

**MERIEM BENCHABANE**

**MODIFICATIONS POST-TRADUCTIONNELLES D'UNE SERPINE  
HUMAINE RECOMBINANTE EXPRIMÉE CHEZ LES PLANTES**

Thèse de doctorat en cotutelle présentée  
à la Faculté des études supérieures de l'Université Laval, Québec  
dans le cadre du programme de doctorat en Biologie végétale  
pour l'obtention du grade de Philosophiae Doctor (Ph. D.)

DÉPARTEMENT DE PHYTOLOGIE  
FACULTÉ DES SCIENCES DE L'AGRICULTURE ET DE L'ALIMENTATION  
UNIVERSITÉ LAVAL  
QUÉBEC

et

U.F.R. DES SCIENCES  
UNIVERSITÉ DE ROUEN  
MONT SAINT AIGNAN, FRANCE  
pour l'obtention du grade de Docteur en Biologie

2007

## REMERCIEMENTS

Je voudrais profiter de ces premières pages pour exprimer ma gratitude envers toutes les personnes, chercheurs, professeurs et amis, qui m'ont aidée et encouragée. Le désir et l'ambition de poursuivre des études supérieures tout comme la motivation lors des difficultés sont bien souvent puisés chez les chercheurs que nous admirons, ceux qui nous inspirent...

Je tiens tout particulièrement à remercier mes directeurs de thèse, Dominique Michaud et Véronique Gomord. Dominique m'a accueillie dans son laboratoire alors que je débutais à peine la Maîtrise, il a su me guider au cours de mes études supérieures. J'ai pu apprécier ses grandes qualités scientifiques. Je me souviendrai de nos innombrables heures passées à commenter des résultats, des articles ou à discuter tout simplement. Enfin, son intérêt et son enthousiasme m'ont souvent soutenue alors que la fatigue et les difficultés s'accumulaient. Je suis touchée de la confiance qu'il m'a témoignée tout au long de ma thèse, tant dans le doctorat que dans la réalisation de projets annexes.

Je ne devais passer à l'université de Rouen que le temps de quelques courts séjours, je m'y suis finalement attardée un peu. La richesse de la France, peut-être, ma retenue, mais surtout c'est au sein du laboratoire de Véronique Gomord que mon projet a pris forme. Je tiens à remercier Véronique d'avoir accepté de travailler sur les serpines et de m'avoir guidée au cours de ces années. Je voudrais la remercier pour la liberté qu'elle m'a accordée mais aussi pour les ré-orientations toujours nécessaires en cours de projet. J'ai beaucoup apprécié sa rigueur scientifique et j'espère avoir acquis un peu de ce raisonnement unique.

J'ai été très sensible à l'intérêt que le Dr Loïc Faye a porté à mon projet et à sa participation aux nombreuses réunions scientifiques qui ont émaillé mon séjour à Rouen. Ses conseils m'ont été précieux, je tiens à l'en remercier. Je lui sais gré de m'avoir accueillie au sein de l'UMR6037, d'avoir facilité ma venue en France et de m'avoir offert des conditions de travail de qualité.

Je voudrais aussi témoigner ma reconnaissance à Claude Saint-Jore Dupas pour les analyses de (et mon initiation au monde de la) microscopie et ainsi que pour son jugement critique des résultats et des articles. J'ai trouvé remarquable la générosité de temps et de connaissance dont elle a fait preuve. À Muriel Bardor, je dois des analyses de spectrométrie de masse qui ont failli ne pas apparaître dans ma thèse, mais qui dans tous les cas sont une part importante de ma formation. Je la remercie pour tout le temps et l'énergie qu'elle a consacrés à m'aider et me conseiller. Je ne peux non plus passer sous silence tous les chercheurs, professionnels et maîtres de conférence de l'UMR6037, de l'INAF et de l'Envirotron, lesquels n'ont jamais été avares de partager leurs connaissances, merci. À Carole et Gaëlle, merci pour le précieux temps que vous passez à assurer le bon roulement du laboratoire. De même à Karine Coenen pour son aide technique et scientifique.

Et puis il y a les coéquipiers, mes amis, Bertrand, Daniel, Marco, Raphaël, Charles, Marie-Claire, Jean-François, Amine, Moustafa, Andreja, François à Québec et Fred, Olivia, David, Christophe, Bruno, Sophie, Pravina, Aurélia, Martial à Rouen. Je leurs suis reconnaissante pour toutes les discussions scientifiques, certes, mais aussi pour les 5 à 7 d'une part et les *pots* de l'autre. D'eux dépendent aussi l'ambiance du labo, les conversations loufoques, les plaisirs et les amitiés inoubliables.

Finalement, à ma famille, mes parents, mes sœurs et Thomas, pour tout le reste et plus encore...

## RÉSUMÉ

Le potentiel des plantes pour la production hétérologue de protéines recombinantes est maintenant bien établi, mais des problèmes de stabilité et de modifications post-traductionnelles inadéquates des protéines persistent et nuisent au développement des plateformes végétales. Bien que les plantes aient la capacité d'effectuer la plupart des modifications post-traductionnelles nécessaires aux propriétés biologiques des protéines produites, il existe des différences subtiles mais significatives entre les modifications post-traductionnelles des protéines chez les mammifères et les plantes. Ces différences peuvent affecter les propriétés biologiques des protéines recombinantes végétales et entraver leur valeur commerciale. Parmi ces modifications, la N-glycosylation et la maturation protéolytique sont celles qui affectent le plus souvent les rendements et la qualité des produits recombinants. Afin d'étudier le potentiel des plantes à produire des glycoprotéines complexes, nous avons exprimé en culture cellulaire de tabac *Nicotiana tabacum* cv BY-2 l'inhibiteur de protéases alpha 1-antichymotrypsine (AACT) humaine, une glycoprotéine d'intérêt clinique. La protéine a été adressée vers le compartiment endomembranaire de sécrétion (RE, vacuole et apoplasme) ou exprimée directement dans le cytosol afin d'établir l'impact de la localisation cellulaire sur sa stabilité et sa glycosylation. Nos études ont montré que, bien qu'exprimé dans tous les compartiments cellulaires, le produit obtenu était plus hétérogène au sein du système endomembranaire, probablement en raison d'une maturation de la N-glycosylation et à une maturation protéolytique induite. Inversement, la protéine recombinante exprimée dans le cytosol, sous forme non glycosylée, montrait une plus grande stabilité. Des études d'immunolocalisation ont par ailleurs démontré que cette forme de l'AACT exprimée dans le cytosol montrait une double localisation nucléaire et cytoplasmique. L'AACT possède un site de liaison à l'ADN et une mutation de ce site a permis ici d'inhiber cette interaction et d'altérer en partie la localisation nucléaire. Ces derniers résultats suggèrent que la liaison à l'ADN est impliquée dans la rétention nucléaire de l'AACT.

**Mots-clés :** Production de protéines recombinantes dans les plantes, moléculture, serpins,  $\alpha$ 1-antichymotrypsine, modifications post-traductionnelles, N-glycosylation, dégradation protéolytique, protéolyse, *Nicotiana tabacum* cv BY-2



## ABSTRACT

Plants represent interesting production platforms for recombinant proteins of therapeutical value. While the proof of concept for protein production in plants is no longer a matter of debate, protein stability and post-translational modifications essential for biological activity pose, on the other hand, serious challenges to the development of plant factories. Plant cells can perform most protein post-translational modifications, but plant and mammalian proteins often present minor but significant differences, with a possible negative impact on their commercial value. To assess the potential of plants for the production of structurally-complex mammalian proteins, we expressed human  $\alpha$ 1-antichymotrypsin (AACT), a glycosylated serine protease inhibitor, in cultured BY-2 tobacco cells. The inhibitor was targeted to different subcellular compartments to assess the impact of cellular final destination on accumulation, stability and glycosylation of the resulting variants. Our results showed that AACT entering the secretory pathway was readily processed to lower molecular weight forms. This post-translational maturation apparently was the result of glycan maturation and proteolytic processing of the polypeptide along the secretory pathway. Intriguingly, cytosolic expression generated more stable proteins, although not glycosylated, that accumulated mostly in the nucleus. We further demonstrated that mutation of AACT DNA binding site partially altered the nucleus distribution thus suggesting a role of DNA binding in recombinant AACT nuclear retention. Taken together, these results illustrate the complex and unpredictable nature of recombinant protein post-translational maturation, and stress the need for additional empirical data on the expression of complex glycoproteins in plant cells.

Keywords: Plant-made recombinant proteins, molecular farming, serpins,  $\alpha$ 1-antichymotrypsin, post-translational maturation, N-glycosylation, protein degradation, proteolysis, *Nicotiana tabacum* cv BY-2

## RÉSUMÉ (version longue)

Au cours des dernières années, la production de protéines thérapeutiques recombinantes dans les plantes est passée, pour plusieurs protéines, du stade expérimental à une production à grande échelle, avec dans certains cas des perspectives de commercialisation tangibles. Les avantages de la production hétérologue de protéines dans les plantes ne sont plus à démontrer. Les plantes permettent de produire des protéines à grande échelle à un coût compétitif, tout en contribuant à minimiser les risques de contamination par des agents pathogènes animaux ou humains. De surcroît, les plantes peuvent produire des protéines complexes subissant lors de leur synthèse des maturations post-traductionnelles élaborées. Pourtant, malgré certains succès et des avancées technologiques évidentes dans le domaine, nombreux sont les exemples de protéines dont la production reste inefficace. Les modifications post-traductionnelles sont souvent en cause, liées plus particulièrement à la glycosylation et aux maturations protéolytiques. Ces modifications, bien que se rapprochant de celles observées chez les mammifères, diffèrent en plusieurs points et peuvent affecter les propriétés biologiques de la protéine produite. En particulier, il est important de mieux cerner le potentiel des plantes pour la production de glycoprotéines complexes et de caractériser ces modifications post-traductionnelles avec plusieurs protéines modèles. Dans ce cadre, l'alpha 1-antichymotrypsine humaine (hAACT) a été choisie comme modèle. L'AACT est une glycoprotéine de la famille des serpins, du groupe fonctionnel des inhibiteurs de protéases de type sérine. *In vivo*, l'AACT est une protéine plasmatique montrant un rôle anti-inflammatoire en inhibant la cathepsine G, la chymotrypsine et la chymase. Elle est impliquée, entre autres, dans plusieurs désordres respiratoires comme les bronchites chroniques et l'emphysème. Bien qu'en pratique, la production hétérologue d'AACT puisse montrer un intérêt médical, son expression n'avait encore jamais été envisagée dans les plantes. Dans cette perspective, l'AACT a été exprimée ici dans des cultures cellulaires de tabac (*Nicotiana tabacum*) et ses étapes de maturation post-traductionnelles ont été analysées.

Afin d'étudier l'impact de la compartimentation cellulaire sur la qualité et le rendement en protéine exogène, l'AACT a été exprimée dans quatre compartiments des cellules de tabac: le cytosol, le réticulum endoplasmique, la vacuole et l'apoplasme. Des analyses de type *pulse-chase* ont permis de mettre en évidence l'étendue des maturations post-traductionnelles de l'AACT dirigée vers le système endomembranaire. D'une part,

l'AACT y était N-glycosylée, avec présence de glycanes complexes typiques des végétaux. D'autre part, l'AACT y subissait une maturation protéolytique et s'est avérée particulièrement sensible à des protéases de type sérine présentes dans le milieu extracellulaire. En pratique, ces modifications pourraient ne pas convenir à une application thérapeutique de la protéine, notamment parce que les glycanes complexes végétaux ont été identifiés antérieurement comme étant immunogènes chez les mammifères, et parce que les clivages protéolytiques peuvent affecter l'intégrité, voire l'activité de la protéine. En revanche, l'AACT exprimée dans le compartiment cytosolique a montré ici un rendement faible en apparence mais la protéine y a montré une stabilité certaine en comparaison aux autres compartiments cellulaires testés. L'AACT cytosolique n'est pas glycosylée et ne pourra par contre être utile que si les glycanes n'affectent ni son activité ni les autres propriétés pharmacologiques nécessaires à son utilisation thérapeutique. Bien que, selon nos études, l'AACT non glycosylée demeure un inhibiteur de protéases actif, l'absence de glycanes pourrait influencer sa clairance *in vivo*, ainsi que d'autres propriétés pouvant dépendre en partie de la glycosylation.

Comme l'ont démontré des expériences d'immunolocalisation, l'AACT dirigée vers le cytosol a montré une double localisation cellulaire, cytosolique et nucléaire. L'adressage nucléaire était inattendu et difficilement explicable, l'AACT n'ayant pas de signaux de localisation nucléaire (NLS) apparents. Toutefois, cette protéine possède un site de liaison à l'ADN et, selon plusieurs études, les NLS et les sites de liaison aux acides nucléiques se chevauchent souvent. Une mutation de ce site de liaison à l'ADN a permis ici de rediriger partiellement la protéine vers le cytosol sans en affecter l'activité inhibitrice, suggérant qu'il pourrait être impliqué dans la rétention nucléaire. Par ailleurs, la liaison à l'ADN semble contribuer à réduire la sensibilité de la protéine à la dégradation protéolytique *in vitro*. Ainsi, dans le cadre d'une production hétérologue d'AACT, la localisation nucléaire et la liaison subséquente à l'ADN pourraient éventuellement s'avérer un atout pour préserver l'intégrité de la protéine.

En somme, ces résultats témoignent de toute la complexité inhérente à l'expression d'une serpine recombinante comme l'AACT en culture cellulaire de tabac. Une meilleure compréhension de la relation structure-fonction de l'AACT et une caractérisation plus approfondie des modifications post-traductionnelles des formes recombinantes seront indispensables dans l'avenir pour adapter le système de production

hétérologue utilisé à cette protéine et pour contrôler ses modifications post-traductionnelles en fonction des besoins. De nouvelles stratégies doivent maintenant être développées pour assurer la stabilité, la qualité et l'homogénéité de l'AACT recombinante. Dans un cadre plus large, chaque protéine étant unique, il est fort peu probable de trouver des approches moléculaires universelles pour assurer une maturation et une stabilité satisfaisantes de l'ensemble des protéines recombinantes en système végétal. L'accent devra plutôt être mis sur le développement de protocoles d'analyse rigoureux des produits recombinants et des innovations technologiques spécifiques à chaque protéine.

## **AVANT-PROPOS**

Cette thèse est rédigée sous la forme d'articles devant être publiés dans des revues scientifiques. La contribution des co-auteurs pour ces articles est détaillée ci-après :

### **Chapitre I –**

#### **Preventing unintended proteolysis in plant protein biofactories**

REVIEW PAPER for *Plant Biotechnology Journal*

Meriem Benchabane, Charles Goulet, Loïc Faye, Véronique Gomord and Dominique Michaud

Cet article a été rédigé par moi-même, avec la participation de C. Goulet et des Drs. D. Michaud, V. Gomord et L. Faye.

### **Chapitre II –**

#### **Targeted expression and maturation of human $\alpha$ 1-antichymotrypsin in tobacco BY-2 cultured cells**

Meriem Benchabane, Claude Saint-Jore Dupas, Muriel Bardor, Loïc Faye, Dominique Michaud and Véronique Gomord

J'ai effectué l'ensemble des travaux présentés dans cet article, sous la supervision des Drs. V. Gomord, L. Faye et D. Michaud. Les analyses de microscopie confocale à fluorescence ont été réalisées avec la collaboration du Dr. C. Saint-Jore Dupas, chercheuse post-doctorale au sein du laboratoire de V. Gomord. Les analyses de spectrométrie de masse ont été effectuées avec l'aide du Dr. M. Bardor. Le manuscrit présenté ici a été rédigé par moi-même, avec la participation des Drs. V. Gomord, L. Faye et D. Michaud.

### **Chapitre III –**

#### **Nucleocytoplasmic transit of human $\alpha$ 1-antichymotrypsin expressed in tobacco leaf epidermal cells**

Meriem Benchabane, Claude Saint-Jore Dupas, Loïc Faye, Véronique Gomord, Dominique Michaud

Les travaux de microscopie confocale présentés dans cet article ont été réalisés par le Dr. C. Saint-Jore Dupas de l'Université de Rouen. J'ai pour ma part réalisé les constructions génétiques, la production des protéines recombinantes, leur caractérisation biochimique subséquente et la rédaction de l'article scientifique présenté. Les Drs. D. Michaud, V. Gomord et L. Faye ont supervisé l'ensemble des expériences et participé à la rédaction finale de l'article.

## TABLE DES MATIÈRES

Remerciements	II
Résumé	IV
Abstract	V
Résumé (version longue)	VI
Avant-propos	IX
Table des matières	XI
Liste des figures	XV
Liste des tableaux	XVI
Glossaire	XVII
<b>Chapitre I – Introduction, hypothèses et objectifs</b>	<b>1</b>
<b>I.0. Introduction générale : des plantes pour la production de protéines recombinantes</b>	<b>2</b>
<b>I.1. La dégradation protéolytique des protéines recombinantes</b> <b>Preventing unintended proteolysis in plant protein biofactories</b>	<b>6</b>
I.1.1. Résumé/Abstract	8
I.1.2 Introduction	10
I.1.2.1 Proteolytic degradation of recombinant proteins in plant systems	11
I.1.3 Stabilizing recombinant proteins <i>in planta</i>	12
I.1.3.1 Tissue-specific transgene expression	13
I.1.3.2 Organelle-specific protein targeting	14
I.1.3.3 Stabilizing fusion partners	23
I.1.3.4 Companion protease inhibitors	25
I.1.4 Stabilizing recombinant proteins <i>ex planta</i>	25
I.1.4.1 Basic precautions during protein extraction	26
I.1.4.2 Protein-stabilizing protease inhibitors	27
I.1.4.3 Non-destructive secretion and recovery	29
I.1.5 Conclusion	30
<b>I.2. La glycosylation des protéines recombinantes</b>	<b>32</b>
I.2.1 Les voies de biosynthèse des glycanes	32
I.2.1.1 La O-glycosylation des protéines	33
I.2.1.2 La biosynthèse des N-glycanes chez les plantes	35
I.2.2 Espèces différentes, N-glycanes différents	39
I.2.3 Les rôles de la N-glycosylation	41
I.2.3.1 Les rôles intramoléculaires	41
I.2.3.2 Les rôles intermoléculaires	42

1.2.4 Les plantes, des bio-usines à glycoprotéines	43
1.2.4.1 Les glycoprotéines végétales ne sont pas sialylées	44
1.2.4.2 Les glycoprotéines végétales comportent des épitopes immunogènes	45
1.2.4.3 La glycosylation végétale, un avantage ?	48
1.2.4.4 Les glycoprotéines recombinantes, des produits non-homogènes	49
<b>1.3. Les serpins</b>	<b>50</b>
1.3.2 La structure tridimensionnelle des serpins	52
1.3.3 Le mécanisme d'inhibition des serpins	54
1.3.4 La glycosylation des serpins	57
1.3.4 Les serpins chez les plantes	58
1.3.5 La production de serpins recombinantes	59
<b>1.4. Hypothèses et objectifs</b>	<b>61</b>
<b>Chapitre II –</b>	<b>64</b>
<b>Targeted expression and maturation of human <math>\alpha</math>1-antichymotrypsin in tobacco BY-2 cultured cells</b>	
<b>II.1 Mise en contexte</b>	<b>65</b>
<b>II.2 Résumé/Abstract</b>	<b>67</b>
<b>II.3 Introduction</b>	<b>69</b>
<b>II.4 Results</b>	<b>71</b>
II.4.1 Expression and subcellular distribution of human AACT in tobacco cells	73
II.4.2 Secreted AACT is subject to extensive post-translational maturation in the tobacco cell secretory pathway	75
II.4.3 Mature forms of plant recombinant AACT harbour complex N-glycans	78
II.4.4 Recombinant AACT is proteolytically matured	82
<b>II.5 Discussion</b>	<b>84</b>
<b>II.6 Materials and Methods</b>	<b>88</b>
II.6.1 Materials	88
II.6.2 Gene constructs	89
II.6.3 Stable transformation of BY-2 tobacco cells	90



II.6.4 AACT transient expression in <i>Nicotiana tabacum</i>	90
II.6.5 Expression and purification of recombinant AACT from <i>E. coli</i>	91
II.6.6 Metabolic pulse-chase labelling and immunoprecipitation	91
II.6.7 Enzymatic deglycosylation of plant recombinant AACT	92
II.6.8 Enzymatic deglycosylation of human AACT	93
II.6.9 Affino- and immunodetection	93
II.6.10 Immunocytolocalisation for confocal microscopy	94
II.6.11 Spiking experiments	94
II.6.12 MALDI-TOF/MS analyses	95
<b>Chapitre III –</b>	<b>96</b>
<b>Nucleocytoplasmic transit of human <math>\alpha</math>1-antichymotrypsin expressed in tobacco leaf epidermal cells</b>	
<b>III.1 Mise en contexte</b>	<b>97</b>
<b>III.2 Résumé / Summary</b>	<b>100</b>
<b>III.3 Introduction</b>	<b>102</b>
<b>III.4 Results and discussion</b>	<b>103</b>
III.4.1 AACT and AACT $\Delta$ K exhibit comparable inhibitory activities against chymotrypsin	105
III.4.2 C-terminal DNA-binding site is involved in the nuclear retention of AACT	107
III.4.3 DNA protects recombinant AACT <i>in vitro</i>	110
<b>III.5 Conclusion</b>	<b>112</b>
<b>III.6 Materials and methods</b>	<b>113</b>
III.6.1 Gene constructs	113
III.6.2 Heterologous expression in <i>E. coli</i>	114
III.6.3 Transient expression in tobacco leaves	115
III.6.4 Confocal fluorescence microscopy	115
III.6.5 DNA gel shift assay	115
III.6.6 Chymotrypsin inhibitory activity	116
III.6.7 Trypsin stability assay	117

<b>Chapitre IV – Discussion et perspectives</b>	<b>118</b>
<b>IV.1 L'AACT peut être exprimée dans un système d'expression hétérologue végétal</b>	<b>119</b>
IV.1.1 La compartimentation cellulaire influence la N-glycosylation de l'AACT humaine	119
IV.1.2 La maturation protéolytique de l'AACT varie en fonction de sa destination	121
<b>IV.2 L'AACT exprimée dans le cytosol a une double localisation, nucléaire et cytoplasmique</b>	<b>123</b>
<b>IV.3 Les perspectives</b>	<b>125</b>
<b>Chapitre V – Références bibliographiques</b>	<b>128</b>

## LISTE DES FIGURES

### Chapitre I - Introduction, hypothèse et objectifs

<b>Figure I.1 :</b>	Subcellular targeting of recombinant proteins in plant cells	15
<b>Figure I.2 :</b>	Stabilization of exogenous proteins in crude plant extracts by phenylmethylsulfonyl fluoride (PMSF) and chymostatin	29
<b>Figure I.3 :</b>	Biosynthèse de glycoprotéines O-glycosylées dans l'appareil de Golgi des mammifères	34
<b>Figure I.4 :</b>	Une IgA est O-glycosylée sur les mêmes séquences lorsqu'exprimé chez l'humain ou chez le maïs	35
<b>Figure I.5 :</b>	Synthèse du N-glycane précurseur sur le transporteur lipidique et son transfert sur une chaîne polypeptidique	36
<b>Figure I.6 :</b>	Voie de maturation des N-glycannes dans la cellule végétale	38
<b>Figure I.7 :</b>	Structure des N-glycannes typiques identifiés dans différentes branches du vivant	40
<b>Figure I.8 :</b>	Structure de l'« <i><math>\alpha</math>1-proteinase inhibitor</i> » ( $\alpha$ 1-PI) en conformation native	53
<b>Figure I.9 :</b>	Les serpines peuvent adopter une variété des conformations spontanément ou suite à des changements du milieu ou à des mutations	54
<b>Figure I.10 :</b>	Mécanisme standard d'inhibition d'une protéase de type sérine par un inhibiteur canonique de même famille	55
<b>Figure I.11 :</b>	Structure et mécanisme d'inhibition des serpines	56

### Chapitre II –

#### Targeted expression and maturation of human $\alpha$ 1-antichymotrypsin in tobacco BY-2 cultured cells

<b>Figure II.1:</b>	Gene constructs for targeted expression of human antichymotrypsin (AACT) in tobacco cells	72
<b>Figure II.2:</b>	Transient expression of AACT in tobacco leaves	72
<b>Figure II.3:</b>	Immunolocalization of recombinant AACT expressed in different subcellular compartments of tobacco BY-2 suspension cultured cells	74
<b>Figure II.4:</b>	Post-translational maturation of recombinant AACT in BY-2 cultured cells	77
<b>Figure II.5:</b>	Human AACT is highly glycosylated, with complex N-glycans containing sialic acids	79
<b>Figure II.6:</b>	Recombinant AACT secreted from AACT-S cells is N-glycosylated and harbours sugar residues typical of plant complex N-glycans	81
<b>Figure II.7:</b>	AACT susceptibility to tobacco extracellular proteases	83

## Chapitre III –

### Nucleocytoplasmic transit of human $\alpha$ 1-antichymotrypsin expressed in tobacco leaf epidermal cells

<b>Figure III.1 :</b> Gene constructs devised for the experiments	104
<b>Figure III.2 :</b> DNA-binding activity of rAACT and rAACT $\Delta$ K	106
<b>Figure III.3 :</b> Antichymotrypsin activity of rAACT and rAACT $\Delta$ K	107
<b>Figure III.4 :</b> Nucleocytoplasmic distribution of rAACT and rAACT $\Delta$ K in tobacco leaf epidermal cells	109
<b>Figure III.5 :</b> Typical cortical views of tobacco leaf epidermal cells expressing GFP-rAACT or GFP-rAACT $\Delta$ K	110
<b>Figure III.6 :</b> Stability of free or DNA linked rAACT, rAACT $\Delta$ K and hAACT following trypsin partial degradation	112

## LISTE DES TABLEAUX

### Chapitre I - Introduction, hypothèses et objectifs

<b>Tableau I.1 :</b> Exemples de protéines recombinantes produites dans les plantes au stade des essais cliniques ou à des stades de commercialisation	5
<b>Tableau I.2 :</b> Impact of subcellular targeting on recombinant protein yield in transgenic plant systems	16
<b>Tableau I.3 :</b> Low-molecular-weight protease inhibitors commonly used for protease characterization and protein stabilization in plant extracts	28
<b>Tableau I.4 :</b> Enzymes et transporteur nécessaires à la synthèse et à l'incorporation du Neu5Ac sur les N-glycannes des glycoprotéines végétales	45
<b>Tableau I.5 :</b> Fonctions et maladies associées aux serpins humaines	51

## GLOSSAIRE

3,4-DCI	3,4-dichloro-isocoumarine
$\alpha$ 1-PI	<i><math>\alpha</math>1-proteinase inhibitor</i>
aa	Acide aminé
AACT	$\alpha$ 1-antichymotrypsine
hAACT	AACT humaine
brAACT	AACT recombinante bactérienne
ADN (DNA)	Acide désoxyribonucléique
AEBSF	4-(2-aminoethyl)-benzenesulfonyl fluoride
ARN	Acide ribonucléique
BET	Bromure d'éthidium
BiP	<i>Binding protein</i>
BSA	<i>Bovin serum albumin</i> , albumine de sérum bovin
BY-2	Cellule de tabac <i>Nicotiana tabacum</i> , cv Bright Yellow 2
cDNA	ADN complémentaire
CHO	<i>Chinese hamster ovary cells</i>
ConA	Lectine Concanavaleine A
CT	Chymotrypsine (E.C. 3.4.21.1)
cv	Cultivar
Dpp	Dolichol pyrophosphate
E-64	<i>trans</i> -époxy succinyl-L-leucylamido-(4-guanidino)butane
EDTA	Acide éthylène-diaminetétraacétique
EM	<i>Extracellular medium</i> , milieu extracellulaire
ER/RE	Réticulum endoplasmique
Fuc	D-fucose
FucT	Fucosyl transférase
Gal	D-galactose
GalNAc	N-acétyl-D-galactosamine
GFP	<i>Green fluorescent protein</i>
Glc	D-glucose
GlcNAc	N-acétyl-D-glucosamine
GNT I	N-acétylglucosaminyltransférase I
GNT II	N-acétylglucosaminyltransférase II
GSH	Glutathione S-transférase réduite
GST	Glutathione S-transférase
His-Tag	Chaîne de 6 histidines
HNE	<i>Human neutrophil elastase</i> , elastase de neutrophile humain (E.C. 3.4.21.37)
HRP	<i>Horseradish peroxidase</i> , peroxydase de raifort
Ig	Immunoglobuline
IM	<i>Intracellular medium</i> , milieu intracellulaire
IP	Inhibiteur de protéases
KDEL	(lys-Asp-Gln-Leu) signal de rétention dans le RE
Lea, Lewis a	Epitope Lewis a, Gal $\beta$ -1,3(Fuc $\alpha$ -1,4)GlcNAc
Man	D-mannose
MCA	Méthylcoumarine
MENT	<i>Myeloid and erythroid nuclear termination stage-specific protein</i>
NAc	Groupe N-acétyl
NES	<i>Nuclear export signal</i> , signal d'export nucléaire
NLS	<i>Nuclear localization signal</i> , signal de localisation nucléaire

NPC	<i>Nuclear pore complex</i> , pore nucléaire
Neu5Ac	Acide N-acétylneuraminique ou acide 2-céto-5-amino-3,5-didésoxy-D-glycero-D-galactononulosonique
Neu5Gc	Acide N-glycolylneuraminique
<i>Nicotiana tabacum</i>	Le tabac
NPTII	Néomycine phosphotransférase II
PAGE	<i>Polyacrylamide gel electrophoresis</i> , électrophorèse sur gel de polyacrylamide
PAI-2	<i>Plasminogen activator inhibitor-2</i>
PBS	<i>Phosphate buffer saline</i> , tampon phosphate salin
PCR	<i>Polymerisation chain reaction</i> , réaction de polymérisation en chaîne
PDB	<i>Protein Data Bank</i>
PDI	Peptide signal de protéine disulfide-isomérase
PEP	Phosphoénolpyruvate
PI-6	<i>Proteinase inhibitor 6</i>
PMSF	Phénylméthane sulfonyl fluoride
PNGaseF	Peptide N-glycosidase F (E.C. 3.2.218)
RCA	<i>Ricinus communis</i> agglutinin I
RCL	<i>Reactive center loop</i> , boucle d'inhibition
rpm	Rotations par minute
RT	Rétro-transcription
SE	<i>Standard error</i> , écart-type
SDS	Sodium dodécyl sulfate
SNA-I	Lectine I de <i>Sambucus nigra</i>
TBS	<i>Tris buffer saline</i> , tampon Tris salin
TLCK	Tosyl lysyl chlorométhyl cétone
TPCK	Tosyl phénylalanyl chlorométhyl cétone
TSP	<i>Total soluble proteins</i> , protéines solubles totales
TTBS	<i>Tris buffer saline</i> supplémenté avec 0,1% Tween20
v/v	Volume/volume
w/v	<i>Weight/volume</i> , poids/volume
Xyl	D-xylose
XylIT	Xylosyl transférase

## **CHAPITRE I**

### **Introduction, hypothèses et objectifs**

## I.0 Introduction générale : des plantes pour la production de protéines recombinantes

Les premières plantes transgéniques ont vu le jour au début des années 1980 (Van Larebeke et al., 1975; De Block et al., 1984) et ont amené de très nombreuses perspectives tant industrielles que thérapeutiques. Alors que cette technologie visait à l'origine à améliorer les caractéristiques agronomiques des plantes, elle a rapidement donné naissance à un nouveau champ d'application, la production de macromolécules recombinantes (Hiatt et al., 1989), discipline désormais connue sous le nom de moléculature. En pratique, l'utilisation de plantes comme bioréacteurs pour la production de protéines recombinantes offre de nombreux avantages, en comparaison des microorganismes, des cellules animales et des animaux transgéniques (Daniell et al., 2001; Gomord et al., 2004). En regard de nos connaissances actuelles, les plantes permettraient une production de protéines complexes en grandes quantités et à un coût relativement faible. De plus, les risques de contamination par des agents pathogènes transmissibles à l'Homme sont pratiquement nuls. La cellule végétale renferme de surcroît toute la machinerie enzymatique et cellulaire nécessaire au repliement des protéines, à l'assemblage de protéines multimériques et aux modifications post-traductionnelles des protéines complexes. Vingt ans après la création des premières plantes transgéniques, de nombreuses protéines recombinantes d'intérêt thérapeutique ont été exprimées avec succès dans des cellules végétales incluant des anticorps, des hormones, des enzymes et des vaccins (Ma et al., 2003; Gomord et al., 2004). Si la capacité des plantes à produire des protéines recombinantes n'est pas remise en question, le nombre de biomolécules commercialisées ou en voie de commercialisation reste par contre restreint (Ma et al., 2003; voir **Tableau I.1**) malgré l'engouement à la fois scientifique et industriel dont la moléculature fait l'objet.

La rentabilité économique des protéines recombinantes dépend essentiellement de l'obtention d'un produit de qualité à un haut rendement. Beaucoup d'efforts ont été déployés au cours des années pour améliorer l'expression des protéines recombinantes dans les plantes (Streatfield, 2007). Les recherches ont porté, entre autres, sur l'usage des codons de la plante hôte, sur le choix de promoteurs forts souvent d'origine virale ou bactérienne, sur la stabilisation des ARN messagers (ARNm) et sur l'élimination des introns. À quelques exceptions près, les résultats obtenus sont restés mitigés. Le niveau



d'accumulation d'une protéine *in vivo* ne dépend pas uniquement des taux de transcription et de traduction de séquences codantes, mais aussi d'une somme de processus survenant en aval de la synthèse peptidique. Certaines maturations post-traductionnelles jouent un rôle capital dans la stabilité des protéines recombinantes, incluant notamment la maturation protéolytique et la glycosylation, qui influent de manière importante à la fois sur la stabilité et l'activité biologique des protéines produites par la cellule.

La cellule est le siège d'une activité protéolytique intense essentielle à son métabolisme (Schaller, 2004) qui, bien qu'importante à l'échelle cellulaire, apparaît souvent comme néfaste à l'accumulation de protéines recombinantes (Doran, 2006; Goulet et Michaud, 2006). En parallèle, des problèmes liés à la glycosylation peuvent aussi réduire la stabilité et l'activité biologique des protéines (Matsumoto et al., 1995). Parmi l'éventail des systèmes de production hétérologue actuellement disponibles, les procaryotes et les levures présentent des patrons de glycosylation inappropriés ou même inexistantes, nuisant à la production de plusieurs protéines d'intérêt thérapeutique. Les levures, par exemple, produisent une hypermannosylation de certaines protéines, défavorable à leur utilisation médicale (Gemmill et Trimble, 1999). Si la N-glycosylation végétale, en revanche, se rapproche de la glycosylation animale, des résidus typiques aux végétaux sont souvent présents sur les glycanes complexes, le  $\beta$ -1,2 xylose et l' $\alpha$ -1,3 fucose. Ces résidus s'avèrent potentiellement immunogènes, ce qui limite encore aujourd'hui l'utilisation thérapeutique des glycoprotéines d'origine végétale (Faye et al., 1993; van Ree et al., 2000; Bardor et al., 2003a; Gomord et al., 2004). Ces modifications post-traductionnelles, associées aux problèmes d'expression et d'accumulation des protéines recombinantes, conduisent à l'obtention d'un produit final souvent hétérogène et à de faibles rendements, rendant leur commercialisation d'autant plus difficile (Faye et al., 2005).

Dans ce contexte, nous avons étudié ici l'expression, la maturation et l'activité des inhibiteurs de protéases thérapeutiques produits en système hétérologue végétal, en fonction de leur niveau de glycosylation et de leur maturation protéolytique. Nous avons, plus spécifiquement, choisi un inhibiteur modèle de la superfamille des serpinés, l' $\alpha$ 1-antichymotrypsine (AACT), exprimé dans des suspensions cellulaires de tabac *Nicotiana tabacum*. Les serpinés sont, pour la plupart, des inhibiteurs de protéases de type sérine impliqués dans divers processus physiologiques chez l'Homme tels la coagulation,

l'inflammation, la fibrinolyse, l'apoptose, le repliement des protéines ou encore l'activation du complément (Silverman et al., 2001). Étant donné leur rôle essentiel dans l'organisme, le dysfonctionnement d'une serpine ou la modification de sa séquence génique entraînent souvent une maladie grave. La production de serpinés recombinantes constitue alors une des perspectives majeures de traitement (Stein et Carrell, 1995; Chang et Lomas, 1998). En introduction à mes travaux de thèse, les sections qui suivent traitent de la maturation et de la dégradation protéolytique des protéines produites en systèmes hétérologues végétaux, de la N-glycosylation des protéines dans les cellules végétales et des caractéristiques et propriétés physiologiques des serpinés. Finalement, les hypothèses et objectifs de mes travaux sont présentés.

**Tableau I.1 Exemples de protéines recombinantes produites dans les plantes au stade des essais cliniques ou à des stades avancés de commercialisation.**

Adapté de (Marshall, 2006)

Protéine	Application	Plante	Étape clinique	Compagnie
<b>Anticorps</b>				
slgA	Prévention de la carie dentaire	<i>Nicotiana tabacum</i>	Phase 2, licence de UE	Planet Biotechnology
Anticorps monoclonaux recombinants	Agent de purification du vaccin de l'hépatite B	<i>N. tabacum</i>	Cuba, approuvé en 2006	CIGB (Cuba)
<b>Vaccins</b>				
Heat-labile toxin E. coli	Diarrhée du voyageur	<i>Zea mays</i>	Phase 1 complétée - interrompu	Prodigene
TGE	Gastroentérite chez le porc	<i>Z. mays</i>	Phase 1 complétée - interrompu	Prodigene
n.i.	Lymphome non-Hodgkin	<i>N. tabacum</i>	Phase 1 réussie en 2002	LSBC
n.i.	Parvovirus félin	<i>N. tabacum</i>	Très avancé, confidentiel	LSBC
Heat-labile toxin E. coli	Diarrhée du voyageur	<i>Solanum tuberosum</i>	Phase 1	Arizona State Univ.
Antigène de surface du virus de l'hépatite B	Hépatite B	<i>S. tuberosum</i>	Phase 1	Arizona State Univ.
Protéine de capsid	Virus Norwalk	<i>S. tuberosum</i>	Phase 1	Arizona State Univ.
Protéine de capsid	Virus Norwalk	<i>N. tabacum</i>	Phase 1 / 2	Arizona State Univ.
n.i.	Maladie Newcastle chez la volaille	Culture cellulaire de <i>N. tabacum</i>	Approuvé USDA en 2006	Dow Agrisciences
n.i.	Maladies chez les chevaux, chiens et volailles	Culture cellulaire de <i>N. tabacum</i>	Confidentiel	Dow Agrisciences
n.i.	Coccidiose de la volaille	<i>Brassica campestris</i>	CFIA phase 2	Guardian Biosciences
Glycoprotéine de la rage	Rage	<i>Spinacia oleracea</i>	Phase 1 complétée en 2002	D. Yusbov et al.
<b>Enzymes</b>				
Lipase gastrique	Fibrose kystique	<i>Z. mays</i>	Phase 2	Meristem Therapeutics
Trypsine	Traitement des blessures	<i>Z. mays</i>	Disponible chez Sigma	Prodigene
Lysozyme	Usage de laboratoire	<i>Oryza sativa</i>	Disponible (Ventria)	Ventria Biosciences
Glucocérébrosidase	Maladie de Gaucher	Culture cellulaire de <i>Daucus carota</i>	Phase 3	Protalix
<b>Autres</b>				
Protéine TGF-Beta	Cancer des ovaires	<i>N. tabacum</i>	Essais chez les animaux	Chlorogen et al.
Aprotinine	Usage non médical	<i>Z. mays</i>	Vente chez Sigma discontinuée	Prodigene
Antigènes viraux	Diagnostic de Hantaan et Puumala	<i>Cucumis melo</i>	Disponible Nexgen Biotech	Nexgen Biotech
Avidine	Usage diagnostique	<i>Z. mays</i>	Disponible chez Sigma	Prodigene
Facteur intrinsèque humain	Déficience en vitamine B12	<i>Arabidopsis thaliana</i>	Approuvé, 37 patients en phase clinique	Cobento AS
β-glucuronidase (GUS)	Usage non medical	<i>Z. mays</i>	Vente chez Sigma discontinuée	Prodigene
ICAM1	Rhume commun	<i>N. tabacum</i>	Phase 1	Planet Biotechnology
Insuline	Diabète	<i>Carthamus tinctorius</i>	Phase 1 prévue en 2008	Symbiosis
Interféron-Alpha	Hépatites "B" & "C", cancer	<i>Lemna</i>	Phase 1	Biorex
Lactoferrine	Désordres gastrointestinaux	<i>Z. mays</i>	Phase 1	Meristem Therapeutics
Lactoferrine	Milieu pour culture cellulaire	<i>O. sativa</i>	Disponible (Ventria Biosciences)	Ventria Biosciences
Récepteur hormonaux stimulant la thyroïde	Diagnostic de la maladie de Graves	<i>C. melo</i>	Disponible Nexgen Biotech	Nexgen Biotech
Transcobalamine	Diagnostic / recherche	<i>A. thaliana</i>	Disponible (Cobento AS)	Cobento AS
Vitronectine	Usage de laboratoire	<i>N. tabacum</i>	Disponible en 2007 (Farmacule)	Farmacule

n.i.: antigène non identifié

## **I.1. La dégradation protéolytique des protéines recombinantes**

La cellule végétale comporte de nombreuses protéases qui assurent diverses fonctions de régulation, de maturation protéolytique et la demi-vie (turnover) des protéines (Schaller, 2004). Bien qu'essentielles au métabolisme cellulaire, ces protéases représentent une menace de taille contre la stabilité des protéines recombinantes qui peuvent subir des clivages partiels, parfois délétères pour leur activité biologique (Outchkourov et al., 2003; Badri et al., 2007a). Le produit final devient alors plus hétérogène, s'éloignant parfois considérablement de sa contrepartie naturelle. Dans d'autres cas, la protéolyse peut être totale, entraînant des rendements faibles (Takaiwa et al., 2007). En système hétérologue, la dégradation des protéines exogènes peut se produire *in planta*, durant les phases d'expression, ou *ex planta*, durant les diverses étapes post-récolte, en cours d'extraction ou de purification. Dans ce dernier cas, la protéolyse peut être importante car l'homogénéisation et le bris des cellules libèrent les protéases et autres composés dénaturants des différents organites (en particulier la vacuole), directement dans l'extrait brut (Rivard et al., 2006). Diverses stratégies ont été envisagées depuis quelques années pour limiter la dégradation protéolytique *in planta* comme *ex planta*. Les sections qui suivent, rédigées en anglais sous la forme d'un article de synthèse, décrivent les principales de ces stratégies. Cette synthèse a été soumise pour publication à la revue *Plant Biotechnology Journal*.

## **Preventing unintended proteolysis in plant protein biofactories**

Meriem Benchabane <sup>1</sup>, Charles Goulet <sup>1</sup>, Loïc Faye <sup>2</sup>, Véronique Gomord <sup>2</sup> & Dominique Michaud <sup>1,\*</sup>

<sup>1</sup> *Département de phytologie, CRH/INAF, Université Laval, Québec (QC) Canada G1K 7P4*

<sup>2</sup> *CNRS UMR 6037, IFRMP 23, GDR 2590, Université de Rouen, Faculté des Sciences Bât. Ext. Biologie, 76821 Mont-Saint-Aignan cedex, France*

\* To whom correspondence should be sent.

### I.1.1 Résumé

De nombreux articles ont été publiés au cours de la dernière décennie sur le potentiel des plantes comme plate-formes d'expression pour la production hétérologue de protéines d'intérêt thérapeutique. Des progrès importants ont été réalisés, en particulier, sur l'optimisation de la transcription et de la traduction des transgènes chez les plantes, et sur l'élucidation des modifications post-traductionnelles spécifiques à la machinerie cellulaire végétale. Nous abordons dans cet article de synthèse l'important phénomène de la dégradation des protéines recombinantes dans les plantes, qui détermine aussi bien le rendement que l'homogénéité et la qualité du produit final. À l'inverse de plusieurs composés biopharmaceutiques plus stables et souvent de structure plus simple, les protéines montrent une tendance naturelle à l'hétérogénéité structurale, résultat en partie de l'instabilité inhérente des chaînes polypeptidiques exprimées en conditions hétérologues. La dégradation protéolytique, notamment, peut altérer de façon marquée l'intégrité structurale et l'accumulation des protéines recombinantes en systèmes d'expression végétaux, aussi bien *in planta* en cours d'expression qu'*ex planta* pendant l'étape d'extraction. Nous décrivons ici les principales stratégies proposées pour minimiser les pertes en protéines recombinantes dans les plate-formes de production végétales, incluant l'expression du transgène dans des organes ou tissus spécifiques, l'adressage de la protéine vers différents organites cellulaires, la fusion de domaines stabilisateurs, la sécrétion dans des fluides extracellulaires et la co-expression d'inhibiteurs de protéases recombinants.

## Summary

Numerous research and review papers have been published over the last decade assessing the potential of plants as useful hosts for the heterologous expression of clinically-useful proteins. Significant progress has been made, in particular, in optimizing transgene transcription and translation in plants, and in elucidating the complex protein post-translational modifications typical of the plant cell machinery. In this review we address the important issue of recombinant protein degradation in plant expression platforms, which directly impacts final yield, homogeneity and overall quality of the resulting protein product. Unlike several more stable –and structurally less complex– biopharmaceuticals, recombinant proteins present a natural tendency to structure heterogeneity resulting in part from the inherent instability of polypeptide chains expressed in heterologous environments. Proteolytic processing, notably, may dramatically alter the structural integrity and overall accumulation of recombinant proteins in plant expression systems, both *in planta* during expression and *ex planta* after extraction. Here we describe the strategies currently proposed to minimize protein loss in plant protein factories, including organ-specific transgene expression, organelle-specific protein targeting, grafting of stabilizing protein domains to labile proteins, protein secretion in natural fluids and co-expression of companion protease inhibitors.

### **I.1.2 Introduction**

A number of heterologous expression systems have been devised over the last decade for the production of clinically-useful recombinant proteins, involving in most cases eucaryotic hosts competent in performing complex post-translational modifications typical of mammalian proteins (Dyck et al., 2003; Ma et al., 2003; Gerngross, 2004; Wurm, 2004). Plant-based expression platforms, in particular, offer unique advantages over traditional expression hosts in terms of cost efficiency, product safety, available expertise and facilities for culture and storage, and scalability easily adaptable to production needs (Daniell et al., 2001; Ma et al., 2003; Twyman et al., 2003; Streatfield, 2007). To date, several proteins of medical interest have been expressed successfully in plants, including a variety of antibodies, vaccine antigens, protein allergens, enzymes and enzyme inhibitors, coagulation factors, cytokines and hormones (Hiatt et al., 1989; Mason et al., 1992; Ma et al., 1995; Ruggiero et al., 2000; Ma et al., 2003; 2005b; Kirk and Webb, 2005; Twyman et al., 2005; Floss et al., 2007; Lienard et al., 2007).

A significant challenge in most systems now is to optimize yield and quality of the recombinant protein product. Significant progress has been achieved over the last fifteen years to optimize transgene transcription and translation in plants (Potenza et al., 2004; Streatfield, 2007) or to elucidate and modulate the complex protein post-translational modifications characteristic of the plant cell machinery (Gomord and Faye, 2004; Faye et al., 2005). Despite these advances, ensuring adequate yields and quality of recombinant proteins often remains a difficult task. Unlike several more stable [and structurally less complex] biopharmaceuticals, proteins present a natural tendency to structure heterogeneity, often giving a complex mixture of variants differing in their primary or tertiary structure (Faye et al., 2005). This natural bias of proteins, that underlines the important issues of product authenticity and heterogeneity, also stresses the importance to correctly understanding the various post-translational steps of the whole protein synthesis–recovery process, from the maturation and assembly of nascent protein backbones in transgenic host cells to extraction and purification of the protein product after biomass harvest.



### **1.1.2.1 Proteolytic degradation of recombinant proteins in plant systems**

One factor strongly influencing recombinant protein quality and yield is the relative inherent stability of polypeptide chains expressed in a heterologous environment (Faye et al., 2005). Living cells are the theatre of numerous proteolytic processes, essential to ensure vital metabolic functions but often a burden for the effective production of biologically active proteins (Doran, 2006; Goulet and Michaud, 2006). Proteolytic enzymes –or proteases– contribute to the overall control of metabolic and transduction pathways by directing the activation or the hydrolysis of proteins implicated in key regulatory processes, or by contributing to the elimination of misfolded proteins and to the selective recycling of amino acids from short-lived proteins (Vierstra, 2003; Schaller, 2004). In plants, these enzymes also initiate the general recycling of proteins in senescing organs and the mobilization of amino acids constituent of seed or tuber storage proteins during germination (Müntz, 2007). Hundreds of genes code for proteins involved in proteolytic processes in plants, with for instance an estimate of 1,900 genes in *Arabidopsis* directly or indirectly implicated in the hydrolysis of peptide bonds (Schaller 2004; Smalle and Vierstra, 2004).

From a practical viewpoint, the ubiquitous nature of proteolytic processes (Schaller, 2004) and the diversity of possible protease forms in the plant genome (Beers et al., 2004) pose a significant challenge to the efficient production of several recombinant proteins. Whereas some proteins accumulate at high levels in plant systems, other proteins apparently undergo extensive hydrolysis, with a strong negative impact on final yields despite easily detectable mRNA transcripts. Accumulation levels lower than 0.01% of total soluble proteins (TSP) in plant tissues have been observed for several proteins of therapeutic value, including for instance serum protein C, interferon  $\beta$ , erythropoietin and epidermal growth factor (Daniell et al., 2001). Low accumulation rates of intact protein products were also observed for other recombinant proteins of practical interest including immunoglobulins, antibodies and enzyme inhibitors (e.g., Stevens et al., 2000; Sharp and Doran, 2001; Outchkourov et al., 2003; Ma et al., 2003; Schillberg et al., 2005), again suggesting a key influence for proteolysis on the overall efficiency of plant-based protein factories.

Proteases may affect the integrity of recombinant proteins in different ways, both *in planta* during protein expression and *ex planta* during extraction and subsequent downstream processing (Michaud et al., 1998; Rivard et al., 2006). Depending on the number of 'susceptible' cleavage sites accessible to endogenous proteases for peptide bond hydrolysis, the protein may undergo complete hydrolysis directly impacting its final yield, or partial trimming altering activity or homogeneity of the final protein product. Even though interesting yields may be obtained in terms of net protein levels, the final product may show in several cases an altered integrity, structure heterogeneity and/or deficient biological activity potentially altering its value for commercialization (Badri et al., 2007a). Several strategies have been considered recently to minimize proteolysis in plant protein factories. The present review summarizes these strategies, taking into account proteolysis observed *in planta* before harvesting, and unwanted proteolytic degradation observed *ex planta* during protein recovery from plant tissues and cells.

