#### **ACKNOWLEDGEMENTS**

We thank Anthony Mathelier, Anthony Fejes, Wyeth W. Wasserman, Carlo Petosa and Patrick Laufs and members of the Floral regulators team for helpful discussions, Pierre Jacob, Adam Round, Guy Schoehn, Daphna Fenel for help with protein characterisation. This work was supported by the ANR (Blanc – SVSE2 – 2011) to F.P., and the Irtelis PhD program (C.S).

#### **EXPERIMENTAL PROCEDURES**

##### **Plant material and growth conditions**

WT, mutants and transgenic lines used are in the *A. thaliana* Columbia-0 background. Seeds were sown on 0.5X Murashige and Skoog (MS; duchefa biochimie) basal salt mixture medium. Plates were kept 3 days at 4°C for stratification, grown at 22°C under long-day conditions and seedlings were transferred to soil. Genotyping to identify the *Ify-12* mutation was performed as described in Benlloch et al., 2011.

##### **Plant transformation**

*A. tumefaciens* C58C1 pMP90 was used for stable transformation of *Ify-12* heterozygous plants by the floral dip method as described by Zhang et al., 2006.

### Plasmid constructions for expression in *Escherichia coli*

GbLFY $\Delta$  (*pETH164*; residues 55-402 from *G. biloba* LFY cDNA) was cloned in *pETM-11* vector (Dummler et al., 2005) to add a 6xHis tag in N-terminal, cleavable by the Tobacco Etch Virus (TEV) protease. AtLFY (*pETH94*; residues 1-420 from *A. thaliana* LFY cDNA) was cloned in *pET-30a* vector (Novagen) to add a 6xHis tag in C-terminal.

GbLFY-SAM (*pETH195*; residues 54-159 from *G. biloba* LFY cDNA) and AtLFY-SAM (*pETH201*; residues 38-151 from *A. thaliana* LFY cDNA) were cloned in *pETM-11* vector.

Mutagenesis was done using the QuichChange site directed mutagenesis (Agilent) strategy with appropriate primers, listed in **Supplementary Table 1**.

GbLFY $\Delta_{TE}$  (*pCA21*), GbLFY $\Delta_{RE}$  (*pCA22*), GbLFY $\Delta_{TERE}$  (*pCA39*) were derived from *pETH164*; AtLFY $_{TE}$  (*pCA23*), AtLFY $_{RE}$  (*pCA24*), AtLFY $_{TERE}$  (*pCA25*) were derived from *pETH94*; GbLFY-SAM $_{TE}$  (*pCA15*), GbLFY-SAM $_{RE}$  (*pCA17*), GbLFY-SAM $_{TERE}$  (*pCA20*) were derived from *pETH195*.

All plasmids were verified by appropriate digestion or by sequencing.

### Binary vector constructions for *A. thaliana* transformation

*pLFY:LFY $_{TERE}$*  (*pCA35*) and *p35S:LFY $_{TERE}$*  (*pCA29*) were obtained by inserting the PstI-Sall digestion fragment, containing the TERE mutations, from *pCA25* in the PstI-Sall digestion backbone of *pETH29* and *pCA26* (Chahtane et al., 2013) respectively.

### Protein expression

Proteins were expressed using *E. coli* Rosetta2 (DE3) strain (Novagen). Cells were grown in LB medium supplemented with Kanamycine (50  $\mu$ g/mL) and Chloramphenicol (34  $\mu$ g/mL) at 37 °C under agitation up to an O.D.<sub>600nm</sub> of 0.6. 2 mM Betaine were then added and cultures were shifted to 17 °C for 1 h before addition of 0.4 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). After overnight incubation growth at 17 °C, cells were pelleted. Production of selenomethionine (Se-Met) GbLFY-SAM for crystallography was performed as described by Robin et al., 2010.

### Protein purification

All purification steps were performed at 4°C. All proteins were solubilised and first purified by an affinity chromatography column. The buffer composition (buffer X) used for each protein is given in **Supplementary Table 2**. Pellets corresponding to 0.5 L culture containing the recombinant protein were sonicated in 50 mL buffer X supplemented by one protease inhibitor cocktail tablet Complete EDTA-free (Roche) and centrifuged for 30 min at 20 000 g. The clear supernatant was transferred on a column containing 1 mL Ni-Sepharose High Performance resin (GE Healthcare), washed with buffer X containing 20 and 40 mM imidazole and eluted with buffer X containing 300 mM imidazole. Eluted fractions were immediately diluted 3 times in buffer X and dialysed overnight against buffer X.

For GbLFY-SAM, GbLFY-SAM $_{TE}$ , GbLFY-SAM $_{RE}$ , GbLFY-SAM $_{TERE}$  and AtLFY-SAM, after dialysis, the 6xHis tag was cleaved overnight by the TEV protease (5% w/w). The cleavage product was loaded on a Ni-Sepharose High Performance resin (GE Healthcare) to remove the 6xHis tag, the remaining non-cleaved and the TEV protease (also carrying a 6xHis tag).

The protein was eluted with buffer A (**Supplementary Table 2**) containing 20 mM imidazole. Eluted fractions were applied to a Hi-load Superdex-200 16/60 prep grade column (GE Healthcare) equilibrated with buffer A. After concentrated using Amicon Ultra Centrifugal filters (Millipore), protein concentration was assessed using a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific Inc). For crystallography, the Se-Met GbLFY-SAM was purified as the native protein.

After dialysis in buffer B (**Supplementary Table 2**), GbLFY $\Delta$ , GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$ , GbLFY $\Delta_{TERE}$  proteins were concentrated using Amicon Ultra Centrifugal filters (Millipore) and applied to a Hi-load Superdex-200 16/60 prep grade column (GE Healthcare) equilibrated with buffer C (**Supplementary Table 2**). After concentrated using Amicon Ultra Centrifugal filters (Millipore), protein concentration was assessed using a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific Inc.).

After dialysis in buffer D (**Supplementary Table 2**), AtLFY, AtLFY $_{TE}$ , AtLFY $_{RE}$ , AtLFY $_{TERE}$  proteins were concentrated using Amicon Ultra Centrifugal filters (Millipore). Because this purification protocol without NaCl led to protein containing some bacterial DNA contamination, the protein concentration could not be determined by spectrometric method and was instead assessed using the Bradford assay (Bradford, 1976).

All proteins were flash-frozen in liquid nitrogen and stored at -80°C, which did not alter their properties.

### Protein crystallisation

Initial crystallisation conditions were identified using the High Throughput Crystallisation (HTX) platform at EMBL Grenoble ([embl.fr/htxlab](http://embl.fr/htxlab)). The optimum condition was obtained at 20°C with the hanging drop vapor diffusion method by mixing 1  $\mu$ L GbLFY-N, native or Se-Met substituted, at 2.5 mg/mL with 1  $\mu$ L of reservoir solution containing 25 mM Tris-HCL pH 8.8, 40mM ammonium sulphate. Crystals were cryoprotected by plunging into liquid nitrogen, after incubation in well conditions to which 20% glycerol had been supplemented.

### Crystal structure determination

The structure was solved by the single wavelength anomalous dispersion (SAD) method from Se-Met substituted protein. A highly redundant dataset was collected on the microfocus beamline ID23-2 (Flot et al., 2010). The selenium substructure was determined by SHELXC/D (Sheldrick, 2010) and this selenium positions were refined in SHARP (Bricogne et al., 2003). Phases were calculated, and solvent flattened maps were obtained with PIRATE (Cowtan, 2002), and a partial model was automatically built with BUCCANEER (Cowtan, 2006). The model was extended into the experimental electron density map and the resultant model was used as a search model for molecular replacement into the native data (preserving and extending the reflections previously selected for the free set). Successive rounds of model improvement in COOT (Emsley and Cowtan, 2004) and refinement in BUSTER (Bricogne et al., 2011) were then performed. The built region in one and the other of the 2 monomers present in the asymmetric unit comprised residues 54-134 and 56-133 respectively.

### Protein sequence and structure analysis

3D structure images were done using PyMOL ([www.pymol.org](http://www.pymol.org)), sequence alignment was done using Multalin (<http://multalin.toulouse.inra.fr/multalin/>) and visualised using ESPrit (<http://esprict.ibcp.fr/ESPrict/ESPrict/>), RMSD were computed using the Dali server ([http://ekhidna.biocenter.helsinki.fi/dali\\_server/](http://ekhidna.biocenter.helsinki.fi/dali_server/)), surface conservation was computed using the Consurf server (<http://consurf.tau.ac.il/>; Ashkenazy et al., 2010).

### EMSA

Oligonucleotides used for EMSA are listed in **Supplementary Table 3**. *AP1*, *AP1m1*, *AP1m2* oligonucleotides were annealed and labelled with Cy3-dCTP as described in Moyroud et al., 2011. *AGI-III* was amplified by PCR using oCA0042-Tamra (CGTTTAAATTTAATCCAATGG) and oCA0043 (CAACCCATTAACACATTGGG) primers.

EMSA and signal quantification were performed as described in Moyroud et al., 2011. Binding reactions were performed in 20  $\mu$ L binding buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% glycerol, 0.25 mM EDTA, 2 mM MgCl<sub>2</sub>, 0.01% Tween-20 and 3 mM TCEP) with 10 nM Cy3-dsDNA, 1 X (28 ng/mL) fish sperm DNA (Roche) as non-specific competitor, and 25 to 500 nM proteins.

Competition assays were performed in duplicates, 1 to 100 X fish sperm DNA (Roche) were used in the binding reaction.

### SEC

The molecular mass of GbLFY-SAM and AtLFY-SAM was estimated using a Superdex-200 10/300GL column (GE Healthcare), equilibrated with buffer A (20 mM Tris-HCl pH 8, 1mM TCEP) and calibrated with low and high molecular weight protein standards (gel filtration calibration kit; GE Healthcare).

### SEC-MALLS

Accurate molecular mass determination was conducted as described in Hamès et al., 2008. GbLFY-SAM, WT and mutants were analysed in buffer A. Protein-DNA complexes containing GbLFY $\Delta$ , WT or mutants and *AP1* DNA were analysed in 20mM Tris-HCl pH 8; 150 mM NaCl; 0,25 mM EDTA; 2 mM MgCl<sub>2</sub>; 1 mM TCEP.

### Western-blot

*Arabidopsis* leaves from 15 day-old seedlings grown under long day conditions were harvested, one leaf from 6 to 8 seedlings were pooled for each transgenic line. The leaves were flash frozen in liquid nitrogen, mechanically crushed, immediately suspended in 80  $\mu$ L denaturing buffer (50 mM Tris-HCl pH 6.8; 10% w/v glycerol; 1% w/v Sodium Dodecyl Sulfate; 0.0025% w/v Bromophenol blue; 0.4% w/v Dithiothreitol) for 10 mg of fresh matter, and denatured for 5 min at 95°C. The sample were run on a classical SDS-PAGE gel and transferred on a nitrocellulose membrane (Immobilon P transfert membrane, Millipore) for blotting. The equivalent of 18.75 mg and 12.5 mg of fresh matter were used for anti-LFY and anti-KARI blots respectively. 5ng of recombinant LFY protein (LFY $\Delta$ , produced in *E. coli*) were used as a positive control. Anti-LFY antibodies (named JA70) were raised in rabbit

against recombinant LFY protein, anti-KARI antibodies were described in Dumas et al., 1989, secondary antibodies against rabbit were used for detection. Revelation was done using the Pierce ECL2 Western Blotting Substrate kit (Thermo Scientific) and scanned on a Typhoon 9400 scanner (Molecular Dynamics).

### Cross-Linking, Chromatin Isolation, and ChIP-seq

The ChIP-seq was done as described in Moyroud et al., 2011. 15 days old seedling from lines CA26-15 (*p35S:LFY*), CA29-17 (*p35S:LFY<sub>TERE</sub>*) and Col-0 were grown on 0.5 X MS medium under long day conditions. The experiment was performed twice to produce biological replicates.

### ChIP-seq analysis

All analyses were performed using the SHORE suite (Shore Software Solutions) as well as python programs using the BioPython ([http://biopython.org/wiki/Main\\_Page](http://biopython.org/wiki/Main_Page)), Matplotlib (<http://matplotlib.org/>) and pybedtools (<http://pythonhosted.org/pybedtools/>) modules. All figures were built using Matplotlib, Inkscape and Adobe Illustrator.

#### - ChIP-Seq mapping and normalisation

ChIP-Seq experiment was performed on 3 different samples in duplicate following the procedure previously described (Moyroud et al., 2011). Read mapping was performed using the SHORE pipeline using -H 1,1 -M 0,4 -X 130 --wpoiss=30 as parameters. The number of reads mapped is indicated in **Supplementary Table 4**.

In order to compensate for differences in number of reads, two normalisation steps were applied. Intra-genotype normalisation was performed based on the height or the coverage on the peaks with highest signal. SHORE peaks was used to detect the 2000 peak with highest confidence in *p35S:LFY* and *p35S:LFY<sub>TERE</sub>* compared to the Col-0 negative samples. A rudimentary python procedure was applied to identify  $\approx 450$  background peaks in the Col-0 negative samples. Peaks height and coverage was computed and a linear fit using the least square method within genotypes was performed to obtain the intra-genotype normalisation coefficients  $m$  (**Supplementary Figure 1A**). Fit based on peak coverage and peak height gave very similar normalisation coefficient and read coverage was used for further analysis. A single normalised signal for each genotype was computed.

Signal(genotype) =  $\frac{1}{2}$  (signal1(genotype)  $m$  + signal2(genotype) for genotype in [Col-0, *p35S:LFY* and *p35S:LFY<sub>TERE</sub>*]

$m = [0.52575494, 1.07670343, 1.74887865]$  for [35S, TERE, Col]

The internormalisation step was proceeded the same way using the normalised background peak signal in Col-0, *p35S:LFY* and *p35S:LFY<sub>TERE</sub>* signal (**Supplementary Figure 1B**). All subsequent analysis were performed on normalised read counts.

tc, sc = 1.08271353333, 1.21276145918

Col = tc \* TERE Col = sc \*35S

#### - Data analyses

Analyses were performed based on peaks read coverage for the 2000 peaks of highest confidence. Similar analyses were performed on peak height or on the 4000 peaks and yielded similar results and identical conclusions.

Read coverage was computed in *p35S:LFY* and *p35S:LFY<sub>TERE</sub>* samples for the 2000 peaks of highest confidence detected in *p35S:LFY* sample. The binding reduction *r* due to the TERE mutation was computed using the following formula:

$$r = [\text{coverage}(p35S:LFY) - \text{coverage}(Col)] / [\text{coverage}(p35S:LFY_{TERE}) - \text{coverage}(Col)]$$

To avoid extreme *r* values, *r* was arbitrarily set to 300 whenever higher than 300 or when the denominator was negative.

The IGB browser (<http://bioviz.org/igb/>) was used for data representation using the normalisation coefficient computed. Samples 2 are shown on Figure 5 and S5. Logos presented in Figure S5 were generated using MEME (Bailey and Elkan, 1994); <http://meme.nbcr.net/meme/>) on the top 500 bound regions of 120 bp centered on the peak maximum of each region. Computing LFY binding site score was performed using the previously described model that takes dependency between positions into account (Moyroud et al., 2011). In order to determine the score threshold that were used in subsequent analyses, two sets of 200 regions bound in *p35S:LFY* sample were compared: the first set corresponded to the regions with the lowest TERE-reduction in *p35S:LFY<sub>TERE</sub>* and the second to the regions with the highest TERE-reduction. We also use a negative set of 2000 non-bound regions selected in the genome that respect the size distribution and the GC content of the 2000 bound regions. Comparison of LFY binding site density in different score ranges in the three samples allowed to the two site score boundaries (-14 and -25).

Analyses involving DNase I sensitivity signal used published list of open regions and quantitative F-seq signal in seedling or flower tissue (Zhang et al., 2012).

## REFERENCES

- Alvarez-Venegas, R., Pien, S., Sadler, M., Witmer, X., Grossniklaus, U., Avramova, Z. (2003). ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Curr Biol* 13, 627-637.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38, W529-533.
- Bailey, T. L., Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2, 28-36.
- Baron, M. K., Boeckers, T. M., Vaida, B., Faham, S., Gingery, M., Sawaya, M. R., Salyer, D., Gundelfinger, E. D., Bowie, J. U. (2006). An architectural framework that may lie at the core of the postsynaptic density. *Science* 311, 531-535.
- Bell, O., Tiwari, V. K., Thoma, N. H., Schubeler, D. (2011). Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12, 554-564.
- Benlloch, R., Kim, M. C., Sayou, C., Thevenon, E., Parcy, F., Nilsson, O. (2011). Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. *Plant J* 67, 1094-1102.
- Boyle, A. P., Guinney, J., Crawford, G. E., Furey, T. S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* 24, 2537-2538.
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-254.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C., Womack, T. O. (2011). BUSTER version 2.8.0. Cambridge, United Kingdom: Global Phasing Ltd.



- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* **59**, 2023-2030.
- Busch, M. A., Bomblies, K., Weigel, D. (1999). Activation of a floral homeotic gene in *Arabidopsis*. *Science* **285**, 585-587.
- Calonje, M., Sanchez, R., Chen, L., Sung, Z. R. (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in *Arabidopsis*. *Plant Cell* **20**, 277-291.
- Carles, C. C., Fletcher, J. C. (2009). The SAND domain protein ULTRAPETALA1 acts as a trithorax group factor to regulate cell fate in plants. *Genes Dev* **23**, 2723-2728.
- Chae, E., Tan, Q. K., Hill, T. A., Irish, V. F. (2008). An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development* **135**, 1235-1245.
- Chahtane, H., Vachon, G., Le Masson, M., Thevenon, E., Perigon, S., Mihajlovic, N., Kalinina, A., Michard, R., Moyroud, E., Monniaux, M., Sayou, C., Grbic, V., Parcy, F., Tichtinsky, G. (2013). A variant of LEAFY reveals its capacity to stimulate meristem development by inducing RAX1. *Plant J* **74**, 678-689.
- Chandler, J. W. (2012). Floral meristem initiation and emergence in plants. *Cell Mol Life Sci* **69**, 3807-3818.
- Chen, L., Cheng, J. C., Castle, L., Sung, Z. R. (1997). EMF genes regulate *Arabidopsis* inflorescence development. *Plant Cell* **9**, 2011-2024.
- Chou, M. L., Haung, M. D., Yang, C. H. (2001). EMF genes interact with late-flowering genes in regulating floral initiation genes during shoot development in *Arabidopsis thaliana*. *Plant Cell Physiol* **42**, 499-507.
- Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279-289.
- Clapier, C. R., Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**, 273-304.
- Cowtan, K. (2002). Generic representation and evaluation of properties as a function of position in reciprocal space. *J Appl Cryst* **35**, 655-663.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011.
- Dumas, R., Joyard, J., Douce, R. (1989). Purification and characterization of acetohydroxyacid reductoisomerase from spinach chloroplasts. *Biochem J* **262**, 971-976.
- Dummler, A., Lawrence, A. M., de Marco, A. (2005). Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact* **4**, 34.
- Emsley, P., Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.
- Flot, D., Mairs, T., Giraud, T., Gujjarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S., Mitchell, E. (2010). The ID23-2 structural biology microfocus beamline at the ESRF. *J Synchrotron Radiat* **17**, 107-118.
- Gietl, A., Grohmann, D. (2013). Modern biophysical approaches probe transcription-factor-induced DNA bending and looping. *Biochem Soc Trans* **41**, 368-373.
- Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M., Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* **386**, 44-51.
- Green, S. M., Coyne, H. J., 3rd, McIntosh, L. P., Graves, B. J. (2010). DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* **285**, 18496-18504.
- Guertin, M. J., Lis, J. T. (2012). Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr Opin Genet Dev* **23**, 116-123.
- Hamès, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F., Martiel, J. L., Benlloch, R., Parcy, F., Muller, C. W. (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J* **27**, 2628-2637.
- Harada, B. T., Knight, M. J., Imai, S., Qiao, F., Ramachander, R., Sawaya, M. R., Gingery, M., Sakane, F., Bowie, J. U. (2008). Regulation of enzyme localization by polymerization: polymer formation by the SAM domain of diacylglycerol kinase delta1. *Structure* **16**, 380-387.
- Hollenhorst, P. C., McIntosh, L. P., Graves, B. J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* **80**, 437-471.
- Ikeda-Kawakatsu, K., Maekawa, M., Izawa, T., Itoh, J., Nagato, Y. (2012). ABERRANT PANICLE ORGANIZATION 2/RFL, the rice ortholog of *Arabidopsis* LEAFY, suppresses the transition from inflorescence meristem to floral meristem through interaction with APO1. *Plant J* **69**, 168-180.

- Irish, V. F. (2010). The flowering of Arabidopsis flower development. *Plant J* 61, 1014-1028.
- Jaeger, K. E., Pullen, N., Lamzin, S., Morris, R. J., Wigge, P. A. (2013). Interlocking feedback loops govern the dynamic behavior of the floral transition in Arabidopsis. *Plant Cell* 25, 820-833.
- Katz, A., Oliva, M., Mosquna, A., Hakim, O., Ohad, N. (2004). FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. *Plant J* 37, 707-719.
- Kim, C. A., Gingery, M., Pilpa, R. M., Bowie, J. U. (2002). The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* 9, 453-457.
- Kim, C. A., Phillips, M. L., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A., Faham, S., Bowie, J. U. (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *Embo J* 20, 4173-4182.
- Kim, C. A., Sawaya, M. R., Cascio, D., Kim, W., Bowie, J. U. (2005). Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *J Biol Chem* 280, 27769-27775.
- Lamb, R. S., Hill, T. A., Tan, Q. K., Irish, V. F. (2002). Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development* 129, 2079-2086.
- Levin, J. Z., Meyerowitz, E. M. (1995). *UFO*: an Arabidopsis gene involved in both floral meristem and floral organ development. *Plant Cell* 7, 529-548.
- Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S., Yanofsky, M. F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* 11, 1007-1018.
- Liu, C., Xi, W., Shen, L., Tan, C., Yu, H. (2009). Regulation of floral patterning by flowering time genes. *Dev Cell* 16, 711-722.
- Lohmann, J. U., Hong, R. L., Hobe, M., Busch, M. A., Parcy, F., Simon, R., Weigel, D. (2001). A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* 105, 793-803.
- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., Weigel, D. (2005). The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260-263.
- Maizel, A., Weigel, D. (2004). Temporally and spatially controlled induction of gene expression in Arabidopsis thaliana. *Plant J* 38, 164-171.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R., Parcy, F. (2010). LEAFY blossoms. *Trends Plant Sci* 15, 346-352.
- Moyroud, E., Minguet, E. G., Ott, F., Yant, L., Pose, D., Monniaux, M., Blanchet, S., Bastien, O., Thevenon, E., Weigel, D., Schmid, M., Parcy, F. (2011). Prediction of Regulatory Interactions from Genome Sequences Using a Biophysical Model for the Arabidopsis LEAFY Transcription Factor. *Plant Cell*.
- Moyroud, E., Tichtinsky, G., Parcy, F. (2009). The LEAFY floral regulators in Angiosperms: Conserved proteins with diverse roles. *J Plant Biol* 52.
- Parcy, F., Bomblies, K., Weigel, D. (2002). Interaction of LEAFY, AGAMOUS and TERMINAL FLOWER1 in maintaining floral meristem identity in Arabidopsis. *Development* 129, 2519-2527.
- Parcy, F., Nilsson, O., Busch, M. A., Lee, I., Weigel, D. (1998). A genetic framework for floral patterning. *Nature* 395, 561-566.
- Posé, D., Yant, L., Schmid, M. (2012). The end of innocence: flowering networks explode in complexity. *Curr Opin Plant Biol* 15, 45-50.
- Qiao, F., Bowie, J. U. (2005). The many faces of SAM. *Sci STKE* 2005, re7.
- Qiao, F., Harada, B., Song, H., Whitelegge, J., Courey, A. J., Bowie, J. U. (2006). Mae inhibits Pointed-P2 transcriptional activity by blocking its MAPK docking site. *Embo J* 25, 70-79.
- Qiao, F., Song, H., Kim, C. A., Sawaya, M. R., Hunter, J. B., Gingery, M., Rebay, I., Courey, A. J., Bowie, J. U. (2004). Derepression by depolymerization; structural insights into the regulation of Yan by Mae. *Cell* 118, 163-173.
- Ratcliffe, O. J., Amaya, I., Vincent, C. A., Rothstein, S., Carpenter, R., Coen, E. S., Bradley, D. J. (1998). A common mechanism controls the life cycle and architecture of plants. *Development* 125, 1609-1615.
- Robin, A. Y., Cobessi, D., Curien, G., Robert-Genthon, M., Ferrer, J. L., Dumas, R. (2010). A new mode of dimerization of allosteric enzymes with ACT domains revealed by the crystal structure of the aspartate kinase from Cyanobacteria. *J Mol Biol* 399, 283-293.
- Saleh, A., Al-Abdallat, A., Ndamukong, I., Alvarez-Venegas, R., Avramova, Z. (2007). The Arabidopsis homologs of trithorax (ATX1) and enhancer of zeste (CLF) establish 'bivalent chromatin marks' at the silent AGAMOUS locus. *Nucleic Acids Res* 35, 6290-6296.
- Sathyamurthy, A., Freund, S. M., Johnson, C. M., Allen, M. D., Bycroft, M. (2011). Structural basis of p63alpha SAM domain mutants involved in AEC syndrome. *Febs J* 278, 2680-2688.

- Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 479-485.
- Siriwardana, N. S., Lamb, R. S. (2012). A conserved domain in the N-terminus is important for LEAFY dimerization and function in *Arabidopsis thaliana*. *Plant J* **71**, 736-749.
- Souer, E., Rebocho, A. B., Blied, M., Kusters, E., de Bruin, R. A., Koes, R. (2008). Patterning of inflorescences and flowers by the F-Box protein DOUBLE TOP and the LEAFY homolog ABERRANT LEAF AND FLOWER of *petunia*. *Plant Cell* **20**, 2033-2048.
- Stafford, R. L., Hinde, E., Knight, M. J., Pennella, M. A., Ear, J., Digman, M. A., Gratton, E., Bowie, J. U. (2011). Tandem SAM domain structure of human Caskin1: a presynaptic, self-assembling scaffold for CASK. *Structure* **19**, 1826-1836.
- Tanahashi, T., Sumikawa, N., Kato, M., Hasebe, M. (2005). Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* **132**, 1727-1736.
- Wagner, D., Sablowski, R. W., Meyerowitz, E. M. (1999). Transcriptional activation of *APETALA1* by *LEAFY*. *Science* **285**, 582-584.
- Webber, J. L., Zhang, J., Cote, L., Vivekanand, P., Ni, X., Zhou, J., Negre, N., Carthew, R. W., White, K. P., Rebay, I. (2013). The relationship between long-range chromatin occupancy and polymerization of the *Drosophila* ETS family transcriptional repressor Yan. *Genetics* **193**, 633-649.
- Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F., Meyerowitz, E. M. (1992). *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843-859.
- Weigel, D., Meyerowitz, E. M. (1993). Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**, 1723-1726.
- Weigel, D., Nilsson, O. (1995). A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495-500.
- Winter, C. M., Austin, R. S., Blanvillain-Baufume, S., Reback, M. A., Monniaux, M., Wu, M. F., Sang, Y., Yamaguchi, A., Yamaguchi, N., Parker, J. E., Parcy, F., Jensen, S. T., Li, H., Wagner, D. (2011). *LEAFY* Target Genes Reveal Floral Regulatory Logic, cis Motifs, and a Link to Biotic Stimulus Response. *Dev Cell* **20**, 430-443.
- Wu, M. F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S. K., Li, Z., Su, Y., Slewinski, T. L., Wagner, D. (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the *LEAFY* and *SEPALLATA3* transcription factors. *Proc Natl Acad Sci U S A* **109**, 3576-3581.
- Wu, X., Dinneny, J. R., Crawford, K. M., Rhee, Y., Citovsky, V., Zambryski, P. C., Weigel, D. (2003). Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development* **130**, 3735-3745.
- Yadon, A. N., Tsukiyama, T. (2013). DNA looping-dependent targeting of a chromatin remodeling factor. *Cell Cycle* **12**, 1809-1810.
- Yamaguchi, N., Wu, M. F., Winter, C. M., Berns, M. C., Nole-Wilson, S., Yamaguchi, A., Coupland, G., Krizek, B. A., Wagner, D. (2013). A molecular framework for auxin-mediated initiation of flower primordia. *Dev Cell* **24**, 271-282.
- Zaret, K. S., Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-2241.
- Zhang, J., Graham, T. G., Vivekanand, P., Cote, L., Cetera, M., Rebay, I. (2010). Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the *Drosophila* ETS family transcriptional repressor Yan. *Mol Cell Biol* **30**, 1158-1170.
- Zhang, W., Zhang, T., Wu, Y., Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* **24**, 2719-2731.
- Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., Chua, N. H. (2006). *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc* **1**, 641-646.

## SUPPLEMENTAL INFORMATION

**Table S1. Molecular mass determination of GbLFY-SAM, WT and mutants, at different concentrations**

Protein	Concentration (mg/mL)	Molecular mass (kDa)	Monomers number
GbLFY-SAM	1.6	78 (+/- 2.2)	6.1
	4.7	95 (+/- 1.6)	7.4
	5.8	97.3 (+/- 2.5)	7.6
	8.4	103.3 (+/- 2.1)	8.1
GbLFY-SAM <sub>TE</sub>	4.6	12.5 (+/- 0.2)	1.0
GbLFY-SAM <sub>RE</sub>	4.4	12.9 (+/- 0.2)	1.0
GbLFY-SAM <sub>TERE</sub>	4.5	13.6 (+/- 0.8)	1.1
GbLFY-SAM <sub>[TE+RE]</sub>	4.7	23 (+/- 1)	1.8

The molecular mass of GbLFY-SAM, GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub>, GbLFY-SAM<sub>TERE</sub> and GbLFY-SAM<sub>[TE+RE]</sub> (equimolar mixture of GbLFY-SAM<sub>TE</sub> and GbLFY-SAM<sub>RE</sub>) was determined using SEC-MALLS. A volume of 50  $\mu$ L was used for each protein. The protein concentration used, the measured molecular mass for each protein (the number in brackets is the standard error), and the calculated number of molecule per complex are indicated. The theoretical molecular mass of a GbLFY-SAM monomer is 12.8 kDa.

**Table S2. Ability of LFY and LFY<sub>TERE</sub> to complement a *lfy-12* null mutant**

	Phenotype	Rescue level			Total
		Low	Average	High	
<i>pLFY:LFY</i>	Plant number	1	2	8	11
	Plant %	9,1	18,2	72,7	100
	Mean silique number	0	2,4	>20	
<i>pLFY:LFY<sub>TERE</sub></i>	Plant number	1	13	3	17
	Plant %	5,9	76,5	17,6	100
	Mean silique number	0	9,5	>20	

*pLFY:LFY* and *pLFY:LFY<sub>TERE</sub>* T1 plant effective for each phenotypic class. Low rescue level corresponds to plants strongly resembling *lfy-12* mutants, average rescue level corresponds to plants with structures lacking proper floral organs and mostly sterile and high rescue level corresponds plants with fertile flowers displaying correctly arranged floral organs.

**Table S3. Effect of LFY and LFY<sub>TERE</sub> constitutive expression**

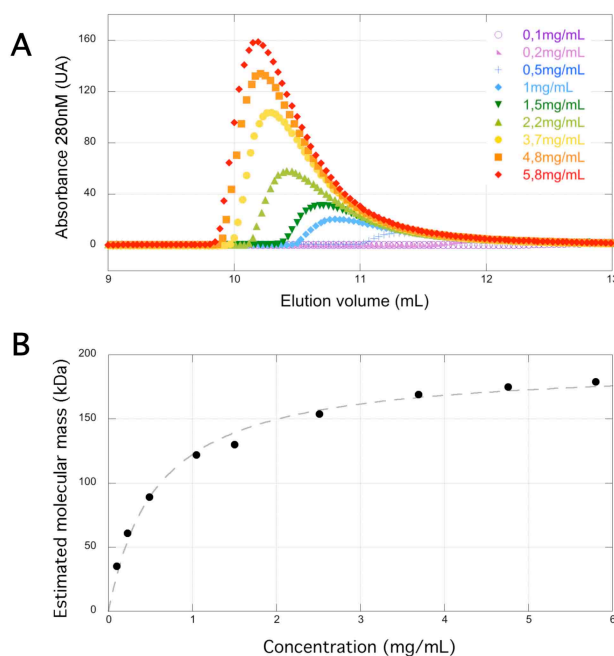
	Phenotype	Ectopic flowers	No ectopic flower	Total
<i>p35S:LFY</i>	Plant number	30	31	61
	Plant %	49,2	50,8	100
<i>p35S:LFY<sub>TERE</sub></i>	Plant number	0	51	51
	Plant %	0	100	100

*p35S:LFY* and *p35S:LFY<sub>TERE</sub>* T1 plant effective for each phenotypic class. Plants were classified in the "ectopic flowers" class when we observed at least one rosette flower was observed and in the "no ectopic flower" class if none were seen.

**Table S4. Molecular mass determination of GbLFY $\Delta$ , WT and mutants**

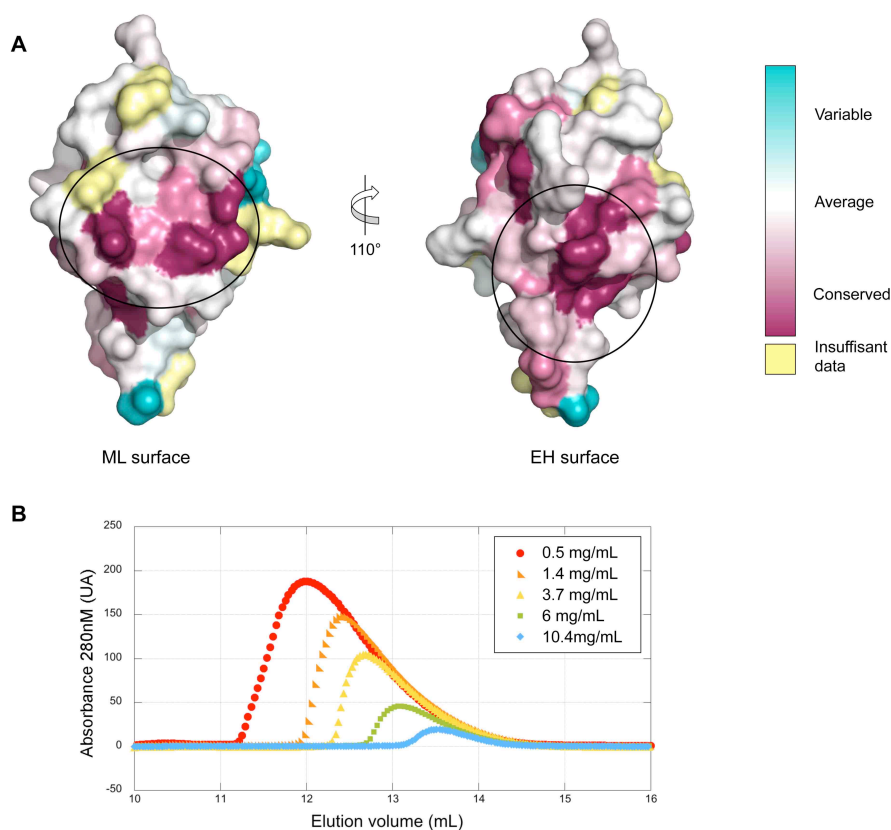
Protein – DNA complex	Molecular mass (kDa)	Monomers number for a DNA molecule
GbLFY $\Delta$ - AP1	98 (+/- 6)	1.8
GbLFY $\Delta_{TE}$ - AP1	97.7 (+/- 4)	1.8
GbLFY $\Delta_{[TE+RE]}$ - AP1	115.5 (+/- 5)	2.2

SEC-MALLS molecular mass analysis of GbLFY $\Delta$ , GbLFY $\Delta_{TE}$  and GbLFY $\Delta_{[TE+RE]}$  in complex with AP1 DNA. 50  $\mu$ L containing 2.6 mg/mL (60  $\mu$ M) protein and 15  $\mu$ M DNA were used for this experiment. The measured molecular mass for each complex (the number in brackets is the standard error) and the calculated number of molecule per complex are indicated. The theoretical molecular mass of a GbLFY $\Delta$  monomer is 43.5 kDa and that of AP1 DNA is 18.4 kDa. The corresponding elution profiles are shown in **Figure 4C**.

**Figure S1. GbLFY-SAM oligomeric state depends on the protein concentration**

(A) SEC analysis of GbLFY-SAM at different concentrations. 50  $\mu$ L of protein at the indicated concentration were used. These curves were employed to estimate the molecular mass shown in (B).

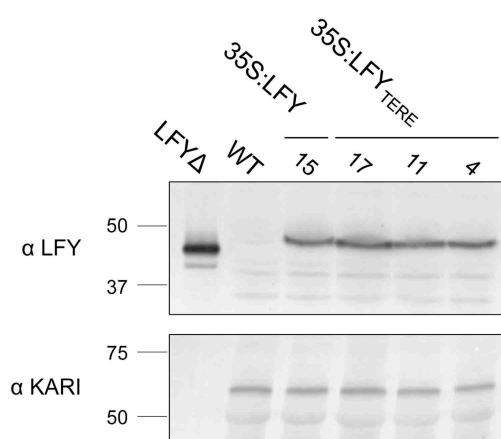
(B) The molecular weight corresponding to each protein concentration was estimated using a calibration curve. The obtained curve was fitted using the equation  $MW = (MW_{max} \times [protein]) / (K_{0.5} + [protein])$ , MW is the molecular weight,  $MW_{max}$  is the MW at the plateau,  $K_{0.5}$  is the concentration required to reach half the  $MW_{max}$  in our experimental conditions. The determined  $K_{0.5}$  was 13.1  $\mu$ M (0.57 mg/mL).



**Figure S2. LFY-SAM oligomerisation is a conserved property**

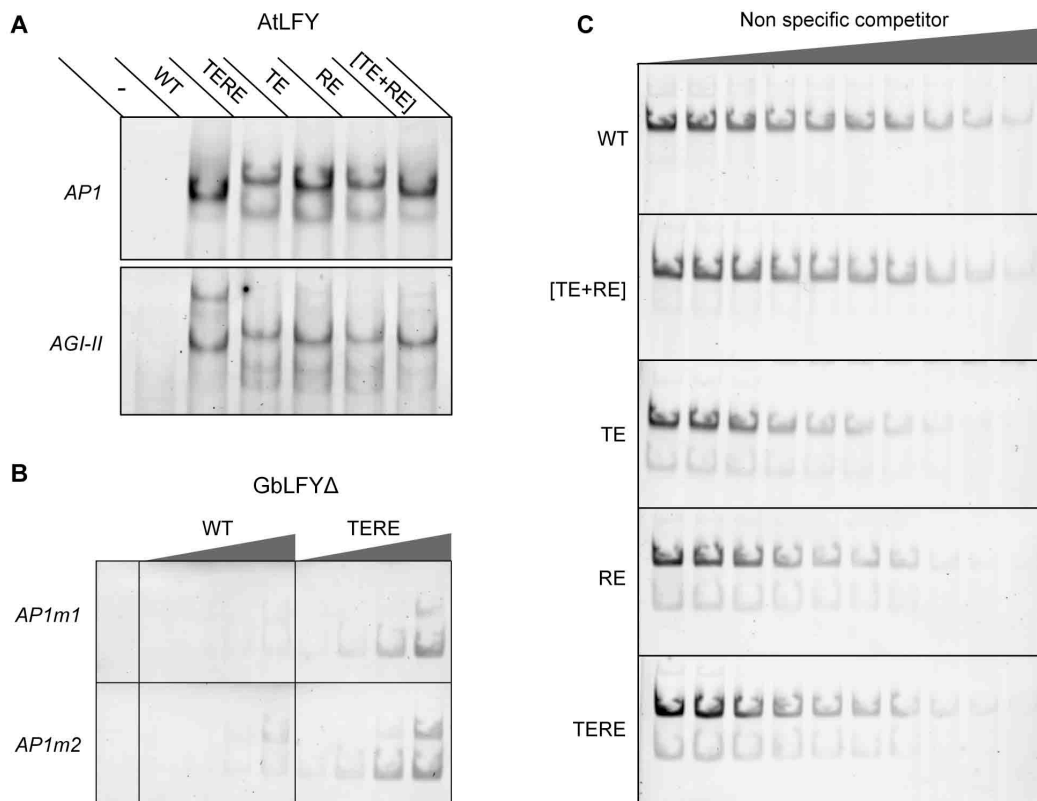
(A) Amino acid conservation of LFY-SAM mapped on the surface of GbLFY-SAM structure. LFY sequences from 18 species from algae to angiosperms (*A. thaliana*, *O. sativa*, *T. aestivum*, *Z. mays*, *P. hybrida*, *A. trichopoda*, *G. biloba*, *P. radiata*, *W. mirabilis*, *M. domestica*, *A. lygodiiifolia*, *B. multifidum*, *C. richardii*, *P. patens*, *N. aenigmaticus*, *Cylindrocystis sp*, *C. scutata*, *K. subtile*) were aligned and the surface conservation was computed using the ConSurf server (<http://consurf.tau.ac.il/>; Ashkenazy et al., 2010).

(B) SEC analysis of AtLFY-SAM at different concentrations. 50  $\mu$ L of protein at the indicated concentration were used.



**Figure S3. LFY and LFY<sub>TERE</sub> proteins are expressed at similar level in overexpressing plants**

Western-blot on *Arabidopsis* leaves extracts from 15 day-old seedlings grown under long day conditions. One leaf from 6 to 8 seedlings were pooled for each transgenic line. 18.75 mg and 12.5 mg of fresh matter were used for anti-LFY and anti-KARI blots respectively. The molecular weight deduced from standard protein migration is indicated on the left. 5ng of recombinant LFY protein (LFY $\Delta$ , produced in *E. coli*) were used as a positive control.

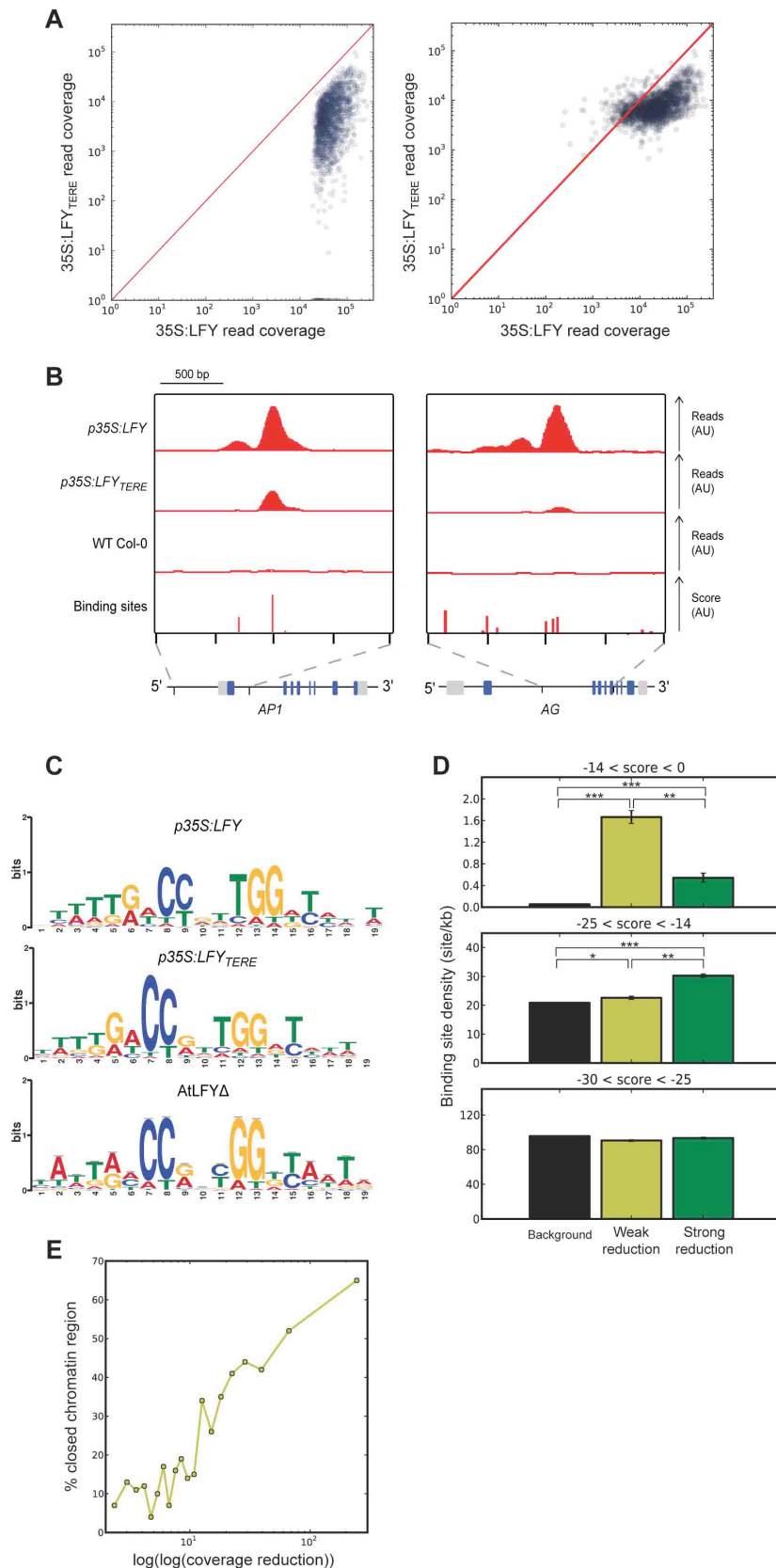


**Figure S4. Oligomerisation through the SAM domain and *in vitro* DNA binding**

(A) EMSA with LFY from *Arabidopsis* with 10 nM AP1 DNA (top) or AGI-II DNA (bottom) and 500 nM AtLFY, AtLFY<sub>[TE+RE]</sub>, AtLFY<sub>TE</sub>, AtLFY<sub>RE</sub> or AtLFY<sub>TERE</sub> protein. Only the protein-DNA complexes are shown.

(B) EMSA with 10 nM AP1m1 or AP1m2 DNA (each carrying mutations on a half-site of the palindromic LFY binding site) and 0, 50, 100, 250, 500 nM GbLFYΔ or GbLFYΔ<sub>TERE</sub> protein. Only the protein-DNA complexes are shown.

(C) Competition assay using EMSA. Binding reactions contained 10 nM AP1 DNA, 500 nM GbLFYΔ, GbLFYΔ<sub>[TE+RE]</sub>, GbLFYΔ<sub>TE</sub>, GbLFYΔ<sub>RE</sub> or GbLFYΔ<sub>TERE</sub> protein and increasing concentration of non-specific unlabelled competitor concentration. Quantification of the % of bound DNA is shown in **Figure 4D**. Only the protein-DNA complexes are shown.



**Figure S5. Oligomerisation through the SAM domain and genome wide DNA binding**

(A) Scatter plot showing the region coverage in *p35SLFY* versus *p35SLFY<sub>TERE</sub>* for the 2000 regions bound with the highest confidence in *p35SLFY* (left panel) and for the 2000 regions bound with the highest confidence in *p35SLFY<sub>TERE</sub>* (right panel). Coverage has been normalised and subtracted for



the background value (coverage in WT Col-0 ChIP-seq). A spot above the red line represents a region with a higher read coverage in *p35SLFY<sub>TERE</sub>* than in *p35SLFY*.

(B) Examples of regions bound by AtLFY identified by ChIP-seq and binding profile of AtLFY<sub>TERE</sub> on those regions. The left panel shows the *AP1* locus (*APETALA1*, AT1G69120), the right panel is the *AG* locus (*AGAMOUS*, AT4G18960). ChIP-seq read coverage is shown for *p35S:LFY*, *p35S:LFY<sub>TERE</sub>* and WT Col-0 (which shows the background signal). The lower panel shows the LFY binding sites (computed with the model from Moyroud et al., 2011. Noncoding and coding sequences in exons are shown on the bottom as grey and blue boxes respectively. AU: Arbitrary Unit.

(C) AtLFY and AtLFY<sub>TERE</sub> DNA binding motifs deduced from *p35S:LFY* and *p35S:LFY<sub>TERE</sub>* ChIP-seq respectively, by aligning the 2000 regions bound with the highest confidence in each ChIP-seq. The SELEX motif obtained using recombinant LFY protein (AtLFY $\Delta$ ; Chahtane et al., 2013) is shown at the bottom. Note that symmetry was imposed on the SELEX obtained motif, and not on the ChIP-seq obtained motifs.

(D) Out of the 2000 regions bound by AtLFY with the highest confidence, we selected the 200 regions with the lowest coverage reduction in the *p35SLFY<sub>TERE</sub>* ChIP-Seq and 200 with the highest one. We also selected a negative set of 2000 regions not bound by LFY and showing the same size distribution and the same GC content as the 2000 bound regions. In these 3 sets of regions, binding site density was calculated for high affinity binding site (score >-14), medium affinity binding sites (-25<score<-14) and unspecific binding sites (score<-25). The error bar is the standard error of mean (SEM). \*, \*\*, \*\*\* mean a p-value lower than  $5.10^{-3}$ ;  $10^{-12}$  and  $10^{-32}$  respectively (Student t-test).

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

**Supplementary Table 1.** Primers used for mutagenesis

Plasmid	Corresponding protein	Matrix vector for mutagenesis	Primers for mutagenesis	
			Name	Sequence
pCA21	GbLFY $\Delta$ <sub>TE</sub>	pETH164	oCA0019	TGGTTGAGATGGGTTTCACCGTCAATGAGTTGGTTAA TATGACTGAACAAGAG
			oCA0020	CTCTTGTTTCAGTCATATTAACCAACTCATTGACGGTG AAACCCATCTCAACCA
pCA22	GbLFY $\Delta$ <sub>RE</sub>	pETH164	oCA0023	AGAAATATGGCATCAAGTCTGCAGTTGAGGCAGAAAA GAGAAGATTGGATGAG
			oCA0024	CTCATCCAATCTTCTCTTTTCTGCCTCAACTGCAGAC TTGATGCCATATTTCT
pCA39	GbLFY $\Delta$ <sub>TERE</sub>	pCA22	oCA0019	TGGTTGAGATGGGTTTCACCGTCAATGAGTTGGTTAA TATGACTGAACAAGAG
			oCA0020	CTCTTGTTTCAGTCATATTAACCAACTCATTGACGGTG AAACCCATCTCAACCA
pCA23	AtLFY <sub>TE</sub>	pETH94	oCA0029	GTTAGGTTTTACGGCTAGCGAGCTTGTGGGTATGAAG GAC
			oCA0030	GTCCTTCATACCACAAGCTCGCTAGCCGTAACCT AAC
pCA24	AtLFY <sub>RE</sub>	pETH94	oCA0027	TACGGTATCAAAGCTGCCGTTGAGGCTGAACGGAGAC GATTGC
			oCA0028	GCAATCGTCTCCGTTTCAGCCTCAACGGCAGCTTTGAT ACCGTA
pCA25	AtLFY <sub>TERE</sub>	pCA24	oCA0031	GTTAGGTTTTACGGCAGCGAGCTCGTGGGTATGAAG GAC
			oCA0032	GTCCTTCATACCACGAGCTCGCTCGCCGTAACCT AAC
pCA15	GbLFY-SAM <sub>TE</sub>	pETH195	oCA0019	TGGTTGAGATGGGTTTCACCGTCAATGAGTTGGTTAA TATGACTGAACAAGAG
			oCA0020	CTCTTGTTTCAGTCATATTAACCAACTCATTGACGGTG AAACCCATCTCAACCA
pCA17	GbLFY-SAM <sub>RE</sub>	pETH195	oCA0023	AGAAATATGGCATCAAGTCTGCAGTTGAGGCAGAAAA GAGAAGATTGGATGAG
			oCA0024	CTCATCCAATCTTCTCTTTTCTGCCTCAACTGCAGAC TTGATGCCATATTTCT
pCA20	GbLFY-SAM <sub>TERE</sub>	pCA17	oCA0019	TGGTTGAGATGGGTTTCACCGTCAATGAGTTGGTTAA TATGACTGAACAAGAG
			oCA0020	CTCTTGTTTCAGTCATATTAACCAACTCATTGACGGTG AAACCCATCTCAACCA

**Supplementary Table 2.** Buffer composition for protein purification

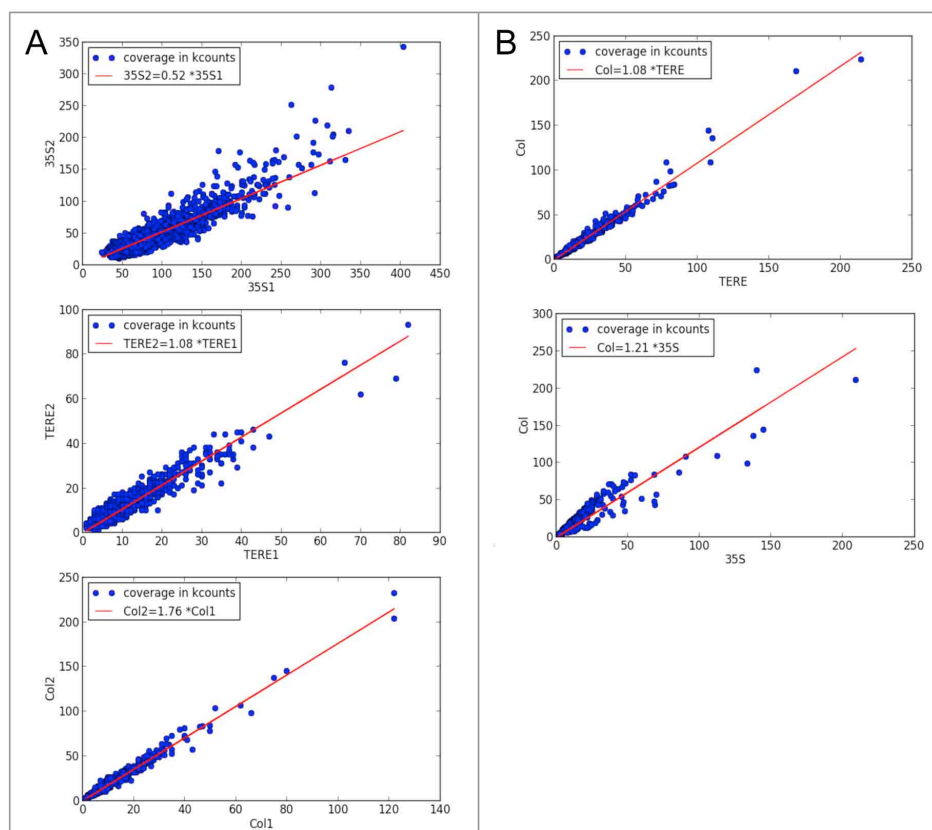
Buffer name	Proteins	Buffer composition
Buffer A	GbLFY-SAM, GbLFY-SAM <sub>TE</sub> , GbLFY-SAM <sub>RE</sub> , GbLFY-SAM <sub>TERE</sub> , AtLFY-SAM	20 mM Tris-HCl pH 8, 1 mM TCEP
Buffer B	GbLFYΔ, GbLFYΔ <sub>TE</sub> , GbLFYΔ <sub>RE</sub> , GbLFYΔ <sub>TERE</sub>	1 M NaCl, 20 mM Tris-HCl pH 8.5, 1 mM TCEP
Buffer C	GbLFYΔ, GbLFYΔ <sub>TE</sub> , GbLFYΔ <sub>RE</sub> , GbLFYΔ <sub>TERE</sub> (SEC only)	0.8 M NaCl, 20 mM Tris-HCl pH 8.5, 1 mM TCEP
Buffer D	AtLFY, AtLFY <sub>TE</sub> , AtLFY <sub>RE</sub> , AtLFY <sub>TERE</sub>	20 mM Tris-HCl pH 8.5, 1 mM TCEP

**Supplementary Table 3.** Oligonucleotides used for EMSA assays

Name	Sequence	Reference
AP1	GTTGGGAAGGACCAGTGGTCCGTACAATGT	LFY bs1 ; Benlloch <i>et al.</i> , 2011
AP1m	GTTGGGAAGGAAAAGTAATCCGTACAATGT	LFY bs1 m3 ; Benlloch <i>et al.</i> , 2011
AP1m1	GTTGGGAAGGAAAAGTGGTCCGTACAATGT	LFY bs1 m1 ; Benlloch <i>et al.</i> , 2011
AP1m2	GTTGGGAAGGACCAGTAATCCGTACAATGT	LFY bs1 m2 ; Benlloch <i>et al.</i> , 2011
AGI-II	CGTTTAAATTTAATCCAATGGTTACAATTTAATACTATCA AATGTCATTGGATTTATACCCAATGTGTTAATGGGTTG	

**Supplementary Table 4.** Oligonucleotides used for EMSA assays

Genotype_replicate	Mapped reads number
Col_1	901815
Col_2	2078019
p35S:LFY <sub>TERE</sub> _1	2370271
p35S:LFY <sub>TERE</sub> _1	2735274
p35S:LFY_1	7164486
p35S:LFY_2	4385809



**Supplementary Figure 1.** ChIP-seq normalisation. (A) Intra-genotype normalisation: peaks coverage and linear fit within genotypes. (B) Inter-genotype normalisation.

### III. Résultats et discussions complémentaires

Les principaux résultats concernant ce chapitre ont été décrits et discutés dans l'**Article 1**. Dans les paragraphes qui suivent, certains points seront précisés, et quelques résultats complémentaires seront annoncés et discutés.

#### 1. Production de protéines recombinantes LFY, domaine N-terminal et protéines entières

##### ▪ Un criblage de constructions pour obtenir les protéines

Notre objectif à long terme est d'étudier les propriétés biochimiques et la structure tridimensionnelle (3D) de LFY. La protéine étant constituée de deux domaines globulaires reliés par un *linker* probablement flexible, obtenir la structure cristallographique de la protéine entière semblait assez difficile. Connaissant déjà la structure du domaine de liaison à l'ADN, nous nous sommes donc plutôt focalisés sur la structure du domaine N-terminal, tout en faisant des tentatives avec la protéine entière. Concernant l'analyse des propriétés biochimiques de LFY, nous avons travaillé à la fois sur le domaine N-terminal isolé et sur la protéine entière.

