Diversité structurale et d'activité biologique des Albumines entomotoxiques de type 1b des graines de Légumineuses

Sandrine Louis

▶ **To cite this version:**

Sandrine Louis. Diversité structurale et d'activité biologique des Albumines entomotoxiques de type 1b des graines de Légumineuses. Biochimie [q-bio.BM]. INSA de Lyon, 2004. Français. NNT : . tel-00007903

HAL Id: tel-00007903
https://theses.hal.science/tel-00007903

Submitted on 4 Jan 2005

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

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

N° d'ordre 04 ISAL 012
Année 2004

Thèse

Diversité structurale et d'activité biologique des Albumines entomotoxiques de type 1b des graines de Légumineuses.

Présentée devant
L'Institut National des Sciences Appliquées de Lyon

Pour obtenir
Le grade de docteur

Formation doctorale
Analyse et Modélisation des Systèmes Biologiques

École doctorale
Evolution, Ecologie Microbienne et Modélisation

Par
Sandrine Louis

Soutenue le 20 février 2004 devant la Commission d'examen

Jury MM.

J.P. Blein	Directeur de recherche INRA, Rapporteur
C. Domoney	Project Leader (JIC, UK), Rapporteur
J. Guéguen	Directeur de recherche INRA
E. Guiderdoni	Directeur de recherche CIRAD, Rapporteur
J.M. Lancelin	Professeur UCB Lyon 1
Y. Rahbé	Directeur de recherche INRA (INSA de Lyon)

Novembre 2003

INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

***Directeur* : STORCK A.**

***Professeurs* :**

AMGHAR Y.	LIRIS
AUDISIO S.	PHYSICOCHIMIE INDUSTRIELLE
BABOT D.	CONT. NON DESTR. PAR RAYONNEMENTS IONISANTS
BABOUX J.C.	GEMPPM***
BALLAND B.	PHYSIQUE DE LA MATIERE
BAPTISTE P.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
BARBIER D.	PHYSIQUE DE LA MATIERE
BASKURT A.	LIRIS
BASTIDE J.P.	LAEPSI****
BAYADA G.	MECANIQUE DES CONTACTS
BENADDA B.	LAEPSI****
BETEMPS M.	AUTOMATIQUE INDUSTRIELLE
BIENNIER F.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
BLANCHARD J.M.	LAEPSI****
BOISSE P.	LAMCOS
BOISSON C.	VIBRATIONS-ACOUSTIQUE
BOIVIN M. (Prof. émérite)	MECANIQUE DES SOLIDES
BOTTA H.	UNITE DE RECHERCHE EN GENIE CIVIL - Développement Urbain
BOTTA-ZIMMERMANN M. (Mme)	UNITE DE RECHERCHE EN GENIE CIVIL - Développement Urbain
BOULAYE G. (Prof. émérite)	INFORMATIQUE
BOYER J.C.	MECANIQUE DES SOLIDES
BRAU J.	CENTRE DE THERMIQUE DE LYON - Thermique du bâtiment
BREMOND G.	PHYSIQUE DE LA MATIERE
BRISSAUD M.	GENIE ELECTRIQUE ET FERROELECTRICITE
BRUNET M.	MECANIQUE DES SOLIDES
BRUNIE L.	INGENIERIE DES SYSTEMES D'INFORMATION
BUFFIERE J-Y.	GEMPPM***
BUREAU J.C.	CEGELY*
CAMPAGNE J-P.	PRISMA
CAVAILLE J.Y.	GEMPPM***
CHAMPAGNE J-Y.	LMFA
CHANTE J.P.	CEGELY*- Composants de puissance et applications
CHOCAT B.	UNITE DE RECHERCHE EN GENIE CIVIL - Hydrologie urbaine
COMBESCURE A.	MECANIQUE DES CONTACTS
COURBON	GEMPPM
COUSIN M.	UNITE DE RECHERCHE EN GENIE CIVIL - Structures
DAUMAS F. (Mme)	CENTRE DE THERMIQUE DE LYON - Energétique et Thermique
DJERAN-MAIGRE I.	UNITE DE RECHERCHE EN GENIE CIVIL
DOUTHEAU A.	CHIMIE ORGANIQUE
DUBUY-MASSARD N.	ESCHIL
DUFOUR R.	MECANIQUE DES STRUCTURES
DUPUY J.C.	PHYSIQUE DE LA MATIERE
EMPTOZ H.	RECONNAISSANCE DE FORMES ET VISION
ESNOUF C.	GEMPPM***
EYRAUD L. (Prof. émérite)	GENIE ELECTRIQUE ET FERROELECTRICITE
FANTOZZI G.	GEMPPM***
FAVREL J.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
FAYARD J.M.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
FAYET M.	MECANIQUE DES SOLIDES
FAZEKAS A.	GEMPPM
FERRARIS-BESSO G.	MECANIQUE DES STRUCTURES
FLAMAND L.	MECANIQUE DES CONTACTS
FLEURY E.	CITI
FLORY A.	INGENIERIE DES SYSTEMES D'INFORMATIONS
FOUGERES R.	GEMPPM***
FOUQUET F.	GEMPPM***
FRECON L.	REGROUPEMENT DES ENSEIGNANTS CHERCHEURS ISOLES
GERARD J.F.	INGENIERIE DES MATERIAUX POLYMERES
GERMAIN P.	LAEPSI****
GIMENEZ G.	CREATIS**
GOBIN P.F. (Prof. émérite)	GEMPPM***
GONNARD P.	GENIE ELECTRIQUE ET FERROELECTRICITE
GONTRAND M.	PHYSIQUE DE LA MATIERE
GOUTTE R. (Prof. émérite)	CREATIS**
GOUJON L.	GEMPPM***
GOURDON R.	LAEPSI****.
GRANGE G.	GENIE ELECTRIQUE ET FERROELECTRICITE
GUENIN G.	GEMPPM***
GUICHARDANT M.	BIOCHIMIE ET PHARMACOLOGIE
GUILLOT G.	PHYSIQUE DE LA MATIERE
GUINET A.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS

GUYADER J.L.	VIBRATIONS-ACOUSTIQUE
GUYOMAR D.	GENIE ELECTRIQUE ET FERROELECTRICITE
HEIBIG A.	MATHEMATIQUE APPLIQUEES DE LYON
JACQUET-RICHARDET G.	MECANIQUE DES STRUCTURES
JAYET Y.	GEMPPM***
JOLION J.M.	RECONNAISSANCE DE FORMES ET VISION
JULLIEN J.F.	UNITE DE RECHERCHE EN GENIE CIVIL - Structures
JUTARD A. (Prof. émérite)	AUTOMATIQUE INDUSTRIELLE
KASTNER R.	UNITE DE RECHERCHE EN GENIE CIVIL - Géotechnique
KOULOUMDJIAN J.	INGENIERIE DES SYSTEMES D'INFORMATION
LAGARDE M.	BIOCHIMIE ET PHARMACOLOGIE
LALANNE M. (Prof. émérite)	MECANIQUE DES STRUCTURES
LALLEMAND A.	CENTRE DE THERMIQUE DE LYON - Energétique et thermique
LALLEMAND M. (Mme)	CENTRE DE THERMIQUE DE LYON - Energétique et thermique
LAUGIER A.	PHYSIQUE DE LA MATIERE
LAUGIER C.	BIOCHIMIE ET PHARMACOLOGIE
LAURINI R.	INFORMATIQUE EN IMAGE ET SYSTEMES D'INFORMATION
LEJEUNE P.	UNITE MICROBIOLOGIE ET GENETIQUE
LUBRECHT A.	MECANIQUE DES CONTACTS
MASSARD N.	INTERACTION COLLABORATIVE TELEFORMATION TELEACTIVITE
MAZILLE H.	PHYSICOCHIMIE INDUSTRIELLE
MERLE P.	GEMPPM***
MERLIN J.	GEMPPM***
MIGNOTTE A. (Mle)	INGENIERIE, INFORMATIQUE INDUSTRIELLE
MILLET J.P.	PHYSICOCHIMIE INDUSTRIELLE
MIRAMOND M.	UNITE DE RECHERCHE EN GENIE CIVIL - Hydrologie urbaine
MOREL R.	MECANIQUE DES FLUIDES ET D'ACOUSTIQUES
MOSZKOWICZ P.	LAEPSI****
NARDON P. (Prof. émérite)	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
NELIAS D.	LAMCOS
NIEL E.	AUTOMATIQUE INDUSTRIELLE
NORMAND B.	GEMPPM
NORTIER P.	DREP
ODET C.	CREATIS**
OTTERBEIN M. (Prof. émérite)	LAEPSI****
PARIZET E.	VIBRATIONS-ACOUSTIQUE
PASCAULT J.P.	INGENIERIE DES MATERIAUX POLYMERES
PAVIC G.	VIBRATIONS-ACOUSTIQUE
PECORARO S.	GEMPPM
PELLETIER J.M.	GEMPPM***
PERA J.	UNITE DE RECHERCHE EN GENIE CIVIL - Matériaux
PERRIAT P.	GEMPPM***
PERRIN J.	INTERACTION COLLABORATIVE TELEFORMATION TELEACTIVITE
PINARD P. (Prof. émérite)	PHYSIQUE DE LA MATIERE
PINON J.M.	INGENIERIE DES SYSTEMES D'INFORMATION
PONCET A.	PHYSIQUE DE LA MATIERE
POUSIN J.	MODELISATION MATHEMATIQUE ET CALCUL SCIENTIFIQUE
PREVOT P.	INTERACTION COLLABORATIVE TELEFORMATION TELEACTIVITE
PROST R.	CREATIS**
RAYNAUD M.	CENTRE DE THERMIQUE DE LYON - Transferts Interfaces et Matériaux
REDARCE H.	AUTOMATIQUE INDUSTRIELLE
RETIF J-M.	CEGELY*
REYNOUARD J.M.	UNITE DE RECHERCHE EN GENIE CIVIL - Structures
RICHARD C.	LGEF
RIGAL J.F.	MECANIQUE DES SOLIDES
RIEUTORD E. (Prof. émérite)	MECANIQUE DES FLUIDES
ROBERT-BAUDOUY J. (Mme) (Prof. émérite)	GENETIQUE MOLECULAIRE DES MICROORGANISMES
ROUBY D.	GEMPPM***
ROUX J.J.	CENTRE DE THERMIQUE DE LYON – Thermique de l'Habitat
RUBEL P.	INGENIERIE DES SYSTEMES D'INFORMATION
SACADURA J.F.	CENTRE DE THERMIQUE DE LYON - Transferts Interfaces et Matériaux
SAUTEREAU H.	INGENIERIE DES MATERIAUX POLYMERES
SCAVARDA S.	AUTOMATIQUE INDUSTRIELLE
SOUIFI A.	PHYSIQUE DE LA MATIERE
SOUROUILLE J.L.	INGENIERIE INFORMATIQUE INDUSTRIELLE
THOMASSET D.	AUTOMATIQUE INDUSTRIELLE
THUDEROZ C.	ESCHIL – Equipe Sciences Humaines de l'Insa de Lyon
UBEDA S.	CENTRE D'INNOV. EN TELECOM ET INTEGRATION DE SERVICES
VELEX P.	MECANIQUE DES CONTACTS
VIGIER G.	GEMPPM***
VINCENT A.	GEMPPM***
VRAY D.	CREATIS**
VUILLERMOZ P.L. (Prof. émérite)	PHYSIQUE DE LA MATIERE

Directeurs de recherche C.N.R.S. **:**

BERTHIER Y.	MECANIQUE DES CONTACTS
CONDEMINE G.	UNITE MICROBIOLOGIE ET GENETIQUE
COTTE-PATAT N. (Mme)	UNITE MICROBIOLOGIE ET GENETIQUE

ESCUDIE D. (Mme)	CENTRE DE THERMIQUE DE LYON
FRANCIOSI P.	GEMPPM***
MANDRAND M.A. (Mme)	UNITE MICROBIOLOGIE ET GENETIQUE
POUSIN G.	BIOLOGIE ET PHARMACOLOGIE
ROCHE A.	INGENIERIE DES MATERIAUX POLYMERES
SEGUELA A.	GEMPPM***
VERGNE P.	LaMcos

Directeurs de recherche I.N.R.A. **:**

FEBVAY G.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
GRENIER S.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
RAHBE Y.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS

Directeurs de recherche I.N.S.E.R.M. **:**

KOBAYASHI T.	PLM
PRIGENT A.F. (Mme)	BIOLOGIE ET PHARMACOLOGIE
MAGNIN I. (Mme)	CREATIS**

* ***CEGELY***	*CENTRE DE GENIE ELECTRIQUE DE LYON*
** ***CREATIS***	*CENTRE DE RECHERCHE ET D'APPLICATIONS EN TRAITEMENT DE L'IMAGE ET DU SIGNAL*
GEMPPM	*GROUPE D'ETUDE METALLURGIE PHYSIQUE ET PHYSIQUE DES MATERIAUX*
****LAEPSI***	*LABORATOIRE D'ANALYSE ENVIRONNEMENTALE DES PROCEDES ET SYSTEMES INDUSTRIELS*

Ecoles Doctorales et Diplômes d'Etudes Approfondies

habilités pour la période 1999-2003

ECOLES DOCTORALES n° code national	RESPONSABLE PRINCIPAL	CORRESPONDANT INSA	DEA INSA n° code national	RESPONSABLE DEA INSA
CHIMIE DE LYON **(Chimie, Procédés, Environnement)** **EDA206**	M. D. SINOU UCBL1 04.72.44.62.63 Sec 04.72.44.62.64 Fax 04.72.44.81.60	**M. R. GOURDON** 87.53 Sec 84.30 Fax 87.17	Chimie Inorganique **910643** Sciences et Stratégies Analytiques **910634** Sciences et Techniques du Déchet **910675**	 M. R. GOURDON Tél 87.53 Fax 87.17
ECONOMIE, ESPACE ET MODELISATION DES COMPORTEMENTS **(E^2MC)** **EDA417**	M.A. BONNAFOUS LYON 2 04.72.72.64.38 Sec 04.72.72.64.03 Fax 04.72.72.64.48	**Mme M. ZIMMERMANN** 60.91 Fax 87.96	Villes et Sociétés **911218** Dimensions Cognitives et Modélisation **992678**	Mme M. ZIMMERMANN Tél 60.91 Fax 87.96 M. L. FRECON Tél 82.39 Fax 85.18
ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE **(E.E.A.)** **EDA160**	**M. D. BARBIER** INSA DE LYON 85.47 Fax 60.82		Automatique Industrielle **910676** Dispositifs de l'Electronique Intégrée **910696** Génie Electrique de Lyon **910065** Images et Systèmes **992254**	M. M. BETEMPS Tél 85.59 Fax 85.35 M. D. BARBIER Tél 85.47 Fax 60.82 M. J.P. CHANTE Tél 87.26 Fax 85.30 Mme I. MAGNIN Tél 85.63 Fax 85.26
EVOLUTION, ECOSYSTEME, MICROBIOLOGIE , MODELISATION **(E2M2)** **EDA403**	**M. J.P FLANDROIS** UCBL1 04.78.86.31.50 Sec 04.78.86.31.52 Fax 04.78.86.31.49	**M. S. GRENIER** 79.88 Fax 85.34	Analyse et Modélisation des Systèmes Biologiques **910509**	M. S. GRENIER Tél 79.88 Fax 85.34
INFORMATIQUE ET INFORMATION POUR LA SOCIETE **(EDIIS)** **EDA 407**	**M. L. BRUNIE** INSA DE LYON 87.59 Fax 80.97		Documents Multimédia, Images et Systèmes d'Information Communicants **992774** Extraction des Connaissances à partir des Données **992099** Informatique et Systèmes Coopératifs pour l'Entreprise **950131**	M. A. FLORY Tél 84.66 Fax 85.97 M. J.F. BOULICAUT Tél 89.05 Fax 87.13 M. A. GUINET Tél 85.94 Fax 85.38
INTERDISCIPLINAIRE SCIENCES-SANTE **(EDISS)** **EDA205**	**M. A.J. COZZONE** UCBL1 04.72.72.26.72 Sec 04.72.72.26.75 Fax 04.72.72.26.01	**M. M. LAGARDE** 82.40 Fax 85.24	Biochimie **930032**	M. M. LAGARDE Tél 82.40 Fax 85.24
MATERIAUX DE LYON **UNIVERSITE LYON 1** **EDA 034**	**M. J. JOSEPH** ECL 04.72.18.62.44 Sec 04.72.18.62.51 Fax 04.72.18.60.90	**M. J.M. PELLETIER** 83.18 Fax 85.28	Génie des Matériaux : Microstructure, Comportement Mécanique, Durabilité **910527** Matériaux Polymères et Composites **910607** Matière Condensée, Surfaces et Interfaces **910577**	M. J.M.PELLETIER Tél 83.18 Fax 85.28 M. H. SAUTEREAU Tél 81.78 Fax 85.27 M. G. GUILLOT Tél 81.61 Fax 85.31
MATHEMATIQUES ET INFORMATIQUE FONDAMENTALE **(Math IF)** **EDA 409**	**M. F. WAGNER** UCBL1 04.72.43.27.86 Fax 04.72.43.00.35	**M. J. POUSIN** 88.36 Fax 85.29	Analyse Numérique, Equations aux dérivées partielles et Calcul Scientifique **910281**	M. G. BAYADA Tél 83.12 Fax 85.29
MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE **(MEGA)** **EDA162**	**M. F. SIDOROFF** ECL 04.72.18.61.56 Sec 04.72.18.61.60 Fax 04.78.64.71.45	**M. G.DALMAZ** 83.03 Fax 04.72.89.09.80	Acoustique **910016** Génie Civil **992610** Génie Mécanique **992111** Thermique et Energétique **910018**	M. J.L. GUYADER Tél 80.80 Fax 87.12 M. J.J.ROUX Tél 84.60 Fax 85.22 M. G. DALMAZ Tél 83.03 Fax 04.78.89.09.80 M. J. F. SACADURA Tél 81.53 Fax 88.11

Un grand merci à :

Yvan, pour m'avoir permis d'effectuer cette thèse, en décrochant un financement et acceptant la lourde tâche d'être mon directeur.

Gérard et Jacques pour m'avoir accueillie dans leur laboratoire respectif,

Tous les membres de mon jury, pour accepter de juger mon travail,

Bernard, Daniel, Didier, Françoise, Fréderic, Khalil, Laurence et Laurence pour leur aide et leurs conseils,

Isa et Gaby pour m'avoir soulagée de certaines manips et m'avoir soutenue par leur sympathie,
Bourriquet, Alex, Cyril, Fédé, Cédric, Vanessa, Hélène, Audrey, Luca, Josette, Chaqué, Marjolaine, Agnès, Caro, André, Michel, Florence et Florence, Béné, Christiane, Fabrice, Claudia, Ludivine, Raul, Géraldine, Maud, Jacky, Marie-Anne, Dominique, Catherine, Jacote, Aziz, Hubert, Patrice, Simon, et tous les autres, pour leur sympathique compagnie durant mes séjours à Lyon et Nantes.

Zouzou qui, étant dans la même galère, me comprends,

Mes proches, sans qui je ne serais peut-être pas restée, pour m'avoir supportée lors des passages difficiles,

Tous ceux qui liront cette thèse et la feront vivre.

Ce travail de thèse a donné lieu à présentation des résultats sous forme de :

- ***Publications :***

Louis S., Delobel B., Gressent F., Rahioui I., Vallier A., Quillien L., Rahbé Y. (2004). *Molecular and biological screening for insect-toxic seed albumins from four legume species*. Plant Science, in revision.

- ***Posters en congrès international :***

Louis S., Jouvensal L., Delobel B., El Morjani K., Gressent F., Vovelle F., Rahbé Y. (2003) *New insecticidal peptide family from legume seeds : structural and biological diversity*. Seventh International Symposium on Plant Molecular Biology, 23-28 Juin 2003, Barcelone (Espagne).

Louis S., Delobel B., Gressent F., Rahbé Y. (2001). *Albumins 1b : a new insecticidal peptides from legume seeds*. Eleventh international Symposium on Insect-Plant relationships, 4-10 Août 2001, Helsingor (Danemark).

- ***Posters en congrès national :***

Louis S., Rahbé Y., Delobel B., Vovelle F., Jouvensal L., Quillen L. (2002). *Characterization of an insecticidal peptides family of legumes seeds : structure-activity relationships studies*. Rencontres de Biologie-Physique du Grand Ouest, 4-5 Juillet 2002, Rennes.

Louis S., Delobel B., Gressent F., Rahbé Y. (2001). *Potentiality of the model lucern* Medicago truncatula *for a genomic approach to characterise plant defence peptides against insects and other pathogens*. Premières Journées Post-Génomique de la Doua, 5-6 Avril 2001, Villeurbanne.

Trois publications sont également prévues :

Louis S., Gressent F., Rahbé Y. (2004). *Plant defence peptides and their roles in resistance to insects*. Article de revue *in prep*.

Louis S., Delobel B., Rahbé Y. (2004). *Broad screening of legumes species for insecticidal albumins 1b*. (Phytochemistry)

Rahbé Y., Louis S., Zhang Y., Le Signor C., Duport G., Delobel B., and Prosperi J.M. (2004), *Screening* Medicago truncatula *genotypes for peptide seed toxicity towards cereal weevils*. *in prep*. (Phytochemistry).

Sommaire

Sommaire **9**
Liste des figures **13**
Liste des tableaux **19**
Abréviations **20**
Avant-propos **21**
Foreword (english version) **22**
Chapitre I : Introduction bibliographique **25**
I. Les défenses chimiques des plantes face aux insectes. 25
A. Les allélochimiques non protéiques à activité entomotoxique. 27
1) Les composés azotés. 28
2) Les terpénoïdes. 36
3) Les composés phénoliques. 44
B. Les protéines végétales de défense 51
1) Les protéines entomotoxiques 51
2) Les peptides de défense (antimicrobiens et insecticides) 67
C. Conclusion 76
II. Contexte de l'étude et présentation des travaux 78
Chapitre II : Définition de l'approche : à la découverte de la famille des A1b **83**
I. Matériels et méthodes 84
A. Biologie : recherche de toxicité analogue 84
1) Les charançons 84
2) Le matériel végétal 85
3) Préparation et fractionnement des farines 85
4) Tests biologiques 86
B. Biochimie : recherche d'homologues protéiques 86
1) Dosage protéique 86
2) Electrophorèse protéique sur gel d'acrylamide (SDS-PAGE) 86
3) Immunodétection 87
4) HPLC (High Performance Liquid Chromatography) 87
5) Spectrométrie de masse 87
6) Tests d'inhibition de liaison 88
C. Biologie moléculaire : recherche de gènes homologues 88
1) Extraction d'ADN 88
2) Amplification de fragments de gènes par PCR 89
3) Electrophorèse sur gel d'agarose 90
4) Clonage 90
5) Minipréparation d'ADN plasmidique par lyse alcaline 91
6) Séquençage 91
7) Recherche des séquences amont des gènes 91
II. Résultats 92
A. Recherche des homologues de PA1b dans les graines des plantes étudiées 92
1) Toxicité des farines entières 92
2) Fractionnement et toxicité des différentes fractions 94
3) Détection électrophorétique et immunologique d'homologues de PA1b 98

4) Résultats des tests d'inhibition de liaison. 100
B. Recherche des gènes homologues au gène codant pour PA1b dans le génome des plantes étudiées. 101
1) Extraction d'ADN des différentes plantes 101
2) Amplification de fragments de gènes homologues à PA1b 101
3) Recherche des séquences amont des gènes 103
4) Analyse des séquences 103
C. Détection de peptides par spectrométrie de masse. 107
III. Discussion 110
1) Toxicité analogue et gènes homologues, cas du pois, du soja et du haricot. 110
2) Cas de *M. truncatula*. 111
IV. Conclusion 112
Chapitre III : La famille des A1b au sein de la famille des Légumineuses 115
I. Les Légumineuses 115
A. Systématique 115
1) Présentation de la famille des Légumineuses. 115
2) Phylogénie 117
B. Choix des espèces testées 120
II. Diversité structurale primaire 121
A. Matériels et Méthodes 121
1) Matériel végétal 121
2) Extraction d'ADN 121
3) Recherche de fragments de gènes homologues par PCR 121
4) Recherche de séquences amont 122
5) Clonage et séquençage 122
6) Northern Blot 122
7) Analyses des séquences 123
B. Résultats 123
1) Extractions d'ADN 123
2) Recherche de gènes homologues 123
3) Marche sur le gène 128
4) Analyse des séquences 129
C. Discussion 134
1) Les A1b au sein des Légumineuses : une évolution récente? 134
2) Les A1b limités à deux clades de Papilionoideae ? 134
III. Variabilité d'activité entomotoxique 136
A. Matériels et Méthodes 136
1) Extractions 136
2) Matériel animal et tests biologiques 136
3) Tests d'inhibition de liaison. 137
4) Electrophorèses 137
B. Résultats 137
1) Extractions 137
2) Variabilité de l'activité biologique 138
3) Compléments biochimiques 142
C. Discussion 144
1) Activité biologique des extraits de graines 144
2) Analyses complémentaires 145

IV. Discussion et conclusions 146
1) Variabilité et histoire évolutive des A1b 146
2) Diversité d'activité des extraits, A1b et autres composés de défense. 147
3) Indications sur le mode d'action? 148
Chapitre IV : Expression hétérologue de PA1b 151
I. Systèmes d'expression hétérologue 151
A. Systèmes eucaryotes 151
1) Les Levures 151
2) cellules d'insectes 152
3) cellules de mammifères 152
4) Transgénèse 153
B. Systèmes procaryotes 154
1) Stratégie générale pour l'expression d'un gène chez *E. coli* 154
2) Expression dans les différents compartiments cellulaires 159
II. Matériel et méthodes 164
A. Construction génétique et expression 164
1) Souches et plasmides 164
2) Milieux de culture 164
3) Extraction d'ADN plasmidique 165
4) digestions et ligations d'ADN double brin 165
5) Transformation d'*E. coli* par un plasmide 165
6) Induction de la surexpression de protéines recombinantes en erlen 166
B. Purification et caractérisation des protéines 166
1) Expression des protéines recombinantes en fermenteur 166
2) Purification des protéines de fusion TRX-PA1b par chromatographie d'affinité 167
3) Purification des protéines fusion par chromatographie d'échange d'ion (IEC). 168
4) Coupure des protéines de fusion 168
5) Purification des protéines recombinantes par chromatographie d'affinité ... 169
6) Purification des protéines recombinantes par HPLC 169
7) Suivi des étapes de purification 169
8) Western-Blotting et révélation immunologique 169
9) Caractérisation des protéines par spectrométrie de masse 169
10) Caractérisation de l'activité des protéines. 170
III. Résultats 170
A. Constructions génétiques 170
1) Stratégie de construction 170
2) Conditions de la surexpression de la protéine fusion 172
B. Purification et caractérisation protéiques 175
1) Purification et caractérisation de la protéine fusion 175
2) Clivages 181
3) Purification de rPA1b après clivage acide 185
4) Purification de rPA1b après clivage enzymatique 189
5) Caractérisation de rPA1b 190
IV. Discussion 193
Conclusions générales et perspectives 197
1) Evolution de la famille des A1b au sein de la famille des Légumineuses. .. 197

2) Origine des A1b ?..... 198
3) Relations structure-fonction..... 200
Conclusions and Prospects (english version)..... 201
1) Evolution of the A1b family among the Legume family. 201
2) The origins of A1b?..... 202
3) Structure – function relationships. 203
List and legends of figures (english version)..... 205
List and legends of tables (english version) 211
Références Bibliographiques..... 215
Annexes 251
I. Annexe 1 : Principe de la marche sur le gène (kit GenomeWalker™, Clontech). 251
II. Annexe 2 : Phylogénie des Légumineuses 252
III. Annexe 3 : Liste des espèces sélectionnées, origine géographique et fournisseur254
IV. Annexe 4 : Représentation des différentes fractions dans la farine d'espèces de Légumineuses..... 256
V. Annexe 5 : Temps létaux 50 257
VI. Annexe 6 : Plasmide et séquences 260
Résumés

Liste des figures

Figure 1 : principales voies de biosynthèse impliquées dans la production des métabolites secondaires végétaux (Panda and Khush, 1995). 28
Figure 2 : classification des alcaloïdes selon leur noyau hétérocyclique. 29
Figure 3 : alcaloïdes de défense des plantes. 30
Figure 4 : DMDP, un analogue azafuranose du fructose. 30
Figure 5 : la nicotine, un alcaloïde insecticide connu depuis le 17ème siècle. 30
Figure 6 : un alcaloïde pyrrolizidine séquestré et métabolisé par les insectes. 31
Figure 7 : Dégradation des glucosides cyanogènes. 31
Figure 8 : hydrolyse des glucosinolates. 33
Figure 9 : ODAP, un AANP neurotoxique du genre *Lathyrus* (Papilionoideae). 34
Figure 10 : AANP à soufre ou sélénium, analogues de la cystéine. 35
Figure 11 : la mimosine, un AANP de Mimosoideae. 36
Figure 12 : l'isoprène et sa forme active l'IPD. 37
Figure 13 : formation des isoprénoïdes par une chaîne réactionnelle commune (Richter, 1993). 37
Figure 14 : exemples de monoterpènes simples et composés 38
Figure 15 : la pyréthrine, un insecticide monoterpènique. 38
Figure 16 : Sesquiterpènes linéaires à activité d'hormone juvénile. 38
Figure 17 : Sesquiterpènes cycliques insecticides. 39
Figure 18 : Précocènes I et II, sesquiterpènes à activité antihormonale. 39
Figure 19 : structure *ent*-gibbérellane 40
Figure 20 : structure du noyau stérane, commune aux triterpènes et premier d'entre eux, le squalène 40
Figure 21 : Cucurbitacine B, un triterpène antiappétant. 41
Figure 22 : Azadirachtine, un puissant insecticide triterpène. 41
Figure 23 : exemples de saponines. 42
Figure 24 : l'ecdysone, un phytostérol à effets anti-insectes. 43
Figure 25 : exemples de caroténoïdes et dérivés. 44
Figure 26 : acide shikimique (a) et acide cinnamique (b). 45
Figure 27 : exemples de phénols simples. 46
Figure 28 : structures générales des deux types de tanins végétaux. 46
Figure 29 : coumarine. 48
Figure 30 : définition des différents types de flavonoïdes à partir du squelette flavane. 48
Figure 31 : la quércétine, un flavonol insecticide. 49
Figure 32 : isoflavonoïdes de défense des plantes. 49
Figure 33 : exemples d'anthocyanidines. 50
Figure 34 : structure tridimensionnelle de la concanavaline de *Canavalia brasiliensis* (Sanz-Aparicio et al., 1997). 53
Figure 35 : structures tridimensionnelles de la thionine viscotoxine A3 du gui (a) (Romagnoli et al., To be published), de la défensine Rs-AFP1 du radis (b) (Fant et al., 1998), de la leginsuline du soja (c) (Yamazaki et al., 2003), de la nsLTP1 du blé (d) (Gincel et al., 1994). En trait fin jaune : pont disulfure. 71
Figure 36: Comparaison des séquences de l'isoforme de PA1b découverte par Higgins et al. (1986) (PsaA1b001), de celle isolée au laboratoire (Delobel et al., 1998)

(PsaA1b005) et de la leginsuline du soja (Watanabe et al., 1994). c représente les remplacements conservatifs, n représente les remplacements non-conservatifs..... 80
Figure 37 : Imago de *Sitophilus oryzae* sur grain de maïs. (photo : C. S. Gorsuch, Clemson University)..... 84
Figure 38 : Protocole d'amplification d'ADN par PCR pour des primers non dégénérés (concentrations finales). A : mélange réactionnel, B : cycle de réaction..... 90
Figure 39 : Principe de la PCR nichée. Les amorces de la première PCR (flèches au trait plein) amplifient un premier fragment sur lequel s'hybrident les amorces de la deuxième PCR (pointillés), qui amplifient alors un fragment (en noir) plus court et interne au premier produit. 92
Figure 40: Mortalités cumulées (pourcentages) au 4ème jour de la souche **sensible** de *S. oryzae* sur différentes concentrations en farine entière du pois, du soja, de *M. truncatula* et du haricot. 93
Figure 41 : Mortalité cumulée (pourcentage) au 4ème jour de *S. oryzae* souches sensible S et résistante R sur différentes concentrations en farine entière de *M. truncatula*. 94
Figure 42 : Importance des différentes fractions dans la farine entière du pois, soja, de *M. truncatula* et du haricot (% p/p). 1 : fraction pentane (lipides), 2 : MeOH, 3 : MeOH60, 4 : H_2O5, 5 : H_2O8, 6 : résidu, 7 : pertes. 95
Figure 43 : Mortalités cumulées, au cours du temps, de la souche sensible S de *S. oryzae* sur les fractions MeOH 60 du pois, du soja, de *M. truncatula* et du haricot, et mortalité cumulée de la souche résistante R sur cette même fraction pour *M. truncatula*. 96
Figure 44: SDS-PAGE révélé à l'argent sur les fractions MeOH60 du pois (1), du soja (2), du haricot (3) et de *M. truncatula* (5). marqueurs de poids moléculaires (M,m); SRA1M60 (4). Environ 10 μg de protéines ont été déposées pour les puits 1, 3 et 5 et 20 μg pour le puit 2..... 98
Figure 45 : SDS-PAGE, révélé à l'argent, des fractions aqueuses de *M. truncatula* (1 : pH 8; 2 : pH 5), et des fractions H_2O5 de *G. max* (3), *P. vulgaris* (4) et *P. sativum* (5). marqueurs de poids moléculaires (M, m); SRA1M60 (6); PA1b (7). 99
Figure 46 : Membranes de transfert marquées par l'anticorps antiPA1b et révélées par la phosphatase alcaline. 100
Figure 47: compétition de liaison de ^{125}I-PA1b sur son site de liaison par les extraits MeOH60 des quatre plantes testées et du blé témoin. ● *P. sativum*, ■ *P. vulgaris*, ○ *G. max*, ▲ *M. truncatula*, ⊄ *T. aestivum*. Les quantités portées en abscisse correspondent aux masses EFE présentes dans le puits de réaction..... 101
Figure 48 : Amplification par PCR avec les amorces For1 et Rev 3 d'un fragment de 130 pb correspondant à une partie du gène de PA1b. 1,2,3 : *P. vulgaris* ; 4,5,6 : *M. truncatula* ; 7,8,9 : *G. max* ; 10,11,12 : *P. sativum* ; 13 : témoin négatif ; M : marqueur 1 Kb. 1,4,7,10 : 500 ng de matrice; 2,5,8,11 : 50 ng; 3,6,9,12 : 10 ng. ... 102
Figure 49 : Amplification par PCR avec les amorces For1 et Rev1 d'un fragment de 220 pb correspondant à une partie du gène de PA1b et du début du gène de PA1a. 1,2,3 : *P. sativum* ; 4,5,6 : *P. vulgaris* ; 7,8,9 : *M. truncatula* ; 10,11,12 : *G. max* ; 13 : témoin négatif ; M : marqueur de poids moléculaire. 1,4,7,10 : 100 ng de matrice; 2,5,8,11 : 50 ng; 3,6,9,12 : 10 ng. 102
Figure 50 : Exemple de résultats de la PCR nichée du protocole GenomeWalker sur les banques d'ADN génomique digéré du haricot (séquence PvuA1b001) (6,7,8,9) et de *M. truncatula* (séquence MtrA1b006) (1,2,3,4). banque *Dra*I (1,6); banque *Eco*RV

(2,7); banque *Stu*I (3,8); banque *Pvu*II (4,9); marqueur 1 kb (M); témoins négatifs (5, 10, 12); témoin positif : ADN humain (11)..... 103
Figure 51 : CC1 à 7 : Segments de PA1b délimités par les cystéines 1 à 6..... 104
Figure 52 : alignement des séquences obtenues lors de notre travail et de celles de la littérature. La première colonne identifie les peptides (accession de Swissprot, d'EST, ou d'EMBL, nom du gène) ainsi que l'organisme source : Psa : *Pisum sativum*, Pvu : *Phaseolus vulgaris*, Gma : *Glycine max*, Mtr : *Medicago truncatula*, Van : *Vigna angularis*, Vra : *Vigna radiata*. La première ligne indique la maturation de type pois de la préproprotéine A1 (Higgins et al., 1986). Les cystéines conservées sont en rouges, les autres acides aminés conservés sont en bleu. 106
Figure 53 : chromatogramme d'HPLC en conditions isocratiques : séparation de pics d'isoformes de PA1b de la fraction MeOH60A80 du pois..... 107
Figure 54 : Diversité des Légumineuses..... 116
Figure 55: Gousses de : à gauche *Teline monspessulanus* (Papilionoideae), à droite : *Cercis siliquastrum* (Caesalpinioideae). 116
Figure 56 : Fleurs de Fabaceae à gauche : *Onobrychis viciifolia* (Papilionoideae), au milieu : *Albizia julibrissin* (Mimosoideae) et à droite : *Caesalpinia gilliesii* (Caesalpinioideae) 117
Figure 57 : Relations phylogénétiques putatives entre tribus au sein des Papilionoideae (Polhill, 1981)..... 117
Figure 58: arbres phylogénétiques simplifiés des Fabaceae (dont Papilionoideae plus détaillées) d'après : à gauche : Chapill (http://www.botany.uwa.edu.au/systematics/paps.html). à droite : Wink and Mohamed (2003) : phylogénie moléculaire (gène *rbc*L)..... 120
Figure 59 : PCR avec les amorces For1 et Rev1 sur ADN génomique d'Ad : *Acacia dealbata*, Af : *Amorpha fruticosa*, Tf : fenugrec (à 10 et ' : 2 ng d'ADN), à différentes Ta. 1: 44°C; 2: 45°C; 3: 46°C; 4: 47°C. 124
Figure 60 : PCR avec les amorces For1 et Rev1 sur ADN génomique de différentes espèces végétales (Ta=40°C). M : marqueur de poids moléculaire ; + : témoin positif (pois) ; 1 : *Albizia julibrissin* ; 2 : *Arachis hypogaea* (pour 10 et ' : 20 ng d'ADN) ; 3 : *Dalbergia purpurescens* ; 4 : *Trigonella foenum-graecum* ; 5 : *Intsia bijuga* ; 6 : *Lotus corniculatus* ; 7 : *Lupinus albus* ; 8 : *Robinia pseudoacacia* ; 9 *: Schotia afra* ; 10 : *Sesbania sesban* ; 11 : *Vigna unguiculata* ; - : témoin négatif H_2O. 124
Figure 61 : PCR For1-Rev5 (5) ou -Rev6 (6), test des amorces sur pois (P) et soja (S). - : témoin négatif H20 ; + : témoin For1-Rev1 pois. 125
Figure 62: PCR nichée (deuxième) de marche sur le gène avec l'amorce Rev5 (R5) sur *M. truncatula* banque *Dra*I. - : témoin négatif (H20); + : témoin positif (amorces spécifiques)..... 125
Figure 63 : résultats des PCRs de marche sur le gène réalisées sur le lupin blanc. M: marqueur, +: témoin positif, -: témoin négatif, D: banque *Dra*I, P: banque *Pvu*II, E: banque *Eco*RV. La bande indiquée par une flèche observée dans la banque *Pvu*II a été clonée..... 126
Figure 64 : résultat des PCRs sur ADN de lupin à feuilles étroites réalisées -selon le protocole décrit par Ilgoutz et al. (1997) : les flèches indiquent les deux produits observés, de taille supérieure à 280 pb. - avec les amorces spécifique du gène de l'ADH. M: marqueur, -: témoin négatif, L: ADN de lupin, P: ADN de pois, mix: ADN de lupin et de pois mélangés**.** 127

Figure 65 : Northern blot avec pour sonde marquée un cDNA d'A1 du pois. M : Mimosoideae, C : Caesalpinioideae, P : Papilionoideae..... 128
Figure 66 : Marche sur le gène. A : PCR nichée (Ta:67°C) sur les trois banques DraI, EcoRV, PvuII de Ca: *Caragana arborescens*, Ao: *Alysicarpus ovalifolius*, Bb: *Bituminaria bituminosa*, Vs: *Vigna subterranea*, La: *Lupinus angustifolius*, Vu: *Vigna unguiculata*, Ce: *Canavalia ensiformis*. B : PCR nichée (TD65) sur les banques *Eco*RV de Vu et Ca, *Pvu*II de Bb, *Dra*I de Vs et La et les trois banques de Ce. + : témoins positifs, - : témoin négatif. M, m : marqueurs. 129
Figure 67 : alignement protéique des séquences caractérisées dans notre travail (avec séquences des amorces) et de séquences de la littérature ou des banques EST. Seuls les acides aminés différant de M13709 sont indiqués : - représente le même résidu que pour la séquence M13709. Les espèces sont : Amo : *Astragalus monspessulanus*, Aov : *Alysicarpus ovalifolius*, Bbi : *Bituminaria bituminosa*, Car : *Caragana arborescens*, Cbr : *Canavalia brasiliensis*, Gma : *Glycine max*, Lan : *Lupinus angustifolius*, Lca : *Lonchocarpus capassa*, Lco : *Lotus corniculatus*, Lla : *Lathyrus latifolius*, Mal : *Melilotus albus*, Mse : *Mundulea sericea*, Mtr : *Medicago truncatula*, Ovi : *Onobrychis viciifolia*, Psa : *Pisum sativum*, Pvu : *Phaseolus vulgaris*, Tfo: *Trifolium foenum-graecum*, Van : *Vigna angularis*, Vhi : *Vicia hirsuta*, Vra : *Vigna radiata*, Vsu : *Vigna subterranea*, Vun : *Vigna unguiculata*. 131
Figure 68 : arbre de phylogénie moléculaire (séquences "incomplètes", BioNJ, distance Galtier et Gouy). En vert : séquences de Phaséoloïdes, en pourpre : Galégoïdes. Seules les valeurs de bootstrap supérieure à 50 sont indiquées. Les noms des séquences sont légendés Figure 67. 132
Figure 69 : Analyse en Composantes Principales (Diol, 2003) rouge : Galégoïdes, vert : Phaséoloïdes, jaune : Génistoïdes, bleu : Aeschynoménoïdes. variables S : souche sensible, R : souche résistante, 0 à 4 : fractions MeOH à résidu**.**..... 139
Figure 70 : Représentation de la distribution des 3 classes de toxicité (C1 à 3) en fonction de la position taxonomique des différentes espèces. 140
Figure 71 : Test d'inhibition de liaison : pourcentage de ligand lié pour deux concentrations en fraction MeOH60 de différentes espèces végétales. Les concentrations sont exprimées en mg d'Equivalent Farine Entière/ puits. Le blé sert de témoin négatif. 143
Figure 72 : Electrophorèse en tris-tricine (révélation au nitrate d'argent) sur fractions MeOH60A80 de quelques espèces. S: SRA1M60, P: PA1b, M: marqueur de poids moléculaire, Aov : *Alysicarpus ovalifolius*, Bbi : *Bituminaria bituminosa*, Bsp : *Bolusanthus speciosus*, Cer : *Crotalaria eremaea*, Lca : *Lonchocarpus capassa*, Mal : *Melilotus albus,* Ovi : *Onobrychis viciifolia*, Tfo : *Trigonella foenum-graecum*, Tre : *Templetonia retusa*, Vsu : *Vigna subterranea*, Vun : *V. unguiculata*. Les flèches indiquent la position de PA1b. 144
Figure 73 : Représentation schématique d'un vecteur d'expression. 156
Figure 74 : Représentation schématique d'une cassette d'expression..... 157
Figure 75 : Représentation schématique des membranes d'*Escherichia coli*. (Extrait de Prescott et al., 1995). 159
Figure 76 : Corps d'inclusion dans des cellules d'*E. coli* visualisés au microscope électronique. 160
Figure 77: schéma du système de fermentation 167
Figure 78 : Stratégie générale de polymérisation d'un gène synthétique. Dans notre cas A= *Nco*I, B= *Mst*I, B'= *Eco*RV, C= *Hin*DIII. 171

Figure 79 : carte de restriction des inserts, du gène codant M-PA1b (PA1b avec une méthionine en N-terminal) et du plasmide pPA1b. 172
Figure 80 : vecteur d'expression pET-DP-PA1b 174
Figure 81 : à gauche : suivi de l'expression (SDS-PAGE, Bleu de Coomassie) ; à droite : immunodétection de TRX-DP-PA1b. T0 à 3 : prélèvements avant et 1 à 3 heures après induction; S: PA1b; M: marqueur. 174
Figure 82 : comparaison des protéines totales (tot) et des protéines solubles (sol) d'*E. coli* après expression (SDS-PAGE, Bleu de Coomassie). 175
Figure 83 : Fractions de chromatographie d'affinité avec élution par palier d'imidazole (SDS-PAGE, Bleu de Coomassie). Sol : protéines solubles du cytoplasme, Nr : fraction non retenue, B : lavage en tampon de fixation, W : tampon de lavage, E : éluat, M : marqueur. 176
Figure 84 : SDS-PAGE des fractions obtenues par chromatographie d'affinité en urée 6M (révélation au Bleu de Coomassie). Tot: protéines totales, Nr: fraction non retenue, B et W: lavages, E: éluat, M: marqueur. 177
Figure 85 : chromatogramme d'IEC post-affinité. 1 à 4 : pics d'élution analysés en électrophorèse (Figure 86). Environ 300 mg de protéines ont été déposées sur la colonne. 177
Figure 86 : SDS-PAGE sur les fractions d'IEC post-affinité (Bleu de coomassie). E: éluat d'affinité (injecté en IEC), Nr : fraction non retenue, 1 à 4 : pics d'élution à 20, 40 et 100% de tampon 2 (Figure 85). 178
Figure 87 : Purification de la protéine de fusion TRX-DP-PA1b par IEC suivie de chromatographie d'affinité (SDS-PAGE, Bleu de Coomassie). sol: protéines solubles du cytoplasme, IEC : éluat d'IEC, E: éluat d'affinité post-IEC, M : marqueur. 178
Figure 88 : chromatographie d'affinité (sur colonne de nickel), élution par gradient linéaire d'imidazole (tampon E). Fractions Nr à E4 analysées par SDS-PAGE Figure 89. Nr: fraction non retenue, B: lavage (10mM), W: lavages (1: 20 mM, 2: 40 mM d'imidazole), E1 à 4: fractions d'élution. 179
Figure 89 : SDS-PAGE (Bleu de Coomassie) des fractions de la chromatographie d'affinité à gradient d'élution linéaire (Figure 88). Sol: protéines solubles du cytoplasme, Nr: fraction non retenue, B: lavage (10mM), W: lavages (1: 20 mM, 2: 40 mM d'imidazole), E1 à 4: fractions d'élution. 179
Figure 90 : chromatogramme d'HPLC sur la protéine fusion TRX-DP-PA1b. 180
Figure 91 : ESI-MS: spectre non déconvolué de la protéine de fusion TRX-DP-PA1b purifiée. Masse moléculaire moyenne observée : 20 650,4 Da (masse théorique réduite : 20658,5 Da). 181
Figure 92 : spectre de masse non déconvolué de rPA1b purifié, provenant de clivage à l'acide formique. 182
Figure 93 : spectre de masse non déconvolué d'isoformes de PA1b natif soumises à l'acide formique 50°C, 24h, composant un pic d'HPLC de temps de rétention identique aux isoformes natives. 183
Figure 94 : suivi de clivages acides. (SDS-PAGE, Bleu de Coomassie). Fus : fusion non clivée A : H : HCl 40mM 50°C, F: acide formique 70% 50°C; durée du clivage : 1 = 4h, 2 = 9h, 3 = 24h, 4 = 48h B : H' : HCl 80°C; durée du clivage: 1 à 3 h. 184
Figure 95 : Gel d'électrophorèse en tris-tricine (révélation à l'$AgNO_3$) : solubilisation différentielle en Méthanol 60% sur mélange de clivage en acide formique 50°C (CT) et sur fraction non retenue de chromatographie d'affinité après clivage (Nr). Ins : culot, sol : fraction soluble, M : marqueur, S: PA1b. 185

Figure 96 : Gel d'électrophorèse tris-tricine (révélation à l'$AgNO_3$): Fractions de chromatographie d'échange d'ion (sur DEAE) sur mélange de clivage HCl 50°C 24 h. M: marqueur, S : PA1b, Nr : fraction non retenue, l1 et 2: lavages, E1 et 2: 20%B, E3: 40% B, E4: 100%B..... 186
Figure 97 : Fractions de chromatographie d'affinité post-clivage (électrophorèse tris-tricine, $AgNO_3$). C: mélange de clivage (acide formique), Nr: fraction non retenue, B: lavage en tampon de fixation, W: lavage en tampon de lavage, E: éluat. 187
Figure 98 : Comparaison des chromatogrammes d'HPLC d'isoformes de PA1b natives (F1), d'un mélange de clivage à l'acide formique (CT) et du soluble MeOH60 de travers de chromatographie d'affinité post-clivage correspondant (Tv)..... 188
Figure 99 : Gel d'électrophorèse en Tris-Tricine (révélation au nitrate d'argent) sur les fractions d'HPLC correspondant à la zone rPA1b (Figure 98) provenant de : **A** : mélange de clivage total **B** : soluble MeOH60 de la fraction non retenue par chromatographie d'affinité post-clivage (Nr). 188
Figure 100 : chromatogrammes d'HPLC d'un soluble MeOH60 de clivage entérokinase (eK) et pour comparaison, d'isoformes de PA1b (F1) et d'un soluble MeOH60 de travers de chromatographie d'affinité post-clivage HCl (Tv)..... 189
Figure 101 : gels d'électrophorèse sur clivages à l'entérokinase : **A** : SDS-PAGE (Bleu de Coomassie) sur le mélange de clivage à l'entérokinase C (24°C), T : témoin fusion non clivée. **B**: Electrophorèse en tris-tricine (révélation au nitrate d'argent) sur les fractions HPLC du soluble MeOH60 (Figure 100) d'un clivé 30°C (clivage quasi total). Les fractions 1 à 4 représentent la "bosse" rPA1b (de 16 à 19'5 min de TR), la fraction 8 le pic à 22-23 min (TRX), les fractions 5 à 7 représentent des pics intermédiaires entre le massif rPA1b et le pic TRX. M: marqueur, S: PA1b**.** 190
Figure 102 : spectres de masse de rPA1b. A : provenant de clivage enzymatique (M théorique oxydée : 3943,6 Da), B: provenant de clivage HCl en restant soluble (M théorique oxydée : 3969,7 Da). 191
Figure 103 : Tests d'inhibition de liaison (Diol, 2003). Pourcentage de ligand lié pour sPA1b 1 et 2 : lots de PA1b synthétisé chimiquement (testés à 10^{-7} M), rPA1b 1 et 2 : lots de PA1b recombinant (clivage enzymatique, testés à 10^{-6} M)..... 192
Figure 104 : Structures tridimensionnelles des inhibiteurs de protéase EETI (Chiche et al., 1989) et CPI (Rees and Lipscomb, 1982), et de PA1b (Jouvensal et al., 2003). 199

Liste des tableaux

Tableau 1 : Principales classes de métabolites secondaires impliqués dans la défense des végétaux face aux insectes et autres ravageurs. Nd : non détérminé (Hartley, 2001 ; Panda and Khush, 1995). 27
Tableau 2 : AANP toxiques et leurs analogues protéiques..... 34
Tableau 3 : Toxicité envers les mammifères et les différents ordres d'insectes des protéines de plante à activité insecticide et des endotoxines de *Bacillus thuringiensis* (*Bt*). 52
Tableau 4: Quelques exemples de plantes transgéniques exprimant des lectines..... 55
Tableau 5 : Exemples de plantes transgéniques exprimant des inhibiteurs de protéases... 61
Tableau 6 : Toxicité des différentes fractions (testées à 100% EFE) vis-à-vis des deux souches de charançons sensible (S) et résistante (R). Temps létaux 50% ± erreur standard. 96
Tableau 7 : Peptides observés par spectrométrie de masse sur des fractions sélectionnées des plantes étudiées..... 109
Tableau 8 : Comparaison des temps létaux 50 des différentes souches de *S. oryzae* isogéniques ou non; pBGW: test de rang de Breslow-Gehan-Wilcoxon (différentiel de mortalité entre les isogéniques (W/I) ou les non isogéniques (B/C)). 141
Tableau 9 : Enzymes et agents chimiques utilisés pour le clivage des protéines de fusion 163
Tableau 10 : Caractéristiques des souches et plasmides utilisés..... 164

Abréviations

°C : degré celsius
αAI : inhibiteur d'alpha-amylase
β-Me : β-mercaptoéthanol
ADN : acide deoxyribonucléique
ARN: acide ribunucléique
CL_{50} : concentration létale 50
Da : dalton (kDa : kilo dalton)
db : double brin
DEAE : diéthylaminoéthyl
DMDP : 2R, 5R-dihydroxyméthyl-3R, 4R-dihydroxypyrrolidine
EDTA : éthylène diamine tetraacétate
F1 : mélange d'isoformes de PA1b extraits du pois
H2O5 : fraction de farine soluble à l'eau pH 5
H2O8 : fraction de farine soluble à l'eau pH 8
HPLC : chromatographie liquide haute pression
IEC : chromatographie d'échange d'ions
IP : inhibiteur de protéase
IPD : isopenténYl diphosphate
ITS : internal transcribed spacers
MeOH : fraction de farine soluble au méthanol 100%
MeOH60 : fraction de farine soluble au méthanol 60 %
MeOH60A80 : fraction soluble à l'acétone 80% obtenue à partir de MeOH60
mM : millimolaire
MWM : molecular weight marker : marqueur de poids moléculaire
nsLTP : protéine de transfert des lipides non spécifique
PA1b : Pea Albumin 1 sous-unité b
pb : paire de base (kb : kilo base)
pH : potentiel hydrogène
pI : point isoélectrique
PMSF : Phenyl-Methyl-Sulfonyl Fluoride (inhibiteur de protéase à sérine)
rPA1b : PA1b recombinant
SRA1M60 : fraction de farine de pois riche en albumine, soluble en méthanol 60%
SDS : sodium dodécylsulfate
SDS-PAGE : SDS polyacrylamide gel electrophoresis
SM : spectrométrie de masse
Ta : annealing temperature
TD : touch down (PCR)
TFA : acide trifluoroacétique
TL_{50} : temps létal 50
Tm : melting temperature
TRX : thiorédoxine
TTGE : Tris-tricine gel electrophoresis

Avant-propos

Ce mémoire de thèse a pour objet d'étude les peptides à activité entomotoxique de type **Albumine 1b** des graines de Légumineuses. Cette famille peptidique, de par sa toxicité pour certains insectes phytophages appartient au groupe très large et varié des **molécules de défense** des plantes. Les graines de Légumineuses sont riches nutrionnellement, d'où leur utilisation importante dans l'alimentation humaine et animale et leur attractivité potentielle pour de nombreux ravageurs. Elles sont en conséquence protégées physiquement et chimiquement. Elles peuvent contenir un arsenal de composés toxiques pour les pathogènes et/ou les herbivores, composés non apparentés aux Albumines 1b et pouvant interférer avec notre étude.

Afin de mieux comprendre la toxicité des plantes envers les ravageurs, et en particulier celle des **graines de Légumineuses** envers les **insectes**, nous nous intéresserons tout d'abord, par une approche bibliographique, aux différents types de molécules de défense des végétaux, par leurs structures et modes d'action, qu'ils soient de nature protéique comme les Albumines 1b ou non. Après situation du contexte de l'étude, nous présenterons nos travaux et résultats concernant la diversité structurale des Albumines 1b et de la toxicité des graines de Légumineuses. Trois parties décriront la mise en place de l'étude, son élargissement, et l'établissement d'un système d'expression hétérologue en vue d'études de relations structure-fonction.

Foreword

This thesis deals with insecticidal Albumin 1b peptides from legume seeds. This peptide family, due to its toxicity to some phytophagous insects, belongs to the large and varied group of plant defence compounds. Legume seeds are nutrionally rich, and so important in human and animal alimentation and potentially attractive to many pests. Consequently they are physically and chemically protected. They contain many toxic compounds towards pathogens and phytophagous animals, compounds not related to Albumins 1b and which may interfere within our study.

To better understand the toxicity of plants to pests and particularly of legume seeds to insects, we will first focus, in a bibliographic introduction, on the different types of plant defence molecules, proteic as A1b or not proteic, on their structure and on their modes of action. After a presentation of the context of our study , we will describe our work and results concerning the structural diversity of A1b, and the variability of legume seed toxicity. Three parts will deal respectively with the methodological approach set up for this study, its widening to screen the whole Fabaceae family, and the development of a heterologous expression system for structure-function relationships studies.

Chapitre 1 : Introduction bibliographique et contexte de l’étude.

Sommaire détaillé du premier chapitre :

Introduction générale 25
I. Les défenses chimiques des plantes face aux insectes. 25
A. Les allélochimiques non protéiques à activité entomotoxique. 27
1) Les composés azotés. 28
(a) Les alcaloïdes 28
(b) Les composés cyanogènes 31
(c) Les glucosinolates 32
(d) Les acides aminés non protéiques 33
2) Les terpénoïdes. 36
(a) Les monoterpènes 37
(b) Les sesquiterpènes 38
(c) Les diterpènes 39
(d) Les triterpènes 40
a) Les cucurbitacines 40
b) Les limonoïdes 41
c) Les saponines 41
d) Les stérols 42
e) Les alcaloïdes stéroïdaux 43
f) Les glucosides cardiotoniques 43
(e) Les tétraterpènes 43
(f) Les polyterpènes 44
3) Les composés phénoliques. 44
(a) Les composés phénoliques simples et les quinones, les acides phénoliques et les tanins 45
(b) Les acides phénylpropanoïques (C6C3), coumarines et lignines. 47
(c) Les flavonoïdes 48
B. Les protéines végétales de défense 51
1) Les protéines entomotoxiques 51
(a) Lectines, hémilectines et autres protéines apparentées 53
a) Les lectines vraies 53
b) Les hémilectines et RIP 55
c) Les arcélines 56
(b) Canatoxines et uréases 57
(c) Inhibiteurs enzymatiques 57
a) Inhibiteurs de protéases (IP) 58
(i) Les inhibiteurs de protéases à sérine 58
(ii) Les inhibiteurs de protéases à cystéine 59
(iii) Les inhibiteurs de métalloprotéases et de protéases acides 59
(iv) Le rôle des inhibiteurs de protéases chez les plantes 60
(v) Effets de l'ingestion d'inhibiteurs de protéases sur les protéases des insectes 61
b) Inhibiteurs d'alpha-amylases (IA) 64
(i) Les IA de type lectine. 64
(ii) Les IA des céréales. 65
(iii) Les IA de type Kunitz 65
(iv) Les IA de type thaumatine. 65

(v) Interêt des inhibiteurs d'enzymes bifonctionnels..... 66
(d) Chitinases et protéines affines de la chitine..... 66
2) Les peptides de défense (antimicrobiens et insecticides)..... 67
(a) Principales familles structurales..... 67
a) Les thionines..... 68
b) Les défensines..... 69
c) Les knottines..... 70
d) Les LTP..... 70
(b) Cibles et mode d'action des peptides végétaux de défense..... 72
a) Phospholipides membranaires..... 72
b) Protéines membranaires..... 73
c) Canaux..... 74
d) Enzymes..... 74
(i) Inhibiteurs d'α-amylases (IA)..... 74
(ii) Inhibiteurs de protéases (IP)..... 75
e) Cibles intracellulaires..... 75
C. Conclusion..... 76
II. Contexte de l'étude et présentation des travaux..... 78

Introduction bibliographique

I. Les défenses chimiques des plantes face aux insectes.

Les relations plantes-insectes font l'objet d'importantes études aux possibles retombées économiques. Si certains insectes sont particulièrement utiles aux plantes en assurant, par exemple, la pollinisation nécessaire à la reproduction du végétal, nombreux sont les nuisibles qui s'alimentent de la plante, provoquant parfois des dégâts irréparables.
Les végétaux subissent la pression des **insectes phytophages** depuis l'apparition de ceux-ci (il y a plus de 145 millions d'années). Bien qu'immobiles, de temps de génération souvent long et de faible taux de recombinaison, les végétaux constituent le taxon le plus important en terme de biomasse sur terre (producteurs primaires). Ils possèdent un efficace système de résistance, basé sur des caractères physiques, chimiques et développementaux, vis-à-vis de ces ravageurs. La **résistance** des plantes face aux insectes définit la capacité de celles-ci à éviter ou réduire les dommages causés par ces derniers (Kogan, 1975). Elle est le plus souvent évaluée de façon relative et au niveau infraspécifique puisque dépend de la comparaison avec une plante sensible plus sévèrement touchée dans les mêmes conditions et constitue une règle générale[1]. Elle varie de plus dans le temps, l'espace (tissu considéré) et selon les conditions environnementales. Trois catégories de résistance ont été définies par Painter (Painter, 1951) :

[1] Le nombre d'espèces de pathogènes ou de ravageurs pour une espèce végétale donnée est très faible par rapport au nombre global de pathogènes et ravageurs des plantes.

- La non-préférence ou **antixénose** définit le type de résistance induite par le rejet de la plante comme hôte de l'insecte. Elle peut être due à la présence de caractères morphologiques ou chimiques repoussant l'insecte ou par l'absence de composés stimulant l'alimentation ou l'oviposition.
- l'**antibiose** indique que l'insecte est touché dans sa physiologie par la plante. Là aussi caractères physiques et chimiques peuvent être impliqués entraînant une inhibition de croissance, une moindre fécondité voire la mort de l'insecte.
- la **tolérance** est la capacité d'une plante à supporter ou se rétablir des dégâts causés par l'herbivore.

Les **protections mécaniques** des plantes sont diverses et nous ne nous étendrons pas sur cette composante. Des épines ou poils peuvent constituer une barrière efficace contre certains herbivores en empêchant l'installation ou l'accès à l'aliment de l'insecte. Les trichomes crochus de *Phaseolus vulgaris* peuvent, par exemple, tuer les nymphes d'*Empoasca fabae* en les empalant (Pillemer and Tingey, 1978). Les cires de surface qui protègent les feuilles de la dessication et des maladies sont également répulsives ou gènent l'adhésion de certains insectes. Toutefois, elles stimulent parfois l'oviposition ou l'alimentation. L'épaisseur des tissus participent également à la résistance de certaines plantes. L'oviposition et l'alimentation des larves de *Chalcodermus aenus* sont en effet limitées par l'épaisseur de la paroi des gousses de certaines variétés de *Vigna unguiculata* (Fery and Cuthbert, 1979).
Toutefois, les **composés chimiques** produits par les plantes sont probablement le facteur le plus important contrôlant le comportement des insectes dans la nature (Schultz, 1988) et l'intérêt porté aux défenses chimiques des plantes n'a cessé de croître depuis le milieu des années 60. Ces composés de structures et natures variées peuvent avoir des effets répulsifs (lors de l'orientation olfactive), antiappétants (lors de l'essai d'alimentation) ou toxiques pour les insectes et d'autres herbivores. Certains composés ont un rôle indirect dans la défense des plantes, attirant des parasitoïdes ou des prédateurs du ravageur. Les composés influant les interactions entre individus d'espèces différentes sont appelés sémiochimiques ou allélochimiques. S'ils sont bénéfiques pour le producteur et le receveur, on les nomme synomones, s'ils ne sont bénéfiques qu'au receveur, on les nomme kairomones, enfin les allomones sont bénéfiques au producteur. Les défenses chimiques peuvent être produites constitutivement ou en réponse à une attaque, localement ou dans toute la plante.
Les différents types de molécules de défense des plantes, protéiques ou non, sont présentés de façon non exhaustive dans cette partie bibliographique. Leurs activités insecticides sont mises en avant bien que d'autres activités, de défense ou non, puissent également en être mentionnées.

A. Les allélochimiques non protéiques à activité entomotoxique.

Composés de défense non protéiques ne participant pas au métabolisme de base commun à toutes les plantes, les allélochimiques appartiennent au **métabolisme** dit **secondaire**. Le Tableau 1 présente les principales classes de métabolites secondaires impliqués dans la résistance des plantes aux insectes. Les métabolites azotés, les polyphénols et les terpénoïdes sont présentés ici de façon plus détaillée bien que non exhaustive.

Tableau 1 : Principales classes de métabolites secondaires impliqués dans la défense des végétaux face aux insectes et autres ravageurs. Nd : non déterminé (Hartley, 2001 ; Panda and Khush, 1995).

Classe	*Origine, biosynthèse*	*Nombre de structures*	*Distribution et activité*
Alcanes, aldéhydes, cires	Acétate-malonate	nd	Feuilles : cires cuticulaires
Lignines, tannins	Acide shikimique	nd	Paroi cellulaire Précipitent les protéines
Terpenoïdes	Acétate-mevalonate		
Monoterpènes		1300	Huiles essentielles (conifères), forte odeur, repoussants
Sesquiterpènes		6500	Répandus, amers, toxiques
Diterpènes		2900	Latex et résine, parfois toxiques
Triterpènes		3600	Répandus, amers, toxiques, colorés
Stéroïdes	Acétate-mevalonate et autres	70	+ de 100 familles végétales
Phytoecdysones			Antihormones
Polyphénols			
Simples	Shikimate	1000	Universels (feuilles)
Coumarines	Shikimate-chorismate	800	70 familles Dicotylédones, toxiques, repoussants
Flavonoïdes	Shikimate- malonate	4000	Universels, colorés, toxiques
Quinones	Shikimate- mevalonate	800	Répandus, colorés
Alcaloïdes	hétérogène	10000	Angiospermes, toxiques, amers
Acides aminés non protéiques	Acides aminés	600	Graines de Légumineuses, répandus, toxiques
Cyanogènes	Acides aminés	60	2500 espèces de 130 familles, toxiques
Glucosinolates	Acides aminés	100	Crucifères, amers, âcres
Amines		100	Angiospermes, repoussants, hallucinogènes

Les principales voies de biosynthèse de ces composés, souvent complexes, peuvent être schématisées comme en Figure 1.

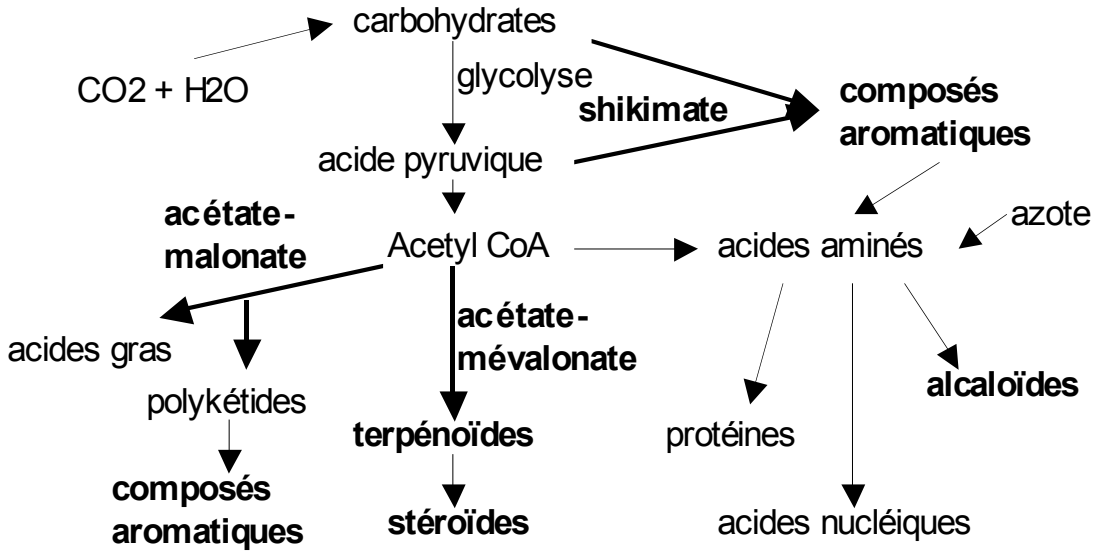


Figure 1 : principales voies de biosynthèse impliquées dans la production des métabolites secondaires végétaux (Panda and Khush, 1995).

On distingue les composés azotés des composés non azotés comme les terpénoïdes et les polyphénols.

1) Les composés azotés.

Très nombreux et diversifiés, les métabolites secondaires azotés de défense ont pour point commun de dériver d'acides aminés protéiques ou non. On distingue les alcaloïdes, les composés cyanogènes, les glucosinolates et les acides aminés non protéiques eux-même.

(a) Les alcaloïdes

Les alcaloïdes sont des bases organiques, présents dans toutes les classes d'organismes vivants mais provenant pour la plupart de plantes supérieures. Ils sont communs chez les Angiospermes, dont 20% en sont riches, et rares chez les végétaux inférieurs, avec toutefois des exceptions comme le pacletaxol de l'if, la lycopodine de *Lycopodium*, la palustrine d'*Equisetum* et même l'ergometrine de *Claviceps* (champignon) (Woolley, 2001).

Ce groupe contient plus de 10 000 composés dont le premier à avoir été découvert, en 1803 par Derosne, est la narcotine (noscapine) de l'opium. Ils sont présents dans la vacuole des cellules sous forme de sels et peuvent en être extraits par de l'eau acidifiée, de l'éthanol, ou par des solvants organiques si la plante est rendue alcaline. Ils sont classifiés selon leur hétérocycle (Figure 2).

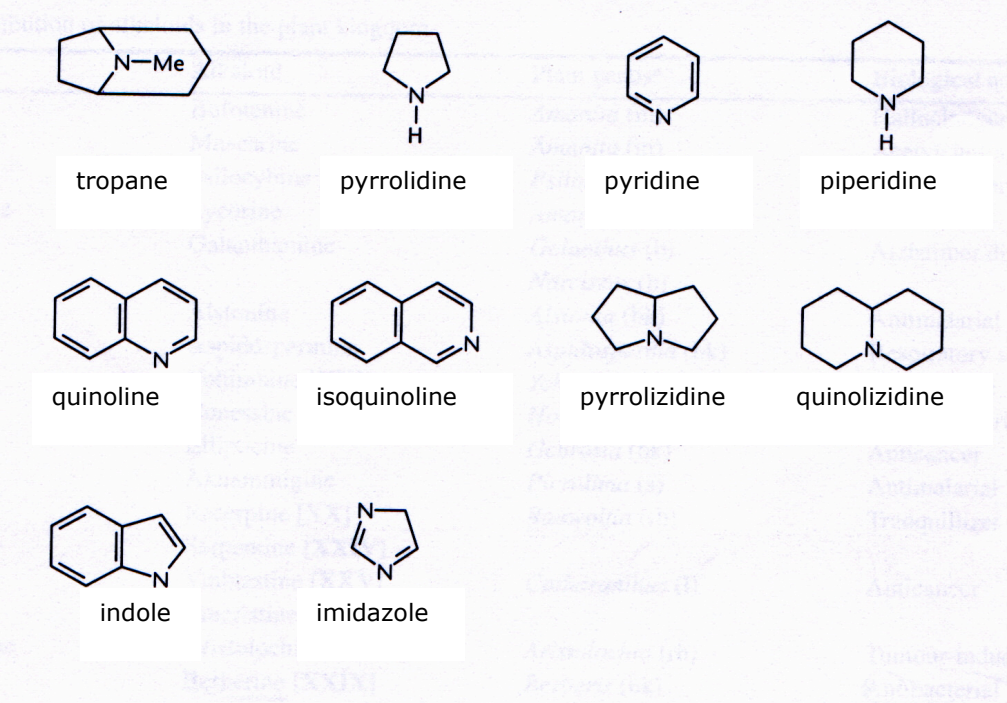


Figure 2 : classification des alcaloïdes selon leur noyau hétérocyclique.

La majorité des alcaloïdes dériveraient des acides aminés tryptophane, tyrosine, lysine, histidine et ornithine par des voies de biosynthèse longues, complexes et coûteuses en énergie. Le site de synthèse de ces composés est difficilement identifiable car il n'est pas forcément le site de stockage. L'identification de gènes spécifiant certaines enzymes de la voie de synthèse d'alcaloïdes a permis dans certains cas de suivre leur expression et de déterminer le site de cette biosynthèse. L'enzyme hyocyamine 6-hydroxylase, qui catalyse deux étapes de la synthèse de la scopolamine (hyoscine) est exprimée dans la région du péricycle des racines des Solanaceae *Atropa* et *Hyoscyamus* (Woolley, 2001).

En plus d'un rôle probable dans le métabolisme de l'azote et son stockage, les alcaloïdes présentent diverses activités biologiques, témoins de leur rôle dans la protection des végétaux. Ainsi, de nombreux alcaloïdes sont des poisons pour l'homme et le bétail comme la strychnine (Figure 3b) des graines de la Loganiaceae *Strychnos nux-vomica,* un alcaloïde de type monoterpène indole, type dérivé du tryptophane présent dans trois familles essentiellement tropicales : les Loganiaceae, les Apocynaceae et les Rubiaceae. D'autres alcaloïdes sont des drogues (cocaïne des feuilles d'*Erythroxylon coca*, Erythroxylaceae), et certains ont des vertus thérapeutiques (morphine des Papaveraceae, quinine de l'écorce des Rubiaceae *Cinchona*). La morphine (Figure 3c) est un alcaloïde de type benzyl-isoquinoline, type qui, bien que représentant la plus large classe avec 2500 composés identifiés, est limité aux Ranunculales.
Certains alcaloïdes présentent des activités antimicrobiennes comme les calystégines antivirales des feuilles de *Morus* (Moraceae) et des racines de *Calystegia* (Convolvulaceae), la canthin-6-one de *Pentacerus* (Rutaceae) et la berbérine (Figure 3a) de l'écorce des Berberidaceae antibactériennes.

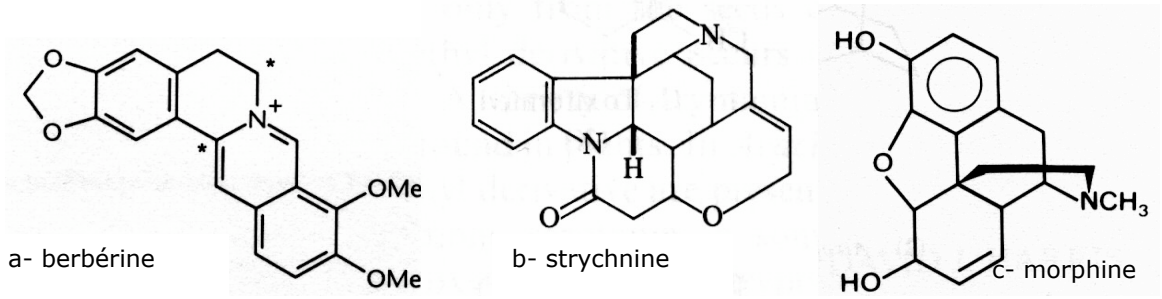


Figure 3 : alcaloïdes de défense des plantes.

Enfin, dans les relations plantes-insectes, les alcaloïdes ont principalement un rôle de défense (répulsifs et toxiques). Les alcaloïdes de type indole quinine et strychnine sont répulsifs pour nombre d'insectes. Les quinolizidines, particulièrement représentées dans le genre *Lupinus,* sont répulsifs pour les mammifères, les insectes et les mollusques. A une concentration de 3 à 50 mM, ils sont mortels pour les insectes et inhibent les croissances bactérienne et fongique dès 1 mM (Wink, 1985). Les alcaloïdes polyhydroxylés, présents dans cinq familles (Légumineuses, Polygonaceae, Aspidiceae, Moraceae et Euphorbiaceae), miment les sucres structurellement. Ils inhibent le métabolisme des sucres et sont d'efficaces antiappétants pour divers insectes (Fellows et al., 1986). Le 2R,5R-dihydroxyméthyl-3R,4R-dihydroxypyrrolidine (DMDP, Figure 4) tue les larves de *Callosobruchus maculatus* à une concentration de seulement 0,03 %, inhibant fortement l'enzyme digestive α-glucosidase (Evans et al., 1985). Les glucosidases sont présentes également dans les sensilles du goût et sont impliquées dans la perception des sucres (Nakashima et al., 1982). Leur inhibition expliquerait l'action antiappétante du DMDP chez certains Lépidoptères (Simmonds et al., 1990). En plus d'inhiber certaines glycosidases, les alcaloïdes polyhydroxylés interfèrent avec la biosynthèse des glycoprotéines intracellulaires. Ce sont des outils prometteurs dans la recherche virale, contre le cancer et en immunologie.

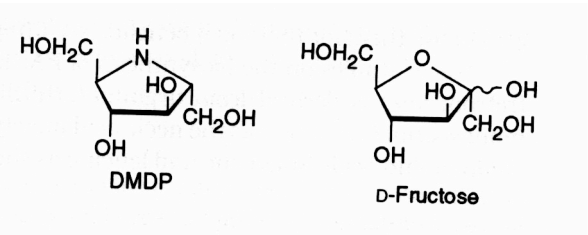


Figure 4 : DMDP, un analogue azafuranose du fructose.

Le plus connu des alcaloïdes entomotoxiques reste cependant la nicotine, utilisée depuis 1690 en France pour lutter contre les pucerons (Figure 5). Elle agit au niveau des synapses du système nerveux central des insectes (agoniste du récepteur à l'acétylcholine).

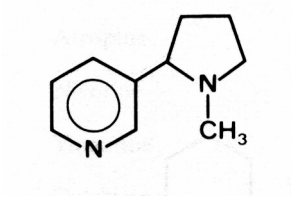


Figure 5 : la nicotine, un alcaloïde insecticide connu depuis le 17ème siècle.

Cependant, si les alcaloïdes sont d'efficaces défenses pour les plantes contre toutes sortes de ravageurs, il existe des herbivores qui se sont adaptés à ces composés et les détournent à leur profit. C'est le cas des genres Arctiidae et Ctenuchidae qui séquestrent des alcaloïdes pyrrolizidines. De nombreux Lépidoptères, Orthoptères, Homoptères et Coléoptères se protègent également ainsi de leur prédateurs. Les pucerons ingèrent les pyrrolizidines de la sève des plantes et les coccinelles se nourrissant des pucerons accumulent ces alcaloïdes. Certains Arctiidae se nourrissant sur des Boraginaceae convertissent même les pyrrolizidines séquestrées en hydroxydanaidal, la phéromone mâle de cour (Figure 6) (Woolley, 2001). Ces alcaloïdes pyrrolizidines dont environ 250 structures sont connues chez certaines Asteraceae, la plupart des Boraginaceae et dans le genre Crotalaria (Légumineuses), sont souvent toxiques pour l'homme et le bétail.

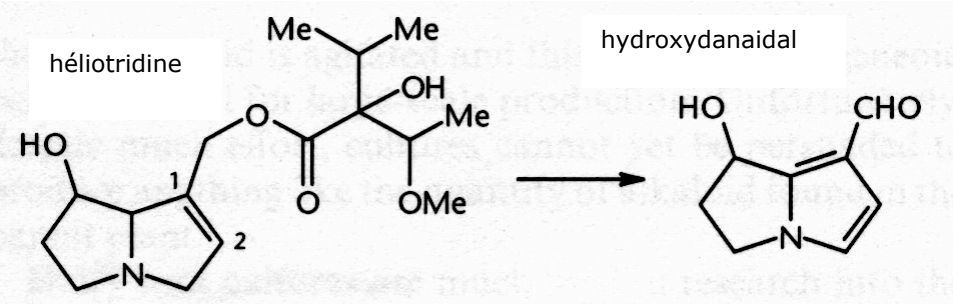


Figure 6 : un alcaloïde pyrrolizidine séquestré et métabolisé par les insectes.

(b) Les composés cyanogènes

La cyanogénèse, formation d'acide cyanhydrique (HCN) libre, est associée aux cyanohydrines (α-hydroxynitriles) dérivés d'acides aminés protéiques ou non et stabilisés par glycosylation, les cyanoglucosides, ou sous forme de cyanolipides. Probablement toutes les plantes ont la possibilité de synthétiser des cyanoglucosides, mais la plupart ne les stockent pas (Schoonhoven et al., 1998). Seul le β-D-glucose est attaché directement à la partie aglycone. Un deuxième sucre peut ensuite être ajouté. Stockés dans les vacuoles, ces composés peuvent être hydrolysés enzymatiquement lors d'une blessure du tissu végétal et du HCN hautement toxique est alors relaché (Figure 7) (Poulton, 1990). Comme pour la famille suivante (glucosinolates), ces composés montrent l'intérêt d'une « arme binaire » par la compartimentalisation du composé de défense stocké sous forme inactive d'une part et de son régulateur (glucosidase) stocké d'autre part.

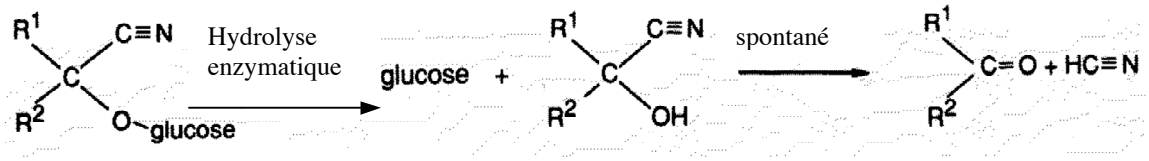


Figure 7 : Dégradation des glucosides cyanogènes.

Le premier cyanoglucoside a été isolé il y a plus de 150 ans et l'intérêt pour ces composés reste grand de par leur variété (60 composés connus), leur distribution (dans 2650 espèces de 130 familles de plantes vasculaires) et leur importance dans certaines plantes alimentaires. Le manioc, *Manihot esculenta*, est la plante cyanogène la plus consommée par l'homme. La linamarine est le cyanoglucoside prédominant dans les feuilles et les tubercules de cette plante, tous deux consommés. La concentration des cyanoglucosides

dans la plante varie énormément selon l'âge, la variété, l'organe, les conditions de croissance. Si les espèces du genre *Sorghum* ne présentent pas de dhurrine (dérivé de la tyrosine plus glucose) dans ses graines, ce composé peut atteindre 30% du poids sec des jeunes pousses (Halkier and Moller, 1989). Le plus connus des cyanoglucosides est l'amygdaline (dérivé de la phénylalanine plus gentiobiose) responsable de la toxicité des graines de nombreuses Rosaceae dont les amandes amères, les pêches, les abricots. La sélection d'amandes douces a permis la réduction de la concentration en cyanogènes et de l'amertume associée à ces composés.
Les glucosides cyanogènes ont été utilisés en taxonomie. Les fougères ne contiennent que des cyanoglucosides dérivés de la phénylalanine comme la prunasine et les Gymnospermes se limitent à un dérivé de la tyrosine, la taxiphylline. Toutefois, certaines familles végétales présentent un profil plus complexe et une diversité plus importante, limitant l'intérêt de ces composés pour des études phylogéniques.
En plus d'un rôle probable dans le **métabolisme général de l'azote**, ces composés jouent un rôle important de **défense contre les herbivores et les pathogènes** (Hruska, 1988). HCN, amer, est en effet un anti-appétant efficace contre divers insectes (Woodhead and Chapman, 1986). La dhurrine est, par exemple, une défense potentielle dans la lutte contre certains orthoptères ravageurs en Afrique de l'Ouest et de l'Inde (Woodhead and Bernays, 1977). Toutefois, certains insectes se sont habitués à l'amertume des cyanoglucosides comme le papillon sud-africain *Acraea horta*, dont les larves ne sélectionnent que des feuilles contenant ces composés lors de tests de choix (Raubenheimer, 1987).
Les cyanolipides sont de distribution plus restreinte et ne sont cyanogéniques que s'ils résultent de la stabilisation d'une cyanohydrine par liaison ester à un acide gras. On les trouve dans les fruits des Sapindaceae et Hippocastanaceae. Ils sont hautement toxiques pour de nombreux insectes dont les bruches des grains stockés, la pyrale du maïs *Ostrinia nubilalis* et les espèces du genre *Lepticoris* (Mikolajczak et al., 1984).

(c) Les glucosinolates

La structure générale de ces composés a été élucidée par Ettlinger et Lundeen (cités par Duncan (1991)) comme consistant en un groupe thioglucose lié à un radical par un groupe sulphonate-oxime (Figure 8). Environ 100 structures ont été décrites depuis. Synthétisés à partir d'un petit nombre d'acide aminés, ils sont présents essentiellement chez les Capparales dont principalement les Crucifères, les Resedaceae et les Capparidaceae. Leur concentration et leur profil varie selon les organes de la plante considérés. Elle est 10 fois plus forte dans les graines que dans les tissus végétatifs, avec par exemple 100 $mmol.kg^{-1}$ de matière sèche dans les graines de *Brassica napus*. Différents facteurs influent sur la concentration en ces composés comme la teneur en azote ou en sulfate du sol, la luminosité.
Ils se dégradent en D-glucose, sulfate et isothiocyanate (Figure 8). Ce dernier composé, volatil et très actif est responsable du goût acre lié aux glucosinolates. Glucosinolates et isothiocyanates sont des composés impliqués dans les relations plantes-insectes (Louda and Mole, 1991).

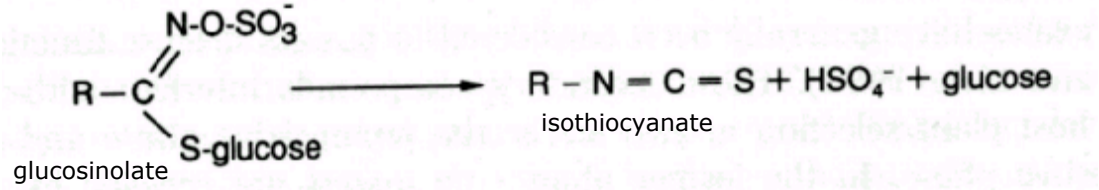


Figure 8 : hydrolyse des glucosinolates.

Ainsi, les dommages causés par des herbivores induisent une augmentation de cette concentration, afin, probablement, d'éviter de nouvelles attaques (Birch et al., 1992). Les glucosinolates affectent le développement larvaire, la métamorphose et la prise alimentaire de nombreux insectes (Wolfson, 1982). Leur action se situe à deux niveaux : ils **repoussent l'insecte** par leur odeur (phase de pré-ingestion) et exercent une activité **toxique** (phase de post-ingestion). Toutefois, certains insectes se sont spécialisés et détoxifient les produits de dégradation des glucosinolates (Duncan, 1991) voire utilisent ces composés pour localiser leur plante hôte (kairomones). Les glucosinolates peuvent alors agir comme **phagostimulant** (Chew and Renwick, 1995). Les produits de dégradation de ces composés ont également une action antifongique et antibactérienne (Duncan, 1991) indiquant un rôle possible dans la défense des plantes face aux pathogènes.

(d) Les acides aminés non protéiques

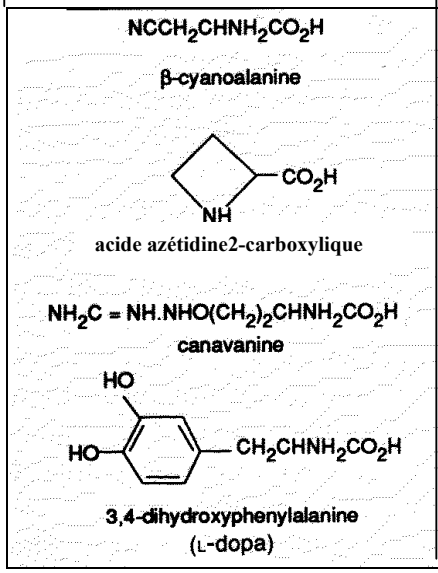
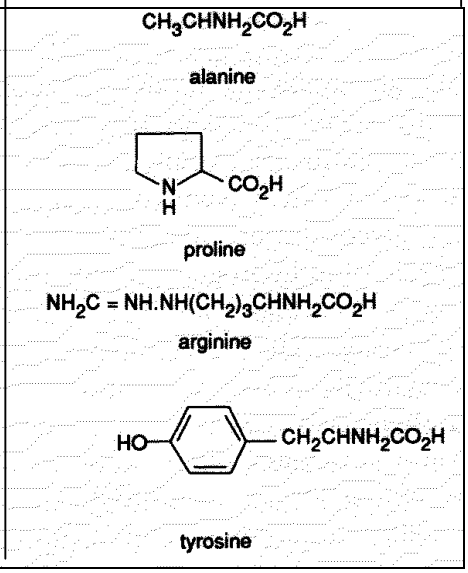
Les acides aminés non protéiques (AANP) sont présents dans un très grand nombre d'espèces végétales, particulièrement chez les Légumineuses, dans les graines et/ou les plantules. Plus de 600 en ont été découverts depuis 1948 et seulement quelques uns (ornithine, homoserine) sont communément impliqués dans le métabolisme primaire en tant qu'intermédiaires (Rosenthal, 1991).

Dans les plantes, ils se présentent majoritairement sous forme libre (contrairement à ceux des champignons qui sont souvent sous forme de petits polypeptides), extractible à l'eau ou à l'éthanol 80%. Ils peuvent être répartis dans toute la plante mais les concentrations les plus élevées sont observées dans les graines. La canavanine atteint une concentration de 127g/kg de poids sec dans les graines de *Dioclea megacarpa* et la mimosine de 145g/kg dans celles de *Leucaena leucocephala* (D'Mello, 1991 ; Rosenthal, 1982a).

La biosynthèse des AANP a fait l'objet de peu d'études mais il est clair qu'ils sont le résultat de plusieurs voies de synthèse. Certains proviennent de modifications directes d'acides aminés protéiques, d'autres de modifications des voies de synthèse de ces derniers et enfin d'autres dérivent de voies originales.

Si leur fonction biochimique fondamentale dans la plante reste encore à préciser, notamment comme réserve d'azote des graines utilisées au début de la germination (Bell, 1960 ; Rosenthal, 1982b), le rôle des AANP dans la **défense contre la prédation, la maladie,** voire la compétition vis-à-vis d'autres plantes, est bien reconnu. De nombreux AANP de plantes sont toxiques pour les vertébrés, invertébrés, plantes supérieures et microorganismes (Bell, 1980). Analogues des acides aminés protéiques (Tableau 2), ils sont parfois incorporés dans les protéines rendues ainsi non fonctionnelles ce qui peut entraîner la mort de l'individu exposé.

Tableau 2 : AANP toxiques et leurs analogues protéiques.

Acides aminés non protéiques	Acides aminés protéiques
	

Selon leur structure, on peut classer les AANP en différents groupes dont voici quelques exemples relatifs à la défense des plantes face aux herbivores.

- On peut distinguer les AANP aliphatiques neutres comme l'acide β-N-oxalyl-α,β-diaminopropionique (**ODAP**, Figure 9), présent chez *Lathyrus sativus* mais aussi chez certaines espèces des genres *Acacia* et *Crotalaria* (Rosenthal, 1982a), qui est probablement le plus puissant neurotoxique des acides aminés végétaux, étant responsable du neurolathyrisme des humains consommant cette espèce.

COOH-CO-NH-CH2-CHNH2-COOH

Figure 9 : ODAP, un AANP neurotoxique du genre *Lathyrus* (Papilionoideae).

- Certains AANP contiennent des atomes de soufre ou des groupements sulfoxides. Les Crucifères et les Alliaceae en sont de riches sources comme par exemple l'homométhionine du chou ou la S-(prop-1-enyl)-cystéine de l'ail. Les AANP soufrés sont souvent accompagnés des sulfoxides correspondants comme le S-(prop-1-enyl)-cysteine sulfoxide découvert chez l'oignon ou le S-Methyl-L-cystéine sulfoxide (SMCO, Figure 10) des *Brassica* hautement toxique pour le bétail. Chez les Papilionoideae, la plupart des AANP soufrés ont été trouvés chez les Mimosoideae et en particulier, les espèces du genre *Acacia* où 8 représentants (dont l'acide djenkolique, Figure 10) s'accumulent dans les graines et permettent, en association avec d'autres AANP de subdiviser le genre en quatre groupes biochimiquement définis (Evans et al., 1977). Dans plusieurs plantes, le sélénium peut remplacer le soufre de certains acides aminés (Figure 10). Toutefois, peu de plantes sont tolérantes à de hautes teneurs en ce composé. Certaines espèces du genre *Astragalus* accumulent le sélénium dans divers AANP non incorporés dans leur propres protéines, mais qui peuvent l'être dans l'animal les consommant, entraînant de nombreux effets toxiques (D'Mello, 1991).

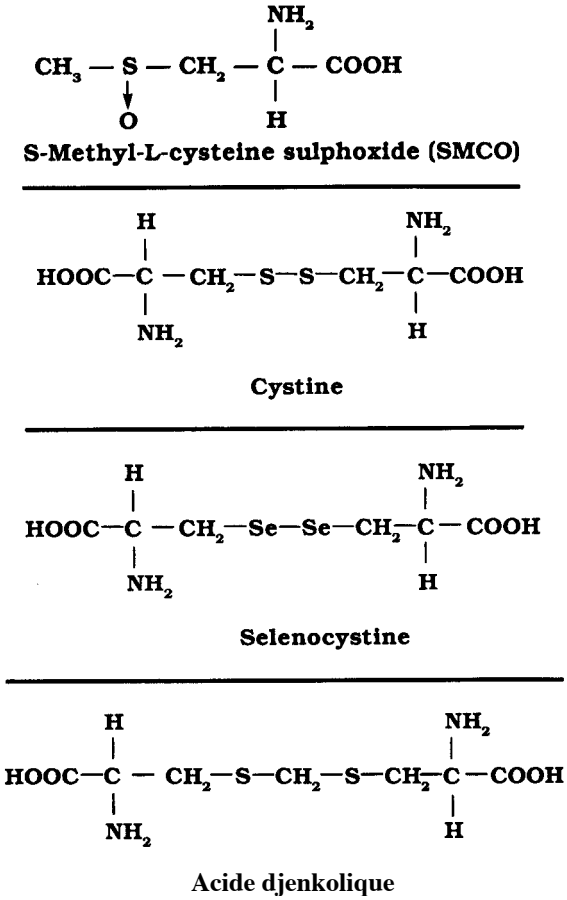


Figure 10 : AANP à soufre ou sélénium, analogues de la cystéine.

- Les AANP acides sont le plus souvent structurellement proches de l'acide glutamique comme l'acide 4-méthylglutamique de *Lathyrus maritimus* ou l'acide 4-méthylèneglutamique de l'arachide, composés que l'on retrouve chez d'autres espèces.
- Il existe également des iminoacides non protéiques comme l'acide azétidine-2-carboxylique qui s'accumule chez certaines espèces de Lilliaceae et de Légumineuses (Tableau 2).
- Parmi les AANP basiques la laminine, dérivé ammonium quaternaire de la lysine a été découvert à la fois chez les algues *Laminaria* et chez la plante supérieure *Reseda luteola*. La **canavanine** largement répandue chez les légumineuses a fait l'objet de nombreuses études. Isolée originellement de *Canavalia ensiformis* (Kitagawa and Tomiyama, 1929), elle est analogue à l'arginine (un atome d'oxygène remplaçant un groupement CH_2, Tableau 2). Présente dans de nombreuses espèces de Papilionoideae (Bell et al., 1978), elle agit par compétition avec cet acide aminé dans les réactions métaboliques (cycle de l'urée). Elle est délétère pour de nombreux lépidoptères dont *Manduca sexta*. A une concentration de 0,05% (p/v) dans l'alimentation de cet insecte, elle entraîne des aberrations de croissance chez l'adulte et une sévère diurèse chez les larves (Rosenthal, 1991).

- Il existe des AANP à hétérocycles avec un atome d'azote et/ou un atome d'oxygène. La **mimosine** (Figure 11), aux propriétés tératogène et dépilatoire, se retrouve dans des espèces des genres *Leucaena* et *Mimosa* (Renz, 1936).

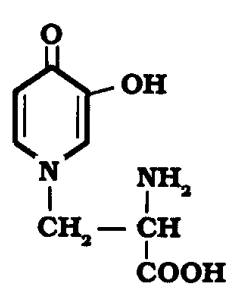


Figure 11 : la mimosine, un AANP de Mimosoideae.

- Les AANP aromatiques se rapprochent structurellement de la phénylalanine, de la tyrosine ou de l'AANP phénylglycine découvert dans le phloème de *Fagus* (Dietrichs and Funke, 1967). La 3,4-dihydroxyphénylalanine ou **L-dopa** (Tableau 2) interfère avec la tyrosinase, enzyme essentielle à la rigidification de la cuticule de l'insecte. Les graines des espèces de *Mucuna*, contenant de 5 à 10% (poids sec) de L-dopa, sont exemptes d'infestation par les bruches de graines de Légumineuses (Bell and Janzen, 1971).

Chaque AANP exerce sa toxicité par divers mécanismes. Une détoxication efficace nécessite donc plusieurs voies. C'est ce qui est observé chez la bruche tropicale *Caryedes brasiliensis* vis-à-vis de la canavanine. En plus d'avoir développé des enzymes capables de distinguer entre l'AANP et l'arginine pendant la synthèse protéique, cet insecte a abandonné la voie uricotélique normale d'élimination de l'azote utilisée chez d'autres invertébrés, en faveur de l'ammoniotélisme et de l'uréotélisme. L'uricotélisme entraîne en effet une dépendance vis-à-vis de l'arginine alimentaire et accroît donc la sensibilité à ses analogues structuraux contrairement aux voies choisies par cette bruche (Rosenthal, 1982a).

Les composés azotés variés et répandus dans le règne végétal peuvent avoir un rôle dans le métabolisme de l'azote comme dans la défense des végétaux qui les produisent. Ils sont complémentaires des composés de défense non azotés de type terpéniques et phénoliques.

2) Les terpénoïdes.

Très variés tant structuralement que du point de vue de leurs activités biologiques, les composés de ce groupe, constitués uniquement des éléments carbone, hydrogène et oxygène, comportent les huiles essentielles, résines, stéroïdes et des polymères comme le caoutchouc.

Leur chaîne carbonée est constituée d'unités isoprène (Figure 12a) à cinq atomes de carbone, assemblées d'abord en une chaîne insaturée qui est ensuite modifiée par oxydation, réduction ou élimination de carbone.

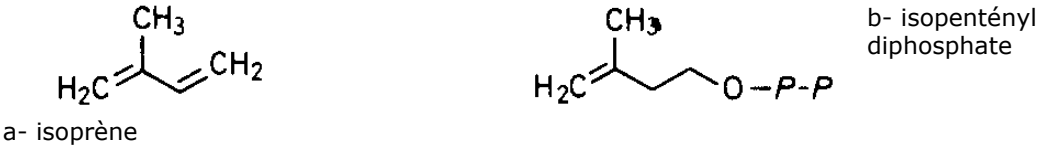


Figure 12 : l'isoprène et sa forme active l'IPD.

La forme activée de l'isoprène qui sert de précurseur à la biosynthèse de ces composés est l'isopentényl diphosphate (IPD, Figure 12b) formé à partir de l'acétyl-CoA. L'isomère prényl diphosphate de ce composé sert d'amorce à la biosynthèse. La Figure 13 indique les voies de synthèse des différents groupes de terpénoïdes à partir de ces composés.

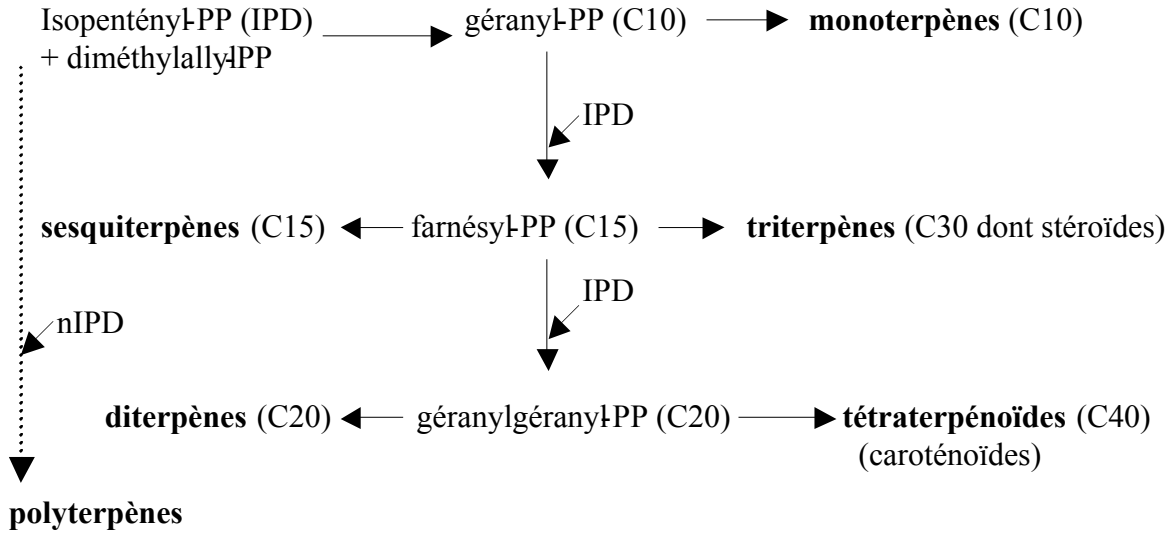


Figure 13 : formation des isoprénoïdes par une chaîne réactionnelle commune (Richter, 1993).

Du point de vue des relations plantes-insectes, ces composés peuvent avoir de nombreux rôles parfois contradictoires, étant tantôt toxiques ou répulsifs, tantôt attractifs. L'importance de chacun des groupes de terpénoïdes dans ces interactions est indiquée.

(a) Les monoterpènes

Ils constituent la majeure partie des huiles essentielles qui sont présentes en quantité appréciable chez environ 2000 espèces de 60 familles végétales. Ils sont rares chez les Légumineuses. Parfois simplement formées dans le cytosol, ces huiles se rassemblent en gouttelettes ou s'accumulent dans les vacuoles des cellules épidermiques ou du mésophylle de nombreux pétales, ou encore dans des cellules oléifères. Quand la température est élevée, ces essences traversent la paroi cellulaire et la cuticule sous forme de vapeur (parfum de fleurs). Mais souvent, des cellules glandulaires les éliminent activement dans des compartiments intercellulaires ou les rejettent vers l'extérieur du végétal.

Les monoterpènes peuvent être linéaires, monocycliques (thymol, menthol) ou bicycliques (camphre). La réaction d'un géranyl diphosphate avec un phénol donne un monoterpène

composé comme le δ9,10-tétrahydrocannabinol, responsable de l'effet psychoactif de *Cannabis sativa* (Figure 14).

