



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*

Discipline ou spécialité : *Biosciences végétales*

Présentée et soutenue par

May HIJAZI

Le 15 décembre 2011

Titre :

Caractérisation structurale et fonctionnelle d'AGP31, une glycoprotéine atypique de la paroi chez Arabidopsis thaliana

JURY

M. LEONARD Renaud - Maître de conférences, Université Lille I

Mme RALET Marie-Christine - Directeur de recherche INRA

Mme ALBENNE Cécile - Maître de conférences, Université Toulouse III

M. ARLAT Matthieu - Professeur, Université Toulouse III

Mme JAMET Elisabeth - Directeur de recherche CNRS (membre invité)

Ecole doctorale : *Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)*

Unité de recherche : *Laboratoire de Recherche en Sciences Végétales (LRSV), UMR 5546-UPS/CNRS*

Directeur(s) de Thèse : *Mme JAMET Elisabeth et Mme ALBENNE Cécile*

Rapporteurs : *M. LEONARD Renaud, Mme RALET Marie-Christine et M. VAN CUTSEM Pierre*

REMERCIEMENTS

Au terme de ces trois années de thèse passées au LRSV, j'ai le plaisir d'exprimer ma reconnaissance et mes remerciements à tous ceux qui ont contribué de près ou de loin au bon déroulement de ma thèse.

*Tout d'abord, mes remerciements s'adressent à mon directeur de thèse, **Elisabeth Jamet**, pour son aide et son encouragement prodigués tout au long de ce travail de recherche, ainsi que pour sa grande disponibilité, sa rigueur scientifique, son enthousiasme et ses précieux conseils qui m'ont permis de travailler dans les meilleures conditions. Je remercie aussi mon co-directeur de thèse, **Cécile Albenne**, pour m'avoir suivie au cours de ces années de thèse avec un encadrement rigoureux et de qualité. Je la remercie pour sa confiance, sa gentillesse ainsi que pour son regard avisé sur mes travaux.*

*Je tiens aussi à remercier ceux qui m'ont fait l'honneur de juger ce travail : **M. Renaud Leonard**, **Mme Marie-Christine Ralet** et **M. Pierre Van Cutsem** pour avoir accepté de juger ce travail et qui ont assumé la tâche de rapporteur de cette thèse, ainsi que les autres membres du jury pour avoir bien voulu prendre part à ce jury.*

*J'adresse mes sincères remerciements à l'ensemble des membres de mon équipe d'accueil, en particulier **Vincent Burlat**, pour ses hautes compétences scientifiques qui m'ont été d'une aide très précieuse pour avancer dans ma thèse. Je le remercie pour toutes les discussions, suggestions et ses précieux conseils. Merci à **David Roujol** qui, pendant mon congé de maternité, m'a apporté un grand soutien. Son aide permanente et ses compétences m'ont permis d'avancer dans ce travail.*

*Un grand merci aux services communs du laboratoire et à la plateforme de microscopie, en particulier **Patrick, Jean-Louis, Patricia, Alain et Gisèle**.*

*Mes remerciements les plus profonds vont à ma famille. Merci à mon mari **Hachem** pour son soutien et ses encouragements et ma fille **Célia** pour sa patience face à une « maman étudiante ». Merci à mes parents qui m'ont offert la possibilité de poursuivre et de terminer mes études. Merci à mes frères **Mahmoud** et **Wassim** et à ma sœur **Mona**, leurs petites attentions au téléphone me donnaient du courage pour mener à bien mon travail.*

Pour finir, il m'est très agréable d'exprimer toute ma gratitude, ma reconnaissance et mes très vifs remerciements à l'association écologique libanaise qui m'a soutenu financièrement.

SOMMAIRE

REMERCIEMENTS	1
SOMMAIRE	2
LISTES DES FIGURES ET DES TABLEAUX	6
ABREVIATIONS	9
CHAPITRE 1. Introduction	12
1.1 Organisation anatomique et fonctions de la paroi végétale	12
1.2 Architecture moléculaire de la paroi primaire	13
1.2.1 Les polysaccharides	13
1.2.1.1 La cellulose	14
1.2.1.2 Les hémicelluloses	14
1.2.1.2.1 Les xyloglucanes	15
1.2.1.2.2 Les xylanes	16
1.2.1.2.3 Les β -(1-4)(1-3) glucanes	16
1.2.1.3 Les pectines	16
1.2.1.3.1 L'homogalacturonane (HG) ou acide homogalacturonique	17
1.2.1.3.2 Les galacturonanes substitués	18
1.2.1.3.3 Le rhamnogalacturonane I (RG I)	18
1.2.1.3.4 Le rhamnogalacturonane II (RG II)	19
1.2.2 Les protéines	19
1.2.3 Organisation de la paroi primaire	21
1.2.4 Les glycoprotéines riches en hydroxyproline (HRGP)	21
1.2.4.1 Les protéines riches en proline (PRP)	21
1.2.4.2 Les extensines	22
1.2.4.3 Les AGP	23
1.3 Structure et fonctions des AGP	24
1.3.1 Structure des AGP	24
1.3.1.1 Le corps protéique	24
1.3.1.2 La partie glycanique des AGP	26
1.3.2 Les outils et stratégies utilisés pour l'étude fonctionnelle des AGP	29
1.3.3 Mécanismes d'action des AGP	30
1.3.4 Fonction biologique des AGP	31
1.4 Contexte et objectifs du travail de thèse	32
1.4.1 Contexte scientifique	32
1.4.2 AGP31 et ses orthologues : état de l'art	32
1.4.3 Objectifs de la thèse	34
CHAPITRE 2. Matériels et méthodes	36
2.1 Matériel	36
2.1.1 Matériel végétal	36
2.1.1.1 Plantes	36
2.1.1.2 Conditions de culture	37
2.1.2 Matériel bactérien	38
2.1.2.1 Plasmides et vecteurs de clonage	38
2.1.2.2 Souches bactériennes	39
2.1.2.3 Conditions de culture	40

2.1.2.4 Milieux de culture	40
2.1.3 Composés chimiques	40
2.2 Méthodes	41
2.2.1 Expression des protéines recombinantes chez les plantes	41
2.2.1.1 Clonage dans le vecteur Gateway® et expression des protéines recombinantes AGP31 et PAC chez <i>A. thaliana</i> et <i>N. benthamiana</i>	41
2.2.1.2 Transformations génétiques	41
2.2.1.3 Etude de la ségrégation du transgène et sélection des lignées homozygotes	43
2.2.1.4 Méthodes d'analyse des transcrits	43
2.2.2 Expression des protéines recombinantes chez <i>E. coli</i>	45
2.2.2.1 Clonage dans le vecteur pBAD-TOPO et expression des protéines recombinantes PAC chez <i>E. coli</i>	46
2.2.2.2 Clonage dans le vecteur pMAL-c2 et expression des protéines recombinantes PAC chez <i>E. coli</i>	46
2.2.3 Méthodes d'extraction des protéines totales	47
2.2.3.1 Extraction rapide de protéines bactériennes ou végétales pour analyse par SDS-PAGE ou <i>lectin blot</i>	47
2.2.3.2 Extraction des protéines pariétales	47
2.2.3.3 Extraction des protéines recombinantes d' <i>E. coli</i>	48
2.2.4 Techniques d'analyse des protéines	48
2.2.4.1 Dosage des protéines	48
2.2.4.2 Détection des protéines	49
2.2.4.3 Identification des protéines par spectrométrie de masse	51
2.2.5 Purification des protéines	51
2.2.5.1 Purification par le système FPLC (<i>Fast Protein Liquid Chromatography</i>)	52
2.2.5.2 Chromatographie d'affinité sur colonne Ni-NTA	53
2.2.5.3 Chromatographie d'affinité sur colonne de lectine	53
2.2.6 Tests d'interactions protéines/polysaccharides	54
2.2.6.1 Extraction des polysaccharides pariétaux	54
2.2.6.2 Digestion des hémicelluloses par la protéinase K	56
2.2.6.3 Mise en œuvre des tests d'interactions sur membrane	56
2.2.7 Protocoles de marquage pour la microscopie	56
2.2.7.1 Préparation des échantillons inclus dans la paraffine	56
2.2.7.2 Hybridation <i>in situ</i>	57
2.2.7.3 Etude cyto-histochimique des coupes de paraffine par la lectine PNA	60
CHAPITRE 3. Caractérisation structurale d'AGP31	61
3.1 Résumé de l'article "Characterization of the <i>O</i> -glycosylation heterogeneity of the arabinogalactan protein 31 (AGP31) of <i>Arabidopsis thaliana</i> : Evidence for new Hyp- <i>O</i> -Gal-rich motifs	61
3.2 Article	63
CHAPITRE 4. Recherche des partenaires d'AGP31 par des tests d'interaction <i>in vitro</i>	81
4.1 Les partenaires possibles d'AGP31 dans les parois végétales	81
4.1.1 Les partenaires des AGP	81
4.1.2 Les protéines à domaine riche en His	82

4.1.3 Les protéines à Cys conservées	82
4.1.4 Analyse bioinformatique de la structure d'AGP31 : recherche de domaines désordonnés	83
4.1.5 Stratégie de recherche de partenaires interagissant avec AGP31 dans les parois végétales	84
4.2 Production et purification d'AGP31 et de son domaine PAC	85
4.2.1 Purification de la protéine AGP31 native et recombinante	85
4.2.2 Production et purification du domaine PAC d'AGP31 recombinant	86
4.2.2.1 Optimisation de la production et de la purification du domaine PAC recombinant exprimé chez <i>E. coli</i>	86
4.2.2.2 Optimisation de la production et de la purification du domaine PAC recombinant exprimé dans des feuilles de <i>Nicotiana benthamiana</i>	89
4.3 Réalisation et validation des puces à polysaccharides	91
4.4 Recherche de partenaires d'AGP31 dans les parois végétales	92
4.4.1 Interactions d'AGP31 avec des polysaccharides pariétaux	93
4.4.2 Interactions du domaine PAC d'AGP31 avec des polysaccharides pariétaux	96
4.4.2.1 Tests d'interaction réalisés avec MBP/PAC/V5/6xHis produite dans <i>E. coli</i>	96
4.4.2.2 Tests d'interaction réalisés avec PAC/V5/6xHis produite dans <i>N. benthamiana</i>	97
4.4.3 Spécificité des interactions entre AGP31 et le galactane et rôle du domaine PAC	98
4.4.4 Mise en évidence d'interactions entre AGP31 et le domaine PAC	100
4.5 Discussion et perspectives	101
CHAPITRE 5. Caractérisation fonctionnelle <i>in planta</i> du gène <i>At1g28290</i> codant AGP31	104
5.1 Caractéristiques de la croissance des plantules étiolées d' <i>A. thaliana</i> et particularités de la mise en place des parois cellulaires	104
5.1.1 Développement des plantules d' <i>A. thaliana</i> à l'obscurité	104
5.1.2 Mise en place de la paroi des cellules des hypocotyles lors de l'élongation cellulaire	105
5.1.3 Stratégies mises en place pour l'étude fonctionnelle d' <i>At1g28290</i>	106
5.2 Analyse du patron d'expression d' <i>At1g28290</i> au cours du développement	107
5.2.1 Obtention et analyse de plantes <i>pAt1g28290::GFP::GUS</i>	107
5.2.2 Analyse <i>in silico</i> du profil d'expression d' <i>At1g28290</i> au cours du développement	107
5.2.3 Analyse de l'accumulation des transcrits d' <i>At1g28290</i> par RT-PCR quantitative au cours de la germination à l'obscurité	108
5.2.4 Mesure de l'accumulation d'AGP31 au cours de la germination	109
5.2.5 Identification des cellules exprimant <i>At1g28290</i> par hybridation d'ARN <i>in situ</i> et par marquage histochimique d'AGP31 dans des plantules sauvages au cours de la germination	110
5.2.6 Conclusion	112
5.3 Caractérisation de lignées sous-expresses d' <i>At1g28290</i> (lignées ARNi) et recherche de phénotypes au cours du développement précoce à l'obscurité	113
5.3.1 Obtention de lignées ARNi sous-expresses d' <i>At1g28290</i>	113

5.3.2 Analyse moléculaire des lignées ARNi sous-expresseurs au cours du développement précoce en conditions d'étiollement	114
5.3.3 Identification des cellules exprimant <i>At1g28290</i> par hybridation d'ARN <i>in situ</i> et par marquage histochimique d'AGP31 dans des plantules ARNi sous-expresseurs au cours de la germination	116
5.4 Caractérisation de lignées sur-expresseurs d' <i>At1g28290</i> et recherche de phénotypes au cours du développement précoce à l'obscurité	117
5.4.1 Obtention de lignées sur-expresseurs d' <i>At1g28290</i>	117
5.4.2 Analyse moléculaire des lignées sur-expresseurs au cours du développement en conditions d'étiollement	118
5.4.3 Identification des cellules exprimant <i>At1g28290</i> par hybridation d'ARN <i>in situ</i> et par marquage histochimique d'AGP31 dans des plantules ARNi sur-expresseurs au cours de la germination à l'obscurité	119
5.5 Analyse phénotypique des plantes ARNi sous-expresseurs et sur-expresseurs d' <i>At1g28290</i> au cours du développement	120
5.6 Discussion	121
CHAPITRE 6. Conclusions et perspectives	125
REFERENCES	130
ANNEXES	149

LISTE DES FIGURES ET DES TABLEAUX

Chapitre 1

Figure 1.1 : Paroi des cellules végétales montrant ses différents niveaux d'organisation	12
Figure 1.2 : Représentation de la molécule de cellulose	14
Figure 1.3 : Structure des xyloglucanes	16
Figure 1.4 : Modèle de structure des pectines	17
Figure 1.5 : Représentation schématique de structures putatives des différentes classes d'AGP chez <i>A. thaliana</i>	25
Figure 1.6 : Modélisation de la partie glycanique des AGP	27
Figure 1.7 : Quelques exemples de structure de Hyp-AG de type II	28
Figure 1.8 : La structure chimique du réactif de Yariv	30
Figure 1.9 : Modèles des différents modes d'action des AGP	31
Figure 1.10 : Structure du corps protéique des protéines partageant des domaines avec AGP31	34

Chapitre 2

Figure 2.1 : A- Carte du vecteur de destination pAGRIKOLA,1 utilisé pour la sous-expression d'At1g28290. B- Localisation du GST (<i>gene sequence tag</i>) sur la région transcrite d'At1g28290	37
Figure 2.2 : Carte du vecteur de destination pK7WG2D.1 utilisé pour la surexpression d'AGP31	39
Figure 2.3 : Amorces nucléotidiques utilisées pour l'analyse de l'expression du gène <i>At1g28290</i> , de sa séquence C-terminale et du gène contrôle <i>At1g49240</i> .	44
Figure 2.4 : Composition des gels de polyacrylamide-SDS utilisés pour la séparation des protéines par SDS-PAGE	49
Figure 2.5 : Polysaccharides utilisés pour les tests d'interaction avec les protéines d'intérêt	55
Figure 2.6 : Linéarisation de l'ADN plasmidique par digestion enzymatique et transcription <i>in vitro</i> des sondes sens contrôles et des sondes anti-sens complémentaires de l'ARNm)	58

Chapitre 3 (Article)

Tableau 1 : MALDI-TOF MS analysis of the tryptic peptides of the AGP31 Pro-rich domain (P0-P4)	73
Tableau 2 : Amino acid sequences of the tryptic peptides of the AGP31 Pro-rich domain (P1-P4)	74
Figure 1 : Identification of AGP31 by PMF	75
Figure 2 : Detection of AGP31 in enriched fractions after separation by SDS-PAGE	76
Figure 3 : Comparison of MALDI-TOF MS spectra of glycosylated and HF-deglycosylated forms of AGP31	77
Figure 4 : Identification of O-glycopeptides of the AGP31 Pro-rich domain using MALDI-TOF MS	78
Figure 5 : Location of Hyp residues within the Pro-rich domain	79
Figure 6 : AGP31 structural models	80

Chapitre 4

Figure 4.1 : Estimation de la propension d'AGP31 à présenter une structure désordonnée par analyse <i>in silico</i>	84
Figure 4.2 : Stratégies de production et de purification de la protéine AGP31 et de son domaine PAC pour les test d'interaction <i>in vitro</i> .	85
Figure 4.3 : Analyse par SDS-PAGE et <i>Western blot</i> des fractions de protéines PAC/V5/6xHis solubles et insolubles exprimées dans <i>E. coli</i>	87
Figure 4.4 : Analyse par SDS-PAGE des fractions obtenues après différents temps d'induction à l'IPTG et à différentes étapes de la purification de la protéine MBP/PAC/V5/6xHis à partir de <i>E. coli</i>	88
Figure 4.5 : Analyse des fractions obtenues aux différentes étapes de la purification de PAC/V5/6xHis à partir de feuilles de <i>Nicotiana benthamiana</i>	90
Figure 4.6 : Validation de la puce à polysaccharides	92
Figure 4.7 : Interactions entre AGP31 et les polysaccharides pariétaux	94
Figure 4.8 : Interactions entre AGP31 et des fractions enrichies en hémicelluloses digérées par la protéinase K	95
Figure 4.9 : Interactions entre AGP31 et des polysaccharides commerciaux	96
Figure 4.10 : Interactions entre MBP/PAC/V5/6xHis et les polysaccharides pariétaux extraits d'hypocotyles étiolés d' <i>A. thaliana</i>	97
Figure 4.11 : Interactions entre PAC/V5/6xHis et des polysaccharides pariétaux	98
Figure 4.12 : Mise en compétition de l'interaction entre AGP31 et les polysaccharides pariétaux	99
Figure 4.13 : Mise en compétition de l'interaction entre PAC/V5/6xHis et les polysaccharides pariétaux	100
Figure 4.14 : Interactions entre la protéine AGP31 et le domaine PAC	101

Chapitre 5

Figure 5.1 : Développement d' <i>A. thaliana</i> à l'obscurité	105
Figure 5.2 : Profil d'accumulation de transcrits d' <i>At1g28290</i> établi par analyse transcriptomique au cours du développement d' <i>A. thaliana</i> à la lumière	108
Figure 5.3 : Mesure de l'accumulation des transcrits d' <i>At1g28290</i> par RT-PCR quantitative et de l'accumulation d'AGP31 par <i>lectin blot</i> au cours de la germination à l'obscurité de plantules d' <i>A. thaliana</i>	109
Figure 5.4 : Coupes longitudinale d'une plantule d' <i>A. thaliana</i> (A) et transversale d'un hypocotyle (B) au cours de la germination à l'obscurité	110
Figure 5.5 : Localisation des transcrits d' <i>At1g28290</i> et d'AGP31 au cours de la germination	112
Figure 5.6 : Quantification par RT-PCR des transcrits d' <i>At1g28290</i> dans les jeunes tiges de différentes lignées de plantes ARNi de la génération T3	114
Figure 5.7 : Mesure de l'accumulation des transcrits d' <i>At1g28290</i> par RT-PCR quantitative et de l'accumulation d'AGP31 par <i>lectin blot</i> au cours de la germination à l'obscurité de plantules d' <i>A. thaliana</i> ARNi sous-expresses d' <i>At1g28290</i>	115
Figure 5.8 : Localisation des transcrits d' <i>At1g28290</i> et d'AGP31 au cours de la germination à l'obscurité de graines de lignées ARNi sous-expresses	117
Figure 5.9 : Quantification par RT-PCR des transcrits d' <i>At1g28290</i> dans les rosettes de différentes lignées de plantes sur-expresses de la génération T3	118
Figure 5.10 : Mesure de l'accumulation des transcrits d' <i>At1g28290</i> par RT-PCR quantitative et de l'accumulation d'AGP31 par <i>lectin blot</i> au cours de la germination à l'obscurité de plantules sur-expresses d' <i>At1g28290</i>	119

Figure 5.11 : Localisation des transcrits d'*Atlg28290* et d'AGP31 au cours de la germination de graines de lignées sur-expressseurs au cours de la germination **120**

Figure 5.12 : Profil de transcription de gènes codant des protéines de structure proche d'AGP31 **123**

Chapitre

Figure 6.1 : Co-régulation au cours du développement des gènes *Atlg28290* codant AGP31 et *Atlg53070* codant une protéine de type lectine légumineuse **127**

Figure 6.1 : Modèle d'interactions non covalentes entre AGP31 et différents constituants pariétaux **129**

ABREVIATIONS

ADNc	: ADN complémentaire
ADN-T	: ADN de Transfert
AG	: ArabinoGalactane
AGP	: Protéine à arabinogalactanes
AGRIKOLA	: <i>Arabidopsis Genomic RnaI Knock-Out Line Analysis</i>
Ara	: Arabinose
ARNi	: ARN interférence
<i>A. thaliana</i>	: <i>Arabidopsis thaliana</i>
ATP	: Adénosine TriPhosphate
<i>A. tumefaciens</i>	: <i>Agrobacterium tumefaciens</i>
BCIP	: 5-bromo-4-chloro-3-indolyl phosphate
<i>B. distachyon</i>	: <i>Brachypodium distachyon</i>
BET	: Bromure d'ETHidium
BSA	: Sérum Albumine Bovine
BY-2	: <i>Bright Yellow-2</i>
CDTA	: Cyclohexane Diamine TétraAcétate
CEC	: Chromatographie d'échange de cations
Col 0	: Ecotype Columbia 0 d' <i>A. thaliana</i>
CRP	: <i>Cysteine-Rich Peptide</i>
Ct	: Cycle seuil
CWP	: <i>Cell Wall Protein</i>
DEPC	: DiEthylPyroCarbonate
DMF	: DiMéthylFormamide
DMSO	: DiMéthylSulfOxyde
dNTP	: DésoxyriboNucléotide Triphosphate (dATP, dCTP, dGTP, dTTP)
DO	: Densité Optique
DTT	: DiThioThréitol
EDTA	: Ethylène Diamine TétraAcétate
<i>E. coli</i>	: <i>Escherichia coli</i>
ETD	: Dissociation par Transfert des Electrons
EXT	: Extensine
FLA	: <i>Fasciclin-like AGP</i>
FPLC	: <i>Fast Protein Liquid Chromatography</i>
Fuc	: Fucose
GA	: Gomme Arabique
GAG	: Glycosyl-Amino Glycane
Gal	: Galactose
GalA	: Acide Galacturonique
GalT	: Galactosyle Transférase
GAX	: GlucuronoArabinoXylane
GC	: Chromatographie en phase Gazeuse
GFP	: <i>Green Fluorescent Protein</i>
GH	: Glycoside hydrolase
Glc	: Glucose
GPI	: Glycosyl-Phosphatidyl-Inositol

GRP	: Protéine Riche en Glycine
GUS	: Glucuronidase
HF	: Fluorure d'Hydrogène
HG	: HomoGalacturonane
HRGP	: GlycoProtéine Riche en Hydroxyproline
Hyp	: Hydroxyproline
IPTG	: IsoPropyl β -D-ThioGalactopyranoside
kpb	: kilo paires de bases
kDa	: kilo Dalton
LC	: Chromatographie en phase Liquide
LC-MS/MS	: LC- <i>tandem Mass Spectrometry</i>
LRR	: <i>Leucine-Rich Repeat</i>
LTP	: Protéines de Transfert des Lipides
MALDI-TOF	: <i>Matrix-Assisted Laser Desorption Ionization-Time Of Flight</i>
Man	: Mannose
MBP	: <i>Maltose Binding Protein</i>
MES	: 4-Morpholino-Ethane-Sulfonate
Milieu LB	: Milieu de Luria Bertani
Milieu MS	: Milieu Murashige et Skoog
MS	: Spectrométrie de Masse
NAC	: Chromatographie d'Affinité aux ions Nickel
<i>N. benthamiana</i>	: <i>Nicotiana benthamiana</i>
NBT	: Nitro Bleu de Tétrazolium
PAC	: <i>PRP and AGP containing Cysteine</i>
PAGE	: PolyAcrylamide Gel Electrophoresis
pb	: paire de bases
PCR	: Réaction de Polymérisation en Chaîne
PG	: PolyGalacturonase
PGA	: Acide PolyGalacturonique
PGIP	: Inhibiteur de PolyGalacturonase
PME	: Pectine MéthylEstérase
PMEI	: Inhibiteur de Pectines MéthylEstérases
PMF	: <i>Peptide Mass Fingerprinting</i>
PNA	: lectine de <i>Peanut Agglutinin</i>
PRP	: Protéine Riche en Proline
PTM	: Modification Post-Traductionnelle
PVDF	: PolyFluorure De Vinylidène
RG I	: RhamnoGalacturonane I
RG II	: RhamnoGalacturonane II
Rha	: Rhamnose
RMN	: Résonance Magnétique Nucléaire
RNAse	: RiboNucléAse
RT-PCR	: <i>Reverse Transcriptase PCR</i>
RT-Q-PCR	: RT-PCR quantitative
SDS	: Dodécyl Sulfate de Sodium
TEMED	: N,N,N',N' - Tétraméthyl Ethylène Diamine
Tris	: Trishydroxyméthylaminométhane, or 2-amino-2-hydroxyméthyl-1,3 propanediol
UHQ water	: Eau <i>Ultra High Quality</i>
vol	: volume

wt	: <i>wild-type</i> (sauvage)
XET	: Xyloglucan EndoTransglycosylase
XG	: XyloGlucane
XTH	: Xyloglucane endoTransglucosylase/Hydrolase
Xyl	: Xylose

Chapitre 1

INTRODUCTION

1. INTRODUCTION

1.1 Organisation anatomique et fonctions de la paroi végétale

La paroi végétale est une structure extracellulaire organisée et complexe, qui enveloppe la membrane cytoplasmique des cellules végétales et qui constitue l'une des caractéristiques anatomiques qui différencie la cellule végétale de la cellule animale. Malgré sa localisation extracellulaire, la paroi cellulaire est aujourd'hui considérée comme un organe à part entière. Sa composition unique ainsi que l'ensemble de ses propriétés physiques et chimiques lui permettent d'assurer un véritable *continuum* entre les cellules. Elle a une double spécificité : elle doit être rigide pour jouer le rôle de squelette et de barrière, mais aussi assurer une certaine élasticité pour permettre la croissance du végétal. Ces exigences, à la fois antagonistes et complémentaires, sont à l'origine de la complexité de la paroi cellulaire végétale.

La paroi végétale est un composite pluristratifié, constituée de différents réseaux de polysaccharides (cellulose, hémicelluloses) enchassés dans une matrice gélatineuse, amorphe et résistante (pectines). Ces polymères sont organisés en couches hétérogènes et dynamiques évoluant selon la plante considérée et son développement. L'ensemble est stabilisé par des protéines et parfois des composés phénoliques dans les parois secondaires lignifiées (Wojtaszek, 2000). Chacun des composants de la paroi possède un rôle et une fonction particuliers.

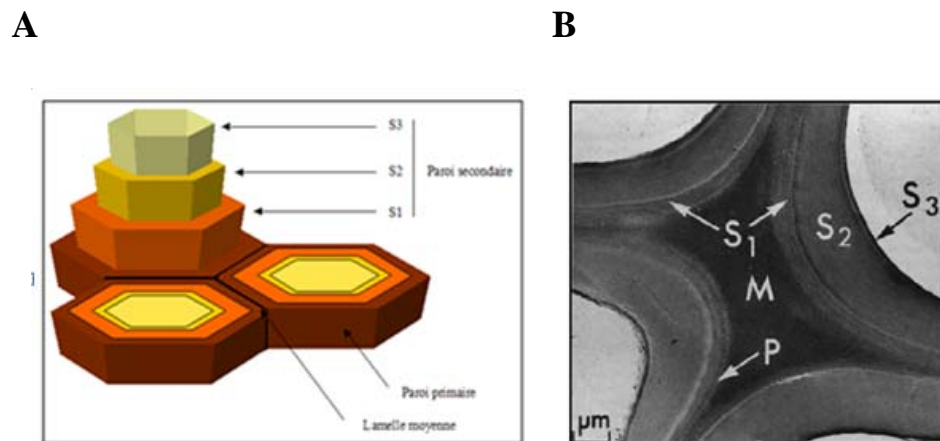


Figure 1.1 : Paroi des cellules végétales montrant ses différents niveaux d'organisation (Sjöström et Westermark, 1999)

A. Représentation schématique. **B.** Coupe vue en microscopie électronique. M : lamelle moyenne ; P : paroi primaire ; S1, S2, S3 : paroi secondaire.

On distingue trois couches : la lamelle moyenne, la paroi primaire et la paroi secondaire (Figure 1.1) :

- La lamelle moyenne : C'est une jonction intercellulaire qui apparaît lors de la mitose, à la suite de la fusion de vésicules golgiennes au niveau du phragmoplaste (Samuels et al., 1995). La lamelle moyenne est constituée principalement de substances pectiques et dépourvue de

cellulose. De la callose est présente transitoirement dans la lamelle moyenne et semble indispensable au bon déroulement de la mitose (Thiele et al., 2009). Le rôle de la lamelle moyenne est d'assurer la cohésion entre les cellules.

- La paroi primaire : C'est un réseau très hydraté qui est mis en place après la formation de la lamelle moyenne. Elle est composée de cellulose, d'hémicelluloses, de pectines et de protéines. Ces différents composants s'organisent et interagissent entre eux. La structure de la paroi primaire et sa composition chimique évoluent au gré de la croissance cellulaire. Elle est à la fois extensible et résistante. Son épaisseur est extrêmement faible relativement à l'épaisseur totale de la paroi.

- La paroi secondaire : Dans certains tissus, les cellules déposent progressivement, une fois leur croissance terminée, des couches pariétales supplémentaires sur la face interne de la paroi primaire. L'ensemble de ces couches constitue la paroi secondaire. Il s'agit d'une structure rigide, inextensible, faiblement hydratée (20% d'eau maximum) et beaucoup plus épaisse que la paroi primaire. Elle est constituée essentiellement de microfibrilles de cellulose cristalline organisées en couches concentriques parallèles entre elles et d'hémicelluloses (Roland, 1980). Cette structure secondaire peut être renforcée par des dépôts hydrophobes de polymères phénoliques, les lignines ; c'est par exemple le cas des tissus conducteurs (xylème) ou de soutien (sclérenchyme).

La paroi cellulaire végétale est donc formée majoritairement de polysaccharides dont les structures et les relations intermoléculaires peuvent expliquer les cinq rôles physiologiques suivants qui lui sont attribuée (Perez et al., 2003): (i) formation d'un réseau intercellulaire, (ii) adhésion des cellules grâce à la lamelle moyenne, (iii) contrôle de la porosité de la paroi cellulaire, (iv) maintien de l'équilibre physico-chimique du symplaste (eau et ions) et (v) transduction de signaux extracellulaires (oligosaccharides ou peptides) vers le milieu intracellulaire (Fry et al., 1993; Meng et al., 2010). Hormis les polysaccharides qui sont les constituants majeurs, la paroi renferme aussi des protéines qui contribuent à son architecture et ses fonctions. L'étude des protéines pariétales est donc primordiale pour comprendre la structure et la dynamique de ce compartiment.

Dans cette introduction, nous nous sommes plus particulièrement intéressés à la structure de la paroi primaire. Après en avoir présenté les constituants, nous verrons comment les différents réseaux s'organisent. Une partie sera réservée à la description de la famille des glycorotéines riches en hydroxyprolines (HRGP) à laquelle appartient AGP31 (*Arabinogalactan Protein*) qui fait l'objet de ce travail de doctorat. La dernière partie traitera des objectifs de la thèse.

1.2 Architecture moléculaire de la paroi primaire

1.2.1 Les polysaccharides

Les polysaccharides sont de longues chaînes de molécules de sucres représentant jusqu'à 90 % en masse des constituants pariétaux. La cellulose, les hémicelluloses et les pectines sont les trois composés polysaccharidiques majeurs présents dans la paroi des cellules végétales (Carpita et Gibaut, 1993). Ces polymères interagissent entre eux et forment de véritables réseaux.

1.2.1.1 La cellulose

La cellulose est une molécule stable, insoluble représentant 15 à 30% de la masse sèche de la paroi primaire. Elle est synthétisée par des organismes aussi différents que les bactéries, les champignons et les plantes. La cellulose est un homopolymère linéaire composé de 2 000 à 25 000 résidus de β -(1-4)-D-glucopyranose, dont le cellobiose (dimère de Glc) est le motif répétitif (Figure 1.2).

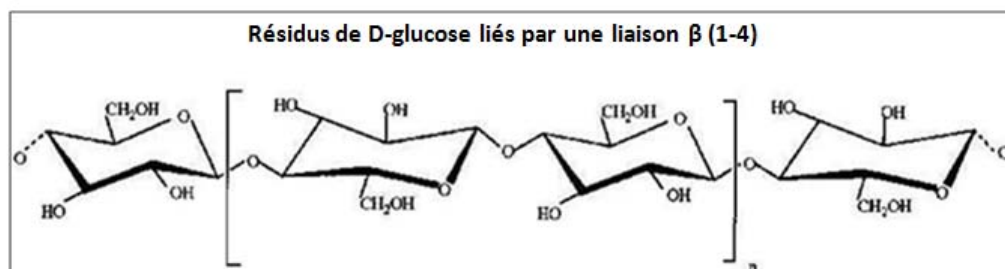


Figure 1.2 : Représentation de la molécule de cellulose

Les chaînes de cellulose ont une forte tendance à s'agréger entre elles en raison de leur constitution chimique et de leur conformation spatiale. La disposition des hydroxyles libres des résidus Glc permet l'établissement de liaisons hydrogènes intramoléculaires, stabilisant la cellulose dans son orientation linéaire, ce qui lui confère une certaine rigidité. L'existence de liaisons hydrogène inter-chaînes conduit à la formation de fibres, ce qui favorise l'établissement d'un état solide ordonné, pseudocristallin et permet la formation de microfibrilles. La structure fibrillaire très condensée de la cellulose explique sa résistance mécanique à la traction, ainsi que son caractère non soluble dans l'eau.

La cellulose est synthétisée et arrangée en microfibrilles par des complexes protéiques appelés rosettes, composés de six sous-unités, contenant chacune six cellulose synthases (CesA) (Doblin et al., 2002). Les rosettes effectuent la synthèse de la cellulose au niveau de la membrane plasmique à partir des résidus UDP-D-Glc. Le dépôt de ces microfibrilles est fait de manière orientée et peut changer au cours de la formation de la paroi secondaire (Brett, 2000; Emons et Mulder, 2000). Plusieurs gènes CesA responsables de la synthèse de la cellulose ont été identifiés par l'étude de mutants déficients en cellulose (Somerville, 2006). Certains sont impliqués dans la synthèse de cellulose dans la paroi primaire, d'autres dans la paroi secondaire.

1.2.1.2 Les hémicelluloses

Les hémicelluloses regroupent tous les polysaccharides présents dans les parois des cellules végétales qui ne sont ni cellulosiques, ni pectiques. Elles possèdent une structure commune constituée par un polymère de résidus de β -(1-4)-D-glycopyranose (ex. Glc, Man, et/ou Xyl) sur laquelle viennent se greffer de courtes chaînes de polysaccharides de natures diverses, composées des unités pyranosyle, furanosyle ou d'acides uroniques (ex. acide glucuronique, acide 4-O-méthylglucuronique). Contrairement à la cellulose, les unités glucidiques peuvent s'agencer de multiples façons dans les hémicelluloses, ce qui aboutit à une grande diversité structurale. Cette

diversité dépend de l'origine botanique, de l'âge des cellules et de leur localisation dans la paroi végétale.

Les polymères hémicellulosiques s'associent à la cellulose par des liaisons non covalentes et permettent ainsi aux microfibrilles cellulosiques d'être indirectement, mais solidement reliées les unes aux autres (McCann et al., 1990). Ces polymères peuvent également établir des liaisons covalentes avec les pectines acides (Thompson et Fry, 2000). Leur extraction de la paroi se fait par des solutions alcalines.

Dans la paroi primaire des dicotylédones, les plus abondants et les plus étudiés des polymères hémicellulosiques sont les xyloglucanes (XG). Selon le stade de développement et l'espèce végétale considérée, d'autres polymères hémicellulosiques peuvent être rencontrés tels que : les xylanes, les β -(1-4)(1-3) glucanes et les mannanes.

1.2.1.2.1 Les xyloglucanes

Les XG sont les principales hémicelluloses de la paroi primaire des dicotylédones où ils représentent 10 à 30% de la masse sèche des parois cellulaires (Fry, 1998). Ils sont présents en plus faible proportion dans les parois des monocotylédones et se distinguent par leur degré de ramification plus important. Les XG sont constitués d'un squelette de β -(1-4)-D-Glc, et jusqu'à 75% de ces Glc peuvent être substitués en α -(1-6) par des résidus D-Xyl.

Les XG présentent une structure conservée, basée sur la répétition de motifs constitués de 7 à 10 monomères (Figure 1.3). Le motif répétitif des XG se compose plus précisément de 4 résidus Glc, dont 3 sont substitués en α -(1-6) par des résidus Xyl, eux-mêmes substitués par des résidus variés (Vincken et al., 1997). Afin de simplifier la dénomination des chaînes latérales de XG, Fry et al. (1993) ont proposé un système simple basé sur un code de lettres permettant de préciser la nature des substituants. Ainsi, la lettre « X » représente une unité Glc substituée en position 6 par une unité Xyl, la lettre « G » représente une unité Glc non ramifiée, la lettre « L » représente une unité Glc substituée en position 6 par une unité Xyl, elle-même substituée en position 2 par une unité Gal et la lettre « F » représente la lettre « X » substituée en position 2 du Gal par une unité Fuc. Il a également été montré que la distribution des chaînes latérales entre deux espèces de plantes n'est pas identique, de même qu'elle diffère légèrement entre deux individus d'une même espèce ayant évolué dans deux milieux distincts. Les XG lient les microfibrilles de cellulose entre elles par des liaisons hydrogène et contribuent de ce fait à la rigidité de la paroi. L'hydrolyse des XG par des enzymes entraîne un relâchement pariétal nécessaire à la croissance des cellules. Par ailleurs, des fragments de XG peuvent avoir un effet sur la croissance. Le nanosaccharide XXFG dérivé du XG inhibe la croissance des segments de la tige de pois stimulé par un analogue d'auxine, le 2,4-D (acide 2,4-D dichlorophénoxyacétique) (York et al., 1984; McDougall et Fry, 1988).

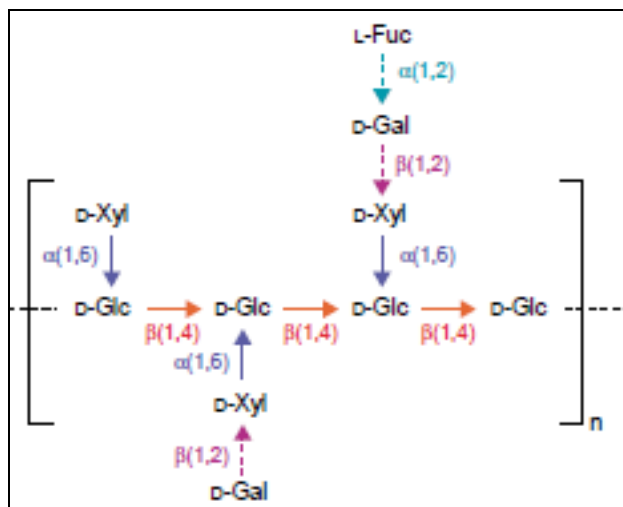


Figure 1.3 : Structure des xyloglucanes (Reiter, 2002)

Le XG de type XLFG chez *A. thaliana* est représenté dans cette figure. Les flèches pleines indiquent des liaisons toujours présentes et les flèches en pointillés indiquent des patrons de substitution.

1.2.1.2.2 Les xylanes

Les xylanes sont formés d'une chaîne principale dont les résidus D-xylopyranosyles (Xyl) sont liés par des liaisons β -(1-4). Ils peuvent être substitués (i) par des résidus de D-glucuronate en α -(1-2) en position C2 du Xyl, ces derniers pouvant être O-méthylés en position 4 (glucuronoxylanes), (ii) par des résidus de L-arabinose en α -(1-2) ou α -(1-3) en position C3 du Xyl (arabinoxylanes), ou (iii) par les deux (glucuronoarabinoxylanes ou GAX). Leur variabilité structurale dépend du type cellulaire et du stade de développement (Buchala et Wilkie, 1973). Le rôle des xylanes dans l'assemblage pariétal est mal compris. Cependant ces polymères de structure complexe modulent la construction de l'édifice pariétal et les propriétés de plasticité, de résistance mécanique et d'hydratation des parois.

1.2.1.2.3 Les β -(1-4)(1-3) glucanes

Les β -(1-4)(1-3) glucanes sont des polymères non ramifiés de Glc qui contiennent en moyenne 30% de liaisons β -(1-3) et 70% de liaisons β -(1-4). Les chaînes sont constituées d'une répétition d'un motif de base de 3 à 4 glucanes liés en β -(1-4) séparés par des liaisons β -(1-3) isolées altérant la rigidité de la chaîne. Ces glucanes se trouvent principalement dans les parois primaires des graminées. Leur existence a été décelée chez le lin par Goubet et al. (1995).

1.2.1.3 Les pectines

Les pectines sont les constituants majeurs de la paroi primaire des dicotylédones (50% de la masse de la matière sèche (O'Neill et al., 1990). Elles sont également abondantes dans la lamelle moyenne. Au niveau de la paroi secondaire, elles sont quasiment absentes (Morvan et al., 1989)

Les rôles des pectines sont très nombreux au cours du développement et en réponse aux contraintes de l'environnement. En effet, la matrice pectique constitue un réseau adapté à

l'expansion cellulaire du fait de son caractère hydrophile qui rend la paroi plastique. Elle régule également la porosité, le pH et la charge globale de la paroi et favorise l'adhésion entre les cellules. Les pectines sont également impliquées dans les mécanismes de défense et de signalisation (Darvill et al., 1992; Aldington et Fry, 1993; Carpita et Gibeaut, 1993)

Quatre polysaccharides pectiques de structures différentes ont été isolés des parois primaires : l'homogalacturonane (HG), les galacturonanes substitués, le rhamnogalacturonane I (RG I) et le rhamnogalacturonane II (RG II) (O'Neill et al., 1990). Les HG, RG I et RG II sont les plus abondants (Willats et al., 2006). Ils peuvent être liés entre eux de façon covalente (Willats et al., 2001; Vincken et al., 2003). Leur biosynthèse dans l'appareil de Golgi nécessite plusieurs enzymes de type glycosyltransférases (Ridley et al., 2001).

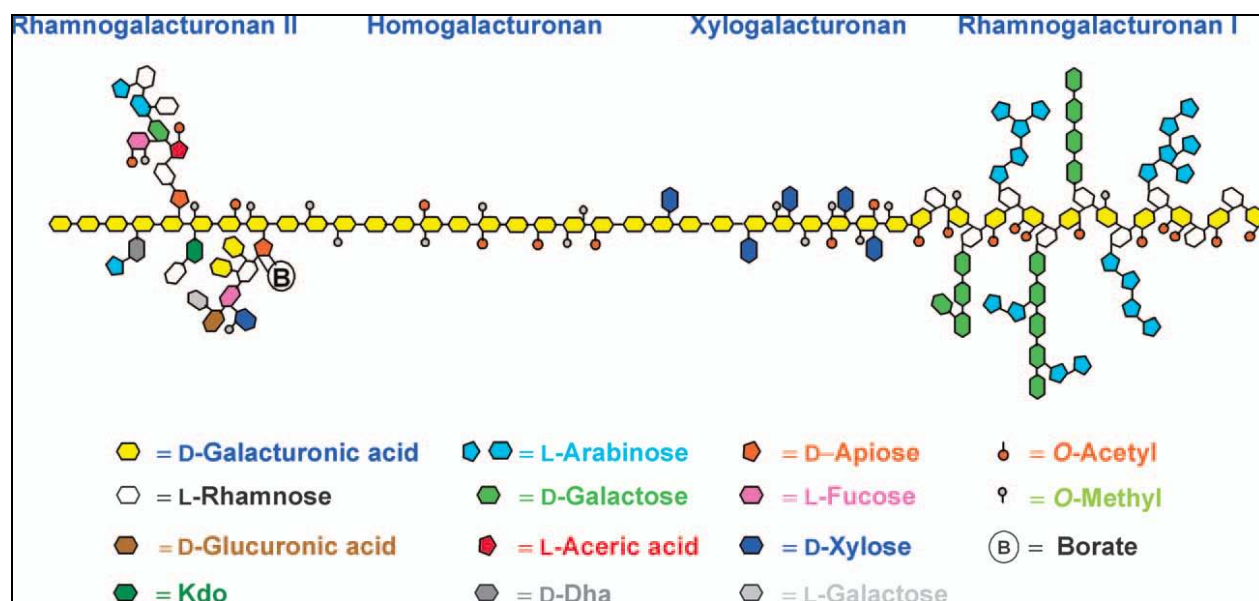


Figure 1.4 : Modèle de structure des pectines (Scheller et al., 2007)

Schéma simplifié montrant les différents domaines pectiques (HG, RG I, RG II et xylogalacturonane) et leurs agencements possibles au sein de la paroi.

1.2.1.3.1 L'homogalacturonane (HG) ou acide homogalacturonique

Le HG peut représenter jusqu'à 60% des pectines de la paroi primaire. C'est une chaîne linéaire d' α -D-GalA liés en 1-4 qui présente le plus souvent plusieurs modifications telles que la méthylestérification des fonctions carboxyliques (Fry et al., 1993) en position C6 ou une acétylation (substitution des GalA par l'acide acétique) sur les fonctions alcools secondaires en position C2 ou C3. En fonction de son taux de méthylestérification, le HG pourra être désigné par pectine (taux de méthylestérification élevé) ou acide pectique (taux de méthylestérification faible). Lorsque deux portions de chaînes sont majoritairement constituées de GalA non méthylestérifiées (désestérification réalisée par les pectines méthylestérases pariétales), elles peuvent interagir en présence de Ca^{2+} selon une structure ordonnée appelée « boîte à œufs » formant ainsi un bloc de jonction, lequel participe à la prise en gel de l'ensemble du réseau (Ridley et al., 2001). L'anticorps monoclonal 2F4 qui reconnaît spécifiquement ces structures montre la

présence de telles structures dans les espaces intercellulaires des cellules du parenchyme sur des coupes de hampes florales d'*A. thaliana*. Ces structures ne sont pas détectées chez le mutant *quasimodo1* qui présente une adhésion cellulaire réduite (Bouton et al., 2002; Orfila et al., 2005). Il est donc possible qu'elles jouent un rôle dans le renforcement de la paroi à des endroits clés pour éviter la séparation des cellules. Ainsi, les propriétés des HG et leurs fonctions biologiques sont souvent corrélées à leur capacité à établir des interactions électrostatiques.

1.2.1.3.2 Les galacturonanes substitués

Des substitutions de l'homogalacturonane existent, notamment par le Xyl en position C3 des GalA, formant ainsi les xylogalacturonanes (Schols et al., 1995; Renard et al., 1997; Vincken et al., 2003), ou par l'apiose (Api) en position C2 et C3, formant alors un domaine apiogalacturonane (Ridley et al., 2001).

Les xylogalacturonanes ont été principalement obtenus à partir de tissus reproducteurs (graines de soja, pépins de pomme, pollen de pin). Des motifs xylooligogalacturonides ont été trouvés dans les tissus végétatifs d'*A. thaliana* (Zandleven et al., 2007). Leur fonction biologique n'est pas connue.

Les apiogalacturonanes ont été décrits uniquement dans les parois de monocotylédones aquatiques des genres *Lemna* et *Zostera* (Golovchenko et al., 2002). Le rôle structural/physiologique des apiogalacturonanes, ainsi que leurs interactions avec les autres polysaccharides de la paroi ne sont pas bien connus. Leur abondance suggère, pour ces plantes aquatiques, un rôle structural important.

1.2.1.3.3 Le rhamnogalacturonane I (RG I)

Le RG I est un polysaccharide dont la chaîne principale est composée de la répétition d'une centaine de motifs disaccharidiques α -(1-4)-GalA- α -(1-2)-Rha (Lerouge et al., 1993; Mohnen, 1999). Entre 20 et 80 % des résidus Rha sont substitués en position C4 par des oligosaccharides acides ou neutres. Ces substitutions peuvent être des résidus uniques de D-galactopyranose en β -(1-4) ou des polymères linéaires ou branchés de résidus α -L-Ara et β -D-Gal. Il s'agit notamment d'arabinanes et/ou de galactanes et d'arabinogalactanes de type I (arabino-4-galactane) ou de type II (arabino-3-6-galactane) (Ridley et al., 2001; Paulsen et Barsett, 2005). Les résidus de GalA ne sont généralement pas substitués. Cependant, une étude menée chez la betterave sucrière a montré la présence de motifs de GlcA liés en position C3 sur environ 2% des résidus GalA de la chaîne principale du RG I (Renard et al., 1999). Les résidus GalA du RG I peuvent être méthylestérifiés ou parfois O-acétylés en position C2 ou C3 (Komalavilas et Mort, 1989). Contrairement aux HG, abondants dans la lamelle moyenne, le RG I est principalement présent dans la paroi primaire (Ermel et al., 2000).

Des données convergentes semblent indiquer que le type et la taille des substitutions du RG I sont variables chez les espèces monocotylédones et dicotylédones (Darvill et al., 1985; York et al., 1986; Thomas et al., 1989). De plus, la structure des chaînes latérales évolue au cours du développement.

1.2.1.3.4 Le rhamnogalacturonane II (RG II)

Bien que le RG II soit quantitativement minoritaire dans la paroi primaire (3 à 4% chez les dicotylédones), les recherches sur ce composé pectique font l'objet d'une attention croissante tant du point de vue de l'élucidation de sa structure complexe que des propriétés qui en découlent. Le RG II comprend environ neuf résidus de GalA auxquels sont liées quatre chaînes latérales complexes (Figure 1.4) (O'Neill et al., 2004). Ces chaînes latérales contiennent des monosaccharides particuliers, tels que le D-apiose (Api), le L-acerate (AceA), le 2-O-méthyl L-Fuc, le 2-O-méthyl D-Xyl, le L-Gal, l'acide 2-keto-3-deoxy-D-lyxo-heptulosarique (Dha), et le 2-keto-3-deoxy-D-manno-octulosonate (Kdo).

Dans les parois, le RG II existe essentiellement sous forme de dimères : deux sous-unités de RG II sont reliées entre elles *via* un atome de bore entre les résidus Api de la chaîne latérale (O'Neill et al., 1996). Cette dimérisation du RG II semble avoir un rôle crucial pour la plante. Des analyses réalisées sur des plantes carencées en bore ont permis d'établir que la dimérisation était non seulement impliquée dans le contrôle de la rigidification mais aussi dans la régulation de la porosité pariétale (Fleischer et al., 1999; Matsunaga et al., 2004), ce qui souligne l'importance de cette liaison pour la structuration correcte du RG II (O'Neill et al., 2001; Ahn et al., 2006).

En revanche, même si quelques différences ont été signalées (Rodriguez-Carvajal et al., 2003), la structure du RG II apparaît très conservée parmi les plantes à travers l'évolution, ce qui laisse supposer qu'ils pourraient avoir une fonction capitale (McNeil et al., 1984).

1.2.2 Les protéines

En plus des polysaccharides précédemment décrits, les parois contiennent des protéines dont la grande diversité reflète la diversité structurale des parois ainsi que la multitude des fonctions qu'elles exercent. En particulier, un très grand nombre d'enzymes interviennent probablement dans les processus de modification et/ou de dégradation des polymères pariétaux nécessaires à la différenciation cellulaire et à la croissance de la plante (Fry, 1995).

Un panorama étendu des protéines pariétales a été révélé par de nombreuses approches protéomiques essentiellement réalisées sur la plante modèle *A. thaliana* (Jamet et al., 2008). Le nombre de gènes codant des protéines pariétales de cette plante est estimé entre 1500 et 2000 (Jamet et al., 2006). A l'heure actuelle, environ 500 d'entre elles ont été identifiées par différentes analyses protéomiques. Une classification en neuf familles fonctionnelles a été proposée (Jamet et al., 2008) à partir de l'analyse bioinformatique des séquences protéiques permettant de prédire des domaines fonctionnels de type PROSITE ou PFAM (Finn et al., 2008; Sigrist et al., 2010).

Ces classes fonctionnelles comportent les protéines suivantes :

- Des protéines agissant sur les polysaccharides (26%) renfermant notamment les glycoside hydrolases (GH) qui sont les plus représentées, les carbohydrates estérases/lyases et les expansines. Les xyloglucanes endotransglycosylases/hydrolases (XTH) permettent la rupture, puis la reformation des liaisons glycosidiques des XG nécessaire à l'expansion cellulaire. Les α -galactosidases, β -galactosidases, β -xylosidases et β -1,4-glucanases, pectinases et α -arabinofuranosidases interviennent dans les modifications des polysaccharides pariétaux. Les β -

1,3-endo-glucanases et les β -1,4-endo-chitinases font partie des protéines PR (*Pathogenesis-Related*) qui jouent des rôles au cours des réactions de défense des plantes (Stintzi et al., 1993).

- Des oxydo-réductases (12,4%) comportant les peroxydases, les *multicopper oxidases*, les *berberine-bridge enzymes* et les *blue copper binding proteins*. Les peroxydases sont impliquées dans l'élongation cellulaire et son arrêt (Passardi et al., 2004). En effet, le cycle hydroxylique conduit à la production de radicaux OH[•] et HOO[•] qui peuvent couper des polysaccharides de manière non enzymatique et contribuer au relâchement de la paroi. La production de H₂O₂ par le cycle catalytique permet la formation de polymères pariétaux comme les lignines et les réseaux de protéines structurales de type extensine. Ceci a pour conséquence de bloquer l'expansion cellulaire.
- Des protéases (11,9%) telles que les subtilisines, les Asp-protéases, les Cys-protéases et les Ser-carboxypeptidases. Ces protéines pourraient intervenir dans la maturation des protéines pariétales, leur dégradation ou la libération de signaux peptidiques (van der Hoorn, 2008; Albenne et al., 2009).
- Des protéines contenant des domaines d'interaction (10,6%) comportant des protéines interagissant avec les sucres (lectines), des protéines interagissant avec d'autres protéines grâce à des domaines LRR (*Leucine-Rich Repeat*), des inhibiteurs d'enzymes comme les inhibiteurs de polygalacturonases (PGIP), les inhibiteurs des pectines méthylestérases (PMEI) et les inhibiteurs de protéases.
- Des protéines de signalisation (7,4%) principalement les protéines à arabinogalactanes (AGP) et les récepteurs de type lectine récepteur kinases ou LRR récepteur kinases identifiés grâce à leur domaine extracellulaire (Shiu et Bleecker, 2001; Seifert et Roberts, 2007).
- Des protéines liées au métabolisme des lipides (5,4%) comportant des protéines homologues aux lipase/acylhydrolases de type GDSL, des glycérophosphodiesterases et des protéines de transfert de lipides (LTP). Les LTP semblent impliquées dans le transport des lipides et la formation des cires de surface. Elles sont étroitement associées à de nombreux processus de développement et de réponse aux stress biotiques et abiotiques (Yeats et Rose, 2008).
- Des protéines structurales qui sont peu représentées dans les protéomes pariétaux (1,8%). Il s'agit de protéines dont la composition en acides aminés est particulière. Elles sont riches en Hydroxyprolines (Hyp) (glycoprotéines riches en Hyp ou HRGP), Pro (protéines riches en Pro ou PRP) ou Gly (protéines riches en Gly ou GRP). Ces trois acides aminés sont également des composants majeurs des protéines structurales des animaux (kératines et collagènes), des araignées et des insectes (présentes dans les toiles d'araignée et les fibres de cocon de soie). Les protéines HRGP seront décrites en détail dans les paragraphes suivants (*cf.* § 1.2.4).
- Des protéines de diverses familles trop peu nombreuses pour former une classe fonctionnelle (11,9%).
- Des protéines de fonction inconnue (12,6%).

Il convient de remarquer que seul un petit nombre de ces protéines a été étudié expérimentalement. Par conséquent, cette classification pourra être révisée en fonction des rôles de ces protéines au cours du développement ou des interactions avec l'environnement et/ou des activités biologiques qui seront démontrés.

1.2.3 Organisation de la paroi primaire

Après avoir présenté les différents composants pariétaux, quelques questions qui concernent l'organisation macromoléculaire des polymères au sein des réseaux pariétaux restent à ce jour sans réponse. Plusieurs modèles ont été établis pour décrire ces assemblages. En 1990, McCann et al. (1990) ont proposé, grâce à des études par microscopie électronique haute résolution et à des techniques d'extraction séquentielle de certains polysaccharides pariétaux, un modèle dans lequel le réseau XG-cellulose est inclus dans une matrice pectique. Carpita et Gibeaut (1993) ont complété ce modèle en considérant un troisième réseau constitué de protéines structurales. Cependant, d'autres interactions doivent être prises en compte. Des études par résonance magnétique nucléaire (RMN) montrent des associations fortes entre des galactanes caractéristiques des RG I et une fraction résiduelle contenant de la cellulose obtenue par solubilisation poussée des hémicelluloses par les alcalis (Willats et al., 2001). Thompson et Fry (2000) ont démontré l'existence de liaisons covalentes entre des XG et les chaînes latérales d'arabinanes et des galactanes des RG I. Concernant le réseau protéique, les interactions qui pourraient s'établir entre les différentes protéines pariétales ou entre les protéines et les polysaccharides seront détaillées dans les paragraphes suivants avec une attention particulière à la famille des HRGP (*cf.* § 1.2.4).

1.2.4 Les glycoprotéines riches en hydroxyproline (HRGP)

Cette superfamille de protéines est largement répandue dans le règne végétal depuis les algues jusqu'aux plantes supérieures. Elle représente entre 5 à 10% de la masse sèche de la paroi des suspensions cellulaires (Wilson et Fry, 1986). Les HRGP sont répertoriées en différentes classes qui vont de protéines peu glycosylées, les PRP, aux modérément glycosylées, les extensines et fortement glycosylées, les AGP (Showalter et al., 2010). Il existe des protéines chimères contenant des domaines HRGP associés à d'autres domaines. Par exemple, la protéine PTL (*Potato Tuber Lectin*) de la famille des lectines de solanacées possède un domaine de type extensine et un domaine riche en Gly et Cys (Showalter, 1993). Une HRGP de 120 kDa spécifique du style de *Nicotiana glauca* possède un domaine extensine en N-terminal et un domaine de type AGP en C-terminal (Lind et al., 1994; Schultz et al., 1997). La protéine AGP31 qui fait l'objet de ce travail, comporte un court domaine AGP en N-terminal et un domaine central de type PRP. Dans les paragraphes suivants, les PRP et les AGP seront décrites en détail.

1.2.4.1 Les protéines riches en proline (PRP)

Les PRP représentent une classe de HRGP faiblement glycosylées caractérisées par la présence du motif pentamérique répétitif (Pro-Hyp-Val-Tyr-Lys)_n ou ses variants (Cassab, 1998) et contenant des quantités équimolaires de Pro et d'Hyp (Datta et al., 1989). Peu de données sur leurs glycosylations sont disponibles dans la littérature. Leur teneur élevée en Lys leur confère un caractère basique. Du fait de leur ressemblance avec les extensines, les PRP pourraient avoir la forme de bâtonnets (Buchanan et al., 2000).

Les PRP sont insolubilisées dans les parois (Datta et al., 1989; Brisson et al., 1994). Leur teneur relativement élevée en Tyr pourrait leur permettre de former des liaisons covalentes avec les autres protéines pariétales sous la forme de liaisons di-tyrosine (Brady et al., 1998).

Les gènes codant pour des PRP sont régulés au cours du développement. Chez le soja, ces gènes sont induits durant le développement de l'hypocotyle et de l'enveloppe de la graine (Wyatt et al., 1992) et chez *A. thaliana* lors du développement des poils absorbants (Bernhardt et Tierney, 2000). L'expression des PRP peut également être induite lors de stress d'origine biotique tels que ceux causés par des éliciteurs fongiques (Sheng et al., 1991; Bradley et al., 1992) ou lors de stress environnementaux tels que la blessure (Sheng et al., 1991; Vignols et al., 1999) et le stress hydrique (Battaglia et al., 2007). Ces protéines pourraient ainsi jouer des rôles dans le renforcement des parois.

1.2.4.2 Les extensines

Les extensines représentent une classe de HRGP modérément à fortement *O*-glycosylée et largement distribuée dans le règne végétal (Lamport, 1965; Lamport, 1967). Les extensines se caractérisent par une richesse en Hyp et en Ser, mais aussi en d'autres acides aminés comme Val, Tyr, Lys et His. Elles sont aussi caractérisées par la présence du motif Ser-(Hyp)_n qui peut représenter jusqu'à 60 % de la chaîne polypeptidique (Showalter, 1993). Ce motif répétitif assure l'agencement de la molécule en hélices alpha gauches de type polyproline II visualisée en microscopie électronique sous forme de bâtonnet (Van Holst et Varner, 1984).

Les *O*-glycanes peuvent représenter jusqu'à 60 % de la masse moléculaire de la protéine. Cette *O*-glycosylation est représentée en majorité (90-97 %) par des courtes chaînes latérales d'Ara liées aux résidus Hyp, allant de 1 à 4 arabinoses (Hyp-(Ara)₁₋₄), avec une forte proportion de tri et tétra-arabinoses (Akiyama et al., 1980), et un à deux Gal liés en alpha sur des résidus Ser (Lamport, 1973). Cette glycosylation par les Ara semble impliquée dans la stabilisation de l'hélice polyproline II grâce à la formation de liaisons hydrogène (Lamport, 1980). En effet, une déglycosylation de la molécule résulte en une distension et une réduction de sa stabilité conformationnelle (Van Holst et Varner, 1984). De plus la formation de liaisons intramoléculaires de type isodityrosine contribue probablement à la stabilité de la molécule (Showalter, 1993).

Les extensines sont rapidement insolubilisées après leur sécrétion dans les parois. Ceci est dû aux liaisons intermoléculaires qui s'établissent entre des Tyr pour former des ponts isodityrosine, ou entre trois Tyr pour former des trimères de pulcherosine, ou encore entre ponts isodityrosine intramoléculaires pour former des liaisons di-isodityrosine (Brady et al., 1998). Il existe également des interactions ioniques et des liaisons covalentes entre les extensines et les pectines pariétales (Qi et al., 1995). Enfin, des données récentes suggèrent que les extensines sont capables d'auto-assemblage pour former des complexes chargés positivement au pH extracellulaire (environ 4,5) du fait de l'abondance des résidus Lys et capables d'interagir avec les domaines pectiques chargés négativement (Cannon et al., 2008). L'ensemble de ces données suggère que ces protéines joueraient un rôle important dans la rigidité de la structure pariétale.

Les extensines sont impliquées dans de nombreux processus physiologiques et développementaux chez les plantes y compris la germination du pollen (Wu et al., 2001), l'abscission et la sénescence (Merkouropoulos et Shirsat, 2003). Plus précisément, il a été montré que le gène RSH (ROOT- SHOOT- HYPOCOTYL-DEFECTIVE) codant un gène d'extensine d'*A. thaliana* joue un rôle dans le positionnement de la plaque cellulaire au cours de la division cellulaire (Hall et Cannon, 2002). Des gènes codant des extensines sont également induits en réponse aux stress mécaniques (Hirsinger et al., 1999; Salva et Jamet, 2001) et à la blessure (Guzzardi et al., 2004).

1.2.4.3 Les AGP

Une classe importante des protéines de la superfamille des HRGP est celle des AGP. Leur étude a été entreprise au début du XIX^{ème} siècle. Elles ont été mises en évidence pour la première fois dans les sécrétions de gommes, surtout de gommes arabiques qui représentent l'une des plus anciennes sources d'arabinogalactanes (AG). Une source plus récente d'AG relativement purs est l'extrait du bois de *Larex occidentalis*. Ces AG sont très étudiés du fait de leur utilisation dans de nombreux domaines comme l'industrie agro-alimentaire et pharmaceutique (Phillips 2009). C'est notamment le cas pour la gomme arabique d'*Acacia senegal* en tant qu'agent de texture, de gélification et comme émulsifiant facilitant le mélange eau-huile.

Les AGP constituent l'une des familles de macromolécules les plus complexes présentes chez les plantes. Cette complexité provient de la diversité des squelettes protéiques, des *O*-glycanes décorant le squelette protéique, et des sucres périphériques décorant les chaînes d'AG. Possédant une partie glucidique pouvant représenter 90% de leur masse moléculaire, elles sont souvent appelées protéoglycanes. Les AGP sont présentes dans tout le règne végétal, aussi bien chez les angiospermes, les gymnospermes que chez les plantes inférieures (bryophytes et algues) (Fincher et al., 1983; Nothnagel, 1997). De par leur composition, elles sont comparées aux mucines et aux glycosyl-amino glycanes (GAG) observés dans le règne animal (Bacic et al., 2000).

Compte tenu de la diversité structurale et fonctionnelle des AGP, et du fait que la protéine qui fait l'objet de ce travail comporte un domaine AGP, nous avons choisi de leur consacrer un paragraphe, présenté ci-dessous.

1.3 Structure et fonctions des AGP

1.3.1 Structure des AGP

1.3.1.1 Le corps protéique

Les AGP sont constituées d'une faible partie protéique, qui ne représente que 1 à 10 % de la masse moléculaire de la glycoprotéine (Osman et al., 1995). Elles sont riches en Hyp, Pro, Ala, Ser et Thr avec des motifs dipeptides Ala-Hyp, Ser-Hyp, Thr-Hyp, Val-Pro, Gly-Pro et Thr-Pro. Les AGP sont très solubles, résistantes à la protéolyse dans leur état natif, probablement en raison de leur richesse en polysaccharides (Showalter, 1993). Elles possèdent une structure en hélice partielle de type polyproline α (Van Holst et Varner, 1984).

En se basant sur la composition du corps protéique, un premier classement a divisé les AGP en deux catégories, les AGP « classiques » et les AGP « non classiques » (Schultz et al. 2000). Le corps protéique des AGP classiques est constitué d'un peptide signal à l'extrémité N-terminale, un domaine AGP riche en résidus Hyp/Pro et d'une ancre Glycosyl-PhosphatidylInositol (GPI) permettant un ancrage des AGP au niveau membranaire par leur extrémité C-terminale hydrophobe. Les AGP non classiques diffèrent des AGP classiques par l'absence d'une ancre GPI et par la présence d'autres motifs riches en Cys, Asn ou Lys en plus du domaine riche en Hyp/Pro.

Une classification plus récente suggérée par Schultz et al. (2002) et complétée par Showalter et al. (2010) a permis de subdiviser les AGP d'*A. thaliana*, à l'aide d'analyses bioinformatiques, en plusieurs catégories sur la base des annotations du génome d'*A. thaliana* et de la composition en acides aminés. Il s'agit de 22 AGP classiques, 3 AGP contenant un court domaine riche en Lys, 16 arabinogalactane (AG) peptides, 21 AGP de type fascicline-like (FLA), 22 *early nodulin-like proteins* (ENODL), et d'AGP chimériques (Figure 1.5).