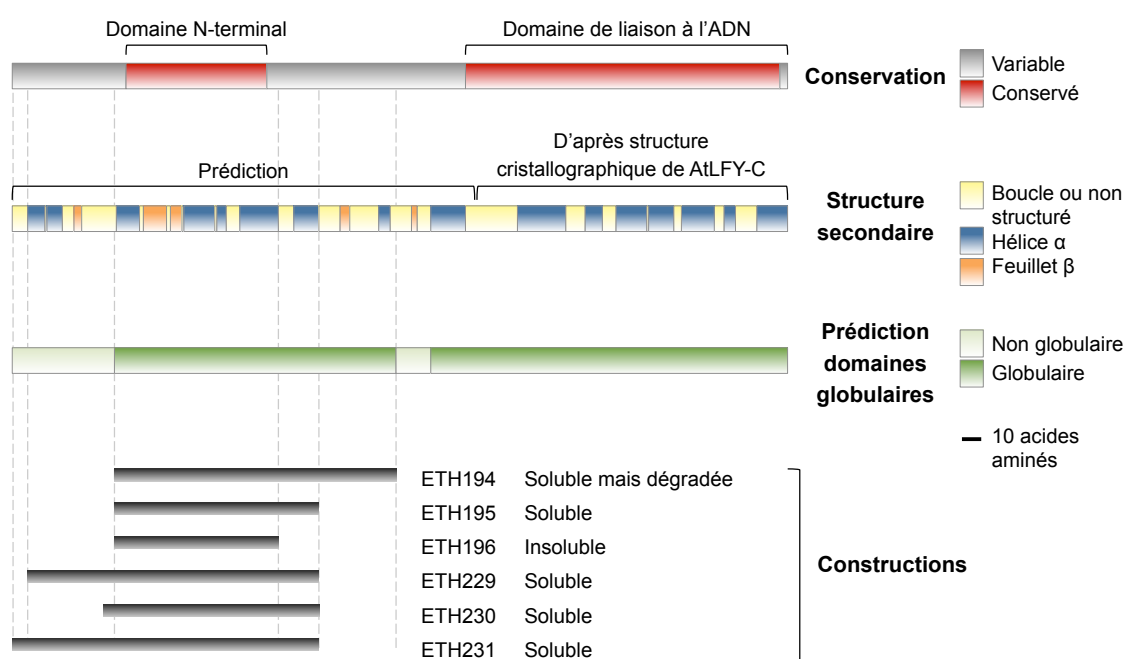
La première étape a donc été l'obtention de protéines recombinantes pures en quantité suffisante, à la fois pour le domaine N-terminal et pour la protéine entière.

La production de la protéine entière, ainsi que celle du domaine N-terminal de la protéine d'*Arabidopsis* avaient été entreprises sans succès dans l'équipe : ces protéines étaient insolubles. Pour tenter de remédier à ce problème et augmenter nos chances de réussite, nous avons réalisé plusieurs constructions de tailles différentes, avec des homologues de LFY issues de plusieurs espèces.

Nous avons travaillé sur trois espèces : *A. thaliana* (AtLFY) car c'est l'espèce modèle (utilisée pour les analyses fonctionnelles, génome séquencé et annoté), *G. biloba* (GbLFY) et *P. patens* (PpLFY1, noté PpLFY par la suite) car des analyses préliminaires suggéraient que ces protéines étaient plus stables que celle d'*Arabidopsis*. En effet, GbLFY a un indice de stabilité relativement bon d'après le serveur ProtParam (<http://web.expasy.org/protparam/>). Cet indice est calculé d'après la séquence primaire de la protéine et est basé sur la corrélation entre la présence de certaines combinaisons d'acides aminés et la stabilité

des protéines. De plus, la spécificité de liaison à l'ADN de GbLFY semblait très proche de celle d'AtLFY, ce qui a été vérifié par la suite (**Article 2**). Chez PpLFY le *linker* entre les deux domaines conservés est relativement court, ce qui pourrait augmenter la stabilité de la protéine (**Figure 8**). Par ailleurs, sa spécificité semble différente de celle de AtLFY (Maizel *et al.*, 2005), ce qui en fait un bon modèle pour l'étude de cette propriété (**Article 2**).

Nous avons dessiné plusieurs constructions pour chaque espèce, en se basant sur la conservation des domaines (alignement de la séquence primaire), sur des prédictions de structures secondaires (serveur GORIV ; [http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_gor4.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html)) et sur des prédictions des domaines globulaires (serveur Globplot ; <http://globplot.embl.de/>).



**Figure 1-1 | Stratégie pour obtenir des constructions donnant des formes solubles du domaine N-terminal de LFY.**

Schéma des constructions dessinées en fonction de la conservation et de la structure secondaire, pour produire le domaine N-terminal de GbLFY. La protéine entière GbLFY fait 402 acides aminés. La conservation est déduite de l'alignement des séquences protéiques de LFY. La structure secondaire de la partie C-terminale est déduite de la structure cristallographique du domaine de liaison à l'ADN de la protéine d'*Arabidopsis* (AtLFY-C), celle de la partie N-terminale a été prédite avec le logiciel GORIV. Les domaines globulaires ont été prédits grâce au logiciel Globplot. Les différentes constructions réalisées, et leur solubilité sont indiquées en bas.

Les prédictions et les différentes constructions dessinées pour le domaine N-terminal de GbLFY (GbLFY-N) sont schématisées en **Figure 1-1**. La même stratégie a été adoptée pour AtLFY-N et PpLFY-N. Entre 4 et 6 constructions ont été réalisées pour chaque espèce.

Un *tag* histidine (6xHis) a été ajouté aux protéines pour permettre leur purification sur colonne d'affinité au nickel. Ce *tag* est placé soit en position N-terminale, précédé d'un site de clivage pour la protéase TEV (Tobacco Etch Virus), soit en position C-terminale.

Des tests d'expression chez *E. coli* et des tests de solubilisation ont montré que la plupart des constructions étaient bien exprimées, certaines solubles, d'autres pas (**Figure 1-1**).

Pour les protéines entières AtLFY et GbLFY, la délétion de la partie non conservée en N-terminal (39 et 53 acides aminés chez AtLFY et GbLFY respectivement) a beaucoup amélioré leur production. Les protéines tronquées sont notées AtLFY $\Delta$  et GbLFY $\Delta$ . Pour PpLFY par contre, nous avons utilisé la protéine pleine longueur car elle est soluble et stable.

Un tableau en **Annexe 1** regroupe les constructions utilisées dans cette étude.

#### ▪ Production du domaine N-terminal de LFY

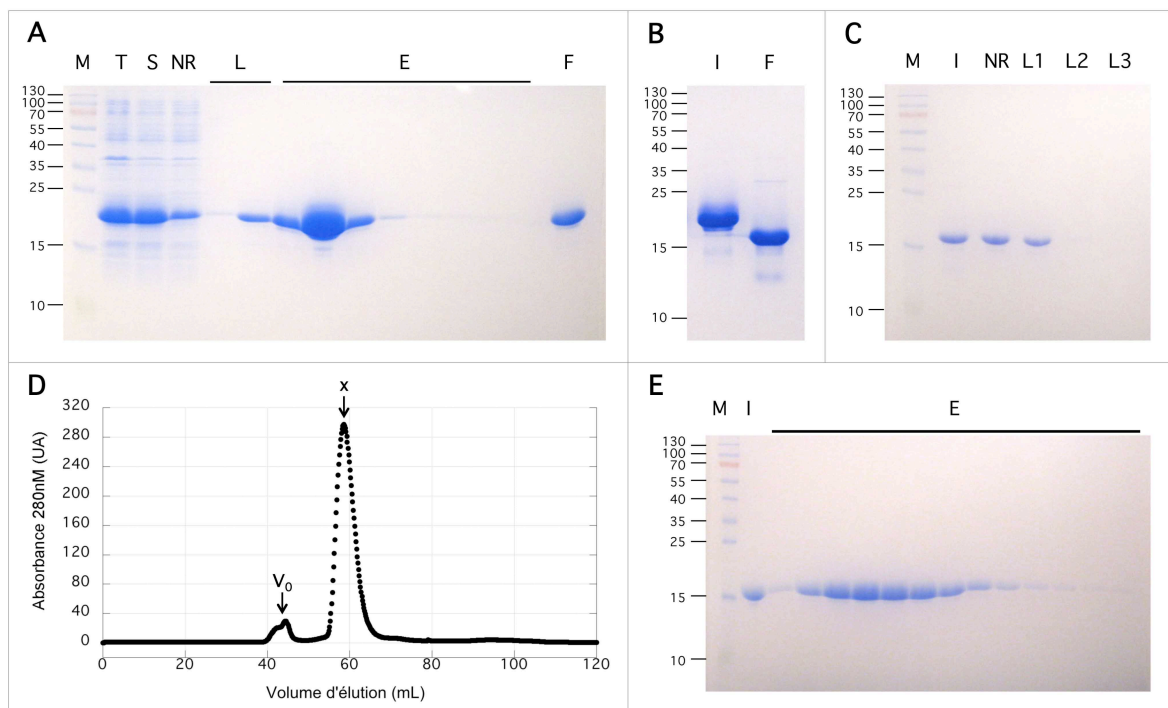
Dans le but d'obtenir la structure cristallographique du domaine N-terminal de LFY, nous avons choisi une construction par espèce pour produire et purifier ce domaine. D'après des résultats préliminaires d'expression et de solubilisation, les constructions nommées ETH201 (AtLFY-N), ETH195 (GbLFY-N) et ETH193 (PpLFY-N) (**Annexe 1**) ont été retenues car elles étaient fortement exprimées et les protéines étaient entièrement solubles.

Nous avons mis au point un protocole d'expression et purification permettant d'obtenir des protéines pures à homogénéité et en quantité relativement élevée. Ces protéines ont été exprimées dans une souche commerciale de bactéries *E. coli* optimisée pour la production de protéines recombinantes, puis solubilisées et purifiées à 4°C dans un tampon contenant du Tris-HCl à pH 8 et un agent réducteur. La purification est réalisée en plusieurs étapes (**Figure 1-2**) :

- Première chromatographie d'affinité (**Figure 1-2A**) : cette première étape a pour but d'enlever le plus gros des contaminants présents dans l'extrait bactérien. La présence d'un *tag* histidine sur la protéine permet sa fixation sur une colonne contenant du nickel (résine de nickel-Sepharose), alors que la plupart des contaminants ne sont pas retenus.
- Clivage du *tag* histidine (**Figure 1-2B**) : même si ce *tag* est relativement petit (3 kDa) et n'influence pas, en principe, le repliement de la protéine, il est préférable de le retirer pour les analyses structurales. Nous produisons au laboratoire la protéase TEV, qui permet de cliver ce *tag* lorsqu'elle est incubée avec la protéine.
- Seconde chromatographie d'affinité (**Figure 1-2C**) : afin de se débarrasser à la fois de la protéine non clivée restante, de la protéase TEV (portant également un *tag*

histidine), et des contaminants ayant été retenus sur la première colonne d'affinité, une seconde chromatographie sur colonne de nickel-Sepharose est réalisée. Cette fois-ci, la protéine n'est pas, ou très peu, retenue par la colonne, alors que les contaminants le sont.

- Chromatographie d'exclusion stérique (SEC pour *Size Exclusion Chromatography*) (**Figure 1-2D et E**) : cette dernière étape sert à éliminer les éventuels agrégats et contaminants restants. La protéine est injectée sur une colonne contenant un polymère (Superdex) qui retient plus ou moins les macromolécules en fonction de leur volume hydrodynamique (ce paramètre dépend de la taille et de la forme de la molécule). Les agrégats sont élués plus tôt que la protéine non agrégée, dans le « volume mort » de la colonne. Cette étape sert également de contrôle-qualité pour vérifier que la protéine est au moins en partie sous forme non agrégée.



**Figure 1-2 | Purification du domaine N-terminal de LFY, exemple de PpLFY-N (construction ETH193).**

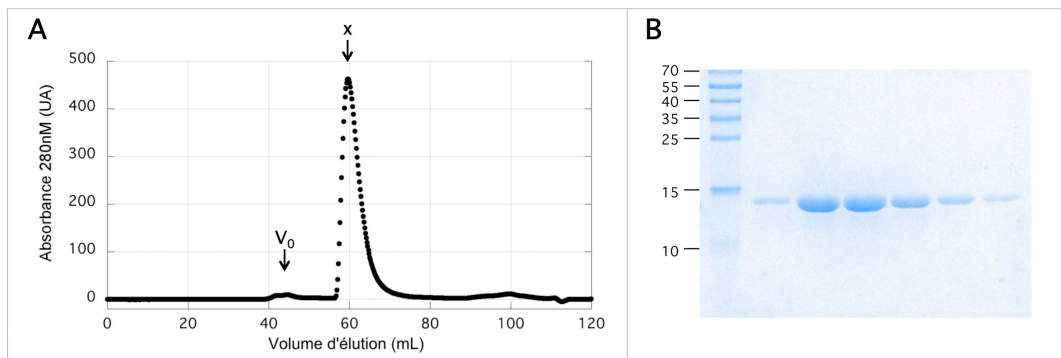
**A, B, C, E** montrent l'analyse des fractions qui correspondent à chaque étape de purification sur gel dénaturant (SDS-PAGE). M : marqueur de poids moléculaire (les tailles en kDa sont indiquées sur la gauche), T : extrait bactérien total, S : extrait soluble; NR : extrait non retenu, L : lavage, E : élution. F : extrait final.

**A.** Première purification sur colonne d'affinité (nickel-Sepharose). Les fractions d'élution contenant la protéine sont groupées pour donner l'extrait final (F), qui est utilisé pour les étapes suivantes. La masse moléculaire attendue est de 20 kDa.

**B.** Clivage du *tag* histidine par la protéase TEV. La masse moléculaire attendue pour la protéine clivée est de 17 kDa.

**C.** Seconde purification sur colonne d'affinité (nickel-Sepharose) pour se débarrasser du *tag* clivé et de la protéase.

**D et E.** SEC (chromatographie d'exclusion stérique). Le profil d'élution sur colonne S200 16/60 est montré en D,  $V_0$  : volume mort, x : pic contenant la protéine. L'analyse des fractions correspondant au pic x est montrée en figure E. A titre indicatif, jusqu'à 20  $\mu\text{g}$  de protéines ont été déposés sur ce gel.



**Figure 1-3 | Dernière étape de purification par chromatographie d'exclusion stérique de GbLFY-N (ETH195).**

**A.** Profil d'élution sur colonne S200 16/60.  $V_0$  : volume mort, x : pic contenant la protéine.

**B.** Analyse sur gel SDS-PAGE des fractions correspondant au pic x. Un marqueur de masse moléculaire, dont les tailles sont indiquées sur la gauche, occupe la première piste du gel. La masse moléculaire attendue pour GbLFY-N est de 12,8 kDa.

Toutes les étapes de purification sont montrées sur la **Figure 1-2** pour PpLFY-N, et la dernière étape (SEC) de purification de GbLFY-N est montrée sur la **Figure 1-3**. AtLFY-N a été purifiée de la même manière.

En partant d'une culture bactérienne initiale d'1 L, le rendement typiquement obtenu pour cette purification est de 90 mg de protéine après la première colonne d'affinité, 65 mg après le clivage du *tag*, 55 mg après la seconde colonne d'affinité, et 30 mg après la SEC. On concentre la protéine à environ 5 mg/mL et on la conserve sous forme d'aliquotes à  $-80^{\circ}\text{C}$ .

Suite à la première étape, on obtient déjà une protéine relativement pure (**Figure 1-2A**). La dernière étape de SEC nous donne un premier indice sur l'état d'oligomérisation de la protéine : le pic correspondant à la protéine (noté x sur les **Figures 1-2D** et **1-3A**) est asymétrique, et le volume d'élution (autour de 60 mL) est beaucoup plus élevé que celui attendu pour un monomère (en se basant sur une calibration faite avec des protéines de masses moléculaires connues). Nous pouvons donc, à ce stade, anticiper que le domaine N-terminal de LFY formait un oligomère.

#### ▪ Production de la protéine LFY pleine longueur

Parallèlement au domaine N-terminal, nous avons également produit des protéines LFY pleine longueur, ou « quasi pleine longueur » (LFY $\Delta$ ). Leur purification a été plus délicate

que celle du domaine N-terminal car elles ont tendance à s'agréger. Selon les protéines et l'utilisation envisagée, nous avons mis au point plusieurs protocoles de purification (**Tableau 1-1**).

**Tableau 1-1 | Différents protocoles de purification pour les protéines LFY entières.**

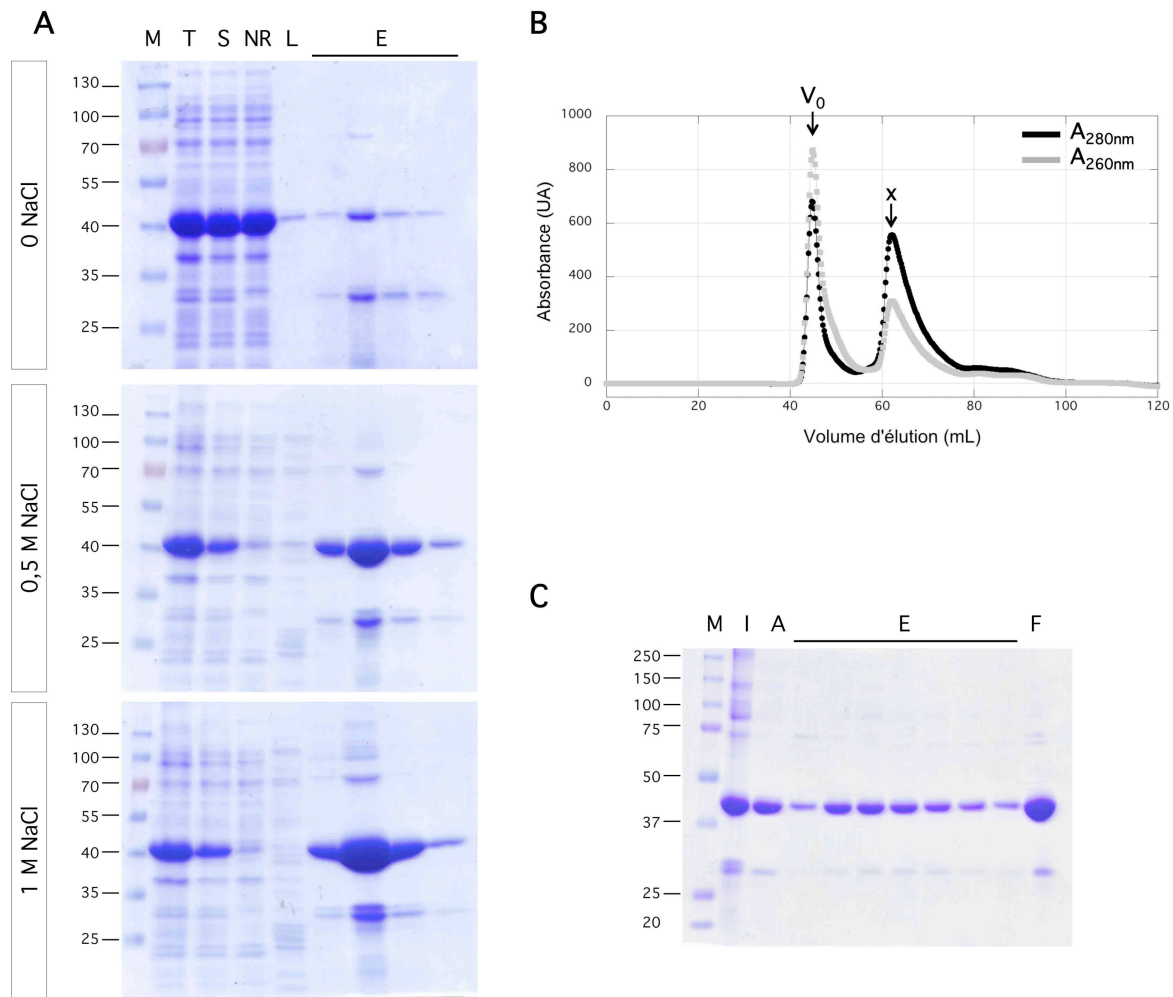
Type de purification	AtLFY	GbLFYΔ	PpLFY	Avantages	Inconvénients	Utilisations
1 étape	✓	✓	✓	Possible pour toutes les protéines, rapide	Pureté limitée, contamination par de l'ADN bactérien, faible rendement	EMSA
2 étapes	✗	✓	✓	Haute pureté	Pas possible pour AtLFY	EMSA, SEC, SEC-MALLS, méthodes biophysiques
4-5 étapes	✗	✓	✓	Haute pureté, tag histidine retiré	Pas possible pour AtLFY, relativement long	EMSA, SEC, SEC-MALLS, cristallographie et autres méthodes biophysiques

Ces protocoles sont détaillés dans le texte ci-dessous. EMSA : Electrophoretic Mobility Shift Assay (méthode d'analyse des interactions ADN-protéine), SEC-MALLS : Size Exclusion Chromatography – Multi Angle Laser Light Scattering (méthode d'analyse de masse moléculaire). Les méthodes biophysiques regroupent la microscopie électronique et le SAXS (Small Angle X-ray Scattering, méthode d'analyse de la structure 3D à basse résolution des protéines).

- Purification par chromatographie d'affinité (une étape, sans NaCl).

Nous nous sommes aperçus que LFY précipitait à faible concentration en sel, ce qui explique en grande partie les échecs des précédentes tentatives de purification, puisque les tampons typiquement utilisés pour manipuler les protéines en contiennent entre 100 et 500 mM. Il est possible de purifier les protéines des trois espèces, AtLFY, GbLFYΔ et PpLFY, dans un tampon sans NaCl sur colonne d'affinité au nickel. Cependant, LFY étant un facteur de transcription, il a de l'affinité pour l'ADN et entraîne avec lui de l'ADN bactérien (les échantillons absorbent fortement à 260 nm, le ratio  $A_{260}/A_{280}$  est souvent supérieur à 1). La présence d'ADN complique le dosage précis des protéines : la mesure d'absorbance à 280 nm est à exclure car l'absorbance de l'ADN interfère. On dose alors par la méthode de Bradford, peu précise, ou sur gel SDS-PAGE contre une gamme de protéine de concentration connue, méthode relativement longue à mettre en œuvre. De plus, dans ces conditions, la protéine n'est que peu retenue par la résine de nickel et le rendement de purification n'est donc pas bon (**Figure 1-4A**).





**Figure 1-4 | Purification en deux étapes de la protéine LFY du *Ginkgo* quasi-entière (GbLFY $\Delta$ , construction ETH164).**

**A et C** montrent l'analyse des fractions qui correspondent à chaque étape de purification sur gel dénaturant (SDS-PAGE).

**A.** Exemple d'optimisation des conditions de purification sur colonne d'affinité (nickel-Sepharose). Un extrait brut a été repris dans un tampon en faisant varier la concentration en NaCl (0 ; 0,5 ; 1 M), ce qui influence la solubilité, l'affinité pour la colonne, et donc la quantité de protéine éluée. M : marqueur de poids moléculaire (tailles en kDa indiquées sur la gauche du gel), T : extrait bactérien total, S : extrait soluble, NR : extrait non retenu, L : lavage, E : élution. La masse moléculaire attendue est de 43,8 kDa. Les fractions d'élution obtenues en 1 M NaCl ont été utilisées pour la purification en SEC.

**B.** Profil d'élution sur colonne SEC S200 16/60 dans un tampon à 0,8 M NaCl.  $V_0$  : volume mort, x : pic contenant la protéine. Les courbes noire et grise correspondent à l'absorbance à 280 et 260 nm respectivement.

**C.** Analyse des fractions correspondant aux pics  $V_0$  et x de la figure B. M : marqueur de masse moléculaire, I : extrait initial avant passage sur la colonne, A : fraction correspondant au pic  $V_0$ , E : fractions correspondant au pic x, F : extrait final après groupement des fractions E.

- Purification par chromatographie d'affinité et SEC (deux étapes, forte concentration en NaCl).

La liaison entre LFY et l'ADN est principalement causée par des interactions entre les charges positives des chaînes latérales des lysines et arginines et les charges négatives de l'ADN. Cette liaison est déstabilisée en présence de sel. Mais LFY étant très sensible au sel, il a fallu trouver des conditions permettant de séparer l'ADN sans que la protéine ne précipite. Cela est possible en travaillant à forte concentration en sel (0,8 à 1 M NaCl) et s'explique par le fait que la liaison ADN-protéine est rompue et les charges apportées par le NaCl compensent celles de l'ADN, stabilisant ainsi la protéine. Dans ces conditions, même si toute la protéine n'est pas soluble, elle est très bien retenue par la résine de nickel, le rendement de purification est donc bien meilleur (**Figure 1-4A**). Cependant, ces conditions sont applicables à GbLFY $\Delta$  et PpLFY, mais pas à AtFLY car on observe de la protéolyse pour cette protéine dans ces conditions. Suite à cette première étape par colonne d'affinité, la protéine est en grande partie débarrassée de l'ADN bactérien (le ratio  $A_{260}/A_{280}$  est de 0,5 à 0,6), il est alors possible de la purifier par SEC (**Figure 1-4B et C**). Même si une partie de la protéine sort dans le volume mort de la colonne, une grande partie de la protéine est non-agrégée et peut être récupérée et concentrée.

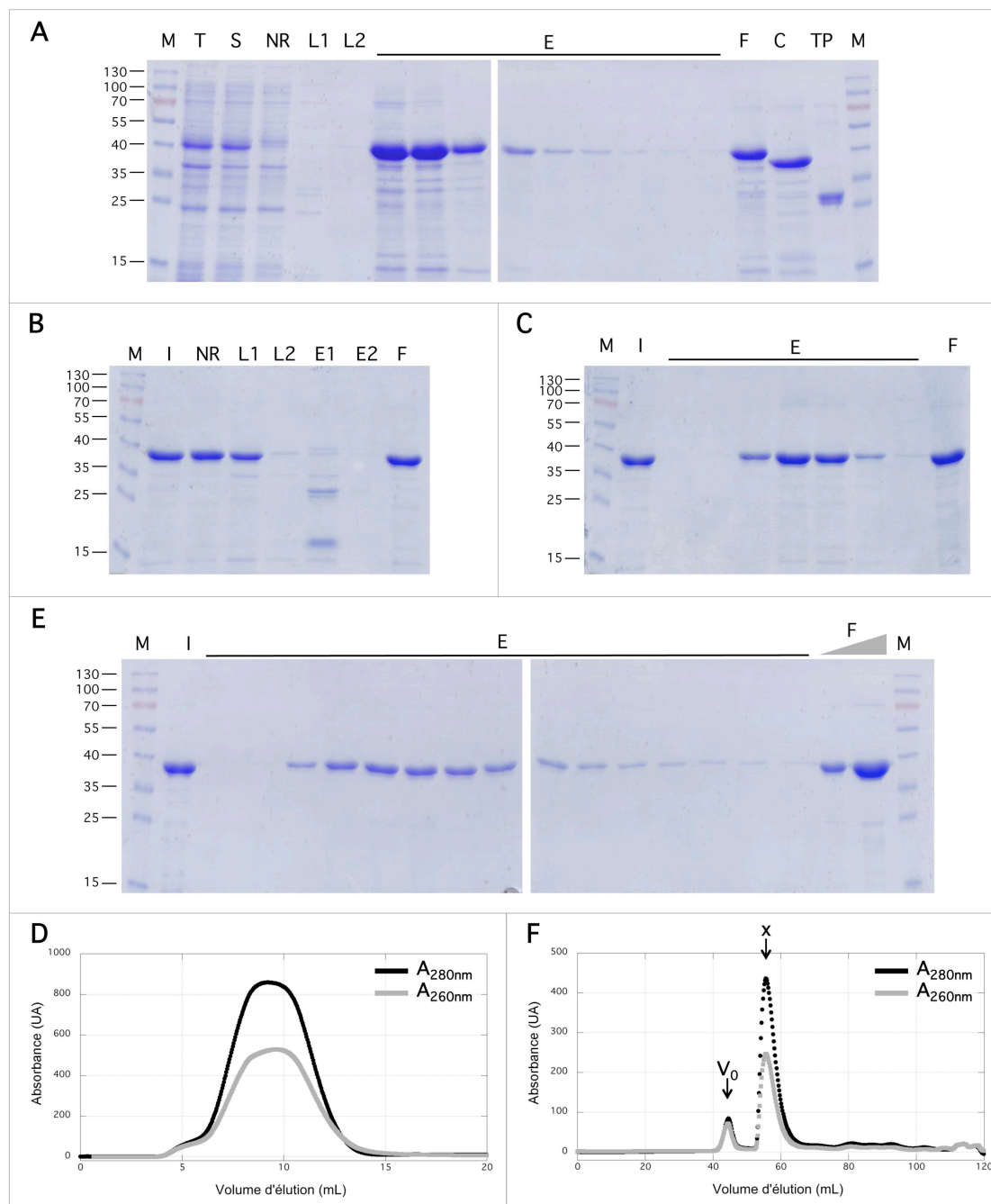
A des concentrations intermédiaires en NaCl, de l'ordre de 0,5 M, la protéine GbLFY $\Delta$  est très instable. Même s'il est possible d'en obtenir des quantités satisfaisantes après passage sur résine de nickel (**Figure 1-4A**), la protéine précipite lors de l'élution.

- Purification pour la cristallographie (quatre à cinq étapes, forte concentration en NaCl).

Une purification plus poussée est préférable pour la cristallographie, afin d'augmenter le degré de pureté, cliver le *tag* histidine et retirer complètement l'ADN bactérien (**Figure 1-5**). Pour se débarrasser totalement de cet ADN contaminant, nous avons mis au point une étape de chromatographie sur colonne échangeuse d'anions Q-Sepharose. L'ADN étant chargé négativement, il a plus d'affinité que la protéine pour ce type de résine. Pour que seul l'ADN soit retenu par la colonne, il est nécessaire d'utiliser un tampon contenant de l'ordre de 0,5 M de NaCl. Or nous avons vu que cette concentration n'était supportée ni par AtLFY, ni par GbLFY $\Delta$ . Par contre, PpLFY peut tolérer de telles conditions et a pu être purifiée de cette manière. La purification est essentiellement la même que pour le domaine N-terminal, mais le tampon utilisé contient 0,5 M de NaCl, et la protéine est injectée sur une colonne échangeuse d'anions avant l'étape finale de SEC (Etape 1 : première chromatographie d'affinité sur colonne de nickel. Etape 2 : clivage du *tag*. Etape 3 : seconde chromatographie d'affinité. Etape 4 : colonne échangeuse d'ions. Etape 5 : SEC) (**Figure 1-5**). Ce protocole

de purification permet d'atteindre un bon degré de pureté, et de réduire le ratio  $A_{260}/A_{280}$  à environ 0,5, signe de l'absence d'ADN.

Pour GbLFY $\Delta$ , nous avons utilisé le même procédé avec un tampon contenant 0,8 à 1 M de NaCl et sans faire la chromatographie échangeuse d'ions (Etape 1 : première chromatographie d'affinité. Etape 2 : clivage du tag. Etape 3 : seconde chromatographie d'affinité. Etape 4 : SEC).



**Figure 1-5 | Purification en cinq étapes de la protéine LFY de *P. patens* entière (PpLFY, construction ETH182).**

**A, B, C, E** montrent l'analyse des fractions qui correspondent à chaque étape de purification sur gel dénaturant (SDS-PAGE). M : marqueur de poids moléculaire (les tailles en kDa sont indiquées sur la

gauche de chaque gel), T : extrait bactérien total, S : extrait soluble; NR : extrait non retenu, L : lavage, E : élution, F : extrait final, C : extrait clivé, TP : TEV protéase purifiée.

**A.** Etapes 1 et 2 : première purification sur colonne d'affinité (nickel-Sepharose) et clivage du *tag* histidine. Les fractions d'élution contenant la protéine sont groupées pour donner l'extrait final F. Après clivage du *tag* histidine par la protéase TEV on obtient l'extrait clivé C. Les masses moléculaires attendues sont de 43,3 kDa avant clivage et 40,3 kDa après clivage.

**B.** Etape 3 : seconde purification sur colonne d'affinité (nickel-Sepharose) pour se débarrasser du *tag* clivé et de la protéase. Les extraits NR et L1 sont groupés pour donner l'extrait F.

**C et D.** Etape 4 : Chromatographie échangeuse d'anion, sur résine Q-Sepharose, pour se débarrasser de l'ADN bactérien. Le profil d'élution est montré en figure D et l'analyse des fractions correspondantes au pic principal est montrée en figure C.

**E et F.** Etape 5 : SEC. Le profil d'élution sur colonne S200 16/60 est montré en figure F,  $V_0$  : volume mort, x : pic contenant la protéine. L'analyse des fractions correspondantes au pic x est montrée en figure E.

Selon les constructions, la protéine entière est plus ou moins soluble et résistante aux conditions de purification, en particulier à la présence de NaCl. AtLFY étant la plus fragile, nous avons réussi à la purifier seulement partiellement, et en quantité plutôt faible. GbLFY $\Delta$  et PpLFY tolérant le NaCl à forte concentration, nous avons pu les purifier de manière plus poussée. Et enfin, PpLFY étant la protéine la plus soluble et la plus résistante aux différentes conditions de sel, il est possible de la purifier de manière très poussée.

Nous avons donc travaillé avec GbLFY $\Delta$  pour toutes les analyses biochimiques. Des analyses ont aussi été réalisées en utilisant AtLFY et PpLFY, notamment pour vérifier que les propriétés déduites de l'analyse de GbLFY $\Delta$  étaient généralisables aux autres espèces. Même si la protéine pleine longueur AtLFY est moins maniable que AtLFY $\Delta$ , nous sommes parvenus à en obtenir un peu, permettant ainsi de s'affranchir du biais éventuel causé par la délétion des quelques résidus en N-terminal. Pour les analyses structurales, en particulier la cristallographie, nous avons utilisé PpLFY et GbLFY $\Delta$ .

À partir de nombreuses constructions dessinées avec l'aide de différentes analyses bioinformatiques, nous avons donc réussi à produire et à purifier des protéines solubles pour le domaine N-terminal et la protéine entière.

## 2. Analyse structurale de LFY

### ▪ Cristallographie du domaine N-terminal

La cristallographie aux rayons X est un moyen d'obtenir la structure à très haute résolution des protéines. Les coordonnées dans l'espace de chacun des atomes composant une

protéine peuvent être déduites de la manière dont cette molécule diffracte les rayons X. Il est nécessaire que les protéines soient sous forme de cristaux, afin qu'elles soient ordonnées et que le signal de diffraction soit suffisamment fort.

La cristallisation consiste à faire passer la protéine d'un état soluble à un état solide et ordonné. Nous avons utilisé les méthodes de goutte assise et de goutte suspendue qui consistent à mélanger la protéine à des agents précipitants (polymères, solvants organiques, sels) puis à déshydrater progressivement ce mélange par évaporation afin d'augmenter la concentration en protéine et en précipitant.

Comme on ne peut pas prédire les conditions dans lesquelles une protéine va cristalliser, on teste un maximum de conditions en espérant en trouver au moins une qui permette d'obtenir des cristaux de bonne qualité. Il existe cependant des moyens d'orienter la recherche des bonnes conditions : des kits commerciaux sont disponibles, avec des conditions optimisées pour certains types de protéines, telles que les protéines solubles, les protéines membranaires ou les protéines qui lient l'ADN. Pour le domaine N-terminal de LFY, nous avons choisi des kits plutôt adaptés pour les protéines à solubilité réduite. Pour pouvoir cribler beaucoup de conditions en un temps raisonnable et en consommant un minimum de protéine, nous avons utilisé le robot de cristallisation disponible à l'EMBL à Grenoble, via la plateforme du High Throughput Crystallisation Laboratory (HTX Lab).

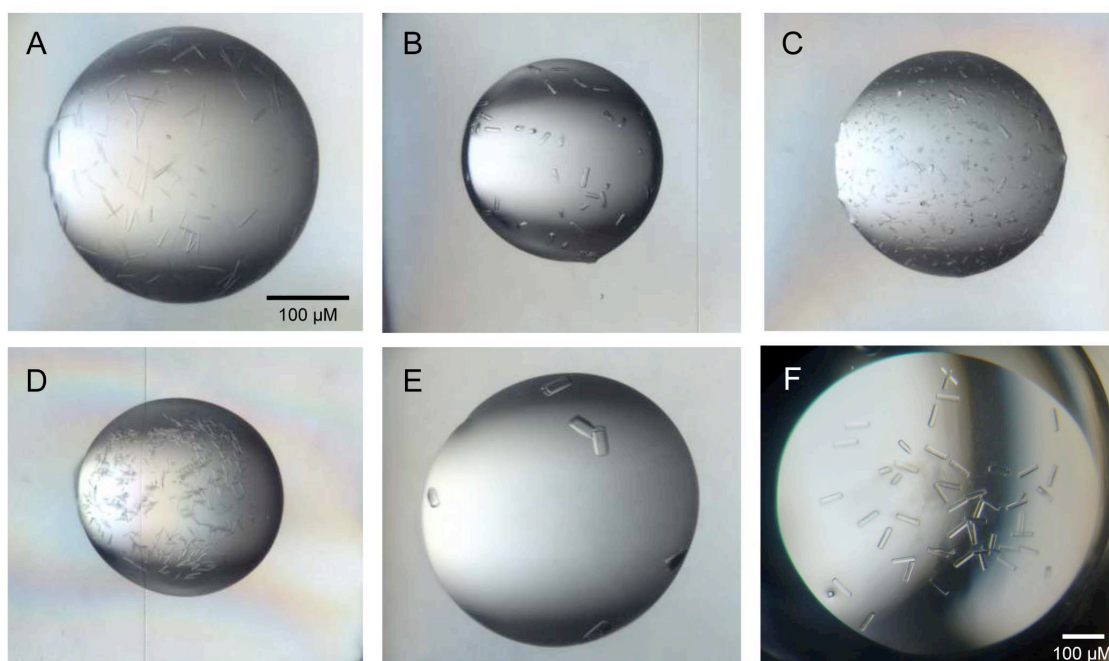
Avant d'envoyer la protéine au robot de cristallisation, nous avons fait quelques tests préliminaires au laboratoire sur GbLFY-N et nous avons remarqué que la présence de 2-methyl-2,4-pentanediol (MPD) dans les gouttes avait tendance à prévenir la précipitation de la protéine. Nous avons alors testé 5 kits (96 conditions chacun) sur GbLFY-N, PpLFY-N et AtLFY-N supplémentées ou non en MPD, soit un total de 2880 conditions (960 conditions par protéine).

En quelques jours nous avons obtenu des cristaux de GbLFY-N pour une quarantaine de conditions (**Figure 1-6**) listées en **Annexe 2**. Nous avons également obtenu plusieurs conditions positives pour AtLFY-N, et quelques-unes pour PpLFY-N.

Les cristaux issus du robot sont très petits (les gouttes font 200 nL), et pas toujours de très bonne qualité (poly-cristaux, micro-cristaux), il est alors nécessaire de reproduire et d'optimiser les conditions manuellement en faisant varier la composition (précipitant, sel, tampon, additifs), la température, la concentration en protéine ou le volume de la goutte. Nous avons donc fait de l'optimisation sur une douzaine de conditions qui nous paraissaient donner les cristaux les plus prometteurs (les cristaux monocristallins, les plus gros) (**Annexe 2**). Nous avons pêché et congelé les cristaux suffisamment gros dans une solution cryo-protectrice et, avec l'aide de Max Nanao, nous avons testé leur diffraction sur la ligne microfocuse ID23-2 au synchrotron de Grenoble (ESRF). Nous avons finalement obtenu des

profils de diffraction exploitables pour l'une des conditions (25 mM Tris-HCL pH 8,8 ; 40 mM sulfate d'ammonium, supplémenté par 20% de glycérol avant congélation).

Pour résoudre le problème de phase, nous avons d'abord incubé les cristaux avec différents métaux lourds, mais sans succès. Nous avons ensuite fait des cristaux à partir de protéine GbLFY-N contenant de la sélénométhionine, ce qui a permis à Max Nanao de pouvoir résoudre la structure à une résolution de 2,3 Å (**Article 1**).



**Figure 1-6 | Exemples de cristaux de GbLFY-N.**

**A-E.** Photos de gouttes contenant des cristaux, prises par le robot du HTX Lab à 20°C. Les gouttes ont un volume de 200 nL, l'échelle est la même pour les 5 images. Le nom du kit, le numéro et la composition de la solution de cristallisation et le type de cristaux sont indiqués ci-dessous.

**A.** MDP Suite n°52 (Qiagen). 0,1 M HEPES pH 7 ; 10% MPD. Aiguilles.

**B.** Crystal Screen light n°39 (Hampton). 0,1 M HEPES de sodium pH 7,5 ; 2% PEG 400 ; 1 M sulfate d'ammonium. Cristaux isolés.

**C.** The Nucleix Suite n°81 (Qiagen). 1 mM spermine ; 15 mM chlorure de magnésium ; 0,05 M HEPES de sodium pH 7,5 ; 1 M sulfate d'ammonium. Aiguilles.

**D.** The Nucleix Suite n°83 (Qiagen). 1 mM chlorure de cobalt (III) hexamine ; 0,2 M acétate de calcium ; 0,05 M cacodylate de sodium pH 6 ; 2 M chlorure de lithium. Aiguilles.

**E.** MBClass n°4 (Qiagen). 0,1 M Tris-HCl pH 8,5 ; 0,5 M sulfate d'ammonium. Plaques.

**F.** Cristaux d'environ 50 x 10 μM, optimisés à partir de la condition MBClass n°4, en goutte pendante d'1 μL. Une partie seulement de la goutte est visible ici.

Nous avons aussi obtenu des profils de diffraction allant jusqu'à une résolution de 3,3 Å avec des cristaux de AtLFY-N (**Figure 1-7**), mais nous n'avons pas pu résoudre le problème de phase par remplacement moléculaire avec la structure de GbLFY-SAM à cause des dimensions trop importantes de la maille (300 Å pour l'un des axes). Nous avons ensuite fait des cristaux à partir de protéine contenant de la sélénométhionine. Ils diffractaient jusqu'à

3,4 Å, mais étaient rapidement endommagés par les radiations, empêchant la collecte d'un nombre suffisant d'images. De plus les dimensions de la maille étaient encore plus grandes (600 Å pour l'un des axes), et les cristaux étaient anisotropes.

La structure de AtLFY-N aurait été un plus à nos résultats mais n'était pas essentielle à l'avancement du projet, nous avons donc décidé de ne pas consacrer plus de temps à sa résolution.



**Figure 1-7 | Exemples de cristaux d'AtLFY-N.**

Photos de cristaux après optimisation de conditions initialement identifiées grâce au robot du HTX Lab. Les trois types de cristaux, bipyramidaux (les deux images de gauche) ou en plaques carrées (l'image de droite) ont poussé à 20°C et ont une taille de l'ordre de 100 x 100 µM.

#### ▪ Vers la structure de la protéine entière

Nous avons tenté d'obtenir la structure de la protéine entière, seule ou en complexe avec de l'ADN, par différentes méthodes, mais sans succès à ce jour. Pour cela, nous avons utilisé PpLFY et GbLFYΔ, mais aussi des variants contenant des mutations dans le domaine N-terminal (mutations TE et RE, voir **Article 1** et paragraphe suivant).

Plusieurs essais de cristallisation ont été menés. Nous avons obtenu quelques petits cristaux, mais soit ils ne diffractaient pas bien, soit la protéine était clivée, donnant à nouveau la structure du domaine N ou C-terminal séparément.

Nous avons fait quelques essais de Small Angle X-ray Scattering (SAXS) avec Adam Round (EMBL, Grenoble). Cette méthode, également basée sur la diffraction des rayons X, peut donner des informations sur la structure 3D de la protéine à basse résolution en solution : forme, organisation des domaines les uns par rapport aux autres, changements de conformation, stœchiométrie des complexes.

Ce type d'informations peut aussi être obtenu par microscopie électronique, méthode que nous avons utilisée avec Guy Schoehn à l'IBS (Institut de Biologie Structurale, Grenoble).

En SAXS et en microscopie électronique nous avons été limités par l'hétérogénéité de nos échantillons qui rend l'interprétation des données très compliquée. Nous envisageons de





**A.** A gauche : structure cristallographique d'un monomère de GbLFY-SAM.  $\alpha$  : hélice alpha. A droite : alignement structural de GbLFY-SAM avec les domaines SAM des protéines listées dans le Tableau 1-2.

**B.** Alignement selon les structures secondaires des protéines, H : hélice  $\alpha$ , L : boucle. La structure la plus fréquente à chaque position est en vert.

**C.** Alignement selon les structures primaires des protéines. Le résidu le plus fréquent à chaque position est surligné en gris.

Malgré leur forte similarité avec GbLFY-SAM, nous avons noté que les autres domaines SAM présentés ici se ressemblent plus entre eux qu'ils ne ressemblent à GbLFY-SAM. En effet, une recherche sur le serveur Dali à partir de l'une des 10 molécules présente dans le **Tableau 1-2** révèle des Z-scores allant jusqu'à 30 et un pourcentage d'identité souvent compris entre 15 et 30, alors que les Z-scores obtenus pour GbLFY-SAM sont autour de 6-7, et l'identité n'atteint pas 20% (**Tableau 1-2**). On remarque notamment que l'hélice  $\alpha_4$  et la boucle L4 sont un peu plus longues chez GbLFY-N que chez les autres domaines SAM (**Figure 1-8A et B**). Le domaine SAM de LFY était donc quasiment impossible à prédire, et semble assez original comparé aux domaines SAM décrits chez les animaux.

**Tableau 1-2 | Similarité structurale entre GbLFY-SAM et d'autres domaines SAM.**

Protéine	Organisme	Z-score	rmsd	lali	nres	%id	N° PDB	Fonction	Référence
Caskin-1	<i>H. sapiens</i>	7.0	3.0	68	142	9	3SEI	Protéine Scaffold Transduction du signal	Stafford <i>et al.</i> , 2011
EPHB2	<i>H. sapiens</i>	6,8	2,7	69	76	10	1B4F	Famille des récepteurs Eph tyrosine kinase Transduction du signal	Thanos <i>et al.</i> , 1999
p63	<i>H. sapiens</i>	6,7	1,9	62	67	13	2Y9U	Famille p53, suppresseur de tumeurs Régulation transcriptionnelle	Sathyamurthy <i>et al.</i> , 2011
EPHA4	<i>M. musculus</i>	6,7	2	65	72	15	1B0X	Famille des récepteurs Eph tyrosine kinase Transduction du signal	Stapleton <i>et al.</i> , 1999
Protein aveugle	<i>D. melanogaster</i>	6,5	2,4	66	75	9	3BS7	Transduction du signal Développement	Rajakulendran <i>et al.</i> , 2008
Harmonin	<i>H. sapiens</i>	6,2	2,6	66	74	17	3K1R	Protéine Scaffold Développement	Yan <i>et al.</i> , 2010
NusA	<i>E. coli</i>	6,1	3,0	69	70	12	1U9L	Facteur d'élongation de la transcription Régulation transcriptionnelle	Bonin <i>et al.</i> , 2004
EPHA1	<i>H. sapiens</i>	6,1	2,3	62	68	16	3KKA	Famille des récepteurs Eph tyrosine kinase Transduction du signal	Walker <i>et al.</i> , non publié
DGK delta	<i>H. sapiens</i>	6,0	2,2	61	68	11	3BQ7	Enzyme kinase diacylglycerol Transduction du signal	Harada <i>et al.</i> , 2008
SAMSN1	<i>M. musculus</i>	5,8	2,5	62	78	11	1V38	Information non disponible	Goroncy <i>et al.</i> , non publié

Tableau des 10 protéines ayant la plus grande similarité structurale avec GbLFY-SAM sur la PDB selon le serveur Dali. Le Z-score est un indicateur de similarité qui tient compte de la distance entre les atomes et de la pertinence statistique de la similarité, une valeur au dessus de 2 signifie que les protéines sont similaires, une valeur au dessus de 8 indique qu'elles sont probablement homologues. Le rmsd (root mean square deviation) est une mesure de distance (en Å) entre les carbones alpha de deux protéines, plus cette valeur est petite, plus les protéines sont similaires. lali est le nombre de résidus alignés pour l'analyse, nres est le nombre total de résidus cristallisés, %id est le pourcentage d'identité au niveau des séquences primaires.

Nous avons vu dans l'**Article 1** que LFY-SAM est un domaine d'oligomérisation. En effet, dans le cristal, les monomères forment des interactions « tête-queue » et s'organisent sous forme de polymère hélicoïdal. D'autres domaines SAM forment des polymères lorsqu'ils sont sous forme de cristaux (**Tableau 1-3**). Pour certains d'entre eux la polymérisation a été observée en solution par microscopie électronique, le plus souvent sous forme de fibres (TEL-SAM, Yan-SAM, Ph-SAM, DGK-SAM), mais aussi parfois sous forme de films (DGK-SAM). De nombreux autres domaines SAM, dont la structure 3D n'a pas encore été résolue, polymérisent en solution (Knight *et al.*, 2011). Cette propriété est d'ailleurs responsable de leur agrégation et donc de la difficulté de les étudier (Knight *et al.*, 2011).

**Tableau 1-3 | Domaines SAM formant des polymères et dont la structure est connue.**

Nom	Famille	Etat d'oligomérisation	Monomères / tour	Distance répétée (Å)	N° PDB	Référence
TEL-SAM	ETS	Polymère (protéines mutantes)	6	53	1JI7	Kim <i>et al.</i> , 2001
TEL-SAM	ETS	Polymère (protéines mutantes) WT et	6	52	1LKY	Tran <i>et al.</i> , 2002
Yan-SAM	ETS	Polymère (modèle)	6	52	1SV4	Qiao <i>et al.</i> , 2004
Scm-SAM	Polycomb	Polymère	6	52	1PK3	Kim <i>et al.</i> , 2005
Ph-SAM	Polycomb	Polymère	6	45	1KW4	Kim <i>et al.</i> , 2002
DGK-SAM	Diacylglycerol kinase	Polymère	6	33	3BQ7	Harada <i>et al.</i> , 2008
Shank3-SAM	Protéine scaffold	Polymère/ film	6	40	2F3N, 2F44	Baron <i>et al.</i> , 2006
Caskin1	Protéine scaffold	Polymère (domaines SAM en tandem)	8	40	3SEI	Stafford <i>et al.</i> , 2011
LFY-SAM	-	Polymère	12	78	-	

Toutes les structures ont été obtenues par cristallographie.

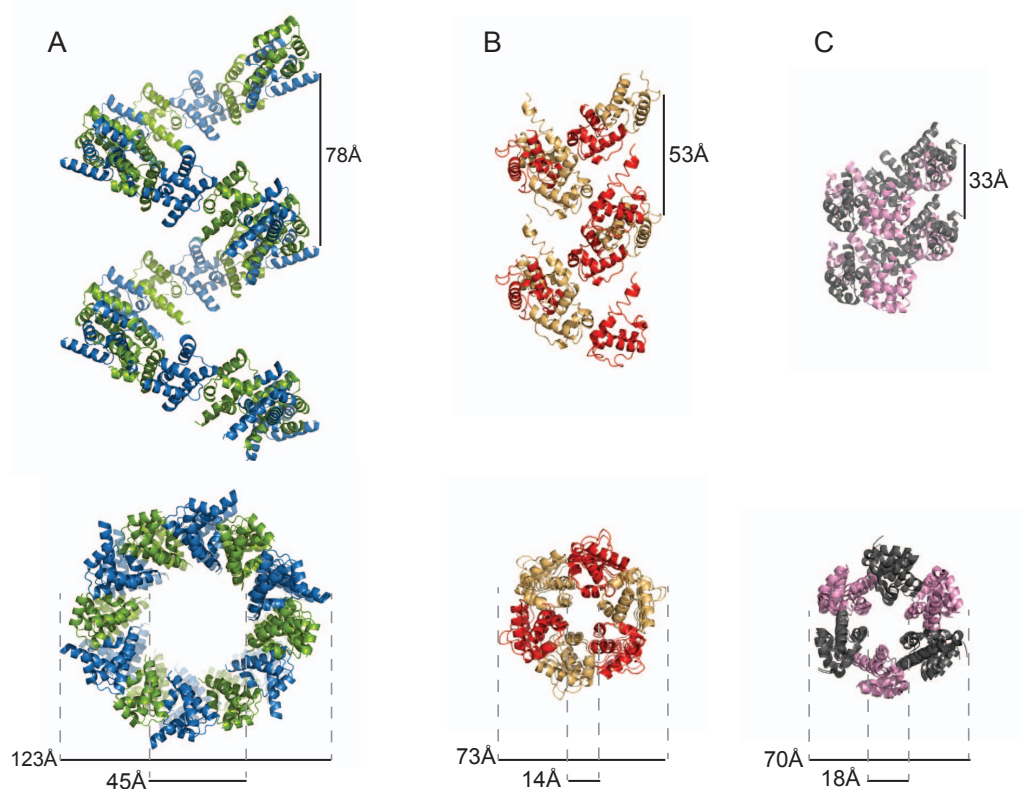
Nous avons montré que l'organisation tête-queue observée dans le cristal s'appliquait aussi en solution pour GbLFY-SAM (ainsi que pour AtLFY-SAM). Mais nous avons observé, en SEC et SEC-MALLS, que l'oligomérisation était limitée à environ 8 molécules (et fonction de la concentration en protéine ; **Article 1**). La polymérisation n'est donc pas infinie comme le suggère la structure mais nous ne savons pas ce qui la limite en solution.

Les autres domaines SAM décrits interagissent plutôt par des contacts hydrophobes (Kim *et al.*, 2002, 2001, 2005 ; Qiao *et al.*, 2004 ; Baron *et al.*, 2006 ; Harada *et al.*, 2008), alors que nous observons des interactions polaires pour GbLFY-SAM. Cela explique peu être que GbLFY-SAM soit relativement soluble.

En plus de la nature différente des interactions, le polymère de GbLFY-SAM présente une forme inédite, par comparaison avec les autres polymères de domaines SAM cristallisés (**Tableau 1-3, Figure 1-9**). GbLFY-SAM compte 12 monomères par tour d'hélice et un pas

d'hélice (la distance entre 2 monomères équivalents) élevé (78 Å), alors qu'on ne compte que 6 à 8 monomères par tour et un pas de 33 à 53 Å chez les autres domaines SAM (**Tableau 1-3**). L'hélice de GbLFY-SAM est donc plus large et moins compacte que celle des autres SAM.

Nous concluons donc que le domaine N-terminal de LFY est un domaine SAM d'oligomérisation présentant plusieurs originalités, comparé aux domaines SAM décrits dans la littérature.



**Figure 1-9 | Exemples de polymères de domaines SAM cristallisés.**

Pour chaque protéine, un nombre de monomères suffisant pour faire deux tours d'hélice est montré. La distance répétée entre chaque tour est indiquée sur l'image du haut ; la largeur et le diamètre de la lumière du polymère sont indiqués sur l'image du bas. Les deux images sont deux vues orthogonales.

**A.** Polymère de GbLFY-SAM, 23 monomères.

**B.** Polymère de TEL-SAM, 11 monomères (Kim *et al.*, 2001 ; PDB 1JI7).

**C.** Polymère de DGK-SAM, 11 monomères (Harada *et al.*, 2008 ; PDB 3BQ7).

Pour valider l'interface d'oligomérisation que nous avons identifiée dans le cristal et pour contrôler le degré d'oligomérisation, nous avons introduit des mutations ponctuelles dans la protéine GbLFY-SAM (**Article 1**). Nous avons créé deux simples mutants, GbLFY-SAM<sub>TE</sub> et GbLFY-SAM<sub>RE</sub>, chacun affecté sur l'une des surfaces et le double mutant GbLFY-SAM<sub>TERE</sub>. Nous avons vu, en mesurant la masse moléculaire en solution de ces protéines par SEC-

MALLS, que GbLFY-SAM formait un oligomère, alors que GbLFY-SAM<sub>TE</sub>, GbLFY-SAM<sub>RE</sub> et GbLFY-SAM<sub>TERE</sub> étaient monomériques. En mélangeant GbLFY-SAM<sub>TE</sub> et GbLFY-SAM<sub>RE</sub> nous obtenons un dimère. Une stratégie similaire et des résultats comparables ont été rapportés pour les autres domaines SAM de structure connue et formant des polymères (Kim *et al.*, 2001 ; Tran *et al.*, 2002 ; Qiao *et al.*, 2004 ; Kim *et al.*, 2005 ; Harada *et al.*, 2008). Ce type de mutations a d'ailleurs été utilisé pour obtenir les protéines sous forme soluble, les formes natives étant agrégées.

La délétion des résidus non conservés en N-terminal n'influence probablement pas l'oligomérisation du domaine SAM, puisqu'une construction démarrant dès le premier résidu (ETH231, résidus 1-159 ; **Figure 1-1** ; **Annexe 1**) est élue à un volume très similaire à celui de la construction utilisée pour la cristallographie et les analyses SEC MALLS (ETH195 ; résidus 55-159) sur une colonne SEC lors de la purification.

À l'occasion du stage M1 de Pierre Jacob, nous avons également purifié et analysé le domaine *linker* de GbLFY (séquence non conservée entre les domaines N et C-terminal). Ce domaine est monomérique en SEC-MALLS et ne participe donc probablement pas à la formation de multimère dans la protéine entière

#### 4. Les diverses fonctions des domaines SAM

Nous avons vu que le domaine N-terminal de LFY était un domaine SAM. Comme plusieurs autres domaines SAM, c'est un domaine d'oligomérisation qui, lorsqu'il est isolé du reste de la protéine, cristallise sous forme de polymère et forme des oligomères de taille limitée en solution.