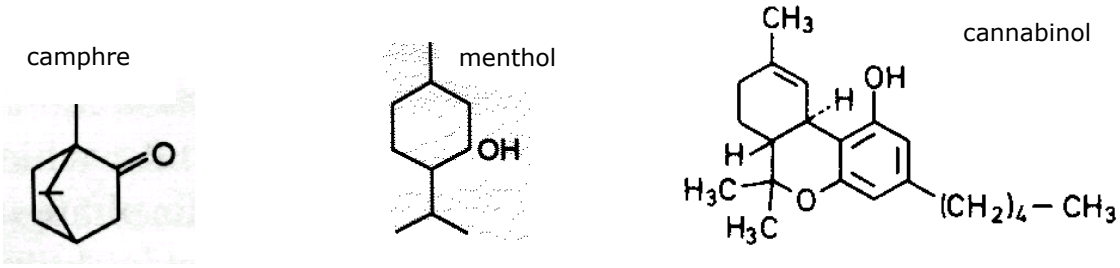


Figure 14 : exemples de monoterpènes simples et composés

Les fonctions des monoterpènes sont multiples. Certains protègent la plante des prédateurs, inhibent la multiplication de bactéries et/ou champignons. Le limonène et le géraniol, constituants de l'odeur florale, attirent les pollinisateurs. Au contraire, le citronellol inhibe l'oviposition d'*Amrasca devastans* (Homoptera : Cicadellidae) (Saxena and Basit, 1982). Un des meilleurs exemples des dérivés de monoterpènes toxiques pour les insectes sont les **pyréthrinoïdes** (Figure 15) des feuilles et fleurs de certains *Chrysanthemum* (Casida, 1973). Ces composés neurotoxiques entraînent une hyperactivité, des mouvements désordonnés voire la paralysie de l'insecte.

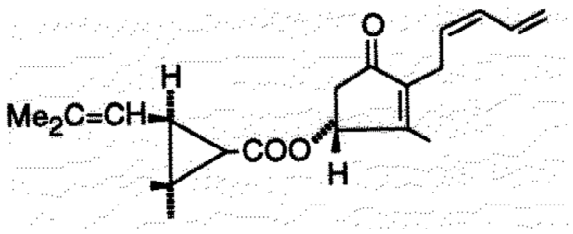


Figure 15 : la pyréthrine, un insecticide monoterpènique.

(b) Les sesquiterpènes

Les sesquiterpènes linéaires sont peu nombreux. Le farnésol (Figure 16), très répandu dans le monde végétal (parfum) en est un exemple. L'hormone juvénile (JH) principale des insectes en est un époxyde de structure très proche (Figure 16 : JHIII), et des dérivés du farnésol, observés chez plusieurs familles de plantes, montrent des effets analogues à ceux de cette hormone, perturbant le processus hormonal de l'insecte en contact avec ces produits.

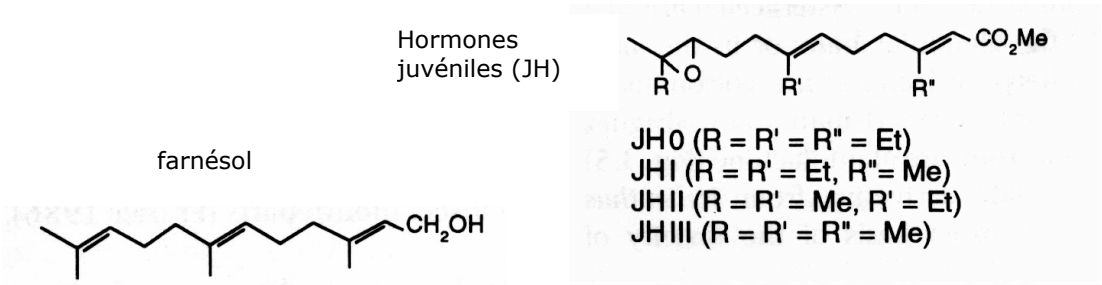


Figure 16 : Sesquiterpènes linéaires à activité d'hormone juvénile.

Plus fréquents sont les sesquiterpènes mono- ou bicycliques. L'acide abscissique, phytohormone présente entre autres dans les bourgeons et les graines en dormance, est un

exemple de sesquiterpène monocyclique. Les dialdéhydes drimanes comme le **polygodial** de *Polygonum hydropiper* (Polygonaceae) (Figure 17a) sont les sesquiterpènes **antiappétants** les plus puissants. Ils inhibent la prise alimentaire de divers insectes dont *Spodoptera littoralis, S. exempta, Leptinotarsa decemlineta*, et *Myzus persicae* (Asakawa et al., 1988). L'activité répulsive de ces terpènes envers certaines larves de Lépidoptères semble due à l'inhibition des effets stimulants du glucose, sucrose et inositol, au niveau des cellules chémoréceptrices des pièces buccales de l'insectes (Frazier, 1986). Le **gossypol**, dimère de sesquiterpène phénolique (Figure 17b) du coton et des genres proches de Malvaceae, est **toxique** pour de nombreux ravageurs dont *Spodoptera littoralis* et *Heliothis virescens*. Sa toxicité résulterait de sa complexation avec des protéines ou enzymes du tube digestif, réduisant ainsi leur digestibilité ou leur activité (Meisner et al., 1977).

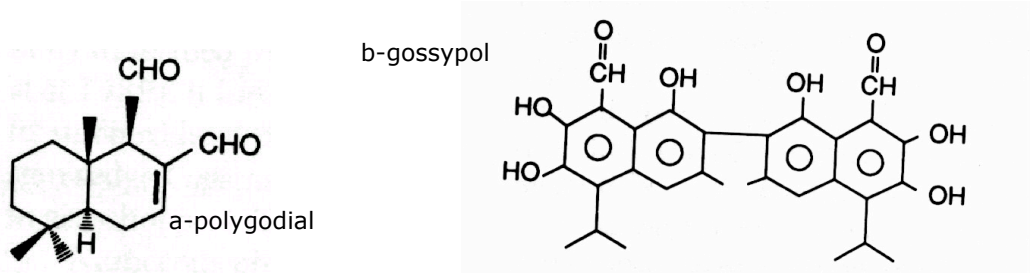


Figure 17 : Sesquiterpènes cycliques insecticides.

Certains composés à activité **anti-HJ** participent à la protection des plantes face aux insectes (Bowers et al., 1976). Les **précocènes** I et II (Figure 18), isolés de la Composée *Ageratum houstoniatum*, agissent directement sur les corps allates des insectes. En en réduisant l'activité neurosécrétrice, ils entraînent un retard de mue chez plusieurs insectes (Unnithan et al., 1978).

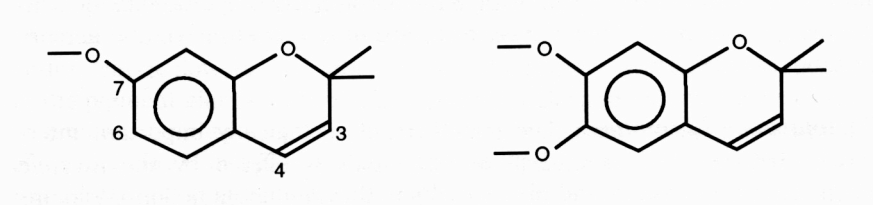


Figure 18 : Précocènes I et II, sesquiterpènes à activité antihormonale.

(c) Les diterpènes

En général non volatils et donc présents dans les résines, les diterpènes peuvent avoir de multiples fonctions.

Le phytol, élément de base de la chlorophylle, en est un représentant linéaire. Il peut également composer la chaîne latérale de la vitamine K1, une polyprénylquinone. Les gibbérellines, de squelette *ent*-gibbérellane tétracyclique (Figure 19), ont également une importante fonction biologique. Ces hormones des plantes supérieures stimulent la croissance en longueur et les divisions cellulaires de certains tissus, entre autres actions. La première gibbérélline découverte le fut par Kurosawa en 1926 chez *Fusarium heterosporum* (Richter, 1993).

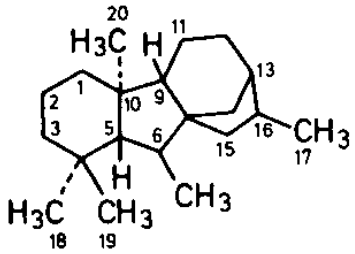


Figure 19 : structure *ent*-gibbérellane

Les diterpènes de la résine du pin sont le facteur clé de la résistance au ravageur *Dendroctonus frontalis* (Hodges et al., 1979).

(d) Les triterpènes

Leur structure de base commune est le stérane, un composé hydrocarboné alicyclique saturé (Figure 20a). Le composé clé de la synthèse des triterpènes est le squalène (Figure 20b), formé par l'assemblage queue-queue de deux farnésyl diphosphate. Leur voie de synthèse est partagée par les animaux et les végétaux.

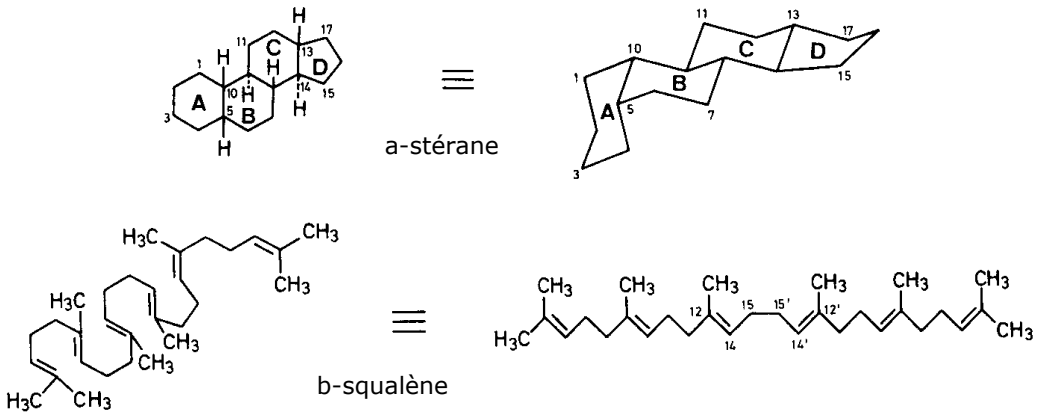


Figure 20 : structure du noyau stérane, commune aux triterpènes et premier d'entre eux, le squalène

Tetra- ou pentacycliques, ils comprennent plusieurs groupes dont un grand nombre de composés possèdent une fonction importante sur le plan biologique. Les principaux groupes impliqués dans la défense des végétaux face aux insectes sont les cucurbitacines, les limonoïdes et les saponines. Toutefois les stérols, les alcaloïdes stéroïdaux et les glucosides cardiotoniques présentent également d'intéressantes activités anti-herbivores. Des exemples de chaque type de composés sont donnés.

a) Les cucurbitacines

Ces triterpènes tetracycliques sont isolés des Cucurbitaceae. On en dénombre 20 dont le principal est la **cucurbitacine** B (CucB, Figure 21). Très amères, se sont des **antiappétants** pour de nombreux arthropodes (Da Costa and Jones, 1971). Elles sont également antagonistes des hormones stéroïdiennes des insectes, agissant au niveau du récepteur aux ecdystéroïdes (Dinan et al., 1997). Toutefois, CucB est une puissante clé de reconnaissance pour les coléoptères de la tribu des Liperini et agit comme stimulant de l'alimentation à très faible concentration (Metcalf et al., 1982).

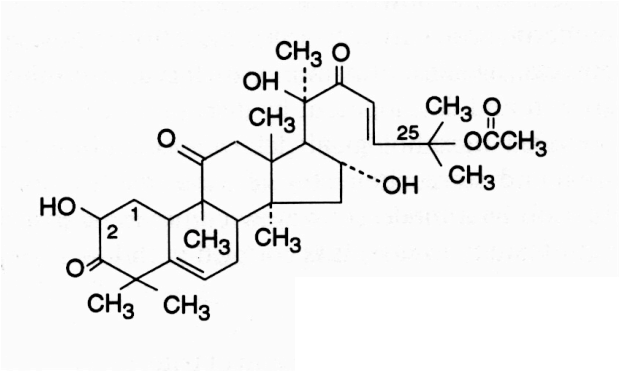


Figure 21 : Cucurbitacine B, un triterpène antiappétant.

b) Les limonoïdes

Ces triterpènes possèdent 26 carbones au lieu de 30 dans leur squelette de base. L'**azadirachtine** du neem *Azadirachta indica* (Meliaceae), très peu toxique pour les mammifères, est très utilisé comme pesticide contre plusieurs insectes, nématodes et pathogènes (Figure 22). En plus de son effet répulsif dès 0,1 ppm sur plus de 100 espèces d'insectes (Saxena, 1989), il accroît la mortalité larvaire, rompt le processus de mue et induit des diformités (Schmutterer, 1990).

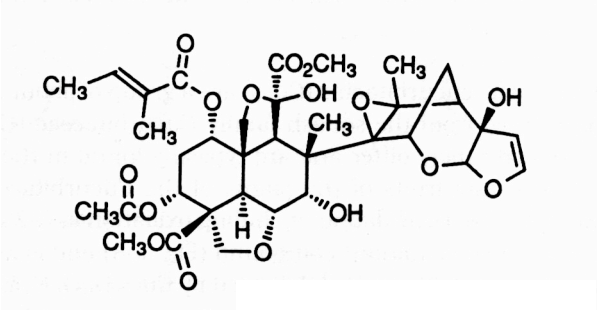


Figure 22 : Azadirachtine, un puissant insecticide triterpène.

c) Les saponines

Ces composés tensioactifs très répandus chez les végétaux, sont des glucosides terpéniques ou stéroïdaux. L'aglycone, nommé sapogénine est un triterpène ou un stérol, définissant ainsi deux groupes de saponines. Les chaînes glucidiques de taille et séquence variées, rattachées par des liaisons glycosidiques ou esters, sont composées à partir d'un petit nombre de sucres (jusqu'à 10).

Les saponines **triterpènes** sont les plus nombreuses (environ 120). Elles sont essentiellement présentes chez les Dicotylédones, la structure de leur aglycone en C30 étant le plus souvent pentacyclique comme la β-amyrine (Figure 23b), rarement tétracyclique.

Les saponines **stéroïdes** sont surtout représentées chez les Monocotylédones. La structure de leur aglycone en C27 dérive uniquement de celle du stérane. La diosgénine en est un exemple (Figure 23a). Elles ont une importance économique comme source de composés stéroïdaux, leur aglycone étant libéré de son oligoholoside pour synthétiser des stéroïdes humains.

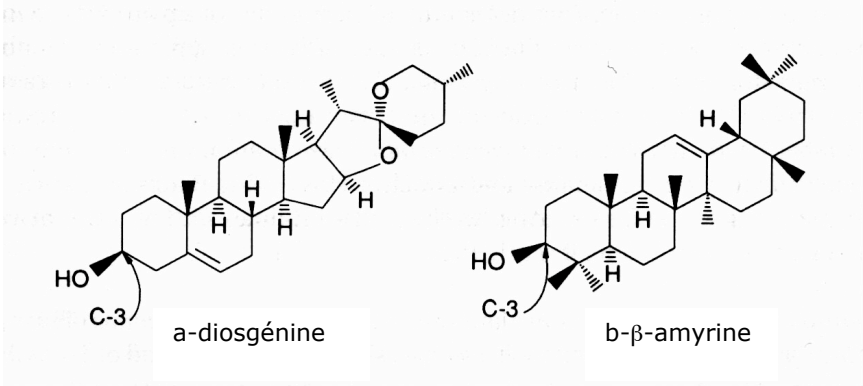


Figure 23 : exemples de saponines.

Les saponines s'avèrent **toxiques et antiappétantes** pour de nombreuses espèces d'insectes (Ishaaya et al., 1969). Leur toxicité serait due à leur liaison avec les stérols libres du tube digestif, réduisant ainsi le taux d'assimilation de ces stérols dans l'hémolymphe ce qui interfére avec le processus de mue. Elles peuvent également interférer avec des fonctions trophiques plus complexes comme les ectosymbioses de fourmis champignonnistes (Rahbé et al., 1988).

d) Les stérols

Chez les végétaux, le premier composé stable après cyclisation du squalène est le cycloarténol, précurseur des phytostérols qui constituent avec divers lipides, les membranes des cellules végétales. Tous les stérols possèdent un groupement hydroxyle caractéristique en C3.

Le cholestérol, précurseur de nombreux stéroïdes animaux, existe à l'état de traces chez les végétaux. Il constitue la structure de base de saponines, glucosides cardiotoniques et d'alcaloïdes stéroïdes. L'ecdysone (Figure 24), précurseur de l'hormone de mue des insectes possède une structure très proche du cholestérol. Sa synthèse dépend des stérols végétaux absorbés par l'insecte phytophage. Environ 200 composés parents de l'ecdysone, dont l'ecdystérone, hormone active de mue et métamorphose, existent dans plus de 100 familles végétales (Dinan, 2001). Les **phytoecdysones inhibent le développement**, la croissance et la reproduction de plusieurs espèces d'insectes (Singh et al., 1982). Leur production peut être induite suite aux dommages causés par un insecte, comme c'est le cas dans les racines d'épinard (Schmelz et al., 1999). Elles ont un effet antiappétant sur certaines espèces dont les récepteurs sensoriels répondent à la présence de 20-hydroxyecdysone (Descoins and Marion Poll, 1999), le plus répandu de ces composés. Toutefois, certains insectes polyphages sont insensibles aux phytoecdysones (Dinan, 1998). L'utilisation de ces composés dans la lutte contre certains ravageurs apparaît possible, vue leur absence de toxicité pour les mammifères, ainsi que la probabilité que de nombreuses plantes ne produisant pas ces composés aient toutefois la capacité génétique de le faire (très large répartition systématique de ces composés).

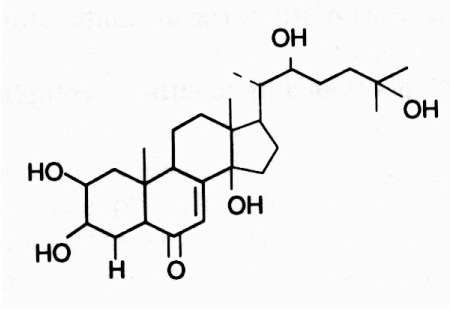


Figure 24 : l'ecdysone, un phytostérol à effets anti-insectes.

e) Les alcaloïdes stéroïdaux

Leur mode de biosynthèse les désigne comme des isoprénoïdes, l'élément azote n'apparaissant que dans les chaînes latérales du squelette carboné, constitué uniquement d'unités en C5. La plupart sont sous forme de glycosides. Leur aglycone possède soit 21 carbones, soit 27. Parmi les composés en C27, deux groupes sont distingués, ceux principalement produits par les Liliaceae et ceux produits principalement par les Solanaceae. Parmi ces derniers, la tomatine, la solanine et la démissine de *Solanum demissum* protègent la plante de la voracité du doryphore (Richter, 1993).

f) Les glucosides cardiotoniques

De structure très proche des saponines qu'ils accompagnent dans de nombreuses espèces de quinze familles végétales, nombre d'entre eux possèdent une action sur le muscle cardiaque et sont utilisés comme médicament, en particulier, ceux issus des genres *Digitalis* et *Strophanthus*.
On distingue deux groupes parmi ces glycostéroïdes, selon leur aglycone. Les cardénolides en C23 possèdent un cycle lactone à 5 chaînons et une liaison insaturée tandis que les bufadiénolides en C24 présentent un cycle lactone à 6 chaînons et deux liaisons insaturées. C'est le cycle lactone qui rend ces composés cardiotoniques.
Comme pour les saponines, plusieurs résidus glucidiques peuvent se trouver dans la molécule en chaîne linéaire terminée par un hexose, le plus souvent le glucose.

(e) Les tétraterpènes

Les caroténoïdes et les acides caroténoïdiques (moins abondants) sont des composés biologiquement importants présents dans les règnes animal et végétal mais uniquement produits par les végétaux, champignons et bactéries. On distingue les caroténoïdes primaires participant à la photosynthèse des secondaires, présents dans les chromoplastes des fleurs et fruits ainsi que chez les bactéries, levures et champignons.
La charpente en C40 se forme par l'assemblage en configuration *cis* de 2 géranylgéranyls diphosphate avec formation d'une nouvelle double liaison centrale. Le produit clé de la biosynthèse des caroténoïdes alors obtenu, dans les plantes supérieures et les champignons, est le 15-cis-phytoène (Figure 25a). Des déshydrogénations successives augmentent le nombre général de doubles liaisons et donc le système de conjugaison ce qui entraîne une intensification de la couleur des composés produits. Le 15-cis-phytoène possédant 3 liaisons conjuguées est incolore alors que le lycopène, produit final des étapes de déshydrogénation, est un caroténoïde caractéristique d'un vif rouge orangé.

Quelques caroténoïdes sont caractérisés par une cyclisation limitée aux extrémités de leur chaîne. Une double β-cyclisation du lycopène conduit au β-carotène (Figure 25b) qui, clivé symétriquement, produit deux molécules de vitamine A.

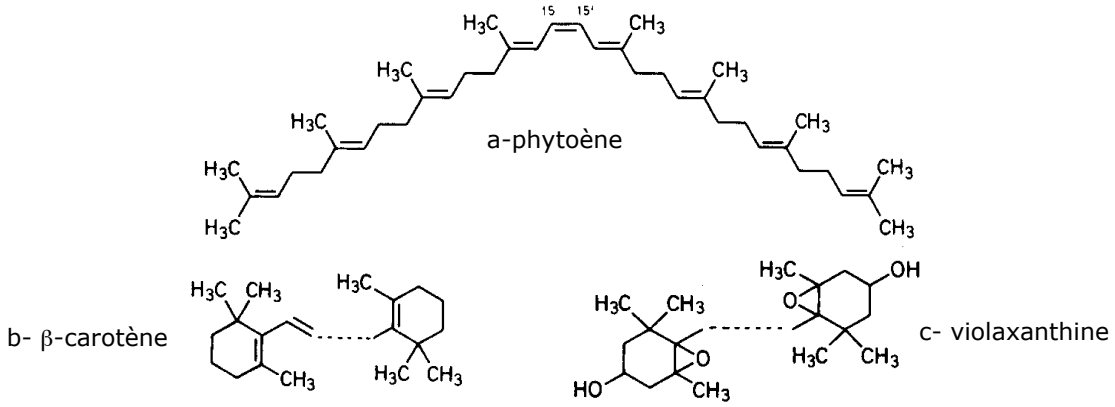


Figure 25 : exemples de caroténoïdes et dérivés.

Les xanthophylles comme la violaxanthine (Figure 25c), formés à partir des caroténoïdes contiennent des atomes d'oxygène.
Lors de la dégradation oxydative des caroténoïdes, la chaîne peut se raccourcir à une de ses extrémités. C'est ainsi que se forme l'acide abscissique à partir de la violaxanthine. La dégradation aux deux extrémités de la chaîne conduit aux acides caroténoïdes comme la crocétine, aglycone de la crocine, substance jaune vif des étamines du safran *Crocus sativus*.
Dans les relations plantes-insectes, ces composés se limitent à attirer les pollinisateurs ou disperseurs de graines, mais ne présentent pas de toxicité.

(f) Les polyterpènes

Le caoutchouc, formé de 500 à plus de 5000 unités isoprènes en chaînes non ramifiées aux doubles liaisons de configuration *cis*, est présent dans environ 2000 espèces végétales dont un quart seulement en possèdent abondamment. C'est le cas d'*Hevea brasiliensis* (Euphorbiaceae) exploitée pour son latex. Les latex sont riches en différents composés actifs pouvant avoir un rôle de défense, en réduisant la prise alimentaire et la fitness du phytophage (Malcom and Zalucki, 1996). De plus, la coagulation du latex peut emprisonner l'animal (Dussourd and Eisner, 1987).

3) Les composés phénoliques.

Les composés phénoliques correspondent à une grande variété de substances possédant un cycle aromatique portant au moins un groupement hydroxyle. Parmi les milliers décrits, les plus nombreux sont les flavonoïdes. Leurs activités biologiques très diversifiées jouent un rôle très important dans les relations plantes-insectes, tantôt attirants, tantôt répulsifs voire toxiques pour ces animaux.

La grande majorité des composés phénoliques dérivent de l'acide cinnamique formé par la voie du shikimate (du nom japonais de l'anis étoilé où il a été découvert) (Figure 26). La

voie de biosynthèse de composés aromatiques à partir du cinnamate est spécifique des végétaux.

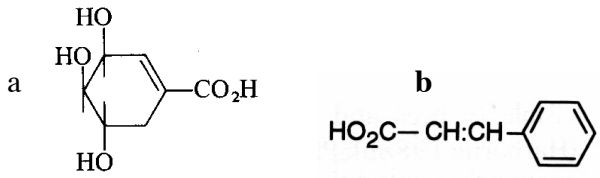


Figure 26 : acide shikimique (a) et acide cinnamique (b).

Souvent acides, les polyphénols sont solubles dans une solution de carbonate de sodium. Chimiquement, ils sont réactifs et peuvent participer à des liaisons hydrogènes, ou chélater des métaux pour les o-dihydroxyphénols (catéchol). Enfin, ils sont sensibles à l'oxydation. Un faible nombre de ces composés (cyanidine, quercetine, acide caféique) sont répandus mais la plupart n'ont qu'une faible occurrence. Leur rôle physiologique est encore incertain. Beaucoup de ces composés ont un effet sur la **croissance** lorsqu'ils sont appliqués sur des tissus à une concentration physiologique mais ils sont tellement variés structuralement qu'il paraît improbable que tous possèdent un rôle universel dans la régulation de la croissance et du développement. Toutefois, certains phénols peuvent avoir de telles propriétés, comme l'acide lunularique, un acide dihydro-stilbène-carboxylique qui aurait le rôle d'hormone de type dormine (= acide abscissique) chez l'hépatique *Lunularia cruciata* (Gorham, 1977). De plus l'équilibre entre l'acide p-coumarique et l'acide caféique pourrait réguler la synthèse d'éthylène.
S'il est difficile d'attribuer un rôle physiologique à la plupart des composés phénoliques, de très nombreuses données témoignent de leur **rôle écologique**. Les pigments flavonoïdes (en association avec les caroténoïdes) contribuent à l'attraction des animaux vis-à-vis des fruits et des fleurs colorés à disperser ou polliniser. De plus, ils sont parfois excrétés de la plante et peuvent affecter la croissance des autres plantes dans l'environnement immédiat (phénomène d'allélopathie) (Rice, 1974). Certains flavonoïdes et surtout les tanins, astringeants, sont des antiappétants protégeant les plantes de la prédation de nombreux animaux (Harborne, 1977 ; Swain, 1977). Ils peuvent également être toxiques pour les insectes : l'acide chlorogénique, très courant, en association avec le flavonoïde rutine des trichomes des feuilles de la tomate est toxique pour *Heliothis zea* (Isman and Duffey, 1982). Enfin certaines classes de phénols agissent comme agents antimicrobien pour divers bactéries, virus et champignons. Certaines phytoalexines phénoliques sont en effet synthétisées suite à une infection fongique chez au moins six familles végétales (Harborne and Ingham, 1978) et d'autres en réaction à une exposition aux rayons ultraviolets (anthocyanes).

On peut distinguer différents types de composés phénoliques :

(a) Les composés phénoliques simples et les quinones, les acides phénoliques et les tanins

Les composés simples comme le catéchol ou le phloroglucinol sont relativement peu communs contrairement aux acides phénoliques universellement répandus chez les végétaux. Ils se caractérisent par la perte au moins partielle des chaînes latérales caractéristiques (par rapport à l'acide cinnamique en C6C3, Figure 26b). La vanilline, aldéhyde en C6-C1 est très répandue chez les angiospermes, gymnospermes et fougères. Il s'agit du principe odorant de *Vanilla planifolia*. L'hydroquinone (C6) est la quinone

simple la plus courante (surtout chez les Rosaceae et les Ericaceae), l'acide gallique (C6C1) est largement répandu, surtout comme constituant des tanins (Figure 27).

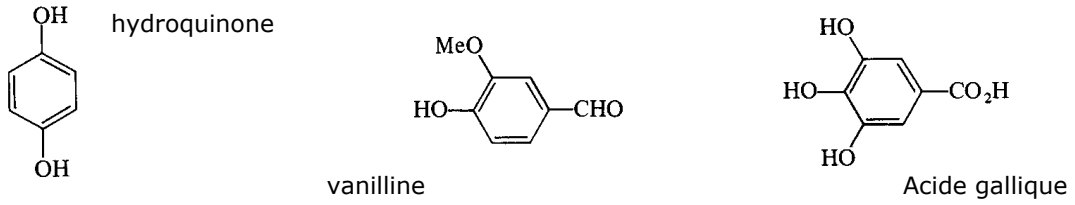


Figure 27 : exemples de phénols simples.

Les **tanins** et les composés simples ont une action **protectrice** contre les attaques de microorganismes en se déposant dans les parois cellulaires. De plus les tanins ayant la propriété de précipiter les protéines réduiraient la valeur nutritionnelle de certains tissus végétaux. Ils affectent également la digestion des insectes en se liant aux mucoprotéines de leur cavité orale (Harborne, 1988), mais la moindre croissance des insectes causée par les tanins serait essentiellement due à une inhibition post-absorption plutôt qu'à une inhibition de la digestion (Butler, 1989). Ils sont classifiés en tanins **hydrolysables** (Figure 28b), les gallotanins (acide gallique et glucose) et ellagitanins (acide hexahydroxydiphénique) présents uniquement chez les Dicotylédones et en tanins **condensés** (non-hydrolysables, Figure 28a), plus répandus dans le règne végétal. Un troisième groupe de tanins, les phlorotanins, a été caractérisé chez diverses algues brunes. Le rôle écologique des tanins condensés est bien illustré par les travaux de Coley (Griffiths, 1991) : les pousses de *Cecropia pelata* L. à fort taux de tanins ont un rendement de production des feuilles inférieur de 30% par rapport aux pousses à faible taux de tanins mais elles sont bien moins sensibles aux dommages d'herbivores. L'aire de feuille, consommée par les larves généralistes de *Spodoptera latifascia*, est cinq fois plus grande sur les plantes à faible teneur en tanins que sur les plantes plus concentrées.

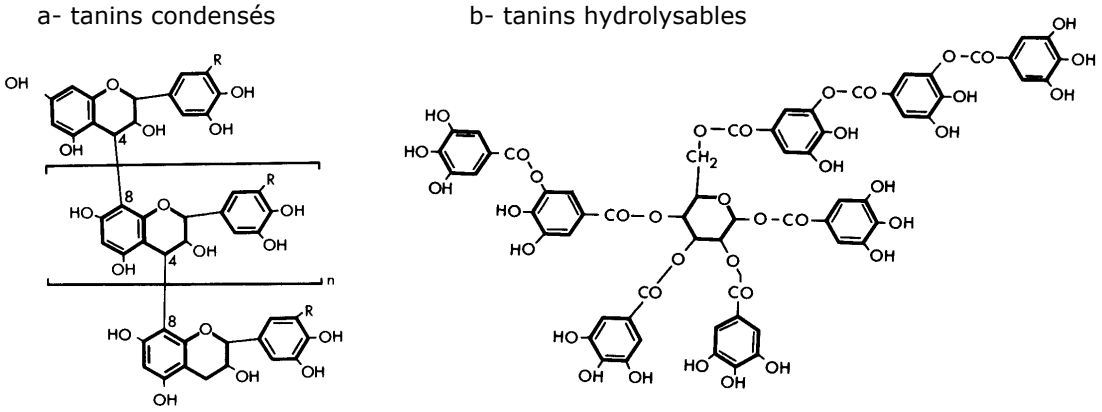


Figure 28 : structures générales des deux types de tanins végétaux.

Les **quinones** sont le plus souvent des hydroquinones aux propriétés phénoliques. Elles peuvent être combinées à des sucres (anthroquinones). Largement mais inégalement réparties chez les végétaux supérieurs, elles proviennent de nombreux précurseurs de différentes voies de biosynthèse. Le potentiel redox du couple quinone-hydroquinone est très important dans de nombreux systèmes biologiques. Les polyprénylquinones (quinones terpénoïdes) dérivent de l'acide 4-hydroxybenzoïque. Leur noyau benzoquinone portent

des chaînes poly-isoprénique. Certains comme la plastoquinone et l'ubiquinone sont des transporteurs d'électrons. En plus de fournir une source de radicaux libres stables, les quinones sont connues pour complexer irréversiblement avec certains acides aminés de protéines qu'elles inactivent ou rendent inutilisables. Ceci explique la variété d'effets antimicrobiens potentiels des quinones. Nombres d'entre elles sont **antiappétantes** et **toxiques**. La naphtoquinone juglone de l'écorce de *Carya ovata* inhibe la croissance de différents champignons et bactéries phytopathogènes et est répulsive pour *Scolytus multistriatus* (Gilbert and Norris, 1968).

(b) Les acides phénylpropanoïques (C6C3), coumarines et lignines.

Formés à partir de l'acide cinnamique ou p-coumarique (Figure 26), ils sont à l'origine des voies de synthèse de nombreuses substances telles que les lignines, les flavonoïdes, les stilbènes.

Les **lignines** des parois cellulaires secondaires leur confèrent une haute résistance à la traction et la pression grâce à leur élasticité. Il s'agit de mélanges de polymères amorphes de 3 constituants fondamentaux, les alcools p-coumarylique (qui domine dans les lignines de Gramineae), coniférylique (dominant dans le bois des Conifères) et sinapylique (bois des feuillus). Le rôle des lignines dans l'évolution des végétaux a été analysé par Raven (1977). Formant une **barrière** mécanique, de goût désagréable et réduisant la digestibilité des sucres de la paroi, les lignines participent à la résistance des plantes aux microorganismes et herbivores. La lignification est une réponse courante à l'infection ou la blessure.

La **coumarine** (Figure 29) et ses dérivés dont plus de 300 structures sont connues, se répartissent dans 9 familles de Monocotylédones et plus de 70 familles de Dicotylédones. Ils participent dans les racines des plantes symbiotiques hébergeant Rhizobium, à la formation des nodules. Elles sont responsables de l'odeur caractéristique de l'aspérule odorante et du mélilot desséchés. Elles possèdent un cycle lactone (cyclisation entre les groupes o-hydroxy et carboxyle). L'ombelliférone, l'aesculetine et la scopoletine sont les trois hydroxycoumarines les plus communes. Les **furanocoumarines**, de structure plus complexe sont beaucoup plus variées mais présentes dans un nombre limité de familles végétales. Elles dérivent principalement de l'ombelliférone par condensation avec un isoprénoïde en C5, et sont souvent liposolubles (Murray et al., 1982). Le cycle furane peut être fusionné au cycle benzénique dans deux positions : linéaire ou angulaire, cette dernière étant plus rare (Ombéllifères et genre *Psoralea* des Légumineuses). Elles présentent un intérêt physiologique, inhibant la croissance des graines en germination. Les coumarines, en présence d'UV, se lient aux bases pyrimidiques de l'ADN (Berenbaum, 1983). Elles sont donc **phototoxiques** pour un large panel d'organismes dont des bactéries, virus, champignons, invertébrés et vertébrés. La simple coumarine est ovicide pour les doryphores et toxique pour une variété d'insectes. La furocoumarine xanthotoxine incorporée à 0,1% dans l'aliment et irradiée d'UV entraîne 100% de mortalité chez *Prodenia eridania* (Berenbaum, 1991).

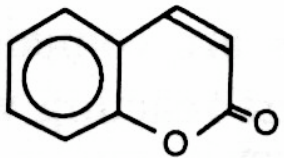


Figure 29 : coumarine.

(c) Les flavonoïdes

Ce groupe, le plus varié, contenant 4000 structures, comprend comme son nom l'indique[2] des composés jaunes mais aussi d'autres couleurs ou incolores. Ils sont universels chez les angiospermes et les gymnospermes. Ils ont tous en commun la structure de la flavane, 3 cycles dont un hétérocycle dont la configuration variée permet la classification en sous-groupes (Figure 30) : flavones, flavanones et flavonols, flavanols, flavanediols et la chalcone dont l'hétérocycle n'est pas formé et qui est un intermédiaire caractéristique de la synthèse des diverses flavonoïdes. Le noyau flavonoïde est souvent lié à un sucre pour former un glycoside hydrosoluble.

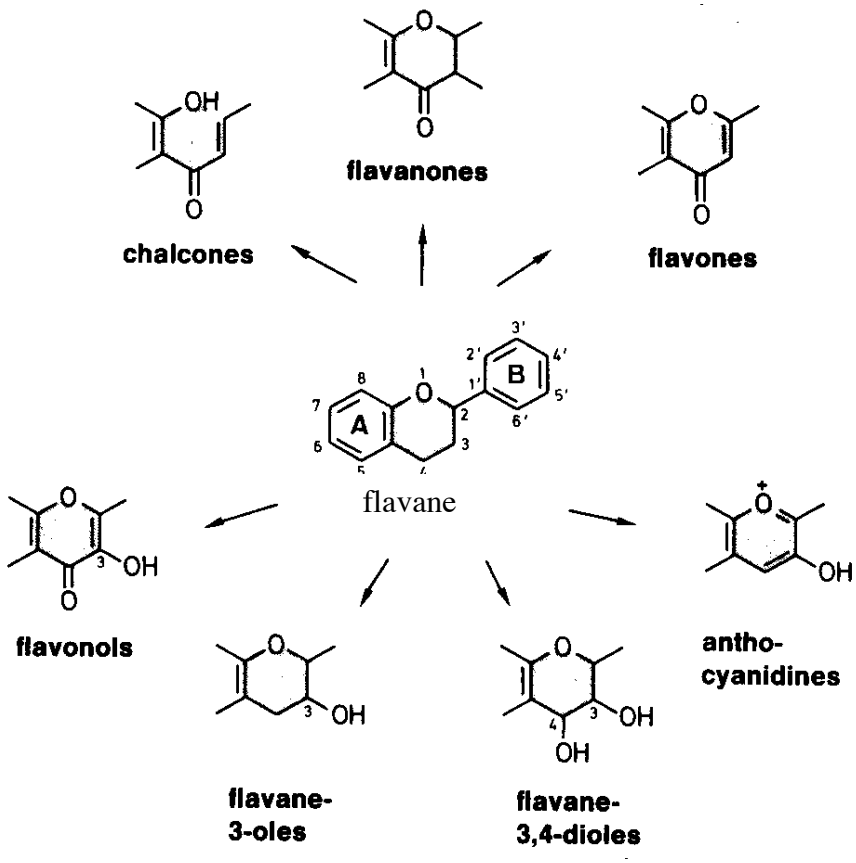


Figure 30 : définition des différents types de flavonoïdes à partir du squelette flavane.

Chaque espèce possède son propre profil de flavonoïdes, stockés principalement dans les vacuoles, et qui absorbent dans les UV et/ou le visible et sont responsables de la couleur

[2]du latin flavus = jaune

jaune ou crème. Ils sont de bons **antiappétants voire toxiques** pour les insectes : à faible concentration, la **quercétine** (Figure 31) et trois de ses dérivés glycosylés inhibent l'alimentation d'*Heliothis zea*, *H. virescens* et *Pectinophora gossypol*. A partir de concentrations de 0,2% ces composés tuent les larves de ces espèces (Shaver and Lukefahr, 1969).

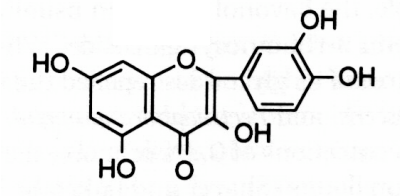


Figure 31 : la quércétine, un flavonol insecticide.

Les **isoflavonoïdes** sont moins répandus taxonomiquement. Ces composés très actifs se trouvent essentiellement chez les **Légumineuses**. Ils sont responsables de la moindre prise alimentaire et de la toxicité envers *Trichoplusia ni* (Lepidoptera : Noctuidae) s'alimentant sur soja (Neupane and Norris, 1990). Certains isoflavonoïdes fonctionnent comme des **phytoalexines**, synthétisées comme défense contre le stress (microorganisme infectieux, froid, UV). La glycéolline (Figure 32b), la pisatine et la phaséolline produites par le soja, le pois et le haricot commun respectivement en sont de bons exemples (Perrin and Cruickshank, 1965). Cette dernière est parmi les antiappétants les plus efficaces jamais déterminés (Lane et al., 1985). Le seul groupe de flavonoïdes connus pour être hautement toxique envers de nombreux insectes sont les isoflavonoïdes **roténoïdes**. Extrait des racines de *Derris elliptica* (Légumineuse), le roténone en est le principe actif (Figure 32a). Ils sont peu toxiques envers les mammifères mais très **toxiques** pour les poissons et les insectes. Leur toxicité est liée à l'inhibition de l'oxydation mitochondriale (transport d'électrons) (Panda and Khush, 1995).

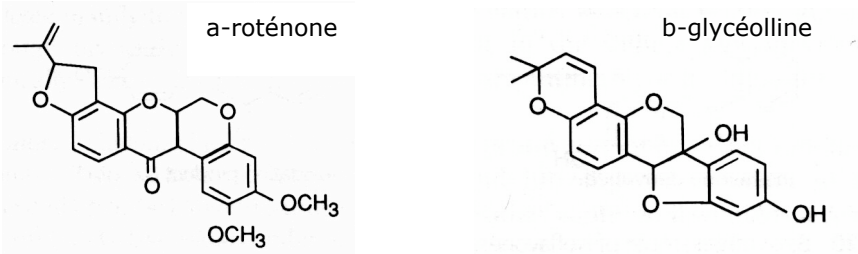


Figure 32 : isoflavonoïdes de défense des plantes.

Les **anthocyanides** sont les aglycones des anthocyanes, pigments vacuolaires rouges ou bleus des végétaux. Plus le nombre de substitutions du cycle B des aglycones est grand, plus la coloration bleue est intense (delphinidine, Figure 33). La méthylation des groupements hydroxyles conduit au contraire au rouge (malvidine, Figure 33). Pour les anthocyanes, d'autres facteurs interviennent pour déterminer les tons, en particulier la formation de chélates avec des ions métalliques et la formation simultanée de complexes de haut poids moléculaires, les chromosaccharides. Ces composés colorés facilitent la **pollinisation** en attirant les abeilles (par absorption des UV) et la dispersion des fruits par les insectes et les animaux.

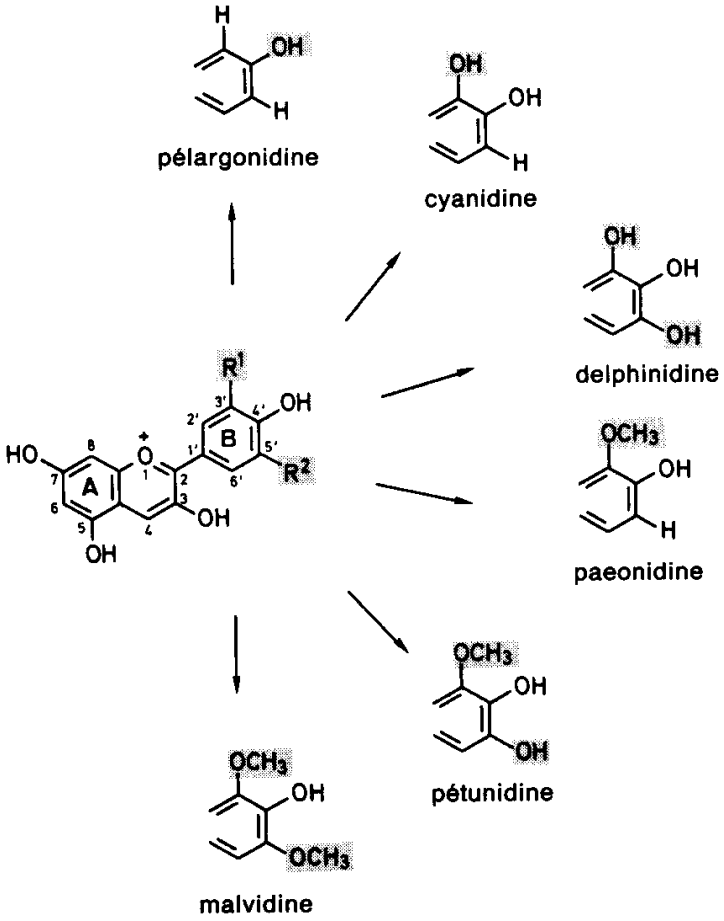


Figure 33 : exemples d'anthocyanidines.

Les métabolites secondaires de défense sont d'efficaces systèmes de résistance aux herbivores. Toutefois, ils résultent de voies de biosynthèse impliquant l'interaction des produits de nombreux gènes. Leur utilisation est donc limitée par les possibilités actuelles d'ingénierie génétique des végétaux, pour lesquelles ces voies apparaissent très souvent trop complexes, malgré quelques travaux qui tentent de les exploiter (Dawson et al., 1989 ; Hallahan et al., 1992). Certains mécanismes de défense des plantes reposent directement sur des protéines, produits d'un seul gène. Ils sont donc particulièrement intéressants pour la création de plantes résistantes par les techniques actuelles de transfert de gènes.

B. Les protéines végétales de défense

Les protéines de défense des plantes sont tout aussi diverses et efficaces que les métabolites secondaires. Nous nous limiterons ici aux principales protéines anti-insectes des végétaux, puis aux principaux peptides (petites protéines de moins de 100 acides aminés), riches en cystéines, et impliqués dans la défense des plantes.

1) Les protéines entomotoxiques

Le Tableau 3 présente différentes protéines végétales insecticides, et indique un ordre de grandeur de leur toxicité pour divers ordres d'insectes et pour les mammifères. Les toxines de *B. thuringiensis* y sont également mentionnées.

La famille des lectines et protéines apparentées, celles des inhibiteurs d'enzymes (protéases et amylases), les canatoxines et les chitinases sont présentées de façon plus détaillée dans cette partie.

Tableau 3 : Toxicité envers les mammifères et les différents ordres d'insectes des protéines de plante à activité insecticide et des endotoxines de *Bacillus thuringiensis* (*Bt*).

	Ordres d'insectes							**Mammifères**
Protéines	*Coléoptères*	*Lépidoptères*	*Homoptères*	*Hémiptères*	*Orthoptères*	*Diptères*	*Hyménoptères*	
Lectines								
Man	Toxique	Toxique	Toxique	nd	nd	nd	Toxique	Toxique?
GlcNAc	Toxique	Toxique	Toxique	nd	nd	Toxique	nd	Toxique
GlcNAc/Gal	Toxique	Toxique	Toxique	nd	nd	Toxique	nd	Toxique
RIP	Toxique	Pas d'effet	nd	nd	Pas d'effet	Pas d'effet	nd	Toxique
Arcélines	Toxique	nd	nd	nd	nd	nd	nd	nd
IP à serine	Toxique	Toxique	Pas d'effet	Pas d'effet	Toxique	nd	Toxique	Toxique
Cystatines	Toxique	Pas d'effet	Pas d'effet	Toxique	nd	nd	nd	Pas d'effet
IA	Toxique	Toxique	Pas d'effet	nd	nd	nd	nd	Toxique
Canatoxine/ uréase	Toxique	Pas d'effet	nd	Toxique	Pas d'effet	Pas d'effet	nd	Pas d'effet
Bt toxins	Toxique (Cry3)	Toxique (Cry 1 et 2)	Toxique?	Pas d'effet	nd	Toxique (Cry2 et 4)	Toxique	Toxique?

Toxique : observations lors d'expériences d'alimentation sur milieu artificiel ou plante transgénique ; Toxique ? : existence de données conflictuelles quant à la toxicité ou non de ces protéines. ; nd : absence de données disponibles. D'après Carlini and Grossi de Sa (2002). RIP: ribosome inactivating protein, IP: inhibiteur de protéase, IA: inhibiteur d'α-amylase, Man = mannose, GlcNAc = N-acétyl-D-glucosamine, Gal = galactose.

(a) Lectines, hémilectines et autres protéines apparentées.

a) Les lectines vraies.

Les lectines ou **hémagglutinines** sont une classe de protéines definies comme d'origine non immune possédant au moins un domaine non catalytique qui leur permet de fixer réversiblement et spécifiquement un mono- ou un oligo-saccharide (Peumans and Van Damme, 1995). Une lectine typique est multivalente et ainsi capable d'agglutiner les cellules. La plupart sont des lipo- ou glyco-protéines, mais certaines sont uniquement protéiques comme la concanavaline A (ConA) isolée de *Canavalia ensiformis* (Bell, 1987). Très largement répandues parmi les végétaux, virus, bactéries, invertébrés et vertébrés, elles présentent toutefois peu de similarités entre ces différentes origines.

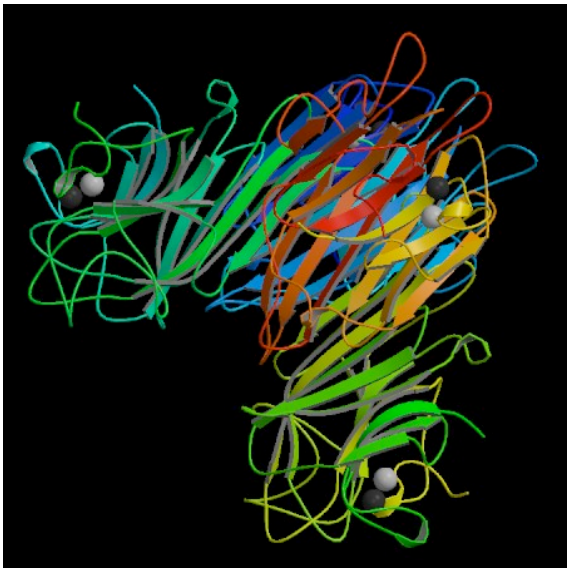


Figure 34 : structure tridimensionnelle de la concanavaline de *Canavalia brasiliensis* (Sanz-Aparicio et al., 1997).

Les **graines** d'**Euphorbiaceae** et **Leguminoseae** sont de riches sources de phytohémagglutinines (Toms and Western, 1971). Plus de 230 structures tridimensionnelles en sont disponibles dans la banque de données 3D des lectines, dont 40% proviennent de légumineuses (www.cermav.cnrs.fr/databank/lectine). Une homologie certaine existe entre lectines de différentes Légumineuses malgré leur différences de spécificité dans la liaison aux sucres. Les phytohémagglutinines PHA du haricot sont, avec la phaséoline (une viciline-like 7S globuline) les protéines de réserve majeures des graines de cette plante. Les PHA lient les glycanes de la muqueuse intestinale des mammifères et agissent comme mitogène. Avec les arcélines (p.56) et les αAI (p.64) elles forment la famille des « lectines » du haricot.

Le rôle biologique des lectines est encore sujet de controverse. Une action dans la nodulation des légumineuses a été proposé (Hirsh, 1999) ainsi que dans les mécanismes de défense des plantes (Chrispeels and Raikhel, 1991 ; Peumans and Van Damme, 1995). Certaines lectines sont très toxiques pour l'homme et le bétail. Des études menées sur souris, cochon, poulet et autres oiseaux ont montré la capacité de certaines phytohémagglutinines à se fixer sur la muqueuse intestinale, empêchant le développement normal des microvillosités. Il s'en suit des difficultés d'absorption des nutriments qui se

traduisent par une perte de poids, un retard de croissance et de développement (Grant, 1991).
De nombreuses lectines possèdent une activité **insecticide** lorsque fournies dans l'alimentation de divers Coléoptères, Homoptères et Lépidoptères. Déjà en 1976, Janzen *et al* montrèrent que les lectines des graines du haricot commun *Phaseolus vulgaris* L. présentent une activité entomotoxique pour la bruche du niébé *Callosobruchus maculatus* F. Les lectines du blé, du ricin et de *Bauhinia purpurea* ont une DL50 de 0,29 ; 0,59 et 0,75 mg/g respectivement lorsqu'elles sont incorporées à l'aliment fourni aux larves d'*Ostrinia nubilalis* Hübner (Czapla and Lang, 1990). Les lectines peuvent se lier à la bordure en brosse des cellules épithéliales de l'intestin de l'insecte ou, pour celles se fixant à la chitine, à la membrane péritrophique. Elles peuvent aussi se lier à des enzymes digestives glycosylées. Toutefois le mode d'action précis de ces protéines chez l'insecte reste à déterminer. Des lectines de spécificités similaires semblent avoir des effets très différents sur des insectes apparentés comme le montrent des réponses contrastées de différents pucerons à diverses lectines (Gatehouse et al., 1995 ; Rahbé et al., 1995).
Des gènes codant pour des lectines entomotoxiques ont été introduits dans différentes plantes cultivées rendues ainsi moins sensibles aux attaques des insectes (pour revue (Hilder and Boulter, 1999 ; Jouanin et al., 1998 ; Schuler et al., 1998) ; Tableau 4). L'activité insecticide envers les pucerons des lectines à mannose comme la GNA du perce-neige *Galanthus nivalis* a été démontrée, il y a une dizaine d'année (Powell et al., 1993 ; Sauvion et al., 1996). Beaucoup moins toxique pour les mammifères que d'autres lectines entomotoxiques, la GNA a été introduite dans plusieurs plantes, leur conférant une protection au moins partielle contre les pucerons (Hilder et al., 1995 ; Down et al., 1996). Certains travaux, sujets à controverse, montrent cependant un possible effet négatif de cette lectine dans des pommes de terre transgéniques données crues à des rats (Ewen and Pusztai, 1999), ainsi qu'un effet négatif sur certains insectes bénéfiques, prédateurs des pucerons intoxiqués (Birch et al., 1999). Ce dernier effet n'est toutefois pas confirmé sur la coccinelle *Adalia bipunctata* L. (Down et al., 2000).

Tableau 4: Quelques exemples de plantes transgéniques exprimant des lectines.

lectine	Insectes cibles	Plantes	Référence
GNA	*Nephotettix virescens* *Nilapavarta lugens*	riz	(Foissac et al., 2000 ; Rao et al., 1998)
	Heliothis virescens	tabac	(Gatehouse and Gatehouse, 1998)
	Myzus persicae		(Hilder et al., 1995)
		pomme de terre	(Down et al., 1996)
	Lacanobia oleceracea		(Gatehouse et al., 1997)
	Aulacorthum solani		(Gatehouse et al., 1996)
GNA + BCH	*M. persicae*		
ConA			(Gatehouse et al., 1999a)
	L. oleceracea		
PLec	*H. virescens*	tabac	(Boulter et al., 1990)
PLec + CpTI			
WGA	*Diabrotica undecimpunctata*	maïs	(Czapla and Lang, 1990)

GNA : lectine du perce-neige, ConA : concanavaline A, Plec : lectine du pois, WGA : lectine du blé, BCH : chitinase du haricot, CpTI : inhibiteur de trypsine de *Vigna unguiculata*.

b) Les hémilectines et RIP

Ces protéines proches des lectines sont souvent concomitantes avec ces dernières. Elles sont toxiques à des concentrations de plusieurs ordres de magnitude plus faible que celles requises pour les lectines. Contrairement aux lectines typiques multivalentes, ces protéines toxiques se comportent comme des lectines monovalentes, désignées comme hémilectines (Carlini and Guimaraes, 1991) ou chimérolectines (Peumans and Van Damme, 1995). Comme de nombreuses toxines bactériennes, elles sont composées de deux chaînes polypeptidiques aux propriétés distinctes. L'haptomère ou chaîne B (Binding) interagit avec un sucre ou composé glycosylé de la surface de la cellule. L'effectomère ou sous-unité A (activity) est l'agent toxique à activité enzymatique. Après internalisation, la protéine est activée par une protéolyse limitée, une réduction ou un changement de conformation par acidification de la vésicule, ce qui entraîne la séparation des deux domaines. L'effectomère activé traverse alors la membrane de la vésicule et agit sur les constituants du cytoplasme provoquant les effets toxiques (Montecucco, 1998 ; Olsnes et al., 1999). Ricine et abrine inhibent la synthèse protéique des cellules eucaryotes par inactivation de la sous-unité 60 S des ribosomes (Olsnes et al., 1974). Ces toxines, aux sous-unités catalytiques et de type lectine de 30 kDa chacune, sont appelées RIP de type 2 pour Ribosome Inactivating Protein à deux chaînes. Elles sont beaucoup plus efficaces que les RIPs de type 1, toxines à une seule chaîne de 27-31 kDa. Bien que puissants inhibiteurs de la synthèse protéique dans des préparations acellulaires, les RIP1 sont beaucoup moins toxiques chez la souris (DL50 de 1 à 40 mg/kg) que les RIP2 (DL50 de la ricine : 2 μg/kg). Certains RIP1 présentent une homologie de séquence avec le domaine A de la ricine et toutes possèdent un mode d'action très proche de celui de ce domaine (Barbieri et al., 1993 ; Barbieri et al., 1997 ; Battelli et al., 1997 ; Endo et al., 1988). Une importante ressemblance structurale entre lectines et RIP2 de certaines graines a été décrite (Barbieri et al., 1993). La plupart des RIP connues sont à chaîne unique, pour seulement une douzaine de type 2.

Bien que capable d'inhiber la synthèse protéique par les ribosomes d'insectes dans une préparation acellulaire, la ricine n'est pas toxique pour une variété d'insectes de différents ordres (Gatehouse et al., 1990 ; Maruniak et al., 1990). Les insectes insensibles à la ricine inactivent probablement la protéine, ou n'internalisent pas l'effectomère.

c) Les arcélines

Les arcélines sont des protéines insecticides présentent dans les graines de certaines variétés sauvages du haricot, *Phaseolus vulgaris*, et y remplaçant la protéine de réserve phaséoline. Elles lui confèrent une résistance aux bruches, en particulier à *Zabrotes subfasciatus*, un des principaux ravageurs des haricots stockés, dont elles inhibent le développement larvaire (Osborn et al., 1988). Une séquence similaire aux arcélines du haricot commun a été découverte chez *P. acutifolius*, suggérant la présence de ces protéines dans le genre *Phaseolus*. Des comparaisons de séquences montrent que les arcélines appartiennent à la famille des protéines de « type lectine » du haricot qui comporte les deux types de sous-unités des phytohémagglutinines PHA-L et PHA-E et les inhibiteurs d'α-amylase de cette plante (Chrispeels and Raikhel, 1991). Les membres de cette famille ont une structure tertiaire semblable[3], mais leurs propriétés biologiques, leur structure quaternaire et leur spécificité sont différentes. Les arcélines lient peu les sucres et sont de pauvres hémagglutinants. Elles existent sous forme de dimères (Arc2), tétramères (Arc3 et 4) ou les deux (Arc1). Arc5 a été observée sous forme de dimère en solution et de monomère en cristal. Les variants d'Arc5 diffèrent également par leur degré de glycosylation, Arc5a contenant deux chaînes glycanes et Arc5c aucune (Goossens et al., 1994). Toutefois aucun lien n'apparaît entre l'état d'oligomérisation, les propriétés de liaison aux sucres et la glycosylation des arcélines.

Sur les sept variants allèliques connus (Arc 1 à 7), allant de 27 à 42 kDa, Arc5 et Arc1 confèrent la meilleure résistance aux bruches (Cardona et al., 1990 ; Fory et al., 1996). Toutefois, chaque variant est en réalité représenté par plusieurs polypeptides spécifiés par différents gènes (Hartweck et al., 1991) et il n'est pas clair que ces diverses isoformes aient la même efficacité contre les ravageurs : L'isoforme Arc5a, exprimée dans les graines de *P. acutifolius* transgénique, n'apparaît pas suffisamment active pour protéger la plante de *Z. subfasciatus* (Goossens et al., 2000).

Le mécanisme d'action des arcélines n'est pas connu. Osborn et ses collaborateurs (1988) les estiment toxiques alors que Minney et ses associés (1990) les considèrent comme inhibiteurs de la digestion, ce qui entraînerait la mort des larves par inanition. Toutefois, les propriétés insecticides des arcélines pourraient être liées à leur interaction avec les glycoprotéines ou autres composants du tube digestif. Si les arcélines sont toxiques pour *Z. subfasciatus*, elles ne semblent pas avoir d'effet sur *Acanthoscelides obtectus*, une autre bruche, ravageur du haricot. Pour comprendre cette différence, des analyses de tissus larvaires ont été effectuées et montrent qu'Arc1 détruit les structures épithéliales de certaines régions de l'intestin moyen de *Z. subfasciatus* mais pas d'*A. obtectus* (Paes et al., 2000). De plus, Arc1 est détectée dans l'hémolymphe de *Z. subfasciatus*, indiquant sa capacité à traverser les cellules du tube digestif.

[3] La structure de l'arcéline Arc5a présente un feuillet de 6 brins β plan et aplati contre un feuillet de 7 brins β incurvé (Hamelryck et al.,1996), comme les lectines de Légumineuses (Famille de la base de structures SCOP : Legume Lectin) (http://scop.berkeley.edu).

(b) Canatoxines et uréases

Les graines de *Canavalia ensiformis* contiennent non seulement la lectine concanavaline A, mais également de l'uréase ainsi que la protéine canatoxine, un puissant neurotoxique. Cette dernière agit sous forme de dimère de chaînes de 95 kDa, liées de façon non covalente, et représente 0,5% du poids sec de la graine (Carlini and Guimaraes, 1981). Elle agit comme une hémilectine mais en interagissant avec des glycoconjugués plus complexes comme les gangliosides (Carlini and Guimaraes, 1981), se rapprochant ainsi plus des hémilectines bactériennes neurotoxiques, comme les toxines botulique et tétanique (Eidels et al., 1983).

La canatoxine présente une forte homologie de séquence avec l'uréase des mêmes graines (Follmer et al., 2001) et en apparaît comme un variant. Elle ne possède toutefois que 30% de l'activité uréolytique de l'uréase. L'uréase présente également certaines propriétés biologiques de la canatoxine, comme la toxicité sur les insectes, l'activité hémilectine ou encore l'activation des plaquettes du sang. L'activité enzymatique de ces protéines n'est pas responsable de leur effet sur les plaquettes ou de la toxicité intrapéritonéale chez la souris de la canatoxine (Follmer et al., 2001). Les diverses propriétés biologiques de ces protéines semblent liées à des domaines structuraux différents.

Les homologues de la canatoxine sont bien répartis au sein des **Légumineuses**, et même chez des espèces comestibles (Carlini et al., 1988). Leur accumulation s'effectue pendant la maturation de la graine (Barcellos et al., 1993). Ces caractéristiques suggèrent un rôle important, probablement lié à la défense. En effet, la canatoxine est mortelle par ingestion pour les insectes ayant pour enzymes digestives principales des cathepsines B et D, comme *Callosobruchus maculatus* (Coleoptera) et *Rhodnius prolixus* (Hemiptera). Au contraire, les insectes dont la digestion repose sur des enzymes trypsiques ne sont pas affectés. Les cathepsines B et D activeraient la toxine en la coupant pour donner un ou plusieurs **peptides insecticides** (Carlini et al., 1997). Toutefois le mode d'action de ces peptides reste inconnu. Dix-huit à 24 heures après le repas, ils ne sont plus présents dans l'intestin des nymphes de *R. prolixus* mais dans son hémolymphe (Ferreira-DaSilva et al., 2000). Cependant, la mortalité ne culmine que 3 jours après l'ingestion, indiquant probablement le temps nécessaire pour atteindre une concentration critique de ces peptides dans l'organe ou le tissu cible, encore non identifié.

(c) Inhibiteurs enzymatiques

Les aliments consommés par les insectes phytophages sont constitués essentiellement de polymères de nature glucidique comme l'amidon, la cellulose et l'hémicellulose et protéique comme les holo- et hétéroprotéines (glyco-, lipo- ou métalloprotéines).

Chez les insectes, la digestion est séquentielle et peut être divisée en trois phases : initiale, intermédiaire et finale. La phase initiale met en jeu des polymères hydrolases (amylases, protéases) qui fragmentent les composés ingérés de grand poids moléculaire en oligomères protéiques ou glucidiques. Ces derniers sont à leur tour découpés par les oligomères hydrolases lors de la phase intermédiaire pour libérer le maltose, le cellobiose et les dipeptides issus respectivement de l'amidon, de la cellulose et des protéines. Dans la phase finale, ces composés de faible poids moléculaire sont dégradés par la maltase, la cellobiase et les dipeptidases, en unités simples directement absorbables, les nutriments.

Il existe des composés d'origine et de nature diverses dont la présence ralentit ou anhile l'acte catalytique des enzymes. Ce sont les **inhibiteurs d'enzymes**. Leur présence conduit à un **dérèglement du métabolisme** de l'organisme concerné, dérèglement qui peut entraîner un retard de croissance, de développement voire la mort des individus.

En plus de l'intérêt de leur étude pour la nutrition et la toxicologie alimentaire, les inhibiteurs d'enzymes sont des outils potentiellement importants pour la protection des végétaux. Parmi les inhibiteurs naturels des plantes, les inhibiteurs protéiques les plus étudiés sont ceux de protéases et d'amylases.

a) Inhibiteurs de protéases (IP)

Les protéases sont les enzymes responsables de la dégradation des protéines. Les endoprotéases sont classées en quatre groupes sur la base de la nature chimique des groupements responsables de l'activité catalytique :

- Les protéases à sérines (EC 3.4.21)[4]
- Les protéases à cystéines (EC 3.4.22)
- Les protéases acides (à aspartate ou glutamate) (EC 3.4.23)
- Les métalloprotéases (avec Zn^{2+}, Ca^{2+} ou Mn^{2+}) (EC 3.4.24)

La protéolyse étant un phénomène cellulaire très contrôlé, un grand nombre d'inhibiteurs naturels sont présents chez les microorganismes, les plantes et les animaux. Ils sont classés par rapport au type de protéase qu'ils inhibent.

(i) Les inhibiteurs de protéases à sérine

Les protéases à sérine constituent sans aucun doute les protéases les plus étudiées et le mieux caractérisées actuellement. Il n'est donc pas étonnant que la majorité des inhibiteurs naturels de protéases connus à ce jour soient des inhibiteurs de protéases à sérine. Une classification de ces inhibiteurs en fonction du mécanisme d'inhibition, permet de distinguer trois principaux groupes (Laskowski and Kato, 1980) dont les **inhibiteurs canoniques** sont les seuls présents chez les végétaux. Ces inhibiteurs canoniques des protéases à sérine sont d'ailleurs les IP à sérine les plus nombreux. Bien que ces molécules aient des modes de repliement complètement différents, elles possèdent toutes en commun une boucle inhibitrice exposée vers l'extérieur, de séquence variable mais dont la conformation P3-P3' (six résidus autour du site inhibiteur)[5] est très similaire, et appelée de ce fait conformation canonique. La structure 3D qui porte cette boucle, comporte un cœur hydrophobe très compact, stabilisé le plus souvent par plusieurs ponts disulfures, ce qui rend ces molécules remarquablement stables. On peut considérer que la formation du complexe protéase-inhibiteur résulte de l'association entre deux molécules rigides. L'interaction est limitée à la région de la boucle inhibitrice avec le site actif de l'enzyme, mais elle est très spécifique du fait de la complémentarité très élevée de forme entre les deux surfaces moléculaires. L'affinité de l'IP pour l'enzyme est très grande mais à la différence d'un substrat, la liaison peptidique sensible P1-P1' est hydrolysée très lentement.

Richardson (1991) en se basant sur l'homologie structurale entre les séquences, sur la localisation des ponts disulfures et sur la position du site actif, a défini sept familles d'IP à sérine. Les deux familles qui renferment le plus grand nombre de séquences connues sont les inhibiteurs de la famille Bowman-Birk ayant deux sites actifs (Birk, 1976), et de la famille Kunitz. Ces derniers d'environ 20 kDa, ne possédent qu'une seule chaîne avec un seul site réactif, renfermant 2 ponts disulfures. Certains inhibiteurs de protéases à sérine sont bifonctionnels, agissant également sur des α-amylases (cf. b)).

[4] Numérotation selon la nomenclature des enzymes (Commission on Biochemical Nomenclature, 1972), telle que décrite dans la base ENZYME (Bairoch**,** 2000).

[5] numérotation des acides aminés selon Schechter and Berger (1967).

(ii) Les inhibiteurs de protéases à cystéine

Les cystatines, ainsi nommées pour leur capacité à bloquer l'action des protéases à cystéine, en sont les principaux inhibiteurs endogènes. Ces molécules sont des inhibiteurs protéiques à forte affinité et réversibles. La super famille des cystatines est divisée en trois familles chez les mammifères, selon des caractéristiques structurales et leur localisation cellulaire (Barrett, 1987) :

- les **stéfines** (famille 1) d'environ 100 acides aminés (11 kDa). Elles ne possèdent pas de ponts disulfure et leur localisation est principalement intracellulaire.
- les **cystatines** (famille 2) sont des molécules un peu plus complexes constituées d'environ 115 acides aminés (13 kDa). Elles renferment en général deux ponts disulfure qui forment deux boucles à proximité de l'extrémité C-terminale de l'IP. Elles sont présentes dans beaucoup de fluides extracellulaires.
- les **kininogènes** (famille 3) apparaissent avec leur structure complexe de 60 à 120 kDa comme les membres les plus évolués de la superfamille des cystatines.

Les inhibiteurs de protéase à cystéine de plantes (phytocystatines) présentent des caractéristiques des familles 1 et 2 (Arai et al., 1998 ; Turk et al., 1997), mais ils ont la particularité de ne pas posséder de ponts disulfures (Nagata et al., 2000). Les phytocystatines, présentes dans diverses mono- et dicotylédones, ont une masse moléculaire de 5 à 87 kDa. Elles ont été purifiées à partir de nombreuses graines (Abe et al., 1987 ; Pernas et al., 1998), mais la présence de ces IP a aussi été révélée dans d'autres tissus tels que les feuilles (Zhao et al., 1996), les fruits (Kimura et al., 1995) et le pollen (Rogers et al., 1993). On distingue deux groupes de phytocystatines. Les premières, à domaine unique, sont majoritaires. Le deuxième type regroupe les inhibiteurs possédant plusieurs domaines, comme les multicystatines des tubercules de pomme de terre ou des graines de tournesol (Kouzuma et al., 2000 ; Walsh and Strickland, 1993).

Les structures 3D obtenues par cristallographie de deux membres représentatifs de la superfamille, la cystatine de poulet (Bode et al., 1988) et la stéfine B humaine complexée avec la papaïne (Stubbs et al., 1990), ont permis de comprendre le mécanisme d'action de ces inhibiteurs. La structure de ces molécules consiste notamment en un feuillet β antiparallèle composé de cinq brins. A l'une des extrémités du feuillet β se trouve une séquence consensus du type QxVxG, flanquée de part et d'autre de la région dipeptide PH pour les familles 2 et 3 et de la région N-terminale possédant une glycine invariante en position 9. Ces trois régions, formant le site inhibiteur, ne sont pas contiguës dans la séquence et donnent à l'IP une forme en "coin", de nature hydrophobe, tout à fait complémentaire du site actif de la papaïne et d'autres protéases à cystéine. A la différence de substrats liés dans le site actif de l'enzyme, la région N-terminale adopte une structure en épingle à cheveux qui lui permet de contourner la cystéine 25 du site catalytique, et ne peut donc pas être clivée (Bode and Huber, 2000).

Les phytocystatines diffèrent des inhibiteurs des animaux par leur très forte activité vis-à-vis des protéases du tube digestif des insectes, les rendant particulièrement attractives en tant qu'agent de contrôle de ces ravageurs (Bode and Huber, 1992).

(iii) Les inhibiteurs de métalloprotéases et de protéases acides

Ces dernières classes d'inhibiteurs n'ont été mises en évidence que dans très peu d'organismes. De manière générale, très peu d'inhibiteurs de métalloprotéases ou de protéases acides sont connus à ce jour.

Un inhibiteur de protéase acide a été caractérisé chez la courge (Christeller et al., 1998). Cette molécule de 10,5 kDa circule dans le phloème mais sa cible physiologique n'est pas connue. Cet IP est capable d'inhiber une protéase acide synthétisée par un champignon pathogène des Cucurbitacées, suggérant un rôle possible dans la défense de la plante contre les pathogènes.

(iv) Le rôle des inhibiteurs de protéases chez les plantes

Les IP dans les plantes, du moins pour ce qui concerne les serpines et cystatines les plus étudiées, semblent remplir plusieurs fonctions.

- *Régulation des protéases endogènes*

Certains IP sont impliqués dans la régulation des protéases endogènes de la plante, lors de la régulation de la mort cellulaire programmée ou **apoptose** (Solomon et al., 1999), la régulation de **la mise en place des réserves dans la graine** (Abe et al., 1987) et la régulation de **la mobilisation des réserves** lors de la germination (Papastoitsis and Wilson, 1990).

- *Protection contre les phytopathogènes et insectes.*

On trouve des IP en grande quantité dans les organes de réserve comme les graines ou les tubercules (10 à 15% des protéines totales), et dans les fruits immatures. En plus du rôle direct de stockage, la prévalence des IP dans les organes de réserve suggère **un rôle défensif contre les herbivores** et en particulier les insectes.
Beaucoup de plantes répondent à l'attaque des herbivores en activant des gènes de défense dont les produits réduisent la qualité nutritionnelle des protéines ingérées (Polyphenol oxydase) et inhibent les protéases digestives de l'insecte (IP), ce qui peut être directement lié à une augmentation de résistance (Ryan, 1990).

La démonstration directe du rôle négatif des IP sur la croissance des larves d'insectes a été faite en 1987. Des lignées de tabac transgéniques résistantes à l'attaque des larves *d'Heliothis virescens* ont été créées par expression constitutive de l'inhibiteur trypsique du niébé (CpTI) (Hilder et al., 1987). Lorsque les larves de premier stade se nourrissent de feuilles dans lesquelles CpTI représente 1% des protéines solubles, on observe après 7 jours, une mortalité de 62,5% des larves. Les larves survivantes présentent un retard de croissance (-70%). Ces plantes transgéniques sont, par conséquent, beaucoup moins défoliées.
Toutefois, dans le cas des inhibiteurs inductibles des Solanacées, on observe rarement ces effets délétères sévères lorsque les insectes sont nourris sur des milieux nutritifs artificiels contenant des concentrations en IP similaires à celles produites par les plantes blessées (Jongsma and Bolter, 1997). Cette observation suggère que les IP agissent sur l'insecte en interaction avec d'autres molécules produites par la plante, comme les composés phénoliques ou les alcaloïdes. Par conséquent, les inhibiteurs de protéases seraient une des composantes de la réaction de défense des plantes, et non l'unique effecteur.
Depuis les travaux de Hilder, une quinzaine de plantes transgéniques exprimant des IP se sont révélées efficaces contre les insectes cibles. Ce sont surtout des inhibiteurs de protéases à sérine (les plus étudiés chez les plantes), efficaces contre les larves de lépidoptères, qui ont été introduits dans ces plantes (Tableau 5). Les bio-essais montrent

que la diminution du poids des larves se développant sur les plantes transgéniques est bien proportionnelle au niveau de sur-expression de l'IP.
Bien que la gamme d'inhibiteurs de protéases, ainsi que celle des plantes d'intérêt agronomique transformées, soient en augmentation, la viabilité commerciale de cette stratégie n'est toujours pas démontrée. Le niveau de protection atteint grâce à la sur-expression d'IP dans les plantes transgéniques est, dans certains cas, supérieur à 50 % en terme de réduction des dégâts, de mortalité des larves ou de diminution de la biomasse. Mais ces niveaux ne sont pas considérés comme étant suffisants commercialement. Ils n'atteignent pas ceux des variétés exprimant des toxines de *Bacillus thuringiensis* (quelques dizaines de microgrammes de toxine par gramme de tissus induisent plus de 95 % de mortalité).

Tableau 5 : Exemples de plantes transgéniques exprimant des inhibiteurs de protéases.

Inhibiteur de protéase	**Insectes cibles**	**Plantes**	**Références**
IP à sérine			
CpTI (Inhibiteur de trypsine de *V. unguiculata*).	*Heliothis virescens*	Tabac	(Hilder et al., 1987)
	Manduca. sexta		
	Chilo suppressalis	Riz	(Xu et al., 1996)
	Spodoptera infestans		
	Pieris rapae	Chou	(Hao and Ao, 1997)
	H. armigera		
	Otiorhynchus sculatus (C)	Fraise	(Graham et al., 1995)
Kunitz du Soja (inhibiteur de trypsine)	*H. virescens*	Tabac	(Gatehouse et al., 1993)
	S. littoralis		(Marchetti et al., 2000)
Pot2 (inhibiteur de trypsine de la Pomme de Terre)	*M. sexta*		(Johnson et al., 1989)
	Plagiodera versicolor (C)	Peuplier	(Klopfenstein et al., 1997)
Inhibiteur multidomaines de *Sinapis alba*	*Plutella xylostella* *Mamestra brassicae*	*Arabidopsis* et Colza	(De Leo et al., 2001)
Cystatines			
OC1 (oryzacystatine du riz) *Arabidopsis thaliana*	*M. persicae* (H)	Colza	(Rahbé et al., 2003)
	Chrysomela tremulae (C)	Peuplier	(Leplé et al., 1995)
	C. populi (C)		(Delledonne et al., 2001)

(C) : Coléoptères, (H) : Homoptères, les autres insectes cibles sont des Lépidoptères.

(v) Effets de l'ingestion d'inhibiteurs de protéases sur les protéases des insectes

De nombreux travaux ont eu pour objet l'identification des enzymes digestives et la détermination de la structure fine du tube digestif chez des insectes d'ordres et de régimes alimentaires différents (pour revue : Terra and Ferreira, 1994).

- *Déséquilibre du pool d'acides aminés*

De nombreuses études ont été menées afin de déterminer l'effet de l'ingestion d'inhibiteurs de protéases sur la survie et le développement de l'insecte. Ces essais ont été effectués *in vivo* grâce à l'incorporation de l'IP dans un milieu nutritif artificiel (Bown et al., 1998 ; Christeller et al., 1992 ; Christeller et al., 1994 ; Liang et al., 1991). En comparaison avec le milieu nutritif contrôle, la présence d'un IP efficace permet d'observer une augmentation de la mortalité des larves, une perte de poids et un retard de développement (Gatehouse et al., 1999b ; Jongsma and Bolter, 1997). Ce sont des substances antimétaboliques et antinutritives responsables d'une réduction du taux de protéolyse et donc d'une **moindre disponibilité en acides aminés** de l'alimentation conduisant à un retard de croissance larvaire (Burgess et al., 1994 ; De Leo et al., 1998 ; Orr et al., 1994). Un enrichissement du milieu nutritif en acides aminés soufrés (méthionine, cystéine) élimine les effets délétères des IP sur la croissance larvaire (Broadway, 1996 ; Gatehouse and Boulter, 1983 ; Oppert et al., 1993). Ces auteurs montrent également que les retards de croissance observés sont corrélés à une hyper-production de protéases digestives en réponse à l'ingestion d'IP. Le taux de digestion des protéines *in vivo* peut ainsi être maintenu malgré la présence de l'IP. Cette hyper-production de protéases mobiliserait des acides aminés soufrés au détriment de la synthèse d'autres protéines essentielles. Le détournement métabolique ainsi induit semble conduire à un déséquilibre du stock d'acides aminés disponibles, qui aboutit finalement à l'apparition de carences en certains d'entre eux (certainement correspondant aux acides aminés faiblement représentés dans la nourriture ingérée). Les IP n'agissent donc pas forcément par simple blocage de l'assimilation des protéines mais aussi par mobilisation des réserves d'acides aminés de la larve en vue de la synthèse de protéines.

- *Stratégie adaptative des insectes face à l'ingestion d'IP*

Malgré les résultats encourageants que l'on observe grâce à l'expression d'inhibiteurs de protéases exogènes pour lutter contre les insectes phytophages, certaines lignées transgéniques ne présentent aucun gain de résistance, voire des effets inverses, les insectes se développent beaucoup mieux (effets probiotiques).
Les effets de l'ingestion d'inhibiteurs de protéases sur les insectes sont très complexes. Chaque espèce d'insecte, voire chaque population d'insecte, réagit différemment à la sur-expression d'inhibiteurs de protéases dans son alimentation (Girard et al., 1998a). Grâce à l'étude de quelques cas, il a été mis en évidence trois stratégies adaptatives :

- La détoxication de l'alimentation **par dégradation de l'IP**. Par exemple, les peupliers transgéniques exprimant constitutivement l'oryzacystatine (OC-I) présentent un gain de résistance aux larves du coléoptère *C. tremulae*. En effet, 40% des larves meurent après ingestion des feuilles de ces peupliers en fin de cycle. Pourtant, aucun effet délétère n'est observé lorsque les larves de la phédone du cresson (*P. cochleaciae*) se nourrissent sur des colzas transgéniques exprimant OC-I (Girard et al., 1998b). De plus, aucun changement au niveau de la nature des protéases ou de leur niveau d'activité n'est observé, aucune réponse de compensation n'est mise en place. La résistance des larves de phédone semble être médiée par la dégradation d'OC-I qui perd rapidement son activité. Ce système de détoxication met en jeu des protéases à sérine en association avec des leucine-aminopeptidases déjà présentes dans le tube digestif de la larve.

- L**'adaptation favorable suite à un faible niveau d'expression de l'IP**. Lorsque l'inhibiteur trypsique de la moutarde MTI-2 est exprimé à un taux de 1,6% des protéines solubles dans les feuilles d'*A. thaliana* ou de tabac transgéniques, des effets délétères sur les larves du lépidoptère *S. littoralis* (20% de mortalité, 50% de réduction de croissance, 56% de réduction de la fertilité) sont observés ainsi qu'une réduction des dommages foliaires (De Leo et al., 1998 ; De Leo and Gallerani, 2002). Par contre, lorsque ces larves se nourrissent de feuilles de plantes transgéniques exprimant l'inhibiteur à un faible niveau (0,5% des protéines solubles), les larves se développent plus rapidement et plus efficacement, provoquant des dégâts beaucoup plus importants sur la plante (+ 26%). Ces observations sont corrélées à une sur-expression de protéases pré-existantes et sensibles à MTI-2 *in vitro*. L'accroissement de la consommation foliaire peut être la conséquence d'une diminution de la qualité nutritive de l'aliment du fait de la présence de l'IP, et/ou d'une activité protéolytique accrue chez ces larves.
- L**'induction de protéases insensibles à l'IP**. De nombreuses équipes ont pu révéler la *néo*-synthèse de protéases insensibles à l'inhibiteur ingéré (Bown et al., 1997 ; Broadway, 1995 ; Jongsma et al., 1995). L'équipe de Jongsma a pu montrer, que contrairement aux résultats encourageants obtenus avec les tests en milieux artificiels, le développement des larves du lépidoptère *S. exigua* n'est pas affecté par l'expression constitutive de PI-II (inhibiteur de protéases à sérine de la pomme de terre) dans des tabacs transgéniques. Ces insectes mettent en place un mécanisme d'adaptation par induction de protéases insensibles à l'inhibiteur. Exposées continuellement à l'inhibiteur, les larves modifient la nature de leurs protéases, réduisant ainsi leur sensibilité et abolissant l'effet antimétabolique du produit du transgène. Bown et al. (1997) ont pu éclaircir le mécanisme moléculaire impliqué dans cette réponse adaptative. Le génome du lépidoptère *H. armigera* renferme au moins vingt huit gènes codant des protéases à sérine. Bien que les résidus du site actif soient conservés, les résidus impliqués dans le contact protéase-inhibiteur sont variables entre les différentes isoformes (démontrant à nouveau la très fine spécificité d'interaction entre l'inhibiteur et sa cible). Ainsi, l'expression de ces différents gènes peut être différentiellement induite selon la nature de l'inhibiteur. Récemment, Mazumdar Leighton et Broadway (2001) ont mis en évidence la cinétique d'induction de ces différentes classes de protéases. La réponse des larves de lépidoptères se décompose en deux phases : tout d'abord, il y a surexpression de protéases sensibles à partir d'un pool d'ARNm pré-existants, et ensuite induction d'une synthèse *de novo* de messagers codant pour des protéases insensibles. Ces protéases insensibles sont détectées deux heures après l'ingestion de l'inhibiteur (Broadway, 1997). Enfin, l'étude de l'interaction entre *H. armigera* et sa plante hôte (pois chiche) montre que les protéases de la larve sont capables de dégrader les inhibiteurs de protéases synthétisés par la plante pour se protéger (Giri et al., 1998). L'induction de protéases insensibles aux inhibiteurs de protéases endogènes, induits lors de la blessure, semble être un mécanisme fréquent grâce auquel les insectes peuvent contourner l'accumulation d'IP. Ces résultats démontrent la rapidité de la réponse adaptative, la plasticité de la physiologie digestive des insectes, ainsi que la complexité de la régulation de tous ces gènes codant des protéases.

b) Inhibiteurs d'alpha-amylases (IA)

Les α-amylases ou α1,4-glucane-4-glucanohydrolases (EC 3.2.1.1) sont des enzymes très répandues dans le monde vivant, hydrolysant les polymères glucidiques liés en α1,4 comme l'amidon et le glycogène, en oligosaccharides. Elles sont les enzymes digestives principales de nombreux insectes se nourrissant exclusivement de graines durant leur vie larvaire et/ou imaginale. Les inhibiteurs d'α-amylases (IA) protéiques font partie des mécanismes de défense de nombreuses plantes et sont particulièrement abondants dans les céréales (Abe et al., 1993 ; Feng et al., 1996 ; Iulek et al., 2000 ; Yamagata et al., 1998) et les légumineuses (Giri and Kachole, 1998 ; Marshall and Lauda, 1975 ; Melo et al., 1999). Ces IA représentent un intéressant potentiel pour la création de plantes transgéniques résistantes à différents ravageurs (Chrispeels et al., 1998 ; Gatehouse and Gatehouse, 1998). Toutefois, la détermination de leur spécificité est une étape essentielle préliminaire à leur utilisation. En général, les IA inhibent plusieurs α-amylases de différents organismes mais certains sont spécifiques des insectes. On distingue six classes d'IA protéiques de plantes, selon leur structure : les lectin-like, les type « céréales », les Kunitz, les thaumatin-like, les knottines et les défensines. Ces deux dernières familles, peptidiques seront abordées au paragraphe 2)(a).

(i) Les IA de type lectine.

Il s'agit d'inhibiteurs d'amylases caractérisés chez différentes variétés de *Phaseolus vulgaris*. La première isoforme découverte, αAI1, a été identifiée comme homologue de la lectine PHA (Moreno and Chrispeels, 1989). Les lectines **PHA, les αAI et les arcélines** constituent une famille de protéines de défense des graines de ***Phaseolus***. Les gènes des membres de cette famille sont liés dans le génome de *P. vulgaris*, leur séquences protéiques sont hautement homologues (45 à 85% d'identité) et ils dérivent probablement d'un gène ancestral par duplication et divergence (Chrispeels and Raikhel, 1991 ; Mirkov et al., 1994).

Les différents variants d'αAI présentent des spécificités différentes. αAI1 inhibe l'α-amylase pancréatique du porc (PPA) ainsi que celles de *Callosobruchus maculatus* et *C. chinensis* mais pas celle de *Z. subfasciatus*. Alpha-AI2 possède un profil d'inhibition contraire (Grossi De Sa et al., 1997).

Leur structure active est un hétérotétramère de deux glycopeptides de chaque sous-unité α et β de 7,8 et 14 kDa respectivement. Chaque tétramère inhibe deux molécules d'α-amylases. Des analyses structurales du complexe entre αAI1 et l'amylase de *Tenebrio molitor* (TMA) montrent que deux boucles en épingle à cheveux (résidus 29-46 et 171-189) sont insérées dans le site actif y formant un réseau de liaisons hydrogènes, empêchant le substrat de l'enzyme de se lier (Nahoum et al., 1998).

Ces inhibiteurs sont **toxiques pour divers insectes** ravageurs (Grossi De Sa et al., 1997 ; Huesing et al., 1991 ; Ishimoto and Kitamura, 1989). Des pois transgéniques exprimant αAI1 dans leur graines ont montré une résistance totale aux bruches *C. chinensis*, *C. maculatus* et *Bruchus pisorum* (Schroeder et al., 1995 ; Shade et al., 1994), même en condition de plein champs (Morton et al., 2000). Les pois transgéniques ont peu d'effets sur la digestion de rats (Pusztai et al., 1999), suggérant la possibilité d'incorporer ces inhibiteurs dans des plantes alimentaires, sans danger pour l'homme ou le bétail.

(ii) Les IA des céréales.

Ces inhibiteurs d'α-amylases sont composés de 120 à 160 acides aminés, 5 ponts disulfure et sont connus comme allergènes des farines de céréales (Buonocore et al., 1977 ; Garcia Casado et al., 1996). L'inhibiteur du blé **0.19** et celui, bifonctionnel, du ragi (*Eleusine coracana*) **RBI** en sont les plus étudiés. L'inhibiteur 0.19, nommé ainsi pour sa mobilité en gel d'électrophorèse, agirait sous forme de dimère sur les α-amylases des oiseaux, bacilles, mammifères et insectes (Feng et al., 1996 ; Franco et al., 2000). Sa structure montre 5 hélices α d'orientation alternée (Oda et al., 1997).

RBI est un monomère inhibant **α-amylase** et **trypsine**. Sa structure tridimensionnelle est très proche de celle de 0.19, étant globulaire avec 4 hélices α alternées et un petit feuillet β antiparallèle. Il inhibe une variété d'amylases dont PPA et TMA. La structure du complexe RBI-TMA montre que l'inhibiteur se lie au site actif de l'enzyme, empêchant la liaison du substrat (Strobl et al., 1998). L'extrémité N-terminale et les résidus P_{52} à C_{55} s'insèrent dans le site de liaison de l'enzyme, les acides aminés de la partie N-terminale formant des liaisons hydrogènes avec les résidus catalytiques de l'enzyme. Les 10 acides aminés N-terminaux sont d'ailleurs suffisants pour l'inhibition de l'enzyme (Alam et al., 2001).

Cette famille d'IA des céréales est multigénique et les différentes séquences fournissent un important champ de spécificités (Garcia Casado et al., 1994).

(iii) Les IA de type Kunitz

Les inhibiteurs d'α-amylases de la famille Kunitz possèdent environ 180 acides aminés dont 4 cystéines et sont présents dans les céréales. Le mieux caractérisé de ces inhibiteurs est **BASI** (Barley Amylase Subtilisin Inhibitor) qui provient de l'orge et inhibe **α-amylase** et **subtilisine** (une protéase à sérine) (Mundy et al., 1983). Sa structure en trèfle de brins β (Vallé et al., 1998) est partagée par l'inhibiteur de trypsine d'*Erythrina caffra* (Onesti et al., 1991) et la chaîne B de la ricine (Rutenber and Robertus, 1991). Le mode d'action de BASI est différent des αAI ou de RBI car il n'interagit pas directement avec les résidus catalytiques de l'enzyme AMY2 (l'α-amylase 2 de l'orge) mais avec des domaines proches du site catalytique, bloquant son accès au substrat (Vallé et al., 1998).

BASI est impliqué dans la **régulation d'AMY2 endogène** lors de la dégradation des sucres des graines (Kadziola et al., 1998). Il possède également un rôle de **défense** des graines vis-à-vis de protéases et amylases externes produites par divers ravageurs et pathogènes (Garcia Olmedo et al., 1992). Ces inhibiteurs bifonctionnels sont spécifiques des amylases endogènes et d'**insectes** mais ne semblent pas être actifs sur celles des mammifères.

(iv) Les IA de type thaumatine.

Ces inhibiteurs d'environ 22 kDa sont proches des protéines de défense PR5 (Pathogenesis Related group 5) et de la thaumatine, une protéine au goût sucré du fruit de *Thaumatococcus danielli* (Cornelissen et al., 1986). Le plus étudié de ces inhibiteurs est la **zéamatine**, bifonctionnelle, de *Zea mays*. La zéamatine est consituée de 13 brins β dont 11 forment un sandwich au cœur de la molécule (Batalia et al., 1996). Plusieurs boucles partent de ce cœur et sont maintenues par un ou plusieurs des 8 ponts disulfures. Cet inhibiteur n'est pas glycosylé et inhibe la **trypsine** pancréatique du porc et les **α-amylases** de *Tribolium castaneum, Sitophilus zeamais* et *Rhyzoperta dominica* (Schimoler O' Rourke et al., 2001). Un homologue de cette protéine (osmotine du tabac) est également

connue pour son activité antifongique qui ne dépend pas de l'inhibition enzymatique mais de la perméabilisation des cellules (Abad et al., 1996).

(v) Interêt des inhibiteurs d'enzymes bifonctionnels.

Les inhibiteurs bifonctionnels, contre amylases et protéases, sont assez courants et leur utilisation pour la protection des végétaux face aux insectes est attrayante. Dans le cas du RBI, la formation d'un complexe stable amylase-RBI-trypsine a été observé (Maskos et al., 1996). Les deux sites d'inhibition sont indépendants : la partie N-terminale pour l'amylase et une boucle canonique de type substrat liant la trypsine. Pour les IA de type Kunitz, la présence de deux sites indépendants est suggérée par les études faites sur l'inhibiteur du blé WASI (Wheat Amylase Subtilisin Inhibitor). Son activité anti-amylasique est conservée après incubation avec la protéinase K qu'il inhibe (Betzel et al., 1988). Pour la zéamatine, il reste à déterminer si elle possède des site indépendants.
Un inhibiteur d'α-amylase d'insectes, isolé des graines de *Coix lachrima jobi* possède une activité chitinase (Ary et al., 1989), connue pour être également une activité de défense des plantes (Bishop et al., 2000) contre insectes et champignons pour lesquels la chitine est un polymère de grande importance. Cette protéine représente donc un autre potentiel pour la protection des végétaux, même s'il requiert une plus ample caractérisation.

L'utilisation de ces inhibiteurs d'enzyme pour la production de plantes transgéniques résistantes nécessite qu'ils soient actifs contre plusieurs ravageurs d'une même plante et qu'ils n'interfèrent pas avec les enzymes endogènes de la plante. L'activité sur les amylases des mammifères est moins rédhibitoire car les IA sont dénaturés par la cuisson avant ingestion. Les inhibiteurs de type Kunitz comme BASI, à forte affinité pour les enzymes des plantes, sont défavorisés. Les inhibiteurs de type céréales, WRP 25 à 27, ainsi que la zéamatine et AAI (type knottine) sont inactifs sur les enzymes de mammifères mais actifs sur celles d'insectes et semblent particulièrement intéressants.
La variabilité fonctionnelle et structurale des IA permet la recherche de variants naturels de spécificités et caractéristiques désirées. Toutefois, la connaissance des bases structurales de l'interaction des inhibiteurs avec leurs enzymes permettrait de créer de nouveaux inhibiteurs.
La coupure des IA par les protéases digestives des insectes peut compliquer le contrôle des ravageurs par ces inhibiteurs. De plus, la présence de plusieurs α-amylases dans un unique insecte (Silva et al., 1999) nécessite l'utilisation d'IA de spécificité assez large sous peine, dans le cas contraire, de n'avoir aucun impact sur la digestion et de ne pas protéger les plantes. Toutefois les IA sont prometteurs pour le contrôle des bruches des grains stockés comme le montrent les essais de plantes transgéniques exprimant αAI1 (Morton et al., 2000 ; Schroeder et al., 1995 ; Shade et al., 1994).

(d) Chitinases et protéines affines de la chitine

La **chitine**, homopolymère linéaire de N-acétyl-D-glucosamine (GlcNAc) liées en β(1,4), est un **composant structural essentiel** des parois cellulaires fongiques, de la cuticule des insectes et de l'enveloppe des œufs de nématodes. Elle est donc une cible spécifique (absente des végétaux et des vertébrés) de grand intérêt pour le développement de pesticides.
Les **lectines** de spécificité GlcNAc interfèrent avec la synthèse de la chitine des insectes et des champignons (Cohen, 1993). Les **vicilines** de légumineuses lient les structures à base de chitine dont celles du tube digestif de *C. maculatus* et *Z. subfasciatus* (Firmino et al.,

1996) et les parois et membranes cellulaires des levures et champignons (Gomes et al., 1998). Les vicilines extraites des graines de niébé résistant à *C. maculatus* et d'autres légumineuses ralentissent le développement larvaire de cet insecte (Sales et al., 2000 ; Yunes et al., 1998). Elles agiraient en réduisant la disponibilité des acides aminés (Firmino et al., 1996). Elles sont par contre inactives contre *Z. subfasciatus*.
D'autres protéines s'attaquent à la chitine déjà formée en l'hydrolysant. Ce sont les **chitinases** (EC 3.2.1.14). Toutes les plantes en possèdent et les expriment constitutivement dans les tissus les plus vulnérables ou en réponse à une attaque de pathogènes, en tant que protéine PR (Pathogenesis Related) (Collinge et al., 1993 ; Samac and Shah, 1991). La surexpression de certaines chitinases dans des plantes transgéniques leur confère une résistance accrue aux pathogènes (Broglie et al., 1991 ; Grison et al., 1996 ; Jach et al., 1995). Leur rôle dans la défense des plantes face aux champignons pathogènes, en association avec les glucanases et chitosanases est bien démontré (Boller, 1986 ; Grenier and Asselin, 1990 ; Mauch et al., 1988 ; Rousseau Limouzin and Fritig, 1991). Les produits de la dégradation de la chitine induisent la production de composés de défense (phytoalexines) et la SAR (systemic acquired resistance) (Brunner et al., 1998). Enfin les chitinases sont une des cibles de la sélection dans la coévolution plante-pathogène, s'adaptant rapidement aux défenses des pathogènes (Bishop et al., 2000). L'importance des chitinases dans la défense des plantes face aux insectes est moins bien documentée. Toutefois, leur utilisation dans la lutte contre les Homoptères en particulier semble intéressante. Des pommes de terre transgéniques exprimant constitutivement la chitinase du haricot (BCH) réduisent légèrement la fécondité du puceron *A. solani* dont les nymphes sont alors significativement plus petites que sur plante non transformée (Down R.E. cité par Gatehouse and Gatehouse, 1998).
La coexpression de la BCH avec la GNA dans des pommes de terre transgéniques montre, par comparaison avec la GNA seule, l'effet négatif de la chitinase sur la production de nymphes du puceron *M. persicae* (diminution des dégâts par baisse de la survie des adultes et de la fécondité totale) (Gatehouse et al., 1996).

Nous nous limitons ici aux protéines de défense des plantes face aux insectes mais il existe de nombreuses autres protéines à activité antimicrobiennes, appelée protéines PR (pathogenesis related). En plus de ces protéines de grosse taille, les plantes disposent d'un arsenal de peptides pour se défendre face aux maladies et prédateurs.

2) Les peptides de défense (antimicrobiens et insecticides)

Les peptides de défense des plantes sont très nombreux et variés, ici nous nous limitons à présenter les principales familles structurales de peptides insecticides et antimicrobiens aux modes d'action divers. De taille maximale d'environ 10 kDa, ils sont sans activité enzymatique, en accord avec la terminologie utilisée pour les peptides antimicrobiens d'animaux (Boman, 1995), et de structures bien définies souvent stabilisées par des ponts disulfures.

(a) Principales familles structurales

Les peptides de défense des plantes présentés ici ont en commun d'être riches en cystéine formant des ponts disulfures qui les rendent compacts et stables. Leurs structures primaires et leurs activités biologiques varient beaucoup d'une famille à l'autre et même au

sein d'une même famille, mais leur structure tridimensionnelle est très conservée, définissant la famille. On distingue parmi ces familles les thionines, les défensines, les protéines de transfert de lipides et les knottines.

a) Les thionines

Les thionines, appelées ainsi à cause de leur riche teneur en soufre (Balls et al., 1942), comprennent les purothionines, premières d'entres elles découvertes dans les grains de blé (Balls et al., 1942), les α et β hordothionines de l'orge et de l'avoine (Gausing, 1987), les viscotoxines responsables, avec la lectine viscumine, de la toxicité du gui (Samuelsson, 1974) et d'autres peptides extraits de diverses espèces végétales. Les thionines sont des peptides de 45-47 acides aminés avec trois ou quatre ponts disulfure et généralement cationiques. Leur structure tridimensionnelle est bien conservée malgré des différences de séquence et de nombre de ponts disulfure. Elles ressemblent à la lettre grecque tau Γ avec deux hélices alpha antiparallèles formant le bras long et un feuillet béta de deux brins antiparallèles, constituant le bras court. Elles sont amphiphiles, leurs résidus hydrophobes étant regroupés sur la surface externe du bras long (Clore et al., 1986 ; Hendrickson and Teeter, 1981).
Selon leur séquence, cinq groupes peuvent être définis mais au niveau structural, trois groupes ont émergé sur la base du nombre et de l'arrangement des ponts disulfure. Le premier groupe contient les puro- et hordothionines ainsi que les peptides de *Pyrularia pubera* (Vernon et al., 1985) qui présentent quatre ponts disulfure en 1-8, 2-7, 3-6, 4-5. Le groupe 2 rapprochent les viscotoxines, les phoratoxines de *Phoradendron tomatum* et les crambines (neutres) de *Crambe abyssinica* qui possèdent toutes trois ponts disulfure connectés en 1-6, 2-5, 3-4. Enfin, le dernier groupe est formé de peptides tronqués des neuf acides aminés C-terminaux (Garcia Olmedo et al., 1989) présentant la connectivité emboitée typique des thionines pour leurs trois ponts. Le consensus Prosite correspondant aux thionines est C-C-X5-(R,K)-X2-(F,Y)-X2-C.
Les thionines montrent diverses activités toxiques sur différents systèmes biologiques (Florack and Stiekema, 1994). Peu après la découverte des purothionines, leurs propriété bactéricide et fongicide étaient caractérisées (Stuart and Harris, 1942). Les thionines de type I et II sont toxiques pour divers mammifères lorsqu'injectées mais pas lorsqu'administrées oralement (Coulson et al., 1942 ; Samuelsson, 1974). Kramer *et al* (1979) ont montré la toxicité de ces thionines par injection sur des larves de *Manduca sexta*. La purothionine de l'orge, incorporée à l'aliment, s'avère toxique pour le puceron (Rahbé Y., communication personnelle). En plus de leur toxicité sur de nombreux organismes, les thionines sont cytotoxiques pour les cultures cellulaires de mammifères ou d'insectes. La toxicité des thionines peut d'ailleurs être mesurée par lyse de cellules de moustiques et de *Choristoneura fumiferana* (Jones et al., 1985).
Elles sont exprimées sous forme de précurseurs avec un peptide signal suivi de la protéine majeure puis d'un domaine C-terminal acide qui jouerait le rôle d'un neutralisateur acide de la toxicité du domaine basique effecteur pendant la synthèse protéique (Bohlmann and Apel, 1991).
Elles peuvent être induites lors d'une infection bactérienne (Bohlmann et al., 1988 ; Epple et al., 1995) et des plantes transgéniques exprimant ces peptides se sont avérées plus résistantes à certaines infections que les plantes non transformées (Carmona et al., 1993 ; Epple et al., 1997).

b) Les défensines.