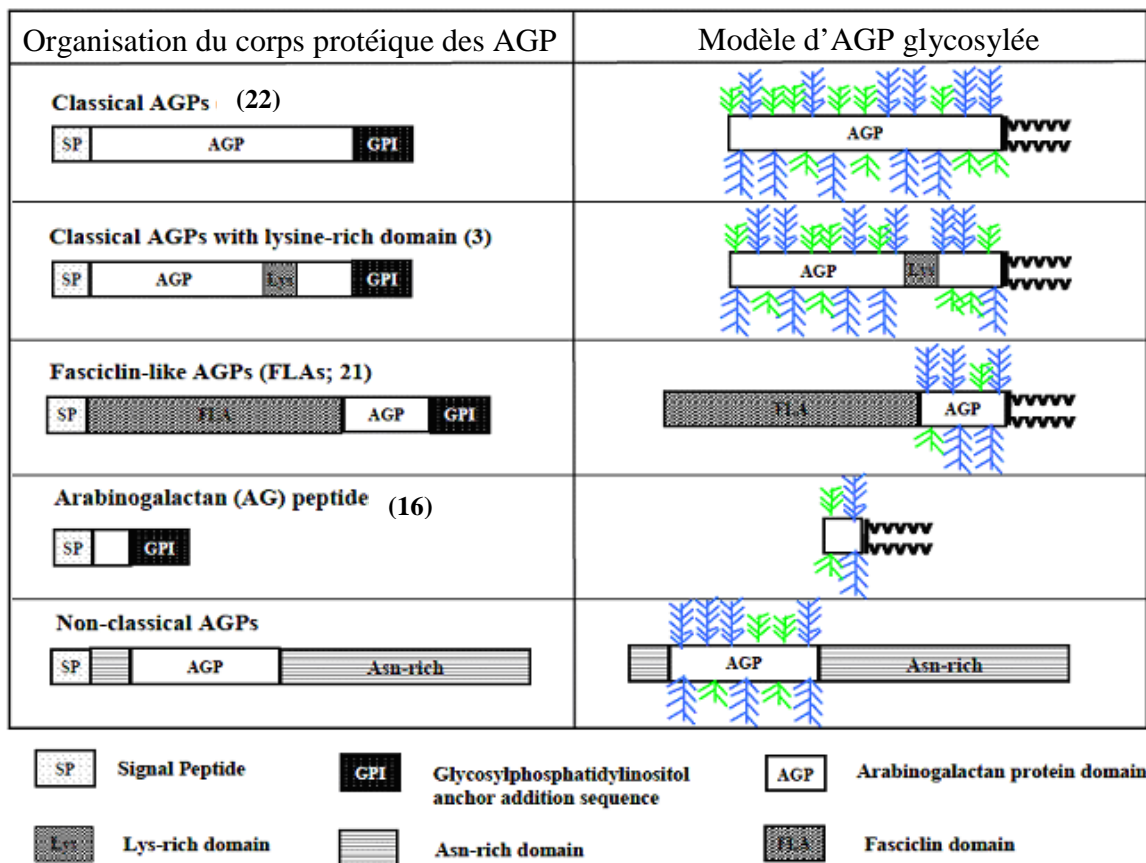


Figure 1.5 : Représentation schématique de structures putatives des différentes classes d'AGP chez *A. thaliana* (d'après Schultz et al., 2000 et Showalter et al. 2010).

Les ramifications bleues indiquent les AG et les vertes indiquent les arabino-oligosaccharides. La queue en zig-zag à l'extrémité des AGP indique l'ancre GPI. Le nombre entre parenthèses indique le nombre de protéines dans chaque classe d'AGP identifiée chez *A. thaliana*.

AGP classiques

Le corps protéique des AGP classiques, comme AGP-Pc1 du pin (Chen et al., 1994), AGP-Na1 du tabac (Du et al., 1994) et AGP1 de la tomate (Li et Showalter, 1996) est constitué de trois domaines, une extrémité N-terminale hydrophobe possédant une séquence signal de sécrétion absente chez la protéine mature, une partie centrale riche en Hyp/Pro fortement glycosylée, une extrémité C-terminale hydrophobe constituant une hélice transmembranaire à laquelle est liée une ancre GPI permettant un ancrage des AGP au niveau membranaire

Les AGP ayant un domaine riche en Lys

Il s'agit d'un petit groupe d'AGP qui se caractérisent par un court domaine de 12 acides aminés riche en Lys à l'intérieur du domaine AGP classique. Chez *A. thaliana*, les AGP riches en Lys représentent la classe la plus étudiée : AtAGP17 (Nam et al., 1999; Gaspar et al., 2004), AtAGP18 (Acosta-Garcia et Vielle-Calzada, 2004; Zhang et al., 2011) et AtAGP19 (Yang et al., 2011). Des protéines homologues ont été trouvées chez d'autres plantes comme LeAGP-1 chez la

tomate (Sun et al., 2004 a; Sun et al., 2004 b), NaAGP4 chez *Nicotiana glauca* (Gilson et al., 2001) et CsAGP-1 chez le concombre (Park et al., 2003).

Les AG peptides

Après leur première identification chez le blé (Fincher et al., 1974), les AG peptides ont ensuite été identifiés chez *A. thaliana*. Il s'agit de petits peptides de 10 à 13 acides aminés formant une classe unique de molécules. AtAGP24 contenant 17 acides aminés est l'un des AG peptides les mieux étudiés. A ce jour, il n'y a pas de données sur la fonction des AG peptides (Schultz et al., 2002).

Les AGP de type fasciclin-like

Les *fasciclin-like* AGP (FLA) présentent en plus de leur domaine AGP, un ou deux domaines fasciclins (Johnson et al., 2003). Ces domaines sont homologues aux fasciclins de la drosophile (Elkins et al., 1990) connus pour jouer un rôle majeur dans l'adhésion cellulaire. Les FLA présentent tous les autres domaines caractéristiques présents chez les AGP classiques.

Les AGP chimériques

Parmi les AGP chimériques, il y a les AGP de type plastocyanine (PAG) et d'autres protéines comme le xylogène, contenant un domaine de transfert de lipide non spécifique et un domaine AGP (Kobayashi et al., 2011).

1.3.1.2 La partie glycanique des AGP

La partie glycanique associée au cœur protéique des AGP constitue 90 à 98% de la masse moléculaire des protéines et est composée majoritairement de *O*-glycanes fixés aux Hyp. Le principal glycanes est un arabinogalactane de type II (AG II) (Bacic et al., 1987). De courtes chaînes d'arabino-oligosaccharides fixées sur des résidus Hyp sont aussi présentes (Figure 1.6). Chaque AGP est *O*-glycosylée par un nombre variable de chaînes d'AG II, conférant ainsi des masses moléculaires extrêmement hétérogènes.

La plupart des informations disponibles sur les conformations des AGP ont été obtenues à partir de la microscopie électronique à transmission qui a permis la visualisation *in vitro* de leurs formes et de leur agrégation. Pour expliquer l'organisation macromoléculaire des AGP, deux modèles probables de structure des chaînes d'AG sur la chaîne protéique ont été proposés (Figure 1.6).

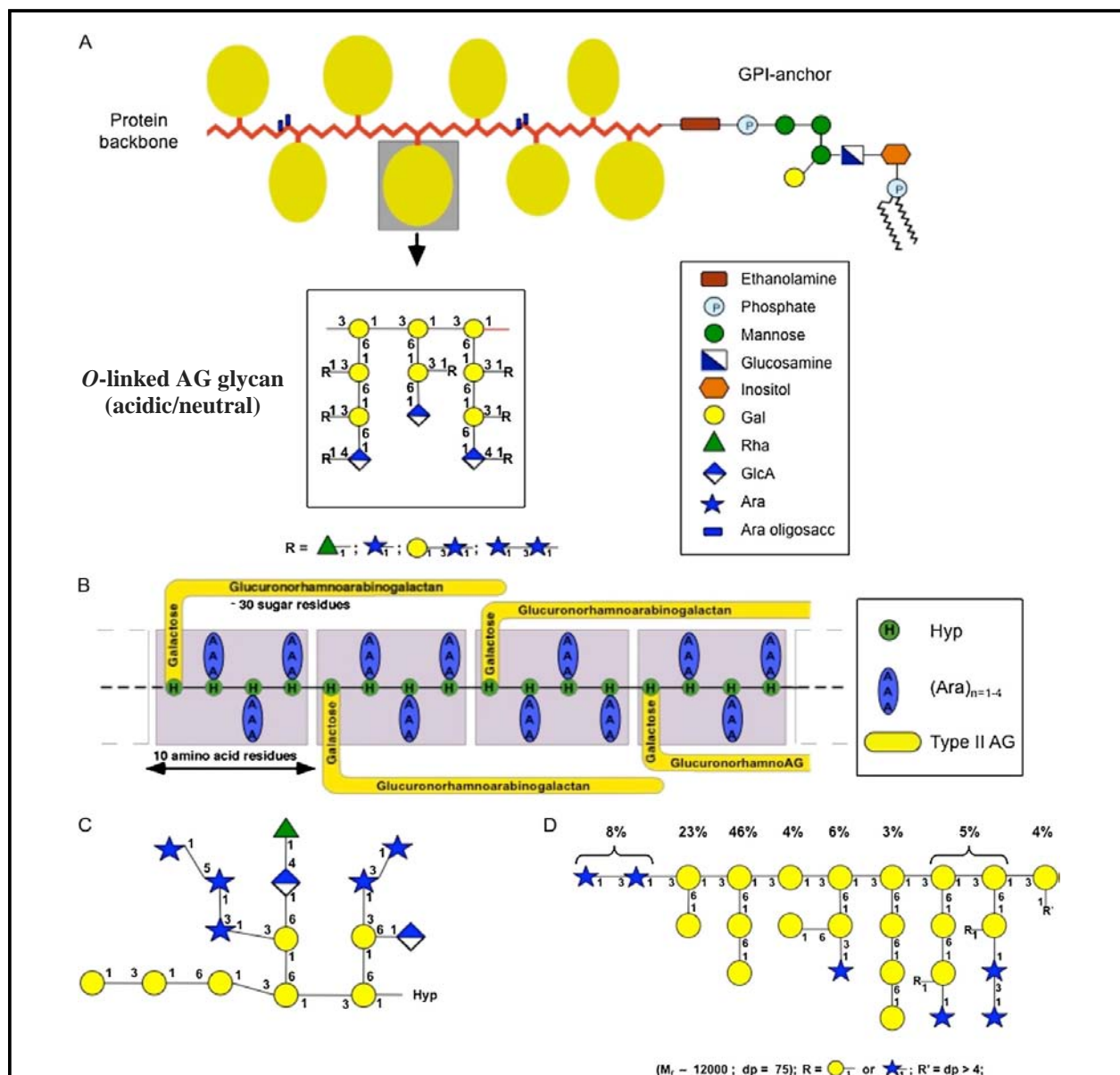


Figure 1.6 : Modélisation de la partie glycanique des AGP d'après Ellis et al. (2010)

A. Le modèle *Wattle blossom* de structure des AGP avec une ancre GPI. Les AG (ovales jaunes) sont liés au corps protéique *via* des résidus Hyp non contigus et constitués en majorité par des unités répétitives de β -(1-3)-Gal. Un petit nombre de résidus Hyp contigus portant de courtes chaînes d'arabino-oligosaccharides pourrait également être présent (bâtonnets bleus). **B.** Le modèle *Twisted hairy rope* de l'AGP de gomme arabique (GAGP) renfermant des motifs AG (jaune) et des courts arabinosides (bleu) attachés aux résidus Hyp. **C.** Structure d'un polysaccharide de type AG. Ce modèle est modifié d'après Tan et al. (2004). **D.** Structure de l'AG du mélèze. Ce modèle est modifié d'après Ponder et Richards (1997).

Le modèle de *Wattle blossom* est caractérisé par des chaînes polysaccharidiques formées d'unités globulaires portées par le corps protéique, ce qui produit une forme macromoléculaire globulaire. Ce modèle propose aussi la présence d'un petit nombre de courts arabino-

oligosaccharides. Dans le modèle *Twister hairy rope*, les AG, unités allongées sous forme de baguettes, et les arabino-oligosaccharides, sont distribués sur le corps protéique en blocs répétitifs ce qui génère une forme macromoléculaire allongée. Ces modèles semblent possibles en raison de la nature hétérogène des AGP. Le modèle *Wattle blossom* est trouvé chez certaines AGP de carotte et les protéines TTS du tabac (Cheung et al., 1995) et le modèle *Twisted hairy rope* est trouvé dans les AGP de la gomme arabique (Qi et al., 1991).

Des études structurales de quelques AGP ont été effectuées en utilisant les techniques de séparation par chromatographie en phase gazeuse (GC) de dérivés acétates d'alditol des sucres, d'analyse d'alditols acétate méthylés en GC suivie de spectrométrie de masse (MS), puis confirmées par RMN (Tan et al., 2004; Gollner et al., 2010; Tan et al., 2010). Plusieurs structures (Figure 1.7) ont été proposées à la suite de ces études, et notamment une structure de type II caractérisée par des liaisons β -1-3 entre les résidus D-Gal densément branchées en position C6 par des chaînes latérales β -e -(1-6)-D-Gal (Tan et al., 2010). Ces dernières chaînes sont substituées par des monosaccharides terminaux de type α -L-Ara ou par d'autres résidus comme le D-Glc, L-Fuc, D-Man, D-Xyl, L-Rha, D-GlcN, Glc A et Gal A (Nothnagel, 1997; Tan et al., 2010).

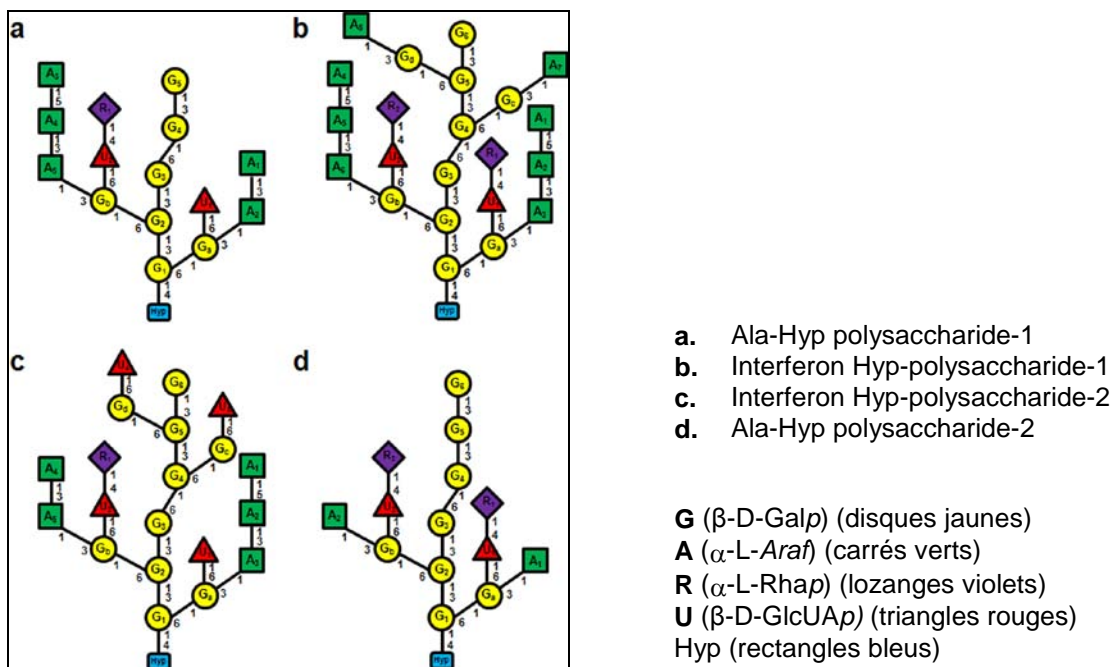


Figure 1.7 : Quelques exemples de structure de Hyp AG de type II (Tan et al., 2010)

Des protéines synthétiques ont été produites dans des suspensions cellulaires de tabac TB2. Ces protéines ont été purifiées et leurs parties glycaniques analysées pour en déterminer la structure. La structure des polysaccharides-1 et -2 est détaillée dans l'article de Tan et al. (2010)

A ce jour, le rôle précis de la partie glycanique des AGP n'est pas connu. Il a été montré que certaines combinaisons de mutants *uge4-rhd1* (UDP-D-Glc 4-épimérase, *root hair deficient*), affectés pour l'épimérisation d'UDP-D-Glc en UDP-D-Gal, avec des mutants de perception ou de synthèse d'hormones présentent des phénotypes sévères et des altérations des polysaccharides des AGP (Seifert et al., 2004). Toutefois il est impossible de déterminer la proportion des phénotypes qui est due à l'altération des AGP de celle qui est la conséquence d'altérations de la paroi.

1.3.2 Les outils et stratégies utilisés pour l'étude fonctionnelle des AGP

Pour étudier les AGP, plusieurs outils ou approches ont été développés. Parmi ceux-ci, l'utilisation de sondes telles que le réactif de Yariv, d'anticorps dirigés contre des épitopes glycaniques ou le squelette peptidique des AGP ou encore le recours à la génétique classique ou inverse.

Le réactif de Yariv

Les réactifs β -D-galactosyl et β -D-glucosyl de Yariv (Figure 1.8) sont utilisés pour leur capacité à se lier sélectivement et à précipiter les AGP (Yariv et al., 1962; Yariv et al., 1967) formant un complexe rouge-orange. Le β -D-mannosyl et le α -D-galactosyl ne peuvent pas lier les AGP et sont souvent utilisés comme contrôles (Nothnagel et Lyon, 1986).

Le réactif de Yariv peut être utilisé comme un réactif histochimique pour détecter les AGP sur des coupes tissulaires (Anderson et al., 1977; Gleeson et Clarke, 1979), détecter et quantifier les AGP dans un extrait tissulaire (Van Holst et Varner, 1984), perturber la distribution normale et la fonction des glycoprotéines *in vivo* dans les plantes et les cultures cellulaires (Serpe et Nothnagel, 1994; Willats et Knox, 1996; Guan et Nothnagel, 2004). Toutefois l'interaction entre AGP et réactif de Yariv reste mal connue (Yariv et al., 1967).

Les anticorps

Plusieurs anticorps polyclonaux et monoclonaux sont couramment utilisés pour étudier la distribution et la fonction des AGP. Certains d'entre eux interagissent avec les fragments de carbohydrate comme JIM13 et JIM14 qui reconnaissent le motif β -D-GlcA-(1,3)- α -D-GalA-(1,2)- α -L-Rha, MAC207 qui reconnaît le motif β -D-GlcA-(1,3)- α -GalA-(1,2)- α -Rha (Pennell et al., 1989; Yates et al., 1996), et LM2 qui reconnaît le β -D-glucuronopyranosyl (Smallwood et al., 1996). D'autres anticorps interagissent avec le corps protéique des AGP comme l'anticorps PAP dirigé spécifiquement contre le domaine riche en Lys de LeAGP1 PAPAPSKGKVKGGKKGKKHNA (Gao et al., 1999).

Le réactif de Yariv et les anticorps monoclonaux dirigés contre les carbohydrates semblent interagir avec 50-100 AGP différentes. De fait, ils permettent d'avoir une vue globale des AGP. Pour étudier une AGP particulière, une approche génétique est nécessaire.

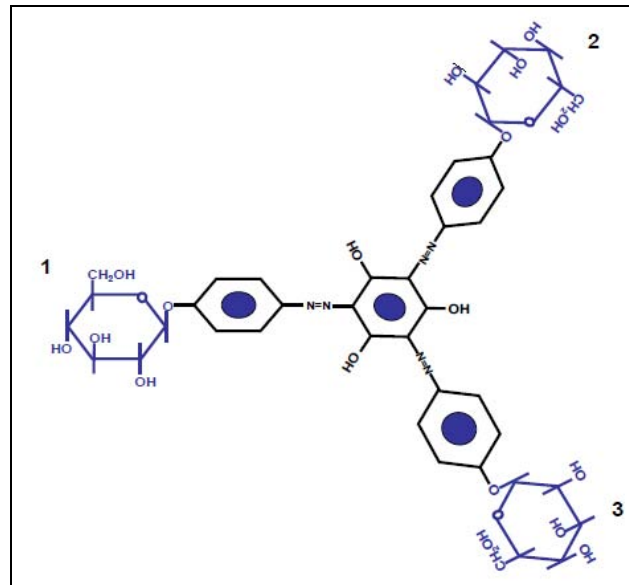


Figure 1.8 : La structure chimique du réactif de Yariv (Showalter, 2001)
Les trois sucres terminaux sont indiqués : 1,2 et 3.

Approches génétiques

Il s'agit de produire des lignées transgéniques comme les mutants d'insertion T-DNA, les plantes sous-expresseurs et surexpresseurs. Ces approches ne sont pas sans difficulté car la famille multigénique des AGP est grande et il existe probablement une redondance fonctionnelle. La production de double mutants ou des études en conditions de stress sont souvent nécessaires pour observer un phénotype.

1.3.3 Mécanismes d'action des AGP

Les AGP sont impliquées dans de nombreuses étapes du développement des végétaux. Cependant, à ce jour, les mécanismes d'action des AGP restent toujours obscurs. Cinq modèles ont été proposés par Showalter (2001) (Figure 1.9) :

- Les AGP libres ou membranaires agiraient par les structures osidiques libérées de leur partie glycanique après hydrolyse par des enzymes pariétales. Ces structures libérées sont considérées comme des activateurs de récepteurs membranaires impliqués dans la transduction de signaux. Ainsi, les motifs *N*-acétylglucosamine et glucosamine libérés par action d'une chitinase participent à l'initiation de l'embryogenèse somatique de la carotte (Hengel et al., 2001) (Figure 1.9A).
- Les AGP pourraient interagir en tant que glycoprotéines directement ou indirectement avec des récepteurs membranaires et participeraient à la transduction du signal (Figure 1.9B).
- Les troisième et quatrième modèles proposent un rôle dans la communication intercellulaire et l'adhésion cellulaire respectivement. Les AGP pourraient se lier soit à des récepteurs membranaires de cellules voisines (Figure 1.9C), soit à d'autres AGP membranaires (Figure 1.9D). Toutefois, la capacité de transduction du signal lors de l'interaction AGP-récepteur n'est pas établie, hormis peut-être par déstabilisation de la membrane et ouverture de canaux calciques.

- La cinquième hypothèse est spécifique aux AGP-GPI. Sous l'action des phospholipases C ou D, l'ancre GPI libérée peut générer du phosphatidyl-inositol mais aussi de l'inositol phosphoglycane et des céramides, molécules connues pour jouer le rôle de seconds messagers (Figure 1.9E).

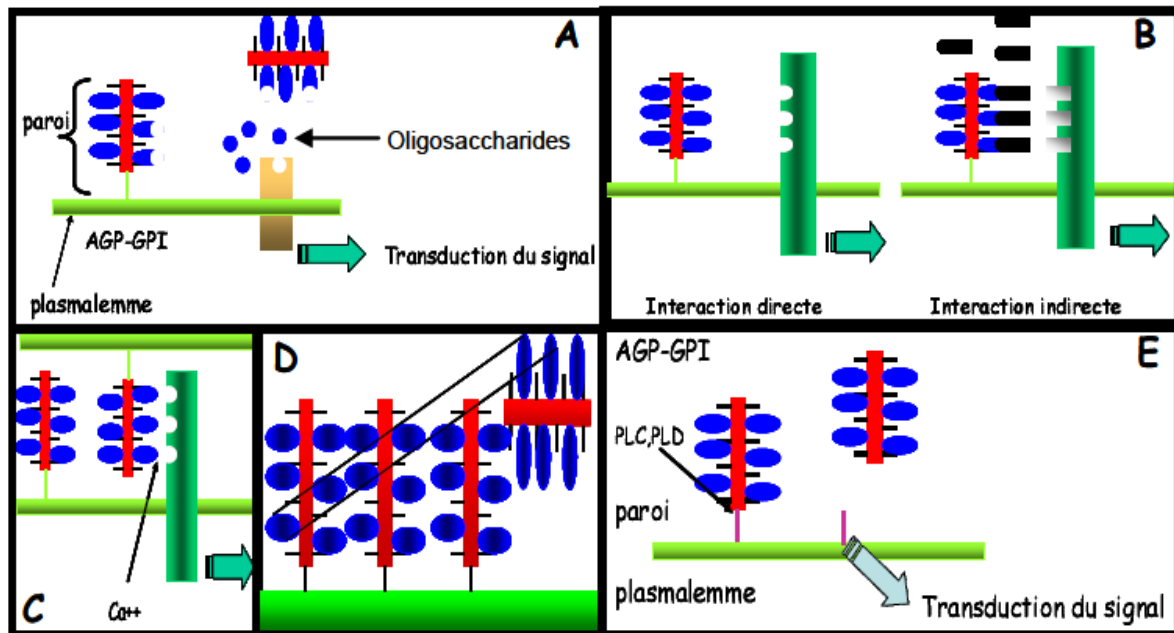


Figure 1.9 : Modèles des différents modes d'action des AGP (d'après Showalter, 2001)

- A. Signalisation par libération d'un oligosaccharide.
- B. Signalisation par une AGP directement ou indirectement après liaison avec un ligand.
- C. Signalisation intercellulaire par une AGP d'une cellule interagissant avec un récepteur d'une autre cellule.
- D. Les AGP peuvent être des molécules d'adhésion cellulaire capables de s'aggréger au niveau de la membrane plasmique et de jouer des rôles dans les interactions membrane/plaie.
- E. Signalisation par libération de l'ancre GPI après l'action d'une phospholipase C ou D.

1.3.4 Fonction biologique des AGP

Les AGP sont trouvées dans tous les types cellulaires (Showalter, 2001). Elles sont localisées au niveau des membranes plasmiques et des espaces intercellulaires (Anderson et al., 1977; Nothnagel, 1997; Börner et al., 2003). Elles sont trouvées dans les exsudats ainsi que dans les milieux de culture liquide (Schultz et al., 2004).

Les AGP semblent impliquées dans tous les processus physiologiques en relation avec la différenciation cellulaire et à la croissance. Ainsi, les AGP ont notamment été montrées comme impliquées dans la croissance et le développement des racines (Van Hengel et Roberts, 2003). L'analyse du mutant *agp30* montre que le gène est impliqué dans la régénération des racines *in vitro* et le déroulement de la germination des graines. Au cours de la différenciation du xylème (Motose et al., 2004), le xylogène, une protéine à arabinogalactane, est localisée dans les parois cellulaires des éléments trachéaires. Un double mutant des gènes codant pour le xylogène chez *A. thaliana* montre une altération dans le développement vasculaire. *AGP18* est essentiel pour la gamétogenèse femelle chez *A. thaliana* (Acosta-Garcia et Vielle-Calzada, 2004). *AGP6* et *AGP11* semblent être deux gènes

partiellement redondants. Leur mutation affecte le développement du pollen (Coimbra et al., 2009). La germination du pollen est réduite ainsi que la croissance du tube pollinique chez le double mutant *agp6-agp11* (Coimbra et al., 2010). Elle peut intervenir précocement dans l'anthère chez le double mutant *agp6-agp11*. Une hypothèse est que *AGP6* et *AGP11* détermineraient la germination du pollen par le biais du contrôle de l'absorption d'eau.

Outre leur rôle important dans le développement des plantes, les AGP ont également été impliquées dans la signalisation hormonale, la défense des plantes, la réaction au stress (Li et Showalter, 1996; Suzuki et al., 2002; Park et al., 2003; Lamport et al., 2006). Le mutant *agp30* codant pour une AGP non classique, est plus résistant à l'inhibition par l'acide abscissique (ABA) de la germination (Van Hengel et Roberts, 2003). Un mutant de *AtAGP17* (*rat1*, *resistant to Agrobacterium tumefaciens*), codant pour une AGP classique riche en Lys, est résistant à la transformation par *Agrobacterium tumefaciens* (Gaspar et al., 2004).

1.4 Contexte et objectifs du travail de thèse

1.4.1 Contexte scientifique

Une question qui demeure sans réponse est de savoir comment les différents composants protéiques et polysaccharidiques de la paroi sont organisés en une structure tri-dimensionnelle pour que la paroi puisse assurer ses fonctions. Une vue dynamique de la paroi végétale est désormais bien admise. De nombreuses observations ont montré que des changements structuraux, pouvant être limités à des microdomaines, se produisent au cours du développement et sont sans aucun doute régulés spécifiquement (Freshour et al., 1996). Le projet de notre équipe « Protéines pariétales et développement » au LRSV rentre dans cette optique et vise à contribuer à une meilleure compréhension de la dynamique de la paroi végétale à travers l'analyse du protéome des parois d'hypocotyles étiolés d'*A. thaliana* (Irshad et al., 2008). Ces travaux ont notamment permis de donner pour la première fois une image globale des protéines présentes à différents stades de leur croissance et d'identifier des protéines d'intérêt parmi lesquelles la protéine AGP31. Des études fonctionnelles fines sont cependant nécessaires pour préciser le rôle de ces protéines au cours du développement des hypocotyles afin de contribuer à une meilleure connaissance de la dynamique de la paroi végétale au cours de l'élongation cellulaire.

1.4.2 AGP31 et ses orthologues : état de l'art

AGP31 est codée par le gène *At1g28290*. Elle est très abondante dans la paroi des hypocotyles étiolés (Irshad et al., 2008) et présente une association de domaines qui la rend unique chez *A. thaliana*. Elle comporte un court domaine N-terminal de type AGP, un domaine riche en His, un domaine central très répétitif de type PRP et un domaine C-terminal appelé PAC (*PRP and AGP containing Cysteine*) qui comporte 6 Cys à des positions particulières (Baldwin et al., 2000). Du fait de la réactivité de la protéine avec le réactif de Yariv, le nom d'AGP a été attribué à cette protéine (Liu et Mehdy, 2007). La structure d'AGP31 a été partiellement caractérisée (Liu et Mehdy, 2007). Sa localisation pariétale a été vérifiée par protéomique (Irshad et al., 2008), et par des approches de biologie moléculaire et cellulaire en regardant la localis

ation d'une protéine de fusion AGP31-GFP (Liu et Mehdy, 2007). *AGP31* est exprimé dans tous les tissus vasculaires, mais à un niveau particulièrement élevé dans les racines et les fleurs (Liu et Mehdy, 2007). La fonction précise d'AGP31 dans la paroi demeure inconnue.

Chez *A. thaliana*, deux protéines partagent des domaines avec AGP31 (Figure 1.10). AGP30 présente une structure très proche de celle d'AGP31, mais n'a pas le domaine riche en His (Van Hengel et Roberts, 2003). Elle intervient dans la régénération des racines *in vitro* et détermine le moment de la germination des graines. Le mutant *agp30* n'est pas sensible à l'acide abscissique (ABA) et présente des modifications de l'expression de gènes régulés par l'ABA. Il a été suggéré qu'AGP30 pourrait médier la réponse à l'ABA. At2g34700 ne comporte que le domaine PAC. Des expériences d'hybridation *in situ* ont montré que At2g34700 est exprimé seulement dans le tégument extérieur de l'inflorescence et pourrait intervenir dans l'expansion cellulaire et la maturation de ces téguments (Skinner et Gasser, 2009).

Une recherche systématique de protéines homologues à AGP31 dans le génome d'*A. thaliana* a permis d'agrandir cette famille de protéines. D'une part les protéines PRP1 (At1g54970) et PRP3 (At3g62680) possèdent un domaine central riche en Pro et un domaine C-terminal comportant 4 Cys conservées (Fowler et al. 1999). D'autre part, huit protéines de petites tailles (159 à 183 acides aminés) présentent des homologies avec les domaines PAC d'AGP31, d'AGP30 et d'At2g34700. Ces protéines ont en commun la présence d'un peptide signal prédit et de 6 Cys à des positions conservées. Il s'agit d'At4g08685, At1g78040, At5g10130, At5g45880, At4g18596, At5g05500 et At3g09925.

La recherche de protéines ayant une structure similaire à AGP31 dans d'autres plantes a permis d'identifier plusieurs protéines (Wu et al., 2001; Liu et Mehdy, 2007). Ce sont notamment PvPRP1 (*French bean*), HyPRP1 (coton), DcAGP1 (carotte), GaRSGP (*Galactose-Rich Style GlycoProtein, Nicotiana alata*), NtTTS1 et NtTTS2 (*Transmitting Tissue-Specific, Nicotiana tabacum*) (Figure 1.10). Ces protéines constituent en fait une famille de protéines associées à la paroi cellulaire et caractérisées par la présence du domaine C-terminal PAC bien conservé et riche en Cys (Baldwin et al., 2001). Par ailleurs, elles peuvent partager d'autres domaines, qu'il s'agisse du domaine AGP, du domaine riche en His ou du domaine PRP. Il faut noter que DcAGP1 est la seule protéine qui partage l'ensemble de ses domaines avec AGP31.

Le premier orthologue d'AGP31 qui a été caractérisé du point de vue structural est la protéine GaRSGP codée par le gène *NaPRP4* et isolée de styles de *N. alata* (Sommer-Knudsen et al., 1996). Cette protéine présente autant de résidus Pro et Hyp, est constituée de 75-85% de carbohydrates, et est *O*-galactosylée par des courtes chaînes très ramifiées de glycanes se terminant principalement par des Gal. Des *N*-glycanes ont été aussi trouvés sur cette protéine. La protéine a pu être localisée dans les tissus de transmission du style. Aucune fonction directe n'a été attribuée à cette protéine. NtTTS1 et NtTTS2 sont glycosylées par 70% de Gal, interagissent avec le réactif de Yariv et l'anticorps monoclonal JIM13 spécifique des AGP et ont tendance à s'oligomériser (Wu et al., 2001). NtTTS1 et NtTTS2 sont présentes, quant à eux, seulement dans les fleurs et peuvent guider la croissance du tube pollinique *in vitro*, alors que leurs fonctions *in vivo* ne sont pas établies (Cheung et al., 1995). Enfin, les transcrits de DcAGP1 sont abondants dans les suspensions cellulaires de carotte et dans les racines des plantules (Baldwin et al., 2001). La fonction de DcAGP1 n'est pas connue.

Par ailleurs, le domaine PAC peut constituer une protéine monodomaine. Chez *Olea europaea* l'allergène du pollen Ole e 1 présente une forte homologie avec le domaine PAC et pourrait avoir des rôles dans le développement reproductif et végétatif (Jiang et al., 2005).

Chez la tomate, LAT52 peut interagir avec un récepteur kinase de type *Leu Repeat Receptor Kinase* et est essentielle pour la germination du pollen (Muschiatti et al., 1994).

Toutes ces protéines apparentées à AGP31 montrent des patrons d'expression différents, ce qui suggère qu'elles peuvent intervenir dans des processus physiologiques différents. Par contre, une activité biochimique similaire liée au PAC a été suggérée par beaucoup d'auteurs.

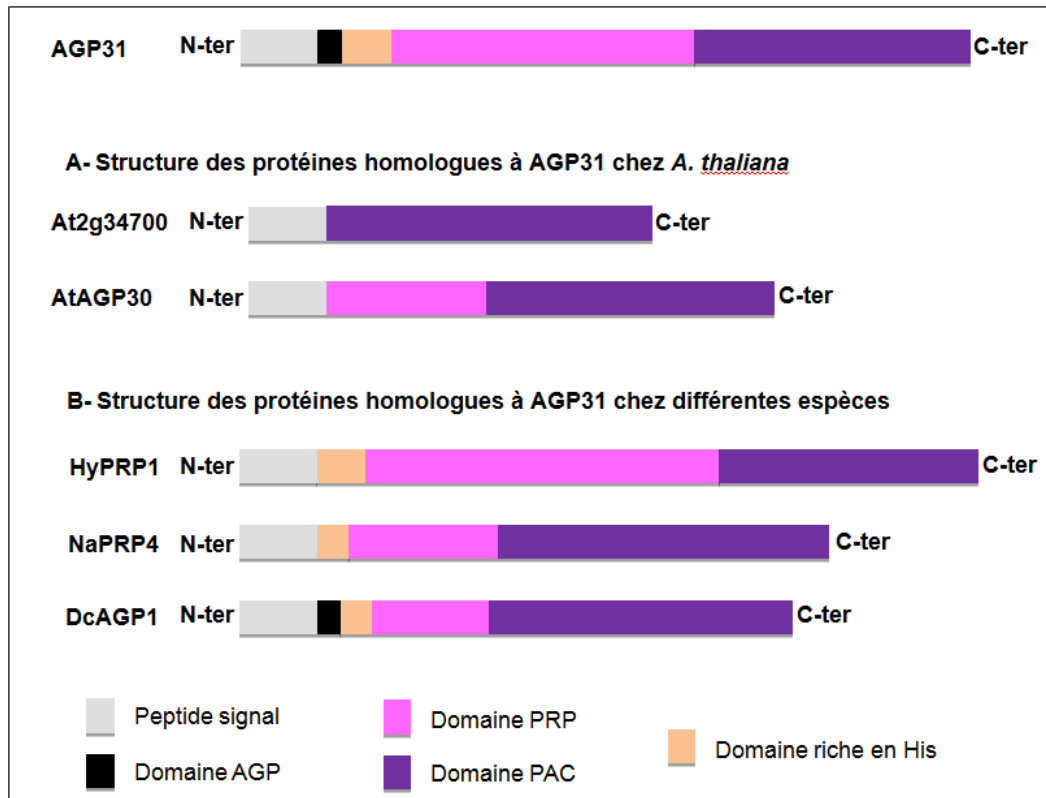


Figure 1.10 : Structure du corps protéique des protéines partageant des domaines avec AGP31

1.4.3 Objectifs de la thèse

Les objectifs poursuivis dans le cadre de ma thèse visent à approfondir la compréhension de la structure et de la fonction d'AGP31 dans son contexte pariétal.

Trois axes de recherche ont été développés :

- Dans un premier temps, nous avons fait une analyse structurale d'AGP31. Nous avons localisé les résidus Hyp du domaine PRP et apporté des informations sur son mode de *O*-glycosylation. Ceci a été réalisé grâce à l'utilisation de plusieurs approches complémentaires de spectrométrie de masse. Ces travaux sont décrits dans le chapitre 3 et font l'objet d'un article soumis pour publication "Characterization of the *O*-glycosylation heterogeneity of the ArabinoGalactan Protein 31 (AGP31) of *Arabidopsis thaliana*: evidence for a new Hyp-*O*-Gal-rich motif" (Hijazi et al., 2011).

- Nous nous sommes attachés par la suite à étudier la fonction d'AGP31 par l'identification de partenaires polysaccharidiques capables d'interagir avec AGP31 *in vitro*. Il s'agissait, d'une part, d'identifier des composants pariétaux interagissant avec AGP31 et d'autre part, de

rechercher des régions spécifiques d'interaction. Dans ce but, nous avons produit et purifié la protéine native ainsi que le domaine PAC sous la forme d'une protéine recombinante étiquetée. Nous avons ensuite mis en œuvre des tests sur membranes afin de mettre en évidence des interactions entre AGP31 entière ou le domaine PAC et des polysaccharides pariétaux de diverses origines. Ces résultats sont décrits dans le chapitre 4.

- Enfin, pour caractériser AGP31 *in planta* et déterminer son rôle physiologique, nous avons décrit l'expression d'AGP31 dans des plantes sauvages et transgéniques (RNAi et surexprimeurs) par des approches d'hybridation *in situ* et de cyto-histochimie. L'ensemble des résultats ainsi obtenus est présenté dans le chapitre 5.

Chapitre 2

MATERIELS ET METHODES

2. MATERIELS ET METHODES

Ne sont indiqués ici que les matériels et méthodes correspondant aux chapitres 4 et 5. Ceux relatifs au chapitre 3 figurent dans celui-ci.

2.1 Matériel

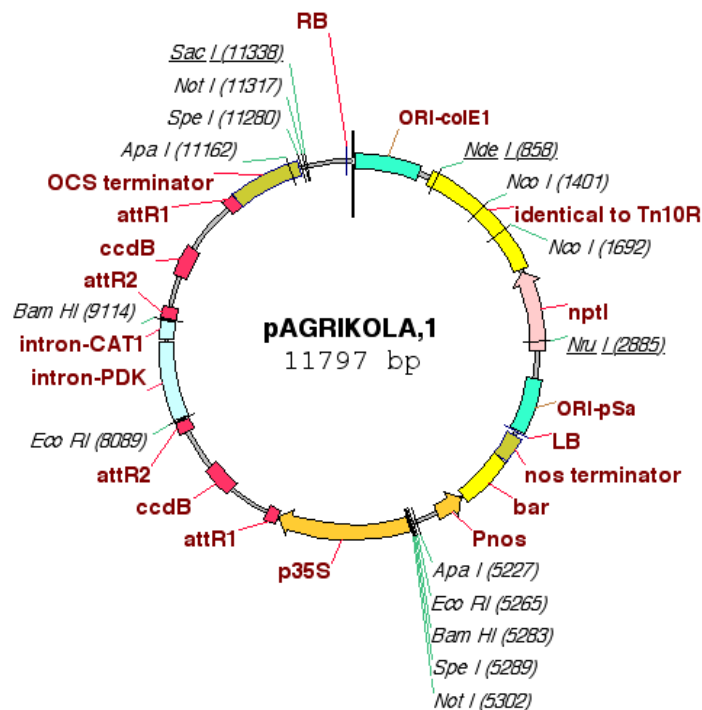
2.1.1 Matériel végétal

2.1.1.1 Plantes

Arabidopsis thaliana

Des plantes d'*A. thaliana* dans un fond génétique Col-0 (Columbia) ont été utilisées dans ces expériences. Des lignées transgéniques, générées au laboratoire ont été utilisées. Il s'agit de lignées ayant intégré dans leur génome des constructions ARNi (AGRIKOLA : *Systematic RNAi knockouts in Arabidopsis*, <http://www.agrikola.org>) qui avaient été obtenues au laboratoire par M. Irshad et D. Roujol et de lignées surexprimant la séquence complète d'AGP31 ou du domaine PAC seul que j'ai obtenues.

A



B

**GST :**

```
GATGATTTGATGCTCTGGTTACTATCTTGATTGTTTCATGTTTAAAAAGATTGAGAGAGAGTG
TTGTTATGCGTCTACTGTACGGTTGAATTTTAATCAATGTTAGCTAGTCTTTAAGTTAAATGC
TTGGGTATTATGTTGTATCTCTTTAAGTGACGTCGTTTGAGTTTCGTTTGA
```

Figure 2.1 : A- Carte du vecteur de destination pAGRIKOLA,1 utilisé pour la sous-expression d'At1g28290. B- Localisation du GST (*gene sequence tag*) sur la région transcrite d'At1g28290

Le GST, sonde gène spécifique d'une taille de 177 pb générée pour *At1g28290* est obtenue à partir de séquence génomique sur la région transcrite localisée à deux nucléotides en aval du codon de terminaison de traduction d'*At1g28290*. Ceci est destiné à l'échelle génomique dans le cadre du projet CATMA européenne («*Complete Arabidopsis Transcriptome biopuces*», <http://www.catma.org/>).

Nicotiana benthamiana

Les expériences d'expression transitoire de protéines dans les cellules végétales ont été réalisées sur *N. benthamiana*.

2.1.1.2 Conditions de culture**Stérilisation des graines**

Pour la culture des hypocotyles étiolés d'*A. thaliana*, les graines ont été vernalisées et stérilisées comme précédemment décrit (Feiz et al., 2006). Pour les autres cultures d'*A. thaliana*, un maximum de 100 mg de graines placées dans un tube Eppendorf sont stérilisées par 1 mL d'eau de javel à 0,4% de chlore actif dans l'éthanol et contenant du Triton X-100 à 0,05%. Les graines sont laissées 10 min sous agitation, puis sont rincées deux fois avec 1 mL d'éthanol à 95%. Après le dernier rinçage et toujours en conditions stériles sous hotte à flux laminaire, les graines sont séchées avant leur utilisation.

Conditions de culture *in vitro****Milieu de culture***

Les graines d'*A. thaliana* stériles sont semées sur milieu Murashige et Skoog (MS). Pour préparer 1 L de milieu, 4,32 g de poudre MS (Sigma M5519) additionnés de 2% de saccharose sont dissous dans de l'eau déminéralisée. Puis le pH est ajusté à 5,7 avec du KOH. Finalement, 1% d'agar (Sigma A1296) sont ajoutés. Pour la sélection de plantules transformées, les antibiotiques, stérilisés par filtration sur filtres Millipore 0,20 µm, sont ajoutés après autoclavage des milieux.

Conditions de culture in vitro

Pour la culture des plantules étiochées d'*A. thaliana*, les graines stérilisées sont semées dans des boîtes Magenta, et placées dans une logette de culture à 24°C, à une humidité de 40%, et à l'obscurité totale. Pour les autres cultures d'*A. thaliana*, les graines stérilisées sont semées en boîtes de Pétri. Ces boîtes sont placées dans une logette réglée dans les conditions

suivantes : cycle jours longs (16 h lumière/ 8 h obscurité), température (22°C lumière/20°C obscurité) et humidité 80%. Pour lever la dormance et obtenir une germination homogène et synchronisée, les graines sont placées à 4°C et à l'obscurité pendant 48 h avant d'être mises dans la logette appropriée.

Conditions de culture en terre

Plantation ou semis

Sur des pastilles de tourbe Jiffy® (Jiffy Products International AS) préalablement imbibées d'eau, des plantules d'*A. thaliana* peuvent être repiquées à partir de boîtes de culture *in vitro*, ou des graines peuvent être directement semées sans stérilisation préalable. Les pots sont ensuite recouverts d'un film en plastique afin de maintenir une atmosphère saturée en eau. Après 7 à 10 jours, les plantules sont transplantées et cultivées dans des pots sur du terreau Tref (Jiffy Products International AS). Les graines de *N. benthamiana* sont directement semées sur terreau.

Conditions en salle de culture

Les plantules et plantes d'*A. thaliana* sont cultivées à une humidité de 70% en conditions de jours longs (16 h lumière/ 8 h obscurité), de température contrôlée (22°C lumière/ 20°C obscurité) et à une humidité de 70% ou en conditions de jours courts (9 h lumière/ 15 h obscurité), de température contrôlée (22°C lumière/ 20°C obscurité). Les plantes de *N. benthamiana*, quant à elles, sont cultivées dans les conditions suivantes : jours longs (16 h lumière/8h obscurité), température contrôlée (25°C lumière/22°C obscurité) et à une humidité de 80%.

Conditions en serre

L'éclairage de la serre est 14 h lumière/10 h obscurité. La température est de 20° C.

Stockage des graines d'A. thaliana

Après leur récolte, les graines sont conservées à 4°C jusqu'à leur utilisation.

2.1.2 Matériel bactérien

2.1.2.1 Plasmides et vecteurs de clonage

Le vecteur pK7WG2D,1

Le vecteur pK7WG2D,1 (Université de Gent, www.psb.ugent.be/gateway) (Figure 2.1), portant la construction promoteur 35S::région codante AGP31::tag V5::tag 6xHis ou 35S::région codante PAC::tag V5::tag 6xHis est utilisé pour transformer les différents souches d'*Agrobacterium tumefaciens* C58/pMP90 et LBA4404. Il contient également la cassette de sélection des plantes et des bactéries transformées conférant la résistance à la kanamycine et à la spectinomycine respectivement.

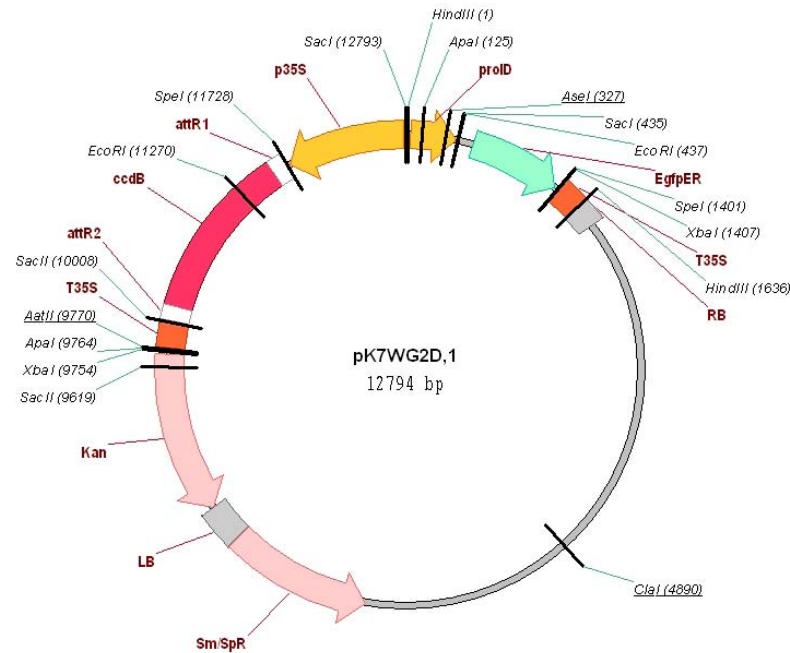


Figure 2.2 : Carte du vecteur de destination pK7WG2D.1 utilisé pour la surexpression d'AGP31
La séquence d'intérêt est sous le contrôle du promoteur 35S. Les gènes codant l'Egfp (gène rapporteur) et la résistance à la kanamycine sont présents entre les bornes RB et LB.

Le vecteur pBAD TOPO® (Invitrogen)

Ce vecteur permet d'exprimer des protéines fusionnées en C-terminal à deux *tag* situés en aval de la cassette de clonage : épitope V5 et 6xHis. Le gène hybride comportant la région codante PAC::*tag* V5::*tag* 6xHis est sous la dépendance du promoteur araBAD, assurant une expression très contrôlée et dose-dépendante en présence de L-arabinose. Ce vecteur est utilisé pour sur-exprimer des protéines, dans des souches *Escherichia coli* TOP10. Les transformants sont sélectionnés par leur résistance à l'ampicilline.

Le vecteur pMAL-c2 (New England Biolabs)

Ce vecteur permet d'exprimer des protéines en fusion avec le *tag* MBP (*Maltose Binding Protein*) située en amont de la cassette de clonage. Le gène hybride comportant la région codante MBP::région codante PAC::*tag* V5::*tag* 6xHis est sous le contrôle du promoteur de l'opéron lactose inductible par l'ajout d'isopropyl β -D-1-thiogalactopyranoside (IPTG). Ce vecteur est utilisé, pour sur-exprimer des protéines, dans des souches *E. coli* TOP10. Les transformants sont sélectionnés par leur résistance à l'ampicilline.

2.1.2.2 Souches bactériennes

La souche d'*Escherichia coli* TOP10 (Invitrogen)

Cette souche est utilisée pour cloner, amplifier un plasmide ou exprimer une protéine recombinante. Elle est déficiente dans la biosynthèse d'arabinose, et peut être utilisée en combinaison avec des vecteurs de type pBAD présentant un promoteur araBAD dans le cas d'une expression inductible au L-arabinose. Son génotype est le suivant : F- mcrA D (mrrhsdRMS-mcrBC) f80lacZDM15 DlacX74 deoR recA1 araD139 D(ara-leu)7697 galU galK rpsL (Str^r) endA1 nupG.

La souche LBA4404 (*A. tumefaciens*)

La souche LBA4404 contient un plasmide Ti (*Tumor Inducing*) pAL 4404 "désarmé", car délété de ses gènes oncogènes, et ne porte plus que les gènes de virulence (*vir*) nécessaires au transfert de l'ADN-T dans le génome de la cellule végétal (Hellens et al., 2000). Elle présente une résistance à la rifampicine et la streptomycine. Cette souche a été utilisée pour la transformation transitoire des cellules des feuilles de *N. benthamiana*.

La souche C58/pMP90 (*A. tumefaciens*)

La souche C58/pMP90 est une souche désarmée issue de la souche nopaline C58 et contient un plasmide Ti pMP90 (Hellens et al., 2000). Elle présente une résistance à deux antibiotiques, la rifampicine et la gentamicine. Cette souche a été utilisée pour la transformation stable des plantes d'*A. thaliana*.

La souche P19 (*A. tumefaciens*)

Cette souche est utilisée associée à la souche LBA4404 pour augmenter la production de protéines recombinantes en expression transitoire dans les feuilles de *N. benthamiana*. Le choix du suppresseur viral P19 vient du fait qu'il s'agit du suppresseur le mieux caractérisé actuellement (Voinnet et al., 1999; Baulcombe et Molnar, 2004), les suppresseurs viraux étant connus pour augmenter la productivité en protéines des feuilles agroinfiltrées (Voinnet et al., 2003).

2.1.2.3 Conditions de culture

Les bactéries sont cultivées sur boîtes de Pétri gélosées ou en milieu liquide sous agitation, à 37°C pour les souches d'*E. coli* et à 28°C pour les souches d'*A. tumefaciens*.

2.1.2.4 Milieux de culture

Les souches bactériennes d'*E. coli* et d'*A. tumefaciens* sont cultivées sur milieu Luria-Bertani (LB) (Sambrook et Russel 2001) (10 g/L Bacto-tryptone, 5 g/L Bacto-yeast extract, 10 g/L NaCl, pH ajusté à 7 avec du NaOH) contenant 15 g/L de Bacto-Agar (pour les cultures en boîte de Pétri) et les antibiotiques appropriés (*cf.* § 1.2.4)

2.1.3 Composés chimiques**Le β -D-glucosyl Yariv**

La synthèse du 1,3,5-tris (4- β -D-Glucopyranosyloxyphenylazo)-2,4,6-trihydroxy-benzène (β -D-Glc-y), un colorant connu sous le nom de β -D-glucosyl Yariv a été effectué dans le laboratoire par H. Canut selon la procédure décrite par Ganjian et Basile (1997). Cette molécule qualifiée d'antigène artificiel de synthèse de la classe des phénylglycosides permet de précipiter spécifiquement les AGP. Toutefois l'interaction entre ces deux molécules reste mal connue (Yariv et al., 1967).

Les Antibiotiques (KALYS)

Différents antibiotiques ont été utilisés pour la sélection des lignées d'*A. thaliana* et de bactéries transgéniques :

- Ampicilline 100 μ g/mL
- Streptomycine 100 μ g/mL
- Spectinomycine 100 μ g/mL
- Kanamycine 50 μ g/mL

- Rifampicine 50 µg/mL

Le Basta (VWR)

Cet herbicide, dont la molécule active est la phosphinothricine, inhibe l'activité de la glutamine synthétase. Pour vérifier que les plantes transgéniques sont résistantes au BASTA, on les asperge 2 à 3 fois par semaine avec cet herbicide à une concentration de 5 mg/mL et on observe les effets de l'herbicide au bout de 15 jours. Les plantes sensibles jaunissent tandis que les plantes résistantes restent vertes.

2.2 Méthodes

2.2.1 Expression des protéines recombinantes chez les plantes

2.2.1.1 Clonage dans le vecteur Gateway® et expression des protéines recombinantes AGP31 et PAC chez *A. thaliana* et *N. benthamiana*

A mon arrivée au laboratoire, les ADNc correspondant à la région codante complète d'*AGP31* et à son domaine PAC seul précédé d'un peptide signal avaient été clonés dans le vecteur Gateway® permettant de surexprimer ces deux constructions en fusion avec un *tag V5* suivie d'un *tag 6xHis* du côté C-terminal de la protéine (constructions réalisées par D. Roujol et L. Cisneros).

La technologie Gateway® est une méthode de clonage universelle basée sur les propriétés de recombinaison site spécifique du bactériophage lambda pour permettre un transfert rapide et efficace d'un gène d'une séquence d'ADN d'intérêt vers de multiples systèmes plasmidiques. Pour exprimer un gène d'intérêt dans les plantes en utilisant la technologie Gateway®, il suffit de cloner ce gène dans un vecteur d'entrée Gateway®, de transférer l'insert présent dans ce vecteur d'entrée vers un vecteur de destination Gateway® en réalisant une réaction de recombinaison, et d'introduire cette construction dans une cellule végétale grâce à *A. tumefaciens* afin d'exprimer la protéine recombinante.

A la fin du clonage, des bactéries *E. coli* TOP10 compétentes transformées par le vecteur pK7WG2D,1 portant la construction « région codante *AGP31::tag V5::tag 6xHis* » ou « région codante *PAC::tag V5::tag 6xHis* » sous la dépendance du promoteur 35S avaient été obtenues. Ces constructions sont ensuite utilisées pour transformer *A. thaliana* et *N. benthamiana*.

2.2.1.2 Transformations génétiques

Transformation d'*A. tumefaciens*

Préparation d'agrobactéries chimiocompétentes

Quelques colonies d'*A. tumefaciens* sont inoculées dans 5 mL de LB supplémenté d'antibiotiques appropriés, et incubées à 28°C sous agitation pendant 24 h. Cette culture est ensuite dosée au spectrophotomètre à 600 nm afin d'ajuster la DO finale entre 0,5 et 1. La culture est alors incubée 10 min dans la glace. Après une centrifugation de 6 min à 1800 g et à 4°C, ce culot est repris délicatement dans 300 µL de CaCl₂ 20 mM à 4°C. Cette suspension est finalement aliquotée par fractions de 100 µL, congelée rapidement dans l'azote liquide, et stockée à -80°C.

Transformation d'A. tumefaciens

De la souche *E. coli* TOP10, l'ADN du plasmide pK7WG2D,1 est extrait au moyen du kit d'extraction de plasmide *QIAprep Spin Miniprep* (QIAGEN) en utilisant le protocole indiqué par le fournisseur. Deux µg d'ADN plasmidique et 100 µL d'agrobactéries compétentes sont mis en présence pendant 5 min dans la glace. Un double choc thermique (1 min dans l'azote liquide puis 4 min à 37°C) achève la transformation. Un mL de LB est alors ajouté, et la suspension laissée à 28°C pendant 4 h sous agitation afin d'initier la réplication du plasmide. Les bactéries sont étalées sur milieu LB gélosé contenant l'antibiotique de sélection de la souche et l'antibiotique de sélection du plasmide recombinant exogène, puis incubées à 28°C pendant 48 h.

Techniques de transformation des plantes

Transformations transitoires des cellules des feuilles de N. benthamiana

Les transformations sont dites transitoires lorsqu'elles n'affectent pas les cellules reproductrices et le patrimoine héréditaire, l'ADN exogène introduit n'est donc pas transmis à la descendance. Cette technique de transformation transitoire par *A. tumefaciens* consiste à infiltrer des bactéries par les stomates de l'épiderme inférieur des feuilles de *N. benthamiana*. Une culture de 10 mL d'*A. tumefaciens* LBA 4404 portant la construction « 35S::région codante PAC::tag V5::tag 6xHis » dans le vecteur pK7WG2D,1 estensemencée à partir de quelques colonies bactériennes dans le milieu LB additionné d'antibiotiques appropriés, et incubée pendant une nuit, à 28°C sous agitation. Une souche d'agrobactérie P19 est encore mise en culture de la même manière. Après une sédimentation par une centrifugation lente de 5 min à 5800 g à température ambiante, les culots sont lavés avec 2 mL d'agromix (MgCl₂ 10 mM ; MES 10 mM, pH 5,6 ; acétosyringone 150 µM), puis centrifugés 5 min à 3000 g. Chaque culot bactérien est finalement repris dans 5 mL d'agromix. La suspension est dosée au spectrophotomètre à 600 nm afin d'ajuster la DO finale à 0,8. Les deux cultures sont mélangées (vol/vol) et infiltrées dans les feuilles à l'aide d'une seringue stérile de 1 mL appliquée perpendiculairement contre la face inférieure de la feuille. Le tampon d'infiltration utilisé place les bactéries dans des conditions iso-osmotiques et l'acétosyringone augmente la virulence de la souche, réunissant ainsi des conditions propices à la transformation de cellules végétales. Un échantillon de feuille infiltrée est analysé 2 à 4 jours après infiltration.

Transformation stable d'A. thaliana

La méthode utilisée est celle de Logemann et al., (2006). Elle est basée sur une transformation des inflorescences et nécessite des plantes en fleurs et en très bon état physiologique. Les plantes sont amenées au stade de floraison en conditions de jours courts (9 h de lumière) afin de permettre le développement de belles rosettes. On passe ensuite les plantes en conditions de jours longs (16 h de lumière). Le nombre de hampes florales est augmenté en coupant les premières hampes afin d'obtenir un maximum de bourgeons. Les *A. tumefaciens* C58/PMP90 portant la construction « région codante AGP31::tag V5::tag 6xHis » dans le vecteur pK7WG2D,1, sont préparées de la même manière que pour la transformation de *N. benthamiana*, mis à part que le culot bactérien est repris dans un petit volume d'une solution de saccharose 5% (p/v) (iso-osmotique). La DO_{600 nm} est alors mesurée contre un témoin saccharose 5%. La suspension est alors diluée dans une solution de saccharose 5% contenant : Silwett L-77 (Lehle Seeds) 0,05% (vol/vol) de façon à obtenir une DO_{600 nm} = 0,8 pour un volume de 100 à 200 mL final. Le surfactant (Silwet L-77) facilite la pénétration des bactéries dans la voie apoplastique des plantes. Les inflorescences sont immergées pendant 3 min dans la solution contenant les agrobactéries, avec une légère agitation pour une meilleure imprégnation. Un léger film bactérien se dépose alors sur les

plantes. Les plantes sont ensuite placées sous un film plastique afin de maintenir une humidité constante, elles doivent également être protégées de toute lumière excessive pendant 2 jours. Le film plastique est alors enlevé et les plantes remises en conditions de culture habituelles jusqu'à maturation des siliques. Les graines sont récoltées et les transformants sélectionnés.

2.2.1.3 Etude de la ségrégation du transgène et sélection des lignées homozygotes

Afin d'étudier la ségrégation du transgène *via* le marqueur de sélection, les graines des plantes transgéniques, de la génération T1, ayant intégré dans leur génome le gène de résistance à la kanamycine sont semées sur milieu MS additionné de kanamycine et celles supprimeuses, intégrant le gène de résistance au Basta® sont semées sur Gifi et aspergées par du Basta® à partir du stade 5 jours. Les plantes de la génération T1 sont des plantes hétérozygotes puisque la cible de transfert de l'ADN-T au cours du processus d'agroinfection au niveau du bourgeon floral est l'ovule. Pour chaque construction, une vingtaine de transformants résistants ont été transférés en serre et les graines ont été récoltées après autofécondation. Le rapport entre le nombre de plantes résistantes et le nombre de plantes sensibles, sur 100 plantules, permet de déterminer le caractère hétérozygote ou homozygote de l'insertion sur les plantules de la génération T2. Par exemple les plantes qui présentent un rapport 3:1 pour la ségrégation Résistant/Sensible sont donc celles ayant potentiellement intégré une seule copie d'ADN-T. Dans la génération T3, les plantes homozygotes, ayant 100% de plantules résistantes ont été utilisées pour l'analyse moléculaire.

2.2.1.4 Méthodes d'analyse des transcrits

Extraction des ARN totaux

Les ARN totaux sont préparés à partir de tiges (pour les lignées ARNi) et de feuilles (pour les lignées surexprimeuses) de jeunes plantules en utilisant le kit *Nucleospin® RNA II* (Clontech), selon le protocole décrit par le fournisseur. Plus de 30 mg de tissus d'*A. thaliana* sont broyés dans un mortier en présence d'azote liquide. La poudre est homogénéisée dans un tube Eppendorf en présence de 350 µL de tampon de lyse additionné de 3,5 µL de β-mercaptoéthanol, agitée au vortex et centrifugée 1 min à 11000 g dans une colonne équipée d'un filtre pour éliminer les débris cellulaires. Après addition de 35 µL d'éthanol à 70% et agitation, les tubes sont centrifugés 30 s à 11000 g ce qui permet de fixer les ARN sur la matrice de la colonne. Ils sont ensuite déssalés, lavés et élués finalement avec 60 µL d'eau ultrapure stérile. La qualité des ARN totaux est contrôlée par électrophorèse en gel d'agarose à 1% où les ARNr 25S et 18S apparaissent comme deux bandes intenses. La bande correspondant aux ARNr 25S doit être approximativement deux fois plus intense que celle correspondant aux ARNr 18S. La quantité d'ARN totaux est évaluée par la mesure de l'absorbance d'une fraction de 1 µL au Nanodrop (Labtech International Ltd). Le rapport de la mesure à des longueurs d'onde de 260 nm et 280 nm doit être supérieur à 1,8 pour une bonne qualité de purification. Afin d'éliminer les éventuelles traces d'ADN génomique, qui parasiteraient les réactions PCR (*Polymerase Chain Reaction*), les échantillons sont soumis à un traitement à la DNase I.

Traitement DNase I

Le protocole utilisé est fourni par le fournisseur du kit RQ1 *RNase-Free DNase* (Promega). Aux ARN obtenus précédemment sont ajoutés 1 µL de tampon et 1 µL de DNase I (1 U/µL). Le mélange ainsi obtenu est placé 30 min à 37°C. Après ce traitement, l'enzyme est inactivée par la chaleur, pendant 10 min à 65 °C. L'efficacité du traitement a été vérifiée en comparant les résultats d'amplification des acides nucléiques entre l'échantillon traité par la DNase I à celle de l'ADN génomique.

Réaction de transcription inverse (RT)

La synthèse des ADN complémentaires (ADNc) est réalisée avec le kit *SuperScript® II reverse Transcriptase* (Invitrogen). Un microgramme d'ARN totaux, sont rétrotranscrits selon le protocole décrit par le fournisseur. Les ARN totaux sont rétrotranscrits grâce à l'action d'une enzyme multifonctionnelle, la transcriptase inverse, qui possède trois activités distinctes : ADN polymérase ARN-dépendante, exoribonucléase hybride-dépendante et ADN polymérase ADN-dépendante. Pour la rétrotranscription *in vitro*, les deux premières activités de l'enzyme sont utilisées pour générer un ADNc simple brin. Un mix réactionnel composé de 1 µL d'oligo (dT)₁₅ à 500 µg/mL (Promega), 1 µL de dNTP à 10 mM, 1 µL d'ARNm de la desmine, 1,7 µM de DMSO est ajouté à 5 µL d'ARN à 1 µg/mL. Le mélange réactionnel est incubé 5 min à 65°C (pour dénaturer les ARN) puis refroidi immédiatement dans la glace. Un autre mix réactionnel composé de 4 µL de 5x *First Strand Buffer* (Invitrogen), 1 µL de DTT à 0,1 M (Invitrogen) et 0,5 µL de *RNasine plus RNase Inhibitor* (Promega) est ajouté et le mélange réactionnel est incubé 2 min à 40°C avant d'ajouter la transcriptase inverse (Invitrogen). Le mélange final est incubé 1 h à 40°C. L'enzyme est inactivée par chauffage à 70°C durant 15 min. Une fraction de la réaction est ensuite utilisée pour analyser par PCR la quantité de transcrits des gènes d'intérêt (Figure 2.2).

Oligonucléotide	Séquence nucléotidique 5' → 3'	Fragment amplifié/Taille
S1 AS1	ATGGGTTTCATTGGTAAGAGTG TTTGGGGCAAGACGGGTTAAAG	- Gène complet/1077 pb - Peptide signal + PAC/477pb (plantes surexprimeurs du domaine PAC)
S2 AS1	TTCAACCGTAGTCTAGTCGCCG TTTGGGGCAAGACGGGTTAAAG	Domaine PAC seul/405 pb
At1g28290_ARNi_F At1g28290_ARNi_R	CTCCAGTTAAGCCCCAAC TTCAACCGTACAGTAGACGCATA	Séquence de 651 pb (plantes ARNi et WT)
S1_qPCR AS1_qPCR	AACCATAAGACTCAAACCCCTTC TGAGGGTGGTGGTGATGAG	Séquence de 95 pb de la partie N-terminale du gène
S2_qPCR AS2_qPCR	GTCTCAAAGCTCTTCCGGTG TTGTTGACGACCACCGTAGA	Séquence de 86 pb du domaine PAC du gène
Actine 8 S Actine 8 AS	CACCCGAGAGGAAGTACAGTG CATACTCTGCCTTAGAGATCCACA	Séquence de 93 pb de <i>At1g49240</i>

Figure 2.3 : Amorces nucléotidiques utilisées pour l'analyse de l'expression du gène *At1g28290*, de sa séquence C-terminale et du gène contrôle *At1g49240*. S : sens ; AS : anti-sens

Amplification de séquences nucléiques

Réaction de polymérisation en chaîne (PCR)

La technique de PCR décrite (Saiki *et al.*, 1985) permet d'amplifier des séquences géniques grâce à plusieurs cycles d'amplification comprenant trois étapes: la dénaturation thermique de l'ADN matriciel bicaténaire, l'hybridation des amorces sur l'ADN matriciel, et la synthèse, à partir des amorces, d'ADN complémentaire grâce à la *Taq DNA polymérase*. Les réactions de PCR sont réalisées dans un mélange réactionnel de 25 µL comprenant le tampon de la *Taq DNA polymérase*, une combinaison d'amorces (0,2 µM chaque) correspondant à la séquence d'intérêt, les dNTP (0,2 mM chaque) et 1 U de *Taq DNA polymérase* (*GoTaq® DNA Polymerase*, Promega). Les différents couples d'amorces utilisés au cours de cette étude sont présentés dans la figure 2.2. Les réactions d'amplification sont

réalisées dans un thermocycleur (*GeneAmp[®] PCR System 2700, Applied Biosystems*). Les échantillons sont soumis à une dénaturation préalable à 94°C pendant 5 min, puis à plusieurs cycles d'amplification comprenant une étape de dénaturation à 94°C pendant 30 s, une étape d'hybridation des amorces sur la matrice à la température adéquate pendant 30 s, et une étape de synthèse d'ADN à 72°C dont la durée varie en fonction de la taille du fragment à amplifier. Une extension finale à 72°C pendant 7 min termine le programme de PCR.