### **1.1.3 Stabilizing recombinant proteins *in planta***

In theory, the easiest way to minimize recombinant protein hydrolysis *in planta* would be to use mutant lines deficient in protease(s) active against the protein of interest. Protease-deficient strains of simple expression hosts such as *Escherichia coli* (Jiang et al., 2002; Rozkov and Enfors, 2004) and yeasts (Sreekrishna et al., 1997; Cereghino and Cregg, 2000; Macauley-Patrick et al., 2005) have been devised several years ago to minimize recombinant protein loss *in vivo*, and such strains are now available commercially and used routinely in most protein biochemistry labs. This approach, however, may be hardly applicable to complex, multicellular organisms such as higher plants, given the central role of proteolysis in growth and development (Schaller, 2004), the expected diversity and variability of protease forms in plant cells and tissues (Beers et al., 2004), and the high number of protease:protein interactions possibly taking place *in vivo*, that will depend both on the structural characteristics of each protein to be expressed and on the metabolic status of each cell compartment or tissue targeted for protein deposition. At this stage, a case-by-case empirical assessment involving the testing of different expression strategies remains the most straightforward way to optimize the production yield and structural integrity of a given protein. Common strategies to overcome unwanted proteolysis *in planta* consist to target transgene expression or protein accumulation to specific tissues or cellular organelles. Approaches involving the grafting of

protein stabilizing fusion domains to recombinant proteins or the co-expression of companion protease inhibitors interfering with endogenous proteases have also been proposed recently.

### **I.1.3.1 Tissue-specific transgene expression**

For several reasons, the specific tissue or organ selected for recombinant protein production has a strong influence on final yields and product quality. Cellular proteases, notably, differ from one tissue to another in terms of quantity and quality [or overall substrate specificity] (Schaller, 2004), with a possible differential impact on the integrity of proteins. Historically, recombinant proteins in plants have often been expressed under the control of strong constitutive promoters such as the cauliflower mosaic virus 35S promoter or the maize ubiquitin 1 promoter, but several tissue-/organ-specific [e.g. seed-specific] promoters have been isolated and are used now to express transgenes in selected tissues and organs (Potenza et al., 2004). In practice, an effective scheme for recombinant protein production should not only provide effective transcription of the transgene sequence, but also involve an accumulation site with low levels of overall proteolytic activity or with endogenous protease species showing no (or little) specific activity against accessible peptides bonds in the protein of interest.

Up to now, green leaves have been the destination of choice for several recombinant proteins, given their rapid growth rate, the possibility in some platforms to harvest leaf material more than once over the growing season, and the availability of numerous regulatory sequences well adapted to transgene expression in the leaf cell environment (Daniell et al., 2001). On the other hand, the highly active metabolism of leaf tissues, that typically exhibit high protein synthesis and turnover rates, may represent a significant hurdle to protein accumulation *in vivo*. In particular, the increased protease levels in senescing leaves (Lin and Wu, 2004; Otegui et al., 2005; Parrott et al., 2005) represent a potential drawback in leaf-based production systems (Stevens et al., 2000; Birch-Machin et al., 2004), both *in vivo* before protein recovery while senescence-associated amino acid recycling is initiated, and *in vitro* during extraction, once cellular proteases have been released in the extraction medium together with the protein(s) of interest. The abundance of poorly-specific proteases and phenolic compounds in green tissues (Michaud and Asselin, 1995), finally, points to the problem of post-harvest protein

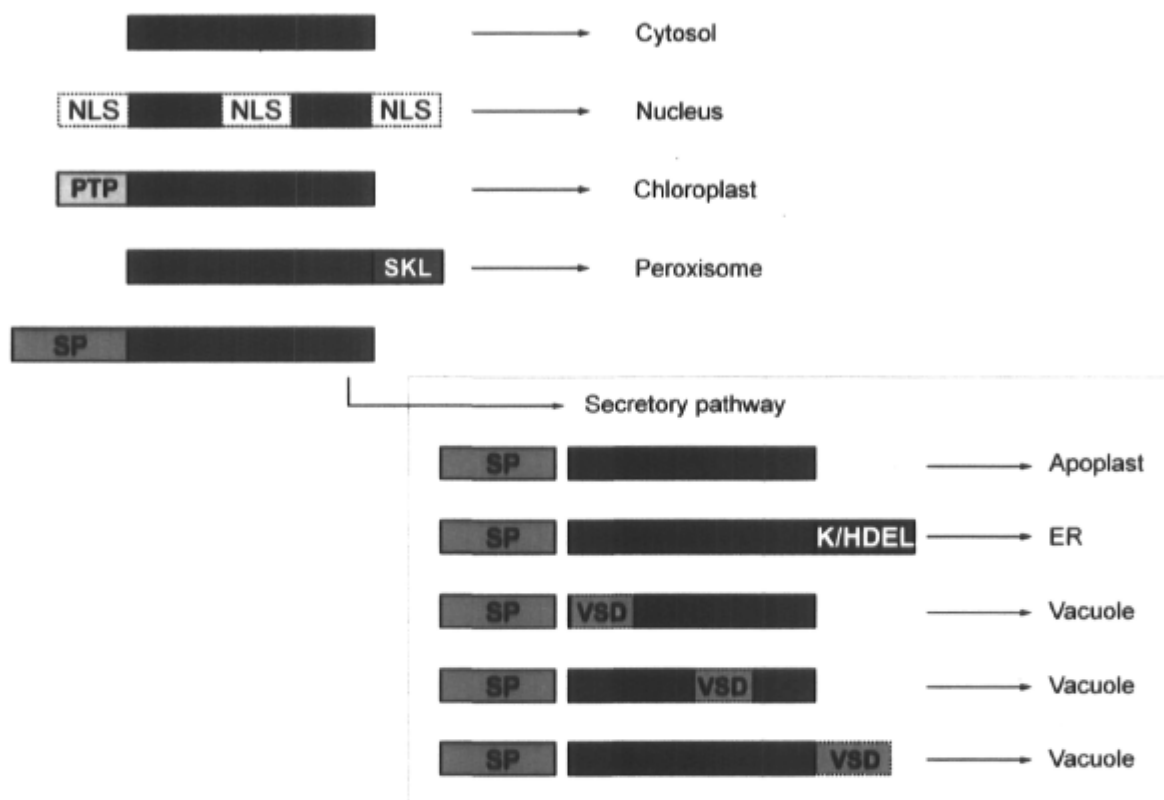
degradation and denaturation in leaves, which can make it necessary to process leaf biomass shortly after harvest, or to store this material at low temperatures until protein extraction (Schillberg et al., 2005).

By contrast with leaf tissues, the seeds and storage organs of several crops appear in general well suited for the efficient deposition and post-harvest storage of recombinant proteins (Stoger et al., 2005). Proteins of clinical interest have been successfully expressed in seeds of several species including major cereal crops (e.g. maize, rice, barley and wheat), soybean and safflower, under the control of either constitutive or seed-specific promoters. Recombinant proteins in seeds often accumulate in storage organelles, mainly in protein bodies and protein storage vacuoles (Streatfield et al., 2003; Yang et al., 2005; Drakakaki et al., 2006). The desiccated nature of mature seeds, along with the low abundance of active proteases in seed tissues during dormancy, prevent extensive proteolysis and promote long-term stability of proteins *in planta* (Fiedler and Conrad, 1995; Stoger et al., 2000). In a similar way, the tubers and stocking roots of some species exhibit reduced metabolic activity and may represent an interesting solution for recombinant protein storage. Artsaenko *et al.* (1998), for instance, documented the long-term stability of a single chain Fv antibody in potato tubers after storage for more than a year. From a technological viewpoint, storage organs also present anatomic features favorable to protein stability, such as the presence of protein and oil bodies in seeds, that may sequester proteins and minimize their exposition to proteases during extraction (Stoger et al., 2005).

#### **1.1.3.2 Organelle-specific protein targeting**

Protein sequestration in –or targeting to– specific cell compartments has also been readily recognized as a key factor determining the overall stability and yield of recombinant proteins *in planta* (Wandelt et al., 1992; Schouten et al., 1996; Gomord et al., 1997). Organelles play specific, complementary functions in the cell and thus harbour their own metabolic machinery, including a protease complement well adapted to their particular enzymatic and physicochemical environment (Callis, 1995). Not surprisingly, the targeting of foreign proteins to different organelles using appropriate peptidic targeting signals has a strong impact on their accumulation rate (**Table 1.2**). In biochemical terms, the stability of a given protein *in vivo* will depend on the relative steric accessibility of peptides or peptide

strings on the protein chain susceptible to the proteases present. In practice, the choice of a suitable cellular destination will also depend on the structural characteristics of the recombinant protein, that will often dictate specific co- or post-translational modifications essential for adequate activity, stability and/or homogeneity (Faye et al., 2005).



**Figure 1.1 Subcellular targeting of recombinant proteins in plant cells.** Recombinant proteins bearing an N-terminal signal peptide (SP) in their primary sequence enter the cell secretory pathway via the endoplasmic reticulum (ER), and then travel through the Golgi system to be secreted in the apoplast [default pathway] or directed to the vacuole if a vacuolar sorting determinant (VSD) is present somewhere in the protein sequence. Proteins secreted into the ER can also be retained in this compartment by the grafting of an ER retention signal –the KDEL (or HDEL) tetrapeptidic motif– at the C terminus. Proteins with no signal peptide accumulate into the cytosol [default location], or migrate to specific organelles when an appropriate peptidic signal is included in the transgene sequence. Peptidic signals used recently in transgenic plant platforms include plastid (e.g. chloroplast) transit peptides (PTP), nuclear localization signals (NLS) and the tripeptidic peroxisome target sequence, SKL.

**Table I.2 Impact of subcellular targeting on recombinant protein yield in transgenic plant systems – Selected examples**

Protein	Transformed species	Plant organ	Yield *						References
			Cytosol	ER	Vacuole	Apoplasm	Plastid	Nucleus	
<b>Antibodies</b>									
ScFv anti-cutinase	<i>Nicotiana tabacum</i>	Leaf	0	100			1		Schouten <i>et al.</i> , 1996
ScFv anti-oxazolone	<i>N. tabacum</i>	Leaf		10-20			1		Fielder <i>et al.</i> , 1997
		Seed		5, 22			1		
ScFv anti-oxazolone	<i>Solanum tuberosum</i>	tuber		1			1		Artsaenko <i>et al.</i> , 1998
ScFv anti-dihydroflavonol 4-reductase	<i>Petunia hybrida</i> **	petal	1	2, 30					De Jaeger <i>et al.</i> , 1999
		leaf	1	20, 60					
BiscFv 2429	<i>N. tabacum</i>	BY-2 cell culture	traces	10			1		Fischer <i>et al.</i> , 1999
FAb MAK33	<i>Arabidopsis thaliana</i>	Leaf /seed		1			1		Peeters <i>et al.</i> , 2001a
scFv anti-carcinoembryonic	<i>N. tabacum</i>	Leaf		25			1		Stoger <i>et al.</i> , 2002
Ab anti-carcinoembryonic	<i>N. tabacum</i> **	Leaf		2-6			1		Vaquero <i>et al.</i> , 2002
Ab 14D9 κ chain	<i>N. tabacum</i>	Leaf		8			1		Petruccelli <i>et al.</i> 2006
Ab 14D9 γ chain				4			1		
<b>Vaccines</b>									
<i>E. coli</i> heat-labile enterotoxin B	<i>Zea mays</i>	Seed	1	100	20 000	3300	7	21	Streatfield <i>et al.</i> , 2003
Hepatis B surface antigen	<i>N. tabacum</i>	BY-2 cell culture	1	1,4		1,8			Sojikul <i>et al.</i> , 2003
Japanese cedar pollen allergens	<i>Oryza sativa</i>	Seed	0	4-6		1			Takagi <i>et al.</i> , 2005
<b>Medical proteins</b>									
Human epidermal growth factor	<i>N. tabacum</i>	Leaf	1			10000			Wirth <i>et al.</i> , 2004
Human growth hormone	<i>Nicotiana benthamiana</i> **	Leaf	1			1000	10		Gils <i>et al.</i> , 2005
<b>Agronomical/industrial proteins</b>									
Vicilin	<i>N. tabacum</i>	Leaf		100			1		Wandelt <i>et al.</i> , 1992
	<i>Medicago sativa</i>	Leaf		20			1		
Equistatin	<i>S. tuberosum</i>	Leaf	1	5					Outchkourov <i>et al.</i> , 2003
Silk-like protein	<i>A. thaliana</i>	Leaf	1	13	0		5		Yang <i>et al.</i> , 2005
		Seed	1	7,8	5,4		0		

\* Relative yield, compared to the yield in 'control' compartment (value of 1).

\*\* Transient expression

Eventual metabolic interference effects on the host plant actively expressing the recombinant protein should also be assessed, given the possible onset of organelle-dependent effects of biologically active proteins *in planta*. Cytosolic or apoplasmic accumulation of the human growth factor, for instance, showed toxic effects in leaves of *Nicotiana benthamiana*, while in contrast no negative effects were observed for the same protein targeted to the chloroplast (Gils et al., 2005). Likewise, the bovine protease inhibitor aprotinin was accumulated at high specific levels when retained into the ER of potato leaf cells, but this positive effect of ER retention was counterbalanced by a general decrease of total protein content in leaves, presumably due to the exogenous inhibitor affecting key steps of protein biosynthesis *in vivo* (Badri et al., 2007b). Considering the complexity of protein maturation processes in plant cells and the often unpredictable nature of pleiotropic effects in transgenic host plants, the most appropriate way to select a suitable cellular destination at this stage consists of empirically testing different possible destinations, taking into account current knowledge on the protein to express and on the physicochemical and enzymatic microenvironment of the different organelles available for protein accumulation. Several subcellular compartments have been considered as possible destinations for recombinant proteins in plant cells, notably including the cytosol, the chloroplast and different subcompartments of the cell secretory pathway (Ma et al., 2003; Daniell, 2006).

*Retention into the cytosol.* In practice, the absence of a targeting signal in the transgene sequence prevents migration of the recombinant protein out of the cytosol following mRNA translation (**Figure I.1**). Recombinant proteins retained into the cytosol are usually detected at very low levels despite good transgene transcription rates, giving in several cases accumulation rates lower than 0.1% TSP (Conrad and Fiedler, 1998). Cytosolic targeting of the tomato mosaic virus antibody 'rAb29' in tobacco leaf cells, for instance, resulted in very weak accumulation rates, whereas the same transgene including a signal peptide for extracellular secretion produced easily detectable amounts of this same protein (Schillberg et al., 1999). Similarly, targeting human growth hormone to the cytosol of *Nicotiana benthamiana* leaf cells led to protein levels of about 0.01% TSP, in contrast with concentrations reaching 10% TSP for the same protein targeted to the apoplast (Gils et al., 2005). Several factors may explain the limited suitability of the cytosol as a destination for recombinant proteins, including (i) the negative redox potential of the cytosolic milieu, unfavorable to the proper folding of proteins that bear disulfide bonds

(Goulet and Michaud, 2006); (ii) the absence of co- and post-translational modification processes such as glycosylation or acylation, that may have a positive impact on the folding, assembly, anchoring and/or structural stability of several nascent and mature proteins (Faye et al., 2005); and (iii) the effective housekeeping activity of the ubiquitin-proteasome proteolytic pathway in this cellular compartment (Vierstra, 1996; 2003), involved notably in the recognition and degradation of incorrectly folded proteins. Although some recombinant proteins remain stable in the cytosol (e.g. Michaud et al., 1998; De Jaeger et al., 1999; Rajabi-Memari et al., 2006; Marusic et al., 2007), alternative destinations such as the chloroplast, the ER or the apoplast appear more appropriate for most proteins.

*Retention into the ER.* Targeting to the cell secretory pathway, in particular, has been proposed for improving the stability and yield of several proteins (Faye and Chrispeels, 1990; Ma et al., 2003; Yoshida et al., 2004; Vitale and Pedrazzini, 2005). In practice, the addition of an N-terminal signal peptide sequence to the protein-encoding transgene triggers a co-translational transfer of the nascent protein to the endomembrane system (Faye and Chrispeels, 1990; **Figure I.1**). Proteins bearing a signal peptide for cellular secretion first enter the ER via the ER protein translocation channel (Galili et al., 1998), and then migrate through this compartment and the Golgi apparatus until reaching the extracellular medium [default pathway] or the vacuole, if a vacuolar sorting signal is found in the protein primary sequence. Recombinant proteins entering the ER may also be retained in this compartment by simple apposition of the tetrapeptidic ER-retention signal (K/H)DEL (Michaud et al., 1998) or grafting of a g-zein proline-rich domain (Mainieri *et al.*, 2004) at the C terminus. The (K/H)DEL motif is a common ER retrieval signal in eucaryotes, believed to redirect tagged proteins to the ER after their recognition by a (K/H)DEL receptor complex in the Golgi apparatus (Lee et al., 1993; Pagny et al., 2000). In contrast with most reticuloplasmins, cereal prolamins accumulate in the ER of seed cells without the involvement of the (K/H)DEL motif, forming high-density protein bodies by a sequence-specific sorting mechanism still to be elucidated (Vitale and Pedrazzini, 2005). In practice, retention of zein (and other prolamins) into the ER and the concomitant formation of protein bodies may be easily reproduced in both storage and vegetative tissues of transgenic plants (Geli et al., 1994; Coleman et al., 1996), which makes this system potentially useful for the high-level accumulation of recombinant proteins *in planta*



and their easy recovery in crude extract preparations by gradient centrifugation procedures (Torrent et al., 2006a, 2006b).

Numerous studies have been published illustrating the positive impact of retaining clinically- or industrially-useful proteins in the ER compartment of plant cells, using in most cases a (K/H)DEL retention signal (Ma et al., 2003). The ER, that constitutes a natural reservoir for some storage proteins in seed cells (Shewry and Halford, 2002), can physically accommodate high levels of recombinant protein product *in planta* (Wandelt et al., 1992). At the biochemical level, the low abundance of proteolytic enzymes and the presence of molecular chaperones into the ER, along with an oxidizing status favoring disulfide bond formation, make this organelle a suitable destination for several proteins susceptible to rapid turnover or showing a complex folding pathway (Nuttall et al., 2002; Faye et al., 2005). Several reports documented the positive impact of ER retention on the production of recombinant antibodies in terms of protein stability, quality or yield (Schouten et al., 1996; Conrad and Fiedler; 1998; Stoger et al., 2002; Gomord et al., 2004). Similar tendencies were observed for several other proteins of medical or industrial interest, including as recent examples human interleukin-4 (Ma et al., 2005c), the SARS coronavirus S protein antigen (Pogrebnyak et al., 2005), the synthetic silk-like protein DR1B (Yang et al., 2005), and a recombinant phytase from *Aspergillus niger* considered as an alternative to inorganic sources of phosphorus in animal regimes (Peng et al., 2006).

Despite these promising developments, the ER cannot be considered a suitable destination for all proteins. To be stable or active, a number of clinically-useful proteins require late post-translational modifications –such as the formation of complex glycans, the addition of a lipid moiety or the proteolytic removal of a propeptide sequence– that may occur downstream of the ER along the secretory pathway, notably in the Golgi, the vacuole or the apoplast (Gomord and Faye, 2004; Faye et al., 2005). Other proteins may exhibit an altered integrity or structure heterogeneity into the ER, as a result of unintended proteolytic processing by ER-resident proteases (Faye et al., 2005). The ER lumen is generally considered a mild environment for labile –e.g. immature– proteins, but proteolytic processing events altering the structure of secreted proteins have been observed in this compartment (Bass et al., 2000; Schmitz and Herzog, 2004). For instance, the bovine plasma protein aprotinin expressed in leaves of transgenic potato lines showed structure heterogeneity when accumulated into the ER, presumably due to the sequential removal of

specific amino acids at the N and C termini by endogenous peptidases (Badri et al., 2007a). A solution to elude unintended protein processing into the ER –or later on along the secretory pathway (Sharp and Doran, 2001)– would be to identify and mutate susceptible amino acid sites by site-directed mutagenesis, as suggested by Outchkourov *et al.* (2003) for the sea anemone protein equistatin, processed to several truncated forms in the secretory pathway of transgenic potato leaf cells by Arg/Lys- and Asn-specific cysteine proteases.

*Targeting to the apoplast.* Simply targeting the protein to an alternative compartment –even downstream in the cell secretory pathway– could also help to prevent unintended processing. For instance, recombinant bovine aprotinin targeted to the apoplast of potato leaf cells could be isolated under a homogenous, unprocessed form similar in size to the native protein purified from bovine pancreas, in contrast with the proteolysis-related microheterogeneity mentioned above for the same protein retained into the ER (Badri et al., 2007a). The extracellular medium typically exhibits a high proteolytic content (Callis, 1995; Hellwig et al., 2004; Schiermeyer et al., 2005), but several recombinant proteins tagged for extracellular secretion were successfully produced in plant or plant cell platforms over the last few years (Gaume et al., 2003; Streatfield et al., 2003; Sojikul et al., 2003; Komarnytsky et al., 2004; Wirth et al., 2004; Gils et al., 2005; Hellwig et al., 2004; Yang et al., 2005). Proteolytic processing in the extracellular medium may even represent an advantage for the correct maturation of certain proteins, as illustrated recently with the dust mite allergenic protein Der p 1, processed to its mature, active form in the culture medium of transgenic BY-2 tobacco cells naturally secreting proteases specifically cleaving the propeptide sequence of this protein (Lienard et al., 2007).

*Targeting to the vacuole.* Similar to the apoplastic medium, the vacuole represents a suitable accumulation site for several recombinant proteins, especially for production platforms relying on seed tissues (Stoger et al., 2005). The vacuole plays several important roles *in planta*, including the control of cell turgor, the turnover of macromolecules, the sequestration of toxic secondary metabolites, and the storage of high-energy compounds in seeds or vegetative storage tissues (Marty, 1999). Recent evidence in the literature suggests the occurrence of two distinct types of vacuoles in plants: the lytic (or vegetative) vacuoles, that present an acidic environment rich in

hydrolytic enzymes; and the protein storage vacuoles, that shows a slightly acidic or neutral pH well adapted to protein storage (Robinson et al., 2005). Targeting to the vacuole, although not yet fully understood, is determined by small stretches of amino acids within the protein primary sequence acting as sorting signals to direct the maturing protein toward the vacuole (Neuhaus and Rogers, 1998; Mackenzie, 2005; Vitale and Hinz, 2005). In general, lytic vacuoles are not considered a suitable destination for recombinant proteins *in planta*, owing to their high proteolytic content (Michaud and Goulet, 2006). By contrast, protein storage vacuoles present a milder environment compatible with protein accumulation (Stoger et al., 2005), especially in seeds where they are the most abundant (Müntz, 1998; Park et al., 2004).

Good accumulation levels were reported for a number of recombinant proteins targeted to the vacuole (**Table I.2**), including for instance the synthetic analogue of spider dragline silk protein, DP1B (Yang et al., 2005), the heat labile enterotoxin B from *E. coli* (Streatfield et al., 2003), the toxic biotin-binding proteins avidin and streptavidin (Murray et al., 2002), and a protein-engineered thermostable  $\beta$ -glucanase of bacterial origin (Horvath et al., 2000). In practice, a proper *in situ* localization of recombinant proteins bearing a sorting sequence for vacuolar targeting should be undergone on a systematic basis, considering the species- or tissue-dependent functionality of some sorting signals (Vitale and Hinz, 2005). A good example of this phenomenon was recently provided for a fungal phytase expressed in rice, that was readily detected in the apoplastic environment of leaf tissues but retained in ER protein bodies and protein storage vacuoles in the seed endosperm (Drakakaki et al., 2006). As illustrated with the silk-like protein DP1B expressed in *Arabidopsis* (Yang et al., 2005), the impact of vacuolar targeting on the stability and yield of recombinant proteins is also tissue-dependent. Whereas this protein was found at levels reaching 8% TSP in seed storage vacuoles, no detectable levels of the same protein could be observed in leaf cell vacuoles. In a similar way, targeting DP1B to the apoplast provided good yields in leaves but poor yields in seeds (Yang et al., 2005), again stressing the need for an empirical, case-by-case assessment with different tissue and cellular destinations for each protein to express.

*Targeting to the chloroplast.* The chloroplast, the peroxisome and the nucleus have been proposed as other cellular destinations for protein production in plant platforms (Daniell et al., 2002, Hyunjong et al., 2006). In practice, recombinant proteins may be sent

to these organelles by inclusion of an appropriate targeting peptide [or localization signal] in the transgene sequence (**Figure I.1**). For instance, the heat-labile toxin Lt-B from enterotoxigenic *E. coli* exhibited increased levels in corn grains when diverted from the cytosol to the nucleus by the addition of a nuclear localization signal from the simian virus 40 large T-antigen in C-terminal position (Streatfield et al., 2003). In a similar way, a fungal xylanase useful in environment-related technologies showed high accumulation levels in *Arabidopsis* leaf tissues when sent to chloroplasts using the rubisco activase transit peptide, or sent to peroxisomes using the tripeptidic targeting motif SKL (Ser-Lys-Leu) grafted at the C terminus (Hyunjong et al., 2006). Another recent example was the improved production of human growth hormone in *Nicotiana benthamiana* leaf cells when transferred from the cytosol to the chloroplast using the transit peptide of rubisco small subunit (Gils et al., 2005).

Efficient procedures have also been devised to insert transgenes in the chloroplastic genome, and then regenerate transplastomic plant lines accumulating high levels of recombinant protein directly into the chloroplast stroma (Daniell et al., 2001; Maliga, 2002). Chloroplast transformation offers several advantages over nuclear transformation, including uniform transgene expression rates, multiple functional copies of the transgene in each cell, co-expression of multiple genes from a same construct, minimal gene silencing, and minimal transgene escape in the environment owing to the maternal inheritance of chloroplastic DNA in several species (Daniell et al., 2002). The chloroplast stroma supports protein post-translational modifications like multimerization and disulfide bridge formation (Daniell, 2006), making it a suitable environment for the expression of proteins not relying on complex modifications –such as glycosylation– typical of the cell secretory pathway. Several transplastomic plant lines have been engineered over the last ten years for recombinant protein expression, providing very high yields for a number of useful proteins of procaryotic or eucaryotic origin including somatotropin, serum albumin, anthrax protective antigen, cholera toxin B subunit and tetanus toxin fragment C (Daniell et al., 2001, 2005; Tregoning et al., 2003).

On the other hand, a number of endogenous proteases are present in the chloroplast (Adam and Clarke, 2002), that can impair the overall stability and accumulation of recombinant proteins. An interesting example was provided for the rotavirus VP6 protein, that showed high accumulation rates in chloroplasts of young tobacco leaves but

negligible rates in older leaves despite comparable levels of mRNA transcript (Birch-Machin et al., 2004). A similar decline in older tissues was observed for a fungal xylanase (Hyunjong et al., 2006) and for the insecticidal Bt toxin Cry2Aa2 (De Cosa et al., 2001), again stressing the importance to carefully assess the stability of each recombinant protein in different physiological or environmental contexts. Recombinant protein degradation by chloroplastic proteases might appear, however, a non-relevant issue in terms of net production yields for some proteins expressed at very high levels (Daniell, 2006). The proteolysis-labile protein human serum albumin, for instance, was found at levels reaching 11% TSP in transplastomic tobacco lines developed using chloroplastic untranslated regions in gene constructs, in sharp contrast with levels lower than 0.02% TSP in lines developed using the commonly used Shine-Delgarno regulatory sequence (Fernandez-San Millan et al., 2003). This dramatic increase, likely due to an increased expression rate of the transgene, was also associated with the formation of large inclusion bodies *in vivo* (Fernandez-San Millan et al., 2003) that presumably sequestered the recombinant protein and prevented its hydrolysis *in planta*, as described earlier for a number of proteins expressed in heterologous environments (Enfors, 1992).

### **1.1.3.3 Stabilizing fusion partners**

Along with strategies based on tissue-specific transgene expression or organelle-specific protein targeting, protein engineering approaches involving site-directed mutagenesis and/or protein domain grafting are currently considered to prevent recombinant protein hydrolysis *in planta*. Site-directed mutagenesis might prove useful, notably, to improve the intrinsic stability of labile proteins in plant cells, as proposed earlier for recombinant proteins expressed microbial systems (Rozkov and Enfors, 2004; Macauley-Patrick et al., 2005). Any recombinant protein, depending on its structural characteristics, potentially includes protease-susceptible amino acid sites sterically accessible to endogenous proteases in the transgenic host organism. In theory, the identification of such susceptible sites in the polypeptide chain, along with a correct knowledge of the host proteolytic machinery, can provide useful information for designing protein variants with improved stability *in vivo* (Rozkov et al., 2000). To our knowledge, this approach has not yet been applied to recombinant proteins expressed in plants, but the characterization of specific proteolytic events implicating recombinant proteins in transgenic plants (e.g. Outchkourov et al., 2003; Badri et al., 2007a) should provide, in the

future, a workable framework for the rational engineering of protein variants with improved resistance to plant proteases.

When structural constraints in the protein do not allow for an alteration of protease-susceptible sites, or when a restrictive regulatory framework for therapeutic proteins dictates maximal identity between the recombinant protein and its original counterpart, expressing the protein attached to a protein-stabilizing fusion partner could prove useful to enhance its stability and/or folding while keeping intact its original primary structure. Early studies with staphylococcal protein A and streptococcal albumin-binding protein demonstrated the potential of fusion partners as promoting agents for recombinant protein accumulation *in vivo* (Nygren et al., 1994, Stahl et al., 1997). In plants, several studies have been published describing the positive impact of protein fusion partners on the accumulation of recombinant proteins. Examples were provided for the marker proteins  $\beta$ -glucuronidase and luciferase and, more recently, for the cholera toxin B subunit, that were accumulated at higher rates in transgenic potato or tobacco leaves when the regulatory protein ubiquitin was fused and transiently present at the N terminus (Garbino et al., 1995; Hondred et al., 1999; Mishra et al., 2006). More recently, the HIV p24 core antigen expressed in tobacco leaves showed an accumulation level increased by more than 10-fold when heavy chains of a human immunoglobulin A were fused at the C terminus (Obregon et al., 2006). In a similar way, the accumulation rates of spider silk proteins, single-chain antibody fragments, gp130 polypeptides or mammalian interleukins in transgenic tobacco lines were significantly improved when elastin-like polypeptides including repeats of the elastin motif VPGXG (Val-Pro-Gly-Xaa-Gly) were fused at the C terminus (Scheller et al., 2004; 2006; Lin et al., 2006; Patel et al., 2007). Whereas work is still required to devise efficient procedures for the proteolytic removal of the fusion partners, and to understand the mechanisms by which the different fusion partners exert their stabilizing effect, these observations clearly suggest the potential of fusion protein technologies for the improvement of recombinant protein yields in plants.

#### **1.1.3.4 Companion protease inhibitors**

Another strategy proposed recently to increase recombinant protein production *in planta* consists to use transgenic hosts with reduced proteolytic capacities. In theory, protease processes affecting recombinant protein accumulation could be contained using antisense or RNA silencing strategies implemented in transgenic host plants (Watson et al., 2005). Alternatively, recombinant protease inhibitors active against specific endogenous proteases could prove functional to modulate proteolytic activities *in situ* (Faye et al., 2005). Recent evidence in the literature suggests that the ectopic expression of protease inhibitors may have a positive impact on protein levels in leaves, with negligible effects on growth and development. The broad-spectrum inhibitor tomato cathepsin D inhibitor, for instance, was shown to increase total soluble protein levels by 20-35% in leaves of transgenic potato lines accumulating this inhibitor in the cytosolic compartment (Michaud et al., 2005). Likewise, the rice cysteine protease inhibitor oryzacystatin led to total protein levels higher than expected in tobacco leaf tissues expressing this inhibitor in the cytosol (van der Vyver et al., 2003). The design of transgenic plant lines deficient in specific protease activities could now represent a challenging, but worthwhile task, that could provide mild cellular environments tailored for the expression and accumulation of fragile proteins. A convincing demonstration supporting the potential of this approach was provided recently by Komarnytsky *et al.* (2006), who successfully stabilized recombinant antibodies secreted by the roots of transgenic tobacco plants also expressing [and co-secreting] a Bowman-Birk trypsin inhibitor from soybean.

#### **1.1.4 Stabilizing recombinant proteins *ex planta***

In parallel to strategies aimed at preserving the integrity of recombinant proteins *in planta*, basic strategies have been devised over the last ten years to protect recombinant proteins *ex planta*, at the time of extraction from plant tissues. Whereas the avoidance of unwanted proteolysis *in planta* is essential to provide sufficient protein levels before harvesting, proteolysis control *ex planta* is essential to prevent protein loss in crude preparations. Protein extraction is often considered as a simple preliminary step for subsequent purification procedures, but it has in fact a strong influence on the overall stability and quality of extracted proteins, and hence on the nature and abundance of



proteins then available for subsequent analysis, enrichment and purification (Michaud and Asselin, 1995). Simple basic precautions may be taken in the lab to prevent excessive proteolysis and preserve recombinant protein integrity in plant samples and protein extracts (Michaud, 1998). Molecular and cellular approaches have been also proposed in recent years to minimize degradation *ex planta*, involving expression of protein-stabilizing recombinant protease inhibitors or non-destructive secretion of the recombinant protein product.

#### **1.1.4.1 Basic precautions during protein extraction**

From a practical viewpoint, plant cells are often seen as a 'recalcitrant' source of material for protein extraction, owing to their particular architecture and content (Saravanan and Rose, 2004; Carpentier et al., 2005). Compared to animal or microbial cells, plant cells from metabolically active tissues typically show a low protein content, and the vacuole(s), which account(s) for most of the cell volume, contain(s) highly reactive compounds potentially deleterious to proteins (Michaud and Asselin, 1995). During extraction, at least some protein loss is usually to be expected given the occurrence of poorly soluble material interacting non-specifically with proteins in crude preparations, and the systematic release of protein denaturing agents including phenolics and proteases after cell disruption (Menkhaus et al., 2004; Saravanan and Rose, 2004). Nevertheless, and although the onset of protein:protease interactions specific to each protein and expression platform will generally dictate the empirical development of adapted protection strategies during protein recovery, general guidelines have been proposed over the years to prevent protein loss in plant biomass and protein extracts (Jervis and Pierpoint, 1989; Michaud and Asselin, 1995; Michaud, 1998; Menkhaus et al., 2004; Schillberg et al., 2005). In brief, these guidelines recommend: (i) to process biomass shortly after harvest for plant material retrieved from metabolically active tissues; and (ii) to adopt laboratory practices preventing or at least minimizing protein loss and denaturation in crude extracts, including the processing of sample tissues at low temperatures, the addition of chelators and phenolics-neutralizing compounds in the extraction buffer, the use of pH conditions unfavorable to endogenous protease activities, and the use of low-molecular-weight protease inhibitors active against the host plant endogenous proteases (Michaud and Asselin, 1995).



#### **I.1.4.2 Protein-stabilizing protease inhibitors**

At the laboratory scale, protease inhibitors with complementary inhibitory functions may prove of particular interest to rapidly assess the inherent (in)stability of recombinant proteins challenged with the host plant's resident proteases, or to retrieve the protein product under a form suitable for further characterization, monitoring and purification (Michaud, 1998; Benchabane et al., 2007). Protease inhibitor "cocktails" composed of inhibitors with different specificities against plant proteases have been devised by several suppliers and are now commonly used as additives to extraction media for the stabilization of proteins in crude preparations. Various diagnostic inhibitors (**Table I.3**) and synthetic peptidic substrates for protease assays (Brunelle and Michaud, 2000) are also available as diagnostic tools to characterize proteases and protein hydrolytic processes in crude extracts, and to devise strategies for the protection of proteins to be expressed in heterologous environments (Benchabane et al., 2007). Whereas the testing of different inhibitors always remains suggested given the general diversity of protease activities in plant crude extracts, small broad-spectrum inhibitors such as chymostatin and phenylmethylsulfonyl fluoride often constitute a good starting point for the stabilization of heterologous proteins *in vitro*, especially in crude extracts prepared from metabolically active tissues such as stems or green leaves (**Figure I.2**).

For large-scale set-ups, protein protease inhibitors co-expressed *in planta* along with the recombinant protein of interest may represent an effective alternative to commercial low-molecular-weight inhibitors (Rivard et al., 2006). Protein protection with exogenous protease inhibitors, which relies on expensive and often toxic compounds, appears useful for diagnostic purposes or small-scale production scenarios, but is hardly applicable to large-scale production schemes involving the extraction and downstream processing of large biomass samples. By contrast, in-built companion inhibitors co-expressed in the host plant are 'included' *de facto* in protein extracts after tissue grounding, then acting as "mouse traps" against the proteases present to reproduce the protein-stabilizing effects of exogenous chemical inhibitors (Benchabane et al., 2007). Broad-spectrum protease inhibitors such as the serine-type inhibitors tomato CDI and bovine aprotinin appear well suited as such effective companion inhibitors, as illustrated recently by the strong stabilization of the marker protein neomycin phosphotransferase II released in leaf crude extracts of tomato CDI- and bovine aprotinin-expressing potato

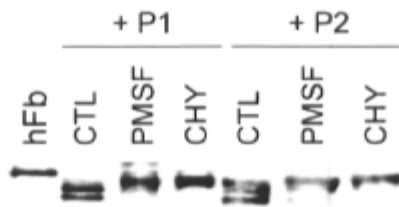
plants (Rivard et al., 2006). Inhibitors of cysteine proteases such as cystatins, or hybrid fusion inhibitors integrating dual [or multiple] inhibitory functions active against cysteine, serine and/or aspartate proteases (e.g. Urwin et al., 1998; Brunelle et al., 2005) also represent interesting candidates, given the general occurrence of serine, cysteine and aspartate protease activities sensitive to such inhibitors in plant crude extracts (Rivard et al., 2006).

**Table I.3 Low-molecular-weight protease inhibitors commonly used for protease characterization and protein stabilization in plant extracts \***

Inhibitor	Mol. Weight (Da)	Target proteases
AEBSF **	240	Serine
Amastatin	475	Metallo- (aminopeptidases)
Antipain	605	Serine (trypsin-like), Cysteine
APMSF	216	Serine (trypsin-like)
Bestatin	308	Metallo- (aminopeptidases)
Chymostatin	605	Serine (chymotrypsin-like), Cysteine
3,4-DCI	215	Serine
Diprotin A	359	Metallo- (aminopeptidases)
Diprotin B	327	Metallo- (aminopeptidases)
E-64	357	Cysteine
EDTA	372	Metallo- (chelating agent)
Elastatinal	513	Serine (elastase-like)
Iodoacetamide	185	Cysteine (may interfere with other enzymes)
Leupeptin	427	Serine (trypsin-like), Cysteine
Pepstatin	686	Aspartate
1,10-Phenanthroline	198	Metallo- (chelating agent)
PMSF	174	Serine, Cysteine in non-reducing conditions
TLCK	333	Serine (trypsin-like)
TPCK	352	Serine (chymotrypsin-like)
Z-Phe-Ala-CHN <sub>2</sub>	395	Cysteine

\* See Michaud (1998) for details on stock solutions and working concentrations.

\*\* Abbreviations: AEBSF, 4-(2-aminoethyl)-benzenesulfonyl fluoride, hydrochloride; 3,4-DCI, 3,4-dichloro-isocoumarin; E-64, L-trans-epoxysuccinyl-leucylamido-(4-guanidino) butane; EDTA, ethylenediamine-tetraacetic acid; PMSF, phenylmethylsulfonyl fluoride; TLCK, tosyl lysyl chloromethyl ketone; TPCK, tosyl phenylalanyl chloromethyl ketone.



**Figure 1.2 Stabilization of exogenous proteins in crude plant extracts by phenylmethylsulfonyl fluoride (PMSF) and chymostatin – The example of human fibronectin (hFb) challenged with alfalfa leaf proteases.** hFb was first incubated for 10 min at 20°C with major protease fractions of alfalfa leaf proteases (P1 or P2) in the presence or absence of 0.1 mM PMSF or 10 µM chymostatin (CHY), and then immunodetected using appropriate anti-hFb polyclonal antibodies. CTL, control extract, mixture, with no inhibitor added.

#### 1.1.4.3 Non-destructive secretion and recovery

An alternative to protease inhibitors for the avoidance of extraction-related protein degradation consists to skip this step by taking advantage of cellular and tissue secretion processes occurring naturally in the host plant. In brief, this strategy consists to direct recombinant proteins in the extracellular medium using appropriate signal peptides for extracellular secretion, and then to recover extracellular proteins in secreted fluids or culture media. Rhizosecretion and leaf guttation represent two such processes allowing for the non-destructive secretion and extraction-free recovery of recombinant proteins (Komarnytsky et al., 2000; Gaume et al., 2003). Rhizosecretion, notably, has been exploited recently for the heterologous expression of several recombinant proteins, including human alkaline phosphatase (Gaume et al., 2003), hepatitis B surface antigen (Kumar et al., 2006) and immunoglobulin G antibodies (Komarnytsky et al., 2006). This approach is based on the capacity of hairy roots to secrete proteins in hydroponic media, that can be easily recovered and used as simple source material for protein enrichment and purification (Komarnytsky et al., 2004). Interestingly, rhizosecretion –and hence recombinant protein release– can be significantly boosted by the initiation of hairy roots following treatment with *Agrobacterium rhizogenes* (Gaume et al., 2003), thus providing an efficient source of recombinant protein throughout the host plant's life.

Plant cell cultures have also been proposed for the continuous secretion of recombinant proteins in culture media (Hellwig et al., 2004). Compared to transgenic plant lines, cultured cells offer a high level of containment, and represent a useful tool for designing simple recovery and purification strategies from source media containing low levels of contaminants. On the other hand, some recombinant proteins may undergo extensive hydrolysis in culture media, owing to the secretion of proteolytic enzymes by growing cells (Lee et al., 2002; Kwon et al., 2003; Schiermeyer et al., 2005). The extracellular concentration of human granulocyte-macrophage colony-stimulating factor, for instance, reached a maximum in tobacco cell cultures at the mid-exponential phase, but then decreased rapidly in the culture medium, along with a dramatic increase of protease activity in the culture medium (Lee et al., 2002). In a similar way, accumulation of the fibrinolytic protease  $\alpha$ 1 plasminogen activator from *Desmodium rotundus* was dramatically hampered in the culture medium of BY-2 tobacco cells, likely as a result of extensive degradation by secreted metallo-proteases (Schiermeyer et al., 2005). The addition of protein-stabilizing agents such as low-molecular-weight protease inhibitors or different polymers (e.g. gelatin, albumin, polyethylene glycol or polyvinylpyrrolidone) in the culture medium has been proposed to prevent protein loss in plant cell cultures (Bateman et al., 1997; LaCount et al., 1997; Tsoi and Doran 2002; Schiermeyer et al., 2005; Soderquist and Lee, 2005). The co-secretion of recombinant protease inhibitors hindering endogenous protease activities along the cell secretory pathway (Komarnytsky et al., 2006), or the development of protease-deficient transgenic plant cells lacking specific, potentially damaging secreted proteases (Schiermeyer et al., 2005) could represent interesting alternatives in the future allowing recombinant proteins secreted in culture media [or natural secretion fluids] to be stabilized *in vivo*, without the need to using exogenous protein-stabilizing chemicals.

### **1.1.5 Conclusion**

In summary, major advances have been made over the last decade to improve the post-translational stability and accumulation of recombinant proteins in plant systems. The use of tissue-specific promoters or the addition of subcellular targeting peptidic signals to the nascent protein, in particular, have proved useful to increase the stability of several proteins *in planta*. Complementary approaches relying on protein-stabilizing fusion partners, companion protease inhibitors or natural secretion processes have also been

proposed recently to protect recombinant proteins, either *in planta* during the production stage and *ex planta* at the time of extraction (or secretion). Whereas a case-by-case, exploratory assessment involving the testing of different strategies could always remain essential given the very own structural and functional characteristics of each protein to express or plant platform to be exploited, the fine-tuning of molecular tools and strategies for recombinant protein stabilization should be made easier in the near future, along with our growing understanding of proteolytic and protein maturation processes in plants. Rapid developments in structural biology and protein engineering should also contribute to the biotechnologist's toolbox in forthcoming years, and help us facing the challenge of recalcitrant protein expression.

## **1.2. La glycosylation des protéines recombinantes**

Les propriétés physico-chimiques des protéines, leur activité biologique et leurs propriétés pharmacologiques sont étroitement liées à leurs modifications post-traductionnelles (Gomord et Faye, 2004; Walsh et Jefferis, 2006). Au-delà de sa simple séquence primaire en acides aminés, une protéine doit ses caractéristiques à diverses transformations et ajouts, incluant son repliement, sa maturation protéolytique et l'ajout de lipides et de glycannes. Comme les modifications post-traductionnelles sont surtout l'apanage des cellules eucaryotes, de nombreuses études ont été publiées récemment sur le développement de systèmes d'expression hétérologue tels les cellules de mammifères, de plantes ou d'insectes, les plantes et les animaux transgéniques, comme alternatives aux systèmes bactériens couramment exploités. Ces nouveaux systèmes de production sont plus aptes à répondre à la demande en protéines recombinantes, qui croît non seulement en quantité mais aussi en complexité (Dove, 2002; Ko et Koprowski, 2005). Des nombreuses modifications post-traductionnelles qui contribuent à la spécificité des protéines, la glycosylation est celle qui semble susciter le plus d'intérêt (Gomord et Faye, 2004). Il est estimé, à cet effet, que plus de 50% des protéines eucaryotiques sont des glycoprotéines (Apweiler et al., 1999). L'abondante littérature scientifique concernant la glycosylation des protéines recombinantes reflète à la fois l'intérêt, la complexité et les enjeux économiques et thérapeutiques de ce processus, en particulier pour les systèmes d'expression hétérologue végétaux. Nous nous intéresserons, dans les sections qui suivent, tout d'abord à la biosynthèse des O- et des N-glycannes chez les végétaux, puis aux rôles de la glycosylation et à la production de glycoprotéines dans les cellules végétales.

### **1.2.1 Les voies de biosynthèse des glycannes**

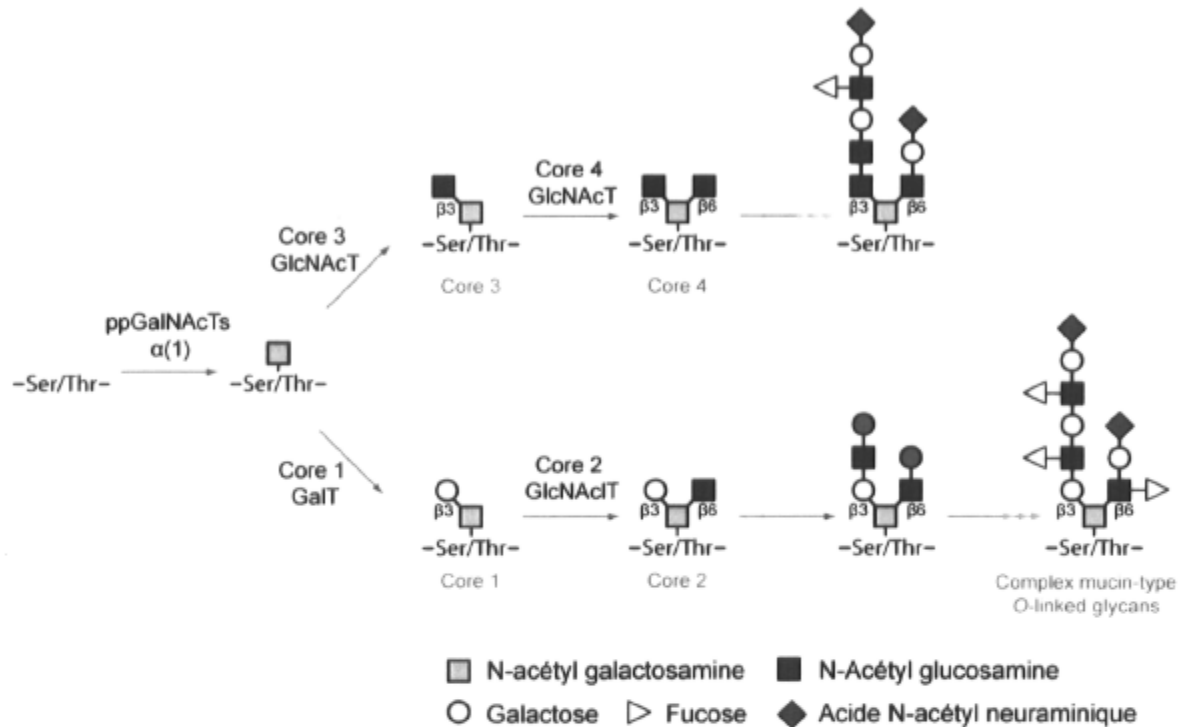
La glycosylation est l'ajout, par une liaison covalente, d'un oligosaccharide (ou glycanne) sur la structure peptidique codée par le génome. L'oligosaccharide peut être lié à l'extrémité hydroxyle d'une sérine, d'une thréonine, d'une tyrosine, d'une hydroxyproline ou d'une hydroxylysine; le processus est alors appelé O-glycosylation. L'oligosaccharide peut aussi être ajouté à l'azote amide d'une asparagine; il s'agit alors du processus de N-glycosylation, de loin le plus caractérisé chez les végétaux. Après un bref aperçu de la O-glycosylation, cette introduction traite ensuite principalement de la N-glycosylation.

### 1.2.1.1 La O-glycosylation des protéines

La O-glycosylation des protéines est l'addition d'un glycanne à l'extrémité hydroxyle d'une sérine (Ser), d'une thréonine (Thr), ou même d'une tyrosine (Tyr), de l'hydroxyproline (Hyp) ou de l'hydroxylysine (Hyl). Chez les mammifères, la O-glycosylation débute dans l'appareil de Golgi par le transfert d'un monosaccharide, plus couramment une N-acétyl galactosamine (GalNAc), sur l'oxygène des groupements hydroxyles, principalement des résidus Ser et Thr (Varki et al., 1999; Hang et Bertozzi, 2005). Il n'existe pas de sites consensus pour la O-glycosylation mais les résidus Ser/Thr sont plus souvent situés dans des séquences riches en résidus prolines. Par la suite, la GalNAc peut-être substituée par des résidus de N-acétyl glucosamine (GlcNAc), de galactose (Gal), de fucose (Fuc) ou d'acide N-acétylneuraminique (Neu5Ac) pour générer une grande diversité de structures de O-glycannes (voir **Figure 1.3**). D'autres O-glycosylations sont possibles, avec l'ajout de mannose (Man), de Fuc, de Gal, de glucose (Glc) ou de GlcNAc directement sur la sérine ou la thréonine. L'ajout de GlcNAc sur les résidus Ser et Thr se distingue des autres types de glycosylation: elle est associée à des protéines nucléaires et cytoplasmiques et peut, dans certains cas, entrer en compétition avec la phosphorylation de ces protéines (Hart et al., 2007).