Les défensines des plantes sont des peptides cationiques de 45-54 acides aminés (environ 6 kDa) apparemment ubiquistes chez les Angiospermes et récemment identifiées chez les Gymnospermes (Fossdal et al., 2003). Autrefois appelées γ-thionines, elles représentent toutefois un groupe structural distinct des thionines. De structure tridimensionnelle proche de celles des défensines d'insectes, elles ont été renommées défensines de plantes (Terras et al., 1995). Elles possèdent 8 cystéines formant quatre ponts disulfures connectés en 1-8, 2-5, 3-6 et 4-7, bien qu'un cinquième pont puisse être présent comme dans les défensines de fleurs de *Petunia hybrida* récemment caractérisées (Janssen et al., 2003 ; Lay et al., 2003). Peu d'acides aminés autre que les cystéines sont conservés parmi les défensines. On distingue différentes classes selon leur structure primaire et leur activité (Osborn et al., 1995). Toutefois, malgré ces différences, elles partagent la même structure tridimensionnelle composée d'un feuillet β à trois brins antiparallèles et d'une hélice α (Almeida et al., 2002 ; Fant et al., 1998 ; Fant et al., 1999). Cette structure contient un motif Csαβ (cystein stabilized α helix - β sheet) qui correspond à la connection de deux ponts disulfure entre la séquence CxxxC de l'hélice et la séquence CxC du brin C-terminal. Ce motif est commun à certaines neurotoxines de scorpions inhibitrices de canaux ioniques (Kobayashi et al., 1991), à la protéine sucrée brazzéine (Caldwell et al., 1998), ainsi qu'aux défensines d'insectes (Cornet et al., 1995).

Globalement, les défensines possèdent des activités biologiques diverses et peuvent être divisées en quatre groupes d'après leur activité antimicrobienne. Le groupe I inhibe les bactéries gram+ et les champignons entraînant un branchement particulier des hyphes (groupe morphogénique). Le deuxième groupe inhibe uniquement les champignons, sans changement de forme des hyphes (non morphogénique). Le groupe III est actif contre gram+ et gram-. Enfin, les peptides du groupes IV sont actifs sur les deux types bactériens et les champignons (action antifongique non morphogénique). Il existe également une classe de défensine inhibant les amylases ou protéases. Enfin, certaines défensines ont montré une activité inhibitrice sur la synthèse protéique (voir paragraphe (b)e)).

L'analyse de nombreuses séquences de défensines connues à ce jour montre que la distinction en groupes basée sur l'activité ne se reflète pas au niveau des séquences (Thomma et al., 2002). Les gènes de défensines codent pour une préprotéine avec un peptide signal permettant l'adressage des peptides. Dans le cas des défensines de fleurs de pétunia et de tabac, ils spécifient en plus une partie C-terminale d'une trentaine de résidus, clivée lors de la maturation, et dont le rôle reste à déterminer (Lay et al., 2003). Les défensines sont une famille multigénique de protéines transcrites avec un peptide signal. Treize gènes putatifs ont été identifiés chez *A. thaliana* qui se regroupent en deux sous-familles (Thomma et al., 2002). L'homologie élevée observée dans la première famille suggère que les duplications et la diversification subséquente de ces peptides s'est effectuée relativement récemment du point de vue évolutif. Les membres de la deuxième famille sont plus divergents indiquant une duplication précédant la diversification des membres de la première famille. Deux autres gènes d'*Arabidopsis thaliana* spécifient une protéine fusion ou précurseur avec un peptide signal, un domaine inconnu et en C-terminal, un domaine défensine (Thomma et al., 2002). Les différents gènes d'*Arabidopsis* possèdent des profils d'expression complémentaires spécifiques de différents organes. L'un d'entre eux est induit dans les feuilles lors d'une attaque par un pathogène, ce qui confirme le rôle des défensines dans la défense des végétaux (Thomma et al., 2002).

c) Les knottines

Le motif cystine-knot, d'abord caractérisé en 1993 dans certains facteurs de croissance animaux (McDonald and Hendrickson, 1993) se retrouve dans de nombreux peptides de différentes origines (plantes, champignons, insectes, venins animaux). Toutefois, la structure de ces peptides ne se superpose pas parfaitement à celle des facteurs de croissance et on les distingue, en référence à leur activité inhibitrice, par le terme d'inhibiteurs à cystine-knot (ICK) ou knottines (Pallaghy et al., 1994). Les ICK animaux auraient une origine commune différente de celles des ICK végétaux et fongiques (Zhu et al., 2003). Toutefois, ces peptides possèdent tous de 26 à 48 résidus dont 6 cystéines formant trois ponts disulfure liés en 1-4, 2-5, 3-6. Le motif cystine-knot est formé d'un feuillet β à trois brins antiparallèles et le nœud consiste en un anneau formé par les ponts 1-4 et 2-5 et le squelette du peptide à travers lequel passe le troisième pont. Certains de ces peptides sont cycliques, généralement de moins de trente résidus, et forment la sous-famille des cyclotides, dont kalata1B d'*Oldenlendia affinis* est le premier représentant dont la configuration des ponts disulfures a été démontrée (Goransson and Craik, 2003).
Le motif cystine-knot confère à ces peptides une structure robuste et compacte. Ce sont souvent des inhibiteurs de nombreuses protéines. Les knottines animales bloquent souvent des canaux tandis que les peptides végétaux agissent sur des enzymes. De nombreuses knottines de plante montrent des activités antibactériennes et antifongiques. Récemment, l'activité insecticide de deux knottines a été caractérisée. Il s'agit de l'albumine PA1b du pois, toxique pour les charançons des céréales (Delobel et al., 1998), et du cyclotide kalata1B, décrit comme antimicrobien et qui réduit la croissance des larves du lépidoptère *Helicoverpa punctigera* (Jennings et al., 2001).

d) Les LTP

Les protéines non spécifiques de transfert de lipides (nsLTP) des plantes ont été identifiées dans de nombreuses espèces (environ 60) dont le maïs, l'épinard, le blé, la carotte. On distingue deux sous-familles de LTP. Celles de type 1, basiques, comportent 90 à 95 acides aminés et ont un point isoélectrique de 8,8 à 10. Les LTP2 ne sont constituées que d'environ 70 résidus (pour revue : Kader, 1996).
Malgré une grande différence de structure primaire, les LTP1 ont une structure tertiaire commune déterminée par des analyses en RMN ou par cristallographie et rayons X sur les LTP1 du maïs, du blé ou du riz (Charvolin et al., 1999 ; Gincel et al., 1994 ; Poznanski et al., 1999 ; Shin et al., 1995). Les LTP1 comprennent 8 cystéines très conservées formant quatre ponts disulfures connectés en 1-6, 2-3, 4-7 et 5-8. Ces ponts stabilisent une structure à 4 hélices α et une queue C-terminale flexible. Les LTP1 forment une poche interne hydrophobe dans laquelle s'insère la chaîne aliphatique des phospholipides qu'elles peuvent transporter (Kader, 1996). Des analyses structurales de la LTP1 du blé suggèrent la présence de deux sites de liaison aux lipides chez les LTP1 (Charvolin et al., 1999). Une récente analyse de mutants de la LTP1 du blé montre l'importance des ponts disulfures pour la présence de la cavité fixatrice de lipides (Sy et al., 2003). La structure tridimensionnelle des LTP2 n'a été élucidée que récemment par RMN sur celle du riz (Samuel et al., 2002). Si les 8 cystéines et la majorité des résidus hydrophobes sont bien conservés entre les LTP1 et les LTP2, la connectivité des ponts disulfures diffère. La LTP2 du riz présente les ponts 1-5, 2-3, 4-7 et 6-8 et ne comporte que trois hélices α dont les deux premières sont bien alignées avec celles des LTP1. La forme de la poche hydrophobe des deux types de LTP diffère également, ce qui est probablement important

pour leur rôle biologique. Elles ne fixent pas les même types de lipides (monomères de cutine ou de subérine pour les type 1 et les type 2 respectivement).
Bien que la fonction biologique des nsLTP soit encore sujet de discussions, les deux types de nsLTP sont considérés comme peptides de défense des plantes de par leur expression induite par des stress tant biotiques qu'abiotiques (Dunn et al., 1991 ; Gomes et al., 2003 ; Molina et al., 1993 ; Treviño and O' Connell, 1998 ; Yubero-Serrano et al., 2003), et par l'activité antibactérienne et antifongique de nombre d'entre elles (Broekaert et al., 1997 ; Garcia Olmedo et al., 1995 ; Kader, 1996), ainsi que par leur participation à la formation des couches de surface protectrices (cutine, subérine) (Douliez et al., 2000). De plus, elles possèderaient un rôle dans l'embryogénèse somatique (Coutos-Thevenot et al., 1993). Les LTP sont transcrites sous forme de préprotéine avec un peptide signal. Elles sont majoritairement excrétées même si certaines peuvent se trouver dans certains compartiments du cytoplasme (Kader, 1996). Elles constituent une famille multigénique (plus de 40 gènes chez *A. thaliana*).

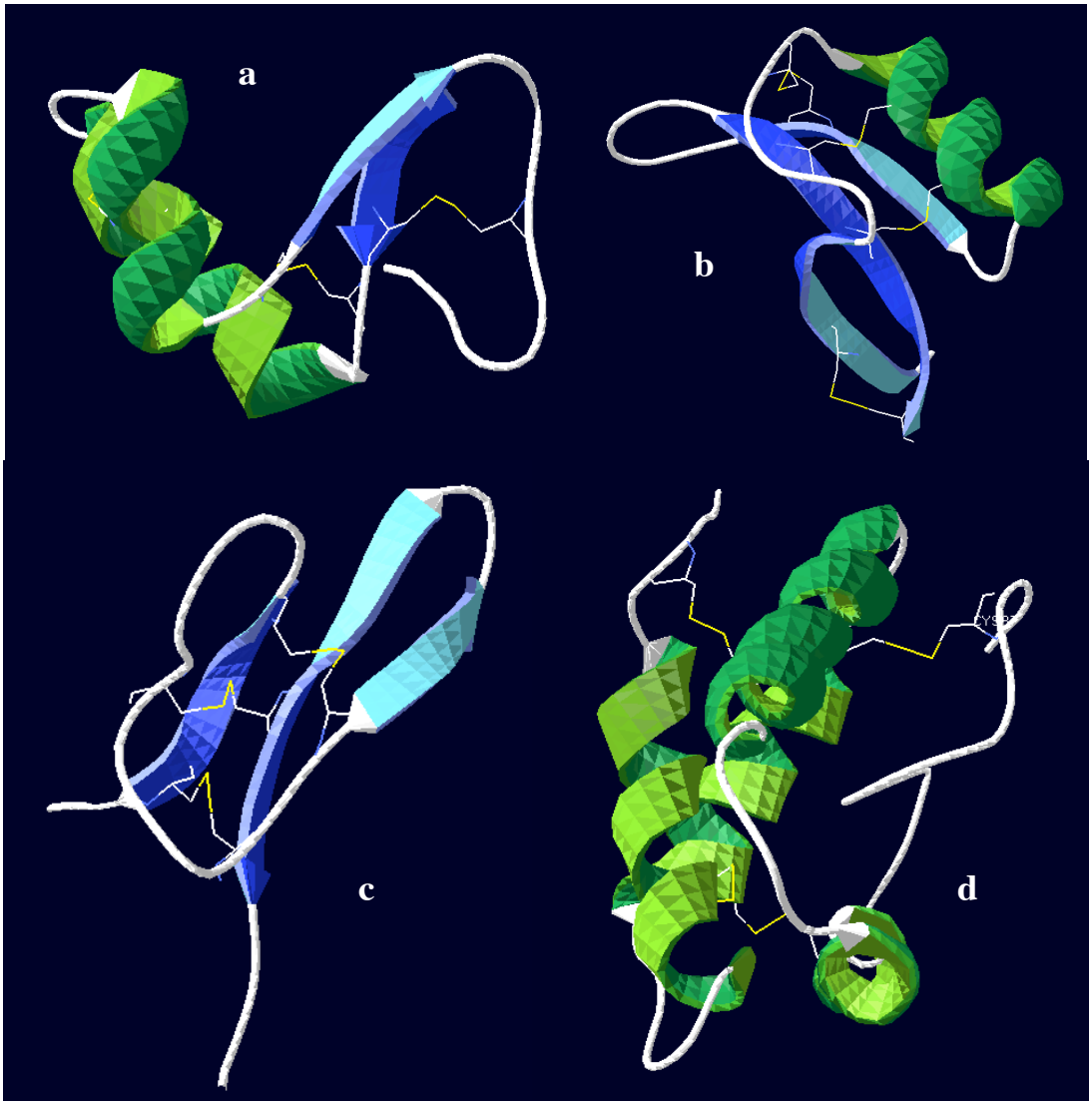


Figure 35 : structures tridimensionnelles de la thionine viscotoxine A3 du gui (a) (Romagnoli et al., To be published), de la défensine Rs-AFP1 du radis (b) (Fant et al., 1998), de la leginsuline du soja (c) (Yamazaki et al., 2003), de la nsLTP1 du blé (d) (Gincel et al., 1994). En trait fin jaune : pont disulfure.

(b) Cibles et mode d'action des peptides végétaux de défense

Les connaissances sur les modes d'action et les cibles des différentes classes de peptides sont encore imparfaites. Certaines familles structurales possèdent des peptides aux activités et cibles variées. Nous présentons ici les connaissances sur les différentes cibles moléculaires possibles de chaque type de peptides étudiés.

a) Phospholipides membranaires

Les propriétés amphiphiles de certains de ces peptides (thionines, PA1b) font des structures membranaires des cibles potentielles.

En effet, la première étape de l'action des **thionines** avec des microbes ou des cellules en cultures consiste en l'interaction électrostatique des peptides positivement chargés avec les phospholipides membranaires négativement chargés. Cette interaction induirait l'agrégation de liposomes et la fuite de leur contenu (Caaveiro et al., 1997). Les thionines ne s'insèrent pas dans la membrane des microorganismes pour former des canaux ioniques, mais désorganiseraient la membrane en se fixant à sa surface (Thevissen et al., 1996 ; Wall et al., 1995), induisant une perméabilisation de celle-ci qui entraînerait l'effet toxique (Thevissen et al., 1999). Des études sur la thionine de *Pyrularia* montrent l'implication d'un site de liaison spécifique sur la membrane d'érythrocytes de mammifères (Osorio e Castro and Vernon, 1989). Ces sites de liaisons nombreux et non protéiques sont probablement des domaines spécifiques de phospholipides comme les phosphoinositides qui transmettrait un signal suite à la liaison d'une thionine. Ce signal résulterait en une série de processus dont l'activation de canaux calcium (Evans et al., 1989).

Contrairement aux défensines de mammifères ou d'insectes (Cociancich et al., 1993 ; Kagan et al., 1990), les **défensines** antimicrobiennes de plante n'interagissent pas directement avec les lipides membranaires pour former des pores multimériques dans des membranes artificielles (Caaveiro et al., 1997 ; Thevissen et al., 1996). Sur des hyphes fongiques, les défensines de plante induisent divers flux ioniques. Elles ont un effet biphasique, dépendant de leur concentration, sur la perméabilisation de leurs membranes (Thevissen et al., 1996). Aux faibles concentrations corrélées à l'inhibition de croissance des champignons, une perméabilisation faible mais indépendante de la force ionique a lieu. Cette perméabilisation est plus spécifique que celle observée à forte dose et pourrait impliquer un site de liaison membranaire qui transmettrait un signal aux canaux ioniques endogènes ou faciliterait l'insertion des défensines dans la membrane fongique, entraînant alors la formation de canaux (Thevissen et al., 1996). En effet, un site de liaison spécifique à haute affinité a été caractérisé sur le plasmalemme de *Neurosporra crassa* pour la défensine Hs-AFP1 d'*Heuchera sanguinea* (Thevissen et al., 1997). La liaison de la défensine sur son site semble liée à l'activité antifongique du peptide. De plus, la liaison d'Hs-AFP1 marqué n'étant réversible qu'en partie par l'ajout d'un excès de ce même peptide non marqué, il est possible que le site de liaison intervienne comme point d'ancrage permettant aux défensines de s'insérer dans la membrane pour former des canaux (Thevissen et al., 1997). De même, un site de liaison spécifique a été caractérisé pour la défensine Dm-AMP1 de *Dahlia merckii* sur *N. crassa* et *Saccharomyces cerevisiae*. La liaison de Dm-AMP1, apparemment irréversible, peut être inhibée par pré-incubation avec d'autres défensines non morphogéniques mais pas avec des défensines plus éloignées ni avec d'autres peptides antimicrobiens riches en cystéines. Deux mutants de la levure résistants à Dm-AMP1 présentent une affinité de liaison très réduite par rapport aux levures sensibles sauvages, suggérant que la liaison de la défensine est

nécessaire à son activité (Thevissen et al., 2000b). Ces souches de levure résistantes sont mutées dans le gène IPT1 codant pour une enzyme impliquée dans la dernière étape de la synthèse du sphingolipide mannose-(inositol-phosphate)-2-céramide (Thevissen et al., 2000a). Des travaux récents montrent que DmAMP1 interagit directement avec les sphingolipides de la levure et que cette interaction est stimulée par la présence d'ergostérol (un stérol spécifique des champignons) (Thevissen et al., 2003). Les défensines morphogéniques comme Hs-AMP1 ont probablement un site de liaison différent, l'absence de l'enzyme IPT1 n'entraînant aucune résistance à ces peptides (Thomma et al., 2002). Malgré la similarité des réponses qu'elles induisent, les défensines antifongiques auraient donc des modes d'action différents.

Les **protéines de transfert des lipides** agissent probablement aussi par l'intermédiaire des lipides membranaires. La LTP2 de l'orge agit, au moins partiellement, par la déstabilisation de la perméabilité de la barrière membranaire (Caaveiro et al., 1997). Un homologue des nsLTP, Ace-AMP1, a été extrait des graines d'oignions et s'avère le membre de cette famille le plus actif envers les champignons, sa concentration minimale pour une inhibition de 50% (IC_{50}) du développement du champignon étant de 0,1 à 1 mM comparée à celle de la nsLTP du radis, supérieure à 2 mM, et à l'absence d'activité antifongique pour les LTP du blé et du maïs (Cammue et al., 1995). Ace-AMP1 n'est toutefois pas capable de s'associer à des lipides en solution contrairement aux autres membres de la famille, certainement du fait qu'il ne possède pas comme eux une poche hydrophobe le traversant. Cependant, Ace-AMP1 interagit avec les phospholipides membranaires, pénétrant lentement au cœur hydrophobe de bicouches lipidiques, même s'il ne désorganise et ne perméabilise guère les membranes. Son activité antifongique ne serait donc probablement pas le résultat d'une interaction directe avec les phospholipides membranaires mais plutôt d'une interaction protéine-lipide médiée par un récepteur membranaire (Tassin et al., 1998).

b) Protéines membranaires

Les **LTP**, affines pour divers lipides, semblent interagir avec les membranes mais leur activité antimicrobienne pourrait également être médiée par un récepteur protéique membranaire. En effet, Buhot et al (2001) ont montré récemment que la LTP du blé se liait aux membranes plasmiques des végétaux par un site de liaison spécifique probablement tetramérique. Ce site de liaison s'avère commun à l'éliciteur cryptogéine. La LTP n'induit pas les réponses de la cryptogéine et pourrait agir comme antagoniste de cet éliciteur. Récemment, un modèle a été proposé par Blein et al. (2002) pour le rôle des LTP dans les interactions plantes-pathogènes sans éliciteurs. Les LTP constitutives fixeraient des monomères de cutine dégradée par un champignon phytopathogène. Le complexe LTP-monomère se lierait au récepteur membranaire de la plante entraînant la réponse de défense de celle-ci par production de composés antifongiques dont des nsLTP induites. Les LTP aideraient, en parallèle, à reformer la cutine dégradée. Enfin, le complexe monomère-LTP pourrait également se lier à un récepteur du champignon, encore inconnu, et entrainer une surproduction de cutinase chez celui-ci. Un déséquilibre de ces interactions entraînerait la mort de l'un des deux partenaires.

La **knottine** PA1b, albumine 1 sous-unité b du pois, est responsable de la toxicité des graines de cette plante pour les charançons des céréales (*Sitophilus spp*.). Il existe toutefois des souches de l'espèce *S. oryzae* contenant des individus résistants à ce peptide. Cette résistance a été étudiée et s'avère monogénique et récessive (Grenier et al., 1997). PA1b possède un site de liaison spécifique de forte affinité dans la fraction microsomale obtenue à partir de charançons sensibles broyés. Ce site, protéique, est absent des

charançons résistants (Gressent et al., 2003). Ceci indique que l'activité insecticide de ce peptide pourrait être liée à la fixation de celui-ci sur son site de liaison membranaire (actuellement en cours de purification). Il reste toutefois à comprendre comment cette fixation au site de liaison induit la mort prématurée des cellules du tube digestif, observée après ingestion de PA1b par les charançons sensibles (Vaublanc, 2001), et à vérifier si cette dégénérescence des cellules de l'épithélium digestif est bien responsable de la mort des insectes intoxiqués.

c) Canaux

Au vu de la structure tridimensionnelle des **défensines**, proche de celle de toxines bloquant des canaux ioniques, ces derniers sont également des cibles potentielles des peptides de plantes.
En effet, les défensines du maïs γ-zeathionine 1 et 2 bloquent de façon réversible la conductance des canaux à sodium voltage-dépendants d'une lignée cellulaire tumorale de rat (Kushmerick et al., 1998). Comme la structure des canaux ioniques est conservée parmi les eucaryotes (Hille, 1992), ces défensines pourraient être dirigées contre des canaux d'insectes ou de champignons phytopathogènes (Kushmerick et al., 1998). Leur activité envers ces ravageurs reste néanmoins à déterminer. Une autre défensine, antifongique, du pois, possèderait également une activité inhibitrice de canaux à potassium. Sa structure est en effet similaire aux autres défensines et neurotoxines. De plus, son potentiel électrostatique de surface présente des caractéristiques communes aux inhibiteurs de canaux K+, dont un résidu lysine essentiel pour l'activité inhibitrice. En revanche, ce potentiel ne présente pas de similitude avec d'autres protéines antifongiques, ni avec les inhibiteurs de canaux à sodium (Almeida et al., 2002).

d) Enzymes

Comme nous l'avons mentionné précédemment (1)(c)), les enzymes comme les α-amylases et les protéases des ravageurs sont des cibles importantes pour la défense des végétaux. Certains peptides présentent des activités inhibitrices envers ces types d'enzymes.

(i) Inhibiteurs d'α-amylases (IA)

Les deux familles d'IA peptidiques sont constituées de peptides des familles structurales knottines et défensines.

L'inhibiteur d'amylase AAI, des graines d'*Amaranthus hypocondriacus*, appartient à la famille des **knottines**. Il inhibe spécifiquement les α-amylases des larves d'insectes (*Tribolium castaneum, Prostephanus truncatus*) et est inactif sur les amylases de mammifères ce qui en fait un bon candidat pour le développement de plantes transgéniques résistantes (Chagolla-Lopez et al., 1994). Ce peptide agit par blocage du site catalytique de l'enzyme. La structure du complexe AAI-TMA (amylase de *T. molitor*) montre que ce peptide pénètre parfaitement dans la crevasse du site actif de l'enzyme, l'interface du complexe comprenant essentiellement des interactions polaires. Le résidu catalytique Asp287 de la TMA forme un pont salin avec l'Arg7 d'AAI et les autres résidus du site actif sont connectés aux groupes fonctionnels des résidus de l'inhibiteur via un réseau de liaisons hydrogènes médiées par des molécules d'eau (Pereira et al., 1999).

Les **défensines** Siα1 à 3 ont été extraites des graines de *Sorghum bicolor*. Elles possèdent une forte activité inhibitrice sur les α-amylases du tube digestif de sauterelles et

de blattes, une faible activité inhibitrice sur les amylases d'*Aspergillus oryzae* et de la salive humaine, et pas d'activité sur les amylases de l'orge, des bactéries du genre *Bacillus* et sur l'amylase pancréatique du porc (Bloch and Richardson, 1991). Leur spécificité les rend, elles aussi, attractives dans le cadre de la production de plantes résistantes aux insectes mais leur site actif et leur mode d'action précis reste à déterminer.

(ii) Inhibiteurs de protéases (IP)

Parmi les **knottines**, on distingue deux classes d'inhibiteurs de protéases. L'inhibiteur de carboxypeptidase de la pomme de terre (CPI) constitue la première classe. Il agit par sa partie C-terminale, sa glycine terminale étant insérée dans la poche de liaison de la carboxypeptidase A (Rees and Liscomp, 1982). Les inhibiteurs de protéases à serine des courges comme le CPTI-II de *Cucurbita pepo* et l'EETI d'*Ecballium elaterium* constituent la deuxième classe. Ils agissent comme des inhibiteurs canoniques, leur région N-terminale formant la boucle inhibitrice dont le résidu P1 est une arginine ou une lysine (Bode et al., 1989 ; Helland et al., 1999 ; Le Nguyen et al., 1990). Récemment, un membre des cyclotides extrait des graines de *Momordica cochinchinensis*, McoTI-II, a révélé une activité antitrypsique (Heitz et al., 2001). Sa conformation est similaire à celle des inhibiteurs de courge non cycliques.

Parmi les **défensines** inhibitrices de protéases, se trouve l'inhibiteur extrait de *Vigna unguiculata*, actif sur la trypsine mais pas sur la chymotrypsine. Dans le modèle d'inhibition proposé, son résidu Lys11 occupe la poche active de la trypsine dans une interaction canonique (Melo et al., 2002). Un peptide de 5,4 kDa extrait de *Cassia fistula* et présentant une homologie de séquence avec les défensines se révèle un inhibiteur correct de la trypsine (IC_{50} : 2μM). De même que pour le peptide du niébé, c'est son acide aminé Lys25 qui est proposé comme résidu P1 (Wijaya et al., 2000). Récemment, la structure d'un troisième peptide, NaD1, a été déterminée. Ce peptide antifongique et insecticide est une défensine des fleurs de *Nicotiana alata*. NaD1 inhibe la croissance de champignons phytopathogènes *in vitro* et retarde le développement des lépidoptères ravageurs *Helicoverpa armigera* et *H. punctigera* lorsqu'il est incorporé dans un milieu artificiel ou exprimé dans une plante transgénique. La capacité de NaD1 à inhiber la trypsine et la chymotrypsine, enzymes digestives majeures d'*Helicoverpa*, pourrait expliquer ses effets sur les larves (Lay et al., 2003). Cependant, son mode d'action sur les insectes et les champignons reste à préciser.

Les protéases et les amylases ne sont pas les seules cibles enzymatiques possibles de ces peptides de défense des plantes. En effet, les thionines sont également connues pour inhiber certaines enzymes comme les β-glucuronidases (Diaz et al., 1992). Toutefois, cette activité serait due à leur activité thioredoxine, probablement par la formation de ponts disulfures intermoléculaires (Pineiro et al., 1995) et non par un mécanisme d'inhibition classique de blocage du site actif de l'enzyme.

e) Cibles intracellulaires

Les **thionines** inhibent la synthèse protéique dans un système eucaryote acellulaire, soit par liaison aux ARN chargés négativement (Garcia Olmedo et al., 1983), soit directement au niveau de l'initiation de la traduction en inactivant les enzymes par réduction de leurs ponts disulfures (Brummer et al., 1994). Une comparaison de la structure du motif hélice-coude-hélice des thionines avec celle des protéines de liaison à l'ADN amène Li et al. (2002) à considérer les thionines comme un nouveau groupe

potentiel de protéines se liant à l'ADN. L'impact des thionines sur cette cible nucléique reste à confirmer.

De même que pour les purothionines, il a été montré que les γ- et ω-hordothionines, des **défensines** de l'endosperme de l'orge, inhibent la synthèse protéique dans des systèmes acellulaire eucaryotes mais aussi procaryotes (Mendez et al., 1990 ; Mendez et al., 1996). La défensine Hv-AMP1 de la légumineuse *Hardenbergia violacea* possède également une activité inhibitrice de la synthèse protéique *in vitro* (Harrison et al., 1997). Ce mode d'action supporte l'hypothèse de cibles intracellulaires des défensines de plante mais l'activité antimicrobienne de ces peptides reste à démontrer. Récemment, une nouvelle défensine de *Vigna radiata* a été caractérisée pour son activité entomotoxique contre les bruches. Des graines artificielles contenant 0,2% (p/p) du peptide recombinant VrCRP-TSP s'avèrent mortelles pour les larves de *Callosobruchus chinensis*. La base moléculaire de cette activité n'est pas encore connue mais ce peptide est lui aussi un inhibiteur de la synthèse des protéines *in vitro* (Chen et al., 2002).

Ainsi les classes structurales des peptides de défense des plantes ne correspondent pas à des modes d'action uniques et distincts. Chaque famille présente diverses activités et vraisemblablement diverses cibles, bien que des travaux soient encore nécessaires pour connaître précisément les modes d'action de la plupart de ces peptides. La présence de diverses cibles pour un même type de peptides est particulièrement intéressante dans l'optique de production de plantes transgéniques résistantes : il est plus difficile pour le ravageur de mettre en place une résistance multiple. Toutefois, le fait que les différentes familles de peptides se recoupent au niveau de leurs activités pourrait limiter leur utilisation, dans le cas où les cibles moléculaires seraient identiques, par l'apparition de résistances croisées.

C. Conclusion

Les plantes disposent de nombreux systèmes de défense basés sur des molécules de natures diverses qui agissent souvent en synergie.
L'**homologie** de structure observée entre différents composés comme les lectines et les αAI du haricot, laisse penser que ces molécules de défense dérivent d'un petit nombre de gènes qui se sont multipliés et diversifiés sous la pression des phytophages. Fraenkel (1959) a été le premier à suggérer que les métabolites secondaires des plantes ont évolués comme défense contre les insectes phytophages, et donc que ceux-ci étaient d'importants agents de sélection pour les plantes. Cette idée est au cœur de la théorie de **coévolution** plante-insecte (Ehrlich and Raven, 1964) qui propose qu'au fur et à mesure que les plantes développent de nouveaux composés de défense, les herbivores s'y adaptent. Nombreux sont les insectes **contournant certaines défenses** des plantes, soit en excrétant simplement les composés toxiques, soit en les transformant en sous-produits non toxiques voire en détournant ces composés toxiques pour se protéger eux-mêmes de leurs prédateurs (séquestration). La deuxième supposition de cette coévolution, c'est-à-dire la sélection de la résistance de la plante par l'herbivore est moins évidente. Tandis que la composition chimique de la plante est probablement un facteur majeur déterminant le choix de la plante hôte par l'insecte, les attaques des insectes ne sont qu'un facteur parmi tant d'autres qui puisse affecter la santé de la plante.

Malgré un arsenal de molécules de défense, protéiques ou non, les espèces végétales sont très généralement toujours sensibles aux attaques d'un certain nombre de pathogènes ou de prédateurs. Lorsque des plantes cultivées d'importance agroalimentaire mondiale sont sensibles à des ravageurs causant de gros dégâts, la recherche de nouveaux moyens de lutte s'impose. Elle peut se faire par synthèse chimique, par recherche de prédateurs ou pathogènes naturels du ravageur ou par détection de nouveaux composés toxiques naturellement présents dans une espèce donnée végétale ou non. C'est dans cette dernière perspective que s'insère le projet dont mon travail de thèse fait partie.

II. Contexte de l'étude et présentation des travaux

Chaque année, environ un quart des céréales récoltées sur terre est détruit par les insectes pendant le stockage (Delobel, 1999). Pour lutter contre ce fléau mondial, de nombreux moyens existent pourtant :

- la **lutte biologique** consiste en l'utilisation d'organismes nuisibles aux ravageurs des denrées : virus, bactéries et champignons entomopathogènes, insectes prédateurs ou parasitoïdes existent naturellement et leur utilisation est prometteuse bien qu'encore peu développée.
- la **lutte physique** repose sur des traitements thermiques, par le froid ou la chaleur, des actions mécaniques comme le tamisage, l'aspiration, les percussions, l'utilisation de rayons ionisants ou encore le maintien sous atmosphère contrôlée. Malheureusement, ces méthodes restent souvent trop coûteuses et difficiles à mettre en œuvre techniquement, surtout pour les pays en voie de développement où les dégâts sont les plus importants.
- la **lutte chimique**, qui s'avère bien souvent efficace, a généré une utilisation intensive de produits chimiques entraînant à terme l'apparition d'insectes résistants, et pouvant également nuire à l'environnement et à la santé humaine et animale par la présence de résidus.
- la **lutte variétale** reste limitée en ce qui concerne les denrées stockées, car on dispose de peu de variétés naturellement résistantes. Toutefois est-il possible, grâce à la sélection variétale ou à la transformation génétique, de rendre les plantes résistantes aux ravageurs par introduction d'un (ou plusieurs) gène(s) d'une plante résistante de la même espèce ou d'une autre espèce, voire d'un organisme entomopathogène. Encore faut-il disposer de tels gènes, acceptables à la fois pour l'environnement et pour le consommateur.
- la **lutte intégrée** fait appel à plusieurs de ces méthodes qui agissent alors en synergie.

Les techniques de lutte chimique ou physique, bien que parfois très efficaces, présentent toutes des inconvénients et restent souvent insuffisantes. Avec le souci grandissant du respect de l'environnement et de la santé humaine et animale, les luttes biologiques et variétales sont donc appelées à se développer. En particulier, elles **ne permettent pas, pour l'instant, de lutter efficacement contre les charançons des céréales** (*Sitophilus oryzae*, *S. granarius, S. zeamais,* Coléoptères, Curculionidés), qui sont les principaux ravageurs des céréales stockées.

Il a été observé que les charançons des céréales peuvent également s'alimenter sur certaines plantes de familles aussi diverses que certaines Fabaceae, Chenopodiaceae, Fagaceae et Clusiaceae (Delobel and Tran, 1993), mais qu'ils meurent rapidement après alimentation sur pois, sauf pour quelques rares souches de *S. oryzae* (Vayssière, 1943). Une étude menée au laboratoire de Biologie Fonctionnelle, Insectes et Interactions (BF2I INRA-INSA de Lyon) en collaboration avec le Laboratoire de Biochimie et Technologie des Protéines de l'INRA de Nantes a permis d'isoler et de purifier le **facteur entomotoxique** responsable de cette mortalité. Celui-ci a fait l'objet d'un dépôt de brevet (Delobel et al., 1998). Il s'agit d'un **peptide** de type albumine 1, précédemment appelé

PA1b (Pea Albumin 1 sous-unité b) (Higgins et al., 1986). Il est constitué de 37 acides aminés dont 6 cystéines engagées dans 3 ponts disulfures intramoléculaires. Sa structure tridimensionnelle et la connectivité des ponts disulfures ont été déterminés par RMN (résonance magnétique nucléaire) du proton au laboratoire de Biophysique du CNRS d'Orléans. Les résultats montrent que ce peptide appartient à la famille structurale des **knottines**, peptides dont le troisième pont disulfure passe au travers de la boucle formée par les deux autres ponts et la chaîne peptidique, formant ainsi un nœud. Cette structure très compacte concorde avec l'observation de sa très grande stabilité. PA1b n'est en effet pas dénaturé par la stérilisation (Delobel et al., 1998), résiste longtemps à la digestion *in vitro* par du fluide de rumen (Spencer et al., 1988) ainsi qu'à la dégradation par la trypsine et la chymotrypsine (Hancock et al., 1994).

L'un des freins à l'utilisation de toxines, par voie chimique ou par transgénèse est l'apparition plus ou moins rapide de souches résistantes dans les populations de nuisibles visées. Ce problème est d'autant plus aigu en ce qui concerne l'éventuelle utilisation de PA1b. En effet, lors du criblage de 90 souches des trois espèces de charançons des céréales pour leur capacité à survivre sur pois, 4 souches, toutes de l'espèce *S. oryzae*, se sont avérées naturellement résistantes au principe actif du pois, bien que présentant un retard de développement sur graines de pois (Delobel and Grenier, 1993). La génétique de cette **résistance** a été étudiée et celle-ci s'avère être **monogénique, autosomale et récessive** (Grenier et al., 1997). Cette caractéristique tend à indiquer qu'une unique protéine du charançon est à l'origine du caractère sensible ou résistant des souches de *S. oryzae*. Ceci a conduit à supposer l'existence chez l'insecte d'un récepteur pour ce peptide végétal toxique. Cependant, aucune activité enzymatique ou inhibitrice d'enzyme chez ce peptide n'a été mise en évidence (Mbaiguinam, 1996) et le mode d'action de PA1b reste encore inconnu. D'autres insectes ont également été testés pour leur sensibilité à PA1b purifié. PA1b s'avère toxique pour les larves de premier et deuxième stade de la teigne de la farine (*Ephestia kuehniellea*) ainsi que pour le puceron du pois (*Acyrthosiphon pisum*) (Delobel et al., 1998), alors qu'il ne semble avoir aucun effet sur la mouche *Drosophila melanogaster* (Gressent, communication personnelle).

PA1b, représentant ainsi **une nouvelle classe d'entomotoxines**, extraite d'une graine largement consommée par l'homme et l'animal sans manifestation de la moindre toxicité ni allergénicité, est potentiellement d'un **grand intérêt pour l'élaboration de plantes transgéniques résistantes à divers insectes ravageurs**. Son gène a été cloné et l'incorporation de celui-ci dans le génome du trèfle a été réalisée, pour augmenter sa teneur en soufre (Ealing et al., 1994), tandis qu'elle est en cours pour le riz, pour évaluer son efficacité insecticide *in planta* (Thèse de J. Petit, CIRAD, Montpellier, France). Cependant, le principal obstacle à l'utilisation de PA1b reste l'existence d'insectes résistants natifs. Dans ce cadre, de nombreuses études sont entreprises sur PA1b, les buts principaux étant de comprendre le mode d'action de cette toxine et de caractériser des homologues toxiques pour des insectes actuellement résistants. Pour déterminer le mode d'action de PA1b, les recherches se sont orientées vers la caractérisation d'un **site de liaison** à cette toxine chez les charançons (dont l'existence est supposée suite aux résultats des expérimentations de génétique). Ce site de liaison, membranaire, a été caractérisé (Gressent et al., 2003) et est en cours de purification au laboratoire BF2I. Connaître le mécanisme d'action de PA1b devrait permettre de déterminer précisément les régions de ce peptide responsables de son activité. De façon complémentaire, l'expression en système hétérologue de gènes homologues à PA1b devrait aider, associée à la mutagénèse dirigée, à la connaissance du mode d'action de PA1b en reliant variations de structure et variations d'activité. Cette approche pourrait permettre d'obtenir des peptides d'activités variées et

particulièrement des peptides plus actifs que PA1b ou actifs sur insectes résistants. Une voie complémentaire d'obtention de tels peptides est la recherche d'homologues naturels. En effet, PA1b appartient à une famille multigénique. Le pois lui-même contient plusieurs isoformes qui présentent cependant toutes la même toxicité (CL_{50} identiques). Une de ces isoformes avait en fait déjà été isolée en 1986 (Higgins et al., 1986) (Figure 36). Étant assez fortement représentés dans la graine (1 % des protéines totales (Rahbé Y., 2003, communication personnelle)), les peptides de cette famille étaient considérés comme des protéines de réserve en soufre pour le pois. Mais le pois n'est certainement pas le seul à produire ces peptides. Des tests sur graines de diverses légumineuses ont montré la présence, dans la plupart de ces plantes, d'une toxicité analogue, définie par une mortalité des souches résistantes largement plus faible que celle des sensibles, voire nulle (Mbaiguinam, 1996). En outre, PA1b présente une forte homologie (65%) avec la leginsuline du soja (Watanabe et al., 1994), les cystéines (rouges), prolines et la plupart des glycines (bleues) se trouvent notamment conservées en position et espacement (Figure 36).

```
PsaA1b001              ASCNGVCSPFEMPPCGTSACRCIPVGLVVGYCRNPSG
                       ----------------------------c--------
PsaA1b005              ASCNGVCSPFEMPPCGTSACRCIPVGLVIGYCRNPSG
                       -n---c-----n---ncnn---c-c--nc-n-nn-c-
Leginsuline du soja    ADCNGACSPFEVPPCRSRDCRCVPIGLFVGFCIHPTG
```

Figure 36: Comparaison des séquences de l'isoforme de PA1b découverte par Higgins et al. (1986) (PsaA1b001), de celle isolée au laboratoire (Delobel et al., 1998) (PsaA1b005) et de la leginsuline du soja (Watanabe et al., 1994). c représente les remplacements conservatifs, n représente les remplacements non-conservatifs.

L'étude de la diversité de la famille de PA1b au sein des Légumineuses apparaît intéressante pour plusieurs raisons. Elle pourrait permettre de déterminer l'importance écologique de ces peptides (rôle de défense et autre rôle ?) et également d'obtenir des indications sur l'évolution de cette nouvelle famille, non identifiée jusqu'ici hors des légumineuses. Peut-être aussi cela apporterait-il des éléments supplémentaires à l'étude taxonomique des légumineuses, toujours en remaniement. De plus, la recherche de variants naturels de PA1b pourrait aider à la compréhension du mode d'action des peptides de type A1b en soulignant de potentiels acides aminés importants pour l'activité. Enfin, il sera peut-être possible de déterminer des homologues de CL_{50} inférieure à celle de PA1b, ou contournant la résistance des insectes insensibles à la toxine du pois.

Mon travail durant ces trois ans de thèse a consisté en la recherche d'homologues de PA1b au sein de différentes espèces de Légumineuses par la voie de la biologie moléculaire (gènes homologues) complémentée par l'analyse de la toxicité des graines des espèces choisies envers *S. oryzae,* souches sensible et résistante. Dans un deuxième temps, afin de relier les variations de séquences aux variations de toxicité obtenues par ces approches, un système d'expression hétérologue devait être mis au point afin de produire certains variants naturels et des mutants choisis de PA1b.

Chapitre 2 : Définition de l’approche : à la découverte de la famille des A1b.

Définition de l'approche : à la découverte de la famille des A1b

L'intérêt de PA1b en tant que nouvelle toxine insecticide, présenté dans la partie précédente, incite à l'étude de sa famille afin de mieux comprendre d'une part l'évolution de celle-ci et d'autre part le mécanisme d'action de ces peptides. Vu l'absence de séquences homologues dans le génome d'*Arabidopsis thaliana* et du riz, cette famille peptidique apparaît pour l'instant limitée aux Légumineuses. Pour une étude large de ces peptides au sein de cette grande famille végétale, il faut des outils efficaces et rapides.

- Les techniques de biologie moléculaire, en particulier l'étude de l'expression des gènes par suivi des ARN messagers (ARNm) semblent attractives. PA1b étant exprimé dans les graines du pois lors de leur maturation (Higgins et al., 1986), il conviendrait d'extraire les ARNm de ces organes végétaux. Malheureusement, comme de nombreuses Légumineuses tropicales sont des arbres, l'obtention de fruits en cours de maturation pour en extraire les ARNm n'est pas envisageable. Nous avons donc décider de travailler sur l'**ADN génomique**, en utilisant des amorces dégénérées.
- De même, la purification d'isoformes à partir des graines de nombreuses espèces serait idéale mais trop laborieuse et coûteuse en temps, c'est pourquoi nous avons essayer de mettre au point une méthode de détection des peptides basée essentiellement sur un **crible biologique spécifique**. Ce crible consiste en un fractionnement grossier des farines et la mesure de leur toxicité sur deux souches de *S. oryzae*, l'une sensible et l'autre résistante (Grenier et al., 1997). Pour confirmer la présence de peptides dans les extraits actifs, une approche par immunodétection, des essais d'inhibition de liaison au site de liaison spécifique de PA1b chez les charançons sensibles (Gressent et al., 2003), et une analyse par spectrométrie de masse ont été réalisés.

Le but de cette première partie est d'évaluer la faisabilité d'une étude des A1b au sein des Légumineuses par la mise au point des techniques nécessaires à une telle étude, en particulier celles spécifiques d'une approche moléculaire, jamais encore abordée sur cette famille. Pour cela, les différentes techniques ont été testées sur 3 Papilionoideae (sous-famille des Légumineuses), en plus du pois (*Pisum sativum* L., tribu des Viciae) témoin :

- le soja (*Glycine max* L., tribu des Phaseolae), pour lequel un homologue, la leginsuline, est déjà connu, mais dont l'activité insecticide n'a pas été étudiée.
- *Medicago truncatula* Gaertner (plante proche de la luzerne, tribu des Trifoliae), où seules des séquences de peptides racinaires ou nodulaires sont connues mais non reliées jusqu'ici à la famille des A1b (Gamas et al., 1996). Cette plante appartient à une tribu parfois considérée comme soeur des Viciae (Young et al., 2003).
- le haricot (*Phaseolus vulgaris* L., tribu des Phaseolae), pour lequel rien n'est connu, mais qui appartient à la même tribu que le soja.

I. Matériels et méthodes

A. Biologie : recherche de toxicité analogue

1) Les charançons

La seule espèce de charançons utilisée ici est *Sitophilus oryzae*, espèce très polyphage pouvant envahir diverses céréales, mais aussi certaines légumineuses pourtant réputées pour leur toxicité liée à la grande diversité des métabolites secondaires qu'elles synthétisent (Bisby et al., 1994 ; Harborne et al., 1971).
De nombreuses souches de cette espèce sont élevées à l'UMR BF2I, dans une enceinte ventilée d'humidité relative (HR) de 70 ± 5%, à 27,5 ± 0,5°C et à l'obscurité, selon le protocole décrit par Laviolette et Nardon (1963). De jeunes adultes matures sont déposés dans des boites en plastique rondes grillagées, de 11 cm de diamètre sur 8 cm de hauteur, remplies aux 2/3 de graines. Ils y restent 2 jours pour pondre puis sont enlevés et placés sur du grain neuf. Ils sont ensuite utilisés pour les expériences de toxicité jusqu'à l'âge de trois semaines. Les graines, qui viennent de recevoir les pontes, sont gardées 6 semaines, temps du développement préimaginal, pour donner la génération suivante d'insectes expérimentaux.
Les souches capables de se développer sur pois cassé, telles que **Chine** utilisée dans ce travail, sont élevées sur cette graine (*Pisum sativum* L.) afin de conserver le caractère résistant, aussi bien que sur blé (*Triticum aestivum* L.). Ces souches sont dites **résistantes** (**R**). Celles ne pouvant se développer sur pois sont élevées sur blé uniquement et sont dites **sensibles** (**S**). C'est le cas de la souche **Bénin** utilisée dans notre travail. L'étude génétique de la résistance à PA1b, monogénique autosomale et récessive, avait déjà été réalisée (Grenier et al., 1997).

Figure 37 : Imago de *Sitophilus oryzae* sur grain de maïs. (photo : C. S. Gorsuch, Clemson University)

2) Le matériel végétal

Les graines destinées à l'élevage (blé et pois cassé non traités) ont été placées à -80°C pendant une semaine pour tuer tout insecte pouvant s'y trouver, puis conservées à 4°C.
Pour la préparation des farines et l'obtention du matériel frais, en plus du pois, plante témoin (*Pisum sativum* L. cv Frisson) à 14 chromosomes (2n) fourni par G. Duc de l'INRA de Dijon, 3 autres légumineuses ont été étudiées. Il s'agit du soja (*Glycine max* L. cv Paoki), espèce diploïde à 40 chromosomes (2n), fourni par P. Sartre de L'INRA de Montpellier, du haricot (*Phaseolus vulgaris* L. cv Contender), espèce à 2n = 22 chromosomes, acheté dans un magasin de jardinage et d'une luzerne méditerranéenne (*M. truncatula* Gaertner cv Salernes), espèce diploïde à 2n = 16 chromosomes. Toutes ces plantes, obtenues sous forme de graines, ont été broyées en farine (pour les tests biologiques et le fractionnement) mais aussi plantées sous serre, afin d'obtenir environ 10 g de feuilles pour l'extraction d'ADN. Quelques plants de pois ont été conservés afin d'obtenir une floraison et la formation de graines, prélevées au cours de leur développement pour en extraire l'ARN.

3) Préparation et fractionnement des farines

Pour chacune des 4 plantes étudiées, environ 200 g de graines ont été broyées dans un moulin de type Warring Blendor pour obtenir de la farine, sassée au tamis AFNOR NF X 11-201 Prolabo de 0,4 mm (0,2 mm pour le blé) d'ouverture de maille : une grande partie des cuticules sont ainsi éliminées.
Les farines ont ensuite été fractionnées par épuisement en 4 étapes successives. Tout d'abord une extraction au pentane (500 mL pour 200 g de farine) a été réalisée : après agitation toute la nuit, le tout est filtré sur fritté (n°3). La fraction soluble (lipides) a été éliminée et la fraction insoluble a été mise à sécher sous hotte (jusqu'à poids constant). On procède de même (sur cette fraction sèche) avec du méthanol (1 L) en récupérant la fraction soluble (acides aminés, petits sucres, sels minéraux, pigments, quelques protéines) d'où le solvant a été éliminé au rotavapor Büchi sous vide (bain marie à 40°C). Le résidu obtenu a été repris dans de l'eau et congelé (-20°C) puis lyophilisé : il s'agit de la fraction **MeOH**. La fraction insoluble dans le méthanol a été séchée sous hotte et a subit une extraction au méthanol 60% (1,5 L, 16 h, 4°C). La solution a ensuite été filtrée sur fritté. La fraction soluble a été récupérée, le méthanol et une partie de l'eau éliminés au rotavapor. La solution concentrée obtenue a été congelée en vue d'une lyophilisation pour obtenir la fraction nommée **MeOH60**. A partir de farine de pois, c'est dans cette fraction que se retrouve PA1b et la majorité de ses isoformes. La fraction insoluble dans le méthanol 60% a été extraite directement à l'eau pH 5 (1,2 L, pH ajusté avec de l'acide chlorhydrique). Après agitation la nuit à 4°C, la solution est centrifugée 20 min à 12 500 *g* et 4°C (centrifugeuse Beckman J2-21 M/E). La fraction soluble obtenue, appelée **H_2O5** et essentiellement protéique, est congelée et lyophilisée. Le culot de centrifugation a finalement été extrait à l'eau pH 8 de la même façon que pour l'eau pH 5. Les 2 fractions récupérées (majoritairement des protéines pour la fraction soluble dite **H_2O8**; amidon et cellulose pour la fraction insoluble dite **résidu**) ont été séparées puis congelées et lyophilisées.
Tous les extraits ont été pesés après lyophilisation.
Des dialyses contre de l'eau distillée (boudins à pores de 3,5 kDa, The Spectrum Company) sur les fractions solubles au méthanol et au méthanol 60 % de *M. truncatula* ont été réalisées, ainsi que sur les fractions MeOH60 des trois autres espèces dans le but d'éliminer des contaminants de faible poids moléculaire comme les sucres, acides aminés, minéraux. De plus, l'extraction des isoformes de PA1b à partir de la fraction MeOH60 du

pois a été réalisée par solubilisation de cette fraction dans de l'acétone 80%, agitation puis passage à –20°C, 40 min., avant centrifugation à 4°C, 20 min., 14 000 g. La fraction soluble, appelée MeOH60A80 a été séchée au speed-vac (Gressent et al., 2003).

4) Tests biologiques

Les tests biologiques ont été réalisés sur les différentes fractions afin de vérifier la présence éventuelle d'une **toxicité analogue** (mortalité différentielle entre charançons sensibles et résistants).
Pour cela, on pèse l'équivalent d'1 g de farine de la fraction à tester et on complète à 1 g avec de la farine de blé. Dans l'aliment artificiel (se présentant sous forme de boulettes), la concentration en fraction testée est ainsi égale à celle de la farine initiale. Toute concentration sera exprimée en cet équivalent farine entière (EFE). Le tout a été mélangé pour homogénéiser et lié avec 0,5 mL d'eau. Un boudin a alors été formé avec cette pâte, qui après séchage 30 min, a été coupé en boulettes aux ciseaux. Celles-ci ont été mises à sécher une nuit avant utilisation. Des boulettes témoins ont été réalisées avec un gramme de farine de blé. L'aliment séché a été séparé équitablement dans 2 tubes plastiques cylindriques de 5 cm de hauteur pour 2 cm de diamètre, fermés à une extrémité par un grillage et de l'autre par un bouchon plastique. Pour chaque fraction, 30 charançons adultes de 15 jours (matures), ont alors été placés dans chacun des 2 tubes : dans l'un de souche sensible, dans l'autre, de souche résistante. Les animaux ont ensuite été placés à l'étuve et les morts ont été dénombrés quotidiennement, à partir du deuxième jour pendant deux semaines.
Les calculs de concentration létale 50 % (CL_{50} : concentration à laquelle la moitié des charançons sont morts, à un jour donné) ont été effectués grâce au logiciel TOXICO^{TM} élaboré au laboratoire (Febvay and Rahbé, 1991).
Les calculs de temps létal 50% (TL_{50}, à une concentration donnée) ont été effectués grâce au logiciel Statview par analyse « actuarielle » (données groupées en classe de temps, ici d'un jour) et statistiques non paramétriques associées. Les méthodes d'analyse de survie permettent d'associer la fréquence et le délai de survenue de l'événement étudié qui est ici la mort des insectes. On étudie le temps qui s'écoule entre le début du traitement (dépôt des charançons sur l'aliment) et la date de dernière observation. Au dernier jour de comptage (quinzième), on note le nombre de survivants qui sont alors censurés. La méthode actuarielle nécessite que les périodes de recul soient prédéterminées et de même durée, ce qui correspond bien à notre cas où le délai est de 15 jours.

B. Biochimie : recherche d'homologues protéiques

1) Dosage protéique

La teneur protéique des différentes fractions a été déterminée par la méthode de Bradford (1976). La gamme étalon est réalisée avec de la BSA (Bovine Serum Albumin). Les échantillons sont analysés au spectrophotomètre à une longueur d'onde de 595 nm.

2) Electrophorèse protéique sur gel d'acrylamide (SDS-PAGE)

Les électrophorèses ont été réalisées en conditions dénaturantes, sur gel de polyacrylamide (gel de séparation à 17,5 % et gel d'espacement à 4,5 %) en présence de SDS 0,1% selon la méthode décrite par Laemmli (1970). Les polypeptides ont été dénaturés 5 min à 100°C dans du tampon de charge (60 mM Tris-HCl, pH 6,8, 2% SDS, 50 mM DTT et 10% glycérol (v/v), 0,001 % Bleu de Bromophénol) puis séparés à voltage constant (200 V). La masse moléculaire des différents polypeptides a été déterminée par référence à des

protéines étalons (M : Low Molecular Weight Marker, Pharmacia) allant de 14,4 kDa à 94 kDa pour le standard de haut poids moléculaires et de 2,5 kDa à 16,9 kDa pour le standard peptides (m : SDS Molecular Weight Marker, Sigma).
Les protéines ont été révélées sur le gel au nitrate d'argent selon le protocole décrit par Blum et al. (1987), avec un temps de fixation de 50 min seulement pour éviter la perte de PA1b, soluble dans le méthanol aqueux.

3) Immunodétection

Après électrophorèse, les polypeptides ont été transférés en milieu liquide (tampon de Towbin modifié sans méthanol), sur membrane de nitrocellulose (Protan BA 83, Schleicher et Schuell) durant 90 min à 40 V (Towbin et al., 1979). Les membranes ont alors été incubées une nuit sous agitation dans une solution de TBS (Tris Buffer Saline, Tris-HCl 50 mM pH 7,4, NaCl 200 mM) contenant 5% de lait écrémé en poudre. Ceci permet de saturer tous les sites de fixation libres qui pourraient fixer de manière aspécifique l'anticorps et générer du bruit de fond. Les membranes ont ensuite été rincées plusieurs fois avec du TTBS (TBS + 0,05 % de Triton X-100) puis placées dans du TBS additionné d'anticorps primaire au 1/1000ème, durant la nuit, à 4°C, sous agitation (anticorps polyclonal anti-PA1b couplé à l'ovalbumine, fourni par L. Quillen, Nantes). Puis les membranes ont été à nouveau lavées par du TTBS, 3 fois 5 min, puis 1 fois 5 min au TBS. L'anticorps secondaire, couplé à la phosphatase alcaline, a été additionné au 1/2000ème dans du TBS durant 3h sous agitation. Les membranes ont ensuite été lavées comme précédemment. Les protéines marquées ont alors été révélées par incubation dans un tampon salin (0,1 M Tris ; 0,1 M NaCl ; 0,05 M MgCl2) en présence de BCIP (5-Bromo-4-Chloro-IndolylPhosphate, substrat de l'activité enzymatique de la phosphatase) et de NBT (NitroBlue Tetrazolium) à l'obscurité pendant 10 min. La réaction a été arrêtée par immersion dans de l'eau distillée.

4) HPLC (High Performance Liquid Chromatography)

La chromatographie de type HPLC en phase inverse a été réalisée, par I. Rahioui, sur colonne C18 Nucléosil 300 (250 x 4,6 mm ; 5 µm de taille de particule ; 300 A de porosité). Les protéines ont été éluées à 1mL/min avec un gradient de 22 min allant de 20% à 60% d'acétonitrile dans l'eau à 0,1% de TFA. Leur détection a été effectuée entre 210 et 350 nm par un détecteur à barette de diodes (DAD 440, Kontron).
La séparation des isoformes de PA1b a été réalisée en conditions isocratique à 38% d'acétonitrile, TFA 0,1%.

5) Spectrométrie de masse

200 µg des fractions de graines présentant une toxicité et contenant potentiellement des peptides de type A1b[6] ont été analysés par spectrométrie de masse sur un spectromètre Voyager DE-PRO (PerSeptive Biosystems, Farmingham, MA, USA) par C. Chambon (INRA, Clermont-Ferrand-Theix). De source MALDI (Matrix Assisted Laser Desorption Ionisation) et d'analyseur Temps de vol (TOF = Time Of Flight), ce spectromètre a été utilisé en mode positif et linéaire. Les spectres de masses ont été calibrés de façon extérieure par un kit de calibration (Pep Mix 2, LaserBio Labs, Sophia Antipolis, France) ce qui offre une précision d'environ 0,05%. Les échantillons ont été mélangés à la matrice (acide sinapinique ou 3,5-diméthoxy-4-hydroxycinnamique, LaserBio Labs) dans des

[6] Il s'agit essentiellement des fractions MeOH60 (cf paragraphe II.C).

proportions allant de 1 :1 à 1 :10. Une fois déposés sur la cible et séchés, ces mélanges ont été analysés par MALDI-TOF.
Pour l'extrait MeOH60 du pois, les fractions obtenues par HPLC (temps de rétention de 16 à 20 min.) ont été analysées par spectrométrie de masse de type ESI (ElectroSpray Ionisation) sur un spectromètre à triple quadrupole API III+SCIEX (Thornihill Ont., Canada) par D. Mollet (INRA, Rennes).

6) Tests d'inhibition de liaison

Une isoforme pure de PA1b (PsaA1b005, MW 3741 Da par MS) avait été marquée à l'iode ^{125}I par F. Gressent. Sa radioactivité spécifique est d'environ 1000 Ci.mmol^{-1}. Ce ligand a été utilisé pour des tests d'inhibition de liaison sur une fraction microsomale (protéines membranaires) contenant le site de liaison spécifique à PA1b (Gressent et al., 2003). Pour préparer cette fraction, un gramme de charançons Bénin (S) a été broyé au mortier en présence d'azote liquide, jusqu'à l'obtention d'une poudre fine. La poudre a à nouveau été broyée à l'Ultraturrax en présence de 20 mL de tampon d'extraction (20 mM Tris-HCl pH8 ; 0,25 M saccharose ; 2 mM $MgCl_2$; 10 µM E-64), à 4°C. Différentes centrifugations ont alors été réalisées à 4°C : 3000 g 10 min puis, sur le surnageant, 10 000 g 15 min et enfin, sur le deuxième surnageant, 45 000 g 45 min. Le culot final a été repris avec 1mL de Tris-HCl 20 mM pH7. Après un dosage des protéines, 40% de glycérol a été ajouté à l'extrait aliquoté et congelé à –80°C.
Les protéines membranaires des charançons (2µg) ont été incubées en présence de 0,4 nM de ^{125}I-PA1b et de tampon de liaison (20 mM Tris-HCl pH7 ; 0,2 % CHAPS ; 10 µM E-64) dans les puits d'une microplaque. Le volume final de la réaction est de 200 µL par puits. Pour déterminer la liaison non spécifique, une isoforme de PA1b non marquée a été utilisée à 1µM final. Les fractions à tester ont été incubées 2h à température ambiante. Les ligands libres et liés ont alors été séparés par filtration sur système Multiscreen (Millipore) et lavage (10 mM Tris-HCl pH7 ; méthanol 60 %), à froid (tampons et cônes dans la glace) et rapidement (pour limiter la dissociation). Les filtres ont été transférés dans des tubes de comptage et la radioactivité γ mesurée (Riastar, Packard Instrument, USA). Chaque point est la moyenne de trois répétitions. Les données ont été analysées à l'aide du logiciel RADLIG 4 (BIOSOFT, Cambridge, UK) et les graphiques ont été dessinés à l'aide du logiciel ORIGIN 5 (Microcal, USA).
Les tests de déplacement de liaison[7] ont été effectués pour chaque plante avec les fractions MeOH60. Pour analyser les données de ces tests, on exprime la masse d'équivalent de farine entière nécessaire à l'inhibition de 50% de la liaison du ligand marqué. De la farine de blé a été extraite comme décrit au paragraphe A.3) et sa fraction MeOH60 a été testée comme témoin négatif de liaison.

C. Biologie moléculaire : recherche de gènes homologues

1) Extraction d'ADN

L'ADN génomique des quatres espèces étudiées a été extrait à partir de jeunes feuilles, selon le protocole utilisant le détergent CTAB (hexadécyltriméthylammonium bromide) décrit par Ausubel et al. (1997). Après extraction des acides nucléiques dans un tampon à 2 % (p/v) de CTAB et 1,4 M NaCl, les protéines sont précipitées au chloroforme/octanol.

[7] Un véritable test de compétition nécessiterait la purification totale des différentes isoformes ce qui n'est pas réalisable dans notre cas.

L'ADN est ensuite précipité sous forme d'un complexe avec le CTAB, complexe insoluble se formant quand la concentration en NaCl est réduite à 0,5 M. Ce complexe est ensuite solubilisé, à des concentrations en sels élevées, ce qui permet de récupérer l'ADN, précipité par de l'isopropanol. La seule modification apportée à ce protocole, est l'ajout, au moment du broyage des feuilles dans l'azote liquide, de polyvinylpolypyrrolidone (PVPP) à 15 % p/p (MW=40000), permettant d'éliminer les composés phénoliques (Page et al., 1997).

2) Amplification de fragments de gènes par PCR

La PCR (Polymerase Chain Reaction) est une méthode d'amplification de l'ADN utilisant les propriétés de réplication de l'ADN.
Différentes amorces ont été utilisées pour ce travail :

- 3 amorces dégénérées (voir la détermination de ces amorces dans la partie résultats).

  For 1: 5' TCY-TCI-CCI-TTY-GAR-RTI-CCI-CCI-TG 3'
  Rev 3: 5' TSR-CAI-ARR-TTI-GGR-TSY-TCI-TC 3'
  Rev 1: 5' CRA-ARC-ACC-AIC-CRT-AIT-CIA-TRT-M 3'

Ces amorces sont dites dégénérées car certaines de leur positions ne sont pas pleinement déterminées, ainsi Y représente C ou T; R, A ou G; M, A ou C et S, C ou G. I, l'inosine s'apparie indifféremment à n'importe laquelle des quatre bases. Plus le degré de dégénérescence est grand, moins l'oligonucléotide est spécifique et plus il doit être concentré lors de la PCR. Les primers For 1, Rev 1 et Rev 3 sont utilisés 10 fois plus concentrés que des primers non dégénérés.

- 2 amorces non dégénérées :

  L 914 : GCT-TAC-TGT-TCT-ATT-TTT-ATG-GT
  U 238 : GCT-TCT-TTG-ATT-GTC-CTG-TTT-GC

Cette paire amplifie un fragment de 699 bp contenant toute la séquence codant pour PA1b. Elle a été utilisée pour tester l'ADN extrait des feuilles de pois lors de la mise au point du protocole d'extraction.

Les réactions de PCR ont été effectuées dans un thermocycleur (Perkin Elmer Gene Ampli PCR 2400) dans un volume réactionnel de 50 µl sur des quantités de matrice variant de 10 à 500 ng, avec 1,25 unités de Taq polymérase et son tampon 10 X (Amersham) selon les schémas de la Figure 38.

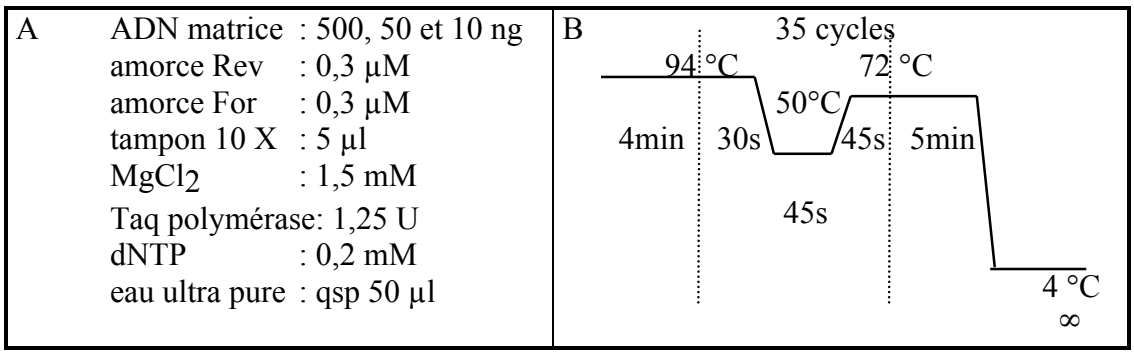


Figure 38 : Protocole d'amplification d'ADN par PCR pour des primers non dégénérés (concentrations finales). A : mélange réactionnel, B : cycle de réaction.

3) Electrophorèse sur gel d'agarose

Pour s'assurer de la qualité de l'amplification, une électrophorèse des produits de PCR a été réalisée sur gel d'agarose à 1,5 %, sous une tension de 80 V. Le tampon de migration utilisé est le TEA (Tris-HCl 1,6 mM pH 8 ; EDTA 4 µM ; acétate de sodium 1,6 mM).
La migration est suivie grâce à une solution de dépôt (bleu de bromophénol 0,25 % (p/v) ; xylène-cyanol 0,25 % (p/v) ; EDTA 25 mM ; glycérol 50 % (v/v)) ajoutée à la solution d'ADN à environ 20 % du volume. Un marqueur de poids moléculaire allant de 0,5 kb à 12 kb (Gibco BRL, 1Kb DNA ladder) a été déposé dans le gel à côté de l'ADN et permet d'évaluer la taille des fragments. Après migration, les fragments d'ADN ont été observés sous UV grâce au bromure d'éthydium (BET) ajouté au gel d'agarose à raison de 0,5 $\mu g.ml^{-1}$.

4) Clonage

Le clonage se fait en plusieurs étapes. Après avoir purifié les fragments d'ADN, ceux-ci sont insérés dans un vecteur plasmidique. Le plasmide est alors introduit par électroporation dans des bactéries compétentes.

(a) Préparation des bactéries électrocompétentes

La bactérie employée ici est *Escherichia coli* (souche NM 522, lac-). A partir d'une préculture d'une nuit dans 5 ml de milieu de culture de Luria Bertoni (LB) (NaCl 10 $g.l^{-1}$; bactotryptone 10 $g.l^{-1}$; extrait de levure 5 $g.l^{-1}$ pH 7) à 37°C, 250 rpm dans un agitateur-incubateur type Innova 4000, un erlen stérile de 50 ml de LB a été ensemencé au $1/100^{ème}$. Après incubation à 37°C pendant 4 ou 5 heures (on se situe alors dans la partie de croissance linéaire de la culture), toujours sous agitation, cette culture a été refroidie dans la glace. Les cellules doivent être maintenues au froid tout au long de la suite de la préparation. Les 50 ml ont été centrifugés à 3000 *g* pendant 15 min et le surnageant a été éliminé. Le culot, repris dans 50 ml de glycérol 10 % stérile, a été centrifugé à 3000 *g* pendant 15 min et le surnageant est éliminé. Cette dernière opération a été répétée 2 fois. Les bactéries compétentes reprises finalement dans 0,4 ml de glycérol 10 %, ont alors été aliquotées par 50 µl et conservées à –80°C.

(b) Préparation des plasmides portant l'insert

Les fragments d'ADN amplifiés à liguer ont tout d'abord été purifiés. Ils ont été concentrés dans des concentrateurs centrifuges (Nanosep 30 K), par une centrifugation de 5 min à 6000 g. Le rétentat a été récupéré et déposé sur gel d'agarose 2% (pour des fragments de 500 bp ou moins). Après migration, les bandes visualisées sous UV ont été

découpées. La purification de l'ADN a alors été effectuée à l'aide du kit QIAquick Gel Extraction de QIAgen selon les recommandations du fournisseur. Avant la ligation, l'ADN purifié, a été dosé par une électrophorèse sur gel d'agarose à 1%.
La ligation a été effectuée grâce au kit pMos*Blue* blunt-end vector (Amersham). Les plasmides sont fournis linéarisés par l'enzyme *Eco*RV et à bouts francs. Le produit de PCR purifié a été placé en contact du plasmide (rapport insert/vecteur de 2,5/1) et soumis à la réaction de ligation d'après les recommandations du kit.
Avant la transformation, les plasmides, issus de la réaction de ligation, subissent une microdialyse afin d'éliminer les sels qui pourraient gêner l'électroporation.

(c) Electroporation

L'électroporation est une technique de transformation par choc électrique de courte durée (entre 5 et 10 millisecondes) mais intense (appareil Biorad, résistance de 400 ohms, capacitance de 25 µFd, différence de potentiel de 2,5 V).
50 µl de bactéries compétentes (4.10^8 bactéries), auxquels ont été ajoutés 1 µl de plasmide (50 ng), ont été déposés dans des cuves à électroporer et soumises à une décharge électrique. 950 µl de milieu LB ont immédiatement été ajoutés, et les cellules ont été laissées à 37°C durant 1 heure. Cette incubation permet aux gènes de résistance aux antibiotiques (ampicilline), portés par le plasmide, de s'exprimer. Les bactéries, après centrifugation 1 min à 3000 *g*, ont été étalées sur milieu sélectif sur lequel seules les bactéries transformées se développent (LB gélosé, ampicilline 50 $\mu g.ml^{-1}$, Xgalactose (Xgal) 12 µM, IsoPropylThio-β-D-Galactoside (IPTG) 200 mM). Après une nuit à 37°C les boites ont été mises à 4°C pendant 4 ou 5 heures. Cette étape permet de différencier les clones transformés avec un plasmide portant un insert (blancs car lac-), des clones au plasmide vide (bleus car lac+). Le site *Eco RV,* qui a permis l'ouverture du plasmide, est contenu dans le polylinker inséré à l'intérieur d'un fragment du gène *lacZ* qui reste fonctionnel malgré cette insertion. Ainsi, lorsque la cellule hôte, lac-, intègre un plasmide vide, elle devient lac+.

5) Minipréparation d'ADN plasmidique par lyse alcaline

Le protocole de préparation de l'ADN plasmidique est celui décrit par Sambrook et al. (1989), basé sur une dénaturation alcaline du chromosome bactérien qui précipite, avec les protéines complexées par le SDS, tandis que l'ADN plasmidique est renaturé et reste soluble. Un traitement à la RNAse et une extraction phénolique ont été réalisés en plus pour une meilleure purification (élimination des ARN et des protéines).

6) Séquençage

Le séquençage de l'ADN plasmidique est réalisé par Génome Express S.A. (Grenoble) sur des aliquotes de 10 µl (500 ng). La réaction de séquençage a été effectuée par amplification par PCR dans un volume final de 20 µl en utilisant 100 ng de produit PCR, 5 pmol de primers et 9,5 µl d'un mélange de ddNTP colorés (Didesoxy 2' 3' Nucléotides Triphosphate) selon le protocole Applied Biosystems.

7) Recherche des séquences amont des gènes

Afin de disposer de l'entière partie des gènes codant pour des homologues de PA1b, les séquences en amont de celles obtenues par PCR ont été recherchées, à l'aide du kit Universal GenomeWalker™ (Clontech, USA), selon les indications du fournisseur (principe en annexe 1). L'ADN génomique a été digéré par 4 enzymes de restriction à bouts francs (*Dra*I, *Eco*RV, *Stu*I, *Pvu*II). Après purification de l'ADN, quatre banques de

digestion ont alors été obtenues. Pour chacune, une réaction de ligation d'adaptateurs fournis dans le kit a été réalisée. L'ADN est alors prêt pour réaliser les 2 PCR consécutives et nichées : les amorces de la seconde amplification s'hybrident, sur le fragment obtenu après la PCR 1, intérieurement aux primers de cette première PCR (Figure 39). Pour chacune, une amorce du kit, s'hybridant sur l'adaptateur, et une amorce spécifique, déterminée d'après les séquences incomplètes obtenues par PCR classique, ont été utilisées.

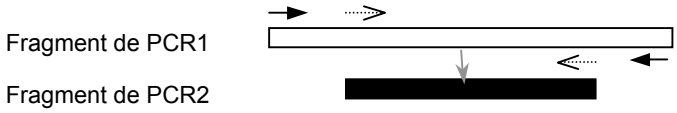


Figure 39 : Principe de la PCR nichée. Les amorces de la première PCR (flèches au trait plein) amplifient un premier fragment sur lequel s'hybrident les amorces de la deuxième PCR (pointillés), qui amplifient alors un fragment (en noir) plus court et interne au premier produit.

Le produit de PCR majoritaire et de taille supérieure à 600 pb obtenu dans l'une des 4 banques a alors été cloné et deux clones en ont été doublement séquencés. Les séquences ont alors été assemblées et analysées avec le logiciel MacMolly, contrôlées vis-à-vis d'éventuelles erreurs de PCR ou séquençage et vis-à-vis des séquences de départ, annotées pour leur partie codante, la position de leur intron et du peptide signal (grâce aux logiciels Signal P, NetGene2, NetStart disponibles au http://www.cbs.dtu.dk et manuellement d'après les séquences génomiques connues du pois et du soja). Elles ont ensuite été soumises à EMBL avec les numéros d'accessions suivants : AJ574789 à AJ574791 plus AJ574793 pour la séquence du pois complétée dans sa partie 3' (PA1a) par la séquence d'un cDNA de graine de pois identique sur la partie commune.

II. Résultats

A. Recherche des homologues de PA1b dans les graines des plantes étudiées

1) Toxicité des farines entières

Pour comparer les 4 plantes, la toxicité globale de leurs graines pour *S. oryzae* a tout d'abord été étudiée. Afin d'éviter l'influence des caractéristiques physiques des graines (taille, forme, aspect), celles-ci ont été réduites en farine et testées à différentes concentrations (80, 60, 40, 20, 10, 5 %), complétées par de la farine de blé.
La mortalité des souches sensible (S) et résistante (R) a été suivie jusqu'à atteindre 100 % sur la souche sensible (environ 15 jours).

(a) Mortalité des charançons sensibles sur farine entière

La toxicité diffère grandement d'une espèce végétale à l'autre et particulièrement sur souche sensible. En effet, bien que dans tous les cas tous les insectes de cette souche meurent en 15 jours quelle que soit la concentration en farine, l'étude des mortalités cumulées au jour 4 montre le mieux cette disparité (Figure 40).

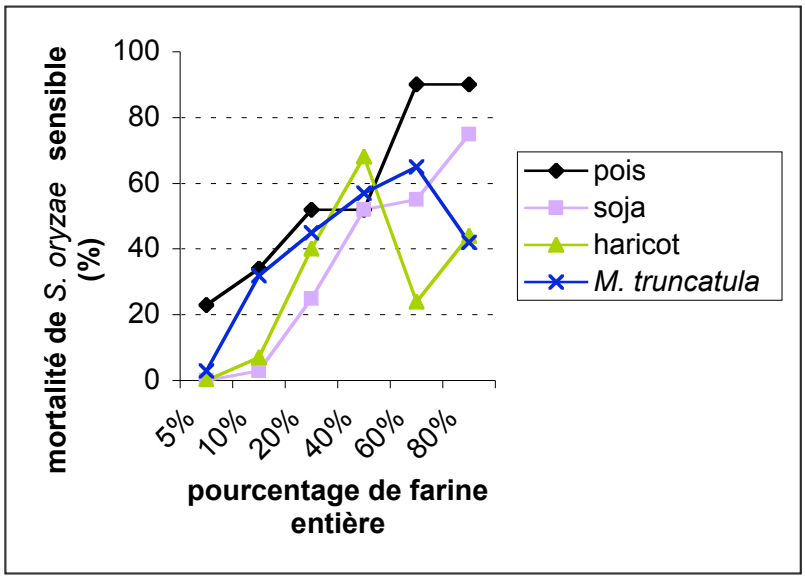


Figure 40: Mortalités cumulées (pourcentages) au 4ème jour de la souche **sensible** de *S. oryzae* sur différentes concentrations en farine entière du pois, du soja, de *M. truncatula* et du haricot.

Le pois, qui s'avère le plus toxique aux hautes concentrations (80 et 60 %), présente un profil de mortalité simple : augmentant linéairement avec la concentration. On retrouve ce même profil pour le soja. Une détermination de la concentration létale 50 a donc été réalisée grâce au logiciel TOXICO™ sur ces deux plantes. Le soja, de CL_{50} au jour 4 égale à 44,3 %, s'avère bien moins toxique que le pois ($CL_{50,J4}$ = 17,4 %).
Au contraire, le haricot apparaît relativement peu toxique à ces fortes concentrations mais bien plus à une concentration moyenne (40 %). Ceci peut s'expliquer par la présence dans la farine du haricot d'anti-appétants, qui, à forte dose, limitent la consommation de l'aliment artificiel par les charançons, et donc l'ingestion de l'éventuel composé toxique d'où une moindre mortalité (les charançons pouvent en effet survivre plus de 7 jours sans s'alimenter). Lors des dénombrements, il a en effet été constaté une faible alimentation sur les boulettes 60 et 80 % du haricot ainsi que sur celles à 80 % de *M. truncatula*, celles-ci n'étant que très partiellement grignotées par rapport aux boulettes de concentrations inférieures. Ces effets anti-appétants ne semblent pas exister chez le pois et le soja où les concentrations les plus fortes correspondent bien aux mortalités les plus élevées. Pour comparer la toxicité vraie des plantes, on peut donc considérer la mortalité à une dose relativement faible (20 %), où les effets anti-appétants sur haricot et *M. truncatula* ne semblent plus intervenir. Le pois apparaît alors le plus toxique suivi de *M. truncatula* et du haricot, le soja étant le moins toxique.

(b) Mortalité des charançons résistants

Sur pois (où on ne s'attend à aucune mortalité, les charançons de souche résistante étant élevés sur pois cassés) comme sur haricot et soja, on n'observe pas de morts parmi les insectes résistants, et ce à toutes les doses. La différence de mortalité entre sensibles et résistants pour le haricot et le soja tend à indiquer la présence d'analogues de PA1b dans ces plantes.
Au contraire, sur *M. truncatula*, les résistants périssent aussi vite que les sensibles pour les concentrations les plus élevées (de 40 à 80 %) (Figure 41). Ceci indique que les mortalités observées ne sont probablement pas le seul fait de la toxicité d'homologues de PA1b

(toxicité observée à faible dose 10-30%) mais bien plutôt d'une conjonction entre ces protéines et d'autres substances toxiques propres à cette plante (toxicité à concentrations plus élevées). La richesse en métabolites secondaires de cette luzerne est en effet bien documentée (Bisby et al., 1994 ; Harborne et al., 1971). Il se pourrait de plus que ses graines contiennent un peptide homologue de PA1b contournant la résistance des charançons de la souche R Chine.

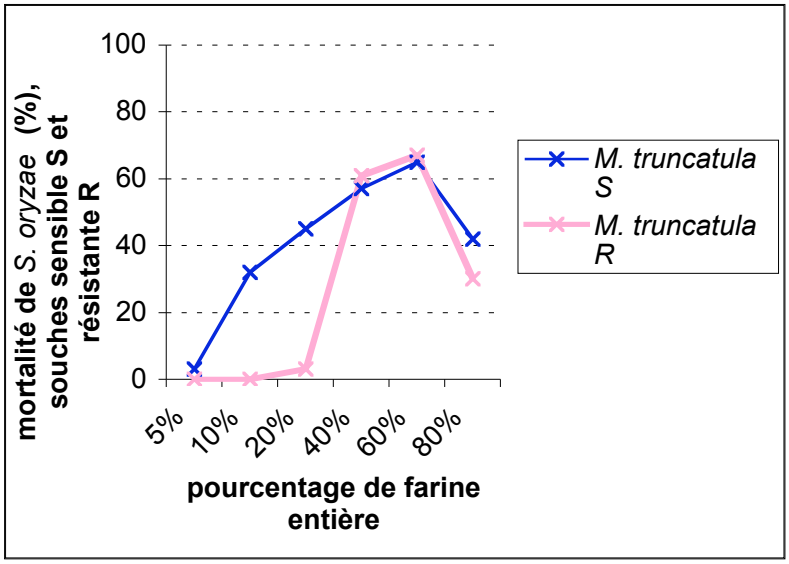


Figure 41 : Mortalité cumulée (pourcentage) au 4ème jour de *S. oryzae* souches sensible S et résistante R sur différentes concentrations en farine entière de *M. truncatula.*

Cette étude sur farine entière montre une mortalité différentielle entre souches sensibles et résistantes, les premiers *mourant* tous dans la durée du test tandis que les seconds survivent pour la plupart. Comme la résistance des charançons envers PA1b est monogénique (Grenier et al., 1997), on peut conclure à la probable présence d'homologues de ce peptide dans les différentes plantes étudiées. On peut de plus conclure de l'analyse de la courbe dose-réponse de cette étude qu'une concentration supérieure à 20 % ne convient pas à une comparaison de la toxicité de farines entières, à cause des effets anti-appétants de certains composés des graines. Ainsi, si un criblage général (variétal ou d'espèces) était à réaliser, il pourrait être conseillé d'utiliser une concentration de 10 à 20 % de farine entière et de relever les mortalités au septième jour (pour exacerber la différence de réponse entre souche témoin et souche résistante). Enfin, pour une étude dose-réponse sur farine entière, il conviendrait de réaliser une gamme de concentrations allant de 60% à 1% et de noter l'index de mortalité entre les jours 4 et 8.
La complexité des résultats sur farine entière est, dans certains cas, source de difficultés d'interprétation. Un fractionnement de ces farines et une étude de la toxicité des fractions obtenues se sont donc avérés nécessaires.

2) Fractionnement et toxicité des différentes fractions

Après avoir étudié la toxicité des farines entières, un fractionnement de celles-ci a été réalisé comme décrit dans Matériel et Méthodes (I.A.3), afin de révéler la présence d'éventuels homologues protéiques de PA1b.

Les proportions des différentes fractions dans la farine totale sont consignées dans la Figure 42. On remarque que l'importance de chaque fraction varie suivant les plantes. En particulier, les proportions des fractions résidu et pentane qui varient inversement. Elles correspondent en effet à deux types de réserves différentes, lipidiques et glucidiques (amidon, cellulose). Le soja, utilisé pour produire de l'huile est bien plus riche en fraction lipidique (21%) que le pois ou le haricot (3%).

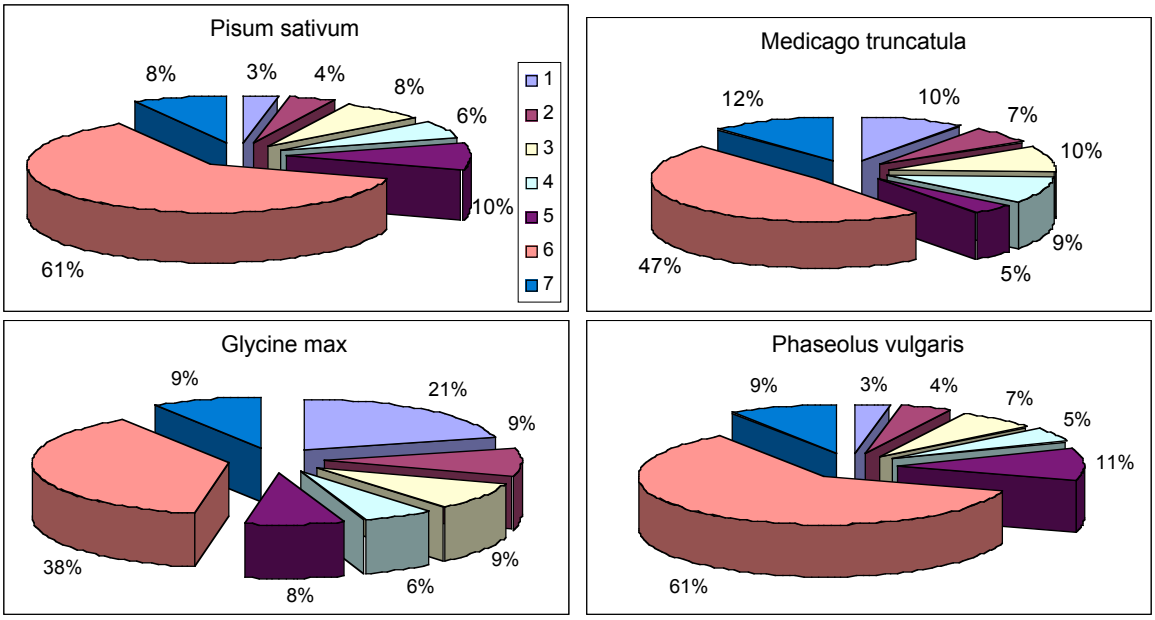


Figure 42 : Importance des différentes fractions dans la farine entière du pois, soja, de *M. truncatula* et du haricot (% p/p). 1 : fraction pentane (lipides), 2 : MeOH, 3 : MeOH60, 4 : H_2O5, 5 : H_2O8, 6 : résidu, 7 : pertes.

Les différentes fractions des quatre plantes étudiées ont ensuite été testées à la concentration qu'elles avaient dans la farine d'où elles ont été extraites (EFE).

(a) *Toxicité des fractions méthanol 60 % sur* S. oryzae

Etant la fraction d'où a été purifiée PA1b à partir du pois, MeOH60 a été la première étudiée. Pour le pois, elle entraîne la mort de 100 % des charançons sensibles en 8 jours mais n'affecte pas les insectes résistants (Figure 43). Un profil similaire est observé pour le soja et le haricot. Ces résultats permettent de supposer la présence d'homologues de PA1b dans ces 2 plantes. Pour *M. truncatula*, la mortalité des résistants comme des sensibles étant totale dès le quatrième jour, on ne peut conclure à la présence de tels homologues.

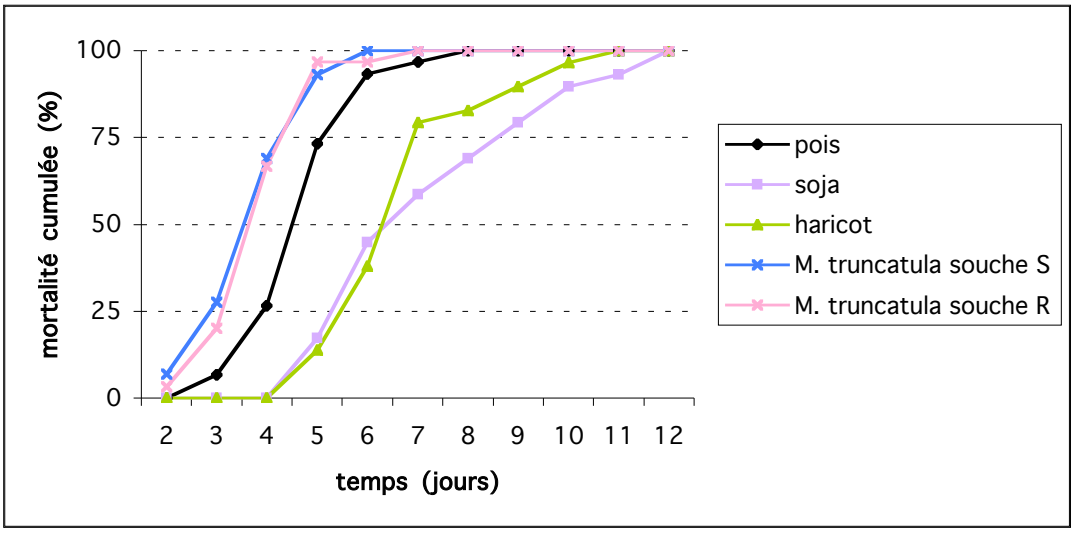


Figure 43 : Mortalités cumulées, au cours du temps, de la souche sensible S de *S. oryzae* sur les fractions MeOH 60 du pois, du soja, de *M. truncatula* et du haricot, et mortalité cumulée de la souche résistante R sur cette même fraction pour *M. truncatula*.

Les TL50 montrent que la fraction MeOH60 de la luzerne méditerranéenne est la plus toxique devant celle du pois et enfin du soja et du haricot (Tableau 6).

(b) *Toxicité des autres fractions sur* S. oryzae

Les différences s'accentuent lorsqu'on considère les autres fractions (Tableau 6).

Tableau 6 : Toxicité des différentes fractions (testées à 100% EFE) vis-à-vis des deux souches de charançons sensible (S) et résistante (R). Temps létaux 50% ± erreur standard.

TL50 (jours) *espèces*	**fractions**									
	MeOH		**MeOH60**		**H_2O5**		**H_2O8**		**résidu**	
	S	R	S	R	S	R	S	R	S	R
Pisum sativum	>20	>20	4,5 ± 0,2	>20	5,7 ± 0,4	>20	5,0 ± 0,4	>20	5,4 ± 0,3	>20
Phaseolus vulgaris	>20	>20	6,3 ± 0,3	>20	>20	>20	8,8 ± 0,9	>20	10,6 ± 0,2	>20
Glycine max	7,3 ± 0,9	>20	6,5 ± 0,9	>20	>20	>20	10,8 ± 0,6	>20	8,0 ± 1,3	>20
Medicago truncatula	3,9 ± 0,4	8,8 ± 0,6	3,5 ± 0,3	3,6 ± 0,2	4,5 ± 0,2	6,5± 0,3	4,7 ± 0,3	5,3 ± 0,2	4,3 ± 0,4	5,2 ± 0,2

L'analyse statistique (Breslow-Gehan-Wilcoxon) utilisée sur cette étude de survie de type actuarielle pour comparer les deux souches dans le cas de *M. truncatula* ne donne de différence significative (p-value<0,0001) que pour les fractions MeOH et H_2O5.

Pour la fraction MeOH du pois ainsi que du haricot, aucune toxicité n'est observée. Celle du soja présente un différentiel de mortalité total (toxicité sur S mais pas sur R) pouvant correspondre à la présence de peptides de type A1b. Pour *M. truncatula* cette fraction s'avère toxique aussi bien pour les sensibles que pour les résistants, avec toutefois un

retard de mortalité des résistants (TL50 significativement plus élevé). Ce retard pourrait être lié à la présence dans cette fraction d'homologues de PA1b, auxquels les charançons de souche résistante seraient insensibles, mais associés à d'autres composés toxiques qui finiraient par entraîner la mort des insectes.
La fraction H_2O5 du pois entraîne un différentiel de toxicité total laissant supposer la présence d'isoformes de PA1b dans cette fraction. Il pourrait s'agir de peptides homologues différents de ceux de la fraction MeOH60 ou d'une contamination de ces derniers dans la fraction aqueuse, les extractions n'étant pas totales. Toutefois, pour le soja et le haricot, aucune toxicité n'est observée, laissant supposer que la contamination de fraction à fraction suivante est relativement faible. Enfin, pour *M. truncatula,* la fraction aqueuse acide est fortement toxique pour les charançons sensibles comme pour les résistants malgré un différentiel significatif.
La fraction H_2O8 du pois est elle aussi toxique pour les charançons sensibles et inoffensive pour les résistants. Un même différentiel de toxicité est observé pour le soja et le haricot bien que les TL50 de ces deux fractions sont moindres que celui du pois (10,8 et 8,8 contre 5,0 respectivement). Une fois de plus, *M. truncatula* s'avère hautement toxique pour les deux souches d'insecte pour cette fraction.
La fraction résidu des 4 plantes tue les sensibles, et les résistants dans le cas de la luzerne. Malgré un aspect physique de l'aliment plus friable et moins homogène, par rapport aux boulettes témoins (farine de blé seule) ou à celles contenant les autres fractions, les fractions résidus sont bien consommées par les insectes des deux souches. Dans le cas de la luzerne méditerranéenne, l'absence de différentiel significatif indique que la fraction résidu contient probablement un ou plusieurs composés toxiques non apparentés à PA1b. Au contraire, les différentiels totaux de toxicité observés dans les résidus des trois autres espèces végétales témoignent certainement de la présence d'A1b dans ces fractions.

Il ne serait pas surprenant de trouver des homologues de PA1b dans différentes fractions. En effet, les peptides A1b sont le produit d'une famille multigénique et, vu leur petite taille, le changement d'un nombre limité d'acides aminés peut suffire à modifier fortement leur hydrophobicité et/ou leur ionisation. Leur présence dans la fraction résidu, indiquant une insolubilité dans les solvants utilisés ici, peut paraître surprenante. Nishizawa et al (1995) ont montré, par immunocytochimie, que la leginsuline se localise autour de la membrane plasmique et des parois des cellules de graines de soja immatures. La liaison des A1b à des portions de membrane ou des polymères de parois cellulaires pourrait expliquer leur présence dans les fractions insolubles.

Au vu des fortes toxicités de toutes les fractions de *M. truncatula*, une purification un peu plus poussée à été réalisée. Les fractions MeOH et MeOH 60 pouvant contenir de nombreuses petites molécules toxiques (acides aminés non protéiques, phénols), ont été dialysées à travers une membrane de pores de 3,5 kDa. Les tests de toxicité ont été réalisés sur le dialysat mais aucune différence notoire n'a pu être observée avec la toxicité des fractions non dialysées.

L'étude de la toxicité des différentes fractions a donc permis de situer plus précisément la présence d'homologues de PA1b solubles (*i.e*. pour lesquels un différentiel de mortalité entre souche témoin sensible et souche résistante est observé) :

- dans les fractions MeOH 60, confirmant les résultats préalablement obtenus sur pois (purification de PA1b à partir de cette fraction)

- mais aussi, probablement, dans les fractions MeOH du soja et de *M. truncatula*, H_2O5 du pois et de *M. truncatula*, et H_2O8 du pois, soja et haricot.

Le cas de *M. truncatula* reste toutefois complexe, puisqu'aucun différentiel total (mort des S sans toxicité pour les R) de mortalité n'a été mise en évidence.

Pour confirmer la présence de peptides de type A1b dans ces fractions, une recherche des analogues immunologiques de ce peptide a été entreprise grâce à la technique de l'immunodétection.

3) Détection électrophorétique et immunologique d'homologues de PA1b

Afin de vérifier la présence de protéines dans la zone de migration de PA1b, et avant de réaliser l'immunodétection, chaque fraction testée précédemment sur charançons a été analysée par électrophorèse sur gel d'acrylamide. Pour cela, sont déposés sur le même gel, PA1b purifiée et/ou une fraction intermédiaire de la purification de ce peptide (SRA1M60 : fraction albumine de la farine de pois extraite au méthanol 60 %, riche d'environ 20 % (p/p) en isoformes de PA1b).
PA1b forme une bande large autour de 4 kDa dans les conditions expérimentales choisies. Ainsi toute trace protéique autour de 4-5 kDa peut être interprétée comme pouvant correspondre à des homologues de PA1b.
On observe de telles traces pour les fractions **MeOH 60** du pois, du soja et du haricot (Figure 44). Ce résultat était attendu pour le pois puisque c'est dans cette fraction que PA1b se situe, et l'on peut conclure que cette trace correspond bien à des isoformes de PA1b. Pour le haricot et le soja, on observe des traces identiques à celles du pois, ce qui permet de penser qu'il s'agit également d'isoformes de PA1b. On observe également une bande à 6 kDa pour le pois qui pourrait correspondre à des A1a[8] non observée pour les autres espèces. La bande du soja d'environ 15 kDa est probablement une protéine de réserve connue pour son hydrophobicité (SHP : soybean hydrophobic protein (Odani, S. et al, 1987)). Pour *M. truncatula*, la fraction MeOH 60 n'a pas pu être visualisée sur gel : malgré un dépôt important de matière, rien n'apparaît sur le gel après migration.

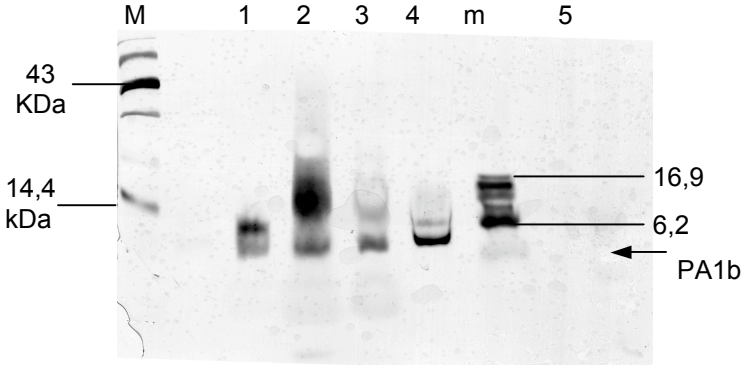


Figure 44: SDS-PAGE révélé à l'argent sur les fractions MeOH60 du pois (1), du soja (2), du haricot (3) et de *M. truncatula* (5). marqueurs de poids moléculaires (M,m); SRA1M60 (4). Environ 10 µg de protéines ont été déposées pour les puits 1, 3 et 5 et 20 µg pour le puit 2.

[8] PA1b est en effet exprimé avec un autre peptide d'environ 6 kDa (PA1a) sous forme d'une préproprotéine maturée par la suite en PA1b et PA1a (Figure 52). De plus, on retrouve dans cette fraction des masses correspondant aux séquences d'isoformes de PA1a clonées (cf II.C

Le même problème est survenu pour les fractions **MeOH** de toutes les plantes. Les concentrations très faibles en protéines nécessitent le dépôt d'une grande quantité de fraction, or celles-ci peuvent contenir de nombreuses molécules interférant avec l'électrophorèse comme des polysucres ou de petites molécules chargées. Pour pouvoir observer les protéines de ces fractions, il faudrait donc les purifier plus avant ce qui n'a pas pu être réalisé jusqu'ici, la purification étant en cours actuellement sur les extraits de *M. truncatula*.

Les fractions **H_2O** n'ont présenté aucune difficulté de ce type, étant riches en protéines solubles. La **Figure 45** montre les profils variés obtenus. On retrouve pour le pois une faible réponse dans la zone de taille de PA1b dans la fraction H_2O5. Pour les autres espèces ainsi que pour les fractions H_2O8, aucun peptide de masse apparente proche de PA1b n'a pu être franchement observé malgré les différentiels de toxicité observés pour certaines d'entre elles.

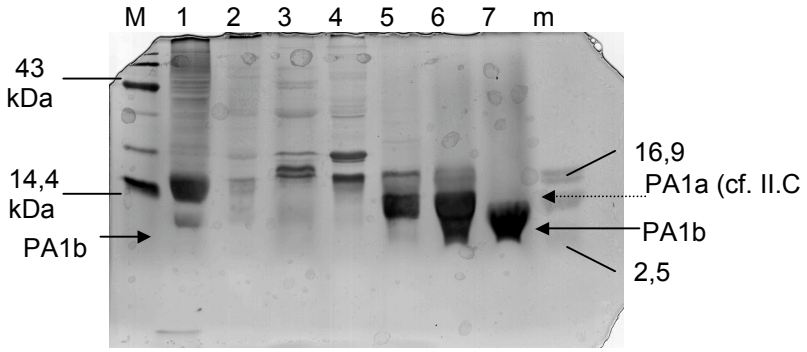


Figure 45 : SDS-PAGE, révélé à l'argent, des fractions aqueuses de *M. truncatula* (1 : pH 8; 2 : pH 5), et des fractions H_2O5 de *G. max* (3), *P. vulgaris* (4) et *P. sativum* (5). marqueurs de poids moléculaires (M, m); SRA1M60 (6); PA1b (7).

Pour vérifier si les protéines observées correspondent bien à des analogues de PA1b, une immunodétection a été réalisée sur les fractions MeOH 60 et H_2O5 (Figure 46). Seuls PA1b purifié, les fractions SRA1M60, MeOH 60 et H_2O5 du pois, et l'ovalbumine[9] du standard protéique fixent l'anticorps. Cette absence apparente de réaction croisée ne permet donc pas de valider immunologiquement la présence d'analogues dans les fractions MeOH 60 de *P. vulgaris* et *G. max,* et H_2O5 de *P. vulgaris*. Il se peut que les isoformes de PA1b potentiellement présentes dans ces fractions ne soient pas reconnues par l'anticorps car différant plus des isoformes du pois que celles-ci ne diffèrent entre elles. En effet, la détection d'une isoforme dans la fraction H_2O5 du pois indique que l'anticorps reconnaîtrait plusieurs isoformes de solubilités différentes.

[9]L'anticorps utilisé ici a été réalisé par injection de PA1b couplé à l'ovalbumine de haut poids moléculaire pour le rendre plus antigénique.

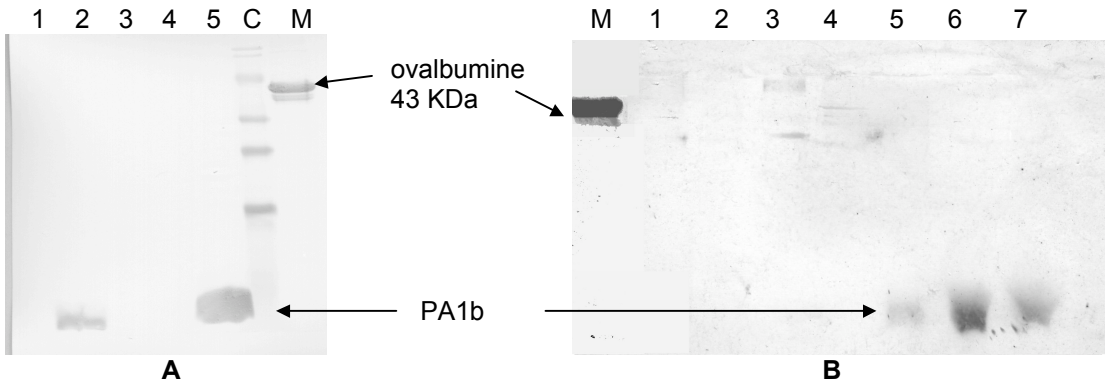


Figure 46 : Membranes de transfert marquées par l'anticorps antiPA1b et révélées par la phosphatase alcaline.
A : fractions MeOH60. *M. truncatula* (1), *P. sativum* (2), *G. max* (3), *P. vulgaris* (4), PA1b (5) ; marqueurs de poids moléculaires (M, C : coloré). **B** : Membrane de transfert correspondant au gel SDS-PAGE de la Figure 45. Fractions aqueuses de *M. truncatula* (1 : pH 8; 2 : pH 5), H_2O5 de *G. max* (3), *P. vulgaris* (4) et *P. sativum* (5) ; SRA1M60 (6); PA1b (7) ; marqueurs de poids moléculaires (M, m).