Les domaines SAM sont en fait très répandus chez les eucaryotes, mais sont aussi présents chez les bactéries. D'après la base de données SMART (<http://smart.embl-heidelberg.de/>), 4169 protéines recensées à ce jour contiendraient un ou plusieurs domaines SAM. Un domaine SAM est prédit chez 8 protéines putatives chez *A. thaliana*, mais aucune n'a été caractérisée jusqu'à présent.

Selon les protéines auxquelles ils appartiennent, ces domaines assurent des fonctions très diverses, comme la régulation transcriptionnelle, la signalisation cellulaire, le contrôle de la localisation subcellulaire, la liaison à l'ARN ou aux lipides (Qiao et Bowie, 2005). Dans tous les cas, le domaine SAM est un domaine d'interaction. Nous nous intéresserons ici aux rôles

des domaines SAM chez les régulateurs transcriptionnels, pour peut-être trouver des indices sur le rôle et le fonctionnement du domaine SAM de LFY.

#### ▪ Le domaine SAM chez les FTs de la famille ETS

La récente revue de Hollenhorst *et al.*, 2011 propose un bilan sur la compréhension actuelle du fonctionnement des protéines ETS, une famille de FTs dont certains possèdent un domaine SAM. Lorsqu'aucune autre référence n'est citée, le texte se réfère à cette revue.

Les membres de la famille ETS sont des régulateurs, activateurs ou répresseurs, de la transcription chez les animaux (8 et 28 paralogues chez la drosophile et chez l'homme, respectivement). Ces facteurs sont caractérisés par la présence d'un domaine ETS de liaison à l'ADN et un tiers d'entre eux possèdent un domaine SAM.

Les protéines Yan et Pointed-P2 (Pnt-P2) font partie de la famille ETS et sont nécessaires au développement de l'oeil chez la drosophile. Elles ont été largement étudiées, ainsi que leurs homologues chez l'homme : TEL (pour Yan) et ETS-1, ETS-2 (pour Pnt-P2). Il ressort de ces analyses qu'un domaine SAM peut influencer l'activité de régulateurs transcriptionnels grâce à divers mécanismes.

#### - Interactions entre plusieurs domaines SAM

Yan fonctionne comme répresseur et est antagoniste à Pnt-P2 qui agit comme activateur. Le domaine SAM de Yan, comme celui de TEL et de LFY, est un domaine d'oligomérisation (**Tableau 1-3**). Le domaine SAM de Pnt-P2, par contre, n'interagit pas avec lui-même. Yan et Pnt-P2 sont toutes deux régulées par phosphorylation par la kinase Rolled (ERK2 chez l'homme). Ces phosphorylations dépendent de leur hétéro-dimérisation avec Mae, une autre protéine possédant un domaine SAM ne formant pas d'homo-oligomère et dépourvue de domaine ETS. Mae interagit avec Yan, Pnt-P2 et avec la kinase Rolled.

Concernant la phosphorylation de Yan par Rolled, deux hypothèses ont été proposées : soit Mae recrute la kinase vers Yan ; soit Mae, qui bloque la polymérisation de Yan, permet à la kinase d'accéder au site de phosphorylation. Au cours du développement, la phosphorylation de Yan entraîne son export du noyau, et par conséquent l'arrêt de son activité de répresseur transcriptionnel.

Au même stade développemental, Pnt-P2 est aussi phosphorylée par Rolled, ce qui lui permet d'activer les gènes précédemment réprimés par Yan. Pnt-P2 interagit avec Rolled et avec Mae via la même surface d'interaction. Mae empêche donc la phosphorylation de Pnt-P2 par Rolled (Qiao *et al.*, 2006). Il a été proposé que ces boucles de régulation assurent un niveau de transcription approprié pour le développement de l'oeil chez la mouche.

Le mécanisme d'activation d'ETS-1 (l'homologue de Pnt-P2) par phosphorylation a été caractérisé. Le site de phosphorylation se trouve dans une région non structurée, à proximité

du domaine SAM et d'une hélice  $\alpha$  décrite comme relativement mobile. La phosphorylation entraîne un positionnement préférentiel de cette hélice, ce qui augmente l'affinité d'ETS-1 pour CBP, un co-activateur général de la transcription.

Il est possible que, comme pour ces facteurs ETS, des interactions entre LFY-SAM et d'autres protéines à domaines SAM régulent l'activité de LFY, en contrôlant son degré d'oligomérisation, ou en permettant des interactions avec des co-régulateurs par exemple.

#### - Oligomérisation et liaison à l'ADN

L'oligomérisation est importante pour la capacité de Yan et TEL à réprimer la transcription de leurs gènes cibles, il est alors possible que cette propriété influence leur liaison à l'ADN (Qiao et Bowie, 2005). Dans le cas de Yan, il a été suggéré que le contrôle de l'oligomérisation via son interaction avec Mae puisse affecter sa capacité à lier l'ADN (Qiao *et al.*, 2004). Il a aussi été proposé que l'oligomérisation permette à la protéine de couvrir de larges régions sur l'ADN. Cette hypothèse a très récemment été testée par ChIP-chip et ChIP-seq en comparant l'occupation sur le génome de Yan avec celle d'un variant monomérique comportant une mutation à la surface EH (Webber *et al.*, 2013). Malgré le fait que les régions liées par la protéine Yan s'étendent sur plusieurs kb, cette caractéristique ne semble pas liée à l'oligomérisation du domaine SAM. En effet, une répartition similaire est observée pour la protéine Yan monomérique. Les auteurs proposent alors que l'oligomérisation puisse avoir un rôle dans la répression transcriptionnelle, plutôt que dans la liaison à des régions étendues sur l'ADN.

Même si l'oligomérisation ne semble pas cruciale pour la liaison à l'ADN à l'échelle de plusieurs kb, il a été montré qu'elle était importante pour lier plusieurs sites de liaison, à l'échelle de quelques dizaines de pb (Green *et al.*, 2010). En effet, des analyses *in vitro* (EMSA), avec des variants monomérique et dimérique de la protéine TEL (la protéine native étant insoluble) ont montrées que l'interaction entre les domaines SAM permettait une liaison coopérative entre deux sites de liaison. Cette coopérativité est possible avec des espacements et des orientations variables entre les sites.

#### ▪ Le domaine SAM chez la famille Polycomb

Les protéines PcG (Polycomb group) sont impliquées dans le maintien de la répression de nombreux gènes du développement, en particulier les gènes *Hox* chez les animaux et les gènes à boîte MADS chez les plantes. Chez les animaux, plusieurs protéines PcG possèdent des domaines SAM, dont Ph (Polyhomeotic) et Scm (Sex comb on midleg) chez la drosophile. Ces protéines font partie du complexe PRC1. Ph et Scm ont été cristallisées sous forme de polymères, comme LFY (**Tableau 1-3** ; Kim *et al.*, 2005, 2002). Des protéines Scm mutantes affectées pour l'oligomérisation ne complètent pas, ou peu, un mutant nul

pour le gène *Scm* chez la mouche, démontrant l'importance de cette propriété pour la fonction de ce FT (Qiao et Bowie, 2005). Il a été proposé, comme pour la protéine Yan, que l'oligomérisation permette à ces protéines de couvrir l'ADN sur de grandes distances.

#### ▪ Le domaine SAM chez p63 et p73

Les protéines p63 et p73 sont les deux autres membres de la famille p53. Ces FTs ne sont pas des suppresseurs de tumeur comme p53 mais sont impliqués dans le développement chez les mammifères. Plusieurs isoformes existent chez l'homme et parmi elles, deux ont un domaine SAM : p63 $\alpha$  et p73 $\alpha$ . La structure 3D de leur domaine SAM est connue (Chi *et al.*, 1999 ; Sathyamurthy *et al.*, 2011) et ces protéines sont monomériques. Plusieurs mutations touchent p63-SAM et perturbent la fonction de la protéine *in vivo*, mais ces mutations semblent déstabiliser le repliement global du domaine et ne sont donc pas très informatives sur son rôle (Sathyamurthy *et al.*, 2011). Pour p73, il a été montré que la présence du domaine SAM avait un effet inhibiteur sur l'activité transcriptionnelle de la protéine, en empêchant son interaction avec le co-activateur CBP (Liu et Chen, 2005). Le domaine SAM de p73 pourrait aussi permettre l'interaction de la protéine avec une ubiquitine ligase, ce qui aboutirait à sa dégradation (Hosoda *et al.*, 2005). Enfin, l'interaction entre p73-SAM et des lipides a été observée *in vitro*, mais la pertinence biologique de cette propriété n'est pas bien connue (Barrera *et al.*, 2003).

#### ▪ Oligomérisation et liaison à l'ADN chez des protéines sans domaine SAM

Nous avons vu qu'il existait un lien entre oligomérisation et liaison à l'ADN chez les protéines ETS à domaine SAM. D'autres domaines protéiques sont aussi capables d'oligomériser. L'importance de cette propriété pour la liaison ADN-protéine a été montrée pour quelques FTs possédant de tels domaines.

Le facteur GAGA chez la drosophile par exemple, comporte un domaine de liaison à l'ADN de type *zinc finger*, un domaine POZ/BTB qui permet l'oligomérisation de la protéine et un domaine riche en glutamines dont la fonction n'est pas clairement définie. Par des expériences d'EMSA et d'empreinte à la DNase, il a été montré que le domaine POZ permettait la liaison coopérative sur plusieurs sites de liaison (Katsani *et al.*, 1999). Beaucoup de gènes contrôlés par le facteur GAGA possèdent de nombreux sites de liaison dans leurs promoteurs et l'oligomérisation du domaine POZ permettrait de pouvoir les lier efficacement. De plus, la présence du domaine POZ entraîne une courbure de l'ADN lors de la liaison (observée par microscopie électronique). Il a été proposé que ce type de liaison déstabilise les contacts histone-ADN et permette le maintien d'une conformation accessible

de la chromatine, après l'intervention de remodeleurs (Katsani *et al.*, 1999 ; Adkins *et al.*, 2006).

Des conclusions similaires ont été tirées de l'étude du FT PRH (Proline-Rich Homeodomain), qui est impliqué dans le développement chez les vertébrés. Grâce à une séquence dans son domaine riche en prolines, PRH forme des complexes de haut poids moléculaires *in vitro* et *in vivo* et lie l'ADN sous forme d'oligomères. Des analyses *in vitro* (EMSA et empreinte à la DNase), avec la protéine entière ou dépourvue de son domaine d'oligomérisation, ont montré que ce domaine permettait la liaison coopérative sur plusieurs sites de liaison et entraînait une torsion de la molécule d'ADN liée (Williams *et al.*, 2008).

Une multitude d'interactions sont décrites dans la littérature pour les domaines SAM au sein de FTs et ces interactions ont des effets très variés. Elles permettent la régulation de l'activité des FTs, leur interaction avec de co-régulateurs, et la modulation de leur capacité à lier l'ADN. L'oligomérisation, chez certaines protéines à domaine SAM mais aussi chez d'autres protéines possédant un domaine d'oligomérisation de nature différente, peut jouer un rôle majeur, notamment pour la liaison à l'ADN.

De nombreuses hypothèses sont donc envisageables pour le rôle du domaine SAM de LFY. Après avoir vérifié que l'oligomérisation du domaine SAM était nécessaire à la fonction de LFY *in planta*, nous nous sommes concentrés, dans l'**Article 1** et dans la suite de ce chapitre, sur le rôle de l'oligomérisation pour la liaison à l'ADN.

## 5. Rôle de l'oligomérisation pour la fonction de LFY comme régulateur floral

Pour connaître l'importance de l'oligomérisation du domaine SAM pour la fonction de LFY chez *Arabidopsis*, nous avons évalué la capacité de LFY<sub>TERE</sub> à prendre en charge cette fonction.

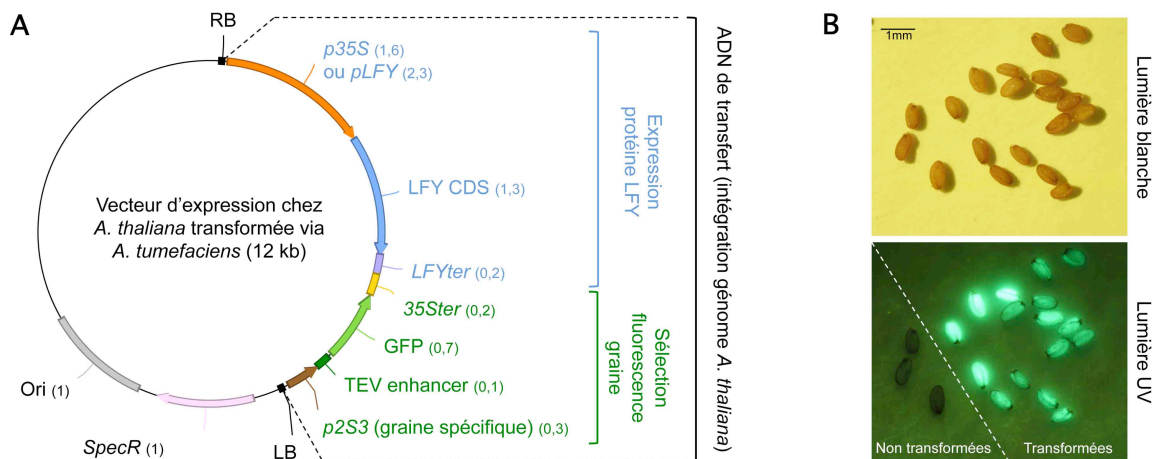
### ▪ Génération de lignées exprimant LFY ou les variants d'oligomérisation chez *Arabidopsis*

Pour cela, nous avons créé des lignées transgéniques permettant d'exprimer la protéine sauvage, ou les variants spécifiquement affectés pour l'oligomérisation.

Nous disposons au laboratoire de vecteurs de type « Alligator » (Bensmihen *et al.*, 2004) contenant la plupart des éléments nécessaires à la création de ces lignées. Il a fallu insérer



la séquence codante pour la protéine LFY, sauvage ou mutées dans ces vecteurs. Les constructions obtenues permettent l'expression des protéines sous le contrôle de *pLFY* pour faire des tests de complémentation fonctionnelle, ou sous le contrôle du promoteur constitutif *p35S* (**Figure 1-10A**). Les vecteurs contiennent aussi la séquence codante pour la GFP (Green Fluorescent Protein) sous le contrôle d'un promoteur spécifique de la graine, ce qui permet la sélection des plantes transgéniques au stade graine (**Figure 1-10A et 1-10B**).



**Figure 1-10 | Génération de lignées exprimant LFY ou les variants d'oligomérisation chez *A. thaliana*.**

**A.** Schéma du vecteur permettant la transformation d'*A. thaliana* par *A. tumefaciens*. Il contient une cassette permettant l'expression de la protéine LFY (en bleu) et une cassette permettant l'émission de fluorescence dans les graines comme moyen de sélection des plantes transgéniques (en vert). La partie se trouvant entre les séquences RB et LB est intégrée dans le génome d'*A. thaliana* lors de la transformation. Le nombre entre parenthèse indique la taille de chaque élément en kb. *p35S* : promoteur constitutif du CaMV (*Cauliflower Mosaic Virus*) ; *pLFY* : promoteur du gène *LFY* ; CDS : séquence codante ; *LFYter* : terminateur du gène *LFY* ; *35Ster* : terminateur du CaMV ; GFP : séquence codante pour la *Green Fluorescent Protein* ; TEV enhancer : séquence favorisant la traduction ; *p2S3* : promoteur du gène *At2S3*, spécifique du stade graine ; RB et LB (*right border* et *left border* respectivement) : séquences permettant la recombinaison avec le génome de la plante transformée ; *SpecR* : gène de résistance à la spectinomycine ; Ori : origine de répllication bactérienne.

**B.** Sélection des graines transformées. Les graines transgéniques ne sont pas distinguables des non-transgéniques sous lumière blanche (image du haut), mais elles émettent de la fluorescence (GFP) sous lumière ultra-violette (UV).

Pour obtenir des plantes transgéniques, l'une des méthodes la plus communément utilisée est la transformation par *Agrobacterium tumefaciens*. Cette bactérie est naturellement capable de transférer un fragment d'ADN vers le génome de la plante hôte, des souches optimisées pour la transformation chez *A. thaliana* sont disponibles. La méthode consiste à mettre en présence les tissus reproducteurs de la plante (les fleurs) avec la bactérie

contenant l'ADN à transférer. Parmi la descendance de ces plantes, une partie (typiquement 0,05%) a intégré le transgène, ce sont les transformants primaires (T1). Ni le lieu d'intégration dans le génome, ni le nombre de copies intégrées ne peuvent être contrôlés. Pour s'affranchir de ce biais, il faut analyser un grand nombre de T1. La transformation a été effectuée sur des plantes hétérozygotes pour l'allèle *lfy-12* (Q32stop), car ces plantes ont un phénotype sauvage et donnent des graines, contrairement au mutant homozygote pour *lfy-12* qui est stérile. Parmi les plantes T1, la moitié est donc en fond hétérozygote (*lfy-12 +/-*), un quart en fond WT et un quart en fond *lfy-12*.

Au total, 8 types de lignées ont été générés et contiennent les séquences codantes pour AtLFY, AtLFY<sub>TE</sub>, AtLFY<sub>RE</sub>, AtLFY<sub>TERE</sub>, soit sous *pLFY*, soit sous *p35S*. Jusqu'à présent, seules celles contenant AtLFY et AtLFY<sub>TERE</sub> ont été exploitées. Dans l'optique de pouvoir faire des croisements entre ces lignées et ensuite vérifier la présence des différents transgènes, des sites de restriction spécifiques pour chaque lignée ont été insérés dans les séquences codantes, via des mutations silencieuses.

#### ▪ Evaluation de la capacité de LFY<sub>TERE</sub> à assumer la fonction florale de LFY

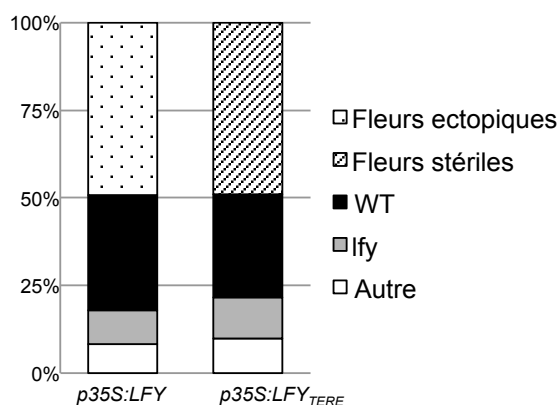
Comme expliqué dans l'Article 1, LFY<sub>TERE</sub> ne complémente que faiblement un mutant *lfy* comparé à LFY, et son expression constitutive n'entraîne pas la formation de fleurs ectopiques comme celle de LFY. Nous avons donc conclu que l'oligomérisation du domaine SAM était importante pour la fonction florale de LFY.

Quelques précisions sont apportées ici sur le phénotype des plantes *p35S:LFY<sub>TERE</sub>* : non seulement ces plantes ne font pas de fleurs de rosette, mais les premières fleurs portées par les hampes florales sont anormales, elles semblent fermées, et ne donnent pas de siliques. Les plantes T1 présentant ce phénotype « fleurs stériles » ont été comptées chez les *p35S:LFY<sub>TERE</sub>* et forment un effectif du même ordre que les plantes ayant des fleurs ectopiques chez les *p35S:LFY* (la moitié des T1 observées) (Tableau 1-4 et Figure 1-11). Nous n'avons pas caractérisé plus en détail ce phénotype, mais il suggère que la protéine LFY<sub>TERE</sub> ait un effet dominant négatif sur la protéine LFY endogène.

**Tableau 1-4 | Classes phénotypiques détaillées des plantes T1 *p35S:LFY* et *p35S:LFY<sub>TERE</sub>*.**

	Phénotype	Fleurs ectopiques	Fleurs stériles	WT	<i>lfy</i>	Autre	Total
<i>p35S:LFY</i>	Nombre de plantes	30	0	20	6	5	61
	% de plantes	49,2	0	32,8	9,8	8,2	100
<i>p35S:LFY<sub>TERE</sub></i>	Nombre de plantes	0	25	15	6	5	51
	% de plantes	0	49,0	29,4	11,8	9,8	100

Effectif pour chaque classe phénotypique. Les classes sont décrites en Figure 1-11.



**Figure 1-11 | Classes phénotypiques détaillées des plantes T1 *p35S:LFY* et *p35S:LFY<sub>TERE</sub>*.**

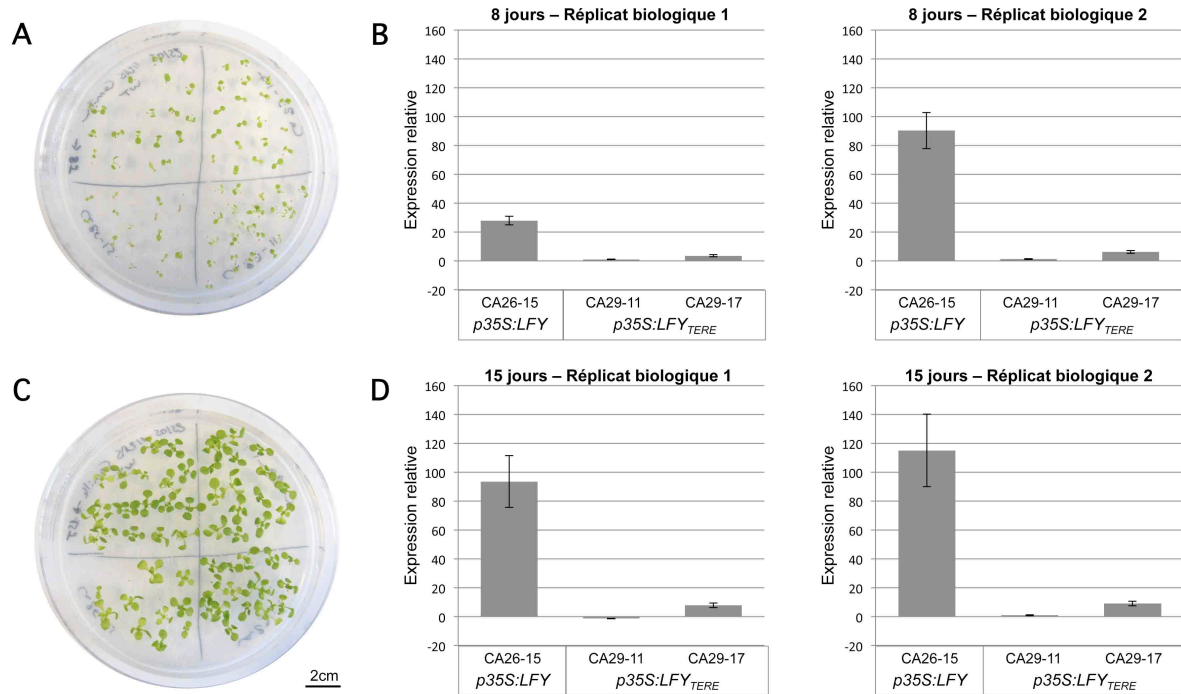
Histogramme des effectifs (en pourcentage) pour chaque classe phénotypique. Les plantes ayant des fleurs de rosettes forment la classe « fleurs ectopiques » ; celles présentant plusieurs fleurs stériles ayant une apparence de fleurs fermées sont comptées dans la classe « fleurs stériles » : les plantes ayant un phénotype sauvage forment la classe « WT » ; celle ayant le phénotype d'un mutant *lfy* forment la classe « *lfy* » ; les plantes n'appartenant à aucune des classes décrites précédemment forment la classe « autre ». Le nombre de plantes appartenant à chacune des classes est indiqué dans la Tableau 1-4.

Avant de connaître les profils de liaison de LFY et LFY<sub>TERE</sub> sur le génome grâce au ChIP-seq, nous avons analysé la capacité de ces deux protéines à réguler l'expression d'*AP1*, gène pour qui l'activation par LFY est bien caractérisée. On sait qu'*AP1* est fortement surexprimé chez *p35S:LFY* par rapport à une plante sauvage (Chahtane *et al.*, 2013 ; Parcy *et al.*, 1998 ; Wagner *et al.*, 1999). Nous avons cherché à savoir si une telle régulation avait lieu chez *p35S:LFY<sub>TERE</sub>*.

Nous avons donc analysé le niveau d'expression d'*AP1* chez *p35S:LFY* (lignée CA26-15) et *p35S:LFY<sub>TERE</sub>* (lignées CA29-11 et CA29-17) par rapport à celui chez des plantes WT, par RT-PCR quantitative (transcription réverse des ARNm en ADN suivie d'une PCR quantitative). Les ARNs ont été extraits à partir de plantules de 8 et 15 jours cultivées en jours courts. Dans ces conditions, les plantules sont au stade végétatif (**Figure 1-12A et C**), la protéine LFY endogène n'est pas ou très peu exprimée et l'on s'attend à une faible expression du gène *AP1* chez une plante sauvage. On constate une très forte expression d'*AP1* dès 8 jours chez *p35S:LFY*, alors que l'expression chez *p35S:LFY<sub>TERE</sub>* est beaucoup plus faible (**Figure 1-12B et D**). Nous avons donc conclu que LFY<sub>TERE</sub> n'était pas capable d'induire l'expression d'*AP1* aussi fortement que LFY dans ces conditions.

Nous avons vu par la suite en ChIP-Seq que la liaison sur la région régulatrice d'*AP1* chez *Arabidopsis* (à 15 jours, jours longs) était réduite chez *p35S:LFY<sub>TERE</sub>* par rapport à celle observée chez *p35S:LFY* (**Article 1** ; la hauteur du pic principal est réduite de l'ordre de 2 fois). On peut se demander si cette faible différence de liaison explique à elle seule la perte de régulation et si la protéine LFY<sub>TERE</sub> liée a la même capacité que LFY à réguler les gènes.

Pour le savoir, il faudrait regarder si des gènes aussi bien liés par LFY<sub>TERE</sub> que par LFY en ChIP-seq sont régulés de manière similaire par les deux protéines.



**Figure 1-12 | Expression relative du gène *AP1* chez les plantes *p35S:LFY* et *p35S:LFY<sub>TERE</sub>*.**

**A.** Aspect des plantules de 8 jours cultivées en jours courts.

**B.** Expression relative d'*AP1* chez *p35S:LFY* (lignée CA26-15) et *p35S:LFY<sub>TERE</sub>* (lignées CA29-11 et CA29-17) par rapport par rapport à celle chez des plantes WT, à 8 jours (jours courts). Deux réplicats biologiques sont montrés.

**C.** Aspect des plantules de 15 jours cultivées en jours courts.

**D.** Expression relative d'*AP1* chez *p35S:LFY* (lignée CA26-15) et *p35S:LFY<sub>TERE</sub>* (lignées CA29-11 et CA29-17) par rapport par rapport à celle chez des plantes WT, à 15 jours (jours courts). Deux réplicats biologiques sont montrés.

## 6. Rôle de l'oligomérisation du domaine SAM pour la liaison à l'ADN *in vitro*

LFY possède une interface de dimérisation au niveau du domaine de liaison à l'ADN. Lorsque ce domaine est isolé (LFY-C) la dimérisation a lieu de manière coopérative sur l'ADN (Hamès *et al.*, 2008). Dans cette étude, nous avons vu que LFY avait aussi une interface d'oligomérisation portée par le domaine SAM en N-terminal de la protéine. Ce domaine est capable d'oligomériser lorsqu'il est isolé (LFY-SAM). Ces deux domaines sont reliés par un *linker* flexible et présentent des symétries différentes : le domaine SAM fait des

interactions « tête-queue » alors que le domaine C-terminal forme un dimère symétrique « tête-tête ».

**Comment ces deux domaines fonctionnent-ils au sein de la protéine entière ? Quel est le rôle du domaine SAM pour la liaison à l'ADN ?**

Pour répondre à ces questions, nous avons cherché à comprendre le comportement de LFY sur l'ADN et en solution, en identifiant les différents complexes formés dans chacune de ces conditions. Les protéines mutantes affectées pour l'oligomérisation du domaine SAM vont d'une part nous permettre d'évaluer l'importance de ce domaine pour la liaison à l'ADN, et d'autre part nous aider à mieux comprendre le fonctionnement de la protéine native.

▪ **Les complexes ADN-LFY**

Nous avons montré dans l'**Article 1**, en EMSA et SEC-MALLS, que l'oligomérisation par le domaine SAM affectait la liaison à l'ADN *in vitro*. Les principales conclusions de ces analyses sont rappelées ci-dessous.

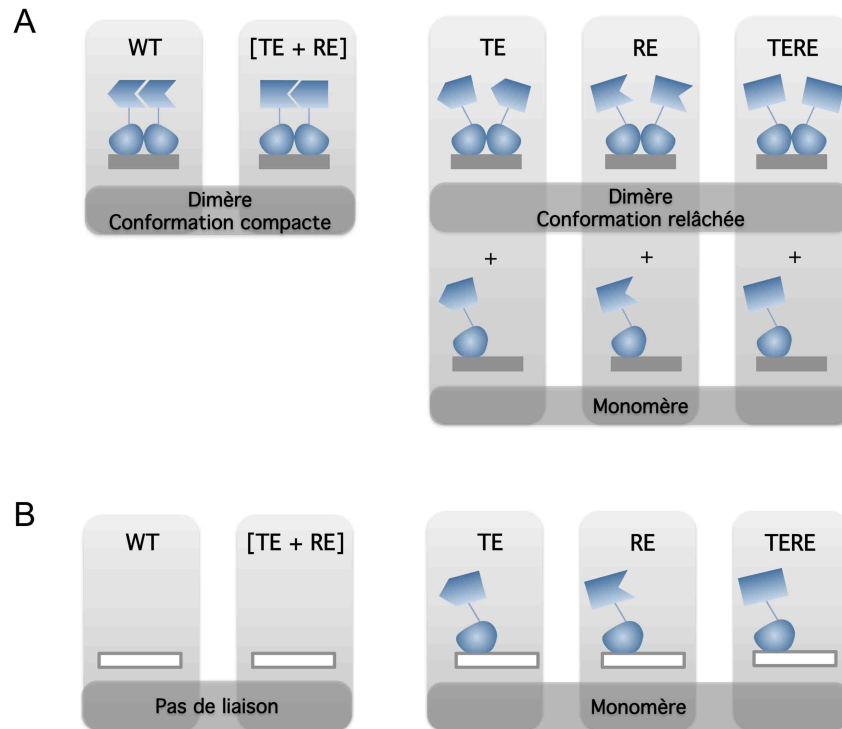
Nous avons vu que LFY pouvait former un dimère sur un site de liaison (oligonucléotide *AP1*), que le domaine SAM soit fonctionnel ou non. Cette observation est en accord avec le fait que le domaine LFY-C isolé peut se lier sous forme de dimère sur un site de liaison (Hamès *et al.*, 2008). Lorsque le domaine SAM a au moins une interface fonctionnelle (LFY et LFY<sub>[TE+RE]</sub>), les domaines SAM s'associent au sein du dimère (dimère compact migrant rapidement sur gel natif) ; lorsque les deux interfaces sont affectées (LFY<sub>TE</sub>, LFY<sub>RE</sub>, LFY<sub>TERE</sub>) les domaines SAM n'interagissent pas (le dimère a une conformation plus relâchée et migre moins rapidement) (**Figure 1-13A**).

Nous avons remarqué que contrairement aux protéines ayant leurs domaines SAM capables de s'associer (LFY et LFY<sub>[TE+RE]</sub>), LFY<sub>TE</sub>, LFY<sub>RE</sub>, LFY<sub>TERE</sub> pouvaient se lier sous forme de monomère sur *AP1* et étaient également capables de reconnaître des oligonucléotides pour lesquels le site de liaison a été muté (*AP1m*, *AP1m1*, *AP1m2*). L'interaction des domaines SAM au sein du dimère réduit donc la liaison non spécifique à l'ADN (**Figure 1-13B**).

Enfin, sur un oligonucléotide comportant deux sites de liaison (*AGI-II*), les deux dimères de LFY, occupant chacun un site, interagissent entre eux facilitant la formation du tétramère (coopérativité). Cette conclusion a été déduite de la comparaison des profils de liaison de LFY et de LFY<sub>[TE+RE]</sub>. En effet, LFY<sub>[TE+RE]</sub>, qui forme un dimère équivalent au dimère de LFY sur un simple site de liaison, mais qui n'est en principe pas capable de former un oligomère plus grand, n'est pas aussi efficace que LFY pour former le tétramère. En complément de l'EMSA présenté dans l'**Article 1** avec l'oligonucléotide *AGI-II*, l'expérience a aussi été réalisée sur l'oligonucléotide *2sites* comportant également deux sites de liaison (**Figure 1-14A**). Le profil est similaire à celui observé sur *AGI-II*. L'oligonucléotide *AGI-II* est issu d'une

région génomique liée par LFY *in planta* (ChIP-seq, Moyroud *et al.*, 2011), alors que *2sites* est un oligonucléotide artificiel dessiné pour cette expérience. Les deux sites sont espacés de 47 pb pour *AGI-II* et de 37 pb pour *2sites* (nombre de pb du centre du premier site au centre du second).

Les complexes observés sur ces oligonucléotides à deux sites sont schématisés sur la **Figure 1-14B**.

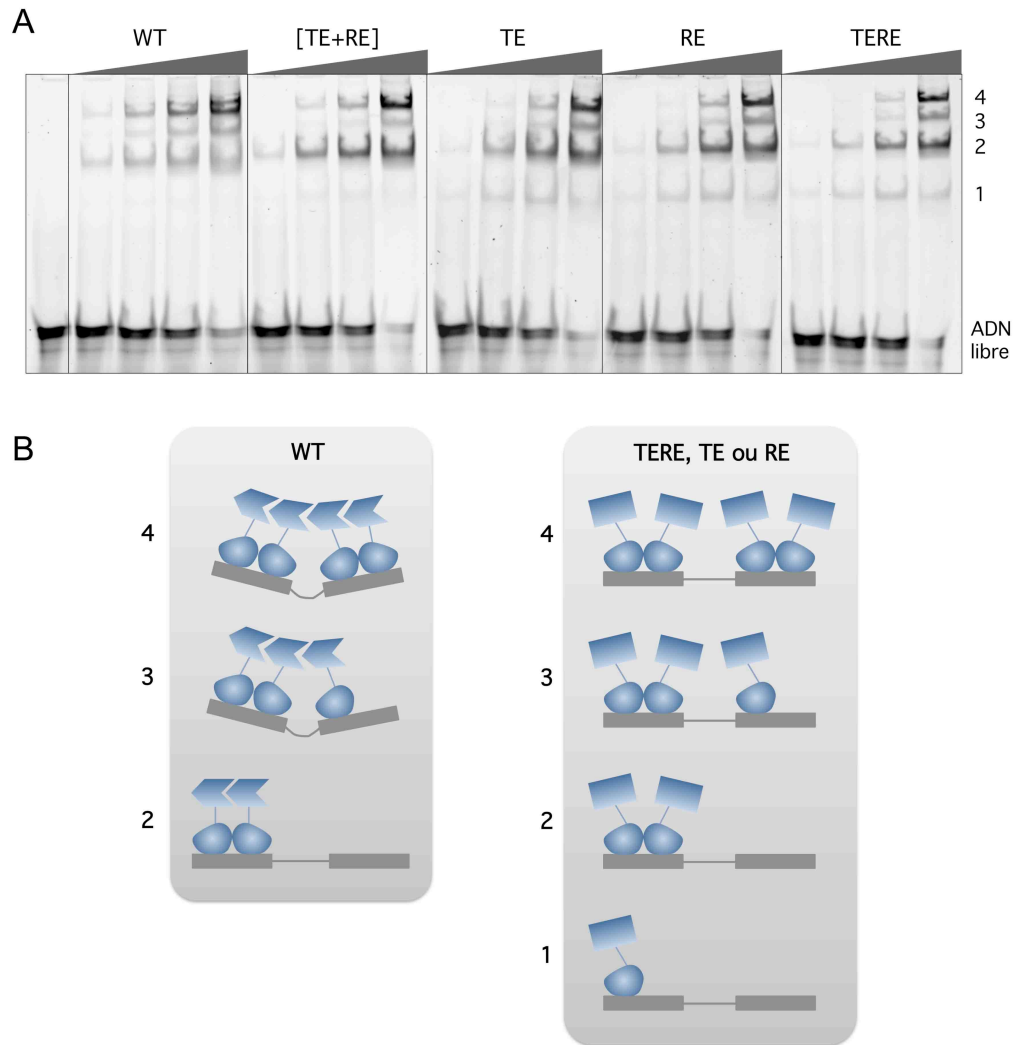


**Figure 1-13 | Comportement de LFY et des mutants d'oligomérisation sur un site de liaison et sur une séquence non spécifique.**

L'ADN est représenté par un rectangle gris (rectangle plein : site de liaison pour LFY, rectangle vide : séquence non spécifique). La protéine LFY est en bleu (forme en contact avec l'ADN : domaine de liaison à l'ADN ; forme en flèche ou en rectangle : domaine SAM ; trait reliant les deux domaines : *linker*).

**A.** Liaison sur un fragment d'ADN contenant un site de liaison.

**B.** Liaison sur un fragment d'ADN contenant une séquence non spécifique.



**Figure 1-14 | Comportement de LFY et des mutants d'oligomérisation sur un ADN contenant deux sites de liaison pour LFY.**

**A.** EMSA avec 10 nM d'ADN *2sites* (deux sites de liaison pour LFY) et 0, 50, 100, 250, 500 nM de protéine GbLFY $\Delta$ , GbLFY $\Delta_{[TE+RE]}$ , GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  ou GbLFY $\Delta_{TERE}$ . 1, 2, 3, 4 indique s'il s'agit d'un monomère, dimère, trimère et tétramère, respectivement.

**B.** Schéma des complexes ADN-protéine observés sur une séquence ADN contenant deux sites de liaison. L'ADN est représenté en gris (rectangle plein : site de liaison pour LFY). La protéine LFY est en bleu (forme en contact avec l'ADN : domaine de liaison à l'ADN, forme en flèche ou en rectangle : domaine SAM, trait reliant les deux domaines : *linker*). 1, 2, 3, 4 indique s'il s'agit d'un monomère, dimère, trimère et tétramère, respectivement.

#### ▪ LFY en solution

Nous avons étudié l'état d'oligomérisation de LFY en présence d'ADN, mais nous avons été limités par la difficulté d'analyser la protéine en solution, sans ADN. En effet, sans ADN, la protéine entière n'est soluble qu'en présence de NaCl à forte concentration (0,8 M), ce qui empêche l'analyse de sa masse moléculaire dans des conditions pertinentes.

Nous avons cependant réalisé des analyses en SEC et SEC-MALLS en présence de 0,8 M NaCl. Nous savons que le NaCl affecte l'oligomérisation de GbLFY-SAM (**Figure 1-15A**, **Tableau 1-5**), nous sommes partis du principe que si la protéine entière était oligomérique en présence de 0,8 M NaCl, alors elle l'était certainement à des concentrations plus faibles. Par contre, si la protéine est monomérique en présence de 0,8 M NaCl, nous ne pourrions pas savoir si elle l'est aussi avec moins de sel.

Comme expliqué précédemment, la méthode de SEC repose sur les propriétés du polymère qui constitue la colonne et dont les pores retiennent plus longtemps les molécules de petite taille. D'après une calibration réalisée avec des protéines standard de masses moléculaires connues, on peut déduire une masse moléculaire apparente pour la protéine analysée. Cette masse moléculaire apparente ne reflète la masse de la protéine que si celle-ci est globulaire, puisque la séparation dépend de leur volume hydrodynamique (dépendant de la masse et de la conformation). Mais elle nous renseigne sur la masse moléculaire relative d'une même protéine analysée dans différentes conditions (concentration, tampon), ou de plusieurs protéines très proches comme des protéines portant des mutations ponctuelles. L'avantage de cette méthode est qu'elle peut être utilisée en routine au laboratoire.

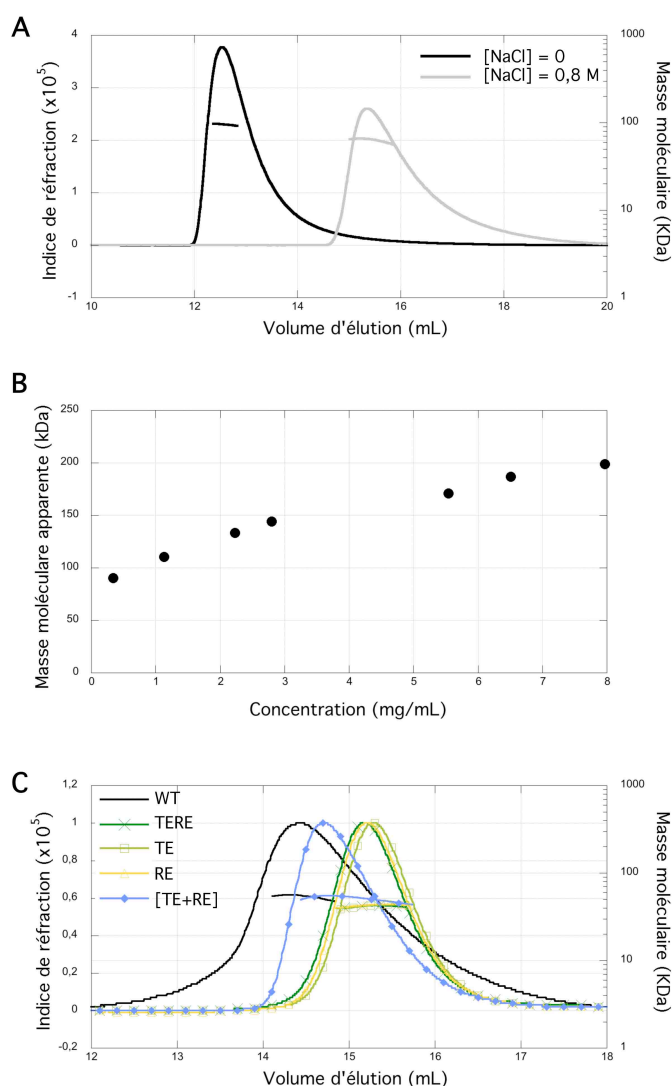
Pour une analyse précise de la masse moléculaire de LFY, nous avons fait appel à la méthode de SEC-MALLS. La SEC est couplée à une mesure de la concentration de l'échantillon par un réfractomètre et à la mesure de la diffusion statique de la lumière. Les données sont soumises à une analyse mathématique qui permet de déduire la masse moléculaire des espèces éluées. En effet, la diffusion de la lumière est proportionnelle au produit de la concentration par la masse moléculaire (Gerard *et al.*, 2007). C'est une méthode dite absolue car elle n'est pas basée sur une calibration. Ces analyses nécessitent un équipement particulier dont ne disposons pas au laboratoire, elles ont été réalisées en collaboration avec le Pr. Marc Jamin (UVHCI, Grenoble).

En SEC, dans un tampon à 0,8 M NaCl, le volume d'éluion de GbLFY $\Delta$  varie en fonction de la concentration en protéine. La masse moléculaire apparente calculée varie du simple au double entre 0,3 et 8 mg/mL de protéine (**Figure 1-15B**). En SEC-MALLS, dans le même tampon avec 4,5 mg/mL de protéine, la masse moléculaire mesurée pour GbLFY $\Delta$  (53,6 kDa) est légèrement supérieure à celle attendue pour un monomère (43,5 kDa) mais n'atteint pas celle attendue pour un dimère (87 kDa), le pic est asymétrique, signe d'un équilibre entre plusieurs espèces. Dans les mêmes conditions, GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$  et GbLFY $\Delta_{TERE}$  sont clairement monomériques (masses moléculaires mesurées comprises entre 41,5 et 42,6 kDa, pic symétrique), et le mélange GbLFY $\Delta_{[TE+RE]}$  a une masse moléculaire mesurée équivalente à celle de GbLFY $\Delta$  (50,9 kDa, pic asymétrique) (**Figure 1-15C** et **Tableau 1-5**).

À forte concentration de sel, condition défavorable aux interactions protéine-protéine, il semble donc que même si GbLFY $\Delta$  ne forme pas un dimère stable, la protéine soit quand



même capable d'interagir grâce à son domaine SAM. Ce résultat semble indiquer que LFY serait capable de s'associer par son domaine SAM en solution, sans ADN. Mais il ne s'agit que d'un indice, cela reste à démontrer.



**Figure 1-15 | Indices sur l'état d'oligomérisation de la protéine entière sans ADN.**

**A.** L'état d'oligomérisation du domaine N-terminal est affecté par la présence de sel. SEC-MALLS sur GbLFY-SAM dans un tampon sans NaCl (20 mM Tris-HCl pH 8 ; 1 mM TCEP) ou à forte concentration en NaCl (0,8 M NaCl ; 20 mM Tris-HCl pH 8 ; 1 mM TCEP). Les masses moléculaires mesurées sont reportées dans le Tableau 1-5.

**B.** Masse moléculaire apparente de GbLFY $\Delta$  en fonction de la concentration en protéine par SEC. Entre 0,3 et 8 mg/mL de protéine ont été injectés sur une colonne équilibrée avec un tampon à forte concentration en sel (0,8 M NaCl ; 20 mM Tris-HCl pH 8 ; 1 mM TCEP).

**C.** SEC-MALLS sur GbLFY $\Delta$ , GbLFY $\Delta_{TE}$ , GbLFY $\Delta_{RE}$ , GbLFY $\Delta_{TERE}$  et GbLFY $\Delta_{[TE+RE]}$  à forte concentration en NaCl (0,8 M NaCl ; 20 mM Tris-HCl pH 8 ; 1 mM TCEP). Les masses moléculaires mesurées sont reportées dans le Tableau 1-5.

**Tableau 1-5 | Indices sur l'état d'oligomérisation de la protéine entière sans ADN.**

Protéine	Concentration (mg/mL)	[NaCl] (M)	Masse moléculaire (kDa)
GbLFY-SAM	4,7	0	95 (+/- 1,6)
GbLFY $\Delta$	4,5	0,8	62,9 (+/- 3)
GbLFY $\Delta_{TE}$	4,2	0,8	42,5 (+/- 1,4)
GbLFY $\Delta_{RE}$	4,2	0,8	42,6 (+/- 1,2)
GbLFY $\Delta_{TERE}$	4,4	0,8	41,5 (+/- 0,8)
GbLFY $\Delta_{[TE+RE]}$	4,5	0,8	50,9 (+/- 3,6)

Masses moléculaires mesurées en SEC-MALLS pour le domaine SAM et la protéine LFY quasi entière du *Ginkgo*, sauvage et mutée, en l'absence ou à forte concentration en NaCl.

Comme discuté dans l'**Article 1**, des données de BiFC (Bimolecular Fluorescence Complementation) publiées par un autre groupe suggèrent l'existence d'un complexe multimérique sans ADN, dans le cytoplasme des cellules végétales (Siriwardana et Lamb, 2012). Cette observation est cohérente avec nos observations. On note que ces analyses ont été réalisées dans des conditions expérimentales éloignées de celles que nous avons utilisées (*in vivo*, protéines de fusion), la comparaison n'est donc pas évidente.

▪ **Par quels mécanismes les complexes LFY-ADN se forment-ils ? Quelles interfaces sont utilisées en solution et sur l'ADN ?**

Pour former un dimère de protéine entière sur un site de liaison, trois scénarios sont possibles :

- Cas 1 : la protéine est monomérique en solution et se dimérise sur l'ADN.
- Cas 2 : la protéine est dimérique en solution. On distingue deux sous-cas : (a) seul le domaine SAM interagit, le domaine de liaison à l'ADN est libre au sein du dimère, son interaction a lieu lors de la liaison à l'ADN. (b) les deux interfaces, en N et C-terminal, sont engagées dans la dimérisation.
- Cas 3 : la protéine en solution est en équilibre entre une forme monomérique (cas 1) et une forme dimérique (cas 2).

Nous ne savons pas de manière certaine si LFY est monomérique ou dimérique en solution. En plus des indices tirés des analyses de masse moléculaire, on trouve d'autres indices en observant les profils de liaison de LFY sur les différents types d'oligonucléotides utilisés lors de cette étude.

LFY<sub>TERE</sub>, pour qui l'interface d'oligomérisation du domaine SAM est mutée, est certainement monomérique en solution. Nous avons vu que sur l'ADN, elle formait aussi un dimère, mais que contrairement à la protéine WT, les domaines SAM n'interagissent pas au sein du dimère (**Figure 1-13A**).

Le complexe monomérique présent chez LFY<sub>TERE</sub> est quasiment invisible chez LFY en EMSA, ce qui peut laisser penser que soit il n'y a pas de monomère en solution, soit la dimérisation via le domaine SAM augmente la coopérativité de fixation entre les deux monomères. Lorsqu'on observe le profil de liaison de LFY sur les oligonucléotides contenant deux sites de liaison (*AGI-II* et *2sites*), on observe, en plus des deux complexes majoritaires, dimère et tétramère, la présence d'un complexe de taille intermédiaire correspondant probablement à un trimère (**Figure 1-14A et B**). Cette observation va plutôt dans le sens des cas 1 ou 3, car la présence d'un trimère impliquerait qu'il y ait du monomère en solution. Cela impliquait aussi que la formation du dimère et celle du trimère ne soient pas

équivalentes. En effet, trois interfaces interviendraient dans la formation du dimère : les surfaces d'interaction ADN-protéine et de dimérisation portées par le domaine de liaison à l'ADN, et la surface d'oligomérisation du domaine SAM ; contre seulement deux interfaces dans la formation du trimère : la surface d'interaction ADN-protéine et la surface d'oligomérisation du domaine SAM.

Toutefois, la présence de monomère en solution pour la protéine WT (cas 1 et 3) est difficilement compatible, à première vue, avec les profils de liaison observés sur les oligonucléotides ayant un site de liaison muté (*AP1m*, *AP1m1*, *AP1m2*) (**Figure 1-13B**). On s'attend à ce qu'un monomère LFY et un monomère LFY<sub>TERE</sub> se comportent de la même manière vis-à-vis de ces séquences. Or on observe une liaison plus importante pour LFY<sub>TERE</sub> que pour LFY. On peut alors imaginer un mécanisme complexe où un premier monomère se fixerait sur l'ADN et l'arrivée d'un second monomère défavoriserait la liaison du premier monomère. Ce mécanisme impliquerait que la dimérisation via le domaine SAM agisse comme un « contrôle qualité » sur la liaison à l'ADN du domaine C-terminal. Des changements de conformation induits par les différentes interactions (protéine-protéine et/ou ADN-protéine) seraient nécessaires pour permettre ce mécanisme.

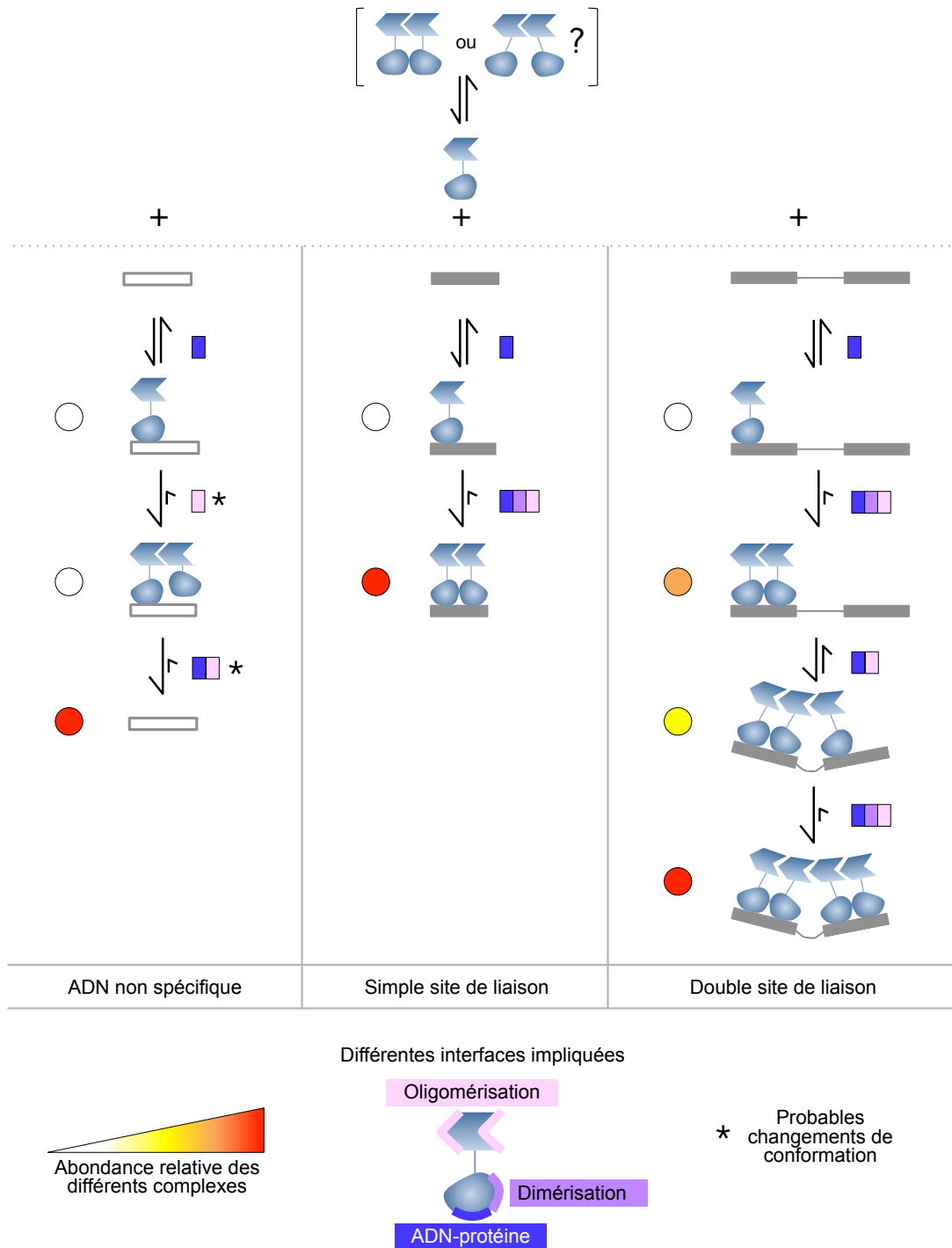
La question du mécanisme de liaison de LFY sur l'ADN n'est donc pas entièrement résolue. Le moyen le plus direct d'y répondre serait de pouvoir mesurer l'état d'oligomérisation de la protéine sans ADN et à faible concentration en sel, mais nous n'y sommes pas parvenu.

Nous proposons un modèle permettant d'expliquer nos données (**Figure 1-16**) :

LFY serait au moins en partie sous forme monomérique en solution, se dimériserait sur un site de liaison grâce au domaine C-terminal, via l'interface de dimérisation et l'interface ADN-protéine et grâce au domaine SAM. La contribution des deux domaines (trois interfaces) pour la formation de ce dimère expliquerait que le complexe monomérique soit quasiment invisible.

Lorsqu'un premier dimère est fixé sur un site de liaison et qu'un second site de liaison se trouve à proximité, alors un nouveau monomère pourrait se fixer sur ce second site formant alors un complexe trimérique. Seuls le domaine SAM et l'interface ADN-protéine contribueraient à la formation de ce trimère (l'interface de dimérisation du domaine C-terminal ne participe pas). Un quatrième monomère viendrait se lier sur le dernier demi-site disponible, avec la contribution des deux domaines N et C-terminal (les trois interfaces).

Un changement de conformation se produirait au sein d'un dimère, défavorisant la liaison à l'ADN dans le cas où le site de liaison ne serait pas satisfaisant (non spécifique), assurant ainsi la sélectivité de la liaison.



**Figure 1-16 | Modèle du mécanisme de liaison de LFY sur l'ADN.**

L'ADN est représenté en gris (rectangle plein : site de liaison pour LFY, rectangle vide : séquence non spécifique). La protéine LFY est en bleu (forme en contact avec l'ADN : domaine de liaison à l'ADN ; forme en flèche : domaine SAM ; trait reliant les deux domaines : *linker*). Le mécanisme de liaison sur trois types d'ADN est schématisé : ADN non spécifique à gauche, un site de liaison au milieu, deux sites de liaison à droite. L'abondance relative de chaque complexe, les interfaces participant à leur formation et l'intervention éventuelle de changements de conformation est représentée par un code couleur légendé sur le bas de la figure. Alternative à ce modèle (non représentée sur ce schéma) : la protéine pourrait se fixer directement sous forme de dimère sur un site de liaison, sans passer par la fixation d'un monomère.

## 7. Oligomérisation et liaison à l'ADN *in planta*, à l'échelle du génome

L'oligomérisation du domaine SAM affecte la liaison à l'ADN de LFY *in vitro*. Pour savoir si elle a aussi un effet *in vivo*, à l'échelle génomique, nous avons comparé les régions liées en ChIP-seq par LFY et LFY<sub>TERE</sub>. Nous avons vu dans l'**Article 1** que LFY<sub>TERE</sub> était globalement moins liée que LFY : les régions couvertes par LFY le sont moins par LFY<sub>TERE</sub> (*read coverage* ou couverture plus faible) et la majorité des régions liées par LFY<sub>TERE</sub> sont mieux liées par LFY.

Comme l'intensité de la réduction de la liaison par LFY<sub>TERE</sub>, comparé à LFY, n'est pas homogène selon les régions, nous avons cherché des règles expliquant que certaines régions aient plus besoin de l'oligomérisation du domaine SAM pour être liées que d'autres.