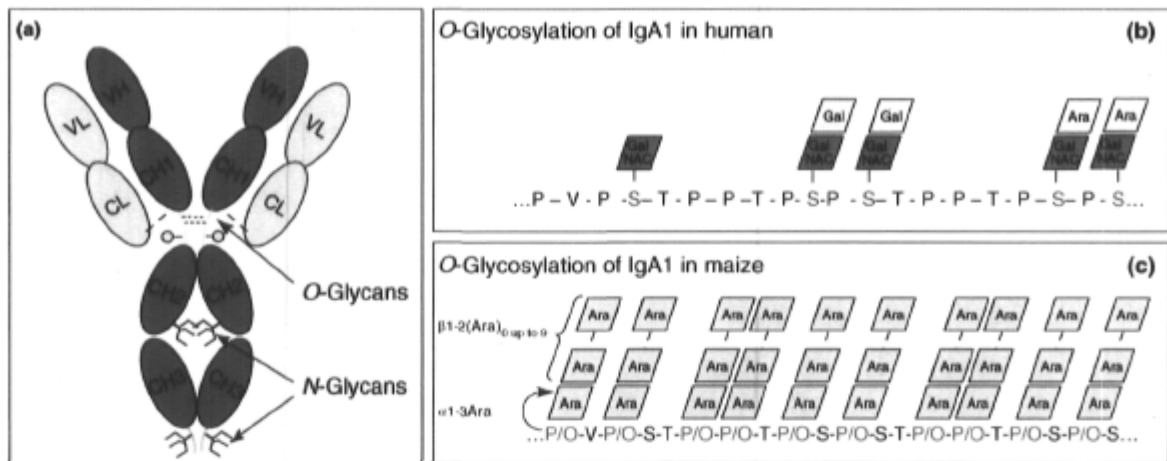
Chez les plantes, la O-glycosylation a surtout été décrite pour les résidus Hyp, Ser et Thr (Gomord et Faye, 2004). Les glycoprotéines riches en hydroxyproline (*Hyp-rich glycoproteins*, ou HRGP) constituants des parois cellulaires et de l'extérieur de la membrane plasmique, en particulier, sont riches en O-glycannes liés aux résidus Hyp et sont constitués d'arabinoses et d'arabinogalactanes (Showalter, 2001). Plus récemment, une étude de Karnoup et al. (2005) a montré qu'une immunoglobuline de type A (IgA) humaine recombinante exprimée chez le maïs transgénique était N et O-glycosylée, tout comme l'IgA endogène chez l'homme (voir **Figure 1.4**). La région charnière de la chaîne longue constituée d'une séquence riche en prolines (Pro) comporte, chez l'homme, des O-glycannes de type mucine sur les résidus Ser/Thr. Après expression chez la plante, ce sont des prolines hydroxylées (Hyp) de cette même région qui sont O-glycosylées, cette fois par des résidus arabinoses. La glycosylation de la région charnière des chaînes longues est supposée essentielle à la stabilité et à une éventuelle protection contre une attaque protéolytique. Des recherches restent encore nécessaires pour déterminer si une O-glycosylation végétale sera suffisante pour assurer la stabilité et l'activité de l'anticorps

recombinant ou si elle risque au contraire de nuire à une potentielle application thérapeutique.



**Figure I.3 Biosynthèse de glycoprotéines O-glycosylées dans l'appareil de Golgi chez les mammifères.** La O-glycosylation débute par la synthèse de structures précurseurs (*Core* 1 à 4) qui subissent ensuite des modifications enzymatiques séquentielles spécifiques au tissu, au stade de développement ou à la protéine. Image adaptée de Hang et Bertozzi, 2005.





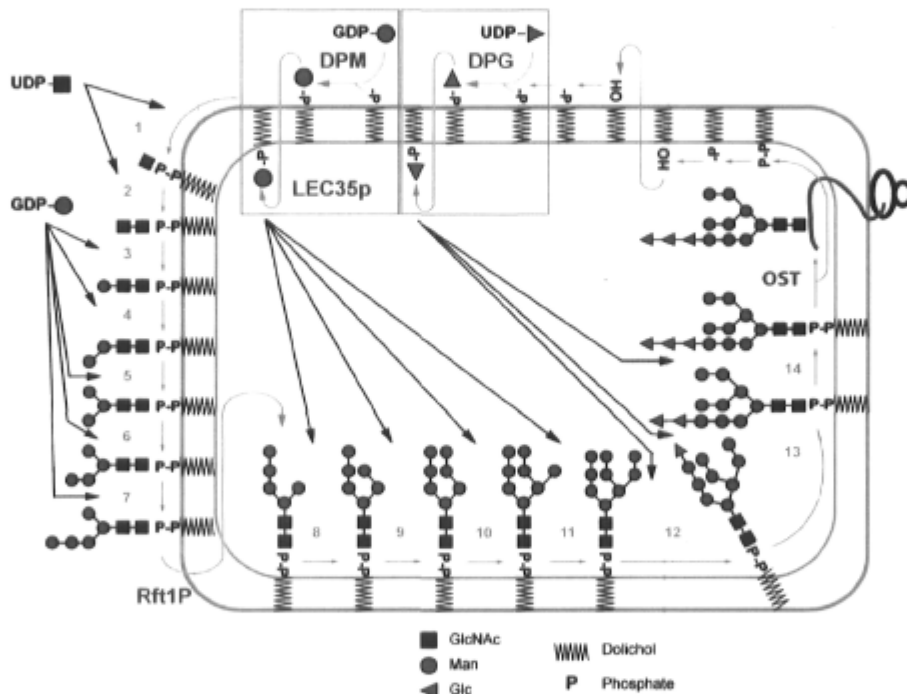
TRENDS in Biotechnology

**Figure 1.4 Une IgA est O-glycosylée sur les mêmes séquences lorsqu'exprimée chez l'humain ou chez le maïs. A.** Représentation schématique d'une IgA; les sites de N- et de O-glycosylation sont indiqués d'une flèche. **B.** Exprimée chez l'humain, la région charnière de la chaîne longue comporte des O-glycannes avec une structure de type mucine sur des résidus sérines. **C.** Chez le maïs, ce sont les hydroxyprolines qui sont O-glycosylées avec des structures typiques des végétaux. Abréviations : Ara, arabinose; Gal, galactose; GalNAc, galactosamine; P/O, hydroxyproline; P, proline; S, sérine; T, thréonine; V, valine. Figure tirée de Saint-Jore Dupas et al., 2007.

### 1.2.1.2 La biosynthèse des N-glycannes chez les plantes

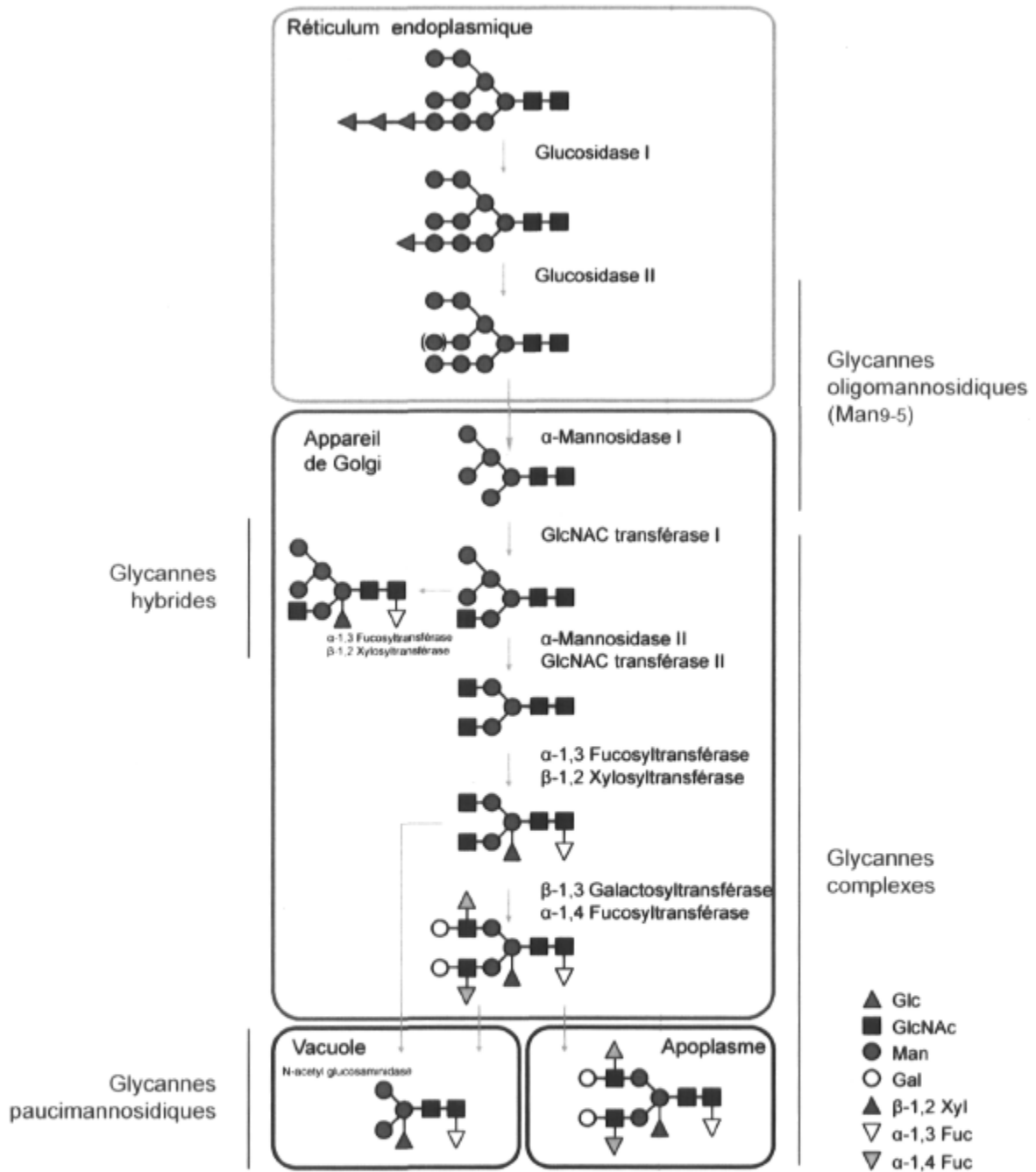
*In vivo*, la glycosylation n'est pas sous le contrôle direct des gènes, mais plutôt le résultat d'un processus enzymatique complexe co- et post-traductionnel. Contrairement aux protéines ou aux acides nucléiques, les glucides forment des polymères qui varient en composition et en structure, notamment dans le nombre et l'agencement des branchements liant leurs différents monosaccharides constituants. Dans la cellule, le processus de N-glycosylation est associé au système endomembranaire de sécrétion. Il débute dans la lumière du réticulum endoplasmique (RE) et se poursuit dans l'appareil de Golgi, puis dans la vacuole ou dans l'apoplasme. Chaque enzyme intervenant au cours de la N-glycosylation possède une localisation précise au sein du système endomembranaire (Saint-Jore-Dupas et al., 2007), de telle sorte que la composition finale du glycanne sera dictée non seulement par la compartimentation et la spécificité de ces enzymes mais aussi par l'adressage final de la protéine.

Le processus de N-glycosylation débute dans le RE avec la synthèse d'un glycanne précurseur, le  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ , attaché à un transporteur lipidique, le dolicholpyrophosphate (ou Dpp), qui permet l'ancrage du glycanne précurseur sur la membrane du RE. Les différents monosaccharides qui composent ce précurseur sont ajoutés sur le Dpp, séquentiellement, par différentes glycosyltransférases de la membrane du RE (Burda et al., 1999; voir **Figure I.5**). Les sept premiers résidus sont synthétisés du coté cytosolique de la membrane du RE, et les sept suivants dans la lumière du RE après translocation dans le RE du  $\text{Dpp-Man}_5\text{GlcNAc}_2$  par une flipase. Ce n'est qu'une fois le dernier résidu de glucose ajouté que cet oligosaccharide précurseur peut être transféré sur une protéine.



**Figure I.5 Synthèse du N-glycanne précurseur sur le transporteur lipidique et son transfert sur une chaîne polypeptidique.** La voie de biosynthèse de l'oligosaccharide précurseur a lieu de part et d'autre de la membrane du RE et implique différentes glycosyltransférases membranaires spécifiques. Du côté cytosolique, une première GlcNAc-1 phosphate est transférée du donneur  $\text{UDP-GlcNAc}$  vers le dolicholpyrophosphate, ou Dpp (1). Une deuxième GlcNAc est ensuite transférée (2), suivie de cinq résidus de mannose (3 à 7). Le  $\text{Man}_5\text{GlcNAc}_2$  transloqué dans la lumière du RE par une flipase (Rft1p) est ensuite allongé par l'ajout de quatre résidus de mannose (8 à 11) et trois résidus de glucose (12 à 14) à partir des donneurs dolichol-P-Man et dolichol-P-Glc. Ces étapes sont conservées chez tous les eucaryotes. Image tirée de Takahashi et Nishikawa, 2002.

Lors de l'insertion co-traductionnelle des protéines en cours d'élongation dans le RE, le  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  est transféré *en bloc* par une oligosaccharide transférase membranaire sur le résidu asparagine de la séquence consensus tripeptidique Asn-X-Ser/Thr, où X représente n'importe quel acide aminé à l'exception de l'acide aspartique ou de la proline. Seuls les sites de glycosylation situés dans une région structurellement exposée du polypeptide, souvent un coude  $\beta$ , peuvent être glycosylés (Faye et al., 1993). Par la suite, le précurseur commun subit des modifications séquentielles plus ou moins élaborées, par l'action de glycosidases et de glycosyltransférases dans le RE puis l'appareil de Golgi, générant des structures variées sur les protéines en cours de maturation. Ainsi, dans le RE, les glucosidases I et II clivent spécifiquement les trois glucoses terminaux du glycanne précurseur pour donner le  $\text{Man}_9\text{GlcNAc}_2$ . Les protéines non résidentes du RE sont ensuite transférées dans l'appareil de Golgi, où se trouvent la majorité des enzymes de maturation des glycanes. À ce niveau, de un à quatre résidus mannoses sont éliminés, conduisant à un N-glycanne de type *oligomannosidique* ( $\text{Man}_5\text{GlcNAc}_2$ ). Des *glycanes complexes* sont générés ensuite par l'élimination spécifique de certains résidus et l'ajout d'autres résidus monosaccharidiques tels la N-acétylglucosamine (GlcNAc), le galactose (Gal), le xylose (Xyl) ou le fucose (Fuc). L'élimination de résidus terminaux des glycanes complexes dans la vacuole par l'action des exoglycosidases génère enfin des glycanes *paucimannosidiques* de structures  $\text{Man}_3\text{XylFucGlcNAc}_2$  et  $\text{Man}_{(1-3)}(\text{Xyl})(\text{Fuc})\text{GlcNAc}_2$ . La **Figure 1.6** schématise la voie de la N-glycosylation des protéines chez les plantes. Selon ce modèle, la structure du N-glycanne porté par une protéine s'avère spécifique aux différents compartiments cellulaires traversés par la protéine, et les glycanes peuvent être utilisés comme marqueurs du transport protéique. Toutefois, des glycanes variés peuvent être identifiés sur les différents sites de glycosylation potentiels d'une même protéine, résultat d'une accessibilité stérique différentielle des glycanes aux différentes enzymes catalysant leur maturation (Faye et al., 1986). Un même site de glycosylation peut, par ailleurs, porter des glycanes différents, en raison d'une catalyse enzymatique ou d'une dégradation différentielle par une exoglycosidase sur le lieu d'accumulation de la protéine. Cette hétérogénéité de la N-glycosylation, observée au sein d'une même cellule, est aussi variable selon les espèces, qui disposent chacune de leurs propres glycanes spécifiques. En pratique, cet état de fait rend la maîtrise du processus de glycosylation d'autant plus difficile et importante, considérant l'impact des glycanes sur les propriétés de la protéine finalement produite.



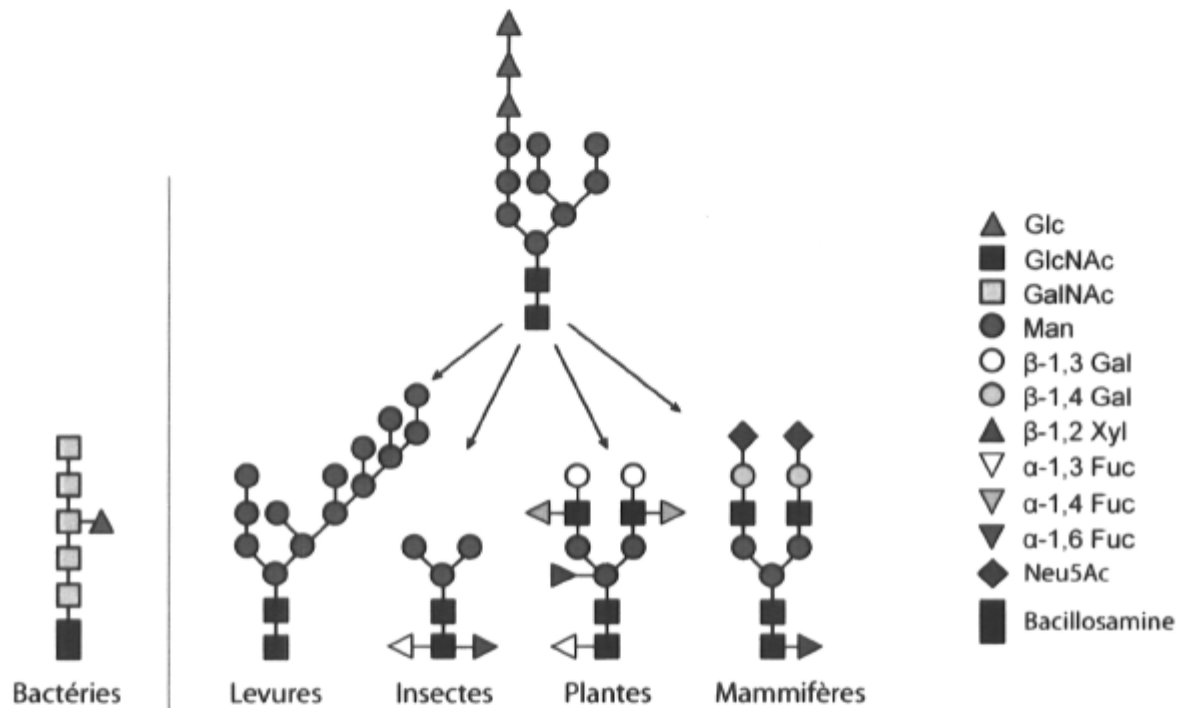
**Figure I.6 Voies de maturation des N-glycannes dans la cellule végétale.** Un glycanne précurseur est ajouté sur la chaîne peptidique naissante dans le RE. Ce glycanne subit ensuite des modifications dans le RE et l'appareil de Golgi, résultant en des glycannes oligomannosidiques, hybrides, complexes et paucimannosidiques. Adapté de Faye et al., 2001.

### 1.2.2 Espèces différentes, N-glycannes différents

Si, sur le plan évolutif, la N-glycosylation est un processus conservé dans sa forme générale chez tous les eucaryotes (Varki et al., 1999), les détails de ce processus diffèrent d'un genre à l'autre, voire d'une espèce à l'autre (**Figure I.7**). La glycosylation a longtemps été considérée comme l'apanage des eucaryotes mais nous savons aujourd'hui que les procaryotes sont, eux aussi, capables de glycosyler leurs protéines (Schaffer et al., 2001; Upreti et al., 2003). La structure de leurs N-glycannes, illustrée à la **Figure I.7**, est par contre plus simple (Young et al., 2002) et leur transfert sur la chaîne polypeptidique a généralement lieu sur des domaines exposés et flexibles de la protéine native (Kowarik et al., 2006a). Les N-glycannes sont transférés sur les asparagines de séquences consensus similaires à celles des eucaryotes, soit D-X-N-Z-S/T et N-X-S/T pour les systèmes bactériens et eucaryotes, respectivement, où X et Z représentent n'importe quel acide aminé sauf la proline (Kowarik et al., 2006b). Ce type de glycosylation s'avère inapproprié, pour l'instant, pour la production de glycoprotéines d'origine eucaryotique (Brooks, 2006), supportant le potentiel des hôtes eucaryotes en moléculture.

Chez tous les eucaryotes, la biosynthèse du glycanne précurseur et les premières étapes de la N-glycosylation au sein du RE sont communes. Levures, insectes, plantes et mammifères transfèrent, dans la lumière du RE, le même glycanne précurseur sur la même séquence consensus. Ce glycanne est ensuite modifié par des glucosidases et des glycannes de type oligomannosidique ( $\text{Man}_{5-9}\text{GlcNAc}_2$ ) ont été identifiés chez toutes ces espèces. C'est au sein de l'appareil de Golgi que les voies de maturation des N-glycannes divergent. Tel qu'illustré à la **Figure I.7**, chez les levures, des résidus mannoses sont ajoutés pour générer des structures hypermannosylées (Gemmill et Trimble, 1999). La majorité des glycoprotéines, endogènes comme recombinantes, synthétisées par les cellules d'insectes comportent des glycannes de type oligomannosidique ( $\text{Man}_{5-9}\text{GlcNAc}_2$ ) ou paucimannosidique ( $\text{Man}_{1-3}\text{GlcNAc}_2$ ; Harrison et Jarvis, 2006). Les plantes, pour leur part, disposent d'enzymes golgiennes spécifiques, des  $\alpha$ -1,3 fucosyltransférases et des  $\beta$ -1,2 xylosyltransférases, qui catalysent respectivement l'ajout d' $\alpha$ -1,3 fucose sur la GlcNAc proximale et de  $\beta$ -1,2 xylose sur le  $\beta$ -mannose, deux résidus absents des glycannes de mammifères (Cabanes-Macheteau et al., 1999; Samyn-Petit et al., 2001; Bardor et al., 2003a). Chez les mammifères, enfin, les protéines peuvent arborer une variété de glycannes complexes bi-, tri- ou tétra-antennés, avec à leurs extrémités des galactoses

liés en  $\beta$ -1,4, ainsi que des résidus d'acides sialiques (Neu5Ac, chez l'être humain), absents des végétaux, des levures et de la plupart des insectes. Un résidu fucose peut également être retrouvé lié en  $\alpha$ -1,6 sur la GlcNAc proximale, plutôt qu'en  $\alpha$ -1,3 comme chez les végétaux.



**Figure I.7 Structure des N-glycannes typiques identifiés dans différentes branches du vivant.** Chez les eucaryotes, les structures des glycannes proviennent de la maturation d'un glycane précurseur présent chez toutes les espèces. Les N-glycannes complexes chez les plantes comportent typiquement un résidu de xylose lié en  $\beta$ -1,2 sur le mannose du core et un résidu fucose lié en  $\alpha$ -1,3 sur le résidu GlcNAc proximal. Les N-glycannes complexes chez les mammifères peuvent être bi- tri- ou tétra-antennés avec des galactoses liés en  $\beta$ -1,4 et des résidus d'acide sialique (le Neu5Ac chez l'être humain) absents des végétaux, des levures et de la plupart des insectes. Chez les mammifères le fucose est lié en  $\alpha$ -1,6 sur le résidu GlcNAc proximal. Chez les bactéries, un résidu typique, le bacillosamine (2,4-diacétamido-2,4,6-tridéoxy-D-glucopyranoside) est ajouté. Adapté de Varki et al. 1999 et de Young et al. 2002.

### **I.2.3 Les rôles de la N-glycosylation**

La N-glycosylation étant un processus ubiquitaire chez les eucaryotes et l'une des modifications post-traductionnelles des protéines les plus courantes, son importance pour le bon fonctionnement cellulaire et le développement d'un organisme multicellulaire apparaît aujourd'hui incontestable. Les nombreuses maladies répertoriées chez l'Homme en lien avec des désordres de la glycosylation (*congenital disorders of glycosylation*, CDG), illustrent d'ailleurs bien cette importance (Grünewald et al., 2002; Eklund et Freeze, 2006). Aucun organisme eucaryote ne peut survivre en absence complète de glycosylation et des mutations d'enzymes-clés du sentier de la glycosylation s'avèrent embryo-létales chez la souris (Ioffe et Stanley, 1994; Schwarzkopf et al., 2002) et chez la plante (Boisson et al., 2001). Ces effets drastiques résultent d'une perte d'intégrité, de stabilité et d'activité des protéines, et d'une altération des fonctions intermoléculaires en l'absence des glycannes.

Les rôles des glycannes sont parfois subtils; l'activité *in vitro* d'une protéine peut ne pas changer en l'absence ou en cas de modification de la structure du N-glycane, alors que certaines de ses propriétés *in vivo* seront perturbées (Coloma et al., 2000). À l'échelle biochimique, les glycannes possèdent des fonctions intramoléculaires et intermoléculaires. Les résidus les plus proximaux des glycannes interviennent le plus souvent à des niveaux structuraux universels comme le repliement et la stabilité des protéines, tandis que les résidus monosaccharidiques plus distaux sont davantage rattachés à des processus spécifiques, comme par exemple la communication intercellulaire, la reconnaissance cellule-cellule ou les interactions protéine-protéine (Varki et al., 1999; Trombetta, 2003; Varki, 2007).

#### **I.2.3.1 Les rôles intramoléculaires**

Un rôle universel de la glycosylation est de promouvoir le repliement des protéines (Ceriotti et al., 1998; Trombetta, 2003; Helenius et Aebi, 2004). Les glycannes sont transférés sur la chaîne polypeptidique naissante avant qu'elle n'adopte sa conformation finale, créant ainsi des contraintes et des interactions supplémentaires qui limitent la mobilité de la chaîne polypeptidique tout en stabilisant certaines structures locales (Imperiali et O'Connor, 1999; Wormald et Dwek, 1999). De plus, l'hydrophilie des

glycannes aide à maintenir en solution les intermédiaires de repliement qui, autrement, risqueraient de former des agrégats insolubles. Les glycannes interagissent aussi avec des lectines du RE, la calnexine et la calréticuline, qui assistent le repliement de la glycoprotéine et en assurent le « contrôle qualité » (Helenius et Aebi, 2001). Ces phénomènes prennent place dans le RE durant les premières étapes de la glycosylation, de façon similaire chez toutes les espèces eucaryotes. Une glycoprotéine recombinante sera donc en général correctement repliée dans un hôte hétérologue eucaryote.

Pour certaines protéines, les glycannes ne sont pas indispensables au repliement mais contribuent favorablement à l'atteinte d'une conformation native et, subséquentement, à l'obtention de rendements intéressants (Ceriotti et al., 1998). Les glycannes forment des « branches » flexibles et polaires exposées à la surface des protéines. Ils influencent la solubilité, le point isoélectrique (dans le cas des glycannes sialylés), la stabilité et la rigidité structurale de la protéine. Des modifications à la structure des glycannes permettent un fin contrôle des propriétés de la protéine sans altérer sa séquence en acides aminés. Une glycoprotéine est souvent plus résistante à la dégradation protéolytique grâce à une meilleure stabilité interne de sa structure, mais aussi grâce à l'encombrement stérique créé par le glycanne, qui limite l'accès aux protéases (Kundra et Kornfeld, 1999). La streptokinase, une protéase d'origine procaryotique naturellement non glycosylée, montre par exemple une plus grande stabilité et une résistance accrue à la dégradation protéolytique lorsqu'elle est exprimée chez la levure *Pichia pastoris* sous une forme recombinante glycosylée (Pratap et al., 2000).

### **1.2.3.2 Les rôles intermoléculaires**

Au niveau intermoléculaire, les glycannes sont souvent impliqués dans la spécificité des interactions protéine-protéine, dans la communication entre les cellules, dans l'adhésion cellulaire et dans l'interaction de la protéine avec des ligands et/ou avec divers agents pathogènes (Haltiwanger et Lowe, 2004; Ohtsubo et Marth, 2006). Dans ces derniers cas, les glycannes des glycoprotéines (de même que les glycolipides) extracellulaires sont reconnus spécifiquement par des récepteurs, les lectines, déclenchant alors une cascade de réactions métaboliques souvent spécifiques à l'interaction. De ce système de glyco-reconnaissance dépendent entre autres, chez les mammifères, la réponse immunitaire, la réponse inflammatoire et la communication entre



les cellules nerveuses (Crocker et Feizi, 1996; Varki, 2007). Il va sans dire qu'un tel système de reconnaissance spécifique dépend non plus des glycanes homologues chez toutes les cellules eucaryotes, mais des processus de maturation golgienne plus ou moins complexes, distincts entre les espèces. La grande diversité de monosaccharides impliqués ainsi que la multitude de leurs embranchements possibles permet de coder une grande quantité d'informations nécessaires à la reconnaissance moléculaire spécifique, de façon compacte, polaire et exposée. Des changements mineurs dans la structure du glycanne risquent alors de grandement altérer sa reconnaissance par une lectine (Weis et Drickamer, 1996).

#### **1.2.4 Les plantes, des bio-usines à glycoprotéines**

Dans une perspective d'application, les glycanes influencent grandement les caractéristiques des protéines, affectant par exemple leurs propriétés pharmacologiques (Sethuraman et Stadheim, 2006). Or, chaque espèce possède ses structures glycaniques spécifiques, plus ou moins différentes de celles de l'être humain. Dans ce contexte, le choix du système hétérologue de production devient primordial et ne doit pas être basé uniquement sur des critères de rentabilité économique. Les glycoprotéines recombinantes devront être analysées pour s'assurer d'un maintien suffisant de leurs caractéristiques biochimiques et immunogéniques. La grande plasticité de la machinerie cellulaire végétale pour la synthèse de glycoprotéines exogènes variées et complexes est illustrée par le nombre de glycoprotéines déjà produites dans les plantes (voir **Tableau I.1**; Ma et al., 2005b). La structure des glycanes complexes, toutefois, diffère de celle des glycoprotéines endogènes de mammifères par l'absence de  $\beta$ -1,4 galactose et d'acide sialique, et par la présence de résidus typiquement végétaux, le  $\beta$ -1,2 xylose et l' $\alpha$ -1,3 fucose. De ces dissemblances peuvent découler des propriétés biologiques et pharmacologiques différentes et le risque de réactions allergiques et immunogènes éventuelles chez le patient traité pour une quelconque maladie.

#### **I.2.4.1 Les glycoprotéines végétales ne sont pas sialylées**

Une glycoprotéine produite sous forme non glycosylée, par exemple dans *Escherichia coli*, risque d'être mal repliée, inactive, instable ou rapidement éliminée de la circulation sanguine lorsqu'administrée en traitement thérapeutique. C'est le cas, notamment, des immunoglobulines G (IgG) qui, en absence de glycanes, voient leur fonctions effectrices et leur demi-vie s'amenuiser (Coloma et al., 2000).

Si l'absence ou la présence d'un N-glycane sont importantes, sa structure même est essentielle pour assurer à la glycoprotéine l'ensemble de ses fonctions. Les acides sialiques, notamment, jouent un rôle prépondérant sur la stabilité *in vivo* des protéines. Ces résidus se retrouvent sur les glycanes de mammifères, l'acide N-acétylneuraminique (Neu5Ac) étant la forme présente chez l'Homme. Par exemple, les propriétés de l'érythropoïétine (EPO) recombinante produite en culture CHO (*chinese hamster ovary cells*) dépendent grandement de la structure de ses glycanes. De simples modifications des conditions de culture des cellules ovariennes ou des procédés de purification distincts ont engendré différentes glycoformes dont les fonctions erythropoïétiques variaient grandement (Yuen et al., 2003). En particulier, les EPO sialylées montraient une augmentation d'activité *in vivo* associée à une augmentation de leur demi-vie. Il s'avère que les protéines non sialylées sont reconnues par des récepteurs hépatiques spécifiques des galactoses exposés en l'absence d'acide sialique (Kelm et Schauer, 1997). Elles sont alors rapidement éliminées de la circulation sanguine par le foie.

En dépit de résultats controversés, les acides sialiques semblent absents du règne végétal (Seveno et al., 2004; Zeleny et al., 2006), si bien que les glycoprotéines recombinantes produites par les plantes, non sialylées, pourraient avoir une utilité limitée en thérapie car elles auraient toutes les chances d'être éliminées rapidement de la circulation sanguine (Matsumoto et al., 1995). Dans ce contexte, plusieurs recherches ont été entamées pour tenter de modifier la voie de maturation des glycanes chez les plantes. À cette fin, les gènes nécessaires non seulement au transfert de l'acide sialique sur l'oligosaccharide, mais aussi ceux indispensables à la synthèse même de ce glucide, doivent être intégrés par transgénése au génome des plantes. Établir la sialylation des protéines *in planta* est une tâche très ardue, la sialylation d'un glycane, de la synthèse du Neu5Ac à son transfert sur le glycane, nécessitant l'incorporation au génome de la plante

d'au moins six nouveaux gènes (voir **Tableau I.4**) ! Plusieurs étapes de la synthèse du Neu5Ac ont déjà été explorées chez les plantes. Le transfert de l'acide sialique sur le glycanne nécessite tout d'abord l'ajout d'un galactose en  $\beta$ -1,4 à l'extrémité du glycanne accepteur par une galactosyltransférase exogène (Palacpac et al., 1999; Bakker et al., 2001). Cette étape permet la formation du glycanne accepteur de l'acide sialique. La condensation de N-acétylmannosamine (ManNAc) avec le phosphoénolpyruvate (PEP) pour former du Neu5Ac a été réalisée récemment (Paccalet et al., 2007), de même que l'activation du Neu5Ac en CMP-Neu5Ac et son transport vers l'appareil de Golgi (Misaki et al., 2006). D'autre part, l' $\alpha$ -2,6 sialyltransférase a, elle aussi, été exprimée dans la plante (Wee et al., 1998). À ce stade, deux obstacles s'opposent encore à la sialylation *in planta* : l'expression d'une UDP-GlcNAc-2 épimérase fonctionnelle essentielle pour la formation dans le cytosol végétal de ManNAc à partir d'UDP-GlcNAc, et la co-expression de toutes ces enzymes à des niveaux satisfaisants au sein d'une même plante.

**Tableau I.4 Enzymes et transporteur nécessaires à la synthèse et à l'incorporation du Neu5Ac sur les N-glycannes des glycoprotéines végétales.**

Enzymes et transporteur	Fonctions
UDP-GlcNAc 2-épimérase	Conversion de l'UDP-GlcNAc en ManNAc
Neu5Ac synthase	Condensation du ManNAc + PEP en Neu5Ac
CMP-Neu5Ac synthase	Activation du Neu5Ac en CMP-Neu5Ac
Transporteur de CMP-Neu5Ac	Transfert du CMP-Neu5Ac du cytosol vers le Golgi
$\beta$ -1,4 galactosyltransférase	Ajout d'un Gal en $\beta$ 1,4 sur le glycanne, accepteur du Neu5Ac
$\alpha$ -2,6 sialyltransférase	Ajout de Neu5Ac sur accepteur galactosylé

#### **1.2.4.2 Les glycoprotéines végétales comportent des épitopes immunogènes**

Une autre limitation importante à l'utilisation des glycoprotéines recombinantes végétales en thérapie réside dans la présence de glycoépitopes spécifiques aux végétaux pouvant éliciter des réponses immunitaires chez l'Homme (Bardor et al., 2003a; Gomord et al., 2005). D'une part, les glycoépitopes peuvent être à l'origine de réactions allergiques et des IgE spécifiques des glycoépitopes  $\beta$ -1,2 xyloses et  $\alpha$ -1,3 fucoses ont été découverts chez des patients allergiques (van Ree et al., 2000; Bardor et al., 2003a; Altmann, 2007). D'autre part, des glycoprotéines végétales s'avèrent immunogènes lorsqu'elles sont administrées à des lapins (Faye et al., 1993), des chèvres (Kurosaka et al., 1991) ou des rongeurs (rat; Bardor et al., 2003a). À l'inverse, la lignée de souris Balb/c ne réagit pas à la présence des résidus monosacharidiques typiques des végétaux

(Chargelegue et al., 2000), ce qui laisse supposer que la réponse immunitaire serait spécifique à chaque espèce. Bien que les réactions allergiques dues à des glycoprotéines recombinantes végétales causent rarement des symptômes cliniques chez l'humain (Altmann, 2007), une glycoprotéine présentant des N-glycannes immunogènes risque d'être plus rapidement éliminée de la circulation sanguine, ce qui réduira d'autant son efficacité.

Devant les risques potentiels d'immunogénicité des glycoépitopes végétaux, certaines solutions ont été proposées (Gomord et al., 2004; Gomord et al., 2005; Saint-Jore-Dupas et al., 2007). Une première possibilité consisterait à retenir la protéine recombinante dans le RE; les premières étapes de la voie de glycosylation étant communes chez toutes les espèces eucaryotes, les glycannes seraient de type oligomannosidique et non immunogènes. En pratique, l'ajout d'un signal de rétention tétrapeptidique K/HDEL à l'extrémité C-terminale de la protéine permet de retenir la protéine dans le RE. Par exemple, la fusion du signal KDEL sur un anticorps spécifique pour la gonadotrophine humaine a permis l'obtention de glycannes oligomannosidiques comportant de six à neuf mannoses (Sriraman et al., 2004). La présence de glycannes de six et sept mannoses témoigne de l'action de l' $\alpha$ -mannosidase I, une enzyme du *cis* Golgi, confirmant l'hypothèse d'un rétrotransport des protéines résidentes du RE à partir de l'appareil de Golgi dans les cellules végétales (Lee et al., 1993; Pagny et al., 2000). Toutefois, en fusionnant le signal KDEL à un anticorps contre la rage, Ko et collaborateurs (2003) ont montré qu'au moins 10% des protéines recombinantes comportaient des glycannes complexes portant les résidus immunogènes  $\beta$ -1,2 xylose et  $\alpha$ -1,3 fucose, des résultats similaires à ceux d'autres équipes s'intéressant à la glycosylation (Mokrzycki-Issartel et al., 2003). Comme la xylosylation et la fucosylation des glycannes sont surtout concentrées au niveau des compartiments médian et *trans* Golgi, respectivement (Fichette-Lainé et al., 1994), l'observation de résidus xylozes et fucoses laisse supposer soit que le système de recyclage des protéines fusionnées au signal K/HDEL n'est pas étanche, une hypothèse déjà émise suite à l'observation de fuites de reticuloplasmines endogènes de la plante hors du RE (Navazio et al., 2002); soit qu'il y a, comme dans les cellules animales, présence d'un rétrotransport à partir des compartiments médian et *trans* Golgi (Miesenbock et Rothman, 1995).

En somme, la rétention des protéines recombinantes dans le RE n'est pas une solution universelle au contrôle de la présence d'épitopes immunogènes. Les données disponibles dans la littérature sont variables d'une protéine à l'autre, si bien que chaque nouvelle protéine devra être analysée afin de détecter la présence éventuelle de glycannes immunogènes. Par ailleurs, les glycannes oligomannosidiques risquent d'interagir avec des récepteurs spécifiques aux mannoses, avec comme conséquence une élimination plus rapide de la protéine circulant dans le sang (Ko et al., 2003).

Une approche alternative pour éviter la production de protéines immunogènes repose sur la modification du sentier de la glycosylation afin d'éviter la synthèse des enzymes générant les glycoépitopes nuisibles. Le mutant d'*Arabidopsis thaliana* *cgl*, par exemple, contient une mutation au niveau du gène de la N-acétylglucosaminyltransférase I (GNT I; voir **Figure I.6**) qui rend impossible la maturation des glycannes au-delà des oligosaccharides  $\text{Man}_5\text{GlcNAc}_2$  non immunogènes (von Schaewen et al., 1993). Toutefois, une telle stratégie, tout comme la rétention des protéines dans le RE, peut poser des problèmes de stabilité *in vivo* de la protéine et d'autres recherches se sont plutôt concentrées sur l'inactivation de l' $\alpha$ -1,3 fucosyltransférase (FucT) et de la  $\beta$ -1,2 xylosyltransférase (XylT) par génie génétique. Par recombinaison homologue, ces gènes ont, par exemple, été réprimés chez la mousse *Physcomitrella patens* (Koprivova et al., 2004). Ces mousses transgéniques ne présentaient aucun phénotype morphologique ou physiologique particulier et leurs glycannes complexes ne comportaient pas de glycoépitopes propres aux glucides transférés par les glycosyltransférases réprimées. De façon similaire, des plantes d'*A. thaliana* mutées au niveau du gène de la XylT et des deux gènes de FucT endogènes ont été isolées à partir d'une banque de mutants d'insertion et les trois mutants ont été croisés pour générer des plantes exprimant des protéines arborant des glycannes complexes exempts de xyloses et de fucoses immunogènes (Strasser et al., 2004; Schähns et al., 2007). Plus récemment, à l'aide des techniques « d'ARN interférence » (de l'anglais : RNAi), une inhibition de l'expression des gènes FucT et XylT a été effectuée partiellement, voire totalement, chez *N. tabacum*, *Medicago sativa* (la luzerne; Sourrouille, 2005) et *Lemna minor* (lentille d'eau; Cox et al., 2006), suggérant le potentiel des approches moléculaires pour la génération d'hôtes végétaux adaptés à la biosynthèse de glycoprotéines non immunogènes.

### 1.2.4.3 La glycosylation végétale, un avantage ?

La structure des glycanes végétaux, avec présence de résidus  $\beta$ -1,2 xylose et  $\alpha$ -1,3 fucose, est, de prime abord, un obstacle majeur à la commercialisation médicale des glycoprotéines végétales. Mais ces mêmes glycanes végétaux pourraient-ils, dans certains cas, apporter un avantage biochimique à la protéine recombinante? Pour la production de vaccins dans les plantes, la présence de N-glycanes végétaux immunogènes pourrait par exemple servir d'adjuvant. Elle pourrait aussi aider à stimuler le système immunitaire en augmentant le caractère immunogène de l'antigène et en facilitant sa capture par les «*antigen-presenting cells*» (van Ree et Aalberse, 1999). Plus encore, la connaissance de plus en plus détaillée de la voie de biosynthèse de la N-glycosylation chez les plantes et le développement d'outils moléculaires pour modifier cette voie biochimique (voir Section 1.2.4.2) permettent maintenant d'envisager de moduler la structure des N-glycanes en fonction des propriétés biochimiques et biophysiques recherchées (Saint-Jore-Dupas et al., 2007). Ainsi, l'obtention de glycanes oligomannosidiques, par rétention de la glycoprotéine dans le RE ou par l'utilisation de mutants de la N-glycosylation (p.ex. mutant *cgl* d'*A. thaliana*), peut stimuler le système immunitaire *via* l'interaction avec les récepteurs à mannose des macrophages ou des cellules dendritiques. Par exemple, la glucocérébrosidase pour le traitement de la maladie de Gaucher nécessite la présence de mannoses à l'extrémité des N-glycanes afin d'être reconnue et internalisée par les récepteurs à mannose des macrophages. Synthétisée dans les cellules de mammifères CHO, cette enzyme doit subir une déglycosylation enzymatique *in vitro* pour assurer toute son activité, ce qui augmente considérablement le coût du médicament (Friedman et al., 1999). En revanche, en adressant la glucocérébrosidase vers la vacuole de cellules de carottes, Shaaltiel et al. (2007) ont produit récemment une enzyme avec des glycanes de type paucimannosidique et une activité biologique similaire au médicament commercialisé.

Les plantes sont envisagées, par ailleurs, pour la production d'anticorps recombinants pour des applications médicales ou vétérinaires (Peeters et al., 2001b; Ma et al., 2005a). Or il s'avère que les N-glycanes sont essentiels aux fonctions effectrices de l'anticorps en participant à l'interaction avec les récepteurs spécifiques (récepteurs Fc) et à la stimulation du système immunitaire (Shields et al., 2001). Il a récemment été démontré que l'élimination des  $\alpha$ -1,6 fucoses des anticorps produits chez les CHO

(Shields et al., 2002) ou des  $\alpha$ -1,3 fucoses des anticorps produits chez la plante aquatique *L. minor* (Cox et al., 2006) augmentait l'interaction de ces anticorps avec les récepteurs Fc et amplifiait la cytotoxicité cellulaire dépendante des anticorps (*antibody-dependent cellular cytotoxicity*, ADCC). Ainsi, en inhibant l'expression des xylosyltransférases et des  $\alpha$ -1,3 fucosyltransférases chez *L. minor*, Cox et al. (2006) ont non seulement diminué le potentiel immunogène de l'anticorps recombinant produit dans cette plante, mais en ont aussi augmenté les fonctions effectrices.

#### **1.2.4.4 Les glycoprotéines recombinantes, des produits non-homogènes**

La N-glycosylation, enfin, est un processus enzymatique conditionné par l'accessibilité de la protéine et du glycanne lui-même aux enzymes de maturation de la N-glycosylation (voir Section 1.2.2). En général, la glycoprotéine et la structure des glycannes sont typiquement hétérogènes, le nombre de glycannes portés par une glycoprotéine et la structure des glycannes sur un même site de glycosylation ou à des sites différents étant variables (Lerouge et al., 1998). Or, un produit pharmaceutique sera d'autant plus facile à commercialiser qu'il sera homogène. Dans les faits, le phénomène d'hétérogénéité des glycannes n'est pas unique aux plantes mais concerne tous les systèmes d'expression. À titre d'exemple, la N-glycosylation de l'anticorps 'Guy 13' produit dans le tabac *N. tabacum* a démontré une diversité de structures oligosaccharidiques supérieure à celle observée sur le même anticorps produit chez la souris (Cabanes-Macheteau et al., 1999). En revanche, l'anticorps C5-1 synthétisé dans la luzerne (*Medicago sativa*) montrait un patron de glycosylation beaucoup plus homogène que les anticorps recombinants provenant de *N. tabacum* (Bardor et al., 2003b). À ce stade, la recherche de systèmes d'expression permettant une meilleure homogénéité des protéines d'intérêt demeure une priorité pour la production de protéines d'intérêt thérapeutique ou industriel, sur laquelle se penchent maintenant plusieurs groupes de recherche.

### **I.3. Les serpins**

Dans ce contexte d'études de l'expression et de la maturation des protéines complexes en systèmes végétaux, nous nous sommes intéressés aux propriétés d'un inhibiteur de protéases de type sérine, l' $\alpha$ 1-antichymotrypsine (AACT) humaine, exprimée dans une suspension cellulaire de tabac *Nicotiana tabacum* cv BY-2. L'AACT fait partie de la superfamille des serpins, qui comprend un large éventail de protéines dans toutes les branches du vivant (bactéries, archéobactéries, eucaryotes et virus; Silverman et al., 2001; Law et al., 2006). Les serpins sont, pour la plupart, des inhibiteurs de protéase (IP) de type sérine (serpin de l'anglais « *serine protease inhibitor* »). Ne mentionner, par contre, que leur rôle d'inhibiteurs de protéase de type sérine serait restreindre les serpins à une seule de leurs diverses fonctions. Plusieurs serpins ont aussi la capacité d'inhiber des protéases de type cystéine comme les caspases ou des protéases de type papaïne (Schick et al., 1998; Irving et al., 2002). D'autres sont des serpins non-inhibitrices aux rôles diversifiés comme le transport hormonal (Zhou et al., 2006a), l'inhibition de tumeurs (Khalkhali-Ellis, 2006) ou le repliement des protéines (chaperones moléculaires). Dans le cadre de cette thèse, nous avons axé notre étude sur le rôle « inhibiteur de protéases de type sérine », considéré comme le plus important pour l'ensemble des serpins.

Chez l'Homme, les serpins sont les inhibiteurs de protéases de type sérine les plus abondants (Rawlings et al., 2004). Elles participent à réguler finement diverses voies protéolytiques impliquées, entre autres, dans la coagulation sanguine, la fibrinolyse, l'activation du complément, le remodelage des tissus, la prolifération des métastases, l'inflammation et l'apoptose (Gettins, 2002). D'un point de vue médical, et vu la prépondérance des serpins dans tous ces processus physiologiques, il va de soi qu'une défaillance dans la synthèse ou l'activité d'une serpine conduit en général à un état pathologique (**Tableau I.5**).

Sur le plan structure/fonction, les serpins se caractérisent par une structure tertiaire typique conservée et par leur capacité à inhiber de manière irréversible leurs protéases-cibles. Suivant l'interaction covalente avec une protéase cible, la serpine subit un changement de conformation drastique essentiel à l'inhibition. Les données concernant la structure des serpins, leur mécanisme d'inhibition irréversible et leur glycosylation sont résumées dans les sections qui suivent.



## Tableau I.5 Fonctions et maladies associées aux serpinines humaines.

Tableau tiré de Law et al., 2006.

Serpin	Alternative name(s)	Protease target or function	Involvement in disease
SERPINA1	Antitrypsin	Extracellular; inhibition of neutrophil elastase	Deficiency results in emphysema; polymerization and retention in the ER results in cirrhosis
SERPINA2	Antitrypsin-related protein	Not characterized, probable pseudogene	
SERPINA3	Antichymotrypsin	Extracellular; inhibition of cathepsin G	Deficiency results in emphysema
SERPINA4	Kallistatin (PI4)	Extracellular, inhibition of kallikrein	
SERPINA5	Protein C inhibitor (PAI-3)	Extracellular; inhibition of active protein C	Angioedema
SERPINA6	Corticosteroid-binding globulin	Extracellular; non-inhibitory; cortisol binding	Deficiency linked to chronic fatigue
SERPINA7	Thyroxine-binding globulin	Extracellular; non-inhibitory, thyroxine binding	Deficiency results in hypothyroidism
SERPINA8	Angiotensinogen	Extracellular; non-inhibitory; amino-terminal cleavage by the protease renin results in release of the decapeptide angiotensin I	Certain variants linked to essential hypertension
SERPINA9	Centerin	Extracellular; maintenance of naive B cells	
SERPINA10	Protein Z-dependent proteinase inhibitor	Extracellular; inhibition of activated factor Z and XI	Deficiency linked to venous thromboembolic disease
SERPINA11	XP_170754.3	Not characterized	
SERPINA12	Vaspin	Extracellular; insulin-sensitizing adipocytokine	
SERPINA13	XM_370772	Not characterized	
SERPINB1	Monocyte neutrophil elastase inhibitor	Intracellular; inhibition of neutrophil elastase	
SERPINB2	Plasminogen activator inhibitor-2 (PAI2)	Intracellular; inhibition of uPA	
SERPINB3	Squamous cell carcinoma antigen-1	Intracellular; cross-class inhibition of cathepsins L and V	
SERPINB4	Squamous cell carcinoma antigen-2	Intracellular; cross-class inhibition of cathepsin G and chymase	
SERPINB5	Maspin	Intracellular; non-inhibitory; inhibition of metastasis through uncharacterized mechanism	Downregulation and/or intracellular location linked to tumor progression and overall prognosis
SERPINB6	Proteinase inhibitor-6 (PI6)	Intracellular, inhibition of cathepsin G	
SERPINB7	Megsin	Intracellular; megakaryocyte maturation	IgA nephropathy
SERPINB8	Cytoplasmic antiproteinase 8 (PI8)	Intracellular; inhibition of furin	
SERPINB9	Cytoplasmic antiproteinase 9 (PI9)	Intracellular, inhibition of granzyme B	
SERPINB10	Bomapin (PI10)	Intracellular; inhibition of thrombin and trypsin	
SERPINB11	Epipin	Intracellular	
SERPINB12	Yukopin	Intracellular; inhibition of trypsin	
SERPINB13	Headpin (PI13)	Intracellular; inhibition of cathepsins L and K	
SERPINC1	Antithrombin	Extracellular; thrombin and factor Xa inhibitor	Deficiency results in thrombosis
SERPIND1	Heparin cofactor II	Extracellular; thrombin inhibitor	May contribute to thrombotic risk when combined with other deficiencies
SERPINE1	Plasminogen activator inhibitor 1 (PAI1)	Extracellular; inhibitor of thrombin, uPA, tPA and plasmin	Abnormal bleeding
SERPINE2	Protease nexin 1 (PI7)	Extracellular; inhibition of uPA and tPA	
SERPINE3	Hs.512272	Not characterized	
SERPINF1	Pigment epithelium derived factor	Non-inhibitory; potent anti-angiogenic molecule	
SERPINF2	Alpha-2-antiplasmin	Extracellular; plasmin inhibitor	Unrestrained fibrinolytic activity, bleeding
SERPING1	C1 inhibitor	C1 esterase inhibitor	Angioedema
SERPINH1	47kDa heat-shock protein	Non-inhibitory molecular Chaperone for collagens	
SERPINI1	Neuroserpin (PI12)	Extracellular; inhibitor of tPA, uPA and plasmin	Polymerization results in dementia
SERPINI2	Myoepithelium-derived serine proteinase inhibitor (PI14)	Extracellular; inhibition of cancer metastasis	

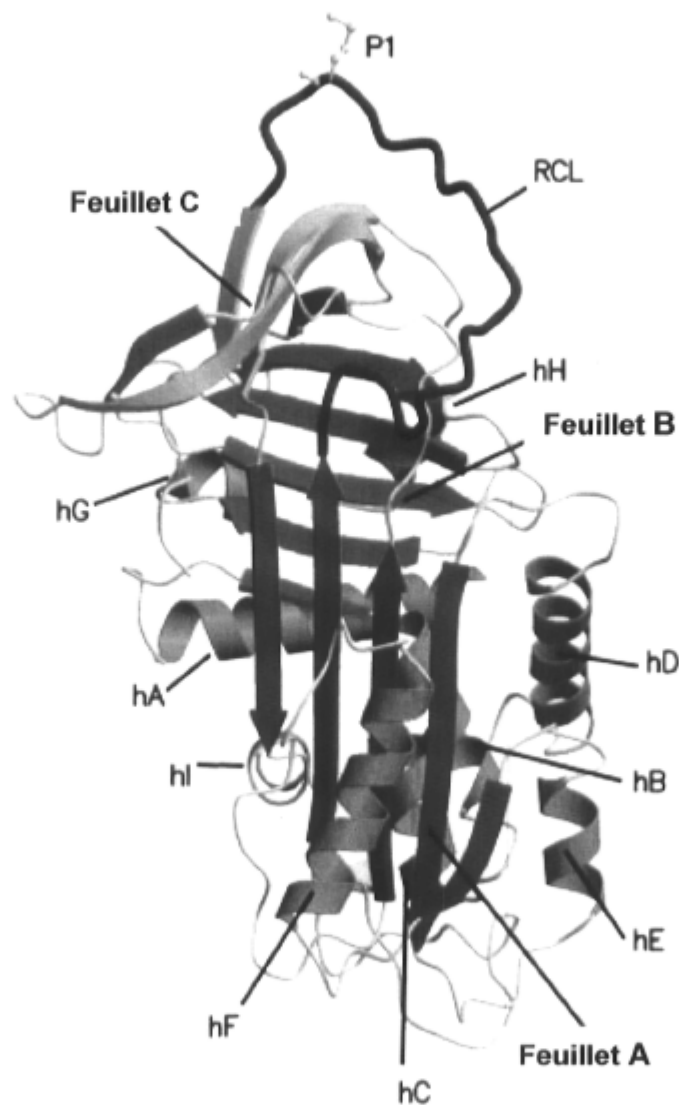
### 1.3.1 La structure tridimensionnelle des serpinés

Les serpinés forment une classe protéique très diversifiée mais conservée sur le plan structurel. Malgré des séquences en acides aminés parfois très divergentes, les serpinés, qu'elles soient ou non inhibitrices de protéases, ont une structure tertiaire hautement conservée (**Figure I.8**), avec présence de trois feuillets bêta, de neuf hélices alpha et d'une boucle mobile exposée à la surface. Cette boucle forme le site d'inhibition, ou RCL (de l'anglais : *reactive center loop*), qui interagit spécifiquement avec la protéase cible. Au-delà de ces éléments de structure secondaire et tertiaire, la plasticité structurale des serpinés est l'un de leurs caractères les plus remarquables. Les serpinés présentent une conformation changeante, ou flexible, qui assure un mécanisme d'inhibition des protéases unique mais qui leur confère aussi une instabilité structurale forçant leur transition vers des formes latentes ou polymériques non actives (Huntington, 2006; Whisstock et Bottomley, 2006).