PCR semi-quantitative

C'est une PCR après laquelle la quantité de produit amplifié est quantifiée sur gel. Il s'agit de mesurer des quantités relatives grâce à des ARN standards. La RT-PCR semi-quantitative doit être tout d'abord rapportée à un contrôle correspondant à l'amplification d'un transcrite exprimé constitutivement. Le transcrite contrôle utilisé ici est celui de la β - tubuline. L'amplification du transcrite du fragment d'intérêt doit être optimisée afin d'atteindre la phase linéaire de la courbe d'amplification. Les réactions de PCR sont multiparamétriques. Les paramètres qu'il est aisé de modifier pour optimiser les réactions de PCR semi-quantitatives sont : le nombre de cycles, la température d'appariement des amorces à la matrice, la quantité de matrice ADN (sachant que sur recommandation du fabricant, le volume de produit de la retrotranscription ne doit pas excéder 1/10^{ème} du volume réactionnel final). Dans le cas de notre étude, différents nombres de cycles (30, 35, 40 cycles) sont testés afin d'obtenir une mesure au cours de la phase exponentielle et non pas au niveau du plateau. Les amorces utilisées pour quantifier les transcrits d'*At1g28290* sont présentées dans la figure 2.2. La quantité d'amplicons est visualisée par coloration au bromure d'éthidium (BET) après électrophorèse en gel d'agarose.

PCR en temps réel

La PCR quantitative en temps réel est utilisée afin de quantifier, de façon relative, l'expression de gènes candidats dans les plantes transgéniques ou sauvages, rapportée à l'expression de gènes constitutifs. Les couples d'amorces utilisés pour analyser le taux de transcrits d'*At1g28290* des plantes sauvages, surexpresseurs d'*At1g28290* et ARNi ainsi que ceux surexpresseurs du domaine PAC sont présentés dans la figure 2.2. Les expériences de PCR quantitative en temps réel sont effectuées avec le système *LightCycler[®] 1.5* (Roche), en utilisant le kit *Light Cycler Fast Start DNA Master^{PLUS} SYBER Green I* » (Roche), selon les recommandations du fabricant. Le mélange réactionnel est constitué de 0,5 μ L de chaque primer (10 μ M final), 1 μ L de la matrice d'ADNc diluée au 1/2^{ème} obtenue à partir de la réaction de RT-PCR, 1 μ L d'un mix SYBR Green (Roche) et d'eau qsp 10 μ L. Ces différentes préparations contiennent toutes le tampon de la réaction ainsi que l'enzyme. Les résultats d'expression relative sont calculés d'après la méthode comparative des Ct (*threshold cycle* ou cycle seuil) (Notice explicative d'Applied Biosystem). L'expression d'un gène d'intérêt est normalisée entre tous les échantillons par rapport à l'expression du gène de référence interne. Nous avons utilisé chaque fois le gène *At1g49240* (actine 8) d'*A. thaliana*. La différence des Ct est donc égale à : $\Delta Ct[\text{échantillon}] = Ct(\text{gène d'intérêt}) - Ct(\text{actine})$. Le résultat est ensuite rapporté à un calibrateur (considéré comme l'expression normale du gène, par exemple dans le cas d'un témoin ce qui donne $\Delta\Delta Ct = \Delta Ct[\text{échantillon}] - \Delta Ct[\text{calibrateur}]$. La quantité de matrice amplifiée doublant à chaque cycle de PCR, l'expression relative du gène est finalement calculée par la formule : $2^{-\Delta\Delta Ct}$.

2.2.2 Expression des protéines recombinantes chez *E.coli*

A mon arrivée au laboratoire, l'ADNc codant pour le domaine PAC d'AGP31 avait été cloné dans le vecteur pBAD-TOPO permettant de fusionner deux *tags* (épitope V5 et 6xHis) à l'extrémité C-terminale de la protéine (construction réalisée par D. Roujol et L. Cisneros).

Cette même construction en fusion avec la *Maltose Binding Protein* (MBP) en N-terminal de la protéine d'intérêt a été ensuite réalisée par D. Roujol.

2.2.2.1 Clonage dans le vecteur pBAD-TOPO et expression des protéines recombinantes PAC chez *E. coli*

Construction du vecteur d'expression

La construction PAC est constituée de la partie C-terminale de la protéine AGP31 suivie de deux tags épitope V5 et 6xHis. Pour préparer cette construction, l'ADNc codant pour le PAC a été obtenu par PCR à partir du clone ADNc RIKEN pda 19846 avec les deux oligonucléotides adéquats (Figure 2.2). Les fragments de PCR ont été analysés par électrophorèse en gel d'agarose 1%, puis purifiés avec le *QIAEX II Gel Extraction Kit* (QIAGEN). Les produits de PCR obtenus à l'aide de la *Pfu DNA polymerase* sont ensuite rallongés d'une adénine puis sont clonés directement dans le vecteur d'expression pBAD-TOPO (Invitrogen). La bonne intégration des séquences et le cadre de lecture ouvert ont été confirmés par séquençage. Cette construction est ensuite utilisée pour transformer des cellules compétentes *E. coli* TOP10 (Invitrogen) par choc thermique (30 s à 42°C). Ce gène hybride est sous la dépendance du promoteur *araB*, assurant une expression très contrôlée et dose-dépendante en présence de L-arabinose. Ces plasmides présentent un gène conférant la résistance à l'ampicilline.

Induction de l'expression des protéines

Les clones bactériens contenant le vecteur d'expression pBAD-TOPO recombinant, sont mis en préculture une nuit à 37°C à partir de stocks glycérol dans 1 mL de milieu LB additionné d'ampicilline (concentration finale de 100 µg/mL). Ces précultures permettent d'ensemencer des cultures de 100 mL. Afin d'optimiser la production, différentes conditions d'expression ont été testées en jouant sur différents paramètres au cours de l'induction : concentration en arabinose (0,2%, 0,02%, 0,002%, 0,0002% et 0,00002%), température (25°C, 30°C, 37°C et 45°C) et temps d'induction (1 à 4 h), sous agitation.

2.2.2.2 Clonage dans le vecteur pMAL-c2 et expression des protéines recombinantes PAC chez *E. coli*

Construction du vecteur d'expression

La séquence correspondant au PAC a été amplifiée par PCR à partir de l'ADN plasmidique pBAD-TOPO, qui contient la séquence codante authentifiée, à l'aide des couples d'amorces dessinées pour l'insertion dans le vecteur pMAL-c2. Les produits des doubles digestions *EcoR* I et *Pst* I du produit PCR ont été clonés entre les sites *EcoR* I et *Pst* I du vecteur d'expression pMAL-c2, en aval du gène *malE* qui code pour la MBP. Pour contrôler le clonage, un criblage par PCR a été effectué, puis le produit PCR a été analysé par électrophorèse en gel d'agarose 1%. Les clones présentant la taille d'insert attendue ont été séquencés pour confirmer la bonne intégration des séquences et le cadre de lecture ouvert de la nouvelle construction. Cette construction a ensuite été utilisée pour transformer des cellules compétentes *E. coli* TOP10. Le gène hybride obtenu est sous la dépendance du promoteur inductible par l'IPTG (isopropyl-1-thio-β-Dgalactopyranoside). Dans la séquence de protéine recombinante, entre le *tag* MBP et la séquence protéique, un site de clivage par le facteur Xa (-Ile-Glu-Gly-Arg-) est introduit lors du clonage permettant de se débarrasser du *tag* MBP.

Induction de l'expression des protéines

Les protéines recombinantes PAC fusionnées à deux *tag* V5 et 6xHis en C-terminal et MBP en N-terminal ont été exprimées dans les bactéries compétentes *E. coli* TOP10. A partir d'une colonie isolée, 1 mL de préculture est réalisée servant ensuite à inoculer 100 mL de milieu LB additionné d'ampicilline (100 µg/mL). Lorsque la DO_{600 nm} atteint 0,5, l'expression est induite par addition d'IPTG (0,3 mM final), à 37 °C sous agitation.

2.2.3 Méthodes d'extraction de protéines totales

2.2.3.1 Extraction rapide de protéines bactériennes ou végétales pour analyse par SDS-PAGE ou *lectin blot*

Le matériel végétal (*A. thaliana* et *N. benthamiana*) est finement broyé dans un mortier à l'aide d'azote liquide. Une partie de la poudre esr récupérée à laquelle on ajoute 2 volumes de tampon de charge 4X (500 mM Tris-HCl pH 6,8, glycérol (vol/vol) 40%, β-mercaptoéthanol 20%, SDS 8% (p/v), bleu de Bromophénol 0,4%). Dans le cas des bactéries, le culot est directement repris dans le tampon de charge. L'échantillon est dénaturé pendant 5 min à 100°C, puis centrifugé brièvement à 16000 g. Le surnageant peut être alors utilisé pour une séparation par électrophorèse en conditions dénaturantes.

Pour les analyses par *lectin blot*, le matériel végétal est finement broyé dans un mortier à l'aide d'azote. On récupère 40 mg de la poudre lyophilisée à laquelle on ajoute 250 µL de tampon d'extraction (Urée 6 M, glycérol 10% (p/v), β-mercaptoéthanol 10% (vol/vol), SDS 5% (p/v)) dilué à moitié. L'échantillon est centrifugé brièvement à 16000 g et 1 µL de surnageant est déposé sur membrane de nitrocellulose pour le test de *lectin blot*.

Cette méthode ne permet pas de doser les protéines, mais elle est très rapide et efficace pour effectuer un contrôle qualitatif des protéines.

2.2.3.2 Extraction des protéines pariétales

Préparation de parois végétales

Après récolte d'environ 50 g d'hypocotyles étiolés de 11 j prélevés dans 55 à 60 boîtes Magenta, une préparation de parois est réalisée en suivant le protocole décrit par Feiz *et al.* (2006). Ce même protocole est adapté pour l'extraction de la paroi des feuilles d'*A. thaliana* et de *N. benthamiana*. Le matériel végétal est lavé sur tamis métallique (taille des pores : 1 mm²) avec 8 L de tampon acétate de sodium 4,7 mM pH 4,2 puis 1 L de tampon acétate de sodium 4,7 mM pH 4,6 / 0,4 M saccharose. Il est alors transféré dans un mixer contenant environ 1,6 L de tampon acétate de sodium 4,7 mM pH 4,6 / 0,4 M saccharose additionné de 2,8 mL de cocktail d'inhibiteurs de protéases (Sigma) et de 8 g de Polyclar. La mixture est broyée en chambre froide 8 fois 1 min, en effectuant des pauses de 30 s entre chaque cycle. Les parois sont ensuite séparées du matériel cytoplasmique par centrifugation (1000 g, 15 min, à 4°C). Le culot est récupéré, puis lavé successivement par 900 mL de tampon acétate de sodium 4,7 mM pH 4,6 additionné de 0,6 M de saccharose puis à nouveau par 900 mL de ce même tampon additionné de 1 M saccharose. Après la dernière centrifugation, le culot est lavé sur tamis de nylon (taille des pores : 25 µm²) avec 10 L de tampon acétate de sodium 4,7 mM pH 4,6, puis broyé dans l'azote liquide à l'aide d'un mortier et d'un pilon ou d'un broyeur à bille Danguomeau afin d'obtenir une poudre homogène qui est lyophilisée jusqu'à dessiccation complète (environ 48 h). Les parois lyophilisées sont stockées à -20°C jusqu'à leur utilisation.

Extraction des protéines pariétales

L'extraction de protéines pariétales est réalisée en suivant le protocole décrit par Feiz et al., (2006). Les protéines sont solubilisées, à raison de 10 mL par g de parois lyophilisées, dans un tampon acétate de sodium 4,7 mM pH 4,6 / 0,2 M CaCl₂ auquel on ajoute 15 µL/g de cocktail d'inhibiteurs de protéases (Sigma). Après avoir vortexé et laissé agir dans la glace 10 min, on centrifuge à 40000 g, à 4°C pendant 15 min. Le culot obtenu est ensuite successivement repris par 5 mL/g de parois de tampon acétate de sodium 4,7 mM pH 4,6 / 0,2 M CaCl₂, puis par 6 mL/g de parois de tampon acétate de sodium 4,7 mM pH 4,6 / 4 M LiCl en présence de 15 µL de cocktail d'inhibiteurs de protéases et enfin par 3,8 mL/g de parois de ce même tampon acétate à 2 M LiCl. Les surnageants contenant les protéines extraites sont récupérés à chaque étape, puis ils sont tous les quatre rassemblés.

2.2.3.3 Extraction des protéines recombinantes d'*E. coli*

Extraction de la protéine PAC/V5/6xHis

Après surexpression de la protéine recombinante PAC/V5/6xHis, la culture bactérienne de 50 mL est centrifugée à 4000 g, pendant 10 min à 4°C. Le culot cellulaire est repris dans 10 mL de tampon de lyse (50 mM NaNa₂PO₄, 500 mM NaCl, 10 mM imidazole, 1 mM DTT). Cinquante mg/mL de lysozyme sont alors ajoutés aux culots remis en suspension qui sont incubés 1 h à 37°C sous agitation douce. Après sonication dans la glace 20 x 30 s à une intensité de 80% (sonicateur Bandelin Sonoplus – HD 2070 avec une sonde de diamètre 3 mm), l'extrait bactérien est centrifugé 20 min à 25000 g. Pour solubiliser les corps d'inclusion, le culot qui contient la protéine d'intérêt est lavé 2 X par du tampon de lyse additionné de Triton X-100 0,5% et est resuspendu dans du tampon de lyse additionné de 8 M d'urée, soniqué 3 x 30 s et centrifugé 20 min à 40000 g. Le surnageant ainsi obtenu constitue la fraction de protéines contenant la protéine d'intérêt en conditions dénaturantes.

Extraction de la protéine MBP/PAC/V5/6xHis

Après surexpression de la protéine recombinante MBP/PAC/V5/6xHis, la culture bactérienne de 100 mL est centrifugée à 4000 g, pendant 10 min à 4°C. Le culot cellulaire est repris dans 5 mL de tampon de lyse (Tris-HCl 20 mM pH 7,5, NaCl 200 mM, EDTA 1 mM, azide de sodium 1 mM, β-mercaptoéthanol 10 mM). Les cellules resuspendues sont soumises à une sonication dans la glace 8 x 15 s à une intensité de 80%. Le lysat acellulaire est centrifugé (9000 g, à 4°C pendant 20 min). Le surnageant ainsi obtenu constitue la fraction de protéines solubles en conditions non dénaturantes qui contient la protéine d'intérêt.

2.2.4 Techniques d'analyse des protéines

2.2.4.1 Dosage des protéines

La quantité de protéines est estimée selon la technique mise au point par Bradford (1976) en utilisant le réactif *Coomassie protein assay reagent* (Uptima) et la serum albumine bovine (BSA) comme standard, selon les recommandations du fournisseur.

2.2.4.2 Détection des protéines

Séparation des protéines par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE)

Préparation du gel de polyacrylamide

La séparation par SDS-PAGE (*Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis*) est réalisée à l'aide d'un gel de séparation à 12,2%. Un gel de concentration (*stacking gel*) à 4,8% d'acrylamide est coulé en haut du gel de séparation pour permettre une entrée homogène de l'échantillon dans le gel de séparation (Figure 2.3). Des gels de taille de 10 x 12 x 0,15 cm sont utilisés. Les échantillons de protéines à séparer sont dilués dans du tampon de charge 1X (cf. § 2.3.1) puis dénaturés pendant 5 min à 100°C. En conditions non dénaturantes, les échantillons de protéines sont dilués dans du tampon de charge en absence de β - mercaptoéthanol (agent réducteur qui réduit les ponts disulfures) et de SDS (détergent anionique fort qui déplie les chaînes polypeptidiques) et sont déposés sur gel sans chauffage préalable.

Solution	Gel de séparation	Gel de concentration
Acrylamide 40%	3,04 mL	583 μ L
Bis-Acrylamide 2%	1,67 mL	322 μ L
Tris-HCl 1,5 M pH 8,8	2,5 mL	-
Tris-HCl 1,25 M pH 6,8	-	0,5 mL
SDS 10% (p/v)	200 μ L	50 μ L
Eau	2,64 mL	3,52 mL
Persulfate d'ammonium 100 mg/mL	50 μ L	25 μ L
TEMED	5 μ L	5 μ L
Eau	2,64 mL	3,52 mL

Figure 2.4 : Composition des gels de polyacrylamide-SDS utilisés pour la séparation des protéines par SDS-PAGE

Séparation des protéines

L'électrophorèse est réalisée dans un tampon de migration (Tris 25 mM ; Glycine 192 mM; SDS 0,1% p/v), à 25 mA jusqu'à ce que le front de migration arrive en bas du gel (cuve d'électrophorèse « Mini-PROTEAN® Tetra Cell », BIO-RAD). Après séparation, les protéines peuvent être colorées ou bien transférées sur une membrane pour une immunodétection. En conditions non dénaturantes, le SDS doit être enlevé du tampon de migration.

Coloration des protéines au Bleu de Coomassie

Après 3 lavages de 10 min chacun par de l'eau désionisée, le gel est coloré par le bleu de Coomassie *PageBlue™ Protein Staining Solution* (Fermentas) une nuit sous agitation lente à température ambiante et conservé à 4°C dans une solution contenant 1 à 2% d'acide acétique.

Electrotransfert des protéines

Après électrophorèse, les protéines sont électrotransférées sur membrane de PVDF (polyvinylidène difluoride) (Schleicher & Schuell) ou de nitrocellulose (Hybond™-C Extra,

Amersham Biosciences) afin de permettre une immunodétection des protéines d'intérêt. Deux méthodes ont été employées :

- La membrane de PVDF est préalablement activée pendant 15 s dans un bain d'éthanol absolu afin d'être ionisée, rincée 5 min dans de l'eau distillée, puis équilibrée 30 s dans le tampon de transfert (48 mM Tris, 39 mM glycine, SDS 0,037% (m/v)). Les protéines du gel sont alors électrotransférées dans la cuve de transfert humide (Mini Trans-Blot Cell, BIO-RAD) pendant 2 h 30, sous voltage constant de 75 V et à 4°C. Pendant le transfert, la cuve est refroidie par un bloc de glace et le tampon agité en permanence.

- La membrane de nitrocellulose est hydratée dans de l'eau distillée puis équilibrée 10 min dans le tampon de transfert additionné de 20% de méthanol. Après équilibration, les protéines du gel sont alors électrotransférées dans un système semi-sec (Trans-Blot SD Semi-Dry Transfer Cell, BIO-RAD) pendant 1 h, sous voltage constant de 20 V.

Le transfert est réalisé en plaçant le gel au contact de la membrane entre deux feuilles Whatman® (Whatman International LTD) et deux éponges imbibées de tampon de transfert (seulement dans le cas du transfert humide), orienté vers la cathode et la membrane orientée vers l'anode du système permettant ainsi aux protéines, toutes chargées négativement par le SDS dans le gel, de se déplacer suivant les courants (électriques et salins) pour se déposer sur la membrane.

Immunodétection des protéines recombinantes

Saturation de la membrane

Afin de saturer les sites de fixation non spécifiques, la membrane est placée dans une solution de saturation composée de 10% de lait écrémé repris dans du tampon TBST (TBS 1X : 20 mM Tris-HCl pH 7,5, 150 mM NaCl additionné de 0,05% Tween 20) sous agitation pendant une nuit à 4°C (ou 1 h à température ambiante).

Incubation de la membrane avec l'anticorps primaire

La membrane est rincée 2 fois pendant 5 min dans un bain de TBST puis incubée dans la solution d'anticorps primaire anti-V5 (Invitrogen) ou anti-MBP (New England Biolabs) dilué à 1/10000^{ème} dans du TBST-BSA 1% pendant 3 h à température ambiante (ou une nuit à 4°C).

Incubation de la membrane avec l'anticorps secondaire

Après 3 rinçages de 10 min dans du TBST, la membrane est incubée pendant 2 h dans la solution d'anticorps secondaire anti-souris (Promega) (anticorps dilué au 1/10000^{ème} dans du TBST-BSA 1%) dirigé contre les immunoglobulines G de l'anticorps primaire et couplé à la phosphatase alcaline.

Révélation

Après incubation avec l'anticorps secondaire, la membrane est rincée 3 fois 10 min avec du tampon TBST. La présence de phosphatase alcaline est révélée par incubation de la membrane avec un substrat chromogène spécifique : 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (Sigma) à 50 mg/mL et nitro bleu de tétrazolium (NBT) (Sigma) à 50 mg/mL en DMF 70% (Diméthylformamide). La révélation se fait par ajout de 33 µL de BCIP additionné de 33 µL de NBT dans 10 mL de TBS à pH 9,5. Après révélation, la membrane est rincée dans de l'eau distillée, séchée et photographiée. Différents temps d'exposition sont réalisés jusqu'à l'obtention de bandes d'intensité satisfaisante, mais non saturante.

Détection des glycoprotéines par *lectin-blot*

Les glycoprotéines fixées sur la membrane sont détectées par la lectine PNA ou GNA en utilisant le kit *DIG Glycan Differentiation Kit* (Roche). Tous les réactifs utilisés sont fournis par le kit sauf le TBS 1X, le tampon 1 (TBS 1X, 1 mM MgCl₂, 1 mM MnCl₂, 1 mM CaCl₂, pH 7,5) et le tampon 2 (0,1 M Tris, 0,1 M NaCl, 0,05 M MgCl₂, pH 9,5). Après transfert ou dépôt direct des échantillons sur la membrane, celle-ci est bloquée pendant une nuit à 4°C sous agitation par environ 30 mL de solution *Blocking reagent* fournie dans le kit. Les étapes suivantes se font à température ambiante, sous agitation, dans les conditions préconisées par le fournisseur. Après le blocage, la membrane est lavée 2 x 10 min dans du TBS 1X et 1 x 10 min dans le tampon 1. La membrane est ensuite incubée pendant 1 h sous agitation par 10 mL de lectine PNA ou GNA couplée à la digoxigénine (DIG) diluée à 1/100^{ème} (cas de PNA) ou à 1/200^{ème} (cas de GNA) dans le tampon 1. Après incubation, la membrane est lavée 3 x 10 min dans environ 50 mL de TBS 1X, puis incubée 1 h dans 10 mL d'anticorps anti-DIG couplé à la phosphatase alcaline diluée au 1/1000^{ème} dans du TBS. La membrane est de nouveau lavée 3 x 10 min dans environ 50 mL de TBS 1X. La révélation se fait par ajout de 200 µL de NBT/BCIP préparés extemporanément dans 10 mL de tampon 2. Après apparition du signal (environ 5 min), la membrane est rincée 2 fois à l'eau UHQ pour arrêter la réaction.

2.2.4.3 Identification des protéines par spectrométrie de masse

Préparation des échantillons et digestion trypsique

Les bandes d'intérêt visualisées sur gels de polyacrylamide sont excisées à l'aide d'une pipette Pasteur. Les morceaux de gel sont ensuite lavés 2 fois par ajout de 100 µL d'une solution de 50% bicarbonate 25 mM, 50% acétonitrile, puis séchés 20 min dans un speed-vac à 45°C. Dix µL d'une solution de trypsine (Promega) à 10 ng/µL sont ajoutés dans chaque tube. Les échantillons sont ensuite placés à l'étuve à 37°C pour la nuit. Après digestion trypsique, 1 µL d'acétonitrile est ajouté à chaque échantillon. Les tubes sont ensuite vortexés 5 min, puis soniqués pendant 5 min.

Analyse par spectrométrie de masse (SM) MALDI-TOF

C'est un spectromètre de masse couplant une source d'ionisation laser assistée par une matrice MALDI (*Matrix-Assisted Laser Desorption/Ionisation*) et un analyseur à temps de vol TOF (*Time-Of-Flight*). Un volume de 1 µL d'échantillon issu de la digestion trypsique additionné de 1 µL de matrice (acide -hydroxycinnamique) est déposé par puits sur une plaque d'analyse pour spectrométrie de masse (MALDI *Sample plate* SS, Applied Biosystems). Les analyses sont effectuées sur la plateforme de spectrométrie de masse de l'IPBS (<http://proteomique.ipbs.fr/plateforme/plateforme.htm>) à l'aide d'un spectromètre de masse MALDI-TOF (Voyager-DeTM STR, Applied Biosystems), en mode réflectron avec les paramètres suivants : tension d'accélération 20 kV, tension de la grille 68%, délai 200 ns. Une calibration externe est réalisée avec un mélange de peptides de masses mono-isotopiques connues. L'intervalle d'acquisition est de 700 à 4000 m/z. Les données de SM MALDI-TOF sont ensuite traitées à l'aide du logiciel Data Explorer version 4 (*Applied Biosystems*).

2.2.5 Purification des protéines

Diverses techniques sont utilisées pour purifier les protéines d'intérêt et dans la majorité des cas la purification est effectuée en deux étapes. La chromatographie d'échange de cations (CEC) ou d'affinité sur colonne (NAC). Entre les deux étapes de purification, un aliquot de chaque fraction est prélevé pour doser les protéines. Le reste est dialysé par le tampon adéquat

ou dessalé sur colonne Econo-Pac 10DG (BIO-RAD), puis concentré par lyophilisation pendant 24 h.

2.2.5.1 Purification par le système FPLC (*Fast Protein Liquid Chromatography*)

Un appareil de chromatographie automatisé (AKTA, GE Healthcare) réfrigéré à 4°C est utilisé. Il contient deux pompes permettant l'injection de tampon A (tampon d'équilibration et de lavage) et de tampon B (tampon d'élution).

Chromatographie d'échange de cations

La chromatographie d'échange de cations est effectuée sur une colonne SP-sepharose, HiTrap (GE Healthcare, 1 mL) mise en œuvre sur l'automate de chromatographie FPLC. La colonne est activée préalablement par 5 mL de tampon d'élution B (50 mM MES, 1,6 M NaCl, pH 5,6) à 1 mL/min, puis équilibrée par 10 mL de tampon de charge (A) (50 mM MES, pH 5,6 ajusté avec KOH) à 1 mL/min. Entre 1 et 2 mg de protéines lyophilisées sont solubilisées dans 10 mL de tampon de charge. Elles sont ensuite déposées à un débit de 0,5 mL/min sur cette colonne et l'effluent est récupéré (10 mL). Après une étape d'élution par un gradient continu de 0 à 50% de tampon d'élution, à 1 mL/min pendant 24 min, un palier à 50% de tampon d'élution est réalisé pendant 3 min à 1 mL/min suivi de deux autres, l'un à 75% et l'autre à 100% de tampon B réalisés pendant 3 min chacun à 1 mL/min. Les fractions d'élution sont collectées puis rassemblées par 3. Un aliquot de chaque mélange de fractions est prélevé pour le dosage des protéines. Le reste est dessalé sur colonne Econo-Pac 10DG (BIO-RAD), puis lyophilisé pendant 24 h.

Purification des protéines sur colonne HisTrap

La chromatographie d'affinité vis-à-vis des résidus His est effectuée sur une colonne *HisTrap FF crude* (GE Healthcare, 1 mL).

Purification de PAC/V5/6xHis à partir de la fraction bactérienne soluble

La première étape de la purification est l'équilibration de la colonne par le tampon A (50 mM NaNa_2PO_4 pH 7,4, 500 mM NaCl, 10 mM imidazole, 1 mM DTT). Vingt mg de l'extrait protéique bactérien soluble dilué 3x dans le tampon A sont fixés sur la colonne (débit : 0,5 mL/min). Des lavages (10 vol de colonne) sont effectués avec du tampon A à 10 mM d'imidazole pour éliminer les protéines non fixées. Suite aux lavages, plusieurs conditions d'élution de 10% à 100% de tampon B (tampon A additionné de 300 mM d'imidazole), par palier ou par gradient en conditions non dénaturantes, ont été testées.

Purification de PAC/V5/6xHis à partir de la fraction bactérienne insoluble

La première étape de la purification est l'équilibration de la colonne par le tampon A constitué de tampon de lyse (50 mM Na_2HPO_4 - NaH_2PO_4 pH 8, 500 mM NaCl, 10 mM Imidazole, 1 mM DTT) additionné de 8 M d'urée. Douze mg de l'extrait protéique de la fraction insoluble dilué 3X dans le tampon A est fixé sur la colonne (débit : 0,5 mL/min). Des lavages (10 vol de colonne) sont ensuite effectués avec du tampon A à 10 mM imidazole pour éliminer les protéines non fixées. Suite aux lavages, un gradient linéaire de tampon A contenant 8 M à 0 M d'urée est effectué en 30 min pour la renaturation des protéines. L'élution des protéines est effectuée par gradient de 0% à 100% de tampon B (tampon A additionné de 500 mM imidazole et sans urée) en 15 min.

Purification de MBP/PAC/V5/6xHis à partir de la fraction bactérienne soluble

La première étape de la purification est l'équilibration de la colonne dans le tampon A (10 mM Na_2HPO_4 - NaH_2PO_4 pH 7,4, 500 mM NaCl). Trente cinq mg de l'extrait protéique

dilué 10X dans le tampon A à 20 mM imidazole sont fixés sur la colonne (débit : 0,5 mL/min). Des lavages (10 vol de colonne) sont effectués avec du tampon A à 20 mM imidazole. Suite aux lavages, plusieurs conditions d'élution de 0% à 100% de tampon B (10 mM Na₂HPO₄-NaH₂PO₄ pH 7,4, 500 mM NaCl, 300 mM d'imidazole), par palier ou par gradient en conditions non dénaturantes sont testées. En conditions dénaturantes, l'extrait protéique soluble dilué 10X dans du tampon A additionné de 8 M d'urée est déposé sur la colonne préalablement équilibrée par le tampon A. Ensuite, un gradient linéaire de tampon A contenant de 8 M à 0 M d'urée est effectué pour la renaturation des protéines en 30 min. L'élution des protéines est effectuée par 3 paliers, à 10%, 50% et 100% de tampon B (10 mM Na₂HPO₄-NaH₂PO₄ pH 7,4, 500 mM NaCl, 300 mM d'imidazole).

Clivage du tag MBP par le facteur Xa

Pour la coupure du tag MBP, l'échantillon purifié est dessalé, lyophilisé et repris par le facteur Xa (New England Biolabs) à 200 µg/mL (20 µg de facteur Xa par mg de protéine de fusion) dans du tampon protéase (20 mM Tris-HCl pH 8, 100 mM NaCl, 2 mM CaCl₂) et incubé à température ambiante. Une cinétique est faite en prélevant des échantillons à 5 h et 20 h, suivie d'une analyse par SDS-PAGE pour visualiser le clivage.

Des essais de clivage ont aussi été tentés sur colonne *HisTrap* lors de la purification des protéines dans les conditions dénaturantes. Après fixation des protéines sur la colonne *HisTrap* et leur renaturation, la colonne est enlevée du dispositif FPLC et lavée à l'aide de 10 vol de tampon protéase et incubée ensuite 24 h par de la protéase à 200 µg/mL à température ambiante. Après incubation, la colonne est lavée 10X par le tampon protéase et éluée comme précédemment décrit.

2.2.5.2 Chromatographie d'affinité sur colonne Ni-NTA

Le kit *Ni-NTA fast start* (QIAGEN) est utilisé. Les fractions lyophilisées issues de la purification par chromatographie d'échange de cation d'intérêt sont solubilisées dans du tampon 50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole pH 8, puis passées par gravité sur une colonne Ni-NTA de 0,5 mL. L'effluent est récupéré et rechargé sur le support. Cette opération est renouvelée 3 fois. La colonne est ensuite lavée par 2 fois 4 mL de tampon 50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole pH 8. Enfin l'élution est réalisée par 3,1 mL de tampon 50 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole pH 8.

2.2.5.3 Chromatographie d'affinité sur colonne de lectine

Une chromatographie d'affinité sur colonne de lectine est réalisée. La lectine utilisée est PNA (*Peanut Agglutinin* d'*Arachis hypogaea*) spécifique du Gal et GalNAC greffée sur un support d'agarose. Un volume de 1 mL de suspension contenant 0,5 mL de résine est prélevé avec 5 mL (10 vol) de tampon de prélavage (20 mM Tris-HCl pH 7,4, 1 M NaCl, 3,33 mM MgCl₂/MnCl₂/CaCl₂). La matrice est ensuite équilibrée avec 5 mL de tampon d'équilibration (20 mM Tris-HCl pH 7,4, 0,15 M NaCl, 1 mM MgCl₂/MnCl₂/CaCl₂). Après chaque étape (prélavage, équilibration), la matrice est laissée décanter, puis le surnageant est prélevé. Environ 1 mg de protéines pariétales dessalées et lyophilisées est repris dans 1 mL du tampon d'équilibration et mis en contact avec la matrice pendant une nuit à 4°C. L'effluent est ensuite récupéré. Après un lavage avec 3 mL de tampon d'équilibration, les protéines fixées sont éluées avec 1,5 mL de tampon d'équilibration additionné de 1 M de galactose.

2.2.6 Tests d'interactions protéines/polysaccharides

2.2.6.1 Extraction des polysaccharides pariétaux

Le protocole d'extraction a été mis au point d'après les publications de Moller et al. (2007) et Fry et al. (2000). A partir de parois d'hypocotyles étiolés d'*A. thaliana* ou de jeunes feuilles de *Brachypodium distachyon* ayant subi un traitement salin d'extraction des protéines (Feiz et al., 2006), 50 mg de matière fraîche sont pesés dans un tube eppendorf. Trois cents μL de CDTA sont ajoutés (50 mM CDTA, pH 7,5) et le mélange est laissé 1 h sous agitation. Après centrifugation 10 min à 2500 g, le surnageant qui constitue la fraction enrichie en pectines est récupéré. De la même manière, à partir du nouveau culot, les hémicelluloses sont extraites par 300 μL d'une solution de 4 M NaOH additionnée de 1% NaBH_4 (m/v). Enfin, à partir du dernier culot, la cellulose est extraite par 300 μL de cadoxen (1,2-diaminoéthane 31% (vol/vol), 0,78 M oxyde de cadmium). Des solutions de polysaccharides commerciaux sont également préparées (Figure 2.4). Les poudres sont diluées (concentration finale 1 mg/mL et 10 mg/mL) dans de l'eau, sauf la cellulose qui est dissoute dans du cadoxen et la pectine 2 où l'on rajoute une demi-pastille de soude pour une concentration finale de 1%.

Nom	Référence	Nom abrégé	Origine
Acide polygalacturonique (PGA)	SIGMA P3850	P1	Agrume
PGA	SIGMA P3889	P2	Orange
Pectine méthylée	Fluka 76282	P3	Pomme
Pectine méthylée	Fluka 76280	P4	Peau de citron
Galactane	Megazyme P-GALPOT	P5	Pomme de terre
Rhamnogalacturonane I	Megazyme P-RHAM1	P6	<i>Potato pectic fiber</i>
Arabinogalactane	Megazyme P-ARGAL	P7	Bois de mélèze
Pectine extraite de <i>B. distachyon</i>		PB	Pectine extraite des parois des feuilles jeunes de <i>B. distachyon</i>
Pectine extraite d' <i>A. thaliana</i>		PA	Pectine extraite des parois d'hypocotyles étiolés d' <i>A. thaliana</i>
Xyloglucane	Megazyme P-WYGLN	H1	Tamarin
β -glucane	Megazyme P-BGBL	H2	Orge
Xylane	SIGMA X0502	H3	Bois de bouleau
Xylane	SIGMA X4252	H4	Bois de hêtre
Arabinoxylane	Megazyme P-WAXYL	H5	Blé
Hémicellulose extraite de <i>B. distachyon</i>		HB	Hémicellulose extraite des parois de feuilles jeunes de <i>B. distachyon</i>
Hémicellulose extraite d' <i>A. thaliana</i>		HA	Hémicellulose extraite des parois d'hypocotyles étiolés d' <i>A. thaliana</i>
Cellulose	SIGMA C6413	C	Cotton
Cellulose extraite de <i>B. distachyon</i>		CB	Cellulose extraite des parois de feuilles jeunes de <i>B. distachyon</i>
Cellulose extraite d' <i>A. thaliana</i>		CA	Cellulose extraite des parois d'hypocotyles étiolés d' <i>A. thaliana</i>

Figure 2.5 : Polysaccharides utilisés pour les tests d'interaction avec les protéines d'intérêt

2.2.6.2 Digestion des hémicelluloses par la protéinase K

Les hémicelluloses extraites sont incubées 1 h à 37°C avec la protéinase K à 0,0005 µg/µL, 0,05 µg/µL ou 1 µg/µL. La réaction est arrêtée par chauffage 5 min à 100°C.

2.2.6.3 Mise en œuvre des tests d'interactions sur membrane

Le protocole mis au point a été adapté de la publication de Moller et al. (2007). Les solutions de sucre préparées (extraits ou commerciaux) sont déposés en double (1 µL par dépôt sauf cellulose : 0,5 µL) sur une membrane de nitrocellulose (8 cm x 11,5 cm) grâce au système Bio-Dot (BIO-RAD) sous vide modéré. Après dépôt, on laisse sécher les membranes 30 min à température ambiante. Les membranes sont ensuite saturées par de la BSA à 3% dans du tampon TBS pendant 1 h sous agitation. Elles sont ensuite rincées dans du tampon TBS (3 x 5 min). Les protéines d'intérêt purifiées sont préparées à une concentration minimale finale de 20 µg/mL pour les protéines de fusion avec la MBP et 10 µg/mL pour les autres protéines. Pour les tests de compétition, les protéines sont préalablement incubées 1 h avec une solution de galactane (Megazyme) à 1 mg/mL, 5 mg/mL, 10 mg/mL ou 33 mg/mL ou une solution de Gal à 1 mg/mL, 10 mg/mL, 100 mg/mL ou 125 mg/mL. Chaque membrane est incubée avec 2 mL de solution de protéine dans du tampon TBS pendant une nuit à 4°C sous agitation. Pour la révélation, et suite à 3 lavages dans du TBS, les différentes membranes sont incubées dans 5 mL d'anticorps primaire anti-V5 (Invitrogen) ou anti-MBP (New England Biolabs) dilué au 1/10000^{ème} dans du TBS-BSA 1% ou dans 5 mL de lectine PNA diluée au 1/100^{ème} dans le tampon 1 (cf. § 2.2.4.2) pendant 2 à 3 h sous agitation douce. Pour détecter les polysaccharides déposés sur la membrane de nitrocellulose, les différentes membranes sont incubées dans des anticorps monoclonaux (PlantProbes, Leeds, UK) dilués au 1/80^{ème} dans du TBS-BSA 1% : JIM5 (anti-homogalacturonane), JIM7 (anti-homogalacturonane), LM5 (anti-(1-4)-β-galactan) et LM15 (anti-xyloglucan). Les anticorps non fixés sont éliminés par 3 lavages dans du TBS avant l'incubation dans 5 mL d'anticorps secondaire anti-souris (Promega) dilué au 1/10000^{ème} dans du TBS-BSA 1% ou dans 5 mL d'anticorps anti-DIG dilué à 1/1000^{ème} en TBS. Après 2 h d'incubation sous agitation douce, les anticorps secondaires en excès sont éliminés par 3 lavages successifs dans du TBS avant d'être révélés par une solution de NBT/BCIP.

2.2.7 Protocoles de marquage pour la microscopie

2.2.7.1 Préparation des échantillons inclus dans la paraffine

Des lignées d'*A. thaliana* sauvages et transgéniques (ARNi et surexprimeurs) sont utilisées pour l'hybridation *in situ* et l'étude histochimique à l'aide de PNA couplée à la digoxygénine (PNA-dig) (Roche). L'hybridation *in situ* ayant pour but la détection d'ARNm spécifiques, il est primordial que dès les premières étapes de préparations des échantillons, des conditions « RNase free » soient strictement respectées. Pour cela, les solutions sont préparées avec H₂O-DEPC : l'eau est traitée avec 0,1% (vol/vol) DEPC (Diéthylpyrocarbonate) en agitation durant 10-24 h puis autoclavée. Du matériel plastique stérile jetable est utilisé chaque fois que cela est possible. Le matériel plastique non jetable (paniers, pinces...) est lavé 20 min dans NaOH 0,2 N et rincé 2 fois 10 min dans de l'eau traitée au DEPC avant d'être séché. Toute la verrerie est préalablement stérilisée 8h à 180°C.

Echantillonnage et fixation

Les semis correspondants aux différentes lignées transgéniques et sauvage, principalement aux stades jeunes de développement sont échantillonnés et rapidement fixés dans le FAA (Formaldéhyde 3,7%, Alcool éthylique 50%, acide acétique 5%) ou dans du formaldéhyde 3,7% dans du PBS (8 g NaCl, 1,44 g Na₂HPO₄, 0,24 g KH₂PO₄ et d'eau qsp 1 L).

L'infiltration du fixateur est facilitée par plusieurs cycles d'infiltration sous vide jusqu'à ce que les échantillons coulent. Enfin, les échantillons infiltrés sont incubés 16 à 20 h à 4°C dans la solution fixatrice.

Déshydratation des échantillons

Le fixateur est éliminé et les échantillons fixés au FAA sont rincés 3x10 min dans du PBS, puis déshydratés par des bains successifs de solutions de concentrations croissantes en alcool ter-butyle (TerBut) et décroissantes en éthanol (10/40, 20/50, 55/45, 75/25, 100/0, vol/vol, 1 h chacun à température ambiante). Lors du passage dans le bain «55/45» on ajoute de l'érythrosine de façon à obtenir une concentration finale de 0,05%, afin de colorer temporairement les échantillons en vue de faciliter leur visualisation lors de l'inclusion dans la paraffine.

Infiltration des échantillons

Le TerBut est progressivement remplacé par du paraplast (mélange de paraffine et de polymères plastiques, CML) par deux bains de paraplast/TerBut (1/1, vol/vol) respectivement durant une nuit et 4 h, puis 3 bains de paraplast pur durant 24 h, le tout à 60°C afin de conserver le paraplast à l'état liquide.

Inclusion des échantillons et microtomie

Les échantillons sont inclus dans des moules en plastique jetables ou métalliques réutilisables remplis par du paraplast maintenu liquide sur une plaque chauffante. L'inclusion se déroule lors du retour progressif de l'ensemble ainsi constitué à température ambiante. Sous cette forme les échantillons peuvent être conservés à 4°C plusieurs années.

Pour réaliser des coupes, les lames sont prétraitées par une solution de 3-aminopropyltriéthoxysilane (Sigma) à 2% (vol/vol) dans l'acétone et disposées sur une plaque chauffante à 40°C, recouvertes avec 4 mL d'eau DEPC dégazée. A partir des blocs, des coupes sériées de 10 µm sont réalisées avec un microtome. Les coupes sériées sont ainsi disposées à la surface de l'eau et les lames sont séchées toute la nuit à 40°C afin que les coupes adhèrent aux lames.

2.2.7.2 Hybridation *in situ*

Cette technique permet de mettre en évidence les sites d'hybridation des sondes anti-sens marquées à la digoxigénine avec les ARNm présents sur des coupes histologiques afin d'étudier l'expression tissu-spécifique des gènes d'intérêt. Cette approche utilise d'une part des techniques de biologie moléculaire afin de produire par transcription *in vitro* des ribosondes marquées à la digoxigénine, et d'autre part l'histologie afin de générer des coupes.

Le protocole d'hybridation a été mis au point au laboratoire (Pernot, 2010), à partir du protocole initial (Burlat et al., 2004) en incorporant certaines modifications (Brewer et al., 2006).

Production de sondes

La production de sondes a été effectuée au laboratoire par C. Pernot (Pernot, 2010). La matrice de transcription (RAFL16-52-IO5, pda19846) qui correspond à l'ADNc de *At1g28290* cloné dans un vecteur disposant de promoteurs d'ARN polymérase (T3 et T7) orientés de part et d'autre du site de clonage a été obtenue auprès du *BioResource Center* RIKEN (<http://www.brc.riken.jp/lab/epd/catalog/cdnaclone.html>) (Seki et al., 1998) puis amplifiée et vérifiée par séquençage.

Linéarisation plasmidique

Le vecteur a été linéarisé au niveau du site de multiclonaage en utilisant *BamH* I (coupure unique en aval de l'extrémité 3' de l'ADNc) ou *EcoR* I (coupure unique en amont de l'extrémité 5' de l'ADNc) afin de créer les matrices de transcription *in vitro* respectivement de la sonde sens (contrôle, identique à l'ARNm) en utilisant l'ARN polymérase T7, et de la sonde antisens (complémentaire de l'ARNm) en utilisant l'ARN polymérase T3 (Figure 2.5).

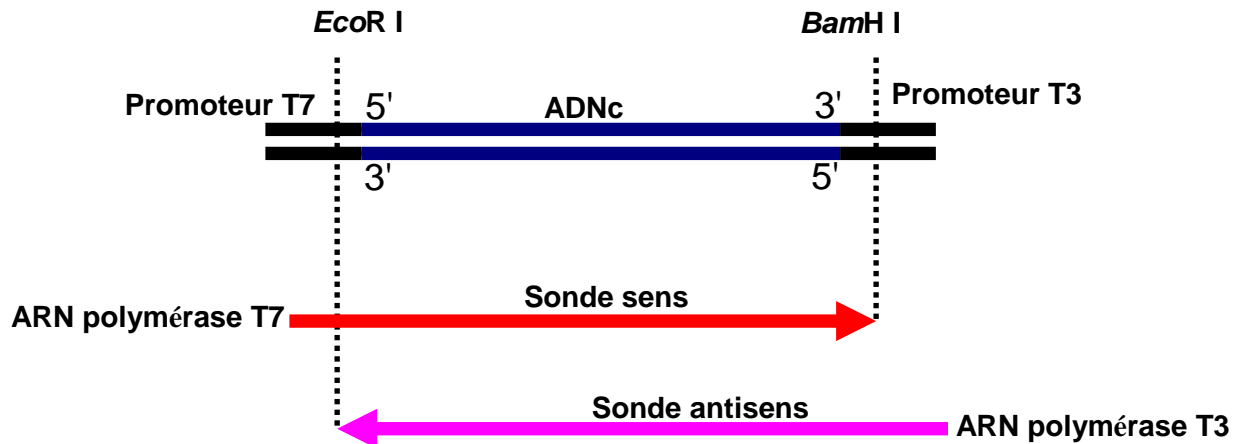


Figure 2.6 : Linéarisation de l'ADN plasmidique par digestion enzymatique et transcription *in vitro* des sondes sens contrôles (ARN polymérase T7 utilisée sur une matrice linéarisée par *BamH* I) et des sondes anti-sens complémentaires de l'ARNm (ARN polymérase T3 utilisée sur une matrice linéarisée par *EcoR* I)

Transcription et hydrolyse des sondes

Les deux transcriptions *in vitro* ont été réalisées en présence d'un mélange des 4 nucléotides ATP, CTP, GTP et UTP marqué à la digoxigénine (Roche). Les sondes sens et anti-sens marquées à la digoxigénine obtenues ont toutes subi une hydrolyse alcaline ménagée de façon à avoir des fragments de sondes d'environ 400 b pouvant accéder plus facilement aux ARNm cibles au niveau des coupes de tissus. Grâce au programme de traitement d'image ImageJ (*National Institute of Health, United States*), les sondes ont pu être dosées par rapport à l'intensité du *smear* d'une sonde déjà dosée permettant ensuite de les normaliser à une concentration dépendante de la taille initiale de la sonde de 25 ng/kb/ μ L.

Déparaffinage des coupes

La paraffine des coupes est éliminée par incubation dans 2 bains de xylène de 5 min chacun. Le xylène est éliminé dans 2 bains successifs d'éthanol absolu de 10 à 15 min chacun. La réhydratation des échantillons est effectuée dans des bains successifs de solutions d'éthanol de concentrations décroissantes dans de l'eau traitée au DEPC (100-100-95-70-50-0-0%) pendant 5 min chacun.

Pré-hybridation

Avant l'hybridation, les protéines pouvant gêner l'accès des sondes à leur cible sont digérées par traitement des coupes réhydratées avec 10 μ g/mL de protéinase K dans du

tampon TE (Tris-HCl 100 mM pH 8, 50 mM EDTA) pendant 30 min à 37°C. Cette étape est suivie d'un rinçage au PBS puis par un blocage de la protéinase K avec de la glycine (10 mg/mL dans du PBS) durant 2 min et par 2 rinçages dans du PBS. Les coupes sont ensuite post-fixées avec 3,7% de formaldéhyde dans du PBS pendant 10 min sous hotte et lavées 2 fois 5 min dans du PBS. Enfin, les coupes sont carbéthoxylées avec du DEPC (0,1% (vol/vol) dans du PBS) 2 fois pendant 15 min, lavées à 2 reprises dans du PBS, et déshydratées avec la série de bains d'éthanol en concentrations croissantes utilisée lors du déparaffinage, jusqu'à l'éthanol 100%.

Hybridation

Les lames déshydratées et séchées sont disposées dans une grande boîte de Petri *RNase Free*, sur du papier filtre Whatman®. Pour l'hybridation, chaque lame reçoit 130 µL d'un mélange contenant 200 ng/kb/mL de sonde antisens ou sens marquée à la digoxygénine, 40% de formamide, 10% de dextran sulfate, 1 µg/mL d'ARN déprotéinisés de levure, 0,3 M NaCl, 0,01 M Tris-HCl (pH 6,8), 0,01 M de phosphate de sodium (pH 6,8), 5 mM d'EDTA et 40 U/mL de RNasin (Promega) inhibitrice de ribonucléases. Cette solution est étalée à l'aide d'une lamelle (25x50 mm) posée soigneusement pour éviter les bulles d'air, afin de créer une « micro-chambre humide ». Le papier filtre est également inbibé avec un tampon SSC 2X (solution saline de citrate de sodium ; SSC 1X : NaCl 0,15 M, citrate de sodium tri-basique 0,015 M) contenant 50 % de formamide afin de stabiliser l'atmosphère dans la boîte de Petri. L'hybridation se fait à 50°C pendant une durée d'environ 16 h dans un four à hybridation.

Lavages post-hybridation

Après incubation, les lamelles sont enlevées une à une dans un bain de SSC 2X à 37°C, puis les lames sont traitées avec de la RNase A (Sigma) à 50 µg/mL dans du TBS-EDTA (10 mM Tris-HCl, 0,5 M NaCl, 1 mM EDTA, pH 7,5) afin d'éliminer les sondes en excès et d'augmenter l'accessibilité ultérieure des anticorps à la digoxygénine incorporée dans les sondes. Les lames sont ensuite lavées par 3 bains successifs sous agitation, 1 h dans SSC 2X et 1 h dans SSC 1X à température ambiante et 1 h SSC 0,1X à 65°C.

Détection immunologique des sondes marquées à la digoxygénine

Pour l'immunolocalisation des transcrits hybridés, les lames sont lavées dans du TTBS (0,3% Triton X100, 0,1 M Tris-HCl pH 7,5, 0,15 M NaCl) pendant 10 min, puis afin de saturer les sites d'interactions non spécifiques entre l'anticorps et les tissus ciblés, les lames sont incubées durant 30 min dans une solution de TTBS-BSA 2%. Des fragments Fab anti-dig couplés à la phosphatase alcaline (Roche) sont dilués au 1/200^{ème} dans une solution de TTBS-BSA 1% et un volume de 100 µL est déposé sur chaque lame. L'incubation a lieu pendant 2 h sous lamelle dans la même chambre humide que précédemment. Les lamelles sont délicatement retirées dans du TTBS, puis les coupes sont rincées par 2 lavages successifs de 15 min dans du TTBS et 2 autres de 10 min dans du tampon AP (0,1 M Tris-HCl, 0,1 M NaCl, 10 mM MgCl₂, pH 9,5). Pour la coloration, les lames sont immergées dans 175µg/ mL de BCIP et 350 µg/mL de NBT dans du Tris-MgCl₂ toute la nuit.

Montage des lames et microscopie

Après coloration, les lames sont lavées dans l'eau, séchées et montées avec de l'huile à immersion entre lame et lamelle. Les observations microscopiques sont faites en champ clair avec un microscope inversé Leica DM IRBE équipé d'un appareil photo numérique Leica DFC 300FX et un logiciel d'acquisition d'images Leica LAS.

2.2.7.3 Etude cyto-histochimique des coupes de paraffine par la lectine PNA

Déparaffinage et déshydratation des coupes

La procédure suivie est la même que celle précédemment décrite pour l'hybridation *in situ*. Des coupes sériées ont été réalisées en parallèle pour hybridation *in situ* et marquage PNA afin de pouvoir comparer les localisations de transcrits et de glycoprotéines.

Blocage des sites protéiques

Afin de bloquer les interactions protéine-protéine non spécifiques sur l'échantillon, les lames sont traitées avec la même solution TTBS utilisée pour hybridation *in situ* additionnée du BSA 1% pendant 30 min à température ambiante.

Incubation de la lectine PNA

Les lames sont rapidement déposées dans les boîtes plastiques ultra-plates contenant de l'eau pour créer une atmosphère humide. Ensuite, 120 µL de solution contenant la lectine PNA couplée à la digoxigénine (*DIG Glycan Differentiation Kit*, Roche), diluée au 1/40^{ème} dans la solution de blocage TTBS-BSA 5% sont déposés sur les lames correspondantes. Chaque lame est recouverte d'une lamelle (25x50 mm) pour que la solution d'incubation s'étale sur la longueur de la lame. La durée d'incubation est de 3 h à température ambiante. Ensuite, les lamelles sont délicatement retirées puis les lames sont rincées dans le tampon TTBS 2 fois 10 min afin d'éliminer l'excès d'anticorps du PNA non couplés et 1 fois dans le tampon TBS-MgCl₂ (0.1 M Tris-HCl, 0.05 M MgCl₂, 0.1 M NaCl, pH 9,5). Les échantillons sont prêts pour une incubation avec l'anticorps secondaire.

Incubation de l'anticorps secondaire

Après les lavages, les lames sont déposées sur les boîtes plastiques ultra-plates et 100 µl de l'anticorps secondaire anti-DIG couplé à la phosphatase alcaline (*Sheep anti-digoxigenin Fab fragments conjugated with alkaline phosphatase*, Roche) diluée au 1/40^{ème} dans le tampon TBS-MgCl₂) puis incubés sous lamelle pendant 1 h dans la chambre humide. Les lames sont ensuite rincées avec de l'eau stérile.

Chapitre 3

CARACTERISATION STRUCTURALE DE AGP31

3. CARACTERISATION STRUCTURALE D'AGP31

3.1 Résumé de l'article : "CHARACTERIZATION OF THE ARABINOGALACTAN PROTEIN 31 (AGP31) OF *ARABIDOPSIS THALIANA*: NEW ADVANCES ON THE HYP-O-GLYCOSYLATION OF THE PRO-RICH DOMAIN".

La diversité structurale des *O*-glycoprotéines est très importante (Spiro, 2002). Chez les végétaux, deux acides aminés, Ser ou Hyp, situés dans des contextes particuliers constituent des sites d'attachement des *O*-glycanes (Faye et al., 2005). Ces sites peuvent être occupés par des *O*-glycanes de différentes structures puisque la nature des sucres (Ara, Gal) et leur enchaînement sont variables (Faye et al., 2005). De plus, une même protéine peut exister sous plusieurs glycoformes. Une des conséquences de cette variabilité structurale pourrait être une diversité de rôles biologiques.

Pour effectuer la caractérisation structurale d'une glycoprotéine, il est nécessaire de connaître sa séquence en acides aminés, de déterminer ses sites de glycosylation et d'établir la structure de ses glycanes (Temporini et al., 2008). Cette caractérisation présente plusieurs types de difficultés du fait de la complexité des glycanes : (i) une glycoprotéine peut posséder plusieurs sites de glycosylation occupés ou non, (ii) chaque site de glycosylation peut porter des glycanes de tailles et de structures différentes, et (iii) les glycosylations peuvent varier de manière qualitative et quantitative en fonction de l'état physiologique des plantes. Ainsi, la caractérisation des glycoprotéines nécessite la mise en place de méthodologies appropriées pour la préparation des protéines et leur analyse et la spectrométrie de masse s'avère un outil particulièrement puissant.

L'analyse de la séquence primaire d'AGP31 fait apparaître la présence de motifs également présents dans des AGP [(X₁₋₃-P)_n, où X=A, S ou T]. En effet, elle comporte un court motif de type AGP à son extrémité N-terminale (APAPA) et un domaine central riche en Pro avec des motifs répétitifs de type XPPX où X= S, K, T, Y, A, V. Il est à noter que les motifs les plus fréquents du domaine central riche en Pro sont de la forme XPPX (Figure 1 de la publication). En référence au code d'hydroxylation des Pro et de *O*-glycosylation des Hyp proposé pour les HRGP (Kieliszewski, 2001), certains résidus Pro du domaine AGP et du domaine riche en Pro pourraient être hydroxylés en Hyp, constituant ainsi des sites potentiels de *O*-glycosylation. Ainsi, le domaine de type AGP situé à l'extrémité N-terminale d'AGP31 est susceptible d'être *O*-glycosylé par des arabinogalactanes (Seifert et Roberts, 2007). Quant au domaine central riche en Pro, il pourrait être *O*-glycosylé par des résidus Gal. En effet, des travaux antérieurs ont montré l'abondance de résidus Gal sur AGP31, ainsi que sur une protéine analogue de *N. alata* (Sommer-Knudsen et al., 1996; Liu et Mehdy, 2007). Cependant, la localisation de ces résidus au niveau du domaine central riche en Pro n'a jamais été démontrée. Le premier objectif de ce travail était de caractériser AGP31 d'un point de vue structural en portant une attention particulière au domaine riche en Pro. Les questions posées étaient les suivantes : où les résidus Hyp sont-ils localisés ? Quels sont les sites *O*-glycosylés et par quels types de glycanes ?

Avant de répondre à ces questions, nous avons dû purifier la protéine AGP31 native à partir des parois des cellules d'hypocotyles étiolés d'*A. thaliana* dans lesquelles elle

s'accumule (Irshad et al. 2008). Deux stratégies différentes ont été mises en œuvre à partir d'un extrait de protéines pariétales préparé suivant Feiz et al. (2006) (Figure 2 de la publication) : (i) une purification en deux étapes comportant une chromatographie d'échange de cations (CEC) (Irshad et al., 2008), suivie par une chromatographie d'affinité sur colonne de nickel (NAC) utilisant la particularité d'AGP31 d'avoir une courte séquence riche en His en aval de son domaine AGP ; (ii) une purification d'affinité en une étape sur une colonne d'agarose couplée à la lectine PNA (*Peanut Agglutinin*) spécifique des résidus Gal, utilisant la remarquable affinité d'AGP31 pour cette lectine (Zhang et al., 2011a). Dans les deux cas, une fraction très enrichie en AGP31 ne comportant qu'un petit nombre de contaminants a été obtenue. La présence de ces contaminants ne nous a pas empêchés de poursuivre la caractérisation d'AGP31 puisque la plupart des analyses par spectrométrie de masse ont été effectuées à partir de fractions prélevées après séparation par SDS-PAGE.

L'observation du profil électrophorétique des deux fractions enrichies montre la présence de larges bandes diffuses correspondant à AGP31 confirmée par spectrométrie de masse MALDI-TOF (Figure 2 de la publication) : deux bandes dans le cas de la purification en deux étapes (CEC + NAC), et trois bandes dans le cas de la purification en une étape (PNA). Ces bandes sont dues à une forte hétérogénéité de masses moléculaires qui pourrait s'expliquer par une hétérogénéité de la *O*-glycosylation d'AGP31 sur un grand nombre de résidus Hyp. La bande de hautes masses moléculaires (> 170 kDa) est révélée par le réactif de Yariv spécifique des arabinogalactanes rencontrés sur les AGP uniquement dans l'échantillon obtenu après la purification en une étape. L'absence de cette bande après la purification en deux étapes signifie que ces glycoformes d'AGP31 ne sont pas retenues sur la colonne d'échange de cations. Notre hypothèse est que la charge de ces glycoformes est masquée par les sucres des motifs arabinogalactanes. La bande de masses moléculaires intermédiaires (50-90 kDa) est présente dans les deux cas et correspond à des formes d'AGP31 portant des chaînes de glycanes de longueurs variables sur le domaine PRP. Cette bande est révélée par *lectin blot* à l'aide de la lectine PNA. Quant à la bande de masses moléculaires plus faibles (30-40 kDa), elle semble correspondre à des formes tronquées de la protéine AGP31 auxquelles manque le domaine PAC C-terminal. Cette première approche a permis de conclure à l'hétérogénéité structurale d'AGP31 se traduisant par la présence de glycoformes différant par le type de *O*-glycosylation et/ou la longueur de la chaîne polypeptidique.

Pour caractériser plus finement la structure du domaine PRP d'AGP31, d'autres approches de spectrométrie de masse adaptées ont été mises en œuvre. Une première approche de spectrométrie de masse MALDI-TOF a permis de montrer la présence de résidus Hyp dans le domaine riche en Pro ainsi que la présence de glycanes comportant des hexoses ou des pentoses. En effets des incréments de masse réguliers de 162 Da ou de 132 Da entre des pics ont pu être visualisés sur les spectres (Figures 3 et 4 de la publication). Le nombre de pentoses n'excédait jamais deux, et le nombre d'hexoses pouvait aller jusqu'à 9. Ces données ont permis de montrer que la glycosylation est hétérogène puisqu'un même peptide existe avec un nombre de sucres variable (Figures 3 et 4, et Figure supplémentaire S1 de la publication). Une analyse en spectrométrie de masse MALDI-TOF/TOF a permis de fragmenter les peptides tryptiques du domaine riche en Pro et de localiser les Hyp partiellement puisqu'il n'est pas possible de fragmenter les doublets de Pro (Tableau 1, et Figure supplémentaire S2 de la publication). Les deux résidus Pro des motifs SPPA et SPPT sont toujours hydroxylés. Seul un des deux résidus Pro des motifs KPPT, KPPV, TPPV et YPPT est hydroxylé. Le résidu Pro des motifs KAPV et KSPV est toujours hydroxylé tandis que les résidus Pro du motif YPPK ne sont pas hydroxylés (Figure 5). Une localisation plus précise de résidus Hyp a pu être déduite des règles proposées pour les codes d'hydroxylation

des HRGP (Kieliszewski et al., 1990; Kieliszewski, 2001). Les résultats nouveaux apportés par nos travaux permettent d'envisager une généralisation pour ces codes. Enfin, une analyse en spectrométrie de masse de type ETD a permis de fragmenter ces mêmes peptides tryptiques en préservant la partie glycanique. Il a ainsi été possible de montrer que des hexoses sont présents sur les Hyp du domaine riche en Pro et qu'ils sont distribués de manière uniforme. Lorsque le nombre d'hexoses est inférieur au nombre de résidus Hyp, il n'est jamais trouvé deux hexoses associés à un Hyp.

L'ensemble de ces résultats a permis de montrer que la glycoprotéine AGP31 présente une complexité remarquable avec la présence de plusieurs glycoformes conduisant à la proposition de modèles structuraux (Figure 6 de la publication). Le recours à la spectrométrie de masse, en particulier l'approche novatrice de type ETD utilisée pour la première fois sur des glycoprotéines végétales, a permis d'apporter des éléments nouveaux sur la structure d'AGP31. Nos résultats, en particulier la spécificité remarquable entre AGP31 et PNA et la présence d'hexoses et de pentoses au niveau du domaine riche en Pro, sont en accord avec la composition en monosaccharides (majoritairement Gal et Ara) préalablement proposée par Liu et Mehdy (2007).

Les glycoformes de hautes masses moléculaires d'AGP31 peuvent être révélées par le réactif de Yariv spécifique des AGP, ce qui signifie que le domaine AGP d'AGP31 porte des arabinogalactanes. Les autres glycoformes d'AGP31 sont dues à la présence de *O*-glycanes au niveau du domaine riche en Pro. Ces motifs, spécifiquement reconnus par la lectine PNA, n'avaient jamais été décrits jusqu'alors et devront, dans l'avenir, faire l'objet d'une caractérisation structurale approfondie.

3.2 Article : “Characterization of the ArabinoGalactan Protein 31 (AGP31) of *Arabidopsis thaliana*: new advances on the Hyp-*O*-glycosylation of the Pro-rich domain”.

May Hijazi, Jessica Durand, Carole Pichereaux, Frédéric Pont, Elisabeth Jamet, Cécile Albenne.

Soumis à *Journal of Biological Chemistry*.

**Characterization of the ArabinoGalactan Protein 31 (AGP31) of *Arabidopsis thaliana*:
new advances on the Hyp-*O*-glycosylation of the Pro-rich domain**

**May Hijazi^{1,2}, Jessica Durand^{1,2}, Carole Pichereaux³, Frédéric Pont⁴, Elisabeth Jamet^{1,2},
Cécile Albenne^{1,2,*}**

From ¹Université de Toulouse; UPS; UMR 5546, Laboratoire de Recherche en Sciences Végétales; BP 42617, F-31326 Castanet-Tolosan, France and ²CNRS; UMR 5546; BP 42617, F-31326 Castanet-Tolosan, France and ³FR3450, Plateforme de Protéomique Toulouse Midi-Pyrénées, Institut de Pharmacologie et Biologie Structurale, Université de Toulouse, F-31077 Toulouse, France, and ⁴INSERM, Plateau technique Interactions et Profils d'expression des Protéines, Plateforme de Protéomique Toulouse Midi-Pyrénées, F-31300 Toulouse, France.

Running head: AGP31 Hyp location and *O*-glycosylation

*Address correspondence to: Cécile Albenne, UPS; CNRS; UMR 5546, Laboratoire de Recherche en Sciences Végétales; BP 42617, F-31326 Castanet-Tolosan, France. Tel: +33(0)534 32 38 27 ; Fax: +33(0)534 32 38 02; E-mail: albenne@lrsv.ups-tlse.fr

Capsule

1. Background: AGP31 is a multi-domain plant cell wall hydroxyproline-rich glycoprotein. The position of Hyp and the distribution of carbohydrates is unknown.
2. Results: Most Hyp of the Pro-rich domain are isolated within repeated motifs and carry glycans of various sizes.
3. Conclusion: AGP31 glycoforms are very heterogeneous.
4. Significance: This might be the first evidence for new Hyp-*O*-glycans.

Proteins are important actors in plant cell walls since they contribute to their architecture and their dynamics. Among them, hydroxyproline (Hyp)-rich glycoproteins (HRGPs) constitute a complex family of *O*-glycoproteins with various structures and functions. In this study, we characterized an atypical HRGP, the ArabinoGalactan Protein 31 (AGP31) which displays a multi-domain organization unique in *Arabidopsis thaliana*, consisting of a short AGP motif, a His stretch, a Pro-rich domain, and a C-terminal PAC (PRP-AGP containing Cys) domain. The use of various mass spectrometry strategies was innovative and powerful: it permitted to locate Hyp residues, to demonstrate the presence of carbohydrates, and to refine their distribution over the Pro-rich domain. Most Hyp were isolated within repeated motifs such as KAOV, KSOV, K(PO/OP)T, K(PO/OP)V, T(PO/OP)V and Y(PO/OP)T. A few extensin-like motifs with contiguous Hyp (SOOA and SOOT) were also found. The Pro-rich domain was shown to carry Gal residues

on isolated Hyp but also Ara residues. The existence of new type Hyp-*O*-Gal/Ara-rich motifs not recognized by the β -glucosyl Yariv reagent, but interacting with the PeaNut Agglutinin (PNA) lectin was proposed. Besides, the N-terminal short AGP motif was assumed to be substituted by arabinogalactans (AGs). Altogether, AGP31 was found to be highly heterogeneous in cell walls since AGs could be absent, Hyp-*O*-Gal/Ara-rich motifs of different sizes were observed and truncated forms missing the C-terminal PAC domain were found, suggesting degradation *in muro* and/or partial glycosylation prior to secretion.

Hydroxyproline (Hyp)-rich glycoproteins (HRGPs)^a are a superfamily of plant cell wall proteins (CWPs) identified several decades ago (1). Their study remains challenging since HRGPs constitute a very complex protein group with various structures and functions whose molecular bases are not fully elucidated (2-6). HRGPs fall into three families: (i) hyperglycosylated arabinogalactan proteins

(AGPs) having diverse functions (5); (ii) moderately glycosylated extensins (EXTs) that can form cell wall scaffolds essential for cytokinesis (7); and (iii) Pro-rich proteins (PRPs) that may be non-, weakly- or highly-glycosylated, and possibly involved in various aspects of development (4,8). Hybrid HRGPs are composed of different types of HRGP domains, and chimeric HRGPs contain at least one HRGP domain within a non-HRGP protein. A recent bioinformatics analysis achieved on the *Arabidopsis thaliana* genome identified 166 HRGPs classified in 85 AGPs, 59 EXTs, 18 PRPs, and 4 AGP/EXT hybrid proteins, highlighting the question why plants need so many HRGPs (9).