### ▪ Importance de l'oligomérisation du domaine SAM pour la liaison sur des régions génomiques à sites multiples

Nous avons établi des corrélations entre les sites de liaison présents dans les régions liées et la dépendance de la liaison pour l'oligomérisation du domaine SAM. En effet, nous avons montré que plus le nombre de sites était élevé dans une région (densité des sites), et plus la liaison dépendait du domaine SAM. Cette observation est cohérente avec la coopérativité de liaison entre plusieurs sites, mise en évidence *in vitro*. Un autre paramètre à prendre en compte est le score des sites, qui est proportionnel au logarithme de leur affinité pour la protéine (Moyroud *et al.*, 2011). Plus le score est bon, moins la liaison repose sur la présence d'un domaine SAM fonctionnel. En tenant compte de ces deux paramètres, la densité et le score des sites de liaison, nous avons vu que les régions nécessitant le plus l'oligomérisation avaient tendance à comporter un grand nombre de sites d'affinité moyenne. Afin d'évaluer plus spécifiquement l'importance de la coopérativité entre plusieurs sites de liaison, nous pourrions envisager de surexprimer LFY<sub>[TE+RE]</sub>. Nous disposons de graines T1 transformées avec *p35S:LFY<sub>TE</sub>* ou *p35S:LFY<sub>RE</sub>*, nous pourrions donc obtenir des plantes exprimant les deux protéines par croisement. Si LFY<sub>[TE+RE]</sub> se montre capable d'assumer la fonction florale de LFY (phénotype fleurs ectopiques), alors cela suggérerait que la coopérativité entre sites n'est pas une propriété majeure pour LFY. Cependant, en cas de résultat négatif (pas de fleur de rosette), nous ne pourrions pas tirer de conclusion définitive.

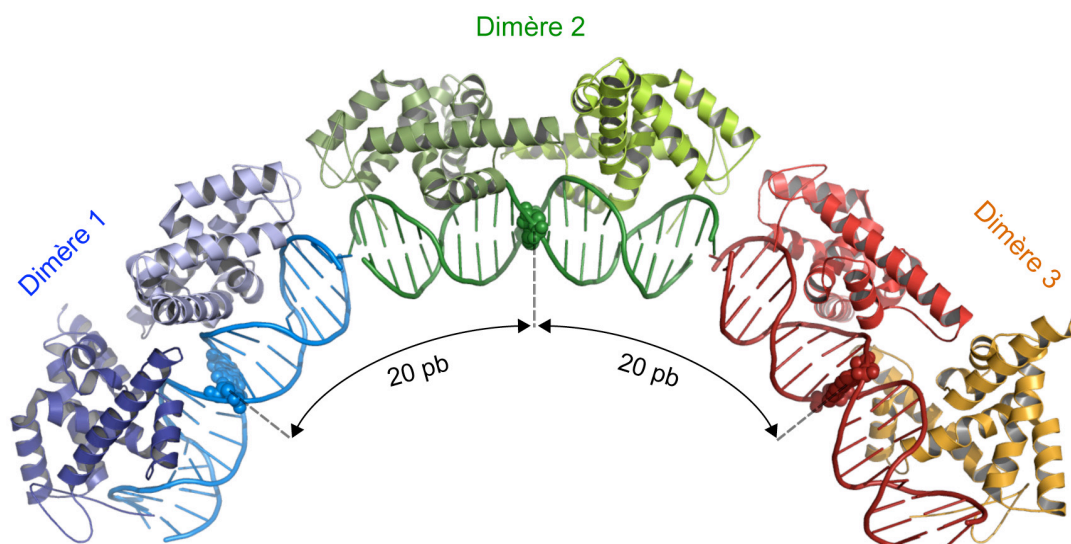
Nous avons recherché s'il existait une distance surreprésentée entre les sites de liaison au sein des régions liées en ChIP-seq. Nous avons trouvé que la distance de 20 pb était spécifiquement surreprésentée dans les régions dont la liaison est la plus dépendante de l'oligomérisation.

Un espacement de 20 pb entre deux sites implique que les dimères liés sur chaque site soient côte à côte, puisqu'un site de liaison fait 19 pb. Cela implique aussi que les sites soient du même côté de la double hélice d'ADN (il faut 10,5 pb pour faire un tour d'hélice).

LFY-C a été cristallisé sous forme de dimère sur un fragment d'ADN cohésif de 19 pb (Hamès *et al.*, 2008). Dans le cristal, les extrémités cohésives de l'ADN s'apparient. Les dimères de LFY-C sont donc espacés de 20 pb (**Figure 1-17**). Bien que cet arrangement soit attribué à des contacts cristallographiques induits par les bouts cohésifs de l'ADN, il représente certainement l'organisation des domaines de liaison à l'ADN sur des sites espacés de 20 pb, tels que ceux mis en évidence grâce aux analyses de CHIP-seq. On peut émettre l'hypothèse que des contacts entre les domaines de liaison à l'ADN, possibles uniquement à cette distance de 20 pb, stabilisent les complexes ADN-protéine.

Cependant, cette organisation du domaine LFY-C n'explique pas que la surreprésentation des sites espacés de 20 pb dans les régions liées en CHIP-seq dépende du domaine SAM. Il est possible que cette distance favorise aussi l'interaction entre les domaines SAM, et augmente ainsi la coopérativité de liaison entre les dimères liés sur chaque site.

Nous chercherons à répondre à cette question, par des expériences d'EMSA sur des fragments d'ADN comportant deux sites espacés d'une distance variable, avec la protéine LFY sauvage et les mutants d'oligomérisation.



**Figure 1-17 | Arrangement du domaine de liaison à l'ADN de LFY dans le cristal.**

Organisation des dimères de AtLFY-C sur un ADN cohésif de 19 bp, dans le cristal (Hamès *et al.*, 2008 ; PDB 2VY1). La paire de base centrale de chaque site de liaison est indiquée par des sphères. On compte 20 pb entre les deux bases centrales de deux molécules d'ADN.

▪ **Importance de l'oligomérisation du domaine SAM pour l'accessibilité à des régions génomiques fermées**

Des expériences de sensibilité à la DNase I réalisées sur les feuilles de plantes sauvages (Col-0) ont été publiées récemment (Zhang *et al.*, 2012). Ces plantes et celles que nous avons utilisées pour le ChIP-seq ont été cultivées dans des conditions similaires (15 jours en jours longs), ce qui permet de pouvoir les comparer. L'accessibilité de la chromatine dans ces tissus nous renseigne sur le contexte auquel est confrontée la protéine LFY dans les échantillons utilisés pour le ChIP-seq. En croisant nos données de ChIP-seq avec ces données nous avons montré que la liaison de LFY sur des régions où la chromatine est fermée (régions résistantes à la digestion par la DNase I) dépendait fortement de l'oligomérisation du domaine SAM.

Cela suggère deux choses : LFY est capable, dans une certaine mesure, de se lier à des régions fermées ; cette liaison est permise par l'oligomérisation du domaine SAM.

Il a été montré que LFY recrute les remodeleurs chromatinienS SPLAYED (SYD) et BRAHMA (BRM), de la famille des SWI2/SNF2, aux loci *AG* et *AP3* (Wu *et al.*, 2012). Ce processus est nécessaire à l'activation de ces gènes lors de la transition florale. L'activation d'*AG* et *AP3* s'accompagne aussi de changements au niveau des marques portées par les histones : les marques répressives H3K27me3 (triméthylation sur la lysine 27 de l'histone H3) sont remplacées par les marques activatrices H3K4me3 (Wu *et al.*, 2012 ; Carles et Fletcher, 2009). La présence de remodeleurs et la dynamique des marques épigénétiques sont le signe d'un changement du niveau de compaction de la chromatine. Le recrutement de remodeleurs chromatinienS par LFY a été mis en évidence sur les régions régulatrices d'*AG* et *AP3*, mais il pourrait s'agir d'un mécanisme plus général pour l'activation des gènes par LFY. La liaison de LFY au niveau de régions fermées pourrait être l'étape qui précède le recrutement des remodeleurs.

Il serait intéressant de vérifier que LFY est bien capable de lier l'ADN dans le contexte de la chromatine, par des tests de liaison sur des nucléosomes reconstitués *in vitro*. Ce type de test, en comparant la protéine sauvage et les mutants d'oligomérisation, permettrait également d'évaluer l'importance du domaine SAM pour ce processus.

Ainsi, LFY pourrait avoir des propriétés de facteur pionnier, comme le FT FoxA présenté dans l'introduction générale. Un facteur pionnier est capable de se lier à des régions inaccessibles à la plupart des autres facteurs de transcription et d'établir un environnement plus favorable, soit par le biais de remodeleurs chromatinienS, soit en ayant lui-même la capacité de changer la conformation de la chromatine (Zaret et Carroll, 2011 ; Guertin et Lis, 2012).

La structure du complexe LFY-C – ADN nous montre que le dimère de LFY-C occupe seulement un côté de la double hélice (**Figure 8**). La liaison de plusieurs dimères sur des sites de liaison espacés de 20 pb leur permet également de rester du même côté de l'ADN (**Figure 1-17**). Ce type de liaison d'un seul côté pourrait permettre à LFY de s'installer sur l'ADN lorsqu'il est enroulé autour du nucléosome et que les histones occupent l'autre côté de la double hélice.



## IV. Bilan et perspectives

### 1. Bilan

Dans ce chapitre, une caractérisation structurale et fonctionnelle du domaine N-terminal de LFY a été présentée.

Nous ne disposions d'aucun d'élément sur la nature et le rôle de ce domaine lorsque nous avons commencé. Nous avons dessiné plusieurs constructions avec des homologues de LFY de plusieurs espèces pour pouvoir exprimer et purifier le domaine N-terminal et analyser sa structure et ses propriétés biochimiques. Nous avons obtenu la structure cristallographique du domaine N-terminal de la protéine LFY du Ginkgo et, de manière inattendue, nous avons découvert qu'il s'agissait d'un domaine SAM, le premier caractérisé chez les plantes. Dans le cristal, les molécules font des interactions tête-queue pour former un polymère en forme d'hélice. Cette organisation surprenante a été décrite pour d'autres domaines SAM, mais LFY-SAM présente des caractéristiques originales, notamment par les dimensions de son polymère dans le cristal.

Afin de valider les interactions observées dans le cristal, nous avons dessiné des mutations sur chacune des deux surfaces formant l'interface d'oligomérisation. L'analyse de la masse moléculaire en solution du domaine SAM, sauvage et muté, a montré que ce domaine formait un oligomère de taille limitée et variable en fonction de la concentration en protéine, et que l'interface observée dans le cristal était bien celle utilisée en solution. Les mutations permettent de contrôler le degré d'oligomérisation du domaine SAM et ont donc été un outil majeur pour comprendre le rôle de cette propriété.

Les domaines SAM de nombreuses protéines, en particulier des régulateurs transcriptionnels, ont été caractérisés. Il apparaît que ce domaine d'interaction à des fonctions très variées selon les FTs. Dans certains cas, comme pour LFY, il permet l'oligomérisation, il est aussi très souvent impliqué dans des interactions avec d'autres domaines SAM ou avec des protéines diverses. Ces interactions peuvent avoir de nombreuses conséquences, telles que des modifications post-traductionnelles, des changements de localisation sub-cellulaire, la modulation de la capacité à réguler la transcription ou le contrôle du degré d'oligomérisation pour les domaines SAM formant des multimères. L'oligomérisation joue un rôle pour la liaison à l'ADN chez plusieurs FTs, permettant notamment une liaison coopérative sur plusieurs sites de liaison.

Nous avons montré, par complémentation fonctionnelle et surexpression, avec la protéine LFY sauvage ou comportant les mutations empêchant son domaine SAM de s'assembler, que l'oligomérisation était nécessaire au fonctionnement de la protéine en tant que régulateur floral chez *A. thaliana*.

Nous avons ensuite cherché à comprendre qu'elle était sa fonction au niveau moléculaire, en particulier pour la liaison à l'ADN. *In vitro*, nous avons montré qu'il permettait à la protéine de lier plusieurs sites de liaison de manière coopérative et d'empêcher la liaison sur des séquences aspécifiques. Ces analyses nous ont permis de proposer un modèle du mécanisme de liaison de LFY sur l'ADN, intégrant les données apportées par l'étude des deux domaines conservés, le domaine de liaison à l'ADN et le domaine SAM.

Le rôle de l'oligomérisation pour la liaison sur le génome *in planta* a été abordé par CHIP-seq sur des plantules exprimant la protéine LFY sauvage ou mutée de manière constitutive. Nous avons vu que cette propriété était nécessaire à LFY pour lier les régions régulatrices de ses gènes cibles, en particulier celles ayant des sites multiples. Ce résultat est cohérent avec l'idée que le domaine SAM permet une coopérativité entre les sites de liaison. De plus, l'oligomérisation du domaine SAM permet à LFY d'accéder à des régions où la conformation de la chromatine est normalement défavorable à la liaison. Cette observation, combinée à des données publiées dans la littérature, suggère que LFY pourrait se lier à des promoteurs inactifs où la chromatine est fermée, recruter des remodeleurs qui ouvriraient ces régions, et permettre ainsi l'accès à d'autres facteurs et l'activation du gène.

Par diverses méthodes combinées dans une approche intégrée, nous avons donc obtenu des résultats permettant de mieux comprendre le fonctionnement de LFY, l'un des régulateurs majeurs du développement floral chez les angiospermes. Il reste cependant plusieurs aspects incompris qui mériteraient d'être éclaircis et de nombreuses perspectives à explorer.

## 2. Perspectives

Par la suite, nous aimerions améliorer la compréhension du fonctionnement moléculaire de la protéine LFY et notamment savoir comment s'organisent les domaines C et N-terminal l'un par rapport à l'autre. Pour cela, nous avons déjà commencé à mettre en place des approches de SAXS, de cristallographie et de microscopie électronique. Même si des optimisations des conditions d'expérimentation seront nécessaires pour que ces méthodes aboutissent, nous pensons pouvoir obtenir des résultats intéressants.

Comme nous l'avons vu, les domaines SAM sont des domaines d'interaction et il est tentant de penser que mis à part l'oligomérisation, d'autres interactions sont possibles. Des partenaires de LFY ont été identifiés génétiquement, et l'interaction physique a été décrite pour certains d'entre eux. Cependant, leur analyse *in vitro* a souvent été limitée par la

difficulté à manipuler la protéine LFY entière. Le domaine SAM isolé, facilement manipulable, pourrait être une alternative pour détecter et étudier des interactions, bien que des validations avec la protéine entière seraient nécessaires par la suite.

Nos résultats attribuent un rôle au domaine SAM pour la liaison de LFY sur plusieurs sites de liaison sur l'ADN. Nous avons étudié ce mécanisme sur des sites adjacents se trouvant sur une même molécule d'ADN. Il est envisageable que des interactions aient lieu entre des sites de liaison situés sur des régions très éloignées sur le génome. Cela entraînerait des phénomènes de *DNA looping*, c'est à dire la formation de boucles via contact entre des régions génomiques distantes, intra ou inter-chromosomique. On commence à mesurer l'importance de ce phénomène pour la régulation transcriptionnelle. En effet, ces boucles peuvent permettre la création « d'îlots » à forte activité transcriptionnelle, ou au contraire, d'isoler des régions réprimées (Hampsey *et al.*, 2011 ; Krivega et Dean, 2012). Nous pensons mettre en place différentes approches pour savoir si LFY est impliqué dans ce type de mécanisme. La microscopie à force atomique (AFM) permettrait de visualiser la formation de boucles sur de l'ADN isolé (Wang *et al.*, 2013 ; Neaves *et al.*, 2009 ; Gietl et Grohmann, 2013). À l'échelle du génome, les interactions longues distances sont détectables par les méthodes de capture de conformation de la chromatine (*Chromosome Conformation Capture* ou 3C). Elles consistent à isoler les complexes ADN-protéine de la même manière que pour un ChIP-seq et à faire une étape de ligation (ou fixation chimique) avant le séquençage afin de lier de manière covalente les régions qui étaient maintenues à proximité par la protéine (Yadon et Tsukiyama, 2013 ; de Wit et de Laat, 2012).

Il est aussi possible que la liaison de LFY induise localement des torsions de la molécule d'ADN, comme le fait la protéine PRH (Williams *et al.*, 2008). Là aussi des méthodes de visualisation directe (microscopie) seraient envisageables pour le vérifier, ainsi que des méthodes chimiques (traitement au permanganate de potassium, agent chimique permettant de détecter les torsions de l'ADN) ou des techniques basées sur l'émission de fluorescence en fonction de la distance entre deux points (comme le FRET, pour Förster Resonance Energy Transfer).

D'après nos analyses, le domaine SAM permet à LFY d'accéder à des régions génomiques où la chromatine est fermée. Nous aimerions comprendre par quels mécanismes LFY accède à ces régions et savoir si cette liaison aboutie à une régulation des gènes. Dans un premier temps, il serait bien de vérifier cette capacité à lier des régions fermées par des tests de liaison *in vitro* sur de la chromatine reconstituée, comme cela a été fait pour FoxA (Cirillo *et al.*, 1998). Nous aimerions aussi comprendre comment LFY recrute des remodeleurs

chromatiniens. L'analyse de l'expression des gènes situés à proximité de ces régions fermées nous indiquerait si la liaison de LFY est associée à une régulation.

Plus généralement, nous avons vu que l'oligomérisation était importante pour la liaison sur le génome, mais nous n'avons pas vraiment abordé le lien entre oligomérisation et régulation. Comme la protéine mutée, affectée pour l'oligomérisation, est quand même capable de se lier sur quelques régions, il serait intéressant de savoir si les gènes concernés sont régulés. Cela nous permettrait de savoir si le domaine SAM a un rôle au-delà de la liaison à l'ADN.

Enfin, nous pourrions intégrer les propriétés de LFY découvertes lors de cette étude au modèle prédictif des régions liées par LFY sur le génome déjà élaboré dans l'équipe (Moyroud *et al.*, 2011). Cela pourrait permettre d'améliorer le pouvoir prédictif du modèle. Un modèle qui prédit précisément les interactions sur le génome peut servir à détecter les gènes potentiellement régulés par LFY chez des espèces non-modèles, sans avoir de données fonctionnelles (Moyroud *et al.*, 2011). On pourrait alors analyser le génome de plantes issues de différents groupes évolutifs, lorsque leur génome sera disponible, afin de connaître l'évolution du réseau de gènes contrôlé par LFY.

# CHAPITRE II

Évolution du mode de liaison  
à l'ADN de LEAFY



# I. Introduction

## 1. Régulation et évolution

D'où vient la grande diversité des êtres vivants ? Quelles sont les bases génétiques et moléculaires de cette diversité et de l'apparition de nouvelles morphologies ?

Depuis que les génomes sont séquencés massivement, il devient évident que le nombre de gènes et la divergence des séquences codantes sont insuffisants pour expliquer la complexité d'un organisme et la variabilité morphologique entre les différentes espèces. On peut d'ailleurs tirer la même conclusion de l'observation des insectes et de certains vertébrés chez qui une métamorphose s'opère : au sein d'une même espèce, deux morphologies bien distinctes, comme celles du têtard et de la grenouille, sont issues d'un même génome.

Il est maintenant admis que les différences phénotypiques et les innovations morphologiques sont dues à des changements dans les réseaux de régulation des gènes du développement. Des plans d'organisation morphologiques différents sont très souvent issus d'un même répertoire de gènes dont la fonction reste constante au cours de l'évolution, mais dont le patron d'expression varie (Carroll, 2000 ; Wagner et Lynch, 2008). C'est le cas chez les animaux où la zone et le moment où s'expriment les gènes *Hox* (codant pour des FTs) le long de l'axe antéro-postérieur au cours du développement déterminent la longueur du thorax et influencent l'émergence des organes latéraux. Par exemple, chez le python, certains gènes *Hox* (*Hoxc8* et *Hoxc6*) s'expriment dans une zone plus étendue que chez d'autres vertébrés comme la poule. Cela explique les centaines de vertèbres et l'absence de membres chez le serpent, comparé à la poule qui n'a que quelques vertèbres et possède des organes latéraux (Cohn et Tickle, 1999).

## 2. Modifications en *cis* ou en *trans* ?

Des changements dans le patron d'expression des gènes peuvent résulter de changements soit dans les éléments *cis*, soit dans les facteurs *trans*, en particulier les FTs. L'évolution est-elle majoritairement dirigée par des modifications dans les éléments *cis*, ou dans les FTs ? Cette question est actuellement vivement débattue par des scientifiques affirmant que les éléments *cis* sont plus à même de contribuer à l'évolution, et d'autres clamant que cette affirmation est trop spéculative et que les facteurs *trans* sont tout aussi aptes à participer à ces processus (Pennisi, 2008).

Plusieurs arguments et de nombreux exemples mettent en avant la contribution des éléments *cis*, plutôt que celle des FTs. En effet, des mutations touchant un FT contrôlant un grand nombre de gènes au cours du développement pourraient avoir des conséquences désastreuses sur l'organisme, alors que des modifications dans une séquence régulatrice contrôlant l'expression d'un gène changeraient seulement l'expression du gène en question. De plus, on constate une variabilité plus importante dans les séquences non codantes que dans les gènes (Carroll, 2008 ; Carroll, 2000 ; Wray, 2007 ; Pennisi, 2008).

La pigmentation chez les insectes est une bonne illustration de la variabilité des éléments *cis*. Chez certaines espèces de drosophiles, les mâles présentent des tâches noires au bout des ailes, alors qu'ils en sont dépourvus chez d'autres espèces. Ces tâches plus foncées sont dues à l'activité de l'enzyme Yellow qui transforme des précurseurs en pigments. Chez les mouches non tachetées, le gène *yellow* est uniformément et faiblement exprimé dans toute l'aile, lui donnant un aspect gris clair. Les espèces ayant la tâche noire ont acquis de nouveaux sites de fixation pour des FTs, dans la région régulatrice du gène *yellow* (Prud'homme *et al.*, 2007). La présence de ces FTs uniquement au bout des ailes explique la très forte expression du gène *yellow* dans cette zone et donc la présence de la pigmentation noire.

L'exemple de la domestication du maïs illustre également le lien entre des différences morphologiques et la variabilité des éléments *cis*. Le locus du gène *teosinte branched1* (*tb1*), qui code pour un régulateur transcriptionnel, est à l'origine de la différence entre l'aspect buissonnant de la téosinte, ancêtre sauvage du maïs, et la tige unique du maïs cultivé (Doebley *et al.*, 2006). En effet, la plus forte expression de *tb1* chez le maïs, comparé à la téosinte, contribue à la forte dominance apicale observée chez le maïs, en réprimant la croissance de branches latérales (Studer *et al.*, 2011). Il n'y a pas de différence majeure dans la séquence de la protéine TB1 entre le maïs et la téosinte, mais il a été montré qu'un élément transposable inséré dans une séquence régulatrice de *tb1* avait modifié le patron d'expression du gène chez le maïs domestique (Studer *et al.*, 2011).

Plusieurs évidences tendent à montrer que les FTs sont également modulables au sein de familles multigéniques et contribueraient à l'évolution au même titre que les éléments *cis* (Wagner et Lynch, 2008 ; Hsia et McGinnis, 2003 ; Hoekstra et Coyne, 2007 ; Pennisi, 2008). L'un des arguments le plus important en faveur de la possible plasticité des FTs est que l'effet néfaste d'un changement dans un régulateur gouvernant un processus biologique essentiel peut être contré par une duplication préalable du gène. En effet, une copie pourrait assurer la fonction initiale tandis que l'autre serait libre d'être modifiée et d'acquiescer ou de perdre certaines propriétés.



Ce processus a probablement eu lieu au sein du réseau de régulation qui contrôle la formation des os : un FT, dont la fonction ancestrale était vraisemblablement la formation du cartilage, aurait été dupliqué et l'une des copies (*Sox9*) aurait conservé la fonction initiale tandis que l'autre (*Runx2*) se serait retrouvée à la tête du réseau contrôlant le développement des os (Fisher et Franz-Odenaal, 2012).

Chez les plantes, la famille des gènes MADS, qui compte une centaine de représentants chez *Arabidopsis*, se serait formée à partir de deux gènes présents chez le dernier ancêtre commun entre les végétaux et les animaux. Les multiples copies ont été générées par des duplications à l'échelle d'un seul gène ou du génome entier (Airoidi et Davies, 2012). Leurs fonctions se sont diversifiées au cours de l'évolution et ces FTs sont impliqués dans de très nombreux processus du développement chez *Arabidopsis* (Smaczniak *et al.*, 2012a).

On note que l'étude du rôle clé joué par les processus associés au développement dans l'évolution est principalement focalisée sur les éléments *cis* et les FTs. Cependant, d'autres pistes n'ont été que peu explorées. Parmi elles, on peut citer l'influence des ARNs non codants, déjà évoqués comme des acteurs potentiels dans l'évolution des réseaux de régulation (Zhang et Su, 2009) ; la contribution de l'épissage alternatif, qui peut conduire à des protéines ayant différentes activités à partir d'un gène unique ; ou encore la participation des éléments contrôlant la stabilité des ARNm et leur degré de traduction (Hoekstra et Coyne, 2007).

### **3. LFY, un bon candidat pour étudier l'évolution des facteurs de transcription**

Les FTs peuvent donc évoluer au sein de familles multigéniques. Mais leur évolution est-elle possible sans duplication ?

Comme évoqué précédemment, le changement de certaines propriétés d'un FT sans qu'il y ait de « copie de secours » conduirait à la perte de régulation d'un grand nombre de gènes et pourrait avoir des conséquences fonctionnelles néfastes. Les exemples de FTs ayant évolué sans se dupliquer sont très peu nombreux dans la littérature, ce phénomène semble donc assez rare.

Au moins deux cas ont cependant été décrits. Ultrabithorax (*Ubx*), un FT codé par un gène *Hox*, comporte une mutation dans le domaine de liaison à l'ADN chez le crustacé Artémia, par rapport à la protéine présente chez la drosophile. Cette mutation aurait modifié la

capacité de *Ubx* à réprimer l'émergence des membres dans certains segments abdominaux, ce qui expliquerait que les crustacés possèdent de nombreuses pattes alors que les insectes en ont seulement six (Ronshaugen *et al.*, 2002). Il n'y a pas de redondance connue entre *Ubx* et d'autres gènes homéotiques. Un autre exemple est le FT *Mata1* chez la levure *S. cerevisiae*, qui régule un groupe de gènes impliqué dans la différenciation du type cellulaire. Son orthologue chez la levure *Candida albicans* ne reconnaît pas les mêmes séquences ADN, et cette divergence a apparemment eut lieu sans duplication du gène (Baker *et al.*, 2011). Il est donc possible d'avoir une modification des propriétés d'un FT sans duplication.

LFY est présent en une seule copie dans la plupart des génomes et dans les rares cas où plusieurs copies existent, en lien avec des événements de polyploïdisation du génome, elles ne semblent pas avoir divergé par néo-fonctionnalisation. La protéine est très conservée, en particulier le domaine de liaison à l'ADN et les résidus impliqués dans les contacts ADN-protéine. LFY a un rôle essentiel chez les angiospermes, en tant que régulateur du développement floral, ainsi que chez la mousse *P. patens* lors de la première division cellulaire après fécondation (Tanahashi *et al.*, 2005).

Il a cependant été montré que PpLFY ne reconnaissait pas les séquences ADN liées par AtLFY *in vitro* et n'était pas capable de compléter un mutant *lfy* chez *A. thaliana* (Maizel *et al.*, 2005). Ces observations laissent penser que soit PpLFY ne lie pas l'ADN, soit sa spécificité est différente. Basé sur une comparaison de séquence, il a été proposé que deux acides aminés appartenant au domaine de liaison à l'ADN (D312 et C345) soient à l'origine de la différence de comportement de PpLFY par rapport à ses homologues chez les autres espèces (**Figure 8**). La protéine PpLFY-D312H, où l'aspartate est remplacé par une histidine, comme dans la séquence de LFY d'*Arabidopsis*, permet la complémentarité du mutant *lfy* et la liaison sur l'ADN (Maizel *et al.*, 2005). D'une manière intéressante, ces résidus ne sont pas impliqués dans des interactions avec l'ADN chez la protéine d'*Arabidopsis*, d'après la structure 3D de LFY-C (Hamès *et al.*, 2008).

Le cas de *P. patens* suggère donc que les propriétés de liaison à l'ADN de LFY aient subi des modifications au cours de l'évolution, malgré sa très forte conservation, son rôle majeur et le fait que le gène ne fasse pas partie d'une famille multigénique.

Plusieurs questions se posent alors : comment expliquer, à l'échelle moléculaire, un changement de spécificité au sein d'un domaine de liaison à l'ADN aussi conservé ? Comment un tel changement a-t-il pu être toléré alors qu'il modifierait le répertoire complet des gènes cibles de LFY ?

LFY nous semble donc être un très bon modèle pour étudier l'évolution des FTs qui ne font pas partie d'une famille multigénique, et tenter de comprendre comment des transitions évolutives peuvent avoir lieu.

Nous avons choisi d'étudier l'évolution de la spécificité de liaison à l'ADN de LFY en déterminant cette propriété par des expériences de SELEX chez différentes espèces représentatives des grands groupes évolutifs. Nous avons utilisé une approche structurale pour expliquer ces changements de spécificité au niveau moléculaire. Cette étude est présentée sous forme d'un article, l'**Article 2**, actuellement en cours de soumission et se trouvant ci-après.

Là encore, de nombreuses personnes ont contribué à ce travail. Nous sommes d'ailleurs quatre co-premiers auteurs à figurer sur l'article : Marie Monniaux, également doctorante dans l'équipe, qui a notamment fait la plupart des expériences de SELEX et a participé aux clonages et analyses biochimiques, Max Nanao qui s'est occupé de la collecte des données de diffraction et de la résolution de la structure cristallographique et Edwige Moyroud qui avait initié le projet lorsqu'elle était doctorante dans l'équipe. Emmanuel Thévenon a ici encore réalisé de nombreux clonages. Hicham Chahtane, également doctorant dans l'équipe a participé à certaines expériences de SELEX. Nous avons collaboré avec l'équipe de Detlef Weigel du Max Planck Institute for Developmental Biology de Tübingen pour le séquençage de nos échantillons de SELEX. Des personnes impliquées dans le projet 1KP, visant à élaborer une base de données transcriptomiques à partir de 1000 espèces végétales, figurent également dans les auteurs, puisqu'ils nous ont permis d'enrichir notre répertoire de séquences de LFY.

## II. Article 2

### A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity

#### AUTHORS

Camille Sayou<sup>‡,1,2,3,4</sup>, Marie Monniaux<sup>‡,1,2,3,4</sup>, Max H. Nanao<sup>‡,5,6,\*</sup>, Edwige Moyroud<sup>‡,1,2,3,4,a</sup>, Emmanuel Thévenon<sup>1,2,3,4</sup>, Hicham Chahtane<sup>1,2,3,4</sup>, Samuel F. Brockington<sup>7</sup>, Norman Warthmann<sup>8, c</sup>, Michael Melkonian<sup>9</sup>, Yong Zhang<sup>10</sup>, Gane Ka-Shu Wong<sup>10, 11</sup>, Detlef Weigel<sup>8</sup>, Francois Parcy<sup>1,2,3,4,b,\*</sup>, Renaud Dumas<sup>1,2,3,4</sup>

<sup>1</sup> CNRS, Laboratoire de Physiologie Cellulaire & Végétale, UMR 5168, 38054 Grenoble, France.

<sup>2</sup> Univ. Grenoble Alpes, LPCV, F-38054 Grenoble, France.

<sup>3</sup> CEA, DSV, iRTSV, LPCV, F-38054 Grenoble, France.

<sup>4</sup> INRA, LPCV, F-38054 Grenoble, France.

<sup>5</sup> European Molecular Biology Laboratory, 6 rue Jules Horowitz, BP 181, 38042 Grenoble, France.

<sup>6</sup> Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France.

<sup>7</sup> Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

<sup>8</sup> Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

<sup>9</sup> Botanisches Institut, Lehrstuhl I, Universität zu Köln, Biozentrum Köln, Zùlpicher Str. 47b, 50674 Köln, Germany

<sup>10</sup> BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

<sup>11</sup> Department of Biological Sciences, Department of Medicine, University of Alberta, Edmonton AB, T6G 2E9, Canada

<sup>a</sup> Present address: Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

<sup>b</sup> Present address: Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, V5Z 4H4, Canada.

<sup>c</sup> Present address: Research School of Biology, The Australian National University, Acton, ACT 0200

<sup>‡</sup> These authors made equal contributions

<sup>\*</sup> Correspondence to: François Parcy (francois.parcy@cea.fr) or Max Nanao (mnanao@embl.fr)

**ABSTRACT**

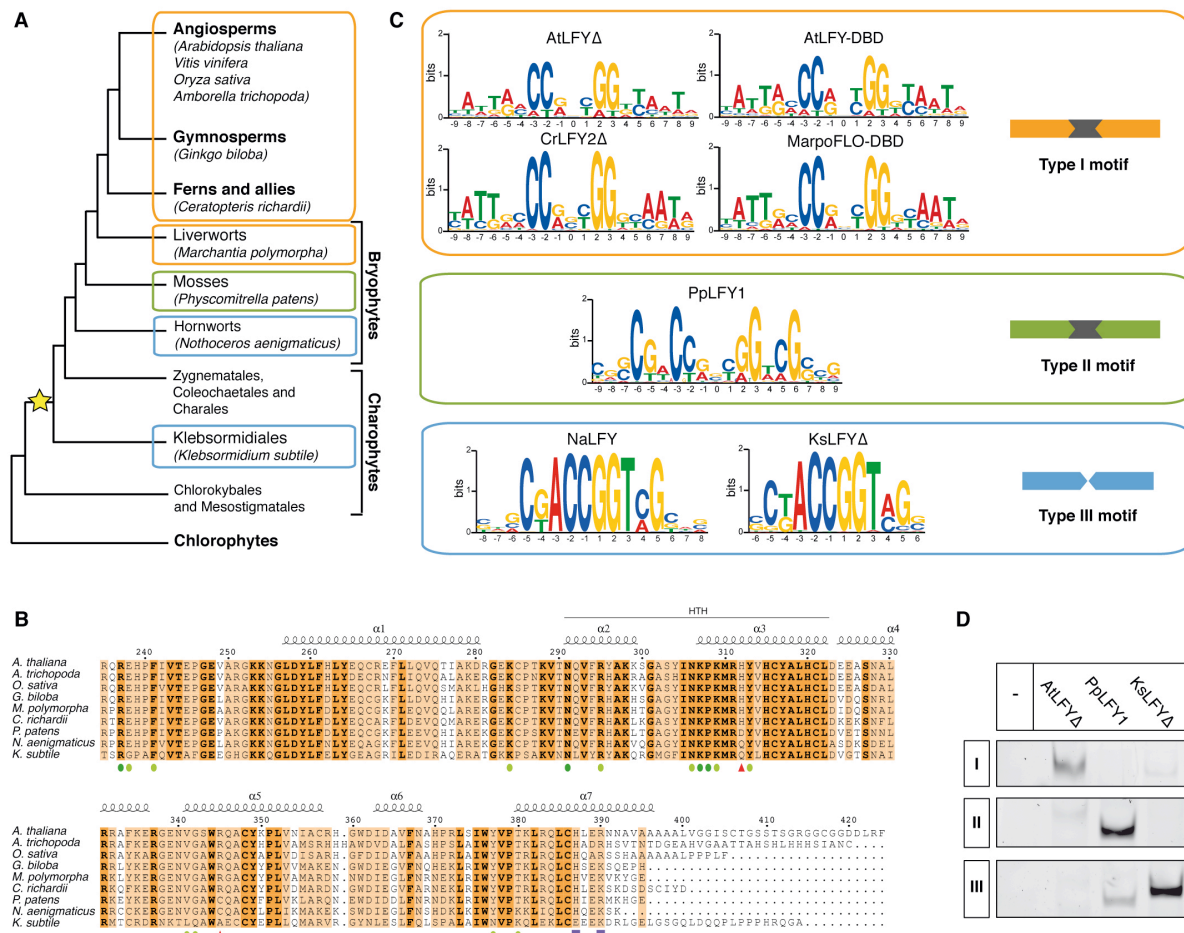
Transcription factors are life architects at the molecular scale and changes affecting their function can yield novel forms that contribute to Nature's diversity. However, such changes can also have deleterious effects. Consequently, transcription factors mostly evolve after gene duplication events that release one gene copy from selective pressure. Here we show that a major regulator involved in flower development and cell division in plants, LEAFY, successfully changed its way of binding DNA despite being a single copy and a highly conserved factor. We explained how these changes occurred at the structural level, and we identified an intermediate LEAFY form in hornworts that can adopt all different specificities and could have smoothed the evolutionary transitions between binding modes.

**One sentence summary**

The single copy, essential and highly conserved LEAFY plant transcription factor changed DNA binding specificity through a promiscuous intermediate.

**MAIN TEXT AND FIGURES**

The rewiring of transcriptional networks provides a rich source of evolutionary novelty (1-3). Whereas variation of *cis*-elements is well documented (3, 4), there is less evidence for changes affecting the coding sequence of transcription factors (TFs). Such changes are expected to be especially delicate, because they could deleteriously affect the expression of many downstream targets. The dangers inherent in changing DNA binding specificity can be mitigated by duplication of TF genes, following which additional gene copies can acquire new functions (5). Indeed TF DNA binding specificity has been shown to diversify within multigene families (6). However some TFs are maintained as single copy genes, and in this scenario the question remains as to how evolution of essential TFs proceeds without gene duplication. To address this paradox, we used the *LEAFY* (*LFY*) gene, which is present in all terrestrial plant genomes, mostly as a single copy (7). *LFY* plays an essential role as the master regulator for floral identity in angiosperms and has a different key function in the moss *Physcomitrella patens* (8), controlling cell division. *LFY* encodes a transcription factor, which binds DNA through a highly conserved dimeric DNA binding domain (DBD) (9). Despite this conservation, a previous report (7) showed that PpLFY1, the *LFY* homolog from *P. patens*, is unable to bind the DNA sequence recognized by *LFY* from *Arabidopsis thaliana* (AtLFY). Whether PpLFY1 binds DNA at all or whether PpLFY1 and AtLFY differ in their DNA binding specificities has remained elusive.



**Fig. 1. Evolution of LFY DNA binding specificity.**

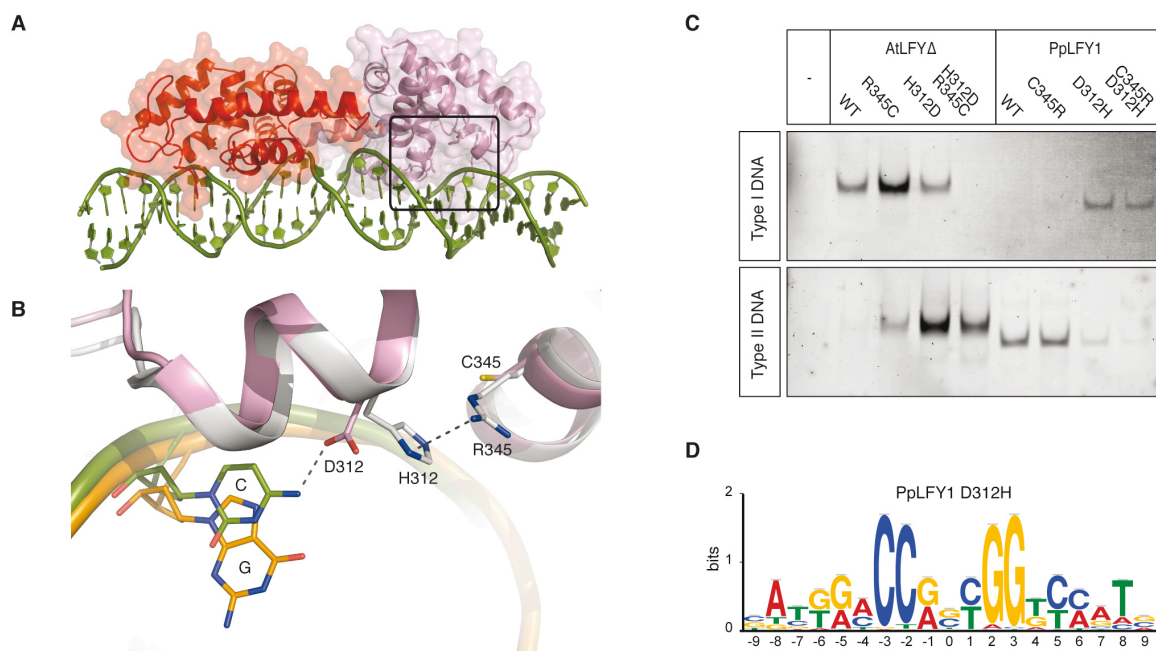
Green plants phylogeny, adapted from (15-17). The yellow star marks the estimated origin of *LFY*. Species names are indicated for *LFY* proteins on which SELEX was performed. Coloured rectangles indicate DNA binding specificity of type I (orange), II (green) or III (blue). **(B)** Alignment of *LFY*-DBDs. Amino acid numbering and secondary structure annotation ( $\alpha$ , alpha helices, HTH, helix-turn-helix domain) is based on *AtLFY*. Dark green dots: DNA base contacts, light green dots: phosphate backbone contacts, red triangles : residues involved in the new DNA contact in *PpLFY1*, purple rectangles : residues involved in the interaction between DBD monomers. **(C)** Examples of SELEX motifs, grouped in three types according to their similarity. *AtLFY* (*Arabidopsis thaliana*), *CrLFY2* (*Ceratopteris richardii*), *MarpoFLO* (*Marchantia polymorpha*), *PpLFY1* (*Physcomitrella patens*), *NaLFY* (*Nothoceros aenigmaticus*), *KsLFY* (*Klebsormidium subtilis*),  $\Delta$ : proteins lacking the first 40 amino acids, -DBD: DNA binding domain only. Cartoons depict binding site organization: with a 3-bp spacer in type I and II motifs, or without in type III motif. **(D)** EMSA with *AtLFY* $\Delta$ , *PpLFY1* and *KsLFY* $\Delta$  proteins on DNA probes carrying one motif type. All proteins are at 10 nM and only the protein-DNA complexes are shown.

To study *LFY* evolution in early terrestrial plants, we searched for *LFY* homologs in recent large-scale transcriptomic data (<http://www.onekp.com/>) from aquatic species whose origin predates the divergence between mosses and tracheophytes. We found *LFY* homologs in several derived streptophyte green algae (**Fig. 1A, fig. S1**). Thus, *LFY* is not specific to land plants as proposed (7) but arose much earlier in the streptophyte green algae. Within this extended family, the *LFY*-DBD sequence remains highly conserved, most notably the amino

acids in direct contact with DNA (**Fig. 1B, fig. S1**). We used HT-SELEX (10) experiments to systematically analyse the DNA binding specificity of recombinant LFY proteins from each plant group (**Fig. 1A**). After optimising the alignments (11), we found that the SELEX motifs fell into three groups (**Fig. 1C, fig. S2**), indicating that LFY changed specificity at least twice. Most LFY proteins from land plants (angiosperms, gymnosperms, ferns and liverworts) bind the same motif (type I) as AtLFY (11). This agrees with the fact that all embryophyte LFY homologs, except PpLFY1, display AtLFY-like activity when expressed in *Arabidopsis* (7). PpLFY1, despite possessing identical amino acids in all 15 positions that contact DNA in AtLFY, bound to a different motif (type II). Type I and II motifs share a similar overall organization, consisting of two 8-bp palindromic half-sites separated by 3 nucleotides, but their peripheral positions differ. The newly identified hornwort and algal LFY proteins, however, bound to a third motif (type III) that resembles type II, but without the central 3-bp spacer (**Fig. 1C**). The pseudo symmetry obtained for all motifs (**fig. S2**) suggested that all LFY proteins bound DNA as dimers. With AtLFY, PpLFY1 and KsLFY (from *Klebsormidium subtile*) as representative proteins of type I, II and III specificities, we confirmed that each protein displayed a strong preference for one sequence type (**Fig. 1D**). The detectable, weak binding of PpLFY1 to a type III probe likely reflects the similarity between type II and type III half sites (**Fig. 1D**).

Given the extensive conservation of the LFY-DBD sequence, we wondered how these different specificities could be explained at the molecular level. The critical determinants of the type I binding mode were previously elucidated by the atomic structure of AtLFY-DBD bound to a type I DNA (9). For comparison, we solved the crystal structure of PpLFY1-DBD bound to a type II DNA (**Fig. 2A, Table S1**). Superimposing the two structures revealed a high similarity between PpLFY1- and AtLFY-DBD/DNA complexes (RMSD of protein backbone atoms of 0.6 Å). However, PpLFY1-DBD makes an additional contact with DNA between aspartic acid 312 (D312) and the cytosine base (C) at position -6 of the DNA binding motif, which is the nucleotide most different between AtLFY and PpLFY1 motifs obtained by SELEX (**Fig. 1C, Fig. 2B**). In AtLFY, position 312 is occupied by a histidine residue (H312), which is shifted away from the DNA by an arginine (R345). This “flipped out” conformation of side chain 312 precludes direct binding of this residue to DNA. By contrast, in PpLFY1, a cysteine residue (C345) replaces R345 and does not affect the positioning of D312. To test the importance of these two positions, we analysed the DNA binding of all four possible variants at positions 312 and 345 for PpLFY1 and AtLFY in EMSA (**Fig. 2C**) and in SELEX for PpLFY1-D312H (**Fig. 2D**). Swapping residues at the critical positions 312 and 345 was sufficient to shift specificity from type I to type II and vice-versa, confirming that these two positions are key determinants of LFY specificity (**Fig. 2C-D**). This result is

consistent with the previous observation that a PpLFY1-D312H mutant can bind a type I sequence and partially complement the *lfy* mutation in *A. thaliana* plants (7).



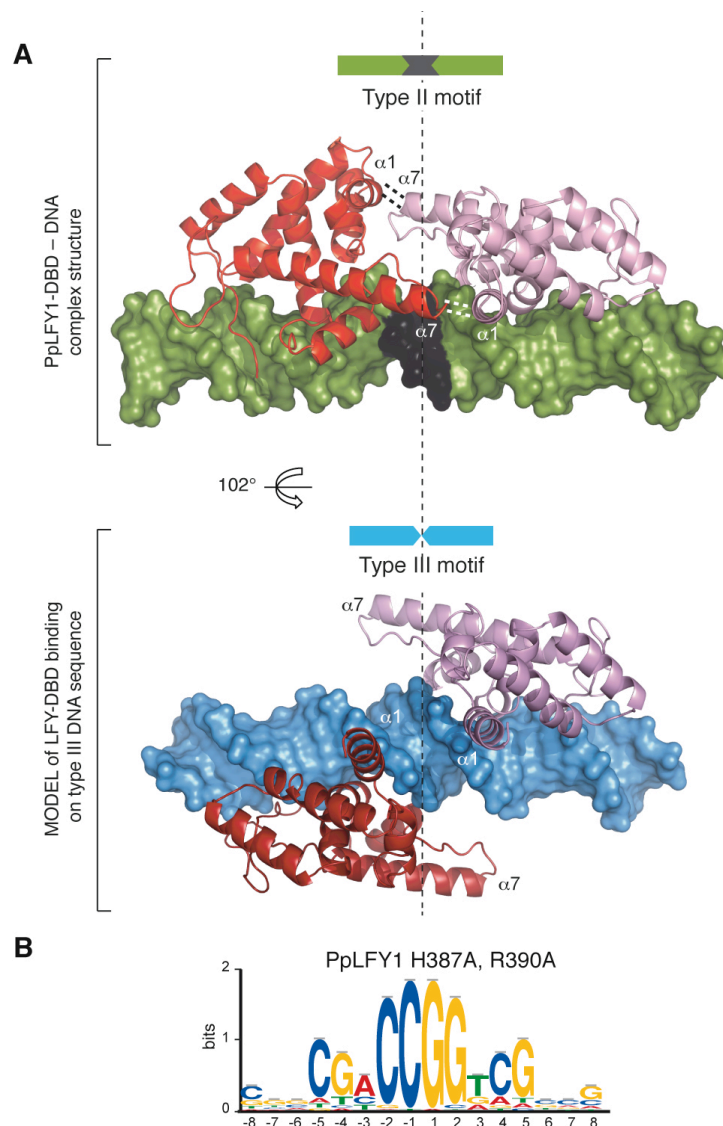
**Fig. 2. Structural basis for type II specificity (PpLFY1-DBD), in comparison with type I (AtLFY-DBD).**

**(A)** Crystal structure of PpLFY1 DNA binding domain in complex with DNA. The two monomers are represented in red and pink, DNA is green. The squared area is detailed in **b** after applying a 70° rotation to improve clarity. **(B)** Superimposition of AtLFY-DBD (grey) - DNA (orange) and PpLFY1-DBD (pink) - DNA (green) complexes. Specificity determinant residues and bases are represented as sticks. For amino acids, H: histidine R: arginine, D: aspartate, C: cysteine. For DNA bases, C: cytosine, G: guanine. **(C)** Effect of specific mutations on the DNA binding specificity of AtLFYΔ and PpLFY1 in EMSA. All proteins are at 25 nM and only the protein-DNA complexes are shown. **(D)** SELEX motif of the PpLFY1-D312H protein, bearing strong resemblance to the type I motif.

We next investigated binding to the type III motif. Type III half-sites are similar to those of type II (**Fig. 1C**). This resemblance can be explained structurally: type III motif proteins (algal and *Nothoceros* LFY) have a glutamine residue at position 312 (Q312), which, like the aspartic acid D312 in PpLFY1, can extend towards and contact DNA without interference from the small residues (cysteine, alanine or serine) at position 345 (**Fig. 1B, fig. S3**). The major difference between type III and type II motifs resides in the absence or presence of the central 3-bp spacer (**Fig. 1C**). Modelling the LFY-DBD binding on a type III motif by removing the 3-bp spacer present in type II DNA sequence (**Fig. 3A**) revealed that the interaction surfaces between helices a1 and a7, which stabilize dimeric AtLFY- and PpLFY1-DBD positioning (9), could not interact with each other in the type III binding mode. Consistent with



this observation, interacting regions of helices  $\alpha 1$  and  $\alpha 7$ , including the key amino acids H387 and R390 on  $\alpha 1$  (9), are highly conserved from bryophytes to angiosperms (type II and I) but are more variable in algae (type III) (**Fig. 1B**, **fig. S1**). To test whether the  $\alpha 1/\alpha 7$  interaction surface plays a key role in binding to 3-bp spaced half sites, we mutated the H387 and R390 residues from PpLFY1. This was sufficient to shift DNA binding preference to type III (**Fig. 3b**). Altogether, these observations strongly suggest that LFY-DBD preferentially binds to 3-bp spaced half sites when the protein has an  $\alpha 1/\alpha 7$  interaction surface.

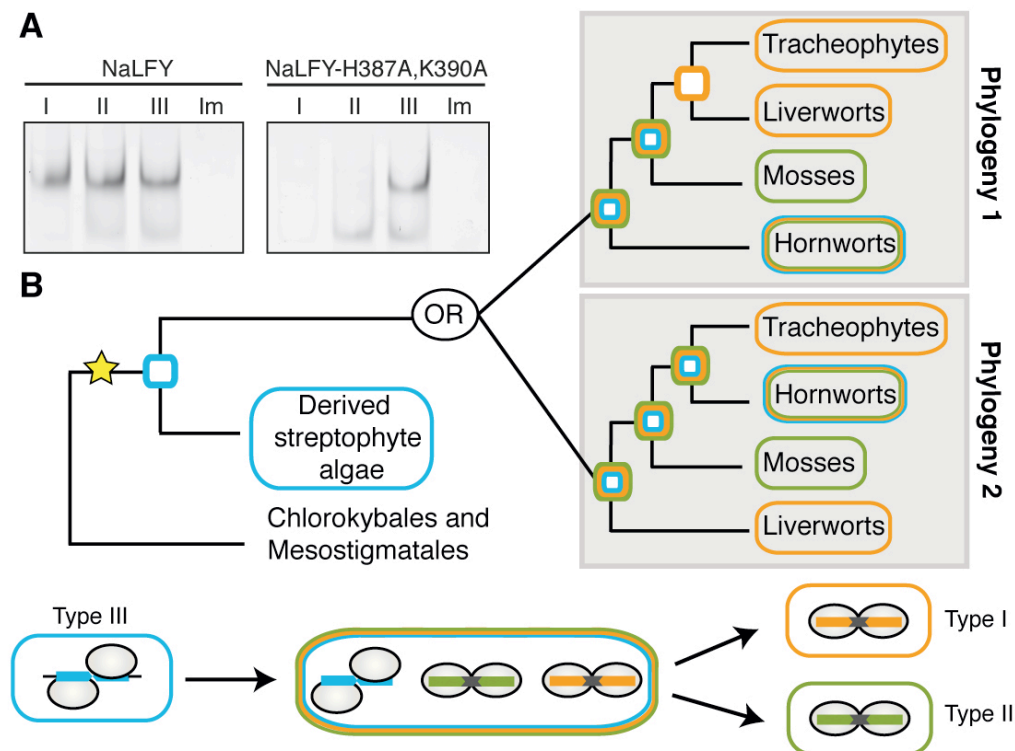


**Fig. 3. Structural model for type III specificity.**

(**A**) Top: Structure of PpLFY1-DBD in contact with DNA, showing the area of contact between monomers (involving alpha helices  $\alpha 1$  and  $\alpha 7$ ) with dashed lines. Monomers are represented in red and pink, DNA is in green except for the 3 central base pairs represented in black. Bottom: Modelled type III binding. DNA is shown in blue. The black arrow symbolizes the rotation of a half-site as compared to the upper structure. (**B**) SELEX motif of the PpLFY1-H387A, R390A protein, showing strong resemblance to the type III motif.

While our analysis provides a molecular basis for DNA specificity changes that occurred during evolution of a single TF, we were left with a critical question: if, as shown in *P. patens* and angiosperms, LFY played a key role throughout plant evolution, how could these changes have been tolerated when they first arose, since they would have drastically and instantly modified the expression of the entire set of LFY target genes? For other types of proteins such as metabolic enzymes or nuclear receptors (12-14), promiscuous intermediates were suggested as essential for the emergence of novel properties. In examining LFY sequences from streptophyte green algae and bryophytes, we realized that NaLFY from the hornwort *N. aenigmaticus* had type III specificity in SELEX despite having the angiosperm-like H387 and K390 interaction residues (**Fig. 1B-C**). We assayed NaLFY DNA binding specificity in EMSA experiments, and found that this protein could bind all three types of motifs and that this increased promiscuity was, as suspected, due to a functional a1/a7 interaction surface (**Fig. 4A**). When we infer the ancestral sequence of key amino-acid residues of LFY-DBD at critical nodes of the plant phylogeny, it appears that such a relaxed LFY form could have mediated the transitions between the different specificities: from an ancestral algal LFY with a type III specificity (Q312 and C345 half-site determinants), the creation of the a1/a7 interaction surface would have generated a more promiscuous LFY intermediate, from which both type I and II specificities would have subsequently evolved through mutations at positions 312 and 345 (**Fig. 4B**). Remarkably, this evolutionary trajectory does not involve any direct shift from type I to type II specificity, a transition that was expected to be deleterious in land plants, where LFY function is critical.

Our work thus demonstrates that a highly conserved and essential single copy TF can undergo radical shifts in DNA binding specificity during evolution. Furthermore, based on a structural characterization of the different modes of DNA binding, we identify a LFY state of increased promiscuity that could have supported a smooth evolutionary path between specificities. While we have focused on the more enigmatic question of evolution in single copy TFs here, it is plausible that the mechanisms we describe could also contribute to the evolution of TFs within multigene families.



**Fig. 4. Proposed evolution of LFY DNA binding specificity in green plants.**

(A) NaLFY interacts with all three types of DNA binding motifs in EMSA, but not with the negative control, a type I mutated probe (Im). Mutating the interacting amino acids H387 and K390 into alanines (NaLFY-H387A,K390A) reduced the binding to type I and type II motifs, but not to type III motif. Both proteins are at 1  $\mu$ M, only the protein-DNA complexes are shown. The similar size of protein-DNA complexes with the three types of sequences indicates that NaLFY likely binds to all sequences as a dimer. (B) The cartoons depict the transitions between the different DNA binding modes adopted by LFY proteins (represented by grey spheres) on the three types of motifs. Two possible phylogenies are shown (15-18). DNA binding specificity in each phylogenetic group is depicted with a color-coded circle (type I in orange, type II in green, type III in blue and relaxed specificity with all three colours). In both phylogenies, a promiscuous intermediate (similar to the current NaLFY protein in hornworts) would mediate specialization to type I or type II binding mode. The yellow star marks the estimated origin of LFY.

## REFERENCES AND NOTES

1. S. B. Carroll, *Cell* **101**, 577 (2000).
2. I. S. Peter, E. H. Davidson, *Cell* **144**, 970 (2011).
3. B. Prud'homme, N. Gompel, S. B. Carroll, *Proc. Natl. Acad. Sci. U.S.A.* **104 Suppl 1**, 8605 (2007).
4. G. A. Wray, *Nature Rev. Genet.* **8**, 206 (2007).
5. H. E. Hoekstra, J. A. Coyne, *Evolution* **61**, 995 (2007).
6. G. Badis *et al.*, *Science* **324**, 1720 (2009).
7. A. Maizel *et al.*, *Science* **308**, 260 (2005).
8. T. Tanahashi, N. Sumikawa, M. Kato, M. Hasebe, *Development* **132**, 1727 (2005).
9. C. Hamès *et al.*, *EMBO J* **27**, 2628 (2008).
10. Y. Zhao, D. Granas, G. D. Stormo, *PLoS Comput. Biol.* **5**, e1000590 (2009).

11. E. Moyroud *et al.*, *Plant Cell* **23**, 1293 (2011).
12. O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **10**, 498 (2006).
13. A. Aharoni *et al.*, *Nat. Genet.* **37**, 73 (2005).
14. G. N. Eick, J. K. Colucci, M. J. Harms, E. A. Ortlund, J. W. Thornton, *PLoS Genet* **8**, e1003072 (2012).
15. C. Finet, R. E. Timme, C. F. Delwiche, F. Marletaz, *Curr. Biol.* **20**, 2217 (2010).
16. D. L. Nickrent, C. L. Parkinson, J. D. Palmer, R. J. Duff, *Mol. Biol. Evol.* **17**, 1885 (2000).
17. S. Wodniok *et al.*, *BMC Evol. Biol.* **11**, 104 (2011).
18. Y. L. Qiu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15511 (2006).