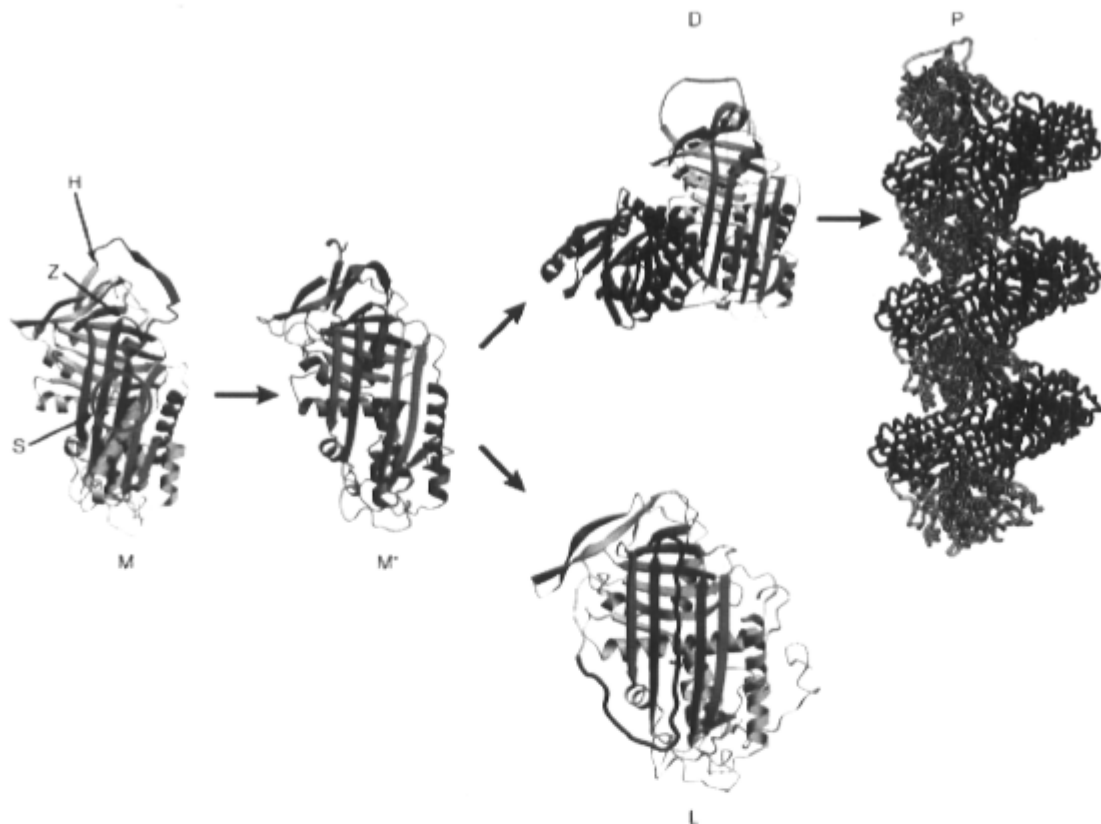
Contrairement au dogme du repliement des protéines qui veut que la chaîne peptidique se replie en divers intermédiaires jusqu'à l'obtention de la structure thermodynamique la plus stable (Anfinsen, 1973), les serpinés, dans leur repliement natif, restent piégées par une barrière cinétique dans un état de plus haute énergie, un état nommé « métastable » (Whisstock et Bottomley, 2006). Cette barrière cinétique est franchie lors de l'interaction de la serpine avec une protéase, conduisant à l'inhibition de cette dernière. Au cours de l'interaction, le RCL est clivé par la protéase et s'insère comme brin supplémentaire du feuillet  $\beta$  A, entraînant à sa suite la protéase liée par une interaction covalente (voir Section 1.3.2, **Figure I.11**). La serpine, en complexe avec la protéase ou dans sa forme clivée, offre alors une nouvelle conformation de plus faible énergie et de meilleure stabilité thermodynamique.

En raison de son état métastable, la structure des serpinés peut être facilement perturbée par des températures élevées ou par des mutations au niveau de la séquence primaire. Il s'ensuit un changement de conformation, qui mène à une transition vers des formes inactives mais plus stables (**Figure I.9**). Certaines serpinés existent ainsi sous une forme latente inactive formée par l'insertion du RCL non clivé dans le feuillet  $\beta$  A (**Figure I.9**). D'autres peuvent former des polymères par l'insertion de la boucle active d'une molécule à l'intérieur du feuillet  $\beta$  d'une seconde molécule (Stein et Carrell, 1995; Benning

et al., 2004; **Figure I.9**). Dans cette situation, les polymères s'agrègent et précipitent dans la cellule, causant un ensemble de maladies nommées serpinopathies (Gooptu et al., 2000). La flexibilité et l'état métastable de la protéine la rendent vulnérable à ces changements de conformation, qui peuvent être accentués par des mutations ponctuelles (Gooptu et al., 2000).



**Figure I.8 Structure de l'«α1-protéinase inhibiteur» (α1-PI) en conformation native** (Protein Data Bank (PDB), code 1QLP). Structure typique des serpinines, avec présence de trois feuillets β (feuillet A/rouge, feuillet B/vert et Feuillet C/jaune), de neuf hélices α (hA-hI/gris) et d'un site d'inhibition ou RCL (mauve). Le résidu qui confère la spécificité, nommé résidu P1, est aussi représenté au niveau du RCL. Image tirée de Pike et al., 2002.

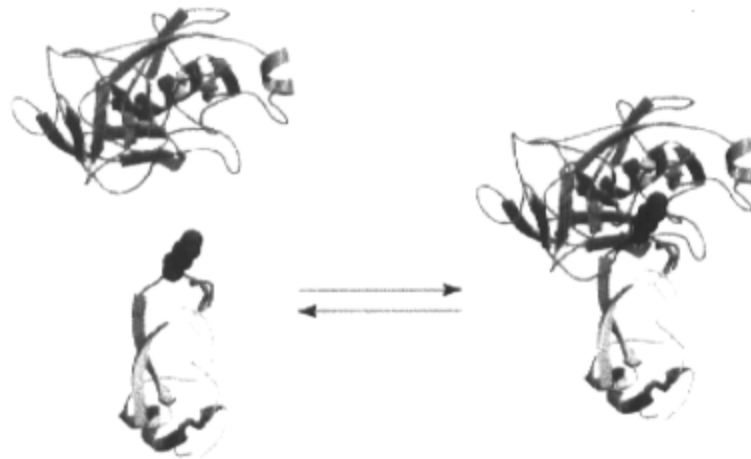


**Figure 1.9 Les serpines peuvent adopter une variété de conformations, spontanément ou suite à des changements du milieu ou à des mutations. Les conformations non natives (illustrées ici par l' $\alpha$ 1-PI) sont inactives. Les changements de conformation débutent par la formation d'un intermédiaire instable suite à l'insertion partielle du RCL (en rouge) dans le feuillet  $\beta$  A (en vert) de la molécule (M\*). Une forme latente (L) peut se former par l'insertion du RCL non clivé (en rouge) dans le feuillet  $\beta$  A (en vert) de la même molécule. Le feuillet  $\beta$  A peut aussi accepter le RCL d'une seconde molécule (en bleu) pour former un dimère (D) qui peut ensuite s'étendre pour former un polymère (P). Les molécules individuelles d' $\alpha$ 1-PI sont représentées en rouge, jaune et bleu. Image tirée de Lomas et Carrell, 2002.**

### 1.3.2 Le mécanisme d'inhibition des serpines

*In vivo*, les cellules synthétisent un grand nombre d'inhibiteurs de protéases qui varient en termes de spécificité et d'activité, pouvant ainsi réguler avec précision une multitude de protéases tout aussi variées. Sur le plan fonctionnel, les inhibiteurs de protéases se distinguent de plus par leur mode d'inhibition (Otlewski et al., 2005). Ainsi retrouve-t-on, entre autres, pour inhiber les protéases de type sérine, les serpines et les

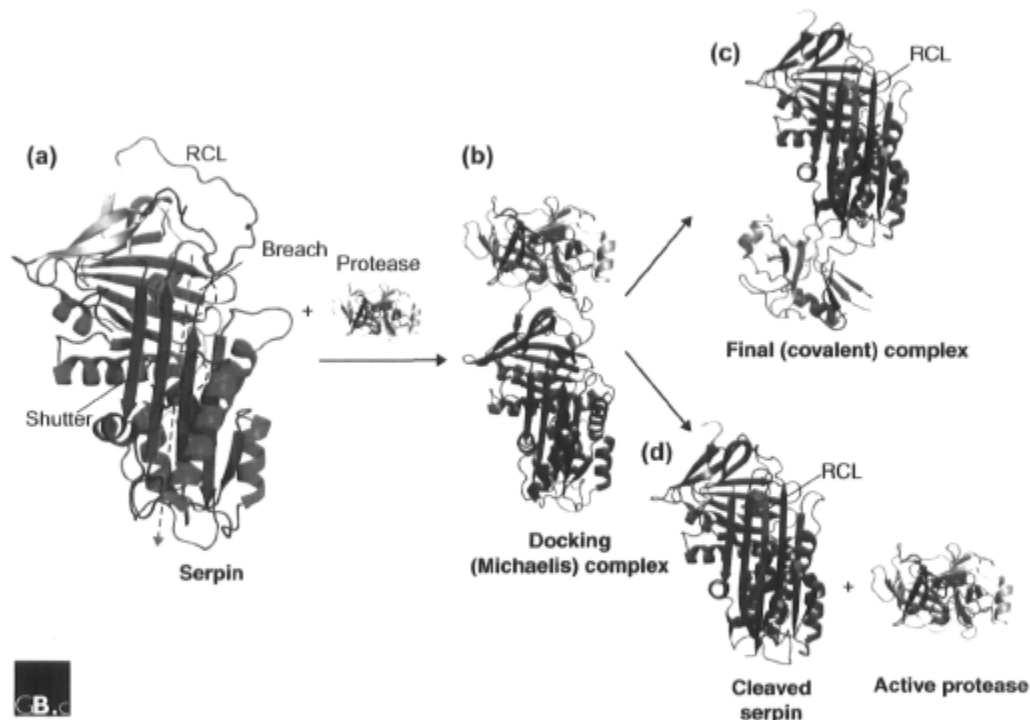
inhibiteurs canoniques. Les inhibiteurs canoniques emploient le « mécanisme standard » des inhibiteurs de protéases, par lequel la boucle d'inhibition de l'inhibiteur est reconnue par la protéase cible au niveau de son site actif, par l'intermédiaire d'une interaction réversible à l'image d'une clé et d'une serrure (Kelly et al., 2005). La très grande complémentarité entre les deux protéines permet de former un complexe très stable (avec souvent une constante de dissociation de l'ordre du nanomolaire) par des liaisons faibles non covalentes. Ces inhibiteurs sont en général rigides et stables et leur boucle d'inhibition n'est pas clivée par les protéases-cibles (**Figure I.10**).



**Figure I.10 Mécanisme standard d'inhibition d'une protéase de type sérine par un inhibiteur canonique de même famille.** Ici, la trypsine (cyan) est inhibée par l'inhibiteur *bovine pancreatic trypsin inhibitor* (BPTI, en gris). La boucle d'inhibition du BPTI (en rouge) a une structure rigide maintenue par un pont disulfure (en vert). Cette boucle a une conformation complémentaire au site actif de la trypsine (de type clé/serrure). Il se forme entre les deux un complexe réversible. La boucle d'inhibition bloque le site actif de la trypsine, et aucune des deux molécules ne subit de changement de conformation. Image tirée de Huntington, 2006.

En comparaison, l'interaction des serpins avec leurs protéases cibles est nettement plus complexe (**Figure I.11**). La boucle d'inhibition des serpins (RCL) est d'abord reconnue comme substrat par la protéase, qui clive le RCL et se lie à la serpine *via* un intermédiaire acyl-covalent (Lawrence et al., 1995). La réaction protéolytique est alors interrompue, le RCL libéré par ce clivage s'insère dans le noyau de la serpine au sein du feuillet  $\beta$  A, et la protéase se retrouve emprisonnée dans le complexe (**Figure I.11 C**). Si l'insertion du RCL dans la molécule n'est pas suffisamment rapide, une déacylation

du complexe survient et la réaction emprunte la voie du substrat (**Figure I.11 D**), avec libération d'une protéase active et d'une serpine clivée inactive (Lawrence et al., 2000). L'insertion du RCL dans le feuillet  $\beta$  entraîne finalement la protéase au pôle opposé de la serpine, sur une distance de 70 Å (Stratikos et Gettins, 1999). Ce déplacement inhibe la protéase en engendrant des distorsions de son site actif (Huntington et al., 2000; Dementiev et al., 2006). La serpine, quant à elle, subit un changement de conformation irréversible et un gain de stabilité thermodynamique. Sur une plus longue période, le complexe se dissocie finalement pour libérer la protéase active et une serpine clivée inactive mais stable.



**Figure I.11 Structure et mécanisme d'inhibition des serpines.** A. L' $\alpha$ 1-PI en conformation native (Protein Data Bank (PDB) code 1QLP) ; le RCL est représenté en magenta. B. Complexe Michaelis entre une protéase (ici la trypsine) et l' $\alpha$ 1-PI. Il s'agit du complexe initial non covalent (PDB 1OPH). La réaction peut ensuite suivre deux voies, celle de l'inhibition (en haut) et celle du substrat (en bas). C. Complexe irréversible entre une serpine et une protéase : la serpine subit un changement de conformation, le RCL est clivé par la protéase et s'insère comme brin supplémentaire dans le feuillet  $\beta$  A. La protéase est déplacée au pôle opposé de la serpine et est inhibée (PDB 1EZK). D. La voie du substrat : le complexe se dissocie, la serpine est clivée et inactivée (PDB 7API), et la protéase reste active. Image tirée de Law et al., 2006.

Comment expliquer la prépondérance des serpins au sein de la nature alors que leur mécanisme d'inhibition est à ce point compliqué? Pourquoi une cellule favoriserait-elle une serpine plutôt qu'un inhibiteur à « mécanisme standard »? L'inhibition irréversible et le changement de conformation des protéines associées dans le complexe permet une régulation très fine des protéases et de ce fait pourrait offrir un avantage sélectif aux serpins (Huntington, 2006). La distorsion de la protéase due à l'interaction avec la serpine lui donne une structure dite de « globule fondu » (*molten globule*; Peterson et Gettins, 2001). La protéase devient alors extrêmement sensible à une dégradation protéolytique, ce qui permet son élimination rapide et localisée. De plus, le complexe serpine-protéase peut être reconnu spécifiquement par des récepteurs lipidiques qui induisent la clairance du complexe. Il est suggéré que le changement de conformation de la serpine et/ou de la protéase découvre des sites reconnus par ces récepteurs, qui dans les formes natives sont cachés (Gettins, 2002; Huntington, 2006).

### **1.3.3 La glycosylation des serpins**

Au niveau structural, la majorité des serpins sont des glycoprotéines contenant un ou plusieurs oligosaccharides. Il s'agit essentiellement de N-glycannes mais des O-glycannes ont aussi été détectés, par exemple sur l'inhibiteur de protéinase C1. Les sites de glycosylation ne sont toutefois pas forcément conservés d'une protéine à l'autre (Gettins et al., 1996). De plus, toutes les serpins ne sont pas glycosylées, en particulier les serpins intracellulaires, cytoplasmiques ou nucléaires, comme l'inhibiteur de protéinase 6 (*proteinase inhibitor 6*) cytoplasmique (Scott et al., 1996), la forme cytoplasmique de la serpine PAI-2 (*plasminogen activator inhibitor 2*) ou la serine MENT (*Myeloid and Erythroid Nuclear Termination stage-specific protein*; Grigoryev et al., 1999).

Chaque serpine montrant un patron de glycosylation particulier tant dans le nombre, la position et la structure des glycannes, la contribution de ces glycannes sur leurs fonctions biologiques risque elle aussi de varier. Si, dans certains cas, les N-glycannes peuvent être supprimés sans affecter de façon notable la fonction inhibitrice de la protéine (Rubin et al., 1990; Kwon et al., 1995b; Lamark et al., 2001), l'absence de N-glycannes peut dans d'autres cas nuire à l'insertion du RCL dans le noyau de la protéine et affecter l'activité inhibitrice ou l'état de latence des serpins à conformation modifiée (Gils et al., 2003). Il est connu, par ailleurs, que les N-glycannes peuvent modifier

certaines propriétés des serpinines en conditions *in vivo*. Par exemple, l'élimination de l' $\alpha$ 1-proteinase inhibitor ( $\alpha$ 1-PI) de la circulation plasmatique est favorisée lorsque la protéine est désialylée ou lorsqu'un site de N-glycosylation est éliminé (Ni et al., 2000). La N-glycosylation préviendrait également l'agrégation de l' $\alpha$ 1-PI en améliorant sa stabilité, surtout si les N-glycannes complexes présents sont ceux de la protéine humaine non recombinante (Kwon et Yu, 1997). En somme, ces expériences démontrent que la présence, l'absence ou l'altération de structure d'un N-glycanne peuvent modifier l'activité inhibitrice et les propriétés des serpinines. Dans un contexte thérapeutique où l'activité et la stabilité peuvent influencer la durée et l'efficacité d'un traitement, il devient primordial de caractériser les N-glycannes et de mieux définir leurs effets sur la stabilité fonctionnelle des serpinines.

### **1.3.4 Les serpinines chez les plantes**

Les serpinines sont répandues dans toutes les branches du vivant et les plantes ne font pas exception. Plusieurs serpinines ont été étudiées chez les plantes et, plus récemment, avec le gain continuuel d'information sur les génomes des végétaux, de nombreuses serpinines putatives ont pu être identifiées (Christeller et Laing, 2005). À titre d'exemple, 12 serpinines putatives ont été annotées dans le seul génome d'*A. thaliana*. Chez les céréales (blé, orge, avoine et seigle), les serpinines forment un groupe de protéines nommées «*protéines Z*» abondantes dans les graines (Hejgaard, 2001; Hejgaard et Hauge, 2002; Roberts et al., 2003; Ostergaard et al., 2004) et pouvant représenter jusqu'à 5% des protéines de l'endosperme (Hejgaard, 1982). Une serpinine, CmPS-1, a aussi été rapportée dans le phloème d'une cucurbitacée *Cucurbita maxima* (Yoo et al., 2000). Toutefois, la fonction des serpinines chez les plantes demeure encore méconnue. Jusqu'à tout récemment, aucune protéase cible n'avait été identifiée et les premières serpinines végétales étaient donc associées à un rôle de défense *via* l'inhibition de protéases des insectes herbivores et des agents pathogènes (Ostergaard et al., 2000), tel que supposé en général, aussi, pour les inhibiteurs de protéases à mécanisme standard. Toutefois, une serpinine d'*A. thaliana*, la protéine AtSerp1, a été reconnue récemment comme un inhibiteur de la métacaspase 9 d'*A. thaliana* (Vercammen et al., 2006). Il est suggéré que les métacaspases, des protéases de type cystéine, sont impliquées dans la mort cellulaire programmée et dans les réactions d'hypersensibilité, élargissant ainsi les fonctions potentielles des serpinines végétales à la régulation de la mort



cellulaire. Par ailleurs, la capacité d'inhiber des protéases de type cystéine, propriété rencontrée chez certaines serpinines de mammifères (Schick et al., 1998; Irving et al., 2002), illustre aussi la grande plasticité fonctionnelle de ces protéines. Les serpinines végétales montrent un mécanisme d'inhibition irréversible, propre aux serpinines, suite à la formation d'un lien acyl-covalent avec la protéase cible (Dahl et al., 1996; Rasmussen et al., 1996). Aucune caractérisation structurale des serpinines végétales n'a encore été réalisée, mais la similitude dans les mécanismes d'inhibition laisse supposer une structure tertiaire similaire. Les éléments de structure secondaire prédites *in silico* pour la protéine AtSerp1 montrent en outre une forte homologie avec la structure secondaire de l' $\alpha$ -1 PI (Vercammen et al., 2006). Les cellules végétales synthétisent donc leurs propres serpinines endogènes. Auraient-elles, dès lors, la capacité de produire des serpinines recombinantes actives?

### **1.3.4 La production de serpinines recombinantes**

La complexité des serpinines, tant dans leur structure que dans leur mécanisme d'inhibition, les rend particulièrement vulnérables sur le plan fonctionnel. Des mutations ponctuelles dans la séquence primaire peuvent altérer directement l'efficacité d'inhibition ou affecter la conformation et conduire à une transition structurelle vers des formes polymériques inactives. Les serpinines étant impliquées dans des processus physiologiques essentiels, une telle altération de leur activité conduit fort souvent à un état pathologique. L'administration de serpinines exogènes est une avenue potentielle de traitement, notamment pour le traitement de l'emphysème ( $\alpha$ 1-PI; Stoller et al., 2002) et de l'angiodème héréditaire (inhibiteur C1; Kaplan, 2001). Les protéines utilisées pour les traitements sont le plus souvent purifiées à partir du plasma, mais elles sont souvent obtenues en faibles quantités et disponibles à un coût élevé en raison de leur purification longue et fastidieuse (Karnaukhova et al., 2006). De plus, de telles préparations posent des risques de contamination par des agents pathogènes transmissibles à l'être humain. À l'échelle expérimentale, plusieurs plateformes pour la production de serpinines hétérologues ont été exploitées (Bos et al., 2003; Jayakumar et al., 2004; Pemberton et Bird, 2004; Karnaukhova et al., 2006), mais elles souffrent souvent de problèmes liés à l'insolubilité des serpinines, à leur dégradation, à des modifications post-traductionnelles absentes ou non optimales, ou à l'obtention de serpinines latentes non actives. En pratique, un système hétérologue adéquat devra non seulement contourner ces contraintes, mais aussi

permettre des rendements importants et une éventuelle baisse des coûts de production (Karnaukhova et al., 2006), des défis technologiques de taille mais néanmoins réalistes pour les plateformes végétales actuellement disponibles.

#### **I.4. Hypothèses et objectifs**

Considérant le potentiel des plantes pour la production de protéines complexes glycosylées, nous avons étudié ici la capacité d'un système végétal à produire des serpins recombinantes stables et actives. Tel que mentionné précédemment, nous avons choisi l' $\alpha$ 1-antichymotrypsine humaine (AACT, ou SERPINA3) comme protéine modèle. L'AACT est une serpine plasmatique synthétisée principalement par les cellules hépatiques et les cellules épithéliales des bronches; chez les mammifères, des isoformes de l'AACT sont aussi produites en plus faibles quantités dans le cerveau (Hwang et al., 1999). *In vivo*, l'AACT joue un rôle anti-inflammatoire en inhibant les protéases chymotrypsine, cathepsine G et chymase. La structure de l'AACT, similaire à celle des autres serpins, contient neuf hélices  $\alpha$  et trois feuillets  $\beta$  (Wei et al., 1994). Sur le plan médical, une déficience en AACT peut conduire à différents troubles physiologiques, notamment des désordres respiratoires (bronchite chronique, emphysème et *chronic obstructive pulmonary disease*) et des problèmes chroniques au foie (Chang et Lomas, 1998; Duranton et al., 2000; Yoon et al., 2002).

L'hypothèse principale de mon projet stipulait que les plantes, comme plate-forme de production hétérologue, permettent la production efficace d'AACT stable et active. Comme la littérature actuelle suggère que tant la qualité que l'abondance des protéines recombinantes exprimées dans les cellules eucaryotes varient en fonction du compartiment cellulaire d'accumulation, j'ai donc formulé des sous-hypothèses de travail, ensuite vérifiées en cours de ce projet. Plus spécifiquement, ces sous-hypothèses stipulaient que :

1. Le site d'accumulation cellulaire influence la qualité et le rendement en AACT recombinante dans les cellules végétales. Cette hypothèse apparaissait d'autant plus intéressante que l'AACT est une glycoprotéine dont la structure des glycannes était destinée à varier en fonction du compartiment cellulaire d'accumulation.
2. Le niveau et la structure de la glycosylation affectent les propriétés biologiques de l'AACT. Environ 25% du poids moléculaire de l'AACT humaine (65 kDa) est

expliqué par la présence de glycannes. Dans sa séquence primaire, l'AACT comporte six sites potentiels de N-glycosylation, mais aucune étude n'a toutefois encore déterminé la localisation, la structure ou la fonction des glycannes présents. Fait à noter, l'AACT a déjà été produite sous une forme active dans *E. coli* (Rubin et al. 1990), malgré l'absence de glycannes. Aucune étude n'a par contre documenté l'impact de la N-glycosylation sur les propriétés de l'AACT, et cette protéine n'a pas été non plus produite dans les plantes. La seule étude portant sur la synthèse de serpins recombinantes dans les plantes visait à produire l' $\alpha$ 1-PI recombinante dans les cellules d'*Oryza sativa* (Terashima et al., 1999b; Huang et al., 2001). L'inhibiteur glycosylé obtenu était actif mais sa demi-vie dans le plasma était grandement réduite, en comparaison de l'inhibiteur exprimé naturellement chez les mammifères.

Afin de vérifier l'ensemble de ces hypothèses, trois objectifs principaux ont été poursuivis :

1. Produire de l'AACT dans des cultures cellulaires de tabac *N. tabacum*, variété BY-2. Ce système d'expression hétérologue a été choisi pour la facilité et la rapidité avec lesquelles il est possible de le transformer à l'aide d'*Agrobacterium tumefaciens*, pour la croissance et la division rapides de ses cellules, et pour la facilité à obtenir une biomasse importante (Hellwig et al., 2004). L'expression de l'AACT a été ciblée ici vers quatre compartiments cellulaires : le cytoplasme, le réticulum endoplasmique (RE), la vacuole et l'apoplasme.
2. Étudier l'expression et la localisation de l'AACT dans les lignées cellulaires modifiées par immunolocalisation et microscopie confocale à fluorescence.
3. Analyser les modifications post-traductionnelles des variantes de l'AACT produites d'un compartiment cellulaire à l'autre, en particulier la glycosylation et la maturation protéolytique. Des expériences de type *pulse-chase* devaient me permettre de suivre les modifications post-traductionnelles au cours du temps.

Au cours de ces expériences, il est apparu clairement que l'AACT exprimée dans le cytoplasme s'accumulait en fait largement dans le noyau. L'AACT n'ayant pas de signaux d'adressage nucléaire connus, cette localisation était inattendue. Comme l'AACT exprimée dans le cytosol montrait aussi une plus grande homogénéité que les formes adressées vers le système endomembranaire de sécrétion, nous nous sommes finalement intéressés plus spécifiquement à cette forme de l'AACT, avec pour but de comprendre son mécanisme de translocation nucléaire et d'altérer, si possible, son transit vers le noyau et de la retenir dans le cytosol. L'AACT possède un site de liaison à l'ADN (Naidoo et al., 1995), ce qui m'a amené à poser une nouvelle hypothèse : le site de liaison à l'ADN est impliqué dans la localisation nucléaire de l'AACT. Pour tester cette hypothèse, une nouvelle variante de l'AACT, mutée au niveau de son site de liaison à l'ADN, a été conçue, introduite par agrotransfection dans des feuilles de *N. tabacum* et exprimée transitoirement pour être localisée dans les cellules par microscopie. Des formes recombinantes de l'AACT, mutées ou non, ont également été exprimées dans *E. coli* afin de comparer leur capacité de liaison à l'ADN et leur efficacité à inhiber la chymotrypsine, une enzyme cible de cette protéine chez les mammifères.

## **CHAPITRE II**

**Targeted expression and maturation of human  $\alpha$ 1-  
antichymotrypsin in tobacco BY-2 cultured cells**

## II.1 Mise en contexte

Produire des protéines recombinantes dans les plantes reste, malgré plusieurs succès, un défi de taille, en particulier pour les protéines complexes nécessitant des modifications post-traductionnelles. Afin d'étudier la capacité des systèmes de production végétaux à synthétiser de telles protéines, nous nous sommes concentrés sur la production d'une serpine recombinante, l' $\alpha$ 1-antichymotrypsine (AACT) humaine, dans une suspension cellulaire de tabac *N. tabacum* cv BY-2. Nous voulions évaluer la possibilité d'exprimer une glycoprotéine complexe dans des cellules végétales et caractériser les modifications post-traductionnelles qui y seraient apportées. Cette première partie de la thèse décrit les étapes menant à l'obtention des lignées cellulaires de tabac transgéniques exprimant l'AACT, et la caractérisation des protéines recombinantes ainsi obtenues.

Plusieurs évidences dans la littérature scientifique démontrent l'importance de la localisation cellulaire de la protéine recombinante sur le rendement et les modifications post-traductionnelles. Dans cette optique, nous avons exprimé l'AACT dans différents compartiments cellulaires (le RE, le cytosol, la vacuole et l'apoplasme) à l'aide de signaux de rétention et d'adressage adjoints à la séquence primaire de la protéine. Par immunolocalisation et microscopie confocale à fluorescence, nous avons ensuite évalué l'expression et la localisation cellulaires des différentes formes d'AACT produites dans les cellules végétales. Enfin, les modifications post-traductionnelles faites sur la protéine ont été évaluées par des études de type *pulse-chase* et par une étude des processus de la glycosylation et de dégradation protéolytique survenus sur la protéine durant son transit cellulaire et son accumulation *in planta*. Les résultats de ces travaux sont présentés ici sous la forme d'un article scientifique qui sera soumis sous peu pour publication dans une revue scientifique.

## Targeted expression and maturation of human $\alpha$ 1-antichymotrypsin in tobacco BY-2 cultured cells

Meriem Benchabane <sup>1,2</sup>, Claude Saint-Jore Dupas <sup>1</sup>, Muriel Bardor <sup>1</sup>, Loïc Faye <sup>1</sup>,  
Dominique Michaud <sup>2</sup> and Véronique Gomord <sup>1,\*</sup>

<sup>1</sup> CNRS UMR 6037, IFRMP 23, GDR 2590, Université de Rouen, Faculté des Sciences  
Bât. Ext. Biologie, 76821 Mont-Saint-Aignan cedex, France

<sup>2</sup> Département de phytologie, CRH/INAF, Université Laval, Québec (QC) Canada G1K  
7P4

\* To whom correspondence should be sent.

**Keywords:**  $\alpha$ 1-antichymotrypsin, serpins, plant-made pharmaceuticals, N-glycosylation, recombinant protein, protein maturation, proteolytic processing, post-translational modifications

**Abbreviations:** AACT,  $\alpha$ 1-antichymotrypsin; hAACT, human AACT; brAACT, recombinant AACT expressed in *E. coli*; CT, chymotrypsin; HNE, human neutrophil elastase; RCL, reactive center loop; ER, endoplasmic reticulum; KDEL, KDEL (Lys-Asp-Gln-Leu) peptidic signal for protein retention in the ER; BY-2, *Nicotiana tabacum* cv. Bright Yellow-2 cells, NPT-II, neomycin phosphotransferase II; PCR, polymerase chain reaction; RT, reverse transcriptase; PNGaseF, peptide N-glycosidase F; BiP, binding protein.



## II.2 Résumé

L'alpha 1-antichymotrypsine (AACT) humaine a été exprimée dans des cellules de tabac BY-2, et ses modifications post-traductionnelles ont été étudiées en fonction du site d'adressage intracellulaire considéré. Tel qu'observé par des analyses d'expression de type *pulse-chase* et des observations en microscopie à immunofluorescence, l'addition d'un signal peptidique N-terminal pour l'insertion co-traductionnelle de l'AACT dans le sentier de sécrétion cellulaire a engendré sa sortie dans le milieu de culture. Des formes d'AACT dirigées vers la vacuole ou le réticulum endoplasmique (RE) ont aussi été sécrétées, présumément en raison d'une élimination par clivage protéolytique des signaux d'adressage adjoints en région C-terminale. De manière inattendue, l'AACT dirigée vers le cytosol s'est accumulée surtout dans le noyau, sous forme stable et non glycosylée. À l'inverse, les formes d'AACT portant un signal de sécrétion ont subi de manière systématique des modifications post-traductionnelles multiples en cours de migration dans la cellule. L'AACT sécrétée était systématiquement N-glycosylée, menant à la présence de N-glycannes de structures complexes comparables aux N-glycannes observés dans les cellules humaines. Des événements de maturation protéolytique ont aussi été observés pour l'ensemble des formes sécrétées, non seulement pendant leur transit cellulaire mais aussi après leur sécrétion dans le milieu de culture. En somme, l'adressage intracellulaire de l'AACT humaine dans le cytosol ou dans le sentier de sécrétion des cellules de tabac BY-2 a permis la production de deux formes protéiques principales: (i) une forme simple non glycosylée, accumulée sous forme stable dans le compartiment nucléaire; et (ii) un ensemble hétérogène de variants sécrétés résultant de modifications post-traductionnelles complexes incluant l'addition de N-glycannes à chaînes complexes et une dégradation protéolytique partielle du polypeptide menant à la sécrétion de versions tronquées dans le milieu extracellulaire. Des études additionnelles sont maintenant requises afin d'évaluer l'activité de ces différents variants AACT et leur efficacité relative en contexte thérapeutique.

## **Summary**

Human  $\alpha$ 1-antichymotrypsin (AACT) was expressed in BY-2 tobacco cultured cells, and its post-translational modifications assessed in relation with the subcellular compartment targeted for accumulation. As assessed by pulse-chase expression experiments and immunofluorescence microscopy, fusing an N-terminal signal peptide to AACT for co-translational transfer into the cell secretory pathway led the protein to be secreted in the cell culture medium. AACT variants targeted to the vacuole or the endoplasmic reticulum (ER) were also secreted in the extracellular milieu, presumably as a result of proteolytic processing events removing ER-retention and vacuole-sorting targeting signals included at the C terminus. Unexpectedly, AACT expressed in the cytosol was found predominantly in the nucleus under a stable, non-glycosylated form. By contrast, AACT variants bearing an N-terminal signal peptide were subject to multiple post-translational modifications during their transit along the cell secretory pathway. Secreted forms of AACT were N-glycosylated in all cases, with the presence of complex glycans comparable to the N-glycans typically observed in human cells. Proteolytic processing events were also observed for all variants, not only during their transit through the secretory pathway, but also in the culture medium after extracellular secretion. Overall, the intracellular targeting of human AACT in different compartments of BY-2 tobacco cells allowed for the production of two protein forms: (i) a single, non glycosylated variant accumulated under a stable form in the nuclear compartment; and (ii) a heterogeneous complement of secreted variants resulting from complex post-translational modifications including N-glycosylation for the attachment of complex glycan chains and proteolytic processing leading to the secretion of truncated polypeptides in the extracellular medium. Additional studies are now required to assess the activity of these different AACT variants, and to understand the molecular determinants for the nuclear localization of AACT expressed in the cytosol.

### **II.3 Introduction**

Heterologous protein expression systems involving different eucaryotic organisms have been devised for the production of several proteins subject to complex post-translational modifications (Yin et al., 2007). Whereas each expression system may perform specific post-translational modifications leading to the accumulation of specific structural variants, most systems allow for the production of protein products with suitable stability and biological activity (Walsh and Jefferis, 2006). In practice, a close resemblance between the recombinant protein and its natural counterpart may represent an advantage as the rapid acceptance of recombinant therapeutic proteins by health authorities relies in part on product equivalence. On the other hand, emphasis must be placed above all on the activity and pharmacokinetic features of the recombinant protein, which suggests that each production platform would have its own advantages and drawbacks depending on each particular protein to express (Sethuraman and Stadheim, 2006; Saint-Jore-Dupas et al., 2007). Glycosylated interferon  $\beta$  produced in Chinese hamster ovary cells, for instance, was ten times more active than a non-glycosylated form produced in *E. coli* (Runkel et al., 1998), while a recombinant, non-glycosylated form of human erythropoietin expressed in bacteria showed high *in vitro* activity compared to a glycosylated homologue targeted to the secretory pathway (Higuchi et al., 1992).

From a clinical perspective, plants show potential as protein expression hosts for a number of reasons, notably including low production costs, the absence of human pathogenic contaminants in protein preparations, and the well documented ability of plant cells to perform post-translational modifications typical of mammalian proteins (Daniell et al., 2001; Ma et al., 2003; Gomord and Faye, 2004; Faye et al., 2005). Practical challenges, now, to support the potential of plants as protein biofactories include a significant improvement of final yields and a better understanding of protein instability and/or deficient activity arising from non-optimal post-translational maturation *in planta* (Ma et al., 2003). Disappointing yields observed in recent years for some proteins expressed *in planta* (e.g., Matsumoto et al., 1995; Sharp and Doran, 2001; Outchkourov et al., 2003), while suggesting a certain inherent inefficiency of plant systems for protein production, maybe simply reflect the high complexity of protein structures and biochemical processes in plant cells. Understanding the complex post-translational modifications of recombinant

proteins *in planta*, in relation with their structures and functions, is essential at this stage to clearly establish the potential of plants as efficient hosts for clinically-useful proteins.

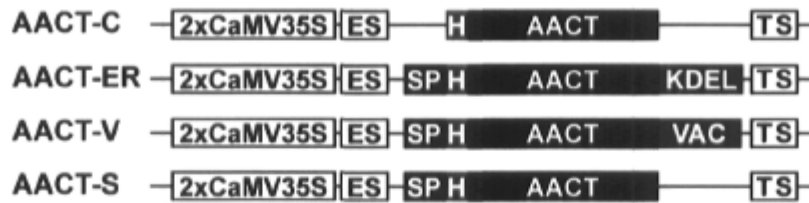
N-glycosylation and proteolysis, in particular, are two post-translational processes requiring special attention (Faye et al., 2005). N-glycosylation is a common protein post-translational process in eucaryotic cells, starting in the ER by the co-translational transfer of an oligosaccharide precursor on the nascent protein backbone at specific Asn residues of the consensus sequence motif Asn–X–Thr/Ser. As the protein migrates through the secretory pathway, the N-glycan is processed by the addition or removal of sugar residues by ER and Golgi processing enzymes. Whereas precursor oligosaccharides and initial maturation steps in the ER are common to all eucaryotes, later steps of N-glycan maturation along the Golgi apparatus are species-specific, giving rise to complex glycans [and protein products] diverging among species. In practice, these differences represent a potential limitation for clinically-useful glycoproteins produced in plants, as plant glycoepitopes are immunogenic in most animal models and elicit the production of glycan-specific antibodies in humans (van Ree et al., 2000; Bardor et al., 2003; Gomord et al., 2005). Such responses in human patients might eventually increase the risk of allergenic reactions or induce a rapid immunological clearance of the therapeutical protein *in vivo*, with a negative impact on its usefulness or effectiveness *in vivo*.

Proteolysis also represents a hurdle for the successful expression of recombinant proteins *in planta* (Doran, 2006; Goulet and Michaud, 2006). Whereas the proteolytic removal of pre- or pro-regions is essential for the correct maturation of secreted proteins, unintended proteolysis may affect the yield and stability of several heterologous proteins (Faye et al., 2005). A number of studies focused in recent years on devising strategies to minimize unintended proteolysis in plants (Rivard et al., 2006; Streatfield, 2007; Benchabane et al., 2007), but empirical data with structurally complex proteins are still needed for the fine-tuning of expression strategies allowing stable and fully active proteins to be isolated. With the aim of further understanding glycosylation and proteolytic processes impacting mammalian proteins in the plant cell environment, we expressed here the model protein, human  $\alpha$ 1-antichymotrypsin (AACT), in different compartments of Bright Yellow-2 (BY-2) tobacco cultured cells.

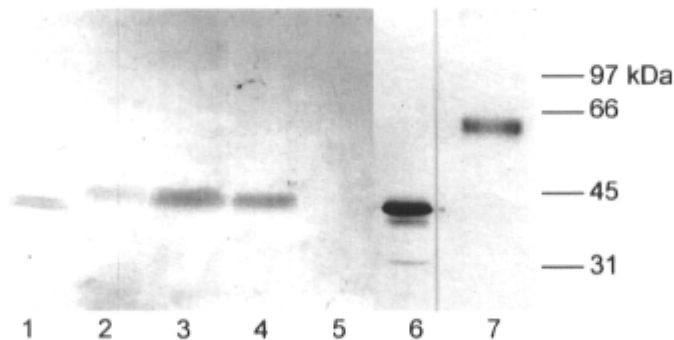
AACT is a member of the serpin protein superfamily, that includes inhibitors of serine proteases involved in diverse physiological, cellular and biochemical processes including coagulation, inflammation, fibrinolysis, apoptosis, neoplasia, viral pathogenesis, protein folding and complement activation (Silverman et al., 2001). Because dysfunctional serpins have been associated with severe diseases (Stein and Carrell, 1995; Chang and Lomas, 1998), recombinant serpins were readily proposed as useful candidates for replacement therapies in humans (Rubin, 1992). AACT is a glycoprotein with six putative N-glycosylation sites and no disulfide bond, showing the typical tertiary structure of serpins with nine  $\alpha$ -helices surrounding three  $\beta$ -sheets and a mobile inhibitory loop referred to as the reactive center loop, or RCL (Rubin et al., 1990; Wei et al., 1994). Glycosylation is non essential for AACT protease inhibitory activity (Rubin et al., 1990), in contrast with the tertiary fold of the polypeptide backbone which is critical for serpin bioactivity and stability. From a practical perspective, AACT is an interesting model to study the potential of plants as expression platforms for the production of structurally complex human proteins in both their glycosylated and unglycosylated forms. We monitored here the maturation, stability and homogeneity of human AACT variants expressed in the cytosol and the secretory pathway of BY-2 tobacco cells.

## **II.4 Results**

Human AACT was transiently expressed in *Nicotiana tabacum* leaf epidermal cells, and then stably in *N. tabacum* cv. BY-2 cultured cells, to assess the potential of plant cells as a suitable platform for the expression of recombinant serpins. Transgenic material was generated by *Agrobacterium tumefaciens*-mediated transformation using different pCambia vectors containing a cDNA coding sequence for AACT and appropriate retention/targeting signals for differential subcellular targeting (**Figure II.1**). Four cell/plant lines were generated for the experiments, targeting AACT to either the cytosol (C), the ER, the vacuole (V) or the apoplast (S, for secreted). The signal peptide of an alfalfa protein disulfide isomerase (PDI) was fused at the N terminus of AACT to direct the co-translational import of the protein into the secretory pathway via the ER membrane. A C-terminal KDEL tetrapeptide was used for retention and accumulation of the protein into the ER. AACT delivery to the vacuole was achieved by fusing the vacuolar sorting signal of tobacco chitinase A at the C terminus of the recombinant inhibitor.



**Figure II.1 Gene constructs for targeted expression of human antichymotrypsin (AACT) in tobacco cells.** Gene constructs were devised for the expression of AACT in the cytosol (AACT-C construct), the ER (AACT-ER), the apoplast (AACT-S) and the vacuole (AACT-V). Each construct include a duplicated version of the cauliflower mosaic virus 35S constitutive promoter (2xCaMV35S), the tobacco etch virus enhancer sequence (ES), a poly-histidine tag (H), the human AACT-encoding sequence and the pA35S terminator sequence (TS). Constructs devised for AACT secretion also included the coding sequence of the protein disulfide isomerase (PDI) signal peptide (SP). A tetrapeptidic sequence KDEL for protein retention into the ER and the tobacco chitinase A vacuolar targeting sequence (VAC) were also added in the AACT-ER and AACT-V constructs, respectively.

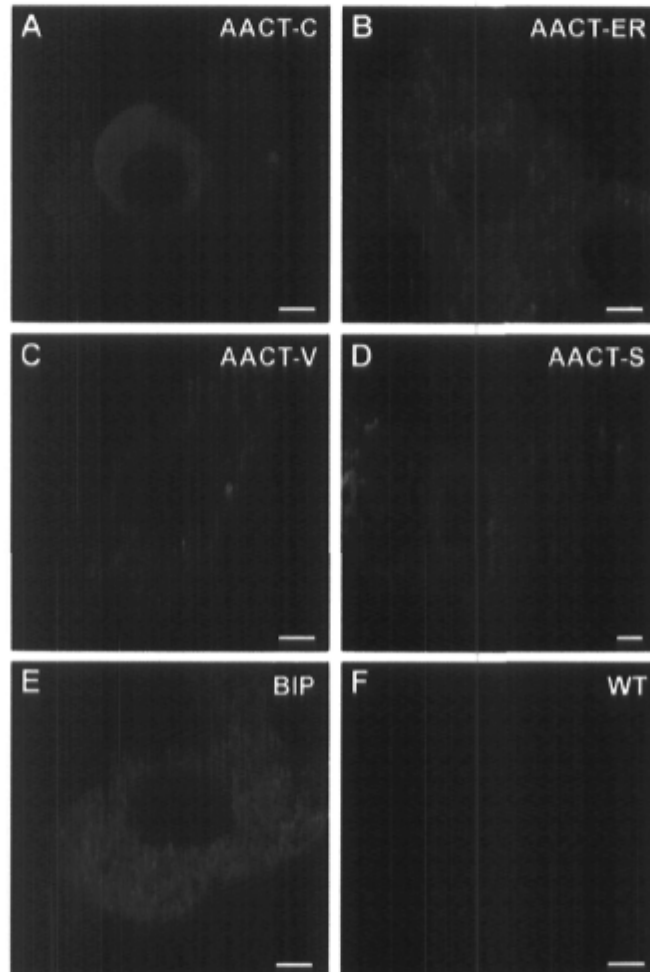


**Figure II.2 Transient expression of AACT in tobacco leaves.** Tobacco leaf epidermal cells were transiently transformed with the *aact* transgene by *Agrobacterium* infiltration. Protein extracts from tobacco leaf epidermal cells transformed by AACT-C (lane 1), AACT-ER (lane 2), AACT-S (lane 3) or AACT-V (lane 4) were resolved by 12% (w/v) SDS-PAGE and immunodetected with monoclonal anti-AACT antibodies after their transfer onto a nitrocellulose membrane. Controls included a protein extract from untransformed tobacco leaves (lane 5), purified bacterial AACT (lane 6) and human glycosylated AACT (lane 7).

#### II.4.1 Expression and subcellular distribution of human AACT in tobacco cells

AACT expression was first monitored following transient expression in *Agrobacterium*-infiltrated tobacco leaves. The recombinant inhibitor was immunodetected as polypeptides of around 45 kDa in crude protein extracts from tobacco leaf epidermal cells transformed with either the AACT-C, AACT-ER, AACT-V or AACT-S constructs (**Figure II.2**, lanes 1-4). Although recombinant forms targeted to the secretory pathway (lanes 2-4) showed an overall electrophoretic mobility closer to the mobility of unglycosylated AACT expressed in bacteria (lane 6) –compared to the glycosylated human AACT (lane 7), their molecular masses remained slightly higher than the mass observed for AACT retained in the cytosol (lane 1), suggesting the onset of glycosylation.

To test this possibility, stably transformed BY-2 tobacco cells expressing the different constructs were generated as described previously (Gomord et al., 1998). Transgenic calli were first screened for AACT expression by RT-PCR and immunodetection with AACT-specific antibodies to confirm the presence of AACT in tobacco cells. In brief, AACT protein patterns for the different BY-2 cell lines were similar to the patterns described above for transient expression in tobacco leaves (data not shown). High-expression calli from each transformation were selected to initiate suspension cultures, and then used for immunolocalization studies by confocal microscopy using anti-AACT antibodies incubated on fixed, permeabilized transgenic BY-2 cells. As illustrated on **Figure II.3**, all transgenic cell lines expressed AACT variants (**Figures II.3A-II.3D**), in contrast with wild-type cells showing no fluorescent signal (**Figure II.3F**). A typical ER labeling was observed for AACT-ER and, to a lesser extent, for AACT-S, thereby confirming the presence of AACT in this compartment following KDEL-dependent accumulation in the ER (AACT-ER) or during its migration toward the apoplast (AACT-S) (**Figures II.3B and II.3D**). Immunostaining patterns obtained for AACT-ER and AACT-S were similar to the pattern observed for BiP, an ER-resident chaperone used here as a specific marker for the ER compartment (**Figure II.3E**). By comparison, vacuolar staining observed in BY-2 cells harboring the AACT-V construct illustrated the efficiency of the chitinase A vacuolar sorting signal for protein targeting to the vacuole (**Figure II.3C**). Unexpectedly, AACT expressed with no signal peptide for cellular secretion was mainly localized in the nucleus, with only a faint staining of the cytosol (**Figure II.3A**).