HRGPs are characterized by the presence of Hyp-rich repetitive motifs, called glycomodules, containing from 3 to 16 amino acids. Glycomodule sequences determine Pro hydroxylation and the subsequent Hyp *O*-glycosylation. These two consecutive sequence-driven post-translational modifications (PTMs) define the so-called Hyp-*O*-glycosylation code (3). AGPs have repetitive variants of (Xaa₁₋₃-Hyp)_n motifs (where Xaa is commonly Ser, Ala or Thr) with *O*-linked arabinogalactan (AG) polysaccharides. EXTs display clusters of Ser-(Hyp)_{2,4} containing short *O*-Hyp arabinose (Ara)-oligosaccharides. Finally, PRPs are probably the most heterogeneous group, regarding their repetitive sequences and their carbohydrate patterns. They are characterized by the repeating occurrence of (Pro)₂ or (Pro)₃ motifs within a variety of larger repeated units. It was shown that PRPs contain approximately equimolar quantities of Pro and Hyp (8,10-13). Many PRPs contain the repeated pentapeptide Pro-Hyp-Xaa-Yaa-Lys, where Xaa and Yaa can be Val, Tyr, His, or Glu (4,14). PRPs were initially assumed to be non- or weakly-glycosylated (11,15). However, the characterization of a Gal (galactose)-rich glycoprotein (GaRSGP) from *N. alata* styles revealed a new class of PRPs with repetitive motifs never described before and a carbohydrate content of 75%, unusual for PRPs (8). Recently, an *A. thaliana* homologue of GaRSGP has been described (12). This glycoprotein mostly contains Gal residues and was called AGP31 because of its positive interaction with the β-glucosyl Yariv reagent. This protein, encoded by *At1g28290*, was shown to be a multi-domain protein and was classified as a chimeric AGP (9).

Here, we provide a detailed characterization of AGP31 extracted from cell walls of *A. thaliana* etiolated hypocotyls. Combining two isolation processes and three detection methods, we demonstrated that the native AGP31 displays a huge heterogeneity with various *O*-glycans and truncated forms. After enrichment, mass spectrometry (MS) analyses permitted to determine the location of Hyp residues within the Pro-rich domain, to demonstrate the presence of carbohydrates, and to refine their distribution.

EXPERIMENTAL PROCEDURES

Plant material and cell wall isolation—*A. thaliana* seeds (ecotype Columbia 0) were grown in the dark as described (16). Etiolated hypocotyls were collected at 11 days and used to isolate cell walls as reported (17).

AGP31 isolation—CWPs were extracted from the cell wall fraction by successive steps of washing with a 0.2 M CaCl₂ solution followed by a 2 M LiCl solution as described (16). About 2 mg of CWPs were usually extracted from 1 g of dry cell walls. Two strategies were employed to isolate AGP31 from a total CWP extract: (i) a combination of cation exchange chromatography (CEC) and Nickel affinity chromatography (NAC); (ii) an affinity chromatography using a PeaNut Agglutinin lectin from *Arachis hypogaea* (PNA)-agarose resin. PNA is a lectin specific for GalNAc, α-Gal and β-Gal (18). For the first isolation strategy, 6 mg of CWPs were separated by CEC as previously described (16). AGP31 was identified by peptide mass fingerprinting (PMF) from Matrix-Assisted Laser Desorption Ionization – Time Of Flight Mass Spectrometry (MALDI-TOF MS) analysis of CEC elution fraction aliquots separated by SDS-PAGE. The elution fractions containing AGP31 were pooled, desalted, lyophilized and subsequently loaded onto a Ni-NTA column supplied in the Ni-NTA Fast Start Kit (Qiagen, Courtaboeuf, France), following the manufacturer's recommendations. For the second isolation strategy, 2 mg of CWPs were incubated with 1 mL of PNA-agarose resin (Sigma, St Louis, MO). Chromatography was performed as previously reported (19). Protein content was estimated using the Bradford method (20).

SDS-PAGE and lectin blot analyses—AGP31-enriched fractions issued from the two isolation processes (CEC/NAC and PNA) were desalted using Econo-Pac 10 DG columns (Bio-Rad, Hercules, CA) and lyophilized prior to SDS-

PAGE. Dried samples were dissolved in 150 μ L of UHQ water. Fifty μ L of each sample were loaded on a 10 x 12 x 0.15 cm SDS-polyacrylamide gel with a concentration of 12.5 % /0.33% of acrylamide/bisacrylamide. Separation was performed according to Laemmli (21). Staining of gels with colloidal blue was carried out using PageBlue™ Protein Staining Solution (Fermentas, Saint-Rémy-lès-Chevreuse, France) or the β -glucosyl Yariv reagent (22). For lectin blots, proteins separated by SDS-PAGE were transferred onto a nitrocellulose membrane (GE Healthcare, Orsay, France) using a semi-dry transfer cell (Trans-Blot SD, Bio-Rad). Transfer was carried out at 4°C in transfer buffer (Tris-HCl 48 mM, glycine 39 mM, SDS 0.0375%, methanol 20%) for 1 h at 20 V. AGP31 bound to the membrane was revealed by a lectin blot method, using the digoxigenin-labeled PNA provided in the DIG Glycan Differentiation Kit, following the manufacturer's recommendations (Roche, Mannheim, Germany).

Monosaccharide composition analyses—AGP31-enriched fractions issued from the two isolation processes (CEC/NAC and PNA) were desalted using Econo-Pac 10 DG columns (Bio-Rad, Hercules, CA) and lyophilized prior to acid hydrolysis carried out for 1 h at 110°C using trifluoroacetic acid 2N. Trimethylsilylated (TMS) derivatives were prepared and analyzed by Gas Chromatography-Mass Spectrometry (GC-MS) as described (23).

Anhydrous Hydrogen Fluoride (HF) deglycosylation—A sample of total CWP extracted from 11-day-old etiolated hypocotyls was HF-deglycosylated for 1 h at 4°C as described (24,25). The HF was blown off under nitrogen gas, and the deglycosylated proteins were then separated by SDS-PAGE.

MS analyses—Sample preparation for all MS analyses was performed as previously described (26). MALDI-TOF MS analyses were performed using a Voyager-DE STR mass spectrometer (Applied Biosystems/MDS, Sciex, USA). Spectra were acquired in reflectron mode as previously reported (26,27). MALDI-TOF/TOF MS/MS analyses were performed using a MALDI-TOF/TOF Voyager 4700 (Applied Biosystems/MDS, Sciex, USA). MS/MS data were recorded using the following parameters: accelerating voltage 8 and 15 kV (sources 1 and 2, respectively) and grid voltage 86%. The mass selection of the precursor ion was achieved using a mass window of ± 5 Da

and collision was performed in CID off mode. Data were acquired with 3750 shoots/spectrum.

For Liquid Chromatography (LC)-MS/MS analyses, chip-based nano-LC separation was performed with an Agilent 1260 nano-pump system consisting of binary capillary-flow and nano-flow pumps, a vacuum degasser, and an autosampler (Agilent Technologies Inc., Palo Alto, CA). Chromatographic separation of tryptic peptides was performed on a C18 chip (75 mm diameter, 43 mm length, 40 nL preconcentration trap, 5 μ m particle size). The chip column was equilibrated at room temperature with 97% eluent A (0.1% formic acid in water) and 3% eluent B (0.1% formic acid in 90% acetonitrile and 10% water) at a flow rate of 600 nL/min. The chip preconcentration trap was equilibrated with 0.1% formic acid in 97 % water and 3% acetonitrile at a flow rate of 4 μ L/min. After injection of the sample, separation was carried out using a linear gradient from 3–55% mobile phase B for 20 min, 55–95% mobile phase B for 1 min and 95% mobile phase B for 2.5 min. The LC system was interfaced to a Bruker Daltonics (Bruker Daltonics GmbH, Bremen, Germany) Amazon ion trap mass spectrometer via a ChipCube interface (Agilent Technologies Inc.). The spraying voltage was set to 1780 V. MS scans were performed in the 'enhanced resolution' mode and MS/MS scans were performed in 'ultrascan' mode covering the 100–3000 m/z range for the analyses of all glycopeptides. For Electron-Transfer Dissociation (ETD) analyses, a negative chemical ionization source generated fluoranthene radical anions, which were transported to the ion trap and used as the electron-transfer agents. To produce the radical anions, the reactant temperature was set at 60°C, while the ionization energy and emission current were set at 70 eV and 4 μ A, respectively. The reaction time for radical anions was set to 100 ms.

PMF was done by comparison to the non-redundant database of *A. thaliana* at NCBI (www.ncbi.nlm.nih.gov/) using ProteinProspector MS-FIT (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard>). The criteria used for the database search were: a monoisotopic mass accuracy of 20 ppm, one missed cleavage, and hydroxylation of Pro residues. MS/MS results were analyzed using Protein Prospector MS-Product

(<http://prospector.ucsf.edu/prospector/cgibin/msform.cgi?form=msproduct>) with Pro hydroxylation and glycosylation added as possible modifications.

RESULTS

Different strategies to isolate AGP31: evidence for a huge heterogeneity—In our previous cell wall proteomics analyses of *A. thaliana* etiolated hypocotyls, we identified AGP31 as an abundant protein (16,19). This protein comprises several domains, from N- to C-terminal (Fig. 1A): a predicted signal peptide, a short AGP motif, a His stretch, a Pro-rich domain, and a C-terminal Cys-rich domain also called PAC (PRP-AGP containing Cys) domain (28).

Following a fractionation step using CEC, AGP31 was eluted at the end of the salt gradient, at about 0.5 M NaCl, as determined by MALDI-TOF MS analyses of elution fraction aliquots separated by SDS-PAGE (Fig. 1B). The fractions containing AGP31 were collected for subsequent NAC. The combination of these two chromatographic steps was very efficient for isolating AGP31 which migrated as a smear from 50 to 90 kDa on a SDS-polyacrylamide gel (Fig. 2, lane 1a). This is much higher than the predicted molecular mass of AGP31, *i.e.* 38 kDa. Besides, an SDS-PAGE analysis of a similar fraction containing three times more proteins than previously (*i.e.* 36 µg) showed an additional smear from 30 to 40 kDa (Fig. 2, lane 1b). MALDI-TOF MS analyses of samples taken all along both smears (30-40 kDa and above 50 kDa) identified AGP31 as unique protein. Up to seven tryptic peptides were used for AGP31 identification by PMF (corresponding to a maximum coverage of 33.3%), among which four belonged to the C-terminal PAC domain and three to the Pro-rich domain (named P1, P2 and P4) (Figs. 1A, 1B and 3A). Truncated forms of AGP31 missing the C-terminal PAC domain were found in the 30-40 kDa smear.

Two minor contaminant proteins, a thaumatin encoded by *At2g28790* and a glycine-rich protein encoded by *At2g05580*, were also found in this fraction, in bands of 26 and 20 kDa, respectively (labeled with stars on Fig. 2, lane 1b). These co-purified proteins may be AGP31 interactants or have similar behavior towards both chromatographic resins used. In particular, *At2g28790* possesses a His stretch that may be

responsible for its retention on the Nickel resin. Altogether, an AGP31-enriched fraction containing about 40 µg of proteins was routinely obtained from 6 mg of CWPs after CEC and NAC.

A PNA-lectin blot analysis of this enriched fraction showed a signal from 34 to 170 kDa, particularly strong above 50 kDa (Fig. 2, lane 2). This result corroborated the fact that AGP31 contains Gal residues, as previously reported (12) and showed that AGP31 displays a heterogeneous pattern of glycosylation. Monosaccharide composition analysis revealed that AGP31-enriched fraction mostly contains Gal (53.2%) but also Ara (39.5%) in a higher proportion than previously reported (12). Lesser amounts of Xyl (2.2%), Fuc (1.9%), Glc (1.8%), Man (1.3%) and GlcUA (0.3%) were also detected.

Finally, no signal was obtained with the β -glucosyl Yariv reagent, specific for AGPs, suggesting that AGP31 isolated by CEC and NAC does not display any AG (Fig. 2, lane 3). Interestingly, the analysis of the flow-through fraction of the CEC showed a positive response with the β -glucosyl Yariv reagent, demonstrating that AGPs did not bind to that resin, probably because their carbohydrate content hide their charge (Fig. 2, lane 4).

An alternative isolation strategy was carried out using a PNA-lectin agarose matrix, capitalizing on the remarkable affinity of AGP31 towards this lectin (19). About 40 µg of proteins were routinely obtained in the PNA elution fraction from 2 mg of CWPs. AGP31 was identified by PMF using MALDI-TOF MS, with the same tryptic peptides as reported above (Fig. 1B). It migrates as two smears, respectively from 30 to 40 kDa and above 50 kDa where AGP31 was the only protein identified (Fig. 2, lane 5). As expected, it was recognized by PNA and the lectin blot signal was similar to that obtained using the CEC/NAC isolation strategy, confirming the large heterogeneity of AGP31 (Fig. 2, lane 6). The presence of residual Gal from the elution buffer (containing 0.5 M Gal) in the desalted AGP31-enriched fraction did not permit to determine a reliable monosaccharide composition of this fraction. Note that the *At5g11420* and *At5g25460* proteins containing a DUF642 domain of unknown function were found in bands of 43 and 40 kDa respectively (labeled with stars on Fig. 2, lane 5). Finally, the β -glucosyl Yariv reagent gave a signal at the top

of the SDS-polyacrylamide gel for the PNA elution fraction (Fig. 2, lane 7). This signal was assumed to be due to AGP31. Note that no other protein has been identified by PMF above 50 kDa in this fraction and that the flow-through fraction of the PNA chromatography showed a positive response with the β -glucosyl Yariv reagent showing that other AGPs were not retained (data not shown).

The Pro-rich domain of AGP31: a domain rich in O-glycosylated Hyp-As mentioned above, among the seven tryptic peptides used for AGP31 identification, three belonged to the Pro-rich domain (named P1, P2 and P4) (Figs. 1A, 1B and 3A). P1, P2 and P4 were predicted to contain 6, 5 and 9 hydroxyl groups, respectively. We assumed that these modifications correspond to hydroxylation of Pro residues, giving rise to Hyp. Additional peaks were observed on MS spectra of samples taken from 50 to 60 kDa, and from 30 to 40 kDa (Fig. 2, lanes 1a, 1b and 5). The masses of these peaks matched those of P1, P2 and P4 with regular mass increments of 162 Da and/or 132 Da, corresponding to the masses of a hexose and a pentose, respectively (Figs. 3A, 4A and 4B). Since these peptides were predicted to contain Hyp residues, we assumed that they carried Hyp-*O*-glycosylations. It should be noted that mass increments of 132 Da corresponding to pentoses were only observed on P1 and P4.

Samples taken along the SDS-polyacrylamide gel lanes, in the 50 - 60 kDa and 30 - 40 kDa ranges, displayed various patterns of glycosylation (Fig. 2, lanes 1a, 1b and 5). First, for a given sample, different glycoforms of each tryptic peptide (P1, P2 and P4) were observed, showing that *O*-glycosylation of the Pro-rich domain is heterogeneous (Figs. 4A and 4B). Second, *O*-glycopeptides are usually larger in the upper part of these mass ranges than in their lower part, consistent with the fact that the smears are due to different degrees of Pro-rich domain *O*-glycosylation (supplemental Figs. S1A and S1B). Note that P1, P2 and P4 were also found as non-glycosylated forms in bands taken at 30 and 50 kDa. Altogether, one to five hexoses were detected by MALDI-TOF MS onto P1 and P2, and one to seven onto P4 (Table 1). One or two pentoses were detected on glycoforms of P1 and P4 carrying various number of hexoses (Figs. 4A and 4B). It should be noted that no glycopeptides were observed on MS spectra of samples taken above 60 kDa where AGP31 was also found, even when

increasing the *m/z* limit, suggesting an *O*-glycosylation pattern not observable by MALDI-TOF MS.

An HF treatment was performed on a total CWP extract in order to remove glycans. After separation on SDS-polyacrylamide gel, bands containing deglycosylated AGP31 were excised, digested with trypsin, and analyzed by MALDI-TOF MS. Spectra obtained revealed that the peaks corresponding to AGP31 glycopeptides from untreated sample (Fig. 3A) disappeared in favor of those corresponding to unmodified peptides, confirming that the mass increments of 162 and 132 Da were due to glycosylations (Fig. 3B). In addition to the increase of intensity of peaks corresponding to P1, P2 and P4, this treatment revealed additional peaks corresponding to peptides of the Pro-rich domain, named P0 and P3 (Fig. 1A), which were not observed in untreated sample. The intensity of P0 was probably too low for detection in MALDI-TOF MS analysis of glycosylated AGP31. The *m/z* of P3 (2486.35) was close to that of P1 with two hexoses (2487.26), making difficult the identification of this peptide and the corresponding *O*-glycoforms in MS analyses of untreated sample. These additional MS data provided a full coverage of the Pro-rich domain. The information concerning the peptides of the Pro-rich region, namely P0 to P4, is summarized in Table 1. Considering the whole sequence of the Pro-rich domain, a total of 42 Hyp and 29 Pro residues were found.

In depth MS/MS analyses of the Pro-rich domain: location of Hyp and hexose residues—Our MALDI-TOF MS data provided a precise determination of Hyp number within the Pro-rich domain. However, the question of the arrangement of the Hyp and Pro residues (respectively 42 and 29) as well as that of the distribution of carbohydrates was not yet addressed.

MALDI-TOF/TOF MS/MS experiments were carried out on deglycosylated AGP31 tryptic peptides obtained following gel excision of the HF sample (Fig. 3B). Fragmentation was successfully performed for P1, P2, P3 and P4 (supplemental Figs. S2A-S2D). Note that it was not possible to perform fragmentation of P0 because of insufficient ion intensity. Analysis of *b* and *y* fragment ion masses allowed determining the number of hydroxylations, *i.e.* the number of Hyp residues on each fragment. These results permitted to locate precisely some Hyp. However, it was not possible to determine

the exact position of Hyp in some Pro-Pro motifs since fragmentation between two Pro (or two Hyp or Pro-Hyp) does not occur. Then, we provided new features about repetitive motifs of the AGP31 Pro-rich domain (Table 2): (i) both Pro were hydroxylated in SPPA and SPPT motifs; (ii) Pro was invariably hydroxylated in KAPV and KSPV motifs; (iii) only one Pro was hydroxylated in KPPT, KPPV, TPPV and YPPT motifs; (iv) Pro was not hydroxylated in the YPPK motif. Note that the KPPV motif is the most widespread in the Pro-rich domain.

Additional LC-MS/MS experiments were performed on *O*-glycopeptides using ETD, a fragmentation mode that keeps PTMs. They were successful on P2 carrying two, three, four or five hexoses, which generated the most intense parent ions (Fig. 4A). Fragmentation patterns displaying c and z fragments were obtained, permitting to determine which Hyp residues carried hexoses (supplemental Figs. S3A-S3E). It was shown from the fragmentation pattern of P2 carrying five hexoses that each of the five Hyp of this peptide was occupied by a hexose, revealing a uniform distribution of hexoses on isolated Hyp (supplemental Fig. S3D). Interestingly, none of the Hyp was shown to carry more than one hexose, demonstrating that the P2 species carrying five hexoses was homogeneous. The fragmentation pattern of P2 carrying two, three or four hexoses indicated heterogeneous distribution of hexoses onto Hyp (supplemental Figs. S3A-S3C). For instance, the fifth Hyp of P2 containing four hexoses was found under glycosylated and non-glycosylated forms (supplemental Fig. S3C). Here again, fragments with several hexoses on a Hyp were never observed.

DISCUSSION

AGP31 is a remarkable cell wall protein with a multi-domain organization unique in *A. thaliana*. Although it was already reported that AGP31 was a glycoprotein, no data about the distribution of Hyp and *O*-glycans on the protein were available so far (12). The use of diverse MS technologies was innovative and provided an in depth characterization of the AGP31 Pro-rich domain. Combining various experimental approaches, we demonstrated that native AGP31 displays a huge heterogeneity in cell walls, with various *O*-glycans and truncated forms of its protein core.

Our results provided for the first time a full MS coverage of the Pro-rich region of a native HRGP, whereas most available data were obtained by Edman sequencing of small peptides (reviewed in (2,4,29) and on recombinant HRGPs (30). Combining literature data with our fragmentation results on tryptic peptides from the Pro-rich domain, we could locate all the Hyp residues in the Pro-rich domain of AGP31. Indeed, the so-called Hyp-*O*-glycosylation code based on protein sequence data assumes that Lys, Phe and Tyr prevent hydroxylation of the following Pro (3). Besides, TPK motifs were found in the maize THRGP (31). Then, we propose that KPOT, KPOV, TPOV and YPOT motifs are present in AGP31 (Fig. 5). The sequences of other repetitive motifs determined experimentally, *i.e.* SOOA, SOOT, KSOV, and KAOV are consistent with the Hyp-*O*-glycosylation code. It should be noted that SOOA/T motifs are similar to those found in EXTs (3,24). The only exception to the Hyp-*O*-glycosylation code is the YPPK motif in which none of the Pro is hydroxylated (Fig. 5). The amino acids preceding Pro are assumed to provide a specific local conformation permitting or preventing the action of prolyl 4-hydroxylases involved in Pro hydroxylation. Enzymatic characterization of *A. thaliana* prolyl 4-hydroxylases provided insight into their substrate specificity (32,33). Recently, the crystal structure of an algal prolyl 4-hydroxylase complexed with a Ser-Pro₅ substrate highlighted the molecular bases of Pro hydroxylation (34). The mechanisms of Pro hydroxylation within Pro-rich domains of PRPs will require further investigation to check whether it follows the same rules as in EXTs and AGPs. As suggested, this code may also depend on the organ and/or plant (35,36).

Our experimental data gave new insight into the *O*-glycosylation of the Pro-rich domain of AGP31. MALDI-TOF MS analyses showed the presence of hexoses and pentoses, assumed to be Gal and Ara according to the PNA-lectin blot detection and the monosaccharide composition of AGP31. The ETD fragmentation technology was employed to describe the distribution of carbohydrates onto the Pro-rich domain, providing the first successful study of a plant protein *O*-glycosylation using this method. ETD MS/MS experiments performed onto P2 *O*-glycopeptides containing hexoses showed that Gal are uniformly distributed on isolated Hyp within the KAOV, KPOT, KPOV, and YPOT

motifs. Since such experiments were not possible on P1 and P4 *O*-glycopeptides containing pentoses, we could not locate Ara on these peptides. However, we suggest that Ara or short Ara-oligosaccharides may be carried by contiguous Hyp in SOOA/T motifs, like in extensins (3,24). Besides, we cannot exclude that Ser residues within these extensin-like motifs could also be *O*-galactosylated (3,24).

An important feature highlighted in this study is the huge heterogeneity of AGP31. After separation by SDS-PAGE and various detection methods, AGP31 - whose predicted molecular mass is 38 kDa - was found as a smear from 30 to about 250 kDa. Combining all our results, we could propose structural models schematized on Fig. 6. The Pro-rich domain on which carbohydrates were detected by MALDI-TOF MS is probably also substituted by larger *O*-glycans which escaped MS analyses because of the large size of the corresponding tryptic glycopeptides. According to the monosaccharide composition of AGP31-enriched fraction obtained using CEC and NAC (53.2% Gal and 39.5% Ara), we propose to call them Hyp-*O*-Gal/Ara-rich motifs. Note that our monosaccharide analysis showed a higher proportion of Ara than previously reported, probably because our expression and purification procedures were different (12). These Pro-rich domain *O*-glycans were specifically recognized by PNA as a smear from 34 to 170 kDa, but not by the β -glucosyl Yariv reagent, suggesting a structure different from that of type II arabino-3,6-galactans predominantly found in AGPs (37). The higher molecular mass glycoforms of AGP31 (Fig. 6, smear above 170 kDa) are also assumed to carry AG motifs on the short N-terminal AGP sequence (APAPAPY) as suggested by the positive signal with the β -glucosyl Yariv reagent. The absence of a signal above 170 kDa by PNA-lectin blot analysis was probably due to electrophoretic transfer limitations. Finally, truncated forms of AGP31 missing the C-terminal PAC domain also exist (Fig. 6, smear of 30-40 kDa after SDS-PAGE). These latter forms were found to be weakly *O*-glycosylated on their Pro-rich domain.

The heterogeneity of AGP31 in cell walls raises the question of its origin. *O*-glycans might be processed by glycoside hydrolases. Galactosidases and arabinosidases were found in *A. thaliana* cell wall proteomes (38,39) and several studies provided evidence for enzymatic digestion of AGP *O*-glycans (40-43). Their

degradation has been assumed to produce signal molecules involved in plant development (6,37). The turnover of AGP *O*-glycans was also described as part of a salvage pathway allowing the recycling of sugars for the synthesis of new polymers (44). This turnover of *O*-glycans could explain the observed low molecular mass glycoforms of AGP31. It should be noted that truncated forms of AGP31 missing the C-terminal PAC domain are (i) not glycosylated on the AGP motif, and (ii) weakly *O*-glycosylated on their Pro-rich domain. We suggest that the PAC domain may be degraded by proteases present in *A. thaliana* cell wall proteomes (45) and that *O*-glycosylation protects AGP31 from proteolysis, as assumed for AGPs (4). Alternatively, the heterogeneity of the Pro-rich domain *O*-glycosylation may reflect a partial *O*-glycosylation of AGP31 along the secretory pathway. Then, our ETD MS/MS data suggest that isolated Hyp of the Pro-rich domain may be first galactosylated before the subsequent elongation step. Galactosylation would require two different galactosyltransferases (GalTs) as recently reported for AGPs in *A. thaliana* and *Nicotiana tabacum* (46,47). Two distinct GalT activities were identified from *in vitro* assays: a Hyp:GalT activity catalyzing the addition of Gal onto peptidyl-Hyp residues, and a Gal:GalT activity extending the sugar chain. Twenty putative β -(1,3)-GalTs were predicted by bioinformatics in *A. thaliana* (48). It will be necessary to characterize the substrate specificity of each of them to elucidate the mechanisms of *O*-galactosylation of AGPs and PRPs.

Altogether, AGP31 is a glycoprotein far more complex than previously assumed. The next challenge will be to perform the structural characterization of the new Hyp-*O*-Gal/Ara-rich motifs of its Pro-rich domain. Such work is mostly performed using biochemical approaches like methylation analysis and nuclear magnetic resonance (NMR) spectroscopy and requires large amounts of pure protein. In this way, several models were proposed for type II AGs isolated from recombinant AGPs (30). Interestingly, Tryfona *et al.* carried out an alternative approach to elucidate the structure of native wheat flour AGP, combining AG enzymatic digestion and product analysis using Polysaccharide Analysis by Carbohydrate gel Electrophoresis (PACE) and MS (49). Note that type III AGs have also been described in Art v 1 and Amb a allergens which have a Pro-rich domain (50,51).

Finally, it will be important to elucidate the function of each of the AGP31 domains and the role of the various glycans. As a first clue, this work highlighted the remarkable affinity between AGP31 and PNA. Hyp-*O*-Gal/Ara-rich motifs of the AGP31 Pro-rich domain may interact with cell wall lectins *in muro*. Legume

lectins, which are the closest homologues of PNA in *A. thaliana*, were found in *A. thaliana* cell wall proteomes (16,39). This trail may be interesting to explore the structure/function relationships of AGP31 and other proteins with *O*-glycosylated Pro-rich domains.

Acknowledgements. This study was funded by the *Université Paul Sabatier* in Toulouse (France), the *Centre National de la Recherche Scientifique* (France), and the Lebanon ecological association (grant to MH). This work has been done in the LRSV, part of the "*Laboratoire d'Excellence*" (LABEX) entitled TULIP (ANR -10-LABX-41). MS analyses were performed on the *Plateforme de Protéomique Toulouse Midi-Pyrénées* (France), except GC-MS analyses achieved in the LRSV. The authors thank Dr S Fournier and Dr A Lemassu for their advices on TMS derivative analyses, Dr G Boudart for HF deglycosylation experiments, late Pr R Pont-Lezica for critical reading of the manuscript and Dr H Canut for stimulating discussions. Their gratitude also goes to Pr M Kielieszewski (University of Ohio) for welcoming Dr G Boudart in her laboratory.

FOOTNOTES

^aThe abbreviations used are: AG: arabinogalactan; AGP: arabinogalactan protein; CEC: cation exchange chromatography; CWP: cell wall protein; ETD: electron-transfer dissociation; EXT: extensin; GalT: galactosyl transferase; GC: gas chromatography; HF: hydrogen fluoride; HRGP: hydroxyproline-rich glycoprotein; Hyp: hydroxyproline; LC: liquid chromatography; MALDI-TOF: matrix-assisted laser desorption ionization – time of flight; MS: mass spectrometry; NAC: nickel affinity chromatography; NMR: nuclear magnetic resonance; PAC: PRP-AGP containing Cys; PACE: polysaccharide analysis by carbohydrate gel electrophoresis; PMF: peptide mass fingerprinting; PNA: peanut agglutinin lectin; PRP: Pro-rich protein; PTM: post-translational modification; TMS: trimethylsilyl.

Les données supplémentaires relatives à cet article sont placées dans les annexes, à la fin de ce mémoire.

REFERENCES

1. Lamport, D., and Northcote, D. (1960) *Nature* **188**, 665-666
2. Kielieszewski, M., and Lamport, D. (1994) *Plant J* **5**, 157-172
3. Kielieszewski, M. (2001) *Phytochemistry* **57**, 319-323
4. Showalter, A. (1993) *Plant Cell* **5**, 9-23
5. Seifert, G., and Roberts, K. (2007) *Annu. Rev. Plant Biol.* **58**, 137-161
6. Nothnagel, E. (1997) *Int Rev Cytol* **174**, 195-291
7. Cannon, M., Terneus, K., Hall, Q., Tan, L., Wang, Y., Wegenhart, B., Chen, L., Lamport, D., Chen, Y., and Kielieszewski, M. (2008) *Proc Natl Acad Sci USA* **105**, 2226-2231
8. Sommer-Knudsen, J., Clarke, A. E., and Bacic, A. (1996) *Plant J.* **9**, 71-83
9. Showalter, A., Keppler, B., Lichtenberg, J., Gu, D., and Welch, L. (2010) *Plant Physiol* **153**, 485-513

10. Averyhart-Fullard, V., Datta, K., and Marcus, A. (1988) *Proc Natl Acad Sci USA* **85**, 1082-1085
11. Datta, K., Schmidt, A., and Marcus, A. (1989) *Plant Cell* **1**, 945-952
12. Liu, C., and Mehdy, M. (2007) *Plant Physiol.* **145**, 863-874
13. Kleis-San Francisco, S., and Tierney, M. (1990) *Plant Physiol* **94**, 1897-1902
14. Fowler, T., Bernhardt, C., and Tierney, M. (1999) *Plant Physiol* **121**, 1081-1092
15. Kieliszewski, M., de Zacks, R., Leykam, J., and Lampport, D. (1992) *Plant Physiol* **98**, 919-926
16. Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R., and Jamet, E. (2008) *BMC Plant Biol.* **8**, 94
17. Feiz, L., Irshad, M., Pont-Lezica, R. F., Canut, H., and Jamet, E. (2006) *Plant Methods* **2**, 10
18. Jeyaprakash, A., Jayashree, G., Mahanta, S., Swaminathan, C., Sekar, K., Surolia, A., and Vijayan, M. (2005) *J. Mol. Biol.* **347**, 181-188
19. Zhang, Y., Giboulot, A., Zivy, M., Valot, B., Jamet, E., and Albenne, C. (2011) *Phytochemistry* **72**, 1109-1123
20. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248-254
21. Laemmli, U. K. (1970) *Nature* **227**, 680-685
22. Willats, W. G., and Knox, J. P. (1996) *Plant J.* **9**, 919-925
23. Merkle, R., and Poppe, I. (1994) *Methods Enzymol* **230**, 1-15
24. Shpak, E., Leykam, J., and Kieliszewski, M. (1999) *Proc Natl Acad Sci USA* **96**, 14736-14741
25. Mort, A., and Lampport, D. (1977) *Anal Biochem* **82**, 289-309
26. Borderies, G., Jamet, E., Lafitte, C., Rossignol, M., Jauneau, A., Boudart, G., Monsarrat, B., Esquerré-Tugayé, M. T., Boudet, A., and Pont-Lezica, R. (2003) *Electrophoresis* **24**, 3421-3432
27. Boudart, G., Jamet, E., Rossignol, M., Lafitte, C., Borderies, G., Jauneau, A., Esquerré-Tugayé, M.-T., and Pont-Lezica, R. (2005) *Proteomics* **5**, 212-221
28. Baldwin, T., Domingo, C., Schindler, T., Seetharaman, G., Stacey, N., and Roberts, K. (2001) *Plant Mol Biol* **45**, 421-435
29. Sommer-Knudsen, J., Bacic, A., and Clarke, A. (1998) *Phytochemistry* **47**, 483-497
30. Tan, L., Varnai, P., Lampport, D., Yuan, C., Xu, J., Qiu, F., and Kieliszewski, M. (2010) *J Biol Chem* **285**, 24575-24583
31. Kieliszewski, M., Leykam, J., and Lampport, D. (1990) *Plant Physiol* **92**, 316-326
32. Hieta, R., and Myllyharju, J. (2002) *J Biol Chem* **277**, 23965-23971
33. Tiainen, P., Myllyharju, J., and Koivunen, P. (2005) *J Biol Chem* **280**, 1142-1148
34. Koski, M., Hieta, R., Hirsila, M., Ronka, A., Myllyharju, J., and Wierenga, R. (2009) *J Biol Chem* **284**, 25290-25301
35. Shpak, E., Barbar, E., Leykam, J., and Kieliszewski, M. (2001) *J Biol Chem* **276**, 11272-11278
36. Estevez, J., Kieliszewski, M., Khitrov, N., and Somerville, C. (2006) *Plant Physiol* **142**, 458-470
37. Gaspar, Y., Johnson, K., McKenna, J., Bacic, A., and Schultz, C. (2001) *Plant Mol Biol* **47**, 161-176
38. Minic, Z., and Jouanin, L. (2006) *Plant Physiol Biochem* **44**, 435-449
39. Jamet, E., Canut, H., Boudart, G., and Pont-Lezica, R. (2006) *Trends Plant Sci.* **11**, 33-39
40. Kotake, T., Tsuchiya, K., Aohara, T., Konishi, T., Kaneko, S., Igarashi, K., Samejima, M., and Tsumuraya, Y. (2006) *J Exp Bot* **57**, 2353-2362

41. Kotake, T., Dina, S., Konishi, T., Kaneko, S., Igarashi, K., Samejima, M., Watanabe, Y., Kimura, K., and Tsumuraya, Y. (2005) *Plant Physiol* **138**, 1563-1576
42. Hata, K., Tanaka, M., Tsumuraya, Y., and Hashimoto, Y. (1992) *Plant Physiol* **100**, 388-396
43. Eudes, A., Mouille, G., Thévenin, J., Goyallon, A., Minic, Z., and Jouanin, L. (2008) *Plant Cell Physiol* **49**, 1331-1341
44. Gibeaut, D., and Carpita, N. (1991) *Plant Physiol* **97**, 551-561
45. Jamet, E., Albenne, C., Boudart, G., Irshad, M., Canut, H., and Pont-Lezica, R. (2008) *Proteomics* **8**, 893-908
46. Oka, T., Saito, F., Shimma, Y.-i., Yoko-o, T., Nomura, Y., Matsuoka, K., and Jigami, Y. (2010) *Plant Physiol* **152**, 332-340
47. Liang, Y., Faik, A., Kieliszewski, M., Tan, L., Xu, W.-L., and Showalter, A. (2010) *Plant Physiol* **154**, 632-642
48. Qu, Y., Egelund, J., Gilson, P., Houghton, F., Gleeson, P., Schultz, C., and Bacic, A. (2008) *Plant Mol Biol* **68**, 43-59
49. Tryfona, T., Liang, H.-C., Kotake, T., Kaneko, S., Marsh, J., Ichinose, H., Lovegrove, A., Tsumuraya, Y., Shewry, P., Stephens, E., and Dupree, P. (2010) *Carbohydr Res* **345**, 2648-2656
50. Léonard, R., Wopfner, N., Pabst, M., Stadlmann, J., Petersen, B., Duus, J., Himly, M., Radauer, C., Gadermaier, G., Razzazi-Fazeli, E., Ferreira, F., and Altmann, F. (2010) *J Biol Chem* **285**, 27192-27200
51. Leonard, R., Petersen, B., Himly, M., Kaar, W., Wopfner, N., Kolarich, D., van Ree, R., Ebner, C., Duus, J., Ferreira, F., and Altmann, F. (2005) *J Biol Chem* **280**, 7932-7940

TABLE 1
MALDI-TOF MS analysis of the tryptic peptides of the AGP31 Pro-rich domain (P0 – P4)

For each peptide, the m/z ratio is indicated as well as the number of Hyp, hexose and pentose residues determined by MALDI-TOF MS analysis of AGP31-enriched fraction issued from CEC/NAC and PNA isolation strategies. Samples were excised from SDS-polyacrylamide gels in the 30-40 kDa and 50-60 kDa mass ranges (Fig. 2, lanes 1a, 1b and 5).

Peptide	m/z	number of Hyp	number of hexoses	number of pentoses
P0	883.52	2	n.d.	n.d.
P1	2163.17	6	1 to 5	1 to 2
P2	2218.25	5	1 to 5	n.d.
P3	2486.35	6	n.d.	n.d.
P4	3476.94	9	1 to 7	1 to 2

n.d.: not detected

TABLE 2**Amino acid sequences of the tryptic peptides of the AGP31 Pro-rich domain (P1 – P4)**

The location of Pro (P) and Hyp (O) within the sequences was deduced from MALDI-TOF/TOF MS/MS results (supplemental Figs. S2A-S2D). Fragmentation experiments were carried out on deglycosylated AGP31 tryptic peptides (P1 – P4) obtained following gel excision of the HF sample (Fig. 3B). Pro and Hyp residues whose position was determined experimentally are written in bold. When it was not possible to determine the exact position of Hyp in some Pro-Pro motifs, both possibilities of sequences (*i.e.* Pro-Hyp or Hyp-Pro) are indicated (PO/OP).

Sequences of the Pro-rich domain tryptic peptides

P1: AOV**SOO**AK(PO/OP)VK(PO/OP)VY(PO/OP)TK

P2: AOVK(PO/OP)TK(PO/OP)VK(PO/OP)VY(PO/OP)TK

P3: AOVK(PO/OP)V**SOO**TK(PO/OP)VT(PO/OP)VY**PP**K

P4: AOVK(PO/OP)TK(PO/OP)VK(PO/OP)V**SOO**AK(PO/OP)VK(PO/OP)VY(PO/OP)TK

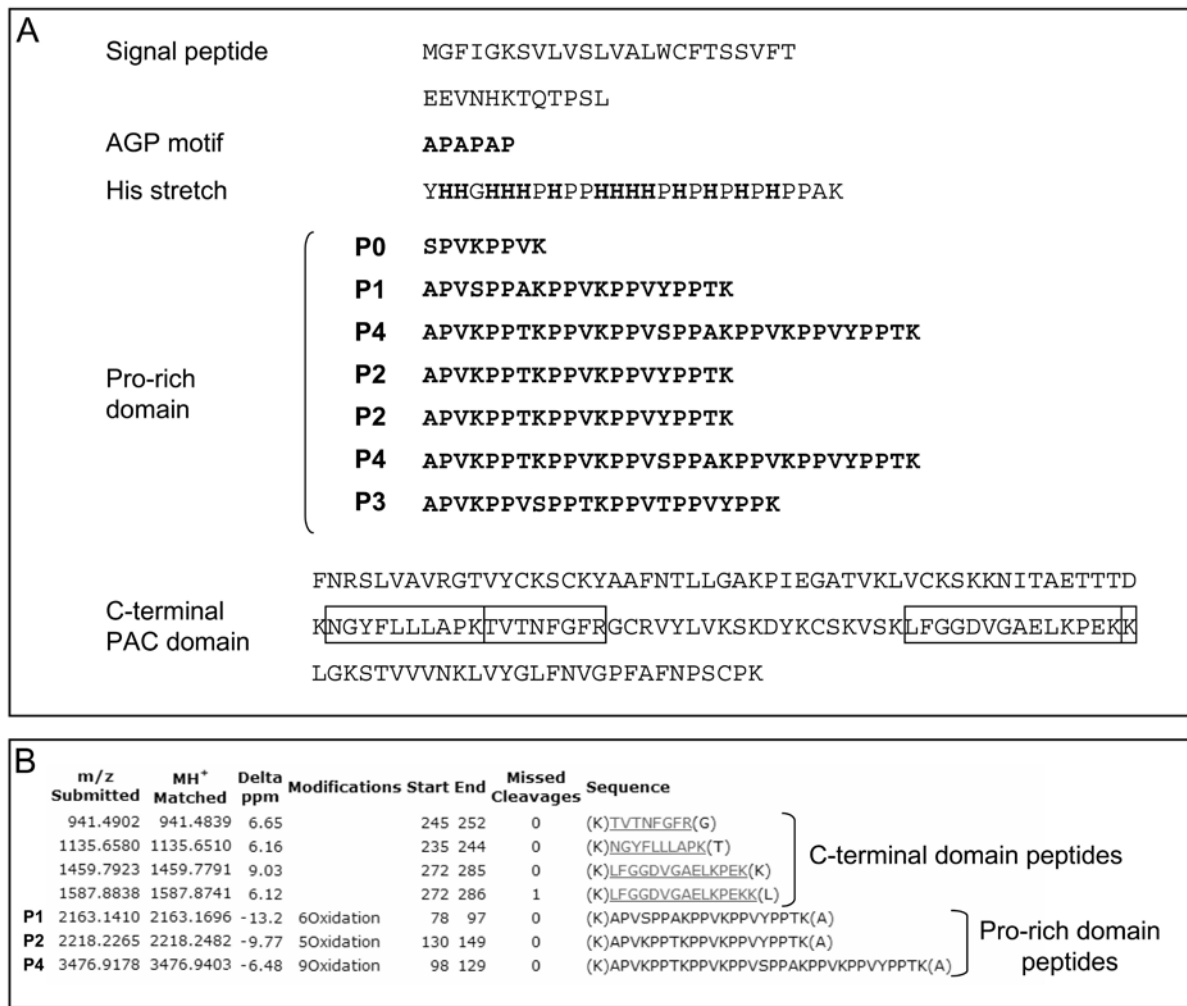


FIGURE 1. Identification of AGP31 by PMF

A- Amino acid sequence of AGP31 displaying the different domains of the protein, *i.e.* signal peptide, AGP motif, His stretch, Pro-rich domain and C-terminal PAC domain. The protein sequence should be read from left to right and from top to bottom. The sequence of the Pro-rich domain is arranged to emphasize the tryptic peptides (P0 to P4) identified in this study. The four tryptic peptides of the C-terminal domain used for protein identification are indicated in frames. Note that P1, P2 and P4 also contributed to protein identification as shown in B.

B- ProteinProspector (<http://prospector.ucsf.edu/prospector/mshome.htm>) search results allowed the identification of AGP31 by PMF using MALDI-TOF MS. Among the seven matching tryptic peptides, four belong to the C-terminal domain and three to the Pro-rich domain (noted P1, P2 and P4). These three peptides are predicted to contain, respectively, 6, 5 and 9 oxidations assumed to correspond to Pro hydroxylations. The table shows submitted and matched m/z of the peptides, the differences in mass between the two (Delta ppm), the proposed modifications, the position of the peptides on the protein (start, end), the number of missed cleavages, and the peptide sequences.

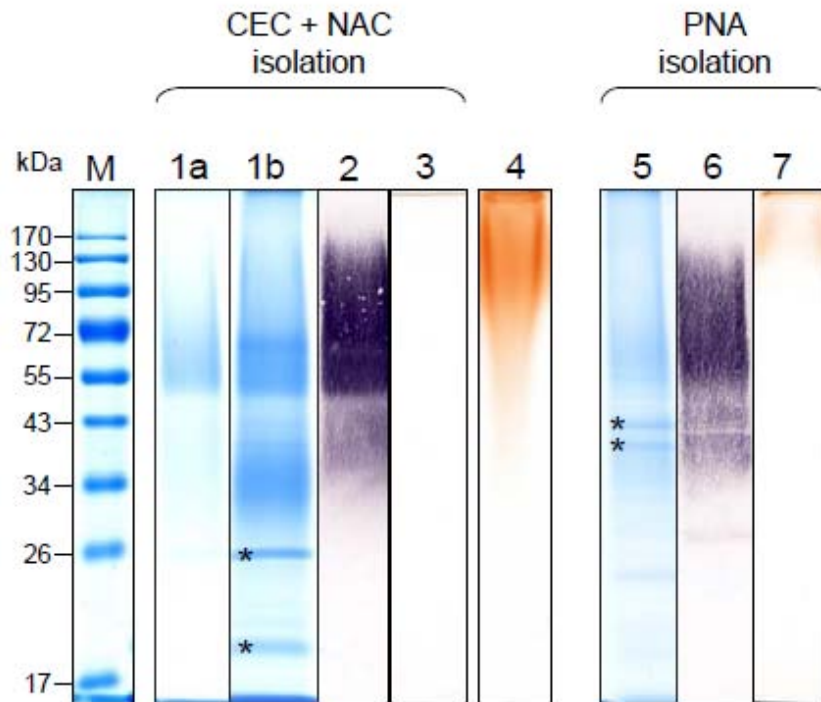


FIGURE 2. Detection of AGP31 in enriched fractions after separation by SDS-PAGE

Two strategies were employed to enrich native AGP31 from a CWP extract of 11-day-old etiolated hypocotyls of *A. thaliana*: (i) a combination of CEC followed by NAC (lanes 1 to 3); (ii) a PNA-lectin affinity chromatography (lanes 5 to 7). AGP31-enriched fractions were separated by SDS-PAGE. About 13 μg of proteins were used for each analysis, excepted for lanes 1b, 2 and 3 where 36 μg of proteins were loaded. Minor contaminant proteins labeled with stars were identified in the elution fractions using both strategies (lanes 1b and 5). Molecular mass markers (M) are shown on the left.

Lanes 1a, 1b and 5: Colloidal blue staining of AGP31-enriched fractions. Note that AGP31 migrated as a smear above 50 kDa but was also present as truncated forms in an additional smear from 30 to 40 kDa (lanes 1b and 5).

Lanes 2 and 6: PNA-lectin blot analysis of AGP31-enriched fractions. After separation by SDS-PAGE, proteins were electrotransferred onto a nitrocellulose membrane and treated as indicated in experimental procedures. PNA labeled a smear from 34 to 170 kDa.

Lanes 3 and 7: β -glucosyl Yariv staining of AGP31-enriched fractions. Note that a signal was observed above 170 kDa in the PNA-fraction.

Lane 4: β -glucosyl Yariv staining of the flow-through fraction from the CEC step.

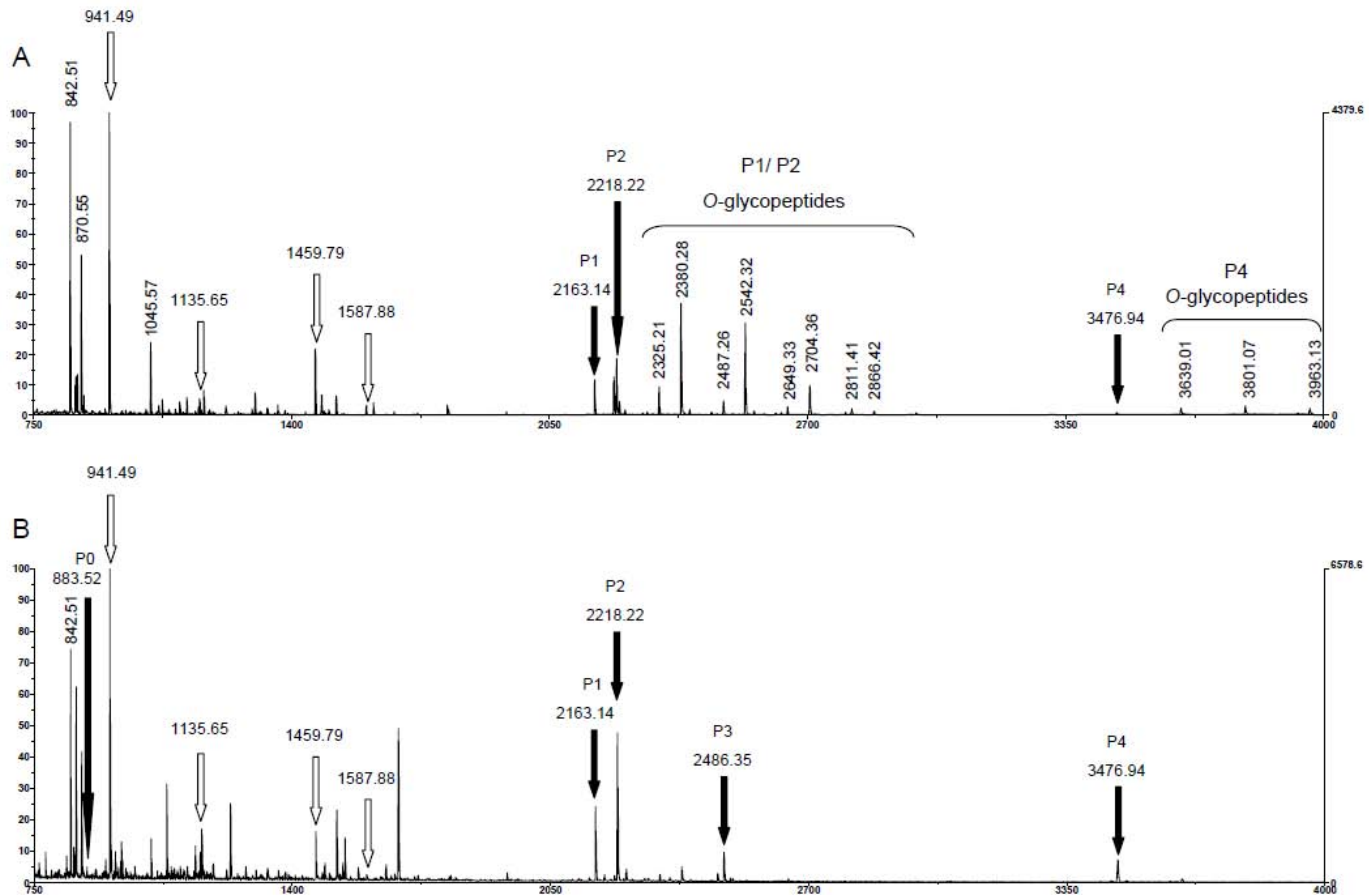


FIGURE 3. Comparison of MALDI-TOF MS spectra of glycosylated and HF-deglycosylated forms of AGP31

A - Identification of AGP31 by PMF using MALDI-TOF MS. After separation by SDS-PAGE of an AGP31-enriched fraction obtained combining CEC and NAC, bands containing glycosylated AGP31 were excised from the 50-60 kDa mass range and digested with trypsin. Unmodified tryptic peptides from the C-terminal domain (see Fig. 1B) are shown with white arrows whereas tryptic peptides of the Pro-rich domain (P1, P2 and P4) are labeled with black arrows. *O*-glycopeptides associated to these peptides are shown with a brace and the masses of most intense ones are indicated. Note that other intense ions (842.51, 870.55 and 1045.57) correspond to trypsin or matrix peptides.

B - Identification of AGP31 after HF deglycosylation. After separation by SDS-PAGE of a total CWP extract, bands containing deglycosylated AGP31 were excised, digested with trypsin and analyzed by MALDI-TOF MS. AGP31 glycopeptides (shown in A) disappeared in favor of the corresponding deglycosylated peptides (P0 – P4, black arrows). As expected, unmodified peptides from the C-terminal domain (white arrows) were not modified by the HF treatment.

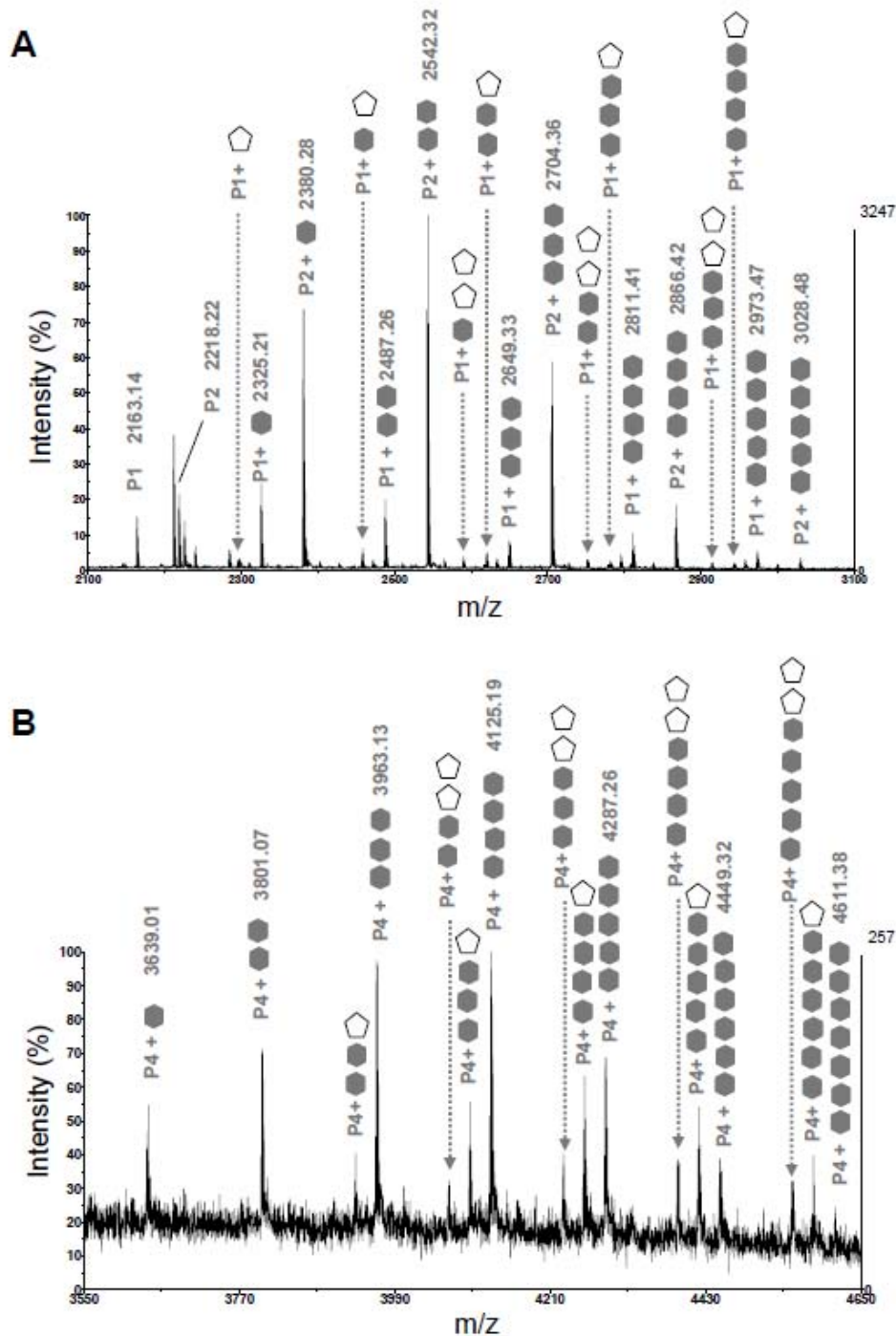


FIGURE 4. Identification of *O*-glycopeptides of the AGP31 Pro-rich domain using MALDI-TOF MS

After separation by SDS-PAGE of an AGP31-enriched fraction obtained combining CEC and NAC, bands containing AGP31 were excised from the 50-60 kDa mass range and digested with trypsin. MALDI-TOF MS spectrum analysis showed the presence of tryptic peptides of the Pro-rich domain (P1, P2 and P4, see Figs. 1 and 3A) and associated glycopeptides, regularly spaced by 162 Da or 132 Da, corresponding to the mass of an hexose (●) or a pentose (◊) respectively. The number of hexoses and pentoses is indicated. For clarity, monoisotopic masses are indicated only for the most intense *O*-glycopeptides.

A - MALDI-TOF MS spectrum zoom (m/z 2100-3100) on P1 and P2 *O*-glycopeptides.

B - MALDI-TOF MS spectrum zoom (m/z 3550-4650) on P4 *O*-glycopeptides.

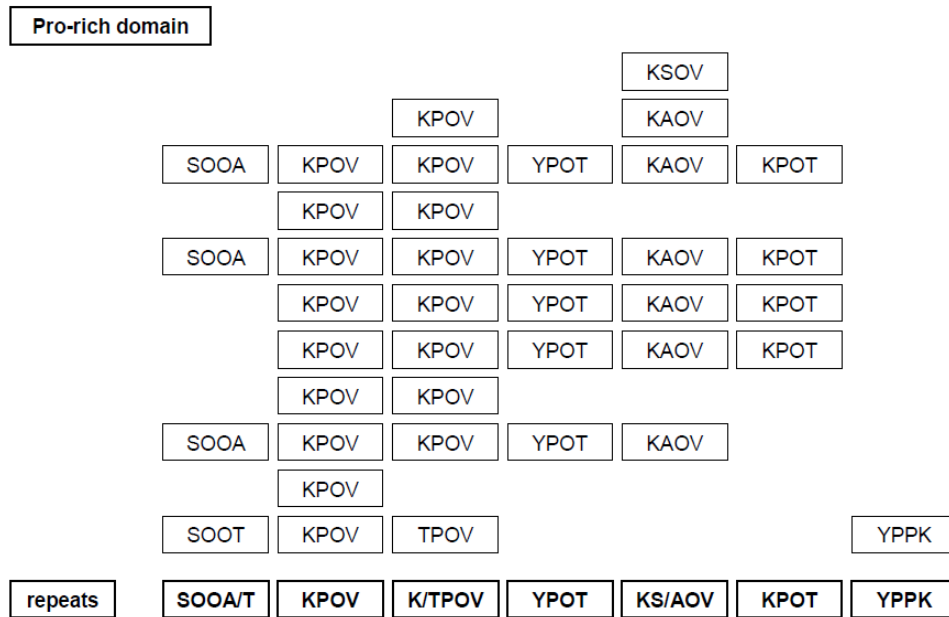


FIGURE 5. Location of Hyp residues within the Pro-rich domain

Hyp residues were located combining MALDI-TOF/TOF MS/MS fragmentation results on deglycosylated tryptic peptides of the AGP31 Pro-rich domain (see supplemental Figs. S2A-S2D) and literature data (3, 29). Sequence should be read from left to right and from top to bottom. The Hyp residues of the AGP31 Pro-rich domain are represented by O. Repeated motifs (repeats) are indicated in bold at the bottom of the figure.

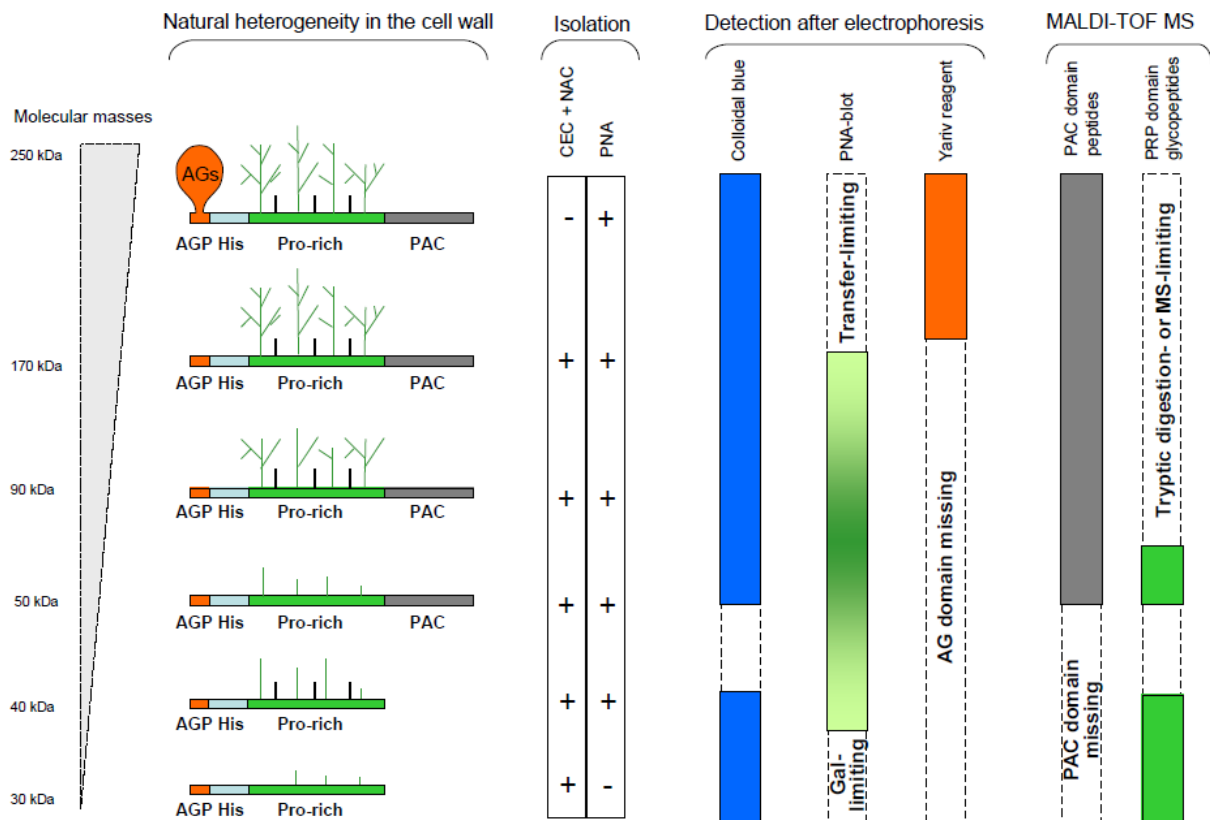


FIGURE 6. AGP31 structural models

The different domains of the mature AGP31 are represented from left to right: N-terminal AGP motif (AGP), His stretch (His), Pro-rich domain (Pro-rich), and C-terminal PAC domain (PAC). The various glycoforms and truncated forms of AGP31 are schematized as well as the results obtained for isolating, detecting and analyzing each of them using the different approaches carried out in this study. Limits encountered at each step are indicated. Arabinogalactan motifs (AGs) assumed to substitute the short N-terminal AGP motif are indicated. The size-heterogeneous Hyp-O-Gal/Ara-rich motifs expected on the Pro-rich domain are represented with green lines and putative short Ara-oligosaccharides with purple lines

Chapitre 4

RECHERCHE DES PARTENAIRES D'AGP31 PAR DES TESTS D'INTERACTIONS IN VITRO

4. RECHERCHE DES PARTENAIRES D'AGP31 PAR DES TESTS D'INTERACTION *IN VITRO*

Parmi les milliers de protéines d'une cellule, si certaines peuvent agir seules, la plupart doivent opérer de concert dans des complexes, des réseaux, de façon à permettre de régir les structures et les fonctions cellulaires. Ainsi au niveau de l'architecture de la paroi primaire, les différentes protéines peuvent interagir entre elles mais aussi avec d'autres constituants polysaccharidiques pour contribuer aux fonctions et à la structure bien ordonnée de la paroi.

4.1 Les partenaires possibles d'AGP31 dans les parois végétales

La protéine AGP31 étant constituée de plusieurs domaines, nous avons recherché dans la bibliographie quels pourraient être ses partenaires dans les parois en nous intéressant successivement au domaine AGP, au domaine riche en His et au domaine PAC. Cette analyse devait nous permettre de guider nos expériences visant à comprendre le mode d'action d'AGP31.

4.1.1 Les partenaires des AGP

Bien que certains mutants d'AGP possèdent des phénotypes spécifiques (*cf.* § 1.3.4), les modes d'interaction moléculaire des AGP et leur intégration dans l'architecture de la paroi cellulaire restent encore à élucider.

Une interaction entre AGP et pectines a été montrée par certains travaux. Après précipitation au réactif de Yariv des AGP de cellules de carotte (*Daucus carota*) en culture, des fractions riches en GalA ont été isolées. Ces fractions sont dégradées par une pectine méthylestérase (PME) et une polygalacturonase (PG), ce qui indique que les AGP pourraient être liées à des HG par des liaisons covalentes (Hengel et al., 2001). De même, une analyse par différents anticorps monoclonaux d'AGP isolés de milieux de culture de suspensions cellulaires de betterave (*Beta vulgaris*) a révélé des épitopes correspondants aux AGP et aux pectines (Wisniewska et Majewska-Sawka, 2007). Li et al. (1995) ont montré une co-localisation des AGP avec les pectines dans l'intine du grain de pollen de tabac (*N. tabacum*) grâce à des anticorps monoclonaux spécifiques, les AGP étant préférentiellement localisées au voisinage de la membrane plasmique. Dans le tube pollinique, les pectines méthylestérifiées ou non co-localisent avec les AGP dans les couches moyennes et externes de la paroi cellulaire (Li et al., 2005).

Kjellbom et al. (1997) ont montré que les AGP de betterave peuvent former des complexes covalents *in vitro* en présence de peroxyde d'hydrogène et de peroxydase. Dans une étude réalisée sur la paroi de cellules de carottes, Baldwin et al. (1993) ont suggéré une interaction *in vitro* et calcium dépendante entre une AGP non classique et des pectines non méthylestérifiées extraites de parois de carottes, mais pas avec des HG commerciaux. Ces auteurs ont également montré une aggrégation entre les molécules de cette AGP. Cependant, dans le cas de cette AGP, ces interactions moléculaires restent à être démontrées *in planta*.

4.1.2 Les protéines à domaine riche en His

Dans plusieurs systèmes biologiques, les résidus His sont associés aux interactions avec des métaux. Les protéines de transport de Fer (ITP) contiennent un domaine riche en His HGDEKGHEHGHGESGE responsable de leur affinité pour les ions métalliques (Kruger et al. 2002). La déhydrine de *Citrus unshiu* comporte le domaine HKGEHHS GDHH qui interagit avec plusieurs métaux, dont les ions Cu^{2+} (Hara et al., 2005). Le domaine riche en His de la protéine Eu3 du soja (*Glycine max*) est nécessaire pour son interaction avec les ions Ni^{2+} et son effet activateur d'une uréase (Freyermuth et al., 2000).

Parmi les protéines pariétales, plusieurs familles comportent des domaines riches en His. Les protéines HRA (*His-Rich AGP*) identifiées par Aizat et al. (2011), présentent des motifs riches en His avec de courts domaines d'AG dont le nombre diffère entre les céréales et les dicotylédones. Chez l'orge (*Hordeum vulgare*), *HvHRA1* code pour une protéine comportant de multiples motifs riches en His entrecoupés par de courts glycomodules d'AGP. Une surexpression de ce gène augmente l'affinité de la fraction d'AGP extraite pour les ions Cu^{2+} . Chez *A. thaliana*, AGP24, un AG peptide de la paroi qui présente un motif N-terminal riche en His suivi d'un motif AGP, HEGHHHHAOAOAOGO (Schultz et al., 2004) pourrait interagir avec des ions métalliques et est présent dans les zones d'abscission des graines (Ellis et al., 2010). D'autres AGP riches en His ont été identifiées chez les dicotylédones. C'est le cas des homologues d'AG31 par exemple GaRSGP, NtTTS1 et DcAGP1 décrites dans l'introduction (cf. § 1.4.2). Le domaine riche en His d'AGP31 interagit avec Ni^{2+} *in vitro* (Liu et Mehdy, 2007); cf Chapitres 3 et 4). Cependant, aucune information n'est disponible concernant le rôle de ces domaines riches en His *in planta*.

4.1.3 Les protéines à Cys conservées

Les protéines à Cys conservés (*Cysteine-Rich Peptide*, CRP) constituent une superfamille caractérisée par la petite taille de ses membres (moins de 160 acides aminés) et la présence d'un peptide signal du côté N-terminal (Silverstein et al., 2007; Marshall et al., 2011). La composition en acides aminés des CRP est très variable, mais ces protéines comportent 4 à 16 résidus Cys conservés. Nous nous intéresserons à certaines d'entre elles dont la structure ou le mode d'action ont été particulièrement étudiés.