## ACKNOWLEDGEMENTS

We thank M. Schmid for help with sequencing, E. Masson, J. Kyojuka, M. Hasebe for various *LFY* cDNAs, C. Scutt and C. Finet for an *Amborella* library and help with its screening, J. C. Villarreal the sequence of *NaLFY*, A. Maizel and S. Rensing for discussion, A. Maizel, R. Worsley-Hunt, A. Mathelier, M. Blazquez, O. Nilsson, C. Zubieta, and J. Chen for critical reading of the manuscript. This work was supported by funds from the Max Planck Society to D.W., the ANR (Blanc – SVSE2 – 2011) to F.P. and the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n° 227764 (P-CUBE) to R.D. The 1000 Plants (1KP) initiative, led by G.K.S.W., is funded by the Alberta Ministry of Enterprise and Advanced Education, Alberta Innovates Technology Futures (AITF) Innovates Centre of Research Excellence (iCORE), Musea Ventures, and BGI-Shenzhen. Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under the accession number 4BHK. The authors declare no competing financial interests.

## SUPPLEMENTARY MATERIALS

This section, available on line, contains the author contributions, the materials and methods, the references for methods, Figures S1-S3 and Tables S1-S3.

## SUPPLEMENTARY MATERIALS

### Author contributions

M.Me, Y.Z. and G.K.S.W. generated the transcriptome data from algae, S.F.B. and E.M. identified LFY cDNA sequences from algae and *Nothoceros*; C.S., M.Mo., E.M. and E.T. constructed the expression clones; C.S., M.Mo., E.M., H.C. and R.D. purified the proteins and performed the SELEX experiments; N.W., D.W. and M.Mo. sequenced the SELEX samples and analysed the sequences; C.S., M.H.N. and R.D. performed the crystallographic experiments; M.Mo. and C.S. performed the EMSAs; R.D. and M.H.N. performed the modelling; F.P. and R.D. designed the study; F.P. wrote the paper with the help of D.W., M.H.N., R.D., C.S., M.Mo., and E.M.. All authors discussed the results and commented on the manuscript.

### Materials and Methods

#### **LFY cDNAs and cloning.**

cDNAs for *VFL*, *RFL*, *PpLFY1* and *CrLFY2* were provided by various colleagues (see Acknowledgments). The coding sequence of previously isolated *LFY* homologs, *GinLFY* (AF108228) and *MarpoFLO* (AF286056), were amplified by PCR, using specific primers and cDNA from mixed tissues as template. The newly identified *LFY* homolog from *Amborella trichopoda*, *AmboLFY*, was isolated in two steps. A partial sequence of *AmboLFY* was first obtained after screening a cDNA library from female flowers following the  $\lambda$  Zap (Stratagene) protocole. The full-length sequence (KF193872) was then recovered using the Marathon<sup>®</sup> cDNA Amplification kit (Clontech). The sequences of *NaLFY* from *Nothoceros aenigmaticus* (KF269532), *KsLFY* from *Klebsormidium subtile* (KF269535), *CsLFY* from *Coleochaete scutata* (KF269533), *CvLFY* from *Chara vulgaris* (KF269536) and *CyLFY* from *Cylindrocystis sp.* (KF269534) were identified among transcriptome data from the 1KP project (<http://www.onekp.com/project.html>) and the corresponding cDNA sequences were synthesized *in vitro* by GeneArt<sup>®</sup> (Life Technologies). All cDNAs were cloned in the pETM-11(1) or pET30a+ (Novagen) vectors to add a 6xHis tag for purification.

#### **Expression and purification of recombinant proteins.**

Recombinant proteins purification for EMSA and SELEX experiments were essentially performed as described (2) except for *NaLFY* for which 1 M NaCl was added in all purification steps to improve the final yield. We verified that adding NaCl to *NaLFY* purification did not affect its DNA binding properties. Detailed buffers used for purification of

all proteins of this study can be found in Table S2. For crystallographic experiments, PpLFY1-DBD protein was purified as previously described (3).

### **EMSA experiments.**

EMSA were performed as previously described (4) but with a modified binding buffer for protein - DNA interaction (HEPES 10 mM pH 7.5, Spermidine 1 mM, EDTA 14 mM, BSA 0.3 mg/mL, CHAPS 0.25%, glycerol 1%, fish sperm DNA 28 ng/mL (Roche) and TCEP 3 mM). Oligonucleotides used for EMSA are listed in Table S3.

### **Systematic Evolution of Ligands by Exponential Enrichment (SELEX) experiments.**

HT-SELEX experiments and sequencing have been described (2). For KsLFY $\Delta$  SELEX, 25 times more fish sperm DNA was added in all incubation steps, to avoid aspecific binding of KsLFY $\Delta$  to the initial random library. The sequencing strategy was optimized for NaLFY and KsLFY for which SELEX sequences were added a 6 bp tag in 5' and a 6 bp index sequence in 3' for identification. For this, a first PCR was performed on beads issued from the SELEX to add the 5' tag, with primers AdF-x and AdR (see Table S3) and Phusion<sup>®</sup> High-Fidelity polymerase (Ozyme). The PCR program was as following: 1' denaturation at 98°C, 3 cycles of 10" denaturation at 98°C, 25" of annealing at 52°C, 13" extension at 72°C, and 13 cycles of 10" denaturation at 98°C, 25" of annealing at 65°C, 13" extension at 72°C. PCR fragments were loaded on a 3% agarose gel and purified with a NucleoSpin<sup>®</sup> Gel and PCR Clean-up kit (Macherey-Nagel). 3 ng of purified PCR product were used in a second PCR with primers Forward-III and Reverse-III-i (see Table S3) to add the 3' index, with 2 cycles of 10" denaturation at 98°C, 25" of annealing at 58°C, 13" extension at 72°C, and 3 cycles of 10" denaturation at 98°C, 25" of annealing at 68°C, 13" extension at 72°C. 3% DMSO (dimethylsulfoxide) was added to improve the amplification. The final PCR product was purified on a 3% agarose gel as previously, and analyzed on the BioAnalyzer 2100 with a DNA 1000 kit (Agilent). Equimolar amounts of all barcoded Selex samples were combined and sequenced on an Illumina Genome Analyzer II (read length 101 bp, single-end). Reads were extracted using the SHORE platform (<http://sourceforge.net/apps/mediawiki/shore/>). Sequences were sorted according to their occurrence, and the 2,000 most-enriched sequences were aligned using the MEME software (5) imposing symmetry and a size of 19 bp for type I and type II motifs, as previously done on AtLFY (4).

### **Crystallography.**

We performed the crystallization screens using the HTX platform at the EMBL outstation (Grenoble). The optimum condition was obtained with the hanging drop vapor diffusion method by mixing 1  $\mu$ l of PpLFY1-DBD – Type II DNA complex with 1  $\mu$ l of reservoir solution

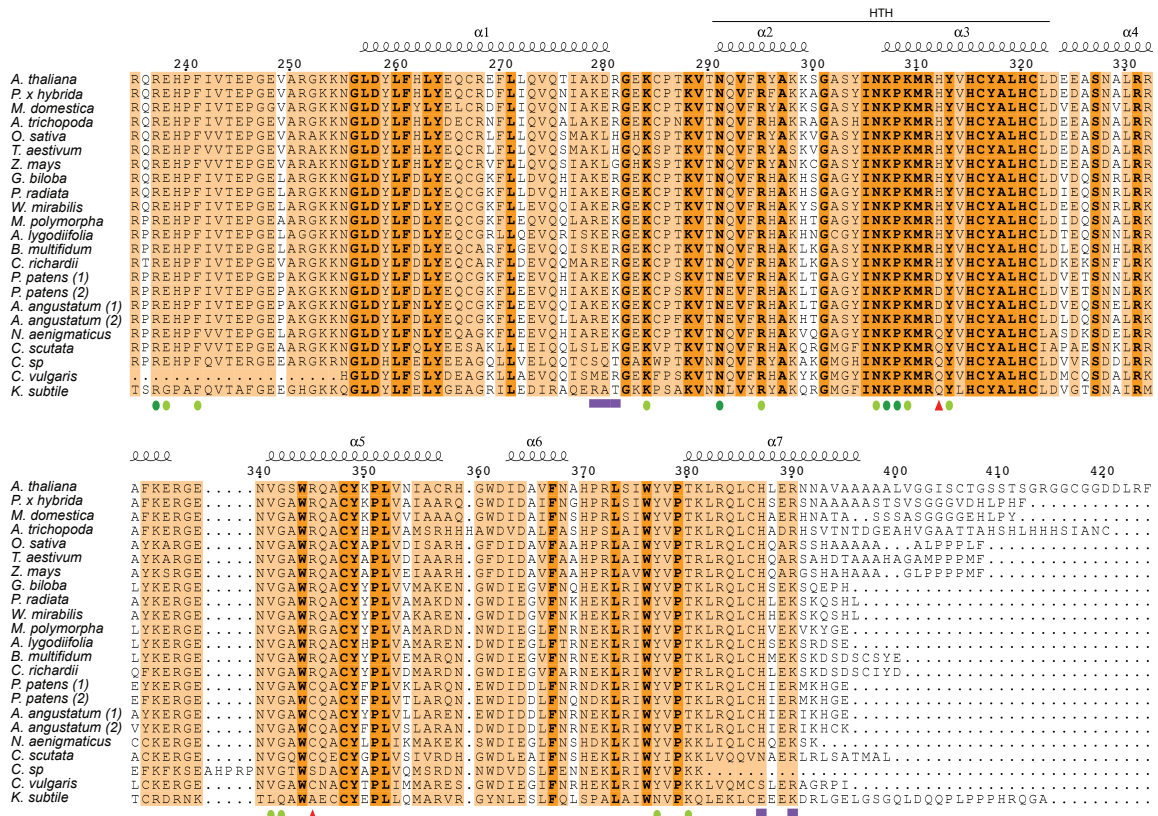
containing 20% MPD, 0.05 M MES pH 5.6 and 0.1 M Magnesium acetate. The type II oligonucleotide sequence is listed in Table S3. Crystals were cryoprotected by plunging into liquid nitrogen, after incubation in well conditions to which 20% glycerol had been supplemented. Diffraction data were collected from crystals maintained in a 100K nitrogen cryostream at a wavelength of 0.873 Å on the ESRF microfocus beamline ID23-2(6) using a helical data collection strategy. The structure was solved by molecular replacement using PHASER (7) using 2VY2 (AtLFY-DBD) as a search model. Two subunits in the asymmetric unit were placed, with rotational Z scores of 7.9 and 7.3 and translational Z-scores of 11.7 and 20.6. The model was improved by successive rounds of manual model building in COOT (8) followed by refinement in BUSTER (9). The final model (PDB ID 4BHK) had the following refinement parameters: crystallographic R-value of 19.6% and free R-value of 24.0% with 94.6%, 5.4%, 0% and 0% of protein residues in the core, allowed, generously allowed and disallowed regions of the Ramachandran graph, respectively.

### 3D structure modelling.

The model of protein binding on a type III motif was created by removing the central 3 base pairs from the PpLFY1-DBD - DNA crystal structure. The model was then split into two sub models, each containing a double stranded DNA half site bound to a single monomer. These models were then aligned to one another using the DNA of the PpLFY1-DBD - DNA structure. Specifically, DNA bases corresponding to positions -1 and -2 (CC) of the Type III SELEX logo and one extra nucleotide from one sub model were aligned to the 3 bp central region of the PpLFY1 DNA.

### **References for Material and Methods**

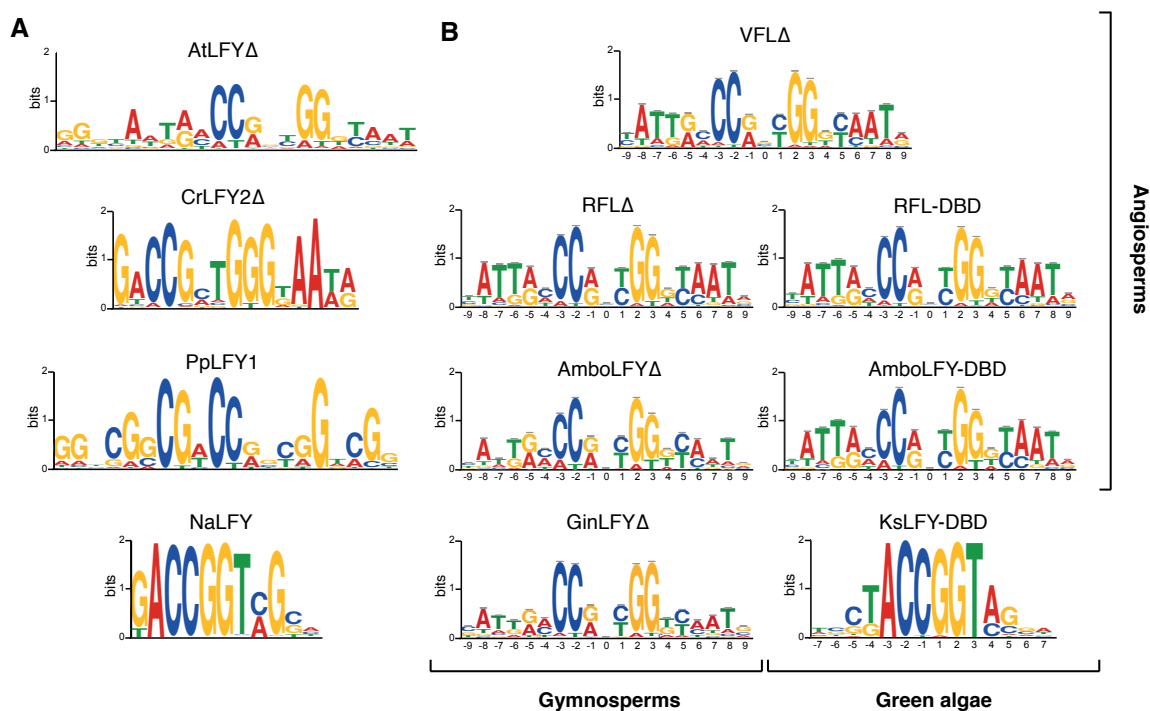
1. A. Dummer, A. M. Lawrence, A. de Marco, *Microb. Cell Fact.* **4**, 34 (2005).
2. H. Chahtane *et al.*, *Plant J.*, (2013).
3. C. Hamès *et al.*, *EMBO J* **27**, 2628 (2008).
4. E. Moyroud *et al.*, *Plant Cell* **23**, 1293 (2011).
5. T. L. Bailey, C. Elkan, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28 (1994).
6. D. Flot *et al.*, *J. Synchrotron Radiat.* **17**, 107 (2010).
7. A. J. McCoy *et al.*, *J. Appl. Crystallogr.* **40**, 658 (2007).
8. P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126 (2004).
9. G. Bricogne, *Acta Crystallogr. D Biol. Crystallogr.* **49**, 37 (1993).



**Fig. S1. Alignment of LFY-DBD sequences from various plant species.**

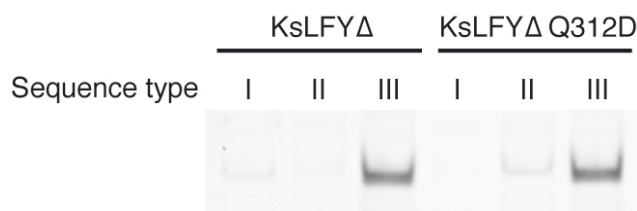
LFY from *Arabidopsis thaliana* (AtLFY) is taken as a reference for amino acids numbering and for the secondary structure annotation on the top of the alignment.  $\alpha$  indicates alpha helices, HTH indicates the helix-turn-helix domain. Dark green disks indicate contact with DNA bases; light green disks: contacts with phosphate backbone; red triangles: residues involved in a new DNA contact in PpLFY1; purple rectangles: residues involved in DBD dimerization (those on  $\alpha 1$  and those on  $\alpha 7$  are involved through their main and side chains respectively). (1) and (2) are indicated when two copies of LFY are shown. *A. thaliana*, *Petunia x hybrida*, *Malus domestica*, *A. trichopoda*, *O. sativa*, *Triticum aestivum*, *Zea mays* belong to angiosperms; *G. biloba*, *Pinus radiata*, *Welwitschia mirabilis* to gymnosperms; *Angiopteris lygodiifolia*, *Botrychium multifidum*, *C. richardii* to ferns; *M. polymorpha* to liverworts; *P. patens* and *Atrichum angustatum* to mosses; *N. aenigmaticus* to hornworts; and *Coleochaete scutata*, *Cylindrocystis* sp., *Chara vulgaris* and *K. subtilis* to derived streptophyte green algae. The sequence from *C. vulgaris* is incomplete in the N-terminal part. *C. sp* sequence is shorter in the C-terminus and might be incomplete.





**Fig. S2. Complementary results from the HT-SELEX experiment.**

(A) Examples of SELEX motifs obtained for AtLFY $\Delta$ , CrLFY2 $\Delta$ , PpLFY1 and NaLFY proteins, without imposing symmetry or size constraints on the motif. (B) SELEX motifs not presented in Fig. 1b, performed on VFL $\Delta$  from *Vitis vinifera*, RFL $\Delta$  and RFL-DBD from *Oryza sativa*, AmboLFY $\Delta$  and AmboLFY-DBD from *Amborella trichopoda*, GinLFY $\Delta$  from *Ginkgo biloba* and KsLFY-DBD from *Klebsormidium subtile*. For all motifs, symmetry is imposed, as well as a size constraint of 19 bp (except for KsLFY-DBD).  $\Delta$  means that around 40 amino acids from the C-terminal extremity have been removed, to allow efficient protein purification. DBD stands for DNA binding domain only.



**Fig. S3. Q and D at position 312 are equivalent for KsLFY $\Delta$  binding to type III sequence.**

KsLFY $\Delta$  preferentially recognizes type III sequence in EMSA, and mutating the glutamine (Q) residue at position 312 into an aspartic acid (D, as in the PpLFY1 protein) does not affect this preference, showing that Q and D at position 312 are essentially equivalent for type III motif recognition. All proteins are at 25 nM and only the protein-DNA complex is shown.

**Table S1. Data collection and refinement statistics.**

Space group	$P2_12_12_1$
Unit cell dimensions:	a=42.7, b=84.5,c=152.4
<b>Data Collection</b>	
Beamline	ESRF ID23-EH2
Wavelength (Å)	0.873
Resolution (Å) <sup>*</sup>	50-2.31(2.46-2.32)
Completeness (%) <sup>*</sup>	99.1 (97.2)
Redundancy <sup>*</sup>	3.8 (3.8)
R <sub>sym</sub> <sup>a</sup> (%) <sup>*</sup>	9.7 (48.9)
$\langle I/s(I) \rangle$ <sup>*</sup>	10.1(2.6)
<b>Refinement statistics</b>	
Resolution <sup>*</sup>	43.6-2.32 (2.42-2.32)
R <sub>work</sub> <sup>b</sup> (%) <sup>*</sup>	19.6(22.5)
R <sub>free</sub> <sup>b</sup> (%) <sup>*</sup>	24.0(24.6)
# protein atoms	2653
# solvent atoms	108
#DNA atoms	1183
Average B value (Å <sup>2</sup> )	
Overall	34.7
Protein	31.6
DNA	41.3
RMSD bond lengths (Å)	0.009
RMSD bond angles (°)	1.20
<b>Ramachandran</b>	
Core/Allowed/Disallowed(%)	94.6/5.4/0
<b>PBD ID</b>	4BHK

$$^a R_{sym} = \frac{\sum_h \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum_h \sum_i I_i(h)}$$

$$^b R_{work,free} = \frac{\sum_h (|F_{obs} - k F_{calc}|)}{\sum_h (|F_{obs}|)}$$

where R<sub>work</sub> is calculated using the working set and R<sub>free</sub> is calculated using the test set (5.1% of the total reflections)

(\*) Values in parentheses correspond to the outermost resolution shell.

**Table S2. Detail of buffers used for protein purifications.**

Proteins	Sonication buffer	Basic buffer	Equilibration buffer	Washing buffer	Elution buffer	Dialysis buffer
RFL-DBD, VFL-DBD, MarpoFLO-DBD	Tris 20 mM pH 8 NaCl 250 mM glycerol 5 % TCEP 5 mM	Tris 20 mM pH 8 NaCl 250 mM glycerol 5 % DTT 1 mM	Basic buffer + 5 mM imidazole	Basic buffer + 50 mM imidazole	Basic buffer + 350 mM imidazole	Tris 20 mM pH 7.5 NaCl 150 mM glycerol 5 % EDTA 0.25 mM MgCl <sub>2</sub> 2 mM DTT 5 mM
RFLΔ, VFLΔ, CrLFY2Δ	Tris 50 mM pH 8 TCEP 3 mM	Tris 20 mM pH 7.5 TCEP 1 mM	Basic buffer + 10 mM imidazole	Basic buffer + 20 mM imidazole	Basic buffer + 300 mM imidazole	Tris 20 mM pH 7.5 glycerol 5 % DTT 5 mM
PpLFY1, AtLFYΔ and mutant proteins from Fig. 2c, KsLFYΔ, NaLFY	Tris 50 mM pH 8 TCEP 3 mM MgCl <sub>2</sub> 1 mM glycerol 10 %	Tris 25 mM pH 8 TCEP 1 mM	Basic buffer + 5 mM imidazole	Basic buffer + 20 mM imidazole	Basic buffer + 300 mM imidazole	Tris 25 mM pH 8 EDTA 0.5 mM DTT 1 mM

**Table S3. List of oligonucleotides used in this study.**

Name	Sequence	Experiment
AdF-x AdR Forward-III Reverse-III-i	ACACTCTTTCCTACACGACGCTCTTCCGATCTxxxxxTGGAGAAGAGGAGAGATCTAGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCCGAATCGAAGAACAAG AATGATACGGCGACCACCGAGATCTACACTTTCCTACACGACGCTCTTCCGATCT CAAGCAGAAGACGGCATACGAGATiiiiiiGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	HT-SELEX sequencing (for NaLFY and KsLFY)
Tag 1 Tag 2 Tag 3 Tag 4	ACGTAT CATGCA TGCAGC GTACTG	Tags in x position for HT-SELEX sequencing
Index 1 Index 2	CGATGT TTAGGC	Indexes in i position for HT-SELEX sequencing
Type I probe (AP1) Type II probe (Moss-54) Type III probe (Notho4) Mutated type I probe (AP1m3)	GTTGGGGAAGGACCAGTGGTCCGTACAATGT GGCCACGGGCGACCAGCGGACGGTGTGAGCAC GCGCACCGACCGTCTGTAGCGCAACCGCGCG GTTGGGGAAGGAAAAGTAATCCGTACAATGT	EMSA

### III. Bilan et perspectives

Afin d'analyser l'évolution de la spécificité de liaison à l'ADN de LFY, nous avons d'abord cherché à enrichir notre collection de séquences d'homologues de LFY issues des grands groupes évolutifs. Nous avons alors recherché des séquences de LFY dans une base de données transcriptomiques récente. Notre première surprise fut de découvrir que LFY existait aussi chez les algues multicellulaires, alors qu'il était jusqu'à présent admis que LFY était spécifique des plantes terrestres.

Nous avons ensuite montré que la spécificité de liaison à l'ADN de LFY avait changé plusieurs fois au cours de l'évolution et nous avons expliqué les bases moléculaires de ces changements dans la lignée verte. Les angiospermes, les fougères et les hépatiques (*liverworts*) reconnaissent le même type de motif (motif I) et adoptent vraisemblablement le même mode de liaison que celui que l'on connaît pour la protéine d'*Arabidopsis*. PpLFY (bryophyte) reconnaît un motif différent (motif II). Cela s'explique par la présence de deux résidus au niveau de l'interface ADN-protéine. Ils permettent à PpLFY de former un contact avec l'ADN qui est absent du complexe AtLFY-ADN. Notre seconde surprise a été de constater que les protéines LFY d'une algue verte et d'une anthocérophyte (*hornwort*) donnaient un motif encore différent (motif III), où les 3 pb centrales du palindrome sont absentes. Nous avons montré que la liaison sur ce type de motif impliquait l'absence du contact protéine-protéine entre les deux monomères, présent chez les protéines liant les motifs I et II.

Afin de consolider nos conclusions sur le mode de liaison de LFY sur le motif III, nous avons initié un projet de cristallographie avec une protéine reconnaissant ce type de séquence. Nous avons actuellement des cristaux qui diffractent à 4,5 Å et qui constituent un point de départ prometteur vers l'obtention de la structure.

Nous avons vu que seulement quelques mutations étaient suffisantes pour faire basculer la spécificité de la protéine d'un type de motif à un autre. Cela montre que la forte conservation généralement observée chez les FTs, argument en défaveur d'une évolution des FTs (Carroll 2008), ne garantit pas que leurs propriétés soient conservées.

Après avoir analysé les changements de spécificité de LFY, la question suivante restait en suspens : comment de telles modifications des propriétés de LFY ont-elles pu être tolérées alors que *LFY* est un gène unique et a des rôles essentiels au moins chez les angiospermes et chez *P. patens* ?

Nous avons montré que la protéine du groupe des anthocérophytes (NaLFY), avait une spécificité plus relâchée que les autres protéines : elle peut s'accommoder des trois types de motif. Nous proposons qu'une protéine ayant des caractéristiques similaires à NaLFY ait pu servir d'intermédiaire et permettre une transition douce entre les différents types de spécificité.

Cette transition graduelle pourrait permettre une adaptation progressive des éléments *cis* (coévolution) si la protéine garde une fonction similaire lors de la transition, ou bien permettre le passage progressif d'une fonction à une autre dans le cas où le changement des propriétés de la protéine serait lié à un changement de fonction.

Nous ne savons pas laquelle de ces différentes solutions s'applique dans le cas de LFY. En effet, dans cette étude nous nous sommes limités à la compréhension moléculaire de l'évolution de la protéine. Il serait intéressant, par la suite, d'étudier l'évolution de sa fonction, en particuliers chez les espèces ayant divergées au moment des changements de spécificité. Nous disposons de données fonctionnelles chez *P. patens*, mais nous ne savons rien sur la fonction de LFY chez les hépatiques, les anthocérophytes et les algues. Les gènes cibles de LFY ne sont connus dans aucun de ces groupes. Ces aspects fonctionnels mériteraient d'être étudiés.

L'histoire évolutive de LFY s'inscrit donc dans le débat sur la contribution des éléments *cis* et facteurs *trans* pour la modification des réseaux de régulation au cours de l'évolution. En effet, cette étude montre d'une part, qu'un FT peut changer de propriétés au cours de l'évolution, et d'autre part, que de tels changements sont possibles même pour un FT n'appartenant pas à une famille multigénique.



# CONCLUSION





Cette étude s'inscrit dans le cadre de la compréhension des mécanismes impliqués dans l'expression du génome au cours du développement et de l'évolution. LFY est un modèle de choix pour ce type de problématique car c'est un FT unique ayant une fonction cruciale chez les plantes.

Nous avons mis en oeuvre différentes méthodes afin de répondre aux objectifs de départ, qui étaient de mieux appréhender le fonctionnement de la protéine au niveau moléculaire (Chapitre I) et de suivre certaines de ses propriétés d'un point de vue évolutif (Chapitre II).

La biologie structurale intégrée a eu une place centrale au cours de ces travaux. Pour l'étude du domaine N-terminal de LFY, l'obtention de la structure 3D a été le point de départ ayant conduit à la mise en place d'une approche alliant biochimie, biologie végétale et analyse à grande échelle avec le ChIP-seq. De plus la structure nous a permis de découvrir le repliement adopté par la protéine (SAM), ce qui n'aurait pas été possible par d'autres approches.

Nous avons mis en évidence un mode de fonctionnement original pour LFY et ses interactions avec l'ADN, qui vient enrichir le répertoire déjà très varié des modes de liaison ADN-protéine. Nos résultats ouvrent également de nouvelles perspectives quant aux interactions plus complexes entre LFY et la chromatine : l'ADN, les protéines associées, et la dynamique des interactions ayant lieu entre tous ces acteurs lors de la régulation transcriptionnelle.

Concernant l'évolution de la spécificité de liaison à l'ADN de LFY, la biologie structurale a fourni de solides explications aux diverses spécificités de liaison mises en évidence par d'autres méthodes (SELEX notamment). Cette étude apporte un exemple inédit aux possibles modes d'évolution des FTs et fournira des pistes pour la compréhension des mécanismes sous-jacents à l'évolution.



# MATERIEL ET METHODES

Les techniques ayant contribué aux résultats exposés dans les articles sont déjà décrites dans les matériels et méthodes des articles correspondants. Des précisions, ainsi que les matériels et méthodes correspondants aux résultats complémentaires sont apportés ici.

## I. Matériel

### 1. Matériel végétal

Les plantes *A. thaliana* utilisées sont de l'écotype Columbia-0 (Col-0). Le mutant *lfy-12*, également en fond génétique Col-0 porte une mutation entraînant la présence d'un codon stop dans la séquence N-terminale de la protéine (Q32stop), seul un court peptide est donc potentiellement synthétisé, la protéine LFY est absente (Maizel *et al.*, 2005).

### 2. Matériel bactérien

Les clonages ont été effectués dans des bactéries *E. coli* de souche DH5 $\alpha$  (Invitrogen). Les expressions de protéines recombinantes ont été effectuées à partir de bactéries *E. coli* de souche Rosetta2 (DE3) (Novagen). Les transformations chez *A. thaliana* ont été effectuées avec des bactéries *A. tumefaciens* C58C1 pMP90.

## II. Méthodes

### 1. Biologie moléculaire

Les extractions d'ADN à partir de bactéries ont été faites à l'aide du kit Minipreps NucleoSpin Plasmid (Machery-Nagel).

- **Constructions des vecteurs d'expression pour la production de protéines recombinantes chez *E. coli***

Les séquences codantes pour les protéines d'intérêt ont été clonées soit dans le vecteur pETM-11 Dummler *et al.*, 2005, qui permet l'addition d'un *tag* 6xHis clivable par la protéase

TEV (Tobacco Etch Virus) en N-terminal de la protéine, soit dans le vecteur pET-30a (Novagen), qui permet l'addition d'un *tag* 6xHis en C-terminal de la protéine.

Typiquement, les séquences à cloner ont été amplifiées par PCR à partir de vecteurs préexistants, disponibles au laboratoire, avec la polymérase Phusion (Thermo Scientific), à l'aide d'oligonucléotides permettant l'insertion de sites de restriction aux deux extrémités du produit PCR. Après purification sur gel d'agarose (kit NucleoSpin Gel and PCR clean-up, Machery-Nagel), le produit est inséré dans le vecteur pCR-Blunt (Invitrogen). A ce stade, la séquence de l'insert est vérifiée par séquençage (compagnie Eurofins) à l'aide d'oligonucléotides placés dans le vecteur (oligonucléotides standard T7 et T7term). L'insert est ensuite extrait du pCR-Blunt par digestion avec des enzymes appropriées, de nouveau purifié sur gel d'agarose, et cloné dans le vecteur final (pETM-11 ou pET-30a). Le vecteur final est validé par obtention de profils de digestion cohérents, avec des enzymes coupant de part et d'autre de l'insert et dans l'insert.

#### ▪ **Mutagenèse dirigée**

Les mutagenèses dirigées ont été réalisées selon le principe QuikChange Site Directed Mutagenesis (Agilent) qui consiste à amplifier la totalité du vecteur avec des oligonucléotides contenant la mutation souhaitée. Les oligonucléotides ont été dessinés à l'aide du serveur QuikChange Primer Design (<http://www.genomics.agilent.com/primerDesignProgram.jsp>). 50 ng de vecteur initial sont utilisés comme matrice dans une réaction PCR contenant 125 ng de chaque oligonucléotide ; 0,2  $\mu$ M de chacun des quatre dNTPs (désoxynucléotide triphosphate) ; la polymérase Phusion accompagnée de son tampon (Thermo Scientific) ; 0 à 10% de DMSO (Diméthylsulfoxyde). Après amplification (30 s à 95°C, (30 s à 95°C, 1 min à 55°C, 30 s/Kb à 72°C) x16 cycles), le produit obtenu est digéré par l'enzyme DpnI qui coupe uniquement le vecteur matrice (l'ADN matrice est méthylé car issu d'une cellule bactérienne alors que l'ADN synthétisé par PRC ne l'est pas), puis est transformé dans des bactéries DH5 $\alpha$ . Après extraction, les vecteurs ont été séquencés (Eurofins) à l'aide d'oligonucléotides placés dans le vecteur.

Les noms des constructions obtenues et des oligonucléotides utilisés sont détaillés dans les Articles 1 et 2.

#### ▪ **Constructions des vecteurs binaires pour la transformation chez *A. thaliana***

Ces constructions sont décrites dans l'Article 1.

## 2. Biochimie des protéines

### ▪ Analyse de séquences et structures des protéines

Les alignements de séquences des protéines ont été faits grâce aux serveurs Multalin (<http://multalin.toulouse.inra.fr/multalin/>) et ESPrit (<http://esprict.ibcp.fr/ESPript/ESPript/>).

Les paramètres physico-chimiques théoriques des protéines (masse moléculaire, coefficient d'extinction molaire, point isoélectrique, indice de stabilité) ont été calculés sur le serveur ProtParam (<http://web.expasy.org/protparam/>).

Les prédictions de domaines globulaires et de structures secondaires ont été faites à l'aide des serveurs Globplot (<http://globplot.embl.de/>) et GORIV ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_gor4.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html)), respectivement.

Les structures 3D ont été visualisées et les images réalisées avec le logiciel PyMOL (Schrodinger). Les alignements de structures 3D et les recherches d'homologies structurales ont été réalisés sur le serveur Dali ([http://ekhidna.biocenter.helsinki.fi/dali\\_server/](http://ekhidna.biocenter.helsinki.fi/dali_server/)). La visualisation de la conservation des acides aminés sur les structures 3D a été faite grâce au serveur Consurf (<http://consurf.tau.ac.il/>).

### ▪ Expression et purification des protéines recombinantes

Ces techniques sont détaillées dans les Articles 1 et 2.

Seule l'étape de Q-Sepharose n'a pas été expliquée. Elle est réalisée à 4°C et consiste à injecter l'extrait protéique sur une colonne contenant 8 mL de résine Q-Sepharose (GE Healthcare). La colonne est reliée à une FPLC (Fast Protein Liquid Chromatography) AKTA (GE Healthcare). Avant l'injection, la colonne est équilibrée avec 5 volumes d'un tampon contenant 0,5 M NaCl ; 20 mM Tris-HCl pH8 ; 1 mM TCEP (Tris(2-carboxyethyl)phosphine) (1 volume correspond au volume de la colonne, soit 8 mL).

### ▪ Suivi des purifications et dosage

#### Gels SDS-page de contrôle de la purification

Les différentes fractions (extrait total, extrait soluble, lavages et élution) sont analysées par électrophorèse sur gel SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) pour repérer la protéine d'intérêt, évaluer sa pureté et le rendement de la purification.

Gel de concentration: 5 % acrylamide/bisacrylamide 37.5:1 (Merck) ; 62,5 mM Tris-HCl pH 6.8 ; 0,1 % (p/v) SDS (Sodium Dodecyl Sulfate) ; 0,1 % (p/v) persulfate d'ammonium (PSA) ; 0,01 % TEMED (Tétraméthyléthylènediamine). Gel de séparation: 10 à 15 % acrylamide/bisacrylamide 37.5:1 ; 375 mM Tris-HCl ; 0,1 % (p/v) SDS ; 0,1 % (p/v) PSA ; 0,01 % TEMED. Les concentrations en acrylamide/bisacrylamide sont ajustées en fonction de la

masse moléculaire de la protéine à observer : 10 à 12 % pour le suivi des étapes de purification des protéines entières (de masse moléculaire supérieure à 20 kDa) ou 15% pour le suivi des étapes de purification des protéines ne contenant que les domaines N ou C-terminal (de masse moléculaire de l'ordre de 12 à 20 kDa).

Des quantités de l'ordre de 5 µg (pour les protéines pures) à 150 µg (pour les extraits bruts) de protéine sont chargées sur le gel dans un tampon de dénaturation (50 mM Tris-HCl pH 6.8 ; 10 % p/v glycérol ; 1% (p/v) SDS ; 0,0025% (p/v) bleu de bromophénol ; 0,4 % (p/v) Dithiothreitol) après dénaturation 5 min à 95°C. Un marqueur de poids moléculaire est chargé dans l'un des puits du gel (Precision Plus Protein Standards Dual Color, Bio-Rad pour la plupart des gels). La migration est effectuée jusqu'à ce que le bleu de bromophénol sorte du gel (environ 1h15) à 20 mA/gel, dans du tampon de type tampon Laemmli (25 mM Tris-HCl ; 192 mM glycine ; 0,1 % (p/v) SDS). Le gel est ensuite coloré dans une solution de bleu de Coomassie (0,25 % (p/v) bleu R 250 ; 45 % éthanol ; 10 % acide acétique) et décoloré dans une solution de décoloration (10 % éthanol ; 10 % acide acétique).

#### Dosage par mesure d'absorbance à 280 nm

Les protéines pures ont été dosées par mesure de l'absorbance à 280 nm. La concentration est calculée suivant la loi de Beer-Lambert :  $A = \epsilon L C$  où A est l'absorbance à 280 nm,  $\epsilon$  est le coefficient d'extinction molaire de la protéine, L est la longueur du trajet optique et C est la concentration en protéine. Les mesures d'absorbance ont été réalisées sur un spectrophotomètre NanoDrop-2000 (Thermo Scientific).

#### Dosage de Bradford

Les protéines issues des purifications sur colonne de nickel-Sepharose ont été quantifiées par dosage au réactif de Bradford (Bio-Rad). Les valeurs d'absorbance à 595 nm de trois dilutions de la protéine sont confrontées à une gamme étalon de BSA (albumine de sérum bovin) titrée réalisée dans les mêmes conditions de mesures.

#### Dosage sur gel d'acrylamide SDS-page

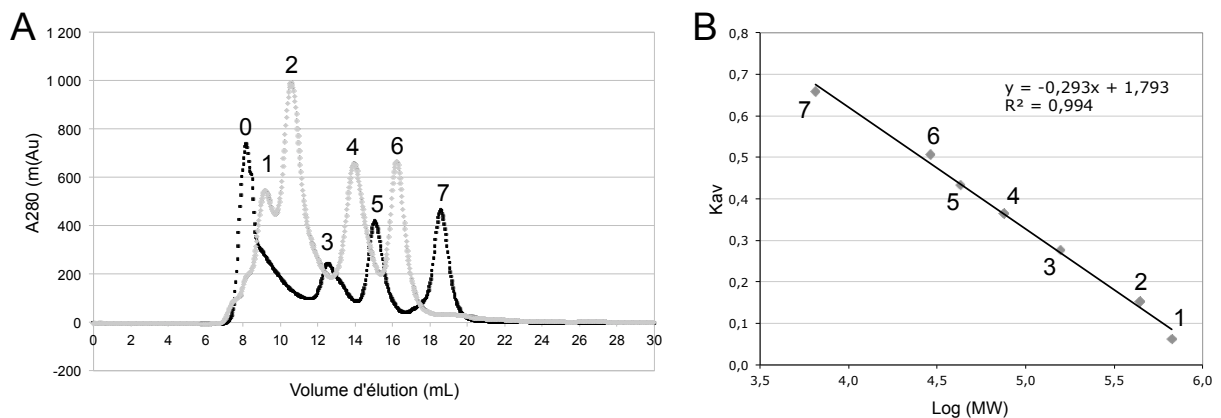
Cette technique de dosage a également été utilisée pour des protéines issues de purification sur colonne de nickel-Sepharose. Elle consiste à déposer un volume croissant de protéines sur gel SDS-PAGE et comparer l'intensité de la bande correspondante à la protéine d'intérêt à celle d'une gamme étalon d'actine titrée (0,5 à 2,5 µg) mise à migrer dans les mêmes conditions. Ces intensités sont estimées à l'aide du logiciel ImageJ (<http://rsbweb.nih.gov/ij/>). Cette technique a été utilisée pour le dosage de protéines non pures à homogénéité, seule la bande correspondante à la protéine d'intérêt a été quantifiée.

### ▪ Analyse des masses moléculaires des protéines par SEC et SEC-MALLS

#### SEC analytique

Les masses moléculaires apparentes des protéines ont été calculées d'après leurs profils d'éluion sur une colonne Superdex-200 10/300GL (GE Healthcare). La colonne est reliée à une FPLC AKTA (GE Healthcare). L'analyse est effectuée à 4°C. Avant l'injection, la colonne est équilibrée avec 2 volumes du tampon approprié pour l'analyse (1 volume correspond au volume de la colonne, soit 24 mL). 50 µL de protéine sont injectés, le débit du tampon sur la colonne est de 0,3 mL/min. Le volume d'éluion correspond au point le plus haut du pic d'absorbance à 280 nm.

La colonne a été calibrée avec des protéines de masses moléculaires connues (Gel filtration calibration kits, GE Healthcare). La calibration en présence de sel (dans un tampon 0,8 M NaCl ; 50 mM Tris-HCl pH 8 ; 1 mM TCEP est montrée ci-dessous et a donné lieu à l'équation  $K_{av} = -0,293 \text{ Log}(MW) + 1,793$ . Une calibration similaire a été réalisée dans un tampon sans sel (20 mM Tris-HCl pH 8 ; 1 mM TCEP). Dans ce cas, l'équation de la droite obtenue était :  $K_{av} = -0,224 \text{ Log}(MW) + 1,338$ . MW est la masse moléculaire.  $K_{av} = (V_e - V_0)/(V_c - V_0)$  ; où  $V_0$  est le volume mort de la colonne et correspond au volume d'éluion du bleu dextran ;  $V_e$  est le volume d'éluion de la protéine et  $V_c$  est le volume de la colonne.



#### Calibration de la colonne SEC S200 10/300GL.

**A.** Profil d'éluion des protéines de masses moléculaires connues et du bleu dextran (Gel filtration calibration kits, GE Healthcare) sur une colonne Superdex-200 10/300GL (GE Healthcare) dans un tampon 0,8 M NaCl ; 50 mM Tris-HCl pH 8 ; 1 mM TCEP. 0 : bleu dextran (volume mort) ; 1 : Thyroglobuline (669 kDa) ; 2 : Ferritine (440 kDa) ; 3 : Aldolase (158 kDa) ; 4 : Conalbumine (75 kDa) ; 5 : Ovalbumin (43 kDa) ; 6 : Anhydrase carbonique (25 kDa) ; 7 : Aprotinine (6,5 kDa). Les protéines ont été injectées en deux fois (courbe noire et courbe grise)

**B.** Courbe de calibration obtenue à partir des profils d'éluion sur la figure A. La numérotation des points correspond à celle des pics en figure A. La droite a été obtenue en ajustant les points avec une droite dont l'équation correspond à  $K_{av} = a \times \text{Log}(MW) + b$ . MW est la masse moléculaire.  $a = -2,293$  et  $b = 1,793$ .  $K_{av} = (V_e - V_0)/(V_c - V_0)$  ; où  $V_0$  est le volume mort de la colonne et correspond au volume d'éluion de bleu dextran ;  $V_e$  est le volume d'éluion de la protéine et  $V_c$  est le volume de la colonne.

SEC-MALLS

50 µL de protéine (typiquement à une concentration de 2 à 5 mg/mL), ou de mélange ADN-protéine ont été injectés sur une colonne Superdex-200 10/300GL (GE Healthcare) équilibrée avec un tampon approprié pour l'analyse effectuée. L'éluion des protéines est conduite à un débit de 0,5 mL/min à 20°C. La concentration en protéine est mesurée par un détecteur RI2000 (Schambeck SFD), la diffusion de la lumière aux angles multiples est mesurée par un détecteur DAWN-EOS (longueur d'onde d'émission = 690nm ; Wyatt Technology Corp.). Le calcul de masse moléculaire est effectué sur le logiciel ASTRA, comme décrit par Gerard *et al.*, 2007.

Les complexes ADN-protéine ont été effectués par mélange des deux espèces à la concentration souhaitée dans un tampon contenant 0,8 M NaCl ; 20 mM Tris-HCl pH 8 ; 1 mM TCEP. Le mélange est ensuite dialysé à 4°C pendant 3 h dans un tampon contenant 0,15 M NaCl ; 20 mM Tris-HCl pH 8 ; 0,25 mM EDTA ; 2 mM MgCl<sub>2</sub> ; 1 mM TCEP, avec une membrane Spectra/Por MWCO 3,500 (Spectrum Labs). La préparation des oligonucléotides est expliquée plus loin (voir « Préparation des oligonucléotides »).

- **Caractérisation des interactions ADN-protéine en EMSA**

Préparation des oligonucléotides

Les séquences des oligonucléotides utilisés dans les Articles 1 et 2 sont données dans les matériels et méthodes correspondants.

La séquence de l'oligonucléotide *2sites* est la suivante : GCTTGGGGAAGGACCAGTGGTCCGTACAATTATCATTGTGTGGTGTGTTGACCGTTGGTCCATTTCTGA.

Les deux sites de liaison (19 pb chacun) sont soulignés.

Les oligonucléotides simples brins sont synthétisés par la compagnie Eurofins, ils sont livrés sous forme lyophilisée et repris dans du tampon TE (10 mM Tris-HCl pH 8 ; 1 mM EDTA).

Les fragments d'ADN double brins utilisés pour les EMSA, le SEC-MALLS et la cristallographie ont été obtenus par hybridation entre deux oligonucléotides simples brins dans un tampon d'hybridation (150 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 7,5) par passage à 95°C pendant 3 minutes et refroidissement progressif jusqu'à température ambiante. Seul l'oligonucléotide *AGI-II* a été obtenu par PCR à partir d'un oligonucléotide simple brin (Eurofins), sa synthèse est expliquée dans l'Article 1.

Pour les EMSA, les oligonucléotides possèdent un dGTP dépassant en 5' permettant un marquage au Cy3-dCTP : 4 pmol d'ADN double brin sont incubés à l'obscurité pendant 2 h à 37°C avec 1 unité de fragment Klenow (Ozyme) et 8 pmol Cy3-dCTP (GE Healthcare) dans le tampon fourni avec la Klenow. La réaction est arrêtée par un passage de 10 min à 65 °C. Les fragments d'ADN obtenus sont stockés à -20°C.



### EMSA

Les gels d'acrylamide en conditions non-dénaturantes (gels natifs) contiennent 6% acrylamide/bisacrylamide 29:1 (Roth) ; 0,5X TBE (45 mM Tris-HCl ; 45 mM acide borique ; 1 mM EDTA pH 8 ; 0,1% (p/v) PSA ; 0,01% TEMED).

Les complexes entre les protéines recombinantes et l'ADN ont été préparés comme décrits dans les Articles 1 et 2. Les gels ont été imagés à l'aide d'un scanner Typhoon 9400 (Molecular Dynamics). Les quantifications ont été faites avec le logiciel ImageQuant (Molecular Dynamics).

#### ▪ **Détermination des séquences ADN liées par les protéines par SELEX**

Les techniques de SELEX et QuMFRA sont décrites dans l'Article 2.

### **3. Cristallographie des protéines**

Les procédures utilisées pour cristalliser les protéines et résoudre leurs structures sont décrites dans les Articles 1 et 2.

### **4. Biologie végétale**

#### ▪ **Culture des plantes *A. thaliana***

Les plantes ont été cultivées en condition standard en jour longs (22°C, 16h de lumière, 8h d'obscurité) ou en jours courts (22°C, 8h de lumière, 16h d'obscurité).

Les graines ont été stérilisées avant semis par un passage de 2h à -80°C, puis par une incubation de 10 min dans une solution contenant 0,4% (p/v) de chlore, 0,02% (p/v) de SDS et 86% d'éthanol. Les graines ont été rincées avec de l'éthanol à 95% et utilisées après évaporation complète de l'éthanol.

Après stratification pendant 3 jours à 4°C à l'obscurité, les graines ont été semées sur milieu 0,5X Murashige and Skoog (MS; duchefa biochemie). Après environ 15 jours, les plantules ont été repiquées sur terre. La terre est un mélange de terreau (2/3) et de vermiculite (1/3) traité par un insecticide biologique à base de *Bacillus thuringiensis* (Vectobac 12-AS, Biobest). Les plantes cultivées dans le but d'être transformées ont été semées directement sur terre.

### ▪ Génotypage de l'allèle *Ify-12*

#### Extraction d'ADN génomique d'*A. thaliana*

L'ADN génomique (ADNg) de la plante est extrait à partir de l'équivalent de 4 mm<sup>2</sup> d'une jeune feuille broyée mécaniquement à l'aide d'un broyeur de tissus (broyeur MM30, Retsch/Qiagen) après congélation dans de l'azote liquide. La suite de l'extraction est réalisée à température ambiante. Le broyat est repris dans 500 µL de tampon d'Edwards (200 mM Tris-HCl pH 7,5 ; 250 mM NaCl ; 25 mM EDTA ; 10% (p/v) SDS ; Edwards *et al.*, 1991) et centrifugé 1 min à 16 000 g. 400 µL de surnageant sont alors mélangés avec 400 µL d'isopropanol et incubés 2 min et centrifugé 5 min à 16 000 g. Le culot est repris par 500 µL d'éthanol 70% et centrifugé 5 min à 16 000 g. Le surnageant est éliminé et le culot est repris dans 30 à 50 µL d'eau après évaporation complète de l'éthanol et conservé à -20°C pendant quelques semaines.

#### Génotypage pour l'allèle *Ify-12*

Le génotypage est réalisée comme décrit par Benlloch *et al.*, 2011. Il est basé sur la méthode du End Point Genotyping LightCycler 480 System (Roche). Les couples d'oligonucléotides *oLFY-2*, *LFYqGenR* et *LFYqGenF*, *oMLM1007* ont été utilisés pour génotyper les plantes transformées par des constructions ayant les promoteurs *pLFY* et *p35S* respectivement. Les sondes Taqman *TAQMANLFYWT* et *TAQMANLFY12* marquées en 5' par les fluorophores FAM et HEX respectivement ont été utilisées. La PCR a été réalisée et suivie sur un appareil Rotor Gene 3000 (Qiagen) avec le Kit LightCycler 480 Probes Master (Roche), en suivant les instructions du kit. Les séquences des oligonucléotides utilisés pour le génotypage sont indiquées dans le tableau ci-dessous :

Nom	Séquence
<i>oLFY-2</i>	CTGTCAATTTCCCAGCAAGACAC
<i>LFYqGenR</i>	CGTCCTTCATACCCACAA
<i>LFYqGenF</i>	ACGAGAGCATTGGTTCAA
<i>oMLM1007</i>	CGTGCGATATTTGATATTTCTATATGTATTCA
<i>TAQMANLFYWT</i>	CGCTGCAGCAACAGC (5'FAM- 3'BHQ)
<i>TAQMANLFY12</i>	CGCTGCAGTAACAGC (5'HEX-3'BHQ)

### ▪ Génération de lignées transgéniques chez *A. thaliana* par transformation par *A. tumefaciens*

La transformation est réalisée comme décrit par Zhang *et al.*, 2006. Le protocole est rappelé ci-dessous.

Préparation des bactéries *A. tumefaciens*

Les *A. tumefaciens* sont cultivées à une température de 28°C. Après transformation du plasmide à transférer par électroporation, les bactéries sont incubées 2 h dans du milieu LB liquide (2% p/v LB Broth Base Lennox ; Invitrogen), puis 2 jours sur un milieu LB solide (contenant 1% d'agar) en présence de 25 µg/mL de rifampicine, 50 µg/mL de gentamicine et 100 µg/mL de spectinomycine. Une colonie est utilisée, après vérification de la présence du plasmide par PCR avec des oligonucléotides appropriés, pour faire une pré-culture dans 1 mL de LB contenant les antibiotiques cités précédemment. Après 24 h, cette pré-culture sert à lancer une culture de 500 mL (LB avec antibiotiques). Après 24h, les bactéries sont récoltées par centrifugation 15 min à 3000 g à 20°C. Le culot est repris dans 50 mL de sucrose 5 % (p/v), puis dilué dans 500 mL de manière à ce que la DO à 600 nm atteigne 0,8. Du Silweet L-77 (un surfactant) est ajouté à 0,05%, et de la benzylaminopurine (hormone favorisant la floraison) est ajoutée à 10 µg/mL.

Transformation d'*A. thaliana* par *A. tumefaciens*

Les hampes florales de plantes adultes (cultivées environ 5 semaines en jours longs) sont trempées environ 10 s dans la solution d'*A. tumefaciens*. Après environ 2 j sous cloche à luminosité réduite, les plantes sont remises en culture (jours longs) jusqu'à récolte des graines.

Sélection des plantes transformées

Les graines transgéniques sont sélectionnées par observation de la fluorescence émise par la GFP sous une loupe binoculaire SZX12 (Olympus).

Les lignées *p35S:LFY* (CA26-15) et *p35S:LFY<sub>TERE</sub>* (CA29-11 et CA29-17) sélectionnées pour les analyses par ChIP-seq, RT-PCR quantitative et Western-blot sont en fond génétique Col-0 et comporte une seule insertion indépendante du transgène. Ce paramètre a été vérifié selon la ségrégation des graines T2 :  $\frac{3}{4}$  de graines transgéniques et  $\frac{1}{4}$  de non transgéniques (déterminé par comptage des graines fluorescentes).

A partir de la T2, ces lignées ont été utilisées à l'état hétérozygote pour le transgène (les plantes *p35S:LFY* homozygotes donnent très peu ou pas de graines). Pour sélectionner ces plantes, la ségrégation de leur descendance est observée (par comptage des graines fluorescentes) et devrait correspondre à :  $\frac{3}{4}$  de graines transgéniques et  $\frac{1}{4}$  de non transgéniques.

- **Détection des protéines dans les tissus végétaux par Western-blot**

Cette technique est expliquée dans l'Article 1.

### ▪ Analyses du niveau d'expression des gènes chez *A. thaliana* par RT-PCR quantitative

Les analyses d'expression des gènes ont été réalisées comme décrit par Chahtane *et al.*, 2013. Le protocole est rappelé ci-dessous.

#### Extraction des ARNs

Les ARNs ont été extraits à partir de plantules de 8 et 15 j cultivées en jours courts sur un milieu 0,5X MS. Ces plantules sont issues de graines dont la plante parentale était hétérozygote pour le transgène, on a donc un mélange de plantules hétérozygotes ( $\frac{3}{4}$ ) et homozygote ( $\frac{1}{4}$ ) pour le transgène. Entre 5 et 13 plantules entières (typiquement entre 10 et 45 mg de masse fraîche) ont été récoltées pour chaque réplicat de chaque génotype, congelées dans de l'azote liquide et broyées à l'aide d'un broyeur de tissus (broyeur MM30, Retsch/Qiagen). Les ARNs ont été extraits avec le kit RNeasy Plant Mini (Qiagen), avec une étape de digestion à la DNase Ambion turbo DNA-free (Invitrogen) pour éviter les contaminations par de l'ADN.

#### RT-PCR

Après quantification par mesure d'absorbance au NanoDrop-2000 (Thermo Scientific). 1 µg d'ARN est incubé avec 2 µL d'oligo dT (1 µg/µL) pendant 5 minutes à 70°C, puis laissé- à reposer 5 minutes dans la glace. Les ADNc sont synthétisés avec 200 unités de Reverse Transcriptase M-MLV (Promega), 1 µL d'inhibiteur de ribonucléase (RNasin, Promega), 0,5 mM de dNTPs et le tampon M-MLV 1X, pendant 1h à 40°C, suivi de 15 minutes à 70°C. 90 µL d'eau sont ensuite ajoutés à la réaction.

#### PCR quantitative

Les oligonucléotides utilisés pour la PCR quantitative en temps réel ont été dessinés à l'aide du serveur QUANTPRIME ([www.quantprime.de/](http://www.quantprime.de/)) et sont listés dans le tableau ci-dessous. Ces oligonucléotides ont été publiés (Chahtane *et al.*, 2013).

Nom	Gène correspondant	Séquence
<i>oGV38</i>	<i>SAND</i> (AT2G28390)	AACTCTATGCAGCATTGATCCACT
<i>oGV39</i>		TGATTGCATATCTTTATCGCCATC
<i>oHC168</i>	<i>AP1</i> (AT1G69120)	TGCACCTGAGTCCGACGTCAATAC
<i>oHC169</i>		GCTCATTGCTTGCAAGTCTTCCC

Les réactions PCR ont été faites à l'aide d'un thermocycleur Rotor Gene 3000 (Qiagen) configuré avec les paramètres par défaut. 0,1 µM de chacun des deux oligonucléotides, 1X de mix SYBR Green JumpStart (Sigma-Aldrich) et 1 µL d'ADNc (produit issu de la RT-PCR

décrite ci-dessus) dilué de 4 à 512 fois ont été mélangés dans un volume final de 15  $\mu$ L. L'expression relative du gène AP1 (AT1G69120) pour le génotype analysé par rapport à l'expression relative chez une plante Col-0 a été calculée par la méthode  $\Delta\Delta$ Ct décrite par Pfaffl, 2001, en utilisant le gène *SAND* (AT2G28390) comme référence (Czechowski *et al.*, 2005; Guenin *et al.*, 2009). Les analyses des données ont été faites avec le logiciel REST 2009 (Pfaffl *et al.*, 2002).

## **5. Caractérisation des interactions ADN-protéine chez *A. thaliana* par ChIP-seq**

Ces analyses sont décrites dans l'Article 1.