**Figure II.3 Immunolocalization of recombinant AACT expressed in different subcellular compartments of tobacco BY-2 suspension cultured cells.** Appropriate retention and targeting signals were used to target AACT in the cytosol (A), the ER (B), the vacuole (C) or the apoplast (D). Immunolocalization was performed on 3-days-old BY-2 cells using polyclonal anti-AACT antibodies and FITC-conjugated secondary antibodies. The protein BiP was immunolocalized with specific anti-BiP antibodies as an ER positive control, in wild type cells, using the same protocol (E). Wild-type cells are presented in (F). Bars = 8 $\mu$ m.



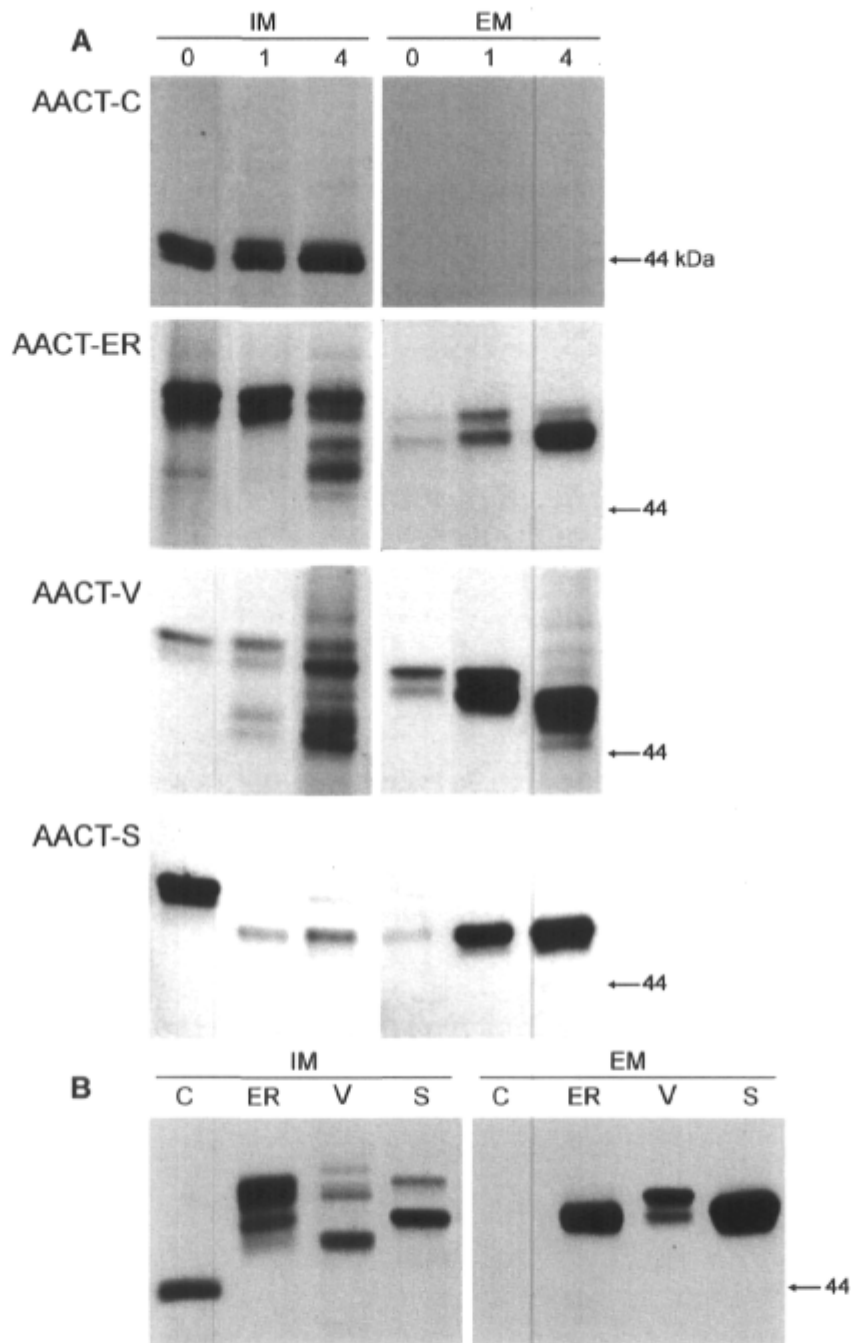
#### **II.4.2 Secreted AACT is subject to extensive post-translational maturation in the tobacco cell secretory pathway**

Immunocytolocalization data were giving an idea of the subcellular compartmentalization of AACT variants in BY-2 cells, but this approach could not allow for the detection of AACT secreted in the culture medium of BY-2 cells. In order to describe more precisely the distribution and maturation of AACT variants at the cell level, we combined here cell fractionation and pulse-chase experiments involving the expression of radio-labeled AACT. Tobacco cells were pulse-labeled for 1 h with [<sup>35</sup>S]-methionine and [<sup>35</sup>S]-cysteine, and then chased for 0, 1 or 4 h before electrophoretic migration and radiodetection (**Figure II.4**). Polypeptides antigenically related to AACT were then immunoprecipitated from the intra- (IM) and extracellular (EM) media (or fractions), and analyzed by fluorography following SDS-PAGE. Two stable polypeptides about 44 kDa in size were detected in the IM of AACT-C cell lines at the end of the labelling period (**Figure II.4A**) or after a 24-h chase period (data not shown). No AACT was detected in the EM of AACT-C calli even after a 4-h chase, clearly suggesting that the EM was not contaminated by intracellular proteins during the fractionation procedure (**Figure II.4A**).

By contrast, recombinant AACT variants targeted to the secretory pathway were synthesized as ~54-kDa polypeptides rapidly matured within the transformed cells. For instance, recombinant AACT expressed in the AACT-S cell line was rapidly matured within the IM to a single isoform of ~48 kDa, then secreted almost completely in the EM after a 4-h chase. The intracellular maturation of AACT targeted to the vacuole or the ER was much more complex, with four major AACT variants appearing successively in the IM and only two intermediate forms detected in the EM. Surprisingly, all AACT variants harbouring a signal peptide for intracellular secretion were detected mainly in the EM at the end of the chase period, independent of the presence or absence of additional targeting signals. Leakage of KDEL-tagged AACT to the EM could have resulted from a low efficiency of the ER retention machinery as observed earlier for other recombinant proteins (Henderson et al., 1997; Okamoto et al., 1999), or from the cleavage of the KDEL tetrapeptide in the ER (see Badri et al., 2007b). AACT was more rapidly translocated to the EM in the AACT-S cell line, which confirms a certain functionality of the ER retrieval tag, as illustrated above from subcellular immunolocalization studies (**Figure II.3**). AACT was also accumulated in the EM of the AACT-V cell line, again suggesting the proteolytic removal of sorting signals

along the secretory pathway. Major forms of AACT with similar electrophoretic mobilities were detected at the end of the chase period in the EM of the AACT-V, AACT-ER and AACT-S cell lines, which could be explained by common processing steps along a unique intracellular pathway. The rapid processing of high-molecular-weight AACT precursors detected in the EM at the end of the pulse period probably explains the absence of AACT variants larger than 44 kDa on immunoblots performed with total protein extracts from tobacco leaves transiently expressing AACT (**Figure II.2**) or with stably transformed BY-2 tobacco cells (data not shown). AACT maturation was also not limited to the IM, as the ~50-kDa form detected in the EM of AACT-ER and AACT-S cell lines tended to disappear with time, along with the concomitant accumulation of a ~48-kDa polypeptide likely corresponding to a truncated variant of glycosylated AACT.

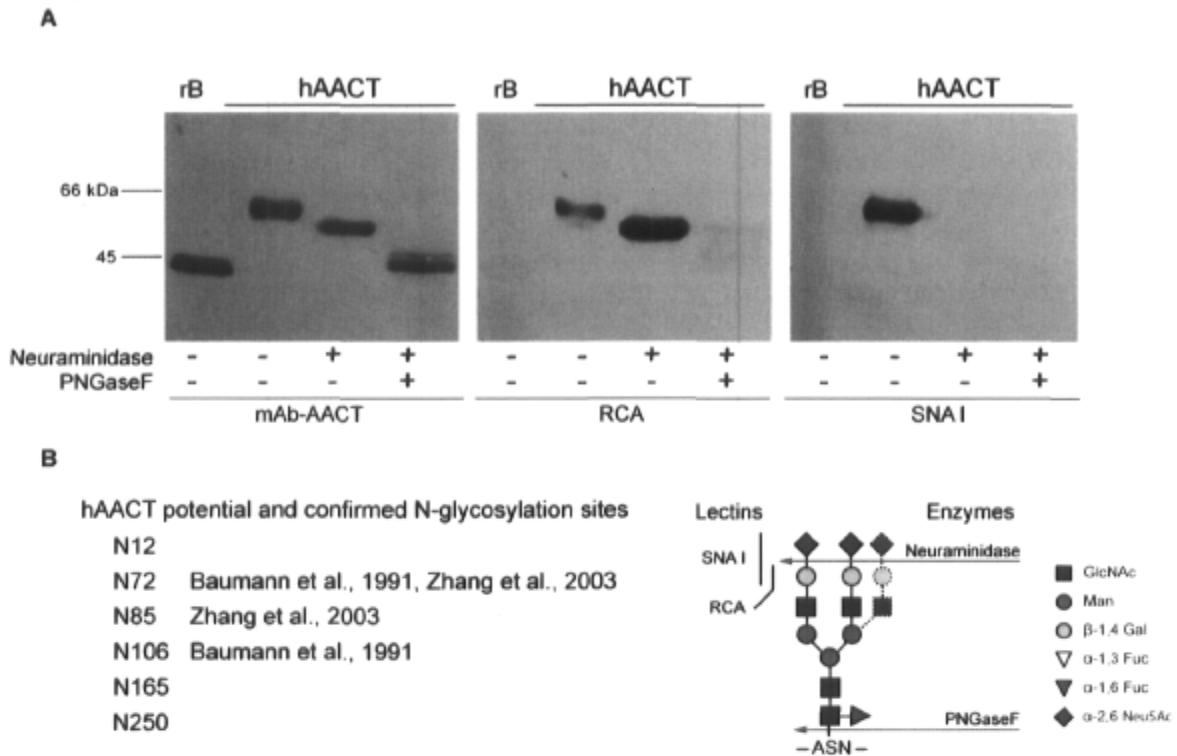
A similar maturation pattern was observed in the IM of AACT-V and AACT-ER cell lines. A retarded transit along the secretory pathway, partial targeting to the vacuole for the AACT-V cell line (**Figure II.3C**), or recycling of the protein to the ER for the AACT-ER cell line probably accounted for the multiple AACT forms detected in the IM of these cell lines. Although a significant amount of AACT was directed to the vacuole (AACT-V cells) or accumulated into the ER (AACT-ER cells) (see **Figure II.3**), AACT was mainly detected in the EM in both cases (**Figure II.4**), which suggests either a lack of efficiency of the targeting machinery, or –more likely– a partial cleavage of the C-terminal targeting signals.



**Figure II.4 Post-translational maturation of recombinant AACT in BY-2 cultured cells.** **A.** Pulse-chase experiment for the expression of AACT in different subcellular compartments of BY-2 cells, including the cytosol (C), the ER (ER), the vacuole (V) and the apoplast (S). The cell lines were subject to metabolic pulse radiolabeling for 60 min ( $^{35}\text{S}$ -Met and  $^{35}\text{S}$ -Cys) and varying chase periods (0, 1 and 4 h). Recombinant AACT was immunoprecipitated from the intracellular medium (IM) or extracellular medium (EM) and analyzed by SDS-PAGE and fluorography. **B.** Maturation of AACT in AACT-C, AACT-ER and AACT-S cell lines after a 60-min pulse production and a 4-h chase period.

### II.4.3 Mature forms of plant recombinant AACT harbour complex N-glycans

Due to the heterogeneous distribution of AACT in AACT-V and AACT-ER cell lines and to the high number of AACT variants detected, we focused on the AACT-S cell line for further characterization of the recombinant inhibitor. Pulse-chase experiments were suggesting that AACT in this line was post-translationally matured by a combination of N-glycosylation and proteolytic processing events. Since little is known about the structure and nature of human AACT, we first assessed the glycosylation status of the native protein purified from human plasma. AACT bears six potential consensus sequences for N-glycosylation, with three or four of them presumably occupied by N-glycans on the human plasmatic protein (Baumann et al., 1991; Gettins et al., 1996; Zhang et al., 2003) (**Figure II.5B**). As observed on **Figure II.5**, human AACT presents an apparent molecular mass of ~58 kDa following SDS-PAGE, with N-glycans presumably accounting for ~25% of this mass. At least some of these glycans bear terminal  $\alpha$ -2,6-N-acetylneuraminic acid (Neu5Ac) residues linked to  $\beta$ -1,4 galactose (Gal) residues, as demonstrated by specific labelling with the *Sambucus nigra* SNA I lectin. In addition, neuraminidase digestion eliminated terminal Neu5Ac residues and exposed  $\beta$ -1,4Gal residues, creating N-glycans on AACT with specific affinity for *Ricinus communis* agglutinin. Together, these results indicate the presence of complex N-glycans typical of mammalian glycoproteins on human AACT. Complete deglycosylation of the protein with peptide N-glycosidase F resulted in the production of a ~44-kDa aglycosylated form of AACT exhibiting electrophoretic mobility similar to the non-glycosylated form of AACT produced in *E. coli*. Considering that N-glycans account for at least 13 kDa of the human AACT molecular mass distributed on three or four glycosylation sites, these observations suggest that N-linked oligosaccharides on the native protein likely present a minimum triantennary structure (**Figure II.5B**).



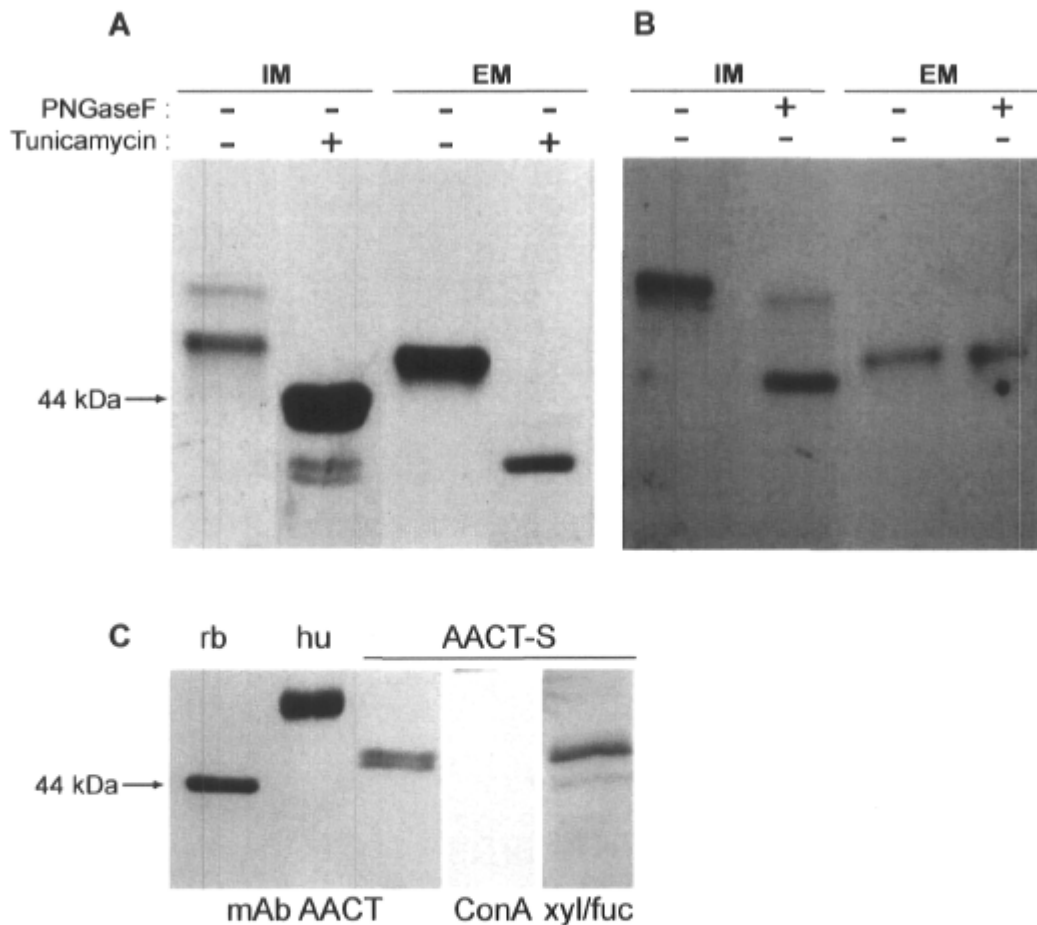
**Figure II.5 Human AACT is highly glycosylated, with complex N-glycans containing sialic acids.** **A.** Human AACT (hAACT) digested with neuraminidase and peptide N glycosidase F (PNGase F). All samples were analyzed by immuno- and affino-detection with anti-AACT specific antibodies (Ab-AACT), *Ricinus communis* lectin A (RCA), and *Sambucus nigra* lectin I (SNA I). Bacterial recombinant AACT (rbAACT) was used as a non-glycosylated control. **B.** Potential N-glycosylation sites of human AACT and putative N-glycan structure of AACT as inferred from exoglycosidase digestion and lectin binding assays.

When expressed in BY-2 cells, AACT targeted to the endomembrane system showed a higher molecular mass than unglycosylated cytoplasmic AACT (**Figures II.2 and II.4**), suggesting that AACT was indeed glycosylated along the secretory pathway. A miscleavage of the signal peptide, although not impossible (Sojikul et al., 2003), is unlikely here as the same signal peptide was correctly trimmed when fused to bovine aprotinin expressed in potato plants (Badri et al., 2007b). It can be observed on **Figure II.2** that plant-expressed recombinant AACT isoforms all exhibited a molecular mass lower than the highly glycosylated human AACT. Pulse-chase experiments on tunicamycin-treated AACT-S cells were carried out (**Figure II.6A**) to investigate the glycosylation pattern of AACT variants. Tunicamycin is an inhibitor of N-glycosylation that prevents the biosynthesis

oligosaccharide precursors. Following this treatment, AACT variants with a lower size were observed, including a major form in the IM showing a molecular mass of 44 kDa, similar to the protein backbone. The detection of two polypeptides with a lower molecular mass could thus not in this case be attributed to N-glycan processing, but instead to proteolytic trimming.

To further characterize the N-glycans of secreted AACT, radiolabelled AACT from AACT-S cells was submitted to enzymatic deglycosylation with peptide N-glycosidase F (PNGase F), which releases N-glycans devoid of  $\alpha$ -1,3 fucose residue on proximal GlcNAc, a residue typically found on plant complex N-glycans (Lerouge et al., 1998). Upon treatment with PNGase F, only the initial ~50-kDa form retained in the IM of AACT-S cell lines (**Figure II.6B**) could be effectively deglycosylated, thus suggesting the presence of oligomannosidic-type N-glycans. By contrast, low-molecular-weight variants detected in the EM after a 4-h chase were not susceptible to PNGase F, clearly showing that N-linked oligosaccharides had been processed by the addition of  $\alpha$ -1,3 fucose in the Golgi apparatus while transiting along the secretory pathway.

To define more precisely the structure of N-glycans on mature AACT from AACT-S cells, the protein was enriched from the culture medium and its N-glycans were analyzed either by affinodetection with the lectin probe concanavalin A or by immunodetection with antibodies specific to  $\beta$ -1,2 xylose and  $\alpha$ -1,3 fucose residues (**Figure II.6C**). Consistent with PNGase F digestion assays (**Figure II.6B**), no glycans were detected with concanavalin A, specific to high mannose-type glycans. On the other hand, the glycans were recognized by antibodies specific to  $\beta$ -1,2 xylose and  $\alpha$ -1,3 fucose residues characteristic of plant complex N-glycans. These results further demonstrate that recombinant AACT forms *in planta* were synthesized as N-glycoproteins in tobacco cells, and that N-glycans were matured to complex N-glycans during the transit of AACT through the cell secretory pathway. Considering that  $\beta$ -1,2 xylose and  $\alpha$ -1,3 fucose residues are immunogenic when administered to laboratory mammals and elicit antibody production in humans (van Ree et al., 2000; Bardor et al., 2003; Gomord et al., 2005), these results may be of considerable significance in a context where plant-made AACT could be used for replacement therapies in humans.



**Figure II.6 Recombinant AACT secreted from AACT-S cells is N-glycosylated and harbours sugar residues typical of plant complex N-glycans.** **A.** AACT-S BY-2 cells were incubated with the N-glycosylation inhibitor tunicamycin (+T), or left untreated (T-), for 90 min prior to a 60 min pulse labelling with  $^{35}\text{S}$ -Met and  $^{35}\text{S}$ -Cys and a 2-h chase in presence or absence of the drug. AACT was immunoprecipitated from the intracellular (IM) or extracellular (EM) protein extracts and analyzed by SDS-PAGE and fluorography. **B.** Peptide N glycosidase F (PNGase F) digestion of AACT-S. AACT-S cells were radiolabelled for 60 min and chased for 2h. AACT was immunoprecipitated from the IM and EM protein extract and submitted to PNGase F digestion (+F) or left untreated (-F). **C.** Bacterial recombinant AACT (rb), human AACT (hu) and tobacco secreted AACT (AACT-S) were immunodetected with anti-AACT antibodies. Polymannosidic N-glycans were affindetected with the specific lectin Concanavalin A (con A) and complex plant N-glycans harbouring  $\beta$ -1,2 xylose and  $\alpha$ -1,3 fucose residues were detected with anti-  $\beta$ -1,2 xylose and anti- $\alpha$ -1,3 fucose antibodies (xyl/fuc).

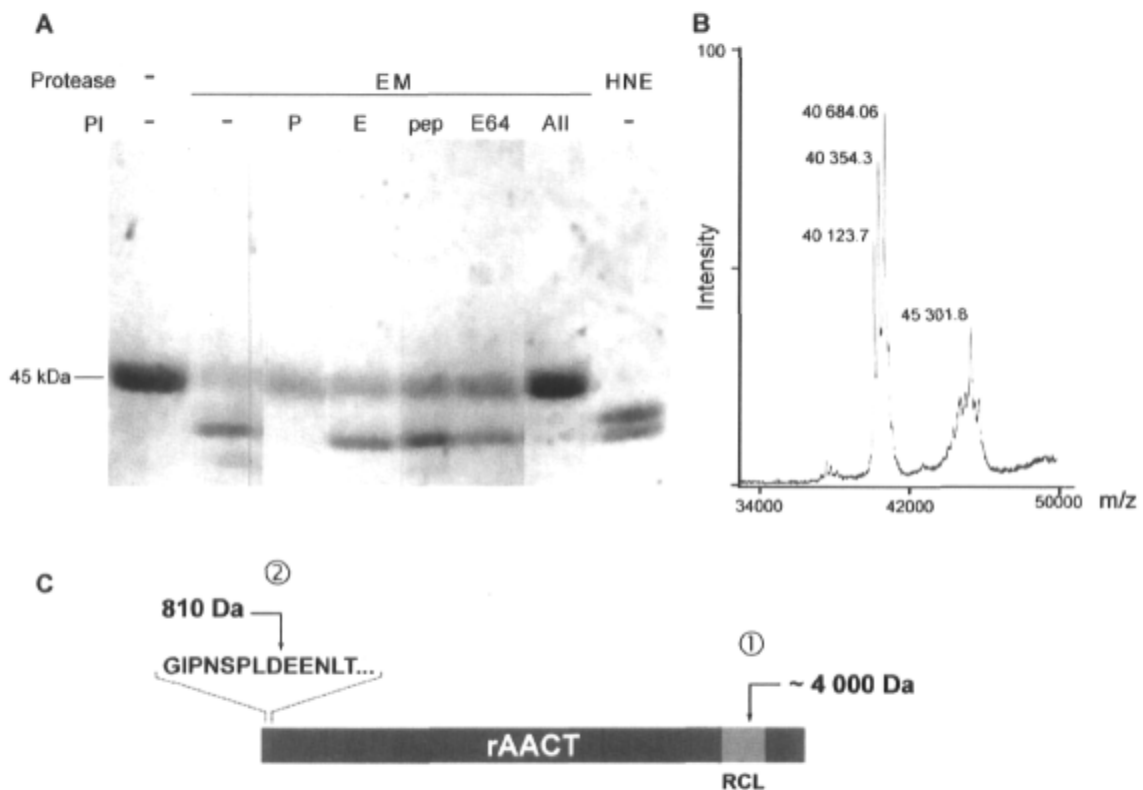
#### II.4.4 Recombinant AACT is proteolytically matured

Whereas pulse-chase analyses have revealed complex N-glycosylation for the different AACT variants, it was unlikely that these variants were solely the result of N-glycan processing. Our results were also suggesting, indeed, a certain role for proteolytic maturation, consistent with the context of a protease-rich plant cell environment. Since most AACT variants targeted to the secretory pathway were found in the EM, spiking experiments with recombinant AACT produced in *E coli* (rbAACT) were conducted to detect proteolytic maturation events that could occur in the culture medium, in the presence of secreted proteases eventually impacting AACT stability. rbAACT was added to the culture medium of a 4-days-old BY-2 cell culture medium, and incubated overnight at room temperature with the recombinant inhibitor to mimic the conditions under which AACT is accumulated in this plant expression platform. A stable truncation product short of 4 to 5 kDa compared to complete AACT was generated during this period (**Figure II.7A**), in contrast with rbAACT spiked in fresh culture medium and incubated under the same conditions, that showed an unaltered migration pattern after the same period. Diagnostic protease inhibitors were used to identify major extracellular protease activities involved in AACT cleavage. The inhibitors E-64, pepstatin and EDTA, respectively specific to cysteine, aspartate and metallo- proteases showed little effect on AACT stability in the culture medium, in sharp contrast with the broad-spectrum inhibitor of serine proteases PMSF, that effectively prevented AACT cleavage. Considering the whole plant secretory pathway, AACT proteolysis might be even more complex, possibly involving over time multiple cleavage steps at the C- and N-termini, a hypothesis supported by the extensive processing observed for the AACT variants along the endomembrane system (**Figure II.4**).

The inhibitory reactive site of serpins is inserted in a protruding loop, the RCL, which acts as a bait to trap a target protease. As the RCL is located in the C-terminal region of the protein, proteolysis at this side of the protein would affect its inhibitory activity. In line with this possibility, N-terminal sequencing of AACT forms detected after treatment with the EM proteases exhibited either an intact N-terminal extremity, or a truncated N-terminal end short of only 810 Da corresponding to the first eight amino acids of the protein (**Figure II.7C**). The molecular mass of plant-cleaved AACT was visually compared to AACT treated with human neutrophil elastase (HNE) (**Figures II.7A and II.7B**), known to process this serpin at position P1-P1', P4-P3 and P6-P5 of the RCL, i.e. 3



and 4 residues upstream of the P1 residue according to the nomenclature of Schechter and Berger, 1967 (Rubin et al., 1994). As observed on gel (**Figure II.7A**), EM proteases and HNE generated AACT polypeptides of similar electrophoretic mobility, about 5 kDa smaller than the complete inhibitor (**Figure II.7B**), which suggests a common cleavage zone likely located near the solvent-exposed RCL. The RCL of inhibitory serpins is very accessible to target (sensitive) proteases but also intrinsically susceptible to hydrolysis by non-target proteases, with a concomitant inactivation of the inhibitor (Potempa et al., 1991).



**Figure II.7 AACT susceptibility to tobacco extracellular proteases.** **A.** Bacterial recombinant rAACT cleaved to lower molecular mass forms by BY-2 proteases secreted in the culture medium. Bacterial recombinant AACT (rAACT) was incubated in 4-days-old filtered BY-2 culture medium (EM) for 16 h at room temperature. As a control, rAACT was incubated in fresh culture medium (lane 1) or with human neutrophil elastase (HNE) known to cleave AACT in the RCL. The addition of various diagnostic protease inhibitors (PMSF (P), EDTA (E), pepstatin (pep), E-64 or a combination of all PIs (All)) had different retarding effects on AACT processing by the extracellular proteases. **B.** MALDI-TOF mass spectrum illustrating the cleavage of rAACT by HNE in the RCL. Molecular  $m/z$  values of 45,301 and ~40,000 were observed for the complete and truncated forms of the serpin, respectively. **C.** N- and C-terminal cleavage of rAACT, as assessed by N-terminal sequencing (step 2) and comparative mobility following SDS-PAGE (step 1), compared to HNE-generated fragments (see panels A and B).

## II. 5 Discussion

The production of recombinant serpins in plants has been limited, until now, to the antitrypsin inhibitor  $\alpha$ 1-proteinase inhibitor (Terashima et al., 1999b; Huang et al., 2001). Here we assessed the potential of BY-2 tobacco cultured cells for the expression of a second serpin, AACT. The sequestration of recombinant proteins in different organelles using targeting/retrieval signals has been proposed by several authors to achieve high yields in various systems, especially in eucaryotic organisms. ER retention using the K/HDEL C-terminal tetrapeptide, in particular, has been largely exploited, notably for the expression of clinically-useful antibodies (Conrad and Fiedler, 1998; Gomord and Faye, 2004). The neutral pH, oxidizing redox environment and low proteolytic content, along with the abundance of molecular chaperones and enzymes facilitating protein folding in the ER, usually make this organelle an interesting location for the accumulation of stable and properly assembled recombinant proteins.

On the other hand, and although heterologous targeting signals are used routinely in protein expression systems, the actual intracellular location of foreign proteins is rarely assessed. Whereas protein sorting signals used for the subcellular targeting of recombinant proteins are usually functional, unexpected protein targeting may arise in some case, as illustrated for instance with Guy's 13 secretory immunoglobulin A (IgA) light and heavy chains in rice seeds, that were accumulated in ER-derived protein bodies and protein storage vacuoles in the endosperm despite the absence of ER-retention or vacuolar sorting peptides (Nicholson et al., 2005). Likewise, the heat-labile enterotoxin antigen B, from *E. coli*, was accumulated in starch granules of transgenic maize kernels despite an expected apoplastic secretion (Chikwamba et al., 2003), while more recently a KDEL-tagged scFv-Fc fusion antibody expressed in *Arabidopsis thaliana* seeds was partially localized in the periplasmic space (Van Droogenbroeck et al., 2007). In the present work, immunolocalization results showed, overall, a proper subcellular compartmentalization for ER- and vacuolar-tagged AACT variants, confirming that heterologous targeting signals included in gene constructs were functional in tobacco cells. On the other hand, data retrieved from pulse-chase experiments clearly showed that AACT-V and AACT-ER cell lines were also accumulating AACT in the culture medium. Most documented examples of recombinant proteins found in unexpected cell compartments were associated with protein expression in seeds (Chikwamba et al., 2003;

Arcalis et al., 2004; Nicholson et al., 2005; Petruccelli et al., 2006; Van Droogenbroeck et al., 2007), then revealing a tissue-specific efficacy of different protein targeting signals (Arcalis et al., 2004; Drakakaki et al., 2006; Petruccelli et al., 2006).

By comparison, the unexpected targeting of AACT in BY-2 tobacco cells was likely the result of intrinsic structural features and modifications of recombinant AACT *in vivo*, rather than an atypical sorting mechanism in cultured cells. KDEL-tagged AACT found in the EM could have resulted from the leakage of K/HDEL tagged proteins from the ER retrieval machinery, as documented earlier for both endogenous (Henderson et al., 1997) and recombinant (Ko et al., 2003) proteins. Alternatively, the KDEL tag could eventually be cleaved off, due to protease-susceptible sites at the C terminus of the protein (e.g., Badri et al., 2007b). As well, the effective retrograde translocation of recombinant proteins from the Golgi back to the ER has been shown also to depend on steric accessibility of the K/HDEL string to specific KDEL receptors in the *cis*-Golgi (Sriraman et al., 2004). The C-terminal structure of recombinant AACT might not be optimally exposing the KDEL signal to the Golgi retrieval machinery, thereby accounting for the (partial) extracellular localization of the protein. Similarly, AACT in AACT-V cells was mainly secreted into the EM, suggesting that the C-terminal vacuolar targeting signal was also either cleaved off or inaccessible for proper receptor-mediated sorting to the vacuole.

Surprisingly, AACT expressed into the cytosol also showed an unexpected subcellular localization, with most fluorescence labelling either confined into the nucleus or showing a dual distribution between this organelle and the cytosol (see **Figure II.3**). Although intriguing, this observation is consistent, on the other hand, with previous studies reporting the accumulation of human plasma AACT in the nucleus of carcinoma cells (Takada et al., 1982; Takada et al., 1986) and the natural ability of this protein to bind DNA (Naidoo et al., 1995) (also see Benchabane et al., 2007: accompanying paper). Small proteins –e.g. proteins smaller than ~40 kDa– may diffuse freely from the cytosol through the nuclear pore complex, while in contrast the nucleocytoplasmic transit of larger proteins generally involves ATP-dependent processes implicating specialized transport proteins after the recognition of specific nuclear localization signals (Merkle, 2004; Pemberton and Paschal, 2005). As AACT likely is too large to easily diffuse into the nucleus, the transit of AACT towards the nucleus still remains to be understood, and the possible existence of still unidentified (or cryptic) localization signal(s) for nuclear import to be confirmed.

One advantage offered by plants for the production of recombinant proteins is their ability to operate complex post-translational modifications required for the activity and stability of several proteins (Gomord and Faye, 2004; Faye et al., 2005). As illustrated by the different variants detected during pulse-chase experiments, recombinant AACT undergone several post-translational modifications in BY-2 tobacco cells (**Figure II.4**). Like its human counterpart, plant-made AACT targeted to the secretory pathway was N-glycosylated, and by extrapolation from the size (50 kDa) of AACT glycosylated precursors at the end of the pulse period (**Figure II.4**), the inhibitor would contain at least four glycan chains in plant cells, as also suggested earlier for human AACT (Baumann et al., 1991; Gettins et al., 1996; Zhang et al., 2003). Because human AACT likely harbours large multi-antennary N-glycans with terminal sialic acid residues (**Figure II.5A, B**), the lower molecular mass of plant-made AACT variants compared to human AACT would likely be the result of less extended glycan structures rather than the result of a smaller number of occupied N-glycans sites. In contrast, plant-made AACT contains N-glycans rapidly processed to PNGase F-resistant complex glycans with  $\alpha$ -1,3 fucose residues on proximal N-acetylglucosamine. To substantiate these results, extracellular AACT was shown to be reactive to  $\alpha$ -1,3 fucose- and  $\beta$ -1,2 xylose-specific antibodies.  $\alpha$ -1,3 fucose and  $\beta$ -1,2 xylose residues, typical of plant N-glycans (Lerouge et al., 1998), have been reported to increase the immunogenicity and allergenicity of plant-derived glycoproteins in mammals (van Ree et al., 2000; Bardor et al., 2003; Gomord et al., 2005). Retaining AACT into the ER using a KDEL C-terminal signal, which represents an efficient strategy to prevent the apposition of immunogenic epitopes in Golgi cisternae (Sriraman et al., 2004; Gomord et al., 2005; Petruccioli et al., 2006), proved unsuccessful however to obtain non-immunogenic AACT, due to a low efficiency of the ER retention signal. Indeed, a large fraction of AACT in AACT-ER cells escaped the ER retrieval machinery and migrated through the Golgi, where N-glycans were matured before secretion of the protein into the EM.

Proteolytic processing also is another cause of protein heterogeneity and unexpected degradation leading to protein loss or subcellular mistargeting in plants (Faye et al., 2005; Doran, 2006). The plant cell contains a complex array of proteases that are essential to ensure important metabolic and cellular functions (Beers et al., 2004; Schaller, 2004), but that may also impair recombinant protein production by directly hydrolyzing nascent or mature polypeptide chains. Several strategies have been proposed to prevent

unwanted proteolysis in plant hosts, including the targeting of recombinant proteins to specific organelles using appropriate sorting signals (Faye et al., 2005; Goulet and Michaud, 2006; Benchabane et al., 2007). Higher yields and improved stability have been obtained, for instance, by sequestering a number of proteins into the ER (Stoger et al., 2002; Gomord et al., 2004; Peng et al., 2006; Petruccioli et al., 2006), but this compartment may not be ideal for all proteins given the presence of proteases as early as in the ER along the endomembrane system, presumably involved in the processing of pre- and proregions, and in the degradation of misfolded proteins (Bass et al., 2000; Schmitz and Herzog, 2004). The bovine plasma protein aprotinin, for instance, showed greater stability when secreted in the apoplast of potato leaves than when retained in the ER using the KDEL signal, probably as a result of proteolytic events at both C and N termini (Badri et al., 2007a). Likewise, the apoplastic milieu has been proposed as an interesting compartment for recombinant protein expression, but plant cells also secrete several proteases that may significantly impair protein accumulation *in vivo* (Hellwig et al., 2004; Schiermeyer et al., 2005; Lienard et al., 2007). For instance, the expression of  $\alpha$ 1-protease inhibitor in the extracellular medium of transgenic rice suspension cultured cells resulted in the accumulation of a truncated protein product with impaired inhibitory activity against target proteases (Terashima et al., 1999b), presumably resulting from the release of sulfhydryl proteases in the cell culture medium (Terashima et al., 1999a). More recently, a recombinant form of  $\alpha$ 1-plasminogen activator from *Desmodus rotundus* in BY-2 tobacco cells was altered in the extracellular medium by metallo-proteases secreted by growing cells (Schiermeyer et al., 2005). In the present study, recombinant AACT was susceptible to PMSF-sensitive, serine proteases found in the EM of tobacco cells, again illustrating the susceptibility of recombinant proteins to specific proteases in plant platforms and stressing the importance of empirically assessing the stability of each protein to express.

As shown with spiking experiments, the tobacco cell culture EM was shown here to contain secreted proteases able to use AACT as a substrate, while pulse-chase experiments were suggesting proteolytic trimming of the same protein within the secretory pathway, before its secretion into the EM. Overall, these data suggest that AACT was sensitive to proteases found in the ER and/or the Golgi apparatus as well as in the apoplast, and that the production of intact AACT in the plant secretory pathway will require innovative production schemes aimed at protecting integrity of the protein backbone (Goulet and Michaud, 2006). The co-secretion of a companion protease inhibitor altering

endogenous proteolytic activities along the secretory pathway could represent an effective strategy, as illustrated recently for recombinant antibodies secreted by transgenic tobacco roots, that were stabilized by the cosecretion of a Bowman-Birk trypsin inhibitor from soybean (Komarnytsky et al., 2006). As illustrated here, simply expressing AACT with no targeting peptide could also provide an effective way of producing active AACT in plants. While many serpins are glycoproteins, the inhibitory activity of these proteins may not be impaired in the absence of N-glycans (Rubin et al., 1990; Kwon et al., 1995a; Lamark et al., 2001). In practice, this observation underlines the potential usefulness of the cytosol for the production of AACT under a nonglycosylated form, as also suggested previously for *E. coli* (Rubin et al., 1990; Bird et al., 2004). Aglycosylated serpins, on the other hand, could exhibit a high plasma clearance rate (Minta, 1981; Travis et al., 1985) and an increased instability to denaturants and thermal denaturation compared to native glycosylated AACT (Powell and Pain, 1992; Kwon et al., 1995b; Kwon and Yu, 1997), which suggests a stabilizing role for glycan chains on these proteins *in vivo*. Work is required, now, to assess the inhibitory potency of non-glycosylated AACT in terms of specific activity compared to human AACT and *in vivo* stability (and efficacy) in a therapeutic context. Work is also needed to understand why cytosol-targeted AACT is found in the nucleus of plant cells, and to assess the impact of such an unexpected targeting on protein accumulation and stability *in planta*.

## **II.6 Materials and Methods**

### **II.6.1 Materials**

Bovine chymotrypsin (E.C.3.4.21.1), human neutrophil elastase (HNE, E.C.3.4.21.37) and human plasma antichymotrypsin were from Sigma (Oakville, ON, Canada). Polyclonal and monoclonal anti-AACT antibodies (nos. A2298 and A2298-01) were from US Biologicals (Swampscott, MA, USA). Polyclonal antibodies against BiP or against  $\beta$ -1,2 xylose and  $\alpha$ -1,3 fucose residues were produced in rabbits as described earlier (Gomord et al., 1997; Faye et al., 1993). All secondary antibodies were from Sigma, and NBC/BCIP reagents for peroxidase color development from BioRad (Mississauga, ON, Canada). The bacterial expression vector pGEX-3X, GSH-Sepharose and protein-A Sepharose were from Amersham Biosciences (Baie d'Ufré, QC, Canada). Factor X<sub>a</sub> and enterokinase cleavage kits were from Novagen (San Diego, CA, USA). The pCR-Blunt

cloning vector was from Invitrogen (La Jolla, Ca, USA), the pUC-19 vector from Fermentas Life Science (Burlington, ON, USA), and the pCambia 2300 binary vector from CAMBIA (Canberra, Australia). The  $^{35}\text{S}$  protein labelling mix was from Perkin Elmer Products (Courtaboeuf, France). Tunicamycin and concavalin A were from Sigma, biotinylated *Sambucus nigra* lectin I (SNA-I) from EY Laboratories (San Mateo CA, USA), and biotinylated *Ricinus communis* agglutinin I (RCA-I) from Vector Laboratories (Burlingame, CA, USA). Horseradish peroxidase was from Sigma and streptavidin horseradish peroxidase conjugate from GE Healthcare (Orsay, France). Peptide N glycosidase F (E.C.3.2.218) was from Roche (Laval, QC, Canada) and *Arthrobacter ureafaciens* neuraminidase (E.C.3.2.1.18) from EY Laboratories. PMSF, E-64, pepstatin and EDTA were from Sigma.

### II.6.2 Gene constructs

Four expression vectors were devised to target recombinant AACT to different subcellular compartments of tobacco cells, namely the cytosol (AACT-C construct), the ER (AACT-ER construct), the vacuole (AACT-V construct) and the apoplast (AACT-S construct) (see **Figure II.1**). The AACT coding sequence (Accession Number J05176) was amplified by reverse transcriptase (RT)-PCR from a hepatocyte mRNA library using the following primers: 5'-AACAG CCCAC TTGAC GAGG-3' and 5'-TTACT GAGAG CCCCA CTGCT TG-3'. The resulting *aact* cDNA was subcloned into the PCR-blunt vector (Invitrogen, La Jolla, CA, USA) and confirmed by automatic DNA sequencing (Service de séquençage de l'Université Laval, Québec, Canada). For targeting to the secretory pathway, the N-terminal signal peptide of protein disulfide isomerase (PDI) (Shorosh and Dixon, 1991) was introduced at the 5' end of the AACT coding sequence. The ER-retention signal KDEL and the tobacco chitinase A vacuolar sorting signal (Neuhaus et al., 1991) were added to the 3' end of the coding sequence by PCR amplification using the following 3'-end primers: 5'-TATCT AGACT AAAGT TCATC CTTGG CTTGC TTGGG ATTGG-3' (for ER retention) and 5'-TAGAC TACAT AGTAT CGACT AAAAG GTCGG CTTGC TTGGG ATTGG-3' (for vacuolar sorting). The four *aact* transgenes were cloned into a modified pUC-19 vector (Fermentas Life Science, Burlington, ON, USA) including the double constitutive CaMV35S promoter, a TEV enhancer sequence and the PA35S terminator sequence. The resulting constructs were transferred into the commercial vector



pCambia 2300 (CAMBIA, Canberra, Australia) and introduced into *Agrobacterium tumefaciens* LBA4404 for tobacco transformation.

For expression in *E. coli*, an AACT/pGEX-3X vector was constructed allowing for the expression of AACT fused to glutathione S-transferase used as an affinity tag for affinity purification. The *aact* sequence was amplified by PCR with the following DNA primers: 5'-AAGGA TCCAA AACAG CCCAC TTGAC G-3' and 5'-TTGAA TTCTA GGCTT GCTTG GGATT G-3'. Unique *Bam*HI and *Eco*RI restriction sites were included in the primer sequences (underlined) to allow ligation into the expression vector pGEX-3X. Following PCR, the hybrid genes were controlled by automatic sequencing and subcloned into pGEX-3X, in frame with the GST-encoding sequence. This vector was introduced into *E. coli*, strain BL21 for heterologous expression and purification.

### **II.6.3 Stable transformation of BY-2 tobacco cells**

Stable transformation of tobacco (*Nicotiana tabacum*) cv. Bright Yellow-2 (BY-2) cells was performed as described previously (Gomord et al., 1998), using *Agrobacterium* cells transformed with the pCambia vectors (see above). Transformants were selected in the presence of kanamycin (100 µg/ml), analyzed for *aact* expression by RT-PCR, and immunodetected with anti-AACT monoclonal antibodies. Microcalli exhibiting high levels of AACT were screened and used to initiate suspension cultures for recombinant protein expression. Tobacco cell cultures were maintained in Murashige and Skoog medium supplemented with 30% (w/v) sucrose, 1 mg/l thiamine, 51 mg/l *myo*-inositol, 0.2 mg/l 2,4-D and 100 mg/l KH<sub>2</sub>PO<sub>4</sub>, on an orbital shaker at 140 rpm, at 20°C in the dark. The cells were transferred to a fresh medium every two weeks with 5% (v/v) inoculums.

### **II.6.4 AACT transient expression in *Nicotiana tabacum***

For transient expression in tobacco leaf cells, a two-days-old *Agrobacterium* culture was washed and resuspended in fresh infiltration medium (50 mM MES, pH 5.6, 0.5% glucose (w/v), 2 mM Na<sub>3</sub>PO<sub>4</sub>, 100 mM acetosyringone). Using a syringe, the bacterial suspension was infiltrated in the abaxial epidermis of *N. tabacum* cv. Xanthi leaves. The plants were grown for two days prior to extraction of total soluble proteins. Leaf-infiltrated



tissues were cut out, frozen in liquid nitrogen, and ground with a mortar and a pestle in protein extraction buffer (75 mM Tris-HCl, pH 6.8, 2% (w/v) SDS, 3% (v/v)  $\beta$ -mercaptoethanol, 1 mM EDTA, 1 mM PMSF). The soluble proteins were recovered by centrifugation for subsequent immunodetection.

### **II.6.5 Expression and purification of recombinant AACT from *E. coli***

Heterologous expression and purification of recombinant AACT in *E. coli* were carried out using the GST gene fusion system (Smith and Johnson, 1988) as described earlier for other protease inhibitors (Michaud et al., 1994). In brief, 1.5 ml of an overnight *E. coli* culture was inoculated into 150 ml of LB medium containing 100  $\mu$ g/ml ampicillin. The culture was incubated at 37°C with vigorous agitation until reaching an OD<sub>600</sub> of 0.4-0.6, prior to the addition of IPTG to a final concentration of 0.2 mM. The culture was incubated for an additional 16 h at 37°C, before harvesting bacteria by centrifugation for 10 min at 14,000g. After discarding the supernatant, the bacterial pellet was submitted to several freeze-and-thaw cycles, and then suspended in a 50 mM Tris lysis buffer, pH 8.0, containing 5% (w/v) sucrose, 50 mM EDTA and 5% (v/v) Triton X-100. After 5 min on ice, the bacterial lysate was centrifuged at 20,000 g for 10 min at 4°C, and the supernatant containing soluble proteins was saved. Purification of the GST fusions was performed by affinity chromatography with reduced glutathione-sepharose beads as described earlier (Michaud et al., 1994), and the GST moiety was removed from the inhibitor using human Factor X<sub>a</sub> (Novagen, La Jolla, CA, USA), according to the supplier's instructions. Purity of the inhibitors was visualized by SDS-PAGE and immunodetection with anti-AACT antibodies. Protein concentrations were determined by densitometry of Coomassie Blue-stained gels after SDS-PAGE, using the PHORETIX 2-D image analysis software (Nonlinear Dynamics, Durham, NC, USA) and bovine serum albumin (Sigma) as a protein standard.

### **II.6.6 Metabolic pulse-chase labelling and immunoprecipitation**

For pulse-chase time course analyses, 300 mg of filtered BY-2 suspension cultured cells were pulse-labelled by resuspension in 900  $\mu$ l of fresh culture medium supplemented with 4 MBq of [<sup>35</sup>S]-protein labelling mix (NEN Life Science Products). After pulse-labelling at 20°C under gentle shaking (140 rpm) for 60 min, the cells were incubated for various

chase periods by adding 110  $\mu$ l of chase medium containing 50 mM methionine and 50 mM L-cysteine. Extra- and intracellular proteins were isolated as described earlier (Gomord et al., 1997). In brief, extracellular proteins were collected by centrifugation at 50 g for 3 min. The cell pellet was washed with TBS (50 mM Tris-HCl, pH 7.5, 150 mM NaCl) prior to harvesting by the addition of 500  $\mu$ l of lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 1 mM EDTA, 1% (w/v) sodium ascorbate) containing the Complete™ protease inhibitor mixture (Roche Diagnostics), and sonicated (2x 20 s 40%). The cell slurry was centrifuged at 12,000 g for 15 min, and the supernatant containing intracellular proteins was recovered. Supernatants from both the cell media and cell lysis mixture were immunoprecipitated with anti-AACT polyclonal antibodies (dilution 1:500) for 2 h at room temperature on an end-over-end shaker before adding protein A Sepharose (50  $\mu$ g) and incubating for an additional hour. Immunocomplexes were washed four times in washing buffer (TBS containing 1% (v/v) Triton X-100 and 0.2% (w/v) SDS), then twice in TBS containing 0.1% (v/v) Triton X-100, and finally twice in TBS. Radiolabelled proteins were recovered by heating at 80°C for 10 min in SDS-PAGE loading buffer, fractionated by SDS-PAGE, and detected by fluorography. The inhibition of N-glycosylation co-translational transfer was carried out by incubating the cells for 90 min at room temperature in fresh culture media supplemented with 18  $\mu$ M tunicamycin, prior to pulse labelling performed as described above.

### **II.6.7 Enzymatic deglycosylation of plant recombinant AACT**

Peptide N glycosidase F (PNGaseF) treatment was performed on the pulse radio-labelled AACT-protein A immunocomplex by heating for 10 min in 60  $\mu$ l of digestion buffer (50 mM Tris-HCl, pH 7.6, 0.2% (w/v) SDS). The proteins recovered were let to stand for a few minutes at room temperature and diluted in 60  $\mu$ l of dilution buffer (50 mM Tris-HCl, pH 7.6, 0.02% (v/v) Nonidet P40). Half of the sample was treated with 1  $\mu$ l of PNGaseF (1 munit/ml), and incubated for 16 h at 37°C. The proteins were resolved by SDS-PAGE and detected by fluorography.