Ainsi la détection des analogues peptidiques de PA1b dans les trois espèces étudiées s'avère pour l'instant peu fructueuse. Un criblage biochimique ou immunologique précoce (sur extraits bruts) des albumines de type 1b sera difficile à obtenir par des techniques électrophorétiques *a priori* adaptées à cet objectif. La nature de ces albumines (compacité, solubilité en milieu non aqueux, ponts disulfures, nombre d'épitopes réduit impliquant une antigénicité médiocre) est probablement responsable de ces difficultés. C'est pour essayer de contourner ce problème que des tests d'inhibition de liaison de ligand marqué ont été utilisés.

4) Résultats des tests d'inhibition de liaison.

Pour confirmer la présence ou non d'Albumines de type 1b dans les extraits MeOH60 et les différentes fractions de *M. truncatula*, des tests de compétition de liaison de ligand ont été réalisés entre ces extraits et la toxine marquée radioactivement. En effet, celle-ci possède un site de liaison spécifique à forte affinité (Kd 6 nM) (Gressent et al., 2003) sur préparations membranaires de charançons sensibles. Les résultats obtenus montrent que toutes les fractions MeOH60, exceptée celle de *M. truncatula,* inhibent fortement la liaison spécifique de ^{125}I-PA1b à son site de liaison (Figure 47). Pour les fractions MeOH, MeOH60 et H_2O5 de *M. truncatula,* l'inhibition observée n'est que 10 fois moindre par rapport à celle des témoins négatifs blé correspondants. La fraction H_2O8 de cette espèce ne présente aucune activité de liaison.
Pour les trois autres espèces, l'inhibition de 50% de la liaison de PA1b marquée est observée pour 10, 18 et 79 µg d'EFE des fractions MeOH60 de *P. sativum, P. vulgaris* et *G. max* respectivement. Ces valeurs, de 2 à 3 ordres de magnitude plus faible que celle du témoin négatif, sont cohérentes avec les résultats des tests biologiques (Figure 43).

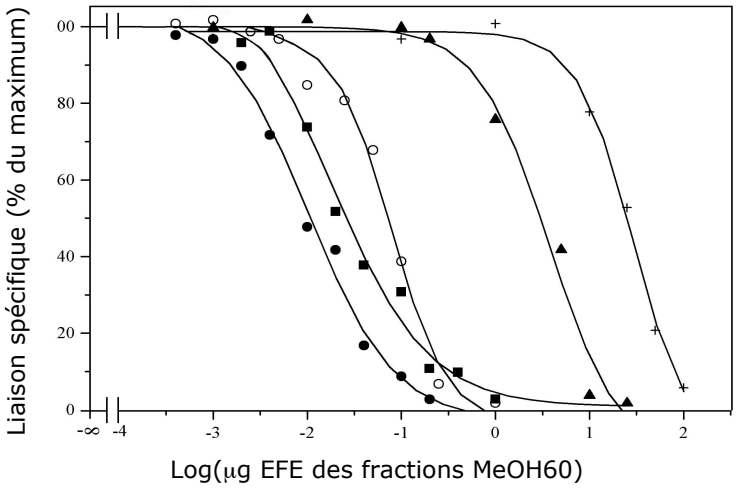


Figure 47: compétition de liaison de ^{125}I-PA1b sur son site de liaison par les extraits MeOH60 des quatre plantes testées et du blé témoin. ● *P. sativum*, ■ *P. vulgaris*, ○ *G. max*, ▲ *M. truncatula*, ⊄ *T. aestivum*. Les quantités portées en abscisse correspondent aux masses EFE présentes dans le puits de réaction.

B. Recherche des gènes homologues au gène codant pour PA1b dans le génome des plantes étudiées.

1) Extraction d'ADN des différentes plantes

Le rendement de l'extraction d'ADN a été évalué à 150 µg par gramme de tissu frais, ce qui entre bien dans l'échelle de rendement attendu (100 à 500 $\mu g.g^{-1}$) (Ausubel et al., 1997).
La qualité de l'ADN extrait a été testée, sur pois, par une PCR avec les amorces non dégénérées U238 et L914. L'amplification ayant été positive, l'ADN génomique obtenu a été aliquoté et conservé à -20°C ou utilisé directement pour la recherche des gènes homologues à celui de PA1b.

2) Amplification de fragments de gènes homologues à PA1b

Le gène codant pour PA1b est constitué en réalité de trois parties, codant pour une préproprotéine. La séquence signal, rapidement clivée lors de la maturation protéique, est suivie de la séquence codant pour PA1b puis de celle codant pour PA1a, une autre albumine de 6 kDa, riche en cystéine mais ne présentant pas d'activité entomotoxique et de fonction encore inconnue (Delobel B et Quillien L, communication personnelle).
L'alignement des séquences des homologues de PA1 auparavant connues chez le pois, le lupin et le soja a permis d'observer différentes régions bien conservées. Une première, au début de la séquence codant pour PA1b, présentant le motif CSPFE$^{V}/_{M}$PPC, a été choisie pour déterminer l'oligonucléotide dégénéré For1. Ce motif est supposé caractéristique de cette nouvelle famille de peptide de graines PA1b car non présent dans les séquences racinaires, à homologie faible mais significative, de *M. truncatula*. Les deux primers reverse dégénérés ont été définis dans la séquence codant pour PA1a, mieux conservée que PA1b, ce qui doit permettre d'obtenir les séquences C-terminales des homologues de PA1b. Rev3 (DEHPNLC) se situe au début de la séquence de PA1a, et Rev1 (DIDYGWCF), en aval d'environ 100 pb de ce dernier (voir Figure 52).

Les deux paires formées par ces 3 primers ont été utilisées. Après avoir déterminé le nombre de cycles (35), la température d'appariement (Ta = 50°C) et la concentration en primers (3 μM) permettant une amplification correcte, les fragments obtenus ont été analysés.

Avec la paire For1/Rev3, on amplifie un fragment d'environ 130 bp, ce qui était attendu d'après les séquences disponibles (Figure 48). L'amplification est positive dans toutes les plantes, même le haricot dont on ne dispose d'aucune séquence et *M. truncatula* pour laquelle les séquences disponibles restent très éloignées, correspondant à des peptides racinaires non exprimés dans les graines. Ce résultat très encourageant nous a amené à vérifier que ces fragments de taille attendue correspondent bien à des séquences analogues au gène de PA1b par leur purification, clonage et séquençage. Chez *M. truncatula*, en plus de ce fragment de 130 bp, une deuxième bande, moins intense, de 400 bp environ est observée. Ce fragment inattendu correspond à une amplification aspécifique (séquence non codante).

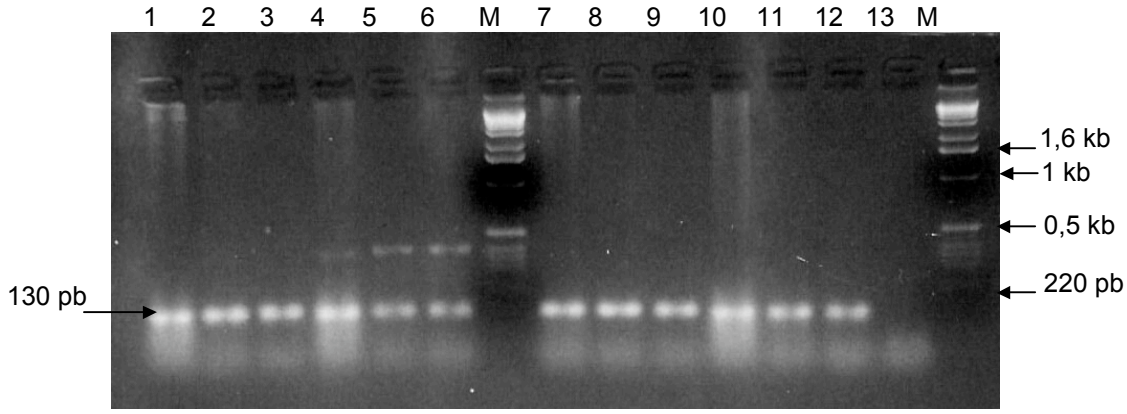


Figure 48 : Amplification par PCR avec les amorces For1 et Rev 3 d'un fragment de 130 pb correspondant à une partie du gène de PA1b. 1,2,3 : *P. vulgaris* ; 4,5,6 : *M. truncatula* ; 7,8,9 : *G. max* ; 10,11,12 : *P. sativum* ; 13 : témoin négatif ; M : marqueur 1 Kb. 1,4,7,10 : 500 ng de matrice; 2,5,8,11 : 50 ng; 3,6,9,12 : 10 ng.

La paire For1/Rev1 a également été utilisée avec succès : chaque plante présente une unique bande d'environ 220 bp, que l'on pouvait attendre d'après les séquences connues (Figure 49).

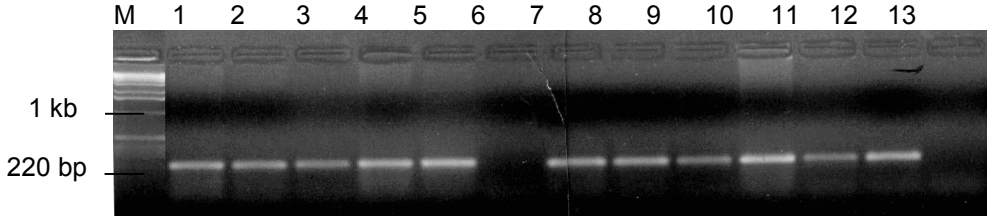


Figure 49 : Amplification par PCR avec les amorces For1 et Rev1 d'un fragment de 220 pb correspondant à une partie du gène de PA1b et du début du gène de PA1a. 1,2,3 : *P. sativum* ; 4,5,6 : *P. vulgaris* ; 7,8,9 : *M. truncatula* ; 10,11,12 : *G. max* ; 13 : témoin négatif ; M : marqueur de poids moléculaire. 1,4,7,10 : 100 ng de matrice; 2,5,8,11 : 50 ng; 3,6,9,12 : 10 ng.

Ces deux amplifications ayant réussi et fourni des fragments de taille identique entre les 4 plantes, on peut en déduire que les séquences entre ces plantes de tribus différentes doivent être assez bien conservées, au moins au niveau des oligonucléotides, et donc pour PA1b, au niveau du motif CSPFE$^{V}/_{M}$PPC. C'est pour confirmer cette hypothèse que les fragments de chacune des plantes obtenus à partir des deux PCR ont été séquencés.

3) Recherche des séquences amont des gènes

L'analyse des séquences des fragments de PCR a abouti à la détermination d'une unique séquence pour le soja, et de deux pour le pois, le haricot et *M. truncatula.* Pour chacune de ces séquences, deux amorces spécifiques ont été déterminées pour la marche sur le gène, excepté pour le pois. Pour cette espèce, une seule séquence a été sélectionnée.
La présence de deux PCR successives permet d'augmenter la spécificité de l'amplification: en effet, pour chacune, une seule amorce est spécifique, l'autre étant fournie par le kit et s'hybridant sur l'adaptateur présent sur tous les fragments de restriction obtenus après digestion et ligation. De plus, il est important de déterminer les primers spécifiques dans les régions variables des séquences, afin qu'ils ne puissent s'hybrider que sur le gène correspondant au clone choisi et dont on veut la séquence amont. En adaptant la température d'amplification (65°C contre 67°C) et le nombre de cycles (25 contre 20 pour la PCR nichée et 35 contre 32 pour la première réaction), des produits majoritaires ont pu être observés lors de la PCR nichée dans certaines banques des quatre plantes, ainsi que le produit attendu sur l'ADN témoin (humain) (Figure 50).

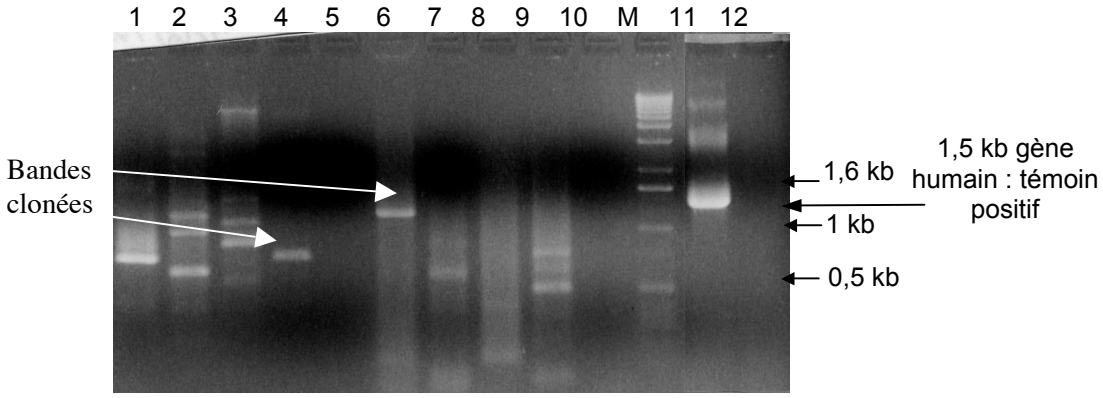


Figure 50 : Exemple de résultats de la PCR nichée du protocole GenomeWalker sur les banques d'ADN génomique digéré du haricot (séquence PvuA1b001) (6,7,8,9) et de *M. truncatula* (séquence MtrA1b006) (1,2,3,4). banque *Dra*I (1,6); banque *Eco*RV (2,7); banque *Stu*I (3,8); banque *Pvu*II (4,9); marqueur 1 kb (M); témoins négatifs (5, 10, 12); témoin positif : ADN humain (11).

Les bandes majoritaires sont purifiées, clonées, séquencées et analysées. Les deux séquences d'A1b de *M. truncatula*, celle du soja, celle du pois et l'une des deux séquences du haricot ont ainsi pu être complétées en 5'. Pour la séquence PvuA1b041, deux séquences 5' divergentes et différentes du fragment obtenu par PCR avec amorces dégénérées ont été obtenues. Ceci indique la présence d'autres allèles ou loci chez le haricot, en accord avec les masses (au moins 3) de peptides pour lesquelles aucune séquence n'a pu être associée, dans l'extrait MeOH60 (voir Tableau 7 et paragraphe C.)
De plus, pour *M. truncatula*, la même démarche a été utilisée avec succès pour rechercher les séquences en 3' des fragments obtenus par les premières PCR, afin de disposer des séquences complètes de ces nouveaux gènes.

4) Analyse des séquences

Les différentes séquences (complètes et incomplètes) obtenues ont été alignées en utilisant le logiciel Multalin™ élaboré à l'INRA de Toulouse et disponible sur Internet

(http://www.toulouse.inra.fr/multalin.html). Pour les séquences de peptides racinaires de *M. truncatula,* l'alignement a été arrangé manuellement afin de conserver l'alignement des cystéines (Figure 52). Les séquences obtenues sont nommées d'après la nomenclature neutre proposée par Y. Rahbé[10].

La structure globale du gène de PA1b est conservé dans toutes les espèces. On retrouve un peptide signal relativement variable coupé d'un intron de taille variable (de 82 à 368 pb pour le pois et le soja respectivement), puis A1b, un court peptide intermédiaire potentiel de longueur relativement variable, et enfin A1a qui apparaît plus conservé (y compris chez les homologues non-graines). La forte conservation de la molécule tend à indiquer qu'elle joue un rôle précis toutefois encore inconnu. Higgins et al. (1986) et Gatehouse et al. (1985) proposent un rôle de stockage du soufre dans les graines, ce qui n'explique ni cette forte conservation, ni la présence d'homologues relativement proche dans les racines.

La comparaison avec les séquences publiées du pois et du soja montre que nos séquences sont de nouveaux allèles avec peu de mutations, toutes synonymes dans le cas de la séquence du soja par rapport à certains EST. Les nouveaux gènes de *M. truncatula* et du haricot apportent de la variabilité dans la famille peptidique, amenant à distinguer certains acides aminés particulièrement bien conservés dans les A1b. Il s'agit des six cystéines il est à noter que la dernière est exceptionnellement absente dans la séquence MtrA1b007), des cinq prolines, des glycines G5 et G30 en plus de l'arginine R21 et de la leucine L27.

Les deux séquences de *M. truncatula*, dont une assez fortement divergente (MtrA1b007), restent toutefois plus proches de la séquence de PA1b que des peptides racinaires et nodulaires auparavant publiés chez cette espèce (Gamas et al., 1996).

L'analyse des séquences complètes déjà connues, ainsi que la structure tridimensionnelle de PA1b, permet de distinguer plusieurs régions plus ou moins conservées (Figure 52 et Figure 51): région de **maturation** N-terminale, segment **nC1** N-terminal, segment **CC2** court, segment **CC3** où a été ancré le primer N-terminal, formant une boucle apparemment souple laissant les résidus F et E accessibles à un pôle de la molécule, segment **CC4** variable formant une boucle opposée à la précédente, segment **CC5** très court (CXC), segment **CC6** replié réalisant un doigt rigide hydrophobe dont l'extrémité jouxte les résidus F et E au même pôle du peptide, puis segment **Cc7** C-terminal et enfin région intermédiaire **I** entre PA1b et PA1a, éliminée à la maturation du peptide.

```
                nC1 CC2    CC3    CC4  CC5    CC6     Cc7
PA1b             ASCNGVCSPFEMPPCGTSACRCIPVGLVIGYCRNPSG
```

Figure 51 : CC1 à 7 : Segments de PA1b délimités par les cystéines 1 à 6.

- La séquence de maturation à clivage post A (motif $^{V}/_{I}$EA) s'avère bien conservée entre peptides racinaires et peptides de graines. Il est probable que l'enzyme responsable de cette maturation soit elle aussi relativement conservée. Il faut toutefois remarquer que les sites de coupure des peptides racinaires et du lupin ne sont pas connus expérimentalement, les séquences obtenues ne provenant pas de séquençage protéique mais nucléique. Seuls les peptides décrits par Higgins et al. (1986) et Watanabe et al. (1994), ainsi que ceux ressortant de notre étude (pois, soja et haricot) sont parfaitement caractérisés d'un point de vue post-traductionnel.

[10] les trois premiers caractères sont la première lettre du nom de genre et les deux premières du nom d'espèce, suivis de trois lettres indiquant le type de peptide (A1b ou A1a) et de trois chiffres (NNN) permettant d'identifier les isoformes d'une même espèce : GesA1xNNN.

- Le segment CC3 s'avère extrêmement conservé (CSPFEM(/V)PPC) dans les séquences de type peptide de graine. Cette extrême conservation et sa position tridimensionnelle dans la protéine (coude tourné vers l'extérieur du peptide, dû à la présence de trois prolines dont une cis (Jouvensal et al., 2003)) permettent de supposer que ce segment jouerait un rôle important dans l'activité du peptide.
- La variabilité de la boucle CC4 se trouve confirmée par les séquences obtenues. En effet, au delà de la variabilité des acides aminés accentuée par l'apport des nouvelles séquences, le nombre même de ceux-ci varie : PvuA1b001 présente 5 acides aminés contre 4 pour les autres séquences de graines.
- Au contraire, le segment CC5 est parfaitement conservé. Dans tous les peptides de graines, le motif est CRC, pour les autres peptides, l'acide aminé central varie fortement (CXC). Cette conservation du motif (un seul acide aminé entre les deux cystéines) est certainement lié à une forte contrainte structurale concernant probablement la connectivité des ponts disulfures.
- Le doigt hydrophobe (CC6) s'avère assez variable, avec toutefois conservation de la proline à la deuxième position, de la leucine à la cinquième et d'une glycine à l'avant-dernière, pour les peptides de graines. Au vu de sa position tridimensionnelle (tourné vers l'extérieur, en face du segment CC3) et de son hydrophobicité conservée, ce segment pourrait également être important pour la toxicité de ce peptide.
- Le segment I, situé entre PA1b et PA1a présente une variabilité de longueur, plus importante et plus attendue que pour CC4, puisque n'influant pas sur la structure des peptides. On peut remarquer que le premier acide aminé de ce peptide, V ou L dans la plupart des séquences, est une proline pour les deux séquences de *M. truncatula*. Jouxtant le site de clivage observé chez le pois (glycine C-terminale remplacée dans cette espèce par une sérine), elle pourrait empêcher la maturation de type pois des peptides de cette plante.

On ne connaît pas le mode d'action des peptides de type A1b et l'observation de motifs conservés ou au contraire fortement variables peut être une aide à la détermination des régions actives. Néanmoins, l'absence de véritable motif protéique structural détecté, ne permet pas de présager du mode d'action et du rôle exact de ces peptides, dans les insectes cibles comme dans la graine. En effet, la caractérisation du site de liaison de PA1b dans les tubes digestifs des charançons est en cours au laboratoire et l'activité entomotoxique éprouvée de ce peptide permet de le considérer comme molécule de défense des graines. Toutefois, les peptides apparentés caractérisés chez le soja et le lupin n'ont pas été testés pour leur activité entomotoxique. La toxicité analogue observée sur la farine de soja au cours de ce travail laisse présager d'une réelle activité toxique de ces peptides mais n'exclut pas un éventuel autre rôle. En effet, il se peut que ces peptides servent de réserve de soufre dans la graine comme proposé initialement par Higgins et al. (1986). Il est également possible qu'ils interviennent dans un mécanisme de transduction de signal, au sein de la plante, comme suggéré pour le lupin par Ilgoutz et al. (1997), et pour le soja par Watanabe et al. (1994).

Figure 52 : alignement des séquences obtenues lors de notre travail et de celles de la littérature. La première colonne identifie les peptides (accession de Swissprot, d'EST, ou d'EMBL, nom du gène) ainsi que l'organisme source : Psa : *Pisum sativum*, Pvu : *Phaseolus vulgaris*, Gma : *Glycine max*, Mtr : *Medicago truncatula*, Van : *Vigna angularis*, Vra : *Vigna radiata*. La première ligne indique la maturation de type pois de la préproprotéine A1 (Higgins et al., 1986). Les cystéines conservées sont en rouges, les autres acides aminés conservés sont en bleu.

C. Détection de peptides par spectrométrie de masse.

Pour appuyer les résultats des tests biologiques, des électrophorèses et des tests de liaison, et de l'approche de biologie moléculaire, les fractions MeOH60 ont été analysées par Spectrométrie de Masse MALDI-TOF. Pour chacune, des peptides autour de 4 kDa ont été détectés, toutefois à un nettement moindre niveau pour *M. truncatula*. Pour le pois, un groupe supplémentaire et abondant de peptides à 6 kDa est également observé.
En considérant une maturation post-traductionnelle semblable à celle de PA1b pour les différents homologues clonés (colonne masses théoriques), nous avons pu trouver les peptides correspondants (colonne masse détectées : valeurs soulignées) dans les fractions MeOH60, sauf pour *M. truncatula* (Tableau 7). De plus, nous avons identifié de probables variants tronqués en C-terminal ne possédant pas la glycine terminale des A1b complets déterminés (-57 Da). Ceci est observé pour les trois plantes positives. D'autres peptides du haricot, de masse 3822, 3866 et 3874 non déterminés par leur séquence, sont probablement également des isoformes d'A1b car leurs équivalents potentiellement tronqués d'une glycine existent. Dans le pois, on retrouve également des pics de masse correspondant aux clones obtenus par recherche de cDNA homologues. Une isoforme supplémentaire non clonée (au niveau génomique comme au niveau des cDNA) mais précédemment identifiée par séquençage protéique se retrouve dans le pic de masse 3758 de la fraction MeOH60 de cette espèce. En outre, les quatre A1a clonées chez cette espèce (ADNc) ont également été identifiés dans les fractions MeOH60 et H_2O5. Pour *M. truncatula* tous les variants possibles de maturation C-terminale des deux séquences nucléiques caractérisées ont été testés pour la présence de pics de masse correspondants, mais sans succès.

Pour la fraction MeOH60 complexe du pois, une analyse plus poussée a été réalisée par séparation en HPLC suivie d'une analyse par spectrométrie de masse de type ESI. Une extraction à l'acétone 80% glacée (-20°C) a été effectuée sur cette fraction. La partie soluble,MeOH60A80, contenant les isoformes de PA1b, a alors été injectée en HPLC amenant à la séparation de 6 groupes de peptides en conditions isocratiques (Figure 53).

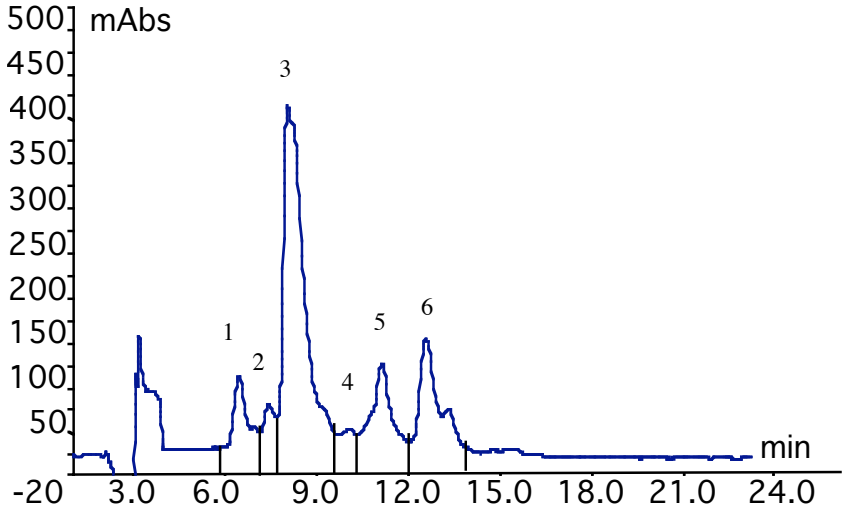


Figure 53 : chromatogramme d'HPLC en conditions isocratiques : séparation de pics d'isoformes de PA1b de la fraction MeOH60A80 du pois.

Pour le pic 1, les masses de 3788,73+-0,08 et de 3741,43+-0.35 sont détectées, pouvant correspondre respectivement aux isoformes PsaA1b015 et aux deux isoformes isomassiques PsaA1b011 et 014. Le pic 2 ne présente qu'une seule masse de 3757,68+-0.60 n'ayant pas de clones correspondant. Le pic 3 n'est également représenté que par une seule masse, de 3788,48+-0.15 qui peut à nouveau correspondre à l'isoforme 015. Cette ambiguïté de position de l'isoforme 015, témoignerait de l'existence d'une isoforme isomassique à 015, plus ou moins hydrophobe que celle-ci, et dont le clone n'a pas été caractérisé. Le pic 4 contient trois masses différentes non reliées à un clone : 3772,53+-0.08 ; 3842,73+-0,08 ; 3819,48+-0,98. Les masses de 3519,78+-0,29 et 3502,53+-0,08 constituent le pic 5. Enfin, la masse de 3789,48+-0,14 détectée dans le pic 6 correspond à celle de la séquence 012. La haute résolution de cette analyse indique qu'au moins dix gènes et peptides (allèles ou loci) sont exprimés dans le pois, bien que nous n'ayons identifié que quatre d'entre eux par la recherche de séquences nucléiques au niveau génomique ou des transcrits. La complexité de cette famille de gènes pour laquelle différents gènes codent pour des peptides peu divergents mais de même masse (isoformes PsaA1b011 et 014, PsaA1b015 virtuellement présent dans deux pics HPLC) est probablement due à une évolution récente de ces homologues dans cette espèce.

Finalement, l'analyse par spectrométrie de masse MALDI des fractions MeOH et H_2O5 du pois a été réalisée afin de déterminer d'éventuelles contaminations entre fractions successives. La comparaison des masses déterminées (tableau) montre que l'extrait MeOH est très peu contaminé par les composants majeurs de la fraction suivante (MeOH60) mais que les fractions MeOH60 et H_2O5 partagent certains composants[11], ce qui pourrait être dû à une extraction incomplète, et/ou à l'existence d'isoformes isomassiques d'hydrophobicité différente.

[11] Les redondances correspondent à des masses différant de moins de 0,05%.

Tableau 7 : Peptides observés par spectrométrie de masse sur des fractions sélectionnées des plantes étudiées

espèce (genotype)	**Masses théoriques** [a] (sequences cibles)	**Fractions**	**Masses détectées** correspondance [b] (intensité relative)	**% intensité de correspondance** [d]
Glycine max (cv. Paoki)	3876.5 (GmaA1b005) 3819.5 (c-1)	MeOH60	3817 (25), 3874 (100 %), 3890 (15), 3928 (16), 7858 (18)	72 %
Medicago truncatula, (cv. Salernes)	3870.4 (MtrA1b006) 3858.4 (MtrA1b007)	MeOH	Pas de masse > 2000 [c]	-
"		MeOH60	2593 (54), 2767 (24), 3443 (23), 4375 (43), 4650 (12), 5114 (23), 5186 (100 %), 6887 (19)	0 %
Phaseolus vulgaris, (cv. Contender)	3994.6 (PvuA1b001) 3937.6 (c-1)	MeOH60	3764 (68), 3809 (100 %), 3820 (89), 3822 (83), 3866 (80), 3868 (75), 3874 (23), 3935 (62), 3993 (25) *2718, 4481, 5439*	14 %
Pisum sativum (cv. Frisson)	3742.4 (PsaA1b011) 3685.3 (c-1)	MeOH	traces de *3745, 3760, 3791, 3808 3819* [c]	-
"	3742.4 (PsaA1b014) 3789.5 (PsaA1b015) 3732.4 (c-1) 3790.4 (PsaA1b012) 3733.4 (c-1) 5908.4 (PsaA1a014) 5923.5 (PsaA1a015)	MeOH60	3731 (16), 3742 (35), 3758 (14), 3789 (86), 3806 (17), 3816 (15), 5541 (14), 5740 (31), 5907 (56), 5921 (53), 5935 (53), 5968 (12), 5983 (100 %), 5998 (14), 6014 (20), 6028 (15) *3685, 3774, 3843* [c]	67 %
"	5971.5 (PsaA1a011) 5984.5 (PsaA1a012)	H_2O pH5	3729 (23), 3743 (24), 3790 (30), 3818 (30), 5005 (10), 5741 (28), 5908 (56), 5923 (50), 5936 (53), 5968 (10), 5984 (100), 6015 (18), 6029 (23), 6070 (13), 6086 (13), 6099 (13), 6114 (11), 6130 (12), 6146 (22), 6191 (17), 6209 (10), 6221 (11), 6809 (18), 7850 (27)	49 %

a : à partir des séquences clonées, masse moyenne MH+ moins 6 (pour les ponts disulfures). C-1 indique les peptides tronqués en C-terminal. **b** : la correspondance avec une masse théorique est considérée (souligné) si une précision de moins de 0,05% est obtenue. **c** : Seuls les pics de masse supérieure à 2000 Da et d'intensité de plus de 10% de celle du pic majoritaire sont reportés (quelques pics mineurs d'intérêt sont indiqués en italique). **d** : pourcentage d'intensité des pics identifiés avec correspondance *versus* l'intensité totale des pics majeurs (définis en c).

III. Discussion

Une toxicité différentielle entre charançons de souche sensible S et souche résistante R a été détectée pour différentes fractions des graines des quatre plantes étudiées. Comme la toxicité aiguë des graines du pois est due à PA1b (Delobel et al., 1998), et que la résistance de *S. oryzae* à cette toxicité est un caractère monogénique (Grenier et al., 1997), nous supposons qu'une toxicité différentielle S/R est due à des peptides de type A1b.

1) Toxicité analogue et gènes homologues, cas du pois, du soja et du haricot.

Cette hypothèse se révèle juste pour trois des plantes testées pour lesquelles un différentiel total a été observé (pas de mortalité pour R). Ce profil de toxicité simple, dû uniquement à des A1b est donc étendu d'une plante de la tribu des Viciae (pois) à deux espèces de Phaseoleae (haricot et soja). La toxicité de ces trois plantes est la plus élevée dans la fraction MeOH60 d'où a été purifiée la première toxine PA1b.
Les tests de compétition de liaison s'avèrent capables de détecter une activité de type A1b dans des extraits relativement bruts de trois plantes, renforçant les résultats des tests biologiques.
Au contraire du cas complexe de *M. truncatula*, la présence de peptides dans les extraits toxiques du pois, du soja et du haricot apparaît simple. Environ 70% des masses des fractions MeOH60 pour lesquelles une séquence nucléique à pu être mise en correspondance appartiennent soit à des A1b (soja et haricot) soit à des A1a et des A1b (pois). Il est surprenant de ne pas détecter d'A1a dans les extraits du soja et du haricot car, d'un point de vue des séquences, ces peptides sont très proches dans les trois espèces. Des temps de renouvellement différents pour ces A1a dans les graines de Phaseoleae peut expliquer de telles différences. Toutefois, peu de choses sont connues sur le contenu des graines de Légumineuses pour de tels peptides et ces A1a ont été reportés jusqu'ici uniquement chez le pois. PA1a purifiée n'est pas toxique pour les charançons (Quillien et Delobel, données non publiées) et sa fonction reste inconnue.
L'absence de signal dans les fractions des autres plantes que le pois lors des essais d'immunodétection avec l'anticorps anti-PA1b est également surprenante. Les peptides présentent en effet une bonne homologie (au moins 60% entre MtrA1b007 et PsaA1b005). Cependant, PA1b s'avère peu antigénique et le titre du sérum ainsi que le seuil de détection sont médiocres. Cette technique n'est donc pas utilisable pour la recherche d'A1b au sein des Légumineuses. La comparaison des séquences et la structure de PA1b indiqueraient un possible épitope pour cet anticorps : le segment CC4 qui forme une boucle exposée, hydrophile et variable dans la structure.
Au contraire, l'approche par spectrométrie de masse MALDI sur des extraits relativement bruts se révèle d'une grande importance pour confirmer les séquences génomiques (en éliminant la faible variabilité pouvant être due à des erreurs de PCR) et leur expression et pour avoir une idée de la complexité peptidique de ces extraits. A part pour *M. truncatula*, toutes les séquences clonées semblent bien exprimées, aboutissant à des A1b canoniques dans les graines. Une analyse précise des masses devrait permettre de déterminer le nombre de gènes exprimés, qui est un caractère assez variable allant d'un minimum d'allèles ou loci estimé à 1-3 pour le soja, 2-4 pour le haricot, 5-10 pour le pois. La faible variabilité individuelle de séquence, ajoutée à l'identification de la localisation génomique principale des loci d'A1b sur le chromosome 6 du pois (EMBL AJ276882) semble

indiquer des évènements de duplication très récents aboutissant à des clusters de gènes d'Albumines 1 dans certaines espèces, possiblement liés à la domestication. L'analyse génétique en cours de ces loci estime à 7 les allèles d'A1b chez le pois, dont 6 seraient polymorphiques entre génotypes et coségrègent avec le groupe de liaison 6 (Domoney, communication personnelle). L'intensité de l'activité entomotoxique des graines de pois a été étudiée sur de nombreux génotypes (Mbaiguinam, 1996) et varie grandement. Toutefois, on ne sait pas dans quelle mesure cette caractéristique peut être déterminée par le niveau d'expression des A1b. Une telle variabilité d'activité entre génotypes et une structure génomique semblable ont été décrites récemment pour les inhibiteurs trypsiques de type Bowman-Birk des graines de pois (Page et al., 2002).

2) Cas de *M. truncatula*.

Pour *M. truncatula*, les résultats des différentes approches sont plus difficiles à interpréter. Concernant les tests biologiques, les fractions MeOH et H_2O5 sont les seules à présenter un différentiel significatif, bien que non total (forte toxicité sur R). Aucun peptide de type A1b n'a été mis en évidence dans les fractions apolaires de cette plante, soulevant deux hypothèses non exclusives : un différentiel de mortalité pourrait exister en l'absence d'A1b (MeOH), ou des homologues de PA1b existeraient dans les fractions H_2O (qu'il reste à analyser en spectrométrie de masse). La dernière hypothèse semble plausible car un peptide de masse supérieure a été détecté dans des extraits totaux de graines de *M. truncatula* par immunodétection avec un anticorps antiPA1a. L'analyse des séquences de *M. truncatula* indique qu'un variant de maturation post-traductionnelle (*ie* sans clivage du propeptide) est probable, car ce sont les seules séquences ne possédant pas la glycine C-terminale identifiée comme variant de maturation par spectrométrie dans les autres espèces mais présentant une proline en aval du site supposé de clivage. Un tel résidu est connu pour inhiber les enzymes endoprotéolytiques lorsque placé en position P'1 de leur substrat.

De plus, pour cette espèce végétale, aucune inhibition significative de liaison n'a pu être observée pour les fractions MeOH60 ainsi que MeOH, fraction qui présente pourtant un différentiel de toxicité. Ceci peut s'expliquer, soit par la présence dans les fractions apolaires d'inhibiteurs du test de liaison et de l'ionisation (pas de détection de peptides par spectrométrie), soit par la présence dans ces extraits d'un composé toxique non apparenté à PA1b, et n'entrant donc pas en compétition avec PA1b pour son site de liaison, mais pour lequel la souche R Chine serait également partiellement résistante. Ces deux hypothèses sont à considérer car les spectres de masse de *M. truncatula* sont à faible teneur en peptides mais présentent des groupes de pics caractéristiques d'oligomères saccharidiques (hexosamine) pouvant interférer avec l'ionisation des peptides. De plus, la souche R avait été récoltée en Chine sur un stock de graines non défini. Une analyse génétique et biochimique de l'interaction entre ces charançons et *M. truncatula* est donc requise pour comprendre leurs relations.

Bien qu'aucun peptide n'ait été détecté dans les fractions de graine de *M. truncatula,* les gènes identifiés dans cette espèce sont probablement fonctionnels, présentant dans leur partie 5' des signaux de régulation de la transcription (potentielle boite TATA). De plus, ces gènes sont probablement exprimés dans les graines de par leur forte homologie de séquence avec PA1b ou la leginsuline (>55% d'identité protéique), et de par l'absence d'étude des EST des graines en maturation dans cette espèce pourtant très étudiée. Cette espèce, appartenant au clade des Galégoïdes comprenant la tribu des Trifolieae et des Vicieae, est plus proche du pois que du haricot et du soja, deux Phaséoloïdes, clade

distinct. Ces relations phylogénétiques associées à la présence démontrée d'A1b chez le pois, le haricot et le soja confortent l'idée que *M. truncatula* exprimerait des homologues de PA1b.

IV. Conclusion

Au cours de ce travail, il a été démontré :

- une certaine difficulté d'analyse (anticorps polyclonal trop spécifique, interprétation des tests biologiques sans différentiel de toxicité R/S),
- mais, malgré tout, la présence de gènes homologues au sein des quatre espèces végétales étudiées,
- ainsi que la forte probabilité de la présence des produits d'expression de gènes homologues à PA1b dans les graines de trois de ces espèces (toxicité analogue et masses des peptides correspondants).

Ceci valide donc l'approche de recherche d'homologues par voie moléculaire, ainsi qu'un début de caractérisation biochimique et biologique des extraits peptidiques de graines. Au vu de ce travail, une étude taxonomique plus approfondie de cette nouvelle famille d'entomotoxines, au sein des Légumineuses, apparaît réalisable : elle laisse présager une potentielle aide à la compréhension de l'histoire évolutive de la famille peptidique, ainsi qu'une intéressante variabilité de structures et d'activités biologiques associées à cette famille. Ceci permettra peut-être d'isoler des homologues de faible CL50 ou toxiques pour les souches de *S. oryzae* résistantes et de mieux comprendre les déterminants structuraux de leur entomotoxicité.

Chapitre 3 : La famille des A1b au sein de la famille des Légumineuses.

La famille des A1b au sein de la famille des Légumineuses

L'approche testée et validée lors de la première partie de cette thèse est ici appliquée à plus large échelle taxonomique, afin de caractériser la variabilité des A1b au sein des Légumineuses et d'essayer de comprendre l'évolution de cette famille peptidique.
Après une rapide présentation des Légumineuses et de leur phylogénie en remaniement permanent, nous présenterons nos résultats sur la variabilité de structure primaire des A1b caractérisées au sein d'une portion de cette famille de plantes, puis sur la variabilité d'activité entomotoxique des graines de nombreuses espèces de cette famille.

I. Les Légumineuses

A. Systématique

1) Présentation de la famille des Légumineuses.

Les Légumineuses ou **Fabaceae** sont classées parmi les Angiospermes, Eudicotylédones, Core Eudicotylédones, Rosides, Eurosides I, Fabales. Elles sont sœurs des Polygalaceae, composant avec les familles Quillajaceae et Surianaceae, les Fabales (Doyle et al., 2000).
Il s'agit de la **troisième plus grande famille d'Angiospermes** en nombre d'espèces (après les Orchidaceae et les Asteraceae) avec plus de 18 000 espèces classées en 750 genres environ (ILDIS, 2001). C'est un des plus importants groupes de plantes pour l'homme, servant de cultures, de fourrages, d'engrais verts et produisant un grand nombre de composés utiles comme des médicaments, des poisons, des teintures ou des parfums.
Les Légumineuses sont représentées dans presque tous les milieux terrestres (excepté en Antarctique). Certaines vivent dans l'eau mais il n'existe pas d'espèces marines. Les espèces vont des herbes naines de l'arctique et des montagnes aux immenses arbres de la forêt tropicale (Figure 54).

Styphnolobium japonicum, une Papilionoideae arborescente

Bituminaria bituminosa, une Papilionoideae herbacée.

Figure 54 : Diversité des Légumineuses

La principale caractéristique unifiant ces espèces est le fruit, la **gousse**, modifiée de diverses façons pour faciliter le transport par le vent, l'eau et les animaux (Figure 55).

Figure 55: Gousses de : à gauche *Teline monspessulanus* (Papilionoideae), à droite : *Cercis siliquastrum* (Caesalpinioideae).

La famille est classiquement divisée en **trois sous-familles** : **Papilionoideae**, **Caesalpinioideae** et **Mimosoideae**, parfois considérées comme trois familles séparées. On les distingue le plus souvent par leurs fleurs (Figure 56).

- Les Papilionoideae consistuent la plus large sous-famille (environ 2/3 des genres) et la plus répandue. Elle contient la plupart des espèces cultivées. La majorité des espèces sont herbacées. Leur fleur est irrégulière, composée de 5 pétales : un étendard, deux ailes et deux pétales partiellement fusionnés en un carène.
- Les Mimosoideae sont pour la plupart des arbres tropicaux. Leurs fleurs sont régulières, petites, groupées souvent sous forme de pompon. Les étamines sont les parties les plus visibles de la fleur.
- Les Caesalpinioideae sont majoritairement des arbres ou arbustes tropicaux ou subtropicaux. Leur fleur irrégulière possède 5 pétales non différenciés et des étamines visibles extérieurement.

Figure 56 : Fleurs de Fabaceae
à gauche : *Onobrychis viciifolia* (Papilionoideae), au milieu : *Albizia julibrissin* (Mimosoideae) et à droite : *Caesalpinia gilliesii* (Caesalpinioideae)

2) Phylogénie

La taxonomie des Légumineuses est en constante évolution. Les sous-familles Papilionoideae et Mimosoideae correspondent véritablement à des groupes monophylétiques, inclus dans les Caesalpinioideae, paraphylétiques (Doyle et al., 2000 ; Doyle et al., 1997).
Chacune des sous-familles est divisée en tribus et sous-tribus définies d'après des caractéristiques morphologiques et sensées regrouper plusieurs genres apparentés. Pour les Papilionoideae, les différentes tribus et leurs relations définies par Polhill (1981) sont représentées Figure 57.

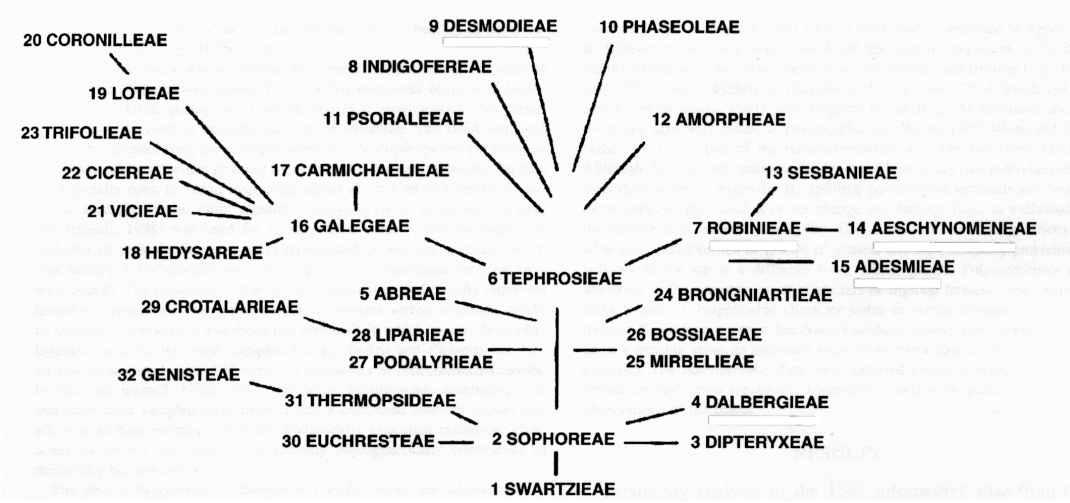


Figure 57 : Relations phylogénétiques putatives entre tribus au sein des Papilionoideae (Polhill, 1981).

Les études de gènes et de la structure de l'ADN pour retrouver les liens phylogénétiques entre espèces, genres ou familles confirment le plus souvent les résultats des études des caractéristiques développementales, morphologiques et biochimiques. Pour les légumineuses comme pour d'autres familles végétales, les gènes chloroplastiques *rbcL*, *matK* et *trnL*, ou les régions ITS nucléaires, sont d'efficaces outils pour la phylogénie au niveau de la famille jusqu'à l'espèce (Bruneau et al., 2001 ; Doyle et al., 2000 ; Hu et al., 2000 ; Kass and Wink, 1997 ; Lavin et al., 2001). Les arbres de gènes obtenus reflètent très souvent les arbres d'espèces (Doyle, 1992). Toutefois, les phylogénie moléculaires amènent parfois à reprendre les données classiques de taxonomie et mettent en évidence la paraphylie de certains groupements. Par exemple, une inversion de 50 kb dans l'ADN

chloroplastique est observée chez la plupart des Papilionoideae, absente chez les deux autres sous-familles, mais également chez certaines Swartzieae et Sophoreae. Cet événement unique chez les Légumineuses tend à montrer la polyphylie de ces deux tribus (Doyle et al., 1996), considérées comme les plus primitives des Papilionoideae. De plus, des analyses du gène chloroplastique *rbcL*, codant pour la grande sous-unité de la ribulose-1,5-diphosphate carboxylase/oxygénase et très utilisé pour les études phylogénétiques au sein des Angiospermes, montrent que les Sophoreae forment deux clades dont l'un se rapprocherait des Genisteae (Wink and Mohamed, 2003). Les Phaseoleae, une importante tribu des Papilionoideae ne semble monophylétique que si l'on y inclus une partie des Desmodieae et des Millettieae (Doyle and Doyle, 1993). Cette dernière tribu (Tephrosieae, Figure 57) était considérée comme basale aux Phaseoleae mais apparaît polyphylétique dans des études de phylogénie moléculaire, avec un cœur interne aux Phaseoleae (Hu et al., 2000).

Chapill a regroupé des données très diverses (morphologiques, développementales, de reproduction) sur les légumineuses et en tire un arbre phylogénétique présenté en annexe 2. Ce travail est en cours et il n'intègre pour l'instant pas de données d'analyses de gènes (http://www.botany.uwa.edu.au/systematics/legumephylogeny.html).
Il confirme cependant la polyphilie des Caesalpinioideae et la monophylie des Mimosoideae et des Papilionoideae. Vu le grand nombre de données intégrées dans l'analyse cladistique de Chapill et la représentation large des tribus et genres, et malgré le fait que ce travail soit en cours, nous considérerons la phylogénie découlant de cette analyse comme base de travail. Chapill définit parmi les Papilionoideae 5 groupes auxquels nous ferons référence :

- Le groupe non monophylétique des **Sophoroïdes**, basal et comprenant la plupart des Sophoreae (SP), les Swartziae (SZ), certaines Dalbergieae (DB) et les Amorpheae.
- Le clade des **Génistoïdes** comportant une partie des Sophoreae dont le groupe Sophora (*Bolusanthus*) qui apparait alors frère des Genisteae (lupin).
- Le clade des **Galégoïdes** dont les tribus ayant divergé le plus récemment seraient d'une part les Vicieae (pois), groupées avec le genre *Cicer* et d'autre part les Trifolieae (*Medicago*) et Loteae (lotiers).
- Le clade des **Aeschynoménoïdes** comportant une partie des Dalbergieae, les genres *Adesmia* et *Abrus,* et les Aeschynomeneae (AE ; arachide).
- Le clade des **Phaséoloïdes** qui comporte les Phaseoleae (PH ; haricot, soja), les Robinieae, la majorité des Desmodieae (DM), et le cœur des Milletieae qui dériveraient de la sous-tribu Diocleineae des Phaseoleae.

L'analyse de Chapill n'est parfois pas congruente avec les analyses de données moléculaires. On peut noter des divergences importantes (Figure 58):

- Tout d'abord au niveau de la "base" de la famille. Pour elle, ce sont trois groupes de Cassieae (CS) qui formeraient le clade le plus primitif des Fabaceae et non pas les Cercideae comme tendent à le montrer différentes analyses moléculaires (Bruneau et al., 2001 ; Doyle et al., 1997 ; Doyle and Luckow, 2003 ; Wink and Mohamed, 2003). Cette dernière tribu serait d'après Chapill la plus basale du clade regroupant la majorité des Caesalpinioideae et les Mimosoideae, clade frère de celui des Papilionoideae. Les Mimosoideae formeraient un clade dérivé, frère d'un clade regroupant la majorité des Amherstieae et la majorité des Detarieae, deux tribus de Caesalpinioideae. Les Detarieae sont au contraire considérées comme primitives par les analyses moléculaires, apparaissant avant la séparation des Papilionoideae d'avec

le clade des Mimosoideae-Caesalpinioideae évoluées (Bruneau et al., 2001 ; Doyle and Luckow, 2003).

- Mais aussi au niveau des Papilionoideae. L'analyse du gène *rbc*L chez les Papilionoideae conduit à une classification différant de l'analyse de Chapill pour la position des Aeschynoménoïdes, plus basales, et la position de certaines tribus dont Abreae, Robinieae et Loteae, regroupées en frères des Galegeae avec qui ils forment le clade connu sous le nom d'Hologalegina, mais aussi Mirbelieae (incluant Bossieae) qui forment un clade à la base des Galégoïdes et Phaséoloïdes.

Malgré ces quelques différences, on retrouve globalement les Papilionoideae tropicales « phaséoloïdes » d'un côté, et de l'autre, les plantes tempérées « galégoïdes » en deux groupes frères dérivés, ainsi que le clade des Génistoïdes (Wink and Mohamed, 2003). Une analyse du gène chloroplastique *trnK/matK* montre également la position basale de la plupart des Aeschynoménoïdes et le regroupement des Galégoïdes en frères des Phaséoloïdes incluant la majorité des Milletieae (Hu et al., 2000). Toutefois ces dernières apparaissent, dans cette étude, très largement polyphylétique contrairement à ce que propose Chapill pour laquelle seuls deux genres de Milletieae se rapprocheraient des Brongniartieae (Génistoïdes), les Wisterieae et les autres Milletieae formant un large clade des Phaséoloïdes.

Ainsi, si la structuration globale de la famille des Fabaceae ne subit plus de remaniements majeurs, les relations phylogénétiques entre certains groupes de tribus ou sous-tribus et l'appartenance de quelques genres à certaines tribus sont encore sujets à controverse. L'analyse d'un ensemble de données tant moléculaires que morphologiques devrait permettre de préciser certaines de ces relations.

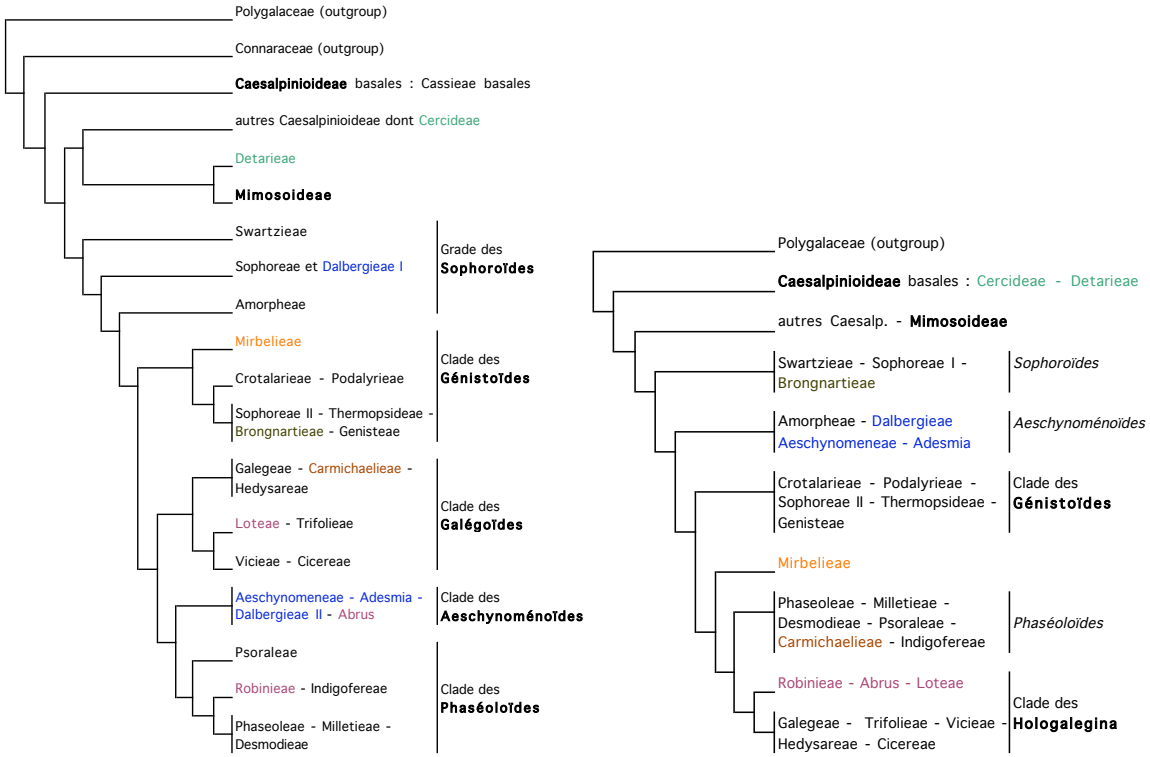


Figure 58: arbres phylogénétiques simplifiés des Fabaceae (dont Papilionoideae plus détaillées) d'après :
à gauche : Chapill (http://www.botany.uwa.edu.au/systematics/paps.html).
à droite : Wink and Mohamed (2003) : phylogénie moléculaire (gène *rbc*L).

B. Choix des espèces testées

Les critères appliqués par B. Delobel lors de la sélection des espèces étudiées sont doubles.

- La position phylogénétique : Les espèces choisies sont réparties au sein des différentes tribus des trois sous-familles selon la classification de Polhill (1994) puis complétées par rapport aux apports constants de la bibliographie.
- La disponibilité : seules des espèces dont nous pouvions nous procurer des graines ont pu être étudiées. Les fournisseurs sont multiples.

La liste des espèces choisies et leur provenance sont indiquées dans le tableau annexe 3.
L'ensemble des espèces sélectionnées n'a pas pu être traité par les deux approches (moléculaire et biologique) pour des raisons techniques (impossibilité d'extraire de l'ADN) et de temps (durée des tests biologiques, élevage). De plus, les Papilionoideae apparaissent bien plus représentées que les deux autres sous-familles. Ceci est dû d'une part au fait que les Papilionoideae sont les plus nombreuses des Fabaceae et d'autre part aux résultats de la biologie moléculaire qui nous a amené à échantillonner plus d'espèces dans les clades Galégoïde et Phaséoloïde de cette sous-famille (cf II.B.4)).

II. Diversité structurale primaire

A. Matériels et Méthodes

Seules les différences d'avec le chapitre correspondant de la première partie de ce travail sont détaillées ici.

1) Matériel végétal

Les fournisseurs des différentes espèces choisies sont indiqués en annexe 3.

2) Extraction d'ADN

Pour la plupart des plantes, les graines mises à germer (par B. Delobel) ont donné des jeunes feuilles d'où l'ADN génomique a été extrait par I. Rahoui pour la majorité des espèces. La méthode utilisant le CTAB (Ausubel et al., 1997) présentée dans la première partie de cette thèse a d'abord été utilisée. Pour certaines plantes, les graines ne germant pas, l'ADN a été extrait directement de celles-ci, par la même méthode.
Pour quelques plantes, tropicales, l'ADN n'a pu être extrait par cette méthode. Un protocole également basé sur l'utilisation du CTAB mais adapté aux plantes tropicales (Scott and Playford, 1996) a été testé pour certaines de ces espèces mais sans succès, malgré une correction de la composition du tampon CTAB (concentration en NaCl et CTAB trop élevées). L'utilisation du kit Promega (Wizard Genomic DNA Purification Kit) pour certaines plantes négatives ayant donné de bons résultats et facilitant la manipulation, celui-ci a été adopté pour extraire les ADN par la suite.

3) Recherche de fragments de gènes homologues par PCR

Les 3 amorces dégénérées déterminées dans la première partie ont été utilisées dans les mêmes conditions de PCR sur les différentes plantes testées, en utilisant de préférence l'amorce reverse R1 puis R3 en cas d'absence de signal dans le premier cas. De plus différentes températures d'appariement (Ta) de PCR ont été testées sur les plantes non réactives (cf II.B.2)(a)).
Deux amorces supplémentaires ont été déterminées. Il s'agit des amorces reverses dégénérées Rev6 et Rev5, placées dans la région C-terminale de PA1a, de séquences suivantes :

Rev6 : $^{5'}$ GCTTCAGAWTYAGAGKCAAARCACCACCRTA $^{3'}$ (32° de dégénérescence)
Rev5 : $^{5'}$ ATTIGGRWAAYGASCRCARAARYAICCACTTYYYTT $^{3'}$ (2048°)

Ces deux nouvelles amorces basées sur les séquences publiées du soja et du pois ainsi que sur les séquences de *M. truncatula* complétées en 3' (cf. deuxième chapitre), ont été réalisées dans le but d'adapter la marche sur le gène à la recherche primaire de séquences homologues dans certaines plantes non réactives aux différentes conditions de PCR classique. En s'affranchissant de l'amorce For1 basée sur un segment de PA1b et en ne se basant que sur PA1a, mieux conservée que PA1b pour les séquences connues, nous espérions obtenir des homologues de PA1b plus variants (dans des espèces phylogénétiquement plus éloignées). Les banques de digestion ont été réalisées de la même façon que pour la marche sur le gène. Les deux PCR nichées ont alors été réalisées avec comme amorces spécifiques Rev6 et Rev5, dans les mêmes conditions que

recommandées pour une marche sur le gène « classique », avec toutefois une concentration en amorces dix fois plus élevée que pour des amorces non dégénérées (3 µM final).
L'amorce Rev6 a également été utilisée en PCR classique combinée avec For1 (mêmes conditions qu'avec Rev1, Ta = 50°C) pour les espèces testées après sa création, car elle permet d'obtenir des séquences plus longues au sein de la partie A1a.

4) Recherche de séquences amont

Pour rechercher les séquences en 5' des fragments de gènes obtenus, le kit Universal GenomeWalker (Clontech, USA) a été utilisé dans les conditions recommandées par le fournisseur, en se limitant toutefois à trois banques de digestion par plante (contre quatre) pour économiser ADN et enzymes. Les enzymes choisies au vu des résultats obtenus sur les quatre premières plantes testées sont *Dra*I, *Eco*RV et *Pvu*II (RocheDiagnostics). La température d'annealing (67°C) préconisée dans le kit a parfois dû être abaissée à 65°C, et le nombre de cycle augmenté à 35 (contre 30), comme déterminé dans la première partie de ce travail.
Pour chaque création de banques de digestion (plusieurs espèces groupées), une banque témoin a été réalisée. Il s'agit de la banque *Dra*I de *M. truncatula*. Celle-ci a ensuite été soumise aux mêmes conditions de PCR que les banques créées simultanément, avec les amorces spécifiques du variant MtrA1b007 (caractérisé dans la première partie de ce travail). Cette banque et ce jeu de primers ont été choisis pour l'intensité et l'unicité de la bande obtenue quelles que soient les conditions des PCR nichées.

5) Clonage et séquençage

Les clonages des fragments de gènes obtenus par PCR et des séquences amont de ceux-ci ont été effectués dans le vecteur PCR 2.1 TOPO et les cellules électrocompétentes TOP10 selon les recommandations du fournisseur (Topo TA Cloning, Invitrogen).
Le séquençage de l'ADN plasmidique a été effectué par Genome Express S.A. (Grenoble).

6) Northern Blot

Différents protocoles ont été testés pour extraire les ARNs de graines en cours de maturation de quelques espèces pour lesquelles on en disposait. Certains se sont révélés infructueux (gélification des échantillons liée à la présence d'amidon dans les graines, qui réagit avec l'isothiocyanate de guanidine). Le protocole proposé par Salzman et al. (1999) a finalement été utilisé avec succès par A. Vallier. Après vérification de la qualité et dosage au spectrophotomètre des ARNs extraits, elle a réalisé une migration des ARNs en vue d'un transfert sur membrane de Nylon, réalisé sur le week-end dans du SSC 20X (standard sodium citraté : NaCl 3M ; citrate de sodium 0,3M) et eau DEPC (diéthylpyrocarbonate). La membrane a ensuite été cuite 24h à 80°C. Un cDNA complet d'une isoforme de PA1 du pois (407 pb), cloné dans un plasmide PCR2.1 TOPO et isolé par PCR avec les amorces M13 reverse et T7 forward (587 pb en tout), a été utilisé comme matrice pour la production de la sonde marquée au ^{32}P par random priming (Prime a Gene Labeling System, Promega). Une préhybridation au sperme de saumon et hareng a été réalisée sur la nuit à 42°C puis la membrane a été rincée au 2X SSC 0,1 % SDS 10 min puis 15 min. La sonde marquée purifiée a alors été placée en contact de la membrane, sur la nuit, à 42°C. Après hybridation, la membrane a été rincée à minima puis mise en contact d'une cassette photosensible (STORM), sur la nuit, avant lecture.

7) Analyses des séquences

Les séquences protéiques complètes et incomplètes spécifiant des homologues de PA1b ont été alignées en utilisant le logiciel Clustal W (Thompson et al., 1994). L'alignement a été ajusté à la main (conservation des positions des cystéines). L'alignement des séquences nucléiques est déduit de celui des séquences protéiques.
L'alignement nucléique a ensuite été analysé en utilisant le logiciel Phylo_win (Galtier et al., 1996) afin d'obtenir les pourcentages d'homologie entre séquences, et une représentation des liens phylogénétiques entre celles-ci (arbre de phylogénie moléculaire) déterminés grâce à l'algorithme BioNJ (Gascuel, 1997) et la méthode de calcul de distances Galtier et Gouy (Galtier et al., 1995).

B. Résultats

1) Extractions d'ADN

57 (+4 non séquencés) nouveaux ADNs génomiques d'espèces de Légumineuses ont été extraits avec succès pour 5 Caesalpinioideae, 5 Mimosoideae, 3 Papilionoideae basales, 12 Génistoïdes, 15 Galégoïdes, 3 Aeschynoménoïdes et 14 Phaséoloïdes. La majorité ont été extraits par la méthode du CTAB à partir de feuilles fraîches ou congelées. Pour le fenugrec, cette méthode a été appliquée avec succès sur les graines. Malheureusement, la même opération pour *Gompholobium scabrum* a échoué.
L'ADN génomique de certaines plantes, non extrait au CTAB, a pu l'être en suivant les recommandations du kit Wizard de Promega (cas d'*Abrus precatorius, Cercis siliquastrum, Templetonia retusa*).
L'ADN de sept des plantes sélectionnées (dont *Gompholobium scabrum* (Papilionoideae)*, Tamarindus indica* (Caesalpinioideae)*, Parkia biglobosa* (Mimosoideae)) n'a pu être extrait, malgré différents protocoles testés.

2) Recherche de gènes homologues

(a) Approche par PCR classique

Sur les 61 ADNs purifiés, 10 n'ont montré aucune amplification quelles que soient les conditions de PCR ou le jeu d'amorces utilisés. Pour de nombreuses plantes pour lesquelles 50°C de température d'hybridation ne permettait pas d'amplification, nous avons testé différentes températures variant entre 40°C et 50°C avec succès dans certains cas (espèces des Galégoïdes et Phaséoloïdes). La Figure 59 montre l'importance de la température d'annealing. A une température trop élevée, on ne détecte pas de signal pour A*morpha fruticosa,* mais en diminuant celle-ci, une bande bien définie, bien que de taille supérieure à celle attendue, apparaît. Dans le cas du fenugrec, pour lequel l'amplification à 50°C donne un signal de la taille attendue bien que faible, une diminution de la température entraîne une augmentation de l'intensité du signal attendu, jusqu'à un certain seuil où celle-ci diminue alors, au profit d'un produit de taille supérieure, qui s'avère aspécifique (clonage et séquençage effectués).

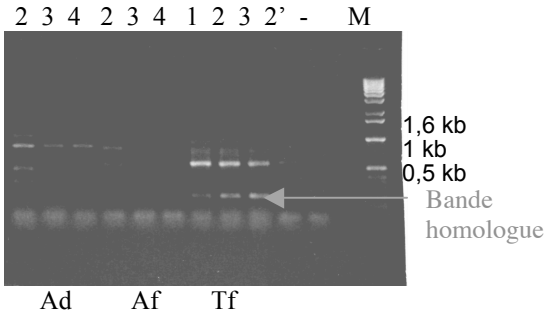


Figure 59 : PCR avec les amorces For1 et Rev1 sur ADN génomique d'Ad : *Acacia dealbata*, Af : *Amorpha fruticosa*, Tf : fenugrec (à 10 et ' : 2 ng d'ADN), à différentes Ta. 1: 44°C; 2: 45°C; 3: 46°C; 4: 47°C.

Certaines espèces au signal non spécifique ont été soumises à une PCR Touch-Down (TD) qui consiste à appliquer une température d'annealing élevée lors des premiers cycles (spécificité) puis à diminuer la température de façon progressive sur quelques cycles et à appliquer pour les cycles suivants majoritaires (amplification) une température plus faible. Toutefois, cette technique ne nous a pas permis d'obtenir de meilleurs résultats. Pour sept de ces espèces (*Robinia pseudoacacia, Amorpha fruticosa, Abrus precatorius, Bolusanthus speciosus, Wisteria sinensis, Lotus tetragonolobus* et *Arachis hypogaea),* réparties dans les 5 groupes de Papilionoideae, et dont des produits de PCR de taille inférieure (rare) ou supérieure à celle attendue ont été obtenus (Figure 60), un clonage et un séquençage ont été réalisés ne révélant à chaque fois que des séquences non codantes.

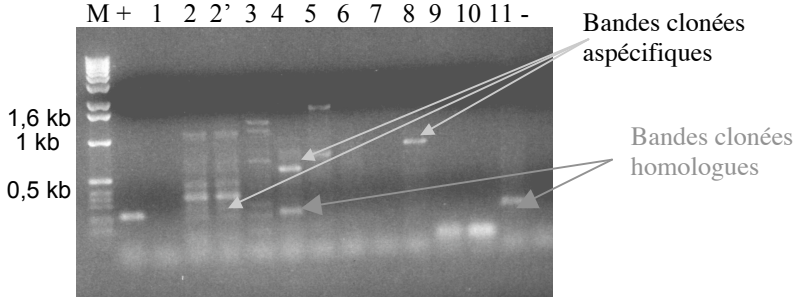


Figure 60 : PCR avec les amorces For1 et Rev1 sur ADN génomique de différentes espèces végétales (Ta=40°C).
M : marqueur de poids moléculaire ; + : témoin positif (pois) ; 1 : *Albizia julibrissin* ; 2 : *Arachis hypogaea* (pour 10 et ' : 20 ng d'ADN) ; 3 : *Dalbergia purpurescens* ; 4 : *Trigonella foenum-graecum* ; 5 : *Intsia bijuga* ; 6 : *Lotus corniculatus* ; 7 : *Lupinus albus* ; 8 : *Robinia pseudoacacia* ; 9 *: Schotia afra* ; 10 : *Sesbania sesban* ; 11 : *Vigna unguiculata* ; - : témoin négatif H_2O.

(b) Approche de type marche sur le gène

Les amorces Rev5 et Rev6 déterminées en vue de tenter une approche de type marche sur le gène sur certaines plantes non réactives aux PCR classiques, ont tout d'abord été testées en PCR simple avec For1 sur le pois et le soja. L'amplification obtenue est correcte (taille et intensité) pour Rev6 mais très faible, bien que spécifique (taille de fragment attendu), pour Rev5.

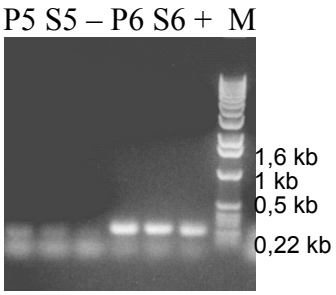


Figure 61 : PCR For1-Rev5 (5) ou -Rev6 (6), test des amorces sur pois (P) et soja (S). - : témoin négatif H20 ; + : témoin For1-Rev1 pois.

Le résultat peu encourageant obtenu pour Rev5 était craint. En effet, l'analyse structurale par le logiciel Oligo (Molecular Biology Insights, Inc.) avait révélé un grand nombre de repliements possibles de l'amorce (en épingles à cheveux), dont un particulièrement stable (de Tm = 25°C). L'existence de formes repliées de l'amorce limite l'amplification. De plus, le degré de dégénérescence de Rev 5 est particulièrement élevé (2048 plus 2 inosines) ce qui rend, là aussi, le nombre d'amorces utiles faible pour une concentration habituelle (3µM pour les amorces Rev1, Rev3 et For1). Toutefois, les caractéristiques nécessaires de ces deux amorces n'ont pas permis de déterminer un remplaçant à Rev5. En effet, elles doivent respecter un certain nombre de critères (longueur, % GC) pour l'utilisation en « marche sur le gène » (température d'appariement élevée) et être placées dans des zones les mieux conservées de PA1a. Les deux amorces ont donc été utilisées en « marche sur le gène », pour laquelle les températures d'appariement (72°C puis 67°C) sont plus élevées que celle de la PCR classique (50°C), ce qui devrait limiter un peu l'effet négatif de l'autohybridation de Rev5. Un premier essai a été réalisé sur une plante témoin, *M. truncatula*, banque *Dra*I. La première PCR laisse apercevoir une trainée, semblable au témoin positif (avec amorces non dégénérées). La deuxième étape a donc été réalisée avec les mêmes températures d'appariement aboutissant à deux bandes, une d'intensité moyenne de 300 pb, et une d'intensité plus faible de 1000 pb (Figure 62). Ce résultat nous a encouragé à tester cette approche sur des plantes pour lesquelles les différentes conditions de PCR classique sont restées vaines et pour lesquelles nous disposons d'assez d'ADN. Il s'agit des espèces *Abrus precatorius*, *Priestleya myrtifolia*, *Wisteria sinensis*, *Podalyria biflora* et *Mundulea sericea*, comme deuxième témoin puisque des séquences homologues ont été caractérisées par PCR chez cette espèce. La première PCR impliquant Rev6 a à nouveau révélé une trainée, mais la deuxième PCR n'a pas permis de détecter d'amplification ailleurs que dans le témoin luzerne.

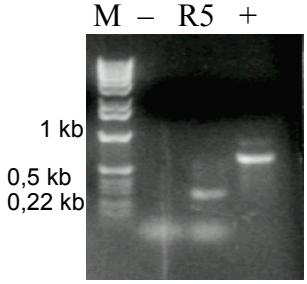


Figure 62: PCR nichée (deuxième) de marche sur le gène avec l'amorce Rev5 (R5) sur *M. truncatula* banque *Dra*I. - : témoin négatif (H20); + : témoin positif (amorces spécifiques)

(c) Cas du lupin

Deux espèces de lupin ont été testées lors de l'approche classique par PCR et aucune n'a révélé d'amplification, quelques soient les conditions de PCR (amorces ou Ta) testées.
Or, en ce qui concerne le lupin à feuilles étroites (*Lupinus angustifolius* cv Unicrop) une séquence homologue à PA1b avait été décrite par Ilgoutz et al. (1997). Nous avons donc testé la qualité de l'ADN de lupin extrait par une amplification d'un fragment de gène de l'ADH des Légumineuses. Le signal obtenu, de la taille attendue, témoigne de l'absence d'inhibiteurs de la polymérase dans cet ADN (Figure 64).

La séquence de lupin publiée a été obtenue avec une amorce For de séquence identique aux acides aminés GACSPFE (Gayler, K.A. communication personnelle) du cDNA de la leginsuline du soja publiée par Watanabe et al. (1994). Elle présente ensuite une isoleucine à la place de la méthionine du soja et de la majorité des autres gènes caractérisés. Cette différence dans la zone utilisée pour ancrer l'amorce For1 de notre travail (cf alignement Figure 67) ne devrait cependant pas empêcher For1 de se fixer (dégénérescences suffisantes). Toutefois, l'absence de la troisième cystéine (après le doublet de proline) correspond à la mutation de la guanine (en position deux du codon) en cytosine. Cette mutation correspondant à l'extrémité 3' de For1 pourrait fortement inhiber l'amplification de ce gène par notre amorce. Afin de contourner ce problème, deux amorces nichées, spécifiques de la séquence publiée par Ilgoutz et al., ont été déterminées et la technique de marche sur le gène a été appliquée aux différentes banques de digestion crées pour les deux espèces de lupin testées. Un produit majoritaire d'environ 1400 pb a été obtenu dans la banque *Pvu*II du lupin blanc alors qu'aucune bande nette n'a pu être observée pour le lupin à feuilles étroites, pourtant de la même variété que celui d'Ilgoutz et al. (Figure 63). La bande obtenue pour le lupin blanc a été clonée et séquencée, mais s'avère résulter d'une amplification non spécifique.

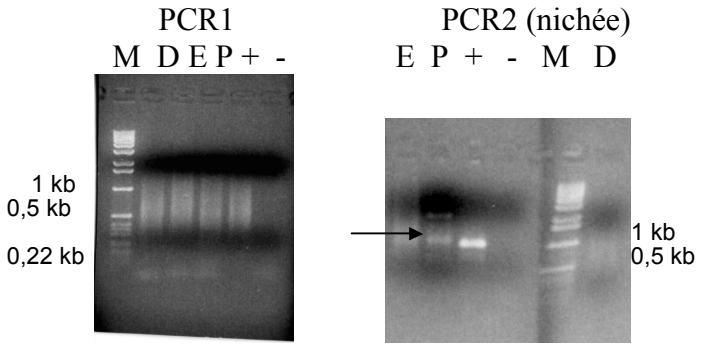


Figure 63 : résultats des PCRs de marche sur le gène réalisées sur le lupin blanc. M: marqueur, +: témoin positif, -: témoin négatif, D: banque *Dra*I, P: banque *Pvu*II, E: banque *Eco*RV.
La bande indiquée par une flèche observée dans la banque *Pvu*II a été clonée.

Finalement, pour conclure au sujet de cette séquence de lupin à feuilles étroites, nous avons réalisé sur cette espèce, une PCR avec les amorces et les conditions exactes décrites par Ilgoutz et ses collaborateurs. Malheureusement, là encore, nous n'avons pas pu obtenir de produit de PCR de la taille définie par les auteurs (280 pb) (Figure 64).

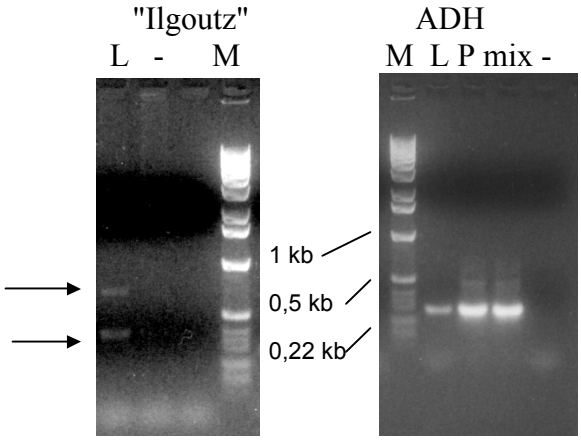


Figure 64 : résultat des PCRs sur ADN de lupin à feuilles étroites réalisées
-selon le protocole décrit par Ilgoutz et al. (1997) : les flèches indiquent les deux produits observés, de taille supérieure à 280 pb.
- avec les amorces spécifique du gène de l'ADH.
M: marqueur, -: témoin négatif, L: ADN de lupin, P: ADN de pois, mix: ADN de lupin et de pois mélangés.

(d) Recherche d'homologues par Northern Blot

Quelques espèces ont été sélectionnées pour leur position taxonomique et la possibilité d'obtenir des graines en cours de maturation (à mi-remplissage). Il s'agit des Mimosoideae *Albizia julibrissin* et *Desmanthus illionensis,* de la Caesalpinioideae *Cercis siliquastrum*, des Génistoïdes *Baptisia australis* et *Lupinus angustifolius* (non montré), des Galégoïdes *Caragana arborescens* et *M. truncatula* (non montré), de la Phaséoloïde *Desmodium canadense,* et du pois témoin. A partir de leurs ARNs, un Northern Blot a été réalisé avec le gène entier codant pour PA1 comme sonde. Le pois, témoin, révèle une bande unique et bien définie alors qu'aucune autre des espèces testées ne présente d'hybridation (Figure 65). Ceci peut paraître étonnant pour les deux Galégoïdes dont des séquences homologues génomiques ont été caractérisées. Toutefois, ces courtes séquences présentent, au niveau nucléique, un pourcentage d'homologie avec les séquences du pois d'environ 70% seulement, ce qui peut expliquer que la sonde PA1 ne s'hybride pas. Une autre hypothèse concerne l'expression des gènes caractérisés. Il est possible que les séquences génomiques ne soient pas du tout exprimées ou exprimées dans d'autres organes des végétaux que la graine. Cette hypothèse tend à être confimée pour l'instant, pour *M. truncatula,* par une RT-PCR réalisée sur l'ARN de graines de cette luzerne méditerranéenne avec des amorces spécifiques des séquences génomiques identifiées par PCR (cf. deuxième chapitre) et qui n'a pas abouti, contrairement à la même expérience sur pois.

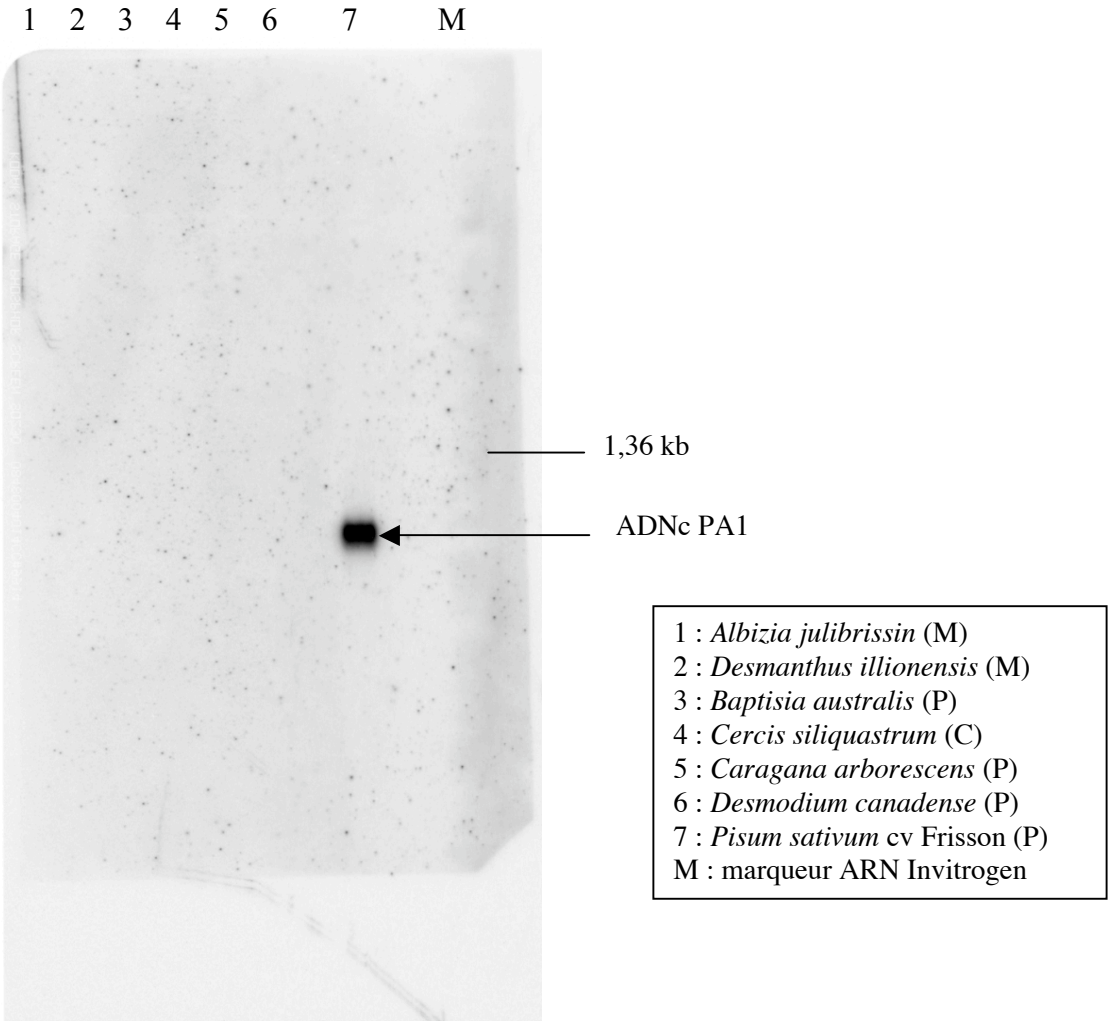


Figure 65 : Northern blot avec pour sonde marquée un cDNA d'A1 du pois. M : Mimosoideae, C : Caesalpinioideae, P : Papilionoideae.

3) Marche sur le gène

Sur les 20 nouvelles séquences obtenues par PCR avec amorces dégénérées, nous n'avons pu compléter en 5' que 11 gènes correspondants à 10 espèces végétales.
Pour chaque plante trois banques ont été réalisées et soumises aux deux PCR nichées avec une amorce spécifique exactement complémentaire d'une fraction des séquences obtenues avec les amorces dégénérées. Les températures d'appariement de 72°C pour les 5 premiers cycles et de 67°C pour les 25 ou 35 suivants ont été utilisées dans la plupart des cas. Toutefois, pour certaines espèces, ces conditions spécifiques (recommandées dans le kit) n'ont pas amené de produits identifiables pour un clonage. Un aménagement des PCR nichées par diminution progressive (1°/cycles sur 7 cycles puis 65°C pour les cycles suivants) de la température d'appariement, élevée au départ (72°C) a été tenté afin de conserver la spécificité des premiers cycles tout en augmentant le rendement d'amplification (= TD65). Les résultats de ces deux approches sont comparés à la Figure 66. On constate que les conditions de Touch Down s'avèrent efficaces pour les espèces *B. bituminosa, C. arborescens, C. ensiformis et V. subterranea*. Pour *V. unguiculata* et *L. angustifolius*, des échelles persistent : aucune des conditions testées n'ont permis d'obtenir un produit identifiable et clonable.

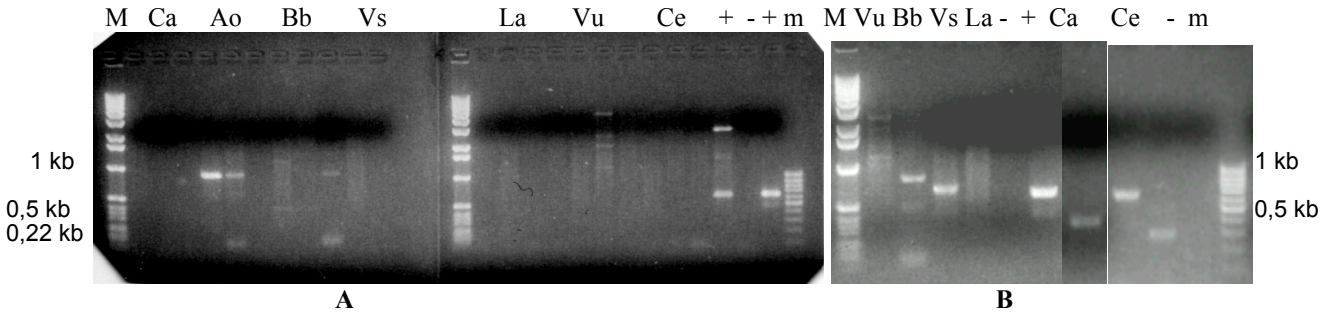


Figure 66 : Marche sur le gène. A : PCR nichée (Ta:67°C) sur les trois banques DraI, EcoRV, PvuII de Ca: *Caragana arborescens*, Ao: *Alysicarpus ovalifolius*, Bb: *Bituminaria bituminosa*, Vs: *Vigna subterranea*, La: *Lupinus angustifolius*, Vu: *Vigna unguiculata*, Ce: *Canavalia ensiformis*.
B : PCR nichée (TD65) sur les banques *Eco*RV de Vu et Ca, *Pvu*II de Bb, *Dra*I de Vs et La et les trois banques de Ce.
+ : témoins positifs, - : témoin négatif. M, m : marqueurs.

L'absence de résultats par la technique de marche sur le gène, pour certaines séquences, peut s'expliquer pour différentes raisons. La présence de plusieurs séquences peu divergentes pour une même espèce ne permet pas de déterminer des amorces suffisamment spécifiques pour l'une ou l'autre des séquences, ce qui amène à des fragments de séquences non compatibles avec celles des clones de départ ou entre elles (deux clones sont séquencés par amplification), indiquant par ailleurs, la présence de plusieurs allèles de séquence proche. De plus, les régions variables choisies de préférence pour ancrer les amorces de la marche sur le gène (spécificité) ont parfois une structure non adaptée au design de ces amorces (formation de structures en épingles à cheveux, parfois stables) ou bien sont incompatibles avec les amorces des adapteurs universels (séquence partiellement complémentaire). Ces problèmes sont particulièrement aigus dans notre cas vu la petite taille des séquences obtenues par PCR (130 à 220 pb) qui limite grandement le choix de zones spécifiques et structuralement correctes pour ancrer deux amorces non chevauchantes, aux caractéristiques spécifiques (longueur, pourcentage de GC, extrémité 3' en C).
De plus, pour certaines séquences, le produit majoritaire obtenu n'a pas permis de compléter la séquence de l'A1b et du peptide signal, soit qu'il se fut agit d'un produit aspécifique non codant (cas de *C. arborescens*), soit que la séquence obtenue présentât un profil correct pour une portion puis un codon stop et une portion non codante, comme c'est le cas pour *Vigna subterranea*.

4) Analyse des séquences

(a) Variabilité des séquences

L'alignement protéique de toutes les séquences obtenues ainsi que de séquences des banques de données est montré Figure 67.
Cet alignement permet de distinguer les zones variables et les zones conservées entre A1b, précisant et renforçant les conclusions à ce sujet tirées de l'étude des séquences des quatre plantes prototypes (cf. deuxième chapitre). On constate en effet que les **acides aminés conservés** entre les séquences des quatre « témoins » le restent lors de l'élargissement de l'étude. Considérons la numérotation des résidus d'après la séquence de l'A1b du pois M13709. Il s'agit en particulier :

- des cystéines (exceptée C32, remplacée par un Y pour une séquence de *Mundulea sericea*, comme observé dans la première partie de ce travail chez *M. truncatula*),
- des prolines (exceptées P9 remplacée par une sérine chez le melilot et P14 transformée en leucine pour AovA1b001),
- de l'arginine 21 uniquement remplacée par une cystéine chez *V. subterranea* et une leucine chez *V. radiata*,
- des leucine 27 et glycine 30 parfaitement conservées et,
- de l'acidité (négativité) du résidu 11 (E, D), qui reste seulement polarité (Q) chez *V. radiata* et *B. bituminosa*.

Différentes variations apparaissent, en nombre et nature de résidus.
En considérant la maturation de type pois des séquences d'A1 (clivage post-alanine en C-terminal du peptide signal et post-glycine en C-terminal d'A1b), le segment nC1 apparaît tronqué de deux acides aminés pour une séquence d'*O. viciifolia*, de *C. brasiliensis* et de *B. bituminosa*. Le segment CC2 présente au contraire deux acides aminés supplémentaires dans les séquences de Milletieae. Ces dernières possèdent également un résidu surnuméraire dans le segment CC4, comme déjà observé chez le haricot. Le peptide intermédiaire putatif (séparant PA1b de PA1a) varie également en longueur par rapport aux séquences du pois, présentant de 1 à 4 résidus supplémentaires pour de nombreuses espèces et un résidu en moins pour *V. hirsuta.*
En plus des mutations citées ci-dessus, on peut remarquer le remplacement de la sérine S8 par une phénylalanine plus encombrante chez *B. bituminosa* ainsi que la présence d'une proline au lieu d'une glycine/sérine en C-terminal (putatif) d'une séquence de *Mundulea sericea*, séquence présentant une autre proline dans ce segment C-terminal (remplaçant N34 et formant un doublet de prolines avec P35). Concernant les prolines, on constate que le fenugrec et le mélilot, tous deux de la même tribu que *M. truncatula*, offrent des séquences présentant la même proline que cette luzerne, en aval du site putatif de clivage de PA1b et du peptide intermédiaire. Ce résidu semblerait donc signer les albumines 1 des Trifolieae, bien qu'une étude plus élargie reste nécessaire pour conclure. Les séquences des Milletieae semblent quant à elles, signées, en plus de leurs résidus surnuméraires, par la présence d'un aspartate et d'une alanine en lieu et place de la sérine-thréonine 17 et du résidu 18 plus variable. On peut également observer la présence d'un résidu aromatique, à la place d'un aliphatique moins encombrant, dans le segment hydrophobe CC6 pour 10 des séquences caractérisées (tryptophane ou tyrosine comme observée chez le haricot).
En comparant les séquences obtenues, on remarque également la variabilité des peptides signal, incompatible avec la détermination d'amorces PCR qui nous auraient permis d'éviter la marche sur le gène. Au contraire, on constate que PA1a reste fortement conservée.

Le pourcentage d'identité au niveau de l'ADN pour nos séquences varie pour les séquences complètes en 5' (du peptide signal au fragment de PA1a inclus) de 82% d'identité entre le pois et *Vicia hirsuta* à 61% entre le soja et *Alysicarpus ovalifolius*. Pour les séquences tronquées, le pourcentage d'homologie entre les deux séquences du pois est de 97%, alors qu'entre la séquence PvuA1b041 et CenA1b001 ou celle de *Caragana arborescens*, il n'est que de 65%.
Au niveau protéique, pour la partie A1b uniquement, le pourcentage d'homologie varie de 64% entre PsaA1b012 et MseA1b003 ou AovA1b001 à 90% entre PsaA1b012 et PsaA1b036.

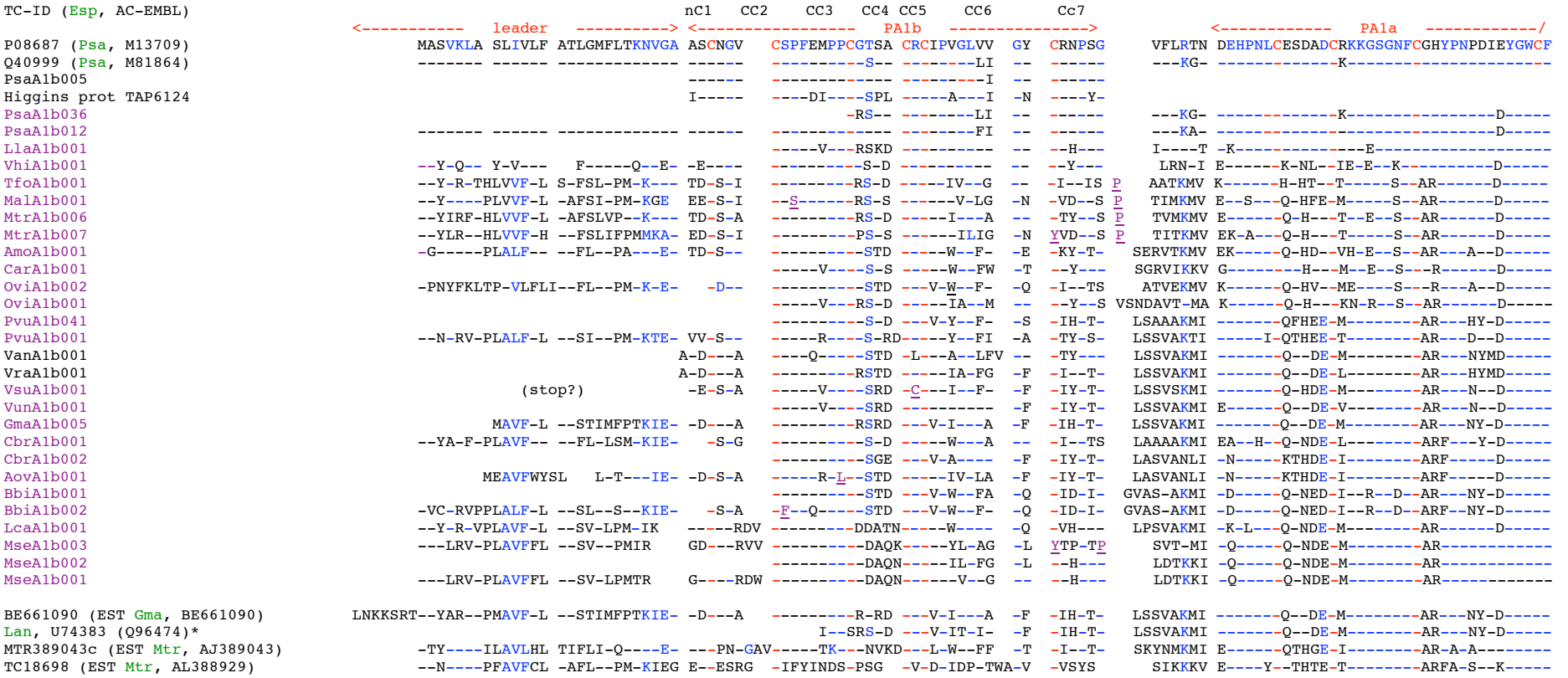


Figure 67 : alignement protéique des séquences caractérisées dans notre travail (avec séquences des amorces) et de séquences de la littérature ou des banques EST. Seuls les acides aminés différant de M13709 sont indiqués : - représente le même résidu que pour la séquence M13709. Les espèces sont : Amo : *Astragalus monspessulanus*, Aov : *Alysicarpus ovalifolius*, Bbi : *Bituminaria bituminosa*, Car : *Caragana arborescens*, Cbr : *Canavalia brasiliensis*, Gma : *Glycine max*, Lan : *Lupinus angustifolius*, Lca : *Lonchocarpus capassa*, Lco : *Lotus corniculatus*, Lla : *Lathyrus latifolius*, Mal : *Melilotus albus*, Mse : *Mundulea sericea*, Mtr : *Medicago truncatula*, Ovi : *Onobrychis viciifolia*, Psa : *Pisum sativum*, Pvu : *Phaseolus vulgaris*, Tfo: *Trifolium foenum-graecum*, Van : *Vigna angularis*, Vhi : *Vicia hirsuta*, Vra : *Vigna radiata*, Vsu : *Vigna subterranea*, Vun : *Vigna unguiculata*.

(b) Phylogénie moléculaire

Une analyse phylogénétique a été réalisée sur les 26 séquences nucléiques incomplètes résultant de l'amplification par les amorces F1R1 (incluant les séquences des quatre espèces « pionnières ») et quelques séquences déjà publiées (deux séquences du pois, une du soja, une du lupin, une de *Vigna radiata* et une de *V. angularis*). Le nombre de sites informatifs est limité par la petite taille des séquences. En effet, seulement 104 sites sur 210 sont informatifs (160 complets). Ils ont été analysés en utilisant l'algorithme BioNJ (Gascuel, 1997) et la méthode de calcul de distances Galtier et Gouy (Galtier et al., 1995). L'arbre consensus obtenu (non raciné), présenté Figure 68, montre la séparation en **deux groupes principaux** des séquences : celles appartenant à des espèces du clade des Phaséoloïdes et celles des espèces de Galégoïdes.

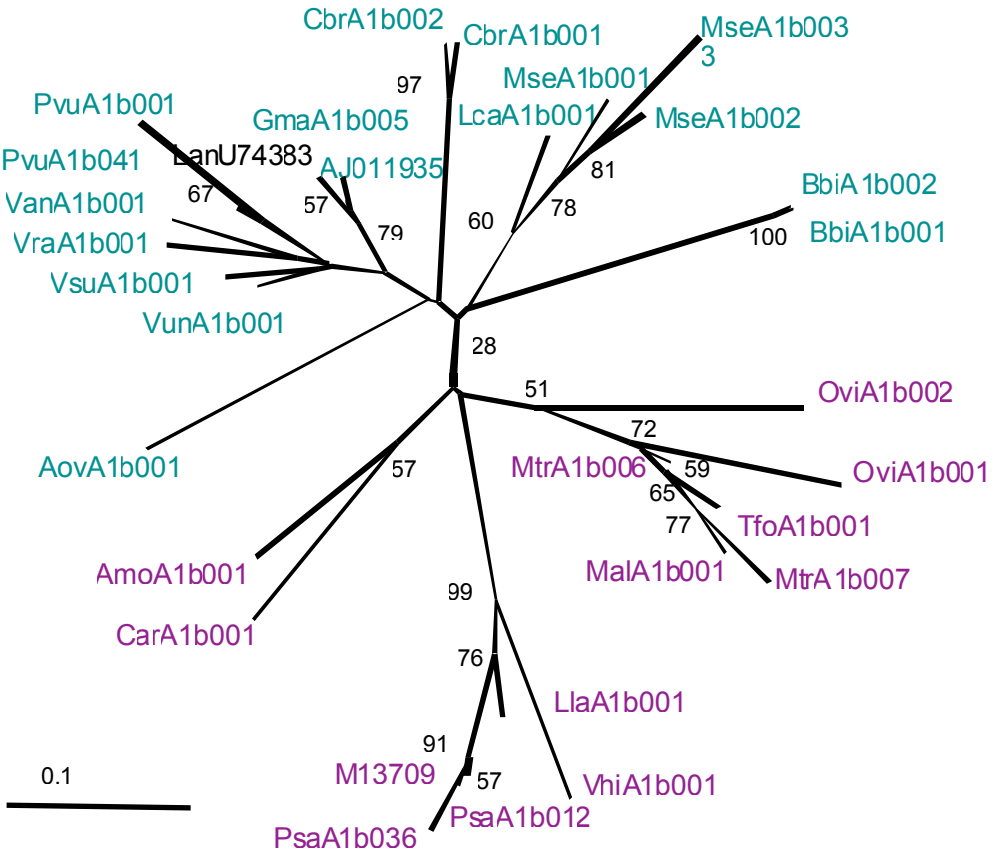


Figure 68 : arbre de phylogénie moléculaire (séquences "incomplètes", BioNJ, distance Galtier et Gouy). En vert : séquences de Phaséoloïdes, en pourpre : Galégoïdes. Seules les valeurs de bootstrap supérieure à 50 sont indiquées. Les noms des séquences sont légendés Figure 67.

- Au sein des Galégoïdes, les séquences des **Vicieae** (Vhi, Lla, Psa) se regroupent (soutien important : bootstrap de 99) ainsi que celles des **Trifolieae** (Mtr, Tfo, Mal) (bootstrap de 59) auxquelles se branchent les deux séquences d'*Onobrychis viciifolia*. Les séquences *d'Astragalus monspessulanus* et de *Caragana arborescens*, deux **Astragalineae**, se regroupent conformément à leur taxon. Toutefois, la position relative de ces trois groupes (Astragalineae, Trifolieae, Vicieae) n'est pas définie. De nombreux travaux suggèrent que les Viciaea et les Trifolieae ont divergé récemment.

Ils sont même parfois considérés comme groupes frères (Young et al., 2003). D'après Chapill (cf. annexe 2), les Hedysareae (*Onobrychis*) auraient un ancêtre commun avec ces deux groupes. Enfin, l'ensemble de ces groupes auraient divergé d'avec les Astragalineae au début de la diversification des Galégoïdes ce qui semble apparaître dans l'analyse réalisée ici, malgré l'absence de soutien au nœud correspondant. On remarque également dans ce clade des Galégoïdes, que les deux séquences de *Medicago truncatula* ne sont pas groupées comme celles du pois mais séparées. Les séquences MtrA1b007 et MalA1b001 semblent avoir un ancêtre commun plus récent que MtrA1b006 et MtrA1b007 (support de 77 pour le nœud MtrA1b007-MalA1b001). Il pourrait s'agir de paralogues (MtrA1b006 et MtrA1b007) et d'orthologues (MtrA1b007 et MalA1b001).