Les Protéines à 12 Cys conservées

La famille GASA (*Giberellic Acid-Stimulated Arabidopsis*) se caractérise par des protéines de petite taille et la présence d'un peptide signal. Parmi les différents membres de cette famille, GASA4, la protéine la mieux caractérisée de cette famille, présente une activité anti-oxydante. Cette activité est attribuée au domaine à Cys conservées, puisqu'une protéine tronquée, ne présentant que le domaine C-terminal, montre cette activité. Au contraire, le remplacement par mutagenèse dirigée de 4 Cys conservées par des Ala inhibe l'activité anti-oxydante de GASA4 (Rubinovich et Weiss, 2010). Ce même effet anti-oxydant avait aussi été montré chez GIP2, GIP4 et GIP5, des protéines de la famille GASA de *P. hybrida* (Wigoda et al., 2006).

Une protéine PRP à domaine GASA a été identifiée chez *P. hybrida*, il s'agit de PRGL, dont l'orthologue chez *A. thaliana* est GASA14. Cette protéine pariétale est impliquée dans la régulation de l'élongation cellulaire durant le développement de la plante (Peng et al., 2010). La nécessité d'ajouter du SDS pour l'extraction de PRGL à partir des parois suggère que PRGL est fortement liée aux polysaccharides pariétaux. Cependant les rôles respectifs des domaines PRP et GASA restent à élucider (Peng et al., 2008).

Les Protéines à 8 Cys conservées

Les membres de la famille des protéines à 8CM (*eight-cysteine motif*) partagent un peptide signal du côté N-terminal et un domaine contenant 8 Cys conservées comportant moins de 100 acides aminés du côté C-terminal. Parmi les membres de cette famille de protéines pariétales, les HyPRP (*Hybrid Proline-Rich Proteins*), sont des protéines ayant un domaine répétitif riche en Pro du côté N-terminal en plus du domaine 8CM (Dvorakova et al., 2007). Le même type de domaine se trouve chez d'autres polypeptides comme les LTP (*Lipid Transfer Proteins*), des protéines pariétales impliquées dans la défense contre les pathogènes, la synthèse de cuticule, et pouvant agir comme un modulateur de la croissance et du développement de la plante (Yeats et Rose, 2007). Une structure tertiaire formée par quatre hélices hydrophobes stabilisées par les Cys conservés a été déterminée par diffraction des rayons X pour une LTP de soja (Baud et al. 1993) et par RMN pour une LTP de blé (Gincel et al., 1994). L'étude structurale suggère une affinité de ces protéines à certains lipides comme les glycérolipides, les acides gras, l'acyl-CoA et les stérols (Yeats et Rose, 2007).

Chez les animaux, des études portant sur la lutropine, une glycoprotéine du sérum, a montré que le domaine à 8 Cys localisé à l'extrémité N-terminale de son récepteur membranaire (*mannose receptor*) reconnaît les *N*-glycanes spécifiques qu'elle porte et qui se terminent par des résidus GalNac-4-SO₄ (Fiete et al., 1998). Des domaines à Cys conservées similaires sont décrits chez trois autres récepteurs à domaines lectines extracellulaires (Fiete et al., 1998).

Les Protéines à 6 Cys conservées (Van Hengel et Roberts, 2003)

Les protéines à domaine C-terminal à 6 Cys conservées définissent chez les plantes la famille des protéines à domaine PAC associées à la paroi (Baldwin et al., 2000). Elle comporte DcAGP1, PvPRP1, NtTTS1, NaPRP4, NtTTS2, NaPRP5, AtAGP30 et AtAGP31. Les résidus Cys sont conservés en nombre et en espacement. Certaines exceptions existent comme NtTTS1 dont les Cys-5 et Cys-6 sont espacées de 20 au lieu de 48 à 53 acides aminés et PvPRP1 à laquelle manquent les Cys-1 et Cys-5. D'autres protéines à 6 Cys conservées spécifiques du pollen ont également été mises en évidence. Ce sont Ole e 1, LAT52, Z16402, U85499 (Baldwin et al., 2000) ainsi que At2g34700 (Van Hengel et Roberts, 2003). Comme mentionné dans l'introduction (*cf.* § 1.4.9) la famille de protéines à domaine PAC pourrait même être plus grande chez *A. thaliana*.

En conclusion, pour ces protéines à Cys conservées, une conservation de structure reflète certainement une fonction commune et/ou un mode d'action similaire. Au vu des résultats expérimentaux, deux modes d'action peuvent être envisagés : des interactions avec des polysaccharides pariétaux ou encore un rôle anti-oxydant.

4.1.4 Analyse bioinformatique de la structure d'AGP31 : recherche de domaines désordonnés

Le logiciel metaPrDOS (*meta Protein DisOrder prediction System*) disponible sur le site <http://prdos.hgc.jp/cgi-bin/meta/top.cgi> estime la propension d'une protéine à présenter des structures organisées ou désorganisées. Les domaines N-terminaux (AGP et domaine riche en His) et le domaine central PRP possèderaient selon la prédiction, une structure secondaire désorganisée alors que le domaine PAC localisé à l'extrémité C-terminale adopterait une structure organisée et globulaire (Figure 4.1). Des protéines dites "non structurées", ne présenteraient pas de structure 3D particulière sauf lorsqu'elles entrent en interaction avec d'autres facteurs (Dyson et Wright, 2005). Plus généralement, environ 40% des protéines eucaryotes possèderaient une région intrinsèquement non structurée (Wright et Dyson, 1999).

Il est possible qu'une telle structure soit un avantage permettant à AGP31 de se lier avec un grand nombre de cibles et d'être impliquée dans l'interaction avec plusieurs composants de la paroi cellulaire. La présence d'une structure désorganisée combinée avec celle de plusieurs domaines capables chacun d'exercer une fonction différente permet de proposer l'hypothèse qu'AGP31 est capable d'interagir avec différents partenaires et de contribuer à des réseaux pariétaux.

A

1	MGFIGKSVLV	SLVALWCFTS	SVFTEEVNHK	TQTPSLAPAP	APYHHGHHHP	50
51	HPPHHHHHPH	HPHPHPPAKS	PVKPPVKAPV	SPPAKPPVKP	PVYPPTKAPV	100
101	KPPTKPPVKP	PVSPPAKPPV	KPPVYPPTKA	PVKPPTKPPV	KPPVYPPTKP	150
151	PVYPPTKAPV	KPPVSPPTKP	PVTPPVYPPK	FNRSLVAVRG	TVYCKSKKYA	200
201	AFNTLLGAKP	IEGATVKLVC	KSKKNITAET	TTDKNGYFLL	LAPKTVTNFG	250
251	FRGCRVYLVK	SKDYKCSKVS	KLFGGDVGAE	LKPEKKLGKS	TVVVNKLVIYG	300
301	LFNVGPFAFN	PSCPK				

B

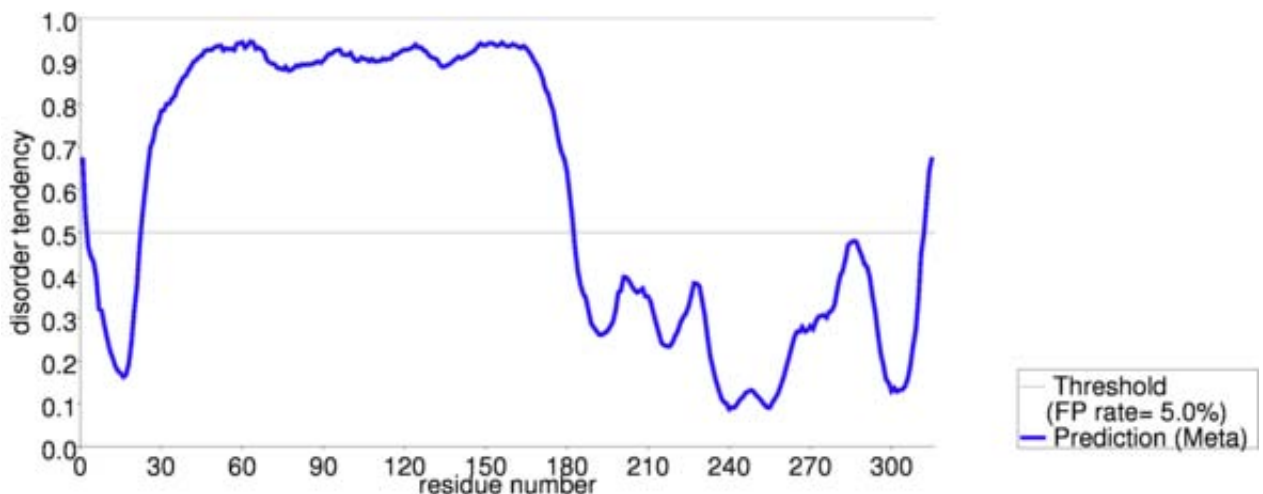


Figure 4.1 : Estimation de la propension d'AGP31 à présenter une structure désordonnée par analyse *in silico*

A. Séquence en acides aminés d'AGP31. La séquence en rouge est celle des domaines N-terminaux (AGP et domaine riche en His) et du domaine central PRP. **B.** Prédiction de structures organisées ou désorganisées par le logiciel metaPrDOS. La prédiction de structures désordonnées est la plus élevée (0.9) pour les séquences indiquées en rouge en A.

4.1.5 Stratégie de recherche de partenaires interagissant avec AGP31 dans les parois végétales

L'ensemble de ces analyses bibliographique et bioinformatique nous a conduits à rechercher des partenaires polysaccharidiques d'AGP31 au sein de la paroi, et à nous focaliser sur le domaine PAC conservé entre différentes protéines.

Pour tester les interactions d'AGP31 et du domaine PAC, des tests d'interaction *in vitro* par *dot-blot* sur des membranes de nitrocellulose ont été effectués. Pour les réaliser, il a été

nécessaire (i) de purifier ces deux protéines grâce notamment à la mise au point d'un système d'expression et de purification du domaine PAC recombinant, (ii) d'extraire des polysaccharides pariétaux et (iii) de tester les interactions dans différentes conditions.

4.2 Production et purification d'AGP31 et de son domaine PAC

Pour les tests d'interaction *in vitro*, la protéine AGP31 et le domaine PAC ont été produits dans des systèmes différents et purifiés par différentes stratégies résumées sur la figure 4.1.

	Domaine PAC d'AGP31			AGP31
Construction				
Système d'expression	<i>E. coli</i>	<i>E. coli</i>	<i>N. benthamiana</i> (expression transitoire dans les feuilles)	<i>A. thaliana</i> sauvages (hypocotyles étiolés) sur-expresses (feuilles, hypocotyles étiolés)
Stratégie de purification	Fraction soluble NAC (palier ou gradient d'imidazole) CEC (gradient de NaCl)	Fraction soluble NAC (gradient d'imidazole) en conditions dénaturantes puis, CEC	Fraction soluble CEC (gradient de NaCl) NAC (palier d'imidazole)	Purification de parois Extraction de protéines avec des solutions salines CEC (gradient de NaCl) puis, NAC ou Chromatographie d'affinité sur PNA
Domaines d'AGP31	domaine AGP	domaine riche en His	domaine PRP	domaine PAC
Étiquettes	MBP	épitope V5	6xHis	

Figure 4.2 : Stratégies de production et de purification de la protéine AGP31 et de son domaine PAC pour les tests d'interaction *in vitro*. Seules les protéines repérées par une étoile bleue ont été utilisées pour ces tests

4.2.1 Purification de la protéine AGP31 native et recombinante

La protéine AGP31 native est purifiée à partir des parois de cellule d'hypocotyles étiolés et sous forme recombinante à partir de jeunes feuilles en utilisant les deux techniques de purification décrites dans le Chapitre 3 : chromatographie d'échange cationique (CEC) suivie d'une chromatographie d'affinité sur colonne de Nickel (NAC), (ii) et chromatographie par affinité sur une matrice d'agarose couplée à la lectine PNA (Figure 4.2).

4.2.2 Production et purification du domaine PAC d'AGP31 recombinant

4.2.2.1 Optimisation de la production et de la purification du domaine PAC recombinant exprimé chez *E. coli*

L'utilisation d'*E. coli* pour l'expression de protéines hétérologues est une méthode efficace. Néanmoins, chaque protéine demeure un cas particulier, nécessitant une optimisation des techniques de production et de purification. Ainsi, la mise au point de la stratégie de production et de purification du domaine PAC d'AGP31 a été laborieuse.

Production d'une protéine de fusion PAC/V5/6xHis (Figure 4.3)

Le premier système d'expression utilisé était *E. coli* TOP10 intégrant le vecteur pBAD-TOPO capable de produire le domaine PAC sous forme de protéine de fusion avec un épitope V5 et une séquence 6xHis en C-terminal (PAC/V5/6xHis). Après induction de la protéine, les fractions de protéines solubles et insolubles obtenues à partir des bactéries ont été analysées par SDS-PAGE et *Western blot* avec un anticorps anti-V5. Une bande de masse moléculaire apparente d'environ 19 kDa correspondant à la masse moléculaire théorique de PAC/V5/6xHis est présente dans les deux fractions analysées, mais en plus grande proportion dans la fraction insoluble (Figure 4.3).

Purification de PAC/V5/6xHis à partir de la fraction soluble

La purification de la protéine a été réalisée par NAC en utilisant l'étiquette 6xHis. Cette méthode n'a pas permis d'enrichir la protéine recombinante. Une optimisation du rendement de purification a été essayée par une augmentation du volume de culture bactérienne à 500 mL et fixation des protéines sur la colonne d'affinité pendant une nuit. Une purification par chromatographie d'échange cationique (CEC) a aussi été tentée (résultats non montrés). Toutes ces modifications sont restées très insuffisantes pour améliorer significativement le rendement en protéine d'intérêt. C'est pourquoi nous avons essayé d'utiliser la fraction de protéines insolubles qui contenait davantage de protéines.

Purification de PAC/V5/6xHis à partir de la fraction insoluble

Une extraction et une purification par NAC dans un tampon comportant de l'urée 8 M a donc été réalisée à partir de 500 mL de culture. L'analyse des fractions d'élution de la colonne de Nickel montre la présence de contaminants résiduels de hauts et de faibles masses moléculaires dans les fractions contenant la protéine d'intérêt et un rendement insuffisant pour poursuivre la purification (résultats non montrés).

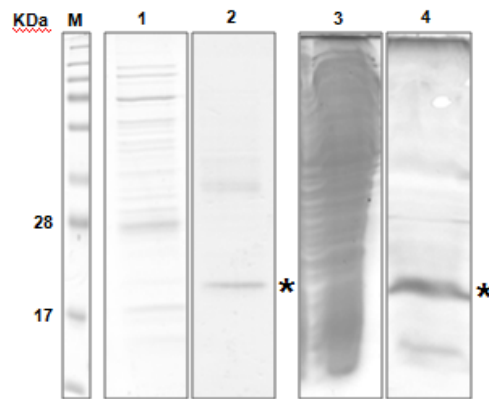


Figure 4.3 : Analyse par SDS-PAGE et *Western blot* des fractions de protéines PAC/V5/6xHis solubles et insolubles exprimées dans *E. coli*

1 et 2 : Fractions de protéines solubles analysées par SDS-PAGE et coloration au bleu de Coomassie (1) ou *Western blot* révélé par un anticorps anti-V5 (2).

3 et 4 : Fractions de protéines insolubles analysées par SDS-PAGE et coloration au bleu de Coomassie (3) ou *Western blot* révélé par un anticorps anti-V5 (4).

Ces premières tentatives infructueuses nous ont amenés à conclure que ce système d'expression ne produisait pas suffisamment de protéine recombinante. Nous avons alors choisi de tester un nouveau système d'expression.

Production d'une protéine de fusion MBP/PAC/V5/6xHis (Figure 4.4)

L'étiquette MBP a été choisie puisqu'elle peut augmenter la solubilité des protéines recombinantes. Ainsi, la production de MBP/PAC/V5/6xHis a été effectuée dans des bactéries *E. coli* en utilisant le vecteur pMAL-c2. Pour optimiser la production de la protéine, le temps d'induction par l'IPTG a été modulé entre 1 h et 3 h. L'analyse par SDS-PAGE des différents extraits bactériens montre que la protéine a bien été induite car une bande de masse moléculaire apparente de 60,7 kDa correspondant à celle qui était attendue, est apparue pour chaque temps testé (Figure 4.4). La durée de l'induction n'a pas eu d'effet sur le niveau d'accumulation des protéines. Enfin, l'analyse par SDS-PAGE et *Western blot* révélé par un anticorps anti-V5 a montré une accumulation significative de la protéine d'intérêt dans la fraction soluble.

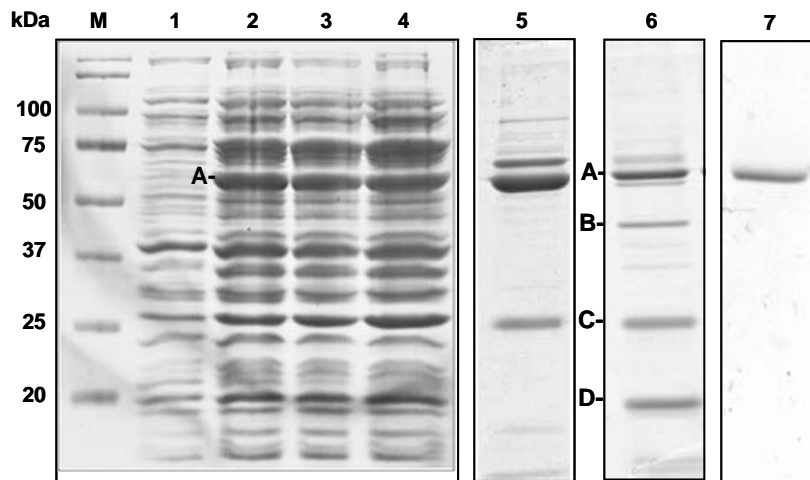


Figure 4.4 : Analyse par SDS-PAGE des fractions obtenues après différents temps d'induction à l'IPTG et à différentes étapes de la purification de la protéine MBP/PAC/V5/6xHis à partir de *E. coli*

1, 2, 3 et 4 : Extraits bactériens obtenus respectivement après 0, 1, 2, 3 h d'induction de la production de la protéine MBP/PAC/V5/6xHis par l'IPTG.

5 et 6 : Fractions éluées de la colonne NAC avant et après digestion par le facteur Xa en conditions dénaturantes

7 : Fraction contenant MBP/PAC/V5/6xHis après CEC.

Les bandes A et B correspondent à MBP/PAC/V5/6xHis et MBP, C et D correspondent à des protéines bactériennes respectivement.

Le gel a été coloré au bleu de Coomassie. La masse moléculaire (en kDa) et la position des marqueurs sont indiquées sur la gauche.

Purification par NAC en conditions dénaturantes

A partir des protéines solubles, la purification de MBP/PAC/V5/6xHis a été réalisée par NAC, l'élution des protéines étant effectuée par un gradient linéaire d'imidazole. L'analyse par SDS-PAGE permettant de contrôler la pureté des protéines des différentes fractions montre la présence de contaminants dans les fractions d'élution contenant MBP/PAC/V5/6xHis. Une optimisation de la NAC a donc été envisagée en utilisant des paliers d'imidazole pour l'élution des protéines. Cette modification a permis de concentrer les protéines d'intérêt, mais également les contaminants (résultats non présentés).

Une amélioration significative de la purification de MBP/PAC/V5/6xHis a été obtenue en travaillant en conditions dénaturantes en présence d'urée 8 M pour réduire les interactions avec les protéines bactériennes. Ces nouvelles conditions ont conduit à l'élimination de la majorité des contaminants (Figure 4.4, piste 5). Ainsi, MBP/PAC/V5/6xHis a été estimée pure à 80%.

Clivage de l'étiquette MBP par le facteur Xa

L'étape suivante consistait à cliver le domaine MBP des protéines de fusion. Une cinétique de réaction a été réalisée sur la protéine fusion à 5 h et à 20 h de digestion par le facteur Xa. Pour cela, des analyses par SDS-PAGE suivies d'analyses par spectrométrie de masse de type MALDI-TOF ont été réalisées pour voir s'il y eu un clivage et pour connaître la nature de chaque fragment protéique généré. Sur la Figure 4.4 (piste 6), on observe une bande A de masse moléculaire apparente de 60,7 kDa qui correspond à MBP/PAC/V5/6xHis et une bande B de 42,5 kDa qui correspond à MBP. Les bandes C et D correspondent à des protéines bactériennes contaminantes. Les autres bandes correspondent probablement à des

contaminants bactériens ou à des produits de dégradation non spécifiques. Par contre, la protéine PAC/V5/6xHis qui a une masse moléculaire théorique de 19,1 kDa n'a pas été retrouvée après le clivage par le facteur Xa.

Nous avons donc envisagé de réaliser les expériences d'interaction avec MBP/PAC/V5/6xHis en utilisant la protéine MBP comme témoin négatif.

Purification par CEC

Une étape supplémentaire de purification par CEC après l'étape de NAC en conditions dénaturantes a été effectuée. La protéine est éluée par 0,5 à 0,75 mM de NaCl avec un niveau de pureté élevé (Figure 4.4, piste 7). Le rendement de purification de MBP/PAC/V5/6xHis était d'environ 20 µg de protéine à partir de 400 mL de milieu de culture, une quantité permettant de réaliser les tests d'interaction *in vitro*.

4.2.2.2 Optimisation de la production et de la purification du domaine PAC recombinant exprimé dans des feuilles de *Nicotiana benthamiana*

Le domaine PAC d'AGP31 a également été produit en système eucaryote par une transformation transitoire des feuilles de *N. benthamiana*. En effet, la protéine possède deux sites potentiels de *N*-glycosylation (NRSL, NITA). Cette modification post-traductionnelle pouvant intervenir dans les interactions testées.

Production de la protéine de fusion PAC/V5/6xHis

Les souches d'*A. tumefaciens* contenant les constructions codantes des protéines de fusion ont été infiltrées dans les feuilles de *N. benthamiana*. Pour vérifier la production de la protéine, des prélèvements sont effectués après 48 h de transformation et analysés par *Western blot*. Sur la membrane de nitrocellulose révélée par un anticorps anti-V5, une bande intense à 19,1 kDa, absente sur la piste du témoin négatif montre que la protéine PAC/V5/6xHis est bien produite (résultat non montré).

Purification de PAC/V5/6xHis

Les parois des cellules des feuilles de *N. benthamiana* surexprimant PAC/V5/6xHis ont été purifiées et les protéines pariétales extraites par des solutions salines. Ensuite, un protocole de purification du domaine PAC a été mis au point. La première étape consistait à éliminer un maximum de protéines pariétales par CEC en utilisant le fait que le pI de la protéine est basique (9,6). La seconde étape consistait en une NAC utilisant l'étiquette 6xHis. Cette deuxième étape ne pouvait pas être réalisée sur l'extrait de protéines totales car les protéines d'intérêt étaient minoritaires, ce qui aurait conduit à une faible efficacité de fixation des protéines sur la matrice.

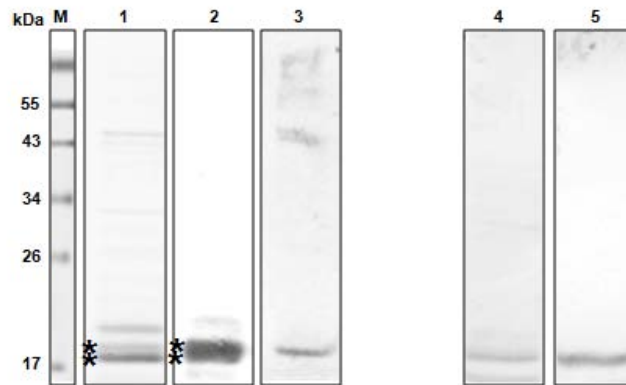


Figure 4.5 : Analyse des fractions obtenues aux différentes étapes de la purification de PAC/V5/6xHis à partir de feuilles de *Nicotiana benthamiana*

1, 2 et 3 : Fractions issues de la CEC analysée par SDS-PAGE et coloration au bleu de Coomassie (1) ou *Western blot* révélé par un anticorps anti-V5 (2) ou *lectin blot* révélé par la lectine GNA (3)

4 et 5 : Fractions issues de la NAC analysée par SDS-PAGE et coloration au bleu de Coomassie (5) ou *Western blot* révélé par un anticorps anti-V5 (6)

La masse moléculaire (kDa) et la position des marqueurs (M) sont indiquées sur la gauche.

* montre la position de deux isoformes de PAC/V5/6xHis.

Purification de PAC/V5/6xHis par CEC

L'éluion des protéines est effectuée par un gradient linéaire de NaCl de 0 à 1,6 M. La protéine PAC/V5/6xHis est éluée par 500 à 560 mM de NaCl, ce qui est cohérent avec son pI de 9,6. Une analyse par SDS-PAGE et immunodétection par un anticorps anti-V5 (Figure 4.5) a montré que les contaminants sont minoritaires. Nous pouvons observer la présence de deux bandes contenant PAC/V5/6xHis à des masses moléculaires apparentes d'environ 19,5 kDa et 19,0 kDa. Des analyses par spectrométrie de masse de type MALDI-TOF ont permis de confirmer l'identité de ces deux protéines. L'isoforme qui correspond à la bande de 19,5 kDa est présente en plus faible quantité et pourrait avoir subi une *N*-glycosylation conduisant à ce décalage de masse. Cette hypothèse est renforcée par la révélation de cette bande par la lectine GNA spécifique des *N*-glycanes mannosylés et la présence d'une seule bande de 19,1 kDa après induction de la production de PAC/V5/6xHis chez *E. coli*. Cependant, différentes analyses effectuées en spectrométrie de masse n'ont pas permis de détecter ces modifications post-traductionnelles.

Purification de PAC/V5/6xHis par NAC

Une deuxième étape de purification par NAC sur la colonne Ni-NTA a permis d'obtenir la protéine à un niveau de pureté très satisfaisant. Comme l'illustre la figure 4.5, l'analyse par SDS-PAGE et *Western blot* montrent une seule bande à 19,5 kDa. La seconde isoforme de PAC/V5/6xHis n'a pu être détectée. Le rendement de purification de PAC/V5/6xHis est d'environ 36 µg de protéine à partir de 2,7 g de paroi lyophilisée.

Cette protéine a été employée pour des tests d'interaction *in vitro*.

4.3 Réalisation et validation des puces à polysaccharides

Nous avons pris comme point de départ pour l'extraction des polysaccharides pariétaux les travaux réalisés par Moller *et al.* (2007). Ceux-ci reposaient sur l'extraction séquentielle des trois classes principales des polysaccharides chez *A. thaliana*, la pectine, les hémicelluloses et la cellulose pour générer une puce à polysaccharides. Dans notre cas, ces différents polysaccharides ont été isolés des parois des cellules des hypocotyles étiolés d'*A. thaliana* après l'extraction des protéines pariétales avec des solutions salines. Pour vérifier la qualité des trois extraits, la puce à polysaccharides a été révélée par différents anticorps monoclonaux spécifiques de certains polysaccharides pariétaux JIM5 (HG pas ou faiblement méthylestérifiées), JIM7 (HG méthylestérifiées), LM15 (XG) et LM5 (4 résidus de Gal liés en β -(1-4), typique des chaînes latérales des RGI). Une puce comportant des polysaccharides commerciaux a été réalisée en parallèle pour servir de contrôle.

La figure 4.6 montre les résultats obtenus. L'anticorps JIM5 reconnaît les fractions de polysaccharides pariétaux enrichies en pectines ainsi que tous les HG commerciaux méthylestérifiés ou non (P1-P4, cf. Figure 2.4). L'anticorps JIM7 révèle les mêmes fractions, avec un signal plus intense au niveau des pectines méthylestérifiées de pomme (P3) et de peau de citron (P4). L'anticorps LM15 révèle la fraction de polysaccharides pariétaux enrichie en hémicelluloses et le xyloglucane commercial (H1). L'anticorps LM5 reconnaît les fractions de polysaccharides pariétaux enrichies en hémicelluloses et en cellulose. Un signal est également obtenu avec le galactane (P5) et le RGI (P6). Pour bien visualiser le signal obtenu avec le RGI, il a été nécessaire d'augmenter la concentration de celui-ci à 10 mg/mL. Il faut noter que l'anticorps LM5 ne révèle pas le galactose.

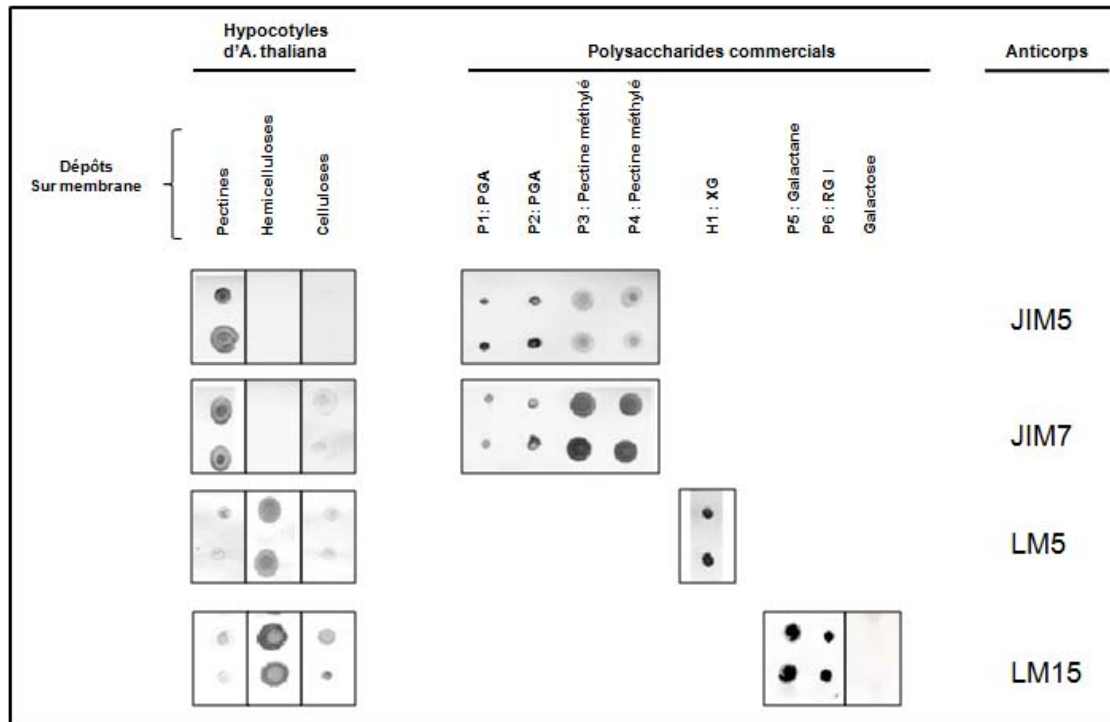


Figure 4.6 : Validation de la puce à polysaccharides

Les différents polysaccharides ont été déposés en duplicats sur la membrane de nitrocellulose : il s'agit de fractions enrichies en pectines, en hémicelluloses ou en cellulose extraites de parois d'hypocotyles étiolés d'*A. thaliana* ; ou de polysaccharides commerciaux. 1 μ L de chacun des polysaccharides a été déposé : P1, P2, P3, P4, P5 et H1 sont en solution à 1 mg/mL, P6 à 10 mg/mL, et le galactose à 33 mg/mL. Les anticorps monoclonaux JIM5, JIM7, LM15 et LM5 ont été utilisés pour révéler les membranes.

Les tests réalisés avec les anticorps monoclonaux ont validé la spécificité de ces anticorps vis-à-vis des polysaccharides commerciaux. Ils ont également permis de montrer que la fraction enrichie en pectines extraites des parois des hypocotyles étiolés d'*A. thaliana* contient des HG méthylestérifiés ou non. La fraction enrichie en hémicelluloses contient des XG et des RGI ayant des chaînes latérales de galactanes. Ces RGI n'ont probablement pas été extraits par le CDTA utilisé pour obtenir la fraction enrichie en pectines. Enfin, la fraction enrichie en cellulose contient également une faible proportion de XG et de RGI. Ces résultats sont en accord avec ceux de Moller *et al.* (2007) et confirment que nous travaillons avec des fractions enrichies en différents polysaccharides, et non pas avec des fractions pures.

L'ensemble de ces résultats valide la puce à polysaccharides et nous permet d'aborder l'étape suivante de recherche de partenaires d'AGP31 dans les parois végétales.

4.4 Recherche de partenaires d'AGP31 dans les parois végétales

Après la validation de la puce à polysaccharides, nous avons procédé aux études d'interaction d'AGP31 avec les fractions enrichies en chacune des trois principales familles de polysaccharides pariétaux et avec des polysaccharides commerciaux. Dans un premier

temps, nous avons travaillé avec la protéine native entière sur des fractions polysaccharidiques extraites de parois d'*A. thaliana*, la plante dicotylédone dont provient AGP31, et de *Brachypodium distachyon*, une plante monocotylédone. La structure des parois des dicotylédones étant très différente de celle des monocotylédones notamment concernant la nature et la proportion des pectines et des hémicelluloses (Carpita et Gibeaut, 1993), nous voulions savoir si les interactions observées étaient particulières à un type de paroi. Dans un deuxième temps, nous avons travaillé avec le domaine PAC recombinant produit dans *E. coli* ou *N. benthamiana* pour localiser le domaine d'interaction entre AGP31 et ses partenaires. Puis, nous avons essayé d'identifier plus précisément le(s) partenaire(s) d'AGP31 en utilisant des polysaccharides commerciaux. Enfin, nous avons testé des interactions entre AGP31 et son domaine PAC.

Pour l'ensemble de ces approches, nous avons utilisé les protéines AGP31, MBP/PAC/V5/6xHis et PAC/V5/6xHis décrites précédemment (cf. § 4.2). Nous avons également tiré parti de la reconnaissance spécifique d'AGP31 par la lectine PNA (cf. Chapitre 3) pour révéler les interactions de la protéine native avec ses partenaires pariétaux en utilisant la lectine PNA couplée à la digoxigénine.

4.4.1 Interactions d'AGP31 avec des polysaccharides pariétaux

Notre premier objectif consistait à savoir s'il existe des interactions entre AGP31 native et les fractions enrichies en pectines, hémicelluloses et cellulose extraites de parois. Deux types de plantes, monocotylédone (*B. distachyon*) et dicotylédone (*A. thaliana*), ont été utilisés pour ces expériences. Ainsi, des polysaccharides pariétaux extraits d'hypocotyles étiolés d'*A. thaliana* et de feuilles de *B. distachyon* ont été fixés sur la membrane de nitrocellulose. Après incubation avec la protéine AGP31, et une révélation par la lectine PNA couplée à la digoxigénine, un fort signal est détecté dans les fractions enrichies en hémicelluloses provenant à la fois d'*A. thaliana* et de *B. distachyon* (Figure 4.7). Un contrôle négatif effectué par incubation avec la lectine PNA n'a pas montré de signal au niveau de dépôt des échantillons.

Il faut noter que sur la membrane, les dépôts de cellulose laissent apparaître une tâche claire qui ne correspond pas à un signal, mais qui est due au cadoxène qui solubilise la nitrocellulose.

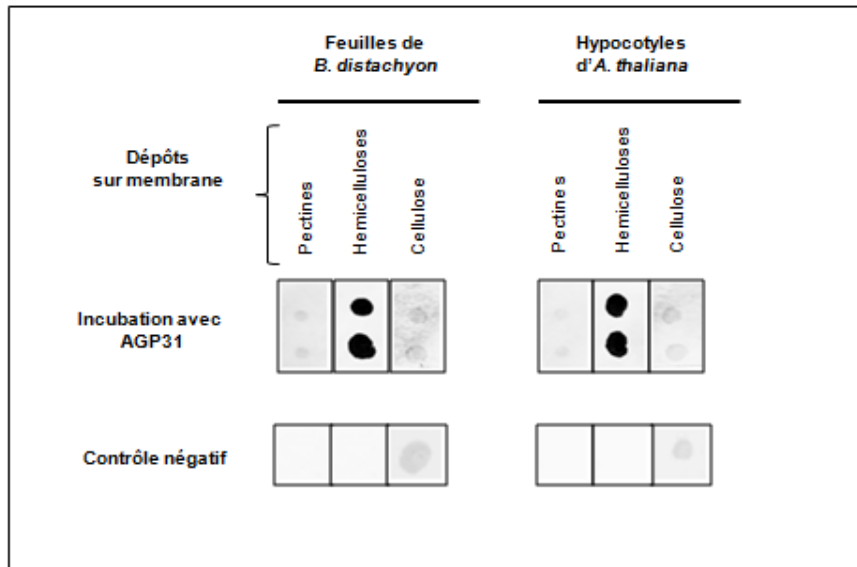


Figure 4.7 : Interactions entre AGP31 et les polysaccharides pariétaux

Un μL de chacune des fractions enrichies en pectines, hémicelluloses et $0,5\mu\text{L}$ de la fraction enrichie en cellulose extraites des parois de feuilles de *B. distachyon* ou d'hypocotyles étiolés d'*A. thaliana* ont déposés sur la membrane de nitrocellulose en duplicats. La protéine AGP31 à $20\ \mu\text{g}/\text{mL}$ ou la lectine PNA (contrôle négatif) à $10\ \mu\text{g}/\text{mL}$ ont été mises en contact avec la membrane pendant une nuit. La révélation des signaux a été effectuée avec la lectine PNA couplée à la digoxigénine.

Bien qu'un grand nombre de protéines aient été préalablement extraites des parois avec des solutions salines, il est possible que certaines protéines soient présentes dans la fraction enrichie en hémicelluloses. Afin de valider le fait que des polysaccharides sont à l'origine des signaux observés, un contrôle supplémentaire a été effectué. Une incubation de la fraction enrichie en hémicelluloses en présence de différentes concentrations de protéinase K a été réalisée pendant 1 h. La figure 4.8 décrit les résultats de cette expérience. Les fractions enrichies en hémicelluloses sont reconnues de la même manière par AGP31, en présence et en absence de digestion par la protéinase K. Ceci montre qu'AGP31 reconnaît bien des polysaccharides présents dans ces fractions, et non pas des protéines. Notons qu'un contrôle d'activité de la protéinase K dans le tampon d'extraction des hémicelluloses ($\text{NaOH}\ 4\ \text{M}$) montre que la protéase reste active dans ces conditions.

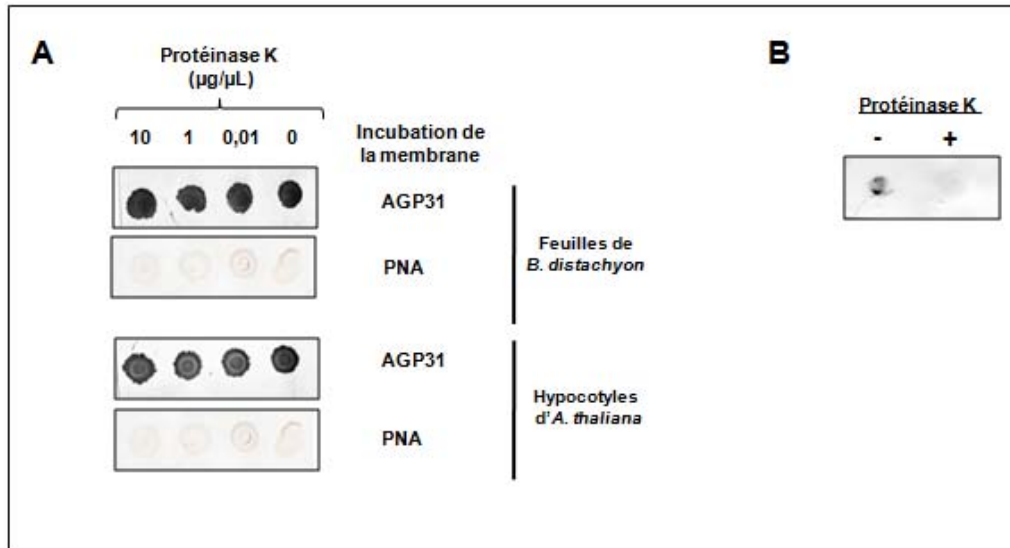


Figure 4.8 : Interactions entre AGP31 et des fractions enrichies en hémicelluloses digérées par la protéinase K

A. Des hémicelluloses extraites de parois de feuilles de *B. distachyon* et d'hypocotyles étioilés d'*A. thaliana* ont été incubées en présence de protéinase K pendant 1 h à différentes concentrations (0, 0,01, 1 et 10 µg/µL de protéinase K). Un µL de chacun de ces échantillons est déposé sur la membrane de nitrocellulose et mis en contact avec AGP31 ou la lectine PNA (contrôle négatif) à 10 µg/mL. La révélation des interactions a été effectuée avec la lectine PNA couplée à la digoxigénine. **B.** Contrôle d'activité de la protéinase K en présence de la solution de NaOH 4 M utilisée pour l'extraction des hémicelluloses. La MBP, digérée par 10 µg/µL de protéinase K (+) ou non (-) est révélée par un anticorps anti-MBP.

Afin de préciser la nature du composé responsable de l'interaction avec AGP31, nous avons réalisé des tests d'interaction avec des polysaccharides commerciaux de différentes origines en utilisant 1 µL d'une solution à 1 mg/mL (Figure 4.9). Aucun signal n'a été observé avec les hémicelluloses dont nous disposons, XG (H1), β -glucane (H2), xylane (H3-H4), arabinoxylane (H5), ni avec la cellulose (C). Un faible signal a été révélé avec les pectines de type HG (P1-P2) et les pectines méthylestérifiées (P3-P4). Lorsque la concentration en pectines méthylestérifiées (P3) est augmentée à 10 mg/mL, un signal plus net apparaît, ce qui n'est pas le cas avec les pectines P1, P2 et P4 ni avec les hémicelluloses H1 à H5 déposées à forte concentration. Par contre, un signal fort a été observé avec le galactane et faible avec le RGI. Lorsque la concentration en RGI est de 10 g/mL, un signal fort est également détecté. Enfin, aucun signal n'a été mis en évidence avec le Gal (résultats non montrés). Notons qu'un contrôle négatif réalisé par incubation avec du PNA n'a montré aucun signal. A l'inverse, un contrôle positif réalisé avec un dépôt d'AGP31 a détecté un fort signal, validant ainsi la méthode de détection.

L'ensemble de ces résultats suggère des interactions entre AGP31 et les pectines méthylestérifiées qui pourraient avoir lieu au niveau du domaine riche en His, chargé positivement au pH légèrement acide de la paroi. Les plus fortes interactions sont cependant observées avec les galactanes et le RGI qui comporte des ramifications de galactanes. L'absence d'interactions avec le Gal suggère que les interactions requièrent la présence de résidus Gal en chaîne.

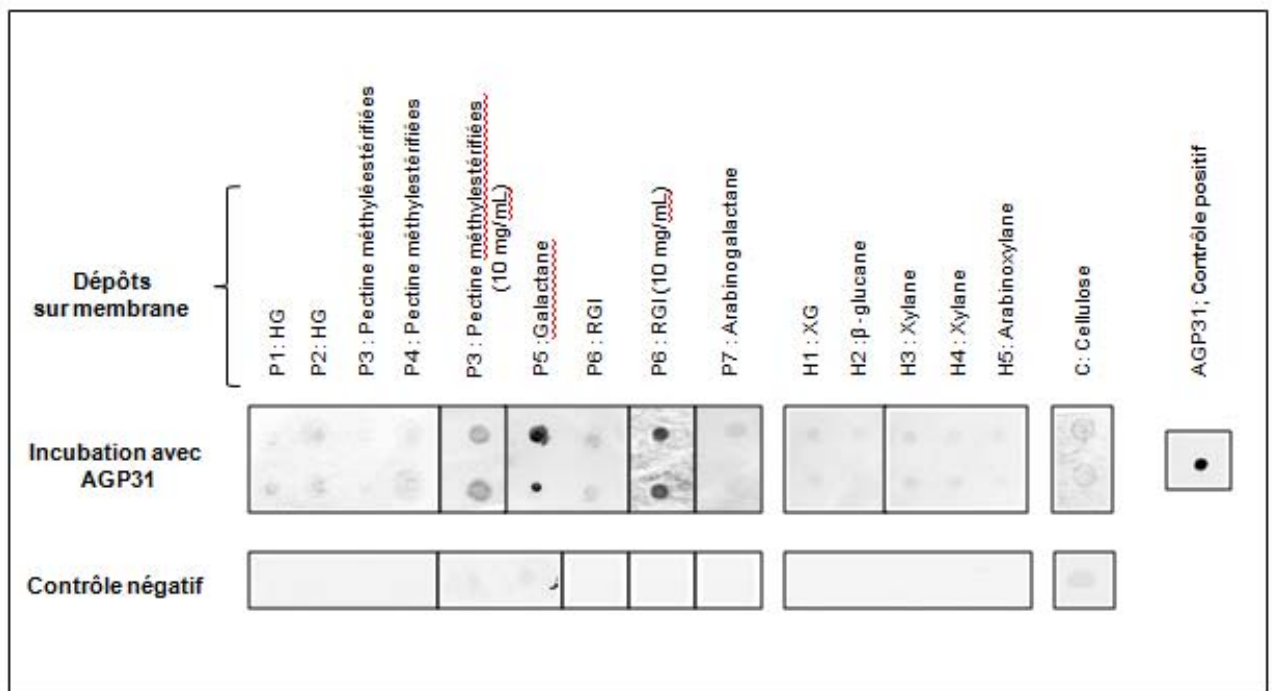


Figure 4.9 : Interactions entre AGP31 et des polysaccharides commerciaux

Un μL d'une solution à 1 mg/mL (ou 10 mg/mL pour P3 et P6) de chacun des polysaccharides commerciaux a été déposé sur la membrane de nitrocellulose en duplicats et mis en contact avec la protéine AGP31 ou la lectine PNA (contrôle négatif) à 20 $\mu\text{g/mL}$. La révélation est effectuée avec la lectine PNA couplée à la digoxigénine. Un témoin positif est réalisé par fixation de la protéine AGP31 à 1 $\mu\text{g}/\mu\text{L}$ sur la membrane.

4.4.2 Interactions du domaine PAC d'AGP31 avec des polysaccharides pariétaux

4.4.2.1 Tests d'interaction réalisés avec MBP/PAC/V5/6xHis produite dans *E. coli*

Nous avons utilisé la protéine de fusion MBP/PAC/V5/6xHis produite dans *E. coli* et purifiée pour réaliser ces tests (cf. § 4.2.2.1). Les contrôles ont été réalisés avec la protéine MBP qui a servi de témoin négatif. Dans la figure 4.10, un signal présent au niveau des fractions enrichies en polysaccharides pariétaux et absent sur le contrôle négatif n'est détecté que sur la fraction enrichie en hémicelluloses. Les autres fractions montrent un signal fort même sur le contrôle négatif ce qui atteste de possibles interactions entre la MBP et les polysaccharides des fractions enrichies en pectines et cellulose. Cette expérience préliminaire réalisée avec la protéine de fusion MBP/PAC/V5/6xHis a permis de mettre en évidence une interaction entre le domaine PAC et les extraits pariétaux enrichis en hémicelluloses. Il faut noter que les interactions observées avec la fraction enrichies en hémicellulose ne semblent pas nécessiter la *N*-glycosylation de la protéine qui n'est pas réalisée dans *E. coli*.

Malheureusement, du fait des interactions de MBP avec les fractions enrichies en pectines et en cellulose, nous ne pouvons pas aller plus loin avec cette forme du domaine PAC. Nous avons donc poursuivi cette étude avec la protéine de fusion PAC/V5/6xHis produite dans des feuilles de *N. benthamiana*.

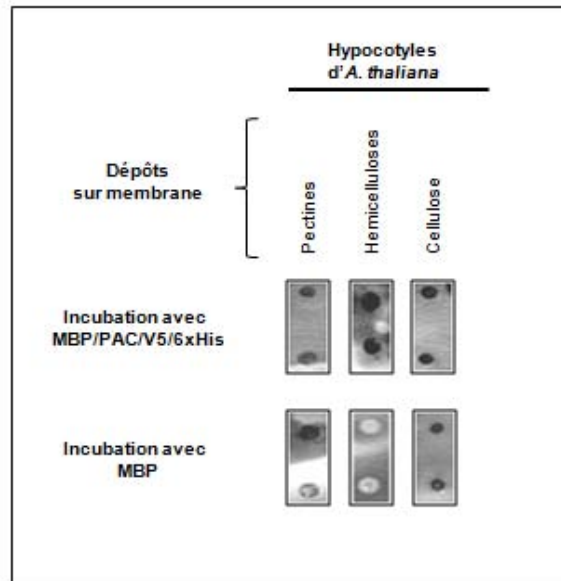


Figure 4.10 : Interactions entre MBP/PAC/V5/6xHis et les polysaccharides pariétaux extraits d'hypocotyles étiolés d'*A. thaliana*

Un μL de chacune des fractions enrichies en polysaccharides pariétaux a été fixé sur la membrane de nitrocellulose en duplicats et mis en contact avec MBP/PAC/V5/6xHis à $17 \mu\text{g/mL}$ ou MBP (contrôle négatif) à $10 \mu\text{g/mL}$. La révélation a été effectuée avec un anticorps anti-MBP.

4.4.2.2 Tests d'interaction réalisés avec PAC/V5/6xHis produite dans *N. benthamiana*

Les tests d'interaction décrits dans ce paragraphe ont été réalisés avec la protéine de fusion PAC/V5/6xHis extraite des feuilles de *N. benthamiana* et purifiée (cf. § 4.2.2.2). Différentes fractions enrichies en pectines, hémicelluloses ou cellulose ont été extraites de parois d'hypocotyles étiolés d'*A. thaliana* ou de feuilles de *B. distachyon*. Différents polysaccharides commerciaux ont également été employés. La figure 4.11 montre une interaction de PAC/V5/6xHis avec les fractions enrichies en hémicelluloses d'*A. thaliana* et de *B. distachyon*, le galactane et le RGI. Contrairement à ce qui a été observé avec la protéine native, aucune interaction n'a été observée avec les pectines méthylestérifiées, même à une concentration de 10 mg/mL . Notons qu'aucune interaction avec des résidus Gal n'a été détectée. Ceci suggère que le domaine PAC d'AGP31 interagit avec les polysaccharides pariétaux et plus particulièrement avec des chaînes de β -(1-4) galactane qui pourraient être des ramifications du RGI.

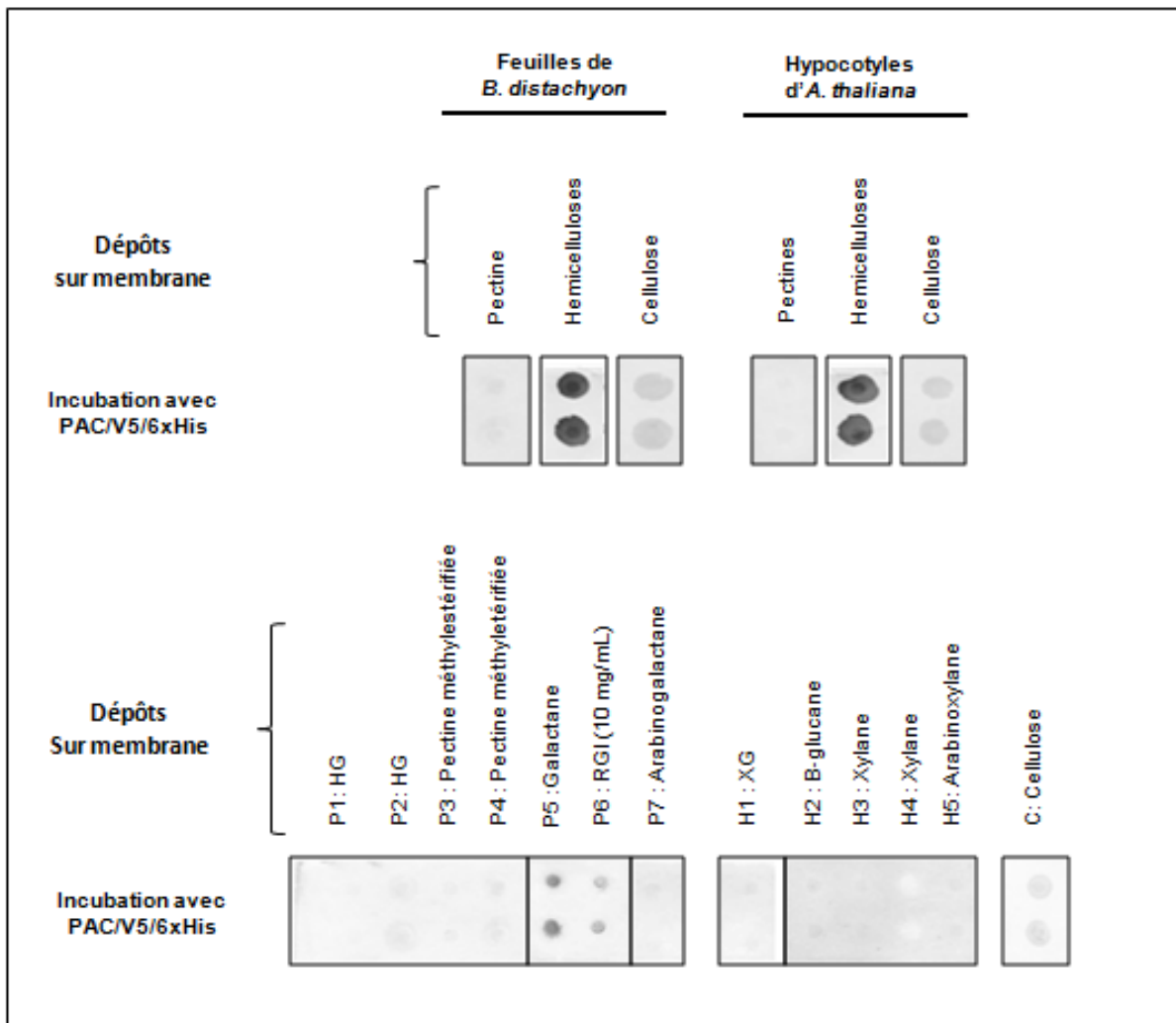


Figure 4.11 : Interactions entre PAC/V5/6xHis et des polysaccharides pariétaux

Un μL de chacune des fractions enrichies en polysaccharides pariétaux et de chacun des polysaccharides commerciaux à 1 mg/mL (ou 10 mg/mL pour RGI) a été fixé sur la membrane de nitrocellulose en duplicats et mis en contact avec PAC/V5/6xHis à 40 $\mu\text{g/mL}$. La révélation a été effectuée avec un anticorps anti-V5

4.4.3 Spécificité des interactions entre AGP31 et le galactane et rôle du domaine PAC

Pour vérifier la spécificité des interactions observées entre les polysaccharides présents dans la fraction enrichie en hémicelluloses, nous avons essayé de supprimer ces interactions en effectuant des tests de compétition avec du galactane commercial. La protéine native AGP31 a été utilisée pour ces tests (cf. § 4.2.1). Elle a été mise en contact avec du galactane ou du Gal avant l'incubation avec la puce à polysaccharides. Quatre concentrations différentes de galactane ont été testées : 1, 5, 10 et 33 mg/mL (Figure 4.12). Les interactions entre AGP31 et la fraction enrichie en hémicelluloses, le galactane et le RGI sont complètement inhibées par 10 mg/mL de galactane. Par contre, l'incubation d'AGP31 avec du Gal, même à forte concentration (100 mg/mL) n'a pas éteint le signal (résultat non montré), ce qui renforce l'idée que AGP31 reconnaît des résidus Gal en chaîne.

Cette même expérience a ensuite été réalisée avec la protéine de fusion PAC/V5/6xHis produite dans des feuilles de *N. benthamiana* (cf. § 4.2.2.2). La figure 4.13 montre la perte du signal lors de la compétition avec le galactane à 10 mg/mL, tandis que le signal persiste lors de l'incubation avec du Gal à 33 mg/mL (résultat non montré). L'incubation de PAC/V5/6xHis avec un agent réducteur, le β -mercaptoéthanol, ne semble pas avoir influencé les interactions. Nous observons que bien que le domaine PAC soit un domaine d'AGP31, il peut exercer sa fonction en tant que protéine à part entière. Cette expression en système hétérologue est accompagnée du bon repliement de la protéine, nécessaire à sa fonction telle que nous pouvons la mettre en évidence *in vitro*.

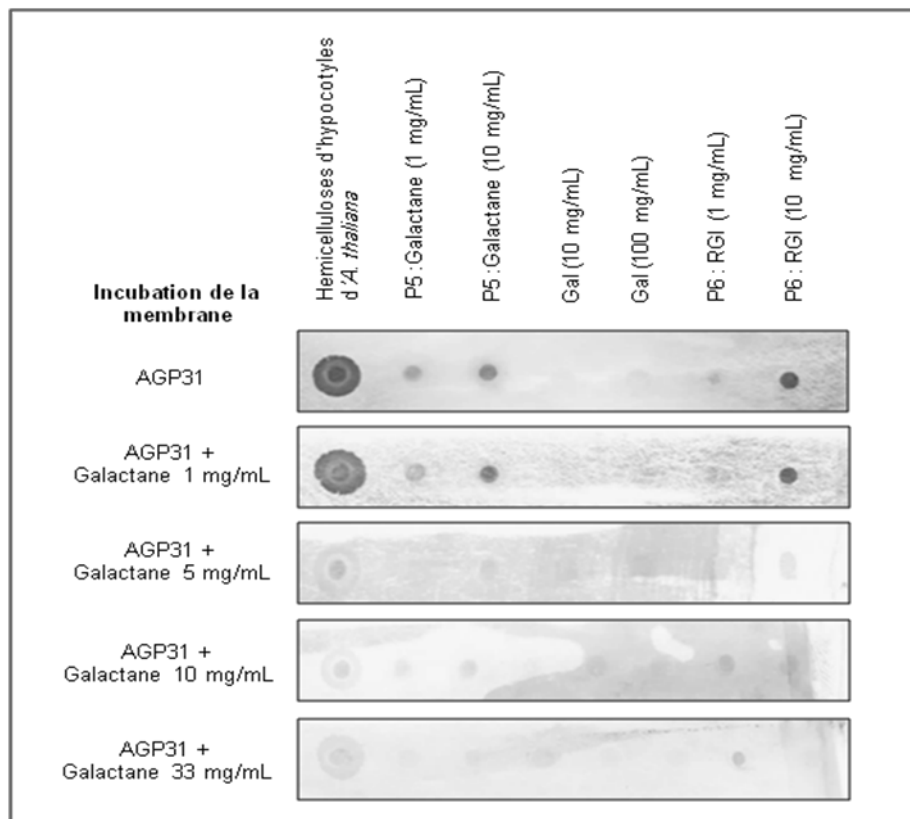


Figure 4.12 : Mise en compétition de l'interaction entre AGP31 et les polysaccharides pariétaux
Des dépôts de la fraction enrichie en hémicelluloses extraites des parois d'hypocotyles étiolés d'*A. thaliana*, de galactane (1 et 10 mg/mL), de RGI (1 et 10 mg/mL) et de Gal (10 et 100 mg/mL) ont été réalisés. Une préincubation d'AGP31 à 10 mg/mL avec du galactane a été réalisée à différentes concentrations en galactane (1, 5, 10 et 33 mg/mL) pendant 1 h avant le test d'interaction sur la membrane. La révélation a été effectuée par la lectine PNA couplée à la digoxigénine.

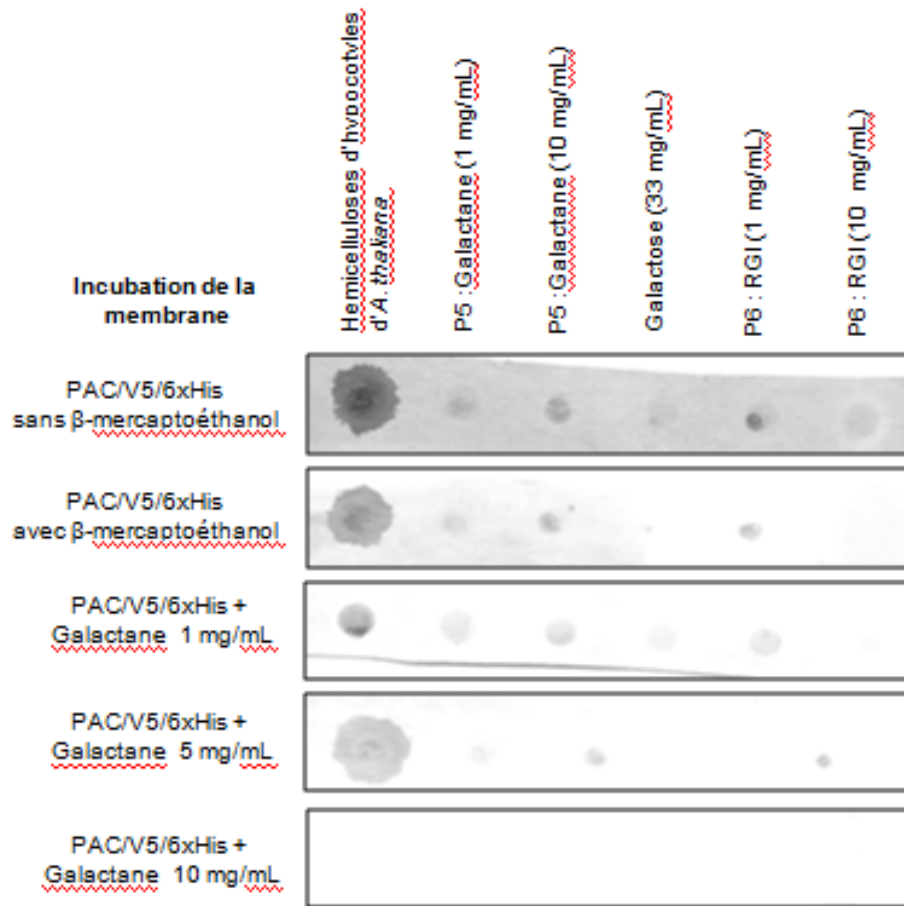


Figure 4.13 : Mise en compétition de l'interaction entre PAC/V5/6xHis et les polysaccharides pariétaux

Des dépôts de la fraction enrichie en hémicelluloses extraites des parois d'hypocotyles étiolés d'*A. thaliana*, de galactane (1 et 10 mg/mL), de RGI (1 et 10 mg/mL) et de Gal (100 mg/mL) ont été réalisés. La protéine de fusion PAC/V5/6xHis à 30 µg/mL a été incubée avec ou sans β-mercaptoéthanol ou une préincubation de PAC/V5/6xHis avec du galactane a été réalisée à différentes concentrations en galactane (1, 5, 10 et 33 mg/mL) pendant 1 h avant le test d'interaction sur la membrane. La révélation a été effectuée par un anticorps anti-V5.

4.4.4 Mise en évidence d'interactions entre AGP31 et le domaine PAC

Nous avons cherché à savoir si le domaine PAC pourrait interagir avec AGP31. Pour cela, l'interaction entre la protéine de fusion PAC/V5/6xHis produite dans des feuilles de *N. benthamiana* (cf. § 4.2.2.2) et AGP31 native (cf. § 4.2.1) a été testée dans les deux sens : (i) la protéine AGP31 a été fixée sur la membrane de nitrocellulose et mise en contact avec le domaine PAC ; (ii) le domaine PAC a été fixé sur la membrane de nitrocellulose et mis en contact avec AGP31. Un signal est observé dans les deux cas, ce qui nous a permis de conclure à une interaction entre le domaine PAC et la protéine entière (Figure 4.14).

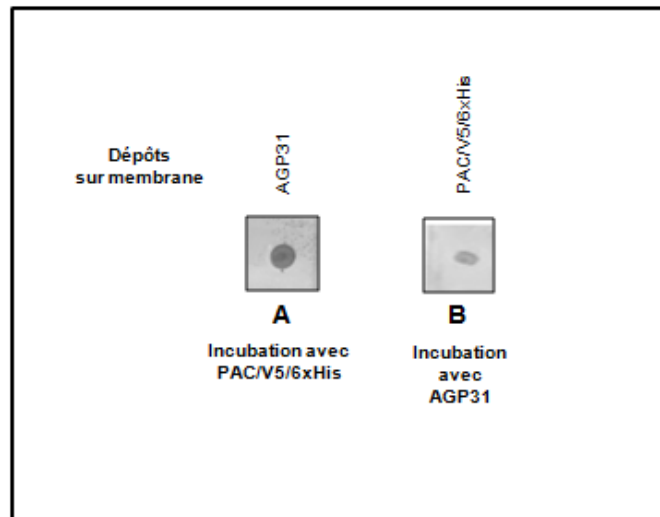


Figure 4.14 : Interactions entre la protéine AGP31 et le domaine PAC

La protéine AGP31 native a été utilisée pour ce test d'interaction ainsi que la protéine de fusion PAC/V5/6xHis. Un μg de chacune d'elles a été fixé sur la membrane. Les interactions ont été réalisées avec des solutions de protéines à $10 \mu\text{g}/\text{mL}$. La membrane A a été révélée par un anticorps anti-V5. La membrane B a été révélée par la lectine PNA couplée à la digoxigénine.

4.5 Discussion et perspectives

Afin d'apporter des éléments nouveaux sur la fonction de la protéine AGP31, nous avons recherché des partenaires polysaccharidiques et protéiques d'AGP31 et de son domaine PAC en réalisant des expériences d'interaction *in vitro* protéine/polysaccharides ou protéine/protéine sur membranes de nitrocellulose.

L'étape préliminaire à cette étude a consisté à produire et purifier la protéine AGP31 et son domaine PAC. Différentes stratégies (schématisées sur la figure 4.2) ont été mises en œuvre, avec plus ou moins de succès. Ainsi, nous avons pu mettre au point différents procédés de production/purification robustes et reproductibles permettant d'obtenir des quantités suffisantes de protéines pour la réalisation des tests d'interaction. Trois constructions ont été testées (MBP/PAC/V5/6xHis exprimée chez *E. coli*, PAC/V5/6xHis exprimée chez *N. benthamiana* et AGP31 exprimée chez *A. thaliana*), permettant de comparer les patrons d'interaction de la protéine AGP31 native et de son domaine PAC et d'apprécier l'influence de possibles modifications post-traductionnelles de ce domaine.

Les résultats des tests d'interaction nous ont permis de montrer que la protéine AGP31 ainsi que son domaine PAC interagissent avec la fraction enrichie en hémicelluloses extraite à partir de jeunes feuilles de *B. distachyon* (monocotylédones) et d'hypocotyles étiolés d'*A. thaliana* (dicotylédones). La recherche dans les bases de données de *B. distachyon* (<http://www.phytozome.net/>) a permis d'identifier un certain nombre de protéines contenant le domaine PAC. L'alignement multiple de séquences a été établi par le logiciel *Clustalw* et nous a permis d'identifier des protéines à domaine PAC putatif ayant un peptide signal prédit, une petite taille (169-209 acides aminés) et 6 Cys conservées qui sont Bradi2g48680, Bradi2g18920, Bradi2g48670, Bradi2g18910, Bradi2g07610, Bradi2g31950, Bradi3g24010, Bradi5g08410, Bradi2g03930, Bradi3g32620.

Ceci atteste que le domaine C-terminal d'AGP31, *i.e.* le domaine PAC, est responsable de cette interaction. Cependant, les essais d'interaction réalisés avec des polysaccharides commerciaux ne montrent aucune interaction avec les hémicelluloses commerciales testées. A l'inverse, une interaction d'AGP31 et du domaine PAC avec des galactanes a été montrée. Ceux-ci pourraient correspondre à des ramifications du RGI, en accord avec le signal d'interaction observé en présence d'une plus forte concentration de RGI. Cette hypothèse soulève la question du contenu de la fraction enrichie en hémicelluloses. L'anticorps LM5 spécifique des galactanes montre qu'on a extrait avec la fraction enrichie en hémicelluloses une fraction importante de chaînes de galactane, en accord avec l'interaction avec le galactane observée. Ceci corrobore aussi les données de Moller et al. (2007) qui ont montré que cette procédure d'extraction des polysaccharides permettait d'obtenir des fractions enrichies en une classe de polysaccharides, mais pas de les séparer finement. Par ailleurs, des données de la littérature montrent que l'extraction des polysaccharides par le CDTA et le Na₂CO₃ entraîne les HG mais aussi les chaînes de RGI ayant des ramifications par des chaînes de galactanes (Schroder et Atkinson, 2006). Par contre il a été montré qu'un traitement alcalin entraîne une importante dégradation des régions HG et préserve les régions RGI substituées par de larges chaînes de galactanes (Zykwinska et al., 2006). Ces données confirment notre hypothèse selon laquelle l'interaction observée entre AGP31 ou PAC et la fraction enrichie en hémicelluloses est due aux galactanes substituant le RGI. Pour compléter ce travail, il serait intéressant de faire une composition en monosaccharides de la fraction enrichie en hémicelluloses afin de déterminer la proportion des résidus Gal la composant.