### II.6.8 Enzymatic deglycosylation of human AACT

Human AACT (hAACT) was treated successively with neuraminidase and peptide N-glycosidase F, and a sample was put aside for gel analysis following each treatment. hAACT (5 µg) was diluted in neuraminidase digestion buffer (90 mM sodium phosphate, pH 5), and incubated with neuraminidase (0,5 units) for 3 h at 37°C. The neuraminidase-treated sample was further digested with PNGase F by diluting the mixture first in one volume of PNGaseF digestion buffer (90 mM sodium phosphate, pH 8, 0,2 % (w/v) SDS), and then heating for 5 min at 100°C. The sample was let to stand for a few minutes, prior to the addition of one volume of buffer (90 mM sodium phosphate, pH 8, 0,2 % (v/v) Nonidet-P40) and PNGase F (0.4 U). The vial was incubated at 37°C for 16h. hAACT samples after simple or double digestion were resolved by SDS-PAGE, electrotransferred onto a nitrocellulose membrane and analyzed by immuno- and affinodetection.

### II.6.9 Affino- and immunodetection

Extracellular proteins from BY-2 tobacco cells were prepared from 3 to 4-days-old suspension cultures. The cells were harvested by filtration and the culture medium containing apoplasmic proteins was put aside. To isolate peripheral proteins ionically bound to the cell membrane, the cell pellet was washed in fresh medium containing 0.5 mM NaCl, incubated for 20 min at 4°C, and filtrated. The washing fluid was pooled with the culture medium and the proteins were precipitated by two subsequent cycles of ammonium sulfate precipitation at 40% (w/v) and 90% (w/v) of saturation. Protein extracts were fractionated on a 12% (w/v) SDS-polyacrylamide slab gel and electrotransferred onto a nitrocellulose membrane. AACT was detected using monoclonal anti-AACT antibodies at a dilution of 1:5000 in TBS (20 mM Tris-HCl, pH 7.5, 500 mM NaCl) containing 1% (w/v) skim milk, or using anti  $\beta(1,2)$ -xylose and  $\alpha(1,3)$ -fucose antibodies diluted at 1:1000 in the same buffer. The appropriate secondary antibodies coupled to alkaline phosphatase were diluted 1:10,000. N-glycans were affinodetected using the lectin concanavalin A (Con A), *Sambucus nigra* agglutinin I (SNA I) and *Ricinus communis* agglutinin I (RCA), specific, respectively, to polymannosidic glycans, terminal  $\alpha$ -2,6 N-acetyl-neuraminic acid  $\beta$ -1,4 galactose N-acetylglucosamine ( $\alpha$ -2,6Neu5Ac  $\beta$ -1,4gal), and terminal  $\beta$ -galactose residues. The membranes were saturated for 1 h in TTBS (TBS supplemented with 0.1% (v/v) Tween 20) prior to incubation with lectins for 90 min at room temperature on an

orbital shaker. Concanavalin A (25 µg/ml) was diluted in Con A buffer (1 mM CaCl<sub>2</sub> + 1 mM MgCl<sub>2</sub> in TTBS), biotin-SNA I (2mg/ml) and biotin-RCA (20ug/ml) were diluted in TTBS. The membranes were washed in the same buffer and incubated for 1 h at room temperature with peroxidase (50 µg/ml) diluted in Con A buffer for the concanavalin A assay, or with streptavidin peroxidase for the biotinylated lectin assays. The proteins were revealed using an appropriate colorimetric peroxidase substrate.

#### **II.6.10 Immunocytolocalisation for confocal microscopy**

For immunocytolocalization, four-days-old BY-2 tobacco cells were harvested by filtration, fixed and permeabilized before staining with anti-AACT antibodies and specific secondary antibodies for confocal microscopy. Unless otherwise indicated, all steps were conducted at room temperature on a end-over-end shaker in 0.1 M PIPES buffer supplemented with specific reactivities. Cell fixation was performed by a 1-h incubation in 4% (w/v) paraformaldehyde. Permeabilization was realized by a 20 min incubation in 1% (w/v) cellulase + 0.1% (w/v) pectinase + 1% (w/v) BSA, followed by a 10 min incubation in 0,5% (v/v) Triton X-100. To eliminate non specific staining, the cells were saturated in 1% (w/v) BSA + 1% (v/v) fish gelatin. The cells were probed with anti-AACT or anti-BiP polyclonal antibodies (dilution 1:100) overnight at 4°C, without agitation, and then washed three times in 1% (w/v) fish gelatin prior to final incubation in a 1:100 dilution of anti-rabbit antibody coupled to FITC, for 1 h. The cells were washed three times in PIPES, twice in water, and mounted on a microscopy slide. Fluorescence images were captured on a Leica TCS SP2 AOBs confocal laser scanning microscope (Leica, Microsystems, Rueil-Malmaison, France). Images were generated using a 488-nm argon ion laser source, and fluorescence was recorded with a photomultiplier setup in the 493–538 nm range. Image patterns were processed using the LCS Lite software, v. 2.64 (Leica, Microsystems).

#### **II.6.11 Spiking experiments**

For spiking experiments, 4 µg of bacterial recombinant AACT (rbAACT) or GST-rbAACT were added to filtered extracellular medium of 4 days-old wild-type BY-2 suspension cells. A negative control was prepared by incubating rbAACT in fresh culture medium. All samples were incubated overnight at room temperature on an end-over-end shaker. The

proteins were resolved by SDS-PAGE. To identify proteases implicated in the proteolytic processing of AACT, the following diagnostic protease inhibitors were added: 1 mM PMSF, 1 µg/ml E-64, 5 mM EDTA, and 1 µg/ml pepstatin A. AACT was cleaved in the reactive center loop (RCL) by partial digestion with human neutrophil elastase (HNE), as described by Rubin et al. (1990). The cleavage was performed with an inhibitor/enzyme ratio of 20:1 in Tris buffer (100 mM Tris-HCl, pH 8.3; 0,025% (v/v) Triton x-100). The reaction was carried out for two h at 37°C, and protein hydrolysis was stopped by the addition 2 mM PMSF.

#### **II.6.12 MALDI-TOF/MS analyses**

For MALDI-TOF/MS, 500 µmoles of rAACT partially digested with HNE was desalted on a Centricon 10 column (Millipore, Molsheim, France) following the suppliers' recommendations, and then lyophilized. The proteins were solubilised in 5 µl of 30% (v/v) acetonitrile containing 0.1% (v/v) trifluoroacetic acid (TFA) prior to spectrometry analysis on a Voyager DE-Pro MALDI-TOF instrument (Applied Biosystems, USA) equipped with a 337-nm nitrogen laser. Mass spectra were acquired in the linear delayed extraction mode using sinapinic acid (Sigma) as a matrix. The matrix solution was freshly prepared by dissolving sinapinic acid at 5 mg/ml in 30% (v/v) acetonitrile containing 0.1% (v/v) TFA, and mixed with solubilized protein samples in 1:1 ratio. The spectra were recorded in the positive mode, using an acceleration voltage of 25,000 V with a delay time of 750 ns, a grid voltage of 92%, a 0.15% guide wire and a 50% laser intensity. Up to a three thousand laser shots were accumulated to obtain an optimal signal-to-noise ratio and spectra were smoothed once. External calibrations were carried out using the commercial protein standards cytochrome B (12 360 Da), apomyoglobin (16 952 Da) and bovine albumin (66 430 Da) (Applied Biosystems).

#### **Acknowledgements**

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the French Ministère de la Recherche, and the National Science and Engineering Research Council (NSERC) of Canada. M.Benchabane was the recipient of an NSERC graduate fellowship.

## **CHAPITRE III**

**Nucleocytoplasmic transit of human  $\alpha$ 1- antichymotrypsin  
expressed in tobacco leaf epidermal cells**

### III.1 Mise en contexte

Lors de la production de protéines recombinantes dans les plantes, l'obtention de hauts rendements et d'un produit structurellement adéquat est fortement tributaire de l'adressage cellulaire de la protéine. Ainsi, dans la première partie de nos travaux, nous avons cherché à produire l' $\alpha$ 1-antichymotrypsine dans différents organites d'une culture cellulaire de tabac BY-2 et à évaluer les modifications post-traductionnelles des formes d'AACT ainsi obtenues. L'adressage vers le système endomembranaire de sécrétion est souvent favorisé en moléculaire car il permet, dans plusieurs cas, l'obtention de rendements intéressants. Pour les protéines sécrétées, une purification facilitée directement à partir du milieu extracellulaire est alors également possible. Dans le cas présent, l'adressage de l'AACT vers le système endomembranaire de sécrétion a, au contraire, donné lieu à une grande hétérogénéité du produit. D'une part, l'AACT y était glycosylée et les formes sécrétées comportaient des glycanes complexes avec présence de résidus typiques des glycanes végétaux, l' $\alpha$ -1,3 fucose et/ou le  $\beta$ -1,2 xylose identifiés antérieurement comme immunogènes lorsqu'administrés chez les mammifères. D'autre part, l'AACT y était dégradée en partie, donnant lieu à une perte d'intégrité, voire d'activité. Adressée vers le cytosol, l'AACT était de toute évidence plus stable mais montrait une double localisation, nucléaire et cytoplasmique. Devant ces résultats inattendus, nous nous sommes penchés davantage sur cette forme cytosolique de l'AACT. Pourrait-elle s'avérer intéressante pour obtenir un produit actif ? L'AACT humaine étant une glycoprotéine, elle pourrait ne pas conserver toutes ses propriétés biochimiques lorsqu'exprimée sous forme non-glycosylée dans le cytosol. Par ailleurs, la localisation nucléaire d'une protéine de la taille de l'AACT est surprenante considérant la taille moindre des pores nucléaires, alors comment expliquer cet adressage ? La translocation nucléaire pourrait-elle, enfin, contribuer à la stabilité apparente dont fait preuve l'AACT lorsqu'elle est exprimée dans le cytosol des cellules de tabac ?

Dans cette deuxième partie des travaux, nous avons donc voulu répondre à certaines de ces nouvelles interrogations. Comme un site de liaison à l'ADN a été identifié auparavant chez l'AACT, nous nous sommes demandé si l'interaction avec l'ADN pouvait être impliquée dans la localisation nucléaire de l'AACT. Nous avons donc créé par génie génétique un mutant AACT ne liant plus l'ADN puis observé sa localisation cellulaire lorsqu'exprimé dans l'épiderme des feuilles de tabac. Nous avons en parallèle mesuré

l'activité inhibitrice des formes glycosylée, non glycosylée et mutante de l'AACT contre la chymotrypsine, une protéase cible majeure pour cet inhibiteur. Nous nous sommes enfin demandé si, dans un contexte de moléculaire, la localisation nucléaire, voire la liaison à l'ADN elle-même, pouvaient contribuer positivement à la stabilité de la protéine recombinante. Les résultats de ces expériences sont décrits ici sous la forme d'un article scientifique qui sera soumis sous peu pour publication.



## **Nucleocytoplasmic transit of human $\alpha$ 1-antichymotrypsin expressed in tobacco leaf epidermal cells**

Meriem Benchabane <sup>1,2</sup>, Claude Saint-Jore Dupas <sup>2</sup>, Loïc Faye <sup>2</sup>, Véronique Gomord <sup>2</sup> & Dominique Michaud <sup>1,\*</sup>

<sup>1</sup> *Département de phytologie, CRH/INAF, Université Laval, Québec (QC) Canada G1K 7P4.*

<sup>2</sup> *CNRS UMR 6037, IFRMP 23, GDR 2590, Université de Rouen, Faculté des Sciences Bât. Ext. Biologie, 76821 Mont-Saint-Aignan cedex, France*

\* To whom correspondence should be sent.

**Keywords:**  $\alpha$ 1-antichymotrypsin; nuclear targeting; plant-made pharmaceuticals, recombinant proteins; serpins.

**Abbreviations:** AACT,  $\alpha$ 1-antichymotrypsin; CT, chymotrypsin; ER, endoplasmic reticulum; GFP, green fluorescent protein; GST, glutathione S-transferase; NLS, nuclear localization signal.

### III. 2 Résumé

Nous avons observé récemment une localisation nucléaire inattendue de l' $\alpha$ 1-antichymotrypsine (AACT) humaine dans le cytosol de cellules transgéniques de tabac BY-2 (voir Benchabane et al., 2007–article conjoint). Nous avons évalué ici si l'activité de liaison naturelle de l'AACT humaine à l'ADN pouvait expliquer sa localisation nucléaire dans les cellules de tabac, et si cette même activité de liaison avait un impact sur sa stabilité et son activité inhibitrice contre la chymotrypsine. Une forme modifiée de l'AACT ne se liant pas à l'ADN, le variant rAACT $\Delta$ K, a d'abord été générée par mutagenèse dirigée, puis comparée à une version non mutée de l'inhibiteur, le variant rAACT, en termes d'activité antichymotryptique, de stabilité et de distribution dans les cellules de tabac. En accord avec les données structurales actuelles montrant des sites distincts pour l'activité inhibitrice et la liaison à l'ADN, des variants rAACT et rAACT $\Delta$ K exprimés dans *E. coli* ont montré des activités inhibitrices similaires envers la chymotrypsine, comparables à l'activité antichymotryptique de l'AACT humaine purifiée du plasma humain. Tel qu'observé antérieurement pour le variant rAACT exprimé dans des cellules de tabac en culture liquide, une fusion GFP-rAACT exprimée transitoirement dans le cytosol de cellules foliaires épidermiques de tabac a été détectée en bonne partie dans le compartiment nucléaire par microscopie laser confocale. À l'inverse, le variant rAACT $\Delta$ K exprimé sous la forme d'une fusion GFP a montré une distribution plus équilibrée entre le cytosol et le noyau, suggérant un effet de rétention de l'ADN sur l'AACT atteignant le compartiment nucléaire. En lien avec des observations qualitatives suggérant des niveaux de fluorescence plus élevés pour le variant GFP-rAACT dans les cellules foliaires de tabac, le variant rAACT $\Delta$ K s'est avéré plus sensible à la digestion tryptique en présence d'ADN, suggérant un effet stabilisateur de l'interaction AACT:ADN sur l'AACT en présence de protéases insensibles, possiblement utile pour l'expression de cette protéine humaine sous forme stable en système hétérologue végétal.

## Summary

We observed recently an unexpected, nuclear localization of human  $\alpha$ 1-antichymotrypsin (AACT) in the cytosol of transgenic BY-2 tobacco cells (see Benchabane et al., 2007—accompanying paper). Here we assessed whether the intrinsic DNA-binding activity of human AACT could explain its nuclear localization in tobacco cells, and whether this same activity could impact its overall stability and inhibitory activity against chymotrypsin. A modified form of AACT with no DNA-binding activity, rAACT $\Delta$ K, was first engineered by site-directed mutagenesis, and then compared to the wild-type polypeptide, rAACT, in terms of chymotrypsin inhibitory potency, overall stability and cellular distribution in tobacco cells. In accordance with available structural data showing distinct sites for protease inhibition and DNA binding, recombinant forms of rAACT and rAACT $\Delta$ K expressed in *E. coli* showed similar inhibitory activity against chymotrypsin, comparable to the antichymotrypsin activity of human AACT purified from human plasma. As observed earlier for rAACT stably expressed in tobacco cultured cells, a green fluorescent protein (GFP)-rAACT fusion protein transiently expressed in the cytosol of tobacco leaf epidermal cells was detected predominantly in the nuclear compartment by confocal laser microscopy. By contrast, rAACT $\Delta$ K expressed as a GFP fusion showed a more balanced distribution between the cytosol and the nucleus, suggesting a nuclear retention effect of DNA on rAACT reaching the nucleus. In line with qualitative observations suggesting higher GFP fluorescence levels for GFP-rAACT in tobacco leaf cells, rAACT $\Delta$ K was more susceptible than rAACT to tryptic digestion in the presence of DNA, suggesting a certain stabilizing effect of the AACT:DNA interaction on AACT challenged with non-target proteases, possibly useful in protecting this human protein expressed in plant heterologous expression systems.

### **III.3 Introduction**

Protein targeting to specific subcellular compartments using appropriate sorting peptides has been readily recognized as a key factor determining the overall stability and yield of recombinant proteins in transgenic plants (Wandelt et al., 1992; Schouten et al., 1996; Gomord et al., 1997; Ma et al., 2003). Several reports have been published over the last decade illustrating the positive impact of sequestering clinically-useful proteins into the endoplasmic reticulum (ER; Ma et al., 2003; Yoshida et al., 2004; Vitale and Pedrazzini, 2005) or the chloroplast (Daniell et al., 2002; Maliga, 2002; Daniell, 2006). Other destinations have been also proposed, notably including the apoplast, the vacuole, the peroxisome and the cytosol (Yoshida et al., 2004; Daniell, 2006; Hyunjong et al., 2006; Marusic et al., 2007). In practice, the targeting of recombinant proteins to specific subcellular compartments is achieved by the grafting of functionally conserved sorting or retention sequences to their primary amino acid sequence (Gnanasambandam et al., 2007). In most cases, the cellular location chosen for protein accumulation has a strong impact on final yield and quality, owing in part to the enzymatic and protein folding machineries specifically characterizing this organelle, and to the relative stability of each recombinant protein expressed in a heterologous environment (Faye et al., 2005; Goulet and Michaud, 2006).

The choice of a suitable cellular destination for protein deposition also depends on the structure/function features of the protein, that may require specific co- or post-translational modifications to show adequate activity, stability or homogeneity (Faye et al., 2005). Additional criteria are the possible metabolic interference [or pleiotropic] effects exerted on the host plant by the expressed protein (Gils et al., 2005; Badri et al., 2007a), and the actual effectiveness of protein sorting signals in organs or tissues considered for protein deposition, that may show species- or tissue-dependent functionality and present unexpected –or aberrant– deposition patterns *in planta* (Vitale and Hinz, 2005). Examples of aberrant cellular targeting have been reported recently for a number of proteins expressed in seeds at high levels (Chikwamba et al., 2003; Arcalis et al., 2004; Nicholson et al., 2005; Drakakaki et al., 2006; Van Droogenbroeck et al., 2007). A fungal phytase bearing a signal peptide for cellular secretion, for instance, could be readily detected in the apoplastic milieu of rice leaf tissues but was retained in ER protein bodies and protein storage vacuoles of the seed endosperm (Drakakaki et al., 2006), clearly illustrating the

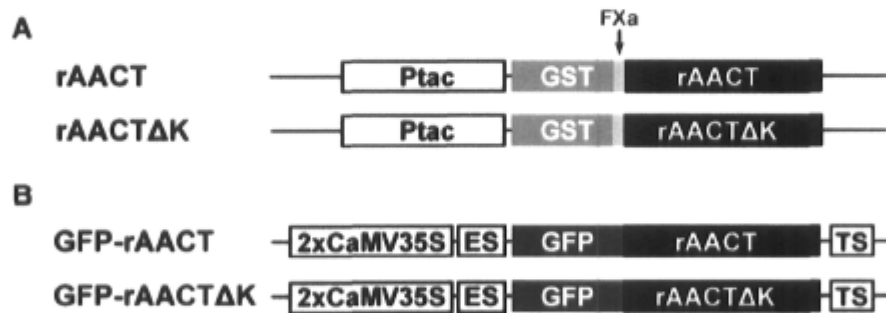
importance of carefully assessing the final distribution profile of each recombinant protein expressed *in planta*.

Cryptic [or uncharacterized] targeting signals in the primary sequence of some recombinant proteins could also impact their final destination in the cell, and hence positively or negatively alter their stability and fate *in planta*. We observed recently a dual cytosolic/nuclear distribution pattern for the serine protease inhibitor human  $\alpha$ 1-antichymotrypsin (AACT) expressed in BY-2 tobacco cells, despite the [assumed] absence of a nuclear localization signal (NLS) in both the primary sequence of this protein and the transgene construct used for transformation (Benchabane et al., 2007; see accompanying paper). Although unexpected *a priori*, this dual distribution of the human protein was consistent, on the other hand, with the previously reported occurrence of AACT in the nucleus of human carcinoma cells (Takada et al., 1982, 1986, 1988b), and with the interfering effects of this protein on DNA synthesis via the inhibition DNA primase and DNA polymerase  $\alpha$  activities (Tsuda et al., 1986; Takada et al., 1988b). Human AACT was also shown earlier to exhibit strong DNA-binding activity *in vitro*, mediated by two positively charged, lysine-rich motifs protruding at the surface of the protein backbone (Rubin et al., 1990; Naidoo et al., 1995). Considering that NLS and DNA-binding regions frequently overlap in the primary sequence of nuclear proteins (LaCasse and Lefebvre, 1995; Cokol et al., 2000), we assessed here whether the ability of AACT to bind DNA was correlated with the nuclear localization of this protein in tobacco cells, and whether this same binding activity could influence the intrinsic activity and stability of the recombinant protein product.

#### **III.4 Results and discussion**

To address these questions, a modified form of human AACT with no DNA-binding activity was first engineered, and then compared to the wild-type, original form of the inhibitor in terms of protease inhibitory potency, overall stability and cellular localization in tobacco cells. Two DNA binding sites have been identified earlier in AACT, each one including a positively charged, lysine-rich cluster physically interacting with DNA *in vitro* (Naidoo et al., 1995). One of these clusters, found at the surface of the protein in the C-terminal region, includes two lysine residues, Lys-391 and Lys-396, essential for DNA-binding activity (Naidoo et al., 1995). Gene constructs were devised here to produce two recombinant forms of AACT in *Escherichia coli* differing in their ability to interact with DNA

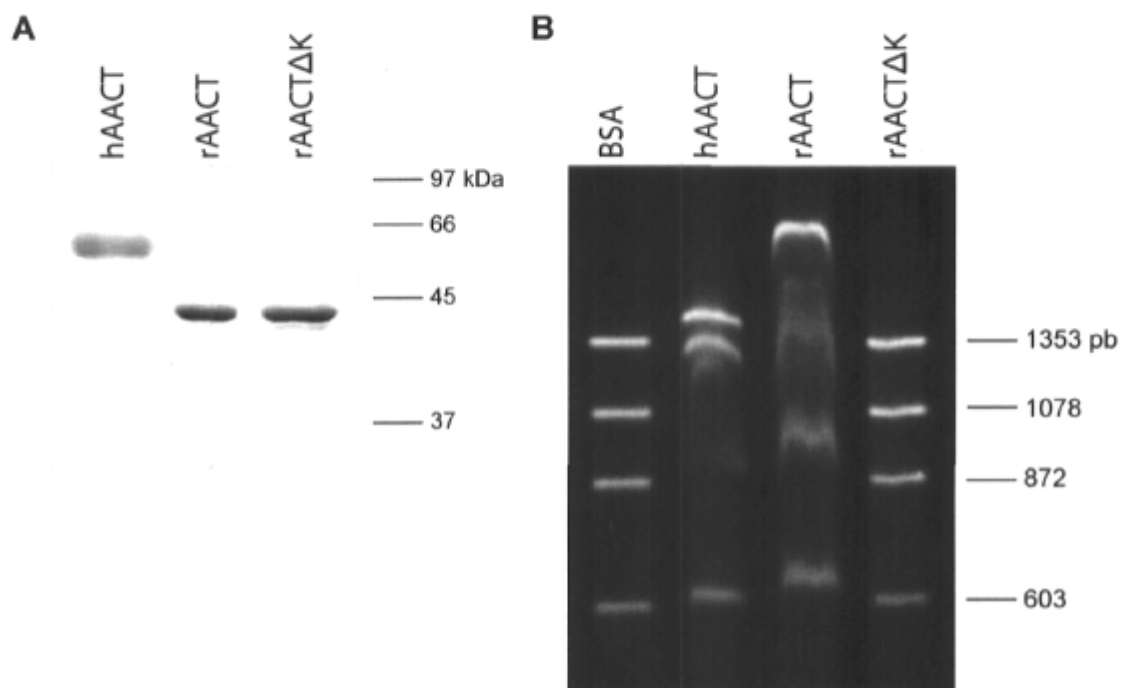
(**Figure III.1A**). The first construct encoded the original polypeptide chain of human AACT (GenBank Accession Number #J05176) fused at the C terminus of the protein fusion partner glutathione *S*-transferase (GST) used as an affinity tag for purification (Smith and Johnson, 1988). The second construct was identical to the first construct, except for two mutations replacing the residues Lys-391 and Lys-396 by threonines. The first inhibitor variant, referred to as rAACT, corresponded to an unglycosylated, cytosolic form of the inhibitor shown earlier to inhibit chymotrypsin *in vitro* (Rubin et al., 1990). The second variant, referred to as rAACT $\Delta$ K, corresponded to a mutant version of rAACT [presumably] unable to bind DNA (Naidoo et al., 1995). Gene constructs coding for the same two inhibitor variants attached to the green fluorescent protein (GFP) were also devised in parallel (**Figure III.1B**), as a tool to monitor the relative distribution of either polypeptides transiently expressed in tobacco leaf cells by confocal microscopy.



**Figure III.1 Gene constructs devised for the experiments. (A)** Gene constructs for heterologous expression of rAACT and rAACT $\Delta$ K in *E. coli*. The serpin variants were expressed as glutathione *S*-transferase (GST) fusion. The GST tag was removed by cleavage with human factor X<sub>a</sub> before enzymatic and stability assays. **(B)** Gene constructs for transit monitoring in tobacco leaf epidermal cells. The serpins were expressed fused to GFP. The constructs include a double version of the CaMV35S promoter (2xCaMV35S), the tobacco etch virus enhancer sequence (ES), the pA35S terminator sequence (TS), and the rAACT- or rAACT $\Delta$ K-encoding sequence.

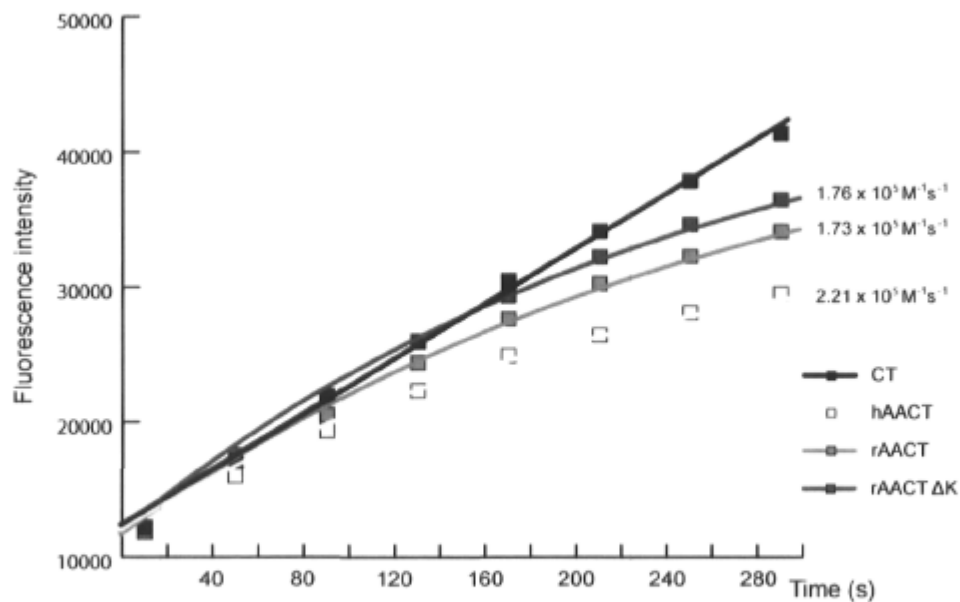
#### III.4.1 AACT and AACT $\Delta$ K exhibit comparable inhibitory activities against chymotrypsin

rAACT and rAACT $\Delta$ K were first expressed in and purified from *E. coli* to compare their DNA-binding activity and chymotrypsin inhibitory activity. Both inhibitors were purified from bacterial crude extracts using the GST gene fusion system (Michaud et al., 1994), and cleaved from the GST moiety with human factor X<sub>a</sub> to give ~45-kDa forms of the protein, about 20 kDa smaller than the original, glycosylated AACT purified from human plasma (**Figure III.2A**). In line with the previously reported ability of AACT to form non-specific complexes with DNA strings (Naidoo et al., 1995), human AACT and rAACT influenced the migration of 'ΦX174-HaeIII' DNA fragments resolved by a gel shift assay, giving a strongly altered DNA band pattern in agarose gels after staining with ethidium bromide, compared to the pattern observed with no protein added (**Figure III.2B**). In sharp contrast, rAACT $\Delta$ K had no visible effect on migration of the same DNA fragments, giving a DNA band pattern similar to the negative control (**Figure III.2B**). On the other hand, rAACT and rAACT $\Delta$ K showed a similar inhibitory efficiency against chymotrypsin (CT), as determined by protease inhibitory assays using the synthetic fluorimetric substrate *succinyl*-Ala-Ala-Pro-Phe-methylcoumarin (**Figure III.3**). Comparable  $k_{inh}$  values of 1.73 and 1.76 M<sup>-1</sup>s<sup>-1</sup> were obtained, respectively, for the two rAACT:chymotrypsin and rAACT $\Delta$ K:chymotrypsin complexes, similar to the  $k_{inh}$  value of 2.21 M<sup>-1</sup>s<sup>-1</sup> measured for human AACT interacting with the same enzyme (**Figure III.3**). These observations, while suggesting a limited steric impact of the C-terminal DNA-binding region on the reactive center loop of AACT involved in protease inhibition, also confirm the documented role of residues Lys-391 and Lys-396 for AACT DNA-binding activity (Naidoo et al., 1995) and the potential of rAACT $\Delta$ K as a model to assess the role of DNA-binding activity in the subcellular distribution of AACT in plant cells.



**Figure III.2 DNA-binding activity of rAACT and rAACTΔK.** (A) rAACT and rAACTΔK expressed in and purified from *E.coli* using the GST gene expression system. The purified serpins were visualized here by Coomassie Blue staining following 12% (w/v) SDS-PAGE. hAACT, human AACT, purified from human plasma. (B) DNA-binding activity of hAACT, rAACT and rAACTΔK, as assessed by gel shift assay. The serpins were incubated with  $\Phi$ X174Hae-III DNA strings (Naidoo et al., 1995), and then resolved in 2% (w/v) agarose gels. Serpin variants interacting with DNA induce an altered DNA band pattern during electrophoretic migration, compared to the unaltered DNA pattern observed in the presence of bovine serum albumin (BSA) used as a negative control.





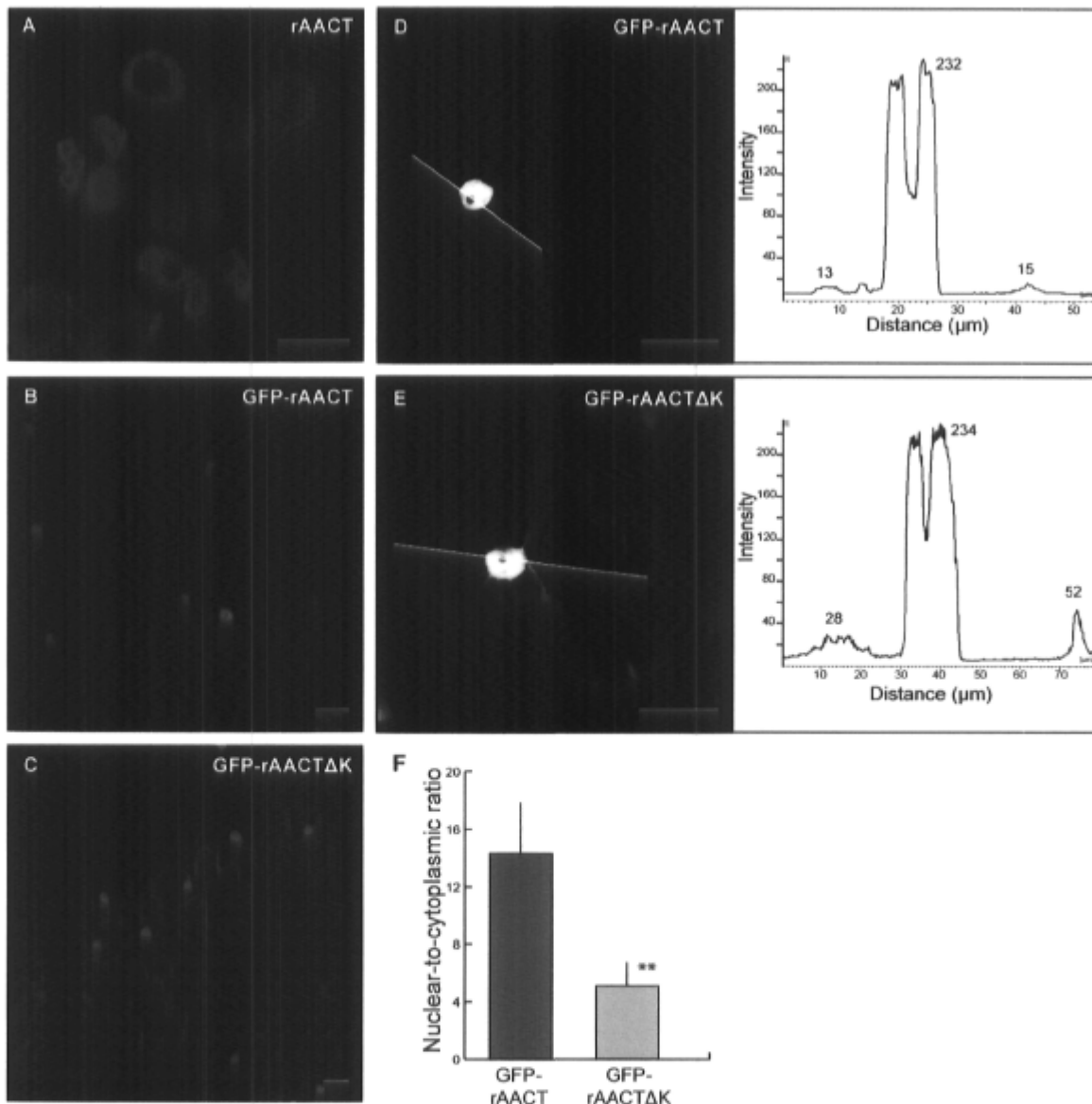
**Figure III.3 Antichymotrypsin activity of rAACT and rAACTΔK.** Data are presented as progress curves of chymotrypsin (CT) activity under pseudo-first order conditions, in the presence of human AACT (hAACT), rAACT or rAACTΔK or with no inhibitor added (positive control). Inhibitory constants ( $k_{inh}$ ) were calculated after empirically estimating  $k_{obs}$  and  $K_M$  constants (see Materials and methods).

#### III.4.2 The C-terminal DNA-binding site is involved in the nuclear retention of AACT

Confocal microscopy experiments were carried out to assess the impact of DNA-binding activity on the distribution of rAACT transiently expressed as a GFP fusion in agro-infiltrated tobacco leaf epidermal cells (**Figure III.4**). We observed recently that rAACT stably expressed in the cytosol of BY-2 tobacco cells was actually found predominantly in the nucleus, giving only a faint fluorescence signal in the cytosolic compartment after immunodetection with FITC-conjugated antibodies (Benchabane et al., 2007; also see **Figure III.4A**). In a similar way, the GFP-rAACT fusion transiently expressed in tobacco leaf epidermal cells was detected here predominantly in the nucleus, giving a merely detectable signal in the cytosol despite the absence of well characterized targeting signals in the transgene sequence (**Figure III.4B**). By comparison, the GFP-rAACTΔK fusion was detected in both the nucleus and the cytosol (**Figure III.4C**), which clearly suggests a direct effect of the lysine to threonine point mutations on cellular transit of the recombinant inhibitor *in planta*. Fluorescence intensity profiles were monitored across a number of

epidermal cells to estimate the relative amounts of AACT in the nuclear and cytosolic compartments of rAACT- and rAACT $\Delta$ K-expressing cells (**Figures III.4D, E**). Overall, the nuclear-to-cytosol fluorescence signal was decreased by about threefold after mutating the DNA-binding site in the inhibitor sequence (**Figure III.4F**), thus suggesting a significant relocalization of the recombinant serpin at the cell scale considering the importance of the cytosolic milieu relative to the overall volume of epidermal cells.

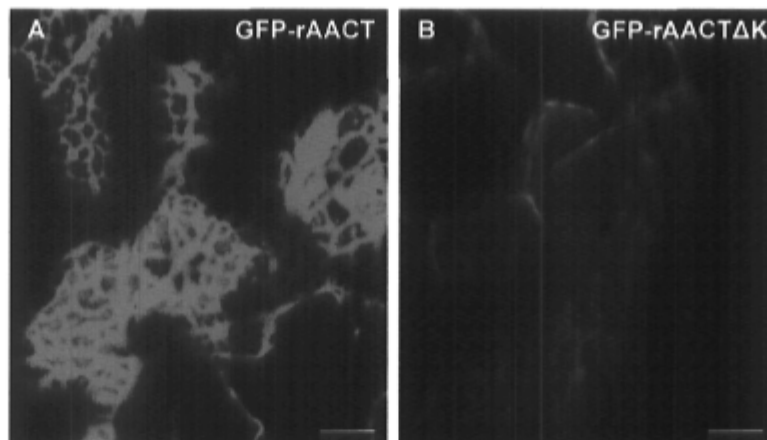
It is not clear at this point how AACT enters in and is retained into the nucleus, or how the C-terminal DNA-binding site of this protein influences its distribution in the cell. Several proteins present in the cytosol freely diffuse into the nucleus through nuclear pore complexes, but the ~40-kDa cut-off generally associated with this phenomenon (Heese-Peck and Raikhel, 1998; Gorlich and Kutay, 1999) may hardly explain the transit of the GFP-AACT fusions, that exhibit a molecular weight higher than 70 kDa. Large nuclear proteins, that often harbour an NLS sequence, typically rely on a facilitated translocation mechanism involving specific protein transporters, the importins, that ensure their efficient passage through the nuclear pores *via* an energy-dependent process (Merkle, 2004; Pemberton and Paschal, 2005). No NLS-like sequence could be identified, however, in the primary sequence of AACT searched against the PSORT (Nakai and Horton, 1999) and NLSdb (Cokol et al., 2000) databases, which suggests an alternative translocation scheme for this protein. Similar to AACT, other proteins of the serpin protein superfamily were shown earlier to exhibit a nuclear or dual nuclear/cytosolic distribution *in vivo*, including serpins with a classical NLS sequence (Irving et al., 2002) and serpins with no well defined nuclear import signal, translocated into the nucleus by a mechanism still to be elucidated (Bird et al., 2001). In a similar way, AACT import into the nucleus, that was not compromised for the rAACT $\Delta$ K mutant (see **Figure III.4C**), could involve a yet to be resolved translocation process not relying on a typical NLS sequence. The DNA-binding activity of AACT, later along the overall translocation pathway, would rather be involved in nuclear retention (Schmidt-Zachmann et al., 1993), thereby contributing to the preponderance of rAACT in the nucleus, in contrast with rAACT $\Delta$ K showing a more balanced dual distribution between the nucleus and the cytosol.



**Figure III.4 Nucleocytoplasmic distribution of rAACT and rAACTΔK in tobacco leaf epidermal cells.** AACT variants were stably expressed in transformed tobacco BY-2 cells (A) or expressed transiently as GFP fusions in tobacco leaf epidermal cells (B-E). (A) rAACT stably expressed in the cytosol of transformed BY-2 cells. The serpin was detected by immunolocalization with specific AACT antibodies (see Benchabane et al., 2007). (B, C) GFP-rAACT (B) and GFP-rAACTΔK (C) transiently expressed in the epidermis of tobacco leaf cells. (D, E) Quantification of nuclear to cytoplasmic distribution of GFP-rAACT (D) and GFP-rAACTΔK (E) transiently expressed in tobacco leaf epidermal cells. Confocal images on left panels show the virtual cross-sections used for quantitation. The right panels illustrate fluorescence intensities measured along this cross-section for either serpins. (F) Nucleus-to-cytoplasm ratios measured for GFP-rAACT and GFP-rAACTΔK transiently expressed in tobacco leaf cells. Data are the mean of nine independent measures  $\pm$  SE. \*\*, statistically significant at  $P < 0.001$ . Bars = 20  $\mu$ m.

### III.4.3 DNA protects recombinant AACT *in vitro*

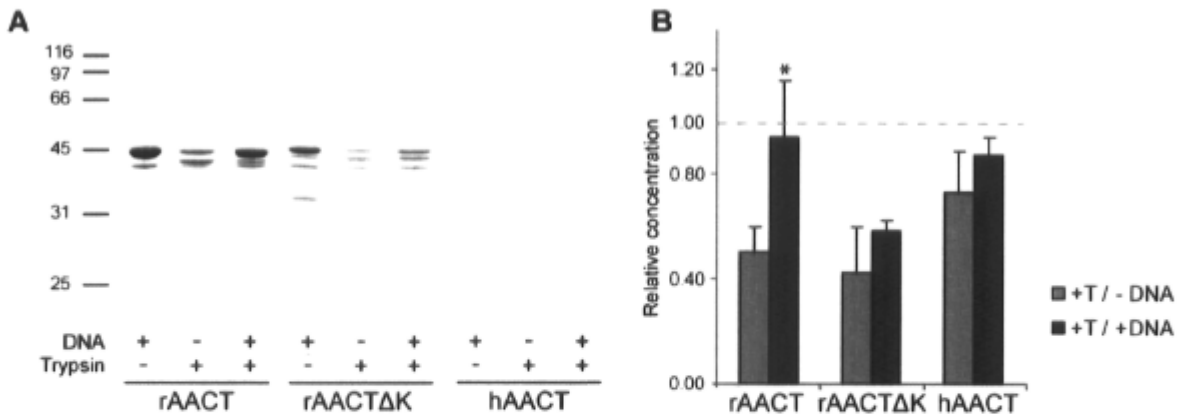
From a practical viewpoint, the DNA-binding activity and nuclear localization of AACT could also contribute to its overall stability *in planta*. Along with the distinct distribution patterns of rAACT and rAACT $\Delta$ K in tobacco cells, microscopy experiments revealed a bias in terms of protein signals detected under similar experimental conditions, that were systematically higher for the non-mutant variant (**Figure III.5**). The cytosol is a reducing, protease-rich milieu poorly compatible with the accumulation of several recombinant proteins (Faye et al., 2005; Goulet and Michaud, 2006), and a translocation diversion toward the nucleus could in some cases lead to higher protein yields by simply providing the protein with a milder microenvironment for folding and accumulation (Streatfield et al., 2003). In theory, the presence of DNA in the nucleus could also have a stabilizing effect on AACT in the presence of nontarget proteases by physically stabilizing the structure of the protein *in vivo* or by masking protease-susceptible sites during the protein:DNA interaction.



**Figure III.5** Typical cortical views of tobacco leaf epidermal cells expressing GFP-rAACT (A) or GFP-rAACT $\Delta$ K (B). Images were captured under identical experimental conditions. Bars = 10  $\mu$ m.

To test this possibility, human AACT, rAACT and rAACT $\Delta$ K were challenged with bovine trypsin, a poorly-specific serine protease insensitive to the inhibitor, in the presence or absence of DNA (**Figure III.6**). The three serpins were incubated for 60 min with the

protease, and then immunodetected with anti-serpin monoclonal antibodies following electrophoretic separation by SDS-PAGE. On a comparative basis, human AACT appeared resistant to tryptic digestion under the assay conditions, unlike the two bacterially-expressed, unglycosylated forms of the serpin readily producing truncated variants in the absence of DNA (**Figure III.6A**). Interestingly, the hydrolytic effect of trypsin on rAACT was strongly inhibited in the presence of DNA, allowing the integrity of the recombinant serpin to be almost fully preserved (**Figures III.6A, B**). By contrast, the presence of DNA had no positive impact on the integrity of rAACT $\Delta$ K, which strongly suggests an important protein stabilizing effect for the rAACT:DNA binding interaction in the presence of potentially deleterious proteases. A structural destabilization of rAACT $\Delta$ K resulting from the lysine to threonine point mutations in the C-terminal region could explain in part the deficient stability of this serpin variant, given the well established structural lability of serpins mutated at the primary structure level (Hopkins et al., 1993; Bottomley and Stone, 1998; Whisstock and Bottomley, 2006). On the other hand, the protease inhibitory potency of rAACT $\Delta$ K was comparable to the activity of rAACT (see **Figure III.3**), which suggests instead a well preserved tertiary structure following mutagenesis. Alternatively, the higher stability of rAACT in the presence of DNA could have been due to a local stabilizing effect on the protein structure at the site of interaction between the two macromolecules, or to a steric hindrance effect of DNA masking protease-susceptible sites on the serpin polypeptide chain. The DNA-binding activity of AACT, from this perspective, likely substituted the apparent protective effect of glycan chains on glycosylated AACT by possibly creating some steric hindrance around protease-susceptible sites, as observed earlier for the glycan moiety of other glycoproteins (Rudd et al., 1994; Kundra and Kornfeld, 1999; Pratap et al., 2000).



**Figure III.6 Stability of free or DNA linked rAACT, rAACTΔK and hAACT following trypsin partial degradation. (A)** SDS-PAGE analysis of partial trypsin proteolysis of AACT in a free form or bounded to DNA. **(B)** Relative AACT concentration linked (+DNA) or not to DNA (-DNA) following partial trypsin degradation (+T) as estimated by densitometry. Each datum is the mean of three independent measures  $\pm$  SE. \* Pairwise comparisons of each variant in presence or not of DNA highlight statistical significant differences  $P < 0.05$ .

### III.5 Conclusion

Numerous studies have been published over the years about the biological roles and structure of AACT (Rubin et al., 1990; Baumann et al., 1991; Baker et al., 2007), but little information is still available about the glycan moiety attached to this protein. A number of biological functions have been attributed to glycan chains on glycoproteins (Crocker and Feizi, 1996; Helenius and Aebi, 2004), that may help ensure the actual effectiveness of recombinant proteins in a clinical context (Gomord et al., 2004; Sethuraman and Stadheim, 2006). N-glycan structures, for instance, contribute to the effector functions of therapeutic antibodies (Ferrara et al., 2006), while the absence of glycosylation was shown to alter the half-life of some proteins and to cause their rapid clearance from biological fluids (Higuchi et al., 1992). Considering the stabilizing effect of DNA on rAACT and the high lability of this protein along the protein secretory pathway of tobacco cells (see Benchabane et al., 2007), the unexpected localization of rAACT in the nucleus of plant cells could represent a useful compromise for the efficient heterologous production of AACT with full protease inhibitory potency. Additional studies are required now to

determine the actual potential of plant protein factories for the efficient production of AACT, and to assess the activity of AACT *ex planta*, in a clinical context.

### III.6 Materials and methods

#### III.6.1 Gene constructs

Four gene constructs encoding either rAACT or rAACT $\Delta$ K fused to the C terminus of GST or GFP were produced for the experiments (see **Figure III.1**), using the plasmids pGEX/GST-rAACT (Benchabane et al., 1997; accompanying paper) and pUC/eGFP (Clontech, Mountain View, CA, USA) as source DNA templates for PCR amplification. The GST-rAACT fusion was expressed in *E. coli* as described recently (Benchabane et al., 2007). The AACT $\Delta$ K-encoding sequence was generated by PCR from pGEX/GST-rAACT using the following primers: 5'-AAGGA TCCAA AACAG CCCAC TTGAC G-3' and 5'-AAGAA TTCCT AGGCT TGCCT GGGAT TGGTG **ACTGT** TGC-3', including *Bam*H1 and *Eco*R1 restriction sites for subsequent cloning (underlined) and alternative G nucleotides for K (Lys) to T (Thr) point mutations (in bold). The *acct $\Delta$ k* amplicon was introduced within the *Eco*RI and *Bam*HI cloning sites of the expression vector pGEX-3X (Amersham Bioscience, Baie d'Urfé, QC, Canada), in frame with the GST-encoding sequence. The fusion constructs were validated by automatic sequencing, and the resulting vectors introduced into *E. coli* strain BL21 for heterologous expression and purification (see below).

For transient expression in tobacco leaves, the *gfp* transgene sequence was first amplified by PCR from the pUC/eGFP vector using the following primers: 5'-ATCCA TGGTG AGCAA GG-3' and 5'-AA TCT AGA CTT GTA CAG CTC GTC-3' including either an *Nco*I or *Xba*I restriction sequence (underlined); and then subcloned into the *Nco*I-*Xba*I cloning site of a modified pUC-19 vector (Fermentas Life Science, Burlington, ON, Canada) integrating a double version of the constitutive cauliflower mosaic virus 35S promoter, an enhancer sequence of the tobacco etch virus, and the PA35S terminator sequence (pUC19-eGPF). The rAACT- and rAACT $\Delta$ K-encoding sequences were PCR-amplified from the pGEX/GST-rAACT vector using the following primers: 5'-AATCT AGAAA CAGCC CACTT GACGA GGAGA ATCTG-3'; and 5'-AATCT AGACT TAGGC

TTGCT TGGGA TTGGT GACTT TGC-3' (for rAACT) or 5'-AATCT AGACT TAGGC TTGCT TGGGA TTGGT GACTT TGC-3' (for rAACT $\Delta$ K), each including an *Xba*I restriction site (underlined). The *aact* amplicons were then ligated within the *Xba*I cloning site of pUC19-eGFP, in frame with the *gfp* coding sequence. The transgene fusions were then validated by automatic sequencing, and finally transferred into the plant expression vector pCambia 2300 (CAMBIA, Canberra, Australia). The resulting constructs were introduced into *Agrobacterium tumefaciens* LBA4404 for transient expression in tobacco leaf cells (see below).

### III.6.2 Heterologous expression in *E. coli*

Heterologous expression of rAACT and rAACT $\Delta$ K in *E. coli* was carried out using the GST gene fusion system (Smith and Johnson, 1988), as described earlier for other protease inhibitors (Michaud et al., 1994). In brief, 1.5 ml of an overnight *E. coli* culture was added to 150 ml of LB medium containing 100  $\mu$ g/ml ampicillin, and then incubated at 37°C with vigorous agitation until reaching an OD<sub>600</sub> of 0.4-0.6. Transgene expression was induced by the addition of isopropyl-thiogalactoside at 200  $\mu$ M final concentration (Sigma, Oakville, ON, Canada), and the bacterial cells were allowed to grow for an additional 16 h at 37°C, before being harvested by centrifugation at 14,000 *g* for 10 min at room temperature. The bacterial pellet was submitted to several freeze-and-thaw cycles, and then suspended in 50 mM Tris lysis buffer, pH 9.0, containing 5% (w/v) sucrose, 50 mM EDTA and 5% (v/v) Triton X-100. After 5 min incubation on ice, the bacterial lysate was centrifuged at 20,000 *g* for 10 min at 4°C, and the supernatant containing soluble proteins was saved for protein purification. Affinity purification of the GST fusions was performed using reduced glutathione-sepharose beads as described (Michaud et al., 1994), and the GST moiety was removed by cleavage with human factor X<sub>a</sub> according to the supplier's instructions (Novagen, San Diego, CA, USA). Purity of the inhibitors was visualized by SDS-PAGE and immunodetection using anti-AACT primary antibodies. Protein concentrations were determined by densitometric analysis of Coomassie-blue-stained gels following SDS-PAGE, using the PHORETIX™ 2-D image analysis software, v. 2004 (Nonlinear Dynamics, Durham, NC, USA) and bovine serum albumin (Sigma) as a protein standard.



### **III.6.3 Transient expression in tobacco leaves**

Transient expression of GFP-rAACT and GFP-rAACT $\Delta$ K in epidermal cells of *Nicotiana tabacum*, cv. Xanthi leaves was performed as described earlier (Saint-Jore et al., 2002). In brief, two-days-old cultures of *Agrobacterium tumefaciens*, strain LBA4404 harboring the AACT-expressing plasmids were washed and resuspended in a 50 mM MES, pH 5.6, infiltration medium containing 0.5% (w/v) sucrose, 2 mM Na<sub>3</sub>PO<sub>4</sub> and 100 mM acetosyringone. The bacterial suspension was infiltrated into the abaxial epidermis of tobacco leaves using a syringe, and the plants were grown for two to three days at room temperature. Recombinant protein expression and subcellular location were monitored by confocal microscopy 24 and 48 h after infiltration.