- En ce qui concerne les Phaséoloïdes, on retrouve deux groupes de séquences correspondant aux clades des **Milletieae** (Lca et Ms*e,* bootstrap de 60) et des **Phaseolineae** (Pvu et *Vigna spp*.). Ce dernier groupe n'est toutefois pas soutenu (bootstrap<50). Hormis ces deux groupes, les relations phylogénétiques entre séquences au sein des Phaséoloïdes sont très mal définies par cette analyse. Selon Chapill, Glycinineae et Phaseolineae partagent un ancêtre commun plus récent qu'avec toute autre sous-tribu du clade. Ce lien n'est toutefois pas soutenu dans notre analyse. La position des séquences de *Canavalia brasiliensis,* espèce appartenant aux Diocleineae plus proches des Milletieae que du groupe Phaseolineae-Glycinineae d'après Chapill, est ici indéterminée. Il en est de même pour la séquence d'*Alysicarpus ovalifolius*, une Desmodiinae dont la place au sein des Phaséoloïdes est également incertaine dans les travaux de Chapill. La position des séquences de *Bituminaria bituminosa* n'est pas soutenue non plus dans notre analyse. Les Psoraleae auxquelles appartient cette espèce, seraient, d'après Chapill, le groupe ayant divergé le plus tôt dans l'évolution des Phaséoloïdes. Vu les valeurs des bootstraps aux nœuds des Phaséoloïdes de cette étude, une représentation en râteaux conviendrait mieux. Dans ce clade Phaséoloïdes, on constate que les différentes séquences d'une même espèce sont regroupées (soja, haricot, Bbi, Cbr et Mse). On ne peut donc pas déterminer les éventuels paralogues et orthologues. Enfin, il faut remarquer la position de la séquence du lupin (LanU74383), très proche de celles du soja. Cette espèce appartient au clade des Génistoïdes, très éloigné des Phaséoloïdes. L'absence d'amplification positive, quelles que soient les conditions de PCR utilisées dans ces travaux, et la position de la séquence au sein des séquences de Phaséoloïdes suggèrent que cet homologue caractérisé par Ilgoutz et al. (1997) pourrait résulter d'une contamination de leurs échantillons par du soja.

L'analyse des séquences nucléiques complétées en 5' confirme la séparation de celles-ci en deux groupes correspondant aux deux clades auxquels appartiennent les espèces dont les séquences sont analysées. Toutefois, le nombre limité de ces séquences plus complètes apporte peu d'information supplémentaire à l'étude de l'évolution de la famille peptidique des Albumines 1. Elles sont néanmoins importantes pour l'étude structurale (au niveau protéique et génétique).

Le rapprochement des séquences d'une même espèce observé ici l'est également lors de la même analyse excluant la troisième base du codon. Ceci est effectué afin d'éliminer l'effet de l'usage du code qui tend à rapprocher les séquences d'une même espèce. Toutefois, le nombre de sites informatifs s'en trouve diminué et les résultats sont à analyser avec précaution.

C. Discussion

Par notre approche de PCR avec amorces dégénérées, 14 nouvelles espèces, toutes Phaséoloïdes ou Galégoïdes, ont révélé un produit de la taille attendue et spécifiant une A1b. Pour certaines d'entre elles, deux (*Onobrychis viciifolia, Bituminaria bituminosa, Canavalia brasiliensis*) ou trois (*Mundulea sericea*) séquences homologues différentes ont pu être caractérisées. Les séquences obtenues sont relativement variables en nature et nombre d'acides aminés. Pour chacune des 12 séquences complétées en 5', la structure du potentiel produit du gène en préproprotéine apparaît conservée, même si la maturation exacte de ces protéines reste inconnue.
Les résultats obtenus nous permettent d'émettre quelques hypothèses quant à la présence et l'évolution des A1b au sein des Légumineuses.

1) Les A1b au sein des Légumineuses : une évolution récente?

L'analyse des relations phylogénétiques entre homologues de PA1b, bien que limitée par la faible longueur des séquences et leur faible nombre, semble montrer une évolution récente de la famille peptidique, essentiellement soutendue par les relations phylogénétiques entre espèces (congruence avec l'analyse de Chapill), et diversifiée au niveau terminal par des duplications récentes. Toutefois, il convient d'être très prudent du fait de la multigénicité de la famille des A1b. Celle-ci, démontrée chez le pois, n'apparaît pas dans notre analyse non exhaustive, sauf pour *Mundulea sericea* espèce pour laquelle 3 séquences différentes ont été caractérisées (présence d'au moins deux gènes). Pour pouvoir interpréter l'évolution de ces A1b multigéniques au sein des Papilionoideae, il conviendrait de déterminer les éventuels orthologues et paralogues. Le regroupement des séquences d'une même espèce impliquerait une évolution essentiellement post-spéciation donc récente, alors que la présence d'homologie plus importante entre séquences d'espèces différentes qu'entre séquences d'une même espèce indiquerait une duplication avant spéciation et donc une histoire évolutive plus ancienne. Lors de notre analyse, nous avons détecté une différence entre Phaséoloïdes et Galégoïdes concernant le regroupement ou non des séquences d'une même espèce. Ceci peut être dû à différents facteurs :

- Notre approche de recherche non exhaustive de séquences. Il est très probable que nous n'ayons pas détecté tous les gènes d'une même plante et il se pourrait que les gènes caractérisés soient artefactuellement proches pour une espèce donnée.
- L'évolution des séquences au sein des deux groupes n'a peut-être pas suivi le même chemin. L'existence d'une évolution concertée chez un des taxons (Phaséoloïdes) pourrait avoir masqué la présence d'orthologues et de paralogues, donnant, à tort, l'impression d'une diversification récente de la famille. Il est possible également que la duplication et la divergence qui s'en est suivie soient des phénomènes plus récents (post-spéciation) chez les Phaséoloïdes que chez les Galégoïdes (pré-spéciation).

2) Les A1b limités à deux clades de Papilionoideae ?

L'absence de détection d'homologues de PA1b au sein des Mimosoideae, des Caesalpinioideae, de 3 des 5 groupes constituant les Papilionoideae et de certaines espèces des deux autres groupes est intrigante.
Si l'on considère la phylogénie des Légumineuses selon Chapill (cf. annexe 2), la présence de séquences dans les espèces reportées ici impliquerait leur présence dans différentes autres tribus. Pour les Galégoïdes, la réponse positive des deux Astragalineae testées, groupe basal, impliquerait la présence de gènes homologues, sauf perte du caractère, chez

les Coluteinae (non testés), les Carmichaelieae (non testés) frères des Hedysareae (dont *Onobrychis viciifolia*), les espèces du genre *Galega* également non testées ainsi que celles du genre *Cicer*, frère des Vicieae, dont *C. arietinum* n'a toutefois pas répondu aux différentes conditions de PCR testées. Seules les espèces du genre *Glycyrrhiza*, non testées ici, sont plus basales et pourraient ne pas présenter de gènes homologues. Il serait donc particulièrement intéressant de tester une de ces espèces.
Pour les Phaséoloïdes, la présence de séquences homologues chez les Psoraleeae (*Bituminaria bituminosa*), groupe le plus basal, indique la probable présence de gènes spécifiant des A1b au sein des différents groupes de ce clade, dont les espèces du genre *Hypocalyptus* (non testées ici), les Robinieae et Indigofereae deux tribus sœurs pour lesquelles les différentes espèces testées se sont toutes révélées négatives, les Ophrestiinae, les Kennediinae et les Cajaninae (non testés), les Lespedezinae frères des Desmodiinae (dont *Alysicarpus ovalifolius*), les Clitoriinae frères des Phaseolinae (haricot, *Vigna spp*.) et formant un des rares clades définis avec les Glycininae (soja), les Erythrininae (non testées), et, enfin, les Wisteriae frères des Milletieae (*Mundulea sericea* et *Lonchocarpus capassa*) mais dont *W. sinensis* n'a pas donné de résultat positif.
De plus, les Phaséoloïdes sont un clade frère de celui des Aeschynoménoïdes et forment avec ceux-ci un groupe frère des Galégoïdes. La présence de gènes homologues à PA1b au sein des deux clades les encadrant indique que les Aeschynoménoïdes pourraient également contenir de tels homologues. Toutefois, les trois espèces testées (*Dalbergia purpurescens, Arachis hypogaea* et *Abrus precatorius*), bien réparties au sein de ce clade, se sont toutes avérées négatives pour l'amplification avec nos amorces.
D'après nos résultats, et si on considère la phylogénie des Papilionoideae déterminée par Chapill, certaines hypothèses sur l'évolution des A1b au sein des Papilionoideae peuvent être considérées :

- Il y aurait eu deux acquisitions des A1b, parallèles, l'une au sein des Galégoïdes après la divergence d'avec le groupe Phaséoloïdes-Aeschynoménoïdes, l'autre au sein des Phaséoloïdes après la divergence de ces deux clades.
- L'acquisition des A1b serait unique, chez un ancêtre commun à ces trois groupes, puis, lors de la divergence des différents clades, une perte du caractère aurait eu lieu dans la branche menant aux Aeschynoménoïdes.
- L'apparition des A1b serait antérieure à la séparation des trois groupes taxonomiques considérés, mais une évolution différentielle entre clades aurait abouti à une divergence trop importante des séquences d'Aeschynoménoïdes pour être reconnues par notre approche.

Les résultats des tests biologiques présentés dans la suite de notre travail tendent à montrer que les Aeshynoménoïdes possèderaient des homologues de PA1b dans leurs graines, ce qui rendrait nulles les deux premières hypothèses.
Toutefois, la position taxonomique réelle des Aeschynoménoïdes est encore mal définie. De nombreuses études de phylogénie moléculaire les placent à la base des Papilionoideae, avec les Amorpheae (Doyle and Luckow, 2003 ; Wink and Mohamed, 2003). Si on considère cette phylogénie, les A1b seraient alors apparus, au plus tard, chez l'ancêtre commun le plus récent des Galégoïdes et Phaséoloïdes. Dans ces analyses, le clade formé de ces deux groupes est frère des Mirbelieae (dont Bossiaeae). Cette large tribu semble donc particulièrement intéressante à explorer pour mieux comprendre l'histoire évolutive des A1b. Malheureusement, les deux espèces que nous avions sélectionnées dans ce groupe (*Goodia latifolia* et *Gompholobium scabrum*) n'ont jamais germé et aucun ADN n'a pu être extrait des graines.

Enfin, avant toute conclusion, il ne faut pas négliger le fait que peu d'Aeschynoménoïdes ont été testées. Il conviendrait d'élargir l'échantillonnage des espèces dans ce large groupe. De plus, la PCR, malgré la dégénérescence des amorces est sensible aux variations de séquences et pourrait ne pas avoir détecté des séquences divergentes mais néanmoins homologues chez ces espèces voire chez les espèces des autres clades ou sous-familles. C'est pour essayer de contourner la sensibilité de la PCR et enrichir l'extrait en séquences recherchées qu'un Northern a été tenté sur quelques espèces sélectionnées, dont le pois en témoin, cette technique étant moins sensible aux variations locales de séquence. Malheureusement, le résultat négatif pour toute autre espèce que le pois ne permet pas de conclure.
La recherche de gène homologues apparaît donc délicate pour certaines espèces phylogénétiquement éloignées des Galégoïdes et Phaséoloïdes. L'approche de biologie-biochimie pourrait permettre d'étendre nos connaissances des A1b (diversité d'activité) et de leur répartition phylogénétique, par détection de toxicité analogue.

III. Variabilité d'activité entomotoxique

Afin de détecter la présence éventuelle de peptides homologues à PA1b au sein des Légumineuses, l'approche par **test biologique spécifique[12] sur extraits** de graines, présentée dans la première partie de ce travail, a été appliquée.

A. Matériels et Méthodes

De même que pour la partie de biologie moléculaire, les techniques utilisées dans cette partie sont essentiellement les mêmes que dans le deuxième chapitre de ce manuscrit. La spectrométrie de masse (MALDI-TOF), bien qu'apparaissant très informative pour la détection de peptides dans les fractions de graines, est une méthode non disponible localement et un peu coûteuse qui n'a pu être appliquée à grande échelle, et n'a donc pas été utilisée ici. De même les tests biologiques sur farine entière sont apparus difficiles à interpréter lors de la validation de l'approche sur les plantes tests et n'ont donc pas été réalisés ici.

1) Extractions

Les extractions successives par 5 solvants des farines ont été réalisées de la même façon mais en adaptant les quantités de solvant aux deux à trois grammes de farine utilisés pour toutes les espèces. La miniaturisation de cette étape a été réalisée par B. Delobel, au vu des faibles quantités de graines obtenues pour certaines plantes et afin de standardiser l'extraction. Une mini-extraction a été réalisée sur le pois dont les fractions, testées sur charançons des deux souches, ont révélé les mêmes activités que celles résultant de l'extraction à partir de 200 g de farine.

2) Matériel animal et tests biologiques

En plus des deux souches Bénin et Chine utilisées pour chaque fraction à tester, deux souches isogéniques ont été élevées et utilisées pour certains extraits. Les tests biologiques

[12] souches sensible et résistante, monogénicité de la résistance

et leurs analyses statistiques ont été réalisés dans les mêmes conditions que décrites précédemment, en grande majorité par B Delobel et MG Duport. Les souches isogéniques, appelées WAA42 (sensible) et ISOR3 (résistant), avaient été obtenues par Grenier et al. après trois générations de rétrocroisement de la souche Chine avec la souche WAA42. Cette dernière avait été choisie, malgré sa faible fertilité, pour son homogénéité génétique, révélée par une homozygotie élevée (Grenier A.M., communication personelle). Ces souches, peu productrices et délicates à élever n'ont été utilisées que sur quelques fractions de graines pour tester une éventuelle multirésistance de la souche Chine.

3) Tests d'inhibition de liaison.

Cette approche *in vitro*[13] a été tentée pour quelques espèces sélectionnées, dans les conditions décrites précédemment, en se limitant à 4 concentrations par fraction choisie (10; 1; 0,1; 0,01 mg d'EFE/puits). Ceci permet d'avoir une idée de la présence de composés inhibant la liaison de ^{125}I-PA1B à son site de liaison, mais n'est pas assez précis pour calculer la concentration minimale nécessaire à 50% d'inhibition de la liaison (IC_{50}).

4) Electrophorèses

Peu de fractions ont été testées par cette approche. Il s'agit toujours de fractions MeOH60A80 (fraction soluble acétone 80% obtenue comme décrit dans la première partie). Le protocole utilisé diffère de la première partie, étant plus adapté à la séparation de peptides de faible poids moléculaire. Il s'agit du protocole d'électrophorèse en Tris-tricine décrit par Schagger et Jagow (1987) qui permet de séparer SDS et peptides améliorant ainsi la résolution de l'électrophorèse. La révélation des protéines s'effectue au nitrate d'argent (AgNO3) (Blum et al., 1987).

B. Résultats

1) Extractions

Les graines de 93 espèces des trois sous-familles ont été broyées puis fractionnées à partir d'environ 2 g de farine. Parmi ces espèces, une Polygalaceae (outgroup), 13 Caesalpinioideae dont 2 basales et 8 Mimosoideae ont été sélectionnées.
Les pourcentages des différentes fractions obtenues sont consignés dans l'annexe 4.
On constate une grande variabilité de la représentation de chaque fraction entre et au sein de chaque sous-famille. Toutefois, on peut remarquer que les Mimosoideae sont globalement pauvres en huiles et en substances non solubles (résidu) mais présentent au contraire une forte représentation des différentes fractions d'intérêt (jusqu'à 54 % de la graine constituée de l'ensemble des extraits MeOH, MeOH60, H_2O5, H_2O8). Parmi celles-ci, *Entada phaseoloides* présente les fractions MeOH60 et MeOH les plus importantes (16,7 et 20,3% respectivement). Les Caesalpinioideae apparaissent également pauvres en huiles (exceptée *Acrocarpus fraxifolius* à 30% de lipides), mais riches en résidus, avec des teneurs en fractions d'intérêt très variables allant de seulement 18% à 47% de la graine. De même les Galégoïdes apparaissent peu riches en lipides, avec l'exception d'*Hippocrepis emerus*. Les Génistoïdes sont moyennement riches en huiles, avec les extrèmes *Daviesa corymbosum* (1%) et *Diplotropis purpurea* (47%). On peut enfin remarquer que *Polygala*

[13] inhibition de la liaison de ^{125}I-PA1b à une fraction microsomale de charançons sensibles (souche Bénin).

myrtifolia, l'espèce hors Fabaceae choisie comme "outgroup" est extrêmement riche en huiles (66,7) et pauvre en fractions d'intérêt (10,6% au total).

2) Variabilité de l'activité biologique

Pour des raisons de temps, les tests biologiques n'ont été effectués que pour 60 des 89 nouvelles plantes dont les farines ont été fractionnées, dont 8 Caesalpinioideae, 4 Mimosoideae, 15 Phaséoloïdes, 15 Galégoïdes, 4 Aeschynoménoïdes, 10 Génistoïdes et 4 Sophoroïdes.
Les TL_{50} des cinq fractions des différentes espèces testées sont consignés dans l'annexe 5. Les fractions présentant un différentiel de toxicité significatif entre les souches de charançons sensibles et résistants sont soulignées.
On constate une grande variabilité de la toxicité, qui est déterminée par trois facteurs:

- L'origine biologique de l'extrait (espèce végétale), soutendue par une structure phylogénétique.
- Les caractéristiques biochimiques de l'extrait (5 modalités du degré de polarité des composants).
- La cible biologique de l'extrait (la souche de charançons : R ou S)

On constate globalement que la fraction MeOH est moins toxique que les autres ce qui est cohérent avec une présence prévalente de la toxicité peptidique *a priori* exclue de cette première fraction. Au contraire, les résidus sont souvent porteurs de toxicité, comme observé sur les quatres plantes "test", ce qui ne correspond pas à une toxicité peptidique soluble attendue. La présence d'un différentiel très net de toxicité dans cette fraction d'espèces de différents groupes dont *A. julibrissin* (Mimosoideae), *L. albus* (Génistoïde), et même *L. tetragonolobus* (Phaséoloïde) et *A. hypogaea* (Aeschynoménoïde), deux plantes sans toxicité par ailleurs, est donc intrigante. Un tel différentiel indiquerait la présence d'homologues de PA1b dans ces résidus. Ceci peut s'expliquer par la liaison de PA1b à des polymères insolubles provenant des parois cellulaires. C'est en effet liée à celles-ci, que la leginsuline a été observée, dans la graine de soja (Nishizawa et al., 1995).
On remarque également que la toxicité est plus forte sur la souche S, ce qui augure d'une bonne présence globale de toxicité différentielle potentiellement liée aux A1b. Une seule espèce présente une toxicité significativement plus grande sur R (TL_{50} 4,3 jours) que sur S (TL_{50} 5,6 j.). Il s'agit de *Crotalaria eremaea*, une Génistoïde (fraction MeOH60).
Toutefois, la toxicité dépend essentiellement de l'espèce végétale. Certaines plantes n'ont en effet aucune toxicité (ex: *Cassia occidentalis, Vigna subterranea*) alors que d'autres présentent une toxicité pour toutes leurs fractions, sur les deux souches de charançons (ex: *Bolusanthus speciosus, Crotalaria eremaea, Robinia pseudoacacia*). La toxicité varie même au sein d'un même genre comme le montrent les résultats obtenus sur *V. unguiculata* (toxicité, différentielle) *et V. subterranea* (pas de toxicité) ou *A. lebbeck* (toxicité totale) et *A. julibrissin* (toxicité différentielle). On constate cependant que la répartition taxonomique de la toxicité n'est pas identique pour tous les groupes. Les Caesalpinioideae, par exemple, apparaissent globalement très peu toxiques (seule *Cercis siliquastrum* présente une toxicité aiguë pour la fraction MeOH60, d'ailleurs différentielle).

(a) Analyse de données

Une analyse de ces données par différents modèles statistiques a été effectuée par Diol (2003) pour comprendre les relations entre la réponse (toxicité) et les trois facteurs (espèce, fraction, souche). Ces travaux montrent que la réponse est à la fois quantitative

(intensité de la toxicité) et qualitative (présence ou absence de toxicité) et confirment les constatations ci-dessus. Les résultats de certaines analyses sont mentionnés ci-dessous.
La réponse quantitative a été analysée par ACP (analyse en composantes principales), révélant comme facteur principal un effet taille (intensité globale de la toxicité) (Diol, 2003). Le second facteur est lié à la nature chimique des fractions, H_2O5 et H_2O8 (protéiques) d'un coté et MeOH-résidu (apolaires, polymères insolubles) de l'autre. On constate par cette analyse que l'effet de la fraction se reproduit à l'identique sur les deux souches (pas d'interactions entre ces deux facteurs). La répartition des espèces dans le plan factoriel est très éclatée (Figure 69). Seules les Galégoïdes de toxicité globale moyenne semblent structurées par une toxicité protéique (fractions H_2O). Les Génistoïdes présentent des espèces très toxiques et leur toxicité domine dans les fractions non protéiques.

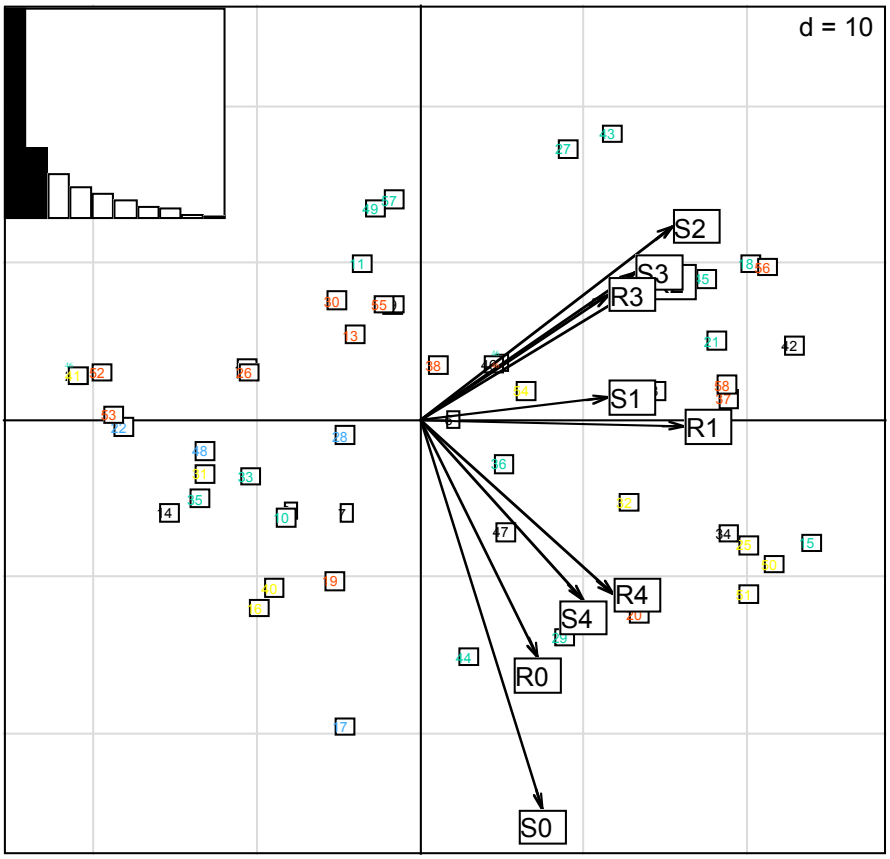


Figure 69 : Analyse en Composantes Principales (Diol, 2003)
rouge : Galégoïdes, vert : Phaséoloïdes, jaune : Génistoïdes, bleu : Aeschynoménoïdes.
variables S : souche sensible, R : souche résistante, 0 à 4 : fractions MeOH à résidu.

Une Analyse en Composantes Multiples (ACM) de la réponse qualitative permet de classer les espèces végétales en trois groupes selon leur toxicité moyenne (Diol, 2003):

- Classe C1: Globalement peu toxique (ni sur R ni sur S)
- Classe C2: Toxicité globale moyenne et différentielle (toxicité pour S plus que pour R)
- Classe C3: Globalement très toxique (sur S comme sur R)

Afin de déterminer l'importance de la composante taxonomique sur la toxicité (aspect qualitatif), une étude utilisant le test d'Abouheif (1999) a été réalisée par Diol avec l'aide

du Pr Chessel (Diol, 2003). Ces tests indiquent une structuration phylogénique globale significative comme le montre la Figure 70.

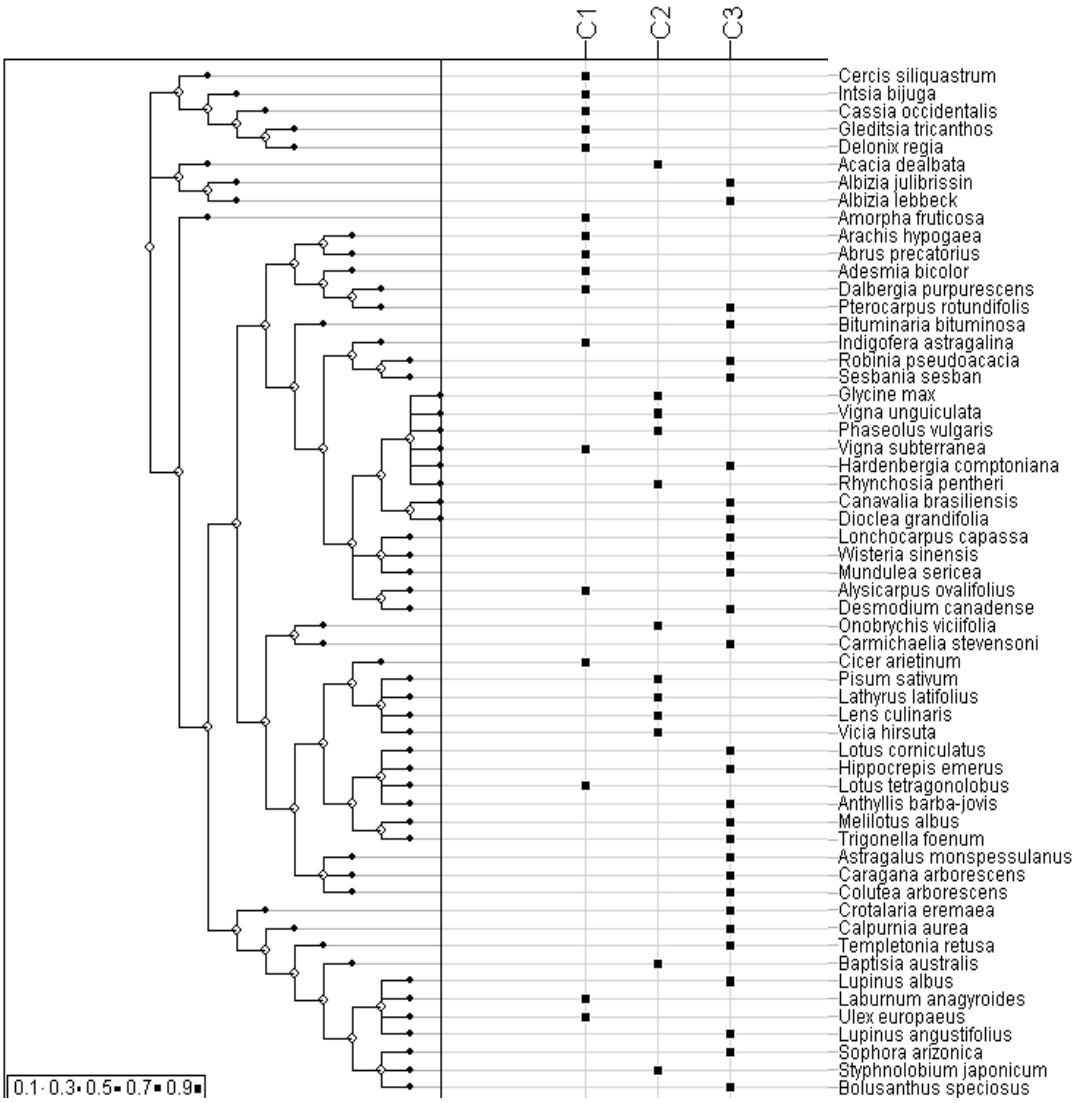


Figure 70 : Représentation de la distribution des 3 classes de toxicité (C1 à 3) en fonction de la position taxonomique des différentes espèces.

On remarque que les plantes d'un même groupe phylogénétique (sous-famille, clade et/ou tribu) ont parfois des toxicités semblables. Ainsi, on observe le regroupement des **Caesalpinoideae** et des **Aeschynoménoïdes** dans la classe des plantes **peu toxiques**. Le groupe phylogénétique des **Milletieae-Wisterieae et Diocleineae** correspond à des plantes de très **forte toxicité**. De même les **Trifolieae et Loteae** (exceptée *L. tetragonolobus*), groupes frères selon Chapill, apparaissent toutes comme des plantes de classe 3. Les Galégoïdes **Vicieae** d'une part et les 2 sous-tribus de Phaséoloïdes **Phaseolinae** (sauf *V. subterranea*) et **Glycininae** d'autre part, sont majoritairement des plantes de **toxicité moyenne et différentielle** ce qui est en accord avec l'observation d'A1b (gènes et/ou peptides) dans les plantes de ces trois clades. La plupart des Galégoïdes hors Vicieae appartiennent à la classe 3 avec les exceptions notables du pois chiche (*Cicer arietinum*) et du lotier corniculé très peu toxiques. On distingue donc dans les clades Galégoïdes et Phaséoloïdes, deux groupes de plantes : celles à toxicité différentielle (C2) où la présence

d'A1b est démontrée et celles de classe C3 qui possèdent probablement également des A1b (gènes caractérisés) plus d'autres composés toxiques pour les insectes résistants et certainement non apparentés aux A1b.
La majorité des Génistoïdes (6/9) testées apparaissent très toxiques. Toutefois on observe également 2 espèces de classe 1 et une espèce, *Baptisia australis*, de toxicité moyenne, dont les fractions méthanoliques et H_2O5 présentent un différentiel net de toxicité indiquant la probable présence d'homologues de PA1b. Quelques plantes isolées dans des groupes autres que Vicieae et Phaseolineae montrent également une toxicité intermédiaire intéressante. Il s'agit de la Sophoreae basale *Styphnolobium japonicum*, d'*Onobrychis viciifolia* qui appartient aux Galégoïdes hors Vicieae, de *Rhynchosia pentheri*, une Phaséoloïde et *d'Acacia dealbata*, une Mimosoideae. Ces 5 plantes ont toutes au moins deux fractions de toxicité différentielle totale (mort de S et pas R), MeOH60 et H_2O5, fractions où l'on retrouve la majorité des isoformes de PA1b à partir du pois. Ce sont donc de bons candidats pour la recherche d'A1b. Chez *Onobrychis*, deux gènes homologues ont d'ailleurs été caractérisés (Figure 67). *A. dealbata, S. japonicum* et *B. australis,* appartenant chacune à des groupes différents ne répondant pas à l'approche moléculaire, sont à considérer tout particulièrement.
Enfin, certaines plantes de classe 1 présentent aussi un différentiel de toxicité (sur une de leurs fractions) pouvant indiquer la présence d'A1b dans ces espèces. Il s'agit d'*A. fruticosa* (Sophoroïde), de *D. purpurescens* (Aeschynoménoïde) et de *C. siliquastrum* (Caesalpinoideae).

(b) Tests de toxicité sur les charançons isogéniques

Quatres fractions au différentiel de toxicité net (total ou non) ont été testées sur les souches isogéniques WAA42 et ISO3R, afin de tester l'hypothèse d'une résistance multiple de la souche Chine, et celle de la présence d'A1b dans ces extraits d'espèces phylogénétiquement éloignées des Galégoïdes et Phaséoloïdes. Les résultats sont consignés dans le Tableau 8.

Tableau 8 : Comparaison des temps létaux 50 des différentes souches de *S. oryzae* isogéniques ou non; pBGW: test de rang de Breslow-Gehan-Wilcoxon (différentiel de mortalité entre les isogéniques (W/I) ou les non isogéniques (B/C)).

TL50 (SE)	Souches sensibles		Souches résistantes		p BGW	
Espèce et fractions	WAA42	Bénin	ISO3R	Chine	W/I	B/C
A. fruticosa MeOH60	4,8 (0,3)	10,5 (1,9)	>20	>20	<0,0001	<0,0001
B. australis MeOH60	3,9 (0,2)	9,0 (1,9)	5,0 (1,0)	>20	<0,0001	<0,0001
C. siliquastrum MeOH60	4,4 (0,2)	5,7 (0,6)	9,5 (0,9)	>15	<0,0001	<0,0001
S. arizonica H_2O5	5,6 (0,3)	5,3 (0,3)	8,3 (1,2)	8,8 (0,7)	0,0014	<0,0001

On constate que les tests biologiques sur insectes isogéniques ne sont pas équivalents à ceux réalisés sur les souches Bénin et Chine, non isogéniques.
Concernant les souches sensibles, les différences sont particulièrement importantes pour les deux premières espèces (*A. fruticosa, B. australis*), WAA42 est une souche de temps de développement plus long, moins féconde (deux ovarioles par femelles contre quatre pour Bénin) et donc difficile à maintenir en élevage. Toutefois son homozygotie, réduisant la variabilité génétique, la rendait particulièrement intéressante pour l'élaboration d'une souche isogénique résistante. Les TL50 de la souche sensible WAA42 sont beaucoup plus

faibles que ceux de la souche Bénin. Il se peut que l'homozygotie élevée de la souche WAA42 la rende plus fragile que Bénin à divers composés, dont les A1b.
Pour les deux autres fractions testées (*S. arizonica, C. siliquastrum*), les différences de toxicité entre souches sensibles sont plus faibles. Il en est de même pour les souches résistantes. Concernant ces dernières, on ne détecte pas de différence pour *A. fruticosa* (pas du tout de toxicité). Ceci indique que cette fraction contient vraisemblablement des A1b typiques, tuant les sensibles Bénin ou WAA42 et n'affectant pas les résistants ISO3R et Chine[14]. Par contre, l'apparition d'une toxicité aiguë pour *B. australis* (TL_{50} de 5 jours / pas de toxicité) est très significative. Ceci pourrait s'expliquer par la présence, dans cet extrait, d'un composé non apparenté aux A1b, auquel la souche Chine résisterait. Lors de la sélection des insectes isogéniques sur pois, le caractère de résistance à ce composé de *B. australis* aurait été perdu, le composé n'étant pas présent dans les grains de pois. La souche Chine serait donc multirésistante.

3) Compléments biochimiques

Pour certaines fractions de quelques espèces, les approches d'inhibition de liaison et de migration électrophorétique ont été appliquées. Les fractions sont choisies pour différentes raisons :

- Pour confirmation, dans le cas d'espèces à toxicité différentielle et dont des séquences homologues ont été caractérisées (*V. unguiculata, O. viciifolia*).
- Pour détecter la présence éventuelle de produits des gènes homologues caractérisés dans le cas d'espèces à toxicité non différentielle (*B. bituminosa, L. capassa, T. foenum-graecum, M. albus*).
- Pour des espèces où l'approche de biologie moléculaire a échoué, qu'elles soient toxiques pour R et S (*T. retusa, B. speciosus, C. eremaea, S. arizonica*) ou à différentiel (*C. siliquastrum*).

Les fractions considérées sont toujours MeOH60.

(a) Tests d'inhibition de liaison.

Les tests de compétition effectués sur des fractions MeOH60 de toxicités variables se sont avérés infructueux, contrairement aux premiers essais effectués sur le pois, soja et haricot. Comme le montre la Figure 71, on ne peut conclure à une inhibition de liaison significativement plus élevée que le témoin négatif blé pour aucune des plantes testées si ce n'est pour *Onobrychis viciifolia*. En effet, si le blé inhibe 50% de la liaison pour un peu plus de 10 mg d'EFE, seules les fractions inhibant 50% de la liaison pour moins de 0,1 mg d'EFE peuvent être considérées comme positives au test. On observe que pour cette plus faible dose, seule *O. viciifolia* n'a pas dépassé les 50% de ligand lié.
Si l'absence d'inhibition de liaison était attendue pour des espèces comme *T. retusa* et *S. arizonica*, ne présentant pas de différentiel de mortalité, ou *V. subterranea* ne présentant aucune toxicité, les résultats négatifs obtenus pour les espèces *B. bituminosa, C. siliquastrum, ou V. unguiculata* dont les fractions testées ici présentent un différentiel net de toxicité et sont donc supposées contenir des A1b, sont plus surprenants. Ceci montre que le test in vitro n'équivaut pas au test in vivo.

[14] Une alternative à la présence d'A1b serait celle d'un autre composé dont le gène de résistance serait lié à celui de PA1b.

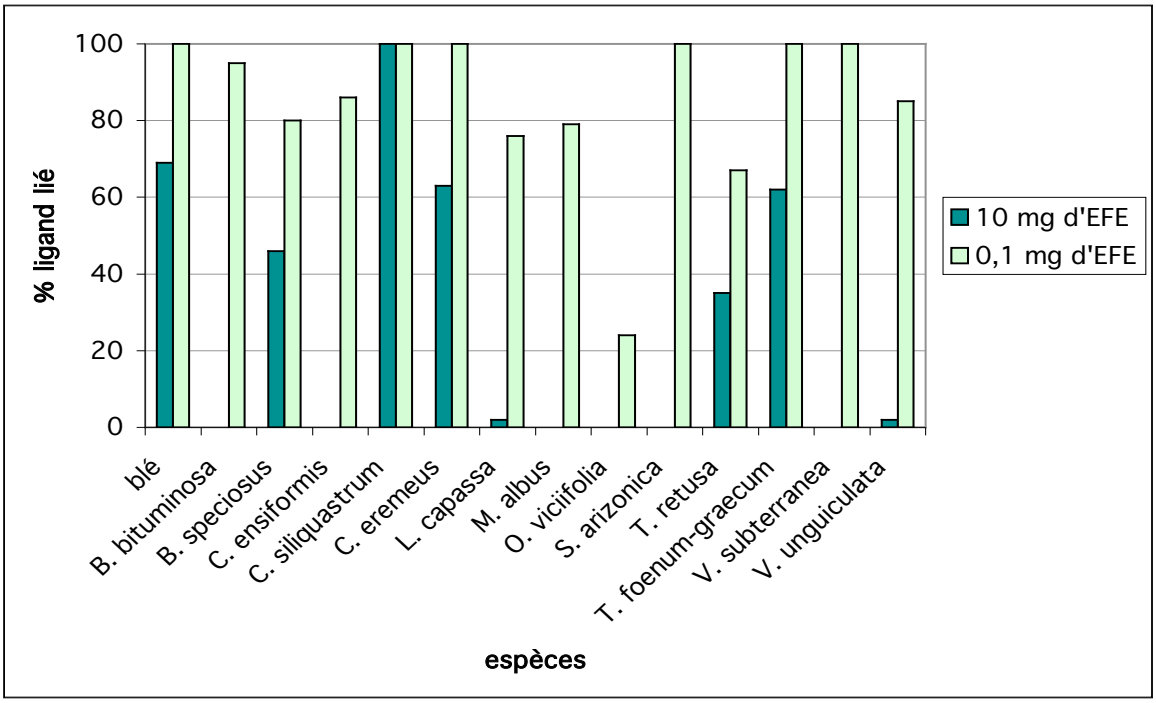


Figure 71 : Test d'inhibition de liaison : pourcentage de ligand lié pour deux concentrations en fraction MeOH60 de différentes espèces végétales. Les concentrations sont exprimées en mg d'Equivalent Farine Entière/ puits. Le blé sert de témoin négatif.

(b) Electrophorèses.

Certaines des fractions testées par inhibition de liaison ont été extraites à l'acétone 80%. Les fractions solubles ont ensuite été déposées sur gel d'électrophorèse afin d'y détecter d'éventuels peptides de même masse moléculaire que PA1b. Les fractions au différentiel de toxicité net et quelques fractions toxiques pour R et S ont été sélectionnées. La Figure 72 montre des peptides de masse apparente équivalente à PA1b pour différentes espèces de Galégoïdes (*O. viciifolia, M. albus*) et de Phaséoloïdes (*V. subterranea, V. unguiculata, A. ovalifolius, B bituminosa*). Toutefois, on n'observe rien de similaire pour *L. capassa*, dont la fraction déposée est toxique pour R et S et qui est une Phaséoloïde possédant au moins un gène homologue (Figure 67). Pour *B. speciosus*, *C. eremaea* et *T. retusa*, Génistoïdes ne présentant pas de différentiel de toxicité net, l'absence de peptides de la taille de PA1b était attendue. Enfin, pour le fenugrec, on observe un peptide de taille supérieure qui pourrait correspondre à une albumine 1 complète (A1a + A1b). Ceci confirmerait l'hypothèse formulée à propos de l'observation d'une proline en aval du site de maturation C-terminal putatif des A1b. Cette proline inhiberait la maturation du peptide et celui-ci pourrait rester sous forme complète. Toutefois, le mélilot présente la même caractéristique pour ses séquences (Figure 67), or on peut observer ici la présence de peptide de masse moléculaire type A1b. Il se peut que cette espèce possède d'autres gènes non identifiés par notre approche ne présentant pas de proline en N-terminal du peptide intermédiaire.

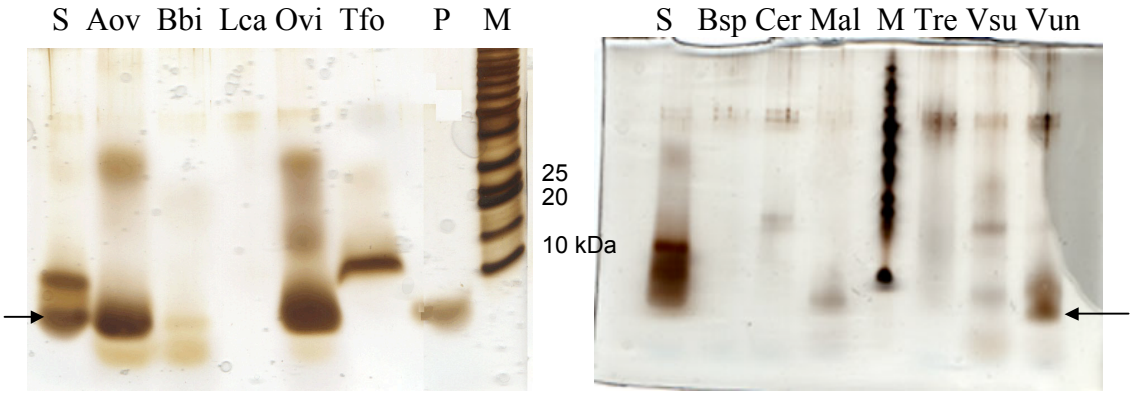


Figure 72 : Electrophorèse en tris-tricine (révélation au nitrate d'argent) sur fractions MeOH60A80 de quelques espèces. S: SRA1M60, P: PA1b, M: marqueur de poids moléculaire, Aov : *Alysicarpus ovalifolius*, Bbi : *Bituminaria bituminosa*, Bsp : *Bolusanthus speciosus*, Cer : *Crotalaria eremaea*, Lca : *Lonchocarpus capassa*, Mal : *Melilotus albus,* Ovi : *Onobrychis viciifolia*, Tfo : *Trigonella foenum-graecum*, Tre : *Templetonia retusa*, Vsu : *Vigna subterranea*, Vun : *V. unguiculata*. Les flèches indiquent la position de PA1b.

L'absence d'anticorps capable de reconnaître des isoformes de PA1b autres que celles du pois ne permet pas de conclure quant à la nature exacte des peptides observés ici.

C. Discussion

1) Activité biologique des extraits de graines

Les résultats obtenus par les tests biologiques sur les souches Chine et Bénin montrent une forte variation d'activité biologique entre plantes, fractions et souche. On détecte tous les cas de figure excepté une toxicité pour Chine sans toxicité sur Bénin.
On peut déterminer trois classes d'espèces végétales, selon leur toxicité moyenne : faible (Classe 1), intermédiaire et différentielle (C2), forte (C3).
Peu d'espèces présentent une toxicité globale différentielle. Toutefois, si la présence d'un différentiel de toxicité total pour certaines fractions dans les plantes de classe 2 est probablement un bon indicateur de la présence d'A1b dans ces plantes, il ne faut pas négliger les plantes très toxiques où la toxicité d'autres composés peut masquer la présence des peptides recherchés. On peut s'intéresser en particulier aux plantes où, si la toxicité globale de la fraction considérée est élevée, elle est essentiellement due à la mortalité des insectes sensibles (différentiel non total mais significatif), comme *Mundulea sericea, Bituminaria bituminosa, Astragalus monspessulanus* et *Melilotus albus*, plantes possédant par ailleurs des gènes homologues (Figure 67), voire des peptides de masse apparente similaire à PA1b (Figure 72). Les plantes pour lesquelles le différentiel de mortalité est moins net sont également à prendre en compte puisque des gènes homologues ont été caractérisés chez certaines d'entre elles comme *Medicago truncatula, Lonchocarpus capassa, Caragana arborescens, Canavalia brasiliensis* et *Lotus corniculatus*. La présence de peptides correspondant aux gènes caractérisés reste à confirmer (éventuellement par spectrométrie de masse), même si pour *Bituminaria bituminaria*, des peptides de taille similaire à PA1b ont été observés en électrophorèse à partir de la fraction MeOH60 et qu'un pic en HPLC de même temps de rétention que PA1b a été caractérisé pour son activité différentielle sur R et S (Diol, 2003).
Les plantes sans activité insecticide sur *S. oryzea* peuvent également nous renseigner sur la famille des A1b. *Vigna subterranea* possède au moins un gène homologue. Ces variants

ne sont peut-être pas exprimés (pseudogène putatif[15]), trop peu exprimés ou bien sont non toxiques. Ce dernier cas pourrait nous renseigner sur les acides aminés impliqués dans l'activité insecticide de ces peptides et indiquerait un autre rôle, expliquant leur conservation, dans la physiologie de la plante, comme le proposent Higgins et al. (1986) pour la leginsuline du soja. Si le produit du gène de *V. subterranea* est exprimé, seul le remplacement de R21 par une cystéine pourrait expliquer la non toxicité du peptide. Toutefois, on ne peut savoir dans ce cas si c'est l'absence du résidu basique, ou un mauvais arrangement des ponts disulfures lié à l'apparition d'une cystéine surnuméraire, qui serait impliqué.

2) Analyses complémentaires

Pour tenter de confirmer la présence d'A1b dans les fractions à différentiel de toxicité ou essayer de détecter d'éventuels homologues dans des extraits fortement toxiques, quelques analyses complémentaires ont été réalisées.
De toutes les fractions testées par inhibition de liaison, aucune n'a permis de conclure à la présence d'homologues de PA1b. La présence d'inhibiteur de liaison dans ces fractions grossières est possible. Ces tests étant conçus pour des composés purifiés ces résultats décevants étaient néanmoins attendus. Toutefois, des analyses réalisées par Diol (2003) montrent que même un fractionnement par HPLC est insuffisant. Les pics de chromatographie, obtenus à partir des extraits MeOH60 des différentes plantes sélectionnées pour leur différentiel de toxicité, ne permettent pas d'observer de déplacement significatif de ^{125}I-PA1b.
Une autre hypothèse pouvant expliquer l'absence de liaison pour des fractions présentant un différentiel de mortalité net serait la présence de facteurs toxiques non apparentés à PA1b mais agissant sur un domaine différent du même site de liaison que notre peptide. De plus, travaillant sur des souches de charançons non isogéniques, on peut émettre l'hypothèse d'un caractère lié au gène portant la résistance ou sensibilité à PA1b et qui jouerait un rôle similaire face à d'autres composés toxiques (résistance multiple de la souche Chine), comme tendent à le montrer les résultats des tests biologiques sur charançons isogéniques.

Les électrophorèses sur fraction à différentiel net tendent à confirmer la présence supposée d'A1b, mais ne sont généralement d'aucune aide pour les extraits sans différentiel et particulièrement pour les espèces hors Galégoïdes et Phaséoloïdes. De plus, pour réellement confirmer que les peptides observés sont des A1b, un anticorps spécifique des A1b et reconnaissant des homologues de différentes espèces serait un outil beaucoup plus efficace. Dans ce contexte, la boucle peptidique correspondant aux acides aminés CSPFEMPPC, conservés dans tous les gènes caractérisés dans cette étude, a été synthétisée. Couplée à une protéine porteuse, elle sera utilisée pour produire des anticorps polyclonaux spécifiques qui devraient éventuellement permettre de détecter des A1b hors du pois (L. Quillien, communication personnelle).

On peut conclure que ces analyses complémentaires ne sont guères informatives, malgré les espoirs fondés dans la première partie de cette thèse, en particulier sur les tests d'inhibition de liaison, assez rapides et peu coûteux en matériel. La spectrométrie de

[15] Un biais dans la composition en AT est attendu pour un pseudogène. Ce n'est pas le cas ici, le pourcentage d'AT de la séquence étant de 54 %, équivalent à ceux des autres séquences caractérisées.

masse, qui s'est avérée un outil très efficace lors de l'étude des quatre plantes "test", doit être tentée, pour rechercher des peptides dans les fractions à toxicité différentielle.

Pour des espèces végétales phylogénétiquement éloignées des deux groupes positifs à l'approche de biologie moléculaire, et toxiques pour R et S (avec ou sans différentiel significatif), l'interprétation des tests biologiques reste difficile.
La purification d'isoformes potentielles chez quelques espèces de toxicité différentielle et de différents groupes, comme *Cercis siliquastrum* et *Dalbergia purpurescens* permettrait de confirmer le lien entre toxicité différentielle et homologues de PA1b, mais aussi de mieux cerner la variabilité de séquence des A1b.

IV. Discussion et conclusions

Les objectifs principaux de l'élargissement de l'étude des A1b au sein des Fabaceae étaient, d'une part, de mieux comprendre cette famille (variabilité d'activité, de structure, répartition taxonomique) et d'autre part, d'en déduire des indications sur son mode d'action.

1) Variabilité et histoire évolutive des A1b

Nous avons montré la présence de gènes homologues à PA1b dans de nombreuses espèces de deux clades de Papilionoideae. De plus, nous avons détecté une toxicité différentielle, entre souches de *S. oryzae* sensible et résistante à PA1b*,* dans diverses espèces des trois sous-familles de Fabaceae. N'ayant pas pu caractériser de gènes homologues (ni par PCR ni par Northern) pour les espèces d'autres clades que Galégoïdes et Phaséoloïdes, la caractérisation d'une toxicité différentielle spécifique est la seule indication de la présence d'homologues de PA1b chez ces espèces. La purification puis le séquençage d'un peptide homologue à partir d'une fraction choisie pour son différentiel de toxicité devrait nous permettre de déterminer de nouvelles amorces dégénérées, plus spécifiques des séquences de la tribu ou sous-famille de l'espèce choisie. Ceci avait été tenté sans succès sur *Cercis siliquastrum*, plante la plus éloignée phylogénétiquement pour laquelle une toxicité analogue avait été observée. Une plante plus proche d'un point de vue taxonomique, au différentiel R/S total, a donc été choisie : il s'agit de *Dalbergia purpurescens* dont la fraction MeOH60 soluble acétone 80% présente un profil d'électrophorèse proche de la fraction équivalente du pois. Le fractionnement de cet extrait par HPLC est en cours.
L'absence de séquence nucléique ou protéique dans les espèces hors Galégoïdes et Phaséoloïdes, malgré la présence de profils de toxicité semblable au pois (MeOH60, différentiel), ne permet pas de conclure définitivement quant à la présence de la famille des A1b au sein des trois sous-familles de Légumineuses. L'approche par spectrométrie de masse s'étant révélée d'une grande efficacité dans la première partie de ce travail, celle-ci pourrait être appliquée sur quelques fractions d'espèces choisies pour leur éloignement phylogénique et leur toxicité différentielle, comme les fractions MeOH de *Parkia biglobosa* (Mimosoideae) ou MeOH60 d'*Amorpha fruticosa ou Baptisia australis*. Ceci permettrait de détecter la présence éventuelle de peptides de masse proche de PA1b, et donc d'envisager quelques microséquençages qui pourraient confirmer l'appartenance de ces peptides à la famille des A1b et permettre de déterminer de nouvelles amorces dégénérées.

2) Diversité d'activité des extraits, A1b et autres composés de défense.

La diversité d'activité observée lors des tests biologiques ne peut pas être attribuée aux seuls A1b. Si la présence de toxicité différentielle totale (pas de toxicité sur Chine) est très probablement liée à la présence d'A1b dans l'extrait considéré, une mortalité sur les charançons résistants peut-être due à de nombreux autres composés. Les Légumineuses ont fait l'objet de nombreuses études de phytochimie. Plusieurs types d'alcaloïdes, d'AANP, de composés phénoliques, de terpènes, d'IP et de lectines, ont été décrits dans cette famille, dans tous les organes des plantes et particulièrement dans les organes reproductifs (cf. premier chapitre) (Bisby et al., 1994 ; Harborne et al., 1971).
Toutefois, en analysant rapidement la répartition de différents composés ou groupe de composés de défense (Bisby et al., 1994 ; Wink and Mohamed, 2003), on ne détecte pas de concordance évidente entre toxicité et présence/absence du composé. En effet, si l'on considère l'AANP canavanine, très étudié, soluble dans le méthanol 60%, on s'aperçoit qu'il est présent dans les graines de plantes dont la fraction MeOH60 est très toxique comme les Phaséoloïdes *Canavalia brasiliensis, Sesbania sesban* et *Robinia pseudoacacia* ou les Galégoïdes *Caragana arborescens, Colutea arborescens* et *Lotus corniculatus*, mais aussi dans les graines de *Laburnum anagyroides* qui ne montre aucune toxicité sur charançons résistants (hors de la fraction résidu). La présence de cet AANP dans les graines ne semble donc pas liée à la forte toxicité des espèces citées. D'autres AANP comme l'albizzine (soluble en méthanol) ou l'acide djenkolique (soluble en milieu aqueux acide ou basique) sont présents dans les graines de la plupart des Mimosoideae, dont *Acacia dealbata* et *Parkia biglobosa* sont deux espèces non toxiques sur la souche R (hors résidu), contrairement aux deux *Albizia* testés.
Les Génistoïdes présentent peu d'AANP mais des alcaloïdes quinolizidines (QA : solubles dans l'eau fortement acide) comme la lupanine des graines de lupin. Ces alcaloïdes interfèrent avec le système nerveux (récepteurs à l'acétylcholine et canaux Na^+ K^+) (Wink, 1993). La lupanine représente 70% des alcaloïdes des graines des deux espèces testées ici (*L. angustifolius* et *L. albus*) (Wink et al., 1995). Toutefois, le lupin blanc n'est quasiment pas toxique contrairement au lupin à feuilles étroites. Parmi les QA, les alcaloïdes α-pyridone comme la cytisine sont connus pour être 10 à 100 fois plus toxiques que la lupanine (Wink et al., 1995). Rare chez les lupins et commune chez les espèces des genres *Baptisia* (dont *B. australis*, non toxique sur Chine) et *Thermopsis*, on la trouve également, sous forme de méthylcytisine, chez *L. anagyroides*. Chez les Crotalarieae, dont *C. eremaea,* très toxique, les QA sont remplacés par des alcaloïdes pyrrolizidines (PA). Cependant, on retrouve de tels alcaloïdes chez *L. anagyroides*, très peu toxique pour les insectes résistants.
Les isoflavonoïdes, restreints aux Papilionoideae s'accumulent dans un grand nombre de taxons. Les anthraquinones, inhibiteurs puissants des Na^+ K^+ ATPase, sont représentées dans tous le genre *Cassia* et dans quelques espèces d'*Abrus* (Wink and Mohamed, 2003). Toutefois, les graines de *Cassia occidentalis* ne montrent aucune toxicité dans les conditions de nos tests biologiques.
Les triterpènes et saponines sont plus répandus dans les Mimosoideae-Caesalpinioideae et les tribus basales des Papilionoideae, mais apparaissent aussi chez les Vicieae, Trifolieae, Cicereae et Phaseoleae ce qui ne permet pas de conclure quant à leur potentielle toxicité pour les charançons. Il en est de même pour les IP (inhibiteurs de trypsine ou chymotrypsine).

Une étude plus poussée de la concentration des divers composés individualisés, dans la graine et au niveau de l'espèce végétale, s'avère nécessaire pour déterminer les composés responsables des toxicités observés sur les charançons résistants. Toutefois, ce n'est pas le propos de notre étude. Quelques tests biologiques avec certains de ces composés, comme la canavanine, un PA ou une anthraquinone, permettraient de répondre en partie à cette question.

3) Indications sur le mode d'action?

Le deuxième objectif de cette large étude était d'essayer de comprendre le mode d'action des A1b et plus particulièrement de déterminer des zones ou résidus impliqués dans l'activité insecticide de ces peptides. Malgré l'absence de lien direct entre les séquences caractérisées et les toxicités observées, la confrontation des données de biologie moléculaire, des tests biologiques et de la structure tridimensionnelle connue de PA1b (Jouvensal et al., 2003) nous permet d'émettre quelques hypothèses.

- La toxicité différentielle observée pour MeOH60 chez Alysicarpus, malgré la mutation surprenante proline-lysine, suggère que cette proline ne serait pas nécessaire à la structuration du peptide actif, et ce bien que les prolines soient d'une grande importance structurale et en général extrêmement conservées.
- La conservation pour toutes les séquences sans exception, de la leucine L27, combinée à sa position externe sur la structure tridimensionnelle, font de ce résidu une cible particulièrement intéressante, *a priori* de grande importance pour l'activité (entomotoxicité ou hormone[16]) et/ou la structure de ces peptides.
- La conservation de l'acide aminé 11 acide (E ou D) ou polaire (Q) et le fait qu'il soit placé à un sommet de la molécule, nous suggère une possible importance de ce résidu dans l'activité du peptide.
- Enfin, l'arginine R21, conservée sauf chez deux espèces de *Vigna* apparaît également intéressante sa basicité pourraient être impliquée dans la réalisation de liaison ionique nécessaires à l'activité des A1b.

Ces hypothèses ont été sélectionnées pour être testées par la réalisation de mutants ciblés de PA1b. La mise au point d'un système d'expression hétérologue du peptide du pois apparaît en effet un moyen rapide et peu coûteux pouvant servir de base à une étude de relations structure-fonction. Une telle approche, combinée à la mutagenèse dirigée et à un criblage de mutants permettra de déterminer les résidus essentiels à l'activité insecticide des A1b.

[16]Rôle dans la physiologie végétales proposé par Higgins et al. (1986), et reprise par Yamazaki et al. (2003) : le doigt hydrophobe serait impliqué dans la liaison au récepteur (Hanada et al., 2003).

Chapitre 4 : Expression hétérologue de PA1b.

Expression hétérologue de PA1b

La troisième partie de ce travail de thèse avait pour objectifs la mise au point d'un système d'expression et de purification de PA1b recombinant et de mutants structuraux, dans le cadre d'une étude des relations structure - activité de ce peptide, afin de relier les variabilités de séquence et de toxicité observée dans les parties précédentes.
La structure de PA1b, petit peptide compact et amphiphile, nécessite des modifications post-traductionnelles spécifiques (structuré par 3 ponts disulfures). Le choix du système d'expression est crucial pour obtenir un produit bien structuré et donc actif.

I. Systèmes d'expression hétérologue

Il existe de nombreux systèmes eucaryotes ou procaryotes dont nous allons présenter, dans cette partie introductive, les différents avantages et inconvénients, en particulier en vue de l'expression de PA1b.

A. Systèmes eucaryotes

1) Les Levures

Les levures sont les premiers organismes eucaryotes à avoir été utilisés pour la production de protéines recombinantes d'origine eucaryote en recherche et en industrie. Ce sont des microorganismes unicellulaires dont la génétique est relativement simple. Les levures poussent rapidement à haute densité sur des milieux simples (Emr, 1990).
De la même manière que les autres organismes eucaryotes, la levure a l'avantage d'effectuer des **modifications post-traductionnelles** nécessaires au repliement correct de certaines protéines d'intérêt.
Deux types de vecteurs ont été développés pour les levures. Il s'agit des vecteurs intégratifs qui s'incorporent de façon stable dans le génome, et des vecteurs épisomiques qui sont des plasmides autonomes, extra-chromosomiques (Emr, 1990).
La levure possède la capacité de **secréter** naturellement un certain nombre de protéines, ceci facilite donc le processus de purification des protéines recombinantes. Il existe deux types de signaux pour les sécrétions : les préséquences (ex : invertase) et les préproséquences (ex : facteur α). Ces séquences sont clivées lors de l'exportation des protéines hors de la cellule. Il a été constaté que la sécrétion fonctionnait très bien pour des peptides mais était plus aléatoire avec des protéines de taille plus importante.
Les deux principaux types de levures utilisés sont la levure boulangère *Saccharomyces cerevisiae* et la levure méthylotrophe *Pichia pastoris*. Des protéines telles que l'insuline, des interférons ou des immunoglobulines (Wood et al., 1985) ont été exprimées dans la levure *S. cerevisiae*. Cependant de nombreuses protéines étaient **hyperglycosylées** dans cette levure (Brake, 1990) et en conséquence biologiquement inactives. Quant à la levure *P. pastoris*, elle possède une capacité de glycosylation inférieure à celle de *S. cerevisiae*. Le système d'expression est basé sur le promoteur fort AOX1 (alcool oxydase), lui-même sous le contrôle d'un inducteur : le méthanol (Chen et al., 1996). De nombreux vecteurs

utilisant ce promoteur ont été développés et ont permis la surexpression de protéines (Muslin et al., 2000 ; Rodriguez et al., 2000).

Ce système semble donc adapté à l'expression de PA1b. Toutefois, des essais de production en système de sécrétion de PA1b recombinant par *P. pastoris* avaient été réalisés, dans l'unité BF2I, sans succès (Zhang, 2001, données non publiées) et les systèmes basés sur des levures n'ont donc pas été choisis pour notre étude.

Les problèmes inhérents à la sécrétion et à la glycosylation ont conduit au développement de systèmes d'expression chez les eucaryotes supérieurs.

2) cellules d'insectes

Depuis son développement dans les années 80, l'utilisation du système d'expression du baculovirus, virus infectant les insectes, pour la synthèse de protéines recombinantes dans les cellules d'insectes s'est largement répandu (Luckow and Summers, 1988 ; Murhammer, 1991). Les gènes étrangers sont insérés dans le génome viral par recombinaison sous le contrôle d'un promoteur fort. L'avantage majeur du vecteur du baculovirus est que son promoteur fort (polyhédrine) permet de produire des protéines recombinantes en **grande quantité**. De plus la manipulation génétique de ce système d'expression eucaryote est simple et rapide. Le baculovirus, comme le virus de la vaccine, a une très grande taille, ce qui permet le clonage de grands fragments d'ADN.

Concernant PA1b, ce système n'a pas été envisagé, le peptide étant toxique pour divers insectes et son mode d'action exact sur les cellules d'insecte inconnu (il entraîne l'apoptose des cellules du tube digestif des charançons sensibles (Vaublanc, 2001). Tout récemment, il a de plus été montré très toxique sur les cellules Sf9, qui sont les cellules les plus courantes d'expression hétérologue du système baculovirus (Gressent F., 2003, communication personnelle).

3) cellules de mammifères

Les cellules de mammifères sont généralement considérées comme le meilleur hôte pour l'expression de protéines eucaryotes nécessitant des **modifications post-traductionnelles** pour être actives. Deux types d'expression sont distinguées : l'expression transitoire et l'expression stable.

- L'expression transitoire est souvent utilisée pour détecter la fonction d'un gène nouvellement cloné. Elle est de courte durée (48-72 heures). L'ADN n'est pas intégré dans le génome, il se maintient dans le noyau pendant quelques jours avant de disparaître. Des cellules sont transfectées avec des plasmides, puis récoltées 48-72 heures plus tard alors qu'elles sont encore en phase d'expression transitoire.
- Bien que les systèmes transitoires produisent suffisamment de protéines pour des études physico-chimiques, l'expression stable est préférée. Elle permet une production à plus grande échelle, même si les **rendements restent limités** par rapport à une expression dans les autres hôtes cellulaires. L'intégration de l'ADN dans le noyau peut se faire par transfection ou par infection par des virus rendus non pathogènes (SV40, adénovirus, virus d'Epstein Barr) ou rétrovirus (Kaufman, 1991). Ces virus sont rendus défectifs, les gènes essentiels à leur réplication ayant été supprimés. Ces virus infectent les cellules de mammifères et y intègrent leurs génomes.

Les principaux inconvénients de ces systèmes sont leur **coût élevé**, une fragilité des cellules, une capacité de production limitée (10 mg par litre), une possible protéolyse des produits d'expression. Ces systèmes sont donc destinés à la production de protéines à usage thérapeutique et à haute valeur ajoutée, comme l'activateur tissulaire du plasminogène (Pennica et al., 1983).

Une alternative aux problèmes rencontrés dans la surexpression des protéines recombinantes dans la levure ou les cellules animales est l'utilisation de la transgénèse animale ou végétale.

4) Transgénèse

(a) La transgénèse animale

Cette technique est devenue quasiment routinière en recherche fondamentale dans la mesure où elle permet d'aborder la fonction des gènes. En transgénèse animale, 90% des animaux transgéniques sont des souris. Plusieurs méthodes peuvent être utilisées pour insérer le gène d'intérêt. La première consiste à microinjecter le gène dans les pronuclei d'embryons. Ceci est applicable à toutes les espèces animales, mais les coûts sont élevés (Hammer et al., 1985). La seconde consiste à obtenir des embryons *in vitro* à partir d'ovaires collectés dans des abattoirs, ce qui nécessite que la maturation des ovocytes, la fécondation, le transfert de gènes et le développement des embryons jusqu'au stade blastocyste soient effectués *in vitro* (Krimpenfort et al., 1991). Pour les souris, une troisième méthode a été mise en place basée sur la recombinaison homologue dans les cellules souches embryonnaires totipotentes (cellules ES) et donc l'intégration du gène d'intérêt. Des vecteurs rétroviraux peuvent être utilisés pour introduire le gène d'intérêt dans des cellules embryonnaires de souris.
Il convient de choisir l'animal en fonction de la quantité de protéines à produire (de quelques centaines de milligrammes par an par souris à une tonne par an pour une vache).
Dans tous les cas, les coûts engendrés par ce type de production, le problème de la reproduction (temps, perte du gène d'intérêt par recombinaison génétique) et les contraintes liées à la manipulation des animaux génétiquement modifiés, limitent l'utilisation de ces systèmes à des protéines de haute valeur ajoutée.

(b) La transgénèse végétale

Les plantes possèdent un fort potentiel pour la production de protéines d'intérêt pharmaceutique. En effet, la culture de plantes est **bon marché** et une immense quantité de protéines peut être obtenue à partir d'un seul champ. De plus, la reproduction du matériel génétiquement modifié est plus rapide que pour les animaux, et le gène d'intérêt est préservé.
Du fait de la paroi cellulosique de la cellule végétale, l'introduction d'un gène dans le génome n'est pas évidente, plusieurs techniques existent :

- par des vecteurs bactériens ou viraux spécifiques des végétaux qui permettent d'introduire la construction génétique dans les cellules hôtes (Hayes et al., 1988),
- par un "canon à gènes" qui permet d'injecter, dans les cellules hôtes embryonnaires, la construction génétique préalablement déposée sur des microbilles d'or ou de tungstène (Klein et al., 1989),

- par électroporation de protoplastes (cellules végétales ne contenant plus qu'une membrane plasmique, la paroi cellulosique ayant été éliminée par des cellulases) (Shimamoto et al., 1989).

Après intégration du gène d'intérêt, les cellules végétales se multiplient sous forme de cals (amas indifférencié de cellules), puis regénèrent des jeunes pousses. Ces jeunes plantes sont stabilisées par rétrocroisement sexué (backcross) puis cultivées dans des serres ou champs de production. Une grande variété de protéines a été produite dans des végétaux tels que le tabac, le maïs, le riz, la pomme de terre (Morassutti et al., 2002 ; Sawahel, 2002 ; Sijmons et al., 1990). Cependant, il reste à **améliorer les niveaux d'expression**, d'extraction et de purification des protéines. A terme, les plantes pourraient représenter le moyen de production de protéines recombinantes le moins coûteux pour l'industrie.

La réalisation de riz transgénique exprimant PA1b est actuellement en cours au CIRAD de Montpellier (thèse de J. Petit, groupe de J.C. Breitler et E. Guiderdoni). L'objectif principal de cette expression hétérologue est d'estimer l'intérêt de PA1b pour la protection des céréales vis-à-vis des charançons des céréales, et non pas de produire et purifier le peptide ou d'éventuels variants. Ce travail constituera cependant le système de validation d'activité le plus proche de l'expression native[17].

B. Systèmes procaryotes

Malgré le fort développement des systèmes eucaryotes ces dernières années, les systèmes d'expression les plus utilisés sont d'origine procaryote. Parmi les différentes bactéries disponibles, l'entérobactérie *Escherichia coli* est l'hôte le plus employé pour la production de protéines hétérologues aussi bien en recherche qu'en production industrielle. Sa génétique est très bien connue et il s'agit d'un très bon producteur de protéines recombinantes, poussant rapidement et à haute densité (Baneyx, 1999). Les taux d'expression obtenus peuvent aller jusqu'à plusieurs grammes par litre. De nombreux vecteurs d'expression et souches mutantes ont été développés. *E. coli* est donc **facile d'utilisation** et constitue un système de production **peu onéreux**, pour des rendements importants. Les gènes d'intérêt peuvent être exprimés dans différents compartiments cellulaires cytoplasmiques (sous forme soluble ou précipitée) ou extra-cytoplasmiques (périplasme ou milieu de culture). Cependant, *E. coli* présente de nombreux inconvénients qui seront développés dans les pages suivantes.
Un autre hôte couramment utilisé est la bactérie *Bacillus subtilis*. Elle possède une capacité de sécrétion supérieure à *E. coli*. Cependant sa génétique est moins bien connue et le niveau de production de protéines est inférieur à *E. coli*. Cet hôte cellulaire ne sera pas détaillé dans cette partie bibliographique.

1) Stratégie générale pour l'expression d'un gène chez *E. coli*

L'approche classique pour la surexpression d'un gène dans la bactérie *E. coli* commence par l'insertion du gène dans un vecteur d'expression contenant des caractéristiques particulières. La seconde étape est la transformation d'une cellule hôte par le vecteur. L'expression est alors effectuée en détournant une partie de l'énergie de la cellule hôte

[17] Expression dans la graine de plusieurs constructions comprenant différents segments de la préproprotéine PA1.

vers la production de la protéine hétérologue. La notion de surexpression est relative. En effet la protéine la plus abondante naturellement chez *E. coli* représente environ 5% des protéines totales. L'objectif de surexpression pourra être considéré comme atteint à partir de 5% des protéines totales. Pour atteindre cet objectif, il faudra prendre en compte les trois partenaires : le gène, le vecteur et la souche (Masson, 1993).

(a) le gène

La nature d'un gène hétérologue n'est pas sans impact sur le taux d'expression dans *E . coli*. Les gènes d'origine eucaryote ne sont pas toujours surexprimés efficacement dans ce microorganisme. Ces problèmes peuvent être dû aux caractéristiques structurales et uniques de la séquence du gène, à la stabilité et à l'efficacité traductionnelle des ARNm, à la présence de protéases, à la toxicité de la protéine d'intérêt pour l'entérobactérie *E. coli* (Makrides, 1996). Mais la contrainte la plus fréquente est un **usage des codons** différents entre *E. coli* et l'organisme d'où provient le gène, qui va jouer à la fois sur la qualité et la quantité de protéines surexprimées (Gouy and Gautier, 1982 ; Ikemura, 1985). Chez *E. coli*, les codons rares ont été définis comme étant utilisés moins de 10% du temps (Kane et al., 1992). L'utilisation de ces codons rares peut affecter drastiquement les rendements protéiques (Makoff et al., 1989). Ceci est la conséquence de stocks limitants en ARNt spécifiques de ces codons (Hannig and Makrides, 1998 ; Pedersen, 1984). Il n'est donc pas surprenant que l'utilisation de tels codons provoquent des retards dans les mouvements des ribosomes et des pauses traductionnelles (Bonekamp et al., 1985). A cause de ces problèmes, les codons rares peuvent générer des changements de cadre de lecture (frameshift) et conduire à la production de protéines aberrantes, ou encore conduire à des sauts traductionnels provoquant une hétérogénéité de protéines surexprimées (Kane et al., 1992). Une solution pour améliorer les niveaux de production hétérologue est le remplacement des codons rares par les codons les plus en usage chez *E. coli* (Buell et al., 1985 ; Hernan et al., 1992 ; Makrides, 1996) ou bien la co-expression des ARNt spécifiques de ces codons rares (Brinkmann et al., 1989).

La construction de gènes synthétiques constitue une autre solution. Lors de l'étape de conception, différents paramètres devront être pris en compte :

- le respect de la structure primaire de la protéine à exprimer,
- l'utilisation des codons les plus en usage chez *E. coli* et ce dans l'optique d'obtenir des rendements de surexpression satisfaisants (Gouy and Gautier, 1982),
- des sites uniques pour des endonucléases de restriction pourront être insérés le long de la séquence du gène. Ils permettront une dérive génétique ultérieure.

Généralement, le gène est construit par assemblage de différents fragments d'ADN double brins les uns à la suite des autres dans un vecteur de construction. Des oligonucléotides correspondant à chaque brin d'ADN du fragment sont synthétisés par voie chimique puis assemblés par complémentarité. Chaque fragment pourra être inséré de manière unidirectionnelle grâce à l'utilisation de sites de restriction compatibles (Howell and Blumenthal, 1989 ; Martin et al., 1995 ; Philips et al., 1989 ; Samal et al., 1995), et parfois non regénérables (Elmorjani et al., 1997 ; Kempe et al., 1985). Une autre méthode qui est également utilisée pour la construction de gènes synthétiques est l'amplification par PCR (Prodromou and Pearl, 1992). Cette technique permet de polymériser des séquences en utilisant la technique de chevauchements multiples. Cependant, la PCR présente un énorme désavantage. En effet, les polymérases utilisées, mêmes celles considérées comme les plus fidèles, peuvent réaliser des erreurs de polymérisation.

Après avoir été synthétisé, le gène est vérifié par un séquençage (Sanger et al., 1977), puis inséré dans des vecteurs d'expression.

(b) le vecteur d'expression

Le choix du vecteur d'expression est un élément important pour la synthèse protéique. Ces vecteurs comportent différents éléments génétiques qui vont gouverner la transcription et la traduction (Figure 73).

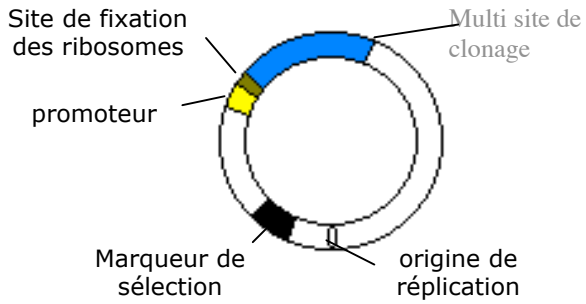


Figure 73 : Représentation schématique d'un vecteur d'expression.

Le marqueur de sélection assure le maintien du plasmide dans la souche. Le plus souvent, il s'agit d'un gène de résistance à un antibiotique. Les marqueurs les plus utilisés sont le gène *bla* ou *Ap* (il code pour la β-lactamase et confère une résistance à l'ampicilline), le gène *cat* (il code pour la chloramphénicolacétyltransférase et confère une résistance au chloramphénicol) et le gène *tet* (il code pour une protéine membranaire et confère une résistance à la tétracycline).

L'origine de réplication, souvent dérivée de ColE1, est un élément essentiel. Elle permet au vecteur de se répliquer de façon autonome par rapport au chromosome bactérien et elle fixe le nombre de copies du plasmide par cellule. Celui-ci varie en fonction du taux de croissance de la cellule hôte. Mais généralement, on considère que l'on a entre 10 et 50 plasmides par cellule. La présence de plasmides multi-copies, donc de multiples copies du gène, doit permettre d'augmenter la production d'ARNm et donc de protéines surexprimées. Ceci est vrai s'il n'y a pas d'autres éléments limitants (Masson, 1993).

Le site de clonage multiple (MCS) ou polylinker permet le clonage du gène étranger. Le gène qui y sera inséré sera sous le contrôle du promoteur de la transcription, situé en amont (Hannig and Makrides, 1998 ; Masson, 1993).

La cassette d'expression contient le promoteur, le site de fixation des ribosomes (RBS) et terminateur de la traduction (Figure 74). L'initiation de la traduction nécessite la présence d'une séquence de fixation du ribosome (RBS ou Shine-Dalgarno), de l'ARNt initiateur, du ribosome et de facteurs d'initiation. Le RBS est une séquence nucléotidique complémentaire de l'extrémité 3' de l'ARN 16S du ribosome, localisée dans la partie non codante de l'ARNm, 5 à 13 nucléotides en amont de l'ATG initiateur de la traduction (Makrides, 1996 ; Shine and Dalgarno, 1974 ; Steiz and Jakes, 1975).
Le promoteur va diriger la transcription en fixant l'ARN polymérase. De la séquence du promoteur va dépendre la force de celui-ci, c'est à dire la fréquence à laquelle la

transcription est initiée à partir de ce promoteur (Masson, 1993). Il existe une grande variété de promoteurs disponibles d'origine bactérienne ou phagique. Le promoteur idéal pour la production de protéines recombinantes est celui qui permet une transcription efficace et donc permet d'obtenir des rendements en protéines surexprimées corrects. Il doit aussi être finement régulé pour minimiser les surcharges métaboliques et les effets toxiques liés à une expression résiduelle (Weickert et al., 1996).

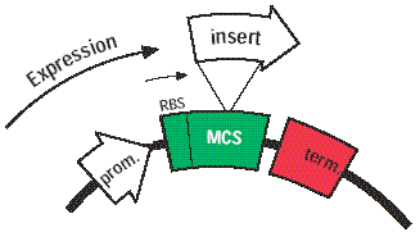


Figure 74 : Représentation schématique d'une cassette d'expression.

Pour l'expression de protéines hétérologues, des promoteurs inductibles sont utilisés. Ce sont des promoteurs forts, activables par un métabolite ou par modification des conditions de croissance. Ce processus permet de dissocier la phase de production de biomasse et celle de production de la protéine d'intérêt, et donc permet d'éviter que la surexpression du gène ait un effet dramatique sur la croissance cellulaire (Weickert et al., 1996). Après une première phase de croissance cellulaire, sans production de la protéine hétérologue, l'ajout d'inducteur au milieu déclenchera la production de cette dernière, généralement aux dépends de la croissance cellulaire (Hannig and Makrides, 1998 ; Masson, 1993). Ces gènes régulés sont sous le contrôle d'un répresseur protéique qui se fixe sur une séquence opérateur située en aval du promoteur, et empêche le passage de l'ARN polymérase. Ce répresseur peut être synthétisé par le génome de la souche. Après ajout de l'inducteur, celui-ci se fixe au répresseur et empêche son interaction avec l'opérateur permettant la transcription. Par exemple, le promoteur p*lac* est contrôlé par le répresseur lac (produit du gène *lac*I) qui se fixe sur l'opérateur. L'induction du promoteur p*lac* est déclenchée par ajout d'IPTG (isopropyl β-D thiogalactopyranoside), un analogue du lactose qui se fixe au répresseur lac et inhibe sa fixation à l'opérateur.
Cependant, il existe un problème majeur dans ce type de régulation. En condition non inductrice, une fuite transcriptionnelle a été observée pouvant entraîner des effets néfastes sur la physiologie de la cellule (Baneyx, 1999 ; Weickert et al., 1996). Elle est due au fait que le répresseur est synthétisé en petite quantité alors que la séquence opératrice située sur le plasmide est présente en plusieurs dizaines de copies par cellule. Pour palier à ce problème, une solution est de cloner sur le vecteur d'expression le gène codant pour le répresseur.
Actuellement, le promoteur du phage T7 est le plus couramment utilisé, et notamment le système pET développé par Studier (Studier and Moffatt, 1986 ; Studier et al., 1990). Ce système a été choisi pour la surexpression de rPA1b. Il sera détaillé dans la partie Résultats (III.A.2)(a)).
Des facteurs génétiques supplémentaires pourront être insérés dans les vecteurs en fonction du type d'expression hétérologue recherché : expression cytoplasmique, périplasmique ou extra-cellulaire. Le choix du vecteur s'effectuera aussi en fonction de l'hôte de surexpression choisi.

(c) la souche

Selon le type de souche utilisée, les rendements en protéines surexprimées seront très variables. Les raisons de ces variations ne sont pas très bien identifiées. Toutefois, la **stabilité des plasmides** et la **dégradation protéique** semblent être deux facteurs responsables de ces modifications.

- Dans certains cas, les souches portant des quantités importantes de vecteurs peuvent être instables. La présence de plasmides dans la cellule nécessite une charge métabolique importante, la cellule aura alors tendance à se développer seule en ségréguant les plasmides (Balbas and Bolivar, 1990).
- De nombreux problèmes de protéolyse ont été mis en évidence lors de la surexpression de protéines hétérologues. Pour tenter de les résoudre, quelques souches mutantes dépourvues en protéases ont été développées. La principale protéase cytoplasmique est codée par le gène *lon* et semble être responsable de la dégradation des protéines naturellement instables ou anormales (Gottesman, 1990). Un autre groupe de protéases important est celui qui induit une réponse à un choc thermique ou un stress. Les souches dépourvues en facteur σ^{32}, facteur qui contrôle la transcription des gènes de choc thermique, ont une faible capacité protéolytique (Baker et al., 1984 ; Gottesman, 1990 ; Murby et al., 1996). Il existe aussi des souches dépourvues en protéase ompT. Cette protéase située sur la membrane externe de la souche *E. coli* serait responsable de la dégradation des protéines dans les extraits cellulaires (Grodberg and Dunn, 1988). Cependant, toutes les protéases d'*E. coli* ne sont pas identifiées et des problèmes protéolytiques peuvent donc toujours exister. Une alternative à la dégradation protéolytique est d'exprimer les protéines d'intérêt en fusion traductionnelle N- ou C-terminale avec une protéine naturelle d'*E. coli*, de les sécréter dans le périplasme ou encore de les faire précipiter en corps d'inclusion (Murby et al., 1996).

Un autre élément qui gouvernera le choix de la souche est la nécessité pour certaines protéines d'avoir des **ponts disulfures** pour être biologiquement actives. Le cytoplasme d'*E. coli* est réducteur. En absence de formation de ponts disulfures, ces protéines vont s'agréger et précipiter dans les corps d'inclusion. Une solution est d'exporter les protéines recombinantes dans le périplasme où les ponts disulfures pourront se former. Récemment, des souches d'*E. coli* ont été génétiquement modifiées au niveau du gène de la thiorédoxine réductase (*trx*) et/ou de la gluthatione réductase (*gor*) afin de permettre la formation de ponts disulfures dans le cytoplasme de la bactérie (Aslund and Beckwith, 1999 ; Derman et al., 1993 ; Lauber et al., 2001 ; Prinz et al., 1997 ; Verheyden et al., 2000). Des études récentes ont montré que les souches Origami, ayant la double mutation, avaient un potentiel supérieur pour la formation des ponts disulfures par rapport aux souches AD494 seulement modifiées au niveau du gène *trx* (Prinz et al., 1997).

Dans l'optique de maximiser les rendements d'expression et en se basant sur les considérations développées ci-dessus, il sera nécessaire d'utiliser des souches adaptées au système d'expression choisi. Pour certains vecteurs, le choix de la souche va s'imposer du fait du génotype très particulier nécessaire à la régulation du système d'expression. Par exemple, pour les vecteurs d'expression de type pET, choisis pour notre étude, il sera nécessaire d'utiliser une souche d'*E. coli* lysogène du bactériophage DE3, c'est à dire portant les éléments nécessaire à la transcription du promoteur T7 (Studier and Moffatt, 1986) (voir III.A.2)(a)).

2) Expression dans les différents compartiments cellulaires

Escherichia coli est une bactérie gram négative. Son enveloppe cellulaire est composée de deux membranes concentriques : les membranes interne et externe séparées par un gel aqueux : le périplasme (Figure 75). La surexpression de protéines recombinantes pourra s'effectuer dans différents compartiments cellulaires : cytoplasme, périplasme ou sécrétion dans le milieu de culture. Depuis quelques années, l'expression sous forme de protéines de fusion s'est largement développée. Le choix du compartiment d'expression s'effectuera en fonction des attentes (nécessité de modifications post-traductionnelles, repliement correct de la protéine d'intérêt...).

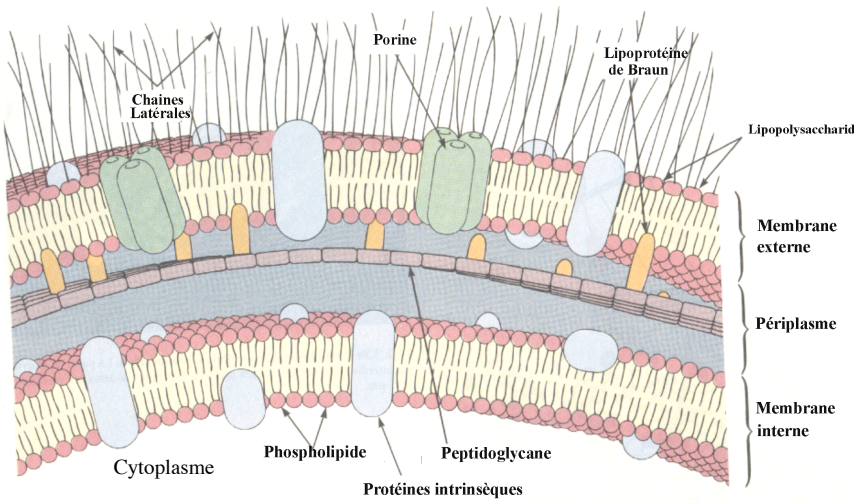


Figure 75 : Représentation schématique des membranes d'*Escherichia coli.* (Extrait de Prescott et al., 1995).

(a) Expression cytoplasmique dans les corps d'inclusion.

La surexpression de protéines recombinantes d'origine eucaryote dans le cytoplasme d' *E. coli* est souvent accompagnée d'un mauvais repliement de celles-ci, ce qui conduit à leur agrégation et à la formation de corps d'inclusion. Ce sont des particules denses de protéines agrégées, dont le diamètre est d'environ 1 µm (Figure 76). Ces corps d'inclusion peuvent contenir de 5 à 40 % des protéines cellulaires totales, mais contiennent en grande partie les protéines surexprimées. Il a été constaté que ces protéines hétérologues avaient malgré tout des structures secondaires partielles (Oberg et al., 1994).

Le mécanisme de formation des corps d'inclusion reste mal connu. Cependant, la nature des protéines exprimées (charge, hydrophobicité, fraction en proline, présence de cystéines) et le niveau d'expression sont des facteurs qui influencent la formation des corps d'inclusion (Mukhopadhyay, 1997). Un taux d'expression élevé ne laisse pas suffisamment de temps au polypeptide naissant pour pouvoir se replier sous une forme native. Ceci combiné avec une forte concentration en protéines hétérologues dans le cytoplasme provoque des agrégations non spécifiques et donc la précipitation des protéines.

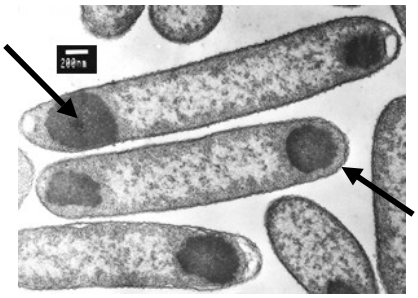


Figure 76 : Corps d'inclusion dans des cellules d'*E. coli* visualisés au microscope électronique.

Différentes approches expérimentales ont été effectuées pour minimiser la formation de ces corps protéiques et améliorer le repliement des protéines :

- Diminution de la température d'expression à 20-25°C. Le taux de synthèse est plus lent, ce qui laisse plus de temps au polypeptide pour se replier (Schein, 1989),
- Utilisation de souches différentes,
- Expression sous forme de protéines de fusion pour augmenter la solubilité des protéines d'intérêt (partie détaillée ultérieurement dans ce chapitre),
- Addition de sucres non métabolisables dans le milieu de culture (Bowden and Georgiou, 1990),
- Co-expression de molécules chaperones. Ces protéines empêchent l'agrégation des polypeptides dans le cytoplasme en aidant au transport et au repliement des protéines. L'agrégation intervient quand les protéines surexprimées ne sont pas capables de se fixer aux chaperones parce que celles-ci sont en quantités insuffisantes (Lee and Olins, 1992). La co-expression des molécules chaperones avec les protéines hétérologues peut résoudre ces problèmes.

Ces paramètres devront être adaptés pour chaque type de protéine surexprimée, puisque la formation des corps d'inclusion dépend entre autre de la nature de la protéine. Cependant ces paramètres ne sont pas applicables dans le cas des protéines à ponts disulfures. En effet, le cytoplasme de la bactérie est un environnement réducteur. Les protéines nécessitant la formation de ponts disulfures pour être actives ne peuvent les former et précipitent. Des cellules souches mutantes, permettant la formation de ces ponts dans le cytoplasme d'*E. coli* ont été isolées (Derman et al., 1993 ; Prinz et al., 1997). Ces mutations inactivent soit le gène de la thiorédoxine réductase, soit le gène de la gluthatione réductase. Les deux thiorédoxines cytoplasmiques sont alors présentes sous une forme oxydée et catalysent la formation des ponts disulfures (Bessette et al., 1999). Ces souches doivent ainsi permettre théoriquement d'obtenir des protéines solubles et correctement repliées.

Malgré tout, les corps d'inclusion présentent quelques avantages. Du fait de la forte concentration en protéines et de leurs formes agrégées, les corps d'inclusion peuvent facilement être isolés des autres protéines cellulaires (Makrides, 1996). De plus, les protéines ainsi précipitées sont protégées de l'action des protéases cytoplasmiques. Et enfin, cette méthode permet la surexpression de protéines, toxiques sous leur conformation native pour la cellule hôte (De Bernadez Clark, 1998), c'est le cas de nombreuses toxines (Howell and Blumenthal, 1989 ; Turkov et al., 1997).

Puisque les corps d'inclusion contiennent de 5-40 % des protéines totales, dont en moyenne 50 % de protéines surexprimées, il est intéressant de les purifier et de les rendre biologiquement actives. La stratégie nécessite trois étapes majeures pour l'obtention de protéines biologiquement actives : l'isolement et la solubilisation des corps d'inclusion, la

purification et la renaturation des protéines dans une forme active, et la séparation des protéines bien repliées de celles mal repliées (Kohno et al., 1991). Dans le cas de PA1b, dont la structure est complexe, l'étape de renaturation apparaît critique et nous n'avons donc pas considéré les corps d'inclusions, mais les protéines solubles dans le cytoplasme.

(b) Expression périplasmique

Le périplasme offre de nombreux avantages pour l'expression de protéines recombinantes. Premièrement, il possède un environnement beaucoup plus oxydant que le cytoplasme. Il est donc particulièrement intéressant pour le repliement correct de nombreuses protéines d'origine eucaryote. En effet, le périplasme contient les enzymes catalysant la formation et le réarrangement des ponts disulfures (Aslund and Beckwith, 1999). Deuxièmement, ce compartiment ne contient que 4 % des protéines totales, soit environ une centaine de protéines. La protéine surexprimée sera donc concentrée et sa purification facilitée. Enfin, le périplasme contient beaucoup moins de protéases que le cytoplasme, les protéines seront donc moins exposées à la dégradation protéolytique (Gottesman, 1990).

Les polypeptides destinés à l'espace périplasmique doivent traverser la membrane interne d'*E. coli*. Une séquence signal située en N-terminal de la protéine hétérologue est nécessaire pour la translocation à travers la membrane interne. La plupart des séquences signal proviennent de protéines naturellement transférées dans l'espace périplasmique, comme OmpA (Ghrayeb et al., 1984), ou la protéine A (Hogset et al., 1990). Le peptide signal a pour but de diriger la protéine et de ralentir son repliement. La protéine chaperonne cytoplasmique SecB interagit et stabilise les protéines destinées à la sécrétion périplasmique, protéines qui doivent être solubles et partiellement repliées pour une translocation efficace (Derman et al., 1993). La suite du processus est ATP-dépendant et fait intervenir différentes chaperonnes membranaires. Les séquences signal sont alors clivées de la protéine d'intérêt par des peptidases liées à la membrane interne. De plus, la méthionine initiatrice de la traduction est généralement clivée en même temps. Ceci permet d'obtenir une extrémité N-terminale identique à celle de la protéine d'intérêt native (Makrides, 1996). Le repliement oxydatif du polypeptide peut ensuite s'effectuer dans le périplasme. Il est catalysé par la protéine DsbA qui est équivalente à la protéine disulfide isomérase (PDI) des systèmes eucaryotes (Baneyx, 1999 ; Mukhopadhyay, 1997) et DsbC qui permet aussi l'isomérisation des ponts disulfures (Cornelis, 2000 ; Lilie et al., 1998).

Cependant dans certains cas, les protéines recombinantes ne sont pas exportées dans le périplasme malgré la présence d'une séquence signal. Ces protéines s'accumulent alors au niveau de la membrane interne et précipitent sous forme de corps d'inclusion ou sont rapidement dégradées par des protéases cytoplasmiques (Baneyx, 1999). Pour tenter d'améliorer cette translocation, différentes solutions peuvent être envisagées. Des composants, intervenant dans le transport et le processus, peuvent être co-exprimés avec la protéine hétérologue (Qiu et al., 1998). Un autre inconvénient possible de la sécrétion périplasmique est le mauvais repliement des protéines d'intérêt et donc la formation de corps d'inclusion ou la dégradation de celles-ci, dans le périplasme. Des stratégies possibles pour résoudre ces problèmes sont la co-expression de chaperonnes moléculaires, de protéines disulfides isomérases ou encore une modification des conditions de culture à plus faible température (Bowden and Georgiou, 1990 ; Makrides, 1996).

Certaines protéines transférées dans l'espace périplasmique peuvent ensuite être sécrétées dans le milieu de culture par une perméabilisation délibérée de la membrane externe.

(c) Sécrétion extra-cellulaire

E. coli est une bactérie qui sécrète naturellement peu de protéines. Cependant la sécrétion présente de nombreux avantages. Premièrement, de nombreuses protéines sont soumises à la dégradation protéolytique dans le cytoplasme. La sécrétion va permettre de résoudre ce problème en séparant physiquement les protéines hétérologues et les protéases. Deuxièmement, la méthionine initiatrice de la traduction, qui peut être à l'origine d'une baisse de l'activité biologique, est clivée durant le processus protéolytique qui accompagne la translocation de la protéine à travers la membrane interne. Troisièmement, l'activité biologique dépend du repliement correct de la protéine. Certaines protéines se replient mal dans le cytoplasme réducteur. La sécrétion de protéines dans le périplasme puis dans le milieu de culture permet d'obtenir un environnement plus favorable pour le repliement des protéines, notamment dans le cas des protéines à ponts disulfures (Better et al., 1988). Enfin, la purification des protéines surexprimées est facilitée car peu de protéines sécrétées sont présentes dans le milieu de culture (Makrides, 1996 ; Stader and Silhavy, 1990).

Cependant, il existe quelques problèmes majeurs pour la sécrétion de protéines. La translocation à travers la membrane interne peut être inefficace ou incomplète. Il se peut aussi que la machinerie de sécrétion soit énergétiquement insuffisante. Tous ces problèmes vont diminuer les rendements en protéines surexprimées. D'autres facteurs, tels qu'un repliement anormal ou des modifications post-traductionnelles incomplètes, peuvent aussi diminuer l'activité biologique des protéines hétérologues. Généralement, les protéines de petites tailles (hormones, facteurs de croissance) sont sécrétées plus efficacement que les protéines plus complexes (Makrides, 1996).

(d) Expression sous forme de protéines de fusion

Depuis quelques années, des stratégies d'expression en fusion traductionnelle des gènes d'intérêt avec un partenaire se développent. Ces systèmes permettent d'augmenter les rendements d'expression et limitent la formation de corps d'inclusion, maintenant leur partenaire soluble dans le cytoplasme. Les partenaires de fusion contiennent les déterminants qui faciliteront la purification, et pourront être facilement clivés de la protéine d'intérêt par des méthodes chimiques ou enzymatiques.

Actuellement quatre systèmes sont principalement utilisés : la protéine A de *Staphylococcus aureus* (SPA) (Nilsson and Abrahmsen, 1990), la glutathion-S-transférase de *Schistosoma japonicum* (GST) (Smith and Johnson, 1988), la Maltose Binding Protein d'*E. coli* (MBP) (Di Guan *et al.,* 1988) et la thiorédoxine de *E. coli* (TRX) (LaVallie et al., 1993). Ces systèmes ont permis la production de protéines recombinantes solubles dans le cytoplasme de la bactérie et correctement repliées. De plus, les niveaux de production ont pu être augmentés. La stabilité et la solubilité des partenaires, ainsi que leurs caractéristiques de repliement, leur permettent d'agir comme des chaperonnes moléculaires covalemment liées à la protéine d'intérêt. Elles donnent une opportunité supplémentaire au polypeptide naissant de se replier plutôt que de s'agréger. Pour assurer une bonne initiation de la traduction, les partenaires sont généralement placés en 5' du gène d'intérêt (LaVallie and McCoy, 1995). De nombreux autres systèmes de protéines fusion existent. Certains d'entre eux vont permettre l'exportation des protéines hétérologues dans le périplasme ou le milieu de culture (cf. I.B.2)(b) et I.B.2)(c)).

Les bons partenaires de fusion possèdent des propriétés qui vont être exploitées pour faciliter la purification. Les protéines porteuses vont servir de « Tag ». Par exemple, les protéines de fusion MBP pourront se fixer sur des résines d'amylose et être éluées avec du

maltose (Di Guan et al., 1988). Chacun des partenaires pourra être purifié par chromatographie d'affinité sur des résines qui leur sont spécifiques. Si les partenaires ne possèdent pas de propriétés favorisant la purification, des étiquettes pourront être ajoutées. Généralement, ce sont des étiquettes polyhistidines (His_6) (Crowe et al., 1994) ou FLAG ($AspTyrLysAsp_4Lys$) (Hopp et al., 1988) qui sont utilisées. Elles sont placées en N-, C-terminal ou entre les deux partenaires de la fusion et permettront une purification par chromatographie d'affinité sur une colonne de métal chélaté (nickel ou cobalt).
Les clivages entre les partenaires de la fusion vont s'effectuer soit par des méthodes chimiques, soit enzymatiques. Pour ce faire, un site de reconnaissance doit être introduit entre les deux protéines. Les principales coupures chimiques et enzymatiques sont résumées dans le Tableau 9. Les méthodes chimiques ne sont pas très spécifiques. Il est fréquent que les protéines d'intérêt contiennent des méthionines ou les dipeptides Asp-Pro et Asn-Gly. Il est alors nécessaire d'effectuer un clivage avec des protéases qui reconnaîtront spécifiquement une séquence. Ces enzymes clivent généralement à l'extrémité C-terminale du site de reconnaissance et il n'y a donc pas d'acide aminé supplémentaire en N-terminal de la protéine d'intérêt. Les protéases sont beaucoup plus sensibles que les agents chimiques à l'encombrement stérique. Le site de clivage devra donc être construit de manière à être accessible aux enzymes.
Après coupure, une étape de purification par chromatographie d'affinité permettra de séparer les deux partenaires. La protéine d'intérêt ne sera pas fixée sur la colonne, tandis que la protéine partenaire le sera par l'intermédiaire de son « tag ».

Tableau 9 : Enzymes et agents chimiques utilisés pour le clivage des protéines de fusion

Méthode de clivage	Séquence de reconnaissance	Référence
Chimique		
Bromure de cyanogène	-Met-	(Kuliopulos and Walsh, 1994)
Clivage acide	-Asp-Pro-	(Szoka et al., 1986)
Clivage à l'hydroxylamine	-Asn-Gly-	(Antorini et al., 2000)
Enzymatique		
Collagénase	-Pro-Val-Gly-Pro-	(Germino and Bastia, 1984)
Entérokinase	-Asp-Asp-Asp-Lys-	(LaVallie et al., 1993)
Facteur Xa	-Ile-Glu-Gly-Arg-	(Zak and Aisen, 2002)
Thrombine	-Gly-Pro-Arg-	(Ahmad et al., 1999)
Trypsine	-Arg-	(Varadarajan et al., 1985)

En conclusion, pour exprimer PA1b :

- Les levures*, a priori* adaptées aux peptides à ponts disulfures, ont été rejetées au vu des essais préliminaires infructueux.
- Les cellules d'insectes n'ont pas été envisagées pour cause de toxicité de PA1b.
- Les cellules de mammifères n'ont pas été retenues essentiellement par la lourdeur de leur manipulation et l'absence de cultures dans l'unité URPVI (Nantes).
- La transgénèse animale comme végétale n'a pas été envisagée, entre autres raisons, pour l'absence de structure adaptée dans l'unité.
- *E. coli* a été choisie pour sa simplicité de manipulation et le fait qu'elle soit *a priori* adaptable à la production de peptides soufrés. Les systèmes bactériens sont, de plus, couramment employés dans l'unité URPVI.

II. Matériel et méthodes

Certaines techniques employées dans cette partie ont déjà été présentées dans les parties précédentes et ne seront par conséquent détaillées que si les protocoles utilisés diffèrent.

A. Construction génétique et expression

1) Souches et plasmides

Les souches et les plasmides utilisés dans cette partie, sont répertoriés dans le tableau suivant :

Tableau 10 : Caractéristiques des souches et plasmides utilisés.

Souches	Génotype	Utilisation
***E. coli* JM109** Source : Stratagène	El4$^-$(*mcr*A$^-$) endA1 gyrA96 thi-1 HsdR17(r_k-m_k+)supE44 relA1 Δ (lac-proAB) F'[traD36 proAB lac 1 ^{q}ZΔM15]	Construction des gènes
E. coli **Origami(DE3)pLysS** Source : Novagen	Δ*ara*$^-$*leu7697* Δ*lacX74* Δ*phoAPvuII phoR araD139 galE galK rspL* F'[*lac*$^+$(*lacI*^{q})*pro*] *gor522* ::Tn*10* (Tc^R) *trxB*::kan (DE3) pLysS Cm^R	Expression hétérologue

Plasmides	Caractéristiques	Utilisation
pPR1 Source : collection du laboratoire	Résistance à l'ampicilline PBluescript modifié : perte du site *Eco*RV, ajout d'un site *Nco*I, *Stu*I et *Sna*BI.	Construction des gènes
pET-DP source : collection du laboratoire	= pET-32b avec modification du site enterokinase (DDDDK) en site DP	Expression hétérologue
pET-32b Source : Novagen	Résistance à l'ampicilline Promoteur T7 Fusion avec la thiorédoxine et une étiquette d'histidines	Expression hétérologue

2) Milieux de culture

Les cultures en routine d'*E. coli*, en erlen, ont été réalisées en milieu riche Luria Bertani liquide (tryptone 10 g/L, extraits de levure 5 g/L et NaCl 10 g/L, pH 7) ou solidifié par l'agar (15 g/L). Le milieu Terrific Broth (TB) liquide (tryptone 12 g/L, extraits de levure 24 g/L, KH_2PO_4 2,2 g/L, K_2HPO_4 9,4 g/L, glycérol 4 mL, pH 7) qui est un milieu plus

riche permettant la croissance cellulaire à plus hautes densités, a été utilisé lors des expériences d'expression en fermenteur.
Des antibiotiques ont été ajoutés aux différents milieux afin de sélectionner les clones résistants : ampicilline (50 µg/mL), kanamycine (15 µg/mL) et chloramphénicol (30 µg/mL) (Sigma).