## REFERENCES

- Adkins, N.L., Hagerman, T.A., Georgel, P.** (2006). GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol* *84*, 559-567.
- Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q.G.S., Roodveldt, C., Tawfik, D.S.** (2005). The 'evolvability' of promiscuous protein functions. *Nat Genet* *37*, 73-76.
- Airoidi, C.A., Davies, B.** (2012). Gene duplication and the evolution of plant MADS-box transcription factors. *J Genet Genomics* *39*, 157-165.
- Alonso, C.R., Wilkins, A.S.** (2005). The molecular elements that underlie developmental evolution. *Nat Rev Genet* *6*, 709-715.
- Alvarez-Venegas, R., Pien, S., Sadler, M., Witmer, X., Grossniklaus, U., Avramova, Z.** (2003). ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Curr Biol* *13*, 627-637.
- Amasino, R.** (2010). Seasonal and developmental timing of flowering. *Plant J* *61*, 1001-1013.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., Iyer, L.M.** (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* *29*, 231-262.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N.** (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* *38*, W529-533.
- Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., Teichmann, S.A.** (2004). Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* *14*, 283-291.
- Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R., Bulyk, M.L.** (2009). Diversity and complexity in DNA recognition by transcription factors. *Science* *324*, 1720-1723.
- Bailey, T.L., Elkan, C.** (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* *2*, 28-36.
- Baker, C.R., Tuch, B.B., Johnson, A.D.** (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proceedings of the National Academy of Sciences of the United States of America* *108*, 7493-7498.
- Baron, M.K., Boeckers, T.M., Vaida, B., Faham, S., Gingery, M., Sawaya, M.R., Salyer, D., Gundelfinger, E.D., Bowie, J.U.** (2006). An architectural framework that may lie at the core of the postsynaptic density. *Science* *311*, 531-535.
- Barrera, F.N., Poveda, J.A., Gonzalez-Ros, J.M., Neira, J.L.** (2003). Binding of the C-terminal sterile alpha motif (SAM) domain of human p73 to lipid membranes. *J Biol Chem* *278*, 46878-46885.
- Barrera, L.O., Ren, B.** (2006). The transcriptional regulatory code of eukaryotic cells--insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr Opin Cell Biol* *18*, 291-298.
- Bell, O., Tiwari, V.K., Thoma, N.H., Schubeler, D.** (2011). Determinants and dynamics of genome accessibility. *Nat Rev Genet* *12*, 554-564.
- Benlloch, R., Kim, M.C., Sayou, C., Thevenon, E., Parcy, F., Nilsson, O.** (2011). Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. *Plant J* *67*, 1094-1102.
- Bensmihen, S., To, A., Lambert, G., Kroj, T., Giraudat, J., Parcy, F.** (2004). Analysis of an activated ABI5 allele using a new selection method for transgenic Arabidopsis seeds. *FEBS Lett* *561*, 127-131.
- Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S.A., Morris, Q.D., Bulyk, M.L., Hughes, T.R.** (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* *133*, 1266-1276.
- Bonin, I., Muhlberger, R., Bourenkov, G.P., Huber, R., Bacher, A., Richter, G., Wahl, M.C.** (2004). Structural basis for the interaction of Escherichia coli NusA with protein N of phage lambda. *Proc Natl Acad Sci U S A* *101*, 13762-13767.
- Bowman, J.L., Meyerowitz, E.M.** (1991). Genetic control of pattern formation during flower development in Arabidopsis. *Symp Soc Exp Biol* *45*, 89-115.

- Bowman, J.L., Smyth, D.R., Meyerowitz, E.M.** (1989). Genes directing flower development in *Arabidopsis*. *Plant Cell* 1, 37-52.
- Boyle, A.P., Guinney, J., Crawford, G.E., Furey, T.S.** (2008). F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* 24, 2537-2538.
- Bradford, M.M.** (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-254.
- Bricogne, G.** (1993). Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. *Acta Crystallogr D Biol Crystallogr* 49, 37-60.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonrhein, C., Womack, T.O.** (2011). BUSTER version 2.8.0 Cambridge, United Kingdom: Global Phasing Ltd.
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W.** (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* 59, 2023-2030.
- Bridgham, J.T., Eick, G.N., Larroux, C., Deshpande, K., Harms, M.J., Gauthier, M.E., Ortlund, E.A., Degnan, B.M., Thornton, J.W.** (2010). Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS biology* 8.
- Bulyk, M.L.** (2003). Computational prediction of transcription-factor binding site locations. *Genome Biol* 5, 201.
- Busch, M.A., Bomblies, K., Weigel, D.** (1999). Activation of a floral homeotic gene in *Arabidopsis*. *Science* 285, 585-587.
- Calonje, M., Sanchez, R., Chen, L., Sung, Z.R.** (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in *Arabidopsis*. *Plant Cell* 20, 277-291.
- Carles, C.C., Fletcher, J.C.** (2009). The SAND domain protein ULTRAPETALA1 acts as a trithorax group factor to regulate cell fate in plants. *Genes Dev* 23, 2723-2728.
- Carles, C.C., Fletcher, J.C.** (2010). Missing links between histones and RNA Pol II arising from SAND? *Epigenetics* 5, 381-385.
- Carpenter, R., Coen, E.S.** (1990). Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. *Genes Dev* 4, 1483-1493.
- Carroll, S.B.** (2000). Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101, 577-580.
- Carroll, S.B.** (2001). Chance and necessity: the evolution of morphological complexity and diversity. *Nature* 409, 1102-1109.
- Carroll, S.B.** (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25-36.
- Castel, S.E., Martienssen, R.A.** (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* 14, 100-112.
- Causier, B., Schwarz-Sommer, Z., Davies, B.** (2010). Floral organ identity: 20 years of ABCs. *Semin Cell Dev Biol* 21, 73-79.
- Chae, E., Tan, Q.K., Hill, T.A., Irish, V.F.** (2008). An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development* 135, 1235-1245.
- Chahtane, H., Vachon, G., Le Masson, M., Thevenon, E., Perigon, S., Mihajlovic, N., Kalinina, A., Michard, R., Moyroud, E., Monniaux, M., Sayou, C., Grbic, V., Parcy, F., Tichtinsky, G.** (2013). A variant of LEAFY reveals its capacity to stimulate meristem development by inducing RAX1. *Plant J* 74, 678-689.
- Chandler, J.W.** (2012). Floral meristem initiation and emergence in plants. *Cell Mol Life Sci* 69, 3807-3818.
- Chen, L., Cheng, J.C., Castle, L., Sung, Z.R.** (1997). EMF genes regulate *Arabidopsis* inflorescence development. *Plant Cell* 9, 2011-2024.
- Chen, L., Glover, J.N., Hogan, P.G., Rao, A., Harrison, S.C.** (1998). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* 392, 42-48.
- Chi, S.W., Ayed, A., Arrowsmith, C.H.** (1999). Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain. *Embo J* 18, 4438-4445.
- Chou, M.L., Haung, M.D., Yang, C.H.** (2001). EMF genes interact with late-flowering genes in regulating floral initiation genes during shoot development in *Arabidopsis thaliana*. *Plant Cell Physiol* 42, 499-507.



- Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K.S.** (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 9, 279-289.
- Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., Zaret, K.S.** (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *Embo J* 17, 244-254.
- Clapier, C.R., Cairns, B.R.** (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78, 273-304.
- Clark, K.L., Halay, E.D., Lai, E., Burley, S.K.** (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364, 412-420.
- Coen, E.S., Meyerowitz, E.M.** (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31-37.
- Coen, E.S., Roemro, J.M., Doyle, S., Elliott, R., Murphy, G., Carpenter, R.** (1990). *Floricaula*: A homeotic gene required for flower development in *Antirrhinum majus*. *Cell* 63, 1311-1322.
- Cohn, M.J., Tickle, C.** (1999). Developmental basis of limblessness and axial patterning in snakes. *Nature* 399, 474-479.
- Cove, D.** (2005). The moss *Physcomitrella patens*. *Annu Rev Genet* 39, 339-358.
- Cowtan, K.** (2002). Generic representation and evaluation of properties as a function of position in reciprocal space. *J Appl Cryst* 35, 655-663.
- Cowtan, K.** (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62, 1002-1011.
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., Scheible, W.R.** (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol* 139, 5-17.
- de Wit, E., de Laat, W.** (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26, 11-24.
- Djordjevic, M.** (2007). SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. *Biomol Eng* 24, 179-189.
- Doebley, J.F., Gaut, B.S., Smith, B.D.** (2006). The molecular genetics of crop domestication. *Cell* 127, 1309-1321.
- Dumas, R., Joyard, J., Douce, R.** (1989). Purification and characterization of acetohydroxyacid reductoisomerase from spinach chloroplasts. *Biochem J* 262, 971-976.
- Dummler, A., Lawrence, A.M., de Marco, A.** (2005). Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb Cell Fact* 4, 34.
- Edwards, K., Johnstone, C., Thompson, C.** (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic acids research* 19, 1349.
- Eick, G.N., Colucci, J.K., Harms, M.J., Ortlund, E.A., Thornton, J.W.** (2012). Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS genetics* 8, e1003072.
- Emamzadah, S., Tropicia, L., Halazonetis, T.D.** (2011). Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-response element. *Mol Cancer Res* 9, 1493-1499.
- Emsley, P., Cowtan, K.** (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Feng, S., Jacobsen, S.E., Reik, W.** (2010). Epigenetic reprogramming in plant and animal development. *Science* 330, 622-627.
- Finet, C., Timme, R.E., Delwiche, C.F., Marletaz, F.** (2010). Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Curr Biol* 20, 2217-2222.
- Fisher, S., Franz-Odenaal, T.** (2012). Evolution of the bone gene regulatory network. *Current opinion in genetics & development* 22, 390-397.
- Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S., Mitchell, E.** (2010). The ID23-2 structural biology microfocuss beamline at the ESRF. *J Synchrotron Radiat* 17, 107-118.
- Fraser, J.S., Gross, J.D., Krogan, N.J.** (2013). From systems to structure: bridging networks and mechanism. *Mol Cell* 49, 222-231.
- Freed-Pastor, W.A., Prives, C.** (2012). Mutant p53: one name, many proteins. *Genes Dev* 26, 1268-1286.

- Friedman, W.E.** (2009). The meaning of Darwin's 'abominable mystery'. *Am J Bot* 96, 5-21.
- Frohlich, M.W.** (2003). An evolutionary scenario for the origin of flowers. *Nat Rev Genet* 4, 559-566.
- Frohlich, M.W., Chase, M.W.** (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184-1189.
- Gajiwala, K.S., Burley, S.K.** (2000). Winged helix proteins. *Curr Opin Struct Biol* 10, 110-116.
- Gerard, F.C., Ribeiro Ede, A., Jr., Albertini, A.A., Gutsche, I., Zaccari, G., Ruigrok, R.W., Jamin, M.** (2007). Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* 46, 10328-10338.
- Gietl, A., Grohmann, D.** (2013). Modern biophysical approaches probe transcription-factor-induced DNA bending and looping. *Biochem Soc Trans* 41, 368-373.
- Glover, J., Harrison, S.** (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* 373, 257-261.
- Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., Coupland, G.** (1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature* 386, 44-51.
- Goroncy, A., Kigawa, T., Koshiba, S., Kobayashi, N., Tochio, N., Inoue, M., Yokoyama, S.** Solution structure of the Sterile Alpha Motif (SAM) domain of mouse SAMS1. non publié.
- Green, S.M., Coyne, H.J., 3rd, McIntosh, L.P., Graves, B.J.** (2010). DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem* 285, 18496-18504.
- Guenin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., Gutierrez, L.** (2009). Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J Exp Bot* 60, 487-493.
- Guertin, M.J., Lis, J.T.** (2012). Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr Opin Genet Dev* 23, 116-123.
- Hamès, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F., Martiel, J.L., Benlloch, R., Parcy, F., Muller, C.W.** (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J* 27, 2628-2637.
- Hampsey, M., Singh, B.N., Ansari, A., Laine, J.P., Krishnamurthy, S.** (2011). Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv Enzyme Regul* 51, 118-125.
- Harada, B.T., Knight, M.J., Imai, S., Qiao, F., Ramachander, R., Sawaya, M.R., Gingery, M., Sakane, F., Bowie, J.U.** (2008). Regulation of enzyme localization by polymerization: polymer formation by the SAM domain of diacylglycerol kinase delta1. *Structure* 16, 380-387.
- Himi, S., Sano, R., Nishiyama, T., Tanahashi, T., Kato, M., Ueda, K., Hasebe, M.** (2001). Evolution of MADS-box gene induction by FLO/LFY genes. *J Mol Evol* 53, 387-393.
- Hoekstra, H.E., Coyne, J.A.** (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61, 995-1016.
- Hollenhorst, P.C., McIntosh, L.P., Graves, B.J.** (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* 80, 437-471.
- Hosoda, M., Ozaki, T., Miyazaki, K., Hayashi, S., Furuya, K., Watanabe, K., Nakagawa, T., Hanamoto, T., Todo, S., Nakagawara, A.** (2005). UFD2a mediates the proteasomal turnover of p73 without promoting p73 ubiquitination. *Oncogene* 24, 7156-7169.
- Hsia, C.C., McGinnis, W.** (2003). Evolution of transcription factor function. *Current opinion in genetics & development* 13, 199-206.
- Huala, E., Sussex, I.M.** (1992). LEAFY interacts with Floral Homeotic Genes to Regulate Arabidopsis Floral Development. *Plant Cell* 4, 901-903.
- Ikeda-Kawakatsu, K., Maekawa, M., Izawa, T., Itoh, J., Nagato, Y.** (2012). ABERRANT PANICLE ORGANIZATION 2/RFL, the rice ortholog of Arabidopsis LEAFY, suppresses the transition from inflorescence meristem to floral meristem through interaction with APO1. *Plant J* 69, 168-180.
- Irish, V.F.** (2010). The flowering of Arabidopsis flower development. *Plant J* 61, 1014-1028.
- Jaeger, K.E., Pullen, N., Lamzin, S., Morris, R.J., Wigge, P.A.** (2013). Interlocking feedback loops govern the dynamic behavior of the floral transition in Arabidopsis. *Plant Cell* 25, 820-833.
- Katsani, K.R., Hajibagheri, M.A., 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.

- Katz, A., Oliva, M., Mosquna, A., Hakim, O., Ohad, N.** (2004). FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. *Plant J* 37, 707-719.
- Kaufmann, K., Wellmer, F., Muino, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueno, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C., Riechmann, J.L.** (2010). Orchestration of floral initiation by APETALA1. *Science* 328, 85-89.
- Khersonsky, O., Roodveldt, C., Tawfik, D.S.** (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. *Current opinion in chemical biology* 10, 498-508.
- Kim, C.A., Gingery, M., Pilpa, R.M., Bowie, J.U.** (2002). The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Biol* 9, 453-457.
- Kim, C.A., Phillips, M.L., Kim, W., Gingery, M., Tran, H.H., Robinson, M.A., Faham, S., Bowie, J.U.** (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *Embo J* 20, 4173-4182.
- Kim, C.A., Sawaya, M.R., Cascio, D., Kim, W., Bowie, J.U.** (2005). Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *J Biol Chem* 280, 27769-27775.
- Knight, M.J., Leettola, C., Gingery, M., Li, H., Bowie, J.U.** (2011). A human sterile alpha motif domain polymerizome. *Protein Sci* 20, 1697-1706.
- Krivega, I., Dean, A.** (2012). Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev* 22, 79-85.
- Krizek, B.A., Fletcher, J.C.** (2005). Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet* 6, 688-698.
- Krylov, D., Vinson, C.** (2001). Leucine Zipper. *Encyclopedia of life sciences*.
- Lamb, R.S., Hill, T.A., Tan, Q.K., Irish, V.F.** (2002). Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development* 129, 2079-2086.
- Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G., Cremer, T.** (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* 8, 104-115.
- Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D.M., Correa, L.G., Reski, R., Mueller-Roeber, B., Rensing, S.A.** (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol Evol* 2, 488-503.
- Levin, J.Z., Meyerowitz, E.M.** (1995). *UFO*: an Arabidopsis gene involved in both floral meristem and floral organ development. *Plant Cell* 7, 529-548.
- Levine, M., Tjian, R.** (2003). Transcription regulation and animal diversity. *Nature* 424, 147-151.
- Li, W., Zhou, Y., Liu, X., Yu, P., Cohen, J.D., Meyerowitz, E.M.** (2013). LEAFY controls auxin response pathways in floral primordium formation. *Sci Signal* 6, ra23.
- Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S., Yanofsky, M.F.** (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* 11, 1007-1018.
- Liu, C., Xi, W., Shen, L., Tan, C., Yu, H.** (2009). Regulation of floral patterning by flowering time genes. *Dev Cell* 16, 711-722.
- Liu, G., Chen, X.** (2005). The C-terminal sterile alpha motif and the extreme C terminus regulate the transcriptional activity of the alpha isoform of p73. *J Biol Chem* 280, 20111-20119.
- Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R., Weigel, D.** (2001). A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* 105, 793-803.
- Maizel, A., Busch, M.A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., Weigel, D.** (2005). The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260-263.
- Maizel, A., Weigel, D.** (2004). Temporally and spatially controlled induction of gene expression in Arabidopsis thaliana. *Plant J* 38, 164-171.
- Man, T.K., Stormo, G.D.** (2001). Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. *Nucleic Acids Res* 29, 2471-2478.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J.** (2007). Phaser crystallographic software. *Journal of applied crystallography* 40, 658-674.
- Meyerowitz, E.M.** (1987). Arabidopsis thaliana. *Annu Rev Genet* 21, 93-111.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R., Parcy, F.** (2010). LEAFY blossoms. *Trends Plant Sci* 15, 346-352.

- Moyroud, E., Minguet, E.G., Ott, F., Yant, L., Pose, D., Monniaux, M., Blanchet, S., Bastien, O., Thevenon, E., Weigel, D., Schmid, M., Parcy, F.** (2011). Prediction of Regulatory Interactions from Genome Sequences Using a Biophysical Model for the Arabidopsis LEAFY Transcription Factor. *Plant Cell*.
- Moyroud, E., Tichtinsky, G., Parcy, F.** (2009). The LEAFY floral regulators in Angiosperms: Conserved proteins with diverse roles. *J Plant Biol* 52.
- Neaves, K.J., Cooper, L.P., White, J.H., Carnally, S.M., Dryden, D.T., Edwardson, J.M., Henderson, R.M.** (2009). Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping. *Nucleic Acids Res* 37, 2053-2063.
- Nickrent, D.L., Parkinson, C.L., Palmer, J.D., Duff, R.J.** (2000). Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular biology and evolution* 17, 1885-1895.
- Parcy, F.** (2005). Flowering: a time for integration. *Int J Dev Biol* 49, 585-593.
- Parcy, F., Bomblies, K., Weigel, D.** (2002). Interaction of LEAFY, AGAMOUS and TERMINAL FLOWER1 in maintaining floral meristem identity in Arabidopsis. *Development* 129, 2519-2527.
- Parcy, F., Nilsson, O., Busch, M.A., Lee, I., Weigel, D.** (1998). A genetic framework for floral patterning. *Nature* 395, 561-566.
- Pennisi, E.** (2008). Evolutionary biology. Deciphering the genetics of evolution. *Science* 321, 760-763.
- Pennisi, E.** (2009). Origins. On the origin of flowering plants. *Science* 324, 28-31.
- Peter, I.S., Davidson, E.H.** (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* 144, 970-985.
- Petty, T.J., Emamzadah, S., Costantino, L., Petkova, I., Stavridi, E.S., Saven, J.G., Vauthey, E., Halazonetis, T.D.** (2011). An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *Embo J* 30, 2167-2176.
- Pfaffl, M.W.** (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research* 29, e45.
- Pfaffl, M.W., Horgan, G.W., Dempfle, L.** (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research* 30, e36.
- Posé, D., Yant, L., Schmid, M.** (2012). The end of innocence: flowering networks explode in complexity. *Curr Opin Plant Biol* 15, 45-50.
- Prud'homme, B., Gompel, N., Carroll, S.B.** (2007). Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A* 104 Suppl 1, 8605-8612.
- Qiao, F., Bowie, J.U.** (2005). The many faces of SAM. *Sci STKE* 2005, re7.
- Qiao, F., Harada, B., Song, H., Whitelegge, J., Courey, A.J., Bowie, J.U.** (2006). Mae inhibits Pointed-P2 transcriptional activity by blocking its MAPK docking site. *Embo J* 25, 70-79.
- Qiao, F., Song, H., Kim, C.A., Sawaya, M.R., Hunter, J.B., Gingery, M., Rebay, I., Courey, A.J., Bowie, J.U.** (2004). Derepression by depolymerization; structural insights into the regulation of Yan by Mae. *Cell* 118, 163-173.
- Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovskaya, O., Lee, J., Kent, L., Rest, J., Estabrook, G.F., Hendry, T.A., Taylor, D.W., Testa, C.M., Ambros, M., Crandall-Stotler, B., Duff, R.J., Stech, M., Frey, W., Quandt, D., Davis, C.C.** (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci U S A* 103, 15511-15516.
- Rajakulendran, T., Sahmi, M., Kurinov, I., Tyers, M., Therrien, M., Sicheri, F.** (2008). CNK and HYP form a discrete dimer by their SAM domains to mediate RAF kinase signaling. *Proc Natl Acad Sci U S A* 105, 2836-2841.
- Ramakrishnan, V., Finch, J., Graziano, V., Lee, P., Sweet, R.** (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 18, 219-223.
- Ratcliffe, O.J., Amaya, I., Vincent, C.A., Rothstein, S., Carpenter, R., Coen, E.S., Bradley, D.J.** (1998). A common mechanism controls the life cycle and architecture of plants. *Development* 125, 1609-1615.
- Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., Yu, G.** (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290, 2105-2110.
- Robin, A.Y., Cobessi, D., Curien, G., Robert-Genthon, M., Ferrer, J.L., Dumas, R.** (2010). A new mode of dimerization of allosteric enzymes with ACT domains revealed by the crystal structure of the aspartate kinase from Cyanobacteria. *J Mol Biol* 399, 283-293.

- Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., Mann, R.S.** (2010). Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* 79, 233-269.
- Ronshaugen, M., McGinnis, N., McGinnis, W.** (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* 415, 914-917.
- Ruiz-Garcia, L., Madueno, F., Wilkinson, M., Haughn, G., Salinas, J., Martinez-Zapater, J.M.** (1997). Different roles of flowering-time genes in the activation of floral initiation genes in Arabidopsis. *Plant Cell* 9, 1921-1934.
- Sabin, L.R., Delas, M.J., Hannon, G.J.** (2013). Dogma derailed: the many influences of RNA on the genome. *Mol Cell* 49, 783-794.
- Saleh, A., Al-Abdallat, A., Ndamukong, I., Alvarez-Venegas, R., Avramova, Z.** (2007). The Arabidopsis homologs of trithorax (ATX1) and enhancer of zeste (CLF) establish 'bivalent chromatin marks' at the silent AGAMOUS locus. *Nucleic Acids Res* 35, 6290-6296.
- Sathyamurthy, A., Freund, S.M., Johnson, C.M., Allen, M.D., Bycroft, M.** (2011). Structural basis of p63alpha SAM domain mutants involved in AEC syndrome. *FEBS J* 278, 2680-2688.
- Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D., Lohmann, J.U.** (2003). Dissection of floral induction pathways using global expression analysis. *Development* 130, 6001-6012.
- Schultz, E.A., Haughn, G.W.** (1991). *LEAFY*, a homeotic gene that regulates inflorescence development in Arabidopsis. *Plant Cell* 3, 771-781.
- Seldeen, K., McDonald, C., Deegan, B., Farooq, A.** (2008). Coupling of folding and DNA-binding in the bZIP domains of Jun-Fos heterodimeric transcription factor. *Arch Biochem Biophys* 473, 48-60.
- Sexton, T., Schober, H., Fraser, P., Gasser, S.M.** (2007). Gene regulation through nuclear organization. *Nat Struct Mol Biol* 14, 1049-1055.
- Sheldrick, G.M.** (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* 66, 479-485.
- Sikorski, T., Buratowski, S.** (2009). The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* 21, 344-351.
- Simpson, G.G., Dean, C.** (2002). Arabidopsis, the Rosetta stone of flowering time? *Science* 296, 285-289.
- Siriwardana, N.S., Lamb, R.S.** (2012). A conserved domain in the N-terminus is important for LEAFY dimerization and function in Arabidopsis thaliana. *Plant J* 71, 736-749.
- Smaczniak, C., Immink, R.G., Angenent, G.C., Kaufmann, K.** (2012a). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139, 3081-3098.
- Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Percy, F., Xu, L., Carles, C.C., Angenent, G.C., Kaufmann, K.** (2012b). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc Natl Acad Sci U S A* 109, 1560-1565.
- Souer, E., Rebocho, A.B., Bliiek, M., Kusters, E., de Bruin, R.A., Koes, R.** (2008). Patterning of inflorescences and flowers by the F-Box protein DOUBLE TOP and the LEAFY homolog ABERRANT LEAF AND FLOWER of petunia. *Plant Cell* 20, 2033-2048.
- Stafford, R.L., Hinde, E., Knight, M.J., Pennella, M.A., Ear, J., Digman, M.A., Gratton, E., Bowie, J.U.** (2011). Tandem SAM domain structure of human Caskin1: a presynaptic, self-assembling scaffold for CASK. *Structure* 19, 1826-1836.
- Stapleton, D., Balan, I., Pawson, T., Sicheri, F.** (1999). The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nat Struct Biol* 6, 44-49.
- Stoltenburg, R., Reinemann, C., Strehlitz, B.** (2007). SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng* 24, 381-403.
- Stormo, G.D.** (2000). DNA binding sites: representation and discovery. *Bioinformatics* 16, 16-23.
- Studer, A., Zhao, Q., Ross-Ibarra, J., Doebley, J.** (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat Genet* 43, 1160-1163.
- Tanahashi, T., Sumikawa, N., Kato, M., Hasebe, M.** (2005). Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* 132, 1727-1736.

- Taube, J.H., Allton, K., Duncan, S.A., Shen, L., Barton, M.C.** (2010). Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells. *J Biol Chem* **285**, 16135-16144.
- Thanos, C.D., Goodwill, K.E., Bowie, J.U.** (1999). Oligomeric structure of the human EphB2 receptor SAM domain. *Science* **283**, 833-836.
- Theissen, G.** (2001). Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* **4**, 75-85.
- Theissen, G., Melzer, R.** (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann Bot (London)* **100**, 603-619.
- Theissen, G., Saedler, H.** (2001). Floral quartets. *Nature* **409**, 469-471.
- Timme, R.E., Bachvaroff, T.R., Delwiche, C.F.** (2012). Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* **7**, e29696.
- Tran, H.H., Kim, C.A., Faham, S., Siddall, M.C., Bowie, J.U.** (2002). Native interface of the SAM domain polymer of TEL. *BMC Struct Biol* **2**, 5.
- Wagner, D.** (2009). Flower morphogenesis: timing is key. *Dev Cell* **16**, 621-622.
- Wagner, D., Meyerowitz, E.M.** (2002). SPLAYED, a novel SWI/SNF ATPase homolog, controls reproductive development in Arabidopsis. *Curr Biol* **12**, 85-94.
- Wagner, D., Sablowski, R.W., Meyerowitz, E.M.** (1999). Transcriptional activation of *APETALA1* by *LEAFY*. *Science* **285**, 582-584.
- Wagner, D., Wellmer, F., Dilks, K., William, D., Smith, M.R., Kumar, P.P., Riechmann, J.L., Greenland, A.J., Meyerowitz, E.M.** (2004). Floral induction in tissue culture: a system for the analysis of *LEAFY*-dependent gene regulation. *Plant J* **39**, 273-282.
- Wagner, G.P., Lynch, V.J.** (2008). The gene regulatory logic of transcription factor evolution. *Trends Ecol Evol* **23**, 377-385.
- Walker, J.R., Yermekbayeva, L., Butler-Cole, C., Weigelt, J., Bountra, C., Arrowsmith, C.H., Edwards, A.M., Bochkarev, A., Dhe-Paganon, S.** Co-Crystal Structure of the SAM Domains of Human Ephrin Type-A Receptor 1 and Human Ephrin Type-A Receptor 2. non publié.
- Wang, H., Dodd, I.B., Dunlap, D.D., Shearwin, K.E., Finzi, L.** (2013). Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res* **41**, 5746-5756.
- Wasserman, W.W., Sandelin, A.** (2004). Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* **5**, 276-287.
- Weake, V.M., Workman, J.L.** (2010). Inducible gene expression: diverse regulatory mechanisms. *Nat Rev Genet* **11**, 426-437.
- Webber, J.L., Zhang, J., Cote, L., Vivekanand, P., Ni, X., Zhou, J., Negre, N., Carthew, R.W., White, K.P., Rebay, I.** (2013). The relationship between long-range chromatin occupancy and polymerization of the Drosophila ETS family transcriptional repressor Yan. *Genetics* **193**, 633-649.
- Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., Meyerowitz, E.M.** (1992). *LEAFY* controls floral meristem identity in Arabidopsis. *Cell* **69**, 843-859.
- Weigel, D., Meyerowitz, E.M.** (1993). Activation of floral homeotic genes in Arabidopsis. *Science* **261**, 1723-1726.
- Weigel, D., Nilsson, O.** (1995). A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495-500.
- Wellmer, F., Riechmann, J.L.** (2010). Gene networks controlling the initiation of flower development. *Trends Genet* **26**, 519-527.
- Williams, H., Jayaraman, P.S., Gaston, K.** (2008). DNA wrapping and distortion by an oligomeric homeodomain protein. *J Mol Biol* **383**, 10-23.
- Winter, C.M., Austin, R.S., Blanvillain-Baufume, S., Reback, M.A., Monniaux, M., Wu, M.F., Sang, Y., Yamaguchi, A., Yamaguchi, N., Parker, J.E., Parcy, F., Jensen, S.T., Li, H., Wagner, D.** (2011). *LEAFY* Target Genes Reveal Floral Regulatory Logic, cis Motifs, and a Link to Biotic Stimulus Response. *Dev Cell* **20**, 430-443.
- Wodniok, S., Brinkmann, H., Glockner, G., Heidel, A.J., Philippe, H., Melkonian, M., Becker, B.** (2011). Origin of land plants: do conjugating green algae hold the key? *BMC evolutionary biology* **11**, 104.
- Wray, G.A.** (2007). The evolutionary significance of cis-regulatory mutations. *Nature reviews Genetics* **8**, 206-216.

- Wu, M.F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S.K., Li, Z., Su, Y., Slewinski, T.L., Wagner, D.** (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc Natl Acad Sci U S A* *109*, 3576-3581.
- Wu, X., Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C., Weigel, D.** (2003). Modes of intercellular transcription factor movement in the Arabidopsis apex. *Development* *130*, 3735-3745.
- Yadon, A.N., Tsukiyama, T.** (2013). DNA looping-dependent targeting of a chromatin remodeling factor. *Cell Cycle* *12*, 1809-1810.
- Yamaguchi, N., Wu, M.F., Winter, C.M., Berns, M.C., Nole-Wilson, S., Yamaguchi, A., Coupland, G., Krizek, B.A., Wagner, D.** (2013). A molecular framework for auxin-mediated initiation of flower primordia. *Dev Cell* *24*, 271-282.
- Yan, J., Pan, L., Chen, X., Wu, L., Zhang, M.** (2010). The structure of the harmonin/sans complex reveals an unexpected interaction mode of the two Usher syndrome proteins. *Proc Natl Acad Sci U S A* *107*, 4040-4045.
- Zaret, K.S., Carroll, J.S.** (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* *25*, 2227-2241.
- Zhang, J., Graham, T.G., Vivekanand, P., Cote, L., Cetera, M., Rebay, I.** (2010). Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the Drosophila ETS family transcriptional repressor Yan. *Mol Cell Biol* *30*, 1158-1170.
- Zhang, R., Su, B.** (2009). Small but influential: the role of microRNAs on gene regulatory network and 3'UTR evolution. *Journal of genetics and genomics = Yi chuan xue bao* *36*, 1-6.
- Zhang, W., Zhang, T., Wu, Y., Jiang, J.** (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *Plant Cell* *24*, 2719-2731.
- Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W., Chua, N.H.** (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat Protoc* *1*, 641-646.
- Zhao, Y., Granas, D., Stormo, G.D.** (2009). Inferring binding energies from selected binding sites. *PLoS Comput Biol* *5*, e1000590.
- Zhu, X., Gerstein, M., Snyder, M.** (2007). Getting connected: analysis and principles of biological networks. *Genes Dev* *21*, 1010-1024.





## ANNEXES

**Annexe 1.** Tableau des constructions ayant permis de produire les protéines recombinantes utilisées dans le Chapitre I. Pour les mutations, la numérotation se réfère à celle de la protéine AtLFY (voir alignement Figure 8).

Espèce	Construction		mutations	Résidus	Tag
<i>A. thaliana</i>	ETH94	AtLFY		1-420	6xHis C-ter
<i>A. thaliana</i>	CA23	AtLFY <sub>TE</sub>	T75E	1-420	6xHis C-ter
<i>A. thaliana</i>	CA24	AtLFY <sub>RE</sub>	R112E	1-420	6xHis C-ter
<i>A. thaliana</i>	CA25	AtLFY <sub>TERE</sub>	T75E, R112E	1-420	6xHis C-ter
<i>A. thaliana</i>	ETH79	AtLFY $\Delta$		40-420	6xHis C-ter
<i>A. thaliana</i>	ETH201	AtLFY-SAM		38-151	6xHis N-ter clivable
<i>G. biloba</i>	ETH164	GbLFY $\Delta$		55-402	6xHis N-ter clivable
<i>G. biloba</i>	CA21	GbLFY $\Delta$ <sub>TE</sub>	T75E	55-402	6xHis N-ter clivable
<i>G. biloba</i>	CA22	GbLFY $\Delta$ <sub>RE</sub>	R112E	55-402	6xHis N-ter clivable
<i>G. biloba</i>	CA39	GbLFY $\Delta$ <sub>TERE</sub>	T75E, R112E	55-402	6xHis N-ter clivable
<i>G. biloba</i>	ETH195	GbLFY-SAM		55-159	6xHis N-ter clivable
<i>G. biloba</i>	CA15	GbLFY-SAM <sub>TE</sub>	T75E	55-159	6xHis N-ter clivable
<i>G. biloba</i>	CA17	GbLFY-SAM <sub>RE</sub>	R112E	55-159	6xHis N-ter clivable
<i>G. biloba</i>	CA20	GbLFY-SAM <sub>TERE</sub>	T75E, R112E	55-159	6xHis N-ter clivable
<i>G. biloba</i>	ETH231	GbLFY-SAM		1-159	6xHis N-ter clivable
<i>P. patens</i>	ETH182	PpLFY		1-348	6xHis N-ter clivable
<i>P. patens</i>	ETH193	PpLFY-N		1-149	6xHis N-ter clivable

**Annexe 2.** Tableau des conditions positives au criblage pour la cristallisation de GbLFY-N avec le robot de l'EMBL (Grenoble), à 20°C. Les conditions choisies pour être optimisées sont notées en bleu. La condition ayant abouti à la structure est en orange. Les kits MPD Suite, Nucleix Suite, MBClass Suite sont commercialisés par Qiagen. Le kit Crystal Screen light est commercialisé par Hampton Research.

Concentration protéine et additif	Type de cristaux	Kit N° condition	Sels	Tampon	Précipitants
5,5mg/mL 20% MPD	Petits cristaux	MPD Suite 20	0.2 M Zinc sulfate		40% MPD
5,5mg/mL 20% MPD	Aiguilles	MPD Suite 52		0.1 M HEPES pH 7.0	10% MPD
5,5mg/mL	Cristaux isolés	Crystal Screen light 2			0,2M Potassium sodium tartrate tetrahydrate
5,5mg/mL 20% MPD	Nombreux cristaux isolés	Crystal Screen light 4		0,1M Tris HCl pH 8,5	1M Ammonium sulfate
5,5mg/mL	Nombreux cristaux isolés	Crystal Screen light 4		0,1M Tris HCl pH 8,5	1M Ammonium sulfate
5,5mg/mL 20% MPD	oursin	Crystal Screen light 44			0.1 M Magnesium formate dihydrate
5,5mg/mL 20% MPD	plaques	Crystal Screen light 29		0,1M Sodium HEPES pH7,5	0,4M Potassium sodium tartrate tetrahydrate
5,5mg/mL	Cristaux isolés	Crystal Screen		0,1M Sodium	2% PEG 400

20% MPD		light 39		HEPES pH7,5	1M Ammonium sulfate
5,5mg/mL	Cristaux isolés	Crystal Screen light 39		0,1M Sodium HEPES pH7,5	2% PEG 400 1M Ammonium sulfate
5,5mg/mL 20% MPD	Cristaux isolés	Crystal Screen light 16		0,1M Sodium HEPES pH7,5	0,75M lithium sulfate monohydrate
5,5mg/mL	Cristaux isolés	Crystal Screen light 16		0,1M Sodium HEPES pH7,5	0,75M lithium sulfate monohydrate
5,5mg/mL 20% MPD	Micro-cristaux	Nucleix Suite 21	0,2M Potassium chloride 0,005M Magnesium chloride	0,05M Sodium cacodylate pH6,5	10% 1,6-hexanediol
5,5mg/mL 20% MPD	Petites aiguilles	Nucleix Suite 36	0,1M Potassium chloride 0,01M Calcium chloride	0,05M Sodium HEPES pH7	10% PEG 400
5,5mg/mL 20% MPD	Micro-cristaux	Nucleix Suite 41	0,1M Potassium chloride 0,015M Magnesium chloride	0,05M Tris HCl pH7,5	10% PEG MME 550
5,5mg/mL 20% MPD	Micro-cristaux	Nucleix Suite 42	0,01M Magnesium acetate	0,05M Tris HCl pH7,5	5% Isopropanol
5,5mg/mL	Petites aiguilles	Nucleix Suite 42	0,01M Magnesium acetate	0,05M Tris HCl pH7,5	5% Isopropanol
5,5mg/mL	Petites plaques	Nucleix Suite 76	0,0025M Spermidine 0,01M Magnesium chloride	0,05M Sodium cacodylate pH6	2,5M sodium chloride
5,5mg/mL 20% MPD	Plaques	Nucleix Suite 77	0,2M tri-Na citrate 0,01M Magnesium chloride	0,05M Sodium cacodylate pH6,5	5% Isopropanol
5,5mg/mL 20% MPD	Aiguilles	Nucleix Suite 81	0,001M Spermine 0,015M Magnesium chloride	0,05M Sodium HEPES pH7,5	1M Ammonium sulfate
5,5mg/mL	Aiguilles	Nucleix Suite 83	0,001M Cobalt(III)hexamine chloride 0,2M Calcium acetate	0,05M Sodium cacodylate pH6	2M Lithium chloride
5,5mg/mL	Petites aiguilles	Nucleix Suite 93	0,005M Cobalt(III)hexamine chloride	0,05M Sodium cacodylate pH6,5	2,5M Potassium chloride
5,5mg/mL	Plaque unique	Nucleix Suite 95	0,002M Cobalt(III)hexamine chloride 0,001M Spermine 0,03M Calcium chloride	0,05M Sodium cacodylate pH6,5	2M Lithium chloride
5,5mg/mL 20% MPD	Aiguilles	MbClass Suite 1	0,1M Sodium citrate	0,1M Sodium citrate pH 5,6	10% Isopropanol
5,5mg/mL	Plaques	MbClass Suite 4		0,1M Tris HCl pH 8,5	0,5M Ammonium sulfate
5,5mg/mL	Micro-cristaux	MbClass Suite 6		0,1M ADA pH6,5	1M Ammonium sulfate
5,5mg/mL	Micro-cristaux	MbClass Suite 7		0,1M Tris HCl pH 8,5	1,2M Ammonium sulfate
5,5mg/mL	Petites aiguilles	MbClass Suite 11		0,1M ADA pH6,5	1M di-Ammonium phosphate
5,5mg/mL 20% MPD	Aiguilles	MbClass Suite 21		0,1M ADA pH6,5	12% MPD
5,5mg/mL	Micro-cristaux	MbClass Suite 27	0,1M Lithium sulfate	0,1M Sodium citrate pH5,6	4% PEG 400
5,5mg/mL 20% MPD	Plaques	MbClass Suite 28	0,3M Lithium sulfate	0,1M ADA pH6,5	4% PEG 400
5,5mg/mL	Micro-cristaux	MbClass Suite 28	0,3M Lithium sulfate	0,1M ADA pH6,5	4% PEG 400
5,5mg/mL	Micro-cristaux	MbClass Suite 35	0,1M Ammonium sulfate	0,1M Sodium HEPES pH7,5	18% PEG 400
5,5mg/mL 20% MPD	Aiguilles	MbClass Suite 88	0,1M Lithium sulfate	0,1M Sodium HEPES pH7,5	0,1M Potassium/Sodium tartrate
5,5mg/mL	Aiguilles	MbClass Suite 88	0,1M Lithium sulfate	0,1M Sodium HEPES pH7,5	0,1M Potassium/Sodium tartrate
5,5mg/mL 20% MPD	Petits cristaux	MbClass Suite 90		0,1M Sodium citrate pH5,6	0,1M Sodium chloride
5,5mg/mL 20% MPD	Micro-plaques	MbClass Suite 93	0,1M Sodium formate		2M Sodium chloride
5,5mg/mL	Micro-cristaux	MbClass Suite 93	0,1M Sodium formate		2M Sodium chloride

## Annexe 3.

the plant journal

SEB  
Society for  
Experimental Biology

The Plant Journal (2011) 67, 1094–1102

doi: 10.1111/j.1365-313X.2011.04660.x

Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of *APETALA1*Reyes Benlloch<sup>1,2,3,4,5,†</sup>, Min Chul Kim<sup>6</sup>, Camille Sayou<sup>1,2,3,4</sup>, Emmanuel Thévenon<sup>1,2,3,4</sup>, Francois Parcy<sup>1,2,3,4,\*</sup> and Ove Nilsson<sup>5</sup><sup>1</sup>CEA, iRTSV, Laboratoire Physiologie Cellulaire et Végétale, F-38054 Grenoble, France,<sup>2</sup>CNRS, UMR5168, F-38054 Grenoble, France,<sup>3</sup>Université Joseph Fourier-Grenoble I, UMR5168, F-38041 Grenoble, France,<sup>4</sup>INRA, UMR1200, F-38054 Grenoble, France,<sup>5</sup>Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 901-83 Umeå, Sweden, and<sup>6</sup>Division of Applied Life Science (BK21 Program), Plant Molecular Biology and Biotechnology Research Center, Gyeongsang National University, Jinju 660-701, Korea

Received 13 April 2011; revised 25 May 2011; accepted 27 May 2011; published online 18 July 2011.

\*For correspondence (fax +04 38 78 50 91; e-mail francois.parcy@cea.fr).

†Present address: Department of Molecular Genetics, Centre for Research in Agricultural Genomics (CRAG), Campus Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), 08193 Barcelona, Spain.

## SUMMARY

The transition to flowering in *Arabidopsis* is characterized by the sharp and localized upregulation of *APETALA1* (*AP1*) transcription in the newly formed floral primordia. Both the flower meristem-identity gene *LEAFY* (*LFY*) and the photoperiod pathway involving the *FLOWERING LOCUS T* (*FT*) and *FD* genes contribute to this upregulation. These pathways have been proposed to act independently but their respective contributions and mode of interaction have remained elusive. To address these questions, we studied the *AP1* regulatory region. Combining *in vitro* and *in vivo* approaches, we identified which of the three putative *LFY* binding sites present in the *AP1* promoter is essential for its activation by *LFY*. Interestingly, we found that this site is also important for the correct photoperiodic-dependent upregulation of *AP1*. In contrast, a previously proposed putative *FD*-binding site appears dispensable and unable to bind *FD* and we found no evidence for *FD* binding to other sites in the *AP1* promoter, suggesting that the *FT*/*FD*-dependent activation of *AP1* might be indirect. Altogether, our data give new insight into the interaction between the *FT* and *LFY* pathways in the upregulation of *AP1* transcription under long-day conditions.

**Keywords:** *APETALA1*, *LEAFY*, *FLOWERING LOCUS T*, *FLOWERING LOCUS D*, floral induction, photoperiod.

## INTRODUCTION

One of the first steps in *Arabidopsis* floral initiation is the localized upregulation of the *APETALA1* (*AP1*) gene in the newly formed floral meristems (Mandel *et al.*, 1992; Gustafson-Brown *et al.*, 1994). This upregulation marks the plant's commitment to its reproductive phase. *AP1* is regulated by several pathways. The flower meristem-identity gene *LEAFY* (*LFY*; Moyroud *et al.*, 2010) is a direct activator of *AP1* expression (Wagner *et al.*, 1999; William *et al.*, 2004). *LFY* expression is weak in young leaves during the vegetative phase and progressively increases as the plant approaches floral induction. Once it reaches a certain threshold, *LFY* induces a sharp upregulation of *AP1* (Blazquez *et al.*, 1997; Hempel *et al.*, 1997). *AP1* is also activated by the complex formed by *FLOWERING LOCUS T* (*FT*) and

*FD* (Ruiz-Garcia *et al.*, 1997; Liljegren *et al.*, 1999; Abe *et al.*, 2005; Wigge *et al.*, 2005). Since *AP1* upregulation still occurs in *ft*, *fd* and *lfy* mutant plants, and there is a strong synergistic interaction between *lfy* and *ft* mutations, *LFY* and *FT* have been proposed to act independently (Ruiz-Garcia *et al.*, 1997; Nilsson *et al.*, 1998). However, this upregulation occurs later in the single mutants compared with wild-type plants (Liljegren *et al.*, 1999; Schmid *et al.*, 2003; Wigge *et al.*, 2005) showing that each of two pathways is not sufficient to properly activate *AP1*. A better mechanistic characterization of *AP1* activation is required to fully understand this interaction.

Previous studies have shown that the most important *AP1* regulatory elements lie within 1.7 kb of its promoter, as this

DNA fragment, when driving the expression of the *AP1* cDNA, is sufficient to complement the *ap1-15* mutant phenotype (Wu *et al.*, 2003). Within this region, a LFY binding site is present (here referred to as bs1; Figure 1a), which has been validated *in vitro* (Parcy *et al.*, 1998; Hamès *et al.*, 2008). This site contains the pseudopalindromic LFY consensus sequence (CCANTG) (Parcy *et al.*, 1998; Busch *et al.*, 1999; Lohmann *et al.*, 2001). The *in vivo* relevance of bs1 has never been tested and two other sites (bs2 and bs3) conforming to the consensus are also present in its vicinity (Abe *et al.*, 2005; Wigge *et al.*, 2005). In addition to these *cis*-elements, a putative binding site for FD (C-box) has also been identified between LFY bs1 and bs2 (Wigge *et al.*, 2005). It is thus important to test whether these putative *cis*-elements are true binding sites, and, as they are in close vicinity to each other, to understand whether they act independently or interact synergistically to trigger the sharp upregulation of *AP1*.

Our analysis of these four putative *cis*-elements allowed us to establish the central importance of LFY bs1 for *AP1* expression in early floral stages and during photoperiod-dependent upregulation of *AP1*. Conversely, the C-box appears neither important nor capable of binding FD, and we could find no evidence for FD binding to other parts of the *AP1* promoter. Our results provide a deeper mechanistic

understanding of how the *LFY* and *FD/FT* pathways act both independently and together to achieve the upregulation of *AP1*.

## RESULTS

### Bs1 is the highest-affinity LFY binding site in the *AP1* promoter

Three putative LFY binding sites, bs1, bs2 and bs3, were identified in the *AP1* promoter based on the presence of the previously described consensus sequence for the LFY binding site (CCANTG) (Parcy *et al.*, 1998; Busch *et al.*, 1999) (Figure 1a,h). A refined definition of this consensus has been proposed (WNNNCCANTGG/TNNNW) based on the structural characterization of LFY/DNA complex (Hamès *et al.*, 2008) and all three putative sites conform well to this new definition (Figure 1h). As actual binding has been observed only for bs1 *in vitro*, we compared the affinity LFY-DNA binding domain (LFY-C) for bs1, bs2 and bs3 sites the using the electrophoretic mobility shift assay (EMSA) (Figure 1). LFY/DNA complex formation was detected at low LFY-C concentrations with bs1, whereas higher LFY-C concentrations were required with bs2 or bs3, illustrating that bs1 has a much higher affinity for LFY than the other putative binding sites (Figure 1b–d). Based on EMSA signal

**Figure 1.** *In vitro* analysis of LFY binding to bs1, bs2 and bs3 sites in the *AP1* promoter.

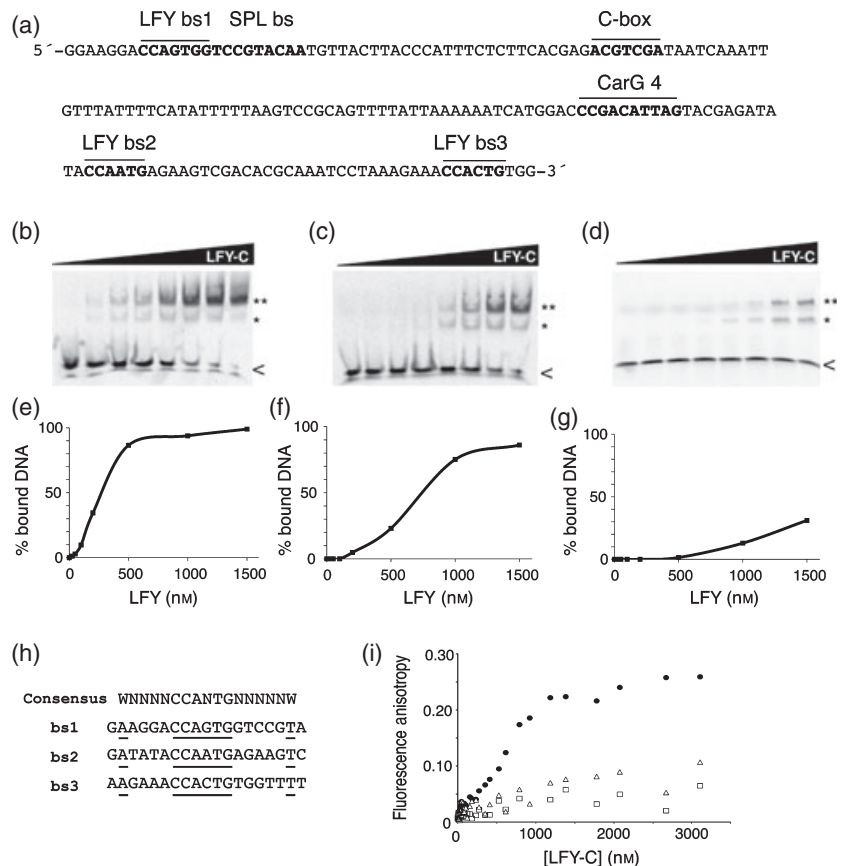
(a) DNA sequence of the *AP1* promoter with different established or putative protein binding sites.

(b–d) EMSA analysis of LFY binding to bs1 (b), bs3 (c) and bs2 (d) oligonucleotides (10 nM). Several LFY-C concentrations were used: 0, 25, 50, 100, 200, 500, 1000 and 1500 nM. Asterisks show the position of the LFY-DNA complexes: (\*) Monomer; (\*\*) Dimer; (<) Free DNA.

(e–g) Percentage of DNA bound to LFY as a function of LFY concentration. Quantifications were based on the EMSA results shown in b–d.

(h) LFY binding site consensus and the bs1, bs2 and bs3 sites sequences.

(i) Analysis of LFY binding to bs1 (●), bs2 (□) and bs3 (△) by fluorescence anisotropy.



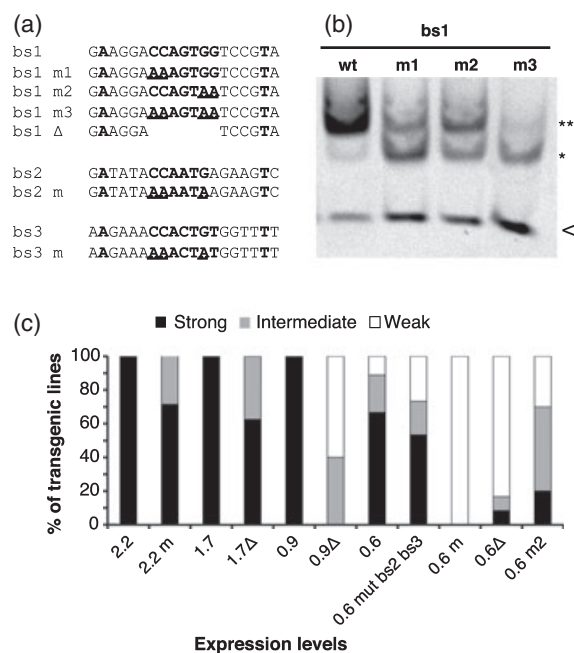
quantification, we estimated an apparent dissociation constant ( $K_{0.5}$ ) of approximately 300 nM for bs1 and approximately 750 nM for bs3, but the signal was too low for  $K_{0.5}$  to be estimated for bs2 (Figure 1e–g).

We also used fluorescent anisotropy as an alternative method to study LFY binding to bs1 *in vitro*. This technique measures the rotational mobility of fluorescently labelled DNA, a characteristic that decreases upon protein/DNA complex formation (Letilly and Royer, 1993). When compared with EMSA, this method enables to study binding in solution, avoiding any possible influence of gel migration. This experiment clearly identified differences between the three sites; the method yielded a  $K_{0.5}$  value of approximately 700 nM for bs1, whereas the binding to bs2 and bs3 was too low for  $K_{0.5}$  to be estimated (Figure 1i). Both techniques consistently showed that bs1 is the best LFY binding site in the *AP1* promoter.

To gain further insight into the importance of key nucleotides in bs1 for LFY binding, we mutagenized this site by introducing mutations either on one (m1 or m2) or both (m3) halves of the palindrome (Figure 2a). Mutating one half of bs1 resulted in a decreased affinity for LFY-C together with a compromised capacity of LFY-C to bind as a dimer (Figure 2b). The mutation affecting both sides of the palindrome dramatically reduced the LFY binding affinity and almost completely inhibited the dimer formation (Figure 2b).

#### *In vivo* activation of *AP1* by LFYVP16

Having shown that bs1 is the best LFY binding site *in vitro*, we tested its importance *in vivo*. We constructed a series of *Arabidopsis* transgenic lines that expressed the *GUS* reporter gene from various versions of the *AP1* promoter (2.2, 1.7 and 0.9 kb, with and without the bs1 site; Table S1). We verified that the *AP1::GUS* (2.2 kb) expression pattern in wild-type, *lfy* and *ft* mutant plants (Figure S1) was consistent with published results on *AP1* mRNA expression (Ruiz-Garcia *et al.*, 1997; Liljegren *et al.*, 1999; Ratcliffe *et al.*, 1999; Wagner *et al.*, 1999; Xu *et al.*, 2010). Next, we analyzed *AP1* activation by LFY in a context where no other co-activators are needed. For this, we produced transgenic plants expressing the LFYVP16 fusion protein (Parcy *et al.*, 1998) under the heat-shock inducible *HSP* promoter (*HSP::LFYVP16* plants). When fused to the VP16 activation domain, LFY possesses constitutive transcriptional activation potential and should, on its own, activate transcription once bound to a regulatory region (Parcy *et al.*, 1998). *AP1::GUS* and *HSP::LFYVP16* plants were crossed and we evaluated the effect of LFYVP16 induction on *GUS* expression in F1 seedlings. Clear *GUS* expression was detected in *HSP::LFYVP16* plants only if a heat shock was performed (Figure 3). This activation occurred with the three *AP1* promoter lengths and was observed in seedlings cotyledons, hypocotyls and roots. It was weak or absent from the young leaves and apex (Figure 3p) in spite of *HSP* promoter activity



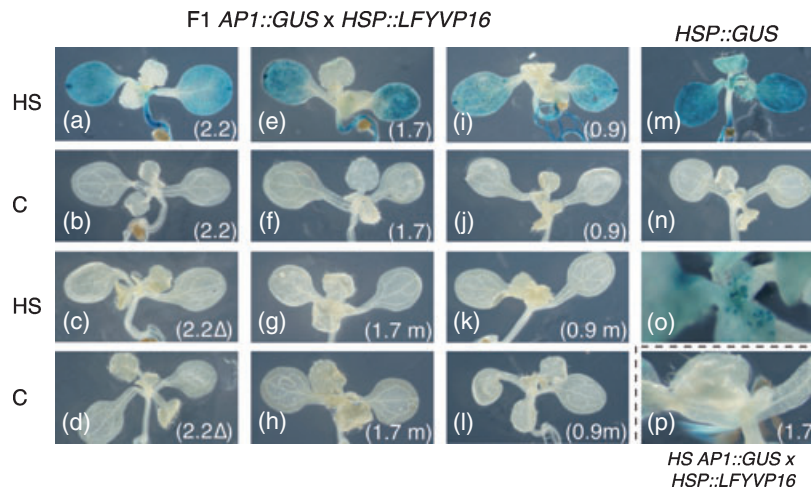
**Figure 2.** *In vitro* and *in vivo* mutation analysis of bs1-3.

(a) Alignment of wild-type and mutated binding site sequences used in the *in vitro* and *in vivo* assays. Three combinations of punctual mutations were designed to affect both sides of LFY binding site (m1, m2, and m3). A deletion covering the bs1 site (bs1 Δ) was used for *in vivo* studies. Mutations affecting both sides of the pseudo-palindrome were introduced for bs2 and bs3. (b) EMSA with wild-type and mutated versions of the bs1 oligonucleotide. LFY = 1500 nM, DNA = 10 nM. Asterisks indicate the position of the LFY-DNA complexes: (\*) Monomer; (\*\*) Dimer; (<) Free DNA. (c) *In vivo* analysis of the promoter-*GUS* fusions. Transgenic lines were generated to contain different versions of the *AP1* promoter with mutations in the LFY binding site: wild-type (2.2, 1.7, 0.9 and 0.6), bs1 mutated m3 (2.2m and 0.6m), bs1 deleted (1.7Δ, 0.9Δ and 0.6Δ), bs1 mutated as m2 (0.6m2) and bs2/bs3 mutations (0.6 mutated bs2 bs3) (the numbers indicate the promoter length in kb). See Table S1 in Supporting online material for details of the promoter-*GUS* fusions. A total of 10–20 lines were analyzed for each construct. Plants were grown in long-day conditions and the *GUS* staining assessed just after floral induction. The percentage of transgenic lines with strong, intermediate or weak *GUS* staining was scored.

in these tissues (Figure 3o). *AP1* induction by LFYVP16 was completely inhibited by the mutation or deletion of the bs1 site (Figure 3c,g,k), indicating that bs1 is essential for activation by LFY *in vivo* and that the bs2 and bs3 sites are unable to compensate for the absence of bs1, probably because of their poor affinity for LFY.