### **III.6.4 Confocal fluorescence microscopy**

Leaf discs from leaves transiently expressing the recombinant fusions were sampled randomly and used for GFP fluorescence monitoring. Fluorescence images were captured with a Leica TCS SP2 AOBS confocal laser scanning microscope (Leica Microsystems, Rueil-Malmaison, France). Images were generated using a 488-nm argon ion laser source, and fluorescence was recorded with a photomultiplier setup in the 493–538 nm range. Image patterns were processed using the LCS Lite software, v. 2.61 (Leica Microsystems). Fluorescence relative intensities in the nucleus and the cytosol were determined along a virtual line crossing the nucleus and the cytosol of each cell observed.

### **III.6.5 DNA gel shift assay**

The DNA-binding activity of AACT was monitored using a gel DNA-shift assay, as described earlier by Naidoo et al. (1995). One  $\mu$ g of AACT was first incubated with 500 ng of  $\Phi$ X174-HaeIII DNA material (New England BioLabs, Pickering, Ontario) for 20 min at room temperature in 10 mM Tris, pH 8.0, containing 1 mM EDTA (total volume = 20  $\mu$ l) to allow for the protein:DNA interaction. The mixture was then resolved by 2% (w/v) agarose gel electrophoresis, and DNA migration bands were visualized for each AACT variant after staining with ethidium bromide. Bovine serum albumin (Sigma) incubated with 500 ng of DNA was used as a negative control.

### III.6.6 Chymotrypsin inhibitory activity

Rate constants for the association of human AACT, rAACT and rAACTΔK with bovine chymotrypsin (E.C.3.4.21.1; Sigma) were determined under pseudo-first order conditions by the monitoring of progress curves. The AACT variants (40 nM) were incubated in protease assay buffer (50 mM Tris-HCl, pH 8.0, containing 50 mM NaCl; total volume of 100 μl) in the presence of chymotrypsin (2 nM) and Suc-AAPF-MCA fluorogenic substrate (70 nM; Peptides International, Louisville, KY, USA). Chymotrypsin activity was monitored by the monitoring of substrate hydrolysis curves using a Fluostar Polstar Galaxy spectrofluorimeter (BMG, Offenburg, Germany). Data were collected over 300 sec using excitation and emission filters of 360 and 450 nm, respectively. Measurements were made in triplicate and averaged for each serpin:chymotrypsin assessed. Apparent inhibition constants ( $k_{obs}$ ) were estimated by a non-linear regression fit of each progress curve to the following model (Schechter and Plotnick, 2004):

$$[P] = [P]_0 + \frac{v_0}{k_{obs}} (1 - e^{-k_{obs}t})$$

where  $[P]$  is the concentration of product at time  $t$ ,  $[P]_0$  the concentration of product at time 0,  $v_0$  the (initial) rate of substrate hydrolysis at time 0, and  $k_{obs}$  the apparent first order rate constant of inhibition. Second order inhibition rate constants ( $k_{inh}$ ) were determined after empirically assessing a  $K_m$  value for the chymotrypsin:substrate interaction, using the following equation:

$$k_{inh} = \frac{k_{obs} (1 + \frac{[S]_0}{K_M})}{[I]_0}$$

where  $[S]_0$  and  $[I]_0$  represent, respectively, the concentration of substrate and the concentration of inhibitor at time 0. A  $K_m$  value of 10 μM was estimated for chymotrypsin interacting with Suc-AAPF-MCA.

### **III.6.7 Trypsin stability assay**

The relative stability of human AACT, rAACT and rAACT $\Delta$ K in the presence of non-target proteases was assessed with bovine trypsin (E.C.3.4.21.4; Sigma) in the presence or absence of DNA. Each serpin was first incubated for 20 min at room temperature in 100 mM Tris-HCl, pH 8.3 (total volume = 20  $\mu$ l) in the presence or absence of 6  $\mu$ g of  $\Phi$ X174-HaeIII DNA (New England BioLabs), and then challenged for 30 min with trypsin (0.03  $\mu$ g) added to the mixture. Proteolysis was stopped by the addition of SDS-PAGE sample buffer (Laemmli, 1970) and heating for 5 min at 100°C. The samples were resolved by 12% (w/v) SDS-PAGE (Laemmli, 1970), electrotransferred onto a nitrocellulose membrane, and immunodetected by standard procedures using anti-AACT monoclonal antibodies (USBiologicals, Swampscott, MA, USA; Benchabane et al., 2007). The amount of intact AACT in immunoblots was estimated by densitometry using Image Scanner (Amersham Pharmacia, Baie d'Urfé, Québec) and the PHORETIX™ 2-D image analysis software, v. 2004 (Nonlinear Dynamics, Durham,NC, USA).

### **Acknowledgements**

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the French Ministère de la Recherche, and the National Science and Engineering Research Council (NSERC) of Canada. M.Benchabane was the recipient of an NSERC graduate fellowship.

## **CHAPITRE IV**

### **Discussion et perspectives**

## **IV.1 L'AACT peut être exprimée dans un système d'expression hétérologue végétal**

De nombreuses protéines recombinantes produites dans des systèmes végétaux montrent des rendements et une qualité insuffisants pour pouvoir être commercialisées. Ces difficultés sont souvent attribuées à des modifications post-traductionnelles inappropriées. Dans le cadre de cette thèse, nous avons voulu étudier le potentiel des plantes à produire une glycoprotéine complexe, l'AACT humaine. Nous nous sommes particulièrement intéressés à l'influence de la compartimentation cellulaire sur le rendement, la qualité et les modifications post-traductionnelles de cette protéine exprimée dans une culture cellulaire de *N. tabacum* cv. BY-2. Comme plusieurs études suggèrent que le rendement en protéines recombinantes et leurs modifications post-traductionnelles varient en fonction de l'adressage cellulaire (Stoger et al., 2002; Streatfield et al., 2003; Peng et al., 2006; Petruccelli et al., 2006), nous avons dirigé l'AACT vers quatre compartiments cellulaires : le cytosol, le RE, la vacuole et l'apoplasme. Nos résultats nous ont permis de mettre en évidence la capacité des cellules BY-2 à produire et à accumuler l'AACT dans ces quatre compartiments. Nous avons aussi démontré, par des expériences de type *pulse-chase*, que l'AACT présentait un patron de maturation différent selon sa localisation cellulaire.

### **IV.1.1 La compartimentation cellulaire influence la N-glycosylation de l'AACT recombinante**

Nos études sur l'AACT humaine ont montré qu'elle présentait une proportion de sucres plus importante que l'AACT recombinante produite dans les plantes. Cette différence s'explique présumément par la présence sur l'AACT humaine de N-glycannes tri-antennés, voire tétra-antennés, des structures typiques des N-glycannes de mammifères. De plus, les N-glycannes de l'AACT humaine comportent des résidus d'acides sialiques terminaux, monosaccharides très importants pour assurer la demi-vie des protéines plasmatiques (Matsumoto et al., 1995; Kelm et Schauer, 1997; Ni et al., 2000).

En ce qui concerne l'AACT recombinante produite dans les cellules de tabac, qu'elle soit adressée vers le système endomembranaire de sécrétion (RE, vacuole,

apoplasme) ou vers le cytosol, elle a pu être détectée dans la cellule végétale, avec des profils d'expression forts différents. L'AACT recombinante produite dans le cytosol est accumulée logiquement sous forme non glycosylée et de manière homogène. En revanche, lorsqu'un signal peptidique provoque le transfert de l'AACT dans le RE, elle est N-glycosylée. Ses glycanes subissent, tout au long de son cheminement dans le système endomembranaire, une maturation menant à l'obtention de glycanes complexes typiques des végétaux, comportant notamment des résidus  $\alpha$ -1,3 fucose et/ou  $\beta$ -1,2 xylose.

Il existe donc des différences importantes dans la structure des N-glycanes de l'AACT humaine et de l'AACT recombinante produite dans la cellule végétale. Ces différences peuvent être préjudiciables à l'utilisation de ces protéines recombinantes en thérapie. La présence de glycanes complexes xylosylés et  $\alpha$ -1,3 fucosylés s'avère potentiellement immunogène chez les mammifères (Faye et al., 1993; van Ree et al., 2000; Bardor et al., 2003a). De plus, les glycoprotéines synthétisées par les cellules végétales ne comportent pas de terminaisons d'acide N-acétylneuraminique (Seveno et al., 2004; Zeleny et al., 2006), or différents travaux ont montré que cette absence d'acides sialiques pouvait accélérer la clairance dans le foie d'une protéine d'intérêt *in vivo* (Matsumoto et al., 1995; Ni et al., 2000).

Nous avons vérifié, *in vitro*, que l'activité inhibitrice de l'AACT ne semblait pas dépendre de la N-glycosylation. En effet, l'AACT recombinante bactérienne, non-glycosylée, et l'AACT humaine présentaient *in vitro* des activités comparables d'inhibition de la chymotrypsine. En revanche, les propriétés biochimiques *in vivo* de l'AACT pourraient être affectées par ces différences de glycosylation. Actuellement, aucune information n'est disponible concernant les effets que des glycanes de structure typiquement végétale pourraient avoir sur l'affinité ou la spécificité d'une serpine recombinante envers ses protéases-cibles. Nos travaux auraient pu nous permettre d'ouvrir la voie dans ce domaine. Malheureusement, nous n'avons pas pu effectuer les analyses nécessaires, faute d'obtenir des variantes stables purifiées de l'AACT adressée vers les compartiments RE, apoplasme et vacuole. De plus, dans chacun de ces trois compartiments, les formes de l'AACT majoritairement accumulées étaient partiellement dégradées par des protéases végétales. Il apparaît donc difficile d'étudier l'activité de cette protéine en fonction de sa seule glycosylation. Pour aller plus loin dans cette voie, il sera

nécessaire de surmonter les problèmes de stabilité et de faible rendement en AACT recombinante dans les cellules végétales.

Plusieurs études récentes ont montré le potentiel du RE et du signal de rétention tétrapeptidique K/HDEL pour l'accumulation, la stabilité et le rendement en protéines recombinantes (Schouten et al., 1996; Stoger et al., 2002; Ma et al., 2003; Peng et al., 2006; Petruccelli et al., 2006), une approche qui favorise de plus l'obtention de glycanes polymannosidiques non-immunogènes (Sriraman et al., 2004; Gomord et al., 2005). Or, dans nos études, les effets escomptés du signal KDEL n'ont pas pu tous être obtenus. L'AACT s'est accumulée dans le milieu intracellulaire, mais elle était aussi détectée dans le milieu extracellulaire des cellules BY-2. Dans l'état actuel de nos connaissances, deux hypothèses nous permettraient d'expliquer ce phénomène. D'une part, le clivage partiel ou total du signal KDEL, au cours du transport de l'AACT dans le RE ou l'appareil de Golgi, pourrait conduire la protéine à poursuivre son transit dans le système endomembranaire de sécrétion. D'autre part, les protéines pourraient échapper au mécanisme de récupération et de rétrotranslocation malgré la présence du signal KDEL. Ce dernier, situé à l'extrémité C-terminale de la protéine, pourrait par exemple être partiellement inaccessible au sein de la structure tridimensionnelle de l'AACT, l'empêchant d'interagir avec son récepteur spécifique. Certains chercheurs ont d'ailleurs proposé l'emploi d'un signal KDEL modifié, présentant un bras de liaison supplémentaire d'acides aminés, pour améliorer son accessibilité (Pagny et al., 2000; Sriraman et al., 2004; Petruccelli et al., 2006).

#### **IV.1.2 La maturation protéolytique de l'AACT varie en fonction de sa destination**

L'AACT exprimée dans les cellules BY-2 s'est également avérée sensible à la protéolyse, un phénomène bien démontré ici dans le milieu extracellulaire. L'environnement cellulaire, en particulier l'apoplasme, est très riche en protéases potentiellement dommageables pour la stabilité des protéines recombinantes (Hellwig et al., 2004; Schaller, 2004; Schiermeyer et al., 2005). Nous avons mis en évidence, dans le milieu extracellulaire, que des protéases de types sérine sont principalement impliquées dans la dégradation de l'AACT recombinante. Toutefois, nos analyses de type *pulse-chase* laissent entrevoir que cette maturation protéolytique imprévue était probablement initiée

dans le milieu intracellulaire, au sein du RE et/ou de l'appareil de Golgi pour se poursuivre dans le milieu extracellulaire et provoquer l'accumulation d'un produit de clivage. Nous n'avons malheureusement pas pu définir les sites de clivage sur l'AACT ou les protéases intracellulaires impliquées. Il est maintenant évident que des protéases jalonnent le système endomembranaire dès le RE (Bass et al., 2000; Schmitz et Herzog, 2004) et une caractérisation plus approfondie de la protéolyse, bien que représentant un défi technique de taille, serait un atout considérable pour la production de l'AACT glycosylée dans le système végétal. Une connaissance plus précise des sites sensibles aux protéases et des protéases impliquées permettrait, en particulier, d'élaborer des stratégies moléculaires pour prévenir cette dégradation (Goulet et Michaud, 2006). Par exemple, des mutations ponctuelles au niveau des sites de clivage (Outchkourov et al., 2003) ou la co-expression de l'AACT avec des IP spécifiques (Komarnytsky et al., 2006) pourraient éventuellement en minimiser la protéolyse.

Naturellement, ces stratégies ne doivent pas interférer avec le métabolisme de la plante ou altérer la stabilité conformationnelle de la serpine. L'activité inhibitrice des serpines est indissociable de leur structure flexible et de leur RCL exposée au solvant, qui agit comme appât pour les protéases à inhiber. En revanche, la nature exposée du RCL la rend très vulnérable à un clivage protéolytique éventuel par des protéases non sensibles, ce qui résulterait systématiquement en une serpine clivée inactive. *In vivo*, l'inactivation d'une serpine par clivage protéolytique permet de contrôler la concentration de serpines fonctionnelles dans un tissu et offre un niveau supplémentaire de régulation de l'activité protéolytique (Gettins, 2002). En outre, le changement de conformation d'une serpine clivée peut amener de nouvelles fonctions physiologiques. Ainsi, l'AACT, qui s'avère être particulièrement sensible aux protéases exogènes microbiennes et mammifères, peut agir comme agent chemo-attractant et stimuler les neutrophiles, lorsqu'elle se présente sous sa forme clivée (Potempa et al., 1991; Gettins, 2002). Dans le cadre d'une production d'AACT recombinante à des fins de thérapie, l'obtention d'une serpine avec une RCL clivée, telle qu'observée dans cette étude pour les formes sécrétées, ne serait évidemment pas souhaitable.



## **IV.2 L'AACT exprimée dans le cytosol a une double localisation, nucléaire et cytoplasmique**

Les analyses de localisation ont révélé que lorsqu'elle était exprimée dans le cytosol, l'AACT, bien qu'en apparence stable, était retrouvée à la fois dans le cytosol et dans le noyau. Cette distribution était surprenante car une protéine de cette taille diffuse difficilement à travers le pore nucléaire (Heese-Peck et Raikhel, 1998; Gorlich et Kutay, 1999) et parce qu'aucun signal de localisation nucléaire n'a été rapporté dans la littérature ou n'a pu être prédit *in silico* pour l'AACT. Une localisation nucléaire avait pourtant déjà été décrite pour l'AACT humaine dans des cellules animales cancéreuses, par un mécanisme de translocation encore inconnu à ce jour (Takada et al., 1986; 1988b). Nous avons donc voulu contribuer à l'exploration de ce phénomène de translocation, qui semblait s'étendre, selon nos résultats, à l'AACT recombinante exprimée en système végétal.

Il est connu que l'AACT a la capacité d'interagir avec l'ADN, un attribut rare parmi les serpines (Rubin et al., 1990; Naidoo et al., 1995). Les fonctions exactes de cette interaction restent mal comprises; comme point de départ pour nos travaux, nous avons émis l'hypothèse que la liaison à l'ADN était impliquée dans la distribution nucléaire de l'AACT lorsque cette protéine n'est pas acheminée vers le système de sécrétion. Aussi avons-nous créé un mutant AACT, l'AACT $\Delta$ K, dont l'affinité pour l'ADN était grandement réduite. Ce mutant, caractérisé antérieurement par Naidoo et al. (1995), ne se lie pas à l'ADN mais son affinité pour la chymotrypsine demeure en revanche inchangée. Après confirmation de l'activité inhibitrice de la protéine mutée, l'AACT $\Delta$ K a été fusionnée à la GFP, après quoi la localisation nucléaire du mutant a pu être étudiée par microscopie confocale à fluorescence. Il s'est avéré que le mutant, toujours détecté dans le noyau, montrait par contre une plus forte accumulation cytosolique que la forme native. Ainsi, en modifiant l'affinité de la protéine pour l'ADN, nous étions parvenus à altérer la rétention nucléaire de l'AACT. Dans un contexte de moléculaire, il deviendrait donc envisageable, en mutant le site de liaison à l'ADN, de relocaliser partiellement l'AACT vers le cytosol sans en altérer l'activité. Mais est-ce réellement un avantage? Les rendements ou le processus de purification, par exemple, seraient-ils simplifiés par une accumulation cytosolique? À l'inverse, la stabilité relative de l'AACT exprimée dans le cytosol pourrait être l'effet de sa séquestration dans le noyau. À titre d'exemple, l'antigène Lt-B de *E. coli*

s'accumulait davantage dans le noyau que dans le cytoplasme de cellules de maïs, probablement grâce à une concentration moindre en protéases dans le premier compartiment (Streatfield et al., 2003). Plus encore, la capacité même de se lier à l'ADN pourrait contribuer à la stabilisation de l'AACT. À cet effet, nous avons pu démontrer que liée à l'ADN, l'AACT non glycosylée est moins sensible à une attaque protéolytique par la trypsine. En comparaison, l'AACT humaine glycosylée s'est avérée relativement résistante à une dégradation par la trypsine, même en l'absence d'ADN. La liaison à l'ADN semblait donc compenser, du moins en partie, la stabilisation naturellement observée en présence de N-glycannes. Ces observations faites *in vitro* ne pourront toutefois pas nécessairement être transposées *in planta*, dans l'environnement complexe de la cellule. Il importera donc, comme suite à ce projet de vérifier si dans sa forme non glycosylée, l'AACT ne montre pas une stabilité altérée *in vivo*.

L'interaction de l'AACT avec l'ADN et la localisation de cette protéine dans le noyau, en plus de stabiliser la protéine exprimée dans la plante, pourraient en outre lui procurer certaines fonctions physiologiques *in vivo*, utiles à des fins thérapeutiques. L'AACT est une protéine multifonctionnelle dont la concentration peut augmenter d'un facteur cinq en phase aigüe dans les sites d'inflammation. Si cette protéine permet la régulation de différentes protéases de type sérine, dont la cathepsine G (Baker et al., 2007), ses fonctions physiologiques associées à son affinité pour l'ADN, par contre, restent inconnues. Au sein du noyau, l'AACT s'est avérée être un modulateur de la réplication de l'ADN *via* l'inhibition de l'ADN primase et l'ADN polymérase  $\alpha$  (Tsuda et al., 1986; Takada et al., 1988a,b), des activités indépendantes de sa capacité d'inhiber des protéases. La mise en évidence de ces nouvelles fonctions, combinée à la présence d'AACT dans le noyau de cellules cancéreuses humaines, suggère que cette protéine pourrait influencer la formation des tumeurs (Tsuda et al., 1986; Takada et al., 1988a,b). Même si les rôles de la liaison à l'ADN et la localisation nucléaire de l'AACT sont encore mal cernés, inhiber ces propriétés pourrait éventuellement compromettre certaines de ses fonctions *in vivo*. Finalement, sur le plan appliqué, cette affinité de l'AACT pour l'ADN, en plus de rendre plus stable face aux protéases, pourrait s'avérer utile pour sa purification à l'aide de techniques de chromatographie d'affinité utilisant l'ADN comme ligand.

### IV.3 Les perspectives

Il est désormais bien connu que les plantes possèdent toute la machinerie cellulaire nécessaire à la synthèse de protéines portant des modifications post-traductionnelles complexes. De fait, plusieurs protéines recombinantes d'origine végétale sont actuellement commercialisées ou en voie de l'être (Ma et al., 2003; Marshall, 2006). Pourtant, l'intégration des séquences codantes de l'AACT au génome de cellules végétales, comme plusieurs autres exemples déjà répertoriés (Matsumoto et al., 1995; Sharp et Doran, 2001; Outchkourov et al., 2003), n'a permis ici d'obtenir qu'un produit hétérogène, avec des modifications post-traductionnelles non optimales. Loin de discréditer les systèmes de production végétaux, ces exemples témoignent de toute la complexité liée à la synthèse protéique. Ils illustrent aussi l'importance d'étudier chaque protéine individuellement et de développer des approches analytiques précises pour en évaluer l'intégrité (Badri et al., 2007a). Une fois les modifications post-traductionnelles caractérisées, il devient ensuite possible d'envisager des stratégies pour moduler ces maturations, en fonction des propriétés biologiques recherchées.

Face aux problèmes de stabilité des protéines recombinantes, plusieurs stratégies moléculaires peuvent être envisagées (Faye et al., 2005; Goulet et Michaud, 2006). En bref, ces stratégies reposent sur des modifications du système d'expression pour diminuer l'activité protéolytique ou sur des améliorations de la protéine recombinante comme, par exemple, des protéines-fusions ou des variantes modifiées au niveau des sites de clivages sensibles aux protéases de la plante. Comme nous l'avons vu dans le cadre de ces travaux, une connaissance minimale de la plante hôte et de la protéine modèle s'imposent. Les serpinines, en particulier, sont des protéines très vulnérables à l'attaque par des protéases non-cibles. En revanche, il existe dans la nature des serpinines avec un certain niveau de résistance aux attaques protéolytiques. À titre d'exemple, l'inhibiteur  $\alpha$ -PI de l'opossum (*Didelphis virginiana*) est résistant aux protéases de venin de serpent (*Crotalus adamanteus*) alors que l'équivalent humain ne l'est pas, malgré une forte homologie de séquence entre les deux inhibiteurs (Catanese et Kress, 1993). Cette résistance proviendrait d'une évolution adaptative de la protéine à l'échelle moléculaire, résultant de mutations au niveau du RCL, qui explique précisément la tolérance accrue de l'opossum aux morsures du serpent. D'une façon similaire, il serait possible de concevoir des serpinines recombinantes résistantes aux protéases exogènes par des approches

courantes de génie génétique. Toutefois, il est important de garder à l'esprit que les serpines ont des conformations natives peu stables (Whisstock et Bottomley, 2006) et que la mutation de certains acides aminés critiques peut mener, à des degrés variables, à une transition vers des formes latentes ou polymériques inactives (Hopkins et al., 1993; Bottomley et Chang, 1997; Bottomley et Stone, 1998; Whisstock et Bottomley, 2006).

L'AACT produite ici sous forme glycosylée dans les cellules de tabac comportait en outre des glycannes complexes, typiques des végétaux, s'éloignant des structures présentes chez l'Homme. Au cours des dernières années, des succès tangibles ont été obtenus dans la modification de la voie de biosynthèse des glycannes chez les plantes (Bakker et al., 2001; Strasser et al., 2004; Bakker et al., 2006; Cox et al., 2006; Paccalet et al., 2007; Schähs et al., 2007) et donnent bon espoir de bientôt pouvoir produire des protéines végétales avec des structures glycaniques de type humain. Mais bien plus que de vouloir imiter les structures des glycannes humains, les recherches actuelles se penchent de plus en plus vers une structure de glycanne spécifique pour chaque protéine permettant de moduler ses propriétés pharmacologiques (Walsh et Jefferis, 2006). À cet égard, une connaissance approfondie de la structure et des fonctions des glucosidases et glycosyltransférases végétales, liée aux avancées technologiques récentes en transgénèse et sur la modification de la voie de biosynthèse des glycannes chez les plantes, laissent entrevoir le potentiel des plantes comme des plateformes appropriées pour l'ingénierie de la glycosylation et l'obtention de glycannes spécifiques à chaque protéine (Saint-Jore-Dupas et al., 2007).

Différentes localisations cellulaires sont par ailleurs souvent exploitées pour améliorer les rendements en protéines, pour en favoriser l'intégrité et pour en modifier les maturations post-traductionnelles. Alors que beaucoup d'études exploitent des signaux d'adressage pour choisir la localisation de protéines modèles, peu d'études vérifient *in situ*, par microscopie, le résultat de leurs efforts d'adressage. Pourtant, il est déjà documenté que certaines protéines exprimées dans des tissus hétérologues peuvent s'accumuler dans des compartiments imprévus (Chikwamba et al., 2003; Hood, 2004; Drakakaki et al., 2006). L'AACT exprimée dans le cytosol s'est aussi accumulée ici dans le noyau; comment expliquer une telle redirection de la protéine ? Un signal de localisation nucléaire (NLS) pourrait exister avec, par exemple, une structure primaire non-classique, défiant les limites des outils de prédiction actuels. La translocation nucléaire pourrait aussi

résulter d'une interaction avec d'autres protéines comportant des NLS (Zhou et al., 2006b), ou encore être dépendante des processus de O-glycosylation (Rondanino et al., 2003; Monsigny et al., 2004), voire peut-être de la SUMOylation (Pichler et Melchior, 2002; Matarasso et al., 2005). Ces deux dernières hypothèses apparaissent peu probables dans le cas de l'AACT, mais elles illustrent bien la multiplicité des mécanismes cellulaires possibles pour un même phénomène.

Finalement, plusieurs évidences suggèrent que l'AACT est une protéine multifonctionnelle (Tsuda et al., 1986; Takada et al., 1988a,b; Baker et al., 2007). Pour l'AACT comme pour de nombreuses autres protéines à plusieurs fonctions, la localisation cellulaire peut être cruciale pour la modulation des différentes activités (Jeffery, 1999; Arnoys et Wang, 2007). Pour cette raison, l'altération d'un signal d'adressage ou de rétention endogène de la protéine, par exemple pour en améliorer le taux de production, risque aussi d'affecter sa localisation cellulaire dans l'organisme cible et de diminuer son potentiel multifonctionnel dans un contexte thérapeutique. De plus, un signal de localisation cryptique, bien que déjouant les plans de production *in planta* initialement prévus, pourrait tout aussi bien, dans certain cas, s'avérer une alternative utile. Produire des protéines recombinantes aussi complexes que l'AACT dans les systèmes végétaux devient un jeu sophistiqué impliquant un ensemble de processus complexes allant d'une prise en compte des relations *structure-fonction* de la protéine d'intérêt à une compréhension adéquate des modifications post-traductionnelles et de la machinerie cellulaire chez la plante hôte.

## **CHAPITRE V**

### **Références bibliographiques**

- Adam, Z., and Clarke, A.K.** (2002). Cutting edge of chloroplast proteolysis. *Trends Plant Sci* **7**, 451-456.
- Altmann, F.** (2007). The role of protein glycosylation in allergy. *Int Arch Allergy Immunol* **142**, 99-115.
- Anfinsen, C.B.** (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.
- Apweiler, R., Hermjakob, H., and Sharon, N.** (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8.
- Arcalis, E., Marcel, S., Altmann, F., Kolarich, D., Drakakaki, G., Fischer, R., Christou, P., and Stoger, E.** (2004). Unexpected deposition patterns of recombinant proteins in post-endoplasmic reticulum compartments of wheat endosperm. *Plant Physiol* **136**, 3457-3466.
- Arnoys, E.J., and Wang, J.L.** (2007). Dual localization: Proteins in extracellular and intracellular compartments. *Acta Histochem* **109**, 89-110.
- Artsaenko, O., Kettig, B., Fiedler, U., Conrad, U., and Doring, K.** (1998). Potato tubers as a biofactory for recombinant antibodies. *Mol Breeding* **4**, 313-319.
- Badri, M.A., Rivard, D., Coenen, K., and Michaud, D.** (2007a). Unintended molecular interactions in transgenic plants expressing clinically-useful proteins - The case of bovine aprotinin traveling the potato leaf cell secretory pathway. *Proteomics in press*.
- Badri, M.A., Rivard, D., Coenen, K., and Michaud, D.** (2007b). A SELDI-TOF MS procedure for the detection, quantification and preliminary characterization of low-molecular-weight recombinant proteins expressed in transgenic plants. *Proteomics in press*.
- Baker, C., Belbin, O., Kalsheker, N., and Morgan, K.** (2007). SERPINA3 (aka alpha-1-antichymotrypsin). *Front Biosci* **12**, 2821-2835.
- Bakker, H., Rouwendal, G.J., Karnoup, A.S., Florack, D.E., Stoopen, G.M., Helsper, J.P., van Ree, R., van Die, I., and Bosch, D.** (2006). An antibody produced in tobacco expressing a hybrid beta-1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes. *Proc Natl Acad Sci U S A* **103**, 7577-7582.
- Bakker, H., Bardor, M., Molthoff, J.W., Gomord, V., Elbers, I., Stevens, L.H., Jordi, W., Lommen, A., Faye, L., Lerouge, P., and Bosch, D.** (2001). Galactose-extended glycans of antibodies produced by transgenic plants. *Proc Natl Acad Sci U S A* **98**, 2899-2904.
- Bardor, M., Faveeuw, C., Fitchette, A.C., Gilbert, D., Galas, L., Trottein, F., Faye, L., and Lerouge, P.** (2003a). Immunoreactivity in mammals of two typical plant glyco-epitopes, core alpha(1,3)-fucose and core xylose. *Glycobiology* **13**, 427-434.
- Bardor, M., Loutelier-Bourhis, C., Paccalet, T., Cosette, P., Fitchette, A.C., Vezina, L.P., Trepanier, S., Dargis, M., Lemieux, R., Lange, C., Faye, L., and Lerouge, P.** (2003b). Monoclonal C5-1 antibody produced in transgenic alfalfa plants exhibits a N-glycosylation that is homogenous and suitable for glyco-engineering into human-compatible structures. *Plant Biotechnol J* **1**, 451-462.
- Bass, J., Turck, C., Rouard, M., and Steiner, D.F.** (2000). Furin-mediated processing in the early secretory pathway: Sequential cleavage and degradation of misfolded insulin receptors. *Proc Natl Acad Sci U S A* **97**, 11905-11909.
- Bateman, K.S., Congiu, M., Tregear, G.W., Clarke, A.E., and Anderson, M.A.** (1997). Bacitracin significantly reduces degradation of peptides in plant cell cultures. *Biotechnol Bioeng* **53**, 226-231.

- Baumann, U., Huber, R., Bode, W., Grosse, D., Lesjak, M., and Laurell, C.B.** (1991). Crystal structure of cleaved human alpha 1-antichymotrypsin at 2.7 Å resolution and its comparison with other serpins. *J Mol Biol* **218**, 595-606.
- Beers, E.P., Jones, A.M., and Dickerman, A.W.** (2004). The S8 serine, C1A cysteine and A1 aspartic protease families in *Arabidopsis*. *Phytochemistry* **65**, 43-58.
- Benchabane, M., Rivard, D., Girard, C., and Michaud, D.** (2007). Companion protease inhibitors to protect recombinant proteins in transgenic plant extracts. *Methods in Biotechnology* **in press**.
- Benning, L.N., Whisstock, J.C., Sun, J., Bird, P.I., and Bottomley, S.P.** (2004). The human serpin proteinase inhibitor-9 self-associates at physiological temperatures. *Protein Sci* **13**, 1859-1864.
- Birch-Machin, I., Newell, C.A., Hibberd, J.M., and Gray, J.C.** (2004). Accumulation of rotavirus VP6 protein in chloroplasts of transplastomic tobacco is limited by protein stability. *Plant Biotechnol J* **2**, 261-270.
- Bird, C.H., Blink, E.J., Hirst, C.E., Buzza, M.S., Steele, P.M., Sun, J., Jans, D.A., and Bird, P.I.** (2001). Nucleocytoplasmic distribution of the ovalbumin serpin PI-9 requires a nonconventional nuclear import pathway and the export factor Crm1. *Mol Cell Biol* **21**, 5396-5407.
- Bird, P.I., Pak, S.C., Worrall, D.M., and Bottomley, S.P.** (2004). Production of recombinant serpins in *Escherichia coli*. *Methods* **32**, 169-176.
- Boisson, M., Gomord, V., Audran, C., Berger, N., Dubreucq, B., Granier, F., Lerouge, P., Faye, L., Caboche, M., and Lepiniec, L.** (2001). *Arabidopsis* glucosidase I mutants reveal a critical role of N-glycan trimming in seed development. *Embo J* **20**, 1010-1019.
- Bos, I.G., de Bruin, E.C., Karuntu, Y.A., Modderman, P.W., Eldering, E., and Hack, C.E.** (2003). Recombinant human C1-inhibitor produced in *Pichia pastoris* has the same inhibitory capacity as plasma C1-inhibitor. *Biochim Biophys Acta* **1648**, 75-83.
- Bottomley, S.P., and Chang, W.-S.W.** (1997). The effects of reactive center loop length upon serpin polymerisation. *Biochem Biophys Res Commun* **241**, 264-269.
- Bottomley, S.P., and Stone, S.R.** (1998). Protein engineering of chimeric serpins: An investigation into effects of the serpin scaffold and reactive centre loop. *Protein Eng* **11**, 1243-1247.
- Brooks, S.A.** (2006). Protein glycosylation in diverse cell systems: Implications for modification and analysis of recombinant proteins. *Expert Rev Proteomics* **3**, 345-359.
- Brunelle, F., and Michaud, D.** (2000). Substrates and inhibitors useful in protease characterization. In *Recombinant protease inhibitors in plants*, D. Michaud, ed (Georgetown TX: Landes Bioscience), pp. 224-227.
- Brunelle, F., Girard, C., Cloutier, C., and Michaud, D.** (2005). A hybrid, broad-spectrum inhibitor of Colorado potato beetle aspartate and cysteine digestive proteinases. *Arch Insect Biochem Physiol* **60**, 20-31.
- Burda, P., Jakob, C.A., Beinhauer, J., Hegemann, J.H., and Aebi, M.** (1999). Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases. *Glycobiology* **9**, 617-625.
- Cabanes-Macheteau, M., Fichette-Laine, A.C., Loutelier-Bourhis, C., Lange, C., Vine, N.D., Ma, J.K., Lerouge, P., and Faye, L.** (1999). N-Glycosylation of a mouse IgG expressed in transgenic tobacco plants. *Glycobiology* **9**, 365-372.
- Callis, J.** (1995). Regulation of protein degradation. *Plant Cell* **7**, 845-857.



- Carpentier, S.C., Witters, E., Laukens, K., Deckers, P., Swennen, R., and Panis, B.** (2005). Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* **5**, 2497-2507.
- Catanese, J.J., and Kress, L.F.** (1993). Opossum serum alpha 1-proteinase inhibitor: Purification, linear sequence, and resistance to inactivation by rattlesnake venom metalloproteinases. *Biochemistry* **32**, 509-515.
- Cereghino, J.L., and Cregg, J.M.** (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* **24**, 45-66.
- Cerioti, A., Duranti, M., and Bollini, R.** (1998). Effects of N-glycosylation on the folding and structure of plant proteins. *J Exp Bot* **49**, 1091-1103.
- Chang, W.S., and Lomas, D.A.** (1998). Latent alpha1-antichymotrypsin. A molecular explanation for the inactivation of alpha1-antichymotrypsin in chronic bronchitis and emphysema. *J Biol Chem* **273**, 3695-3701.
- Chargelegue, D., Vine, N.D., van Dolleweerd, C.J., Drake, P.M., and Ma, J.K.** (2000). A murine monoclonal antibody produced in transgenic plants with plant-specific glycans is not immunogenic in mice. *Transgenic Res* **9**, 187-194.
- Chikwamba, R.K., Scott, M.P., Mejia, L.B., Mason, H.S., and Wang, K.** (2003). Localization of a bacterial protein in starch granules of transgenic maize kernels. *Proc Natl Acad Sci U S A* **100**, 11127-11132.
- Christeller, J., and Laing, W.** (2005). Plant serine proteinase inhibitors. *Protein Pept Lett* **12**, 439-447.
- Cokol, M., Nair, R., and Rost, B.** (2000). Finding nuclear localization signals. *EMBO Rep* **1**, 411-415.
- Coleman, C.E., Herman, E.M., Takasaki, K., and Larkins, B.A.** (1996). The maize gamma-zein sequesters alpha-zein and stabilizes its accumulation in protein bodies of transgenic tobacco endosperm. *Plant Cell* **8**, 2335-2345.
- Coloma, M.J., Clift, A., Wims, L., and Morrison, S.L.** (2000). The role of carbohydrate in the assembly and function of polymeric IgG. *Mol Immunol* **37**, 1081-1090.
- Conrad, U., and Fiedler, U.** (1998). Compartment-specific accumulation of recombinant immunoglobulins in plant cells: An essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. *Plant Mol Biol* **38**, 101-109.
- Cox, K.M., Sterling, J.D., Regan, J.T., Gasdaska, J.R., Frantz, K.K., Peele, C.G., Black, A., Passmore, D., Moldovan-Loomis, C., Srinivasan, M., Cuisson, S., Cardarelli, P.M., and Dickey, L.F.** (2006). Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat Biotechnol* **24**, 1591-1597.
- Crocker, P.R., and Feizi, T.** (1996). Carbohydrate recognition systems: Functional triads in cell-cell interactions. *Curr Opin Struct Biol* **6**, 679-691.
- Dahl, S.W., Rasmussen, R., and Hejgaard, J.** (1996). Heterologous expression of three plant serpins with distinct inhibitory specificities. *J Biol Chem* **271**, 25083-25088.
- Daniell, H.** (2006). Production of biopharmaceuticals and vaccines in plants via the chloroplast genome. *Biotechnol J* **1**, 1071-1079.
- Daniell, H., Streatfield, S.J., and Wycoff, K.** (2001). Medical molecular farming: Production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci* **6**, 219-226.
- Daniell, H., Khan, M.S., and Allison, L.** (2002). Milestones in chloroplast genetic engineering: An environmentally friendly era in biotechnology. *Trends Plant Sci* **7**, 84-91.

- Daniell, H., Chebolu, S., Kumar, S., Singleton, M., and Falconer, R.** (2005). Chloroplast-derived vaccine antigens and other therapeutic proteins. *Vaccine* **23**, 1779-1783.
- De Block, M., Herrera-Estrella, L., Van Montagu, M., Schell, J., and Zambryski, P.** (1984). Expression of foreign genes in regenerated plants and in their progeny. *Embo J* **3**, 1681-1689.
- De Cosa, B., Moar, W., Lee, S.B., Miller, M., and Daniell, H.** (2001). Overexpression of the *Bt cry2Aa2* operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnol* **19**, 71-74.
- De Jaeger, G., Buys, E., Eeckhout, D., De Wilde, C., Jacobs, A., Kapila, J., Angenon, G., Van Montagu, M., Gerats, T., and Depicker, A.** (1999). High level accumulation of single-chain variable fragments in the cytosol of transgenic *Petunia hybrida*. *Eur J Biochem* **259**, 426-434.
- Dementiev, A., Dobo, J., and Gettins, P.G.** (2006). Active site distortion is sufficient for proteinase inhibition by serpins: Structure of the covalent complex of alpha1-proteinase inhibitor with porcine pancreatic elastase. *J Biol Chem* **281**, 3452-3457.
- Doran, P.M.** (2006). Foreign protein degradation and instability in plants and plant tissue cultures. *Trends Biotechnol* **24**, 426-432.
- Dove, A.** (2002). Uncorking the biomanufacturing bottleneck. *Nat Biotechnol* **20**, 777-779.
- Drakakaki, G., Marcel, S., Arcalis, E., Altmann, F., Gonzalez-Melendi, P., Fischer, R., Christou, P., and Stoger, E.** (2006). The intracellular fate of a recombinant protein is tissue dependent. *Plant Physiol* **141**, 578-586.
- Duranton, J., Boudier, C., Belorgey, D., Mellet, P., and Bieth, J.G.** (2000). DNA strongly impairs the inhibition of cathepsin G by alpha(1)- antichymotrypsin and alpha(1)-proteinase inhibitor. *J Biol Chem* **275**, 3787-3792.
- Dyck, M.K., Lacroix, D., Pothier, F., and Sirard, M.A.** (2003). Making recombinant proteins in animals--Different systems, different applications. *Trends Biotechnol* **21**, 394-399.
- Eklund, E.A., and Freeze, H.H.** (2006). The congenital disorders of glycosylation: A multifaceted group of syndromes. *NeuroRx* **3**, 254-263.
- Enfors, S.O.** (1992). Control of *in vivo* proteolysis in the production of recombinant proteins. *Trends Biotechnol* **10**, 310-315.
- Faye, L., Johnson, K.D., and Chrispeels, M.I.** (1986). Oligosaccharide side chains of glycoproteins that remain in the high-mannose form are not accessible to glycosidases. *Plant Physiol* **81**.
- Faye, L., Fitchette-Laine, A.C., Gomord, V., Chekkafi, A., and Driouich, A.** (1993). Detection, biosynthesis and some functions of glycans N-linked to plant secreted proteins. *Planta* **213-242**.
- Faye, L., Boulaflous, A., Benchabane, M., Gomord, V., and Michaud, D.** (2005). Protein modifications in the plant secretory pathway: Current status and practical implications in molecular pharming. *Vaccine* **23**, 1770-1778.
- Fernandez-San Millan, A., Mingo-Castel, A., Miller, M., and Daniell, H.** (2003). A chloroplast transgenic approach to hyper-express and purify Human Serum Albumin, a protein highly susceptible to proteolytic degradation. *Plant Biotechnol J* **1**, 71-79.
- Ferrara, C., Brunker, P., Suter, T., Moser, S., Puntener, U., and Umana, P.** (2006). Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-N-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. *Biotechnol Bioeng* **93**, 851-861.

- Fichette-Laine, A.C., Gomord, V., Chekkafi, A., and Faye, L.** (1994). Distribution of xylosylation and fucosylation in the plant Golgi apparatus. *Plant J* **5**, 673-682.
- Fitchette-Lainé, A.-C., Gomord, V., Cabanes, M., Michalski, J.-C., Saint Macary, M., Foucher, B., Cavelier, B., Hawes, C., Lerouge, P., and Faye, L.** (1997). N-glycans harboring the Lewis a epitope are expressed at the surface of plant cell. *Plant J* **12**, 1411-1417.
- Fiedler, U., and Conrad, U.** (1995). High-level production and long-term storage of engineered antibodies in transgenic tobacco seeds. *Biotechnology (N Y)* **13**, 1090-1093.
- Fiedler, U., Phillips, J., Artsaenko, O., and Conrad, U.** (1997). Optimization of scFv antibody production in transgenic plants. *Immunotechnology* **3**, 205-216.
- Fischer, R., Schumann, D., Zimmermann, S., Drossard, J., Sack, M., and Schillberg, S.** (1999). Expression and characterization of bispecific single-chain Fv fragments produced in transgenic plants. *Eur J Biochem* **262**, 810-816.
- Floss, D.M., Falkenburg, D., and Conrad, U.** (2007). Production of vaccines and therapeutic antibodies for veterinary applications in transgenic plants: An overview. *Transgenic Res* **16**, 315-332.
- Friedman, B., Vaddi, K., Preston, C., Mahon, E., Cataldo, J.R., and McPherson, J.M.** (1999). A comparison of the pharmacological properties of carbohydrate remodeled recombinant and placental-derived beta-glucocerebrosidase: Implications for clinical efficacy in treatment of Gaucher disease. *Blood* **93**, 2807-2816.
- Galili, G., Sengupta-Gopalan, C., and Ceriotti, A.** (1998). The endoplasmic reticulum of plant cells and its role in protein maturation and biogenesis of oil bodies. *Plant Mol Biol* **38**, 1-29.
- Garbino, J.E., Oosumi, T., and Belknap, W.R.** (1995). Isolation of a polyubiquitin promoter and its expression in transgenic potato plants. *Plant Physiol* **109**.
- Gaume, A., Komarnytsky, S., Borisjuk, N., and Raskin, I.** (2003). Rhizosecretion of recombinant proteins from plant hairy roots. *Plant Cell Rep* **21**, 1188-1193.
- Geli, M.I., Torrent, M., and Ludevid, D.** (1994). Two structural domains mediate two sequential events in [gamma]-zein targeting: Protein endoplasmic reticulum retention and protein body formation. *Plant Cell* **6**, 1911-1922.
- Gemmill, T.R., and Trimble, R.B.** (1999). Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim Biophys Acta* **1426**, 227-237.
- Gerngross, T.U.** (2004). Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat Biotechnol* **22**, 1409-1414.
- Gettins, G.W., Patston, P.A., and Olson, S.T.** (1996). Serpins: Structure, fonction and biology. R.G. Landes Texas.
- Gettins, P.G.** (2002). Serpin structure, mechanism, and function. *Chem Rev* **102**, 4751-4804.
- Gils, A., Pedersen, K.E., Skottrup, P., Christensen, A., Naessens, D., Deinum, J., Enghild, J.J., Declerck, P.J., and Andreasen, P.A.** (2003). Biochemical importance of glycosylation of plasminogen activator inhibitor-1. *Thromb Haemost* **90**, 206-217.
- Gils, M., Kandzia, R., Marillonnet, S., Klimyuk, V., and Gleba, Y.** (2005). High-yield production of authentic human growth hormone using a plant virus-based expression system. *Plant Biotechnol J* **3**, 613-620.
- Gnanasambandam, A., Polkinghorne, I.G., and Birch, R.G.** (2007). Heterologous signals allow efficient targeting of a nuclear-encoded fusion protein to plastids and endoplasmic reticulum in diverse plant species. *Plant Biotechnol J* **5**, 290-296.
- Gomord, V., and Faye, L.** (2004). Posttranslational modification of therapeutic proteins in plants. *Curr Opin Plant Biol* **7**, 171-181.

- Gomord, V., Chamberlain, P., Jefferis, R., and Faye, L.** (2005). Biopharmaceutical production in plants: Problems, solutions and opportunities. *Trends Biotechnol* **23**, 559-565.
- Gomord, V., Denmat, L.A., FichetteLaine, A.C., SatiatJeunemaitre, B., Hawes, C., and Faye, L.** (1997). The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. *Plant J* **11**, 313-325.
- Gomord, V., Fichette-Lainé, A.-C., Denmat, L.-A., Michaud, D., Faye, L., and eds (: ), p.** (1998). Production of foreign proteins in tobacco cell suspension culture. *Meth Biotechnol* **3**, 155–164.
- Gomord, V., Sourrouille, C., Fichette, A.C., Bardor, M., Pagny, S., Lerouge, P., and Faye, L.** (2004). Production and glycosylation of plant-made pharmaceuticals: The antibodies as a challenge. *Plant Biotechnol J* **2**, 83-100.
- Goputu, B., Hazes, B., Chang, W.S., Dafforn, T.R., Carrell, R.W., Read, R.J., and Lomas, D.A.** (2000). Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: Implications for inhibitory function and conformational disease. *Proc Natl Acad Sci U S A* **97**, 67-72.
- Gorlich, D., and Kutay, U.** (1999). Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* **15**, 607-660.
- Goulet, C., and Michaud, D.** (2006). Degradation and stabilization of recombinant proteins in plants. In *Floriculture, ornamental and plant biotechnology*, T.d. Silva, ed., Global Science Books, London UK, pp. 35-40.
- Grigoryev, S.A., Bednar, J., and Woodcock, C.L.** (1999). MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member. *J Biol Chem* **274**, 5626-5636.
- Grünewald, S., Matthijs, G., and Jaeken, J.** (2002). Congenital disorders of glycosylation: A review. *Pediatric Res* **52**, 618-624.
- Haltiwanger, R.S., and Lowe, J.B.** (2004). Role of glycosylation in development. *Annu Rev Biochem* **73**, 491-537.
- Hang, H.C., and Bertozzi, C.R.** (2005). The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* **13**, 5021-5034.
- Harrison, R.L., and Jarvis, D.L.** (2006). Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Adv Virus Res* **68**, 159-191.
- Hart, G.W., Housley, M.P., and Slawson, C.** (2007). Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017-1022.
- Heese-Peck, A., and Raikhel, N.V.** (1998). The nuclear pore complex. *Plant Mol Biol* **38**, 145-162.
- Hejgaard, J.** (1982). Purification and properties of protein Z - A major albumin of barley endosperm. *Physiol Plant* **54**, 174-182.
- Hejgaard, J.** (2001). Inhibitory serpins from rye with glutamine as P1 and P2 residues in the reactive center. *FEBS Lett* **488**, 149-153.
- Hejgaard, J., and Hauge, S.** (2002). Serpins of oat (*Avena sativa*) grain with distinct reactive centres and inhibitory specificity. *Physiol Plant* **116**, 155-163.
- Helenius, A., and Aebi, M.** (2001). Intracellular functions of N-linked glycans. *Science* **291**, 2364-2369.
- Helenius, A., and Aebi, M.** (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **73**, 1019-1049.
- Hellwig, S., Drossard, J., Twyman, R.M., and Fischer, R.** (2004). Plant cell cultures for the production of recombinant proteins. *Nat Biotechnol* **22**, 1415-1422.

- Henderson, J., Baulry, J.M., Ashford, D.A., Oliver, S.C., Hawes, C.R., Lazarus, C.M., Venis, M.A., and Napier, R.M.** (1997). Retention of maize auxin-binding protein in the endoplasmic reticulum: Quantifying escape and the role of auxin. *Planta* **202**, 313-323.
- Hiatt, A., Cafferkey, R., and Bowdish, K.** (1989). Production of antibodies in transgenic plants. *Nature* **342**, 76-78.
- Higuchi, M., Oh-eda, M., Kuboniwa, H., Tomonoh, K., Shimonaka, Y., and Ochi, N.** (1992). Role of sugar chains in the expression of the biological activity of human erythropoietin. *J Biol Chem* **267**, 7703-7709.
- Hondred, D., Walker, J.M., Mathews, D.E., and Vierstra, R.D.** (1999). Use of ubiquitin fusions to augment protein expression in transgenic plants. *Plant Physiol* **119**, 713-724.
- Hood, E.E.** (2004). Where, oh where has my protein gone? *Trends Biotechnol* **22**, 53-55.
- Hopkins, P.C.R., Carrell, R.W., and Stone, S.R.** (1993). Effects of mutations in the hinge region of serpins. *Biochemistry* **32**, 7650-7657.
- Horvath, H., Huang, J., Wong, O., Kohl, E., Okita, T., Kannangara, C.G., and von Wettstein, D.** (2000). The production of recombinant proteins in transgenic barley grains. *Proc Natl Acad Sci U S A* **97**, 1914-1919.
- Huang, J., Sutliff, T.D., Wu, L., Nandi, S., Bengel, K., Terashima, M., Ralston, A.H., Drohan, W., Huang, N., and Rodriguez, R.L.** (2001). Expression and purification of functional human alpha-1-Antitrypsin from cultured plant cells. *Biotechnol Progr* **17**, 126-133.
- Huntington, J.A.** (2006). Shape-shifting serpins--Advantages of a mobile mechanism. *Trends Biochem Sci* **31**, 427-435.
- Huntington, J.A., Read, R.J., and Carrell, R.W.** (2000). Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407**, 923-926.
- Hwang, S.R., Steineckert, B., Kohn, A., Palkovits, M., and Hook, V.Y.** (1999). Molecular studies define the primary structure of alpha1-antichymotrypsin (ACT) protease inhibitor in Alzheimer's disease brains. Comparison of ACT in hippocampus and liver. *J Biol Chem* **274**, 1821-1827.
- Hyunjong, B., Lee, D.S., and Hwang, I.** (2006). Dual targeting of xylanase to chloroplasts and peroxisomes as a means to increase protein accumulation in plant cells. *J Exp Bot* **57**, 161-169.
- Imperiali, B., and O'Connor, S.E.** (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* **3**, 643-649.
- Ioffe, E., and Stanley, P.** (1994). Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc Natl Acad Sci U S A* **91**, 728-732.
- Irving, J.A., Shushanov, S.S., Pike, R.N., Popova, E.Y., Bromme, D., Coetzer, T.H., Bottomley, S.P., Boulyenko, I.A., Grigoryev, S.A., and Whisstock, J.C.** (2002). Inhibitory activity of a heterochromatin-associated serpin (MENT) against papain-like cysteine proteinases affects chromatin structure and blocks cell proliferation. *J Biol Chem* **277**, 13192-13201.
- Jayakumar, A., Cataltepe, S., Kang, Y., Frederick, M.J., Mitsudo, K., Henderson, Y., Crawford, S.E., Silverman, G.A., and Clayman, G.L.** (2004). Production of serpins using baculovirus expression systems. *Methods* **32**, 177-184.
- Jeffery, C.J.** (1999). Moonlighting proteins. *Trends Biochem Sci* **24**, 8-11.
- Jenkins, N., and Curling, E.M.** (1994). Glycosylation of recombinant proteins: Problems and prospects. *Enzyme Microb Technol* **16**, 354-364.
- Jervis, L., and Pierpoint, W.S.** (1989). Purification technologies for plant-proteins. *J Biotechnol* **11**, 161-198.