3) Extraction d'ADN plasmidique

(a) *Extraction à partir d'une culture d'*E. coli

La technique de purification d'ADN plasmidique est basée sur la technique dite de lyse alcaline qui joue sur la renaturation différentielle de l'ADN chromosomique et des plasmides (Birnboim and Doly, 1979). Nous avons utilisé le kit « QIAprep Spin Miniprep Kit » (Qiagen) selon les recommandations du fournisseur.

(b) Extraction à partir d'un gel d'agarose

Cette méthode permet d'extraire et de purifier des fragments d'ADN, préalablement séparés par électrophorèse sur gel d'agarose, grâce au kit « QIAquick gel extraction kit » utilisé selon les recommandations du fournisseur (Qiagen).

4) digestions et ligations d'ADN double brin

Les digestions d'ADN ont été effectuées par des endonucléases selon les recommandations des fournisseurs (Boeringher Manheim, Biolabs, Amersham).
Les ligations d'ADN ont été effectuées avec l'enzyme T4 DNA ligase (Biolabs), en utilisant le tampon recommandé par le fournisseur, pendant une nuit à 16°C.
La stratégie de construction du gène synthétique développée par K. Elmorjani est détaillée dans la partie Résultats. Les séquences des deux oligonucléotides utilisés sont soulignées au niveau des sites de restriction insérés (*Nco*I, *Kpn*I, *Mst*I, *Eco*RV, *Hin*DIII), les codons de PA1b sont représentés en majuscule.
Insert 1 :
atg**ccATGG**CTTCTTGCAACGGTGTTTGCTCTCCGTTCGAAATGCCGCCGTGC**GGTACC**TCTGCTTGCCGT**TGCgcaaagctt**atg
Insert 2 :
atg**gatATC**CCGGTTGGTCTGGTTATCGGTTACTGCCGTAACCCGTCTGGTT**AAgctt**atg

5) Transformation d'*E. coli* par un plasmide

Les transformations génétiques d'*E. coli* ont été effectuées avec des cellules rendues compétentes selon la méthode du « TSS » (Transformation and Storage Solution) légèrement modifiée (Seidman, 1997). Les bactéries cultivées en milieu LB à 37°C jusqu'à une DO_{600}= 0,5-0,6 sont centrifugées. Puis le culot est resuspendu dans 1/10ème du volume de départ dans du TSS fraîchement préparé (10% PEG 3350, 5% DMSO, 50 mM $MgCl_2$, milieu LB). Les cellules sont alors transformées de suite ou aliquotées et conservées à –80°C. Pour la transformation, les cellules sont mises en présence du mélange de ligation pendant 20 minutes dans la glace, puis elles subissent un choc thermique à 42°C pendant 1 minute. Du milieu LB est alors ajouté et le mélange est incubé à 37°C, sous agitation, pendant 1 heure. Les cellules (généralement 100 µl) sont

ensuite étalées sur un milieu solide LB plus antibiotique(s) et incubées à 37°C pendant une nuit.

6) Induction de la surexpression de protéines recombinantes en erlen

Les cellules d'*E. coli* Origami contenant la construction plasmidique ont été mises en culture en milieu LB à 37°C, sous agitation jusqu'à obtention d'une DO_{600}= 0,8. L'induction de la synthèse de protéines recombinantes est déclenchée par l'addition d'IPTG (isopropyl β-D-thiogalactopyranoside) 0,4 mM.
Après surexpression, les cultures ont été centrifugées à 10000 g, à 4°C pendant 5 minutes. Les culots cellulaires ont été remis en suspension, dans le dixième du volume de départ, soit dans un tampon de fixation non dénaturant (Tris-HCl 20 mM, pH 7,9 ; imidazole 10 mM ; NaCl 0,5 M) permettant d'extraire les protéines solubles, soit dans le même tampon dénaturant (Tris-HCl 20 mM, pH 7,9 ; imidazole 10 mM ; NaCl 0,5 M ; urée 7 M) permettant d'obtenir les protéines totales (extraction pour contrôle de l'expression et non pour purification). Les cellules resuspendues ont été soumises à une sonication (4°C, 2x1min30s), et ce lysat acellulaire a été centrifugé (20 000 g, à 4°C pendant 15 minutes). Le surnageant ainsi obtenu constitue la fraction de protéines solubles (en conditions non dénaturantes) ou totales (conditions dénaturantes).
Ces fractions ont été analysées par électrophorèse.

B. Purification et caractérisation des protéines

1) Expression des protéines recombinantes en fermenteur

Une préculture est démarrée dans du milieu LB à partir de la souche étudiée. Les cellules sont ensuite diluées au 1/200^{e} dans du milieu TB contenant les antibiotiques (ampicilline, kanamycine et chloramphénicol). Quand les cultures ont atteint la DO_{600} =1, l'induction de la synthèse des protéines est déclenchée par addition de l'isopropyl β-D thiogalactopyranoside (IPTG) à une concentration de 0,5 mM.

Un schéma du système de fermentation de type « batch » utilisé est présenté Figure 77. Le fermenteur est un bioréacteur d'une capacité de 2 à 5 litres piloté par une Bio-Console ADI 1035 (Applikon® de Dependable Instruments) reliée à un micro-ordinateur. Le logiciel Bioxpert est utilisé pour contrôler le système et acquérir les données. Le fermenteur est équipé pour contrôler l'oxygène dissout, le pH, la température, l'agitation et l'ajout d'antimousse.
Le pH est maintenu à 7 automatiquement par addition d'HCl 1N ou de NaOH 1N (sonde pH autoclavable). La température est maintenue à 37°C par un système de chauffage et refroidissement d'un circuit d'eau. La mousse est supprimée par l'addition automatique d'un antimousse Polymekon 1488 à 0,5 % (Goldschmidt).
Le système, qui est alimenté par de l'air, permet de contrôler le niveau d'oxygène dissout dans le milieu en modulant à la fois le débit d'air et l'agitation. L'agitation peut varier de 200 à 800 rpm (rotations par minute). L'oxygène dissout est mesuré grâce à une électrode polarographique. Un taux relatif d'oxygène dissout à 80 % de la saturation du milieu en oxygène est maintenu. L'aération se fait en fond de cuve grâce à un débimètre d'une capacité de 1 à 10 litres par minute. L'air est filtré (0,2 µm) afin de maintenir la stérilité. Une sortie d'air filtré (0,2 µm) permet de prévenir l'augmentation de pression à l'intérieur du fermenteur. Des échantillons sont prélevés régulièrement pendant la fermentation afin

de mesurer la DO. Le volume utile est de 4 litres. L'ensemencement du milieu ainsi que l'addition d'IPTG se font avec une seringue stérile.

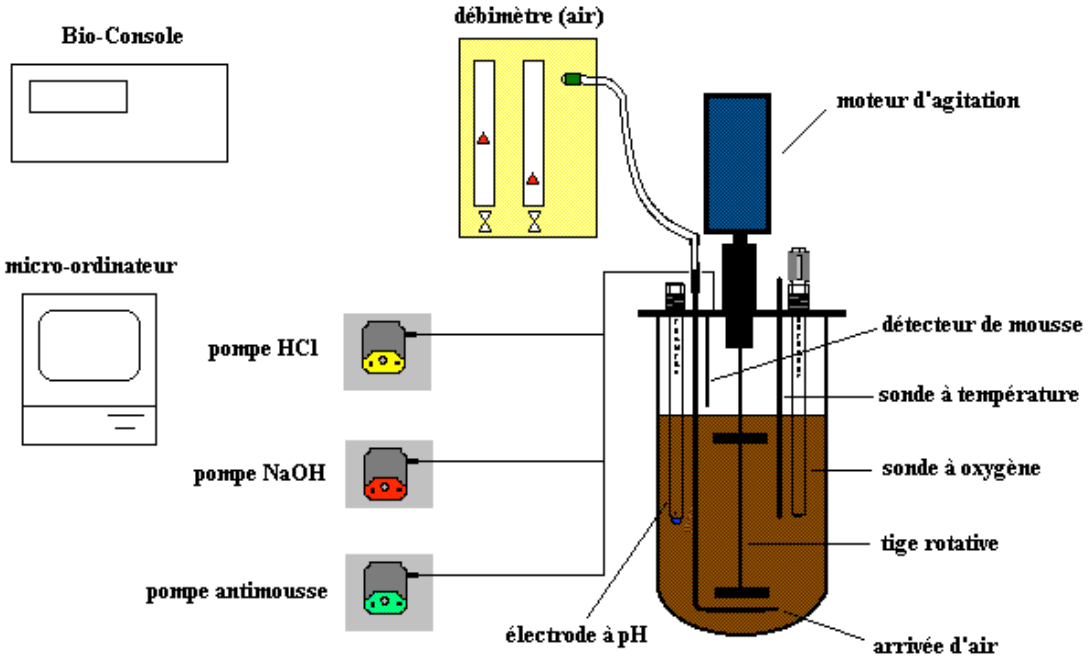


Figure 77: schéma du système de fermentation

2) Purification des protéines de fusion TRX-PA1b par chromatographie d'affinité

Les vecteurs utilisés contiennent une séquence codant pour une « étiquette » de six histidines située entre le gène d'intérêt et celui de la thiorédoxine. Les histidines sont affines pour les ions nickels. Les techniques de purification par chromatographie d'affinité utilisent une résine adaptée (Pharmacia Biotech) sur laquelle ont été fixés des ions nickels.

Après surexpression des protéines de fusion TRX-PA1b chez Origami, une partie de la protéine recombinante produite reste sous forme soluble. Toutes les expériences de purification ont été conduites sur cette fraction de protéines solubles. Les tampons utilisés ne sont donc pas dénaturants (pas d'urée).
Après expression hétérologue en fermenteur contrôlé, les cultures ont été centrifugées à 10000 g, 4°C pendant 15 minutes. Les culots cellulaires ont été remis en suspension, dans le dixième du volume de départ, dans le tampon de fixation (Tris-HCl 20 mM, Imidazole 10 mM, NaCl 0,5 M, β-mercaptoéthanol 5 mM pH 7,9) permettant l'extraction des protéines solubles. Une sonication de quelques minutes, est réalisée dans la glace. Après une centrifugation à 20000 g, 4°C pendant 30 minutes, le surnageant contenant la fraction de protéines solubles est prélevé et déposé sur la résine (avec ions Ni^{2+}) préalablement équilibrée avec le tampon de fixation. Après passage du lysat à travers la colonne, et dans le but d'éliminer les protéines ayant une faible affinité pour le nickel, un premier lavage

est effectué avec ce même tampon puis les lavages et l'élution sont effectués avec des concentrations croissantes en imidazole (20 et 40 mM pour les lavages, jusqu'à 0,5 M pour l'élution). Plusieurs conditions ont été testées (cf. Résultats).
Les fractions éluées sont dialysées à 4°C (membranes de porosités 6000 Da, the Spectrum Companies) contre de l'eau ou un tampon adéquat (3 bains successifs).

3) Purification des protéines fusion par chromatographie d'échange d'ion (IEC).

Afin d'augmenter la pureté de la protéine de fusion avant clivage, une IEC a été effectuée. Le pI (point isoélectrique) théorique de la protéine de fusion TRX-DP-PA1b étant de 5,4, nous avons choisi un tampon basique pour une colonne échangeuse d'anions[18].
En sortie d'affinité, l'éluat contenant les protéines TRX-DP-PA1b est dialysé contre 3 x 5 litres de tampon 1 (Tris-HCl 20mM, pH 8,5 ; β-mercaptoéthanol 5mM). Il est ensuite chargé sur la colonne échangeuse préalablement équilibrée en tampon 1 puis élué par un gradient en sel : par palier de 20, 40 et enfin 100% de tampon 2 (Tris-HCl 20mM pH 8,5 ; β-mercaptoéthanol 5mM; NaCl 1M).

4) Coupure des protéines de fusion

(a) Coupure chimique

Le gène codant pour la protéine de fusion TRX-DP-M-PA1b a été modifié en 5' du premier acide aminé de PA1b. Deux codons spécifiant un dipeptide Asp-Pro particulièrement labile en condition acide ont été insérés grâce aux outils génétiques en amont de la méthionine initiatrice de la traduction de PA1b. La coupure s'effectue au niveau de l'extrémité C-terminale de l'acide aspartique. La protéine recombinante libérée contient donc une proline et une méthionine supplémentaires en N-terminal : PM-PA1b.
Plusieurs conditions de clivage ont été testées, soit avec de l'acide chlorhydrique (HCl), soit avec de l'acide formique (cf. Résultats). Ce dernier est toujours utilisé à 70% et éliminé par lyophilisation après dilution du mélange de clivage par de l'eau (pour amener la concentration en acide sous les 20% et congeler).

(b) Coupure enzymatique

En raison de problèmes de précipitation et des mauvais rendements de la coupure acide de la protéine de fusion TRX-DP-PA1b (voir Résultats), une coupure enzymatique a été envisagée, sur de petites quantités, pour valider le système d'expression. L'entérokinase est une protéase qui reconnaît spécifiquement la séquence « DDDDK » et la clive au niveau de l'extrémité C-terminale de la lysine. Cette séquence spécifique est présente sur le vecteur originel pET-32b (Novagen) et est située juste en amont du site NcoI du vecteur (cf. annexe 6). Le gène synthétique spécifiant PA1b a donc été introduit dans ce vecteur de la même façon que dans le vecteur pET-DP, en phase avec la thioredoxine. Cette coupure enzymatique libère la protéine recombinante PA1b flanquée d'une extention N-terminale AM : AM-PA1b.

[18] Pour pH>pI + 1, la charge nette de la protéine est négative et celle-ci sera donc retenue sur une matrice échangeuse d'anions.

La protéine de fusion TRX-DDDDKA-M-PA1b a été reprise (à 1 mg/mL) dans le tampon Tris-HCl 50 mM, $CaCl_2$ 2 mM pH 7,9. La coupure enzymatique a été effectuée à 21°C pendant 4 heures par addition de l'entérokinase (une unité d'enzyme pour 50 µg de protéines) (Invitrogen). L'enzyme est inactivée par l'ajout de PMSF 1mM.

5) Purification des protéines recombinantes par chromatographie d'affinité

Après coupure, le mélange de clivage a été repris dans (ou mis en dialyse contre) le même tampon de fixation que pour les protéines fusion, puis déposé sur la résine d'affinité préalablement équilibrée avec le même tampon. Un court lavage avec ce tampon a été effectué, ceci afin de bien éluer toutes les protéines d'intérêt. L'élution des protéines de fusion et du partenaire de fusion a été effectuée avec le tampon d'élution (Tris-HCl 20 mM, pH 7,9 ; Imidazole 0,5 M ; NaCl 0,5 M ; β-mercaptoéthanol 5 mM).
Les fractions non-retenues et les fractions de lavage ont été dialysées contre de l'eau, puis congelées et lyophilisées.

6) Purification des protéines recombinantes par HPLC

Les protéines lyophilisées ont été resuspendues en méthanol 60% (MeOH60), filtrées puis injectées sur colonne C18 préalablement équilibrée à 20% de tampon B (Acétonitrile-TFA 0,8%) – 80% de tampon A (TFA 0,1%). Le gradient réalisé est le même que décrit précédemment.

7) Suivi des étapes de purification

Les échantillons prélevés en sortie des différentes étapes de purification ont été analysés par éléctrophorèse :

- par SDS-PAGE à 15% d'acrylamide (Laemmli, 1970) pour ceux contenant la protéine sous forme de fusion, révélée avec une solution de Bleu de Coomassie (BC R250 Serva 0,18% (p/v) ; méthanol 45% ; acide acétique 10%).
- par électrophorèse en Tris-tricine (TTGE) (Schagger and Jagow, 1987) pour ceux contenant le peptide clivé. PA1b fixant médiocrement le bleu de coomassie, les échantillons post-clivage ont été révélés par coloration au nitrate d'argent ($AgNO_3$) (Blum et al., 1987).

8) Western-Blotting et révélation immunologique.

Les protéines ont été séparées par électrophorèse en gel dénaturant et électrotransférées sur une membrane de nitrocellulose comme décrit précédemment. Seule la mise en contact de la membrane avec le premier anticorps varie légèrement : elle s'effectue au $1/500^{ème}$ dans du tampon TBS, pendant 1 heure. L'anticorps polyclonal anti-PA1b utilisé dans la première partie de notre travail est utilisé ici après épuisement sur broyat d'*E. coli* afin d'éliminer les anticorps pouvant reconnaître les protéines de la bactérie.

9) Caractérisation des protéines par spectrométrie de masse.

Après purification, les masses et la pureté des protéines de fusion et de rPA1b ont été caractérisées par spectrométrie de masse sur une trappe d'ions LCQ Advantage (Thermo-Finnigan, San Jose, CA), équipée d'une source électrospray à pression atmosphérique. L'échantillon est infusé en continu à un débit de 5 µL/min. Les spectres sont enregistrés

entre 200 et 2000 M/z et l'acquisition des données est pilotée par le logiciel X-Calibur v 1.3.

10) Caractérisation de l'activité des protéines.

Les échantillons de masses attendues ont été testés pour leur activité biologique sur charançons sensibles (Bénin) et résistants (Chine), par incorporation dans de la farine de blé à une concentration PA1b-équivalente de 100 µg/250 mg d'aliment artificiel (1X : dose pour laquelle le TL_{50} des insectes sensibles est de 4,1 jours) ou 7 fois plus élevée (7X) (voir chap2).
De plus, le peptide recombinant clivé et purifié a été analysé en test d'inhibition de liaison de ^{125}I-PA1b à son site de liaison du charançon. Ce test a été effectué à la concentration 10^{-6} M (concentration pour laquelle PA1b natif inhibe 100% de la liaison)

III. Résultats

A. Constructions génétiques

1) Stratégie de construction

Lors de l'étape de conception du gène synthétique spécifiant PA1b, différents paramètres ont été pris en compte:

- Pour optimiser l'expression des protéines recombinantes dans l'hôte choisi, les oligonucléotides ont été conçus en utilisant les codons les plus en usage chez *E. coli.*
- L'insertion d'un site pour l'endonucléase KpnI dans la séquence du gène. Ceci permet d'envisager une dérive génétique raisonnée par insertion-délétion. C'est le seul site qui a pu être ainsi introduit tout en conservant ouverte la phase de lecture.

La stratégie générale de construction d'un gène synthétique est schématisée sur la Figure 78. Cette approche a été choisie car elle permet une maîtrise totale de la phase de polymérisation. Pour des raisons de limites techniques et de fidélité de la synthèse chimique, cette stratégie est basée sur l'insertion de plusieurs fragments d'ADN les uns à la suite des autres dans le vecteur de construction. Les différents fragments doivent disposer de sites de restriction pour des endonucléases remplissant des conditions précises. Ils doivent être uniques, et leur coupure doit générer des extrémités d'ADN, compatibles deux à deux, afin que le clonage des différents fragments soit **unidirectionnel**.

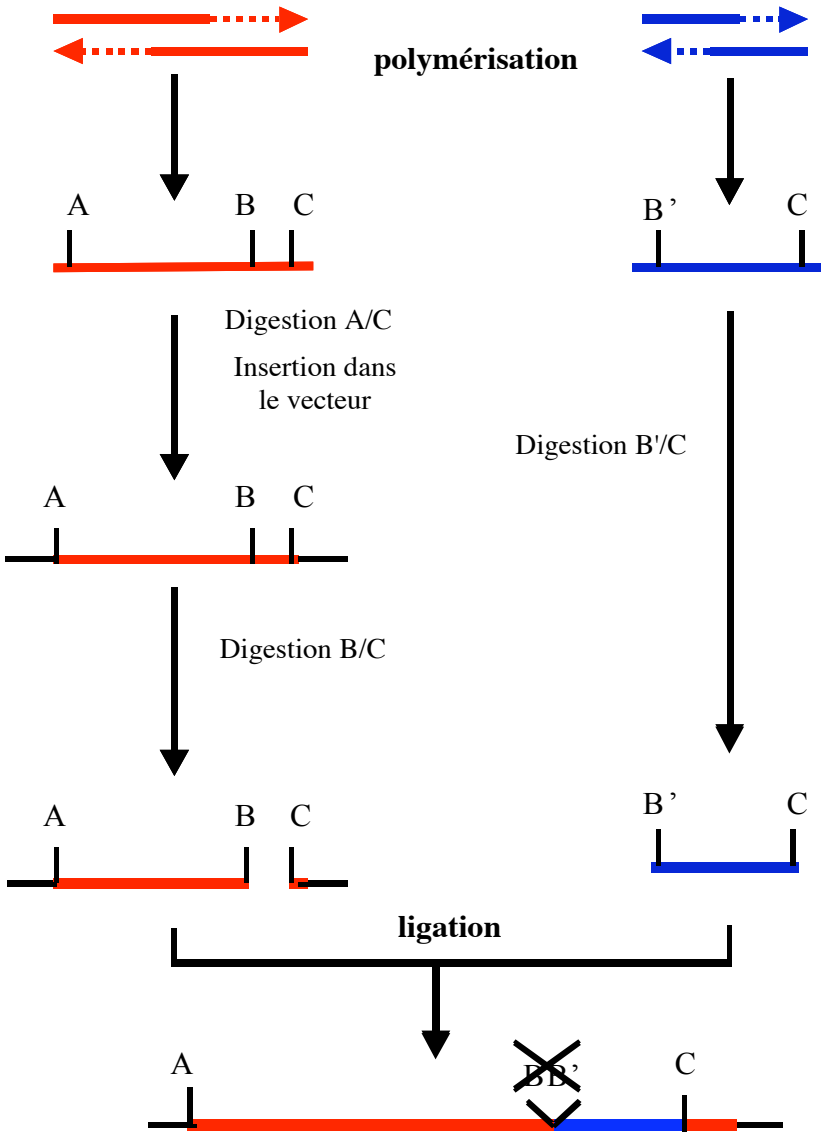


Figure 78 : Stratégie générale de polymérisation d'un gène synthétique. Dans notre cas A= *Nco*I, B= *Mst*I, B'= *Eco*RV, C= *Hin*DIII.

Le gène *pa1b* étant de taille réduite, seuls 2 oligonucléotides, sous forme double brin (db), ont été nécessaires à la construction du gène synthétique. L'étape de polymérisation n'est pas appliquée dans notre cas.
L'appariement en double brin et l'assemblage des deux fragments l'un à la suite de l'autre permet la construction d'une phase de lecture ouverte spécifiant un polypeptide de structure primaire identique à la protéine PA1b extraite du pois, plus une méthionine N-terminale supplémentaire.
Pour les besoins de la construction, les deux fragments double brin contiennent, au niveau de leur extrémité 3', un site unique pour l'endonucléase de restriction *Hin*DIII, permettant leur insertion dans le vecteur de construction pPR1 (plasmide bluescript portant une séquence supplémentaire de 39 paires de bases (pb) entre les sites *Hin*DIII et *Eco*RI, en phase avec le gène *lac*Z (Elmorjani et al., 1997)).
L'insert n°1 possède, à son extrémité 5', un site de reconnaissance pour l'endonucléase *Nco*I dont le clivage génère un bout cohésif. Ce site chevauche un codon ATG spécifiant la méthionine initiatrice de la traduction. Une fois la construction achevée, ce site servira à

l'insertion du gène dans les différents vecteurs d'expression. L'insert n°1 possède également un site pour l'endonucléase *Mst*I à son extrémité 3'. La coupure génère une extrémité « bout franc » qui est compatible avec l'extrémité « bout franc » obtenue après clivage par l'endonucléase *EcoR*V au niveau de l'extrémité 5' de l'insert n°2. En plus de la nécessité de compatibilité des extrémités d'ADN, le choix des enzymes est dicté par le fait que le triplet en 5' des six paires de bases du site de reconnaissance de *Mst*I (TGC) et en 3' du site de *Eco*RV (ATC) sont des codons faisant partie intégrante de la séquence de PA1b et spécifiant deux acides aminés successifs, Cys22 et Ile23.
Grâce à leur structure, les deux fragments du gène *pa1b* vont pouvoir être insérés de façon unidirectionnelle l'un à la suite de l'autre dans le vecteur de construction.
L'insert n°1 et le plasmide de construction pPR1 ont été digérés par les endonucléases de restriction *Nco*I et *Hin*DIII. Après purification à partir d'un gel d'agarose, une réaction de ligation de l'insert db n°1 et du plasmide linéarisé a été effectuée. Des cellules compétentes d'*E. coli* JM109 ont été transformées avec ce mélange de ligation.
La sélection des clones positifs a été effectuée par la résistance à l'antibiotique ampicilline, par la vérification de la taille de l'insert et l'acquisition de sites de restriction uniques. Le plasmide ayant intégré l'insert n°1 a été nommé p*pa1b*1.
L'insert n°2 a été digéré par les endonucléases de restriction *Eco*RV et *Hin*DIII générant respectivement une extrémité à « bout franc » en 5' et une extrémité à bout cohésif en 3'. Le vecteur p*pa1b*1 a été digéré par les enzymes de restriction *Mst*I et *Hin*DIII. Ainsi l'extrémité 5' « bout franc » de l'insert est compatible avec l'extrémité 3' « bout franc » du plasmide linéarisé. Après purification, ligation (unidirectionnelle) de p*pa1b*1 et de l'insert n°2, et transformation de cellules JM109, les clones positifs ont été testés pour la taille de l'insert et la présence/absence des sites uniques de restriction (absence des sites *Mst*I et *Eco*RV, présence des sites *Nco*I et *Hin*DIII). Enfin, un séquençage a permis de confirmer l'obtention d'un vecteur, appelé pPA1b, portant un fragment d'ADN de 114 pb et spécifiant un polypeptide de 38 résidus, identique à PA1b avec une méthionine supplémentaire en N-terminal (M-PA1b).

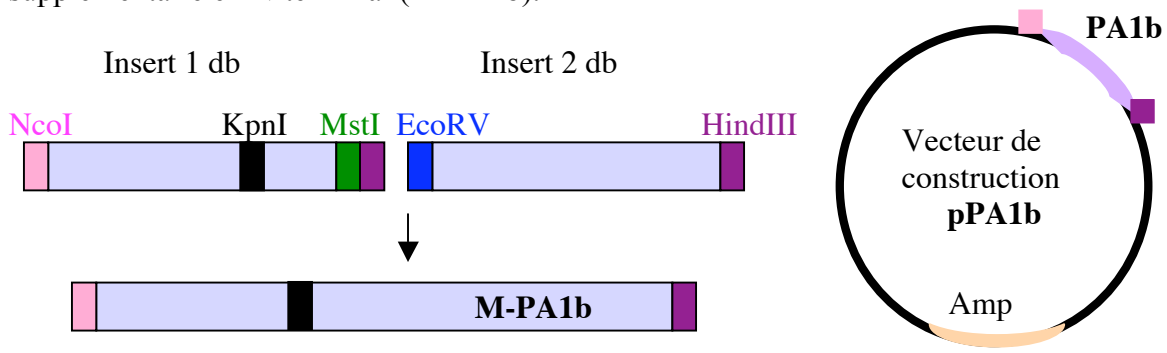


Figure 79 : carte de restriction des inserts, du gène codant M-PA1b (PA1b avec une méthionine en N-terminal) et du plasmide pPA1b.

2) Conditions de la surexpression de la protéine fusion

(a) Choix du système d'expression

Le choix de l'hôte d'expression est crucial, en particulier pour des peptides nécessitant des modifications post-traductionnelles. En effet, si le système d'expression bactérien est couramment utilisé pour sa simplicité et son efficacité, les cellules procaryotes ne peuvent

normalement pas réaliser de modifications post-traductionnelles parfois nécessaires à l'activité de protéines eucaryotes. Le cytoplasme d'*E. coli* est réducteur et ne permet normalement pas la formation de ponts disulfures (Derman and Beckwith, 1991). Toutefois, les mutations des gènes codant pour la glutathione reductase (*gor*) et la thioredoxine réductase (*trxB*) dans la souche Origami rend le cytoplasme de ces cellules plus oxydant et permet la formation de ponts disulfures (Bessette et al., 1999) et donc l'obtention de plus de protéines actives que dans une souche à cytoplasme sauvage (Prinz et al., 1997). C'est donc l'hôte ***E. coli*** (rendement d'expression important, facilité technique d'utilisation) souche **Origami** (cytoplasme permettant la formation de ponts disulfures) qui a été choisi.
Le système pET, distribué par Novagen, est l'un des systèmes les plus puissants développés à ce jour pour l'expression de protéines recombinantes chez *E. coli*. Ce système est basé sur trois niveaux de contrôle de l'expression des gènes d'intérêt. En effet, dans les plasmides pET, le gène synthétique est sous le contrôle du promoteur T7*lac* (Studier et al., 1990). Ce promoteur hybride ne peut être transcrit que par l'ARN polymérase du phage T7 (T7pol) et est inhibé par le répresseur lacR dont le gène est porté à la fois par le plasmide et par le chromosome bactérien. Le gène codant pour la T7pol a été introduit sur le chromosome de la bactérie (DE3 lysogène), lui-même sous le contrôle du promoteur lacUV5, réprimé par lacR et inductible par l'IPTG. Après addition d'IPTG, la polymérase T7 est donc synthétisée, permettant de transcrire le gène d'intérêt. Ce système de contrôle au niveau du promoteur T7*lac* limite l'expression basale du gène d'intérêt liée à la fuite de synthèse de la T7pol observée même en absence d'IPTG. Pour accroître encore la stabilité des plasmides spécifiant des protéines potentiellement toxiques pour la bactérie, les souches, adaptées pour ce système, portent un plasmide secondaire à faible nombre de copies, le pLys, qui leur confère une résistance au chloramphénicol et porte le gène spécifiant le T7 lysozyme. Cet inhibiteur de la T7 ARN polymérase (fuite d'expression) permet un contrôle plus stringeant de l'expression des gènes recombinants (Moffatt and Studier, 1987). Ce système est très puissant puisque la régulation porte sur une ARN polymérase exclusive en plus du promoteur.
Dans notre cas, la souche Origami DE3 choisie porte le plasmide pLysS qui permet une faible expression du lysosyme et ne limite donc pas la croissance bactérienne.

Les essais préliminaires d'expression du gène *pa1b* directement sous le contrôle du promoteur T7 ont montré que les niveaux de cette expression étaient extrêmement faibles. Pour améliorer ces niveaux, ce gène a été cloné dans le vecteur pET32b en fusion traductionnelle avec la thiorédoxine. Le choix de la thiorédoxine comme partenaire de fusion a également été dicté par le fait que cette protéine permet de garder solubles nombres de protéines dont la précipitation dans le cytoplasme d'*E. coli* est quasi totale. C'est une condition *sine qua none* pour permettre la mise en place de ponts disulfures. De plus, elle joue probablement un rôle actif dans la formation des ponts disulfures de son partenaire (Stewart et al., 1998).
Le vecteur pET32b permet l'expression de protéines en fusion avec la thiorédoxine, une étiquette de 6 cystéines, et un polypeptide présentant le site de clivage à l'entérokinase DDDK (codé au début du site de clonage multiple). Vu le coût d'un clivage enzymatique, ce plasmide avait été modifié par insertion d'un fragment d'ADN entre les sites *Bgl*II et *Nco*I, remplaçant les acides aminés D, K et A par une seule proline mais conservant les sites de restriction et la phase de lecture. Un dipeptide DP est alors codé par ce plasmide modifié appelé pET-DP. La liaison entre ces deux acides aminés est particulièrement sensible aux conditions acides ce qui permet un clivage chimique.

L'insert codant M-PA1b a été extrait du plasmide pPA1b par restriction par *Nco*I et *Hin*DIII et inséré dans le plasmide pET-DP préalablement digéré par ces mêmes enzymes. Après transformation de bactéries compétentes JM109, le vecteur a été testé pour la taille de l'insert et la présence des sites de restriction attendus. Après les vérifications, il a été appelé pET-DP-PA1b (Figure 80) et utilisé pour transformer des bactéries compétentes Origami. Deux clones positifs ont été sélectionnés (vérification de la séquence clonée). Les souches ont été alors conservées à –80°C avant vérification de l'expression.

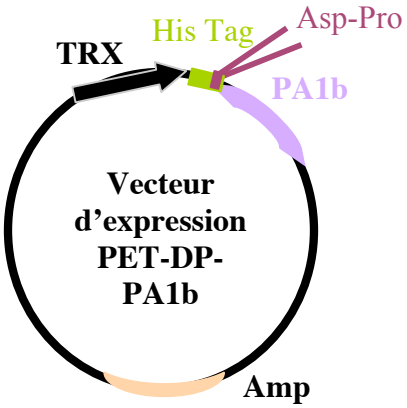


Figure 80 : vecteur d'expression pET-DP-PA1b

(b) Conditions d'expression

a) Vérification en erlenmeyer

La vérification de l'expression de la protéine recombinante chez la souche Origami transformées a été effectuée en erlen sous agitation normale, en milieu LB à 37°C. L'induction a été déclenchée à une DO_{600} de 1 et l'expression a été suivie pendant trois heures. L'apparition d'une bande protéique de masse moléculaire apparente d'environ 25 kDa après induction témoigne de l'expression de la protéine fusion (masse réelle 20 kDa). Cette protéine est reconnue, après transfert sur membrane de nitrocellulose, par les anticorps polyclonaux anti-PA1b épuisés sur *E. coli* (Figure 81).

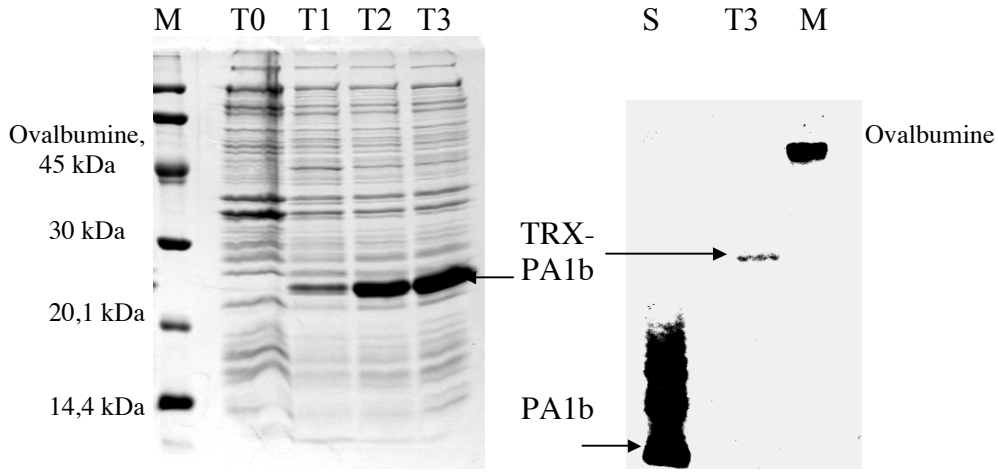


Figure 81 : à gauche : suivi de l'expression (SDS-PAGE, Bleu de Coomassie) ; à droite : immunodétection de TRX-DP-PA1b. T0 à 3 : prélèvements avant et 1 à 3 heures après induction; S: PA1b; M: marqueur.

Le gel d'électrophorèse présenté Figure 82 montre qu'environ 50% des protéines de fusion totales synthétisées restent sous forme soluble dans le cytoplasme d'*E. coli*.

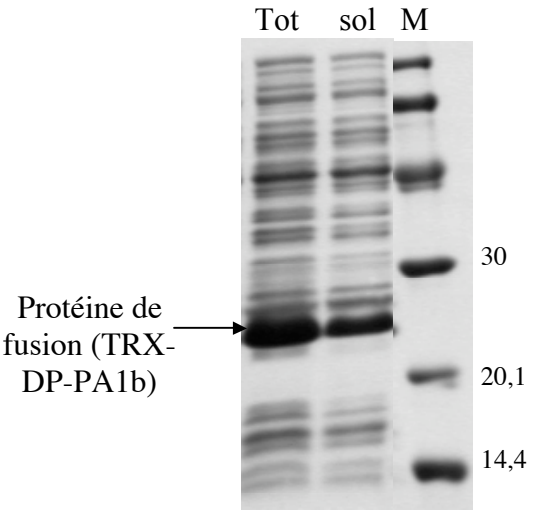


Figure 82 : comparaison des protéines totales (tot) et des protéines solubles (sol) d'*E. coli* après expression (SDS-PAGE, Bleu de Coomassie).

Ces rendements significatifs nous ont permis de ne considérer, par la suite, que cette fraction soluble de protéines extraites.

b) fermentation

Différentes conditions de fermentation, à basses densités cellulaires, avaient été testées au laboratoire sur d'autres protéines recombinantes. Les meilleurs rendements étaient observés avec :

- Un milieu de culture TB, riche, qui permet une forte croissance de la biomasse bactérienne et une meilleure expression.
- Une DO d'induction de 1, pour laquelle les cellules sont encore en phase exponentielle de croissance.
- Une durée d'induction de 3h pour laquelle la quantité maximale de protéines solubles est obtenue.
- Une température d'induction de 37°C.
- Une aération continue à 80 % de saturation en air. Ce facteur est crucial en particulier pour l'expression de la protéine recombinante.

B. Purification et caractérisation protéiques

1) Purification et caractérisation de la protéine fusion

(a) Purification de TRX-PA1b

Une fois les protéines solubles du cytoplasme récupérées, la purification des protéines fusion (TRX-DP-PA1b ou TRX-KA-PA1b, nommées ci-après TRX-PA1b si les deux sont concernées) est réalisée en utilisant l'étiquette de six histidines affines pour le Nickel (Ni^{2+}). Différentes conditions natives d'affinité et d'échange d'anions ont été testées afin d'optimiser la purification des protéines TRX-PA1b.

a) Chromatographie d'affinité par palier.

Les conditions standard appliquées au laboratoire et éprouvées sur d'autres polypeptides recombinants consistent en un premier lavage en tampon de fixation (B) suivi d'un deuxième lavage en tampon (W) de même composition excepté l'imidazole qui s'y trouve deux fois plus concentré (20 mM). Ces lavages sont suivis de l'élution des protéines retenues par un tampon d'élution (E) de même composition, mais de concentration en imidazole de 0,5 M. Ces conditions d'augmentation de la concentration en imidazole par palier nous ont permis de concentrer les protéines d'intérêt et d'éliminer un grand nombre de protéines bacteriennes. On observe toutefois sur gel d'électrophorèse la présence de contaminants résiduels de hauts et surtout de faibles poids moléculaires dans les fractions d'élution contenant la protéine fusion (Figure 83).

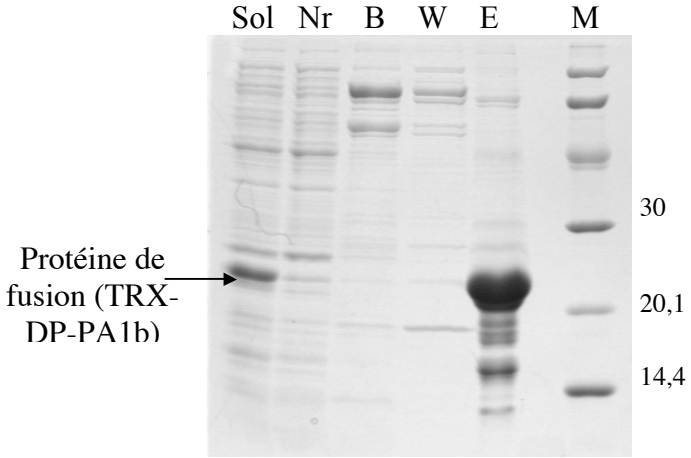


Figure 83 : Fractions de chromatographie d'affinité avec élution par palier d'imidazole (SDS-PAGE, Bleu de Coomassie). Sol : protéines solubles du cytoplasme, Nr : fraction non retenue, B : lavage en tampon de fixation, W : tampon de lavage, E : éluat, M : marqueur.

Ces contaminants interfèrent avec les étapes suivantes de clivage et de purification et nous avons donc cherché à améliorer la pureté des protéines d'intérêt avant clivage. Pour cela, une étape supplémentaire de lavage à 40 mM d'imidazole a été ajoutée avant l'élution. Ce lavage permet d'éliminer une partie des protéines contaminantes de haut poids moléculaire (Figure 89) mais reste très insuffisant pour améliorer significativement le rendement de purification.

Il a été souvent observé qu'en conditions natives, l'efficacité de purification par affinité de protéines à étiquette d'histidines est plus faible qu'en conditions dénaturantes. Afin d'essayer de réduire les interactions protéiques sans dénaturer nos protéines, nous avons réalisé une affinité en présence de 2 M d'urée dans les tampons de fixation et de lavages. Après le troisième lavage, un lavage supplémentaire à 40 mM d'imidazole mais sans urée a été réalisé afin d'éliminer l'urée avant l'élution. L'élution s'est effectuée comme précédemment. Malheureusement, ces conditions peu dénaturantes n'ont pas permis d'améliorer la pureté de la fraction éluée. Une extraction et une affinité en urée 6 M (dans chacun des tampons) à été réalisée, pour vérifier l'efficacité de purification en conditions dénaturantes. Le gel Figure 84 montre que les contaminants représentent alors une moindre portion des protéines de la fraction d'élution. En particulier, le contaminant d'environ 13 kDa semble éliminé au lavage.

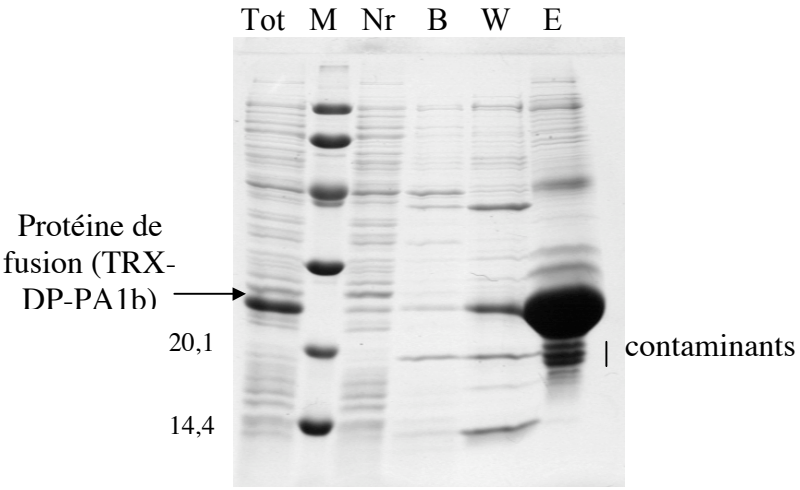


Figure 84 : SDS-PAGE des fractions obtenues par chromatographie d'affinité en urée 6M (révélation au Bleu de Coomassie). Tot: protéines totales, Nr: fraction non retenue, B et W: lavages, E: éluat, M: marqueur.

Toutefois, nous n'avons pas continué en conditions dénaturantes, estimant que les protéines fusion extraites des corps d'inclusion lors de la resuspension et du cassage des cellules en urée poseraint des problèmes de renaturation que les protéines solubles seules n'étaient pas sensées soulever.

b) Chromatographie d'échange d'ion (IEC).

Nous avons donc envisagé une étape supplémentaire de purification en conditions natives par chromatographie d'échange d'ion (IEC). Plusieurs matrices échangeuses d'anion et conditions de chromatographie ont été testées. Le meilleur rapport temps de manipulation sur efficacité de purification a été obtenu sur une colonne de 50mL de résine DEAE Sepharose Fast Flow (Amersham Pharmacia), éluée à 20 mL/min.
Ces conditions optimisées permettent une légère amélioration de la pureté de la protéine fusion par élimination d'une partie des contaminants de poids moléculaire inférieur à 20 kDa (Figure 85 et Figure 86).

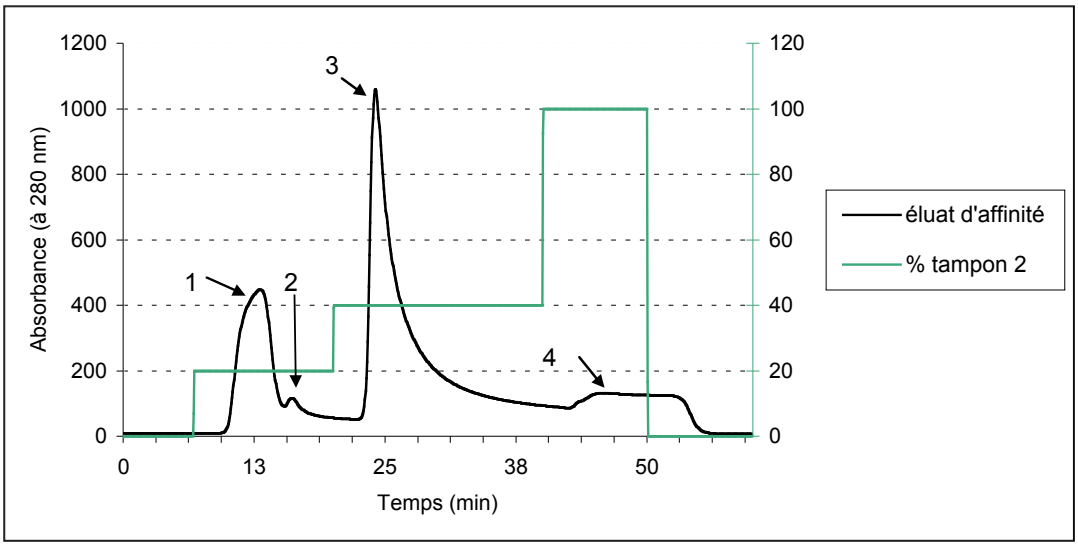


Figure 85 : chromatogramme d'IEC post-affinité. 1 à 4 : pics d'élution analysés en électrophorèse (Figure 86). Environ 300 mg de protéines ont été déposées sur la colonne.

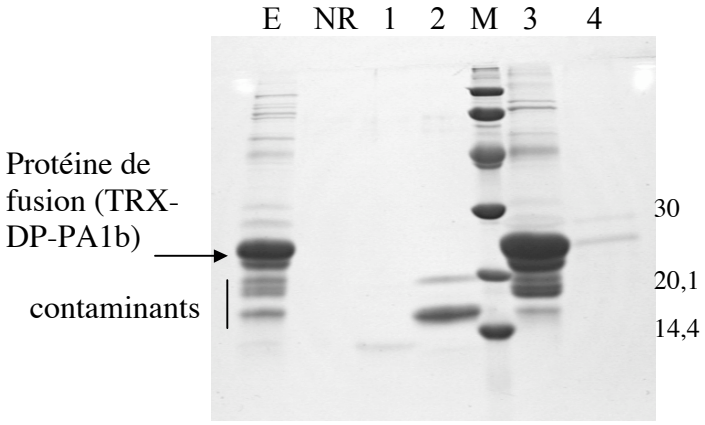


Figure 86 : SDS-PAGE sur les fractions d'IEC post-affinité (Bleu de coomassie). E: éluat d'affinité (injecté en IEC), Nr : fraction non retenue, 1 à 4 : pics d'élution à 20, 40 et 100% de tampon 2 (Figure 85).

L'IEC étant une technique de chromatographie plus grossière que l'affinité qui peut donc intervenir en dernière étape de purification, l'inversion de ces deux techniques a été tentée. Malheureusement, l'échantillon de départ (protéines solubles des bactéries), relativement complexe, n'est guère purifié par l'échange d'anions et l'affinité qui la suit n'aboutit pas à une meilleure purification que sans l'IEC (Figure 87).

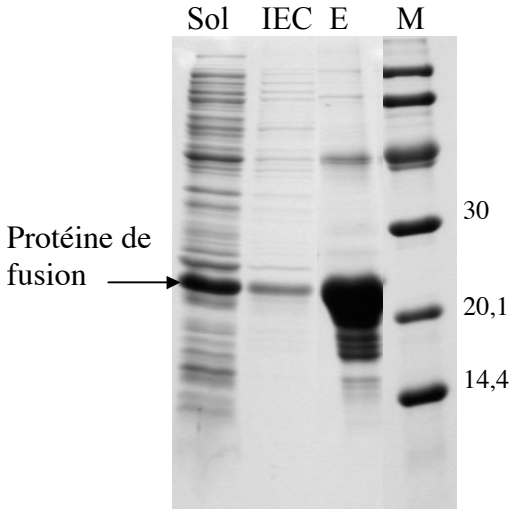


Figure 87 : Purification de la protéine de fusion TRX-DP-PA1b par IEC suivie de chromatographie d'affinité (SDS-PAGE, Bleu de Coomassie). sol: protéines solubles du cytoplasme, IEC : éluat d'IEC, E: éluat d'affinité post-IEC, M : marqueur.

c) Chromatographie d'affinité par gradient linéaire.

Finalement, le rapport temps de manipulation/amélioration de la purification est faible pour l'IEC. Une optimisation de la chromatographie d'affinité a donc été envisagée par un gradient linéaire en tampon d'élution en remplacement du passage par paliers à 0,5 M d'imidazole.

La meilleure purification obtenue utilise pour l'élution un simple gradient linéaire de 8% à 100% de tampon E en 8 volumes de colonne. Le suivi de l'élution des protéines est représenté Figure 88. Le gel d'électrophorèse permettant de contrôler la sortie et la pureté des protéines des différentes fractions est présenté Figure 89. On observe que ce simple

changement permet l'élimination d'un pic important de contaminants de poids moléculaires inférieurs à celui de la fusion (E1). Ces protéines de faible poids moléculaire étant les plus gênantes pour le suivi du clivage et la purification du peptide recombinant, nous avons considéré que ce protocole de purification de la protéine fusion était satisfaisant.

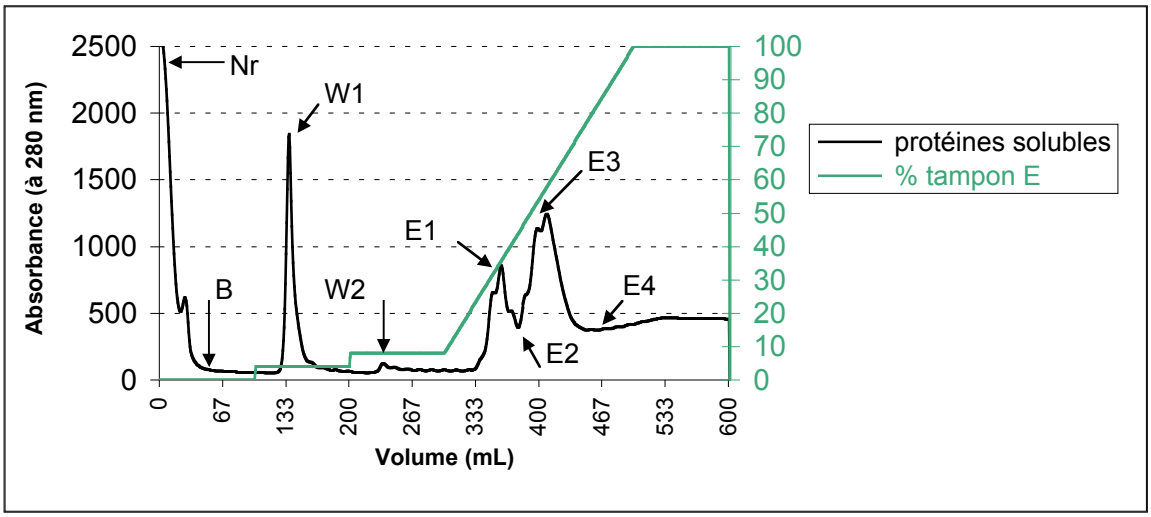


Figure 88 : chromatographie d'affinité (sur colonne de nickel), élution par gradient linéaire d'imidazole (tampon E). Fractions Nr à E4 analysées par SDS-PAGE Figure 89. Nr: fraction non retenue, B: lavage (10mM), W: lavages (1: 20 mM, 2: 40 mM d'imidazole), E1 à 4: fractions d'élution.

Par observation sur gel d'électrophorèse, les protéines fusion semblent pures à 85-90% (Les fractions E1 à E3 sont déposées sur gel sans dilution préalable, pour détecter les éventuels contaminants).

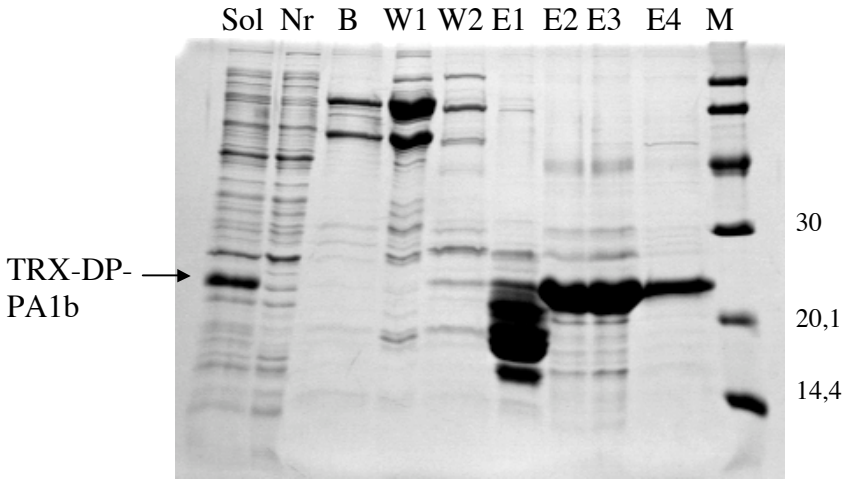


Figure 89 : SDS-PAGE (Bleu de Coomassie) des fractions de la chromatographie d'affinité à gradient d'élution linéaire (Figure 88). Sol: protéines solubles du cytoplasme, Nr: fraction non retenue, B: lavage (10mM), W: lavages (1: 20 mM, 2: 40 mM d'imidazole), E1 à 4: fractions d'élution.

Les fractions correspondant au pic de protéine fusion (E2 à E4) sont collectées, réunies et dialysées contre de l'eau avant lyophilisation ou clivage.

On obtient un rendement correct d'expression - purification de 50 mg de fusion/Litre de culture (pesée après lyophilisation).

(b) Caractérisation de TRX-PA1b

La protéine fusion a été passée en HPLC afin de déterminer son temps de rétention. Les contaminants ne sont pas visiblement séparés par HPLC (un seul pic, Figure 90). D'environ 26 min, le temps de rétention de TRX-PA1b est plus élevé que pour le peptide PA1b natif (environ 21 min). Les restes de fusion non clivée ne devraient donc pas gêner la purification de rPA1b par HPLC.

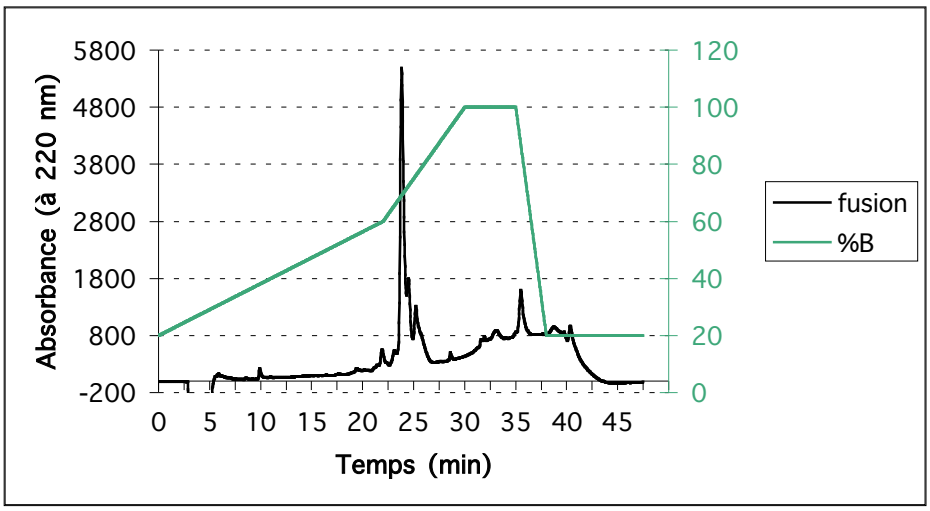


Figure 90 : chromatogramme d'HPLC sur la protéine fusion TRX-DP-PA1b.

Les deux types de protéines fusion purifiées ont été analysées par spectrométrie de masse et testées biologiquement sur charançons sensibles et résistants.

L'analyse ESI-MS de la protéine fusion TRX-DP-PA1b purifiée par affinité à gradient linéaire d'imidazole témoigne de son degré de purification et permet de calculer une masse moyenne de 20 650,4 Da (Figure 91). La masse moyenne réduite théorique (calculée sur Expasy) est de 20 658,49. La différence de 8,09 indique que les quatres ponts disulfures de la protéine fusion seraient bien formés (six cystéines de PA1b et deux de TRX oxydées). Toutefois, la précision de la mesure de masse (200 ppm environ) ne permet pas d'affirmer que les cystéines sont oxydées à 100 %. Cette forme de la protéine est largement dominante, mais on observe également des pics liés de masse plus élevée d'environ 155 Da. Un tel supplément de masse pourrait correspondre à la fixation de deux molécules de β-mercaptoéthanol sur deux cystéines (ouverture d'un pont).

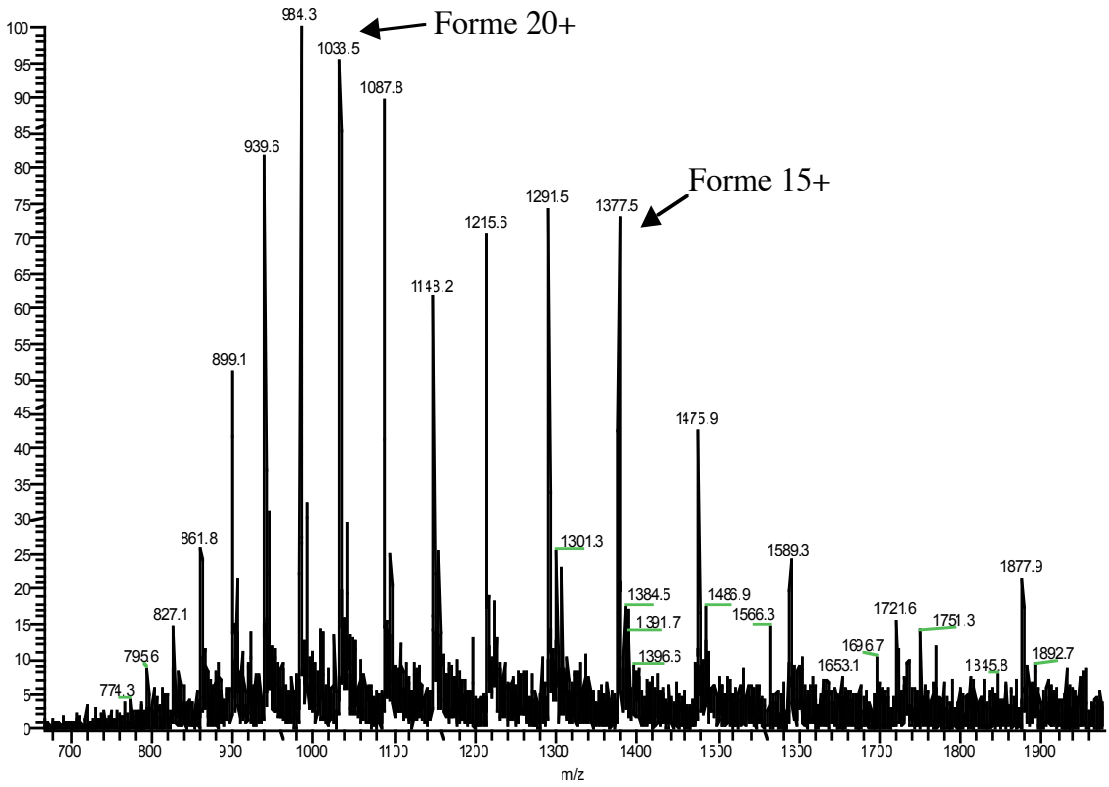


Figure 91 : ESI-MS: spectre non déconvolué de la protéine de fusion TRX-DP-PA1b purifiée. Masse moléculaire moyenne observée : 20 650,4 Da (masse théorique réduite : 20658,5 Da).

Des tests d'activité biologique sur fusion entière ont été réalisés sur différents lots de protéine TRX-PA1b. Ni TRX-DP-PA1b (purifiée en conditions natives ou dénaturantes), ni TRX-KA-PA1b ne montrent de toxicité sur charançons à dose 1X et 7X.
Les fécès des deux souches correspondantes aux tests sur TRX-DP-PA1b 1X ont été récoltées. La fraction soluble au méthanol 60% en a été injectée en HPLC. Les pics observés dans la zone de rétention de PA1b natif ont été analysés en électrophorèse. On retrouve des peptides similaires à PA1b dans l'échantillon provenant des résistants, mais pas dans celui des sensibles. Le même résultat est également obtenu à partir d'un aliment contenant du PA1b natif du pois. PA1b et rPA1b seraient donc retenus dans le tube digestif des insectes sensibles. De plus, ces résultats indiquent qu'un clivage partiel des protéines de fusion a lieu dans ce même tube digestif. Toutefois, l'absence de toxicité peut provenir d'une absence d'activité du peptide rPA1b libéré et/ou de la trop faible dose libérée (dose sublétale).
Ces résultas ne permettent donc pas de conclure quant à l'activité du peptide recombinant : la séparation des deux membres de la fusion s'avère nécessaire.

2) Clivages

La liaison peptidique Asp-Pro est la moins stable de toutes les liaisons peptidiques en conditions acides (Poulsen et al., 1972). De nombreuses publications témoignent de sa labilité dans diverses conditions acides (revues par Landon, 1977). Dans le cadre d'expression hétérologue avec partenaire de fusion, Gram et al. (1994) ont démontré qu'en insérant une telle liaison entre la protéine porteuse et la protéine d'intérêt, cette dernière pouvait être libérée de son partenaire de fusion par traitement acide des corps d'inclusion,

sans solubilisation de ceux-ci par des détergents. L'utilisation d'HCl et d'acide formique a ici été envisagée sous diverses conditions.

(a) Clivage à l'acide formique

L'acide formique à 70% est couramment utilisé car il permet un bonne solubilisation des protéines à cliver (Kawakami et al., 1997). Déjà utilisé sur d'autres polypeptides au laboratoire, il s'avère bien plus efficace que l'acide chlorhydrique. Il a donc d'abord été utilisé, notre protéine ne possédant pas, *a priori*, de résidus sensibles à la formylation (lysine).

Deux conditions de clivage à 70% d'acide formique ont été testées :

- 37°C 72h
- 50°C 24h

Le rendement de clivage à 50°C est plus élevé qu'à 37°C (presque total). La Figure 94 montre l'efficacité de ce clivage, comparé à différentes conditions en acide chlorhydrique.

Suite à ces clivages, rPA1b a été purifié (comme décrit ci-dessous) et analysé par spectrométrie de masse (SM).

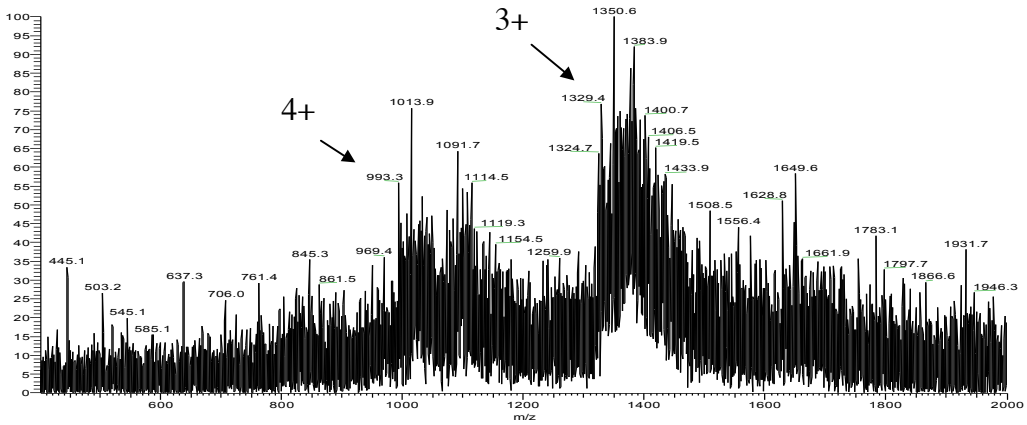


Figure 92 : spectre de masse non déconvolué de rPA1b purifié, provenant de clivage à l'acide formique.

Si on remarque la présence de la masse attendue (3970 Da : pics de m/z à 993,3 et 1324,7), l'hétérogénéité du spectre (Figure 92) alors observée pour les peptides provenant de clivage à 37°C comme à 50°C, et ce malgré une purification poussée, nous a amené à vérifier l'inocuité attendue de l'acide formique sur PA1b natif. Un mélange d'isoformes pures du pois a été soumis aux mêmes conditions de clivage que la fusion (70% d'acide formique, 24h, 50°C) puis analysé par HPLC et spectrométrie de masse. Une hétérogénéité accrue pour le peptide ayant subi les conditions acides a été constatée, d'une part en HPLC (apparition de pics de temps de rétention inférieurs à celui attendu, se révélant de masses apparentées à PA1b) et d'autre part, en spectrométrie de masse avec la présence nette de formes portant un ou plusieurs groupements formyl (28 Da/groupe de plus) (Figure 93). Cette formylation inattendue du peptide natif nous a amené à reconsidérer un clivage acide par l'acide chlorhydrique, celui-ci n'ayant pas d'effets (visibles en HPLC et SM) sur les isoformes de PA1b natif, quelques soient les conditions utilisées ici.

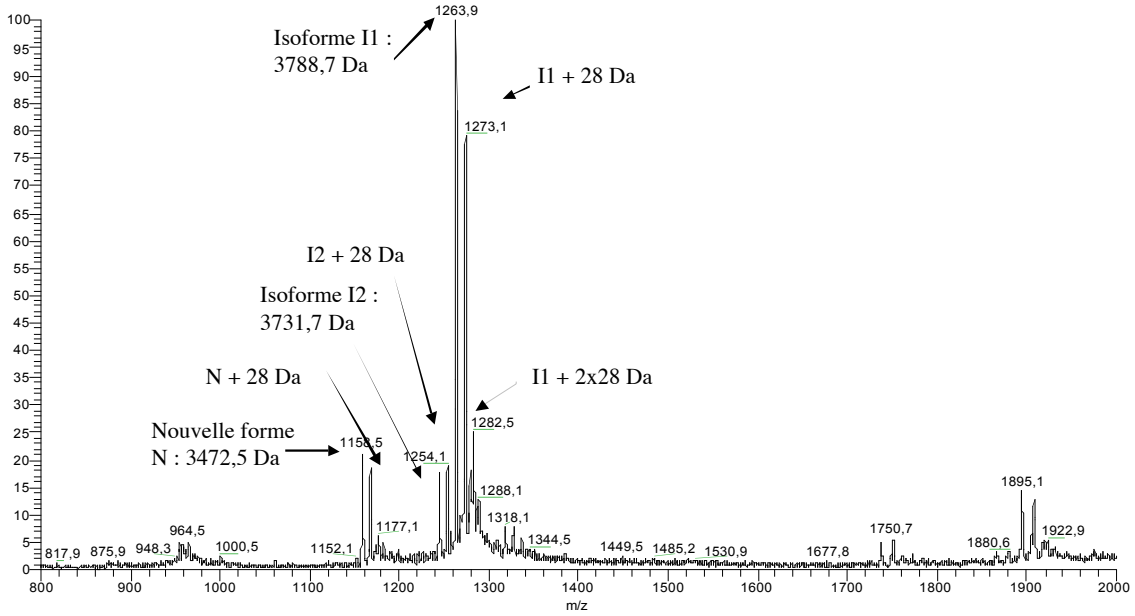


Figure 93 : spectre de masse non déconvolué d'isoformes de PA1b natif soumises à l'acide formique 50°C, 24h, composant un pic d'HPLC de temps de rétention identique aux isoformes natives.

(b) Clivage à l'acide chlorhydrique

Différentes conditions de clivage à l'HCl ont été testées. Le pH de 1,6 obtenu avec HCl, déjà validé sur d'autres protéines fusion produites au laboratoire, s'avère ici également satisfaisant. La concentration de TRX-DP-PA1b, testée entre 1 et 5 mg/mL, n'influe guère sur le rendement de clivage (évalué par SDS-PAGE). La température est par contre un élément crucial. A 37°C, la réaction est très lente et peu efficace. A 50°C, condition couramment utilisée au laboratoire, le clivage reste partiel, même au bout de 48h de réaction (Figure 94). Gavit et Better (2000) ont utilisé avec succès un clivage HCl (pH<2,6) à 85°C 3h, pour simultanément lyser les cellules et hydrolyser la liaison DP de leur protéine fusion, directement dans le fermenteur. De même, Wang et al. (2002) ont démontré l'efficacité d'un clivage HCl 25 mM, 2h à 85°C, sur une fusion TRX-neurotoxine de cobra, préalablement purifiée. Comme PA1b est, *a priori*, résistant à de telles conditions, nous avons testé sur TRX-DP-PA1b un clivage à 80°C, pH 1,6, dont le suivi sur 3h est présenté Figure 94. On observe dans ces conditions drastiques, une dégradation importante de la fusion et de la TRX ("trainée" protéique de faible masse apparente).

Le suivi des différents clivages de TRX-DP-PA1b, par acide formique ou acide chlorhydrique, a été réalisé par HPLC. Toutefois, l'apparition de nombreux pics de temps de rétention variés et la sortie simultanée de la TRX clivée et du reste de fusion ne permettent pas un suivi efficace. Nous avons donc préféré le suivi par SDS-PAGE avec révélation au Bleu de Coomassie. Ceci ne permet pas de visualiser rPA1b libéré de son partenaire TRX mais permet d'évaluer le rendement du clivage. Un gel révélé à l'argent montre trop de bandes "parasites", contaminants et leurs produits de clivages et/ou produits de clivages aspécifiques, qui empêchent une bonne visualisation des trois protéines d'intérêt (fusion, TRX, rPA1b) (cf. puit C Figure 97).

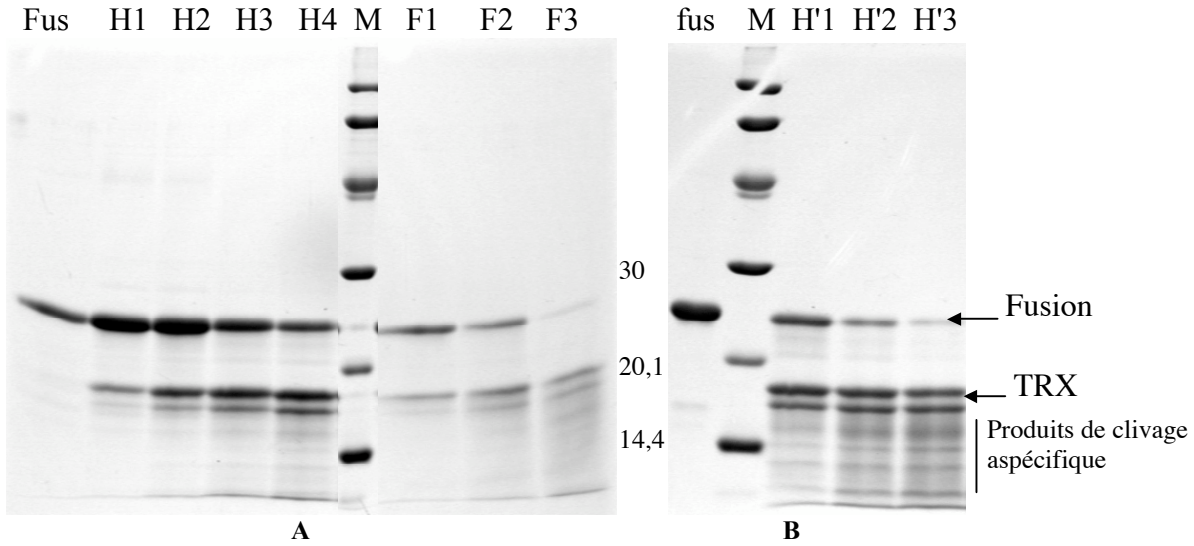


Figure 94 : suivi de clivages acides. (SDS-PAGE, Bleu de Coomassie). Fus : fusion non clivée
A : H : HCl 40mM 50°C, F: acide formique 70% 50°C; durée du clivage : 1 = 4h, 2 = 9h, 3 = 24h, 4 = 48h
B : H' : HCl 80°C; durée du clivage: 1 à 3 h.

De nombreuses publications font part de l'importance de l'accessibilité du site de clivage. L'acide formique tend à déstructurer les protéines (et les solubilise donc plus efficacement), ce que ne fait pas l'acide chlorhydrique. La conformation native pourrait protéger les protéines du clivage (Jauregui Adell and Marti, 1975). Des conditions dénaturantes (ajout de Guanidine-HCl 6 ou 7 M) sont donc reconnues pour améliorer le rendement de clivage dans certains cas, même s'il a été observé que la perte des structures secondaires des protéines pouvait inhiber le clivage (Ségalas et al., 1995). Toutefois, de telles conditions ne semblent pas influer dans le cas de notre protéine, et nous avons donc conservé les conditions non dénaturantes qui évitent l'étape de renaturation.
Pour éviter les problèmes de resolubilisation incomplète de TRX-DP-PA1b après lyophilisation, nous avons envisagé de rester en solution après chromatographie d'affinité. L'éluat est donc dialysé contre de l'eau puis le pH est rapidement amené à 1,6 par ajout d'HCl concentré. Une dialyse complémentaire sur la journée est effectuée contre HCl 40 mM puis l'échantillon est placé à 50°C 24h. La réaction est arrêtée par retour à température ambiante et ajout du mélange de clivage dans du tampon Tris-HCl 50 mM pH8,0. La remontée du pH doit s'effectuer doucement afin d'éviter la précipitation des protéines. Une dialyse et/ou un ajustement du tampon est effectué afin de procéder aux étapes de purification de rPA1b. Ces conditions, suivies d'une purification par affinité puis HPLC comme décrit ci-dessous, ont permis d'obtenir la meilleure purification, bien qu'imparfaite, de rPA1b après clivage acide (cf. spectre de masse correspondant Figure 102)

(c) Clivage à l'entérokinase

Au vu des difficultés (rendement, effets secondaires) des clivages acides, un clivage enzymatique a également été réalisé sur la fusion TRX-KA-PA1b (avec site enterokinase). La Figure 101 A montre le rendement obtenu avec l'enterokinase à 24°C. On constate que le clivage n'est pas total, toutefois, l'échantillon semble plus homogène (moins de "bandes parasites", formant une trainée lors de clivages acides). Une élévation de la température à 30°C ou 37°C augmente l'efficacité de clivage, mais conduit à la formation de deux

peptides correspondant à rPA1b tronqué en N-terminal de 2 ou 4 acides aminés (visibles en SM après purification par HPLC). Nous avons donc choisi les conditions à température ambiante pour produire une quantité de rPA1b suffisante pour les tests biologiques.

3) Purification de rPA1b après clivage acide

La mise au point de la purification de rPA1b en sortie de clivage acide a nécessité de nombreux essais.

(a) Propriétés de PA1b natif

Nous avons tout d'abord essayé d'exploiter les caractéristiques connues de la protéine native qui sont d'être thermostable, soluble dans le méthanol 60 % et l'acétone 80 %, ou encore de ne pas être retenue sur colonne d'échange d'anion à groupements DEAE à pH 9,0 (Delobel et al., 1998) ou pH 8,0 (Higgins et al., 1986). La solubilisation différentielle (par solvant ou choc thermique) à l'avantage d'être une technique simple et rapide. Malheureusement, aucune des trois solubilisations ne permet une réelle purification, même si un certain enrichissement du surnageant en peptides de masse proche de rPA1b est observé par solubilisation au MeOH60 (Figure 95). Le mélange de clivage est trop complexe, malgré une bonne purification de la fusion avant clivage, et rPA1b interagit probablement avec d'autres protéines présentes dans le mélange.

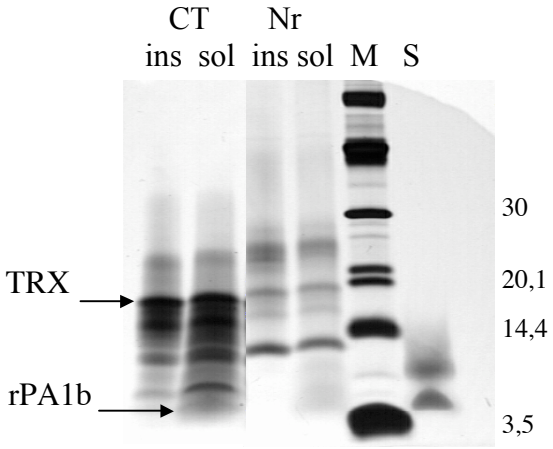


Figure 95 : Gel d'électrophorèse en tris-tricine (révélation à l'$AgNO_3$) : solubilisation différentielle en Méthanol 60% sur mélange de clivage en acide formique 50°C (CT) et sur fraction non retenue de chromatographie d'affinité après clivage (Nr). Ins : culot, sol : fraction soluble, M : marqueur, S: PA1b.

La purification des isoformes de PA1b à partir du pois fait appel à une IEC sur DEAE. Celle-ci a donc été tentée sur le mélange de clivage. Malheureusement, contrairement aux observations de Delobel et al. (1998) et d'Higgins et al. (1986), rPA1b semble retenu sur le gel (Figure 96). Les protéines du mélange initial sont très différentes entre l'extrait du pois et l'extrait de la bactérie et il se peut que des interactions moléculaires différentes expliquent ce résultat surprenant.

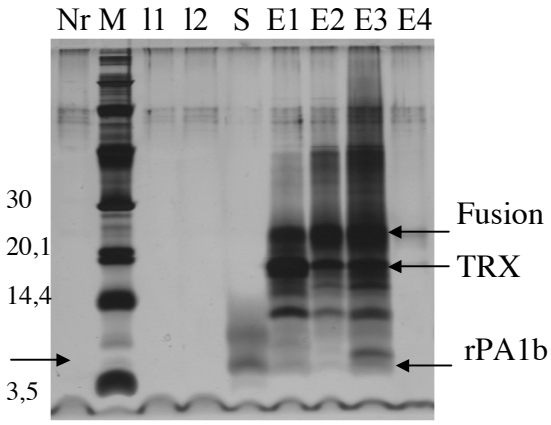


Figure 96 : Gel d'électrophorèse tris-tricine (révélation à l'$AgNO_3$): Fractions de chromatographie d'échange d'ion (sur DEAE) sur mélange de clivage HCl 50°C 24 h. M: marqueur, S : PA1b, Nr : fraction non retenue, l1 et 2: lavages, E1 et 2: 20%B, E3: 40% B, E4: 100%B.

Une étape d'HPLC en sortie de clivage n'est également pas suffisante pour purifier rPA1b, comme observé lors du suivi du clivage (spectre Figure 98 + gel Figure 99A).

(b) Chromatographie d'affinité post-clivage

Une deuxième étape d'affinité paraît donc nécessaire. En repassant le mélange de clivage sur colonne de nickel, rPA1b est récolté dans la fraction non retenue, tandis que la thiorédoxine et la fusion non clivée sont retenues par leur étiquette d'histidines. L'inconvénient de cette méthode est que rPA1b n'est pas concentré par passage sur la colonne. De plus, PA1b natif est connu pour intéragir avec la verrerie et les plastiques, en particulier lorsqu'il est pur. La chromatographie d'affinité ne peut s'effectuer avec du méthanol et il est à craindre qu'une partie des peptides recombinants clivés ne soit perdue. Malgré tout, l'analyse par électrophorèse des fractions obtenues lors de cette deuxième affinité montre, d'une part, la présence de peptides de masse attendue parmi les protéines non retenues et, d'autre part, la rétention de la thiorédoxine et du reste de fusion non clivée (Figure 97). Des essais d'affinité avec 10 % d'acétonitrile dans le tampon de fixation ont été réalisés dans le but de réduire les interactions du peptide et du support et/ou des autres protéines, sans amélioration visible (sur gel d'électrophorèse tris-tricine).

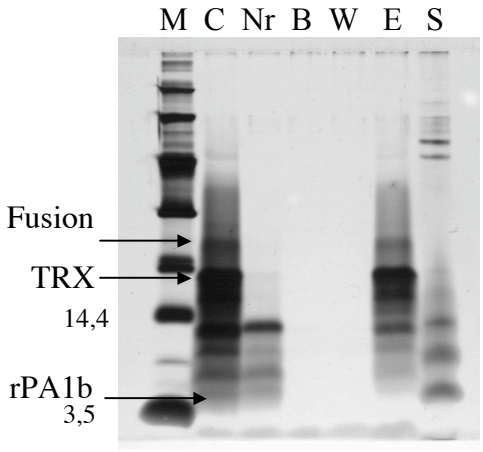


Figure 97 : Fractions de chromatographie d'affinité post-clivage (électrophorèse tris-tricine, $AgNO_3$). C: mélange de clivage (acide formique), Nr: fraction non retenue, B: lavage en tampon de fixation, W: lavage en tampon de lavage, E: éluat.

En sortie d'affinité, le travers contenant rPA1b contient toujours quelques contaminants comme des traces de fusion et de thioredoxine mais également d'autres peptides inconnus, en particulier, un peptide majeur d'environ 13 kDa. Une autre étape de purification s'impose donc.

(c) HPLC

Une solubilisation différentielle au MeOH60 permet un enrichissement en peptide recombinant (Figure 95) et l'injection directe en HPLC de la fraction soluble (profil Tv Figure 98). Le peptide rPA1b sort en HPLC sous forme d'un pic s'étalant de 15' à 19' (même profil en électrophorèse : Figure 99). La comparaison avec le profil d'isoformes de PA1b du pois (F1 Figure 98), qui forme un pic bien défini de temps de retention d'environ 21' dans les conditions utilisées ici, laisse présager d'une hétérogénéité du peptide recombinant, probablement mal replié. La comparaison avec le profil obtenu sur mélange de clivage sans affinité intermédiaire (CT Figure 98) montre combien cette étape est essentielle. De nombreux autres peptides sortent en même temps que rPA1b dans le cas du mélange brut, comme en témoignent les photos d'électrophorèse réalisées sur le massif "rPA1b" des deux profils. (Figure 99).

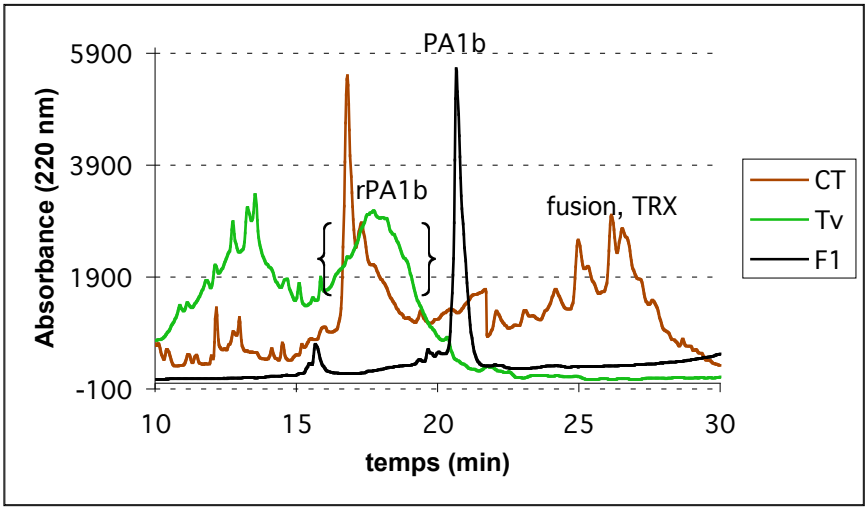


Figure 98 : Comparaison des chromatogrammes d'HPLC d'isoformes de PA1b natives (F1), d'un mélange de clivage à l'acide formique (CT) et du soluble MeOH60 de travers de chromatographie d'affinité post-clivage correspondant (Tv).

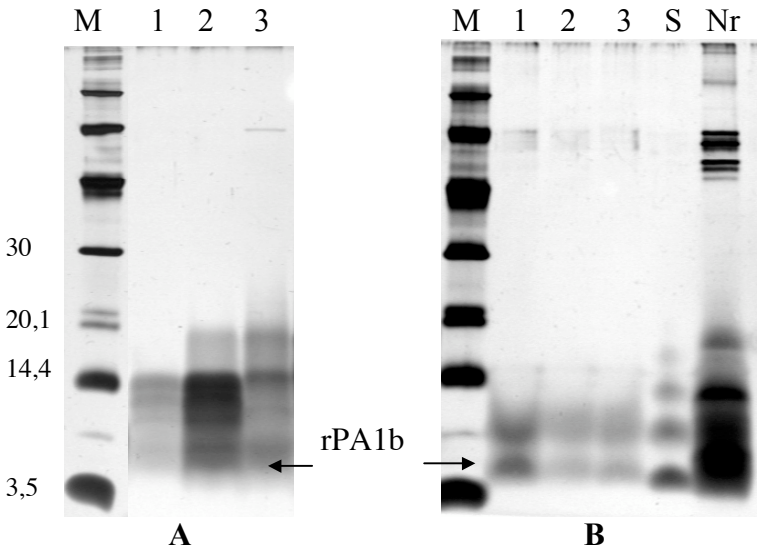


Figure 99 : Gel d'électrophorèse en Tris-Tricine (révélation au nitrate d'argent) sur les fractions d'HPLC correspondant à la zone rPA1b (Figure 98) provenant de :
A : mélange de clivage total
B : soluble MeOH60 de la fraction non retenue par chromatographie d'affinité post-clivage (Nr).

Un gradient isocratique est utilisé pour séparer les isoformes de PA1b à partir d'extraits de pois. Appliqué ici, il ne permet pas de séparer les pics du massif (données non montrées). Une réduction (peptide à 2mg/mL dans Tris HCl 0,1 M pH 8 ; DTT 20 mM : 10 min. à 100°C) suivie d'une alkylation (Tris HCl 0,1 M pH 8 ; Iodoacétamide 0,1M : 60 min. à température ambiante, obscurité) a été réalisée sur ce massif récolté après HPLC, et ensuite réinjecté en HPLC. Malgré un temps de rétention plus faible, on observe toujours une forte hétérogénéité, ce qui tend à montrer que les différents pics ne sont pas des formes plus ou moins oxydées au niveau des ponts disulfures. Les différences de masse de 16 Da observées entre différents pics d'un même massif d'HPLC indique la présence d'adduits de sodium (+ 22 Da) et/ou de modifications oxydatives (+ 16 Da), qui peuvent contribuer à l'hétérogénéité globale de l'échantillon.

L'HPLC constitue la dernière étape de purification du peptide recombinant. Afin de tester la présence de contaminants non visibles en électrophorèse et vérifier que les peptides observés sur gel correspondent bien à rPA1b, les différentes fractions du massif observé en HPLC ont été analysées en spectrométrie de masse, et montrent toutes un même profil (Figure 102).

4) Purification de rPA1b après clivage enzymatique

Le mélange de clivage enzymatique est beaucoup plus propre que celui provenant d'un clivage acide. L'absence de clivage aspécifique ou des contaminants par l'entérokinase rend le mélange plus facile à analyser par HPLC directement après solubilisation au MeOH60.

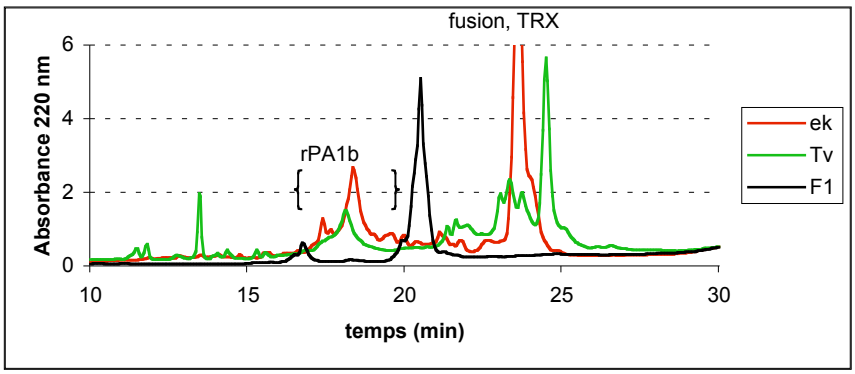


Figure 100 : chromatogrammes d'HPLC d'un soluble MeOH60 de clivage entérokinase (eK) et pour comparaison, d'isoformes de PA1b (F1) et d'un soluble MeOH60 de travers de chromatographie d'affinité post-clivage HCl (Tv).

Le même profil d'HPLC étalé est observé pour rPA1b obtenu par clivage enzymatique que par clivage acide (Figure 100). Une électrophorèse (tampons tris-tricine, révélation au nitrate d'argent) de ce massif analysé montre la pureté du peptide dans toutes les fractions le composant, mais aussi la variation de présence d'une forme dimèrisée (Figure 101). On constate aussi que rPA1b migre un peu moins loin que PA1b du pois. Enfin, on peut observer la présence de rPA1b dans des fractions de temps de rétention plus élevé, en association avec la TRX.

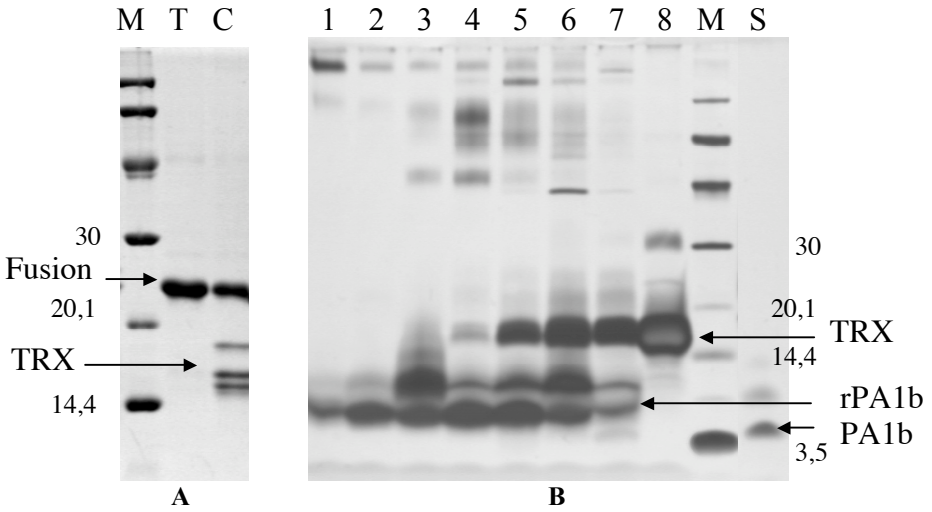


Figure 101 : gels d'électrophorèse sur clivages à l'entérokinase :
A : SDS-PAGE (Bleu de Coomassie) sur le mélange de clivage à l'entérokinase C (24°C), T : témoin fusion non clivée. **B**: Electrophorèse en tris-tricine (révélation au nitrate d'argent) sur les fractions HPLC du soluble MeOH60 (Figure 100) d'un clivé 30°C (clivage quasi total). Les fractions 1 à 4 représentent la "bosse" rPA1b (de 16 à 19'5 min de TR), la fraction 8 le pic à 22-23 min (TRX), les fractions 5 à 7 représentent des pics intermédiaires entre le massif rPA1b et le pic TRX. M: marqueur, S: PA1b**.**

5) Caractérisation de rPA1b

Les pics d'HPLC correspondant au peptide recombinant ont été analysés par spectrométrie de masse. Les échantillons provenant de clivage HCl montrent une certaine hétérogénéité, même si la masse majoritaire observée correspond bien au peptide PM-PA1b (3969,6 contre 3969,7 de masse moyenne oxydée théorique). Les échantillons de clivage enzymatique sont beaucoup plus nets et l'on y retrouve quasi uniquement la masse attendue pour AM-PA1b (3942,9 contre 3943,6 de masse moyenne oxydée théorique) (Figure 102).

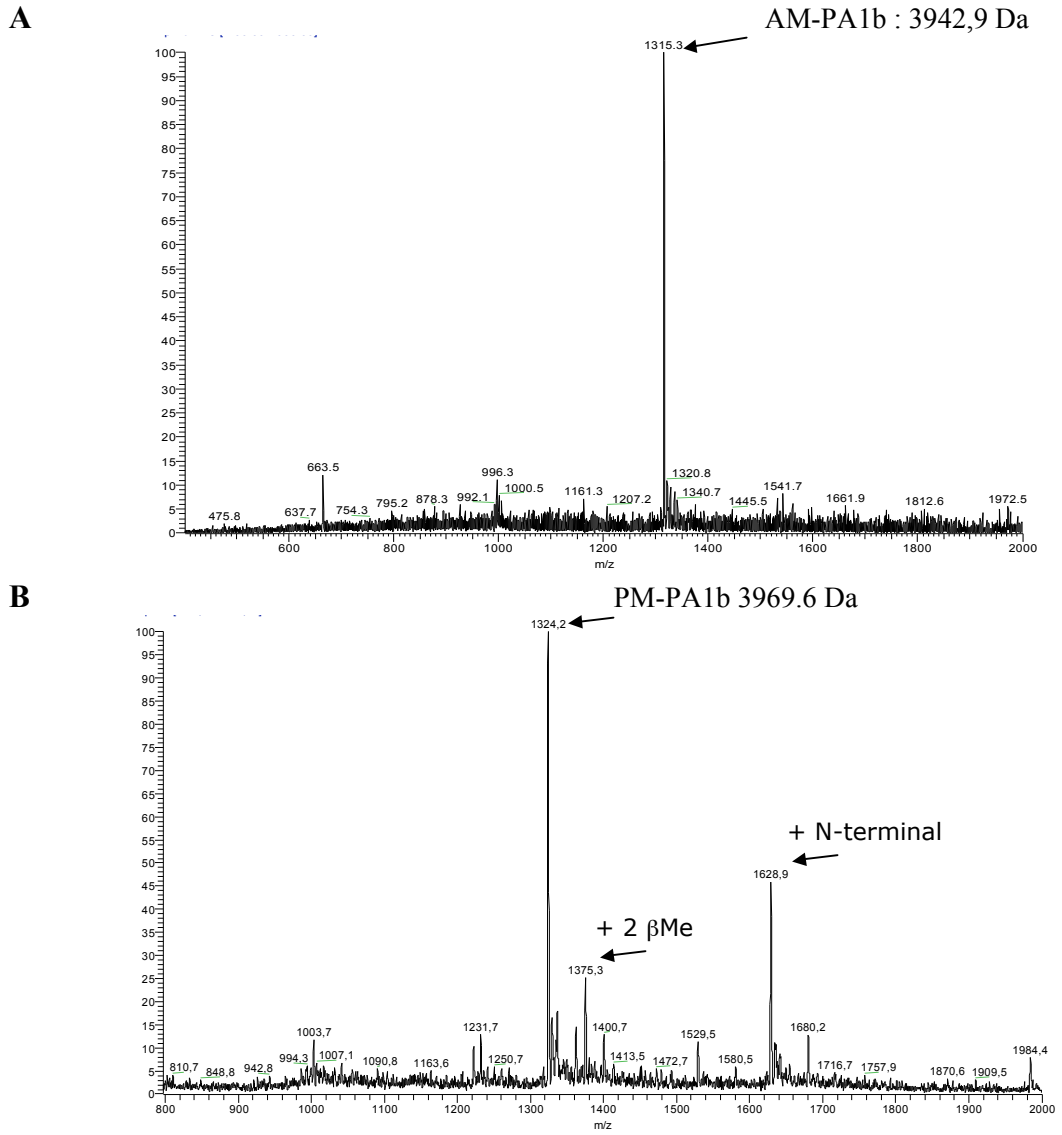


Figure 102 : spectres de masse de rPA1b.
A : provenant de clivage enzymatique (M théorique oxydée : 3943,6 Da), B: provenant de clivage HCl en restant soluble (M théorique oxydée : 3969,7 Da).

Dans le cas du produit de clivage acide, on observe également, et ce quelle que soit la fraction du massif de rPA1b, des pics de masses à 1375,3 et 1628,9 pouvant correspondre respectivement à des masses de 4122,9 Da et 4883,7 Da. Le fait que ces contaminants sortent comme PM-PA1b après les différentes étapes de purification laisse penser qu'ils sont probablement des formes modifiées du peptide recombinant. Une différence de 153,3 Da (4122,9-3969,6) pourrait correspondre à la fixation de deux groupements β-mercaptoéthanol (β-Me) sur deux cystéines, à la place d'un pont disulfure. Le β-Me est utilisé lors des différentes étapes de purification afin de protéger les éventuelles cystéines libres. La faible concentration utilisée ne devrait pas réduire les ponts formés, comme le confirme la masse d'AM-PA1b. Il se peut que l'étape du clivage acide rompe parfois un pont et que le β-Me se fixe sur les cystéines réduites lors de la chromatographie d'affinité suivant le clivage. La masse de 4883,7 Da pourrait correspondre à un peptide résultant du clivage aspécifique de la liaison D-S en amont de la liaison DP dans la thioredoxine. Le

peptide SPDLGTDDDPM-PA1b présente en effet une masse théorique oxydée de 4885,5 Da.
La massif observé en HPLC correspond donc bien au peptide d'intérêt. Dans les deux cas (clivage enzymatique ou acide), on retrouve un peptide de masse correspondant à la forme oxydée. Le pic HPLC, étalé et de temps de rétention plus faible que PA1b natif dans les deux cas, n'est donc certainement pas dû à la coexistence de formes partiellement oxydées. Il pourrait correspondre à un mauvais repliement, et fait donc craindre une absence d'activité biologique.
Le peptide AM-PA1b, plus pur que PM-PA1b, a été testé sur charançons à dose 1X[19], mais ne montre aucune activité. De même, en test d'inhibition de liaison, il ne montre pas de déplacement significatif du ligand marqué (Figure 103). Sa structure tridimensionnelle n'est certainement pas conforme à PA1b natif. Le très faible rendement de purification n'a pas permis de produire suffisamment de peptide pour effectuer des tests à doses plus élevées, ni des analyses de structure secondaire (dichroïsme circulaire ou spectrométrie infra-rouge).
Les conditions d'expression permettraient donc d'oxyder le peptide mais sous forme inactive, les ponts formés n'ayant peut-être pas la bonne connectivité, ce qui pourrait être vérifié par analyse des produits de digestion par des enzymes comme la trypsine. Toutefois, sur le peptide natif, cette approche est sans succès, aucune protéase classique, hormis le mélange enzymatique pronase E, ne libérant de produits de clivage (Gressent F., communication personnelle).

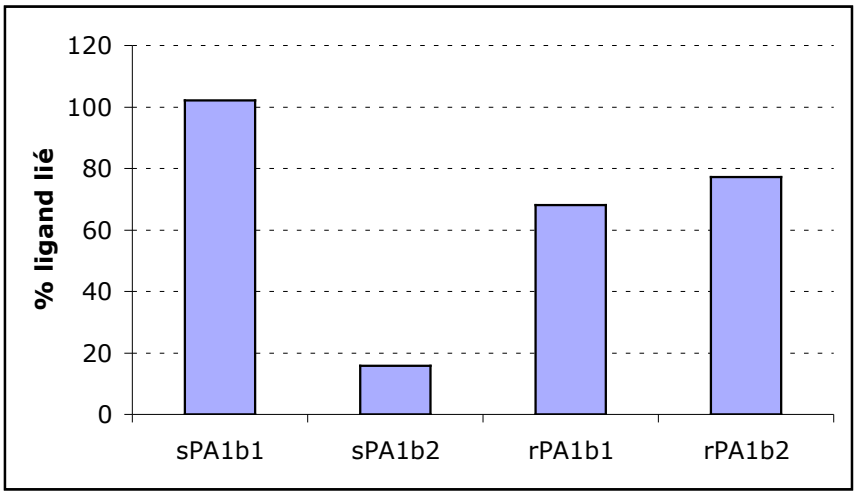


Figure 103 : Tests d'inhibition de liaison (Diol, 2003). Pourcentage de ligand lié pour sPA1b 1 et 2 : lots de PA1b synthétisé chimiquement (testés à 10^{-7} M), rPA1b 1 et 2 : lots de PA1b recombinant (clivage enzymatique, testés à 10^{-6} M).

[19] 1X : dose équivalente à 400μg de PA1b par gramme d'aliment, dose de TL_{50} de 4 jours.