#### *AP1* activation during the floral transition

We also assayed the importance of the different binding sites during and after the floral transition. For this, we analyzed the expression of various versions of the *AP1* promoter under long-day (LD) conditions. We found that mutations affecting bs1 weakened the expression of the *AP1* promoter, especially in shortened versions of the promoter (Figure 2c). Mutations in bs2 and bs3 had little or no impact on the activation of *AP1* in response to LD conditions, confirming



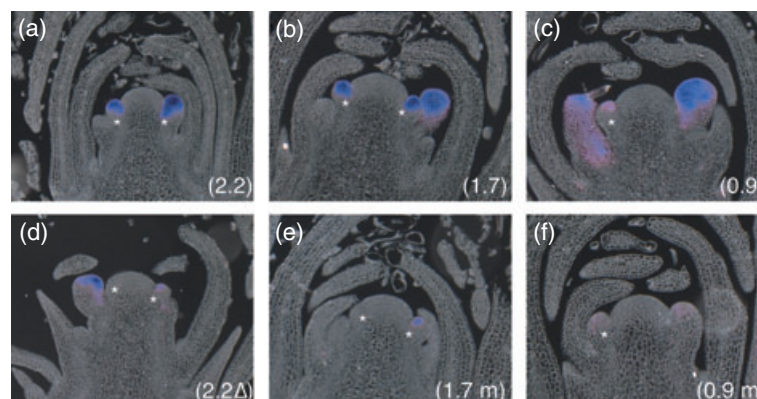
**Figure 3.** *AP1* activation by *HSP::LFYVP16* in seedlings. *AP1::GUS* transgenic plants were crossed with *HSP::LFYVP16* and the activation of *AP1* was evaluated on the F1 population. Seedlings were grown for 11 days in short-day conditions on MS plates and then heat shocked as detailed in Experimental Procedures. Controls correspond to non heat shocked seedlings from the same F1 population. The length of the promoter is shown in parentheses along with the presence of punctual mutations (m) or the deletion ( $\Delta$ ) of the bs1 site. HS, heat shocked; C, control.

- (a) Heat shocked *AP1(2.2)::GUS*.  
 (b) *AP1(2.2)::GUS* control.  
 (c) Heat shocked *AP1(2.2 $\Delta$ )::GUS*.  
 (d) *AP1(2.2 $\Delta$ )::GUS* control.  
 (e) Heat shocked *AP1(1.7)::GUS*.  
 (f) *AP1(1.7)::GUS* control.  
 (g) Heat shocked *AP1(1.7m)::GUS*.  
 (h) *AP1(1.7m)::GUS* control.  
 (i) Heat shocked *AP1(0.9)::GUS*.  
 (j) *AP1(0.9)::GUS* control.  
 (k) heat shocked *AP1(0.9m)::GUS*.  
 (l) *AP1(0.9m)::GUS* control.  
 (m) Heat shocked *HSP::GUS* seedling.  
 (n) Control *HSP::GUS* seedling.  
 (o) Apex from a heat shocked *HSP::GUS* seedling.  
 (p) Apex from a heat shocked seedling from an *AP1(1.7)::GUS*  $\times$  *HSP::LFYVP16* F1 population.

that the bs1 LFY binding site is an essential *cis*-element for *AP1* activation under LD conditions.

An efficient way to evaluate the photoperiodic pathway effect on *AP1* activation is to perform 'shift' experiments in

which plants are grown in short-day (SD) conditions for long periods and then shifted to LD conditions. We analyzed *AP1* promoter expression in the apex and in the emergent floral meristems of shifted plants (Figure 4) and observed a clear



**Figure 4.** Photoactivation of different versions of the *AP1* promoter. Transgenic plants harboring the wild-type or mutated version of the *AP1* promoter fused to the *GUS* reporter gene were grown in SD conditions for 25 days before being shifted to LD conditions. The activation of the promoter was monitored by localization of the *GUS* activity in the apex 3 days after the shift. Asterisks denote the position of newly formed floral meristems at stages 1–2.

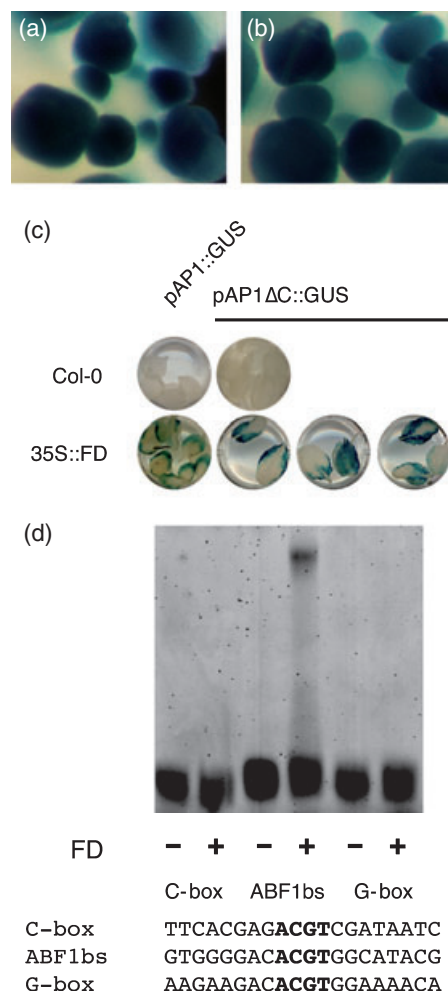
- (a–c) Wild-type promoters (2.2, 1.7 and 0.9 kb, respectively).  
 (d–f) bs1 mutated/deleted promoter (2.2, 1.7 and 0.9 kb, respectively). The promoter length and presence of mutations (m) or the deletion ( $\Delta$ ) of bs1 are indicated.

1098 Reyes Benlloch et al.

*AP1* induction for the 2.2 and 1.7 kb wild-type promoter versions; stage 1 and 2 floral meristems displayed strong GUS staining (Figure 4a,b). For plants carrying the 0.9 kb *AP1* promoter, the GUS staining was highly reduced during these stages (Figure 4c). These results suggest that elements important for photoperiodic activation are present in the promoter region between 1.7 and 0.9 kb upstream of the ATG. For long promoter versions (2.2 and 1.7 kb), mutation of *bs1* resulted in a strong reduction in the GUS signal during the early stages, while there was very little staining in the whole apex for the 0.9 kb *AP1* promoter (Figure 4d–f). These results confirm the importance of the *LFY bs1* for *AP1* expression both in the initial response after photoperiodic induction and the recurrent *AP1* expression in early floral stages on the inflorescence. They also show that the photoperiodic activation of *AP1* by FD/FT cannot immediately compensate for the lack of induction by *LFY*.

#### The putative FD-binding C-box is not needed for *AP1* activation

In addition to *LFY*, several regulators contribute to *AP1* photoperiodic induction (Parcy, 2005; Giakountis and Coupland, 2008; Liu *et al.*, 2009; Albani and Coupland, 2010; Irish, 2010) but the *cis*-elements involved have not yet been identified. An FD-response region has been mapped in the *AP1* promoter close to the *LFY bs1* site, including a putative binding site for FD (C-box) (Wigge *et al.*, 2005) (Figure 1a). We mutated this C-box in the 2.2 kb *AP1* promoter context and did not observe any notable modification of *AP1::GUS* expression (Figure 5a,b). This result suggested either that the C-box is not a genuine FD-binding site or that other *cis*-elements or pathways can compensate its absence. To assay the authenticity of this C-box as an FD-binding site, we performed both *in vivo* and *in vitro* assays. For the *in vivo* experiments, we used young *Arabidopsis* leaves from 1-month-old plants, a tissue where *AP1* is not normally expressed but can be induced by the constitutive *FD* expression (*35S::FD*) (Wigge *et al.*, 2005). We introduced an *AP1::GUS* transgene, with or without the C-box, in wild-type and *35S::FD* transgenic plants and observed that *FD* activates *AP1* promoter independently of the presence of the C-box (Figure 5c). This result suggests that this *cis*-element might not be a *bona-fide* FD-binding site. To test this hypothesis, we produced a recombinant version of the *FD* protein and tested its DNA binding properties using EMSA. *FD* DNA binding specificity is unknown but it belongs to the same bZIP group (Jakoby *et al.*, 2002) as the *ABF1* protein for which this property has been established (Choi *et al.*, 2000). We observed a clear *FD* binding to the canonical *ABF1* binding site (*ABF1bs*) (Figure 5d). However, *FD* was not able to bind the C-box. We also tested the binding to a G-box present in the *AP1* promoter but did not observe any binding either. Altogether, our data indicate that the C-box is unlikely to be the *FD*-binding site and that there are no other obvious



**Figure 5.** Analysis of the putative FD-binding site.

(a, b) *AP1::GUS* activity in *Arabidopsis* inflorescence apices transformed with a reporter construct with an intact (a) or mutated (b) C-box.

(c) Transactivation of the *AP1* promoter by *FD* overexpression in leaves does not require the C-box. *AP1::GUS* reporter construct (1.6 kb with or without the C-box) was transformed into wild-type and *35S::FD* backgrounds and stained for GUS activity. *AP1::GUS* expression is detected in leaves even in the absence of the C-box.

(d) EMSA showing that the recombinant *FD* protein binds to a canonical *ABF1* binding site but not to the G-box and the C-box present in *AP1* promoter.

*FD*-binding sites in the *AP1* promoter, suggesting that either the FT/*FD* complex interacts with other atypical elements within this promoter region, or the FT-dependent activation of *AP1* transcription is not mediated by a direct binding of *FD* to the *AP1* promoter.

#### DISCUSSION

##### *Bs1* is the essential *LFY* binding site required for *AP1* activation

*LFY* was shown to directly activate *AP1* *in planta* and to bind *in vitro* to an element present in the *AP1* promoter region (Parcy *et al.*, 1998; Wagner *et al.*, 1999). However, the

relevance of this site had never been tested *in vivo* and several other potential binding sites were identified in the *AP1* promoter. Here, we combined *in vitro* and *in vivo* experiments to show that bs1 is the highest-affinity LFY binding site *in vitro* and that mutating this site has a strong effect on the activation of *AP1* transcription, both by a LFY protein with a constitutive transcriptional activation potential (LFYVP16) and by the LD pathway. These results are consistent with two recent analyses of LFY-DNA binding preferences *in vitro* and *in vivo* that showed that the CCANTG consensus represents only the central portion of the LFY binding site (Moyroud *et al.*, 2011; Winter *et al.*, 2011). The presence of this small consensus sequence is neither necessary nor sufficient for LFY binding and can be advantageously replaced by a Position Specific Scoring Matrix that better captures LFY-DNA binding specificity (Moyroud *et al.*, 2011). This improved tool indeed predicts that bs1 has a much higher affinity than bs2 and bs3, as shown here experimentally.

#### Bs1 is important for photoperiodic induction

*APETALA1* is known to be induced upon photoinduction (Hempel *et al.*, 1997) and this regulation was shown to involve the FT/FD pathway (Ruiz-Garcia *et al.*, 1997; Liljegren *et al.*, 1999; Wigge *et al.*, 2005). The importance of LFY in this process has been a matter of debate. Data in Arabidopsis or *Sinapis alba* showed that LFY activation precedes *AP1* upregulation (Hempel *et al.*, 1997; D'Aloia *et al.*, 2009). Other studies describe *AP1* as primarily induced by FD and subsequently activating LFY through a positive feedback loop (Giakountis and Coupland, 2008). LFY expression is not abolished in *ap1* or *ap1 cal* backgrounds (Ferrandiz *et al.*, 2000) indicating that it does not entirely depend on AP1/CAL even if the positive feedback loop involving these three genes is essential for their sustained expression. Our finding that *AP1* promoter activation is strongly reduced in the absence of LFY binding site bs1 is in favour of a primary role of LFY in *AP1* photoinduction. The requirement for bs1 is particularly striking with the 0.9 kb version of *AP1* promoter: no *AP1* activation is observed in the absence of bs1 after the long-day shift. It is however likely that a LFY independent *AP1* induction will occur a few days later as observed for *AP1* mRNA (Schmid *et al.*, 2003). In our opinion, these results and the previously published data in Arabidopsis and *Sinapis alba* are thus in accordance with a model in which both FD/FT and LFY pathways act synergistically to induce *AP1* even if each of them can act on its own, when the other pathway is defective.

#### What are the *cis*-elements responsible for *AP1* induction?

The nature of the other *cis*-elements required for *AP1* photoperiodic induction remains elusive. FD belongs to the bZIP class of transcription factors, known to bind elements with an ACGT core sequence (Izawa *et al.*, 1993; Foster *et al.*,

1994). Depending on the flanking nucleotides, these sequences are classified as A-box (TACGTA), C-box (GAC-GTC) or G-box (CACGTG). A C-box is present in the proximal *AP1* promoter region (Figure 1a) that was proposed as potential FD-binding site (Wigge *et al.*, 2005). Our experiments performed *in vitro* with recombinant FD or *in vivo* using ectopic FD expression in leaves show that this C-box is not required for *AP1* activation by FD and does not bind FD *in vitro*. As shown for ABF1 (Choi *et al.*, 2000), a bZIP belonging to the same subfamily as FD [Group A; (Jakoby *et al.*, 2002)], FD appears to rather have a higher affinity for a G-box-like motif (Figure 5d). Our observation that early upregulation of *AP1* was reduced in the short promoter (0.9 kb) compared to the longer versions (1.7 and 2.2 kb) suggests that some element(s) that lie between 0.9 and 1.7 kb are important for the photoperiodic response. We found one G-box in this region (Figure S2) but it was not bound by FD *in vitro* in our experimental conditions, suggesting that nucleotide positions outside of the G-box might be important for FD binding.

Overall, our results raise the question whether FD binds at all to the *AP1* promoter. By using chromatin immunoprecipitation with antibodies against FT in a *35S::FD* background, it was shown that FT is recruited to a 130-bp fragment of the *AP1* promoter in an FD-dependent manner (Wigge *et al.*, 2005). However, since the antibodies were directed against FT and not FD it is possible that FT was recruited to the *AP1* promoter, not by FD itself, but by another protein induced by FD overexpression. FD might also be recruited to the *AP1* promoter in complex with other proteins and through a *cis*-element different from the G-box recognized by the FD homodimer. The BLADE-ON-PETIOLE 1 (BOP1) protein, for example, was reported to be recruited to the *AP1* promoter, possibly via interaction with a TGA bZIP transcription factor (Xu *et al.*, 2010). There are several TGA binding sites in *AP1* promoter including two sites adjacent to LFY bs1 (Figure S2). FD could thus interact with some bZIPs or other transcription factors and bind *AP1* promoter through elusive *cis*-elements. The mechanism for *AP1* expression by FD and FT therefore remains an open question.

Recently it was suggested that the miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factors define a new endogenous flowering pathway (Wang *et al.*, 2009; Yamaguchi *et al.*, 2009). SPL3 and SPL9 have been identified as direct *AP1* regulators (Wang *et al.*, 2009; Yamaguchi *et al.*, 2009). SPL proteins bind DNA sequences containing the core sequence GTAC (Cardon *et al.*, 1997, 1999; Yamaguchi *et al.*, 2009) and several of these sequences have been identified in the *AP1* promoter (Figure S2) including one next to LFY bs1 (Figure 1). Important nucleotides for LFY binding have been found expanding into the SPL binding site and we cannot completely exclude that the mutations introduced into bs1



also affect SPL binding. This hypothesis would explain why mutations in *bs1* have even stronger effects on *AP1* expression than *lfy* mutations.

Other factors have also been shown to contribute to *AP1* upregulation. MADS-box proteins, such as SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (*SOC1*), FRUIT-FULL (*FUL*), SEPALLATA3 (*SEP3*) and AGL17, have key roles in controlling *AP1* expression. *SOC1* upregulation in the apex occurs very early in response to different environmental signals, such as photoperiod, gibberellins or vernalization (Moon *et al.*, 2003; Yoo *et al.*, 2005). *FUL* can promote floral transition and can have a redundant function with *AP1* and CAULIFLOWER (*CAL*) (Hempel *et al.*, 1997; Ferrandiz *et al.*, 2000), while AGL17 regulates *AP1* expression in response to photoperiod in a FT-independent manner (Han *et al.*, 2008). *SEP3* is bound to *AP1* promoter *in vivo*, likely as part of MADS-box protein complexes controlling *AP1* expression (Kaufmann *et al.*, 2010). MADS-box proteins bind DNA containing the consensus sequences CC(A/T)<sub>6</sub>GG (CarG-box) (Riechmann *et al.*, 1996). Motif searches along the *AP1* promoter region allowed us to identify four putative CarG-boxes (allowing one mismatch in the consensus sequence) (Figure S2). These CarG-boxes could represent upstream regulatory elements with a role in the photoperiodic response and it would be interesting to further investigate their role in *AP1* regulation.

Recent work has revealed that *AP1* is a crucial factor that orchestrates a network of genes involved floral meristem-identity specification and floral development (Kaufmann *et al.*, 2010). It is therefore not surprising that such a key gene contains many different regulatory elements along its promoter; these elements allow the fine-tuning of its expression in response to different environmental cues (Figure S1). In our study, we examined the role of these factors (focusing on LFY, FT and FD) in the regulation of *AP1*, but their physical interactions at the promoter level still require further investigation.

To summarize, we have provided strong evidence for *bs1* being the most functionally important LFY binding site in the *AP1* promoter. This conclusion provides a substantial basis for studying the contribution of LFY and other regulatory factors to *AP1* expression pattern. We and others have also shown that under LD conditions LFY and FT actions are tightly coupled to achieve the correct upregulation of *AP1* at the correct time and at the right place, but that we have found no evidence for a direct binding of FD to the *AP1* promoter. We thus propose that either FD does not bind to *AP1* promoter directly, or that it is recruited to this promoter through interactions with other proteins. Further studies are now needed in order to determine how other factors and signals integrate with LFY and FT to achieve the correct temporal and spatial transcriptional regulation of *AP1* transcription.

## EXPERIMENTAL PROCEDURES

### Plant material and growth conditions

All transgenic plants as well as the *lfy-12* and *ft-10* mutants are in the *Arabidopsis thaliana* Col-0 background. Plants were grown in LD (16 h light) or SD (9 h light) conditions at 22°C/18°C (light/dark). The light source was composed of cool white fluorescence tubes (colour code 840) with a maximum irradiance of 150 μmol m<sup>-2</sup> sec<sup>-1</sup>. The FR/R ratio was 1.2. The relative humidity was 60%.

### Plasmid constructions

pEDW5 was constructed by introducing the β-glucuronidase (*GUS*) coding sequence into the pFP100 binary vector (Bensmihen *et al.*, 2004). pRB14 was generated by the introduction of the Gateway RfA cassette into pEDW5 allowing the introduction of promoters upstream of the *GUS* gene by LR recombination. The different promoter versions were amplified by polymerase chain reaction (PCR), using mutagenic oligonucleotides (listed in Table S2) whenever needed. The m3 LFYbs1 mutation was generated by gene synthesis (Millegen, <http://www.millegen.com>). Amplified promoters were cloned into the pENTR™/D-TOPO® vector (Invitrogen, <http://www.invitrogen.com>) and shuttled into the pRB14. The HSP::LFYVP16 vector (pRB13) was generated by cloning the LFYVP16 sequence from pDW245 (Parcy *et al.*, 1998) into the pK2GW7-HSP vector, a modified version of pK2GW7 (Karimi *et al.*, 2002) in which the promoter from the soybean *Gmhsp 17.6L* heat-shock protein gene (Severin and Schoffl, 1990) replaces the 35S promoter. pALS-5 was built by inserting *FD* cDNA into the pETM-11 vector (Hamès *et al.*, 2008).

### Electrophoretic mobility shift assays

Fluorescence EMSA with LFY-C and FD were performed as described by Hamès *et al.* (2008) except for the FD binding buffer (Tris 20 mM pH 7.9, KCl 40 mM, MgCl<sub>2</sub> 2 mM, Ficoll 5%, bovine serum albumin (BSA) 200 μg ml<sup>-1</sup>, arginine 60 mM, dithiothreitol (DTT) 1 mM, 28 ng μl<sup>-1</sup> fish sperm DNA 2.8 μg ml<sup>-1</sup>). The oligonucleotides used are listed in Table S2.

### Fluorescence anisotropy (FA)

Fluorescence anisotropy measurements were performed on a Safire2 fluorimeter (Tecan, <http://www.tecan.com>) according to the manufacturers' instructions using double stranded DNA oligonucleotides end-labeled with tetra-methylcarboxy-rhodamine (TAMRA) on a single strand (Sigma, <http://www.sigmaaldrich.com>). *K*<sub>0.5</sub> was estimated from the resulting binding curve. Normalized FA values shown in Figure 1 were obtained by subtracting the anisotropy of free DNA.

### Heat-shock treatment in seedlings

Three different *AP1*::*GUS* (RB47, RB40 and RB44) reporter lines were crossed with a homozygous HSP::LFYVP16 (RB13) line. F1 seeds selected to contain both transgenes (*GUS* reporter and LFYVP16 effector) were grown on MS petri plates for 11 days in SD conditions and then heat shocked three times (three consecutive days; 1 h each day). The *GUS* assay was performed after the last heat-shock treatment.

### GUS assay

Seedlings were prefixed in cold 90% acetone for 20 min, washed with X-Gluc Buffer (0.2% Triton, 50 mM NaPO<sub>4</sub> pH 7.2, 2 mM ferrocyanide, and 2 mM ferricyanide) on ice and finally stained with X-Gluc Buffer containing 2 mM of X-GlcA (Duchefa

Biochemie, <http://www.duchefa.com>). Seedlings were infiltrated for 5 min under a vacuum and incubated overnight at 37°C. The staining reaction was stopped using an ethanol series (20, 35 and 50%). Samples were post-fixed with FAA (50% EtOH, 5% formaldehyde and 10% acetic acid) for 30 min and dehydrated in an ethanol series, stained in eosin overnight at 4°C and incubated in a series of ethanol/histoclear mixtures (3:1, 1:1, 1:3), followed by an incubation in 100% histoclear. The samples were then embedded in paraffin and cut into 8-µm sections with a Microm GmbH microtome (Zeiss, <http://www.zeiss.fr>). Samples were examined and images taken with a light microscope (Axioplan2; Zeiss).

### qPCR

Apices were microdissected from wild-type and *lfy-12* mutants harboring the 2.2 kb version of the *AP1* promoter. *lfy-12* mutant plants were genotyped using End Point Genotyping in a LightCycler® 480 (Roche, <http://www.roche.com>). Primers and probes for the genotyping are listed in Table S2. Plants were grown in SD conditions for 25 days and then moved to LD conditions. Three independent biological replicates were analyzed. The expression of *AP1* and *GUS* mRNA was measured by quantitative PCR (qPCR) using a LightCycler® 480 SYBR Green I Master. The PCR profile was as follows: primary denaturation at 95°C for 5 min followed by 40 amplification cycles of 10 sec at 95°C, 15 sec at 60°C and 15 sec at 72°C. Melting curve analysis was performed to ensure there was no formation of primer dimers. Each sample was loaded in triplicate. Results were normalized to the expression of actin. Primer sequences are listed in Table S2.

### Histochemical analysis of pAP1ΔC::GUS plants

The pAP1ΔC::GUS construct, harboring 1.6 kb of *AP1* promoter with a deleted C-box, was transformed into Col-0 and *35S::FD* transgenic plants, respectively. T<sub>1</sub> lines of pAP1ΔC::GUS/Col-0 and pAP1ΔC::GUS/35S::FD transgenic plants were grown under LD conditions for 1 month and GUS staining was performed on the leaves.

### Protein purification

LFY-C expression and purification were carried out as described by Hamès *et al.* (2008). FD expression was induced overnight by 0.4 mM IPTG at 17°C in *Escherichia coli* Rosetta2 strain (Novagen, <http://www.merck-chemicals.fr>) transformed with the pALS-5 plasmid. Bacterial pellet was resuspended in 100 ml of buffer per liter of culture [Tris 50 mM pH 8.0, protease inhibitor cocktail tablet Complete EDTA-free (Roche Diagnostics)], sonicated for 15 min at 4°C on a Sonifier® 250 ultrasonic cell disruptor (Branson Ultrasonics S.A., <http://www.bransonultrasonics.com>). Insoluble proteins were purified by precipitation (30 min, 30 000 g, 4°C) followed by solubilisation in arginine 0.2 M and additional centrifugation (30 min, 30 000 g, 4°C). The resulting supernatant contained the purified FD protein (20 mg L<sup>-1</sup> of *E. coli* cells, 0.4 µg µl<sup>-1</sup> final concentration, 95% homogeneity).

### ACKNOWLEDGEMENTS

We are grateful to R. Dumas for advice on FD protein production, E. Moyroud for the pEDW5 plasmid, A-L Schinke and M. Schmid for the pALS-5 vector and to D. Weigel for experiments performed in his laboratory. This work was supported by funding from the Centre National de la Recherche Scientifique (ATIP) to FP and RB, the Agence Nationale de la Recherche (ANR, Plant-TFcode) to FP, the ANR and the Biotechnology and Biological Sciences Research Council (Flower Model) to FP, the Next-Generation BioGreen 21

Program (PJ008025), Rural Development Administration, Republic of Korea to MCK, a post-doctoral stipend from the Kempe foundation to RB, and grants from the Swedish research Council (VR) and the Swedish Governmental Agency for Innovation Systems (VINNOVA) to ON.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Photoactivation of the *AP1* promoter (2.2 kb) in *ft-10* and *lfy-12* mutant backgrounds.

**Figure S2.** Nature and position of putative transcription factors binding sites in *AP1* promoter. SPL and TGA binding sites have been located using the Athamap tool (Steffens *et al.*, 2004).

**Table S1.** Information on the promoter-GUS fusions used in the experiments.

**Table S2.** Sequence of the primers and probes.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

### REFERENCES

- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K. and Araki, T. (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science*, **309**, 1052–1056.
- Albani, M.C. and Coupland, G. (2010) Comparative analysis of flowering in annual and perennial plants. *Curr. Top. Dev. Biol.* **91**, 323–348.
- Bensmihen, S., To, A., Lambert, G., Kroj, T., Giraudat, J. and Parcy, F. (2004) Analysis of an activated ABI5 allele using a new selection method for transgenic *Arabidopsis* seeds. *FEBS Lett.* **561**, 127–131.
- Blazquez, M.A., Soowal, L.N., Lee, I. and Weigel, D. (1997) LEAFY expression and flower initiation in *Arabidopsis*. *Development*, **124**, 3835–3844.
- Busch, M.A., Bomblies, K. and Weigel, D. (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science*, **285**, 585–587.
- Cardon, G.H., Hohmann, S., Nettekheim, K., Saedler, H. and Huijser, P. (1997) Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. *Plant J.* **12**, 367–377.
- Cardon, G., Hohmann, S., Klein, J., Nettekheim, K., Saedler, H. and Huijser, P. (1999) Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene*, **237**, 91–104.
- Choi, H., Hong, J., Ha, J., Kang, J. and Kim, S.Y. (2000) ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.* **275**, 1723–1730.
- D'Aloia, M., Tamseddak, K., Bonhomme, D., Bonhomme, F., Bernier, G. and Perilleux, C. (2009) Gene activation cascade triggered by a single photoperiodic cycle inducing flowering in *Sinapis alba*. *Plant J.* **59**, 962–973.
- Ferrandiz, C., Gu, Q., Martienssen, R. and Yanofsky, M.F. (2000) Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development*, **127**, 725–734.
- Foster, R., Izawa, T. and Chua, N.H. (1994) Plant bZIP proteins gather at ACGT elements. *FASEB J.* **8**, 192–200.
- Giakountis, A. and Coupland, G. (2008) Phloem transport of flowering signals. *Curr. Opin. Plant Biol.* **11**, 687–694.
- Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1994) Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell*, **76**, 131–143.
- Hamès, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F., Martiel, J.L., Benlloch, R., Parcy, F. and Muller, C.W. (2008) Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* **27**, 2628–2637.
- Han, P., Garcia-Ponce, B., Fonseca-Salazar, G., Alvarez-Buylla, E.R. and Yu, H. (2008) AGAMOUS-LIKE 17, a novel flowering promoter, acts in a FT-independent photoperiod pathway. *Plant J.* **55**, 253–265.
- Hempel, F.D., Weigel, D., Mandel, M.A., Ditta, G., Zambryski, P.C., Feldman, L.J. and Yanofsky, M.F. (1997) Floral determination and expression of floral regulatory genes in *Arabidopsis*. *Development*, **124**, 3845–3853.

- Irish, V.F. (2010) The flowering of *Arabidopsis* flower development. *Plant J.* **61**, 1014–1028.
- Izawa, T., Foster, R. and Chua, N.H. (1993) Plant bZIP protein DNA binding specificity. *J. Mol. Biol.* **230**, 1131–1144.
- Jakoby, M., Weisshaar, B., Droge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. and Parcy, F. (2002) bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* **7**, 106–111.
- Karimi, M., Inze, D. and Depicker, A. (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195.
- Kaufmann, K., Wellmer, F., Muino, J.M. et al. (2010) Orchestration of floral initiation by *APETALA1*. *Science*, **328**, 85–89.
- Letilly, V. and Royer, C.A. (1993) Fluorescence anisotropy assays implicate protein-protein interactions in regulating trp repressor DNA-binding. *Biochemistry*, **32**, 7753–7758.
- Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S. and Yanofsky, M.F. (1999) Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell*, **11**, 1007–1018.
- Liu, C., Thong, Z. and Yu, H. (2009) Coming into bloom: the specification of floral meristems. *Development*, **136**, 3379–3391.
- Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R. and Weigel, D. (2001) A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell*, **105**, 793–803.
- Mandel, M.A., Gustafson-Brown, C., Savidge, B. and Yanofsky, M.F. (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature*, **360**, 273–277.
- Moon, J., Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C., Kim, S.G. and Lee, I. (2003) The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in *Arabidopsis*. *Plant J.* **35**, 613–623.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R. and Parcy, F. (2010) *LEAFY* blossoms. *Trends Plant Sci.* **15**, 346–352.
- Moyroud, E., Minguet, E.G., Ott, F. et al. (2011) Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* *LEAFY* Transcription Factor. *Plant Cell*, **23**, 1293–1306.
- Nilsson, O., Lee, I., Blazquez, M.A. and Weigel, D. (1998) Flowering-time genes modulate the response to *LEAFY* activity. *Genetics*, **150**, 403–410.
- Parcy, F. (2005) Flowering: a time for integration. *Int. J. Dev. Biol.* **49**, 585–593.
- Parcy, F., Nilsson, O., Busch, M.A., Lee, I. and Weigel, D. (1998) A genetic framework for floral patterning. *Nature*, **395**, 561–566.
- Ratcliffe, O.J., Bradley, D.J. and Coen, E.S. (1999) Separation of shoot and floral identity in *Arabidopsis*. *Development*, **126**, 1109–1120.
- Riechmann, J.L., Wang, M. and Meyerowitz, E.M. (1996) DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins *APETALA1*, *APETALA3*, *PISTILLATA* and *AGAMOUS*. *Nucleic Acids Res.* **24**, 3134–3141.
- Ruiz-Garcia, L., Madueno, F., Wilkinson, M., Haughn, G., Salinas, J. and Martinez-Zapater, J.M. (1997) Different roles of flowering-time genes in the activation of floral initiation genes in *Arabidopsis*. *Plant Cell*, **9**, 1921–1934.
- Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann, J.U. (2003) Dissection of floral induction pathways using global expression analysis. *Development*, **130**, 6001–6012.
- Severin, K. and Schoffl, F. (1990) Heat-inducible hygromycin resistance in transgenic tobacco. *Plant Mol. Biol.* **15**, 827–833.
- Steffens, N.O., Galuschka, C., Schindler, M., Bülow, L. and Hehl, R. (2004) AthaMap: an online resource for in silico transcription factor binding sites in the *Arabidopsis thaliana* genome. *Nucleic Acids Res.* **32**, D368–D372.
- Wagner, D., Sablowski, R.W. and Meyerowitz, E.M. (1999) Transcriptional activation of *APETALA1* by *LEAFY*. *Science*, **285**, 582–584.
- Wang, J.W., Czech, B. and Weigel, D. (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell*, **138**, 738–749.
- Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U. and Weigel, D. (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science*, **309**, 1056–1059.
- William, D.A., Su, Y., Smith, M.R., Lu, M., Baldwin, D.A. and Wagner, D. (2004) Genomic identification of direct target genes of *LEAFY*. *Proc. Natl Acad. Sci. USA*, **101**, 1775–1780.
- Winter, C.M., Austin, R.S., Blanvillain-Baufume, S. et al. (2011) *LEAFY* Target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell*, **20**, 430–443.
- Wu, X., Dinneny, J.R., Crawford, K.M., Rhee, Y., Citovsky, V., Zambryski, P.C. and Weigel, D. (2003) Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development*, **130**, 3735–3745.
- Xu, M., Hu, T., McKim, S.M., Murmu, J., Haughn, G.W. and Hepworth, S.R. (2010) *Arabidopsis* *BLADE-ON-PETIOLE1* and 2 promote floral meristem fate and determinacy in a previously undefined pathway targeting *APETALA1* and *AGAMOUS-LIKE24*. *Plant J.* **63**, 974–989.
- Yamaguchi, A., Wu, M.F., Yang, L., Wu, G., Poethig, R.S. and Wagner, D. (2009) The microRNA-regulated SBP-Box transcription factor *SPL3* is a direct upstream activator of *LEAFY*, *FRUITFULL*, and *APETALA1*. *Dev. Cell*, **17**, 268–278.
- Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S. and Ahn, J.H. (2005) *CONSTANS* activates *SUPPRESSOR OF OVER-EXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in *Arabidopsis*. *Plant Physiol.* **139**, 770–778.



## Annexe 4.

## A variant of LEAFY reveals its capacity to stimulate meristem development by inducing *RAX1*

Hicham Chahtane<sup>1,2,3,4</sup>, Gilles Vachon<sup>1,2,3,4</sup>, Marie Le Masson<sup>1,2,3,4</sup>, Emmanuel Thévenon<sup>1,2,3,4</sup>, Sophie Périgon<sup>1,2,3,4,†</sup>, Nela Mihajlovic<sup>5</sup>, Anna Kalinina<sup>5</sup>, Robin Michard<sup>1,2,3,4</sup>, Edwige Moyroud<sup>1,2,3,4,‡</sup>, Marie Monniaux<sup>1,2,3,4</sup>, Camille Sayou<sup>1,2,3,4</sup>, Vojislava Grbic<sup>5</sup>, Francois Parcy<sup>1,2,3,4,\*§</sup> and Gabrielle Tichtinsky<sup>1,2,3,4</sup>

<sup>1</sup>Laboratoire Physiologie Cellulaire et Végétale, CEA (Commissariat à l'Energie Atomique et aux Energies Alternatives), iRTSV (Institut de Recherches en Technologies et Sciences pour le Vivant), Grenoble F-38054, France,

<sup>2</sup>Centre National de la Recherche Scientifique, UMR5168, Grenoble F-38054, France,

<sup>3</sup>Université Joseph Fourier-Grenoble I, UMR5168, Grenoble F-38041, France,

<sup>4</sup>Institut National de la Recherche Agronomique, USC1359, Grenoble F-38054, France, and

<sup>5</sup>Department of Biology, University of Western Ontario, London, ON N6A 5B8, Canada

Received 28 November 2012; revised 13 February 2013; accepted 19 February 2013.

\*For correspondence (e-mail francois.parcy@cea.fr).

†Present address: Laboratoire d'Ecologie Alpine, Centre National de la Recherche Scientifique, UMR 5553, Grenoble F-38041, France.

‡Present address: Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK.

§Present address: Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC V5Z 4H4, Canada.

### SUMMARY

In indeterminate inflorescences, floral meristems develop on the flanks of the shoot apical meristem, at positions determined by auxin maxima. The floral identity of these meristems is conferred by a handful of genes called floral meristem identity genes, among which the LEAFY (LFY) transcription factor plays a prominent role. However, the molecular mechanism controlling the early emergence of floral meristems remains unknown. A body of evidence indicates that LFY may contribute to this developmental shift, but a direct effect of LFY on meristem emergence has not been demonstrated. We have generated a LFY allele with reduced floral function and revealed its ability to stimulate axillary meristem growth. This role is barely detectable in the *lfy* single mutant but becomes obvious in several double mutant backgrounds and plants ectopically expressing LFY. We show that this role requires the ability of LFY to bind DNA, and is mediated by direct induction of *REGULATOR OF AXILLARY MERISTEMS1* (*RAX1*) by LFY. We propose that this function unifies the diverse roles described for LFY in multiple angiosperm species, ranging from monocot inflorescence identity to legume leaf development, and that it probably pre-dates the origin of angiosperms.

**Keywords:** meristem development, flower, LEAFY, *RAX1*, *Arabidopsis thaliana*.

### INTRODUCTION

In plants bearing raceme inflorescences such as *Arabidopsis thaliana*, flowers arise on the flanks of the shoot apical meristem. They initially emerge as small groups of cells that acquire first meristematic identity and then floral identity, to eventually produce flower buds. Meristematic identity is first evident in the shape and organization of this domed group of dividing cells (stages 1–2), and later in the molecular markers that they express, such as *CLAVATA3* (*CLV3*), *WUSCHEL* (*WUS*) and *UNUSUAL FLORAL ORGANS* (*UFO*) (stages 2–3). Floral meristem identity genes determine floral identity (Liu *et al.*, 2009; Wagner, 2009; Irish, 2010; Moyroud *et al.*, 2010; Pose *et al.*, 2012). This class includes *LEAFY* (*LFY*) and *APETALA1* (*AP1*),

which control the development of flowers and their patterning through induction of organ identity genes such as *AGAMOUS* (*AG*) and *APETALA3* (*AP3*). *LFY* and *AP1* act very early in meristem development (stages 0–1) and share many direct target genes (Kaufmann *et al.*, 2010; Moyroud *et al.*, 2011; Winter *et al.*, 2011). Although determination of floral meristem identity has been intensively studied and is well understood, how the initial group of cells acquires its meristematic identity remained unknown. It is known that auxin maxima determine where floral meristems arise (Kuhlemeier, 2007; Vernoux *et al.*, 2010), but the link between auxin signaling and induction of meristematic features has not yet been fully established.

## 2 Hicham Chahtane et al.

Several observations indicate that LFY is a possible candidate to mediate this action. First, LFY expression is controlled by auxin as shown in *pin* mutants (Vernoux *et al.*, 2000; Lebedeva *et al.*, 2005; Blazquez *et al.*, 2006). Second, scattered data suggest that LFY may play a role in conferring meristematic identity to the floral meristem (Moyroud *et al.*, 2009, 2010). Finally, LFY is able to induce meristematic identity in other developmental contexts, for instance to create compound leaves in some legumes, tillers in rice (*Oryza sativa*), or complex inflorescences in rice and maize (*Zea mays*) (Moyroud *et al.*, 2009, 2010). LFY may thus provide the link between auxin signaling and meristematic growth. However, LFY is not required for meristem development *per se*: in the *lfy* single mutant, flowers are replaced by cauline leaves that bear a meristematic structure in their axil. This structure gives rise to shoots or shoot/flower intermediates. It is only when the *lfy* mutation is combined with defects in genes such as *PINOID* (*PID*) (Ezhova *et al.*, 2000), *CLV1* (Clark *et al.*, 1993), *FILAMENTOUS FLOWERS* (*FIL*) (Chen *et al.*, 1999; Sawa *et al.*, 1999a), *ENHANCED RESPONSE TO ABA* (*ERA*) (Running *et al.*, 1998; Yalovsky *et al.*, 2000), *FUSED FLORAL ORGANS* (*FFO*) (Levin *et al.*, 1998), *FLOWERING LOCUS T* (*FT*), *FWA* (Ruiz-Garcia *et al.*, 1997) or *BLADE ON PETIOLE1/2* (*BOP1/2*) (Norberg *et al.*, 2005) that leaves lacking axillary meristems develop at some positions on the inflorescence (referred to as empty cauline leaves).

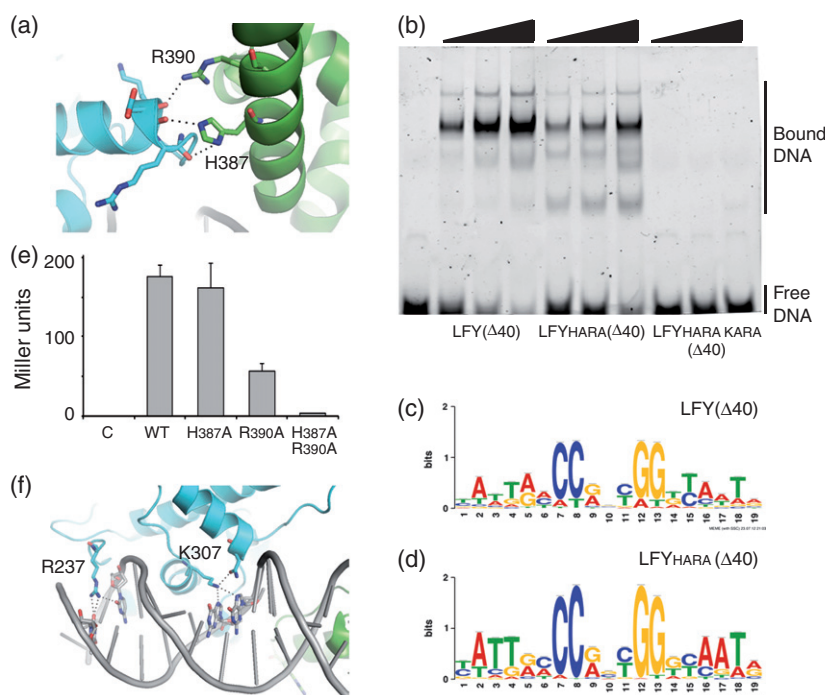
Here we provide direct evidence in *A. thaliana* that LFY has the capacity to trigger meristem growth. We obtained this evidence by altering the properties of the LFY protein.

LFY has two conserved functional domains: an N-terminal dimerization domain and a C-terminal DNA-binding domain (DBD) (Hamès *et al.*, 2008; Siriwardana and Lamb, 2012). The DBD dimerizes upon DNA binding. Taking advantage of its crystallographic structure, we engineered a form of the LFY DBD (called LFY<sub>HARA</sub>) that has a compromised dimerization interface. When expressed in plants, this allele shows reduced floral function, thereby revealing the ability of LFY to stimulate the emergence of meristems.

## RESULTS

## The HARA mutation reduces the affinity of LFY for DNA without affecting its sequence specificity

Two amino acid residues (H387 and R390) involved in dimerization of the LFY DBD (Figure 1a) have been identified previously, but their importance was only established *in vitro* using the isolated LFY DBD (Hamès *et al.*, 2008); their role within the context of a full-length LFY protein, which contains the conserved N-terminal dimerization domain (Siriwardana and Lamb, 2012), has not been determined. We thus analyzed the properties of the near full-length LFY<sub>HARA</sub>( $\Delta$ 40) protein (see Experimental procedures), in which both the H387 and R390 residues are mutated to alanine, thereby suppressing the hydrogen bonds between the DBD monomers (Figure 1a). Using EMSA analysis (Figure 1b), we found that LFY<sub>HARA</sub>( $\Delta$ 40) binds AP1bs1, a binding site present in the AP1 promoter (Benlloch *et al.*, 2011), but with weaker affinity than wild-type LFY( $\Delta$ 40). This result confirms the importance of



**Figure 1.** Biochemical characterization of the LFY<sub>HARA</sub> and LFY<sub>HARAKARA</sub> proteins.

(a) Details of the structure of a LFY DBD dimer (Hamès *et al.*, 2008) showing the hydrogen bonds (dotted lines) between monomers (shown in blue and green) involving His387 (H387) and Arg390 (R390).

(b) Electrophoretic mobility shift assay (EMSA) using the same increasing quantities of LFY ( $\Delta$ 40), LFY<sub>HARA</sub>( $\Delta$ 40) and LFY<sub>HARAKARA</sub>( $\Delta$ 40) mixed with the AP1bs1 binding site (Benlloch *et al.*, 2011).

(c, d) Logos illustrating position-specific scoring matrices obtained by high-throughput SELEX experiments performed with LFY( $\Delta$ 40) (c) and LFY<sub>HARA</sub>( $\Delta$ 40) (d).

(e) Yeast one-hybrid assay. LFY proteins were fused to the viral VP16 activation domain. Their ability to bind the regulatory sequences of AG was followed using the activity of the reporter gene *AG:LACZ* ( $\beta$ -galactosidase activity). C: Control (LFY unfused to VP16); WT: wild-type LFY-VP16; H387A: LFY<sub>H387A</sub>-VP16; R390A: LFY<sub>R390A</sub>-VP16; H387A R390A: LFY<sub>H387A R390A</sub>-VP16. Error bars indicate the standard error of the mean ( $n = 10$ ).

(f) Detail of the LFY/DNA interaction (Hamès *et al.*, 2008) showing hydrogen bonds (dotted lines) between Lys307 (K307) and Arg237 (R237) and the DNA bases.

contacts between the DBDs even in the presence of the N-terminal dimerization domain. To test whether the HARA mutations affect LFY DNA-binding specificity, we performed a high-throughput Systematic evolution of ligands by exponential enrichment (SELEX) experiment (Zhao *et al.*, 2009). The resulting logo, representing the DNA-binding preferences of LFY<sub>HARA</sub>( $\Delta$ 40) (Figure 1c,d), was found to be extremely similar to the one obtained for LFY( $\Delta$ 40), indicating that the HARA mutations do not greatly affect LFY DNA-binding specificity. We also tested the importance of LFY DBD dimerization in a yeast one-hybrid assay. Because LFY displays no transcriptional activity in yeast, full-length LFY and LFY<sub>HARA</sub> proteins were fused to the VP16 transcriptional activation domain (Parcy *et al.*, 1998; Lohmann *et al.*, 2001), and their activity were tested using *AGAMOUS* (*AG*) regulatory sequences cloned upstream of the *lacZ* reporter gene. As shown in Figure 1(e), the activity of LFY<sub>HARA</sub> was dramatically reduced compared to LFY, showing that HARA mutations strongly weaken the ability of LFY to bind to *AG* regulatory sequences.

#### LFY<sub>HARA</sub> is impaired in its floral meristem identity function

We next tested the functional relevance of LFY DBD dimerization in plants. LFY and LFY<sub>HARA</sub> coding sequences were placed under the control of the LFY endogenous promoter (*ProLFY*) in a *Ify-12* null mutant background. In 9 of 10 *Ify-12*<sup>-/-</sup> primary transformants, the *ProLFY:LFY* construct rescued the *Ify* mutant phenotype almost fully: flowers were present all along the main inflorescence and most displayed a wild-type appearance and correctly arranged floral organs (Figure 2c,e and Table S1). Although the *ProLFY:LFY<sub>HARA</sub>* transgene also induced flower formation in eight of nine *Ify-12*<sup>-/-</sup> primary transformants (Figure 2d and Table S1), flower development was systematically abnormal compared to wild-type or *ProLFY:LFY* plants. Petals and stamens were either missing or distorted, frequently resulting in sterile flowers (Figure 2e and Table S1). Occasionally, the shoot apical meristem produced a few fertile flowers on the lower positions of the inflorescence, and shoot-like structures at higher positions on the stem (Figure S1), revealing the weakened ability of LFY<sub>HARA</sub> to trigger flower development. These results indicate that the capacity of LFY to control flower formation is reduced (but not abolished) when the two amino acids, H387 and R390, are mutated.

#### LFY<sub>HARA</sub> over-expression triggers precocious meristem development in the axil of rosette leaves

Upon floral transition, several partially independent pathways converge to build flowers at the inflorescence meristem (Fornara *et al.*, 2010). To assess the function of LFY<sub>HARA</sub> independently of these other pathways, we ectopically expressed LFY and LFY<sub>HARA</sub> under control of the 35S

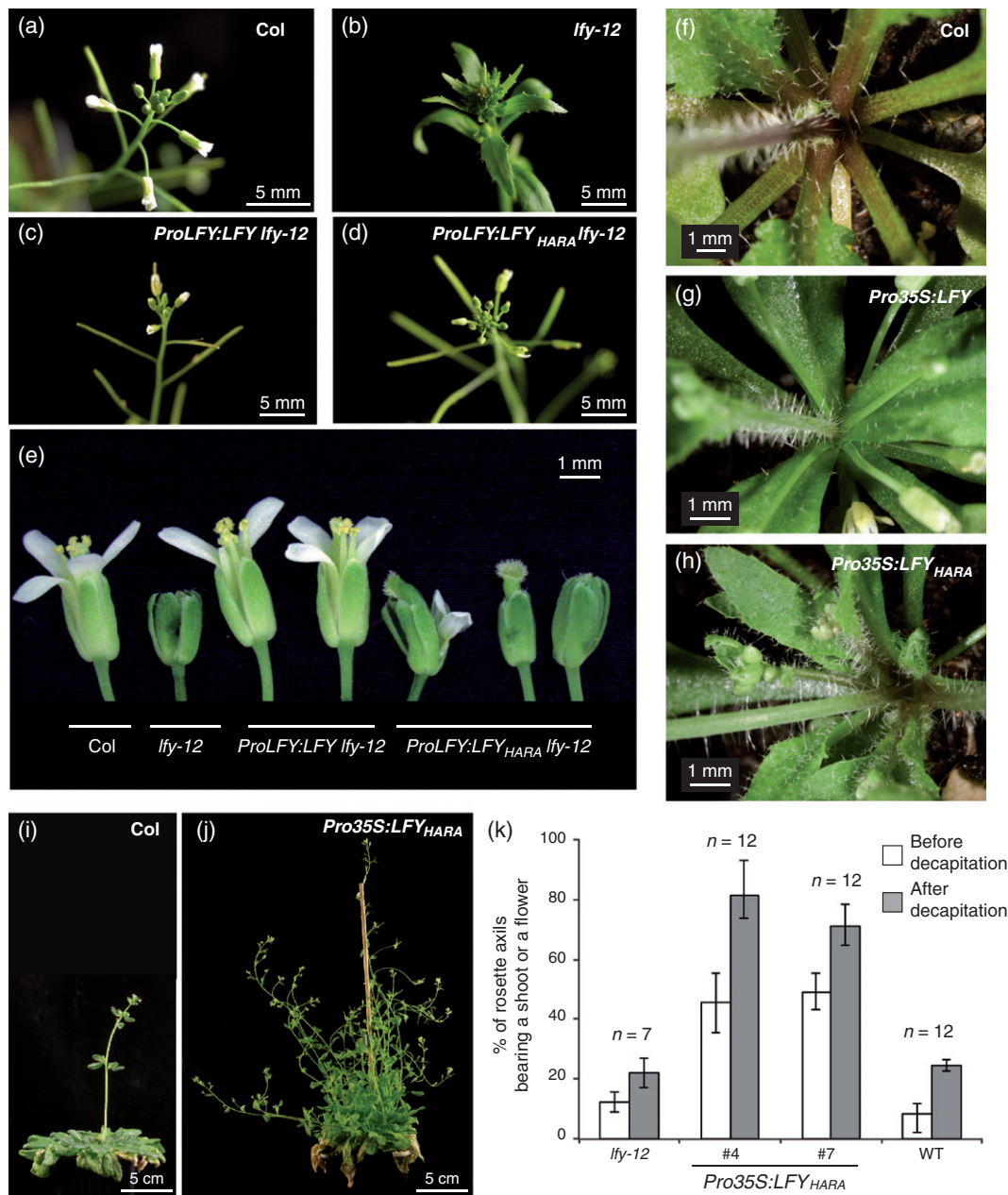
CaMV promoter (*Pro35S*). Under long-day conditions, 5-week-old wild-type plants displayed no visible axillary structures at the base of rosette leaves (Figure 2f). As previously described (Weigel and Nilsson, 1995), over-expression of the wild-type LFY cDNA induced formation of ectopic flowers in the axils of rosette and cauline leaves, as well as termination of the main shoot in a solitary flower (Figure 2g). In contrast, *Pro35S:LFY<sub>HARA</sub>* plants displayed no or very few ectopic flowers; instead, they showed precocious emergence of axillary shoots compared to wild-type plants (Figure 2h and Table S2). This effect was even more evident under short-day conditions, where *Pro35S:LFY<sub>HARA</sub>* plants became so bushy that the primary shoot was indistinguishable from the numerous secondary inflorescences (Figure 2i,j).

The growth of rosette axillary shoots in *Pro35S:LFY<sub>HARA</sub>* plants may be due to accelerated development of their meristems or insensitivity to the apical dominance imposed by the shoot apical meristem. In order to distinguish between these possibilities, we performed a decapitation experiment: we pruned the apical meristem to reduce apical dominance and counted visible axillary shoots after 10 days. In wild-type and *Ify-12* mutant plants, decapitation resulted in a two- to threefold increased proportion of rosette leaves bearing visible shoots at their axil (Figure 2k). *Pro35S:LFY<sub>HARA</sub>* plants also responded to decapitation to a similar extent, despite the fact they already displayed a high number of axillary shoots before decapitation (Figure 2k). This result shows that *Pro35S:LFY<sub>HARA</sub>* plants are still sensitive to apical dominance. Next we studied the time course of axillary meristem initiation. To do this, we monitored the expression of the *CLV3:GUS* meristematic marker (Brand *et al.*, 2002) during seedling development (Figure 3a–c). We found that, on average, the *CLV3:GUS* signal appeared 3.5 days earlier at the axil of rosette leaves in *Pro35S:LFY<sub>HARA</sub>* seedlings than in wild-type seedlings (Figure 3a). In representative 16-day-old wild-type seedlings, only the apical meristem was stained (Figure 3b), whereas in *Pro35S:LFY<sub>HARA</sub>* seedlings, several axillary meristems also showed *CLV3:GUS* expression (Figure 3c).

#### Wild-type LFY over-expression also accelerates axillary meristem development

To determine whether the wild-type LFY gene (and not only the LFY<sub>HARA</sub> allele) influences rosette axillary meristem development, we analyzed *Pro35S:LFY* plants. We used short-day conditions to delay precocious flowering, and included *Pro35S:AP1* plants, which share many features with *Pro35S:LFY* plants, such as conversion of axillary meristems into flowers (Mandel and Yanofsky, 1995). To determine the developmental rate of meristem initiation and growth at the axil of rosette leaves, we scored the developmental stage of the axillary meristem at each inter-

## 4 Hicham Chahtane et al.



**Figure 2.** *ProLFY:LFY* and *Pro35S:LFY* transgenic Arabidopsis.

(a–d) Inflorescences of wild-type Col-0 (a), *lfy-12* mutant (b), *ProLFY:LFY lfy-12* (c) and *ProLFY:LFY<sub>HARA</sub> lfy-12* (d) plants.

(e) Corresponding floral phenotypes. Some floral organs of *ProLFY:LFY lfy-12* flowers are distorted. Petal and stamen number are frequently reduced in the flowers of *ProLFY:LFY<sub>HARA</sub> lfy-12* plants.

(f–h) Close-up view of the rosette of 5-week-old wild-type and *Pro35S:LFY* plants. The axils of wild-type plants do not show any macroscopic structures (f), whereas flowers and shoots developed at the axils of *Pro35S:LFY* (g) or *Pro35S:LFY<sub>HARA</sub>* (h) rosette leaves, respectively.

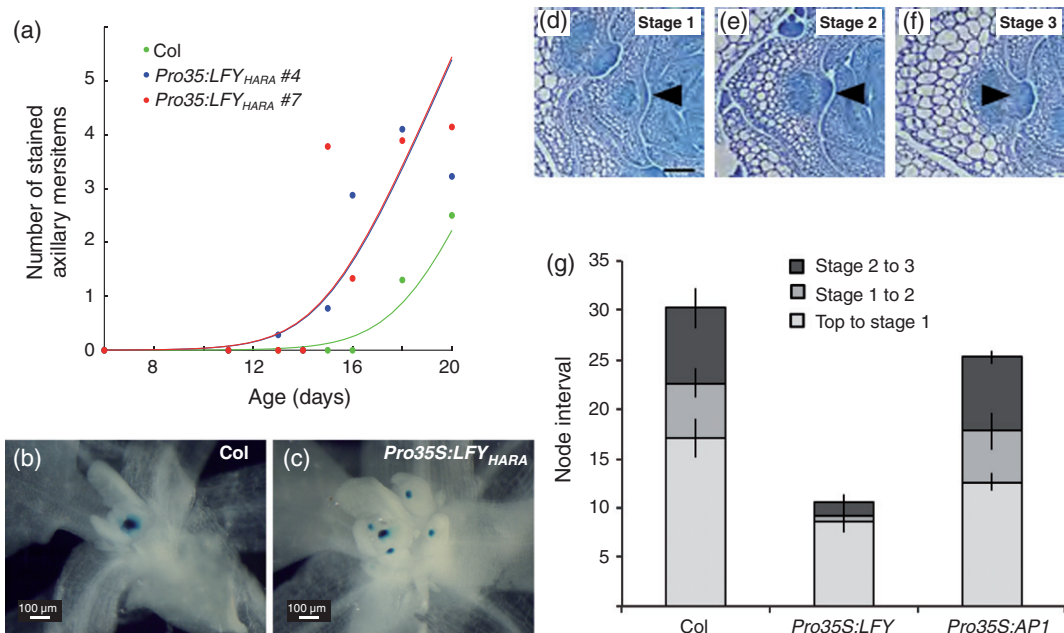
(i, j) Wild-type plant (i) and *Pro35S:LFY<sub>HARA</sub>* plant (j) cultivated for 2 months under short-day conditions.