Par ailleurs, nous avons mis en évidence une interaction entre la protéine AGP31 et le domaine PAC. La protéine AGP31 étant galactosylée au niveau de son domaine PRP (*cf.* Chapitre 3), nous pouvons suggérer que ses motifs *Hyp-O-Gal-rich* permettent d'établir une interaction entre son domaine PRP et son domaine PAC. L'interaction entre PAC et les chaînes de Gal ne serait pas spécifique du type de liaisons entre les résidus Gal. En effet, elle pourrait reconnaître des chaînes β de $-(1,3)(1,6)$ -Gal supposées pour le domaine PRP d'AGP31, par analogie avec GaRSGP (Sommer-Knudsen et al., 1996). Les galactanes des ramifications du RGI sont quant à eux caractérisés par des liaisons $-(1,4)$. Il serait maintenant utile de pouvoir quantifier ces interactions afin de déterminer l'affinité du domaine PAC vis-à-vis de galactanes de différentes structures pour affiner la spécificité de reconnaissance de ce domaine.

Enfin, il a été suggéré que des *stretches* d'His pourraient former un site de liaison des métaux ou former des ponts salins avec les groupes carboxyles portés par les résidus galacturonosyles des pectines (Fry, 1988). Une interaction a été suggérée entre AGP31 et des pectines méthylestérifiées, non observée avec le domaine PAC seul. Cette interaction pourrait impliquer le domaine riche en His présent uniquement sur la forme entière d'AGP31. Cette hypothèse a été aussi suggérée par Baldwin et al. (1993). Cependant, dans notre cas, une interaction entre AGP31 et des pectines méthylestérifiées à une concentration élevée a été montrée. Il serait intéressant d'approfondir l'étude des interactions entre AGP31 et des HG de structure et de degré de méthylestérification contrôlés afin de confirmer la nature électrostatique des interactions. De plus, disposer d'une forme de la protéine délétée du domaine riche en His permettrait de valider l'implication de ce domaine.

Un point central mis en évidence par cette série d'expériences concerne le fait que le domaine PAC semble fonctionnel sans le reste de la protéine. En effet, il est capable de lier les galactanes, issus des RGI et/ou des *O*-glycosylations du domaine PRP d'AGP31. Il reste à

savoir quelles régions du domaine PAC sont impliquées dans l'interaction. La protéine PAC présente deux sites putatifs de *N*-glycosylation. Cependant, la *N*-glycosylation ne semble pas nécessaire à l'interaction puisqu'une expression de PAC dans *E. coli* n'a pas empêché l'interaction. De plus, la protéine PAC présente 6 Cys conservées. La persistance en conditions réductrices de l'interaction entre PAC et les polysaccharides (résultat non montré) laisse penser que si les ponts disulfures existent, ils n'ont aucun effet sur l'interaction. Afin de comprendre le rôle éventuel de ces Cys conservées et d'élucider leur contribution éventuelle dans l'interaction, nous pourrions remplacer ces résidus Cys par d'autres acides aminés par mutagenèse dirigée et déterminer le profil d'interaction de la protéine mutée. La capacité du domaine PAC à lier des polysaccharides soulève la question fondamentale de l'organisation multi-domaines d'AGP31, alors qu'il existe dans les parois d'*A. thaliana* des protéines composées uniquement du domaine PAC. A ce jour, rien n'est connu sur l'activité biochimique de ce domaine ni sur les protéines à Cys conservées qui ont des structures similaires.

L'analyse *in silico* d'AGP31 a montré que son domaine central PRP est dépourvu d'éléments de structure secondaire, contrairement au domaine PAC qui est bien structuré. Il est possible qu'une telle structure soit un avantage lui permettant de se lier avec un grand nombre de cibles. Parmi les hypothèses avancées sur les propriétés spécifiques des protéines déstructurées, leur flexibilité locale ou totale est souvent citée par les biologistes (Gohlke et al., 2004; Radivojac et al., 2004; Rambo et Tainer, 2011). Cette particularité leur permet de modifier leur structure en réponse à des cibles moléculaires, favorisant alors la capacité d'une protéine donnée d'interagir avec de nombreux partenaires cellulaires différents. Ceci permet donc une régulation fine et transitoire des affinités de fixation lors de processus biologiques, ce qui offre d'importants avantages dans les voies de signalisation et de régulation cellulaires.

Les résultats obtenus dans ce chapitre nous ont permis d'identifier *in vitro* certains partenaires d'AGP31, de définir certaines régions d'interactions d'AGP31 et de montrer des spécificités de reconnaissance. Au regard de l'ensemble de ces résultats, nous pouvons conclure qu'AGP31 est une protéine au cœur d'un réseau d'interactions avec des polysaccharides impliquant ses différents domaines. La possibilité d'un auto-assemblage d'AGP31 pour former un réseau dans la paroi est également suggérée. Il reste maintenant à décrypter plus finement les bases moléculaires de ces interactions et à corrélérer ces données *in vitro* aux mécanismes d'assemblages macromoléculaires qui prennent place dans la paroi.

Chapitre 5

***CARACTERISATION FONCTIONNELLE IN
PLANTA DU GENE AT1G28290 CODANT
AGP31***

5. CARACTERISATION FONCTIONNELLE IN PLANTA DU GENE AT1G28290 CODANT AGP31

Outre l'étude structurale de la glycoprotéine AGP31 (*cf.* Chapitre 3) et la mise en évidence d'interactions avec des polysaccharides pariétaux (*cf.* Chapitre 4), ce travail comporte un troisième volet consistant à comprendre sa fonction *in planta*. Partant du fait que nous avons trouvé cette protéine en abondance dans des hypocotyles étiolés, nous nous sommes intéressés plus particulièrement aux phases précoces du développement des plantules d'*A. thaliana*.

Nous avons orienté nos recherches de l'élucidation de la fonction d'*At1g28290* selon une approche génétique. Dans un premier temps, nous avons cherché à caractériser le patron d'expression d'*At1g28290* en combinant des approches bioinformatiques, de RT-PCR, d'hybridation *in situ* et d'immunocytochimie. Des mutants ARNi (ARN interférence) et des mutants sur-expressseurs affectés dans l'expression d'*At1g28290* ont été générés. Les plantes transgéniques ont été utilisées pour la recherche de phénotypes au cours du développement.

5.1 Caractéristiques de la croissance des plantules étiolées d'*A. thaliana* et particularités de la mise en place des parois cellulaires

La croissance est l'ensemble des processus biologiques qui conduisent à une augmentation irréversible des dimensions et de la masse d'un individu. Cette croissance associe trois phénomènes : (i) la division cellulaire (mitose au niveau des méristèmes), (ii) le grandissement cellulaire représenté par une augmentation de volume de la cellule dans toutes les directions (croissance isodiamétrique) ou dans une direction précise (élongation cellulaire) et (iii) la différenciation cellulaire. Au cours de ces processus, la paroi évolue fortement du point de vue de sa composition, mais aussi de sa structure.

5.1.1 Développement des plantules d'*A. thaliana* à l'obscurité

A l'obscurité, le développement des plantules d'*A. thaliana* est marqué par un très fort allongement de l'hypocotyle. L'hypocotyle est l'organe qui relie les cotylédons à la racicule. Il se caractérise par une extrémité constituée d'une crosse apicale supportant des cotylédons jaunes et fermés (Figure 5.1-A). Il présente une organisation cellulaire simple, en partie similaire à celle des racines (Figure 5.1-B). Il est constitué d'un épiderme, de deux couches de cellules corticales, d'un endoderme et d'un cylindre central dans lequel se trouvent les tissus conducteurs

L'hypocotyle est formé au cours de l'embryogenèse. Sa croissance est rapide au cours de la germination. Elle est essentiellement due à un phénomène d'élongation cellulaire chez *A. thaliana* et ne fait pas intervenir de divisions cellulaires (Cosgrove, 1994; Wei et al., 1994; Reid et Howell, 1995; Gendreau et al., 1997; Smalle et al., 1997; Collett et al., 2000; Saibo et al., 2003; Refrégier et al., 2004). A l'obscurité, l'hypocotyle s'allonge dix fois plus que lorsque les plantules sont cultivées à la lumière (Gendreau et al., 1997).

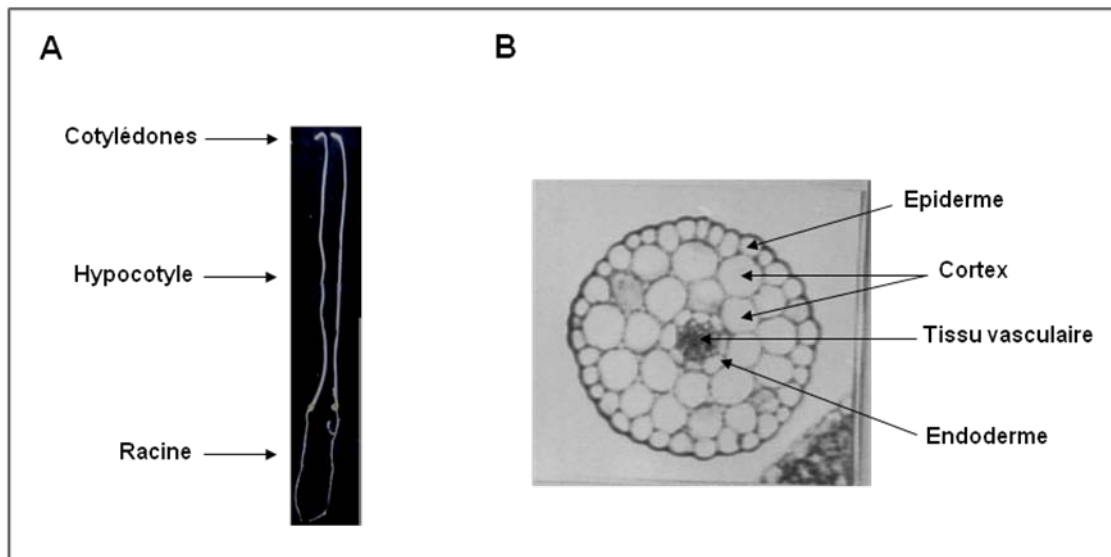


Figure 5.1 : Développement d'*A. thaliana* à l'obscurité

A. Phénotype étiolé d'une plantule d'*A.thaliana* ayant germé à l'obscurité. **B.** Coupe transversale de l'hypocotyle. Adapté de Gendreau et al. (1997).

5.1.2 Mise en place de la paroi des cellules des hypocotyles lors de l'élongation cellulaire

Dans l'hypocotyle en élongation, l'épaississement de la paroi est extrêmement complexe et dynamique et dépend du type des cellules et des conditions de croissance (Gendreau et al., 1997; Derbyshire et al., 2007). La paroi des cellules est mise en place en deux phases successives séparées au cours du temps : (i) une synthèse active de constituants pariétaux ayant pour conséquence une augmentation de l'épaisseur de la paroi ; puis (ii) une phase de réorganisation de la paroi ayant pour conséquence la réduction de l'épaisseur de la paroi.

La synthèse des constituants pariétaux prend place durant la phase tardive de l'embryogenèse ou après l'imbibition des graines, avant l'accélération de la croissance de la jeune plantule (Gendreau et al., 1997). Cependant, l'activité transcriptionnelle de gènes impliqués dans la biogenèse des parois est maintenue au cours de la croissance de l'hypocotyle en conditions d'étiollement (Jamet et al., 2009). Au début de la phase d'élongation, la réduction de l'épaisseur de la paroi commence dans les cellules de la base de l'hypocotyle tandis que les cellules du sommet ont une paroi épaisse. Ensuite le processus de l'élongation progresse graduellement vers le sommet (direction acropétale) pour couvrir l'ensemble de l'hypocotyle (Gendreau et al., 1997 ; Derbyshire et al., 2007).

Les propriétés de la paroi doivent nécessairement subir des évolutions en fonction du stade du développement. L'élongation cellulaire nécessite le relâchement de la paroi primaire pour la rendre extensible. Sous la pression de turgescence, les composés pariétaux glissent les uns sur les autres tandis que d'autres constituants viennent renforcer la structure et ceci grâce à plusieurs types de protéines pariétales (Cosgrove, 2005). En raison de la complexité de la structure de la paroi, il existe peu de données claires sur les mécanismes moléculaires qui sont responsables de la réorganisation de ces constituants. Chez *A. thaliana*, les hémicelluloses principalement composées de xyloglucanes, peuvent être la cible des *xyloglucan endotransglucosylases/hydrolases* (XTH). Les XTH rompent les liaisons entre les glucanes et la cellulose grâce à leur double activité de xyloglucane endotransglucosylase (XET) et

endohydrolase (XEH) (Rose et al., 2002; Cosgrove, 2005). Les endo 1,4- β -glucanases, nommées également cellulases, réalisent la dégradation des régions non cristallines de la cellulose (Ohmiya et al., 1995; Ohmiya et al., 2000). Le relâchement de la paroi primaire est aussi sous le contrôle de protéines pH dépendantes, appelées expansines. Les expansines n'hydrolysent pas les polymères de la paroi, mais jouent un rôle dans l'affaiblissement des liaisons hydrogène entre les hémicelluloses et les microfibrilles de cellulose, ce qui permet le glissement des microfibrilles les unes par rapport aux autres (McQueen-Mason et Cosgrove, 1994; Cosgrove, 2005). Les pectine méthylestérases (PME) sont des enzymes dont l'impact sur la rigidité pariétale dépendrait de la présence d'autres effecteurs. Lorsque les HG sont déposés au niveau de la paroi primaire, ils sont méthylestérifiés à raison de 70% à 80% (Ridley et al., 2001). Les PME catalysent l'hydrolyse des groupements méthylester permettant ainsi des interactions ioniques avec des ions calcium, la formation de boîtes à oeufs et la rigidification de la paroi cellulaire (cf. § 1.2.1.3.1). Ainsi, l'action des PME bloque l'expansion du tube pollinique et des hypocotyles (Jiang et al., 2005; Derbyshire et al., 2007; Rockel et al., 2008). La sur-expression d'un inhibiteur de PME (PMEI4) conduit à un retard d'accélération de croissance des hypocotyles en conditions d'étiollement (Pelletier et al., 2010).

Plusieurs travaux ont montré l'implication des AGP dans l'expansion cellulaire. L'ajout de β -glucosyl Yariv (cf. § 1.3.2) à une suspension cellulaire de carotte bloque l'élongation des cellules (Willats et Knox, 1996). Chez *A. thaliana*, l'application de β -glucosyl Yariv entraîne une réduction de la croissance en longueur des racines et plus particulièrement une réduction de l'élongation des cellules épidermiques dont la croissance devient radiale (Willats et Knox, 1996). De même, l'AGP1 de *Physcomitrella patens* joue un rôle dans l'expansion des cellules apicales (Lee et al., 2005). Des mutants d'*A. thaliana* ayant une extension réduite des hypocotyles (Takahashi et al., 1995) et une croissance racinaire réduite (Ding et Zhu, 1997) présentent des niveaux réduits ou altérés d'AGP. Le mutant *murl* d'*A. thaliana* ayant des AGP non fucosylées présente aussi une réduction de la croissance de la racine (van Hengel et Roberts, 2002). Enfin, le gène *CsAGP1*, (*Giberellin-Responsive Gene*) intervient dans l'élongation des cellules de l'hypocotyle du concombre tandis que la surexpression ectopique de ce gène chez *N. tabacum* induit la croissance de la plante, suggérant un rôle dans l'élongation de la tige (Park et al., 2003).

5.1.3 Stratégies mises en place pour l'étude fonctionnelle d'*At1g28290*

Pour mieux comprendre la fonction d'*At1g28290*, nous avons pensé qu'il était important de définir son patron d'expression dans la plante au cours du développement. Notre première approche a consisté à obtenir des plantes transgéniques comportant une construction « promoteur *At1g28290* (p*At1g28290*)-GFP-GUS ». Comme nous le verrons, cette approche n'ayant pas donné des résultats cohérents, nous nous sommes tournés vers des approches de PCR quantitatives, d'hybridation d'ARN *in situ* et des recherches dans les banques de données transcriptomiques pour définir le patron d'expression d'*At1g28290* au cours de la germination.

Une deuxième approche génétique a consisté à obtenir des plantes transgéniques sous-exprimant (plantes ARNi) ou sur-exprimant (sous contrôle du promoteur 35S du CaMV) *At1g28290*. Des lignées homozygotes ont été isolées et caractérisées pour rechercher un phénotype au cours du développement.

5.2 Analyse du patron d'expression d'*At1g28290* au cours du développement

5.2.1 Obtention et analyse de plantes *pAt1g28290::GFP::GUS*

Peu avant le démarrage de ce travail de thèse, une autre équipe de recherche a publié des données sur le gène *At1g28290* (Liu et Mehdy, 2007). Des plantes transgéniques exprimant le gène rapporteur GUS sous le contrôle du promoteur de *At1g28290* (1,4 kb) ont été produites. L'analyse des plantes transgéniques de 1 à 4 semaines révèle que *pAt1g28290(1,4 kb)::GUS* a un patron d'expression spécifique des tissus conducteurs partout dans la plante à l'exception des stigmates. Dans les racines, *pAt1g28290(1,4 kb)::GUS* n'est pas exprimé au niveau de l'apex. Une coupe transversale de la racine montre que la construction est exprimée dans le phloème et le xylème primaire. Cependant dans ce travail, le patron d'expression du gène rapporteur dans des plantules étiolées n'a pas été analysé.

Afin de déterminer le patron d'expression d'*At1g28290* pendant la germination, 20 lignées T2 homozygotes indépendantes d'*A. thaliana* Col0 transgéniques comportant la construction *pAt1g28290(2,5 kb)::GFP::GUS* ont été testées. Le profil d'expression de cette construction a été analysé à la lumière et à l'obscurité à différents stades du développement des plantules jusqu'à 11 jours. Les résultats obtenus n'étaient pas cohérents entre les différentes lignées et ne montraient pas d'expression forte dans les hypocotyles des plantules étiolées (résultat non montré). Le test GUS appliqué aux lignées homozygotes de la génération T3 a montré un profil similaire à celui de la génération T2.

Non seulement les résultats n'étaient pas cohérents entre les différentes lignées, mais encore ils n'étaient pas cohérents avec les données de transcriptomique et de protéomique obtenues par ailleurs dans l'équipe (Irshad et al., 2008 ; Jamet et al., 2009). En effet, deux stades de développement des hypocotyles étiolés avaient été comparés : 5 jours et 11 jours après le début de la germination, ce qui correspondait à un stade d'élongation et à un stade de fin de croissance. Cette analyse détaillée avait montré que les transcrits d'*At1g28290* et la protéine AGP31 s'accumulent fortement à ces deux stades du développement. D'autre part, des données transcriptomiques récemment publiées, montrent un niveau d'accumulation faible de transcrits d'*At1g28290* dans des plantules 6 h après le début de la germination à l'obscurité, et une augmentation de ce niveau d'accumulation de transcrits à partir de 30 h de germination (Pelletier et al., 2010).

L'ensemble de ces résultats suggérait que la construction *pAt1g28290(2,5 kb)::GFP::GUS* ne permettait pas de décrire le patron d'expression d'*At1g28290*. Ceci nous a amené à choisir d'autres stratégies pour poursuivre ce travail.

5.2.2 Analyse *in silico* du profil d'expression d'*At1g28290* au cours du développement

La disponibilité de plusieurs milliers d'expériences de microarrays, traitant la grande majorité des gènes d'*A. thaliana* pour différents organes, au cours du développement et suite à différents stress, permet d'avoir une vue relativement exhaustive de la régulation spatio-temporelle de l'expression d'un gène d'intérêt (Hruz et al., 2008). Le profil d'accumulation de transcrits *At1g28290* au cours du développement est représenté sur la figure 5.2. Le niveau de transcrits d'*At1g28290* est relativement élevé dans la racine et dans l'hypocotyle. Plus précisément, il est élevé au niveau de l'apex, la zone méristématique et la zone d'élongation de la racine primaire. Dans la racine secondaire, le niveau de transcrits est élevé dans la coiffe. Un niveau de transcrits élevé a aussi été détecté dans l'apex caulinaire et la zone

méristémoïde de la tige. Par contre, aucun transcrite n'a été détecté au niveau de la stèle de la racine primaire en contradiction avec les données obtenues avec la construction *pAt1g28290(1,4 kb)::GUS* (Liu et Mehdy, 2007).

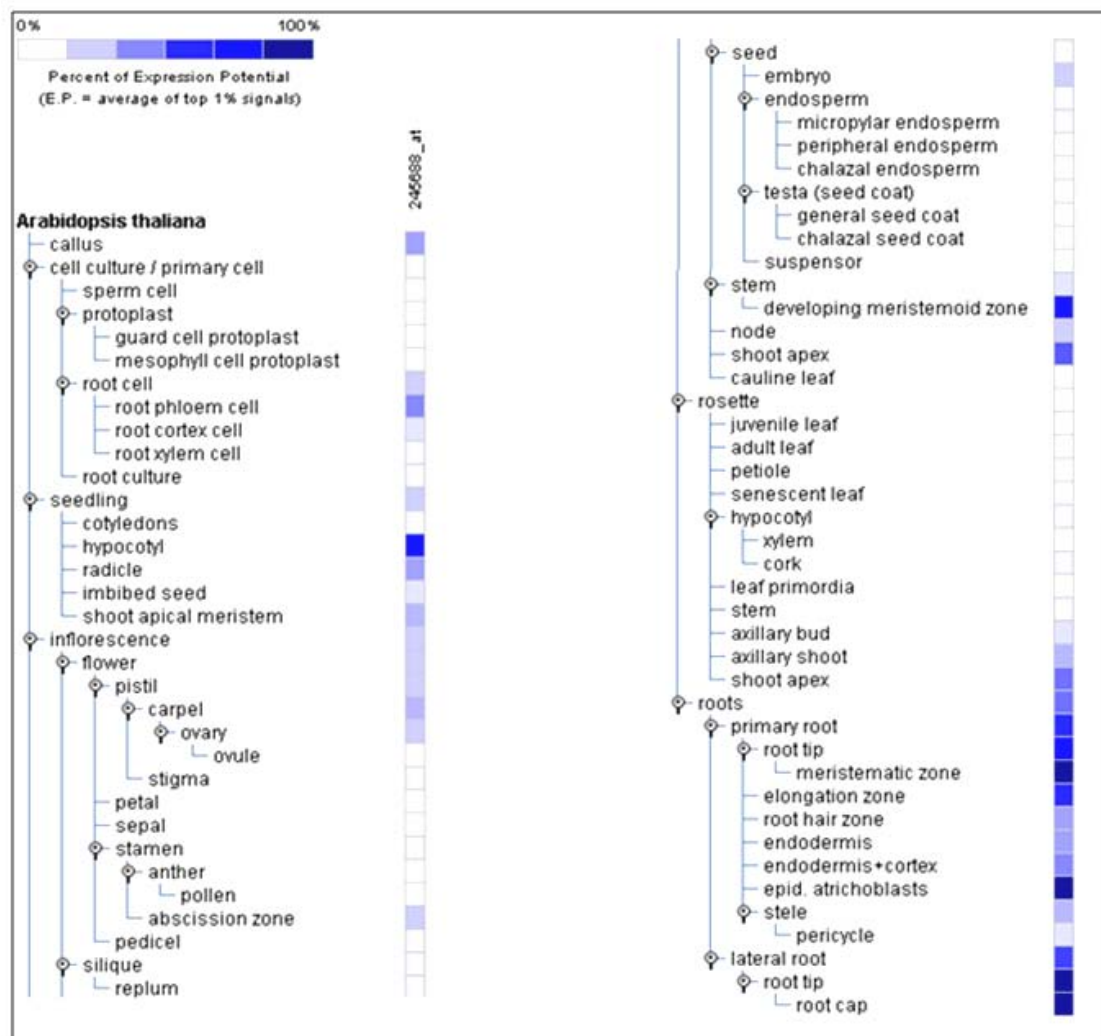


Figure 5.2 : Profil d'accumulation de transcrits d'*At1g28290* établi par analyse transcriptomique au cours du développement d'*A. thaliana* à la lumière (données <https://www.genevestigator.com/gv/>)

5.2.3 Analyse de l'accumulation des transcrits d'*At1g28290* par RT-PCR quantitative au cours de la germination à l'obscurité

Des graines d'*A. thaliana* sauvages ont été semées sur milieu gélosé et les plantules étioilées ont été récoltées à différents stades de développement à l'obscurité après 3 h d'imbibition des graines, 72 h de vernalisation et une exposition de 4 h à la lumière : 0, 1, 2, 3 et 5 jours. Des ARN totaux ont été extraits des plantules et des expériences de RT-PCR quantitatives ont été réalisées. Les résultats d'amplification ont été normalisés par rapport à l'amplification des transcrits du gène codant l'actine 8 (*At1g49240*) dans chaque échantillon, ce qui doit permettre d'éliminer les fluctuations liées à des différences dues à des quantités d'ADNc différentes dans les échantillons utilisés pour la PCR. Enfin, les quantités de transcrits sont exprimées relativement à la quantité détectée au stade 1 jour qui constitue notre référence interne. L'amplification a été réalisée en duplicat avec un seul lot d'ADNc.

Les résultats sont représentés sur la figure 5.3. Les transcrits sont présents à tous les stades du développement. L'accumulation des transcrits n'a pas pu être mesurée au stade 0 jour du fait de la faible quantité de matériel obtenue. L'accumulation des transcrits augmente progressivement au cours des 3 premiers jours avec une accumulation rapide et importante au stade 5 jours qui correspond à environ 16 fois plus de transcrits par rapport au stade de référence (1 jour).

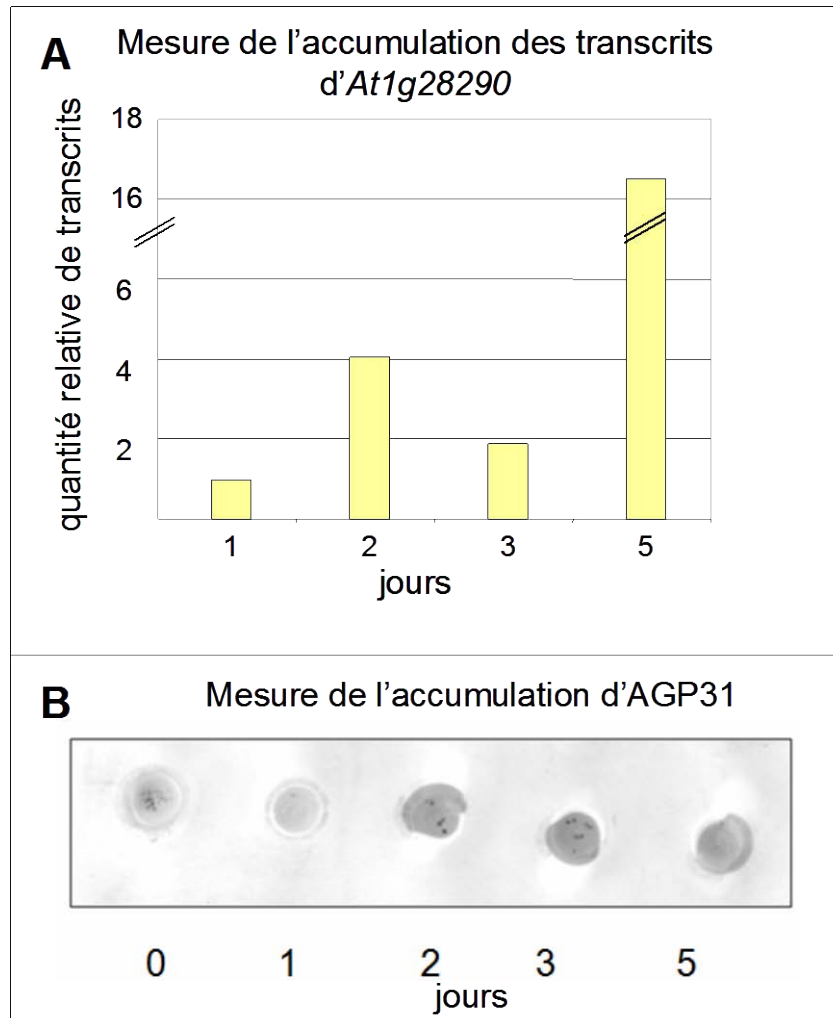


Figure 5.3 : Mesure de l'accumulation des transcrits d'At1g28290 par RT-PCR quantitative et de l'accumulation d'AGP31 par lectin blot au cours de la germination à l'obscurité de plantules d'*A. thaliana*

A. Les quantités de transcrits sont exprimées relativement à la quantité détectée au stade 1 jour qui constitue la référence interne. **B.** La protéine AGP31 est révélée par lectin blot par PNA-dig.

5.2.4 Mesure de l'accumulation d'AGP31 au cours de la germination

Après avoir mesuré l'accumulation des transcrits d'At1g28290 au cours de la germination, il nous semblait intéressant de regarder si le niveau d'accumulation de la protéine était cohérent avec ces données transcriptomiques. Les analyses ont été effectuées sur le même lot de plantules que celles ayant servi à l'analyse de l'accumulation des transcrits. Pour le lectin blot présenté dans cette partie, les protéines totales ont été extraites des plantules puis déposées sur membrane de nitrocellulose et révélées par incubation avec la lectine PNA-dig en utilisant la propriété de PNA de révéler spécifiquement AGP31 (cf. Chapitre 3).

Un signal faible est observé au stade 1 jour (Figure 5.3-B). Puis, il augmente à 2 et 3 jours et se stabilise à 5 jours, ce qui est cohérent avec les mesures de niveaux de transcrits.

5.2.5 Identification des cellules exprimant *At1g28290* par hybridation d'ARN *in situ* et par marquage histochimique d'AGP31 dans des plantules sauvages au cours de la germination

L'utilisation de broyats pour l'analyse transcriptomique et protéomique ne permet pas de discriminer les cellules qui expriment le gène d'intérêt. Nous avons choisi d'approfondir ce point en analysant l'accumulation des transcrits et des protéines par les techniques d'hybridation d'ARN *in situ* et de marquage histochimique des protéines à l'aide de PNA-dig.

L'étude de l'accumulation des transcrits d'*At1g28290* a été réalisée en utilisant des sondes anti-sens et sens (contrôle négatif) et la localisation tissulaire d'AGP31 a été réalisée par marquage avec ou sans PNA-dig (contrôle négatif) sur des coupes sériées de plantules étiolées à différents stades de germination. Du fait que ces expériences sont réalisées sur des coupes sériées, la morphologie des tissus est globalement préservée, facilitant ainsi les comparaisons de profils dans les quatre conditions de marquage. Afin de faciliter la lecture des images obtenues, la figure 5.4 présente de façon détaillée l'identification des différents tissus observés sur des coupes sériées identiques à celles ayant servi aux expériences.

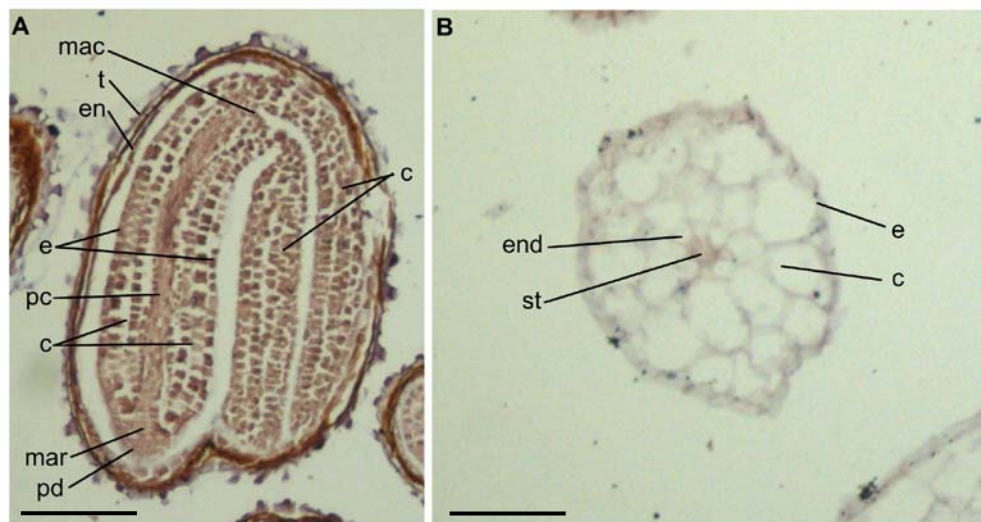


Figure 5.4 : Coupes longitudinale d'une plantule d'*A. thaliana* (A) et transversale d'un hypocotyle (B) au cours de la germination à l'obscurité

Les échantillons sont prélevés au stade 0 jour. c : cortex, e : épiderme, end : endoderme, en : endosperme, mac : méristème apical caulinaire, mar : méristème apical racinaire, pc : procambium, pd : protoderme, st : stèle, t : tégument. Barres = 100 μ m.

A la fin de l'embryogenèse, la plantule a l'organisation de base de la plante avec une symétrie radiale par rapport à l'axe embryonnaire. Les futurs tissus conducteurs (procambium) se trouvent au centre et le parenchyme cortical et le protoderme à la périphérie. L'axe embryonnaire possède des tissus méristématiques à ses extrémités : les méristèmes apicaux caulinaire et racinaire qui renferment des amas cellulaires en prolifération préfigurant les primordia d'organes (Barton et Poethig, 1993). Ces méristèmes assurent la croissance de la plante en longueur, respectivement au niveau de la tige, des feuilles et des racines. L'embryon

présente aussi deux cotylédons qui sont des organes de réserve au moment de la germination. Les réserves des cotylédons sont constituées à partir de l'endosperme lors de l'embryogenèse. Au stade mature de la graine, l'endosperme se trouve réduit à une assise cellulaire située sous les téguments. Les cotylédons sont reliés à l'axe embryonnaire juste en dessous du méristème apical caulinaire et l'enserrent. Après la phase d'embryogenèse, la graine entre en maturation pour se préparer à subir la phase de quiescence. La croissance de l'embryon reprend au moment de la germination.

La comparaison des résultats obtenus avec les sondes ARN anti-sens et sens au niveau de coupes longitudinales sériées de la graine portant l'embryon cotylédonaire au stade 0 jour montre un marquage spécifique du méristème apical caulinaire et du protoderme de la radicule et, dans une moindre mesure, de certaines zones de l'endosperme (Figure 5.5-A, B). Les coupes sériées suivantes révélées avec PNA-dig permettent d'observer un marquage comparable du protoderme et dans une moindre mesure du méristème apical caulinaire, mais surtout un très fort marquage de l'endosperme et un fort marquage constitutif de l'embryon (Figure 5.5-C, D). A ce stade, la correspondance entre les deux types de marquage ciblant respectivement les transcrits et la protéine est donc partielle, suggérant un problème de spécificité de reconnaissance d'AGP31 par PNA. Par contre, à partir de 1 jour, les deux types de marquage sur des coupes sériées montrent une bonne correspondance. En effet, le marquage spécifique observé parfois à 0 jour dans le méristème apical caulinaire et le protoderme de la radicule s'intensifie et se généralise à 1 jour avec les deux techniques de marquage (Figure 5.5 E-H). Dans le même temps on observe une disparition nette du marquage des transcrits dans l'endosperme. Le marquage PNA constitutif de l'embryon subsiste. Enfin, à partir du stade 2 jours, ce marquage généralisé est observé avec les deux techniques avec une intensité diminuant graduellement du fait de l'augmentation du volume des cellules conduisant à une dilution du signal (Figure 5.5 I-L).

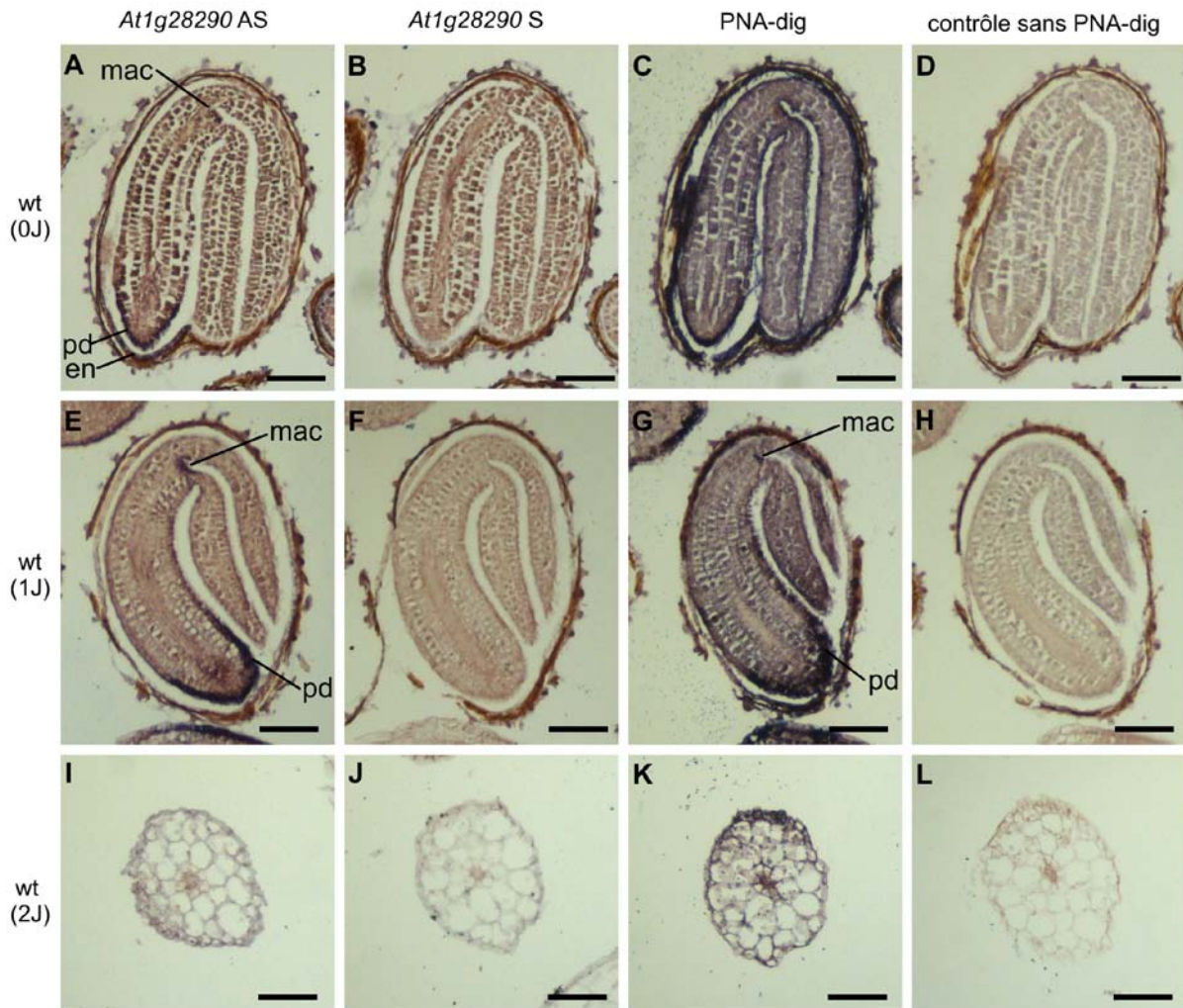


Figure 5.5 : Localisation des transcrits d'*At1g28290* et d'*AGP31* au cours de la germination

Les coupes sériées d'embryons d'*A. thaliana* sauvage (wt) à 0 (0J), 1 (1J) et 2 (2J) jours ont été hybridées avec 400 ng/mL de sondes anti-sens (AS) ou sens (S) marquées à la digoxigénine (dig) produites à partir de l'ADNc d'*At1g28290* ou incubées avec ou sans PNA-dig comme indiqué. A 0 jour, on peut voir, pour les transcrits, un marquage pourpre/violet spécifique (signalé par des traits) au niveau du méristème apical caulinaire (mac), du protoderme de la racine (pd) et de l'endosperme (en) (A). Ce même marquage est trouvé après incubation par PNA mais il semble beaucoup plus fort au niveau de l'endosperme, et il est constitutif dans l'embryon (C). A 1 jour, le marquage des sondes AS et par PNA du méristème apical caulinaire s'intensifie, ainsi qu'au niveau du protoderme où le marquage devient plus diffus. Le marquage de l'endosperme disparaît complètement et le marquage constitutif par PNA persiste (E, G). A 2 jours, un marquage constitutif est observé (I, K). La spécificité de ces marquages peut être appréciée par l'absence de tels marquages chez les contrôles sens et sans PNA-dig (B, D, F, H, J, L). Noter la coloration brune naturelle des téguments présents sur certaines coupes ne correspondant pas à un marquage spécifique.

Barres = 100 μ m.

5.2.6 Conclusion

L'ensemble de ces résultats permet d'avoir une idée du patron d'expression d'*At1g28290* au cours de la germination à l'obscurité. Les données de transcriptomique obtenues avec des plantules entières (Pelletier et al., 2010) montrent que le niveau de transcrits augmente progressivement à partir de 6 h de germination. Il atteint sa valeur maximale après 30 h de germination et la conserve jusqu'à 50 h après le début de la germination. Les données

obtenues avec les hypocotyles étiolés à 5 et 11 jours indiquent un fort niveau d'accumulation de transcrits (Jamet et al., 2009). Les contrôles effectués par RT-PCR quantitative sur des plantules entières confirment ce profil d'accumulation de transcrits. L'approche par hybridation d'ARN *in situ* apporte des informations plus précises pour l'identification des cellules exprimant *At1g28290* : aux stades précoces (0 et 1 jour), il s'agit des cellules du méristème apical caulinaire, des cellules du protoderme et de certaines cellules de l'endoderme. Ensuite, toutes les cellules semblent exprimer *At1g28290* (2 jours).

Les informations concernant le niveau d'accumulation d'AGP31 sont cohérentes avec les données transcriptomiques aux stades tardifs de la germination à l'obscurité (5 et 11 jours) avec une forte accumulation de protéines (Irshad et al., 2008). Aux stades précoces, l'utilisation de la lectine PNA couplée à la digoxigénine permet d'apporter des précisions. D'une part, l'approche immunocytochimique montre une localisation identique à celle des transcrits aux jours 1 et 2. D'autre part, l'approche *lectin blot* montre que la protéine commence à s'accumuler dès le jour 2. Les marquages constitutifs observés aux temps précoces (0 et 1 jour) pourraient être dus à une reconnaissance d'autres constituants pariétaux par PNA. Des expériences complémentaires devraient être effectuées pour comprendre ce marquage qui n'est pas cohérent avec les résultats de l'hybridation d'ARN *in situ*.

5.3 Caractérisation de lignées sous-expresssurs d'*At1g28290* (lignées ARNi) et recherche de phénotypes au cours du développement précoce à l'obscurité

5.3.1 Obtention de lignées ARNi sous-expresssurs d'*At1g28290*

Nous avons choisi d'analyser des mutants susceptibles d'avoir des niveaux réduits de transcrits d'*At1g28290*, en exploitant le mécanisme d'ARNi. Des plantes transgéniques ayant intégré dans leur génome des constructions ARNi (AGRIKOLA, cf. § 2.1.1.1) avaient été obtenues au laboratoire par M. Irshad et D. Roujol. Parmi celles-ci, 4 lignées hémizygotés de la génération T1 (lignées 14, 10, 11 et 24) sélectionnées par résistance au Basta présentaient des niveaux de réduction différents des transcrits d'*At1g28290*. A la génération T2 et du fait de la ségrégation, 5 lignées transgéniques issues de ces 4 lignées T1 (14-19, 10-17, 11-17, 24-14, 24-19) avec un rapport de ségrégation (3/4 ; 1/4), (1/2 ; 1/2), (1 ; 0) ou (15/16 ; 1/16) ont été sélectionnées pour passer à la génération T3 sur la base du niveau d'accumulation de transcrits d'*At1g28290*. A la génération T3, 9 lignées homozygotes représentées sur la figure 5.6 ont été sélectionnées par résistance au Basta.

Pour vérifier si le mécanisme de répression d'*At1g28290* par ARNi s'est mis en place chez les transformants, nous avons analysé l'accumulation des transcrits d'*At1g28290* dans des jeunes tiges de plantes par RT-PCR semi-quantitative pour les lignées T1 et T2 et par RT-PCR quantitative pour les lignées T3. Deux gènes exprimés constitutivement ont été utilisés comme contrôles de la quantité d'ADNc dans les échantillons : le gène codant la β -tubuline (*At5g44340*) pour la RT-PCR semi-quantitative, et le gène codant l'actine 8 (*At1g49240*) pour la RT-PCR quantitative. Les résultats concernant les lignées homozygotes T3 sont présentés sur la figure 5.6. Les quantités de transcrits d'*At1g28290* sont exprimées relativement à la quantité détectée chez le sauvage (valeur arbitrairement fixée à 1) après normalisation des mesures par rapport au niveau d'accumulation des transcrits d'*At1g49240*.

Des transcrits d'*At1g28290* sont détectés par RT-PCR quantitative chez les plantes sauvages et transformées. Néanmoins, les transcrits s'accumulent moins dans la majorité des lignées analysées que dans la plantule sauvage. On remarque une grande variabilité dans les niveaux de répression d'*At1g28290*, qui pourrait être liée à un effet de position des insertions d'ADN-T, à des insertions multiples d'ADN-T, ou à des régulations transcriptionnelles ou post-transcriptionnelles. Les lignées 24.14.1.3 et 24.19.8.4 présentant une forte réduction de l'accumulation des transcrits d'*At1g28290* (27,3 et 10 fois respectivement par rapport aux plantes sauvages) ont été choisies pour les analyses suivantes.

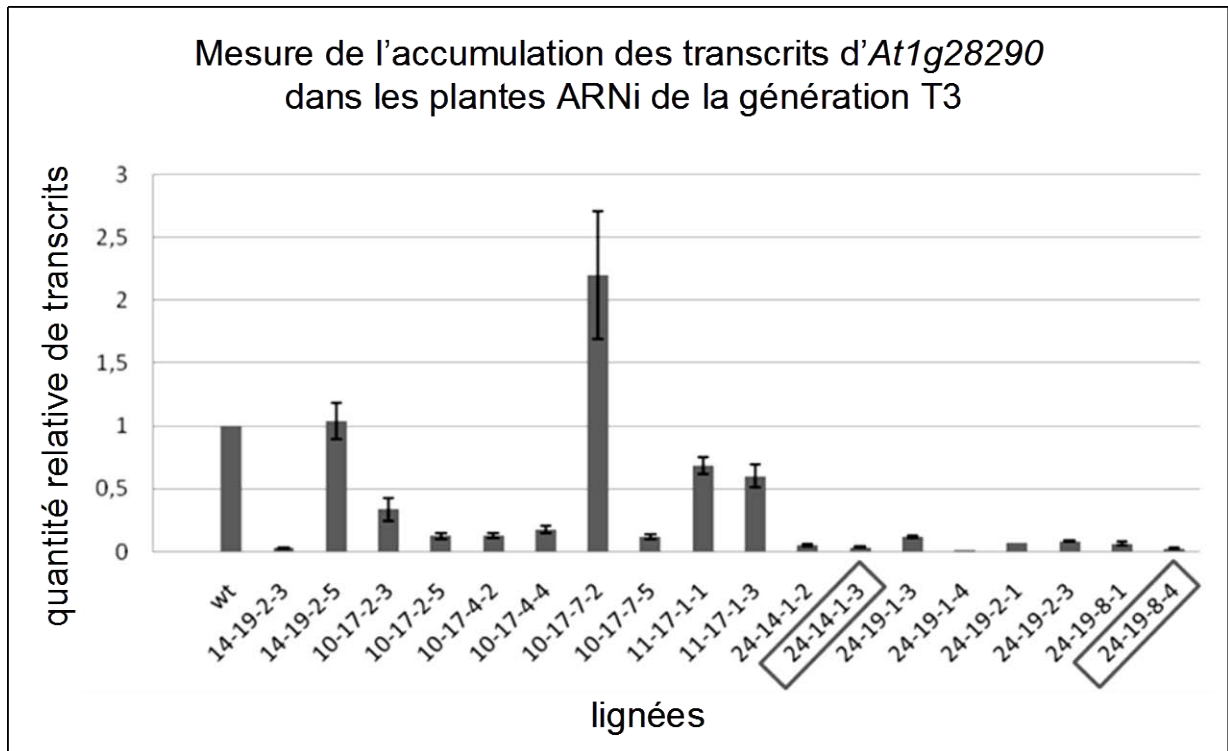


Figure 5.6 : Quantification par RT-PCR des transcrits d'*At1g28290* dans les jeunes tiges de différentes lignées de plantes ARNi de la génération T3

Les niveaux de transcrits d'*At1g28290* dans 18 lignées ARNi ont été mesurés par RT-PCR quantitative. Les quantités de transcrits dans les plantes transgéniques sont rapportées à celle de la plante sauvage (wt), arbitrairement fixée à 1. Le gène codant l'actine 8 sert de contrôle interne. Son expression est considérée comme constitutive dans toutes les plantes analysées.

5.3.2 Analyse moléculaire des lignées ARNi sous-expresseurs au cours du développement précoce en conditions d'étiollement

Une série d'expériences a été réalisée dans le but de déterminer si les mutants présentaient un phénotype de développement. Afin de vérifier si le niveau de transcrits est réduit dans les plantules transgéniques au cours de la germination à l'obscurité, des mesures ont été effectuées par RT-PCR quantitative en duplicat et AGP31 a été recherchée par *lectin blot* sur des extraits totaux de protéines de façon similaire à l'étude effectuée chez les plantes sauvages (Figure 5.7). Le niveau des transcrits d'*At1g28290* est très faible pour les lignées ARNi. Il est inférieur à celui des plantules sauvages d'un facteur 10 dès le stade 1 jour et se maintient à ce niveau aux jours suivants. Il est fortement corrélé avec le niveau d'accumulation des protéines entre 1 et 5 jours. Par contre, au stade 0 jour la forte dilution de

l'échantillon n'a pas permis d'obtenir des résultats de RT-PCR quantitative corrects. Pour ce stade, on ne dispose donc que du résultat du *lectin blot*. La mesure du niveau d'accumulation d'AGP31 par *lectin blot* montre une réduction forte de cette accumulation dès 0 jour confirmant l'efficacité de la stratégie ARNi.

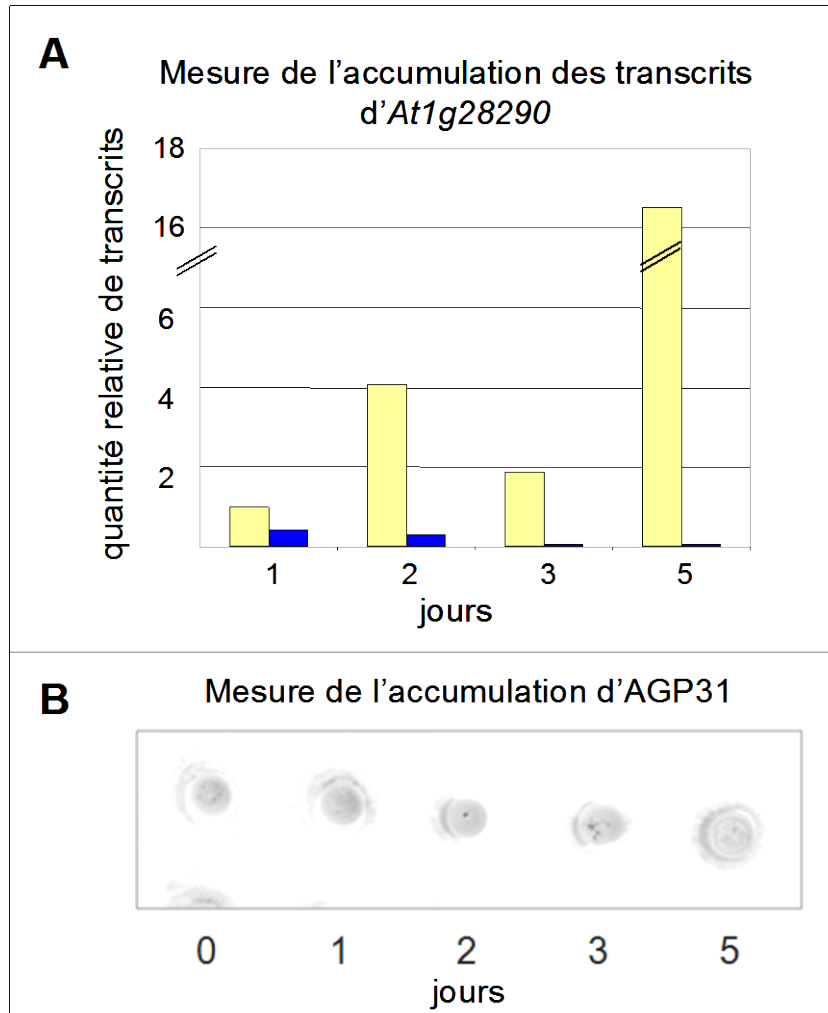


Figure 5.7 : Mesure de l'accumulation des transcrits d'*At1g28290* par RT-PCR quantitative et de l'accumulation d'AGP31 par *lectin blot* au cours de la germination à l'obscurité de plantules d'*A. thaliana* ARNi sous-expresses d'*At1g28290*

A. Les quantités de transcrits sont exprimées relativement à la quantité détectée au stade 1 jour chez le sauvage qui constitue la référence interne. Les barres jaunes et bleues correspondent respectivement aux mesures effectuées sur les plantules sauvages et ARNi sous-expresses. **B.** La protéine AGP31 est révélée par *lectin blot* par PNA-dig. Les traces observées correspondent à du bruit de fond.

5.3.3 Identification des cellules exprimant *At1g28290* par hybridation d'ARN *in situ* et par marquage histochimique d'AGP31 dans des plantules ARNi sous-expressseurs au cours de la germination

De même que pour les plantules sauvages (*cf.* § 5.2.5) des expériences d'hybridation d'ARN *in situ* et d'immunomarquage ont été effectuées sur des plantules des lignées ARNi sous-expressseurs au cours de la germination à l'obscurité. La forte majorité des graines a montré une absence de signal par hybridation *in situ* au niveau du méristème apical caulinaire et du protoderme de la radicule aux jours 0 et 1 (Figure 5.8 A, E). Par contre, l'ensemble des cellules présente un marquage avec du PNA au jour 0, particulièrement au niveau de l'endoderme (Figure 5.8 C). A ce stade, il n'y a pas de correspondance entre les deux types de marquage, ciblant respectivement les transcrits et la protéine, comme observé avec les graines sauvages (*cf.* figure 5..5). Au stade 1 jour, aucun marquage n'est observé avec PNA (Figure 5.8 G).

Ces résultats montrent une grande spécificité des sondes ARN sens et antisens, mais confirment la reconnaissance par PNA de constituants pariétaux différents d'AGP31 aux stades très précoces de la germination.

Enfin, aucun phénotype n'a pu être observé sur ces plantules en germination. Le suivi de stades plus tardifs de germination à la lumière et à l'obscurité n'a pas non plus permis de visualiser des différences macroscopiques (résultats non montrés).

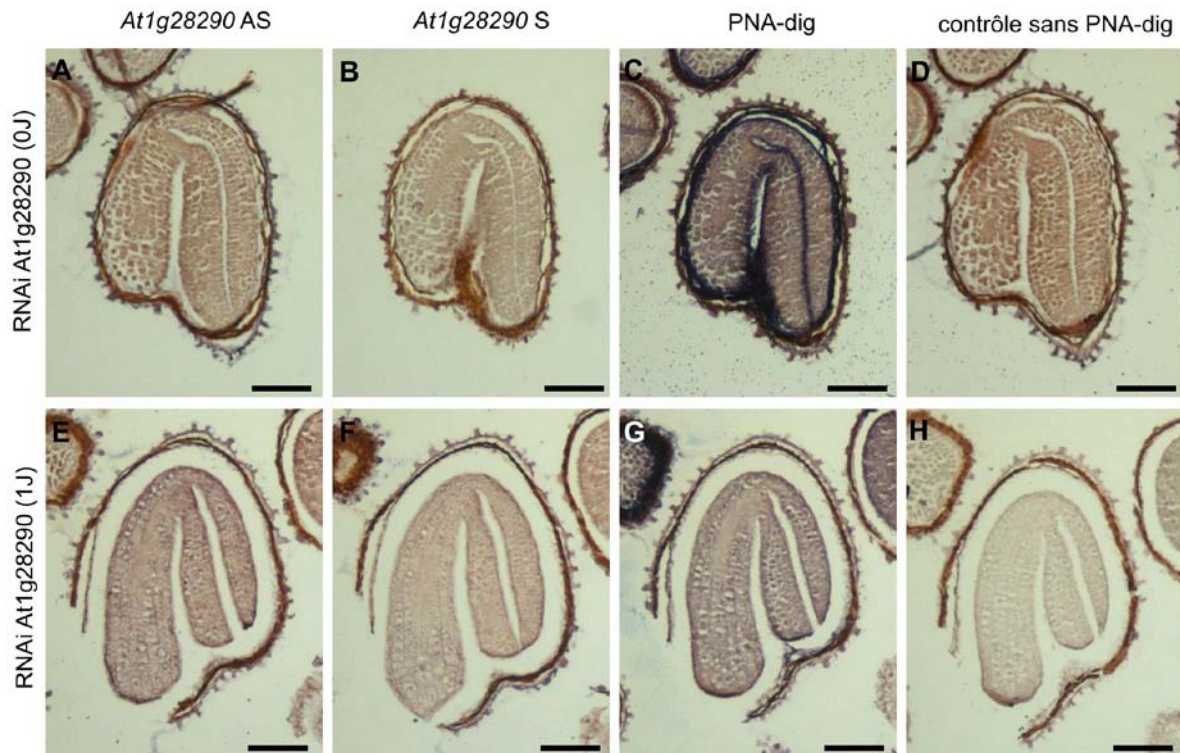


Figure 5.8 : Localisation des transcrits d'*At1g28290* et d'*AGP31* au cours de la germination à l'obscurité de graines de lignées ARNi sous-expressseurs

Les coupes sériées d'embryons ARNi sous-expressseurs d'*At1g28290* effectuées aux stades 0 jour (0J) et 1 jour (1J) ont été hybridées avec 200 ng/mL de sondes anti-sens (AS) ou sens (S) marquées ou incubées avec ou sans PNA-dig comme indiqué. Au stade 0 jour, un marquage PNA de toutes les cellules est visible, et de manière plus marquée au niveau de l'endosperme (C). Ce marquage est absent dans le cas des hybridations réalisées avec les sondes sens (A). Au stade 1 jour, le marquage disparaît complètement de toutes les cellules (E, G). Les contrôles sens et sans PNA-dig ne révèlent aucun marquage (B, D, F, H).

Barres = 100 μ m.

5.4 Caractérisation de lignées sur-expressseurs d'*At1g28290* et recherche de phénotypes au cours du développement précoce à l'obscurité

5.4.1 Obtention de lignées sur-expressseurs d'*At1g28290*

Le transfert de la construction 35S::région codante d'*At1g28290*::tag V5::tag 6xHis chez *A. thaliana* Col-0 a été effectué par trempage des fleurs dans une suspension d'*A. tumefaciens*. Nous avons obtenu des transformants primaires après sélection des plantes sur kanamycine. Afin de faciliter le travail de criblage, il est important d'utiliser des lignées transgéniques homozygotes et comportant un seul locus d'insertion d'ADN-T. Nous avons donc testé la ségrégation des transgènes dans la descendance des lignées transgéniques afin de déterminer s'il y a eu un ou plusieurs événements d'insertion. Pour cela, un test de germination a été effectué sur un milieu MS en présence de kanamycine.

Onze lignées hémizygotiques T2 issues des plantes T1 sélectionnées par résistance à la kanamycine ont été obtenues. Parmi ces 11 lignées T2, 4 lignées (4I.1'.4, 4I.1'.5, 4I.1'.8 et 4I.1'.12) ont été sélectionnées après une analyse par PCR qui a montré une sur-expression d'*At1g28290* en comparant avec des plantes sauvages. A la génération T2 et du fait de la

ségrégation, une quarantaine de lignées transgéniques issues de ces 4 lignées T2 avec un rapport de ségrégation (3/4 ; 1/4) ont été sélectionnées pour passer à la génération T3. Neuf lignées T3 représentées sur la figure 5.9 présentant 100% de graines germées résistantes à la kanamycine ont été obtenues : elles étaient donc homozygotes pour le transgène.

Nous avons analysé l'accumulation des transcrits d'*At1g28290* dans ces lignées sur-expresses homozygotes T3 et chez une plante sauvage par RT-PCR quantitative. Un contrôle de la quantité des ADNc utilisés est obtenu en amplifiant les ADNc de l'actine 8. Les profils d'expression d'*At1g28290* dans les jeunes feuilles de rosette d'*A. thaliana* sont présentés dans la figure 5.10. Les quantités de transcrits sont exprimées relativement à la quantité détectée chez le sauvage au stade 1 jour.

Nous remarquons une variabilité dans les niveaux de sur-expression d'*At1g28290* dans les différentes lignées. La lignée 4l.1'.12.7 présentant un niveau modéré de sur-expression de 21,6 fois par rapport au sauvage a été choisie pour les analyses, du fait qu'une sur-expression modérée peut limiter le phénomène de *silencing* qui peut intervenir lorsque le niveau de transcrits d'un gène est trop élevé (Vaucheret et al., 1998).

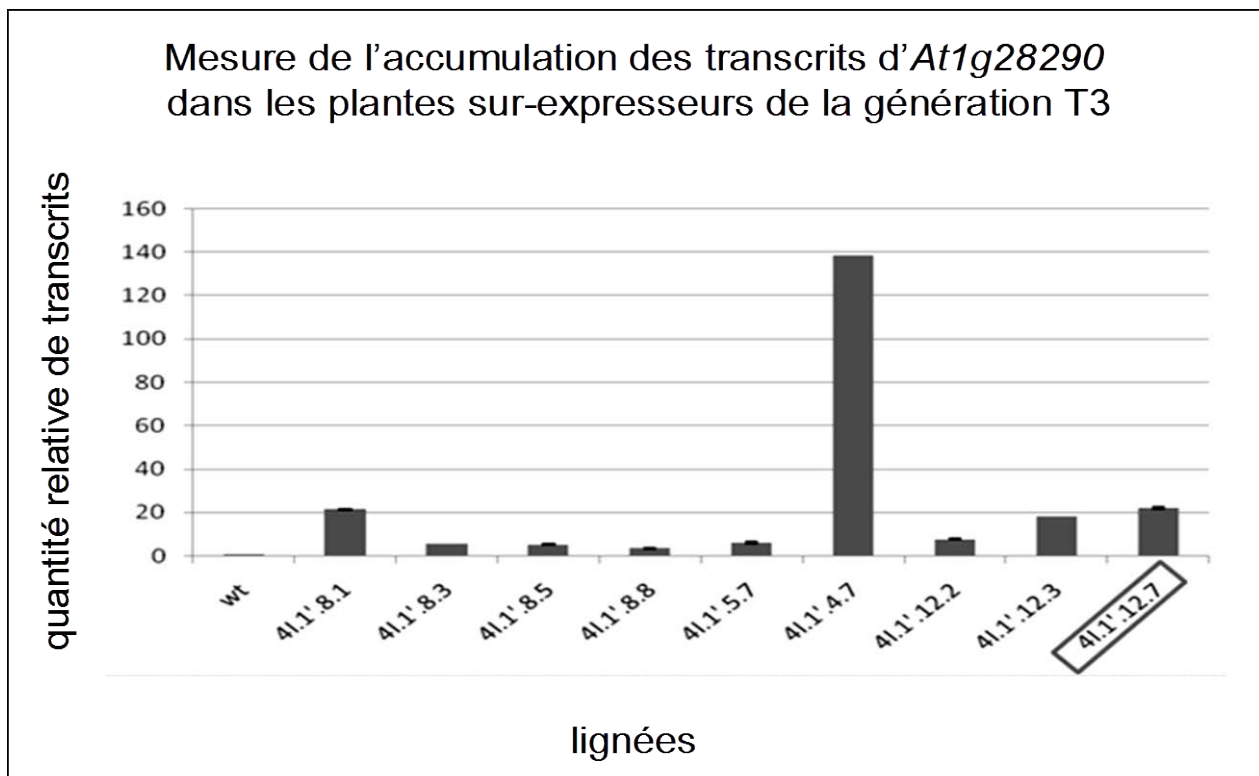


Figure 5.9 : Quantification par RT-PCR des transcrits d'*At1g28290* dans les rosettes de différentes lignées de plantes sur-expresses de la génération T3

Les niveaux de transcrits d'*At1g28290* dans 9 lignées sur-expresses ont été mesurés par RT-PCR quantitative. Les quantités de transcrits dans les plantes transgéniques sont rapportées à celle de la plante sauvage (wt), arbitrairement fixée à 1. Le gène codant l'actine 8 sert de contrôle interne. On considère son expression comme constitutive dans toutes les plantes analysées.

5.4.2 Analyse moléculaire des lignées sur-expresses au cours du développement en conditions d'étiollement

Comme précédemment, les niveaux d'accumulation des transcrits d'*At1g28290* ont été mesurés par RT-PCR quantitative en duplicat et ceux des protéines par *lectin blot* sur des

extraits totaux de protéines prélevés à différents stades de germination à l'obscurité (Figure 5.11). La forte dilution de l'échantillon 0 jour n'a pas permis une mesure correcte du niveau d'accumulation des transcrits par RT-PCR quantitative. Les niveaux de transcrits d'*At1g28290* sont plus élevés que ceux mesurés chez les plantules sauvages à partir du stade 2 jours (environ 30 fois). Ils se maintiennent ensuite autour de 40 fois les quantités mesurées dans les plantules sauvages aux mêmes stades de développement. Ils sont très bien corrélés au niveau d'accumulation d'AGP31 (Figure 5.10 B).

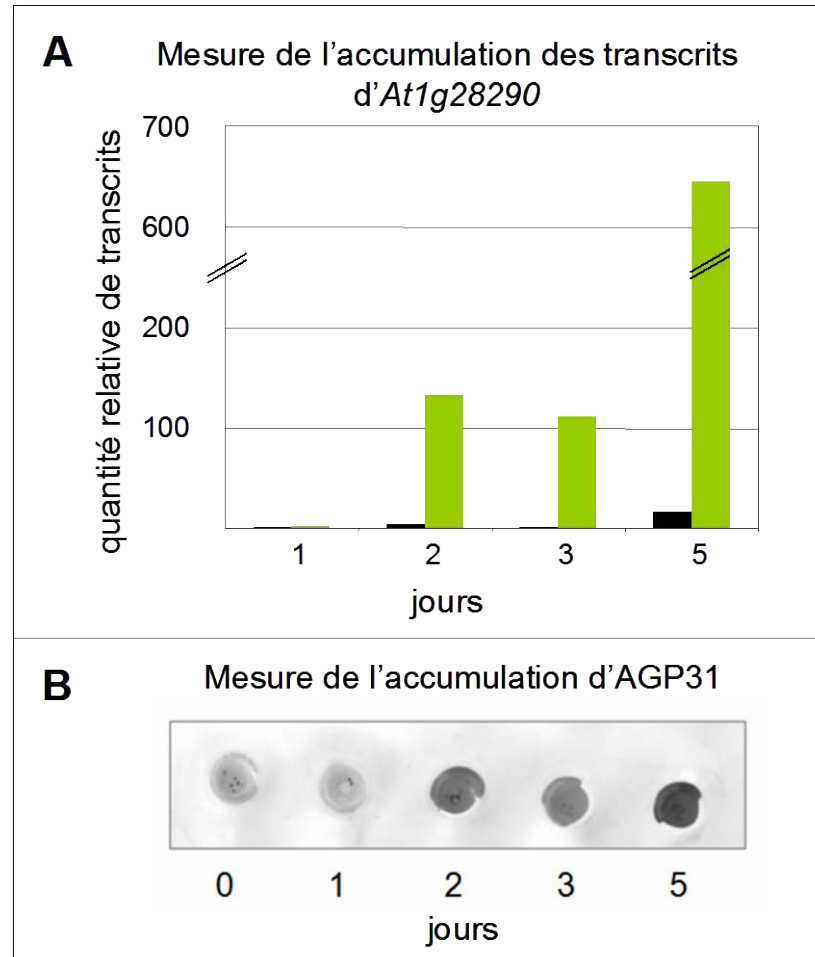


Figure 5.10 : Mesure de l'accumulation des transcrits d'*At1g28290* par R-PCR quantitative et de l'accumulation d'AGP31 par lectin blot au cours de la germination à l'obscurité de plantules sur-expressseurs d'*At1g28290*

A. Les quantités de transcrits sont exprimées relativement à la quantité détectée au stade 1 jour dans les plantules sauvages qui constitue notre référence interne. Les barres noires et vertes correspondent respectivement aux mesures effectuées sur des plantules sauvages et sur-expressseurs. **B.** La protéine AGP31 est révéllé par lectin blot par PNA-dig (B).