IV. Discussion

L'objectif de la mise au point d'un système d'expression hétérologue de PA1b était la production de mutants pour une étude des relations structure-fonction. Toutefois, les difficultés rencontrées lors de cette mise au point et le temps limité imparti n'ont pas permis d'aborder la production des mutants. En effet, il convenait de vérifier avant tout que le peptide non muté était bien produit sous forme active dans le système choisi et développé, ce qui s'avère ne pas être le cas.
On constate à nouveau combien la structure de PA1b est source de difficultés. Le système choisi, *a priori* adapté pour un peptide à ponts disulfures comme PA1b, s'avère incapable de produire un peptide actif.
Toutefois, nous avons pu montrer que l'expression d'un gène spécifiant PA1b chez *E. coli* est possible, en fusion avec la TRX. Nous avons également déterminé des conditions de purification de cette protéine fusion efficaces. Le rendement d'expression obtenu est correct pour la souche utilisée. Des conditions de clivage enzymatique et acide ont été déterminées, ainsi que les méthodes de purification qui s'en suivent. Le peptide recombinant obtenu dans le cas d'un clivage enzymatique est pur, et dans les deux cas, apparaît oxydé. La souche choisie, mutée pour permettre la formation de ponts disulfures dans son cytoplasme, est donc efficace. Toutefois, il apparaît que dans les deux cas, le peptide est mal structuré et donc inactif. Ce problème de structure de rPA1b explique également les problèmes rencontrés lors de la purification du peptide.
Ce problème de formation d'une structure active de PA1b avait également été observé par Zhang (2001, résultats non publiés) lors d'essais de production hétérologue dans la levure *P. pastoris*. Le peptide était apparemment bien produit, bien qu'en faible quantité, mais ne présentait pas d'activité biologique. Comme le gène s'était avéré correctement transcrit, deux hypothèses avaient été évoquée pour expliquer cet échec : soit un mauvais processus de repliement et/ou d'oxydation du peptide recombinant chez la levure, soit une interférence de la forte hydrophobicité de PA1b avec le système de sécrétion. En absence de confirmation par spectrométrie de masse de la nature et de l'état du peptide produit, cet échec n'a pas pu être totalement interprété.
D'autre part, la structuration de PA1b après synthèse chimique a également posé des problèmes très importants de rendement de synthèse et de repliement. Le peptide produit par synthèse FMOC automatisée (Altergen, Strasbourg) n'a pu être oxydé en peptide structuré (contrôle en HPLC par rapport au peptide natif) par une stratégie standard d'oxydation par dilution à pH 8. Malgré l'expérience affichée du chimiste d'Altergen (Dr R. Amiri) dans la synthèse de peptides soufrés, l'oxydation en peptide actif n'a pu être obtenue dans des conditions satisfaisantes de rendement et de reproductibilité. Des activités biologiques partielles (toxicité) ont été obtenues sur des lots non conformes en HPLC, et un lot de faible importance (insuffisant pour un test de toxicité) s'est avéré contenir deux pics bien résolus en HPLC dont l'un possédait une activité biologique conforme en test de compétition de liaison avec ^{125}I-PA1b (Figure 103) (Diol, 2003).
Par ailleurs, l'équipe du Dr Hirano travaille sur le rôle de la leginsuline du soja (homologue de PA1b) dans la plante, et a montré sa liaison à un récepteur protéique des membranes plasmiques et parois cellulaires de soja. Cette équipe a réussit, d'une part, à produire dans *E. coli* la leginsuline sous forme active (*ie* liant le récepteur végétal) (Hanada et al., 2003), et d'autre part, à la synthétiser chimiquement (Yamazaki et al., 2003). Les conditions d'expression hétérologue, en souche BL21*trxB*(DE3) simple mutée (gène de la thioredoxine réductase) et fusion avec la thioredoxine, ne sont guère

différentes de notre système (souche Origami(DE3)pLysS, double mutée, et fusion TRX), et les différences (moins de protéases, une seule mutation) n'expliquent *a priori* pas leur succès et notre échec. Concernant les conditions de refolding après synthèse chimique, Yamazaki *et al.* (2003) décrivent un protocole plus complexe que les conditions employées couramment pour structurer un peptide à 3 ponts disulfures, témoignant ainsi de la particularité de la structure de cet homologue de PA1b. Le peptide obtenu possède les mêmes caractéristiques de liaison au récepteur que la molécule native du soja. Ces conditions de renaturation et oxydation ont donc été testées au laboratoire sur PA1b préalablement réduit (Rahbé Y., communication personnelle). On obtient alors de 30 à 50% de peptide restructuré (temps de rétention d'HPLC identique au natif). L'autre partie de peptide, résultant de ce protocole particulier, semble également oxydé mais sous des conformations non natives (temps de rétention inférieurs, rappelant ceux de rPA1b). Suite à cette validation du protocole de réoxydation, notre peptide recombinant a été réduit et soumis aux mêmes conditions d'oxydation. Malheureusement, très peu de peptide semble prendre la bonne conformation. En oxydant sur une à deux semaines, on augmente le rendement, en peptide conforme, qui reste toutefois très bas (moins de 1%). Une hypothèse pouvant expliquer cette différence, entre renaturation du peptide natif et du peptide recombinant, serait la configuration des prolines (Rahbé Y., communication personnelle). En effet, dans la structure de PA1b, la proline 13 est en configuration cis tandis que les quatres autres sont en trans (Jouvensal et al., 2003). Le peptide recombinant pourrait ne pas présenter cette configuration, et sa renaturation serait donc limitée par l'isomérisation lente des prolines. Des expériences sont prévues pour confirmer ou infirmer l'importance des prolines dans la structuration de PA1b.

Pour essayer de résoudre le problème de structure du peptide recombinant, sans passer par une dénaturation et une réoxydation *in vitro*, certaines améliorations du système d'expression peuvent être envisagée. La coexpression d'une protéine chaperonne et/ou d'une isomérase de ponts disulfures, facilitant le repliement du peptide, a permis dans de nombreux cas d'améliorer le rendement de production de molécules actives (Maskos et al., 2003 ; Nishihara et al., 1998 ; Schäffner et al., 2001). Dans ce sens, l'expression d'une séquence spécifiant tout PA1 est envisagée, avec ou sans fusion à la thiorédoxine. En effet, PA1b est suivi d'un polypeptide plus long, PA1a, dont la fonction est pour l'instant inconnue[20]. La conservation de la structure du gène (PA1b- peptide intermédiaire -PA1a) dans différentes espèces (cf. troisième chapitre) témoigne de l'importance de celle-ci. Un des rôles possible de PA1a pourrait être d'aider à la conformation de PA1b.

[20] PA1a na pas d'activité insecticide sur les charançons des céréales (Delobel B et Quillien, L communication personnelle).

Conclusions générales.

Conclusions générales et perspectives.

PA1b est le peptide responsable de la toxicité des graines du pois vis-à-vis des charançons des céréales (Delobel et al., 1998). Ce peptide et son gène avaient été caractérisés comme protéine de stockage du soufre chez *Pisum sativum* (Higgins et al., 1986), et un homologue en avait été décrit dans le soja (leginsuline), sans que leur activité entomotoxique ne soit soupçonnée (Watanabe et al., 1994). L'intérêt de PA1b dans la lutte contre les importants ravageurs des céréales stockées que sont les espèces du genre *Sitophilus*, nous ont amenés à étudier la distribution de ses homologues dans les espèces de la même famille que le pois, les Fabaceae.

1) Evolution de la famille des A1b au sein de la famille des Légumineuses.

Lors de notre travail de thèse, la caractérisation de nouveaux gènes par PCR génomique avec amorces dégénérées, associée à la détermination de la toxicité d'extraits de graines pour deux souches de *S. oryzae*, sensible et résistante à PA1b, nous a permis de montrer que les homologues de PA1b constituent une véritable **famille peptidique** répartie au sein de la famille des Légumineuses. Les A1b ne sont pas restreintes à la leginsuline du soja et aux quelques isoformes de PA1b du pois, connus jusqu'ici, mais sont présents dans de nombreuses espèces de Légumineuses.

- Des gènes homologues, provenant de 18 espèces et représentant deux clades de Papilionoideae (Galégoïdes et Phaséoloïdes), ont en effet été caractérisés au cours de ce travail. La limitation de la répartition de ces gènes est très probablement due à une limitation technique, les amorces dégénérées utilisées pour rechercher les gènes homologues ayant été déterminées d'après les séquences du pois et du soja, deux espèces appartenant à ces deux clades. Les gènes des A1b des autres clades ou sous-espèces de Fabaceae sont peut-être trop variants pour être reconnus par nos amorces.
- De plus, un différentiel de toxicité net et caractéristique de PA1b a été déterminé dans certaines espèces des trois sous-familles, dont les Mimosoideae *Albizia julibrissin* et *Acacia dealbata*, les Caesalpinioideae *Cercis siliquastrum* et *Tamarindus indica*, et les Papilionoideae *Abrus precatorius* (Aeschynoménoïde), *Laburnum anagyroides* (Génistoïde), *Styphnolobium japonicum* (Sophoroïde) ainsi que *Phaseolus vulgaris* (Phaséoloïde) et *Onobrychis viciifolia* (Galégoïde). Ceci indiquerait la présence d'homologues de PA1b au sein des trois sous-familles de Fabaceae. Ainsi, les A1b apparaissent largement présentes dans les graines de légumineuses, mais elles ne sont certainement pas les seuls composés de défense de ces graines (cf Introduction bibliographique). Notre criblage biologique nous a permis de caractériser une très forte variabilité d'activité entomotoxique dans les graines de nombreuses espèces de Fabaceae et ouvre donc la voie à des recherches d'autres molécules de défense, protéiques ou non.

La complexité du profil de représentativité et d'expression des A1b est toutefois à souligner. En effet, si Galégoïdes et Phaséoloïdes sont les seuls clades regroupant des

espèces ayant répondu à la PCR génomique, ils présentent également des espèces « négatives », dont celles de la tribu des Loteae. La position phylogénétique de cette tribu au sein des Galégoïdes est encore incertaine, se rapprochant des Trifolieae selon Chapill ou formant, avec les Robinieae et Abreae, un clade frère du clade des Galegeae-Trifolieae-Vicieae, selon les analyses du gène *rbc*L (cf. Figure 58). Il se peut que les Loteae, ainsi que les Abreae et les Robinieae, dont les espèces testées n'ont également pas répondu à la PCR, forment effectivement un clade, intermédiaire entre Phaséoloides et les autres Galégoïdes, où les gènes des A1b auraient été éteints ou seraient trop divergents pour avoir été reconnus par nos amorces. *Medicago truncatula*, *Glycine max* et *Lotus japonicus* sont l'objet d'analyses d'**EST** intenses (Alkharouf and Matthews, 2004 ; Bell et al., 2001 ; Kawasaki and Murakami, 2000 ; Lamblin et al., 2003). Si le transcrit de la leginsuline a bien été détecté dans les banques de soja (expression racinaire faible, et dans les graines, forte), aucun transcrit correspondant aux deux gènes de *M. truncatula* caractérisés dans ce travail n'a encore été décrit dans les banques publiées (fin 2003). De même, aucun transcrit d'homologue de PA1b n'a été observé chez *L. japonicus*. L'absence d'EST de graines en maturation chez ces deux espèces peut expliquer cette absence. Toutefois, concernant *L. japonicus*, qui appartient aux Loteae, l'absence d'EST d'homologues confirmerait nos résultats négatifs sur deux autres espèces de lotier (*L. corniculatus et L. tetragonolobus*). Ces deux espèces ne présentent guère de toxicité différentielle, excepté dans le résidu de *L. tetragonolobus*, faiblement toxique pour les charançons sensibles. L'hypothèse de variants trop divergents et d'expression spécifique dans les graines, et très faible ou nulle ailleurs, pourrait expliquer leur absence des banques d'EST. Chez *M. truncatula* et *L. japonicus*, la couverture EST croissante devrait répondre à cette question dans un avenir proche.

Si la présence de toxicité différentielle dans les extraits de graines est une indication de la possible présence d'A1b, elle n'en n'est pas une démonstration absolue. Afin de poursuivre et compléter cette étude des A1b, la purification de peptides homologues à partir d'extraits de graines montrant une toxicité différentielle totale, et provenant d'espèces phylogénétiquement éloignées des Galégoïdes et Phaséoloïdes s'avère nécessaire. Ceci permettra, d'une part, de délimiter l'étendue de la famille peptidique des A1b et, d'autre part, par le séquençage d'un petit nombre de ces peptides, de déterminer de nouvelles amorces dégénérées qui seront utilisées pour rechercher les gènes homologues d'espèces du même clade, élargissant ainsi la diversité de structures caractérisées.

2) Origine des A1b ?

Les A1b semblent limités aux Légumineuses (pas de gène homologue chezle riz, ni chez *A. thaliana*). L'absence d'homologie de séquence avec d'autres familles de peptides existantes ne permet pas de déterminer l'origine de ces peptides. Toutefois, plusieurs autres caractères peuvent les rapprocher de quelques familles peptidiques végétales de répartition taxonomique large.

- Au sein des groupes qui ont répondu à la PCR génomique, nous avons pu noter la conservation de la structure du gène de PA1b spécifiant une **préproprotéine**. Cette structure est celle des Albumines 2S, protéines de réserve des graines, comme la napine du colza. Les deux polypeptides codés par les gènes de ces 2S forment généralement un hétérodimère. Les sous-unités de 30-40 et 60-90 résidus sont liées par deux ponts disulfures (Shewry and Pandya, 1999). Dans les cas du pois, PA1a et PA1b ne sont pas associées par un tel pont, bien que toutes deux possèdent des cystéines. Les thionines sont également exprimées sous forme de préproprotéines

avec un peptide C-terminal acide qui neutraliserait la basicité du domaine thionine lors de la synthèse protéique (Bohlmann and Apel, 1991). PA1b est légèrement basique (Pi 7,8) et PA1a, bien qu'acide (Pi 4,9) n'aurait donc pas un rôle de neutralisateur. Si aucune interaction entre A1a et A1b n'a été démontrée jusqu'à présent (ni vis-à-vis de l'activité entomotoxique de PA1b, ni vis-à-vis de son rôle putatif d'hormone végétale), la conservation des A1a observée lors de cette étude témoigne de leur importance. Il est possible que PA1a intervienne lors du repliement de PA1b, en tant que chaperonne.

- La **structure du gène** à deux exons associant un intron variable inséré dans le peptide signal N-terminal conservé (avec une alanine conservée en P1 du site putatif de clivage), est un trait commun aux différents gènes caractérisés dans cette famille. Cette structure rappelle celle d'un groupe de peptides riches en cystéines de la moutarde et d'*Arabidopsis thaliana* (Ceci et al., 1995). Les gènes de ces inhibiteurs trypsiques sont groupés par quatre en un cadre serré dans le génome d'*Arabidopsis*. Ils sont exprimés dans les graines et également induits par blessure (De Leo et al., 2001) ou par une attaque de nématodes (Vercauteren et al., 2001), constituant ainsi un complexe de gènes défensifs.
- La **structure de la protéine** n'est connue que pour PA1b (Jouvensal et al., 2003) et la leginsuline (Yamazaki et al., 2003). Toutefois, l'homologie de séquence entre ces deux peptides et les séquences caractérisées dans notre travail est grande et la structure tridimensionnelle doit donc être conservée au sein des A1b. Cette structure, caractérisée par un feuillet β et une connectivité des ponts disulfures en cystine-knot, est proche de celles d'un certain nombre d'autres knottines, et en particulier des inhibiteurs EETI (inhibiteur trypsique d'*Ecballium*) et CPI (inhibiteur de carboxypeptidase de *Solanum*).

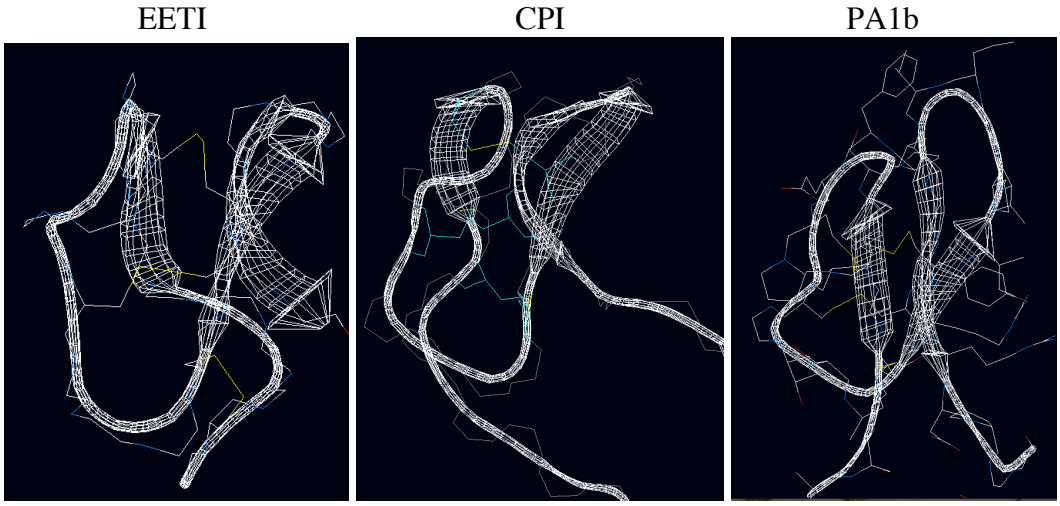


Figure 104 : Structures tridimensionnelles des inhibiteurs de protéase EETI (Chiche et al., 1989) et CPI (Rees and Lipscomb, 1982), et de PA1b (Jouvensal et al., 2003).

Ces différents caractères pourront être étudiés de façon plus systématique, afin de déterminer l'origine possible des A1b, et en particulier, chez les Légumineuses modèles génomiques.

3) Relations structure-fonction.

Nous avons également montré combien la structure de PA1b, exposant ses acides aminés hydrophobes, possédant une proline *cis* et trois ponts disulfures imbriqués en cystine-knot (Jouvensal et al., 2003), peut être source de difficultés. Certaines méthodes couramment employées pour étudier des familles protéiques posent ici problème. C'est le cas de la détection d'homologues par anticorps, ici de spécificité trop étroite (peptide peu immunogène), de la spectrométrie de masse (peptide soluble dans des fractions où d'autres composés non protéiques pourraient l'être également, qui inhiberaient l'ionisation). De même, l'obtention de cette structure particulière, que ce soit après synthèse chimique ou expression hétérologue en système eucaryote ou procaryote, apparaît complexe, bien que la réussite du groupe d'Hirano avec la leginsuline ouvre un perspective intéressante (Hanada et al., 2003 ; Yamasaki et al., 2003).

Grâce aux données des tests biologiques et des séquences peptidiques obtenues lors de notre travail, nous avons pu émettre des hypothèses quant aux acides aminés potentiellement importants pour l'activité. Malheureusement, les mutants prévus n'ont pu être produits faute de validation du système d'expression et purification. Le doigt hydrophobe de la leginsuline semble impliqué dans son interaction avec un récepteur potentiel des cellules végétales (Hanada et al., 2003). Un autre domaine du peptide pourrait donc être impliqué dans sa liaison au site de liaison du tube digestif des charançons. La boucle CSPFEMPPC, conservée dans nos séquences (car point d'ancrage d'une de nos amorce), est un candidat possible. Cette boucle, synthétisée chimiquement, ne présente pas d'activité de liaison au récepteur de PA1b (tests d'inhibition de liaison) (Diol, 2003). Toutefois sa conformation n'est pas native, le pont formé entre les deux cystéines de la boucle ne correspondant pas à un pont de la structure native de PA1b, et les configurations des prolines (1 cis et 2 trans) n'étant pas déterminées spécifiquement lors de la synthèse chimique.
Pour aborder les relations structure-activité des A1b, la production de PA1b actif est une étape à franchir. Plusieurs voies sont possibles qui aideront à la compréhension du mécanisme de structuration du peptide et ainsi permettront d'obtenir un peptide actif.

- Il se peut que le système d'expression de l'équipe japonaise, basé sur une souche d'*E. coli* B (contrairement à celle utilisée dans notre travail, dérivant d'*E. coli* K12) soit mieux adapté à la structuration de PA1b.
- L'expression de chaperonne dans le système bactérien utilisé ici est également à envisager, et en particulier, l'expression du gène de PA1 entier afin de tester le possible rôle de PA1a dans le repliement de PA1b.
- La synthèse chimique est également une voie possible de l'étude des relations structure-fonction, mais nécessite la validation de l'étape de renaturation du peptide. Le rôle putatif de chaperone de PA1a pourra également être testé *in vitro*, et tout d'abord sur PA1b de pois réduit.

Enfin, si les mutants ciblés doivent permettre de cerner l'importance d'un acide aminé donné dans l'activité, l'expression de variants entiers de PA1b, comme MtrA1b007 ou son potentiel homologue non maturé (MtrA1007), permettra de faire le lien entre les séquences et les activités mesurées dans les extraits de graines, et ainsi, de mieux comprendre la relation entre structure et activité des A1b.

Conclusions and Prospects (english version)

PA1b is the factor responsible for the toxicity of pea seeds to cereal weevils (Delobel et al., 1998). This peptide and its gene had been characterized previously as sulfure storage protein in *Pisum sativum* (Higgins et al., 1986), and a homologue (leginsulin) had been described in soybean (Watanabe et al., 1994). As PA1b is of great potential interest for the control of *Sitophilus* species, which are the major pests of stored cereal products, we studied the distribution of its homologues in selected species of the pea family, the Fabaceae.

1) Evolution of the A1b family among the Legume family.

Through our work, the characterization of new genes by genomic PCR with degenerated primers, associated with the determination of the toxicity of seed extracts to two *S. oryzae* strains, one susceptible and one resistant to PA1b, shows that PA1b homologues constitute a true peptidic family distributed among Leguminosae. A1b peptides are not restricted to leginsulin and the few pea isoforms known so far, but are present in many legumes species :

- Homologous genes from 18 species of two clades of Papilionoideae (Galegoid and Phaseoloid) have indeed been characterized in our work. The limited repartition of these genes is probably linked to a methodological limitation. Our degenerated primers, used to search for homologous genes, had been determined according to pea and soybean sequences, two species belonging to the two “positive” clades. A1b from other Papilionoideae clades or other subfamilies may be too divergent to be recognized by our primers.
- Furthermore, a clear differential toxicity, typical of PA1b, has been determined in some species from the three legumes subfamilies, including Mimosoideae such as *Albizia julibrissin* and *Acacia dealbata*, Caesalpinioideae such as *Cercis siliquastrum* and Papilionoideae such as *Abrus precatorius* (Aeschynomenoid), *Laburnum anagyroides* (Genistoid) or *Styphnolobium japonicum* (Sophoroid). This result indicates the presence of PA1b homologues among the three Fabaceae subfamilies. A1b seemed broadly present in legume seeds but they certainly are not the only defence compounds in these seeds (see Part 1). Our biological screening allowed us to characterise a very variable insecticidal activity in seeds of many Fabaceae species. This work opens the way to research on other defense compounds, either of proteic nature or not.

The distribution and expression pattern of A1b appears complex. Indeed, if Galegoid and Phaseoloid are the only clades including PCR-positive species, they also display “negative” species, in particular, within the tribe Loteae. The phylogenetic position of this tribe among Galegoid is still uncertain, being either sister of the Trifolieae according to Chapill, or forming together with Robinieae and Abreae, a brother clade to the Galegeae-Trifolieae-Vicieae, according to *rbc*L gene analyses (cf. Figure 58). Loteae may form with Robinieae and Abreae, from which the tested species did not respond to PCR either, a clade between Phaseoloid and all the other Galegoid, in which the A1b genes could lack

or have been lost, or could be too divergent to be detected by our primers. *Medicago truncatula, Glycine max* and *Lotus japonicus* are extensively studied by EST analyses (Alkharouf and Matthews, 2004 ; Bell et al., 2001 ; Kawasaki and Murakami, 2000 ; Lamblin et al., 2003). The transcript of leginsulin was detected in soybean libraries (weak root and high seed expression), but no transcript corresponding to the two *M. truncatula* genes characterized in our work has yet been described in published EST libraries (end 2003). Also, no homologous transcript has been observed in *L. japonicus*. The absence of EST from late maturing seeds in these two species may explain this absence. However, concerning *L. japonicus*, which belongs to Loteae, the absence of any EST homologue to PA1b might confirm our negative results on two other *Lotus* species (*L. tetragonolobus* and *L. corniculatus*). These two species did not display differential toxicity, but in *L. tetragonolobus,* the residue fraction was shown to be weakly toxic for susceptible weevils. The hypothesis of too divergent homologues with weak seed specific expression (and no expression in other organs), might explain their absence in EST libraries. In *M. truncatula* and *L. japonicus*, the increasing EST coverage will probably answer soon to this question.
If the presence of differential toxicity in seed extracts is an indication of a probable presence of A1b peptides, it does not demonstrate it properly. To continue and complete this study of the A1b, the purification of homologous peptides from seed extracts displaying a totally differential toxicity, and from species phylogenetically not related to Galegoid and Phaseoloid, appears necessary. It will then be possible to delimit the extent of the peptidic A1b family and, through sequencing of a few peptides, to determine new primers, which would be used to search homologous genes in the corresponding clades, enlarging the diversity of known structures from the A1b family.

2) The origins of A1b?

A1b appear to be limited to Fabaceae, as no homologues were detected in rice nor in *A. thaliana* or any other non legume species. As they display no sequence homology with other known peptide families, it is not possible to determine with confidence the origin of the family. However, several other features may affiliate A1b to some plant peptide families with wider taxonomic repartition.

- The structure of the gene encoding a preproprotein is conserved among the sequences obtained in our work. This structure is one feature of the 2S albumins seed storage protein family, such as the rape napin. Both peptides encoded by these 2S genes, generally form an heterodimer. The subunits, of 30-40 and 60-90 residues, are linked by two disulfide bridges (Shewry and Pandya, 1999). Concerning pea, PA1b and PA1a are not associated by such a bridge, even if both possess cysteins. Thionins are also expressed as preproprotein with a C-terminal acidic peptide, which may neutralize the basicity of the thionin domain during protein synthesis (Bohlmann and Apel, 1991). PA1b is slightly basic (Pi = 7,8) and PA1a, even if acidic (Pi = 4,9), may probably not display such a neutralizing function. No interaction between A1a and A1b has been demonstrated to date, neither for the insecticidal activity nor for the putative plant hormonal function of PA1b. Yet, the conservation of A1a, observed in our study, gives evidence of their importance. It is possible that PA1a intervenes in PA1b folding, as a chaperone.
- The structure of A1 genes with two exons and one variable intron, inserted in the conserved N-terminal signal peptide (with a conserved alanine at P1 position of the putative cleavage site), is a common feature of the genes characterized in this family. This structure is shared by a group of cystein-rich peptides from mustard

and *Arabidopsis thaliana* (Ceci et al., 1995). These trypsin inhibitor genes are clustered as a four genes tandem group in the *Arabidopsis thaliana* genome. They are expressed in seeds and also induced by wounding (De Leo et al., 2001) or by nematode attack (Vercauteren et al., 2001), constituting a defence genes complex.

- The structure of the protein is only known for PA1b (Jouvensal et al., 2003) and the leginsulin (Yamazaki et al., 2003). However, the sequence homology between these two peptides and the sequences we characterized is high, and the three-dimensional structure is probably conserved among A1b. This structure, with β-sheet and cystine-knot disulfide bridges connectivity, is similar to the structures of different other plant knottins, and in particular of the inhibitors EETI (trypsin inhibitor from *Ecballium*) and CPI (carboxypeptidase inhibitor from *Solanum*) (Figure 103).

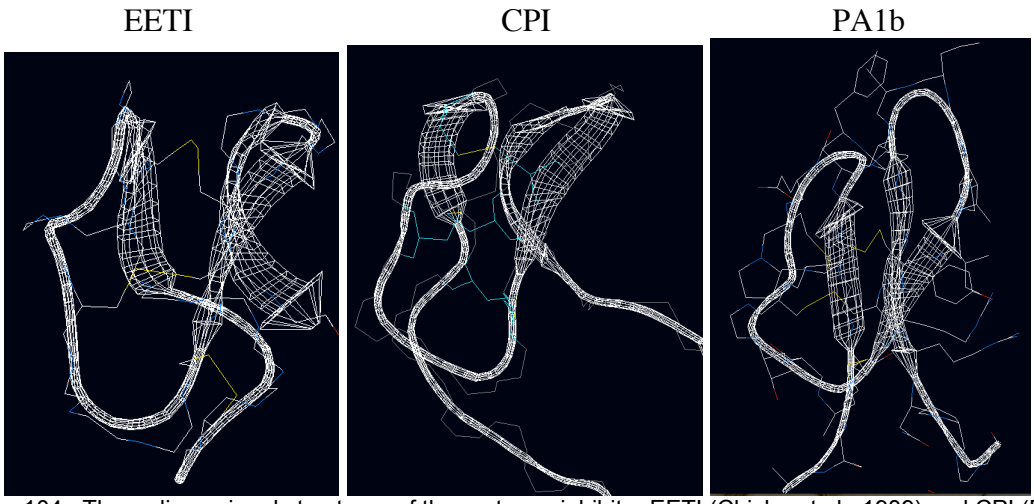


Figure 104 : Three-dimensional structures of the protease inhibitor EETI (Chiche et al., 1989) and CPI (Rees and Lipscomb, 1982), and of PA1b (Jouvensal et al., 2003).

These different characteristics could be studied more systematically, in order to determine the possible origin of A1b, and particularly in the three available legume genomic species.

3) Structure – function relationships.

In this study we showed how PA1b structure may be source of difficulties, exposing its hydrophobic amino acids, displaying a *cis*-proline and three disulfide bridges forming a cystine-knot (Jouvensal et al., 2003). Some techniques usually used to study protein families appear to be problematic in our study. It was the case of the immunodetection of homologous peptides (PA1b is weakly immunogenic, and its antibody has a very narrow specificity), of the mass spectrometry (PA1b is soluble in fractions in which other non protein compounds may be soluble, which could inhibit ionisation). In the same way, the folding of PA1b, after either chemical synthesis or when expressed in prokaryotic (as tested in this work) or eukaryotic system (Y. Zhang, 2001 not published), appears complex, even if the success of Hirano's group with leginsulin opens an interesting prospect (Hanada et al., 2003 ; Yamazaki et al., 2003).

Thanks to data from biological tests and peptidic sequences obtained in our work, we could raise some hypotheses concerning a few amino acids which might be involved in PA1b insecticidal activity. Still, the mutant peptides were not produced due to the lack of validation of the expression and purification system. The hydrophobic loop of PA1b seems to be involved in its interaction with a potential receptor in plant cells (Hanada et al., 2003). An other domain of the peptide might then be involved in its binding to its binding site in weevil guts. The CSPFEMPPC loop, conserved in all our sequences (as the anchor site of one primer), is a possible candidate. This loop, chemically synthesised, does not display any binding activity to the PA1b weevil receptor (ligand binding competition assay) (Diol, 2003). Nevertheless, its structure as a simple loop is not the native one, as the disulfide bridge formed between the two cysteines does not correspond to any native bond of the PA1b structure and as the prolines configuration (1 cis and 2 trans) was not specifically determined during the chemical synthesis.
To study the structure-function relationships of A1b, the production of an active synthetic PA1b is essential. Several ways are possible which would help to understand the peptide folding mechanism, and therefore would make it possible to obtain active peptide.

- It is possible that Hirano's group expression system, based on an *E. coli* B strain (contrary to the K12 derived strain we used), is more adapted to PA1b folding.
- Expressing a chaperone in the bacterial system we used is also a possibility, and particularly, expressing the whole gene PA1 to test the possible role of PA1a in assisting the folding of PA1b.
- Chemical synthesis is another way to study the structure-function relationships, but it needs a validation of the refolding step of the peptide, as was done with leginsulin (Hanada et al., 2003). The role of PA1a (putative chaperone) could also be tested *in vitro*, for example on pea-purified PA1b, chemically reduced *in vitro*.

Finally, if the pinpoint mutants would help to determine the importance of one specific amino acid in PA1b activity, expression of whole homologues, such as MtrA1b007 or its potential non-matured homologues (MtrA1007), would help to link sequences and toxicity detected in seed extracts, and therefore to better understand the relationships between the structure and the insecticidal activity of A1b.

List and legends of figures (english version)

Figure 1 : Major biochemical pathways involved in the synthesis of plant secondary metabolites (Panda and Khush, 1995). 28
Figure 2 : Alkaloid classification according to the heterocyclic ring system. 29
Figure 3 : Some plant defence alkaloids. 30
Figure 4 : DMDP, an azafuranose analogue of fructose. 30
Figure 5 : Nicotine, an insecticidal alkaloid used since the 17th century. 30
Figure 6 : A pyrrolizidine alkaloid sequestered and metabolized by insects. 31
Figure 7 : Enzymatic degradation of cyanogenic glucosides. 31
Figure 8 : Hydrolysis of glucosinolates. 33
Figure 9 : ODAP, a neurotoxic NPAA from the genus *Lathyrus* (Papilionoideae). 34
Figure 10 : NPAA with sulfure or selenium : cysteine analogues. 35
Figure 11 : Mimosine, a NPAA from Mimosoideae. 36
Figure 12 : Isoprene and its active form IPD. 37
Figure 13 : Biosynthetic pathway of terpenoids (Richter, 1993). 37
Figure 14 : Some monoterpenoids. 38
Figure 15 : Pyrethrin, an insecticidal monoterpenoid. 38
Figure 16 : Plant linear sesquiterpenoids with insect juvenile hormone activity. 38
Figure 17 : Some cyclic insecticidal sesquiterpenoids. 39
Figure 18 : Structures of plant anti-juvenile hormones. 39
Figure 19 : *Ent*-gibberellane structure 40
Figure 20 : Structure of the sterane core, shared by all triterpenoids, among which squalene 40
Figure 21 : Cucurbitacin B, a feeding deterrent triterpenoid. 41
Figure 22 : Azadirachtin, a promising feeding deterrent and insecticidal triterpenoid. 41
Figure 23 : Structures of some saponins. 42
Figure 24 : Plant ecdysone, a phytosteroid analogue to insect molting hormone. 43
Figure 25 : Some carotenoids. 44
Figure 26 : Shikimic (a) and cinnamic (b) acids 45
Figure 27 : Some simple phenolic compounds. 46
Figure 28 : Structures of the two categories of plant tannin. 46
Figure 29 : Coumarin. 48
Figure 30 : Categories of flavonoids, derived from the flavane skeleton. 48
Figure 31 : Quercetin, an insecticidal flavonol. 49
Figure 32 : Some plant defence isoflavonoids. 49
Figure 33 : Some anthocyanidins. 50
Figure 34 : Three-dimensional structure of concanavalin from *Canavalia brasiliensis* (Sanz-Aparicio et al., 1997). 53
Figure 35 : Three-dimensional structures of the A3 viscotoxin from mistletoe (thionin) (a) (Romagnoli et al., To be published), the radish Rs-AFP1 defensin (b) (Fant et al., 1998), the soybean leginsulin (c) (Yamazaki et al., 2003), the wheat nsLTP1 (non specific lipid transfer protein) (d) (Gincel et al., 1994). In yellow: disulfure bridges. 71

Figure 36: Sequences of PA1b isoform described by Higgins et al. (1986) (PsaA1b001), of the one isolated by Delobel et al. (1998) (PsaA1b005) and the soybean leginsulin (Watanabe et al., 1994). c is for conservative replacement, n, for non-conservative replacement. 80
Figure 37 : Sitophilus oryzae (adult) on a maize seed (picture : CS Gorsuch, Clemson University). 84
Figure 38 : PCR protocol for non degenerated primers (final concentrations). A : reaction mix, B : amplification cycles..... 90
Figure 39 : Principle of nested PCR. Primers of the first PCR (plain arrows) amplify a first fragment (white) on which primers of the second PCR (dotted arrows) hybridize, amplifying a shorter nested fragment (black)..... 92
Figure 40: Acute toxicity (at day 4) to weevils of seed flour from pea (pois), soybean (soja), bean (haricot) and *M. truncatula*. Dose response curves (% legume in wheat diet) on susceptible S strain..... 93
Figure 41 : Acute toxicity to susceptible S and resistant R weevils of seed flour from *M. truncatula*. Dose response curve. 94
Figure 42 : Representation of the different fractions in whole flour from pea, soybean, bean and *M. truncatula* (% w/w). 1 : pentane extracted fraction (lipids); 2 : Methanol 100%; 3 : Methanol 60%; 4 : H_2O pH5; 5 : H_2O pH8; 6 : residue; 7 : loss. 95
Figure 43 : Time-course of weevil mortality on MeOH60 fractions from the four legume species (dosed at 100% original seed equivalent, in wheat). All data from susceptible strain plus resistant strain for *M. truncatula*. 96
Figure 44: SDS-PAGE revealed with $AgNO_3$ on MeOH60 extracts from pea (1), soybean (2), bean (3) and *M. truncatula* (5). Molecular weight markers (M,m); PA1b (4). ... 99
Figure 45 : SDS-PAGE, revealed with $AgNO_3$ on aqueous fractions from *M. truncatula* (1 : pH 8; 2 : pH 5), and $H_2$05 fraction from *G. max* (3), *P. vulgaris* (4) and *P. sativum* (5). Molecular weight markers MWM (M, m); SRA1M60 (6); PA1b (7).... 99
Figure 46 : Immunodetection of A1bs with anti-PA1b antibody on : **A** : MeOH60 fractions from *M. truncatula* (1), *P. sativum* (2), *G. max* (3), *P. vulgaris* (4), PA1b (5) ; MWM (M, C : coloré). **B** : Aqueous fractions from *M. truncatula* (1 : pH 8; 2 : pH 5), H_2O5 de *G. max* (3), *P. vulgaris* (4) et *P. sativum* (5) ; SRA1-60 (6); PA1b (7) ; MWM (M, m) (cf gel Figure 44). 100
Figure 47: Competitive inhibition of ^{125}I-PA1b binding to its binding protein by MeOH60 extracts from the four test plant species plus wheat control. X-axis : dose of extract (total meal equivalent, TME) present in binding wells. ● *P. sativum*, ■ *P. vulgaris*, ○ *G. max*, ▲ *M. truncatula*, ∉ *T. aestivum*. 101
Figure 48 : Results of PCR with For1 et Rev 3 primers : amplification of a 130 bp fragment corresponding to a portion of PA1b gene. 1,2,3 : *P. vulgaris* ; 4,5,6 : *M. truncatula* ; 7,8,9 : *G. max* ; 10,11,12 : *P. sativum* ; 13 : negative control ; M : MWM 1 Kb. 1,4,7,10 : 500 ng of genomic DNA ; 2,5,8,11 : 50 ng; 3,6,9,12 : 10 ng..... 102
Figure 49 : Results of PCR with For1 et Rev1 primers : amplification of a 220 bp fragment corresponding to a portion of PA1b plus PA1a gene. 1,2,3 : *P. sativum* ; 4,5,6 : *P. vulgaris* ; 7,8,9 : *M. truncatula* ; 10,11,12 : *G. max* ; 13 : negative control ; M : MWM. 1,4,7,10 : 100 ng of DNA; 2,5,8,11 : 50 ng; 3,6,9,12 : 10 ng. 102
Figure 50 : Results of nested PCR (GenomeWalker protocol) on the digested DNA libraries of *P. vulgaris* (sequence PvuA1b001) (6,7,8,9) and *M. truncatula* (sequence MtrA1b006) (1,2,3,4). *Dra*I (1,6); *Eco*RV (2,7); *Stu*I (3,8); *Pvu*II (4,9) libraries ; MWM 1 kb (M); negative controls (5, 10, 12); positive control (11). 103

Figure 51 : CC1 to 7 : Segments of PA1b delimited by the cysteines 1 to 6..... 104
Figure 52 : Sequences alignment of reference PA1b peptides and cloned genes from the four studied species. First column indicates peptide identification (Swissprot, EST or EMBL accession, gene name) and source plant species : Psa : *Pisum sativum*, Pvu : *Phaseolus vulgaris*, Gma : *Glycine max*, Mtr : *Medicago truncatula*, Van : *Vigna angularis*, Vra : *Vigna radiata*. First line indicates the pea-like processing of the prepropeptide (Higgins et al., 1986). Conserved cysteines are in red, other conserved amino acids are in blue..... 106
Figure 53 : Separation of six groups of peptides by isocratic HPLC on pea MeOH60A80 fraction (soluble part of MeOH60 fraction in 80% acetone). 107
Figure 54 : Legume diversity 116
Figure 55 : Pods of: left: *Teline monspessulanus* (Papilionoideae) and right: *Cercis siliquastrum* (Caesalpinioideae). 116
Figure 56 : Flowers of Fabaceae: left: *Onobrychis viciifolia* (Papilionoideae), middle: *Albizia julibrissin* (Mimosoideae) and right: *Caesalpinia gilliesii* (Caesalpinioideae) 117
Figure 57 : Putative phylogenetic relationships between tribes within Papilionoideae (Polhill, 1981)..... 117
Figure 58: Simplified phylogenetic trees of Fabaceae (particularly Papilionoideae) according to : left : Chapill (http://www.botany.uwa.edu.au/systematics/paps.html). right : Wink and Mohamed (2003) : molecular phylogeny (*rbc*L gene). 120
Figure 59 : Results of PCR amplification with For1 and Rev1 primers on genomic DNA from Ad : *Acacia dealbata*, Af : *Amorpha fruticosa*, Tf : *Trigonella foenum-graecum* (at 10 and ’ : 2 ng of DNA), at different annealing temperatures (Ta) : 1: 44°C; 2: 45°C; 3: 46°C; 4: 47°C. 124
Figure 60 : Results of PCR amplification with For1 and Rev1 primers on genomic DNA from different species at Ta=40°C. M : MWM; + : positive control (pea) ; 1 : *Albizia julibrissin* ; 2 : *Arachis hypogaea* (with 10 and ’ : 20 ng of DNA) ; 3 : *Dalbergia purpurescens* ; 4 : *Trigonella foenum-graecum* ; 5 : *Intsia bijuga* ; 6 : *Lotus corniculatus* ; 7 : *Lupinus albus* ; 8 : *Robinia pseudoacacia* ; 9 : *Schotia afra* ; 10 : *Sesbania sesban* ; 11 : *Vigna unguiculata* ; - : negative control. 124
Figure 61 : PCR with For1-Rev5 (5) or For1-Rev6 (6) primers : validation of these new degenerated reverse primers on pea (P) and soybean (S). - : negative control ; + : positive control (For1-Rev1 on pea)..... 125
Figure 62: Gene walking nested PCR with Rev5 (R5) primer on the *M. truncatula Dra*I test library. - : negative control ; + : positive control (specific primers). 125
Figure 63 : Gene walking PCRs on *L. albus*. Arrow indicates a cloned fragment in *Pvu*II library. M: MWM, +: positive control, -: negative control, D: *Dra*I, P: *Pvu*II, E: *Eco*RV libraries. 126
Figure 64 : PCRs on *L. angustifolius* DNA : left : following Ilgoutz et al. (1997) protocol : arrows indicates two non expected products (no fragment at expected size 280 bp). Right : with ADH gene specific primers. M: MWM, -: negative control, L : DNA from lupine, P : DNA from pea, mix : mixed DNA from pea and lupine. 127
Figure 65 : Northern blot with a labelled pea PA1 cDNA. M : Mimosoideae, C : Caesalpinioideae, P : Papilionoideae. 128
Figure 66 : Gene walking. **A** : nested PCR (Ta=67°C) on the three libraries *Dra*I, *Eco*RV, *Pvu*II of Ca: *Caragana arborescens*, Ao: *Alysicarpus ovalifolius*, Bb: *Bituminaria bituminosa*, Vs: *Vigna subterranea*, La: *Lupinus angustifolius*, Vu: *Vigna*

unguiculata, Ce: *Canavalia ensiformis*. **B** : nested PCR (Touch Down 65°C) on the libraries *Eco*RV from Vu and Ca, *Pvu*II from Bb, *Dra*I from Vs and La, and the three libraries from Ce. + , - : positive and negative controls. M, m : MWM. 129
Figure 67 : Protein sequences alignment from our work (with sequences of the primers) and from literature or EST libraries. Only the amino acids different from M13709 are indicated : - indicates the same residue as M13709 sequence. First line : PA1b segments. Second line : pea-like maturation of the preproprotein. Plant species are : Amo : *Astragalus monspessulanus*, Aov : *Alysicarpus ovalifolius*, Bbi : *Bituminaria bituminosa*, Car : *Caragana arborescens*, Cbr : *Canavalia brasiliensis*, Gma : *Glycine max*, Lan : *Lupinus angustifolius*, Lca : *Lonchocarpus capassa*, Lco : *Lotus corniculatus*, Lla : *Lathyrus latifolius*, Mal : *Melilotus albus*, Mse : *Mundulea sericea*, Mtr : *Medicago truncatula*, Ovi : *Onobrychis viciifolia*, Psa : *Pisum sativum*, Pvu : *Phaseolus vulgaris*, Tfo: *Trifolium foenum-graecum*, Van : *Vigna angularis*, Vhi : *Vicia hirsuta*, Vra : *Vigna radiata*, Vsu : *Vigna subterranea*, Vun : *Vigna unguiculata*. 131
Figure 68 : Molecular phylogeny on the “partial” sequences (For1-Rev1 PCR fragments) (BioNJ algorithm and Galtier & Gouy distance). green : sequences from Phaseoloids species; crimson : from Galegoids species. Names of the sequences are indicated in Figure 67. 132
Figure 69 : Principal component analysis (Diol, 2003) red : Galegoids, green : Phaseoloids, yellow : Genistoids, blue : Aeschynomenoids. Variables : S : susceptible strain, R : resistant strain, 0 to 4 : MeOH to residue fractions**.** 139
Figure 70 : Distribution of the 3 toxicity classes (C1 low toxicity to C3 high toxicity) among Fabaceae. Phylogenetic relationships between species are indicated in the first column. 140
Figure 71 : Binding competition assay : percentage of labelled ligand bound to its binding site for two concentrations of MeOH60 extracts competitor from different species. Concentrations are expressed in mg of TME (total meal equivalent) per binding well. Wheat is a negative control. 143
Figure 72 : Tris-Tricine Gel Electrophoresis (revealed with AgNO3) on MeOH60A80 fractions from some selected species. S: SAR1M60, P: PA1b, M: MWM, Aov : *Alysicarpus ovalifolius*, Bbi : *Bituminaria bituminosa*, Bsp : *Bolusanthus speciosus*, Cer : *Crotalaria eremaea*, Lca : *Lonchocarpus capassa*, Mal : *Melilotus albus,* Ovi : *Onobrychis viciifolia*, Tfo : *Trigonella foenum-graecum*, Tre : *Templetonia retusa*, Vsu : *Vigna subterranea*, Vun : *V. unguiculata*. Arrows indicate PA1b position. .. 144
Figure 73 : Diagram of an expression vector. 156
Figure 74 : Diagram of an expression cassette. 157
Figure 75 : Diagram of *Escherichia coli* membranes. (From Microbiology, Prescott *et al.*, 1995). 159
Figure 76 : Inclusion bodies in *E. coli* cells. 160
Figure 77: Diagram of the bioreactor. 167
Figure 78 : General strategy to polymerize synthetic genes. In our work : A= *Nco*I, B= *Mst*I, B'= *Eco*RV, C= *Hin*DIII. 171
Figure 79 : Restriction maps of the inserts used to obtain the synthetic gene encoding M-PA1b (PA1b with a N-terminal methionin) and diagram of the pPA1b plasmid..... 172
Figure 80 : Expression vector pET-DP-PA1b. 174
Figure 81 : left: expression of the fusion protein TRX-DP-PA1b in Origami cells from time of induction (T0) to 3h after induction (T3) (SDS-PAGE, Coomassie blue

coloration) ; right: western blotting with antiPA1b antibody : detection of the fusion protein TRX-DP-PA1b. S: PA1b; M: MWM..... 174
Figure 82 : Total (tot) and soluble (sol) proteins of *E. coli* after 3h expression (SDS-PAGE, Coomassie blue). 175
Figure 83 : Purification of the fusion protein TRX-DP-PA1b by affinity chromatography using stepwise increase in imidazole concentration (SDS-PAGE, Coomassie blue). Sol : soluble protein (from cytoplasm), Nr : flow-through, B : wash with binding buffer, W : wash with washing buffer, E : elute, M : MWM. 176
Figure 84 : Purification of fusion protein by "stepwise" affinity chromatography in 6M urea buffer (SDS-PAGE , Coomassie blue). Tot: total proteins, Nr: flow-through, B and W: washes, E: elute, M: MWM. 177
Figure 85 : Anion exchange chromatography on eluted fraction of "stepwise" affinity chromatography. 1 to 4 : elution peaks analyzed by electrophoresis (Figure 86). Around 300 mg of proteins have been loaded on the column..... 177
Figure 86 : Purification of fusion protein by post-affinity ion exchange chromatography (SDS-PAGE, Coomassie blue). E: affinity elute, Nr: IEC flow-through, 1 to 4 : elution peaks at 20, 40 et 100% of buffer 2 (Figure 85). 178
Figure 87 : Purification of fusion protein TRX-DP-PA1b by post-IEC affinity chromatography (SDS-PAGE, Coomassie blue). sol: soluble proteins from cytoplasm, IEC : IEC elute, E: elute of affinity post-IEC, M : MWM..... 178
Figure 88 : Affinity chromatography (on Ni^{2+} column) with elution by a linear gradient in imidazole concentration (optimized conditions). Fractions Nr to E4 were analysed using SDS-PAGE (Figure 89). Nr: flow-through, B and W: washes (10; 20; 40 mM imidazole), E1 to E4: elution fractions 179
Figure 89 : SDS-PAGE (Coomassie blue) on the different fractions from the "linear gradient" affinity chromatography (Figure 88). Sol: soluble proteins from cytoplasm, Nr: flow-through, B and W: washes (10; 20; 40 mM imidazole), E1 to E4: elution fractions..... 179
Figure 90 : HPLC profile of the purified fusion protein TRX-DP-PA1b..... 180
Figure 91 : ESI-MS spectrum of the purified fusion protein. Observed molecular mass : 20 650,4 Da (theoretical reduced mass : 20658,5 Da)..... 181
Figure 92 : Mass spectrum of purified rPA1b obtained by formic acid cleavage..... 182
Figure 93 : Mass spectrum of pea PA1b isoforms submitted to 70% formic acid, 24h at 50°C, and forming a peak by HPLC with retention time identical to that of native isoforms..... 183
Figure 94 : Acidic cleavages (SDS-PAGE, BC). Fus : fusion protein **A** : H : with 40 mM HCl, 50°C, or F: with 70% formic acid, 50°C; cleavage duration : 1 = 4h, 2 = 9h, 3 = 24h, 4 = 48h **B** : H' : with HCl at 80°C; cleavage duration : 1 to 3 h..... 184
Figure 95 : Tris-tricine electrophoresis gel (revealed with $AgNO_3$): Differential solubilisation in methanol 60% of the mixture obtained after formic acid cleavage (CT) and of the flow-through of a post-cleavage affinity chromatography (Nr). Ins : insoluble fraction, sol : soluble fraction, M : MWM, S: PA1b. 185
Figure 96: Tris-tricine electrophoresis gel (revealed with $AgNO_3$) on fractions from an anion exchange chromatography on a mixture obtained after cleavage with 40 mM HCl 50°C 24 h. M: MWM, S: PA1b,Nr: flow-through, l1 and 2: washes, E1 and 2: 20% buffer B, E3: 40% B, E4: 100%B..... 186
Figure 97 : Tris-tricine electrophoresis gel (revealed with $AgNO_3$): purification of the recombinant peptide by post-cleavage affinity chromatography (on Ni^{2+} column).

C: cleavage mix (formic acid), Tv: flow-through, B: wash with binding buffer, W: wash with washing buffer, E: elute..... 187
Figure 98 : HPLC profiles of native PA1b isoforms (F1), a formic acid cleavage mix (CT) and the methanol 60% soluble fraction of the corresponding flow-through obtained by post-cleavage affinity chromatography (Tv). 188
Figure 99 : Tris-tricine electrophoresis gel (revealed with $AgNO_3$) on different HPLC fractions corresponding to the rPA1b peak (Figure 97) from : **A** : cleavage mix **B** : MeOH60 soluble fraction from affinity flow-through (Nr). 188
Figure 100 : HPLC profiles of a MeOH60 soluble fraction from an enterokinase cleavage (eK), of pea PA1b isoforms (F1) and of a MeOH60 soluble flow-through from post HCl cleavage affinity (Tv). 189
Figure 101 : Electrophoresis gels on enterokinase cleavages : **A** : SDS-PAGE on a cleavage mix (C) (cleavage at 24°C) and the control fusion protein (T) **B**: TTGE on HPLC fractions from MeOH60 soluble fraction of an enzymatic cleavage mix (Figure 100) (cleavage at 30°C). Fractions 1 to 4 are part of the large rPA1b peak (from 16 to 19'5 min of retention time), fraction 8 is for the 22-23 min peak (TRX). Fractions 5 to 7 stands for intermediate pics. M: marker, S: PA1b**.**..... 190
Figure 102 : Mass spectra of rPA1b. **A** : from enzymatic cleavage (theorical oxidised mass : 3943,6 Da), **B**: from a soluble HCl cleavage (theorical oxidised mass : 3969,7 Da). 191
Figure 103 : Ligand binding inhibition assays (Diol, 2003). Percentage of bound labelled pea PA1b for competition by : sPA1b 1 and 2 : two chemically synthesized PA1b fractions (tested at 10^{-7} M), rPA1b 1 and 2 : two recombinant PA1b fractions from enzymatic cleavage (tested at 10^{-6} M)..... 192
Figure 104 : Three-dimensional structures of the protease inhibitor EETI (Chiche et al., 1985) and CPI (Rees and Liscomp, 1982), and of PA1b (Jouvensal et al., 2003). .. 199

List and legends of tables (english version)

Table 1 : Major classes of secondary metabolites involved in plant resistance to insects. Nd : not determined (Hartley, 2001 ; Panda and Khush, 1995). 27
Table 2 : Toxic nonprotein amino acids (NPAA) and their protein amino acids analogues. 34
Table 3 : Toxicity to mammals and spectrum of entomotoxic activity of plant proteins and *Bacillus thuringiensis* (*Bt*) endotoxins..... 52
Table 4: Some transgenic plants expressing lectins. GNA : snowdrop lectin, ConA : concanavalin A, Plec : pea lectin, WGA : wheat lectin BCH : bean chitinase, CpTI : cowpea trypsin inhibitor..... 55
Table 5 : Some transgenic plants expressing protease inhibitors. (C) : Coleoptera, (H) : Homoptera, other target insects are Lepidoptera..... 61
Table 6: Toxicity of seed fractions to susceptible and resistant to PA1b weevil strains. Median survival time (+-SE) on diets incorporating 100% seed equivalent of each fraction. Test of actuarial survival analysis were used to compare LT50 between both strains on each fraction from *M. truncatula*. Significative differences were detected only in MeOH and H_2O5 fractions (p-value of Breslow-Gehan-Wilcoxon <0,0001) 96
Table 7 : Mass spectrometry peptide matches on selected solvent extracts from the four plant species analysed. **a** : expressed as MH+ (from nucleic sequences obtained, average mass minus 6 for full cysteine bridge bonding), when a hit was discovered with a truncated sequence, the theoretical mass of the truncated sequence is reported (C-1 : c-terminal truncation, based on a canonical pea-type pre- and propeptide processing). **b** : matching was scored positive when a precision better than 0,05% was obtained with the theoretical mass (linear detection mode). **c** : we only reported peaks with m/z>2000, and of intensity exceeding 10% of the highest observed peak (other significant but minor peaks are listed in italics). **d** : expressed as % of matched peak intensities *vs* total intensity of major peaks (as defined in c)..... 109
Table 8 : Lethal time 50 of weevils from two non isogenic strains (Bénin and Chine) and two isogenic strains (WAA42, ISO3R) ; p-value (survival test Breslow-Gehan-Wilcoxon) are indicated in the last column to compare LT50 of the two susceptible or the two resistant strains on each fraction.. 141
Table 9 : Enzymes and chemical products used to cleave fusion proteins. 163
Table 10 : *E. coli* strains and plasmids used. 164

Bibliographie.

Références Bibliographiques

Abad, L. R., D'Urzo, M. P., Lin, D., Narasimhan, M. L., Renveni, M., Zhu, J. K., Niu, X., Singh, N. K., Hasegawa, P. M. and Bressan, R. A. Antifungal activity of tobacco osmotin has specificity and involves plasma membrane permeabilization. *Plant Science* (1996) **vol. 118**, pp. 11-23.

Abe, J. I., Sidenius, U. and Svensson, B. Arginine is essential for the alpha-amylase inhibitory activity of the alpha-amylase/subtilisin inhibitor (BASI) from barley seeds. *Biochem. J.* (1993) **vol. 293**, pp. 151-155.

Abe, K., Emori, Y., Kondo, H., Suzuki, K. and Arai, S. Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds *J. Biol. Chem.* (1987) **vol. 262**, pp. 16793-16797.

Abouheif, E. A method for testing the asumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* (1999) **vol. 1**, pp. 895-909.

Ahmad, S., Akbar, P. K., Wiker, H. G., Harboe, M. and Mustafa, A. S. Cloning, expression and immunological reactivity of two mammalian cell entry proteins encoded by the mce 1 operon of *Myobacterium tuberculosis*. *Scand. J. Immunol.* (1999) **vol. 50**, pp. 510-518.

Alam, N., Gourinath, S., Dey, S., Srinivasan, A. and Singh, T. P. Substrate-inhibitor interactions in the kinetics of alpha-amylase inhibition by ragi alpha-amylase/trypsin inhibitor (RATI) and its various N-terminal fragments. *Biochemistry* (2001) **vol. 40**, pp. 4229-4233.

Alkharouf, N. W. and Matthews, B. F. SGMD: the Soybean Genomics and Microarray Database *Nucleic Acids Res.* (2004) **vol. 32**, pp. D398-400.

Almeida, M. S., Cabral, K. M. S., Kurtenbach, E., Almeida, F. C. L. and Valente, A. P. Solution structure of Pisum sativum defensin 1 by high resolution NMR: plant defensins, identical backbone with different mechanisms of action. *J. Mol. Biol.* (2002) **vol. 315**, pp. 749-757.

Antorini, M., Breme, U., Caccia, P., Grassi, C., Lebrun, S., Orsini, G., Taylor, G., Valsasina, B., Marengo, E., Todeschini, R. et al. Hydroxylamine-induced cleavage of the asparaginyl-glycine motif in the production of recombinant proteins : the case of insuline-like growth factor I. *Prot. Expr. Purif.* (2000) **vol. 11**, pp. 135-147.

Arai, S., Matsumoto, I. and Abe, K. Phytocystatins and their target enzymes : from molecular biology to practical application : a review. *J. Food Biochem.* (1998) **vol. 22**, pp. 287-299.

Ary, M. B., Richardson, M. and Shewry, P. R. Purification and Characterization of an Insect Alpha-Amylase Inhibitor Endochitinase from Seeds of Job Tears (Coix-Lachryma-Jobi) *Biochim. Biophys. Acta* (1989) **vol. 999**, pp. 260-266.

Asakawa, Y., Dawson, G. W., Griffiths, D. C., Lallemand, J. Y., Ley, S. V., Mori, K., Mudd, A., Pezechik-Leclaire, M., Pickett, J. A., Watanabe, H. et al. Activity of drimane antifeedants and related compounds against aphids and comparative biological effects and chemical reactivity of (-) and (+)polygodial. *J. Chem. Ecol.* (1988) **vol. 14**, pp. 1845-1855.

Aslund, F. and Beckwith, J. The thioredoxin superfamily : redundacy, specificity and gray-area genomics. *J. Bacteriol.* (1999) **vol. 181**, pp. 1375-1379.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. Current Protocols in Molecular Biology. New York: John Wiley. (1997) p.

Bairoch, A. The ENZYME database in 2000. *Nucleic Acids Research* (2000) **vol. 28**, pp. 304-305.

Baker, T. A., Grossman, A. D. and Gross, C. A. A gene regulating the heat shock response in *Escherichia coli* also affects proteolysis. *Proc. Natl. Acad. Sci.* (1984) **vol. 81**, pp. 6779-6783.

Balbas, P. and Bolivar, F. Design and construction of expression plasmid vectors in *Escherichia coli. Methods Enzymol.* (1990) **vol. 185**, pp. 14-37.

Balls, A. K., Hale, W. S. and Harris, T. H. A crystalline protein obtained from a lipoprotein of wheat flour. *Cereal Chemistry* (1942) **vol. 19**, pp. 279-288.

Baneyx, F. Recombinant protein expression in *Escherichia coli. Curr. Opin. Biotechnol.* (1999) **vol. 10**, pp. 411-421.

Barbieri, L., Battelli, M. G. and Stirpe, F. Ribosome-Inactivating proteins from plants *Biochim. Biophys. Acta* (1993) **vol. 1154**, pp. 237-282.

Barbieri, L., Valbonesi, P., Bonora, E., Gorini, P., Bolognesi, A. and Stirpe, F. Polynucleotide: adenosine glycosidase activity of ribosome-inactivating proteins: effects on DNA, RNA and poly(A). *Nucl. Acids Res.* (1997) **vol. 25**, pp. 518-522.

Barcellos, G. B. S., Almeida, L. M., Cavada, B. S., Oliveira, J. T. A. and Carlini, C. R. Canatoxin-, concanavalin A- and canavalin-cross-reactive materials during maturation of Canavalia brasiliensis seeds. *Planta* (1993) **vol. 189**, pp. 397-402.

Barrett, A. J. The cystatins : a new class of peptidase inhibitors. *TIBS* (1987) **vol. 12**, pp. 193-196.

Batalia, M. A., Monzingo, A. F., Ernst, S., Roberts, W. and Robertus, J. D. The crystal structure of the antifungal protein zeamatin, a member of the thaumatin-like, PR-5 protein family. *Nat. Struct. Biol.* (1996) **vol. 3**, pp. 19-23.

Battelli, M. G., Barbieri, L., Bolognesi, A., Buonamici, L., Valbonesi, P., Polito, L., Van Damme, E. J. M., Peumans, W. J. and Stirpe, F. Ribosome-inactivating lectins with polynucleotide; adenosine glycosidase activity. *FEBS Letters* (1997) **vol. 408**, pp. 355-359.

Bell, C. J., Dixon, R. A., Farmer, A. D., Flores, R., Inman, J., Gonzales, R. A., Harrison, M. J., Paiva, N. L., Scott, A. D., Weller, J. W. et al. The Medicago Genome Initiative: a model legume database *Nucl. Ac. Res.* (2001) **vol. 29**, pp. 114-117.

Bell, E. A. Canavanine in the Leguminosae. *Biochem. J.* (1960) **vol. 75**, pp. 618-620.

Bell, E. A. Non-Protein Amino Acids in Plants. In *Secondary Plant Products.*, (eds. E. A. Bell and B. V. Charlwood): Springer-Verlag, (1980), pp. 403-432.

Bell, E. A. Secondary compounds and insect herbivores. In *6th International Symposium of Insect-Plant Relationships*, eds. V. Labeyrie G. Fabres and D. Lachaise, (1987) pp. 19-23. Pau: Dordrecht: Junk, W.

Bell, E. A. and Janzen, D. H. Medical and ecological considerations on L-DOPA and 5-HTP in seeds. *Nature* (1971) **vol. 229**, pp. 136-137.

Bell, E. A., Lackey, J. A. and Polhill, R. M. The systematic significance of canavanine in the Papilionoideae (Faboideae). *Biochem. Syst. Ecol.* (1978) **vol. 6**, pp. 201-212.

Berenbaum, M. Coumarins and caterpillars: a case for coevolution. *Evolution* (1983) **vol. 37**, pp. 163-179.

Berenbaum, M. R. Coumarins In *Herbivores: their interactions with secondary plant metabolites*., (eds. G. A. Rosenthal and M. R. Berenbaum) New York: Academic Press, (1991) vol. I, pp. 221-249.

Bessette, P. H., Aslund, F., Beckwith, J. and Georgiou, G. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci.* (1999) **vol. 96**, pp. 13703-13708.

Better, M., Chang, C. P., Robinson, R. and Horwitz, A. H. *Escherichia coli* secretion of an active chimeric antibody fragment. *Science* (1988) **vol. 240**, pp. 1041-1043.

Betzel, C., Pal, G. P. and Saenger, W. Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1,5 A resolution. *Acta Crystallogr.* (1988) **vol. 44**, pp. 163-172.

Birch, A. N. E., Geoghegan, I. E., Majerus, M. E. N., Mcnicol, J. W., Hackett, C. A., Gatehouse, A. M. R. and Gatehouse, J. A. Tri-trophic interactions involving pest aphids, predatory 2-spot ladybirds and transgenic potatoes expressing snowdrop lectin for aphid resistance *Mol. Breed.* (1999) **vol. 5**, pp. 75-83.

Birch, A. N. E., Griffiths, D. W., Hopkins, R. J., Macfarlane Smith, W. H. and McKinlay, R. G. Glucosinolate responses of swede, kale, forage and oilseed rape to root damage by turnip root fly (*Delia floralis*) larvae. *J. Sci. Food Agric.* (1992) **vol. 60**, pp. 1-9.

Birk, Y. Proteinase inhibitors from plant sources. *Methods Enzymol.* (1976) **vol. 45**, pp. 695-697.

Birnboim, H. C. and Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* (1979) **vol. 7**, pp. 1513-1523.

Bisby, F. A., Buckingham, J. and Harborne, J. B. Phytochemical dictionary of the Leguminosae.: Chapman and Hall. (1994) 854 p.

Bishop, J. G., Dean, A. M. and Mitchell-Olds, T. Rapid evolution in plant chitinases : molecular targets of selection in plant-pathogen co-evolution. *Proc. Natl. Acad. Sci. USA* (2000) **vol. 97**, pp. 5322-5327.

Blein, J. P., Coutos-Thevenot, P., Marion, D. and Ponchet, M. From elicitins to lipid-transfer proteins : a new insight in cell signalling involved in plant defence mechanisms. *Trends in Plant Science* (2002) **vol. 7**, pp. 293-296.

Bloch, C. J. and Richardson, M. A new family of small (5kDa) protein inhibitors of insect alpha-amylases from seeds of sorghum (Sorghum bicolor (L) Moench) have sequence homologies with wheat gamma-purothionins. *FEBS Letters* (1991) **vol. 279**, pp. 101-104.

Blum, H., Beier, H. and Gross, H. J. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide *Electrophoresis* (1987) **vol. 8**, pp. 93-99.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. The 2.0 A X-ray cristal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO Journal* (1988) **vol. 7**, pp. 2593-2599.

Bode, W., Greyling, J. G., Huber, R., Otlewski, J. and Wilusz, T. The refined 2.0 A X-ray crystal structure of the complex formed between bovine beta-trypsin and CMTI-I, a trypsin inhibitor from squash seeds (Cucurbita maxima). Topological similarity of the squash seed inhibitors with the carboxipeptidase A inhibitor from potatoes. *FEBS Letters* (1989) **vol. 242**, pp. 285-292.

Bode, W. and Huber, R. Natural protein proteinase inhibitors and their interactions with proteinases *Eur. J. Biochem.* (1992) **vol. 204**, pp. 433-451.

Bode, W. and Huber, R. Structural basis of the endoproteinase-protein inhibitor interaction. *Bioch. Biophy. Acta* (2000) **vol. 1477**, pp. 241-252.

Bohlmann, H. and Apel, K. Thionins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* (1991) **vol. 42**, pp. 227-240.

Bohlmann, H., Clausen, S., Behnke, S., Giese, H., Hiller, C., Reimann Philipp, U., Schrader, G., Barkholt, V. and Apel, K. Leaf-specific thionins of barley : a novel class of cell wall proteins toxic to plant-pathogenic fungi and possibly involved in the defense mechanisms of plants. *EMBO Journal* (1988) **vol. 7**, pp. 1559-1565.

Boller, T. Chitinase : a defense of higher plants against pathogens. In *Chitin in Nature and Technology.*, (eds. R. Muzzarelli C. Jeuniaux and G. W. Gooday) New York: Plenum, (1986), pp. 223-230.

Boman, H. G. Peptide Antibiotics and their Role in Innate Immunity. *Annu. Rev. Immunol.* (1995) **vol. 13**, pp. 61-92.

Bonekamp, F., Andersen, H. D., Christensen, T. and Jensen, K. F. Codon-defined ribosomal pausing in *Eschericha coli* detected by using the pyrE attenuator to probe the coupling between transcription and translation. *Nucleic Acids Res.* (1985) **vol. 13**, pp. 4113-4123.

Bouckaert, J., Poortmans, F., Wyns, L. and Loris, R. Sequential structural changes upon zinc and calcium binding to metal-free concanavalin A. *J. Biol. Chem.* (1996) **vol. 271**, pp. 16144-16150.

Boulter, D., Edwards, G. A., Gatehouse, J. A., Gatehouse, A. M. R. and Hilder, V. A. Additive protective effects of incorporating two different higher plant derived insect resistance genes in transgenic tobacco plants. *Crop Protec.* (1990) **vol. 9**, pp. 351-354.

Bowden, G. A. and Georgiou, G. Folding and aggregation of beta-lactamase in the periplasmic space of *Escherichia coli. J. Biol. Chem.* (1990) **vol. 265**, pp. 16760-16766.

Bowers, W. S., Ohta, T., Cleere, J. S. and Marsella, P. A. Discovery of insect antijuvenile hormones in plants. *Science* (1976) **vol. 119**, pp. 419.

Bown, D. P., Wilkinson, H. S. and Gatehouse, J. A. Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, Helicoverpa armigara, are members of complex multigene families *Insect Biochem. Molec. Biol.* (1997) **vol. 27**, pp. 625-638.

Bown, D. P., Wilkinson, H. S. and Gatehouse, J. A. Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: enzyme characterisation, cDNA cloning and expression *Insect Biochem. Molec. Biol.* (1998) **vol. 28**, pp. 739-749.

Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding *Anal. Biochem.* (1976) **vol. 72**, pp. 248-254.

Brake, A. J. Alpha-factor leader directed-secretion of heterologous proteins from yeast. *Methods Enzymol.* (1990) **vol. 185**, pp. 408-421.

Brinkmann, U., Mattes, R. E. and Buckel, P. Hign-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the DNA Y gene product. *Gene* (1989) **vol. 85**, pp. 109-114.

Broadway, R. M. Are insects resistant to plant proteinase inhibitors? *J. Insect Physiol.* (1995) **vol. 41**, pp. 107-116.

Broadway, R. M. Dietary proteinase inhibitors alter complement of midgut proteases *Arch. Insect Biochem. Physiol.* (1996) **vol. 32**, pp. 39-53.

Broadway, R. M. Dietary regulation of serine proteinases that are resistant to serine proteinase inhibitors *J. Insect Physiol.* (1997) **vol. 43**, pp. 855-874.

Broekaert, W. F., Cammue, B. P. A., De Bolle, M. F. C., Thevissen, K., De Samblanx, G. W. and Osborn, R. W. Antimicrobial peptides from plants. *Critical Reviews in Plant Sciences.* (1997) **vol. 16**, pp. 297-323.

Broglie, K., Chet, I., Holliday, M., Cressman, R., Biddle, P., Knowlton, S., Mauvais, C. J. and Broglie, R. Transgenic Plants with Enhanced Resistance to the Fungal Pathogen Rhizoctonia-Solani *Science* (1991) **vol. 254**, pp. 1194-1197.

Brummer, J., Thole, H. and Kloppstech, K. Hordothionins inhibit protein synthesis at the level of initiation in the Wheat-Germ system *Eur. J. Biochem.* (1994) **vol. 219**, pp. 425-433.

Bruneau, A., Forest, F., Herendeen, P. S., Klitgaard, B. B. and Lewis, G. P. Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trn*L intron sequences. *Syst. Bot.* (2001) **vol. 26**, pp. 487-514.

Brunner, F., Stintzi, A., Fritig, B. and Legrand, M. Substrate specificities of tobacco chitinases. *Plant J.* (1998) **vol. 14**, pp. 225-234.

Buell, G., Schulz, M. F., Selzer, G., Chollet, A., Movva, N. R., Semon, D., Escanez, S. and Kawashima, E. Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-C (IGF-I). *Nucleic Acids Res.* (1985) **vol. 13**, pp. 1923-1938.

Buhot, N., Douliez, J. P., Jacquemard, A., Marion, D., Tran, V., Maume, B. F., Milat, M. L., Ponchet, M., Mikes, V., Kader, J. C. et al. A lipid transfer protein binds to a receptor involved in the control of plant defence reponses. *FEBS Letters* (2001) **vol. 509**, pp. 27-30.

Buonocore, V., Petrucci, T. and Silano, V. Wheat protein inhibitors of alpha-amylase. *Phytochemistry* (1977) **vol. 16**, pp. 811-820.

Burgess, E. P. J., Main, C. A., Stevens, P. S., Christeller, J. T., Gatehouse, A. M. R. and Laing, W. A. Effects of protease inhibitor concentration and combinations on the survival, growth and gut enzyme activities of the black field cricket, Teleogryllus commodus *J. Insect Physiol.* (1994) **vol. 40**, pp. 803-811.

Butler, L. G. Effects of condensed tannin on animal nutrition. In *Chemistry and significance of condensed tannins.*, (eds. R. W. Hemingway and J. J. Karchesy) New York: Plenum Press, (1989), pp. 391-402.

Caaveiro, J. M. M., Molina, A., Gonzales-Manas, J. M., Rodriguez-Palenzuela, P., Garcia Olmedo, F. and Goni, F. M. Differential effects of five types of antipathogenic plant peptides on model membranes *FEBS Letters* (1997) **vol. 410**, pp. 338-342.

Caldwell, J. E., Abildgaard, F., Dzakula, Z., Ming, D., Hellekant, G. and Markley, J. L. Solution structure of the thermostable sweet-tasting protein brazzein. *Nat. Struct. Biol.* (1998) **vol. 5**, pp. 427-431.

Cammue, B. P. A., Thevissen, K., Hendricks, M., Eggermont, K., Goderis, I. J., Proost, P., Van Damme, J., Osborn, R. W., Guerbette, F., Kader, J. C. et al. A potent antimicrobial protein from onion seeds showing sequence homology to plant lipid transfer proteins. *Plant Physiol.* (1995) **vol. 109**, pp. 445-455.

Cardona, C., Kornegay, J., Posso, C. E., Morales, F. and Ramirez, H. Comparative value of four arcelin variants in the development of dry bean lines resistant to the mexican bean weevil *Entomol. Exp. Appl.* (1990) **vol. 56**, pp. 197-206.

Carlini, C. R., Barcellos, G. B. S., Baeta Neves, A. D. V. and Guimaraes, J. A. Immunoreactivity for canatoxin and concanavalin A among proteins of legume seeds. *Phytochemistry* (1988) **vol. 27**, pp. 25-30.

Carlini, C. R. and Grossi de Sa, F. Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides *Toxicon* (2002) **vol. 40**, pp. 1515-39.

Carlini, C. R. and Guimaraes, J. A. Isolation and characterization of a toxic protein from Canavalia ensiformis (jack bean) seeds, distinct from concanavalin A *Toxicon* (1981) **vol. 19**, pp. 667-75.

Carlini, C. R. and Guimaraes, J. A. Plant and microbial toxic proteins as hemilectins: emphasis on canatoxin *Toxicon* (1991) **vol. 29**, pp. 791-806.

Carlini, C. R., Oliveira, A. E., Azambuja, P., Xavier-Filho, J. and Wells, M. A. Biological effects of canatoxin in different insect models: evidence for a proteolytic activation of the toxin by insect cathepsinlike enzymes *J Econ Entomol* (1997) **vol. 90**, pp. 340-8.

Carmona, M. J., Molina, A., Fernandez, J. A., Lopez Fando, J. J. and Garcia Olmedo, F. Expression of the alpha-thionin gene from barley in tobacco confers enhanced resistance to bacterial pathogens. *Plant Journal* (1993) **vol. 3**, pp. 457-462.

Casida, J. E. Pyrethrum-The Natural Insecticide. New York: Academic Press. (1973) p.

Ceci, L. R., Spoto, N., de Virgilio, M. and Gallerani, R. The gene coding for the mustard trypsin inhibitor-2 is discontinuous and wound-inducible *FEBS Letters* (1995) **vol. 364**, pp. 179-81.

Chagolla-Lopez, A., Blanco-Labra, A., Patthy, A., Sanchez, R. and Pongor, S. A novel alpha-amylase inhibitor from Amaranth (Amaranthus hypocondriacus) seeds. *J. Biol. Chem.* (1994) **vol. 269**, pp. 23675-23680.

Charvolin, E., Douliez, J. P., Marion, D., Cohen Abbad, C. and Pebay Peyroula, E. The crystal structure of a wheat nonspecific lipid transfer protein (ns-LTP1) complexed with two molecules of phospholipid at 2.1 A resolution. *Eur. J. Biochem.* (1999) **vol. 264**, pp. 562-568.

Chen, K. C., Lin, C. Y., Kuan, C. C., Sung, H. Y. and Chen, C. S. A novel defensin encoded by a mungbean cDNA exhibits insecticidal activity against bruchid *J. Agric. Food Chem.* (2002) **vol. 50**, pp. 7258-7263.

Chen, Y., Cino, J., Hart, G., Freedman, D., White, C. and Komives, E. A. High protein expression in fermentation of recombinant *Pichia pastoris* by a fed-batch process. *Process Biochemistry* (1996) **vol. 32**, pp. 107-111.

Chew, F. S. and Renwick, J. A. A. Host plant choice in *Pieris* butterflies. In *Chemical Ecology of Insects.*, (eds. R. T. Cardé and W. J. Bell) New York: Chapman & Hall, (1995) vol. 2, pp. 214-238.

Chiche, L., Gaboriaud, C., Heitz, A., Mornon, J. P., Castro, B. and Kollman, P. A. Use of restrained molecular dynamics in water to determine three-dimensional protein structure: prediction of the three-dimensional structure of Ecballium elaterium trypsin inhibitor II. *Proteins* (1989) **vol. 6**, pp. 405-17.

Chrispeels, M. J., Desa, M. F. G. and Higgins, T. J. V. Genetic engineering with alpha-amylase inhibitors makes seeds resistant to bruchids *Seed Sci. Res.* (1998) **vol. 8**, pp. 257-263.

Chrispeels, M. J. and Raikhel, N. V. Lectins, Lectin Genes, and Their Role in Plant Defense *Plant Cell* (1991) **vol. 3**, pp. 1-9.

Christeller, J. T., Farley, P. C., Ramsay, R. J., Sullivan, P. A. and Laing, W. A. Purification, characterization and cloning of an aspartic proteinase inhibitor from squash phloem exudate *Eur. J. Biochem.* (1998) **vol. 254**, pp. 160-167.

Christeller, J. T., Laing, W. A., Markwick, N. P. and Burgess, E. P. J. Midgut protease activities in 12 phytophagous lepidopteran larvae - dietary and protease inhibitor interactions *Insect Biochem. Molec. Biol.* (1992) **vol. 22**, pp. 735-746.

Christeller, J. T., Markwick, N. P. and Burgess, E. P. J. Midgut proteinase activities of three keratinolytic larvae, hofmannophila pseudospretella, tineola bisselliella, and anthrenocerus australis and the effect of proteinase inhibitors on proteolysis *Arch. Insect Biochem. Physiol.* (1994) **vol. 25**, pp. 159-173.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brunger, A. T., Karplus, M. and Gronenborn, A. M. The three dimensional structure of alpha-1-purothionin in solution : combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO Journal* (1986) **vol. 5**, pp. 2729-2735.

Cociancich, S., Ghazi, A., Hetru, C., Hoffmann, J. A. and Letellier, L. Insect defensin, an inducible antibacterial peptide, forms voltage-dependant channels in *Micrococcus luteus*. *J. Biol. Chem.* (1993) **vol. 268**, pp. 19239-19245.

Cohen, E. Chitin synthesis and degradation as targets for pesticide action. *Arch. Insect Biochem. Physiol.* (1993) **vol. 22**, pp. 245-261.

Collinge, D., Kragh, K., Mikkelsen, J., Nielsen, K., Rasmussen, U. and Vad, K. Plant chitinases. *Plant J.* (1993) **vol. 3**, pp. 31-40.

Cornelis, P. Expressing genes in different *Escherichia coli* compartments. *Curr. Opin. Biotechnol.* (2000) **vol. 11**, pp. 450-454.

Cornelissen, B. J. C., Hooft Van Huijsduijnen, R. A. M. and Bol, J. F. A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin. *Nature* (1986) **vol. 231**, pp. 531-532.

Cornet, B., Bonmatin, J. M., Hetru, C., Hoffmann, J. A., Ptak, M. and Vovelle, F. Refined three-dimensional solution structure of insect defensin A. *Structure* (1995) **vol. 3**, pp. 435-448.

Coulson, E. J., Harris, T. H. and Axelrod, B. Effect on small laboratory animals of the injection of the crystalline hydrochloride of a sulfur protein from wheat flour. *Cereal Chemistry* (1942) **vol. 19**, pp. 301-307.

Coutos-Thevenot, P., Jouenn, e. T., Maes, O., Guerbette, F., Grosbois, M., Le Caer, J. P., Boulay, M., Deloire, A., Kader, J. C. and Guern, J. Four 9-kDa proteins excreted by somatic embryos of grapevine are isoforms of lipid-transfer proteins. *Eur. J. Biochem.* (1993) **vol. 217**, pp. 885-889.

Crowe, J., Dobeli, H., Gentz, R., Hochuli, E., Struber, D. and Henco, K. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol. Biol.* (1994) **vol. 31**, pp. 371-387.

Czapla, T. H. and Lang, B. A. Effect of Plant Lectins on the Larval Development of European Corn Borer (Lepidoptera, Pyralidae) and Southern Corn Rootworm (Coleoptera, Chrysomelidae) *J. Econ. Entomol.* (1990) **vol. 83**, pp. 2480-2485.

D'Mello, J. P. F. Toxic amino acids. In *Toxic substances in crop plants*., (eds. J. P. F. D'Mello C. M. Duffus and J. H. Duffus) Cambridge: The Royal Society of Chemistry., (1991), pp. 22-48.

Da Costa, C. P. and Jones, C. M. Cucumber beetle resistance and mite susceptibility controlled by the bitter gene in *Cucumis sativus* L. *Science* (1971) **vol. 172**, pp. 1145-1146.

Dawson, G. W., Hallahan, D. L., Mudd, A., Patel, M. M., Pickett, J. A., Wadhams, L. J. and Wallsgrove, R. M. Secondary plant metabolites as targets for genetic modification of crop plants for pest resistance. *Pestic. Sci.* (1989) **vol. 27**, pp. 191-201.

De Bernadez Clark, E. Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* (1998) **vol. 9**, pp. 157-163.

De Leo, F., Bonadé-Bottino, M., Ceci, L. R., Gallerani, R. and Jouanin, L. Opposite effects on *Spodoptera littoralis* larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants *Plant Physiol.* (1998) **vol. 118**, pp. 997-1004.

De Leo, F., Bonadé-Bottino, M. A., Ceci, L. R., Gallerani, R. and Jouanin, L. Effects of a mustard trypsin inhibitor expressed in different plants on three lepidopteran pests. *Insect Biochem. Mol. Biol.* (2001a) **vol. 31**, pp. 593-602.

De Leo, F., Ceci, L. R., Jouanin, L. and Gallerani, R. Analysis of mustard trypsin inhibitor-2 gene expression in response to developmental or environmental induction *Planta* (2001b) **vol. 212**, pp. 710-717.

De Leo, F. and Gallerani, R. The mustard trypsin inhibitor 2 affects the fertility of *Spodoptera littoralis* larvae fed on transgenic plants. *Insect Biochem. Mol. Biol.* (2002) **vol. 32**, pp. 489-496.

Delledonne, M., Allegro, G., Belenghi, B., Balestrazzi, A., Picco, F., Levine, A., Zelasco, S., Calligari, P. and Confalonieri, M. Transformation of white poplar (Populus alba L.) with a novel Arabidopsis thaliana cysteine proteinase inhibitor and analysis of insect pest resistance *Mol. Breed.* (2001) **vol. 7**, pp. 35-42.

Delobel, A. and Tran, M. Les Coléoptères des denrées alimentaires entreposées dans les régions chaudes. Paris: ORSTOM. (1993) 424 p.

Delobel, B. Une protéine du petit pois contre les charançons. In *Presse info.*, INRA, disponible sur: http: //www.inra.fr/PRESSE/MAI99/nb1.htm. (1999).

Delobel, B. and Grenier, A. M. Effect of non cereal food on cereal and tamarin pod weevils (Coleoptera: Curculionidae) *J. stored Prod. Res.* (1993) **vol. 29**, pp. 7-14.

Delobel, B., Grenier, A. M., Gueguen, J., Ferrasson, E. and Mbaiguinam, M. Utilisation d'un polypeptide dérivé d'une albumine PA1b de légumineuse comme insecticide (Brevet-98/05877),, (1998) pp. 1-25. Paris.

Derman, A. I. and Beckwith, J. *Escherichia coli* alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm. *J. Bacteriol.* (1991) **vol. 173**, pp. 7719-7722.

Derman, A. I., Prinz, W. A., Belin, D. and Beckwith, J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* (1993) **vol. 262**, pp. 1744-1747.

Descoins, J. C. and Marion Poll, F. Electrophysiological responses of gustatory sensilla of Mamestra brassicae (Lepidoptera, Noctuidae) larvae to three ecdysteroids : ecdysone, 20-hydroxyecdysone and ponasterone A. *Journal of Insect Physiology* (1999) **vol. 45**, pp. 871-876.

Di Guan, C., Li, P., Riggs, P. D. and Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* (1988) **vol. 67**, pp. 21-30.

Diaz, I., Carmona, M. J. and Garcia Olmedo, F. Effects of thionins on beta-glucuronidase *in vitro* and in plant protoplasts. *FEBS Lett.* (1992) **vol. 269**, pp. 279-282.

Dietrichs, H. H. and Funke, H. Free amino acids in phloem and spring sap of red beeches (*Fagus sylvatica*). *Holzforschung* (1967) **vol. 21**, pp. 102-107.

Dinan, L. A strategy towards the elucidation of the contribution made by phytoecdysteroids to the deterrence of invertebrate predators on plants *Russian J. Plant Physiol.* (1998) **vol. 45**, pp. 296-305.

Dinan, L. Phytoecdysteroids : biological aspects. *Phytochemistry* (2001) **vol. 57**, pp. 325-339.

Dinan, L., Whiting, P., Girault, J. P., Lafont, R., Dhadialla, T. S., Cress, D. E., Mugat, B., Antoniewski, C. and Lepesant, J. A. Cucurbitacins are insect steroid hormone antagonists acting at the ecdysteroid receptor *Biochem. J.* (1997) **vol. 327**, pp. 643-650.

Diol, O. Caractérisation de l'activité biologique *in vitro* de peptides A1b synthétiques et d'homologues purifiés à partir de graines de Légumineuses. Test de binding au récepteur du charançon et corrélation des activités *in vitro* / *in vivo*., *DEA/ Masters thesis,* Lyon: UCB Lyon 1 / INSA Lyon. (2003) 30 p.

Douliez, J. P., Michon, T., Elmorjani, K. and Marion, D. Structure, biological and technological functions of lipid transfer proteins and indolines, the major lipid binding proteins from cereal kernels. *J. Cereal Sci.* (2000) **vol. 30**, pp. 1-20.

Down, R. E., Ford, L., Woodhouse, S. D., Raemaekers, R. J. M., Leitch, B., Gatehouse, J. A. and Gatehouse, A. M. R. Snowdrop lectin (GNA) has no acute toxic effects on a beneficial insect predator, the 2-spot ladybird (*Adalia bipunctata* L.) *J. Insect Physiol.* (2000) **vol. 46**, pp. 379-391.

Down, R. E., Gatehouse, A. M. R., Hamilton, W. D. O. and Gatehouse, J. A. Snowdrop lectin inhibits development and decreases fecundity of the glasshouse potato aphid (*Aulacorthum solani*) when administered in vitro and via transgenic plants both in laboratory and glasshouse trials *J. Insect Physiol.* (1996) **vol. 42**, pp. 1035-1045.

Doyle, J. J. Gene trees and species trees : molecular systematics as one-character taxonomy. *Syst. Bot.* (1992) **vol. 17**, pp. 144-163.

Doyle, J. J., Chapill, J. A., Bailey, D. C. and Kajita, T. Towards a comprehensive phylogeny of legumes : evidence from *rbc*L sequences and non-molecular data. In *Advances in Legume systematics*., (eds. P. S. Herendeen and A. Bruneau) Kew: Royal Botanic Gardens, (2000), pp. 1-20.

Doyle, J. J. and Doyle, J. L. Chloroplast DNA phylogeny of the papilionoid legume tribe Phaseoleae. *Syst. Bot.* (1993) **vol. 18**, pp. 309-327.

Doyle, J. J., Doyle, J. L., Ballenger, J. A., Dickson, E. F., Kajita, T. and Ohashi, H. A phylogeny of the chloroplast gene *rbc*L in the Leguminosae : Taxonomic correlations and insights into the evolution of nodulation. *American J. Botany* (1997) **vol. 84**, pp. 541-554.

Doyle, J. J., Doyle, J. L., Ballenger, J. A. and Palmer, J. D. The distribution and phylogenetic significance of a 50kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol. Phylo. Evol.* (1996) **vol. 5**, pp. 429-438.

Doyle, J. J. and Luckow, M. A. The rest of the Iceberg. Legume diversity and Evolution in a phylogenetic context. *Plant Physiol.* (2003) **vol. 131**, pp. 900-910.

Duncan, A. J. Glucosinolates In *Toxic substances in crop plants*., (eds. J. P. F. D'Mello C. M. Duffus and J. H. Duffus) Cambridge: The Royal Society of Chemistry., (1991), pp. 126-147.

Dunn, M. A., Hughes, M. A., Zhang, L., Pearce, R. S., Quigley, A. S. and Jack, P. L. Nucleotide sequence and molecular analysis of the low temperature induced cereal gene, BLT4. *Mol. Gen. Genet.* (1991) **vol. 229**, pp. 389-394.

Dussourd, D. E. and Eisner, T. Vein-cutting behaviour : insect counterplay to the latex defense of plants. *Science* (1987) **vol. 237**, pp. 898-901.

Ealing, P. M., Hancock, K. R. and White, D. W. Expression of the pea albumin 1 gene in transgenic white clover and tobacco *Transgenic Res.* (1994) **vol. 3**, pp. 344-54.

Ehrlich, P. R. and Raven, P. H. Butterflies and plant : a study in coevolution. *Evolution* (1964) **vol. 18**, pp. 586-608.

Eidels, L., Proia, R. L. and Hart, D. A. Membrane receptors for bacterial toxins. *Microbiol. Rev.* (1983) **vol. 47**, pp. 596-620.

Elmorjani, K., Thievin, M., Michon, T., Popineau, Y., Hallet, J.-N. and Gueguen, J. Synthetic genes specifying periodic polymers modelled on the repetitive domain of wheat gliadins : Conception and expression. *Biochem. Biophys. Res. Com.* (1997) **vol. 239**, pp. 240-246.

Emr, S. D. Heterologous gene expression in yeast. *Methods Enzymol.* (1990) **vol. 185**, pp. 231-233.

Endo, Y., Tsurugi, K. and Lambert, J. M. The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: the RNA *N*-glycosidase activity of the proteins. *Biochem. Biophys. Res. Commun.* (1988) **vol. 150**, pp. 1032-1036.

Epple, P., Apel, K. and Bohlmann, H. An *Arabidopsis thaliana* thionin gene is inducible via a signal transduction pathway different from that for pathogenesis related proteins. *Plant Physiology* (1995) **vol. 109**, pp. 813-820.

Epple, P., Apel, K. and Bohlmann, H. Overexpression of an endogenous thionin gives enhanced resistance of *Arabidopsis thaliana* against *Fusarium oxysporium*. *Plant Cell* (1997) **vol. 9**, pp. 509-520.

Evans, C. S., Qureshi, Y. and Bell, E. A. Free amino acids in the seeds of Acacia species. *Phytochemistry* (1977) **vol. 16**, pp. 565-670.

Evans, J., Wang, Y. D., Shaw, K. P. and Vernon, L. P. Cellular responses to Pyrularia thionin are mediated by Ca2+ influx and phospholipase A2 activation and are inhibited by thionin tyrosine iodination. *Proc. Natl. Acad. Sci. U.S.A.* (1989) **vol. 86**, pp. 5849-5853.

Evans, S. V., Fellows, L. E., Shing, T. K. M. and Fleet, G. W. J. Glycosidase inhibition by plant alkaloids which are structural analogues to monosaccharides. *Phytochemistry* (1985) **vol. 24**, pp. 1953-1955.

Ewen, S. W. and Pusztai, A. Effect of diets containing genetically modified potatoes expressing Galanthus nivalis lectin on rat small intestine [letter] [see comments] *Lancet* (1999) **vol. 354**, pp. 1353-4.

Fant, F., Vranken, W., Broekaert, W. F. and Borremans, F. Determination of the three-dimensional solution structure of *Raphanus sativus* antifungal protein 1 by 1H NMR. *J. Mol. Biol.* (1998) **vol. 279**, pp. 257-270.

Fant, F., Vranken, W. F. and Borremans, F. A. The three-dimensional solution structure of *Aesculus hippocastanum* antimicrobial protein 1 determined by 1H nuclear magnetic resonance. *Proteins* (1999) **vol. 37**, pp. 388-403.

Febvay, G. and Rahbé, Y. "Toxicologie", un programme pour l'analyse des courbes de mortalité par la méthode des probits sur MacIntosh *Cah. Techn. INRA* (1991) **vol. 27**, pp. 77-78.

Fellows, L. E., Evans, S. V., Nash, R. J. and Bell, E. A. Polyhydroxy plant alkaloids as glucodidase inhibitors and their possible ecological role. In *Natural resistance*

of plants to pests., (eds. M. B. Green and P. A. Hedin) Washington, DC: American Chemical Society, (1986), pp. 72-78.

Feng, G. H., Richardson, M., Chen, M. S., Kramer, K. J., Morgan, T. D. and Reeck, G. R. alpha-amylase inhibitors from wheat: amino acid sequences and patterns of inhibition of insect and human alpha-amylases *Insect Biochem. Mol. Biol.* (1996) **vol. 26**, pp. 419-426.

Ferreira-DaSilva, C. T., Gombarovits, M. E., Masuda, H., Oliveira, C. M. and Carlini, C. R. Proteolytic activation of canatoxin, a plant toxic protein, by insect cathepsin-like enzymes [In Process Citation] *Arch. Insect Biochem. Physiol.* (2000) **vol. 44**, pp. 162-71.

Fery, R. L. and Cuthbert, J. F. P. Measurement of podwall resistance to the cowpea cucurlio in the southern pea, *Vigna unguiculata* (L.). *Walp. Hortic. Sci.* (1979) **vol. 14**, pp. 29-30.

Firmino, F., Fernandes, K. V. S., Sales, M. P., Gomes, V. M., Miranda, M. R. A., Domingues, S. J. S. and Xavier, J. Cowpea (*Vigna unguiculata*) vicilins associate with putative chitinous structures in the midgut and feces of the bruchid beetles *Callosobruchus maculatus* and *Zabrotes subfasciatus*. *Braz. J. Med. Biol. Res.* (1996) **vol. 29**, pp. 749-756.

Florack, D. E. A. and Stiekema, W. J. Thionins: properties, possible biological roles and mechanisms of action. *Plant Molecular Biology* (1994) **vol. 26**, pp. 25-37.

Foissac, X., Loc, N. T., Christou, P., Gatehouse, A. M. R. and Gatehouse, J. A. Resistance to green leafhopper (Nephotettix virescens) and brown planthopper (Nilaparvata lugens) in transgenic rice expressing snowdrop lectin (Galanthus nivalis agglutinin; GNA) *J. Insect Physiol.* (2000) **vol. 46**, pp. 573-583.

Follmer, C., Barcellos, G. B. S., Zingali, R. B., Machado, O. L. T., Alves, E. W., Barja Fildago, C., Guimaraes, J. A. and Carlini, C. R. Canatoxin, a toxic protein of jack beans (Canavalia ensiformis), is a variant form of urease (EC 3.5.1.5). Biological effects of urease independant of its ureolytic activity. *Biochem. J.* (2001) **vol. 360**, pp. 217-224.

Fory, L. F., Finardi, F., Quintero, C. M., Osborn, T. C., Cardona, C., Chrispeels, M. J. and Mayer, J. E. alpha-Amylase inhibitors in resistance of common beans to the mexican bean weevil and the bean weevil (Coleoptera: bruchidae) *J. Econ. Entomol.* (1996) **vol. 89**, pp. 204-210.

Fossdal, C. G., Nagy, N. E., Sharma, P. and Lonneborg, A. The putative gymnosperm plant defensin polypeptide (SPI1) accumulates after seed germination, is not readily released, and the SPI1 levels are reduced in Pythium dimorphum-infected spruce roots. *Plant Mol. Biol.* (2003) **vol. 52**, pp. 291-302.

Fraenkel, G. S. The raison d'être of secondary compounds. *Science* (1959) **vol. 129**, pp. 1466-1470.

Franco, O. L., Rigden, D. J., Melo, F. R., Bloch, C., Silva, C. P. and Grossi de Sa, M. F. Activity of wheat alpha-amylase inhibitors towards bruchid alpha-amylases and structural explanation of observed specificities *Eur. J. Biochem.* (2000) **vol. 267**, pp. 2166-2173.

Frazier, J. L. The perception of plant allelochemicals that inhibit feeding. In *Molecular aspects of insect-plant associations*., (eds. L. B. Brattsten and S. Ahmad) New York: Plenum Press, (1986), pp. 1-42.

Galtier, N., Gouy, M. and Gautier, C. Inferring phylogenies from DNA sequences of unequal base compositions *Proc. Natl. Acad. Sci. USA* (1995) **vol. 92**, pp. 11317-11321.

Galtier, N., Gouy, M. and Gautier, C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny *Comput. Applic. Biosci.* (1996) **vol. 12**, pp. 543-548.

Gamas, P., Niebel, F. d. C., Lescure, N. and Cullimore, J. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development *Mol. Plant Microbe Interact.* (1996) **vol. 9**, pp. 233-42.

Garcia Casado, G., Sanchez Monge, R., Chrispeels, M. J., Armentia, A., Salcedo, G. and Gomez, L. Role of complex asparagine-linked glycans in the allergenicity of plant glycoproteins. *Glycobiology* (1996) **vol. 6**, pp. 471-477.

Garcia Casado, G. L., Sanchez Monge, R., Lopez Otin, C. and Salcedo, G. Rye inhibitors of animal alpha-amylases shown different specificities, aggregative properties and IgE-binding capacities than their homologues from wheat and barley. *Eur. J. Biochem.* (1994) **vol. 224**, pp. 525-531.

Garcia Olmedo, F., Carbonero, P., Hernandez Lucas, C., Paz Ares, J., Ponz, F., Vicente, O. and Sierra, J. M. Inhibition of eukaryotic cell-free protein synthesis by thionins from wheat endosperm. *Biochim. Biophys. Acta* (1983) **vol. 750**, pp. 52-56.

Garcia Olmedo, F., Molina, A., Segura, A. and Moreno, A. The defensive role of nonspecific lipid-transfer proteins in plants. *Trends in Microbiology* (1995) **vol. 3**, pp. 72-74.

Garcia Olmedo, F., Rodriguez-Palenzuela, P., Hernandez Lucas, C., Ponz, F., Marana, C., Carmona, M. J., Lopez Fando, J. J., Fernandez, J. A. and Carbonero, P. The thionins : a protein family that includes purothionins, viscotoxins and crambin. In *Oxford Surveys of Plant Molecular and Cell Biology 6.*, ed. B. J. Miflin):, (1989), pp. 31-60.

Garcia Olmedo, F., Salcedo, G., Sanchez Monge, R., Hernandez Lucas, C., Carmona, M. J., Lopez Fando, J. J., Fernandez, J. A., Gomez, L., Royo, J., Garcia Maroto, F. et al. Trypsin/alpha-amylase inhibitors and thionins : possible defense proteins from barley. In *Barley : Genetics, Biochemistry, Molecular Biology and Biotechnology.*, ed. P. R. Shewry) Wallingford, UK: CAB International, (1992), pp. 335-350.

Gascuel, O. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* (1997) **vol. 14**, pp. 685-695.

Gatehouse, A. M., Barbieri, L., Stirpe, F. and Croy, R. D. D. Effects of ribosome-inactivating proteins on insect development, differences between Lepidoptera and Coleoptera. *Entomol. Exp. Appl.* (1990) **vol. 54**, pp. 43-51.

Gatehouse, A. M., Powell, K. S., Peumans, W. J., Van Damme, E. J. and Gatehouse, J. A. Insecticidal properties of plant lectins: their potential in plant protection. In *Lectins: Biomedical perspectives.*, (eds. A. Pusztai and S. Bardocz) London: Taylor & Francis, (1995), pp. 35-58.

Gatehouse, A. M. R. and Boulter, D. Assessment of antimetabolic effects of trypsin inhibitors from Cowpea (*Vigna unguiculata*) and other legumes on development of the bruchid beetle *Callosobruchus maculatus. J. Sci. Food Agric.* (1983) **vol. 34**, pp. 345-350.

Gatehouse, A. M. R., Davidson, G. M., Newell, C. A., Merryweather, A., Hamilton, W. D. O., Burgess, E. P. J., Gilbert, R. J. C. and Gatehouse, J. A.

Transgenic potato plants with enhanced resistance to the tomato moth, Lacanobia oleracea: growth room trials *Mol. Breed.* (1997) **vol. 3**, pp. 49-63.

Gatehouse, A. M. R., Davison, G. M., Stewart, J. N., Galehouse, L. N., Kumar, A., Geoghegan, I. E., Birch, A. N. E. and Gatehouse, J. A. Concanavalin A inhibits development of tomato moth (Lacanobia oleracea) and peach-potato aphid (Myzus persicae) when expressed in transgenic potato plants *Mol. Breed.* (1999a) **vol. 5**, pp. 153-165.

Gatehouse, A. M. R., Down, R. E., Powell, K. S., Sauvion, N., Rahbé, Y., Newell, C. A., Merryweather, A., Hamilton, W. D. O. and Gatehouse, J. A. Transgenic potato plants with enhanced resistance to the peach-potato aphid *Myzus persicae Entomol. Exp. Appl.* (1996) **vol. 79**, pp. 295-307.

Gatehouse, A. M. R. and Gatehouse, J. A. Identifying proteins with insecticidal activity: Use of encoding genes to produce insect-resistant transgenic crops *Pestic. Sci.* (1998) **vol. 52**, pp. 165-175.

Gatehouse, A. M. R., Norton, E., Davison, G. M., Babbe, S. M., Newell, C. A. and Gatehouse, J. A. Digestive proteolytic activity in larvae of tomato moth, Lacanobia oleracea; effects of plant protease inhibitors in vitro and in vivo *J. Insect Physiol.* (1999b) **vol. 45**, pp. 545-558.

Gatehouse, A. M. R., Shi, Y., Powell, K. S., Brough, C., Hilder, V. A., Hamilton, W. D. O., Newell, C. A., Merryweather, A., Boulter, D. and Gatehouse, J. A. Approaches to insect resistance using transgenic plants *Phil. Trans. R. Soc. Lond. B* (1993) **vol. 342**, pp. 279-286.

Gatehouse, J. A., Gilroy, J., Hoque, M. S. and Croy, R. R. Purification, properties and amino acid sequence of a low-Mr abundant seed protein from pea (*Pisum sativum* L.) *Biochem. J.* (1985) **vol. 225**, pp. 239-47.

Gausing, K. Thionin genes specifically expressed in barley leaves. *Planta* (1987) **vol. 171**, pp. 241-246.

Gavit, P. and Better, M. Production of antifungal recombinant peptides in *Escherichia coli. Journal of Biotechnology* (2000) **vol. 79**, pp. 127-136.

Germino, J. and Bastia, D. Rapid purification of a cloned gene product by genetic fusion and site-specific proteolysis. *Proc. Natl. Acad. Sci.* (1984) **vol. 81**, pp. 4692-4696.

Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y. and Inouye, M. Secretion cloning vectors in *Escherichia coli. EMBO J.* (1984) **vol. 3**, pp. 2437-2442.