- Jiang, X., Oohira, K., Iwasaki, Y., Nakano, H., Ichihara, S., and Yamane, T.** (2002). Reduction of protein degradation by use of protease-deficient mutants in cell-free protein synthesis system of *Escherichia coli*. *J Biosci Bioeng* **93**, 151-156.
- Kaplan, A.P.** (2001). C1 inhibitor deficiency: Hereditary and acquired forms. *J Investig Allergol Clin Immunol* **11**, 211-219.
- Karnaikhova, E., Ophir, Y., and Golding, B.** (2006). Recombinant human alpha-1 proteinase inhibitor: Towards therapeutic use. *Amino Acids* **30**, 317-332.
- Karnoup, A.S., Turkelson, V., and Anderson, W.H.K.** (2005). O-Linked glycosylation in maize-expressed human IgA1. *Glycobiology* **15**, 965-981.
- Kelly, C.A., Laskowski, M., Jr., and Qasim, M.A.** (2005). The role of scaffolding in standard mechanism serine proteinase inhibitors. *Protein Pept Lett* **12**, 465-471.
- Kelm, S., and Schauer, R.** (1997). Sialic acids in molecular and cellular interactions. *Int Rev Cytol* **175**, 137-240.
- Khalkhali-Ellis, Z.** (2006). Maspin: The new frontier. *Clin Cancer Res* **12**, 7279-7283.
- Kirk, D.D., and Webb, S.R.** (2005). The next 15 years: Taking plant-made vaccines beyond proof of concept. *Immunol Cell Biol* **83**, 248-256.
- Ko, K., and Koprowski, H.** (2005). Plant biopharming of monoclonal antibodies. *Virus Res* **111**, 93-100.
- Ko, K., Tekoah, Y., Rudd, P.M., Harvey, D.J., Dwek, R.A., Spitsin, S., Hanlon, C.A., Rupprecht, C., Dietzschold, B., Golovkin, M., and Koprowski, H.** (2003). Function and glycosylation of plant-derived antiviral monoclonal antibody. *Proc Natl Acad Sci U S A* **100**, 8013-8018.
- Komarnytsky, S., Borisjuk, N.V., Borisjuk, L.G., Alam, M.Z., and Raskin, I.** (2000). Production of recombinant proteins in tobacco guttation fluid. *Plant Physiol* **124**, 927-933.
- Komarnytsky, S., Gaume, A., Garvey, A., Borisjuk, N., and Raskin, I.** (2004). A quick and efficient system for antibiotic-free expression of heterologous genes in tobacco roots. *Plant Cell Rep* **22**, 765-773.
- Komarnytsky, S., Borisjuk, N., Yakoby, N., Garvey, A., and Raskin, I.** (2006). Cosecretion of protease inhibitor stabilizes antibodies produced by plant roots. *Plant Physiol* **141**, 1185-1193.
- Koprivova, A., Stemmer, C., Altmann, F., Hoffmann, A., Kopriva, S., Gorr, G., Reski, R., and Decker, E.L.** (2004). Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic N-glycans. *Plant Biotechnol J* **2**, 517-523.
- Kowarik, M., Numao, S., Feldman, M.F., Schulz, B.L., Callewaert, N., Kiermaier, E., Catrein, I., and Aebi, M.** (2006a). N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148-1150.
- Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., Wacker, M., and Aebi, M.** (2006b). Definition of the bacterial N-glycosylation site consensus sequence. *Embo J* **25**, 1957-1966.
- Kumar, G.B.S., Ganapathi, T.R., Srinivas, L., Revathi, C.J., and Bapat, V.A.** (2006). Expression of hepatitis B surface antigen in potato hairy roots. *Plant Sci* **170**, 918-925.
- Kundra, R., and Kornfeld, S.** (1999). Asparagine-linked oligosaccharides protect Lamp-1 and Lamp-2 from intracellular proteolysis. *J Biol Chem* **274**, 31039-31046.
- Kurosaka, A., Yano, A., Itoh, N., Kuroda, Y., Nakagawa, T., and Kawasaki, T.** (1991). The structure of a neural specific carbohydrate epitope of horseradish peroxidase recognized by anti-horseradish peroxidase antiserum. *J Biol Chem* **266**, 4168-4172.

- Kwon, K.S., and Yu, M.H.** (1997). Effect of glycosylation on the stability of alpha1-antitrypsin toward urea denaturation and thermal deactivation. *Biochim Biophys Acta* **1335**, 265-272.
- Kwon, K.S., Song, M., and Yu, M.H.** (1995a). Purification and characterization of alpha 1-antitrypsin secreted by recombinant yeast *Saccharomyces diastaticus*. *J Biotechnol* **42**, 191-195.
- Kwon, K.S., Lee, S., and Yu, M.H.** (1995b). Refolding of alpha 1-antitrypsin expressed as inclusion bodies in *Escherichia coli*: Characterization of aggregation. *Biochim Biophys Acta* **1247**, 179-184.
- Kwon, T.H., Kim, Y.S., Lee, J.H., and Yang, M.S.** (2003). Production and secretion of biologically active human granulocyte-macrophage colony stimulating factor in transgenic tomato suspension cultures. *Biotechnol Lett* **25**, 1571-1574.
- LaCasse, E.C., and Lefebvre, Y.A.** (1995). Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res* **23**, 1647-1656.
- LaCount, W., An, G.H., and Lee, J.M.** (1997). The effect of polyvinylpyrrolidone (PVP) on the heavy chain monoclonal antibody production from plant suspension cultures. *Biotechnol Lett* **19**, 93-96.
- Lamark, T., Ingebrigtsen, M., Bjornstad, C., Melkko, T., Mollnes, T.E., and Nielsen, E.W.** (2001). Expression of active human C1 inhibitor serpin domain in *Escherichia coli*. *Protein Expr Purif* **22**, 349-358.
- Law, R.H., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., and Whisstock, J.C.** (2006). An overview of the serpin superfamily. *Genome Biol* **7**, 216.
- Lawrence, D.A., Ginsberg, D., Day, D.E., Berkenpas, M.B., Verhamme, I.M., J-O., K., and Shore, J.D.** (1995). Serpin-protease complexes are trapped as stable acyl-enzyme intermediates. *J Biol Chem* **270**, 25309-25312.
- Lawrence, D.A., Olson, S.T., Muhammad, S., Day, D.E., Kvassman, J.O., Ginsburg, D., and Shore, J.D.** (2000). Partitioning of serpin-proteinase reactions between stable inhibition and substrate cleavage is regulated by the rate of serpin reactive center loop insertion into beta-sheet A. *J Biol Chem* **275**, 5839-5844.
- Lee, H.I., Gal, S., Newman, T.C., and Raikhel, N.V.** (1993). The *Arabidopsis* endoplasmic reticulum retention receptor functions in yeast. *Proc Natl Acad Sci U S A* **90**, 11433-11437.
- Lee, J.H., Kim, N.S., Kwon, T.H., Jang, Y.S., and Yang, M.S.** (2002). Increased production of human granulocyte-macrophage colony stimulating factor (hGM-CSF) by the addition of stabilizing polymer in plant suspension cultures. *J Biotechnol* **96**, 205-211.
- Lerouge, P., Cabance-Macheteau, M., Rayon, C., Fischette-Lainé, A.-C., Gomord, V., and Faye, L.** (1998). N-glycoprotein biosynthesis in plants: Recent developments and future trends. *Plant Mol Biol* **38**, 31-48.
- Lienard, D., Tran Dinh, O., van Oort, E., Van Overtvelt, L., Bonneau, C., Wambre, E., Bardor, M., Cosette, P., Didier-Laurent, A., de Borne, F.D., Delon, R., van Ree, R., Moingeon, P., Faye, L., and Gomord, V.** (2007). Suspension-cultured BY-2 tobacco cells produce and mature immunologically active house dust mite allergens. *Plant Biotechnol J* **5**, 93-108.
- Lin, J.F., and Wu, S.H.** (2004). Molecular events in senescing *Arabidopsis* leaves. *Plant J* **39**, 612-628.
- Lin, M., Rose-John, S., Grotzinger, J., Conrad, U., and Scheller, J.** (2006). Functional expression of a biologically active fragment of soluble gp130 as an ELP-fusion

- protein in transgenic plants: Purification via inverse transition cycling. *Biochem J* **398**, 577-583.
- Lomas, D.A., and Carrell, R.W.** (2002). Serpinopathies and the conformational dementias. *Nat Rev Genet* **3**, 759-768.
- Ma, J.K., Drake, P.M., and Christou, P.** (2003). The production of recombinant pharmaceutical proteins in plants. *Nat Rev Genet* **4**, 794-805.
- Ma, J.K., Drake, P.M., Chargelegue, D., Obregon, P., and Prada, A.** (2005a). Antibody processing and engineering in plants, and new strategies for vaccine production. *Vaccine* **23**, 1814-1818.
- Ma, J.K., Hiatt, A., Hein, M., Vine, N.D., Wang, F., Stabila, P., van Dolleweerd, C., Mostov, K., and Lehner, T.** (1995). Generation and assembly of secretory antibodies in plants. *Science* **268**, 716-719.
- Ma, J.K.C., Barros, E., Bock, R., Christou, P., Dale, P.J., Dix, P.J., Fischer, R., Irwin, J., Mahoney, R., Pezzotti, M., Schillberg, S., Sparrow, P., Stoger, E., Twyman, R.M., and Pharma, E.U.F.** (2005b). Molecular farming for new drugs and vaccines - Current perspectives on the production of pharmaceuticals in transgenic plants. *Embo Reports* **6**, 593-599.
- Ma, S., Huang, Y., Davis, A., Yin, Z., Mi, Q., Menassa, R., Brandle, J.E., and Jevnikar, A.M.** (2005c). Production of biologically active human interleukin-4 in transgenic tobacco and potato. *Plant Biotechnol J* **3**, 309-318.
- Macauley-Patrick, S., Fazenda, M.L., McNeil, B., and Harvey, L.M.** (2005). Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* **22**, 249-270.
- Mackenzie, S.A.** (2005). Plant organellar protein targeting: A traffic plan still under construction. *Trends Cell Biol* **15**, 548-554.
- Mainieri, D., Rossi, M., Archinti, M., Bellucci, M., De Marchis, F., Vavassori, S., Pompa, A., Arcioni, S., and Vitale, A.** (2004). Zeolin. A new recombinant storage protein constructed using maize gamma-zein and bean phaseolin. *Plant Physiol* **136**, 3447-3456.
- Maliga, P.** (2002). Engineering the plastid genome of higher plants. *Curr Opin Plant Biol* **5**, 164-172.
- Marshall.** (2006). PMPs in clinical trials and advanced PMIPs - Research findings presented at FinMed 2006, on 30th March 2006 - <http://www.molecularfarming.com/PMPs-and-PMIPs.html>
- Marty, F.** (1999). Plant vacuoles. *Plant Cell* **11**, 587-600.
- Marusic, C., Nuttall, J., Buriani, G., Lico, C., Lombardi, R., Baschieri, S., Benvenuto, E., and Frigerio, L.** (2007). Expression, intracellular targeting and purification of HIV Nef variants in tobacco cells. *BMC Biotechnol* **7**, 12.
- Mason, H.S., Lam, D.M., and Arntzen, C.J.** (1992). Expression of hepatitis B surface antigen in transgenic plants. *Proc Natl Acad Sci U S A* **89**, 11745-11749.
- Matarasso, N., Schuster, S., and Avni, A.** (2005). A novel plant cysteine protease has a dual function as a regulator of 1-aminocyclopropane-1-carboxylic acid synthase gene expression. *Plant Cell* **17**, 1205-1216.
- Matsumoto, S., Ikura, K., Ueda, M., and Sasaki, R.** (1995). Characterization of a human glycoprotein (erythropoietin) produced in cultured tobacco cells. *Plant Mol Biol* **27**, 1163-1172.
- Menkhaus, T.J., Bai, Y., Zhang, C.M., Nikolov, Z.L., and Glatz, C.E.** (2004). Considerations for the recovery of recombinant proteins from plants. *Biotechnol Progr* **20**, 1001-1014.



- Merkle, T.** (2004). Nucleo-cytoplasmic partitioning of proteins in plants: Implications for the regulation of environmental and developmental signalling. *Current Genetics* **45**, 231-260.
- Michaud, D.** (1998). Gel electrophoresis of proteolytic enzymes. *Analytica Chimica Acta* **372**, 173-185.
- Michaud, D., and Asselin, A.** (1995). Application to plant-proteins of gel-electrophoretic methods. *J Chromatogr A* **698**, 263-279.
- Michaud, D., Nguyen-Quoc, B., and Yelle, S.** (1994). Production of oryzacystatin I and II in *Escherichia coli* using the glutathione S-transferase gene fusion system. *Biotechnol Progr* **10**, 155-159.
- Michaud, D., Anguenot, R., and Brunelle, F.** (2005). Method for increasing protein content in plant cells. U.S. Patent application 2005, 2005/0055746.
- Michaud, D., Vrain, T.C., Gomord, V., and Faye, L.** (1998). Stability of recombinant proteins in plants. *Meth Biotechnol* **3**, 177-188.
- Miesenbock, G., and Rothman, J.E.** (1995). The capacity to retrieve escaped ER proteins extends to the trans-most cisterna of the Golgi stack. *J Cell Biol* **129**, 309-319.
- Minta, J.O.** (1981). The role of sialic acid in the functional activity and the hepatic clearance of C1-INH. *J Immunol* **126**, 245-249.
- Misaki, R., Fujiyama, K., and Seki, T.** (2006). Expression of human CMP-N-acetylneuraminic acid synthetase and CMP-sialic acid transporter in tobacco suspension-cultured cell. *Biochem Biophys Res Commun* **339**, 1184-1189.
- Mishra, S., Yadav, D.K., and Tuli, R.** (2006). Ubiquitin fusion enhances cholera toxin B subunit expression in transgenic plants and the plant-expressed protein binds GM1 receptors more efficiently. *J Biotechnol* **127**, 95-108.
- Mokrzycki-Issartel, N., Bouchon, B., Farrer, S., Berland, P., Laparra, H., Madelmont, J.C., and Theisen, M.** (2003). A transient tobacco expression system coupled to MALDI-TOF-MS allows validation of the impact of differential targeting on structure and activity of a recombinant therapeutic glycoprotein produced in plants. *FEBS Lett* **552**, 170-176.
- Monsigny, M., Rondanino, C., Duverger, E., Fajac, I., and Roche, A.C.** (2004). Glyco-dependent nuclear import of glycoproteins, glycoplexes and glycosylated plasmids. *Biochim Biophys Acta* **1673**, 94-103.
- Muntz, K.** (1998). Deposition of storage proteins. *Plant Mol Biol* **38**, 77-99.
- Muntz, K.** (2007). Protein dynamics and proteolysis in plant vacuoles. *J Exp Bot*.
- Murray, C., Sutherland, P.W., Phung, M.M., Lester, M.T., Marshall, R.K., and Christeller, J.T.** (2002). Expression of biotin-binding proteins, avidin and streptavidin, in plant tissues using plant vacuolar targeting sequences. *Transgenic Res* **11**, 199-214.
- Naidoo, N., Cooperman, B.S., Wang, Z.M., Liu, X.Z., and Rubin, H.** (1995). Identification of lysines within alpha 1-antichymotrypsin important for DNA binding. An unusual combination of DNA-binding elements. *J Biol Chem* **270**, 14548-14555.
- Nakai, K., and Horton, P.** (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**, 34-36.
- Navazio, L., Miuzzo, M., Royle, L., Baldan, B., Varotto, S., Merry, A.H., Harvey, D.J., Dwek, R.A., Rudd, P.M., and Mariani, P.** (2002). Monitoring endoplasmic reticulum-to-Golgi traffic of a plant calreticulin by protein glycosylation analysis. *Biochemistry* **41**, 14141-14149.
- Neuhaus, J.M., and Rogers, J.C.** (1998). Sorting of proteins to vacuoles in plant cells. *Plant Mol Biol* **38**, 127-144.

- Neuhaus, J.M., Sticher, L., Meins, F., and Boller, T.** (1991). A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. *Proc Natl Acad Sci U S A* **88**, 10362-10366.
- Ni, H., Blajchman, M.A., Ananthanarayanan, V.S., Smith, I.J., and Sheffield, W.P.** (2000). Mutation of any site of N-linked glycosylation accelerates the in vivo clearance of recombinant rabbit antithrombin. *Thromb Res* **99**, 407-415.
- Nicholson, L., Gonzalez-Melendi, P., van Dolleweerd, C., Tuck, H., Perrin, Y., Ma, J.K.C., Fischer, R., Christou, P., and Stoger, E.** (2005). A recombinant multimeric immunoglobulin expressed in rice shows assembly-dependent subcellular localization in endosperm cells. *Plant Biotechnol J* **3**, 115-127.
- Nuttall, J., Vine, N., Hadlington, J.L., Drake, P., Frigerio, L., and Ma, J.K.** (2002). ER-resident chaperone interactions with recombinant antibodies in transgenic plants. *Eur J Biochem* **269**, 6042-6051.
- Nygren, P.-A., Stahl, S., and Uhlén, M.** (1994). Engineering proteins to facilitate bioprocessing. *Trends Biotechnol* **12**, 181-188.
- Obregon, P., Chargelegue, D., Drake, P.M., Prada, A., Nuttall, J., Frigerio, L., and Ma, J.K.** (2006). HIV-1 p24-immunoglobulin fusion molecule: A new strategy for plant-based protein production. *Plant Biotechnol J* **4**, 195-207.
- Ohtsubo, K., and Marth, J.D.** (2006). Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867.
- Okamoto, T., Minamikawa, T., Edward, G., Vakharia, V., and Herman, E.** (1999). Posttranslational removal of the carboxyl-terminal KDEL of the cysteine protease SH-EP occurs prior to maturation of the enzyme. *J Biol Chem* **274**, 11390-11398.
- Ostergaard, H., Rasmussen, S.K., Roberts, T.H., and Hejgaard, J.** (2000). Inhibitory serpins from wheat grain with reactive centers resembling glutamine-rich repeats of prolamin storage proteins. *J Biol Chem* **275**, 33272-33279.
- Ostergaard, O., Finnie, C., Laugesen, S., Roepstorff, P., and Svensson, B.** (2004). Proteome analysis of barley seeds: Identification of major proteins from two-dimensional gels (pl 4-7). *Proteomics* **4**, 2437-2447.
- Otegui, M.S., Noh, Y.S., Martinez, D.E., Vila Petroff, M.G., Staehelin, L.A., Amasino, R.M., and Guimmet, J.J.** (2005). Senescence-associated vacuoles with intense proteolytic activity develop in leaves of *Arabidopsis* and soybean. *Plant J* **41**, 831-844.
- Otlewski, J., Jelen, F., Zakrzewska, M., and Oleksy, A.** (2005). The many faces of protease-protein inhibitor interaction. *Embo J* **24**, 1303-1310.
- Outchkourov, N.S., Rogelj, B., Strukelj, B., and Jongsma, M.A.** (2003). Expression of sea anemone equistatin in potato. Effects of plant proteases on heterologous protein production. *Plant Physiol* **133**, 379-390.
- Paccalet, T., Bardor, M., Rihouey, C., Delmas, F., Chevalier, C., D'Aoust, M.A., Faye, L., Vezina, L., Gomord, V., and Lerouge, P.** (2007). Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes. *Plant Biotechnol J* **5**, 16-25.
- Pagny, S., Cabanes-Macheteau, M., Gillikin, J.W., Leborgne-Castel, N., Lerouge, P., Boston, R.S., Faye, L., and Gomord, V.** (2000). Protein recycling from the Golgi apparatus to the endoplasmic reticulum in plants and its minor contribution to calreticulin retention. *Plant Cell* **12**, 739-756.
- Palacpac, N.Q., Yoshida, S., Sakai, H., Kimura, Y., Fujiyama, K., Yoshida, T., and Seki, T.** (1999). Stable expression of human beta1,4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns. *Proc Natl Acad Sci U S A* **96**, 4692-4697.

- Park, M., Kim, S.J., Vitale, A., and Hwang, I.** (2004). Identification of the protein storage vacuole and protein targeting to the vacuole in leaf cells of three plant species. *Plant Physiol* **134**, 625-639.
- Parrott, D., Yang, L., Shama, L., and Fischer, A.M.** (2005). Senescence is accelerated, and several proteases are induced by carbon "feast" conditions in barley (*Hordeum vulgare* L.) leaves. *Planta* **222**, 989-1000.
- Patel, J., Zhu, H., Menassa, R., Gyenis, L., Richman, A., and Brandle, J.** (2007). Elastin-like polypeptide fusions enhance the accumulation of recombinant proteins in tobacco leaves. *Transgenic Research* **16**, 239-249.
- Peeters, K., De Wilde, C., and Depicker, A.** (2001a). Highly efficient targeting and accumulation of a F(ab) fragment within the secretory pathway and apoplast of *Arabidopsis thaliana*. *Eur J Biochem* **268**, 4251-4260.
- Peeters, K., De Wilde, C., De Jaeger, G., Angenon, G., and Depicker, A.** (2001b). Production of antibodies and antibody fragments in plants. *Vaccine* **19**, 2756-2761.
- Pemberton, L.F., and Paschal, B.M.** (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6**, 187-198.
- Pemberton, P.A., and Bird, P.I.** (2004). Production of serpins using yeast expression systems. *Methods* **32**, 185-190.
- Peng, R.H., Yao, Q.H., Xiong, A.S., Cheng, Z.M., and Li, Y.** (2006). Codon-modifications and an endoplasmic reticulum-targeting sequence additively enhance expression of an *Aspergillus* phytase gene in transgenic canola. *Plant Cell Rep* **25**, 124-132.
- Peterson, F.C., and Gettins, P.G.** (2001). Insight into the mechanism of serpin-proteinase inhibition from 2D [1H-15N] NMR studies of the 69 kDa alpha 1-proteinase inhibitor Pittsburgh-trypsin covalent complex. *Biochemistry* **40**, 6284-6292.
- Petrucelli, S., Otegui, M.S., Lareu, F., Dinh, O.T., Fitchette, A.C., Circosta, A., Rumbo, M., Bardor, M., Carcamo, R., Gomord, V., and Beachy, R.N.** (2006). A KDEL-tagged monoclonal antibody is efficiently retained in the endoplasmic reticulum in leaves, but is both partially secreted and sorted to protein storage vacuoles in seeds. *Plant Biotechnol J* **4**, 511-527.
- Pichler, A., and Melchior, F.** (2002). Ubiquitin-related modifier SUMO1 and nucleocytoplasmic transport. *Traffic* **3**, 381-387.
- Pike, R.N., Bottomley, S.P., Irving, J.A., Bird, P.I., and Whisstock, J.C.** (2002). Serpins: Finely balanced conformational traps. *IUBMB Life* **54**, 1-7.
- Pogrebnyak, N., Golovkin, M., Andrianov, V., Spitsin, S., Smirnov, Y., Egolf, R., and Koprowski, H.** (2005). Severe acute respiratory syndrome (SARS) S protein production in plants: Development of recombinant vaccine. *Proc Natl Acad Sci U S A* **102**, 9062-9067.
- Potempa, J., Fedak, D., Dubin, A., Mast, A., and Travis, J.** (1991). Proteolytic inactivation of alpha-1-anti-chymotrypsin. Sites of cleavage and generation of chemotactic activity. *J. Biol. Chem.* **266**, 21482-21487.
- Potenza, C., Aleman, L., and Sengupta-Gopalan, C.** (2004). Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *In Vitro Cell Dev-PI* **40**, 1-22.
- Powell, L.M., and Pain, R.H.** (1992). Effects of glycosylation on the folding and stability of human, recombinant and cleaved alpha 1-antitrypsin. *J Mol Biol* **224**, 241-252.
- Pratap, J., Rajamohan, G., and Dikshit, K.L.** (2000). Characteristics of glycosylated streptokinase secreted from *Pichia pastoris*: Enhanced resistance of SK to proteolysis by glycosylation. *Appl Microbiol Biotechnol* **53**, 469-475.
- Rajabi-Memari, H., Jalali-Javaran, M., Rasee, M.J., Rahbarizadeh, F., Forouzandeh-Moghadam, M., and Esmaili, A.** (2006). Expression and characterization of a

- recombinant single-domain monoclonal antibody against MUC1 mucin in tobacco plants. *Hybridoma (Larchmt)* **25**, 209-215.
- Rasmussen, R., Dahl, S.W., Norgard, A., and Hejgaard, J.** (1996). A recombinant wheat serpin with inhibitory activity. *Plant Mol Biol* **30**, 673-677.
- Rawlings, N.D., Tolle, D.P., and Barrett, A.J.** (2004). Evolutionary families of peptidase inhibitors. *Biochem J* **378**, 705-716.
- Rivard, D., Anguenot, R., Brunelle, F., Le, V.Q., Vezina, L.P., Trepanier, S., and Michaud, D.** (2006). An in-built proteinase inhibitor system for the protection of recombinant proteins recovered from transgenic plants. *Plant Biotechnol J* **4**, 359-368.
- Roberts, T.H., Marttila, S., Rasmussen, S.K., and Hejgaard, J.** (2003). Differential gene expression for suicide-substrate serine proteinase inhibitors (serpins) in vegetative and grain tissues of barley. *J Exp Bot* **54**, 2251-2263.
- Robinson, D.G., Oliviusson, P., and Hinz, G.** (2005). Protein sorting to the storage vacuoles of plants: A critical appraisal. *Traffic* **6**, 615-625.
- Rondanino, C., Bousser, M.T., Monsigny, M., and Roche, A.C.** (2003). Sugar-dependent nuclear import of glycosylated proteins in living cells. *Glycobiology* **13**, 509-519.
- Rozkov, A., and Enfors, S.O.** (2004). Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. *Adv Biochem Eng Biotechnol* **89**, 163-195.
- Rozkov, A., Schweder, T., Veide, A., and Enfors, S.** (2000). Dynamics of proteolysis and its influence on the accumulation of intracellular recombinant proteins. *Enzyme Microb Technol* **27**, 743-748.
- Rubin, H.** (1992). The biology and biochemistry of antichymotrypsin and its potential role as a therapeutic agent. *Biol Chem Hoppe Seyler* **373**, 497-502.
- Rubin, H., Plotnick, M., Wang, Z.M., Liu, X., Zhong, Q., Schechter, N.M., and Cooperman, B.S.** (1994). Conversion of alpha 1-antichymotrypsin into a human neutrophil elastase inhibitor: Demonstration of variants with different association rate constants, stoichiometries of inhibition, and complex stabilities. *Biochemistry* **33**, 7627-7633.
- Rubin, H., Wang, Z.M., Nickbarg, E.B., McLarney, S., Naidoo, N., Schoenberger, O.L., Johnson, J.L., and Cooperman, B.S.** (1990). Cloning, expression, purification, and biological activity of recombinant native and variant human  $\alpha_1$ -antichymotrypsins. *J Biol Chem* **265**, 1199-1207.
- Rudd, P.M., Joao, H.C., Coghill, E., Fiten, P., Saunders, M.R., Opdenakker, G., and Dwek, R.A.** (1994). Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **33**, 17-22.
- Ruggiero, F., Exposito, J.Y., Bournat, P., Gruber, V., Perret, S., Comte, J., Olganier, B., Garrone, R., and Theisen, M.** (2000). Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. *FEBS Lett* **469**, 132-136.
- Runkel, L., Meier, W., Pepinsky, R.B., Karpusas, M., Whitty, A., Kimball, K., Brickelmaier, M., Muldowney, C., Jones, W., and Goelz, S.E.** (1998). Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta). *Pharm Res* **15**, 641-649.
- Saint-Jore-Dupas, C., Faye, L., and Gomord, V.** (2007). From planta to pharma with glycosylation in the toolbox. *Trends Biotechnol* **25**, 317-323.
- Saint-Jore, C.M., Evins, J., Batoko, H., Brandizzi, F., Moore, I., and Hawes, C.** (2002). Redistribution of membrane proteins between the Golgi apparatus and endoplasmic reticulum in plants is reversible and not dependent on cytoskeletal networks. *Plant J* **29**, 661-678.

- Samyn-Petit, B., Gruber, V., Flahaut, C., Wajda-Dubos, J.P., Farrer, S., Pons, A., Desmaizieres, G., Slomianny, M.C., Theisen, M., and Delannoy, P.** (2001). N-glycosylation potential of maize: The human lactoferrin used as a model. *Glycoconjugate J* **18**, 519-527.
- Saravanan, R.S., and Rose, J.K.** (2004). A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics* **4**, 2522-2532.
- Schaffer, C., Graninger, M., and Messner, P.** (2001). Prokaryotic glycosylation. *Proteomics* **1**, 248-261.
- Schähs, M., Strasser, R., Stadlmann, J., Kunert, R., Rademacher, T., and Steinkellner, H.** (2007). Production of a monoclonal antibody in plants with a humanized N-glycosylation pattern. *Plant Biotechnol J* **5**.
- Schaller, A.** (2004). A cut above the rest: The regulatory function of plant proteases. *Planta* **220**, 183-197.
- Schechter, N.M., and Plotnick, M.I.** (2004). Measurement of the kinetic parameters mediating protease-serpin inhibition. *Methods* **32**, 159-168.
- Scheller, J., Leps, M., and Conrad, U.** (2006). Forcing single-chain variable fragment production in tobacco seeds by fusion to elastin-like polypeptides. *Plant Biotechnol J* **4**, 243-249.
- Scheller, J., Henggeler, D., Viviani, A., and Conrad, U.** (2004). Purification of spider silk-elastin from transgenic plants and application for human chondrocyte proliferation. *Transgenic Res* **13**, 51-57.
- Schick, C., Pemberton, P.A., Shi, G.-P., Kamachi, Y., Çataltepe, S., Bartuski, A.J., Gornstein, E.R., Brömme, D., Chapman, H.A., and Silverman, G.A.** (1998). Cross-class inhibition of the cysteine proteinase cathepsins K, L and S by the serpin squamous cell carcinoma antigen 1: A kinetic analysis. *Biochemistry* **37**, 5258-5266.
- Schiermeyer, A., Schinkel, H., Apel, S., Fischer, R., and Schillberg, S.** (2005). Production of *Desmodus rotundus* salivary plasminogen activator alpha1 (DSPAlpha1) in tobacco is hampered by proteolysis. *Biotechnol Bioeng* **89**, 848-858.
- Schillberg, S., Twyman, R.M., and Fischer, R.** (2005). Opportunities for recombinant antigen and antibody expression in transgenic plants--Technology assessment. *Vaccine* **23**, 1764-1769.
- Schillberg, S., Zimmermann, S., Voss, A., and Fischer, R.** (1999). Apoplastic and cytosolic expression of full-size antibodies and antibody fragments in *Nicotiana tabacum*. *Transgenic Research* **8**, 255-263.
- Schmidt-Zachmann, M.S., Dargemont, C., Kuhn, L.C., and Nigg, E.A.** (1993). Nuclear export of proteins: The role of nuclear retention. *Cell* **74**, 493-504.
- Schmitz, A., and Herzog, V.** (2004). Endoplasmic reticulum-associated degradation: exceptions to the rule. *Eur J Cell Biol* **83**, 501-509.
- Schouten, A., Roosien, J., van Engelen, F.A., de Jong, G.A., Borst-Vrensse, A.W., Zilverentant, J.F., Bosch, D., Stiekema, W.J., Gommers, F.J., Schots, A., and Bakker, J.** (1996). The C-terminal KDEL sequence increases the expression level of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco. *Plant Mol Biol* **30**, 781-793.
- Schwarzkopf, M., Knobloch, K.P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R.** (2002). Sialylation is essential for early development in mice. *Proc Natl Acad Sci U S A* **99**, 5267-5270.
- Scott, F.L., Coughlin, P.B., Bird, C., Cerruti, L., Hayman, J.A., and Bird, P.** (1996). Proteinase inhibitor 6 cannot be secreted, which suggests it is a new type of cellular serpin. *J Biol Chem* **271**, 1605-1612.



- Sethuraman, N., and Stadheim, T.A.** (2006). Challenges in therapeutic glycoprotein production. *Curr Opin Biotechnol* **17**, 341-346.
- Seveno, M., Bardor, M., Paccalet, T., Gomord, V., Lerouge, P., and Faye, L.** (2004). Glycoprotein sialylation in plants? *Nat Biotechnol* **22**, 1351-1352.
- Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., Dym, O., Boldin-Adamsky, S.A., Silman, I., Sussman, J.L., Futerman, A.H., and Aviezer, D.** (2007). Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol J* **5**, 579-590.
- Sharp, J.M., and Doran, P.M.** (2001). Characterization of monoclonal antibody fragments produced by plant cells. *Biotechnol Bioeng* **73**, 338-346.
- Shewry, P.R., and Halford, N.G.** (2002). Cereal seed storage proteins: Structures, properties and role in grain utilization. *J Exp Bot* **53**, 947-958.
- Shields, R.L., Lai, J., Keck, R., O'Connell, L.Y., Hong, K., Meng, Y.G., Weikert, S.H., and Presta, L.G.** (2002). Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fc gamma RIII and antibody-dependent cellular toxicity. *J Biol Chem* **277**, 26733-26740.
- Shields, R.L., Namenuk, A.K., Hong, K., Meng, Y.G., Rae, J., Briggs, J., Xie, D., Lai, J., Stadlen, A., Li, B., Fox, J.A., and Presta, L.G.** (2001). High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. *J Biol Chem* **276**, 6591-6604.
- Shorrosh, B.S., and Dixon, R.A.** (1991). Molecular cloning of a putative plant endomembrane protein resembling vertebrate protein disulfide-isomerase and a phosphatidylinositol-specific phospholipase C. *Proc Natl Acad Sci U S A* **88**, 10941-10945.
- Showalter, A.M.** (2001). Arabinogalactan-proteins: Structure, expression and function. *Cell Mol Life Sci* **58**, 1399-1417.
- Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G.W., Irving, J.A., Lomas, D.A., Luke, C.J., Moyer, R.W., Pemberton, P.A., Remold-O'Donnell, E., Salvesen, G.S., Travis, J., and Whisstock, J.C.** (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins - Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem* **276**, 33293-33296.
- Smalle, J., and Vierstra, R.D.** (2004). The ubiquitin 26S proteasome proteolytic pathway. *Annu Rev Plant Biol* **55**, 555-590.
- Smith, D.B., and Johnson, K.S.** (1988). Single-step purification of polypeptides expressed in *Escherichia Coli* as fusions with glutathione S-transferase. *Gene* **67**, 31-40.
- Soderquist, R.G., and Lee, J.M.** (2005). Enhanced production of recombinant proteins from plant cells by the application of osmotic stress and protein stabilization. *Plant Cell Rep* **24**, 127-132.
- Sojikul, P., Buehner, N., and Mason, H.S.** (2003). A plant signal peptide-hepatitis B surface antigen fusion protein with enhanced stability and immunogenicity expressed in plant cells. *Proc Natl Acad Sci U S A* **100**, 2209-2214.
- Sourrouille, C.** (2005). Inactivation de l'alpha (1,3)-fucosyltransférase et de la beta (1,2)-xylosyltransférase, en vue de la production de protéines recombinantes d'intérêt thérapeutique chez la luzerne. Département de Biologie, Université de Rouen. Thèse de doctorat.

- Sreekrishna, K., Brankamp, R.G., Kropp, K.E., Blankenship, D.T., Tsay, J.T., Smith, P.L., Wierschke, J.D., Subramaniam, A., and Birkenberger, L.A.** (1997). Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* **190**, 55-62.
- Sriraman, R., Bardor, M., Sack, M., Vaquero, C., Faye, L., Fischer, R., Finnern, R., and Lerouge, P.** (2004). Recombinant anti-hCG antibodies retained in the endoplasmic reticulum of transformed plants lack core-xylose and core-alpha(1,3)-fucose residues. *Plant Biotechnol J* **2**, 279-287.
- Stahl, S., Nygren, P.-A., and Uhlén, M.** (1997). Detection and isolation of recombinant proteins based on binding affinity of reporter: Protein A *Methods Mol Biol* **63**, 103-118.
- Stein, P.E., and Carrell, R.W.** (1995). What do dysfunctional serpins tell us about molecular mobility and disease? *Nat Struct Biol* **2**, 96-113.
- Stevens, L.H., Stopen, G.M., Elbers, I.J., Molthoff, J.W., Bakker, H.A., Lommen, A., Bosch, D., and Jordi, W.** (2000). Effect of climate conditions and plant developmental stage on the stability of antibodies expressed in transgenic tobacco. *Plant Physiol* **124**, 173-182.
- Stoger, E., Ma, J.K., Fischer, R., and Christou, P.** (2005). Sowing the seeds of success: Pharmaceutical proteins from plants. *Curr Opin Biotechnol* **16**, 167-173.
- Stoger, E., Sack, M., Perrin, Y., Vaquero, C., Torres, E., Twyman, R.M., Christou, P., and Fischer, R.** (2002). Practical considerations for pharmaceutical antibody production in different crop systems. *Mol Breeding* **9**, 149-158.
- Stoger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P., and Fischer, R.** (2000). Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Mol Biol* **42**, 583-590.
- Stoller, J.K., Rouhani, F., Brantly, M., Shahin, S., Dweik, R.A., Stocks, J.M., Clausen, J., Campbell, E., and Norton, F.** (2002). Biochemical efficacy and safety of a new pooled human plasma alpha(1)-antitrypsin, Respitin. *Chest* **122**, 66-74.
- Strasser, R., Altmann, F., Mach, L., Glossl, J., and Steinkellner, H.** (2004). Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett* **561**, 132-136.
- Stratikos, E., and Gettins, P.G.** (1999). Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into beta-sheet A. *Proc Natl Acad Sci U S A* **96**, 4808-4813.
- Streatfield, S.J.** (2007). Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnol J* **5**, 2-15.
- Streatfield, S.J., Lane, J.R., Brooks, C.A., Barker, D.K., Poage, M.L., Mayor, J.M., Lamphear, B.J., Drees, C.F., Jilka, J.M., Hood, E.E., and Howard, J.A.** (2003). Corn as a production system for human and animal vaccines. *Vaccine* **21**, 812-815.
- Takada, S., Tsuda, M., Yamamura, M., and Katsunuma, T.** (1988a). Effect of alpha-1-antichymotrypsin on activity of DNA primase isolated from human stomach adenocarcinoma cells. *Biochem Int* **16**, 949-954.
- Takada, S., Tsuda, M., Fujinami, S., Yamamura, M., Mitomi, T., and Katsunuma, T.** (1986). Incorporation of alpha-1-antichymotrypsin into carcinoma cell nuclei of human stomach adenocarcinoma transplanted into nude mice. *Cancer Res* **46**, 3688-3691.
- Takada, S., Tsuda, M., Matsumoto, M., Fujinami, S., Yamamura, M., and Katsunuma, T.** (1988b). Incorporation of alpha-1-antichymotrypsin into human stomach

- adenocarcinoma cell nuclei and inhibition of DNA primase activity. *Tokai J Exp Clin Med* **13**, 321-327.
- Takada, S., Tsuda, M., Mitomi, T., Nakasaki, H., Noto, T., Ogoshi, K., and Katsunuma, T.** (1982). Analysis of the tissue and cellular localization of alpha-1-antichymotrypsin by an immunohistochemical technique. *Gann* **73**, 742-747.
- Takagi, H., Hiroi, T., Yang, L., Tada, Y., Yuki, Y., Takamura, K., Ishimitsu, R., Kawauchi, H., Kiyono, H., and Takaiwa, F.** (2005). A rice-based edible vaccine expressing multiple T cell epitopes induces oral tolerance for inhibition of Th2-mediated IgE responses. *Proc Natl Acad Sci U S A* **102**, 17525-17530.
- Takahashi, T., and Nishikawa, Y.** (2002). Human inherited diseases that have defects in the assembly step of lipid-linked oligosaccharide, *GlycoWord*, ed. <http://www.glycoforum.gr.jp/science/word/glycoprotein/GPC01E.html>
- Takaiwa, F., Takagi, H., Hirose, S., and Wakasa, Y.** (2007). Endosperm tissue is good production platform for artificial recombinant proteins in transgenic rice. *Plant Biotechnol J* **5**, 84-92.
- Terashima, M., Ejiri, Y., Hashikawa, N., and Yoshida, H.** (1999a). Effect of osmotic pressure on human alpha(1)-antitrypsin production by plant cell culture. *Biochemical Engineering Journal* **4**, 31-36.
- Terashima, M., Murai, Y., Kawamura, M., Nakanishi, S., Stoltz, T., Chen, L., Drohan, W., Rodriguez, R.L., and Katoh, S.** (1999b). Production of functional human alpha 1-antitrypsin by plant cell culture. *Appl Microbiol Biotechnol* **52**, 516-523.
- Torrent, M., Bastida, M., Marzabal, P., Llompарт, B., and Ludevid, D.** (2006a). Production of proteins US 2006/0121573.
- Torrent, M., Bastida, M., Marzabal, P., Llompарт, B., and Ludevid, D.** (2006b). Protein isolation and purification US 2006/0123509.
- Travis, J., Owen, M., George, P., Carrell, R., Rosenberg, S., Hallewell, R.A., and Barr, P.J.** (1985). Isolation and properties of recombinant DNA produced variants of human alpha 1-proteinase inhibitor. *J. Biol. Chem.* **260**, 4384-4389.
- Tregoning, J.S., Nixon, P., Kuroda, H., Svab, Z., Clare, S., Bowe, F., Fairweather, N., Ytterberg, J., van Wijk, K.J., Dougan, G., and Maliga, P.** (2003). Expression of tetanus toxin Fragment C in tobacco chloroplasts. *Nucleic Acids Res* **31**, 1174-1179.
- Trombetta, E.S.** (2003). The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology* **13**, 77R-91R.
- Tsoi, B.M., and Doran, P.M.** (2002). Effect of medium properties and additives on antibody stability and accumulation in suspended plant cell cultures. *Biotechnol Appl Biochem* **35**, 171-180.
- Tsuda, M., Masuyama, M., and Katsunuma, T.** (1986). Inhibition of human DNA polymerase alpha by alpha 1-antichymotrypsin. *Cancer Res* **46**, 6139-6142.
- Twyman, R.M., Schillberg, S., and Fischer, R.** (2005). Transgenic plants in the biopharmaceutical market. *Expert Opin Emerg Drugs* **10**, 185-218.
- Twyman, R.M., Stoger, E., Schillberg, S., Christou, P., and Fischer, R.** (2003). Molecular farming in plants: Host systems and expression technology. *Trends Biotechnol* **21**, 570-578.
- Upreti, R.K., Kumar, M., and Shankar, V.** (2003). Bacterial glycoproteins: Functions, biosynthesis and applications. *Proteomics* **3**, 363-379.
- Urwin, P.E., Mcpherson, M.J., and Atkinson, H.J.** (1998). Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs. *Planta* **204**, 472-479.



- van der Vyver, C., Schneidereit, J., Driscoll, S., Turner, J., Kunert, K., and Foyer, C.H.** (2003). Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances chilling tolerance. *Plant Biotechnol J* **1**, 101-112.
- Van Droogenbroeck, B., Cao, J., Stadlmann, J., Altmann, F., Colanesi, S., Hillmer, S., Robinson, D.G., Van Lerberge, E., Terry, N., Van Montagu, M., Liang, M., Depicker, A., and De Jaeger, G.** (2007). Aberrant localization and underglycosylation of highly accumulating single-chain Fv-Fc antibodies in transgenic *Arabidopsis* seeds. *Proc Natl Acad Sci U S A* **104**, 1430-1435.
- Van Larebeke, N., Genetello, C., Schell, J., Schilperoort, R.A., Hermans, A.K., Van Montagu, M., and Hernalsteens, J.P.** (1975). Acquisition of tumour-inducing ability by non-oncogenic agrobacteria as a result of plasmid transfer. *Nature* **255**, 742-743.
- van Ree, R., and Aalberse, R.C.** (1999). Specific IgE without clinical allergy. *J Allergy Clin Immunol* **103**, 1000-1001.
- van Ree, R., Cabanes-Macheteau, M., Akkerdaas, J., Milazzo, J.P., Loutelier-Bourhis, C., Rayon, C., Villalba, M., Koppelman, S., Aalberse, R., Rodriguez, R., Faye, L., and Lerouge, P.** (2000). Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J Biol Chem* **275**, 11451-11458.
- Vaquero, C., Sack, M., Schuster, F., Finnern, R., Drossard, J., Schumann, D., Reimann, A., and Fischer, R.** (2002). A carcinoembryonic antigen-specific diabody produced in tobacco. *Faseb J* **16**, 408-410.
- Varki, A.** (2007). Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **446**, 1023-1029.
- Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J.** (1999). *Essentials of glycobiology*, Cold Spring Harbor Laboratory Press, New York.
- Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J.A., De Rycke, R., Brackenier, A., Inze, D., Harris, J.L., and Van Breusegem, F.** (2006). Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J Mol Biol* **364**, 625-636.
- Vierstra, R.D.** (1996). Proteolysis in plants: Mechanisms and function. *Plant Mol Biol* **32**, 275-302.
- Vierstra, R.D.** (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci* **8**, 135-142.
- Vitale, A., and Hinz, G.** (2005). Sorting of proteins to storage vacuoles: How many mechanisms? *Trends Plant Sci* **10**, 316-323.
- Vitale, A., and Pedrazzini, E.** (2005). Recombinant pharmaceuticals from plants: The plant endomembrane system as bioreactor. *Mol Interv* **5**, 216-225.
- von Schaewen, A., Sturm, A., O'Neill, J., and Chrispeels, M.J.** (1993). Isolation of a mutant *Arabidopsis* plant that lacks N-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans. *Plant Physiol* **102**, 1109-1118.
- Walsh, G., and Jefferis, R.** (2006). Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* **24**, 1241-1252.
- Wandelt, C.I., Khan, M.R., Craig, S., Schroeder, H.E., Spencer, D., and Higgins, T.J.** (1992). Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants. *Plant J* **2**, 181-192.
- Watson, J.M., Fusaro, A.F., Wang, M., and Waterhouse, P.M.** (2005). RNA silencing platforms in plants. *FEBS Lett* **579**, 5982-5987.

- Wee, E.G., Sherrier, D.J., Prime, T.A., and Dupree, P.** (1998). Targeting of active sialyltransferase to the plant Golgi apparatus. *Plant Cell* **10**, 1759-1768.
- Wei, A., Rubin, H., Cooperman, B.S., and Christianson, D.W.** (1994). Crystal structure of an uncleaved serpin reveals the conformation of an inhibitory reactive loop. *Nat Struct Biol* **1**, 251-258.
- Weis, W.I., and Drickamer, K.** (1996). Structural basis of lectin-carbohydrate recognition. *Annu Rev Biochem* **65**, 441-473.
- Whisstock, J.C., and Bottomley, S.P.** (2006). Molecular gymnastics: Serpin structure, folding and misfolding. *Curr Opin Struct Biol* **16**, 761-768.
- Wirth, S., Calamante, G., Mentaberry, A., Bussmann, L., Lattanzi, M., Baranao, L., and Bravo-Almonacid, F.** (2004). Expression of active human epidermal growth factor (hEGF) in tobacco plants by integrative and non-integrative systems. *Mol Breeding* **13**, 23-35.
- Wormald, M.R., and Dwek, R.A.** (1999). Glycoproteins: Glycan presentation and protein-fold stability. *Structure* **7**, R155-160.
- Wurm, F.M.** (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **22**, 1393-1398.
- Yang, J., Barr, L.A., Fahnestock, S.R., and Liu, Z.B.** (2005). High yield recombinant silk-like protein production in transgenic plants through protein targeting. *Transgenic Res* **14**, 313-324.
- Yin, J., Li, G., Ren, X., and Herrler, G.** (2007). Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Biotechnol* **127**, 335-347.
- Yoo, B.-C., Aoki, K., Xiang, Y., Campbell, L.R., Hull, R.J., Xoconostle-Cazares, B., Monzer, J., Lee, J.-Y., Ullman, D.E., and Lucas, W.J.** (2000). Characterization of *Cucurbita maxima* phloem serpin-1 (CmPS-1). *J Biol Chem* **275**, 35122-35128.
- Yoon, D., Kueppers, F., Genta, R.M., Klintmalm, G.B., Khaoustov, V.I., and Yoffe, B.** (2002). Role of alpha-1-antichymotrypsin deficiency in promoting cirrhosis in two siblings with heterozygous alpha-1-antitrypsin deficiency phenotype SZ. *Gut* **50**, 730-732.
- Yoshida, K., Matsui, T., and Shinmyo, A.** (2004). The plant vesicular transport engineering for production of useful recombinant proteins. *J Mol Catal B-Enzym* **28**, 167-171.
- Young, N.M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C.M.** (2002). Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J Biol Chem* **277**, 42530-42539.
- Yuen, C.T., Storing, P.L., Tiplady, R.J., Izquierdo, M., Wait, R., Gee, C.K., Gerson, P., Lloyd, P., and Cremata, J.A.** (2003). Relationships between the N-glycan structures and biological activities of recombinant human erythropoietins produced using different culture conditions and purification procedures. *Br J Haematol* **121**, 511-526.
- Zeleny, R., Kolarich, D., Strasser, R., and Altmann, F.** (2006). Sialic acid concentrations in plants are in the range of inadvertent contamination. *Planta* **224**, 222-227.
- Zhang, H., Li, X.J., Martin, D.B., and Aebersold, R.** (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**, 660-666.
- Zhou, A., Wei, Z., Read, R.J., and Carrell, R.W.** (2006a). Structural mechanism for the carriage and release of thyroxine in the blood. *Proc Natl Acad Sci U S A* **103**, 13321-13326.

**Zhou, Y., Niu, H., Brandizzi, F., Fowke, L.C., and Wang, H. (2006b).** Molecular control of nuclear and subnuclear targeting of the plant CDK inhibitor ICK1 and ICK1-mediated nuclear transport of CDKA. *Plant Mol Biol* **62**, 261-278.