(k) Percentage of rosette leaves bearing a flower or a shoot at their axil just before decapitation and 10 days after decapitation of the primary inflorescence. The number of individuals tested is indicated for each genotype, i.e. wild-type plants, *lfy-12* mutants and two independent homozygous *Pro35S:LFY<sub>HARA</sub>* lines (#4 and #7). Error bars represent 95% confidence intervals.

node as described previously (Stirnberg *et al.*, 2002). Stage 1 is when a group of densely stained cells become visible at the leaf axil; stage 2 is when meristematic dome forms at the leaf axil; stage 3 is when axillary meristems commence primordial initiation (Figure 3d–f). Plastochron for-

mation was used as the developmental time unit. Stage 1 was observed after 17 plastochrons in Col plants, but only after nine plastochrons (on average) in *Pro35S:LFY* plants. The difference between *Pro35S:LFY* and Col is even more striking for progression to stages 2 and 3. Axillary meris-





**Figure 3.** Characterization of axillary meristem development in *Pro35S:LFY*, *Pro35S:LFY<sub>HARA</sub>* and *Pro35S:AP1* plants. (a–c) *CLV3:GUS* expression in *Pro35S:LFY<sub>HARA</sub>* and wild-type seedlings grown under long-day conditions. (a) Kinetics of *CLV3:GUS* activity in axillary meristems during seedling development. The mean number of stained axillary meristems is indicated as green, blue and red circles for wild-type plants and *Pro35S:LFY<sub>HARA</sub>* lines (#4 and #7), respectively ( $n = 5–12$ ). A mathematical model was constructed to describe these data (see Experimental procedures). The corresponding green, blue and red fitting curves are shown. According to this model, the plants entered the phase of axillary meristem production at day 18.3 (17.2–19.5) for wild-type plants versus day 15.0 (14.2–15.8) and 14.9 (14.1–16.8) for *Pro35S:LFY<sub>HARA</sub>* lines #4 and #7, respectively (95% confidence intervals in parentheses). Representative 16-day-old wild-type (b) and *Pro35S:LFY<sub>HARA</sub>* (c) seedlings are shown. (d–g) Development of the axillary meristem in *Pro35S:LFY*, *Pro35S:AP1* and wild-type (Col) plants grown under short-day conditions. (d–f) Transverse sections of individual shoots illustrating the stages of axillary meristem development (arrowheads) according to Stirnberg *et al.* (2002). Scale bar = 20  $\mu\text{m}$ . (g) Mean number of node intervals, or plastochrons, between the shoot apex and the first stage 1 axillary meristem (top to stage 1), between the first stage 1 and the first stage 2 axillary meristems (stage 1 to 2), and between the first stage 2 and the first stage 3 axillary meristems (stage 2 to 3). Error bars indicate the standard error of the mean ( $n = 3–12$ ).

tems at stage 2 were frequently absent in *Pro35S:LFY* plants, and axillary meristems at stages 1 and 3 were seen in adjacent plastochrons. This contrasts with the slower development of axillary meristems in Col and *Pro35S:AP1* plants, where it took over five plastochrons (on average) to progress from stage 1 to stage 2, and 7.5 additional plastochrons to reach stage 3 (Figure 3g). This experiment showed that axillary meristems develop faster in *Pro35S:LFY* plants than in wild-type or *Pro35S:AP1* plants. Given that the plastochron rate is not slower in *35S:LFY* plants than in wild-type plants (Figure S2), this clearly demonstrates that LFY accelerates axillary meristem development. A further inspection of *Pro35S:LFY* plants revealed the occasional presence of meristems in the axils of cotyledons (Table S3), a structure commonly observed in the *Pro35S:LFY<sub>HARA</sub>* plants (Figure S3) or in mutants such as *branched1* (Aguilar-Martinez *et al.*, 2007), but extremely rare in wild-type plants.

#### The meristematic function of LEAFY requires DNA binding

The experiments described above suggest that LFY may have two distinct and sequential functions: first, LFY may trigger the development of meristems, and then convert

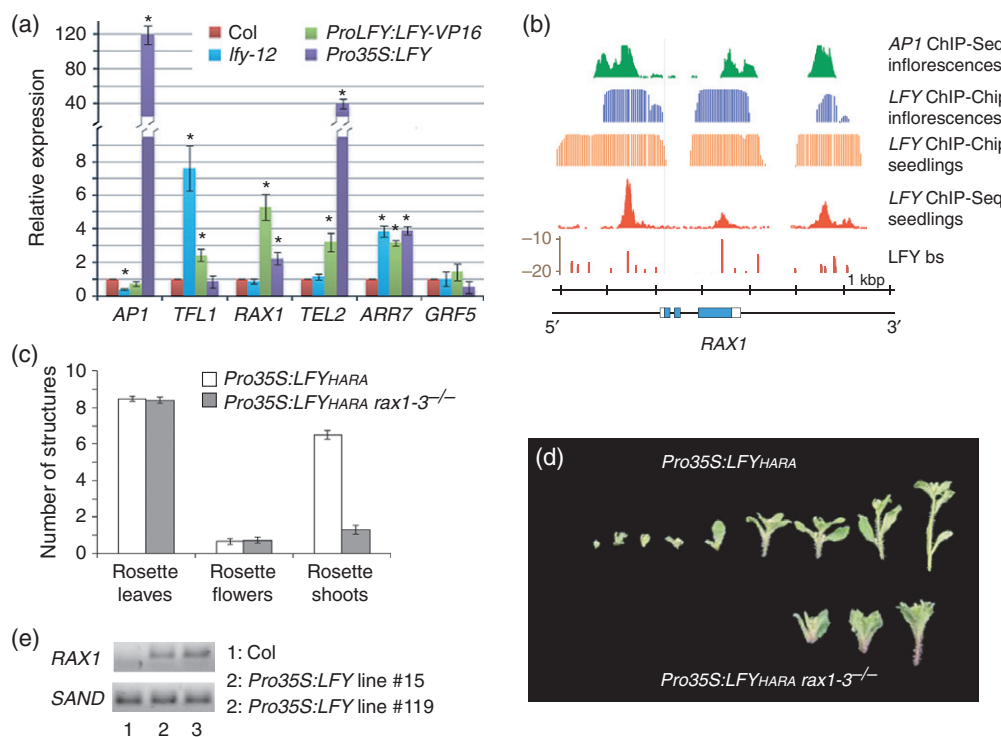
them into flowers. LFY DNA binding is essential for the later function, but whether it is required for the first role has never been tested. To answer this question, we took advantage of the crystallographic structure of LFY DBD (Hamès *et al.*, 2008) and specifically disrupted the interaction of LFY with DNA: in addition to H387 and R390 involved in dimerization, we mutated the two key amino acids (K307 and R237), which contact DNA bases directly, into alanines (Figure 1f). We confirmed *in vitro* that the resulting LFY<sub>HARAKARA</sub> mutant protein had lost its ability to specifically bind DNA sequences (Figure 1b). Next, we ectopically expressed LFY<sub>HARAKARA</sub> in transgenic Arabidopsis: these plants were indistinguishable from wild-type plants (Table S2), demonstrating that, as for the floral function, the meristematic function of LFY requires LFY DNA binding. This result suggests that LFY fulfils this function by directly regulating downstream target genes.

#### The LFY meristematic function involves up-regulation of RAX1

We next assessed through which pathway LFY mediates this function. Among genes bound by LFY in ChIP experiments (Moyroud *et al.*, 2011; Winter *et al.*, 2011), we found

several that are known to contribute to meristem or axillary meristem development: *REGULATOR OF AXILLARY MERISTEMS 1 (RAX1)* (Keller *et al.*, 2006; Muller *et al.*, 2006), participating in axillary meristem development, *TERMINAL EAR LIKE 2 (TEL2)* (Anderson *et al.*, 2004), which is probably involved in meristem maintenance, *ARABIDOPSIS RESPONSE REGULATOR (ARR7)*, which is involved in cytokinin signaling (Lee *et al.*, 2007), and *GROWTH REGULATOR FACTOR 5 (GRF5)* (Kim *et al.*, 2003; Horiguchi *et al.*, 2005). To determine whether these genes are genuinely regulated by LFY (and not just bound), we analyzed their transcript levels in various tissues and genetic backgrounds where LFY expression or activity was modified relative to wild-type plants. We used inflorescences from *lfy-12* and *ProLFY:LFY-VP16* plants (Parcy *et al.*, 1998) to assess regulation in floral tissues. We also used young seedlings of *Pro35S:LFY* transgenic plants to monitor gene expression in a context that minimizes the influence of

inflorescence-specific redundant pathways. In this assay, known LFY direct target genes such as *AP1*, an induced target, or *TERMINAL FLOWER 1 (TFL1)*, a repressed one, served as positive controls (Figure 4a) and behaved as expected from the published data (Parcy *et al.*, 1998, 2002; Ratcliffe *et al.*, 1999). We found that *RAX1* and *TEL2* were up-regulated in *ProLFY:LFY-VP16* and *Pro35S:LFY*, showing that LFY induces expression of these two genes in both floral and seedling tissues. *ARR7* was up-regulated in all three backgrounds (seedlings of *Pro35S:LFY* and inflorescences of *ProLFY:LFY-VP16* and *lfy-12*), suggesting that it is regulated by LFY (repressed in inflorescence tissues and induced in seedlings). However, *GRF5* was not affected in any of the three genetic backgrounds, despite strong binding by LFY. Among these candidates, *RAX1* was particularly interesting for several reasons: (i) the *RAX1* locus is bound by LFY in three regions (promoter, coding sequence and 3' untranslated region) in both *Pro35S:LFY* seedling



**Figure 4.** Identification of *RAX1* as a LFY target involved in the stimulation of the axillary meristem in *Pro35S:LFY<sub>HARA</sub>* plants.

(a) Expression of *AP1*, *TFL1*, *RAX1*, *TEL2*, *ARR7* and *GRF5* in wild-type, *lfy-12* mutant and *ProLFY:LFY-VP16* inflorescences, and in *Pro35S:LFY* and wild-type whole seedlings. Expression relative to wild-type is shown. Error bars indicate standard errors, and asterisks indicate statistically significant differences compared with wild-type ( $P$  value < 0.05) according to a randomization test (Pfaffl *et al.*, 2002).

(b) LFY and AP1 binding profiles at the *RAX1* locus. Coding and untranslated regions of the *RAX1* gene are shown as blue or white boxes, respectively. ChIP-Seq read coverage for LFY (Moyroud *et al.*, 2011) and AP1 (Kaufmann *et al.*, 2010) is shown in red and green, respectively, and ChIP/Chip data for LFY binding in inflorescences and seedlings (Winter *et al.*, 2011) are indicated in blue and orange, respectively. The LFY-binding sites (LFY bs) predicted by the LFY binding model (Moyroud *et al.*, 2011) are indicated as red bars (for scores above -20). The figure was created using the Integrated Genome Browser (Nicol *et al.*, 2009).

(c, d) A mutation in the *RAX1* gene partially reduces the number of precocious axillary structures in *Pro35S:LFY<sub>HARA</sub>*. (c) Rosette axillary structures of *Pro35S:LFY<sub>HARA</sub>* plants ( $n = 35$ ) and *Pro35S:LFY<sub>HARA</sub> rax1-3<sup>-/-</sup>* plants ( $n = 38$ ). Error bars indicate standard errors of the mean. (d) Representative *Pro35S:LFY<sub>HARA</sub>* and *Pro35S:LFY<sub>HARA</sub> rax1-3<sup>-/-</sup>* rosettes were dissected to show all axillary structures growing at the base of the rosette leaves.

(e) RT-PCR analysis of *RAX1* expression in rosette leaves of 1-month-old wild-type and *Pro35S:LFY* plants grown under long-day conditions. The *SAND* gene (*At2g28390*) was used as a constitutive control.

and wild-type inflorescence samples (Figure 4b) (Moyroud *et al.*, 2011; Winter *et al.*, 2011), (ii) the spatio-temporal expression pattern of *RAX1* in floral meristems correlates with *LFY* expression (Keller *et al.*, 2006; Muller *et al.*, 2006), (iii) *RAX1* is required for the emergence of axillary meristems in various conditions and genetic backgrounds (Keller *et al.*, 2006; Muller *et al.*, 2006; Yang *et al.*, 2012), and (iv) *RAX1* up-regulation in isolated *Pro35S:LFY* leaves (Figure 4e) indicates that *LFY* is sufficient for this induction: this regulation does not require any additional flower-specific co-factor and is direct enough to occur independently of meristem development. To determine whether *RAX1* contributes to meristem stimulation in *Pro35S:LFY<sub>HARA</sub>* plants, we characterized *Pro35S:LFY<sub>HARA</sub> rax1-3* plants. Under long-day conditions, the number of shoots emerging from *rax1-3 Pro35S:LFY<sub>HARA</sub>* rosette leaves was markedly decreased compared to *Pro35S:LFY<sub>HARA</sub>* plants (Figure 4c, d), showing that *LFY* induces the formation and growth of axillary meristems at least in part through induction of the *RAX1* gene.

## DISCUSSION

### LFY promotes the development of axillary meristems

*LEAFY* is a prominent floral meristem identity gene. It is highly expressed in the youngest incipient floral primordium to which it confers a floral fate. How meristems are turned into flowers is relatively well understood, but the molecular mechanism controlling their early emergence on the flanks of the shoot apical meristem has remained more elusive.

In this paper, we combine a variety of approaches to demonstrate that *LFY* is able to stimulate the growth of axillary meristems. We took advantage of knowledge of the *LFY* DBD 3D structure to modify its properties in a subtle and precise manner. By altering the capacity of the *LFY* DBD to dimerize on DNA, we engineered a modified version, *LFY<sub>HARA</sub>*, that partially fulfils the *LFY* floral function, revealing the ability of *LFY* to stimulate axillary bud outgrowth. This property is obvious in plants ectopically expressing *LFY<sub>HARA</sub>*: their architecture is profoundly altered (particularly under short-day conditions), and the *CLV3* meristematic marker is precociously expressed in the leaf axils.

Various lines of evidence demonstrate that this property is a native characteristic of wild-type *LFY* rather than a neomorphic feature of the *LFY<sub>HARA</sub>* allele. First, the accelerated meristem development was also observed in plants ectopically expressing the wild-type *LFY* gene. These plants sometimes develop meristems in the axil of the cotyledons, a feature that is not present in the wild-type plants. Second, we established that *LFY<sub>HARA</sub>* not only binds to the *AP1bs1* sequence in EMSA, but to the same set of DNA sequences as *LFY* in high-throughput SELEX. Third, we showed that a *LFY<sub>HARA</sub>* version that does not bind DNA (*LFY<sub>HARAKARA</sub>*) loses its capacity to stimulate meristem

growth. It is therefore likely that *LFY<sub>HARA</sub>* acts by binding to genuine *LFY* targets.

### LFY stimulates meristem growth through induction of RAX1

Our data indicate that the *RAX1* gene is an important direct *LFY* target that is required for stimulation of axillary meristem growth. Indeed, *RAX1* is strongly bound by *LFY* in ChIP experiments (both in *Pro35S:LFY* seedlings and wild-type inflorescences) (Moyroud *et al.*, 2011; Winter *et al.*, 2011), and its transcript level is increased in whole seedlings or isolated leaves of *Pro35S:LFY* plants and in *ProLFY:LFY-VP16* inflorescences. Moreover, the *rax1* mutation reduces the emergence of axillary shoots in *Pro35S:LFY<sub>HARA</sub>* plants, demonstrating that *LFY* induces the growth of axillary shoots through activation of *RAX1*.

### LFY meristematic function is cryptic but important during flower meristem development

The *lfy* mutation compromises the emergence of a wild-type floral meristem but does not lead to a naked pin-like shoot because leaf primordia with associated axillary meristems grow instead of flowers. Similarly, during plant vegetative growth, axillary meristems emerge in the absence of *LFY* expression: *LFY* is therefore not required for meristem development on the flanks of the shoot apical meristem as other pathways exist that also drive this process. This redundancy may explain why the role of *LFY* in the emergence of floral meristems has remained unnoticed. However, when the *lfy* mutation is combined with defects in other genes, the importance of *LFY* in this process becomes clear. *FIL* and *CLV1* are two genes that act during floral meristem development (Clark *et al.*, 1993; Chen *et al.*, 1999; Sawa *et al.*, 1999a,b). *fil* and *clv1* single mutants show some floral meristem initiation defects, with occasional empty cauline leaves or filamentous structures. The *lfy* mutation greatly enhances the *fil* and *clv1* floral defects. In double mutant plants, most floral primordia are replaced by empty cauline leaves or filaments, which are determinate structures lacking meristematic potential (Sawa *et al.*, 1999a). *PINOID* (*PID*) is another regulator (from the auxin signaling cascade) that contributes to meristem development (Bennett *et al.*, 1995; Christensen *et al.*, 2000). *pid* floral meristem defects are also strongly enhanced in the *pid lfy* double mutant even under growth conditions where the *pid* single mutant only displays a very mild phenotype (Ezhova *et al.*, 2000). Another striking example comes from analysis of *BOP* genes (Norberg *et al.*, 2005): empty cauline leaves replace some flowers in the *bop1 bop2 lfy* triple mutant but not in the *bop1 bop2* double mutant. Thus, *LFY* does stimulate floral meristem outgrowth, but this role is masked by redundant pathways.

Ectopic expression of *LFY* at a time when these pathways are less active revealed the importance of *RAX1*

induction by LFY for precocious emergence of meristems in the axils of rosette leaves. There is evidence that this induction also occurs in flowers: (i) *RAX1* mRNA is present in early stage 1 flower meristems just after LFY expression starts (Keller *et al.*, 2006; Muller *et al.*, 2006), (ii) LFY strongly binds the *RAX1* locus at three locations in inflorescence tissues (Winter *et al.*, 2011), and (iii) expression of LFY-VP16 induces *RAX1* expression in inflorescences (Figure 4a). However, expression of *RAX1* is not lower in the single *lfy* mutant, probably because it is also induced by other regulators such as the *LATERAL ORGAN FUSION* (*LOF*) genes (Lee *et al.*, 2009). Similarly, the single *rax1* mutation does not affect development of flowers but only suppresses some axillary meristems at the base of leaves from plants grown under short-day conditions. Genes such as *LATERAL SUPPRESSOR* (*LAS*) (Greb *et al.*, 2003) and *REGULATOR OF AXILLARY MERISTEM FORMATION* (*ROX*) (Yang *et al.*, 2012) have been shown to act in parallel with *RAX1* to stimulate meristem growth. Interestingly, these two genes are required for axillary meristem emergence but not for flowers, although they are both expressed in floral primordia: they are thus prominent candidates to act redundantly with LFY during the reproductive phase.

Down-regulation of *ARR7* expression by LFY (as indicated by analysis of the *lfy-12* mutant) may also contribute to meristem emergence. *ARR7* is a negative regulator of cytokinin signaling (Lee *et al.*, 2007). Cytokinin is a phytohormone that is known to promote axillary meristem growth (Tantikanjana *et al.*, 2001) and is directly repressed by *WUS* during early flower development (Leibfried *et al.*, 2005). As LFY and *WUS* interact to induce *AG* expression (Lohmann *et al.*, 2001), it is possible that they also collaborate to repress *ARR7*. A direct interaction between LFY and the *ARR7* promoter was observed both *in vitro* (Leibfried *et al.*, 2005) and *in vivo* in seedlings ectopically expressing LFY (Moyroud *et al.*, 2011; Winter *et al.*, 2011). In addition to the cytokinin pathway, auxin signaling may also be important downstream of LFY: several genes from the auxin pathway are bound by LFY in ChIP experiments (Moyroud *et al.*, 2011; Winter *et al.*, 2011), and *LAS* has been shown to act upstream of auxin components (Greb *et al.*, 2003). Meristem emergence may thus result from both convergence of LFY and *LAS* action on hormones and stimulation of the *RAX1* pathway.

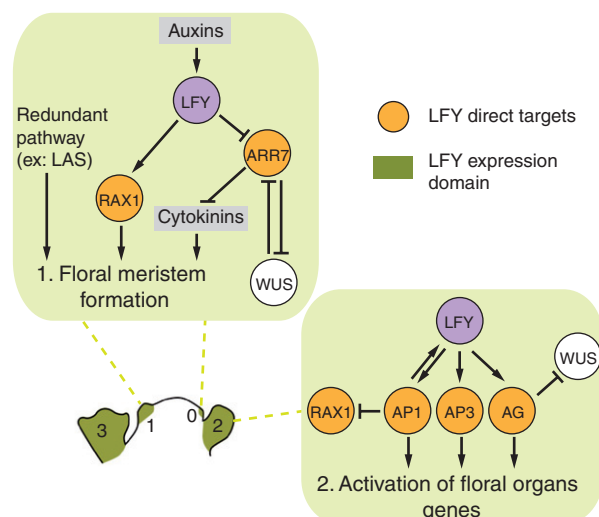
#### AP1 and LFY share the floral function, but only LFY stimulates meristem development

We observed that constitutive *AP1* expression only mildly affects meristem development compared to *LFY*. The fact that *AP1* does not participate in meristem outgrowth is supported by the phenotypes of double mutants such as *ap1 clv1*, *ap1 fil* or *ap1 bop1 bop2* that do not display the empty cauline leaves or filamentous structures observed in

double or triple mutants that include the *lfy* mutation (Clark *et al.*, 1996; Chen *et al.*, 1999; Xu *et al.*, 2010). This may seem surprising given that LFY and AP1 share many floral functions and bind to a common set of target genes, including the *RAX1* locus (Figure 4). However, the fact that *RAX1* is not induced but instead is repressed by AP1 (Kaufmann *et al.*, 2010) provides a striking explanation for the differential effect of these two floral meristem identity genes on meristem stimulation: AP1 probably contributes to the loss of meristematic character of the floral meristem established by LFY.

#### A developmental scenario for meristem emergence and floral fate determination

The data provided here and elsewhere suggests the following sequence of events for early flower development (Figure 5). The site of leaf or flower meristem emergence is determined by the auxin maximum (Kuhlemeier, 2007; Vernoux *et al.*, 2010). After the floral transition, auxin accumulation stimulates LFY expression in the floral anlage (stage 0). LFY contributes to outgrowth of the meristem and expression of meristematic genes such as *WUS*, *CLV3* and *UFO* by inducing *RAX1* and probably repressing *ARR7* together with *WUS*. Together with some of its own targets such as AP1 or LMI1 (Saddic *et al.*, 2006), LFY also confers floral fate to the nascent meristem (stage 2). Acquisition of floral identity involves the induction of floral homeotic genes as well as loss of inflorescence meristem genes through repression of *RAX1*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*), *AGAMOUS-LIKE24* (*AGL24*) and *SHORT VEGETATIVE PHASE* (*SVP*) by AP1 (Liu *et al.*, 2007; Kaufmann *et al.*, 2010) and of *WUS* by AG (Lohmann *et al.*, 2001). This scenario provides another



**Figure 5.** Speculative model in which the links between LFY, *RAX1* and *ARR7* are placed in the context of flower meristem development and integrated with other known regulation pathways. The numbers on the inflorescence section indicate floral stages.

illustration of the importance of the timing of these feed-forward loops for proper development of the flower. Controlling both the emergence of the meristem and its floral determination by the same regulator (LFY) may be an efficient way to couple both events and ensure they do not occur independently, when environmental conditions affect the cell division rate for example. Firm demonstration of the role of the LFY-RAX1 module in flowers will require further investigation in genetic contexts with reduced redundancy between pathways.

### The action of LFY on meristems explains the *lfy* phenotypes in various plants

Homologs of *LFY* are found in all angiosperms, but the phenotypes of *lfy* loss-of-function mutants differ in some species. Moyroud *et al.* (2009, 2010) proposed that a meristematic function of *LFY* provides a common explanation for the phenotypes observed in various species, such as the reduced number of flowers and tillers of the *rice flo-ricaula leafy (rfl)* mutant (Kyojuka *et al.*, 1998; Rao *et al.*, 2008; Ikeda-Kawakatsu *et al.*, 2012) or simple leaf development when the *LFY* ortholog is mutated in pea (*Pisum sativum*) or *Medicago truncatula* (Hofer *et al.*, 1997; Champagne *et al.*, 2007; Wang *et al.*, 2008). It will be interesting to determine whether the molecular mechanism unraveled in Arabidopsis is shared by other species, and whether homologs of *RAX1* contribute to rice tiller or inflorescence development or to leaf dissection in some legumes. In tomato, two *RAX1* co-orthologs, named *BLIND (BL)* and *POTATO LEAF (C)*, control shoot architecture and leaf development (Schmitz *et al.*, 2002; Busch *et al.*, 2011). Mutations of these genes reduce plant branching and leaf complexity. Interestingly, mutations in *FALSIFLORA (FA)*, the *LFY* ortholog in tomato (*Solanum lycopersicum*), also affect both processes (Molinero-Rosales *et al.*, 1999). These common phenotypes, together with the overlapping expression patterns of *FA* and *BL/C*, are consistent with possible regulation of *BL/C* genes by *FA* in tomato. In the moss *Physcomitrella patens*, *LFY* homologs were shown to act in the first cell division after fertilization (Tanahashi *et al.*, 2005). It is therefore possible that the ancestral function of *LFY* was to control of cell division, and that *LFY* specialized in the emergence of meristems before being co-opted to confer floral fate to the meristems it produced.

## EXPERIMENTAL PROCEDURES

### Plant material and growth conditions

The *Pro35S:AP1* seeds were a gift from Detlef Weigel (Max Planck Institute, Tübingen, Germany). The *ProCLV3:GUS* line (Brand *et al.*, 2002) and the *rax1-3* mutant T-DNA line (SALK\_071748) (Muller *et al.*, 2006) were obtained from the Nottingham Arabidopsis Stock Centre (<http://arabidopsis.info>). All mutants and transgenic lines used are in the *A. thaliana* Columbia-0 accession.

Seeds were sown on soil or surface-sterilized and grown in Petri dishes on Murashige and Skoog basal salt mixture medium (Sigma-Aldrich, [www.sigmaaldrich.com](http://www.sigmaaldrich.com)). Plants were transferred to soil and grown at 22°C under long-day conditions (16 h of 100  $\mu$ E light) or short-day conditions (8 h of 100  $\mu$ E light). The *ProLFY:LFY-VP16* plants have been described previously (Parcy *et al.*, 1998).

### Plant genotyping

Plants carrying the *lfy-12* mutation were genotyped using the end-point genotyping program with a Light Cycler 480 (Roche, <http://www.roche.com>) as described by Benlloch *et al.* (2011) or using the following dCAPS primers: forward (5'-TCAAGCACCTCCGGTTCCACCTCCA-3') and reverse (5'-CGGACGAAAACCTACGCTGAACCACCA-3'). Only the wild-type 85 bp PCR product is cleaved by the *Fnu4HI* restriction enzyme. The *RAX1* locus was identified by PCR using primers oHC150 (5'-TGTGAAAAGACCAACCTCACC-3') and oHC151 (5'-TCGGACATTTTCAGTTTGAAG-3') for *RAX1*, and the *rax1-3* allele was identified using primers oHC150 and oHC152 (5'-ATTTTGCCGATTTCCGGAAC-3').

### Plant vectors and transformation

The pFP100 binary vector (Bensmihen *et al.*, 2004), which contains the *ProAt2S3:GFP in planta* selection marker, was used as a backbone for the following binary constructs: *ProLFY:LFY* (pETH29), *ProLFY:LFY<sub>HARA</sub>* (pETH39), *Pro35S:LFY* (pCA26) and *Pro35S:LFY<sub>HARA</sub>* (pSP3). Details of the cloning procedures can be provided upon request. *Agrobacterium tumefaciens* C58 pMP90 was used for stable transformation of wild-type or *lfy-12* heterozygous plants.

### Yeast vectors and one-hybrid assay

For yeast experiments, the p424 vector with the *GAL1* promoter was used to clone *LFY<sub>HA</sub>-VP16* (pETH32), *LFY<sub>RA</sub>-VP16* (pETH46) and *LFY<sub>HARA</sub>-VP16* (pETH31) fusions either by conventional cloning or by using the megaprimer strategy (Kirsch and Joly, 1998). Yeast vectors containing *LFY* and *LFY:VP16* (pFP13 and pFP14, respectively) and the *AGAMOUS:LACZ* reporter construct (pFP50) have been described previously (Parcy *et al.*, 1998; Lohmann *et al.*, 2001). Detailed maps for all constructs are available on request. For one-hybrid experiments, effectors (pFP13, pFP14, pETH31, pETH32 or pETH46) and reporter plasmids (pFP50) were co-transformed into *Saccharomyces cerevisiae* strain EGY48 (Invitrogen, [www.lifetechnologies.com](http://www.lifetechnologies.com)).  $\beta$ -galactosidase measurements were performed as described previously (Lohmann *et al.*, 2001), except that they were adapted to the Safire2 microplate reader (Tecan, [www.tecan.com](http://www.tecan.com)).

### Expression and purification of recombinant LFY proteins

The pET-30a vector (Novagen, [www.merckmillipore.com](http://www.merckmillipore.com)) was used to clone recombinant near full-length LFY [LFY( $\Delta$ 40), pETH79], LFY<sub>HARA</sub>( $\Delta$ 40) (pETH156) and LFY<sub>HARAKARA</sub>( $\Delta$ 40) (pETH180). Because it greatly facilitated expression and purification of these LFY recombinant proteins, the non-conserved first 39 amino acids were omitted in these constructs. The proteins were expressed in the *Escherichia coli* Rosetta™2 (DE3) strain (Novagen, <http://www.emdmillipore.com/>), and purified on Ni Sepharose™ high-performance resin (GE Healthcare, [www.gehealthcare.com](http://www.gehealthcare.com)) as described in Methods S1.

### EMSA experiments

Electrophoretic mobility shift assay were performed as previously described (Moyroud *et al.*, 2011). AP1 oligonucleotides (5'-

GTTGGGAAGGACCAGTGGTCCGTACAATGT-3' and 5'-ACATTGTACGGACCACTGGTCTCCCAA-3') labeled with the Cy5 fluorophore were prepared as described by Moyroud *et al.* (2011). The binding reactions were performed over 15 min using 100, 500 or 1500 nM LFY in 20  $\mu$ l binding buffer (10 mM HEPES, 1 mM spermidine, 14 mM EDTA, 30  $\mu$ g BSA, 0.25% CHAPS, 3% glycerol, 3 mM tris(2-carboxyethyl)phosphine (TCEP), pH 8, and 28 ng/ml fish sperm DNA), and loaded on a native gel (6% acrylamide, 0.5  $\times$  TBE). Gels were electrophoresed at 90 V for 70 min at 4°C, and scanned on the Typhoon 9400 scanner (Amersham Biosciences, www.gehealthcare.com).

### Systematic evolution of ligands by exponential enrichment experiments

Systematic evolution of ligands by exponential enrichment experiments were performed using fluorescent 73-mers and LFY ( $\Delta$ 40) or LFY<sub>HARA</sub>( $\Delta$ 40) proteins. Initially, a random library was synthesized using 73-mers 5'-TGGAGAAGAGGAGATCTAGC(N)<sub>30</sub>CTTGTCTTCTCGATTCCGG-3' as template with a fluorescent TAMRA-labelled forward primer (SElex-F, 5'-TGGAGAAGAGGAGAG ATCTAG-3') and a non-labeled reverse primer (SElex-R, 5'-CCGG AATCGAAGAAGAACAA-3'), as previously described (Moyroud *et al.*, 2011).

For each selection cycle, 1  $\mu$ M protein was mixed with 10 nM fluorescent dsDNA in 225  $\mu$ l Selex buffer (20 mM Tris pH 8, 250 mM NaCl, 2 mM MgCl<sub>2</sub>, 5 mM TCEP, 60  $\mu$ g/ml fish sperm DNA and 1% glycerol). After a 15 min incubation, 25  $\mu$ l Ni-NTA magnetic beads (Qiagen, www.qiagen.com), previously equilibrated in Selex buffer, were added to the reaction mix to immobilize DNA/protein complexes via the His tag of the protein. After 30 min incubation, the reaction mix was placed on a tube magnet for 1 min, and the supernatant was removed from the separated beads to eliminate the unbound DNA. Five washes were subsequently performed at 4°C, each of them consisting of adding 250  $\mu$ l Selex buffer containing 20  $\mu$ g/ml fish sperm DNA, followed by 2 min incubation and 1 min on the tube magnet to discard the supernatant. Finally, the magnetic beads were resuspended in 50  $\mu$ l Selex buffer without fish sperm DNA. Selected 73-mers were amplified by PCR as described previously (Moyroud *et al.*, 2011) using 1  $\mu$ l of the magnetic beads solution as template. PCR products were quantified as previously described (Moyroud *et al.*, 2011), and the selection cycle was repeated three times, each time using the newly synthesized fluorescent DNA as a library. The 73-mers libraries obtained after four cycles of selection were subsequently used for barcoding and high-throughput sequencing.

### Barcoding for Illumina sequencing of the SELEX libraries

Oligonucleotides allowing hybridization with the sequencing primer fused to a 6 bp barcode (one unique barcode per library) were added by PCR (PCR#1) to 73-mers of each library selected, using Phusion<sup>®</sup> DNA polymerase (Ozyme, http://www.ozyme.fr) and 0.5–2  $\mu$ l of the magnetic beads solution as template, extending the 73-mers into 146 bp dsDNA. Three PCR cycles (PCR#2) were subsequently performed, using 5  $\mu$ l of PCR#1 product as template, Phusion<sup>®</sup> DNA polymerase (Ozyme, http://www.ozyme.fr) and Illumina oligonucleotides (http://www.illumina.com/), allowing fixation of the dsDNA to the Illumina sequencing chip. PCR products from PCR#2 (171 bp) were purified using a NucleoSpin Extract II kit (Macherey-Nagel, http://www.mn-net.com), and sequenced on an Illumina platform at the Max Planck Institute for Developmental Biology, Tübingen, Germany.

The 2000 unique most frequent sequences were aligned using MEME software version 4.3.0 (Bailey and Elkan, 1994) using default parameters and imposing symmetry to construct the LFY ( $\Delta$ 40) and the LFY<sub>HARA</sub>( $\Delta$ 40) matrix.

### Gene expression analysis

RNA extractions were prepared from 50 mg plant tissues using and RNeasy plant RNA extraction kit (Qiagen). Extractions were obtained from 21-day-old seedlings (Col-0 and *Pro35S:LFY* plants cultivated under short-day conditions) or young apices from which older flowers were removed (Col-0, *lfy-12* and *ProLFY:LFY-VP16* plants cultivated under long-day conditions). Total RNAs were treated with Ambion<sup>®</sup> turbo DNA-free<sup>™</sup> DNase (Invitrogen) to reduce genomic DNA contamination. Reverse transcription was performed using MMLV reverse transcriptase (Sigma-Aldrich) from 1  $\mu$ g of total RNA.

Primers used for real-time PCR were designed using QUANTPRIME software (www.quantprime.de/) and are listed in Table S4. PCR reactions were performed in a 15  $\mu$ l final volume using 0.1  $\mu$ M for each primer, 1  $\times$  mix SYBR Green JumpStart<sup>®</sup> (Sigma-Aldrich) and 3.125 ng cDNA in a Rotor Gene 3000 (Qiagen) with default cycle parameters. The relative expression of a target gene was calculated by the  $\Delta\Delta C_t$  method as described by Pfaffl (2001) using the reference gene At2g28390 as previously described (Czechowski *et al.*, 2005; Guenin *et al.*, 2009). Statistical analyses were performed using REST 2009 software (Pfaffl *et al.*, 2002).

### Decapitation experiment

Plants were grown for 1 month under short-day conditions and then transferred to long-day conditions. The primary inflorescence was cut once it reached 10–15 cm, i.e. after 20 days for *Pro35S:LFY<sub>HARA</sub>* or 26 days for wild-type and *lfy-12* plants. The number of secondary inflorescences was scored on the day of decapitation and 10 days later. Parametric statistics for the percentage of rosette leaves bearing a shoot or flower were applied after arcsine root transformation; 95% confidence intervals were found with the usual *t* distribution and then back-transformed to percentages.

### Histology

Tissues were fixed in formaldehyde/acetic acid/ethanol (FAA), and infiltration with paraffin was performed in an ASP300 embedding automat (Leica, http://www.leica.com). Thin sections (9–12  $\mu$ m) were prepared on an EG1160 microtome (Leica) and transferred to glass slides. Staining was performed with toluidine blue. Images were taken on an Axioplan2 (Zeiss, www.zeiss.com) equipped with an AxioCam HRc digital camera (Zeiss). GUS staining was performed as described previously (Parcy *et al.*, 1998).

### Model describing the appearance of the axillary meristem staining

We assumed that the probability of a plant having entered the phase of meristematic development (production of axillary meristem at the axil of rosette leaves) was a sigmoid function of time, characterized by its mean  $d_{\text{switch}}$  and its temporal uncertainty  $\tau$ ; thus this probability on day  $d$  was modeled as  $P$  (meristematic |  $d$ ) =  $\sigma((d - d_{\text{switch}})/\tau)$ , where  $\sigma$  is the logistic function:  $\sigma(x) = 1/[1 + \exp(-x)]$ . In the meristematic development phase, we assumed that a Poisson process of rate  $\lambda$  governed the appearance of meristems: thus the probability of observing  $n$  meristems  $t$  days after the switch to the meristematic phase is  $\pi(n, \lambda, t) = (\lambda t)^n/n! \exp(-\lambda t)$ ; before this meristematic phase,  $P$

( $n = 0$ ) = 1. With these assumptions, the overall probability of observing  $n > 0$  meristems on day  $d$  is

$$p(n|d) = 1/\tau \int_0^d \sigma^t \left[ \frac{t - d_{\text{switch}}}{\tau} \right] \pi(n, \lambda, d - t) dt$$

By maximizing the likelihood of the experimental meristem counts in this model, we derived the parameters  $d_{\text{switch}}$ ,  $\tau$  and  $\lambda$  for each strain. Likelihood ratio tests were then used to test the statistical significance of the parameter differences across strains. As  $\tau$  and  $\lambda$  were not significantly different across strains ( $P > 0.1$  for both), the model illustrated on Figure 3(a) was constrained to use a single value for each of these two parameters, with three values of  $d_{\text{switch}}$  (one for each strain).

## ACKNOWLEDGEMENTS

We thank R. Dumas, R. Benlloch (UPSC, Umeå, Sweden), O. Nilsson (UPSC, Umeå, Sweden), A. Mathelier (CMMT, Vancouver, Canada), W. Wasserman (CMMT, Vancouver, Canada), P. Baraduc (CNC, Lyon, France) and M. Nanao (EMBL, Grenoble, France) for help with experiment or data analysis, D. Weigel (MPI, Tuebingen, Germany) and N. Warthman (MPI, Tuebingen, Germany) for SELEX sequencing, P. Cubas (CNB, Madrid, Spain) and P. Doerner for discussion, and V. Pautot (INRA, Versailles, France), L. Yan (Dept. of Organismic and Evolutionary Biology Harvard University, USA) and P. Laufs (INRA, Versailles, France) for critical reading of the manuscript. This work is supported by grants from the Charmful ANR (Agence Nationale de la Recherche) program and by the Natural Sciences and Engineering Research Council of Canada to V.G.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Method S1.** Expression and purification of LFY proteins from bacteria.

**Figure S1.** Inflorescence phenotype of *ProLFY:LFY lfy-12* and *ProLFY:LFY<sub>HARA</sub> lfy-12* plants.

**Figure S2.** Production rate of rosette leaves in *Pro35S:LFY* and *Pro35S:AP1* plants.

**Figure S3.** Structures developing at the axils of cotyledons of *Pro35S:LFY* and *Pro35S:LFY<sub>HARA</sub>* plants.

**Table S1.** Floral phenotype of *lfy-12* plant complemented by *LFY* or *LFY<sub>HARA</sub>* cDNA.

**Table S2.** Distribution of phenotypic classes among *Pro35S:LFY* and *Pro35S:LFY<sub>HARA</sub> T<sub>1</sub>* plants.

**Table S3.** *CLV3:GUS* activity in axillary meristems of *Pro35S:LFY* seedlings.

**Table S4.** Oligonucleotides used for the quantitative PCR experiments.

## REFERENCES

- Aguilar-Martinez, J.A., Poza-Carrion, C. and Cubas, P. (2007) Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell*, **19**, 458–472.
- Anderson, G.H., Alvarez, N.D., Gilman, C., Jeffares, D.C., Trainor, V.C., Hanson, M.R. and Veit, B. (2004) Diversification of genes encoding mei2-like RNA binding proteins in plants. *Plant Mol. Biol.* **54**, 653–670.
- Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36.
- Benlloch, R., Kim, M.C., Sayou, C., Thevenon, E., Parcy, F. and Nilsson, O. (2011) Integrating long-day flowering signals: a LEAFY binding site is essential for proper photoperiodic activation of APETALA1. *Plant J.* **67**, 1094–1102.
- Bennett, S.R.M., Alvarez, J., Bossinegr, G. and Smyth, D.R. (1995) Morphogenesis in *pinoid* mutants of *Arabidopsis thaliana*. *Plant J.* **8**, 505–520.

- Bensmihen, S., To, A., Lambert, G., Kroj, T., Giraudat, J. and Parcy, F. (2004) Analysis of an activated *ABI5* allele using a new selection method for transgenic Arabidopsis seeds. *FEBS Lett.* **561**, 127–131.
- Blazquez, M.A., Ferrandiz, C., Madueno, F. and Parcy, F. (2006) How floral meristems are built. *Plant Mol. Biol.* **60**, 855–870.
- Brand, U., Grunewald, M., Hobe, M. and Simon, R. (2002) Regulation of CLV3 expression by two homeobox genes in Arabidopsis. *Plant Physiol.* **129**, 565–575.
- Busch, B.L., Schmitz, G., Rossmann, S., Piron, F., Ding, J., Bendahmane, A. and Theres, K. (2011) Shoot branching and leaf dissection in tomato are regulated by homologous gene modules. *Plant Cell*, **23**, 3595–3609.
- Champagne, C.E., Goliiber, T.E., Wojciechowski, M.F., Mei, R.W., Townsley, B.T., Wang, K., Paz, M.M., Geeta, R. and Sinha, N.R. (2007) Compound leaf development and evolution in the legumes. *Plant Cell*, **19**, 3369–3378.
- Chen, Q., Atkinson, A., Otsuga, D., Christensen, T., Reynolds, L. and Drews, G.N. (1999) The Arabidopsis *FILAMENTOUS FLOWER* gene is required for flower formation. *Development*, **126**, 2715–2726.
- Christensen, S.K., Dagenais, N., Chory, J. and Weigel, D. (2000) Regulation of auxin response by the protein kinase PINOID. *Cell*, **100**, 469–478.
- Clark, S.E., Running, M.P. and Meyerowitz, E.M. (1993) CLAVATA1, a regulator of meristem and flower development in Arabidopsis. *Development*, **119**, 397–418.
- Clark, S.E., Jacobsen, S.E., Levin, J.Z. and Meyerowitz, E.M. (1996) The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in Arabidopsis. *Development*, **122**, 1567–1575.
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. and Scheible, W.R. (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* **139**, 5–17.
- Ezhova, T.A., Soldatova, O.P., Kalinina, A. and Medvedev, S.S. (2000) Interaction of ABRUPTUS/PINOID and LEAFY genes during floral morphogenesis in *Arabidopsis thaliana* (L.) Heynh. *Genetika*, **36**, 1682–1687 (in Russian).
- Fornara, F., de Montaigu, A. and Coupland, G. (2010) SnapShot: control of flowering in Arabidopsis. *Cell*, **141**, 550, 550.e2
- Greb, T., Clarenz, O., Schafer, E., Muller, D., Herrero, R., Schmitz, G. and Theres, K. (2003) Molecular analysis of the *LATERAL SUPPRESSOR* gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. *Genes Dev.* **17**, 1175–1187.
- Guenin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C. and Gutierrez, L. (2009) Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J. Exp. Bot.* **60**, 487–493.
- Hames, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F., Martiel, J.L., Benlloch, R., Parcy, F. and Muller, C.W. (2008) Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* **27**, 2628–2637.
- Hofer, J., Turner, L., Hellens, R., Ambrose, M., Matthews, P., Michael, A. and Ellis, N. (1997) *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Curr. Biol.* **7**, 581–587.
- Horiguchi, G., Kim, G.T. and Tsukaya, H. (2005) The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J.* **43**, 68–78.
- Ikeda-Kawakatsu, K., Maekawa, M., Izawa, T., Itoh, J. and Nagato, Y. (2012) *ABERRANT PANICLE ORGANIZATION 2/RFL*, the rice ortholog of Arabidopsis *LEAFY*, suppresses the transition from inflorescence meristem to floral meristem through interaction with *AP01*. *Plant J.* **69**, 168–180.
- Irish, V.F. (2010) The flowering of Arabidopsis flower development. *Plant J.* **61**, 1014–1028.
- Kaufmann, K., Wellmer, F., Muino, J.M. et al. (2010) Orchestration of floral initiation by APETALA1. *Science*, **328**, 85–89.
- Keller, T., Abbott, J., Moritz, T. and Doerner, P. (2006) Arabidopsis REGULATOR OF AXILLARY MERISTEMS1 controls a leaf axil stem cell niche and modulates vegetative development. *Plant Cell*, **18**, 598–611.
- Kim, J.H., Choi, D. and Kende, H. (2003) The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. *Plant J.* **36**, 94–104.
- Kirsch, R.D. and Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.* **26**, 1848–1850.

## 12 Hicham Chahtane et al.

- Kuhlemeier, C. (2007) Phyllotaxis. *Trends Plant Sci.* **12**, 143–150.
- Kyoizuka, J., Konishi, S., Nemoto, K., Izawa, T. and Shimamoto, K. (1998) Down-regulation of RFL, the FLO/LFY homolog of rice, accompanied with panicle branch initiation. *Proc. Natl Acad. Sci. USA*, **95**, 1979–1982.
- Lebedeva, O.V., Ondar, U.N., Penin, A.A. and Ezhova, T.A. (2005) Effect of the *ABRUPTUS/PINOID* gene on expression of the *LEAFY* gene in *Arabidopsis thaliana*. *Genetika*, **41**, 559–565 (in Russian).
- Lee, D.J., Park, J.Y., Ku, S.J., Ha, Y.M., Kim, S., Kim, M.D., Oh, M.H. and Kim, J. (2007) Genome-wide expression profiling of *ARABIDOPSIS RESPONSE REGULATOR 7 (ARR7)* overexpression in cytokinin response. *Mol. Genet. Genomics*, **277**, 115–137.
- Lee, D.K., Geisler, M. and Springer, P.S. (2009) LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in *Arabidopsis*. *Development*, **136**, 2423–2432.
- Leibfried, A., To, J.P., Busch, W., Stehling, S., Kehle, A., Demar, M., Kieber, J.J. and Lohmann, J.U. (2005) WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. *Nature*, **438**, 1172–1175.
- Levin, J.Z., Fletcher, J.C., Chen, X. and Meyerowitz, E.M. (1998) A genetic screen for modifiers of UFO meristem activity identifies three novel *FUSED FLORAL ORGANS* genes required for early flower development in *Arabidopsis*. *Genetics*, **149**, 579–595.
- Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T. and Yu, H. (2007) Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. *Development*, **134**, 1901–1910.
- Liu, C., Thong, Z. and Yu, H. (2009) Coming into bloom: the specification of floral meristems. *Development*, **136**, 3379–3391.
- Lohmann, J.U., Hong, R.L., Hobe, M., Busch, M.A., Parcy, F., Simon, R. and Weigel, D. (2001) A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell*, **105**, 793–803.
- Mandel, M.A. and Yanofsky, M.F. (1995) A gene triggering flower formation in *Arabidopsis*. *Nature*, **377**, 522–524.
- Molinero-Rosales, N., JAMILINA, M., Zurita, S., Gomez, P., Capel, J. and Lozano, R. (1999) *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant J.* **20**, 685–693.
- Moyroud, E., Tichtinsky, G. and Parcy, F. (2009) The *LEAFY* floral regulators in angiosperms: conserved proteins with diverse roles. *J. Plant Biol.* **52**, 177–185.
- Moyroud, E., Kusters, E., Monniaux, M., Koes, R. and Parcy, F. (2010) *LEAFY* blossoms. *Trends Plant Sci.* **15**, 346–352.
- Moyroud, E., Minguet, E.G., Ott, F. et al. (2011) Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* *LEAFY* transcription factor. *Plant Cell*, **23**, 1293–1306.
- Muller, D., Schmitz, G. and Theres, K. (2006) Blind homologous R2R3 Myb genes control the pattern of lateral meristem initiation in *Arabidopsis*. *Plant Cell*, **18**, 586–597.
- Nicol, J.W., Helt, G.A., Blanchard, S.G. Jr, Raja, A. and Loraine, A.E. (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, **25**, 2730–2731.
- Norberg, M., Holmlund, M. and Nilsson, O. (2005) The *BLADE ON PETIOLE* genes act redundantly to control the growth and development of lateral organs. *Development*, **132**, 2203–2213.
- Parcy, F., Nilsson, O., Busch, M.A., Lee, I. and Weigel, D. (1998) A genetic framework for floral patterning. *Nature*, **395**, 561–566.
- Parcy, F., Bomblies, K. and Weigel, D. (2002) Interaction of *LEAFY*, *AGAMOUS* and *TERMINAL FLOWER1* in maintaining floral meristem identity in *Arabidopsis*. *Development*, **129**, 2519–2527.
- Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45.
- Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36.
- Pose, D., Yant, L. and Schmid, M. (2012) The end of innocence: flowering networks explode in complexity. *Curr. Opin. Plant Biol.* **15**, 45–50.
- Rao, N.N., Prasad, K., Kumar, P.R. and Vijayraghavan, U. (2008) Distinct regulatory role for RFL, the rice LFY homolog, in determining flowering time and plant architecture. *Proc. Natl Acad. Sci. USA*, **105**, 3646–3651.
- Ratcliffe, O.J., Bradley, D.J. and Coen, E.S. (1999) Separation of shoot and floral identity in *Arabidopsis*. *Development*, **126**, 1109–1120.
- Ruiz-Garcia, L., Madueno, F., Wilkinson, M., Haughn, G., Salinas, J. and Martinez-Zapater, J.M. (1997) Different roles of flowering-time genes in the activation of floral initiation genes in *Arabidopsis*. *Plant Cell*, **9**, 1921–1934.
- Running, M.P., Fletcher, J.C. and Meyerowitz, E.M. (1998) The *WIGGUM* gene is required for proper regulation of floral meristem size in *Arabidopsis*. *Development*, **125**, 2545–2553.
- Saddic, L.A., Huvermann, B., Bezhani, S., Su, Y., Winter, C.M., Kwon, C.S., Collum, R.P. and Wagner, D. (2006) The *LEAFY* target LMI1 is a meristem identity regulator and acts together with *LEAFY* to regulate expression of *CAULIFLOWER*. *Development*, **133**, 1673–1682.
- Sawa, S., Ito, T., Shimura, Y. and Okada, K. (1999a) FILAMENTOUS FLOWER controls the formation and development of *Arabidopsis* inflorescences and floral meristems. *Plant Cell*, **11**, 69–86.
- Sawa, S., Watanabe, K., Goto, K., Liu, Y.G., Shibata, D., Kanaya, E., Morita, E.H. and Okada, K. (1999b) *FILAMENTOUS FLOWER*, a meristem and organ identity gene of *Arabidopsis*, encodes a protein with a zinc finger and HMG-related domains. *Genes Dev.* **13**, 1079–1088.
- Schmitz, G., Tillmann, E., Carriero, F., Fiore, C., Cellini, F. and Theres, K. (2002) The tomato *Blind* gene encodes a MYB transcription factor that controls the formation of lateral meristems. *Proc. Natl Acad. Sci. USA*, **99**, 1064–1069.
- Siriwardana, N.S. and Lamb, R.S. (2012) A conserved domain in the N-terminus is important for *LEAFY* dimerization and function in *Arabidopsis thaliana*. *Plant J.* **71**, 736–749.
- Stirnberg, P., van De Sande, K. and Leyser, H.M. (2002) MAX1 and MAX2 control shoot lateral branching in *Arabidopsis*. *Development*, **129**, 1131–1141.
- Tanahashi, T., Sumikawa, N., Kato, M. and Hasebe, M. (2005) Diversification of gene function: homologs of the floral regulator *FLO/LFY* control the first zygotic cell division in the moss *Physcomitrella patens*. *Development*, **132**, 1727–1736.
- Tantikanjana, T., Yong, J.W., Letham, D.S., Griffith, M., Hussain, M., Ljung, K., Sandberg, G. and Sundaresan, V. (2001) Control of axillary bud initiation and shoot architecture in *Arabidopsis* through the *SUPERSHOOT* gene. *Genes Dev.* **15**, 1577–1588.
- 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.
- Vernoux, T., Besnard, F. and Traas, J. (2010) Auxin at the shoot apical meristem. *Cold Spring Harb. Perspect. Biol.* **2**, a001487.
- Wagner, D. (2009) Flower morphogenesis: timing is key. *Dev. Cell*, **16**, 621–622.
- Wang, H., Chen, J., Wen, J., Tadege, M., Li, G., Liu, Y., Mysore, K.S., Ratet, P. and Chen, R. (2008) Control of compound leaf development by *FLORICAULA/LEAFY* ortholog *SINGLE LEAFLET1* in *Medicago truncatula*. *Plant Physiol.* **146**, 1759–1772.
- Weigel, D. and Nilsson, O. (1995) A developmental switch sufficient for flower initiation in diverse plants. *Nature*, **377**, 495–500.
- Winter, C.M., Austin, R.S., Blanvillain-Baufume, S. et al. (2011) *LEAFY* target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell*, **20**, 430–443.
- Xu, M., Hu, T., McKim, S.M., Murmu, J., Haughn, G.W. and Hepworth, S.R. (2010) *Arabidopsis* *BLADE-ON-PETIOLE1* and 2 promote floral meristem fate and determinacy in a previously undefined pathway targeting *APETALA1* and *AGAMOUS-LIKE24*. *Plant J.* **63**, 974–989.
- Yalovsky, S., Kulukian, A., Rodriguez-Concepcion, M., Young, C.A. and Griessem, W. (2000) Functional requirement of plant farnesyltransferase during development in *Arabidopsis*. *Plant Cell*, **12**, 1267–1278.
- Yang, F., Wang, Q., Schmitz, G., Muller, D. and Theres, K. (2012) The bHLH protein ROX acts in concert with RAX1 and LAS to modulate axillary meristem formation in *Arabidopsis*. *Plant J.* **71**, 61–70.
- Zhao, Y., Granas, D. and Stormo, G.D. (2009) Inferring binding energies from selected binding sites. *PLoS Comput. Biol.* **5**, e1000590.





## Résumé

LEAFY (LFY) est un facteur de transcription central pour le développement des plantes, en particulier pour la floraison chez les angiospermes. LFY est très conservée, même chez les espèces ne portant pas de fleurs. On dispose de nombreuses données génétiques sur LFY et son réseau de régulation chez la plante modèle *Arabidopsis thaliana*, mais les mécanismes moléculaires impliqués dans son fonctionnement ne sont pas entièrement élucidés. LFY possède deux domaines conservés : un domaine de liaison à l'ADN et un domaine de fonction inconnue en position N-terminal. L'objectif a été de comprendre le rôle du domaine N-terminal et d'étudier l'évolution de la spécificité de liaison à l'ADN de LFY.

Nous avons obtenu la structure cristallographique du domaine N-terminal de LFY et découvert qu'il s'agissait d'un domaine SAM (Sterile Alpha Motif) permettant l'oligomérisation de la protéine. Nous avons validé l'importance de cette propriété pour la fonction florale de LFY chez *A. thaliana*. Nous avons ensuite montré, par des analyses *in vitro* et *in vivo* en CHIP-seq que l'oligomérisation influençait la liaison à l'ADN en permettant une liaison coopérative sur plusieurs sites de liaison, en assurant la sélectivité de la protéine vis-à-vis de l'ADN et en permettant l'accès de la protéine à des régions génomiques où la conformation de la chromatine est normalement défavorable à la liaison. Cette étude intégrative a permis de mieux comprendre le fonctionnement de LFY.

Des modifications dans les réseaux de régulation de l'expression des gènes sont source de nouveauté et d'évolution. LFY étant très conservée et ne faisant pas partie d'une famille multigénique, nous nous sommes demandé si sa spécificité de liaison à l'ADN avait évolué. Nous avons montré que LFY était apparue chez les algues multicellulaires et que sa spécificité avait connue au moins deux changements majeurs au cours de l'évolution. Nous avons expliqué ces modifications au niveau moléculaire par des approches de biologie structurale et de biochimie. Nous avons identifié une espèce chez qui LFY a une spécificité relâchée et nous proposons qu'une telle forme ait pu permettre les transitions d'une spécificité à une autre.

## Abstract

LEAFY (LFY) is a key transcription factor for plant development, particularly for flowering in angiosperms. LFY is highly conserved in plants, including non-flowering species. Despite a wealth of genetic data about LFY and its regulatory network in the model plant *Arabidopsis thaliana*, how the protein works at the molecular level is not fully understood. It has two conserved domains: a DNA binding domain and a N-terminal domain of unknown function. My two main projects were to understand the role of the N-terminal domain and to study LFY DNA binding specificity evolution.

We obtained LFY N-terminal domain crystal structure and discovered it was a Sterile-Alpha-Motif (SAM) mediating LFY oligomerisation. We validated the importance of that property for flower development in *A. thaliana*. Using both *in vitro* analyses and a ChIP-seq experiment, we pointed out that oligomerisation is required for proper DNA binding. It enables cooperative binding on several LFY binding sites, increases the protein selectivity towards DNA and allows LFY to access genomic regions where the chromatin conformation normally prevents binding. This integrative study provides a better understanding of how LFY works.

The rewiring of transcriptional networks provides a rich source of evolutionary novelty. As LFY is highly conserved and single copy in most plant genomes, we asked whether its DNA binding specificity had evolved. We showed that LFY was present since multicellular algae and that it underwent at least two major shifts in DNA-binding specificity during plant evolution. We provided a structural explanation for the two newly identified DNA binding modes and we identified a LFY form with a relaxed specificity that could have served as an intermediate between evolutionary transitions.