5.4.3 Identification des cellules exprimant *At1g28290* par hybridation d'ARN *in situ* et par marquage histochimique d'AGP31 dans des plantules sur-expressseurs au cours de la germination à l'obscurité

Au stade 0 jour, les lignées sur-expressseurs présentent le même patron d'expression d'*At1g28290* que les embryons sauvages avec les sondes ARN anti-sens. Le marquage par PNA-dig est également similaire dans le méristème apical caulinaire et le protoderme de la racicule, avec une intensité hétérogène dans l'endosperme (Figure 5.11 A, C). Les mêmes

résultats ont été obtenus au stade 1 jour (non montrés). Par contre, à partir du stade 2 jours, une expression ectopique d'*At1g28290* a été détectée (Figure 5.11 E-G) confirmant ainsi les données de RT-PCR quantitative (cf. § 5.4.2).

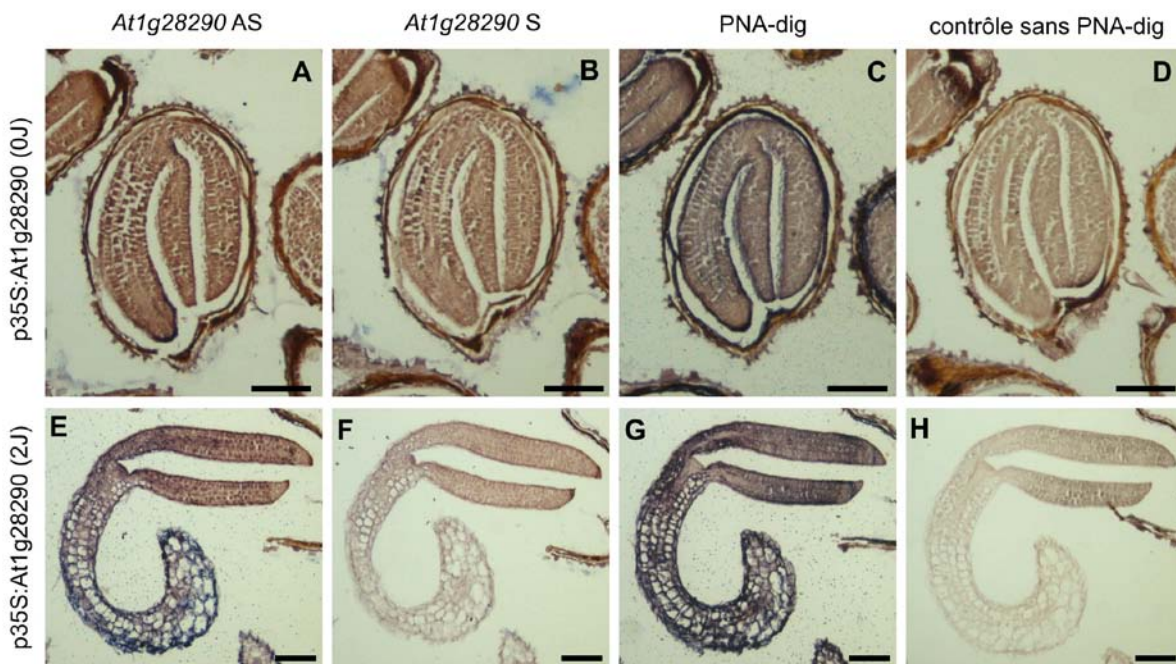


Figure 5.11 : Localisation des transcrits d'*At1g28290* et d'*AGP31* au cours de la germination de graines de lignées sur-expressseurs au cours de la germination

Les coupes sériées d'embryons d'*A. thaliana* sur-expressseurs effectuées aux stades 0 jour (0J) et 2 jours (2J) ont été hybridées avec 400 ng/mL de sondes anti-sens (AS) ou sens (S) marquées à la digoxigénine (dig) ou incubées avec ou sans PNA-dig comme indiqué. Au stade 0 jour, un marquage faible au niveau des cellules du méristème apical caulinaire et du protoderme sont visibles sur les coupes A et C. Ce patron d'expression est similaire à celui du sauvage. Au stade 2 jours, un marquage ectopique est détecté au niveau de la plantule (E, G). Les contrôles sens et sans PNA-dig ne révèlent aucun marquage (B, D, F, H).

Barres = 100 μ m.

Du point de vue phénotypique, il n'y a pas de différence au niveau de la morphologie de l'embryon dans les lignées sur-expressseurs comparées aux sauvages. Par contre, il est remarquable que seules des coupes effectuées sur des plantules sur-expressseurs soient préservées correctement au-delà du stade 1 jour (Figure 5.11 E-H). Des modifications subtiles des parois conduisant à leur renforcement pourraient favoriser cette préservation. Des observations complémentaires devraient être effectuées pour vérifier cette hypothèse.

5.5 Analyse phénotypique des plantes ARNi sous-expressseurs et sur-expressseurs d'*At1g28290* au cours du développement

Les lignées transgéniques ARNi sous-expressseurs et sur-expressseurs d'*At1g28290* ont été suivies au cours de leur développement en logette de culture et en serre. Les lignées transgéniques manifestent une germination tardive à l'obscurité par rapport aux lignées sauvages. Il semble nécessaire de répéter l'expérience et de faire des tests de germination sur davantage de graines pour vérifier que cette germination précoce n'est pas liée aux conditions de culture.

Au niveau du développement de l'appareil végétatif, malgré des niveaux élevés de répression ou de sur-expression d'*At1g28290*, nous n'avons pas observé de variations phénotypiques macroscopiques avec les plantes sauvages (résultats non montrés). Les plantes transgéniques présentent la même stature au niveau de la rosette et de la hampe florale. La croissance de l'appareil racinaire a été analysée sur *Araponics* (<http://www.araponics.com/>), un système de culture hydroponique, qui utilise des substances nutritives à la place du sol, permettant d'observer facilement les racines au cours de leur développement. Aucune différence n'a pu être mise en évidence avec les racines sauvages (longueur des racines, ramifications). Le délai de floraison a été étudié en dénombrant les feuilles de la rosette au cours de la croissance jusqu'à l'apparition de la hampe florale. De même, aucune différence remarquable n'a été notée.

5.6 Discussion

L'ensemble des résultats présentés dans ce chapitre permet de proposer un patron d'expression pour *At1g28290* au cours de la germination en conditions d'étiollement, et notamment pour ses phases précoces. Différentes approches expérimentales ont été combinées puisque l'analyse du patron d'expression de la construction *pAt1g28290 (2,5 kb)::GFP::GUS* n'a pas donné de résultats cohérents. Ainsi, *At1g28290* est exprimé dans les cellules du méristème apical, du protoderme de la radicule et de l'endoderme dès les premières heures de la germination. Après cinq jours de germination, les niveaux élevés d'accumulation des transcrits et d'AGP31 dans les hypocotyles laissent penser que l'ensemble des cellules expriment *At1g28290*. C'est la première fois que le patron d'expression précoce de ce gène est décrit. Par ailleurs, des plantes transgéniques homozygotes ARNi sous-expresses d'*At1g28290* ainsi que des plantes transgéniques sur-expresses d'*At1g28290* ont été obtenues et caractérisées au niveau moléculaire. Pour les plantes étudiées plus en détail, les niveaux d'accumulation de transcrits d'*At1g28290* sont respectivement réduits d'environ 10 fois et augmentés d'environ 20 fois. Par contre, aucun phénotype macroscopique évident n'a été observé au cours du développement précoce, ni chez les plantes adultes.

L'absence de cohérence entre les résultats obtenus avec les constructions *pAt1g28290(2,5 kb)::GFP::GUS* et les autres approches expérimentales a posé problème puisqu'il n'a pas été possible de décrire l'ensemble du patron d'expression du gène au cours du développement comme prévu. En effet, l'approche par hybridation *in situ* a été limitante du fait de l'impossibilité d'obtenir des coupes de bonne qualité après deux jours de germination à l'obscurité. Une explication à cet échec pourrait être l'absence d'éléments de régulation dans la construction réalisée. De tels éléments pourraient être localisés dans la région codante du gène, dans l'intron ou dans sa région 3' non codante. Cependant, la construction *pAt1g28290 (1,5 kb)::GFP::GUS* avait permis de décrire un patron d'expression dans des plantes adultes, même si ce patron d'expression n'est pas tout à fait cohérent avec les données transcriptomiques (Liu et Mehdy, 2007). Une autre hypothèse serait que la construction *pAt1g28290(2,5 kb)::GFP::GUS* comporte des éléments de régulation négatifs situés entre -2,5 et -1,5 kb en amont du codon d'initiation de la traduction. Il sera donc nécessaire de réaliser des constructions comportant des longueurs intermédiaires de la région promotrice pour essayer de comprendre la régulation d'*At1g28290*.

La régulation de la transcription d'*At1g28290* au cours du développement précoce de l'embryon, *i.e.* de manière très précise dans certaines cellules (méristème apical caulinaire, protoderme et endoderme), puis de manière plus généralisée au moins dans l'hypocotyle, est à rapprocher du mode de croissance de la plantule au cours de la germination. Une première phase correspond à la croissance initiale de la plantule jusqu'à l'émergence de la radicule. Les cellules du protoderme et de l'endosperme sont soumises à un fort stress mécanique correspond à la pression exercée par la radicule pour traverser l'endosperme et le tégument de la graine. Un tel profil d'expression a déjà été décrit pour un gène de *N. tabacum* codant l'extensine *Ext1.4* (Salvà et Jamet, 2001). Une seconde phase d'expression suit la croissance de l'embryon avec notamment l'élongation subie par les cellules de l'hypocotyle (Gendreau et al., 1997). Cette expression accompagne le passage de l'étape d'épaississement des parois à celle de leur amincissement accompagnant l'élongation cellulaire (Derbyshire et al., 2007). On pourrait imaginer qu'au cours de la première phase, AGP31 participe au renforcement des parois des cellules soumises à des contraintes mécaniques au moment de l'émergence de la radicule, tandis qu'elle participe à la structuration des parois au cours de la phase d'élongation rapide des cellules de l'hypocotyle.

Concernant la spécificité de reconnaissance d'AGP31 par PNA qui avait été mise à profit notamment pour purifier la protéine à partir des hypocotyles étiolés (*cf.* Chapitre 3), et pour détecter AGP31 lors des tests d'interaction *in vitro* avec des polysaccharides (*cf.* Chapitre 4), elle doit être remise en question dans le cas des embryons. En effet, un signal uniforme est observé dans toutes les cellules de l'embryon après le marquage par PNA. Ce signal n'est pas observé au niveau des marquages des extraits totaux de protéines effectué par *lectin blot* qui ne contiennent pas de polysaccharides. Il semble donc que d'autres constituants pariétaux sont reconnus par PNA aux phases précoces de la germination. Il pourrait s'agir de polysaccharides dont la distribution ou l'accessibilité sont susceptibles d'être modulés au cours du développement (Freshour et al., 1996 ; Marcus et al., 2010). Des expériences complémentaires devraient être réalisées pour comprendre l'origine des signaux observés. Aux stades plus tardifs de la germination, une bonne correspondance est retrouvée entre les résultats des hybridations d'ARN *in situ* et des marquages par PNA.

Malgré l'obtention de lignées transgéniques homozygotes ARNi sous-expresses et sur-expresses, il n'a pas été possible de mettre en évidence un phénotype macroscopique net au cours du développement. La seule indication concerne les plantes sur-expresses pour lesquelles il a été possible d'observer une phase plus tardive de la germination sur des coupes réalisées dans la paraffine avec une bonne conservation des tissus. Les parois des cellules pourraient être plus résistantes aux traitements nécessaires à l'obtention de ces coupes que celles des plantes sauvages ou sous-expresses. *At1g28290* code une protéine unique dans le génome d'*A. thaliana* avec l'agencement de ses différents domaines : un peptide signal prédit à l'extrémité N-terminale, suivi d'une courte région de type AGP, une région riche en His, un domaine de type PRP, et un domaine C-terminal de type PAC (*cf.* § 1.4.2). Cependant, il existe dans le génome d'*A. thaliana* des gènes qui codent pour certains de ces domaines (*cf.* § 1.4.2). Ces gènes pourraient être co-exprimés avec *At1g28290* et/ou leur expression pourrait être modulée dans les plantes sous- ou sur-expresses. La figure 5.12 présente des données transcriptomiques trouvées dans des banques de données pour un certain nombre de ces gènes. Leur classification hiérarchique montre une co-régulation dans certains organes, dont les hypocotyles pour *At2g33790* codant AGP30 (comportant un peptide signal, des domaines AGP, PRP et PAC) et *At4g08665* (codant une protéine comportant un domaine PAC putatif). Dans les racines, ce sont les gènes *At2g33790* et *At5g10130* (codant une protéine comportant un domaine PAC putatif) qui sont co-régulés.

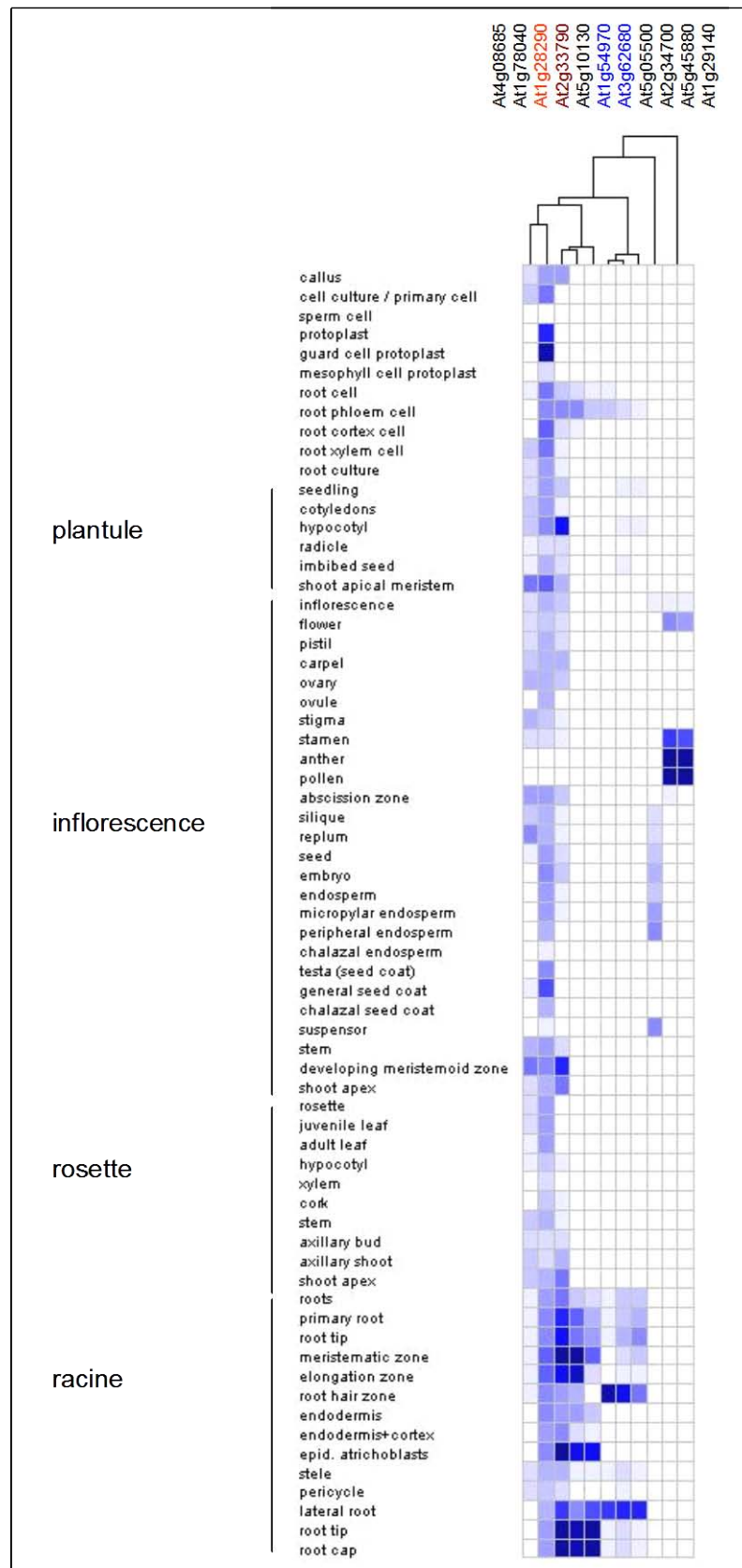


Figure 5.12 : Profil de transcription de gènes codant des protéines de structure proche d'AGP31
At2g33790 code AGP30 ; *At1g54970* et *At3g62680* codent respectivement PRP1 et PRP3 ; *At4g08665*,
At1g78040, *At5g10130*, *At5g05500*, *At5g45880* et *At1g29140* codent des protéines à domaine PAC
 putatif. (données <https://www.genevestigator.com/gv/>).

Concernant la possibilité de dé-régulation de certains de ces gènes dans les plantes sous- ou sur-expressseurs d'*At1g28290*, nous disposons de données préliminaires et partielles qui ne permettent pas de conclure. Là encore, des expériences complémentaires seront nécessaires pour analyser plus finement les plantes sous- ou sur-expressseurs d'*At1g28290*. D'une part, il s'agira de mettre en évidence un éventuel phénotype subtil au niveau de la composition des parois par exemple par des approches de microscopie utilisant des anticorps monoclonaux spécifiques de certains constituants pariétaux (<http://www.ccruc.uga.edu/~mao/wallmab/Antibodies/antib.htm>). D'autre part, il sera nécessaire de regarder si le niveau d'expression des gènes codant des protéines de structure proche d'AGP31 est modifié.

En conclusion, ce travail a permis une première description du patron d'expression d'*At1g28290* au cours de la germination, et l'obtention de lignées homozygotes ARNi sous-expressseurs et de lignées sur-expressseurs qui seront des outils essentiels pour comprendre la fonction d'AGP31 dans les parois. Les premiers résultats suggèrent une fonction de renforcement des parois en réponse à des stress mécaniques et un rôle dans la structuration des parois au cours de l'élongation cellulaire rapide dans les hypocotyles étioilé.

Chapitre 6

CONCLUSIONS ET PERSPECTIVES

6. CONCLUSIONS ET PERSPECTIVES

Ce travail de thèse a porté sur l'étude d'AGP31, une glycoprotéine remarquable de la paroi végétale, codée par le gène *Atlg28290*. L'intérêt pour cette protéine est apparu sur la base d'une triple observation : (i) elle présente une organisation multi-domaine unique chez *A. thaliana*, *i.e.* un peptide signal à l'extrémité N-terminale, un court motif de type AGP, un domaine riche en résidus His, un domaine central riche en résidus Pro/Hyp et un domaine PAC à l'extrémité C-terminale ; (ii) elle est fortement accumulée dans les parois des cellules d'hypocotyles étiolés qui subissent une élongation rapide ; et (iii) sa fonction est inconnue. Nous avons entrepris la caractérisation structurale et fonctionnelle d'AGP31 au travers d'approches expérimentales très diversifiées, telles que la biochimie, la génétique inverse, l'hybridation *in situ* et l'immunocytochimie. La combinaison de ces différentes approches a permis d'apporter des éléments nouveaux sur la structure et la fonction d'AGP31, et de proposer un modèle d'interactions entre AGP31 et des glycoprotéines ou des polysaccharides pariétaux. L'ensemble de ces résultats permet une meilleure compréhension du rôle de cette glycoprotéine complexe dans son contexte pariétal.

Nous avons tout d'abord entrepris **une caractérisation structurale d'AGP31** native, extraite de parois d'hypocotyles étiolés. Deux stratégies de purification, associées à trois méthodes de détection après SDS-PAGE et à différentes approches de spectrométrie de masse, ont permis de révéler la complexité structurale d'AGP31 en caractérisant son profil de *O*-glycosylation. AGP31 est trouvée sous des formes très hétérogènes, avec : (i) un domaine AGP pouvant être occupé ou non par des chaînes d'AG ; (ii) un domaine central riche en Pro/Hyp substitué par des chaînes de galactanes de différentes tailles distribuées de façon homogène sur des résidus Hyp (O) au niveau de séquences répétitives de type KAOV, KSOV, KPOT, KPOV, TPOV et YPOT ; (iii) des motifs SOOA et SOOT de type extensine dans le domaine riche en Pro/Hyp, qui pourraient porter de courts arabinooligosaccharides. Enfin, le domaine PAC C-terminal est absent des formes de plus faibles masses moléculaires d'AGP31. La caractérisation structurale des domaines *O*-glycosylés pourrait être affinée en utilisant une forme tronquée d'AGP31 ne comportant que le domaine PRP. Des analyses biochimiques complémentaires (type de liaisons par perméthylation, digestion enzymatique ou résonance magnétique nucléaire, la localisation putative des résidues Ara à l'extrémité des chaînes de glycanes par digestion alcaline, hydrolyse acide limitée ou incubation par l'arabinofuranosidase) devraient être envisagées pour décrire finement la structure des *O*-glycanes.

Des tests d'interaction *in vitro* entre AGP31 native ou son domaine PAC recombinant et différents polysaccharides ont été mis en œuvre. Après purification des protéines d'intérêt, des interactions avec des β -(1,4)-galactanes ont été mises en évidence. Ces galactanes miment des ramifications du RGI. Ce polymère ramifié est probablement contenu dans les fractions enrichies en hémicelluloses que nous avons extraites de parois végétales. En outre, nous avons montré des interactions avec des pectines méthylestérifiées dans lesquelles il y a sûrement un peu de galactanes, uniquement avec la protéine AGP31 entière. Il pourrait s'agir d'interactions électrostatiques avec le domaine riche en His, absent de la protéine recombinante qui ne comporte que le domaine PAC. Enfin, des interactions entre la protéine AGP31 et le domaine PAC ont été mises en évidence. Ces interactions pourraient impliquer les galactanes du domaine central riche en Pro/Hyp. Ces résultats suggèrent la formation de réseaux non covalents entre AGP31 et différents constituants pariétaux. Ce travail pourrait être poursuivi avec l'obtention de nouvelles formes tronquées d'AGP31 dépourvues du

domaine AGP ou du domaine riche en His par exemple. De même, les Cys conservées du domaine PAC pourraient être mutagénéisées pour évaluer leur rôle dans les interactions avec les galactanes, même si les résultats des tests de liaison aux polysaccharides *in vitro* ne sont pas modifiés en conditions réductrices.

Une approche fonctionnelle *in vivo* a été menée avec deux volets principaux : (i) la caractérisation du patron d'expression d'*At1g28290* en combinant des approches de RT-PCR, d'hybridation *in situ* et d'immunocytochimie ; (ii) l'obtention de lignées transgéniques homozygotes ARNi sous-expressseurs ou sur-expressseurs et la recherche de phénotype au cours du développement. *At1g28290* est exprimé dans les cellules du méristème apical, du protoderme de la racine et de l'endoderme dès les premières heures de la germination. Après cinq jours de germination, les niveaux élevés d'accumulation des transcrits et d'AGP31 dans les hypocotyles étioilés laissent penser que l'ensemble des cellules exprime *At1g28290*. Ces résultats permettent d'envisager deux fonctions pour AGP31 : (i) un renforcement des parois des cellules soumises à des contraintes mécaniques au moment de l'émergence de la racine à un stade précoce de la germination ; puis, (ii) du fait de ses nombreuses possibilités d'interactions avec elle-même ou des constituants pariétaux, un rôle dans l'assemblage ou l'organisation des polymères dans la paroi des hypocotyles en croissance rapide à l'obscurité. L'analyse des lignées ARNi sous-expressseurs et des lignées sur-expressseurs n'a pas permis de mettre en évidence un phénotype macroscopique clair. Cependant, l'analyse de coupes de plantules sur-expressseurs a permis de visualiser une phase plus tardive de la germination avec une bonne conservation des tissus, suggérant une meilleure résistance des parois ayant une sur-accumulation d'AGP31. Enfin, il est possible que cette absence de phénotype clair soit due à une redondance fonctionnelle entre *At1g28290* et *At2g33790* codant AGP30, une protéine pariétale homologue, d'une part, et entre *At1g28290* et des gènes codant des protéines à domaine PAC putatif d'autre part. L'analyse de doubles mutants pourrait être envisagée.

L'ensemble de ces résultats a été obtenu grâce à la **mise en place de nouveaux outils** qui seront très utiles dans le cadre de la poursuite des recherches sur AGP31. Nous disposons aujourd'hui de constructions permettant la production d'AGP31 entière ou de son domaine PAC dans des systèmes procaryotes et/ou eucaryotes. Les protocoles de production et de purification sont bien maîtrisés au laboratoire et pourront être utilisés pour obtenir des lots de protéines purifiées en quantités suffisantes pour des analyses structurales ou fonctionnelles complémentaires. Aussi, nos travaux ont permis de révéler une remarquable affinité entre AGP31 et la lectine PNA qui a été mise à profit d'une part pour purifier la protéine native en une seule étape chromatographique et d'autre part pour la détecter. PNA, couplée à la digoxigénine, a ainsi pu être utilisée pour les analyses *lectin blot* d'extraits protéiques, pour les essais d'interaction sur membrane *in vitro* ainsi qu'en immunocytochimie. La détection d'AGP31 par PNA s'est avérée très spécifique pour les essais *in vitro*, en accord avec le faible nombre de protéines contaminantes obtenus après chromatographie d'affinité sur colonne d'agarose couplée à PNA d'extraits de protéines pariétales. Aucun des polysaccharides utilisés pour les tests d'interaction sur membrane ne générerait de signal. A l'inverse, des signaux non spécifiques sont apparus sur des coupes de plantules à des stades très précoces de la germination, suggérant des interactions entre PNA et d'autres constituants cellulaires, présents ou accessibles uniquement à ces stades de développement. Cette reconnaissance d'AGP31 par PNA suggère des interactions entre AGP31 et des lectines dans les parois. Une recherche de protéines homologues à PNA dans le génome d'*A. thaliana* permet d'ailleurs d'identifier plusieurs protéines ayant un domaine lectine putatif, dont celle codée par le gène *At1g53070* qui possède un domaine fonctionnel prédit de type « lectine de légumineuse »

(PF00139) et qui, de façon très intéressante, est co-régulé avec *At1g28290* (Figure 6.1). De telles lectines pourraient être testées pour leur capacité de reconnaissance d'AGP31.

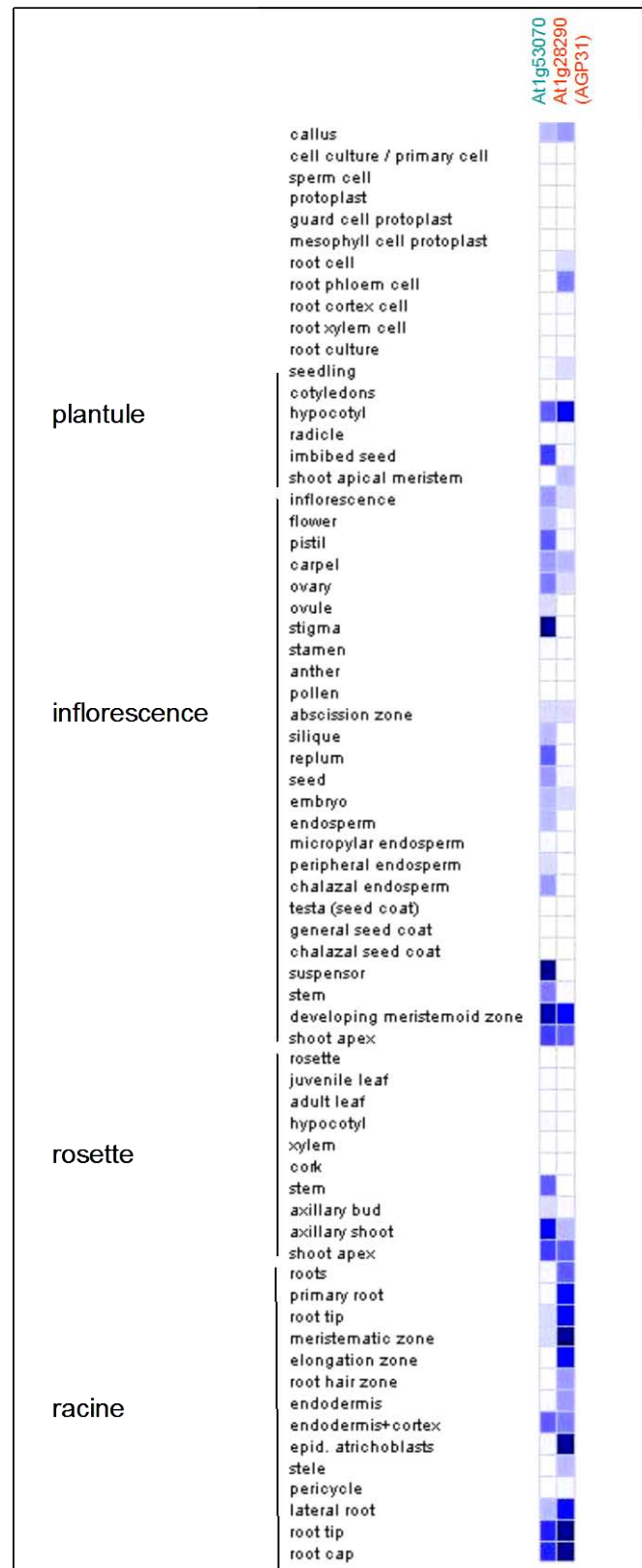


Figure 6.1: Co-régulation au cours du développement des gènes *At1g28290* codant AGP31 et *At1g53070* codant une protéine de type lectine légumineuse (données <https://www.genevestigator.com/gv/>).

Enfin, l'obtention des lignées ARNi sous-expresseurs et sur-expresseurs a représenté un travail de longue durée qui nous permet de disposer aujourd'hui d'un ensemble de lignées bien caractérisées affectées différemment dans l'expression d'*At1g28290*. Ces lignées seront précieuses pour la poursuite de l'étude fonctionnelle.

L'ensemble des données apportées par ce travail permet de proposer **un modèle d'interactions non covalentes entre AGP31 et d'autres constituants pariétaux**. Ainsi, les différents domaines d'AGP31 pourraient interagir avec ceux-ci : le domaine riche en His avec des HG méthylestérifiés ; le domaine riche en Pro/Hyp avec des lectines pariétales et/ou le domaine PAC ; le domaine PAC avec les *O*-galactanes du domaine riche en Pro/Hyp et/ou des galactanes du RGI. Toutes ces interactions contribueraient à former un réseau non covalent dans les parois. Ce type de réseau faisant intervenir plusieurs acteurs est très différent des réseaux covalents décrits pour des protéines structurales de types extensine (Lamport et al., 2011). En effet, ces protéines sont liées entre elles par des liaisons di-tyrosine ou iso-dityrosine (Brady et al., 1998) et semblent liées à des pectines par des liaisons non précisément caractérisées (Qi et al., 1995). Ces assemblages forment des *dendritic scaffolds* visualisables en microscopie à force atomique (Cannon et al., 2008). Les réseaux formés par AGP31 pourraient être beaucoup plus flexibles et plus diversifiés. Ils pourraient faire intervenir AGP30, une protéine très proche d'AGP31 et plusieurs types de lectines. Ce modèle pourra être approfondi par des approches *in vitro* en construisant des protéines de structures moins complexes qu'AGP31 et en identifiant des lectines pariétales interagissant avec ces protéines.

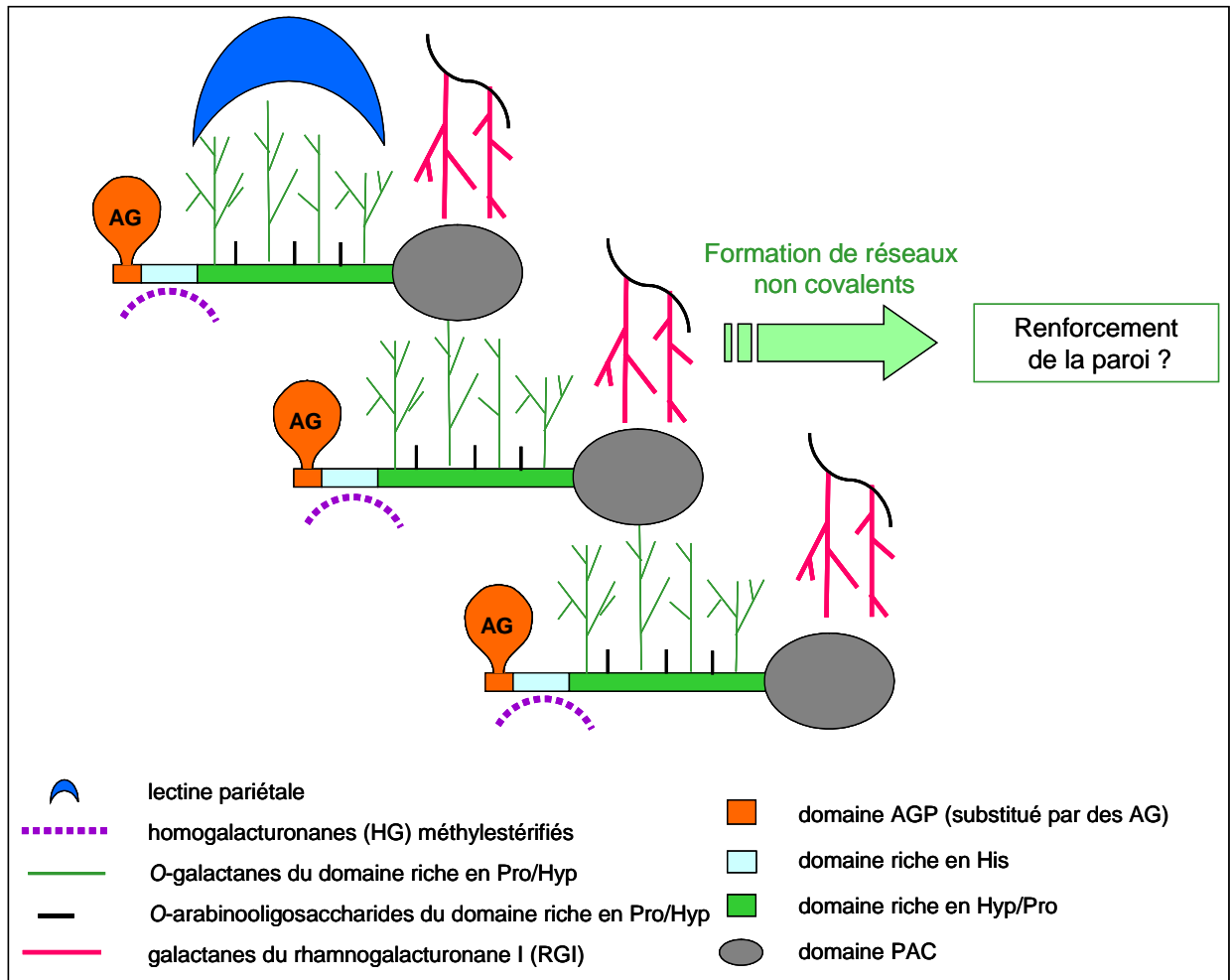


Figure 6.2: Modèle d'interactions non covalentes entre AGP31 et différents constituants pariétaux

La représentation schématique des différents domaines de la protéine AGP31 tient compte de leur possible conformation spatiale prédite *in silico*, à savoir une forme désordonnée pour tous les domaines excepté le domaine PAC supposé adopter une structure plus globulaire.

REFERENCES
BIBLIOGRAPHIQUES

- Acosta-Garcia G, Vielle-Calzada JP** (2004) A classical arabinogalactan protein is essential for the initiation of female gametogenesis in *Arabidopsis*. *Plant Cell* **16**: 2614-2628
- Ahn JW, Verma R, Kim M, Lee JY, Kim YK, Bang JW, Reiter WD, Pai HS** (2006) Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. *J Biol Chem* **281**: 13708-13716
- Aizat W, Preuss JM, Johnson A, Tester M, Schultz C** (2011) Investigation of a His-rich arabinogalactan-protein for micronutrient biofortification of cereal grain. *Physiologia Plantarum* DOI: [10.1111/j.1399-3054.2011.01499.x](https://doi.org/10.1111/j.1399-3054.2011.01499.x)
- Akiyama Y, Mori M, Kato K** (1980) C-13-NMR analysis of hydroxyproline arabinosides from *Nicotiana tabacum*. *Agric Biol Chem* **44**: 2487-2489
- Albenne C, H C, Boudart G, Zhang Y, San Clemente H, Pont-Lezica R, Jamet E** (2009) Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. *Mol Plant* **2**: 977-989
- Aldington S, Fry SC** (1993) Oligosaccharins. *Adv Bot Res* **19**: 1-101
- Anderson RL, Clarke AE, Jermyn MA, Knox RB, Stone BA** (1977) Carbohydrate-binding arabinogalactan-protein from liquid suspension cultures of endosperm from *Lolium multiflorum*. *Austr J Plant Physiol* **4**: 143-158
- Averyhart-Fullard V, Datta K, Marcus A** (1988) A hydroxyproline-rich protein in the soybean cell wall. *Proc Natl Acad Sci USA* **85**: 1082-1085
- Bacic A, Churms SC, Stephen AM, Cohen PB, Fincher GB** (1987) Fine structure of the arabinogalactan-protein from *Lolium multiflorum*. *Carbohydr Res* **162**: 85-93
- Bacic A, Currie G, Gilson P, Mau SL, Oxley D, Schultz C, Sommer-Knudsen J, Clarke AE** (2000) Structural classes of arabinogalactan-proteins. In E Nothnagel, A Bacic, A Clarke, eds, *Cell and Developmental Biology of Arabinogalactan-Proteins*. Kluwer Academic/Plenum Publishers, New York, pp 11-23
- Baldwin T, McCann M, Roberts K** (1993) A novel hydroxyproline-deficient arabinogalactan protein secreted by suspension-cultured cells of *Daucus carota*: purification and partial characterization. *Plant Physiol* **103**: 115-123
- Baldwin T, Hengel A, Roberts K** (2000) The C-terminal PAC domain of a secreted arabinogalactan protein from carrot defines a family of basic proline-rich proteins. In *Cell and Developmental Biology of Arabinogalactan Proteins*. Kluwer Academic Publishers, New York, pp 43-50
- Baldwin T, Domingo C, Schindler T, Seetharaman G, Stacey N, Roberts K** (2001) DcAGP1, a secreted arabinogalactan protein, is related to a family of basic proline-rich proteins. *Plant Mol Biol* **45**: 421-435
- Barton MK, Poethig RS** (1993) Formation of the shoot apical meristem in *Arabidopsis*

- thaliana*, an analysis of development in the wild type and in the shoot meristemless mutant. *Development* **119**: 823-831
- Battaglia M, Solorzano RM, Hernandez M, Cuellar-Ortiz S, Garcia-Gomez B, Marquez J, Covarrubias AA** (2007) Proline-rich cell wall proteins accumulate in growing regions and phloem tissue in response to water deficit in common bean seedlings. *Planta* **225**: 1121-1133
- Baud F, Pebay-Peyroula E, Cohen-Addad C, Odani S, Lehmann M** (1993) Crystal structure of hydrophobic protein from soybean; a member of a new cysteine-rich family. *J Mol Biol* **231**: 877-887
- Baulcombe DC, Molnar A** (2004) Crystal structure of p19 - a universal suppressor of RNA silencing. *Trends Biochem Sci* **29**: 279-281
- Bernhardt C, Tierney ML** (2000) Expression of AtPRP3, a proline-rich structural cell wall protein from *Arabidopsis*, is regulated by cell-type-specific developmental pathways involved in root hair formation. *Plant Physiol* **122**: 705-714
- Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerré-Tugayé MT, Boudet A, Pont-Lezica R** (2003) Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis. *Electrophoresis* **24**: 3421-3432
- Borner GH, Lilley KS, Stevens TJ, Dupree P** (2003) Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. *Plant Physiol* **132**: 568-577
- Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé M-T, Pont-Lezica R** (2005) Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics. *Proteomics* **5**: 212-221
- Bouton S, Leboeuf E, Mouille G, Leydecker MT, Talbotec J, Granier F, Lahaye M, Hofte H, Truong HN** (2002) *QUASIMODO1* encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in *Arabidopsis*. *Plant Cell* **14**: 2577-2590
- Bradford MM** (1976) Rapid and sensitive method for quantification of microgram quantities of protein utilizing principle of protein-dye binding. *Anal Biochem* **72**: 248-254
- Bradley DJ, Kjellbom P, Lamb CJ** (1992) Elicitor-induced and wound induced oxidative cross-linking of a proline rich plant cell wall protein: a novel, rapid defense response. *Cell* **70**: 21-30
- Brady J, Sadler I, Fry S** (1998) Pulcherosine, an oxidatively coupled trimer of tyrosine in plant cell walls: its role in cross-link formation. *Phytochemistry* **47**: 349-353
- Brett CT** (2000) Cellulose microfibrils in plants: Biosynthesis, deposition, and integration into the cell wall. *Int Rev Cytol* **199**: 161-199

- Brewer PB, Heisler MG, Hejatko J, Friml J, Benkova E** (2006) *In situ* hybridization for mRNA detection in Arabidopsis tissue sections. *Nat Protoc* **1**: 1462-1467
- Brisson LF, Tenhaken R, Lamb C** (1994) Function of oxidative cross linking of cell wall structural proteins in plant disease resistance. *Plant Cell* **6**: 1703-1712
- Buchala A, Wilkie K** (1973) Total hemicellulose from wheat at different stages of growth. *Phytochemistry* **12**: 499-505
- Buchanan B, Gruissem W, Jones R** (2000) *Biochemistry & Molecular Biology of Plants*. American Society of Plant Physiologists, Rockville, Maryland
- Burlat V, Oudin A, Courtois M, Rideau M, St-Pierre B** (2004) Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J* **38**: 131-141
- Cannon M, Terneus K, Hall Q, Tan L, Wang Y, Wegenhart B, Chen L, Lamport D, Chen Y, Kieliszewski M** (2008) Self-assembly of the plant cell wall requires an extensin scaffold. *Proc Natl Acad Sci USA* **105**: 2226-2231
- Carpita NC, Gibeaut DM** (1993) Structural models of primary cell walls in flowering plants, consistency of molecular structure with the physical properties of the walls during growth. *Plant J* **3**: 1-30
- Cassab GI** (1998) Plant cell wall proteins. *Annu Rev Plant Physiol Plant Mol Biol* **49**: 281-309
- Chen CG, Pu ZY, Moritz RL, Simpson RJ, Bacic A, Clarke AE, Mau SL** (1994) Molecular cloning of a gene encoding an arabinogalactan-protein from pear (*Pyrus communis*) cell suspension culture. *Proc Natl Acad Sci USA* **91**: 10305-10309
- Cheung AY, Wang H, Wu HM** (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. *Cell* **82**: 383-393
- Coimbra S, Costa M, Jones B, Mendes MA, Pereira LG** (2009) Pollen grain development is compromised in Arabidopsis *agp6 agp11* null mutants. *J Exp Bot* **60**: 3133-3142
- Coimbra S, Costa M, Mendes MA, Pereira AM, Pinto J, Pereira LG** (2010) Early germination of Arabidopsis pollen in a double null mutant for the arabinogalactan protein genes *AGP6* and *AGP11*. *Sex Plant Reprod* **23**: 199-205
- Collett CE, Harberd NP, Leyser O** (2000) Hormonal interactions in the control of Arabidopsis hypocotyl elongation. *Plant Physiol* **124**: 553-561
- Cosgrove D** (1994) Photomodulation of growth by blue light. *Plant Physiol* **67**: 584-590
- Cosgrove DJ** (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol* **6**: 850-861

- Darvill AG, Albersheim P, McNeil M, Lau JM, York WS, Stevenson TT, Thomas J, Doares S, Gollin DJ, Chelf P, Davis K** (1985) Structure and function of plant cell wall polysaccharides. *J Cell Sci Suppl* **2**: 203-217
- Darvill A, Augur C, Bergmann C, Carlson RW, Cheong JJ, Eberhard S, Hahn MG, Lo VM, Marfa V, Meyer B, Mohnen D, O'Neill MA, Spiro MD, Vanhalbeek H, York WS, Albersheim P** (1992) Oligosaccharins-oligosaccharides that regulate growth, development and defense responses in plants. *Glycobiology* **2**: 181-198
- Datta K, Schmidt A, Marcus A** (1989) Characterization of two soybean repetitive proline-rich proteins and a cognate cDNA from germinated axes. *Plant Cell* **1**: 945-952
- Derbyshire P, Findlay K, McCann MC, Roberts K** (2007) Cell elongation in Arabidopsis hypocotyls involves dynamic changes in cell wall thickness. *J Exp Bot* **58**: 2079-2089
- Ding L, Zhu JK** (1997) A role for arabinogalactan-proteins in root epidermal cell expansion. *Planta* **203**: 289-294
- Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP** (2002) Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiol* **43**: 1407-1420
- Du H, Simpson RJ, Moritz RL, Clarke AE, Bacic A** (1994) Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana glauca* and characterization of a corresponding cDNA. *Plant Cell* **6**: 1643-1653
- Dvorakova L, Cvrckova F, Fischer L** (2007) Analysis of the hybrid proline-rich protein families from seven plant species suggests rapid diversification of their sequences and expression patterns. *BMC Genomics* **8**: 412
- Dyson HJ, Wright PE** (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**: 197-208
- Elkins T, Hortsch M, Bieber AJ, Snow PM, Goodman CS** (1990) Drosophila fasciclin-I is a novel homophilic adhesion molecule that along with fasciclin-III can mediate cell sorting. *J Cell Biol* **110**: 1825-1832
- Ellis M, Egelund J, Schultz CJ, Bacic A** (2010) Arabinogalactan-proteins: key regulators at the cell surface? *Plant Physiol* **153**: 403-419
- Emons AMC, Mulder BM** (2000) How the deposition of cellulose microfibrils builds cell wall architecture. *Trends Plant Sci* **5**: 35-40
- Ermel FF, Follet-Gueye ML, Cibert C, Vian B, Morvan C, Catesson AM, Goldberg R** (2000) Differential localization of arabinan and galactan side chains of rhamnogalacturonan 1 in cambial derivatives. *Planta* **210**: 732-740
- Estevez J, Kieliszewski M, Khitrov N, Somerville C** (2006) Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in Arabidopsis. *Plant Physiol* **142**: 458-470

- Eudes A, Mouille G, Thévenin J, Goyallon A, Minic Z, Jouanin L** (2008) Purification, cloning and functional characterization of an endogenous beta-glucuronidase in *Arabidopsis thaliana*. *Plant Cell Physiol* **49**: 1331-1341
- Faye L, Boulaflous A, Benchabane M, Gomord V, Michaud D** (2005) Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* **23**: 1770-1778
- Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E** (2006) Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls. *Plant Methods* **2**: 10
- Fiete DJ, Beranek MC, Baenziger JU** (1998) A cysteine-rich domain of the "mannose" receptor mediates GalNAc-4-SO₄ binding. *Proc Natl Acad Sci USA* **95**: 2089-2093
- Fincher GB, Sawyer WH, Stone BA** (1974) Chemical and physical properties of an arabinogalactan-peptide from wheat endosperm. *Biochem J.* **139**: 535-545
- Fincher GB, Stone BA, Clarke AE** (1983) Arabinogalactan protein: structure, biosynthesis and function. *Annu Rev Plant Physiol Plant Mol Biol* **34**: 47-70
- Finn R, Tate J, Mistry J, Coghill P, Sammut J, Hotz H, Ceric G, Forslund K, Eddy S, Sonnhammer E, Bateman A** (2008) The PFAM protein families database. *Nucleic Acids Res* **36**: D281-D288
- Fleischer A, O'Neill MA, Ehwald R** (1999) The pore size of non-graminaceous plant cell walls is rapidly decreased by borate ester cross-linking of the pectic polysaccharide rhamnogalacturonan II. *Plant Physiol* **121**: 829-838
- Fowler TJ, Bernhardt C, Tierney ML** (1999) Characterization and expression of four proline-rich cell wall protein genes in *Arabidopsis* encoding two distinct subsets of multiple domain proteins. *Plant Physiol* **121**: 1081-1092
- Freshour G, Clay RP, Fuller MS, Albersheim P, Darvill AG, Hahn MG** (1996) Developmental and tissue-specific structural alterations of the cell-wall polysaccharides of *Arabidopsis thaliana* roots. *Plant Physiol* **110**: 1413-1429
- Freyermuth SK, Bacanamwo M, Polacco JC** (2000) The soybean *Eu3* gene encodes an Ni-binding protein necessary for urease activity. *Plant J* **21**: 53-60
- Fry SC** (1988) *The Growing Plant Cell Wall: Chemical and Metabolic Analysis*. Logman, London
- Fry SC, York WS, Albersheim P, Darvill A, Hayashi T, Joseleau JP, Kato Y, Lorences EP, Maclachlan GA, McNeil M, Mort AJ, Reid JSG, Seitz HU, Selvendran RR, Voragen AGJ, White AR** (1993) An unambiguous nomenclature for xyloglucan derived oligosaccharides. *Physiologia Plantarum* **89**: 1-3
- Fry SC** (1995) Polysaccharide modifying enzymes in the plant cell wall. *Annu Rev Plant Physiol Plant Mol Biol* **46**: 497-520

- Fry SC** (1998) Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem J* **332**: 507-515
- Fry SC, Willis SC, Paterson AEJ** (2000) Intraprotoplasmic and wall-localised formation of arabinoxylan-bound diferulates and larger ferulate coupling-products in maize cell-suspension cultures. *Planta* **211**: 679-692
- Ganjian I, Basile DV** (1997) Reductive syntheses of p-aminophenyl-beta-D-glucoside and its conversion to beta-glucosyl Yariv reagent. *Anal Biochem* **246**: 152-155
- Gao MG, Kieliszewski MJ, Lamport DTA, Showalter AM** (1999) Isolation, characterization and immunolocalization of a novel, modular tomato arabinogalactan-protein corresponding to the *LeAGP-1* gene. *Plant J* **18**: 43-55
- Gaspar Y, Johnson K, McKenna J, Bacic A, Schultz C** (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Mol Biol* **47**: 161-176
- Gaspar YM, Nam J, Schultz CJ, Lee LY, Gilson PR, Gelvin SB, Bacic A** (2004) Characterization of the *Arabidopsis* lysine-rich arabinogalactan-protein *AtAGP17* mutant (*rat1*) that results in a decreased efficiency of *Agrobacterium* transformation. *Plant Physiol* **135**: 2162-2171
- Gendreau E, Traas J, Desnos T, Grandjean O, Caboche M, Hofte H** (1997) Cellular basis of hypocotyl growth in *Arabidopsis thaliana*. *Plant Physiol* **114**: 295-305
- Gibeaut D, Carpita N** (1991) Tracing cell wall biogenesis in intact cells and plants. *Plant Physiol* **97**: 551-561
- Gilson P, Gaspar YM, Oxley D, Youl JJ, Bacic A** (2001) NaAGP4 is an arabinogalactan protein whose expression is suppressed by wounding and fungal infection in *Nicotiana glauca*. *Protoplasma* **215**: 128-139
- Gincel F, Pebay-Peyroula F, Cohen-Addad C, Odani S, Lehmann M** (1994) Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional ¹H-NMR data. *Eur J Biochem* **226**: 413-422
- Gleeson PA, Clarke AE** (1979) Structural studies on the major component of *Gladiolus* style mucilage, an arabinogalactan-protein. *Biochem J* **181**: 607-621
- Gohlke H, Kuhn LA, Case DA** (2004) Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* **56**: 322-337
- Gollner EM, Blaschek W, Classen B** (2010) Structural investigations on arabinogalactan-protein from wheat, isolated with Yariv reagent. *J Agric Food Chem* **58**: 3621-3626
- Golovchenko VV, Ovodova RG, Shashkov AS, Ovodov YS** (2002) Structural studies of the pectic polysaccharide from duckweed *Lemna minor* L. *Phytochemistry* **60**: 89-97

- Goubet F, Bourlard T, Girault R, Alexandre C, Vandeveld MC, Morvan C** (1995) Structural features of galactans from flax fibers. *Carbohydr Polymers* **27**: 221-227
- Guan Y, Nothnagel EA** (2004) Binding of arabinogalactan proteins by Yariv phenylglycoside triggers wound-like responses in *Arabidopsis* cell cultures. *Plant Physiol* **135**: 1346-1366
- Guzzardi P, Genot G, Jamet E** (2004) The *Nicotiana sylvestris* extensin gene, *Ext 1.2A*, is expressed in the root transition zone and upon wounding. *Biochim Biophys Acta* **1680**: 83-92
- Hall Q, Cannon MC** (2002) The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in *Arabidopsis*. *Plant Cell* **14**: 1161-1172
- Hara M, Fujinaga M, Kuboi T** (2005) Metal binding by citrus dehydrin with histidine-rich domains. *J Exp Bot* **56**: 2695-2703
- Hata K, Tanaka M, Tsumuraya Y, Hashimoto Y** (1992) α -L-Arabinofuranosidase from radish (*Raphanus sativus* L.) seeds. *Plant Physiol* **100**: 388-396
- Hellens R, Mullineaux P, Klee H** (2000) A guide to *Agrobacterium* binary Ti vectors. *Trends Plant Sci* **5**: 446-451
- Hengel AV, Tadesse Z, Immerzeel P, Schols H, Kammen AV, Vries SD** (2001) *N*-acetylglucosamine and glucosamine-containing arabinogalactan proteins control somatic embryogenesis. *Plant Physiol* **117**: 1880-1890
- Hieta R, Myllyharju J** (2002) Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. *J Biol Chem* **277**: 23965-23971
- Hirsinger C, Salva I, Marbach J, Durr A, Fleck J, Jamet E** (1999) The tobacco extensin gene *Ext 1.4* is expressed in cells submitted to mechanical constraints and in cells proliferating under hormone control. *J Exp Bot* **50**: 343-355
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P** (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* **2008**: 420747
- Irshad M, Canut H, Borderies G, Pont-Lezica R, Jamet E** (2008) A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biol* **8**: 94
- Jamet E, Canut H, Boudart G, Pont-Lezica R** (2006) Cell wall proteins: a new insight through proteomics. *Trends Plant Sci* **11**: 33-39
- Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R** (2008) Recent advances in plant cell wall proteomics. *Proteomics* **8**: 893-908
- Jamet E, Roujol D, San Clemente H, Irshad M, Soubigou-Taconnat L, Renou J-P, Pont-Lezica R** (2009) Cell wall biogenesis of *Arabidopsis thaliana* elongating cells:

- transcriptomics complements proteomics. *BMC Genomics* **10**: 505
- Jeyaprakash A, Jayashree G, Mahanta S, Swaminathan C, Sekar K, Surolia A, Vijayan M** (2005) Structural basis for the energetics of jacalin-sugar interactions: promiscuity versus specificity. *J Mol Biol* **347**: 181-188
- Jiang LX, Yang SL, Xie LF, Puah CS, Zhang XQ, Yang WC, Sundaresan V, Ye D** (2005) *VANGUARDI* encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract. *Plant Cell* **17**: 584-596
- Johnson KL, Jones BJ, Schultz CJ, Bacic A** (2003) Non enzymic cell wall (glyco)proteins. In JK Rose, ed, *The Plant Cell Wall*. Blackwell Publishing, CRC Press, Boca Raton, pp 111-154
- Kieliszewski M, Leykam J, Lamport D** (1990) Structure of the threonine-rich extensin from *Zea mays*. *Plant Physiol* **92**: 316-326
- Kieliszewski M, de Zacks R, Leykam J, Lamport D** (1992) A repetitive proline-rich protein from the gymnosperm douglas fir is a hydroxyproline-rich glycoprotein. *Plant Physiol* **98**: 919-926
- Kieliszewski M, Lamport D** (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J* **5**: 157-172
- Kieliszewski M** (2001) The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **57**: 319-323
- Kjellbom P, Snogerup L, Stohr C, Reuzeau C, McCabe PF, Pennell RI** (1997) Oxidative cross-linking of plasma membrane arabinogalactan proteins. *Plant J* **12**: 1189-1196
- Kleis-San Francisco S, Tierney M** (1990) Isolation and characterization of a proline-rich cell wall protein from soybean seedlings. *Plant Physiol* **94**: 1897-1902
- Kobayashi Y, Motose H, Iwamoto K, Fukuda H** (2011) Expression and genome-wide analysis of the xylogen-type gene family. *Plant Cell Physiol* **52**: 1095-1106
- Komalavilas P, Mort AJ** (1989) The acetylation at O-3 of galacturonic acid in the rhamnose-rich portion of pectins. *Carbohydr Res* **189**: 261-272
- Koski M, Hieta R, Hirsila M, Ronka A, Myllyharju J, Wierenga R** (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem* **284**: 25290-25301
- Kotake T, Dina S, Konishi T, Kaneko S, Igarashi K, Samejima M, Watanabe Y, Kimura K, Tsumuraya Y** (2005) Molecular cloning of a β -galactosidase from radish that specifically hydrolyzes β -(1-3)- and β -(1-6) galactosyl residues of arabinogalactan protein. *Plant Physiol* **138**: 1563-1576
- Kotake T, Tsuchiya K, Aohara T, Konishi T, Kaneko S, Igarashi K, Samejima M, Tsumuraya Y** (2006) An α -L-arabinofuranosidase/ β -D-xylosidase from immature seeds

- of radish (*Raphanus sativus* L.). J Exp Bot **57**: 2353-2362
- Kruger C, Berkowitz O, Stephan U, Hell R** (2002) A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of *Ricinus communis* L. J Biol Chem **277**: 25062-25069
- Laemmli UK** (1970) Cleavage of the structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685
- Lampert D, Northcote D** (1960) Hydroxyproline in primary cell walls of higher plants. Nature **188**: 665-666
- Lampert D** (1965) The protein component of primary cell walls. Adv Bot Res **2**: 151-218
- Lampert DTA** (1967) Hydroxyproline-*O*-glycosidic linkage of the plant cell wall glycoprotein extensin. Nature **216**: 1322-1324
- Lampert DTA** (1973) Galactosylserine in extensin. Biochem J **133**: 125-131
- Lampert DTA** (1980) Structure and function of plant glycoproteins. In J Preiss, ed, The Biochemistry of Plants, Vol 3. Academic Press, New York, pp 501-541
- Lampert DTA, Kieliszewski MJ, Showalter AM** (2006) Salt stress upregulates periplasmic arabinogalactan proteins: using salt stress to analyse AGP function. New Phytol **169**: 479-492
- Lampert DT, Kieliszewski M, Chen Y, Cannon MC** (2011) Role of the extensin superfamily in primary cell wall architecture. Plant Physiol **156**: 11-19
- Lee KJD, Sakata Y, Mau SL, Pettolino F, Bacic A, Quatrano RS, Knight CD, Knox JP** (2005) Arabinogalactan proteins are required for apical cell extension in the moss *Physcomitrella patens*. Plant Cell **17**: 3051-3065
- Lerouge P, O'Neill MA, Darvill AG, Albersheim P** (1993) Structural characterization of endo-glycanase-generated oligoglycosyl side chains of rhamnogalacturonan I. Carbohydr Research **243**: 359-371
- Léonard R, Petersen B, Himly M, Kaar W, Wopfner N, Kolarich D, van Ree R, Ebner C, Duus J, Ferreira F, Altmann F** (2005) Two novel types of *O*-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. J Biol Chem **280**: 7932-7940
- Léonard R, Wopfner N, Pabst M, Stadlmann J, Petersen B, Duus J, Himly M, Radauer C, Gadermaier G, Razzazi-Fazeli E, Ferreira F, Altmann F** (2010) A new allergen from ragweed (*Ambrosia artemisiifolia*) with homology to Art v 1 from mugwort. J Biol Chem **285**: 27192-27200
- Li YQ, Faleri C, Geitmann A, Zhang HQ, Cresti M** (1995) Immunogold localization of arabinogalactan proteins, unesterified and esterified pectins in pollen grains and pollen tubes of *Nicotiana tabacum* L. Protoplasma **189**: 26-36

- Li SX, Showalter AM** (1996) Cloning and developmental/stress-regulated expression of a gene encoding a tomato arabinogalactan protein. *Plant Mol Biol* **32**: 641-652
- Li YC, Yang YC, Hsu JSF, Wu DJ, Wu HH, Tzen JTC** (2005) Cloning and immunolocalization of an antifungal chitinase in jelly fig (*Ficus awkeotsang*) achenes. *Phytochemistry* **66**: 879-886
- Liang Y, Faik A, Kieliszewski M, Tan L, Xu W-L, Showalter A** (2010) Identification and characterization on *in vitro* galactosyltransferase activities involved in arabinogalactan-protein glycosylation in tobacco and Arabidopsis. *Plant Physiol* **154**: 632-642
- Lind JL, Bacic A, Clarke AE, Anderson MA** (1994) A style-specific hydroxyproline rich glycoprotein with properties of both extensins and arabinogalactan proteins. *Plant J* **6**: 491-502
- Liu C, Mehdy M** (2007) A nonclassical arabinogalactan protein gene highly expressed in vascular tissues, *AGP31*, is transcriptionally repressed by methyl jasmonic acid in Arabidopsis. *Plant Physiol* **145**: 863-874
- Logemann E, Birkenbihl RP, Ulker B, Somssich IE** (2006) An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. *Plant Methods* **2**: 16
- MacKinnon IM, Sturcova A, Sugimoto-Shirasu K, His I, McCann MC, Jarvis MC** (2006) Cell-wall structure and anisotropy in *procuste*, a cellulose synthase mutant of *Arabidopsis thaliana*. *Planta* **224**: 438-448
- Marcus S, Blake A, Benians T, Lee K, Poyser C, Donaldson L, Leroux O, Rogowski A, Petersen H, Boraston A, Gilbert H, Willats W, Knox J** (2010) Restricted access of proteins to mannan polysaccharides in intact plant cell walls. *Plant J* **64**: 191-203
- Marshall E, Costa LM, Gutierrez-Marcos J** (2011) Cysteine-Rich Peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J Exp Bot* **62**: 1677-1686
- Matsunaga T, Ishii T, Matsumoto S, Higuchi M, Darvill A, Albersheim P, O'Neill MA** (2004) Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants. *Plant Physiol* **134**: 339-351
- McCann MC, Wells B, Roberts K** (1990) Direct visualization of cross-links in the primary plant cell wall. *J Cell Sci* **96**: 323-334
- McDougall GJ, Fry SC** (1988) Inhibition of auxin-stimulated growth of pea stem segments by a specific nonasaccharide of xyloglucan. *Planta* **175**: 412-416
- McNeil M, Darvill AG, Fry SC, Albersheim P** (1984) Structure and function of the primary cell walls of plants. *Annu Rev Biochem* **53**: 625-663
- McQueen-Mason S, Cosgrove DJ** (1994) Disruption of hydrogen-bonding between plant

- cell wall polymers by proteins that induce wall extension. *Proc Natl Acad Sci USA* **91**: 6574-6578
- Meng L, Wong JH, Feldman LJ, Lemaux PG, Buchanan BB** (2010) A membrane-associated thioredoxin required for plant growth moves from cell to cell, suggestive of a role in intercellular communication. *Proc Natl Acad Sci USA* **107**: 3900-3905
- Merkouropoulos G, Shirsat AH** (2003) The unusual Arabidopsis extensin gene *atExt1* is expressed throughout plant development and is induced by a variety of biotic and abiotic stresses. *Planta* **217**: 356-366
- Minic Z, Jouanin L** (2006) Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol Biochem* **44**: 435-449
- Mohnen D** (1999) Biosynthesis of pectins and galactomannans. *In* B Pinto, ed, Carbohydrates and their Derivatives Including Tannins, Cellulose, and Related Lignins. Elsevier Science Ltd, Oxford, pp 497-527
- Moller I, Sorensen I, Bernal AJ, Blaukopf C, Lee K, Obro J, Pettolino F, Roberts A, Mikkelsen JD, Knox JP, Bacic A, Willats WGT** (2007) High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *Plant J* **50**: 1118-1128
- Mort AJ, Lamport DT** (1977) Anhydrous hydrogen fluoride deglycosylates glycoproteins. *Anal Biochem* **82**: 289-309
- Morvan O, Jauneau A, Morvan C, Voreux H, Demarty M** (1989) Biosynthesis of pectins and differentiation of cellulose fibers during growth of flax. *Can J Bot* **67**: 135-139
- Motose H, Sugiyama M, Fukuda H** (2004) A proteoglycan mediates inductive interaction during plant vascular development. *Nature* **429**: 873-878
- Muschietti J, Dircks L, Vancanneyt G, McCormick S** (1994) LAT52 protein is essential for tomato pollen development, pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilization. *Plant J* **6**: 321-338
- Nam J, Mysore KS, Zheng C, Knue MK, Matthyse AG, Gelvin SB** (1999) Identification of T-DNA tagged Arabidopsis mutants that are resistant to transformation by Agrobacterium. *Mol Gen Genet* **261**: 429-438
- Nothnagel EA, Lyon JL** (1986) Structural requirements for the binding of phenylglycosides to the surface of protoplasts. *Plant Physiol* **80**: 91-98
- Nothnagel EA** (1997) Proteoglycans and related components in plant cells. *Int Rev Cytol* **174**: 195-291
- O'Neill MA, Albersheim P, Darvill A** (1990) The pectic polysaccharides of primary cell walls. *Methods Plant Biochem* **2**: 415-441
- O'Neill MA, Warrenfeltz D, Kates K, Pellerin P, Doco T, Darvill AG, Albersheim P**

- (1996) Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester - *In vitro* conditions for the formation and hydrolysis of the dimer. *J Biol Chem* **271**: 22923-22930
- O'Neill MA, Eberhard S, Albersheim P, Darvill AG** (2001) Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth. *Science* **294**: 846-849
- O'Neill MA, Ishii T, Albersheim P, Darvill AG** (2004) Rhamnogalacturonan II: Structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* **55**: 109-139
- Ohmiya Y, Takeda T, Nakamura S, Sakai F, Hayashi T** (1995) Purification and properties of a wall-bound endo-1,4-beta-glucanase from suspension-cultured poplar cells. *Plant Cell Physiol* **36**: 607-614
- Ohmiya Y, Samejima M, Shiroishi M, Amano Y, Kanda T, Sakai F, Hayashi T** (2000) Evidence that endo-1,4-beta-glucanases act on cellulose in suspension-cultured poplar cells. *Plant J* **24**: 147-158
- Oka T, Saito F, Shimma Y-i, Yoko-o T, Nomura Y, Matsuoka K, Jigami Y** (2010) Characterization of endoplasmic reticulum-localized UDP-D-galactose: hydroxyproline *O*-galactosyltransferase using synthetic peptide substrates in *Arabidopsis*. *Plant Physiol* **152**: 332-340
- Orfila C, Sorensen SO, Harholt J, Geshi N, Crombie H, Truong HN, Reid JSG, Knox JP, Scheller HV** (2005) *QUASIMODO1* is expressed in vascular tissue of *Arabidopsis thaliana* inflorescence stems, and affects homogalacturonan and xylan biosynthesis. *Planta* **222**: 613-622
- Osman ME, Menzies AR, Martin BA, Williams PA, Phillips GO, Baldwin TC** (1995) Characterization of gum arabic fractions obtained by anion exchange chromatography. *Phytochemistry* **38**: 409-417
- Park MH, Suzuki Y, Chono M, Knox JP, Yamaguchi I** (2003) *CsAGPI*, a gibberellin-responsive gene from cucumber hypocotyls, encodes a classical arabinogalactan protein and is involved in stem elongation. *Plant Physiol* **131**: 1450-1459
- Passardi F, Penel C, Dunand C** (2004) Performing the paradoxical: how plant peroxidases modify the cell wall. *Trends Plant Sci* **9**: 534-540
- Paulsen BS, Barsett H** (2005) Bioactive pectic polysaccharides. *Adv Polym Sci* **186**: 69-101
- Pelletier S, Van Orden J, Wolf S, Vissenberg K, Delacourt J, Ndong Y, Pelloux J, Bischoff V, Urbain A, Mouille G, Lemonnier G, Renou J, Höfte H** (2010) A role for pectin de-methylesterification in a developmentally regulated growth acceleration in dark-grown *Arabidopsis* hypocotyls. *New Phytol* **188**: 726-739
- Peng J, Lai L, Wang X** (2008) PRGL: A cell wall proline-rich protein containing GASA domain in *Gerbera hybrida*. *Sci China C Life Sci* **51**: 520-525

- Peng J, Lai L, Wang X** (2010) Temporal and spatial expression analysis of PRGL in *Gerbera hybrida*. *Mol Biol Rep* **37**: 3311-3317
- Pennell RI, Knox JP, Scofield GN, Selvendran RR, Roberts K** (1989) A family of abundant plasma membrane associated glycoproteins related to the arabinogalactan proteins is unique to flowering plants. *J Cell Biol* **108**: 1967-1977
- Perez S, Rodriguez-Carvajal MA, Doco T** (2003) A complex plant cell wall polysaccharide: rhamnogalacturonan II. A structure in quest of a function. *Biochimie* **85**: 109-121
- Pernot C** (2010) Cell wall proteome of *Arabidopsis thaliana* etiolated hypocotyls: Towards a functional analysis through systematic *in situ* hybridization of candidate genes. M2R Biosciences Végétales, Université Paul Sabatier, Toulouse
- Phillips GO** (2009) Molecular association and function of arabinogalactan protein complexes from tree exudates. *Struct Chem* **20**: 309-315
- Ponder GR, Richards GN** (1997) Arabinogalactan from Western larch, Part III: alkaline degradation revisited, with novel conclusions on molecular structure. *Carbohydr Polym* **34**: 251-261
- Qi W, Fong C, Lamport DTA** (1991) Gum arabic glycoprotein is a twisted hairy rope, a new model based on *O*-galactosylhydroxyproline as polysaccharide attachment site. *Plant Physiol* **96**: 848-855
- Qi XY, Behrens BX, West PR, Mort AJ** (1995) Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension-cultures, evidence for a covalent cross-link between extensin and pectin. *Plant Physiol* **108**: 1691-1701
- Qu Y, Egelund J, Gilson P, Houghton F, Gleeson P, Schultz C, Bacic A** (2008) Identification of a novel group of putative *Arabidopsis thaliana* beta-(1,3)-galactosyltransferases. *Plant Mol Biol* **68**: 43-59
- Radivojac P, Obradovic Z, Smith DK, Zhu G, Vucetic S, Brown CJ, Lawson JD, Dunker AK** (2004) Protein flexibility and intrinsic disorder. *Protein Sci* **13**: 71-80
- Rambo RP, Tainer JA** (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**: 559-571
- Refrégier G, Pelletier S, Jaillard D, Hofte H** (2004) Interaction between wall deposition and cell elongation in dark-grown hypocotyl cells in *Arabidopsis*. *Plant Physiol* **135**: 959-968
- Reid B, Howell S** (1995) The functioning of hormones in plant growth and development. In P Davies, ed, *Plant Hormones: Physiology, Biochemistry and Molecular Biology*. Dordrecht, The Netherlands, pp 448-485
- Reiter WD** (2002) Biosynthesis and properties of the plant cell wall. *Curr Opin Plant Biol* **5**: 536-542

- Renard C, Weightman RM, Thibault JF** (1997) The xylose-rich pectins from pea hulls. *Int J Biol Macromol* **21**: 155-162
- Renard C, Crepeau MJ, Thibault JF** (1999) Glucuronic acid directly linked to galacturonic acid in the rhamnogalacturonan backbone of beet pectins. *Eur J Biochem* **266**: 566-574
- Ridley BL, O'Neill MA, Mohnen DA** (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **57**: 929-967
- Rockel N, Wolf S, Kost B, Rausch T, Greiner S** (2008) Elaborate spatial patterning of cell-wall PME and PME1 at the pollen tube tip involves PME1 endocytosis, and reflects the distribution of esterified and de-esterified pectins. *Plant J* **53**: 133-143
- Rodriguez-Carvajal MA, du Penhoat CH, Mazeau K, Doco T, Perez S** (2003) The three-dimensional structure of the mega-oligosaccharide rhamnogalacturonan II monomer: a combined molecular modeling and NMR investigation. *Carbohydr Res* **338**: 651-671
- Roland J** (1980) Visualisation des polysaccharides au niveau ultrastructural. *In* B Monties, AM Catesson, JC Roland, F Barnoud, eds, *Les Polymères Végétaux: Polymères Pariétaux et Alimentaires non Azotés*. Gauthier-Villars, Paris, pp 30-45
- Rose JK, Braam J, Fry SC, Nishitani K** (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant Cell Physiol* **43**: 1421-1435
- Rubinovich L, Weiss D** (2010) The Arabidopsis cysteine-rich protein GASA4 promotes GA responses and exhibits redox activity in bacteria and *in planta*. *Plant J* **64**: 1018-1027
- Saibo NJ, Vriezen WH, Beemster GT, Van Der Straeten D** (2003) Growth and stomata development of Arabidopsis hypocotyls are controlled by gibberellins and modulated by ethylene and auxins. *Plant J* **33**: 989-1000
- Saiki R, Scharf S, Faloona F, Mullis K, Horn G, Erlich H** (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354
- Salvà I, Jamet E** (2001) Expression of the tobacco *Ext1.4* extensin gene upon mechanical constraints and localization of regulatory regions. *Plant Biol* **3**: 1-10
- Sambrook J, Russel D** (2001) *Molecular Cloning: A Laboratory Manual*, Vol 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)
- Samuels AL, Giddings TH, Staehelin LA** (1995) Cytokinesis in tobacco BY-2 and root tip cells: a new model of cell plate formation in higher plants. *J Cell Biol* **130**: 1345-1357
- Scheller HV, Jensen JK, Sorensen SO, Harholt J, Geshi N** (2007) Biosynthesis of pectin. *Physiologia Plantarum* **129**: 283-295
- Schols HA, Bakx EJ, Schipper D, Voragen AGJ** (1995) A xylogalacturonan subunit present in the modified hairy regions of apple pectin. *Carbohydr Res* **279**: 265-279

- Schroder R, Atkinson RG** (2006) Kiwifruit cell walls: Towards an understanding of softening? *New Zealand J Forestry Sci* **36**: 112-129
- Schultz CJ, Hauser K, Lind JL, Atkinson AH, Pu ZY, Anderson MA, Clarke AE** (1997) Molecular characterisation of a cDNA sequence encoding the backbone of a style-specific 120 kDa glycoprotein which has features of both extensins and arabinogalactan proteins. *Plant Mol Biol* **35**: 833-845
- Schultz CJ, Johnson KL, Currie G, Bacic A** (2000) The classical arabinogalactan protein gene family of arabidopsis. *Plant Cell* **12**: 1751-1767
- Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YG, Bacic A** (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol* **129**: 1448-1463
- Schultz CJ, Ferguson KL, Lahnstein J, Bacic A** (2004) Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. Endoplasmic reticulum and glycosylphosphatidylinositol-anchor signal cleavage sites and hydroxylation of proline. *J Biol Chem* **279**: 45503-45511
- Seifert GJ, Barber C, Wells B, Roberts K** (2004) Growth regulators and the control of nucleotide sugar flux. *Plant Cell* **16**: 723-730
- Seifert G, Roberts K** (2007) The biology of arabinogalactan proteins. *Annu Rev Plant Biol* **58**: 137-161
- Seki M, Carninci P, Nishiyama Y, Hayashizaki Y, Shinozaki K** (1998) High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. *Plant J* **15**: 707-720
- Serpe MD, Nothnagel EA** (1994) Effects of Yariv phenylglycosides on rosa cell suspensions, evidence for the involvement of arabinogalactan-proteins in cell proliferation. *Planta* **193**: 542-550
- Sheng JS, Dovidio R, Mehdy MC** (1991) Negative and positive regulation of a novel proline rich protein messenger-RNA by fungal elicitor and wounding. *Plant J* **1**: 345-354
- Shiu SH, Bleecker AB** (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci USA* **98**: 10763-10768
- Showalter AM** (1993) Structure and function of plant cell wall proteins. *Plant Cell* **5**: 9-23
- Showalter AM** (2001) Arabinogalactan-proteins: structure, expression and function. *Cell Mol Life Sci* **58**: 1399-1417
- Showalter A, Keppler B, Lichtenberg J, Gu D, Welch L** (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol* **153**: 485-513
- Shpak E, Leykam J, Kieliszewski M** (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-*O*-glycosylation codes. *Proc Natl Acad Sci USA* **96**: 14736-

14741

- Shpak E, Barbar E, Leykam JF, Kieliszewski MJ** (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *J Biol Chem* **276**: 11272-11278
- Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N** (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**: D161-D166
- Silverstein KAT, Moskal WA, Jr., Wu HC, Underwood BA, Graham MA, Town CD, VandenBosch KA** (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J* **51**: 262-280
- Sjöström E, Westermark U** (1999) Chemical composition of wood and pulps: Basic constituents and their distribution. *In* E Sjöström, R Alén, eds, *Analytical Methods in Wood Chemistry, Pulping, and Papermaking*. Springer Series in Wood Science. Springer-Verlag, Heidelberg, pp 1-19
- Skinner DJ, Gasser CS** (2009) Expression-based discovery of candidate ovule development regulators through transcriptional profiling of ovule mutants. *BMC Plant Biol* **9**: 29
- Smalle J, Haegman M, Kurepa J, Van Montagu M, Van Der Straeten D** (1997) Ethylene can stimulate Arabidopsis hypocotyl elongation in the light. *Proc Natl Acad Sci USA* **94**: 2756-2761
- Smallwood M, Yates EA, Willats WGT, Martin H, Knox JP** (1996) Immunochemical comparison of membrane-associated and secreted arabinogalactan-proteins in rice and carrot. *Planta* **198**: 452-459
- Somerville C** (2006) Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol* **22**: 53-78
- Sommer-Knudsen J, Clarke A, Bacic A** (1996) A galactose-rich, cell-wall glycoprotein from styles of *Nicotiana glauca*. *Plant J* **9**: 71-83
- Sommer-Knudsen J, Bacic A, Clarke A** (1998) Hydroxyproline-rich plant glycoproteins. *Phytochemistry* **47**: 483-497
- Spiro RG** (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R-56R
- Stintzi A, Heitz T, Prasad V, Wiedemannmerdinoglu S, Kauffmann S, Geoffroy P, Legrand M, Fritig B** (1993) Plant pathogenesis-related proteins and their role in defense against pathogens. *Biochimie* **75**: 687-706
- Sun WX, Kieliszewski MJ, Showalter AM** (2004a) Overexpression of tomato LeAGP-1 arabinogalactan-protein promotes lateral branching and hampers reproductive development. *Plant J* **40**: 870-881
- Sun WX, Zhao ZD, Hare MC, Kieliszewski MJ, Showalter AM** (2004b) Tomato LeAGP-1

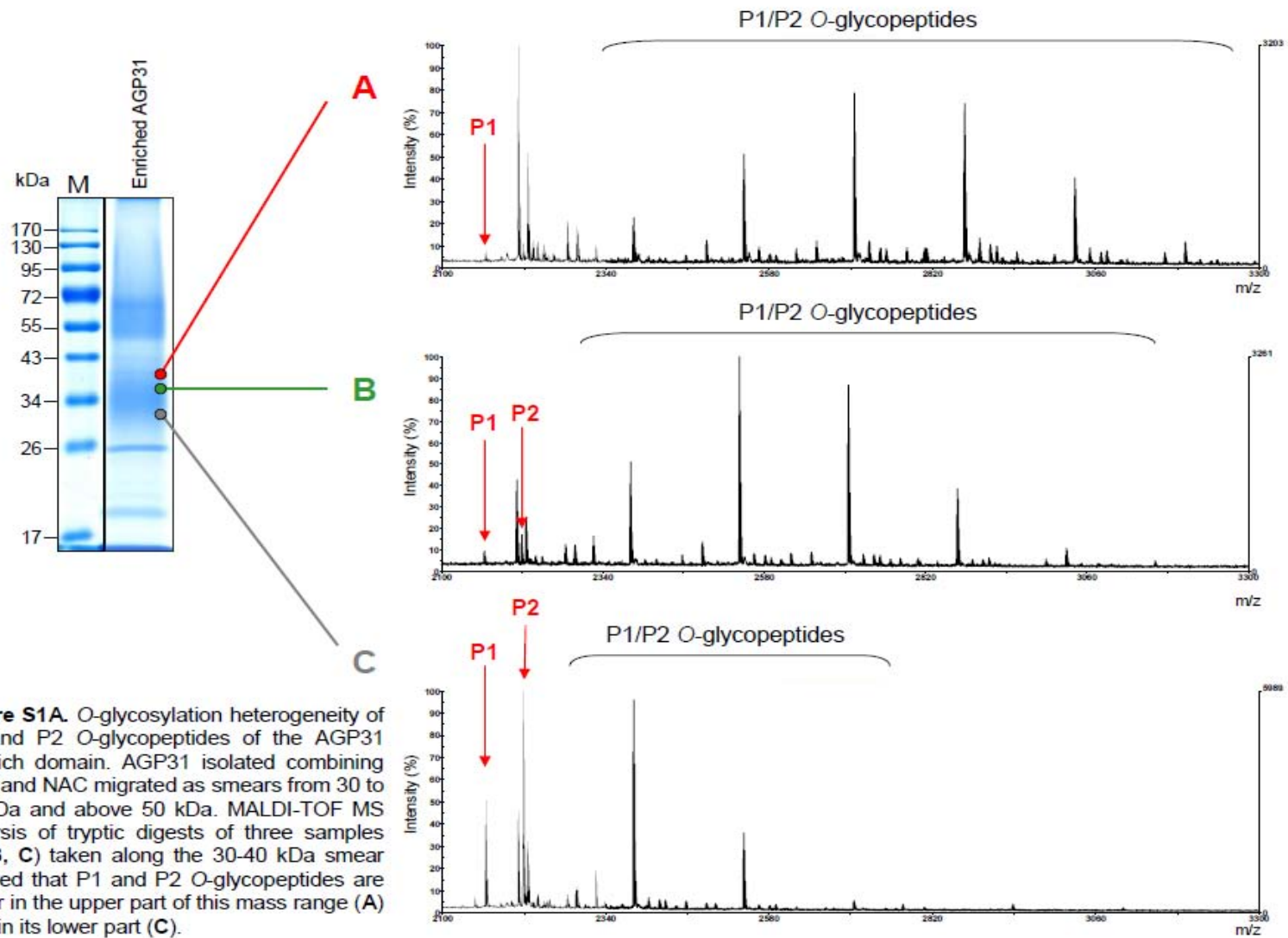
- is a plasma membrane-bound, glycosylphosphatidylinositol-anchored arabinogalactan-protein. *Physiologia Plantarum* **120**: 319-327
- Suzuki Y, Kitagawa M, Knox JP, Yamaguchi I** (2002) A role for arabinogalactan proteins in gibberellin-induced alpha-amylase production in barley aleurone cells. *Plant J* **29**: 733-741
- Takahashi T, Gasch A, Nishizawa N, Chua NH** (1995) The *diminuto* gene of Arabidopsis is involved in regulating cell elongation. *Genes Dev* **9**: 97-107
- Tan L, Qiu F, Lamport D, Kieliszewski M** (2004) Structure of a hydroxyproline (Hyp)-arabinogalactan polysaccharide from repetitive Ala-Hyp expressed in transgenic *Nicotiana tabacum*. *J Biol Chem* **279**: 13156-13165
- Tan L, Varnai P, Lamport D, Yuan C, Xu J, Qiu F, Kieliszewski M** (2010) Plant O-hydroxyproline arabinogalactans are composed of repeating trigalactosyl subunits with short bifurcated side chains. *J Biol Chem* **285**: 24575-24583
- Temporini C, Calleri E, Massolini G, Caccialanza G** (2008) Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins. *Mass Spectrom Rev* **27**: 207-236
- Thiele K, Wanner G, Kindzierski V, Jurgens G, Mayer U, Pahl F, Assaad FF** (2009) The timely deposition of callose is essential for cytokinesis in Arabidopsis. *Plant J* **58**: 13-26
- Thomas JR, Darvill AG, Albersheim P** (1989) Rhamnogalacturonan I, a pectic polysaccharide that is a component of monocot cell walls. *Carbohydr Res* **185**: 279-305
- Thompson JE, Fry SC** (2000) Evidence for covalent linkage between xyloglucan and acidic pectins in suspension-cultured rose cells. *Planta* **211**: 275-286
- Tiainen P, Myllyharju J, Koivunen P** (2005) Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *J Biol Chem* **280**: 1142-1148
- Tryphona T, Liang H-C, Kotake T, Kaneko S, Marsh J, Ichinose H, Lovegrove A, Tsumuraya Y, Shewry P, Stephens E, Dupree P** (2010) Carbohydrate structural analysis of wheat flour arabinogalactan protein. *Carbohydr Res* **345**: 2648-2656
- van der Hoorn R** (2008) Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol* **59**: 191-223
- van Hengel AJ, Roberts K** (2002) Fucosylated arabinogalactan-proteins are required for full root cell elongation in arabidopsis. *Plant J* **32**: 105-113
- van Hengel AJ, Roberts K** (2003) AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination. *Plant J* **36**: 256-270

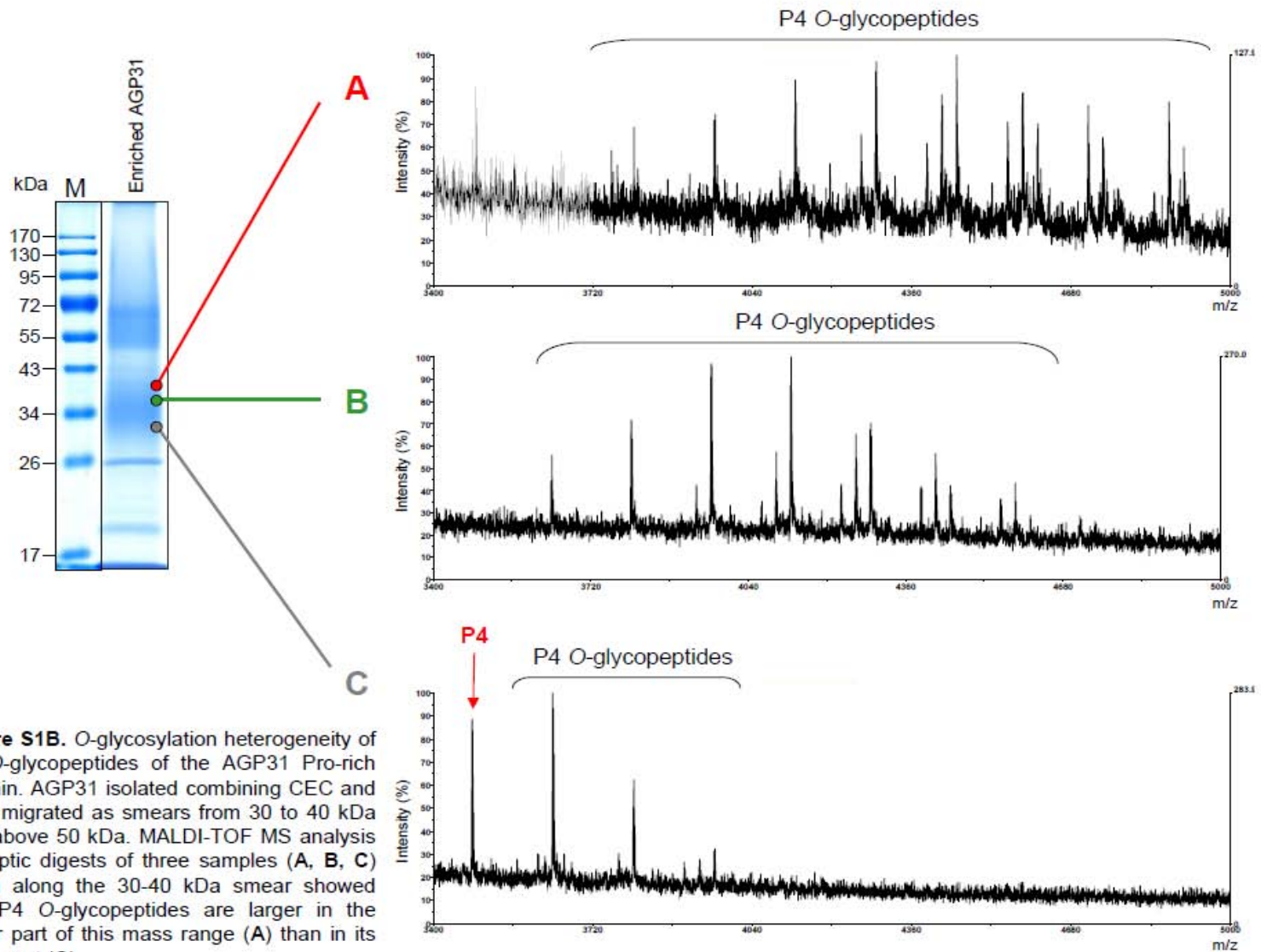
- van Holst GJ, Varner JE** (1984) Reinforced polyproline-II conformation in a hydroxyproline rich cell wall glycoprotein from carrot root. *Plant Physiol* **74**: 247-251
- Vaucheret H, Béclin C, Elmayan T, Feuerbach F, Godon C, Morel J, Mourrain P, Palauqui J, Vernhettes S** (1998) Transgene-induced gene silencing in plants. *Plant J* **16**: 651-659
- Vignols F, Jose-Estanyol M, Caparros-Ruiz D, Rigau J, Puigdomenech P** (1999) Involvement of a maize proline-rich protein in secondary cell wall formation as deduced from its specific mRNA localization. *Plant Mol Biol* **39**: 945-952
- Vincken JP, York WS, Beldman G, Voragen AGJ** (1997) Two general branching patterns of xyloglucan, XXXG and XXGG. *Plant Physiol* **114**: 9-13
- Vincken JP, Schols HA, Oomen RJ, McCann MC, Ulvskov P, Voragen AGJ, Visser RGF** (2003) If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. *Plant Physiol* **132**: 1781-1789
- Voinnet O, Pinto YM, Baulcombe DC** (1999) Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proc Natl Acad Sci USA* **96**: 14147-14152
- Voinnet O, Rivas S, Mestre P, Baulcombe D** (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**: 949-956
- Wei N, Kwok SF, Vonarnim AG, Lee A, McNellis TW, Piekos B, Deng XW** (1994) Arabidopsis *COP8*, *COP10* and *COP11* genes are involved in repression of photomorphogenic development in darkness. *Plant Cell* **6**: 629-643
- Wigoda N, Ben-Nissan G, Granot D, Schwartz A, Weiss D** (2006) The gibberellin-induced, cysteine-rich protein GIP2 from *Petunia hybrida* exhibits in planta antioxidant activity. *Plant J* **48**: 796-805
- Willats WG, Knox JP** (1996) A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of β -glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. *Plant J* **9**: 919-925
- Willats WG, McCartney L, Mackie W, Knox JP** (2001) Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol* **47**: 9-27
- Willats WGT, Knox P, Mikkelsen JD** (2006) Pectin: new insights into an old polymer are starting to gel. *Trends Food Sci Technol* **17**: 97-104
- Wilson LG, Fry JC** (1986) Extensin, a major cell wall glycoprotein. *Plant Cell Environ* **9**: 239-260
- Wisniewska E, Majewska-Sawka A** (2007) Arabinogalactan-proteins stimulate the organogenesis of guard cell protoplasts-derived callus in sugar beet. *Plant Cell Rep* **26**: 1457-1467

- Wojtaszek P** (2000) Genes and plant cell walls: a difficult relationship. *Biol Rev* **75**: 437-475
- Wright PE, Dyson HJ** (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**: 321-331
- Wu H, de Graaf B, Mariani C, Cheung AY** (2001) Hydroxyproline-rich glycoproteins in plant reproductive tissues: structure, functions and regulation. *Cell Mol Life Sci* **58**: 1418-1429
- Wyatt RE, Nagao RT, Key JL** (1992) Patterns of soybean proline-rich protein gene expression. *Plant Cell* **4**: 99-110
- Yang J, Zhang Y, Liang Y, Showalter AM** (2011) Expression analyses of AtAGP17 and AtAGP19, two lysine-rich arabinogalactan proteins, in Arabidopsis. *Plant Biol* **13**: 431-438
- Yariv J, Rapport M, Graf L** (1962) The interaction of glycosides and saccharides with antibody to the corresponding phenylazo glycosides. *Biochem J* **85**: 383-388
- Yariv J, Lis H, Katchalski E** (1967) Precipitation of arabic acid and some seed polysaccharides by glycosylphenylazo dyes. *Biochim J* **105**: 1C-2C
- Yates EA, Valdor JF, Haslam SM, Morris HR, Dell A, Mackie W, Knox JP** (1996) Characterization of carbohydrate structural features recognized by anti-arabinogalactan-protein monoclonal antibodies. *Glycobiology* **6**: 131-139
- Yeats TH, Rose J** (2007) The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). *Protein Sci* **17**: 191-198
- York WS, Darvill AG, Albersheim P** (1984) Inhibition of 2,4-dichlorophenoxyacetic acid-stimulated elongation of pea stem segments by a xyloglucan oligosaccharide. *Plant Physiol* **75**: 295-297
- York WS, Darvill AG, McNeil M, Stevenson TT, Albersheim P** (1986) Isolation and characterization of plant cell walls and cell wall components. *Methods Enzymol* **118**: 3-40
- Zandleven J, Sorensen SO, Harholt J, Beldman G, Schols HA, Scheller HV, Voragen AJ** (2007) Xylogalacturonan exists in cell walls from various tissues of *Arabidopsis thaliana*. *Phytochemistry* **68**: 1219-1226
- Zhang Y, Giboulot A, Zivy M, Valot B, Jamet E, Albenne C** (2011a) Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* **72**: 1109-1123
- Zhang YZ, Yang J, Showalter AM** (2011b) AtAGP18 is localized at the plasma membrane and functions in plant growth and development. *Planta* **233**: 675-683
- Zykwinska A, Rondeau-Mouro C, Garnier C, Thibault J-F, Ralet M-C** (2006) Alkaline extractability of pectic arabinan and galactan and their mobility in sugar beet and potato cell walls. *Carbohydr Polymers* **65**: 510-520

ANNEXES

Supplemental Figure S1: Evidence for *O*-glycosylation heterogeneity of the AGP31 Pro-rich domain.





Supplemental Figure S2: Hyp location within the Pro-rich domain of AGP31 using MALDI-TOF/TOF MS/MS.

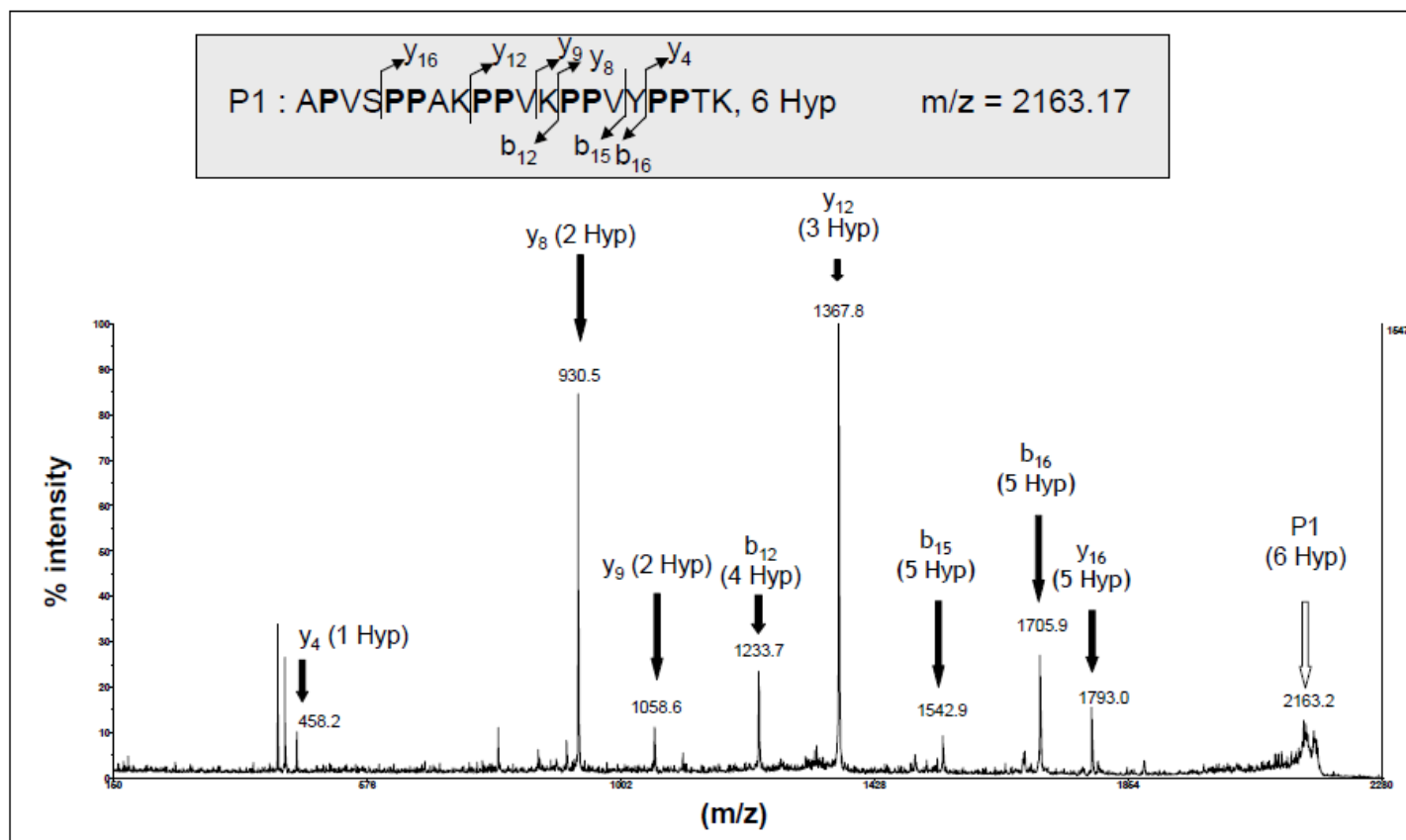


Figure S2A. MALDI-TOF/TOF MS/MS spectrum of the P1 peptide (m/z 2163.17, white arrow) of the AGP31 Pro-rich domain.

Fragmentation was carried out on the deglycosylated AGP31 P1 tryptic peptide obtained following gel excision of an HF sample (see Fig. 3B). Peaks corresponding to b and y fragment ions are labeled with black arrows and the number of Hyp residues from the fragment masses is indicated. The fragmentation pattern is reported onto the sequence in which Pro residues possibly hydroxylated are in bold (in frame). The number of Hyp residues predicted within each fragment allowed partial Hyp location. Non-assigned peaks mostly corresponded to internal fragments.

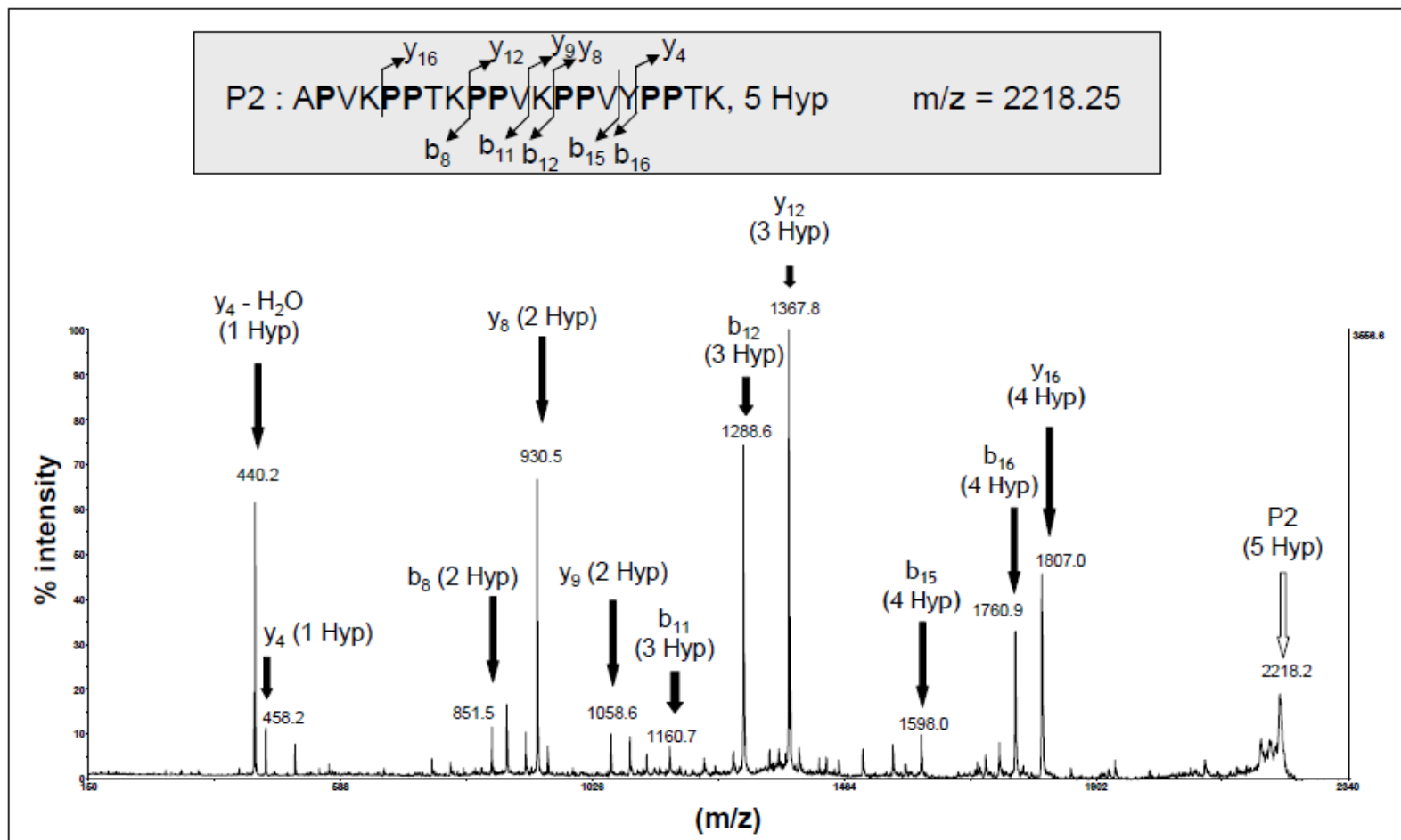


Figure S2B. MALDI-TOF/TOF MS/MS spectrum of the P2 peptide (m/z 2218.25, white arrow) of the AGP31 Pro-rich domain.

Fragmentation was carried out on the deglycosylated AGP31 P2 tryptic peptide obtained following gel excision of an HF sample (see Fig. 3B). Peaks corresponding to b and y fragment ions are labeled with black arrows and the number of Hyp predicted from the fragment masses is indicated. The fragmentation pattern is reported onto the sequence in which Pro residues possibly hydroxylated are in bold (in frame). The number of Hyp residues predicted within each fragment allowed partial Hyp location. Non-assigned peaks mostly corresponded to internal fragments.

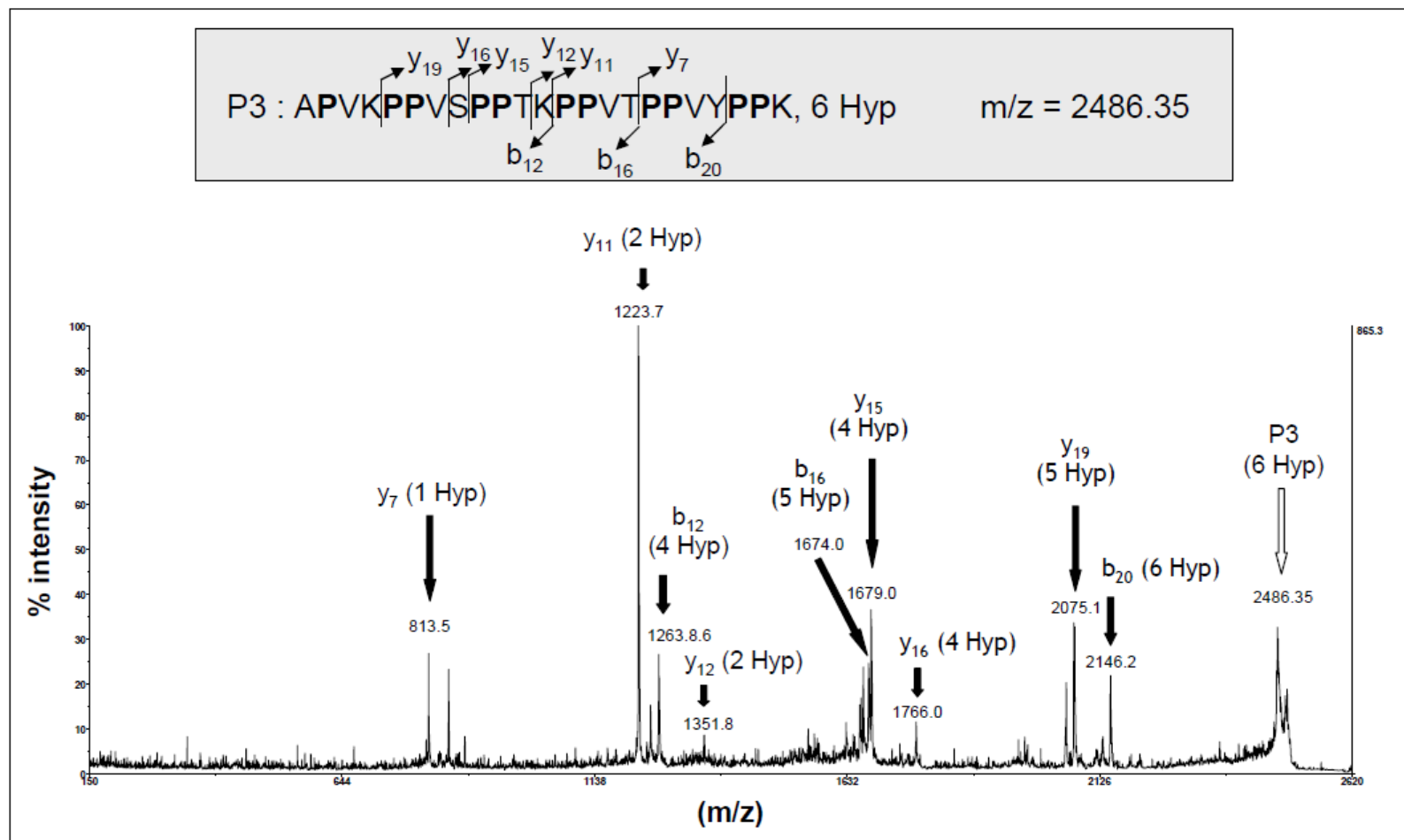


Figure S2C. MALDI-TOF/TOF MS/MS spectrum of the P3 peptide (m/z 2486.35, white arrow) of the AGP31 Pro-rich domain.

Fragmentation was carried out on the deglycosylated AGP31 P3 tryptic peptide obtained following gel excision of an HF sample (see Fig. 3B). Peaks corresponding to b and y fragment ions are labeled with black arrows and the number of Hyp predicted from the fragment masses is indicated. The fragmentation pattern is reported onto the sequence in which Pro residues possibly hydroxylated are in bold (in frame). The number of Hyp residues predicted within each fragment allowed partial Hyp location. Non-assigned peaks mostly corresponded to internal fragments.

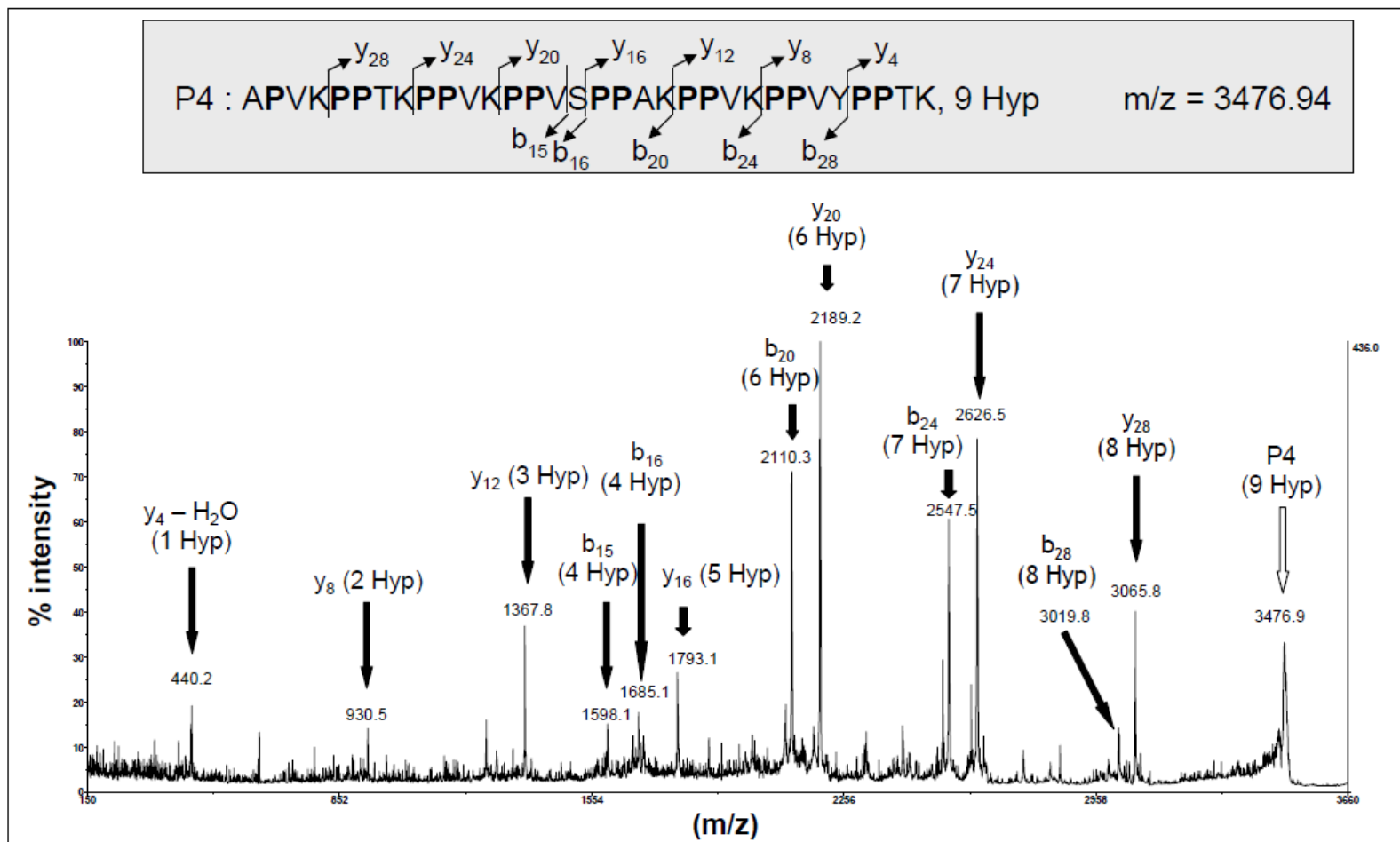


Figure S2D. MALDI-TOF/TOF MS/MS spectrum of the P4 peptide (m/z 3476.94, white arrow) of the AGP31 Pro-rich domain.

Fragmentation was carried out on the deglycosylated AGP31 P4 tryptic peptide obtained following gel excision of an HF sample (see Fig. 3B). Peaks corresponding to b and y fragment ions are labeled with black arrows and the number of Hyp predicted from the fragment masses is indicated. The fragmentation pattern is reported onto the sequence in which Pro residues possibly hydroxylated are in bold (in frame). The number of Hyp residues predicted within each fragment allowed partial Hyp location. Non-assigned peaks mostly corresponded to internal fragments.

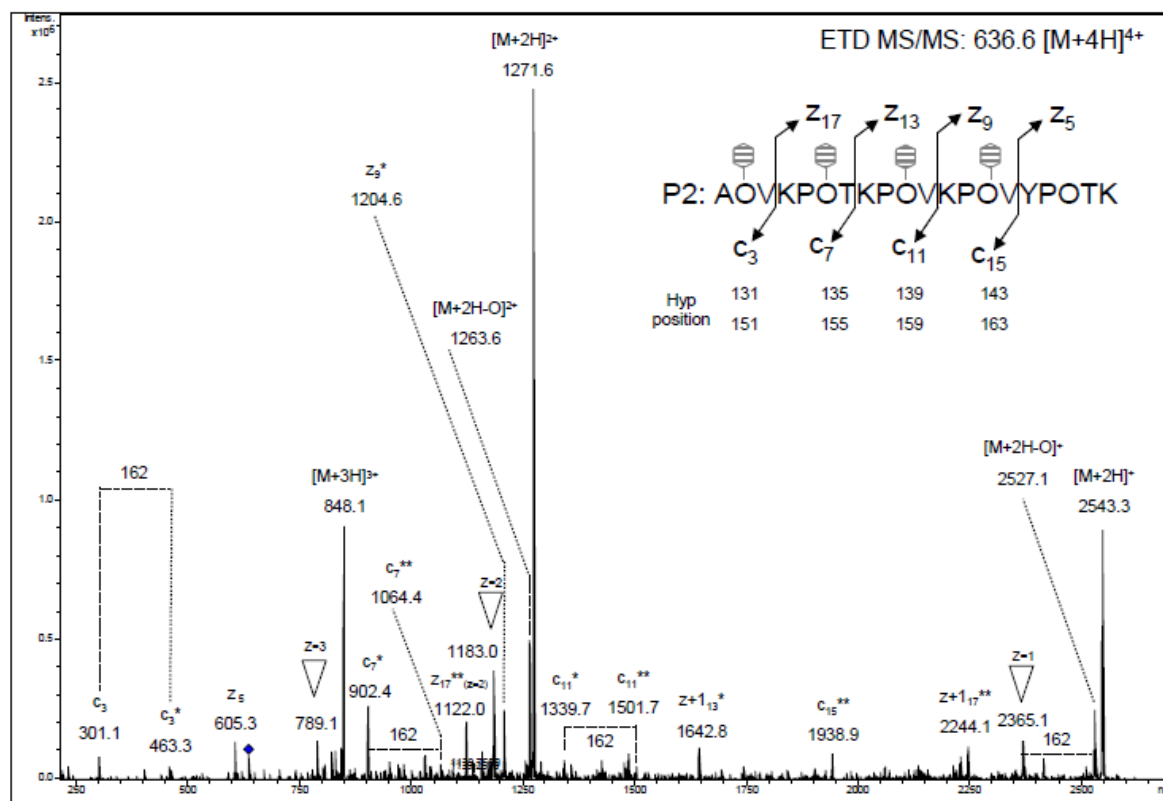
Supplemental Figure S3: Distribution of hexoses on AGP31 P2 O-glycopeptides using ETD MS/MS


Figure S3A. ETD MS/MS analysis of the P2 glycopeptide containing 2 hexoses (636.6 [M+4H]⁴⁺).

For simplification, Hyp residues (O) of the P2 peptide were located according to our experimental results and literature data. Fragmentation was carried out on the P2 O-glycopeptide containing 2 hexoses obtained following gel excision of an AGP31 enriched fraction in the 50-60 kDa mass range (see Figs. 2, 3A and 4). Sites of O-glycosylation were identified from c and z series of differentially glycosylated fragment ions. The number of hexose residues (⊖) found on each fragment is indicated by stars (*). Fragmentation pattern showed that different glycoforms exist with possible sites of O-glycosylation on Hyp131/151, Hyp135/155, Hyp139/159 and Hyp143/163 of the AGP31 Pro-rich domain. Peaks corresponding to the loss of one hexose and one hydroxyl moiety were found (▽). Parent ion fragmented is labeled with ♦.

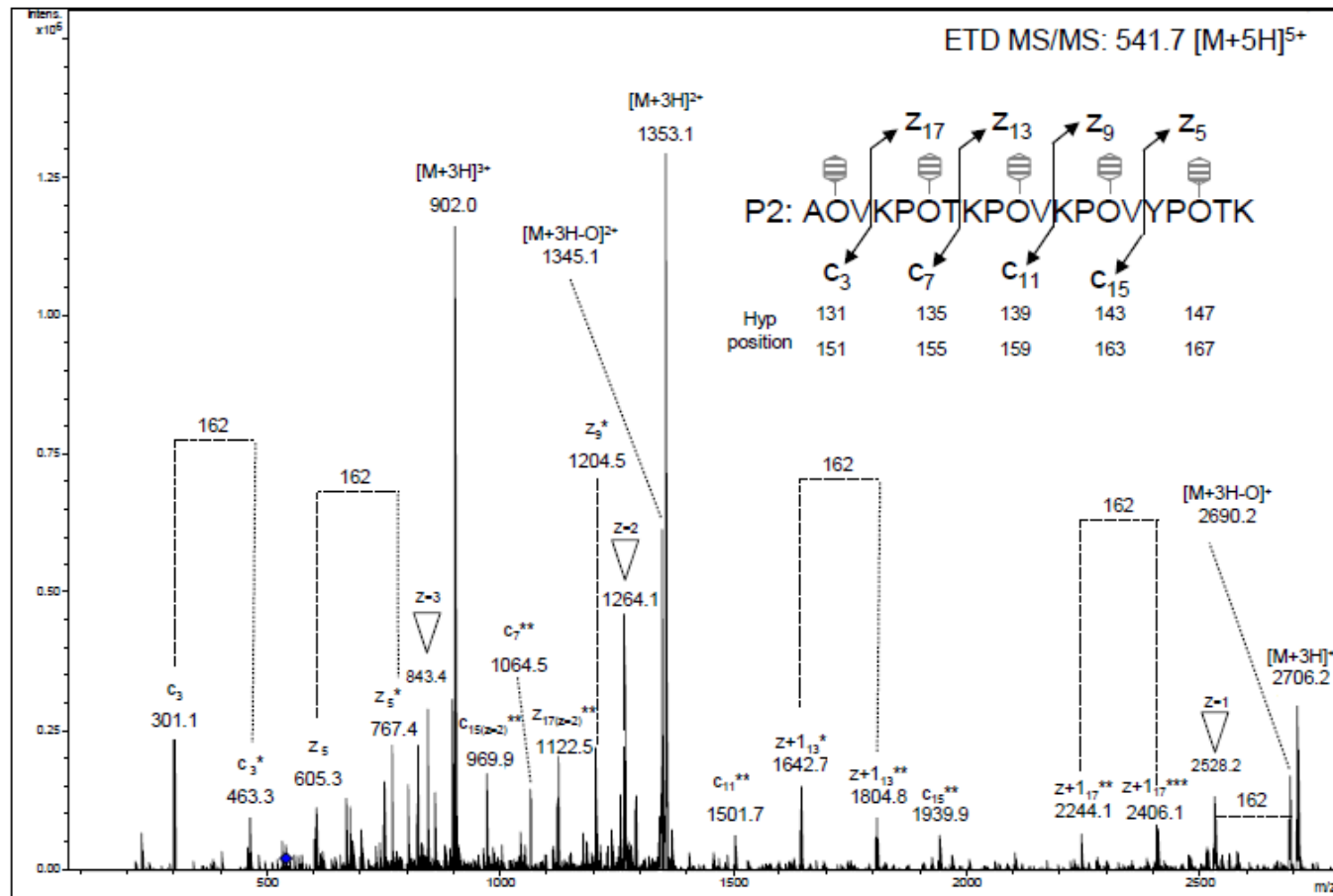


Figure S3B. ETD MS/MS analysis of the P2 glycopeptide containing 3 hexoses (541.7 [M+5H]⁵⁺).

For simplification, Hyp residues (O) of the P2 peptide were located according to our experimental results and literature data. Fragmentation was carried out on the P2 O-glycopeptide containing 3 hexoses obtained following gel excision of an AGP31 enriched fraction in the 50-60 kDa mass range (see Figs. 2, 3A and 4). Sites of O-glycosylation were identified from c and z series of differentially glycosylated fragment ions. The number of hexose residues (⊖) found on each fragment is indicated by stars (*). Fragmentation pattern showed that different glycoforms exist with possible sites of O-glycosylation on Hyp131/151, Hyp135/155, Hyp139/159, Hyp143/163 and Hyp147/167 of the AGP31 Pro-rich domain. Peaks corresponding to the loss of one hexose and one hydroxyl moiety were found (▽). Parent ion fragmented is labeled with ♦.

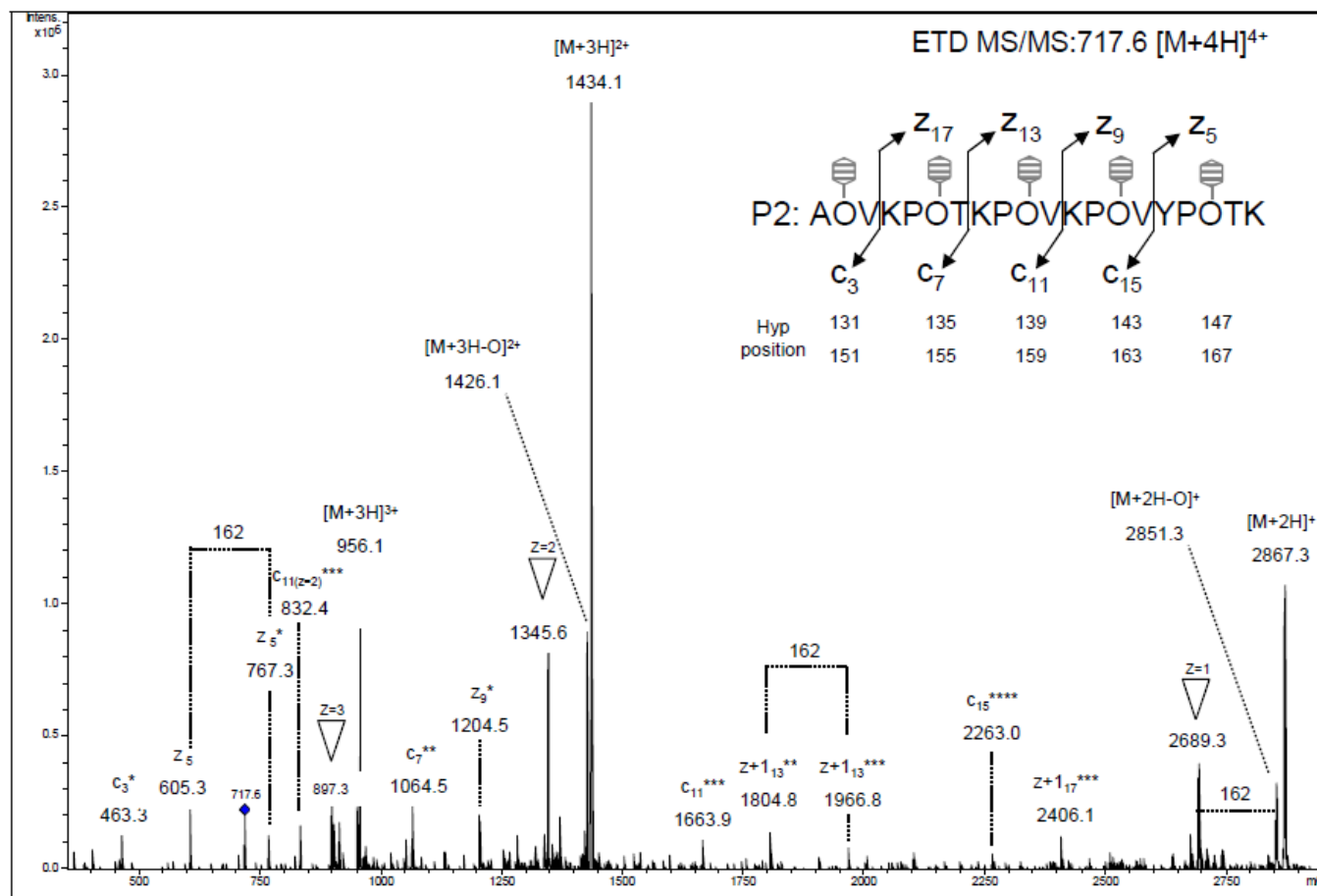


Figure S3C. ETD MS/MS analysis of the P2 glycopeptide containing 4 hexoses (717.6 [M+4H]⁴⁺).

For simplification, Hyp residues (O) of the P2 peptide were located according to our experimental results and literature data. Fragmentation was carried out on the P2 O-glycopeptide containing 4 hexoses obtained following gel excision of an AGP31 enriched fraction in the 50-60 kDa mass range (see Figs. 2, 3A and 4). Sites of O-glycosylation were identified from c and z series of differentially glycosylated fragment ions. The number of hexose residues (⊖) found on each fragment is indicated by stars (*). Fragmentation pattern showed that different glycoforms exist with possible sites of O-glycosylation on Hyp131/151, Hyp135/155, Hyp139/159, Hyp143/163 and Hyp147/167 of the AGP31 Pro-rich domain. Peaks corresponding to the loss of one hexose and one hydroxyl moiety were found (▽). Parent ion fragmented is labeled with ♦.

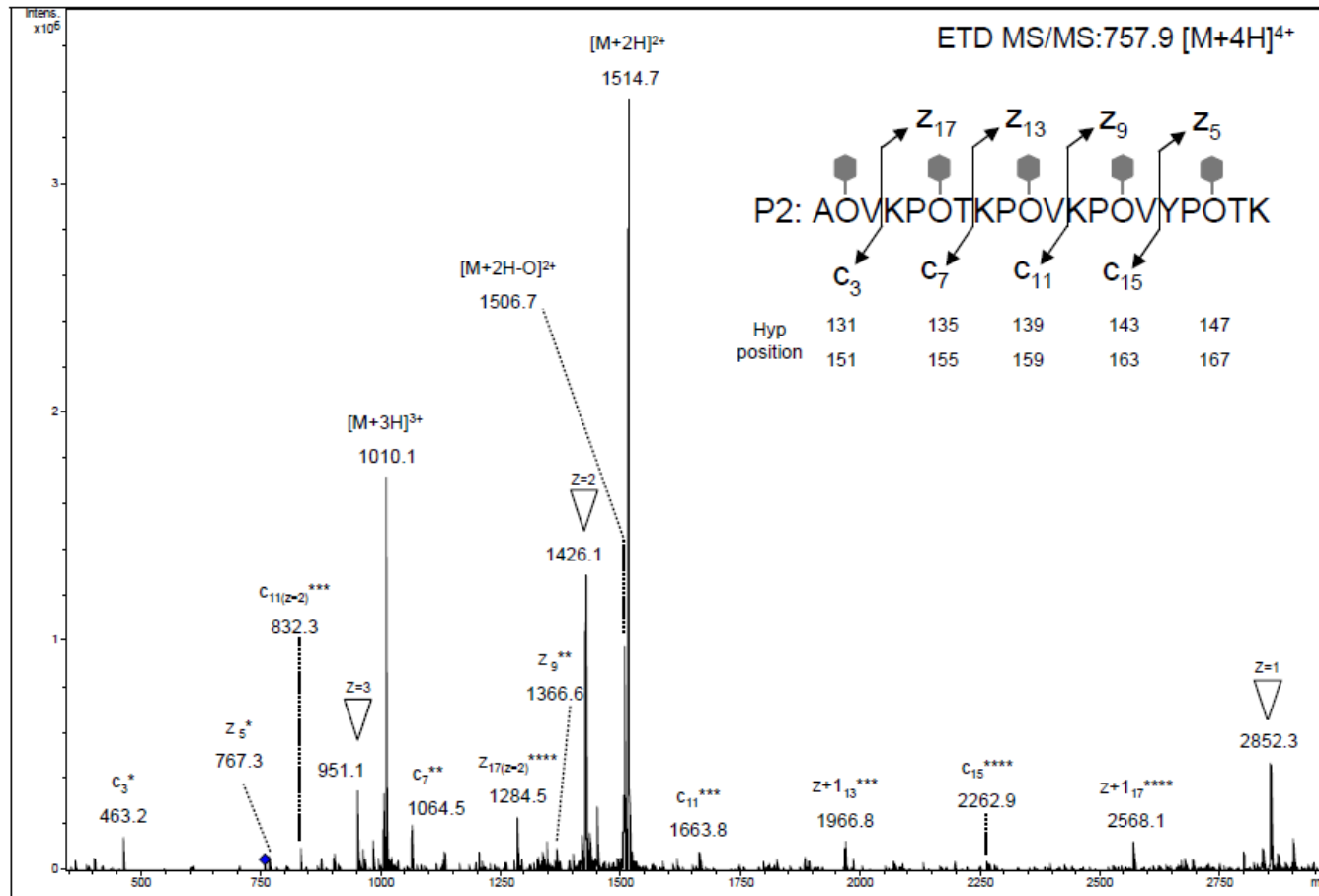
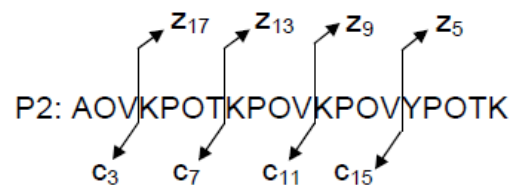


Figure S3D. ETD MS/MS analysis of the P2 glycopeptide containing 5 hexoses (757.9 [M+4H]⁴⁺).

For simplification, Hyp residues (O) of the P2 peptide were located according to our experimental results and literature data. Fragmentation was carried out on the P2 O-glycopeptide containing 5 hexoses obtained following gel excision of an AGP31 enriched fraction in the 50-60 kDa mass range (see Figs. 2, 3A and 4). Sites of O-glycosylation were identified from c and z series of glycosylated fragment ions. The number of hexose residues (●) found on each fragment is indicated by stars (*). Fragmentation pattern showed that each Hyp residue of the P2 peptide of the AGP31 Pro-rich domain carries only one hexose. Peaks corresponding to the loss of one hexose and one hydroxyl moiety were found (▽). Parent ion fragmented is labeled with ◆.

Figure S3E. Number of hexose residues found on c and z fragment ions obtained by ETD MS/MS of the P2 O-glycopeptide containing 2, 3, 4 or 5 hexoses.

Fragmentations were carried out on P2 O-glycopeptides obtained following gel excision of an AGP31 enriched fraction in the 50-60 kDa mass range (see Figs. 2, 3A and 4). For simplification, Hyp residues (O) of the P2 peptide were located according to our experimental results and literature data.



	c series				z / z+1 series			
	3	7	11	15	5	9	13	17
P2 + 2 hexoses	0	-	-	-	0	-	-	-
	1	1	1	-	-	1	1	-
	-	2	2	2	-	-	-	2
P2 + 3 hexoses	0	-	-	-	0	-	-	-
	1	-	-	-	1	1	1	-
	-	2	2	2	-	-	2	2
	-	-	-	-	-	-	-	3
P2 + 4 hexoses	-	-	-	-	0	-	-	-
	1	-	-	-	1	1	-	-
	-	2	-	-	-	-	2	-
	-	-	3	-	-	-	3	3
	-	-	-	4	-	-	-	-
P2 + 5 hexoses	-	-	-	-	-	-	-	-
	1	-	-	-	1	-	-	-
	-	2	-	-	-	2	-	-
	-	-	3	-	-	-	3	-
	-	-	-	4	-	-	-	4
	-	-	-	-	-	-	-	

AUTHOR : May HIJAZI

TITLE : STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF AGP31, AN ATYPICAL CELL WALL GLYCOPROTEIN IN *ARABIDOPSIS THALIANA*

PhD SUPERVISORS : Dr. Elisabeth JAMET and Dr. Cécile ALBENNE

ABSTRACT

Plant cell wall is a dynamic structure consisting of polysaccharide and protein networks with a complex supramolecular organization. AGP31, encoded by *At1g28290*, was identified as a abundant multi-domain protein in *Arabidopsis thaliana* etiolated hypocotyls cell walls. My thesis project consisted in elucidating the structure and the function of AGP31. Structural characterization of the Pro rich domain was performed: Hyp residues and *O*-galactans were located using mass spectrometry. AGP31 was found as several glycoforms differing by the type and the size of *O*-glycans as well as by the polypeptidic length. *In vitro* assays showed interactions between AGP31 domains and cell wall polysaccharides (methylsterified HG, galactans from RGI). A remarkable affinity between AGP31 *O*-galactans and the PNA lectin was shown, indicating possible interactions with cell wall lectins. The expression pattern of *At1g28290* suggests a role in cell wall strengthening during radicle emergence and fast elongation of etiolated hypocotyls. However, study of RNAi underexpressors and overexpressors of *At1g28290* did not permit to find a phenotype during development, probably because of functional redundancy with *At1g28290* homologues. This work constitutes the first structural characterization of a Pro rich domain of a cell wall protein and permits to propose that AGP31, *via* its different domains, could make scaffolds in the cell wall through interactions with itself, polysaccharides and/or lectins.

KEY WORDS : AGP31, *Arabidopsis thaliana*, cell wall, glycoprotein, hydroxyproline, *O*-glycosylation, protein/polysaccharide interactions.

AUTEUR : May HIJAZI

TITRE : CARACTERISATION STRUCTURALE ET FONCTIONNELLE D'AGP31, UNE GLYCOPROTEINE ATYPYQUE DE LA PAROI CHEZ *ARABIDOPSIS THALIANA*

DIRECTEURS DE THESE : Dr. Elisabeth JAMET et Dr. Cécile ALBENNE

RESUME

La paroi cellulaire végétale est une structure dynamique constituée de réseaux de polysaccharides et de protéines dont l'organisation supramoléculaire est complexe. AGP31, codée par *At1g28290*, a été identifiée comme une protéine multi-domaines abondante dans la paroi des cellules des hypocotyles étiolés d'*Arabidopsis thaliana*. Mon travail de thèse a consisté à élucider la structure et la fonction d'AGP31. La caractérisation structurale du domaine riche en Pro d'AGP31 a été effectuée : des résidus Hyp et des *O*-galactanes ont été localisés par spectrométrie de masse. AGP31 a été trouvée sous plusieurs glycoformes différant par le type et la taille des *O*-glycanes et/ou la longueur de la chaîne polypeptidique. Des tests *in vitro* ont montré des interactions entre différents domaines d'AGP31 et des polysaccharides pariétaux (HG méthylestérifiés, galactanes du RGI). Une remarquable affinité entre les *O*-galactanes d'AGP31 et la lectine PNA a été montrée, indiquant de possibles interactions avec des lectines pariétales. Le patron d'expression d'*At1g28290* suggère un rôle de renforcement des parois lors de l'émergence de la racicule et de l'élongation rapide des hypocotyles étiolés. Cependant, l'étude de plantes ARNi sous-expresses et sur-expresses d'*Atg28290* n'a pas permis de trouver un phénotype au cours du développement, probablement du fait d'une redondance fonctionnelle avec des gènes proches d'*At1g28290*. Ce travail constitue la première caractérisation structurale d'un domaine riche en Pro d'une protéine pariétale et permet de proposer qu'AGP31, *via* ses différents domaines, interagisse en réseau dans les parois avec elle-même, des polysaccharides ou des lectines.

MOTS CLEFS : AGP31, *Arabidopsis thaliana*, glycoprotéine, paroi, hydroxyproline, interactions protéine/polysaccharides, *O*-glycosylation.

DISCIPLINE : Biosciences végétales

INTITULE ET ADRESSE DU LABORATOIRE : Laboratoire de Recherche en Sciences Végétales (LRSV), UMR 5546 – UPS/CNRS, Chemin de Borde Rouge, BP 42617, 31326 Castanet-Tolosan Cedex, France