Gilbert, B. L. and Norris, D. M. A chemical basis for bark beetle (*Scolytus*) distinction between host and non-host trees. *J. Insect Physiol.* (1968) **vol. 14**, pp. 1063-1068.

Gincel, E., Simorre, J. P., Caille, A., Marion, D., Ptak, M. and Vovelle, F. Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional 1H-NMR data. A new folding for lipid carriers. *Eur. J. Biochem.* (1994) **vol. 226**, pp. 413-422.

Girard, C., Bonadé-Bottino, M., Phamdelegue, M. H. and Jouanin, L. Two strains of cabbage seed weevil (Coleoptera: Curculionidae) exhibit differential susceptibility to a transgenic oilseed rape expressing oryzacystatin I *J. Insect Physiol.* (1998a) **vol. 44**, pp. 569-577.

Girard, C., Le Metayer, M., Bonadé-Bottino, M., Phamdelegue, M. H. and Jouanin, L. High level of resistance to proteinase inhibitors may be conferred by proteolytic cleavage in beetle larvae *Insect Biochem. Molec. Biol.* (1998b) **vol. 28**, pp. 229-237.

Giri, A. P., Harsulkar, A. M., Deshpande, V. V., Sainani, M. N., Gupta, V. S. and Ranjekar, P. K. Chickpea defensive proteinase inhibitors can be inactivated by podborer gut proteinases *Plant Physiol.* (1998) **vol. 116**, pp. 393-401.

Giri, A. P. and Kachole, M. S. Amylase inhibitors of pigeonpea (Cajanus cajan) seeds *Phytochemistry* (1998) **vol. 47**, pp. 197-202.

Gomes, E., Sagot, E., Gaillard, C., Laquitaine, L., Poinssot, B., Sanejouand, Y. H., Delrot, S. and Coutos-Thevenot, P. Nonspecific lipid-transfer protein genes expression in grape (Vitis sp.) cells in response to fungal elicitor treatments. *Mol. Plant Microbe Interac.* (2003) **vol. 16**, pp. 456-464.

Gomes, V. M., Da Cunha, M., Miguens, F. C., Fernandes, K. V. S., Rose, T. L. and Xavier Filho, J. Ultrastructure and immunolabelling of *Vigna unguiculata* vicilins (7S storage proteins) associated to fungi cells. *Plant Sci.* (1998) **vol. 138**, pp. 81-89.

Goossens, A., Geremia, R., Bauw, G., Van Montagu, M. and Angenon, G. Isolation and characterisation of arcelin-5 proteins and cDNAs *Eur J Biochem* (1994) **vol. 225**, pp. 787-95.

Goossens, A., Quintero, C., Dillen, W., De Rycke, R., Valor, J. F., De Clercq, J., Van Montagu, M., Cardona, C. and Angenon, G. Analysis of bruchid resistance in the wild common bean accession G02771: no evidence for insecticidal activity of arcelin 5 *J. Exp. Bot.* (2000) **vol. 51**, pp. 1229-36.

Goransson, U. and Craik, D. J. Disulfide mapping of the cyclotide kalata B1. Chemical proof of the cystic cystine knot motif. *J. Biol. Chem.* (2003) **vol. 278**, pp. 48188-48196.

Gorham, J. Lunularic acid and related compounds in liverworts, algae and Hydrangea. *Phytochemistry* (1977) **vol. 16**, pp. 249-253.

Gottesman, S. Minimizing proteolysis in *Escherichia coli* : genetic solutions. *Methods Enzymol.* (1990) **vol. 185**, pp. 119-129.

Gouy, M. and Gautier, C. Codon usage in bacteria : correlation with gene expressivity. *Nucleic Acids Res.* (1982) **vol. 10**, pp. 7055-7074.

Graham, J., McNicol, R. J. and Greig, K. Towards genetic based insect resistance in strawberry using the cowpea trypsin inhibitor. *Ann. Appl. Biol.* (1995) **vol. 127**, pp. 163-173.

Gram, H., Ramage, P., Memmert, K., Gamse, R. and Kocher, H. P. A novel approach for high level production of a recombinant human parathyroid hormone fragment in *Escherichia coli*. *Bio/technol.* (1994) **vol. 12**, pp. 1017-1023.

Grant, G. Lectins In *Toxic Substances in Crop Plants.*, (eds. J. P. F. D'Mello M. C. Duffus and J. H. Duffus) Cambridge: Royal Society of Chemistry, (1991), pp. 49-67.

Grenier, A. M., Mbaiguinam, M. and Delobel, B. Genetical analysis of the ability of the rice weevil Sitophilus oryzae (Coleoptera, Curculionidae) to breed on split peas *Heredity* (1997) **vol. 79**, pp. 15-23.

Grenier, J. and Asselin, A. Some pathogenesis-related proteins are chitosanases with lytic activity against fungal spores. *Mol. Plant Microbe Interac.* (1990) **vol. 3**, pp. 401.

Gressent, F., Rahioui, I. and Rahbé, Y. Characterization of a high affinity binding site for the pea albumin 1b (PA1b) entomotoxin in the weevil *Sitophilus Eur. J. Biochem.* (2003) **vol. 270**, pp. 2429-2435.

Griffiths, D. W. Condensed Tannins In *Toxic substances in crop plants.*, (eds. J. P. F. D'Mello C. M. Duffus and J. H. Duffus) Cambridge: The Royal Society of Chemistry., (1991), pp. 180-201.

Grison, R., Grezes-Besset, B., Schneider, M., Lucante, N., Olsen, L., Leguay, J. J. and Toppan, A. Field tolerance to fungal pathogens of *Brassica napus* constitutively expressing a chimeric chitinase gene. *Nat. Biotechnol.* (1996) **vol. 14**, pp. 643-656.

Grodberg, J. and Dunn, J. J. OmpT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* (1988) **vol. 170**, pp. 1245-1253.

Grossi De Sa, M. F., Mirkov, T. E., Ishimoto, M., Colucci, G., Bateman, K. S. and Chrispeels, M. J. Molecular characterization of a bean alpha-amylase inhibitor that inhibits the alpha-amylase of the Mexican bean weevil Zabrotes subfasciatus *Planta* (1997) **vol. 203**, pp. 295-303.

Halkier, B. A. and Moller, B. L. Biosynthesis of the cyanogenic glucoside dhurrin in seedlings of *Sorghum bicolor* (L.) Moench. and partial purification of the enzyme system involved. *Plant Physiol.* (1989) **vol. 90**, pp. 1552-1559.

Hallahan, D. L., Pickett, J. A., Wadhams, L. J., Wallsgrove, R. M. and Woodcock, C. M. Potential of secondary metabolites in genetic engineering of crops for resistance. In *Plant genetic manipulation for crop protection - Biotechnology in Agriculture.*, (eds. A. M. R. Gatehouse V. A. Hilder and D. Boulter) Wallingford, Oxon, UK: C.A.B. International, (1992) vol. 7, pp. 215-248.

Hamelryck, T. W., Poortmans, F., Goossens, A., Angenon, G., Montagu, M. v., Wyns, L. and Loris, R. Crystal structure of arcelin-5, a lectin-like defense protein from Phaseolus vulgaris *Journal of Biological Chemistry* (1996) **vol. 271**, pp. 32796-32802.

Hammer, R. E., Pursel, V. G., Rexroad, C. E. J., Wall, R. J., Bolt, D. J., Ebert, K. M., Palmiter, R. D. and Brinster, R. L. Production of transgenic rabbits, sheep and pigs by microinjection. *Nature* (1985) **vol. 315**, pp. 680-683.

Hanada, K., Nishiuchi, Y. and Hirano, H. Amino acid residues on the surface of soybean 4-kDa peptide involved in the interaction with its binding protein *Eur. J. Biochem.* (2003) **vol. 270**, pp. 2583-92.

Hancock, K. R., Ealing, P. M. and White, D. W. Identification of sulphur-rich proteins which resist rumen degradation and are hydrolysed rapidly by intestinal proteases *British. J. Nutr.* (1994) **vol. 72**, pp. 855-63.

Hannig, G. and Makrides, S. C. Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends Biotechnol.* (1998) **vol. 16**, pp. 54-60.

Hao, Y. and Ao, G. Transgenic cabbage plants harbouring cowpea trypsin inhibitor (CpTI) gene showed improved resistance to two major insect pests, *Pieris rapae* and *Heliothis armigera*. *FASEB J.* (1997) **vol. 11**, pp. A868.

Harborne, J. B. Introduction to ecological biochemistry. London: Academic Press. (1977) p.

Harborne, J. B. Introduction to ecological biochemistry. New York: Academic Press. (1988) p.

Harborne, J. B., Boulter, D. and Turner, B. L. Chemotaxonomy of the Leguminosae,, (1971) pp. London (GR): Academic Press.

Harborne, J. B. and Ingham, J. L. Biochemical aspects of the coevolution of higher plants with their fungal parasites In *Biochemical aspects of plant and animal coevolution. Annual proc. phytochem. Soc. Europe*, ed. H. J. B. (Ed)): Academic press 436 p., (1978), pp. 343-405.

Harrison, S. J., Marcus, J. P., Goulter, K. C., Green, J. L., Maclean, D. J. and Manners, J. M. An antimicrobial peptide from the australian native Hardenbergia

violacea provides the first functionally characterised member of a subfamilly of plant defensins. *Aust. J. Plant. Physiol.* (1997) **vol. 24**, pp. 571-578.

Hartley, S. E. Plant Interactions with Biotic Factors. In *Encyclopedia of life sciences (www.els.net).*,: Nature Publishing Group, (2001), pp. 1-10.

Hartweck, L. M., Vogelzang, R. D. and Osborn, T. C. Characterization and comparison of arcelin seed protein variants from common bean. *Plant Physiol.* (1991) **vol. 97**, pp. 204-211.

Hayes, R. J., Petty, I. T. D., Coutts, R. H. A. and Buck, K. W. Gene amplification and expression in plants by a replicating geminivirus vector. *Nature* (1988) **vol. 334**, pp. 179-182.

Heitz, A., Hernandez, J. F., Gagnon, J., Hong, T. T., Pham, T. T. C., Nguyen, T. M., Le-Nguyen, D. and Chiche, L. Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. *Biochemistry* (2001) **vol. 40**, pp. 7973-7983.

Helland, R., Berglund, G. I., Otlewski, J., Apostoluk, W., Andersen, O. A., Willassen, N. P. and Smalas, A. O. High-resolution structures of three new trypsin-squash-inhibitor complexes : a detailed comparison with other trypsins and their complexes. *Acta Crystallogr. D. Biol. Crystallogr.* (1999) **vol. 55**, pp. 139-148.

Hendrickson, W. A. and Teeter, A. M. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* (1981) **vol. 290**, pp. 107-113.

Hernan, R. A., Hui, H. L., Andracki, M. E., Noble, R. W., Sligar, S. G., Walder, J. A. and Walder, R. Y. Human hemoglobin expression in *Escherichia coli* : importance of optimal codon usage. *Biochemistry* (1992) **vol. 31**, pp.

Higgins, T. J. V., Chandler, P. M., Randall, P. J., Spencer, D., Beach, L. R., Blagrove, R. J., Kortt, A. A. and Inglis, A. S. Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds *J. Biol. Chem.* (1986) **vol. 261**, pp. 11124-11130.

Hilder, V. A. and Boulter, D. (1999). Genetic engineering of crop plants for insect resistance - a critical review *Crop Protec.* **18**, 177-191.

Hilder, V. A., Gatehouse, A. M. R., Sheerman, S., Barker, R. F. and Boulter, D. A novel mechanism of insect resistance engineered into tobacco *Nature* (1987) **vol. 330**, pp. 160-163.

Hilder, V. A., Powell, K. S., Gatehouse, A. M. R., Gatehouse, J. A., Gatehouse, L. N., Shi, Y., Hamilton, W. D., Merryweather, A., Newell, C. A., Timans, J. C. et al. Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. *Transgenic Res.* (1995) **vol. 4**, pp. 18-25.

Hille, B. Ionic channels of excitable membranes. (1992) p.

Hirsh, A. M. Role of lectins (and rhizobial exopolysaccharides) in legume nodulation. *Curr. Opin. Plant Biol.* (1999) **vol. 2**, pp. 320-326.

Hodges, J. D., Elam, W. W., Watson, W. F. and Nebeker, T. E. Oleoresin characteristics and susceptibility of four southern pines to southern pine beetle (Coleoptera: Scolytidae) attacks. *Can. Entomol.* (1979) **vol. 111**, pp. 889-896.

Hogset, A., Blingsmo, O. R., Saether, O., Gautvik, V. T., Holmgren, E., Hartanis, M., Josephson, S., Gabrielsen, O. S., Gordeladze, J. O., Alestrom, P. et al. Expression and characterization of a recombinant human parathyroid hormone secreted by *Escherichia coli* employing the staphylococcal protein A promoter and signal sequence. *J. Biol. Chem.* (1990) **vol. 265**, pp. 7338-7344.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L. and Conlon, P. J. A short polypeptide marker sequence useful for recombinant protein identification an purification. *Bio/technol.* (1988) **vol. 6**, pp. 1204-1210.

Howell, M. L. and Blumenthal, K. M. Cloning and expression of a synthetic gene for *Cerebratulus lacteus* neurotoxin B-IV. *J. Biol. Chem.* (1989) **vol. 264**, pp. 15268-15273.

Hruska, A. J. Cyanogenic glucosides as defense compounds : a review of the evidence. *J. Chem. Ecol.* (1988) **vol. 14**, pp. 2213-2217.

Hu, J. M., Lavin, M., Wojciechowski, M. F. and Sanderson, M. J. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *TRNK/MATK* sequences and its implications for evolutionary patterns in Papilionoideae. *American J. Botany* (2000) **vol. 87**, pp. 418-430.

Huesing, J. E., Shade, R. E., Chrispeels, M. J. and Murdock, L. L. Alpha-Amylase Inhibitor, Not Phytohemagglutinin, Explains Resistance of Common Bean Seeds to Cowpea Weevil *Plant Physiol.* (1991) **vol. 96**, pp. 993-996.

Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* (1985) **vol. 2**, pp. 13-34.

ILDIS. Legumes of the World. International Legume Database and Information Services., The University of Reading, disponible sur: (2001).

Ilgoutz, S. C., Knittel, N., Lin, J. M., Sterle, S. and Gayler, K. R. Transcription of genes for conglutin gamma and a leginsulin-like protein in narrow-leafed lupin *Plant molec. Biol.* (1997) **vol. 34**, pp. 613-27.

Ishaaya, I., Birk, Y., Bondi, A. and Tencer, Y. Soybean saponins. IX. Studies of their effects on birds, mammals and cold-blooded organisms. *J. Sci. Food Agric.* (1969) **vol. 20**, pp. 433-436.

Ishimoto, M. and Kitamura, K. Growth inhibitory effects of an alpha-amylase inhibitor from kidney bean, *Phaseolus vulgaris* (L.) on three species of bruchids (Coleoptera: Bruchidae). *Appl. Entomol. Zool.* (1989) **vol. 24**, pp. 281-286.

Isman, M. B. and Duffey, S. S. Toxicity of tomato phenolic compounds to the fruitworm *Heliothis zea*. *Entomol. Exp. Appl.* (1982) **vol. 31**, pp. 370-376.

Iulek, J., Franco, O. L., Silva, M., Slivinski, C. T., Bloch, C., Rigden, D. J. and de Sa, M. F. G. Purification, biochemical characterisation and partial primary structure of a new alpha-amylase inhibitor from Secale cereale (rye) *Inter. J. Biochem. Cell Physiol.* (2000) **vol. 32**, pp. 1195-1204.

Jach, G., Gornhardt, B., Mundy, J., Logemann, J., Pinsdorf, P., Leah, R., Schell, J. and Maas, C. Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco *Plant J.* (1995) **vol. 8**, pp. 97-109.

Janssen, B. J., Schirra, H. J., Lay, F. T., Anderson, M. A. and Craik, D. J. Structure of Petunia hybrida defensin 1, a novel plant defensin with five disulfide bonds. *Biochemistry* (2003) **vol. 42**, pp. 8214-8222.

Janzen, D. H., Juster, H. B. and Liener, I. E. Insecticidal action of the phytohemagglutinin in black beans on a bruchid beetle *Science* (1976) **vol. 192**, pp. 795-796.

Jauregui Adell, J. and Marti, J. Acidic cleavage of the aspartyl-proline bond and the limitations of the reaction. *Anal. Biochem.* (1975) **vol. 69**, pp. 468-473.

Jennings, C., West, J., Waine, C., Craik, D. and Andersen, M. Biosynthesis and insecticidal properties of plant cyclotides : the cyclic knotted proteins from *Oldendia affinis*. *Proc. Natl. Acad. Sci. U.S.A.* (2001) **vol. 98**, pp. 10614-10619.

Johnson, R., Narvaez, J., An, G. H. and Ryan, C. A. Expression of Proteinase Inhibitor-I and Inhibitor-II in Transgenic Tobacco Plants - Effects on Natural Defense Against Manduca-Sexta Larvae *Proc. Natl. Acad. Sci. USA* (1989) **vol. 86**, pp. 9871-9875.

Jones, B. L., Lookhart, G. L. and Johnson, D. E. Improved separation and toxicity analysis methods for purothionins. *Cereal Chem.* (1985) **vol. 62**, pp. 327-331.

Jongsma, M. A., Bakker, P. L., Peters, J., Bosch, D. and Stiekema, W. J. Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition *Proc. Natl. Acad. Sci. USA* (1995) **vol. 92**, pp. 8041-8045.

Jongsma, M. A. and Bolter, C. The adaptation of insects to plant protease inhibitors. *J. Insect Physiol.* (1997) **vol. 43**, pp. 885-895.

Jouanin, L., Bonadé-Bottino, M., Girard, C., Morrot, G. and Giband, M. (1998). Transgenic plants for insect resistance *Plant Sci.* **131**, 1-11.

Jouvensal, L., Quillien, L., Ferrasson, E., Rahbé, Y., Gueguen, J. and Vovelle, F. PA1b, an insecticidal protein extracted from pea seeds (*Pisum sativum*): (1)H-2D-NMR study and molecular modeling *Biochemistry* (2003) **vol. 42**, pp. 11915-11923.

Kader, J. C. Lipid-transfer proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* (1996) **vol. 47**, pp. 627-654.

Kadziola, A., Sogaard, M., Svensson, B. and Haser, R. Molecular structure of a barley alpha-amylase inhibitor complex : implications for starch binding and catalysis. *J. Mol. Biol.* (1998) **vol. 278**, pp. 205-217.

Kagan, B. L., Selsted, M. E., Ganz, T. and Lehrer, R. Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. *PNAS* (1990) **vol. 87**, pp. 210-214.

Kane, J. F., Violand, B. N., Curran, D. F., Staten, N. R., Duffin, K. L. and Bogosian, G. Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*. *Nucleic Acids Res.* (1992) **vol. 20**, pp. 6707-6712.

Kass, E. and Wink, M. Phylogenetic relationships in the Papilionoideae (family Leguminosae) based on nucleotide sequences of cpDNA (*rbc*L) and ncDNA (ITS1 and 2). *Mol. Phylogen. Evolution* (1997) **vol. 8**, pp. 65-88.

Kaufman, F. J. Vectors used for expression in Mammalian cells. *Methods Enzymol.* (1991) **vol. 185**, pp. 487-511.

Kawakami, T., Kamo, M., Takamoto, K., Miyazaki, K., Chow, L. P., Ueno, Y. and Tsugita, A. Bond-specific chemical cleavages of peptides and proteins with perfluoric acid vapors : novel peptide bond cleavages of glycyl-threonine, the amino side of serine residues and the carboxyl side of aspartic acid residues. *J. Biochem.* (1997) **vol. 121**, pp. 68-76.

Kawasaki, S. and Murakami, Y. Genome analysis of Lotus japonicus *J. Plant Res.* (2000) **vol. 113**, pp. 497-506.

Kempe, T., Kent, S. B., Chow, F., Peterson, S. M., Sundquist, W. I., L'Italien, J. J., Harbrecht, D., Plunkett, D. and DeLorbe, W. J. Multiple-copy genes : production and modification of monomeric peptides from large multimeric fusion proteins. *Gene* (1985) **vol. 39**, pp. 239-245.

Kimura, M., Ikeda, T., Fukumoto, D., Yamasaki, N. and Yonekura, M. Primary structure of a cysteine proteinase inhibitor from the fruit of avocado (*Persea americana* Mill). *Biosc. Biotechnol. Biochem.* (1995) **vol. 59**, pp. 2328-2329.

Kitagawa, M. and Tomiyama, T. A new amino compound in the Jack bean and a corresponding new enzyme. *J. Biochem. (Tokyo)* (1929) **vol. 11**, pp. 265-271.

Klein, T. M., Roth, B. A. and Fromm, M. E. Regulation of anthocyanin genes introduced into intact maize tissues by microprojectiles. *Proc. Natl. Acad.Sci.* (1989) **vol. 86**, pp. 6681-6685.

Klopfenstein, N. B., Allen, K. K., Avila, F. J., Heuchelin, S. A., Martinez, J., Carman, R. C., Hall, R. B., Hart, E. R. and McNabb, H. S. Proteinase inhibitor II gene in transgenic poplar : chemical and biological assays. *Biomass Bioenergy* (1997) **vol. 12**, pp. 299-311.

Kobayashi, Y., Takashima, H., Tamaoki, H., Kyogoku, Y., Lambert, P., Kuroda, H. and *al*, *e.* The cysteine stabilized alpha-helix : a common structural motif of ion channel blocking neurotoxic peptides. *Biopolymers* (1991) **vol. 31**, pp. 1213-1220.

Kogan, M. Plant resistance in pest management. In *Introduction to insect pest management.*, ed. W. a. Sons) New York:, (1975), pp. 103-146.

Kohno, T., Carmichael, D. F., Sommer, A. and Thompson, R. C. Refolding of recombinant proteins. *Methods Enzymol.* (1991) **vol. 185**, pp. 187-195.

Kouzuma, Y., Inanaga, H., Doi Kawano, K., Yamazaki, N. and Kimura, M. Molecular cloning and functional expression of cDNA encoding the cysteine proteinase inhibitor with three cystatin domains from sunflower seeds. *J. Biochem.* (2000) **vol. 128**, pp. 161-166.

Kramer, K. J., Klassen, L. W., Jones, B. L., Speirs, R. D. and Kammer, A. E. Toxicity of purothionin and its homologues to the tobacco hornworm, *Manduca sexta* (L.) (Lepidoptera: Sphingidae). *Toxicology and applied pharmacology* (1979) **vol. 48**, pp. 179-183.

Krimpenfort, P., Rademakers, A., Eyestone, W., Van der Schans, A., Van der Broek, S., Kooiman, P., Kootwijk, E., Platenburg, G., Pieper, F., Strijker, R. et al. Generation of transgenic dairy cattle using *in vitro* embryo production. *Bio/technol.* (1991) **vol. 9**, pp. 844-847.

Kuliopulos, A. and Walsh, C. T. Production, purification and cleavage of tandem repeats of recombinant peptides. *J. Am. Chem. Soc.* (1994) **vol. 116**, pp. 4599-4607.

Kushmerick, C., De Souza Castro, M., Cruz, J. S., Bloch, C. J. and Beirao, P. S. L. Functional and structural features of gamma-zeathionins, a new class of sodium channels blockers. *FEBS Letters* (1998) **vol. 440**, pp. 302-306.

Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4 *Nature* (1970) **vol. 227**, pp. 680-685.

Lamblin, A. F., Crow, J. A., Johnson, J. E., Silverstein, K. A., Kunau, T. M., Kilian, A., Benz, D., Stromvik, M., Endre, G., VandenBosch, K. A. et al. MtDB: a database for personalized data mining of the model legume Medicago truncatula transcriptome *Nucleic Acids Res.* (2003) **vol. 31**, pp. 196-201.

Landon, M. Cleavage at Aspartyl-Prolyl Bonds. *Methods Enzymol.* (1977) **vol. 47**, pp. 145-149.

Lane, G. A., Biggs, D. R., Sutherland, O. W. R., Williams, E. M., Maindonald, J. M. and Donnell, D. J. Isoflavonoid feeding deterrents for *Costelytra zealandica*, Structure activity relationships. *J. Chem. Ecol.* (1985) **vol. 11**, pp. 1713-1735.

Laskowski, M. J. and Kato, I. Protein inhibitors of proteinases. *Annu. Rev. Biochem.* (1980) **vol. 49**, pp. 593-626.

Lauber, T., Marx, U., Schulz, A., Kreutzmann, P., Rosch, P. and Hoffmann, S. Accurate disulfide formation in *Escherichia coli* : overexpression and characterization of the first domain (HF6478) of the multiple Kazal-type inhibitor LEKTI. *Protein Expr. Purif.* (2001) **vol. 22**, pp. 108-112.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio/technol.* (1993) **vol. 11**, pp. 187-193.

LaVallie, E. R. and McCoy, J. M. Gene fusion expression systems in *Escherichia coli*. *Curr. Opin. Biotechnol.* (1995) **vol. 6**, pp. 501-506.

Lavin, M., Pennington, R. T., Klitgaard, B. B., Sprent, J. I., Cavacante de Lima, H. and Gasson, P. E. The dalbergioid legumes (Fabaceae) : delimitation of a pantropical monophyletic clade. *American J. Botany* (2001) **vol. 88**, pp. 503-533.

Laviolette, P. and Nardon, P. Action des Rayons gamma du Cobalt 60 sur la mortalité et la fertilité des adulte d'un charançon du riz *Biol. Bull.* (1963) **vol. XCVII**, pp. 305-333.

Lay, F. T., Schirra, H. J., Scanlon, M. J., Anderson, M. A. and Craik, D. J. the three-dimensional solution structure of NaD1, a new floral defensin from Nicotiana alata and its application to a homology model of the crop defense protein alfAFP. *J. Mol. Biol.* (2003) **vol. 325**, pp. 175-188.

Le Nguyen, D., Heitz, A., Chiche, L., Castro, B., Boigegrain, R. A., Favel, A. and Coletti-Previero, M. A. Molecular recognition between serine proteases and new bioactive microproteins with a knotted structure. *Biochimie* (1990) **vol. 72**, pp.

Lee, S. C. and Olins, P. O. Effect of overproduction of heat shock chaperones GroESL and DnaK on human procollagenase production on *Escherichia coli*. *J. Biol. Chem.* (1992) **vol. 267**, pp. 2849-2852.

Leplé, J. C., Bonadé-Bottino, M., Augustin, S., Pilate, G., Letan, V. D., Delplanque, A., Cornu, D. and Jouanin, L. Toxicity to *Chrysomela tremulae* (Coleoptera: chrysomelidae) of transgenic poplars expressing a cysteine proteinase inhibitor *Mol. Breed.* (1995) **vol. 1**, pp. 319-328.

Li, S. S., Gullbo, J., lindholm, P., Larsson, R., Thunberg, E., Samuelsson, G., Bohlin, L. and Cleson, P. Ligatoxin B, a new cytotoxic protein with a novel helix-turn-helix DNA-binding domain from the mistletoe *Phoradendron liga*. *Biochem. J.* (2002) **vol. 366**, pp. 405-413.

Liang, C., Brookhart, G., Feng, G. H., Reeck, G. R. and Kramer, K. J. Inhibition of Digestive Proteinases of Stored Grain Coleoptera by Oryzacystatin, a Cysteine Proteinase Inhibitor from Rice Seed *FEBS Letters* (1991) **vol. 278**, pp. 139-142.

Lilie, H., Schwarz, E. and Rudolph, R. Advances in refolding of proteins produced in *E. coli*. *Curr. Opin. Biotechnol.* (1998) **vol. 9**, pp. 497-501.

Louda, S. and Mole, S. Glucosinolates: chemistry and ecology. In *Herbivores: their interactions with secondary plant metabolites*., (eds. G. A. Rosenthal and M. R. Berenbaum) New York: Academic press, (1991) vol. I, pp. 124-164.

Luckow, V. A. and Summers, M. D. Signals important for high-level expression of foreign genes in *Autographa californica* nuclear polyhedrosis virus expression vectors. *Virology* (1988) **vol. 167**, pp. 56-71.

Makoff, A., Parry, N. and Dicken, L. Translational fusions with fragments of the trpE gene improve the expression of a poorly expressed heterologous gene in *Escherichia coli*. *J. Gen. Microbiol.* (1989) **vol. 135**, pp. 11-24.

Makrides, S. C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* (1996) **vol. 60**, pp. 512-538.

Malcom, S. B. and Zalucki, M. P. Milkweed latex and cardenolide induction may resolve the lethal plant defense paradox. *Entomol. Exp. Appl.* (1996) **vol. 80**, pp. 193-196.

Marchetti, S., Delledonne, M., Fogher, C., Chiaba, C., Chiesa, F., Savazzini, F. and Giordano, A. Soybean Kunitz, C-II and PI-IV inhibitor genes confer different levels of insect resistance to tobacco and potato transgenic plants *Theor. Appl. Genet.* (2000) **vol. 101**, pp. 519-526.

Marshall, J. J. and Lauda, C. M. Purification and properties of phaseolamin, an inhibitor of alpha-amylase, from the kidney bean *Phaseolus vulgaris*. *J. Biol. Chem.* (1975) **vol. 250**, pp. 8030-8037.

Martin, S. L., Vrhovski, B. and Weiss, A. S. Total synthesis and expression in *Escherichia coli* of a gene encoding human tropoelastin. *Gene* (1995) **vol. 154**, pp. 159-166.

Maruniak, J. E., Fiesler, S. E. and Mcguire, P. M. Susceptibility of Insect Cells and Ribosomes to Ricin *Comp. Biochem. Physiol.* (1990) **vol. 96B**, pp. 543-548.

Maskos, K., Huber Wunderlich, M. and Glockshuber, R. RBI, a one domain alpha-amylase/trypsin inhibitor with completely independent binding sites. *FEBS Letters* (1996) **vol. 397**, pp. 11-16.

Maskos, K., Huber Wunderlich, M. and Glockshuber, R. DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns *in vitro* and *in vivo*. *J. Mol. Biol.* (2003) **vol. 325**, pp. 495-513.

Masson, J. M. Surexprimer chez *Escherichia coli*. In *Expression des Proteines Recombinantes*,: Les cahiers IMABIO, (1993) vol. 9, pp. 5-12.

Mauch, F., Mauch-Mani, B. and Boller, T. Antifungal hydrolases in pea tissue. II. Inhibition of fungal growth by combinations of chitinase and B-1,3-glucanase *Plant Physiol.* (1988) **vol. 88**, pp. 936-942.

Mazumdar Leighton, S. and Broadway, R. M. Transcriptional induction of diverse midgut trypsins in larval *Agrotis ipsilon* and *Helicoverpa zea* feeding on the soybean trypsin inhibitor. *Insect Biochem. Mol. Biol.* (2001) **vol. 31**, pp. 645-657.

Mbaiguinam, M. Détermination des causes de la résistance des légumineuses aux charançons des céréales du genre *Sitophilus*, *PhD,* Lyon: Univ. Claude Bernard Lyon1. (1996) 1-164 p.

McDonald, N. Q. and Hendrickson, W. A. A structural superfamily of growth factors containing a cystine-knot motif. *Cell* (1993) **vol. 73**, pp. 421-424.

Meisner, J., Navon, A., Zur, M. and Ascher, K. R. S. The response of *Spodoptera littoralis* larvae to gossypol incorporated in an artificial diet. *Environ. Entomol.* (1977) **vol. 6**, pp. 243-244.

Melo, F. R., Ridgen, D. J., Franco, O. L., Mello, L. V., Ary, M. B., Grossi de Sa, M. F. and Bloch, C. J. Inhibition of trypsin by cowpea thionin : characterization, molecular modeling and docking. *Proteins* (2002) **vol. 48**, pp. 311-319.

Melo, F. R., Sales, M. P., Silva, L. S., Franco, O. L., Bloch, C. J. and Ary, M. B. Alpha-amylases from cowpea seeds. *Prot. Pept. Lett.* (1999) **vol. 6**, pp. 387-392.

Mendez, E., Moreno, A., Colilla, F., Pelaez, F., Limas, G. G., R., M., Soriano, F., Salinas, M. and de Haro, C. Primary structure and inhibition of protein synthesis in

eukaryotic cell-free system of a novel thionin, gamma-hordothionin, from barley endosperm. *Eur. J. Biochem.* (1990) **vol. 194**, pp. 533-539.

Mendez, E., Rocher, A., Calero, M., Girbes, T., Citores, L. and Soriano, F. Primary structure of omega-hordothionin, a member of a novel family of thionins from barley endosperm, and its inhibition of protein synthesis in eukaryotic and prokaryotic cell-free systems. *Eur. J. Biochem.* (1996) **vol. 239**, pp. 67-73.

Metcalf, R. L., Rhodes, A. M., Metcalf, R. A., Ferguson, J., Metcalf, E. R. and Lu, P. Cucurbitacin contents and diabroticite (Coleoptera: Chrysomelidae) feeding upon *Cucurbita* spp. *Environ. Entomol.* (1982) **vol. 11**, pp. 931-937.

Mikolajczak, K. L., Madrigal, R. V., Smith, J. C. R. and Reed, D. K. Insecticidal effects of cyanolipids on three species of stored product insects, European corn borer (Lepidoptera: Pyralidae) larvae, and striped cucumber beetle (Coleoptera: Chrysomelidae). *J. Econ. Entomol.* (1984) **vol. 77**, pp. 1144-1148.

Minney, B. H. P., Gatehouse, A. M. R., Dobie, P., Dendy, J., Cardona, C. and Gatehouse, J. A. Biochemical Bases of Seed Resistance to Zabrotes-Subfasciatus (Bean Weevil) in Phaseolus-Vulgaris (Common Bean) - A Mechanism for Arcelin Toxicity *J. Insect Physiol.* (1990) **vol. 36**, pp. 757.

Mirkov, T. E., Wahlstrom, J. M., Hagiwara, K., Finardifilho, F., Kjemtrup, S. and Chrispeels, M. J. Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-alpha-amylase inhibitor family of the common bean and its relatives *Plant molec. Biol.* (1994) **vol. 26**, pp. 1103-1113.

Moffatt, B. A. and Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell.* (1987) **vol. 49**, pp. 221-227.

Molina, A., Segura, A. and Garcia-Olmedo, F. Lipid transfer proteins (nsLTPs) from barley and maize leaves are potent inhibitors of bacterial and fungal plant pathogens. *FEBS Lett.* (1993) **vol. 316**, pp. 119-122.

Montecucco, C. Protein toxins and membrane transport. *Curr. Opin. Cell Biol.* (1998) **vol. 10**, pp. 530-536.

Morassutti, C., De Amicis, F., Skerlavaj, B., Zanetti, M. and Marchetti, S. Production of a recombinant antimicrobial peptide in transgenic plants using a modified VMA intein expression system. *FEBS Lett.* (2002) **vol. 519**, pp. 141-146.

Moreno, J. and Chrispeels, M. J. A lectin gene encodes the Alpha-amylase inhibitor of the common bean *Proc. Natl. Acad. Sci. USA* (1989) **vol. 86**, pp. 7885-7889.

Morton, R. L., Schroeder, H. E., Bateman, K. S., Chrispeels, M. J., Armstrong, E. and Higgins, T. J. V. Bean alpha -amylase inhibitor 1 in transgenic peas (*Pisum sativum*) provides complete protection from pea weevil (*Bruchus pisorum*) under field conditions. *Proc. Natl. Acad. Sci. USA* (2000) **vol. 97**, pp. 3820-3825.

Mukhopadhyay, A. Inclusion bodies and purification of proteins in biologically active forms. *Adv. Biochem. Engin. Biotechnol.* (1997) **vol. 56**, pp. 61-109.

Mundy, J., Svendsen, I. and Hejgaard, J. Barley alpha-amylase/subtilisin inhibitor. Isolation and characterization. *Carlsberg Res. Commun.* (1983) **vol. 48**, pp. 81-90.

Murby, M., Uhlen, M. and Stahl, S. Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*. *Protein Expr. Purif.* (1996) **vol. 7**, pp. 129-136.

Murhammer, D. W. Review and patents and literature. The use of insect cell cultures for recombinant protein synthesis : Engineering aspects. *Appl. Biochem. Biotechnol.* (1991) **vol. 31**, pp. 283-310.

Murray, R. D. H., Mendez, J. and Brown, S. A. The natural coumarins. Chichester, UK: John Wiley and Sons. (1982) p.

Muslin, E. H., Kanikula, A. M., Clark, S. E. and Henson, C. A. Overexpression, purification, and characterization of a barley alpha-glucosidase secreted by *Pichia pastoris*. *Protein Expr. Purif.* (2000) **vol. 18**, pp. 20-26.

Nagata, K., Kudo, N., Abe, K., Arai, S. and Tanokura, M. Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. japonica *Biochemistry* (2000) **vol. 39**, pp. 14753-60.

Nahoum, V., Farisei, F., Le-Berre-Anton, V., Egloff, M. P., Rouge, P., Poerio, E. and Payan, F. A plant seed inhibitor of two classes of alpha-amylases : X-ray analysis of *Tenebrio molitor* larvae alpha-amylase in complex with the bean *Phaseolus vulgaris* inhibitor. *Acta Biol. Crystal. D.* (1998) **vol. 55**, pp. 360-362.

Nakashima, M., Enomoto, K., Kijima, H. and Morita, H. Discrepency between the affinities of certain inhibitors for the membrane-bound alpha-glucosidases and those for the sugar receptor of flies. *Insect Biochem.* (1982) **vol. 12**, pp. 579-585.

Neupane, F. P. and Norris, D. M. Indo-acetic acid alteration of soybean resistance to the cabbage looper (Lepidoptera: Noctuidae). *Environ. Entomol.* (1990) **vol. 19**, pp. 215-221.

Nilsson, B. and Abrahmsen, L. Fusions to staphylococcal protein A. *Methods Enzymol.* (1990) **vol. 185**, pp. 144-161.

Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. and Yura, T. Chaperone coexpression plasmids : differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* (1998) **vol. 64**, pp. 1694-1699.

Nishizawa, N. K., Mori, S., Kajiwara, H., Komatsu, S. and Hirano, H. Subcellular localization of leginsulin in the immature seeds of soybean. *Plant Cell Physiol.* (1995) **vol. 36**, pp. 42.

Oberg, K., Chrunyk, B. A., Wetzel, R. and Fink, A. L. Nativelike secondary structure in interleukin-1 beta inclusion bodies by attenuated total reflectance FTIR. *Biochemistry* (1994) **vol. 33**, pp. 2628-2634.

Oda, Y., Matsunaga, T., Fukuyiama, K., Miyasaki, T. and Morimoto, J. T. Tertiary and quaternary structures of 0.19 alpha-amylase inhibitor from wheat kernel determined by X-ray analysis at 2.06 A resolution. *Biochemistry* (1997) **vol. 36**, pp. 13503-13511.

Odani, S., Koide, T., Ono, T., Seto, Y., Tanaka, T. Soybean hydrophobic protein. Isolation, partial characterization and the complete primary structure. *Eur. J. Biochem.* (1987) **vol.162** (3), pp. 485-491

Olsnes, S., Refsnes, K. and Pihl, A. Mechanism of action of the toxic lectins abrin and ricin. *Nature* (1974) **vol. 249**, pp. 627-631.

Olsnes, S., Wesche, J. and Falnes, P. Binding, uptake, routing and translocation of toxins with intracellular sites of action. In *The comprehensive Sourcebook of bacterial protein toxins*., (eds. J. E. Alouf and J. H. Freer) London: Academic Press, (1999), pp. 73-93.

Onesti, S., Brick, P. and Blow, D. M. Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds. *J. Mol. Biol.* (1991) **vol. 217**, pp. 153-176.

Oppert, B., Morgan, T. D., Culbertson, C. and Kramer, K. J. Dietary mixtures of cysteine and serine proteinase inhibitors exhibit synergistic toxicity toward the red flour beetle, Tribolium castaneum *Comp. Biochem. Physiol.* (1993) **vol. 105C**, pp. 379-385.

Orr, G. L., Strickland, J. A. and Walsh, T. A. Inhibition of Diabrotica larval growth by a multicystatin from potato tubers *J. Insect Physiol.* (1994) **vol. 40**, pp. 893-900.

Osborn, R. W., De Samblanx, G. W., Thevissen, K., Goderis, I. J., Torrekens, S., Van Leuven, F., Attenborough, S., Rees, S. B. and Broekaert, W. F. Isolation and characterisation of plant defensins from seeds of Asteraceae, Fabaceae, Hippocastanaceae and Saxifragaceae. *FEBS Letters* (1995) **vol. 368**, pp. 257-262.

Osborn, T. C., Alexander, D. C., Sun, S. S., Cardona, C. and Bliss, F. A. Insecticidal activity and lectin homology of arcelin seed protein *Science* (1988) **vol. 240**, pp. 207.

Osorio e Castro, V. R. and Vernon, L. P. Hemolytic activity of thionin from Pyrularia pubera nuts and snake venom toxins of Naja naja species : Pyrularia thionin and snake venom cardiotoxin compete for the same membrane site. *Toxicon* (1989) **vol. 27**, pp. 511-517.

Paes, N. S., Gerhardt, I. R., Coutinho, M. V., Yokoyama, M., Santana, E., Harris, N., Chrispeels, M. J. and de Sa, M. F. G. The effect of arcelin-1 on the structure of the midgut of bruchid larvae and immunolocalization of the arcelin protein *J. Insect Physiol.* (2000) **vol. 46**, pp. 393-402.

Page, D., Aubert, G., Duc, G., Welham, T. and Domoney, C. Combinatorial variation in coding and promoter sequences of genes at the Tri locus in Pisum sativum accounts for variation in trypsin inhibitor activity in seeds *Mol. Genet. Genomics* (2002) **vol. 267**, pp. 359-369.

Page, D., Delclos, B., Aubert, G., Bonavent, J. F. and Mousset-Declas, C. Sclerotinia rot resistance in red clover: identification of RAPD markers using bulked segregant analysis *Plant Breed.* (1997) **vol. 116**, pp. 73-78.

Painter, R. H. Insect resistance in crop plants.: University of Kansas Press. (1951) 520 p.

Pallaghy, P. K., Nielsen, K. J., Craik, D. J. and Norton, R. S. A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet in toxic and inhibitory polypeptides *Prot. Sci.* (1994) **vol. 3**, pp. 1833-9.

Panda, N. and Khush, G. S. Secondary Plant Metabolites for Insect Resistance. In *Host Plant Resistance to Insects.*, Wallingford, UK: CAB International, (1995), pp. 22-66.

Papastoitsis, G. and Wilson, K. A. Initiation of the degradation of the soybean Kunitz and Bowman-Birk inhibitors by cysteine protease. *Plant Physiol.* (1990) **vol. 96**, pp. 11086-11092.

Pedersen, S. *Escherichia coli* ribosomes translate *in vivo* with variable rate. *EMBO J.* (1984) **vol. 3**, pp. 2894-2898.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L. et al. Cloning and expression of human tissue-type plasminogene activator cDNA in *E. coli*. *Nature* (1983) **vol. 301**, pp. 214-221.

Pereira, P. J. B., Lozanov, V., Patthy, A., Huber, R., Bode, W., Pongor, S. and Strobl, S. Specific inhibition of insect alpha-amylases: yellow meal worm alpha-amylase in complex with the Amaranth alpha-amylase inhibitor at 2.0 A resolution *Structure* (1999) **vol. 7**, pp. 1079-1088.

Pernas, M., Sanchezmonge, R., Gomez, L. and Salcedo, G. A chestnut seed cystatin differentially effective against cysteine proteinases from closely related pests *Plant molec. Biol.* (1998) **vol. 38**, pp. 1235-1242.

Perrin, D. R. and Cruickshank, I. A. M. Studies on the phytoalexins. VII. Chemical stimulation of pisatin formation in *Pisum sativum*. *Aust. J. Biol. Sci.* (1965) **vol. 18**, pp. 803-816.

Peumans, W. J. and Van Damme, E. J. M. Lectins as plant defense proteins *Plant Physiol.* (1995) **vol. 109**, pp. 347-352.

Philips, J. A., Lopez, A. F., Milton, S. E., Vadas, M. A. and Shannon, M. F. Synthesis and expression of the gene encoding human interleukin-3. *Gene* (1989) **vol. 84**, pp. 501-507.

Pillemer, E. A. and Tingey, W. M. Hooked trichomes and resistance of *Phaseolus vulgaris* to *Empoasca fabae* (Harris). *Entomol. Exp. Appl.* (1978) **vol. 24**, pp. 83-94.

Pineiro, M., Diaz, I., Rodriguez Palenzuala, P., Titarenko, E. and Garcia Olmedo, F. Selective disulfide linkage of plant thionins with other proteins. *FEBS Lett.* (1995) **vol. 369**, pp. 239-242.

Polhill, R. M. Papilionoideae. In *Advances in legume systematics.*, (eds. R. M. Polhill and J. A. Raven) Kew: Royal Botanical Gardens, (1981) vol. 1, pp. 192-208.

Polhill, R. M. Classification of the Leguminosae. In *Phytochemical dictionary of the Leguminosae.*, (eds. F. A. Bisby J. Buckingham and J. B. Harborne): Chapman and Hall, (1994) vol. 1, pp. xxxv-lvii.

Poulsen, K., Fraser, K. J. and Haber, E. An active derivative of rabbit antibody light chain composed of the constant and the variable domains held together only by a native disulfide bond. *Proc. Natl. Acad. Sci. USA* (1972) **vol. 69**, pp. 2495-2499.

Poulton, J. E. Cyanogenesis in plants. *Plant Physiol.* (1990) **vol. 94**, pp. 401-405.

Powell, K. S., Gatehouse, A. M. R., Hilder, V. A. and Gatehouse, J. A. Antimetabolic effects of plants lectins and plant and fungal enzymes on the nymphal stages of two importante rice pests, Nilaparvata lugens and Nephotettix nigropictus. *Entomol. Exp. Appl.* (1993) **vol. 66**, pp. 119-126.

Poznanski, J., Sodano, P., Won Suh, S., Young Lee, J., Ptak, M. and Vovelle, F. Solution structure of a lipid transfer protein extracted from rice seeds. *Eur. J. Biochem.* (1999) **vol. 259**, pp. 692-708.

Prescott, Harley and Klein. Microbiologie. Bruxelles: De Boeck-Wesmael. (1995) 1014 p.

Prinz, W. A., Aslund, F., Holmgren, A. and Beckwith, J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* (1997) **vol. 272**, pp. 15661-15667.

Prodromou, C. and Pearl, L. H. Recursive PCR : a novel technique for total gene synthesis. *Protein Eng.* (1992) **vol. 5**, pp. 827-829.

Pusztai, A., Bardocz, G. G. S., Alonso, R., Chrispeels, M. J., Schroeder, H. E., Tabe, L. M. and Higgins, T. J. V. Expression of the insecticidal bean alpha-amylase inhibitor transgene has minimal detrimental effect on the nutritional value of peas fed to rats at 30% of the diet *J. Nutr.* (1999) **vol. 129**, pp. 1597-1603.

Qiu, J., Swartz, J. and Georgiou, G. Expression of active human tissue type plasminogen activator in *Escherichia coli*. *Appl. Environ. Microbiol.* (1998) **vol. 64**, pp. 4891-4896.

Rahbé, Y., Deraison, C., Bonadé-Bottino, M., Girard, C., Nardon, C. and Jouanin, L. Effects of the cysteine protease inhibitor oryzacystatin (OC-I) on different aphids and reduced performance of *Myzus persicae* on OC-I expressing transgenic oilseed rape *Plant Sci*. (2003) **vol. 164**, pp. 441-450.

Rahbé, Y., Febvay, G. and Kermarrec, A. Foraging activity of the attine ant Acromyrmex octospinosus (Reich) (Hymenoptera: Formicidae) on resistant and susceptible yam varieties *Bull. entomol. Res.* (1988) **vol. 78**, pp. 329-337.

Rahbé, Y., Sauvion, N., Febvay, G., Peumans, W. J. and Gatehouse, A. M. R. Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum Entomol. Exp. Appl.* (1995) **vol. 76**, pp. 143-155.

Rao, K. V., Rathova, K. S., Hodges, T. K., Fu, X., Stoger, E., Sudhakar, D., Williams, S., Christou, P., Bharathi, M., Bown, D. P. et al. Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper *Plant J.* (1998) **vol. 15**, pp. 469-477.

Raubenheimer, D. Cyanogenesis and the feeding preference of *Acraea horta* (L) (Lepidoptera : Acraeinae)., *M.S thesis,*: University of Capetown. (1987) p.

Raven, J. A. Evolution of vascular land plants. *Adv. Bot. res.* (1977) **vol. 154**, pp. 129.

Rees, D. C. and Lipscomb, W. N. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 A resolution. *J. Mol. Biol.* (1982) **vol. 160**, pp. 475-98.

Renz, J. Ueber das Mimosin. *Z. Physiol. Chem.* (1936) **vol. 244**, pp. 153-158.

Rice, E. L. Allelopathy. New York: Academic Press. (1974) p.

Richardson, M. Seed Storage Proteins - The Enzyme Inhibitors In *Methods in Plant Biochemistry*, New York: Academic Press, (1991) vol. 5, pp. 259-305.

Richter, G. Isoprénoïdes. In *Métabolisme des Végétaux.*, ed. G. Richter): Presses Polytechniques et Universitaires Romandes., (1993), pp. 287-315.

Rodriguez, E., Mullaney, E. J. and Lei, X. G. Expression of the *Aspergillus fumigatus* phytase gene in *Pichia pastoris* and characterization of recombinant enzyme. *Biochem. Biophys. Res. Com.* (2000) **vol. 268**, pp. 373-378.

Rogers, B. L., Pollock, J., Klapper, G. and Griffith, I. J. Sequence of the proteinase-inhibitor cystatin homologue from the pollen of *Ambrosia artemisiifolia* (short ragweed). *Gene* (1993) **vol. 133**, pp. 219-221.

Romagnoli, S., Ugolini, R., Fogolari, F., Schaller, G., Urech, K., Giannattasio, M., Ragona, L. and Molinari, H. NMR structural determination of viscotoxin A3 from Viscum album L. (To be published) **vol.**, pp.

Rosenthal, G. A. Plant non-protein amino and imino acids., New York: Academic Press, (1982a) vol. chapter 3, pp. 57.

Rosenthal, G. A. Secondary plant metabolites-round table discussion In *5th International Symposium on Insect Plant Relationships.*, eds. J. H. Visser and A. K. Minks, (1982b) pp. 331-334. Wageningen: Centre for Agricultural Publishing and Documentation.

Rosenthal, G. A. Non protein amino acids as protective allelochemicals. In *Herbivores: their interactions with secondary plant metabolites.*, (eds. G. A. Rosenthal and M. R. Berenbaum) New york: Academic Press, (1991) vol. I, pp. 1-34.

Rousseau Limouzin, M. and Fritig, B. Induction of chitinases, 1,3-beta-glucanases and other pathogenesis-related proteins in sugar beet leaves upon infection with *Cercospora beticola*. *Plant Physiol. Biochem.* (1991) **vol. 29**, pp. 105.

Rutenber, E. and Robertus, J. D. Structure of ricin B chain at 2,5 A resolution *Proteins* (1991) **vol. 10**, pp. 260-269.

Ryan, C. A. Protease inhibitors in plants: genes for improving defenses against insect and pathogens *Annu. Rev. Phytopathol.* (1990) **vol. 28**, pp. 425-449.

Sales, M. P., Gerhardt, I. R., Grossi de Sa, F. and XavierFilho, J. Do legume storage proteins play a role in defending seeds against bruchids? *Plant Physiol.* (2000) **vol. 124**, pp. 515-22.

Salzman, R. A., Fujita, T., Zhu-Salzman, K., Hasegawa, P. M. and Bressan, R. A. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates *Plant Molecular Biology Reporter* (1999) **vol. 17**, pp. 11-17.

Samac, D. A. and Shah, D. M. Developmental and Pathogen-Induced Activation of the *Arabidopsis* Acidic Chitinase Promoter *Plant Cell* (1991) **vol. 3**, pp. 1063-1072.

Samal, B. B., Arakawa, T., Boone, T. C., Jones, T., Prestrelski, S. J., Narhi, L. O., Wen, J., Stearns, G. W., Crandall, C. A., Pope, J. et al. High level expression of human leukemia inhibitory factor (LIF) from a synthetic gene in *Escherichai coli* and the physical and biological characterization of the protein. *Biochim. Biophys. Acta* (1995) **vol. 1360**, pp. 27-34.

Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular cloning, a laboratory manual, vol. 3V, (1989) pp. New York: Cold Spring Harbor Laboratory Press.

Samuel, D., Liu, Y. J., Cheng, C. S. and Lyu, P. C. Solution structure of plant nonspecific lipid transfer protein-2 from rice (Oryza sativa). *J. Biol. Chem.* (2002) **vol. 277**, pp. 35267-35273.

Samuelsson, G. Mistletoe toxins *Systematic Zoology* (1974) **vol. 22**, pp. 566-569.

Sanger, F., Nicklen, S. and Coulson, A. R. DNA sequencing wiht chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* (1977) **vol. 74**, pp. 5463-5467.

Sauvion, N., Rahbé, Y., Peumans, W. J., Vandamme, E. J. M., Gatehouse, J. A. and Gatehouse, A. M. R. Effects of GNA and other mannose binding lectins on development and fecundity of the peach-potato aphid Myzus persicae *Entomol. Exp. Appl.* (1996) **vol. 79**, pp. 285-293.

Sawahel, W. A. The production of transgenic potato plants expressing human alpha-interferon using lipofectin-mediated transformation. *Cell. Mol. Biol. Lett.* (2002) **vol. 7**, pp. 19-29.

Saxena, K. N. and Basit, A. Inhibition of oviposition by volatiles of certain plants and chemicals in the leafhopper *Amrasca devastans* (Distant). *J. Chem. Ecol.* (1982) **vol. 8**, pp. 329-338.

Saxena, R. C. Insecticides from neem. In *Insecticides of plant origin.*, (eds. J. T. Arnason B. J. R. Philogene and P. Morand) Washington DC: American Chemical Society, (1989), pp. 110-135.

Schäffner, J., Winter, J., Rudolph, R. and Schwarz, E. Cosecretion of chaperones and low-molecular-size medium additives increases the yield of recombinant disulfide-bridged proteins. *Appl. Environ. Microbiol.* (2001) **vol. 67**, pp. 3994-4000.

Schagger, H. and Jagow, G. V. Tricine-Sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa *Anal. Biochem.* (1987) **vol. 166**, pp. 368-379.

Schechter, I. and Berger, A. On the size of the active site in proteases. *Biochem. Biophys. Res. Commun.* (1967) **vol. 27**, pp. 157-162.

Schein, C. H. Production of soluble recombinant proteins in bacteria. *Bio/technol.* (1989) **vol. 7**, pp. 1141-1149.

Schimoler O' Rourke, R., Richardson, M. and Selitrennikoff, C. P. Zeamatin inhibits trypsin and alpha-amylase activities *Appl. Environ. Microbiol.* (2001) **vol. 67**, pp. 2365-2366.

Schmelz, E. A., Grebenok, R. J., Galbraith, D. W. and Bowers, W. S. Insect-induced synthesis of phytoecdysteroids in spinach, *Spinacia oleracea J. Chem. Ecol.* (1999) **vol. 25**, pp. 1739-1757.

Schmutterer, H. Properties and potential of natural pesticides from the neem tree, *Azadirachta indica*. *Annu. Rev. Entomol.* (1990) **vol. 35**, pp. 271-297.

Schoonhoven, L. M., Jermy, T. and Van Loon, J. J. A. Plant chemistry : endless variety. In *Insect-Plant Biology*, (eds. L. M. Schoonhoven T. Jermy and J. J. A. Van Loon) London: Chapman & Hall, (1998), pp. 31-82.

Schroeder, H. E., Gollash, S., Moore, A., Tabe, L. M., Craig, S., Hardie, D. C., Chrispeels, M. J., Spencer, D. and Higgins, T. J. V. Bean alpha-amylase inhibitor confers resistance to the pea weevil (Bruchus pisorum) in transgenic peas (Pisum sativum L.) *Plant Physiol.* (1995) **vol. 107**, pp. 1233-1239.

Schuler, T. H., Poppy, G. M., Kerry, B. R. and Denholm, I. Insect-resistant transgenic plants *TIBTECH* (1998). **vol 16**, pp. 168-175.

Schultz, J. C. Many factors influence the evolution of herbivore diets, but plant chemistry is central. *Ecology* (1988) **vol. 69**, pp. 896-897.

Scott, K. D. and Playford, J. DNA Extraction Technique for PCR in Rain Forest Plant Species. *BioTechniques* (1996) **vol. 20**, pp. 974-978.

Ségalas, I., Thai, R., Ménez, R. and Vita, C. A particularly labile Asp-Pro bond in the green mamba muscarinic toxin MTX2. Effect of protein conformation on the rate of cleavage. *FEBS Lett.* (1995) **vol. 371**, pp. 171-175.

Seidman, C. E. Escherichia coli, plasmids and bacteriophages. In *Current Protocols in Molecular Biology.*,: John Wiley & Sons, (1997) vol. 1, pp. Unit1.

Shade, R. E., Schroeder, H. E., Pueyo, J. J., Tabe, L., Murdock, L. L., Higgins, T. J. and Chrispeels, M. J. Transgenic pea seeds expressing the alpha-amylase inhibitor of the common bean are resistant to bruchid beetles *BioTechnology* (1994) **vol. 12**, pp. 793-796.

Shaver, T. N. and Lukefahr, M. J. Effect of flavonoid pigments and gossypol on growth and development of bollworm, tobacco budworm and pink bollworm. *J. Econ. Entomol.* (1969) **vol. 62**, pp. 643-646.

Shewry, P. R. and Pandya, M. J. The 2S Albumin Storage Proteins In *Seed Proteins.*, (eds. P. R. Shewry and R. Casey): Kluwer Academic Publishers, (1999), pp. 563-586.

Shimamoto, K., Terada, R., Izawa, T. and Fujimoto, H. Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* (1989) **vol. 338**, pp. 274-276.

Shin, D. H., Lee, J. Y., Hwang, K. Y., Kim, K. K. and Suh, S. W. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* (1995) **vol. 3**, pp. 189-199.

Shine, J. and Dalgarno, L. The 3'-terminal sequence of *Escherichia coli* 16S riobosomal RNA : complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci.* (1974) **vol. 71**, pp. 1342-6.

Sijmons, P. C., Dekker, B. M., Schrammeijer, B., Verwoerd, T. C., Van den Elzen, P. J. and Hoekema, A. Production of correctly processed human serum albumin in transgenic plants. *Bio/technol.* (1990) **vol. 8**, pp. 217-221.

Silva, C. P., Terra, W. R., Xavier-Filho, J., de Sa, M. F. G., Lopes, A. R. and Pontes, E. G. Digestion in larvae of Callosobruchus maculatus and Zabrotes subfasciatus (Coleoptera: Bruchidae) with emphasis on alpha -amylases and oligosaccharidases *Insect Biochem. Molec. Biol.* (1999) **vol. 29**, pp. 355-366.

Simmonds, M. S. J., Blaney, W. M. and Fellows, L. E. Behavioral and electrophysiological study of antifeedant mechanisms associated with polyhydroxy alkaloids. *J. Chem. Ecol.* (1990) **vol. 16**, pp. 3167-3196.

Singh, P., Russell, G. B. and Fredericksen, S. The dietary effects of some ecdysteroids on the development of housefly. *Entomol. Exp. Appl.* (1982) **vol. 32**, pp. 7-12.

Smith, D. B. and Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* (1988) **vol. 67**, pp. 31-40.

Solomon, M., Belenghi, B., Delledonne, M., Menachem, E. and Levine, A. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants *Plant Cell* (1999) **vol. 11**, pp. 431-443.

Spencer, D., Higgins, T. J. V., Freer, M., Dove, H. and Coombe, J. B. Monitoring the fate of dietary proteins in rumen fluid using gel electrophoresis *British. J. Nutr.* (1988) **vol. 60**, pp. 241-247.

Stader, J. A. and Silhavy, T. J. Engineering *Escherichia coli* to secrete heterologous gene products. *Methods Enzymol.* (1990) **vol. 185**, pp. 166-187.

Steiz, J. A. and Jakes, K. How ribosomes select initiator regions in mRNA : base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli. Proc. Natl. Acad. Sci.* (1975) **vol. 72**, pp. 4734-4738.

Stewart, E. J., Aslund, F. and Beckwith, J. Disulfide formation in the *Escherichia coli* cytoplasm : an *in vivo* role reversal for the thioredoxins. *EMBO J.* (1998) **vol. 17**, pp. 5543-5550.

Strobl, S., Maskos, K., Wiegand, G., Huber, R., F.X., G. R. and Glockshuber, R. A novel strategy for inhibition of alpha-amylases : yellow meal worm alpha-amylase in complex with the ragi bifunctional inhibitor at 2.5 A resolution. *Structure* (1998) **vol. 6**, pp. 911-921.

Stuart, L. S. and Harris, T. H. Bactericidal and fungicidal properties of a crystalline protein isolated from unbleached wheat flour. *Cereal Chemistry* (1942) **vol. 19**, pp. 288-300.

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. and Turk, V. The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO Journal* (1990) **vol. 9**, pp. 1939-1947.

Studier, F. W. and Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* (1986) **vol. 189**, pp. 113-130.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* (1990) **vol. 185**, pp. 60-89.

Swain, T. Secondary compounds as protective agents. *Annu. Rev. Plant Physiol.* (1977) **vol. 28**, pp. 479-501.

Sy, D., Le Gravier, Y., Goodfellow, J. and Vovelle, F. Protein stability and plasticity of the hydrophobic cavity in wheat ns-LTP. *J. Biomol. Struct. Dyn.* (2003) **vol. 21**, pp. 15-29.

Szoka, P. R., Schreiber, A. B., Chan, H. and Murthy, J. A general method for retrieving the components of a genetically engineered fusion protein. *DNA* (1986) **vol. 5**, pp. 11-20.

Tassin, S., Broekaert, W. F., Marion, D., Acland, D. P., Ptak, M., Vovelle, F. and Sodano, P. Solution structure of Ace-AMP1, a potent antimicrobial protein extracted from onion seeds. Structural analogies with plant nonspecific lipid transfer proteins. *Biochemistry* (1998) **vol. 37**, pp. 3623-3637.

Terra, W. R. and Ferreira, C. Insect digestive enzymes: properties, compartmentalization and function *Comp. Biochem. Physiol.* (1994) **vol. 109B**, pp. 1-62.

Terras, F. R., Eggermont, K., Kovaleva, V., Raikhel, N. V., Osborn, R. W., Kester, A., Rees, S. B., Torrekens, S., Van Leuven, F. and Vanderleyden, J. Small cystein-rich antifungal proteins from radish : their role in host plant defense. *Plant Cell* (1995) **vol. 7**, pp. 573-588.

Thevissen, K., Cammue, B. P. A., Lemaire, K., Winderickx, J., Dickson, R. C., Lester, R. L., Ferket, K. K., Van Even, F., Parret, A. H. and Broekaert, W. F. A gene encoding a sphingolipid biosynthesis enzyme determines the sensitivity of *Saccharomyces cerevisiae* to an antifungal plant defensin from dahlia (*Dahlia merckii*). *Proc. Natl. Acad. Sci. U.S.A.* (2000a) **vol. 97**, pp. 9531-9536.

Thevissen, K., Francois, I. E., Takemoto, J. Y., Ferket, K. K., Meert, E. M. and Cammue, B. P. A. DmAMP1, an antifungal plant defensin from dahlia (Dahlia merckii), interacts with sphingolipids from Saccharomyces cerevisiae. *FEMS Microbiol. Lett.* (2003) **vol. 226**, pp. 169-173.

Thevissen, K., Ghazi, A., De Samblanx, G. W., Brownlee, C., Osborn, R. W. and Broekaert, W. F. Fungal membrane responses induced by plant defensins and thionins. *J. Biol. Chem.* (1996) **vol. 271**, pp. 15018-15025.

Thevissen, K., Osborn, R. W., Acland, D. P. and Broekaert, W. F. Specific, high affinity binding sites for an antifungal plant defensin on Neurospora crassa hyphae and microsomal membranes. *J. Biol. Chem.* (1997) **vol. 272**, pp. 32176-32181.

Thevissen, K., Osborn, R. W., Acland, D. P. and Broekaert, W. F. Specific binding sites for an antifungal plant defensin from dahlia (Dahlia merckii) on fungal cells are required for antifungal activity. *Mol. Plant Microbe Interac.* (2000b) **vol. 13**, pp. 54-61.

Thevissen, K., Terras, F. R. G. and Broekaert, W. F. Permeabilization of fungal membranes by plant defensins inhibits fungal growth. *Applied and Environmental Microbiology* (1999) **vol.**, pp. 5451-5458.

Thomma, B. P. H. J., Cammue, B. P. A. and Thevissen, K. Plant defensins *Planta* (2002) **vol. 216**, pp. 193-202.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice *Nucleic Acids Res* (1994) **vol. 22**, pp. 4673-80.

Toms, G. C. and Western, A. Phytohemagglutinins. In *Chemotaxonomy of the Leguminosae*., (eds. J. B. Harborne D. Boulter and B. L. Turner) New York: Academic Press, (1971), pp. 367-462.

Towbin, H., Staehelin, T. and Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications *Proc. Natl. Acad. Sci. USA* (1979) **vol. 76**, pp. 4350-4354.

Treviño, M. B. and O' Connell, M. A. Three Drought-Responsive Members of the Nonspecific Lipid-Transfer Protein Gene Family in Lycopersicon pennellii Show Different Developmental Patterns of Expression. *Plant Physiol.* (1998) **vol. 116**, pp. 1461-1468.

Turk, B., Turk, V. and Turk, D. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol. Chem.* (1997) **vol. 378**, pp. 141-150.

Turkov, M., Rashi, S., Noa, M., Gordon, D., Khalifa, R., Stankiewicz, M., Pelhate, M. and Gurevitz, M. *In vitro* folding and functional analysis of an anti-insect selective scorpion depressant neurotoxin produced in *Escherichia coli*. *Protein Expr. Purif.* (1997) **vol. 10**, pp. 123-131.

Unnithan, G. C., Nair, K. K. and Kooman, C. J. Effects of precocene II and juvenile hormone III on the activity of neurosecretory A-cells in *Oncopeltus fasciatus*. *Experientia* (1978) **vol. 34**, pp. 411-412.

Vallé, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K. W., Svensson, B. and Haser, R. Barley alpha-amylase bound to its endogenous protein inhibitor BASI : crystal structure of the complex at 1,9 A resolution. *Structure* (1998) **vol. 6**, pp. 649-659.

Varadarajan, R., Szabo, A. and Boxer, S. G. Cloning, expression in *Escherichia coli*, and reconstitution of human myoglobin. *Proc. Natl. Acad. Sci.* (1985) **vol. 82**, pp. 5681-5684.

Vaublanc, G. V. d. Histopathologie de l'appareil digestif de Sitophilus oryzae après intoxication par une protéine du pois PA1b, *Masters thesis,* Lyon (Fra): Université Catholique de Lyon. (2001) 36 p.

Vayssière, M. P. Les charançons du riz et les légumes secs. *C.R. Acad. Agri. F2* (1943) **vol. 29**, pp. 449-450.

Vercauteren, I., Van Der Schueren, E., Van Montagu, M. and Gheysen, G. Arabidopsis thaliana genes expressed in the early compatible interaction with root-knot nematodes *Mol Plant Microbe Interact* (2001) **vol. 14**, pp. 288-99.

Verheyden, G., Volckaert, G. and Engelboghs, Y. Expression of chymotrypsin (ogen) in the thioredoxin reductase deficient mutant strain of *Escherichia coli* AD494(DE3) and purification via a fusion product with a hexahistidine-tail. *J. Chromatogr. B. Biomed. Sci. App.* (2000) **vol. 737**, pp. 213-224.

Vernon, L. P., Evett, G. E., Zeikus, R. D. and Gray, W. R. A toxic thionin from *Pyrularia pubera* : purification, properties and amino acid sequence. *Archives of Biochemistry and Biophysics* (1985) **vol. 238**, pp. 18-29.

Wall, J., Golding, C. A., Van Veen, M. and O'Shea, P. The use of fluoresceinphosphatidylethanolamine (FPE) as a real-time probe for peptide-membrane interactions. *Mol. Membr. Biol.* (1995) **vol. 12**, pp. 183-192.

Walsh, T. A. and Strickland, J. A. The 85-kd crystalline cysteine protease inhibitor from potato contains 8 cystatin domains. *Protein Engineering* (1993) **vol. 6**, pp. 53-56.

Wang, Y., Jing, L. and Xu, K. A unique approach for high level expression and production of a recombinant cobra neurotoxin in *Escherichia coli*. *Journal of Biotechnology* (2002) **vol. 94**, pp. 235-244.

Watanabe, Y., Barbashov, S. F., Komatsu, S., Hemmings, A. M., Miyagi, M., Tsunasawa, S. and Hirano, H. A peptide that stimulates phosphorylation of the plant

insulin-binding protein. Isolation, primary structure and cDNA cloning *Eur. J. Biochem.* (1994) **vol. 224**, pp. 167-172.

Weickert, M. J., Doherty, D. H., Best, E. A. and Olins, P. O. Optimizing of heterologous protein production in *Escherichia coli*. *Curr. Opin. Biotechnol.* (1996) **vol. 7**, pp. 494-499.

Wijaya, R., Neumann, G. M., Condron, R., Hughes, A. B. and Polya, G. M. Defense proteins from seed of Cassia fistula include a lipid transfer protein homologue and a protease inhibitory plant defensin. *Plant Science* (2000) **vol. 159**, pp. 243-255.

Wink, M. Chemische Verteidigung der Lupinen: zur biologischen Bedeutung der Chinolizidinalkaloide *Plant System. Evol.* (1985) **vol. 15O**, pp. 65-81.

Wink, M. Quinolizidine alkaloids. In *Methods in Plant Biochemistry*, ed. P. G. Waterman) London: Academic Press, (1993) vol. 8, pp. 197-239.

Wink, M., Meissner, C. and Witte, L. Patterns of quinolizidine alkaloids in 56 species of the genus *Lupinus*. *Phytochemistry* (1995) **vol. 38**, pp. 139-153.

Wink, M. and Mohamed, G. I. A. Evolution of chemical defense traits in the Leguminosae: mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred from nucleotide sequences of the rbcL gene *Biochem. Syst. Ecol.* (2003) **vol. 31**, pp. 897-917.

Wolfson, J. L. Developmental responses of *Pieris rapae* and *Spodoptera eridania* to environmentally induced variation in *Brassica nigra*. *Environ. Entomol.* (1982) **vol. 11**, pp. 207-213.

Wood, R. W., Boss, M. A., Kenten, J. H., Calvert, J. E., Roberts, N. A. and Emtage, J. S. The synthesis and *in vivo* assembly of functional antibodies in yeats. *Nature* (1985) **vol. 314**, pp. 446-449.

Woodhead, S. and Bernays, E. Changes in release rates of cyanide in relation to palatability of Sorghum to insects. *Nature* (1977) **vol. 270**, pp. 235-236.

Woodhead, S. and Chapman, R. F. Insect behaviour and the chemistry of plant surface waxes. In *Insect and the plant surface.*, (eds. B. Juniper and T. R. E. Southwood) London: Edward Arnold, (1986), pp. 123-135.

Woolley, J. G. Plant alkaloids. In *Encyclopedia of life sciences (www.els.net).*,: Nature Publishing Group, (2001), pp. 1-11.

Xu, D. P., Xue, Q. Z., McElroy, D., Mawal, Y., Hilder, V. A. and Wu, R. Constitutive expression of a cowpea trypsin-inhibitor gene, CpTI, in transgenic rice plants confers resistance to two major rice insect pests. *Mol. Breed.* (1996) **vol. 2**, pp. 167-173.

Yamagata, H., Kunimatsu, K., Kamasaka, H., Kuramoto, T. and Iwasaki, T. Rice bifunctional alpha-amylase/subtilisin inhibitor : characterization, localization and changes in developing and germinating seeds. *Biosc. Biotechnol. Biochem.* (1998) **vol. 62**, pp. 978-985.

Yamazaki, T., Takaoka, M., Katoh, E., Hanada, K., Sakita, M., Sakata, K., Nishiuchi, Y. and Hirano, H. A possible physiological function and the tertiary structure of a 4-kDa peptide in legumes *Eur. J. Biochem.* (2003) **vol. 270**, pp. 1269-76.

Young, N. D., Mudge, J. and Ellis, T. H. Legume genomes: more than peas in a pod *Curr. Opin. Plant Biol.* (2003) **vol. 6**, pp. 199-204.

Yubero-Serrano, E. M., Moyano, E., Medina-Escobar, N., Munoz-Blanco, J. and Caballero, J. L. Identification of a strawberry gene encoding a non-specific lipid transfer protein that responds to ABA, wounding and cold stress. *J. Exp. Bot.* (2003) **vol. 54**, pp. 1865-1877.

Yunes, A. N. A., Andrade, M. T., Sales, M. P., Morais, R. A., Fernandes, K. V. S., Gomes, V. M. and Xavier Filho, J. Legume seed vicilins (7S storage proteins) interfere with the development of the cowpea weevil (*Callosobruchus maculatus* F.). *J. Sci. Food Agric*. (1998) **vol. 76**, pp. 111-116.

Zak, O. and Aisen, P. A new method for obtaining human transferrin C-lobe in the native conformation : preparation and properties. *Biochemistry* (2002) **vol. 41**, pp. 1647-1653.

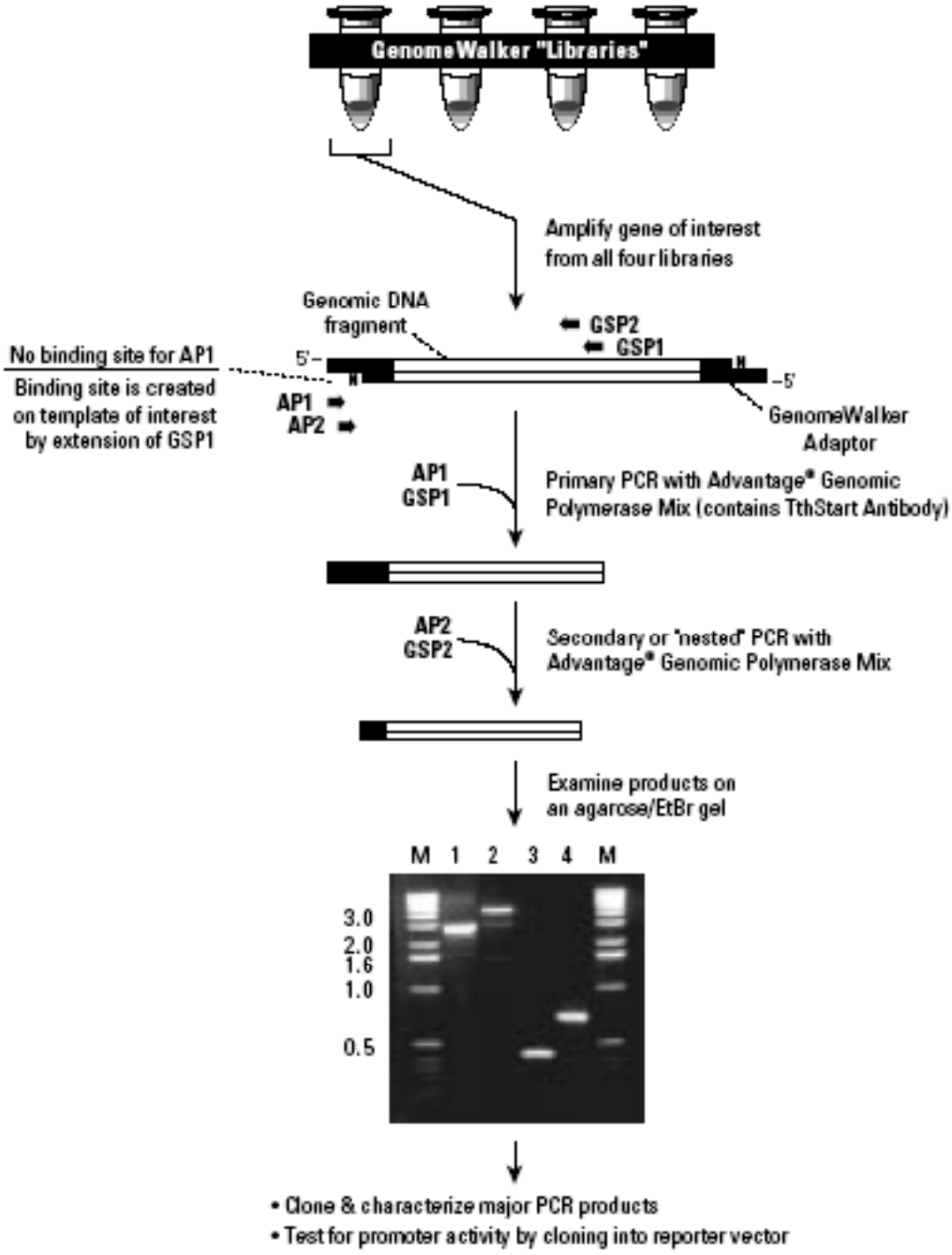
Zhao, Y., Botella, M. A., Subramanian, L., Niu, X. M., Nielsen, S. S., Bressan, R. A. and Hasegawa, P. M. Two wound-inducible soybean cysteine proteinase inhibitors have greater insect digestive proteinase inhibitory activities than a constitutive homolog *Plant Physiol*. (1996) **vol. 111**, pp. 1299-1306.

Zhu, S., Darbon, H., Dyason, K., Verdonck, F. and Tytgat, J. Evolutionary origin of inhibitor cystine knot peptides. *FASEB J*. (2003) **vol. 17**, pp. 1765-1767.

Annexes.

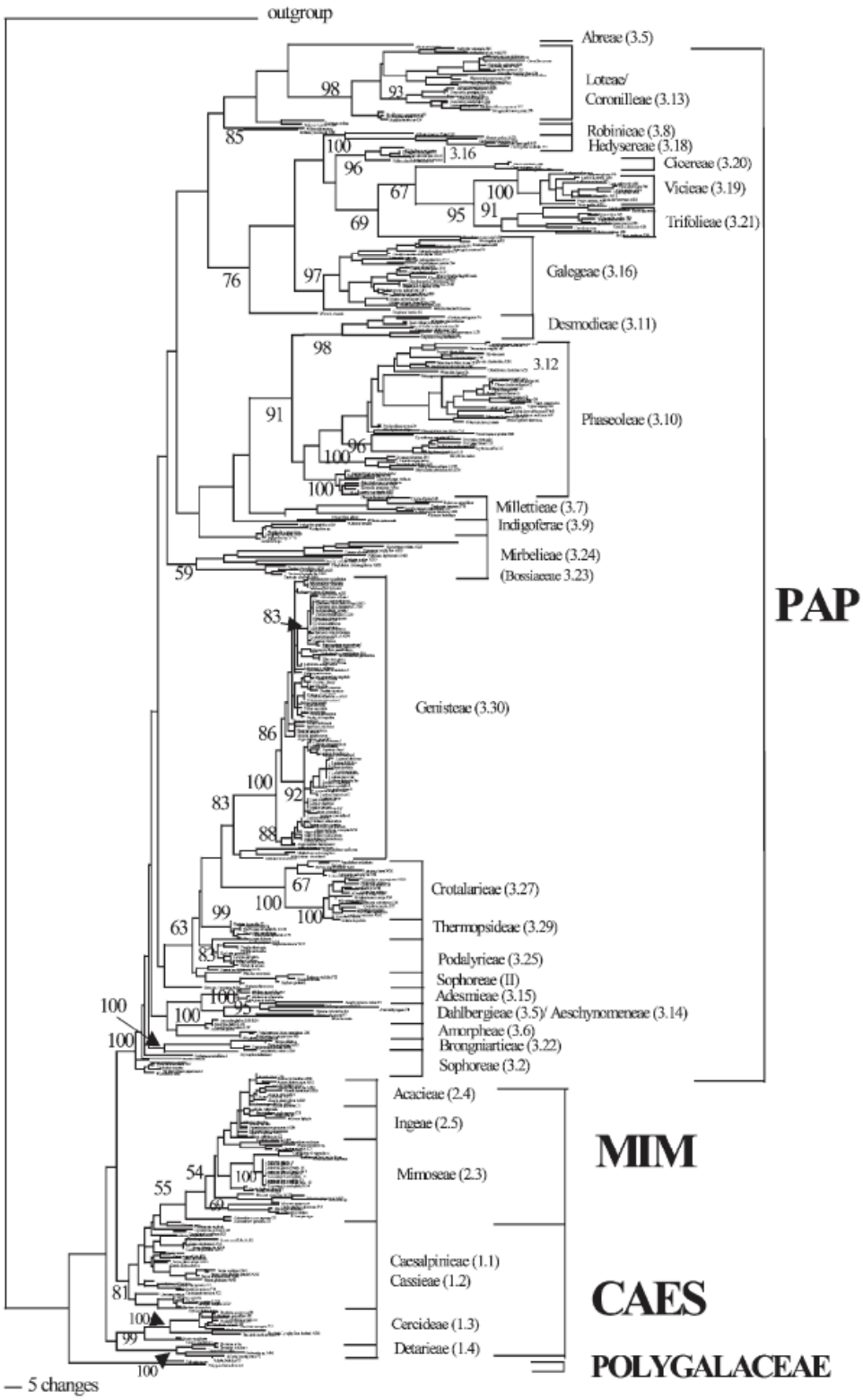
I. Annexe 1 : Principe de la marche sur le gène (kit GenomeWalker^{TM}, Clontech).

Après digestion de l'ADN génomique par quatres enzymes différentes, des adaptateurs sont liés aux fragments de restriction. On obtient 4 « banques » alors soumises à deux réactions de PCR nichées, avec les amorces fournies AP (adaptor primer) et les amorces spécifiques GSP (gene specific primer).

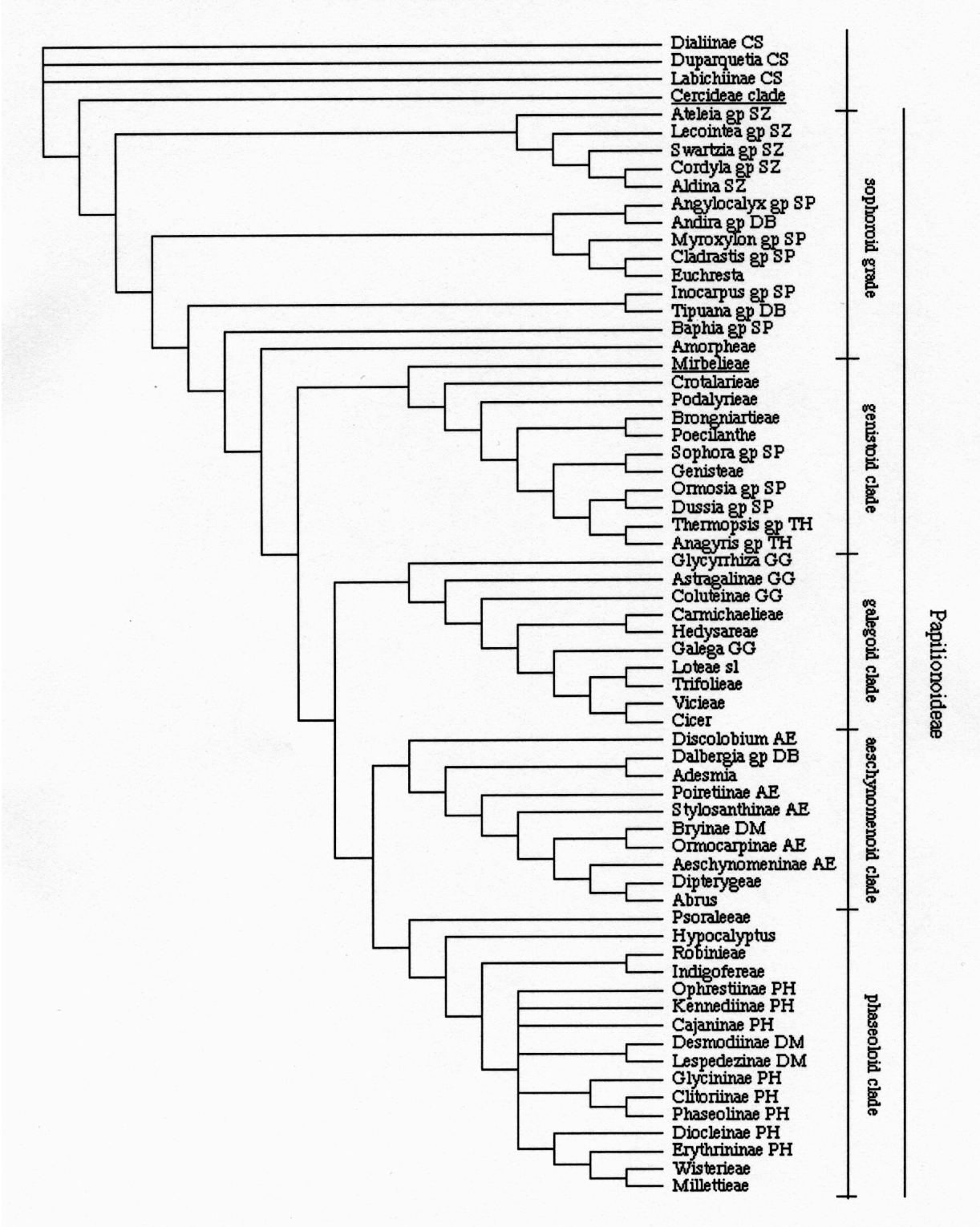


II. Annexe 2 : Phylogénie des Légumineuses

A : Analyse du gène *rbc*L (Wink, M. 2003). Pap : Papilionoideae, Mim : Mimosoideae, Caes : Caesalpinioideae.



B: Analyse de multiples données (non moléculaires) d'après Chapill (http://www.botany.uwa.edu.au/systematics/paps.html)



III. Annexe 3 : liste des espèces sélectionnées, origine géographique et fournisseur

Sous-famille	Tribu	Genre	Espèce	Origine géographique	Fournisseur
Caesalpinioideae		Acrocarpus	fraxinifolius		Sandeman seeds
Caesalpinioideae	Amherstieae	Tamarindus	indica L.	Sénégal	P Delobel
Caesalpinioideae	Caesalpinieae	Caesalpinia	bonduc (L.) Roxb.	Afrique	commerce
Caesalpinioideae	Caesalpinieae	Delonix	regia (Hook) Raf.	Guadeloupe	Y Rahbé
Caesalpinioideae	Caesalpinieae	Gleditsia	tricanthos L.	Lyon	B Delobel
Caesalpinioideae	Caesalpinieae	Parkinsonia	aculeata L.	Sénégal	A Delobel
Caesalpinioideae	Caesalpinieae	Vouacapoua	americana Aublet	Guyane	
Caesalpinioideae	Cassieae	Cassia	occidentalis L.		
Caesalpinioideae	Cassieae	Ceratonia	siliqua L.	Portugal (Algarve)	Luis Neto
Caesalpinioideae	Cassieae	Dialium	guineense Willd.	sénégal	A Delobel
Caesalpinioideae	Cassieae	Dicorynia	guianensis Amshoff	Guyane	
Caesalpinioideae	Cercideae	Bauhinia	natalensis Hook.	RSA	Silverseeds
Caesalpinioideae	Cercideae	Cercis	siliquastrum L.	France (Lyon)	B Delobel
Caesalpinioideae	Detarieae	Brachystegia	spiciformis Benth	RSA	Silverseeds
Caesalpinioideae	Detarieae	Detarium	senegalense Gmelin	Sénégal	A Delobel
Caesalpinioideae	Detarieae	Eperua	falcata Aublet	Guyane	
Caesalpinioideae	Detarieae	Hymenaea	courbaril L.	Guyane	
Caesalpinioideae	Detarieae	Intsia	bijuga (Colebr.) Kuntz	Madagascar	SNGF
Caesalpinioideae	Detarieae	Schotia	afra (L.) Thunb.	RSA	silverseeds
Mimosoideae	Acacieae	Acacia	dealbata Link	Madagascar	SNGF
Mimosoideae	Acacieae	Acacia	nilotica (L.) Del	Sénégal	A Delobel
Mimosoideae	Acacieae	Acacia	raddiana Save	Sénégal	A Delobel
Mimosoideae	Ingeae	Albizia	julibrissin Durazz.	Lyon	parc
Mimosoideae	Ingeae	Albizia	lebbeck(L.) Benth.		
Mimosoideae	Ingeae	Faidherbia	albida Del.	RSA	Silverseeds
Mimosoideae	Ingeae	Pithecellobium	dulce (Roth) Benth	Sénégal	A Delobel
Mimosoideae	Mimoseae	Desmanthus	illinoensis (Michaux) Mac Millan	Lyon	parc
Mimosoideae	Mimoseae	Dichrostachys	cinerea (L.) Wight & Arn	Sénégal	B Delobel
Mimosoideae	Mimoseae	Entada	rheedii Sprengel = phaseoloides	Viet-nam	H Delobel
Mimosoideae	Mimoseae	Leucaena	leucocephala (Lam) De Witt	Guadeloupe	Y Rahbé
Mimosoideae	Mimoseae	Neptunia	oleracea Lour	Sénégal	A Delobel
Mimosoideae	Mimoseae	Prosopis	africana (G. et P.) Taubert	Sénégal	A Delobel
Mimosoideae	Parkieae	Parkia	biglobosa (Jacq.) Don	Senegal	A Delobel
Papilionoideae	Abreae	Abrus	precatorius L.	Australie	Nindethana seeds
Papilionoideae	Adesmieae	Adesmia	bicolor (Poiret)DC	Uruguay	Fac Sci Montevideo
Papilionoideae	Aeschynomeneae	Aeschynomene	indica L		
Papilionoideae	Aeschynomeneae	Arachis	hypogaea L.		commerce
Papilionoideae	Aeschynomeneae	Zornia	glochidiata DC		
Papilionoideae	Amorpheae	Amorpha	fruticosa L.	Lyon	parc
Papilionoideae	Bossiaeae	Goodia	lotifolia Salisb.	Australie	Nindethana seeds
Papilionoideae	Brongniartieae	Hovea	acanthoclada F. Muell	Australie	Nindethana seeds
Papilionoideae	Brongniartieae	Templetonia	retusa (Vent) R. Br.	Australie	Nindethana seeds
Papilionoideae	Carmichaelieae	Carmichaelia	stevensoni	Nouvelle zelande	
Papilionoideae	Cassieae	Storckiellia	pancheri Baill.	Nouvelle Calédonie	Endemia
Papilionoideae	Cicereae	Cicer	arietinum L.		graines Baumeaux
Papilionoideae	Crotalarieae	Crotalaria	eremaea (F.) Muell.	Australie	Nindethana seeds
Papilionoideae	Crotalarieae	Crotalaria	podocarpa DC	Sénégal	A Delobel
Papilionoideae	Crotalarieae	Lotononis	sp	RSA	Silverseeds
Papilionoideae	Dalbergieae	Dalbergia	purpurescens	Madagascar	SNGF
Papilionoideae	Dalbergieae	Pterocarpus	rotundifolis (Sonder)Druce	RSA	Silverseeds
Papilionoideae	Desmodieae	Alysicarpus	ovalifolius (Schum) Leonard		
Papilionoideae	Desmodieae	Desmodium	canadense (L.) DC	Lyon	parc
Papilionoideae	Desmodieae	Desmodium	glutinosum (Willd.) Wood	Canada	
Papilionoideae	Galegeae	Alhagi	graecorum Boiss.	Egypte	A.Delobel
Papilionoideae	Galegeae	Astragalus	monspessulanus L.	Lyon,	parc
Papilionoideae	Galegeae	Caragana	arborescens L.	Villeurbanne	B Delobel
Papilionoideae	Galegeae	Colutea	arborescens L.	Lyon	B Delobel
Papilionoideae	Galegeae	Galega	officinalis L.	vaulx en velin	B Delobel
Papilionoideae	Galegeae	Galega	officinalis L. cv lady wilson	monsols	B Delobel
Papilionoideae	Galegeae	Glycyrrhiza	glabra L.	Lyon	parc
Papilionoideae	Genisteae	Cytisus	sessilifolium (L)Lang	ardeche (Crussols)	B Delobel
Papilionoideae	Genisteae	Laburnum	anagyroides Med.	Lyon	B Delobel
Papilionoideae	Genisteae	Lupinus	albus L. cv lublan		
Papilionoideae	Genisteae	Lupinus	angustifolius L. cv unicrop		INRA
Papilionoideae	Genisteae	Spartium	junceum	corse	B Delobel
Papilionoideae	Genisteae	Ulex	europaeus L.	monsols	B Delobel
Papilionoideae	Hedysareae	Onobrychis	viciifolia Scop	Lyon	parc
Papilionoideae	Indigofereae	Indigofera	astragalina DC	senegal	A Delobel
Papilionoideae	Indigofereae	Indigofera	circinata Harvey	RSA	Silverseeds
Papilionoideae	Indigofereae	Indigofera	dosua	parc de la tete d'or	B Delobel
Papilionoideae	Liparieae	Hypocalyptus	coluteoides (Lam) Dahlgren	RSA	silverhill seeds
Papilionoideae	Liparieae	Priestleya	myrtifolia (Thunb.)DC	RSA	silverseeds
Papilionoideae	Loteae	Anthyllis	barba-jovis L.	Lyon	parc
Papilionoideae	Loteae	Hippocrepis	emerus (L.) Lassen	Lyon	B Delobel
Papilionoideae	Loteae	Lotus	corniculatus L. cv leo		
Papilionoideae	Loteae	Lotus	tetragonolobus L.		graines Baumaux
Papilionoideae	Loteae	Securigera	varia L.	Lyon	parc
Papilionoideae	Millettieae	Lonchocarpus	capassa Rolfe	RSA	Silverseeds
Papilionoideae	Millettieae	Lonchocarpus	sericeus (Poiret) DC	Sénégal	A Delobel
Papilionoideae	Millettieae	Millettia	grandis (Meyer)Skeel	RSA	Silverseeds
Papilionoideae	Millettieae	Mundulea	sericea (Willd.) Chev	RSA	Silverseeds
Papilionoideae	Millettieae	Tephrosia	bracteolata Guill. & Perr.		
Papilionoideae	Millettieae	Wisteria	sinensis	Monsols	B Delobel
Papilionoideae	Mirbelieae	Daviesia	corymbosa Smith	Australie	Nindethana seeds
Papilionoideae	Mirbelieae	Gompholobium	scabrum Smith	Australie	Nindethana seeds
Papilionoideae	Mirbelieae	Viminaria	juncea Schroder) Hoffsgg	Australie	Nindethana seeds
Papilionoideae	Phaseoleae	Apios	americana Med.	Lyon	parc
Papilionoideae	Phaseoleae	Cajanus	minima (L.) DC		

Papilionoideae	Phaseoleae	Canavalia	brasiliensis Benth	Bresil	Prof Benildo, Fortaleza
Papilionoideae	Phaseoleae	Dioclea	grandifolia Benth	Bresil	sousa cavada
Papilionoideae	Phaseoleae	Erythrina	crista-galli L.	?	Beaumaux
Papilionoideae	Phaseoleae	Glycine	max (L.) Merr. cv Paoki		INRA
Papilionoideae	Phaseoleae	Hardenbergia	comptoniana (Andrews) Benth	Australie	Nindethana seeds
Papilionoideae	Phaseoleae	Kennedia	nigricans Lindley	Australie	Nindethana seeds
Papilionoideae	Phaseoleae	Phaseolus	vulgaris L. cv Contender	France	Vilmorin
Papilionoideae	Phaseoleae	Rhynchosia	pentheri Zahlbr.	RSA	silverhill seeds
Papilionoideae	Phaseoleae	Strongylodon	macrobotrys Gray	phillipines	parc
Papilionoideae	Phaseoleae	Teramnus		Vietnam	H Delobel
Papilionoideae	Phaseoleae	Vigna	radiata (L.) Wilczek		
Papilionoideae	Phaseoleae	Vigna	subterranea (L.) Verdc.	Tchad	Mbailao
Papilionoideae	Phaseoleae	Vigna	unguiculata (L.) Walp. cv Dall'oro	France	graines Baumeaux
Papilionoideae	Podalyrieae	Calpurnia	aurea (Aiton)Benth.	RSA	Silverhill seeds
Papilionoideae	Podalyrieae	Cyclopia	pubescens Ecklon&Zeyher	RSA	Silverhill seeds
Papilionoideae	Podalyrieae	Podalyria	biflora (Retz) Lam.	RSA	silverseeds
Papilionoideae	Psoraleeae	Bituminaria	bituminosa L.	sud-ardèche	B Delobel
Papilionoideae	Psoraleeae	Psoralea	oligophylla Eklon & Zegher	RSA	Silverseeds
Papilionoideae	Robinieae	Robinia	pseudoacacia L.	Lyon	B Delobel
Papilionoideae	Robinieae	Sesbania	pachycarpa DC	Sénégal	A Delobel
Papilionoideae	Robinieae	Sesbania	sesban (L.) Merr.	Madagascar	Nindethana seeds
Papilionoideae	Sophoreae	Baphia	massaiensis Taubert	RSA	Silverhillseeds
Papilionoideae	Sophoreae	Bolusanthus	speciosus (Bolus)Harms	RSA	silverhill seeds
Papilionoideae	Sophoreae	Cladrastis	kentukea (D-C.) Rudd.	Lyon	parc
Papilionoideae	Sophoreae	Diplotropis	purpurea (Rich)Amshoff	guyane	Inra ecol forest (yves Goret)
Papilionoideae	Sophoreae	Pericopsis	angolensis (Baker) van Meeuwen	RSA	Silverseeds
Papilionoideae	Sophoreae	Sophora	arizonica Watson	USA	desert legume program, arizona uni.
Papilionoideae	Sophoreae	Sophora	secundiflora (Ortega) DC	USA	desert legume program, arizona uni.
Papilionoideae	Sophoreae	Styphnolobium	japonicum (L) Schott	Lyon	parc
Papilionoideae	Swartzieae	Bobgunnia	madagascariensis (Desv) Kirkbr.&	Wiersem RSA	silverhill seeds
Papilionoideae	Thermopsideae	Baptisia	australis (L.) Br.	Lyon	parc
Papilionoideae	Trifolieae	Medicago	lupulina	Monsols	B Delobel
Papilionoideae	Trifolieae	Medicago	sativa L.		
Papilionoideae	Trifolieae	Medicago	truncatula Gaertner cv salernes		
Papilionoideae	Trifolieae	Melilotus	albus Medikus	Lyon	parc
Papilionoideae	Trifolieae	Trifolium	fragiferum L.	Villeurbanne	B Delobel
Papilionoideae	Trifolieae	Trigonella	foenum-graecum L.	?	commerce
Papilionoideae	Vicieae	Lathyrus	latifolius L.	Monsols	B Delobel
Papilionoideae	Vicieae	Lens (2)	culinaris Med.cv "du puy"		commerce
Papilionoideae	Vicieae	Pisum	sativum arvense L.	Bresil	Sousa cavada
Papilionoideae	Vicieae	Pisum	sativum L. cv frisson	France	INRA
Papilionoideae	Vicieae	Vicia	hirsuta (L)Gray	Villeurbanne	B Delobel
Polygalaceae		Polygala	myrtifolia	Corse	B Delobel

IV. Annexe 4 : Représentation des différentes fractions dans la farine d'espèces de Légumineuses.

espèce	sous-famille/tribu	lipides%	MeOH%	MeOH60%	H2O5%	H2O8%	residu%	somme	non réserve
Abrus precatorius	Aeschynomenoide	6,4	4,3	6,6	3,5	5,1	65,5	91,5	19,6
Adesmia bicolor	Aeschynomenoide	8	7,3	5,6	3,8	9,2	54,2	88,2	25,9
Arachis hypogea	Aeschynomenoïde	46,4	2,6	2,1	2,3	0,0	39,5	92,8	6,9
Dalbergia purpurescens	Aeschynomenoïde	5,7	16,9	10,1	4,4	5,0	47,2	89,3	36,4
Acrocarpus fraxifolius	Caesalpinioideae	30,7	3,0	4,5	3,9	7,0	46,8	95,8	18,3
Bauhinia natalensis	Caesalpinioideae	11	11,7	4,1	2,6	5,1	57,8	92,3	23,5
Brachystegia spiciformis	Caesalpinioideae	16,6	4,3	4,4	3,2	6,9	54,2	89,6	18,8
Caesalpinia bonduc	Caesalpinioideae	23	11,1	14,2	6,0	7,1	29,8	91,2	38,4
Ceratonia siliqua	Caesalpinioideae	4,9	1,9	1,6	9,6	6,7	65,0	89,7	19,8
Cercis siliquastrum	Caesalpinioideae	7,1	8,2	6,2	2,5	2,0	62,5	88,5	18,9
Delonix regia	Caesalpinioideae	7,9	8,2	9,7	9,5	19,7	36,7	91,7	47,1
Dialium guineense	Caesalpinioideae	2,3	4,6	5,1	12,1	11,9	55,2	91,3	33,7
Gleditsia triacanthos	Caesalpinioideae	8	11,2	11,4	10,9	13,3	32,1	86,9	46,8
Intsia bijuga	Caesalpinioideae	12,7	6,4	4,6	3,2	7,4	62,2	96,5	21,6
Schotia afra	Caesalpinioideae	9,8	3,3	4,5	7,5	29,6	34,8	89,6	45,0
Storckiella	Caesalpinioideae	0,9	5,0	4,7	12,0	6,1	61,3	90,0	27,8
Tamarindus indica	Caesalpinioideae	7,2	2,8	4,9	3,7	8,0	63,3	89,8	19,3
Alhagi graecorum	Galegoide	10,6	8,8	12,6	7,8	3,3	46,1	89,2	32,5
Anthyllis barba-jovis	Galegoide	15,5	6,7	4,4	6,6	10,3	43,6	87,0	27,9
Astragalus monspessulanus	Galegoide	10	6,1	6,7	6,8	10,1	50,6	90,3	29,7
Caragana arborescens	Galegoide	15,1	5,7	9,0	5,5	10,0	46,2	91,3	30,1
Carmichaelia stevensoni	Galegoide	10,7	5,9	8,2	7,1	11,1	49,2	92,2	32,3
Cicer arietinum	Galegoide	5,3	6,4	6,4	4,9	9,3	53,7	86,1	27,0
Colutea arborescens	Galegoide	2,4	6,8	12,0	6,7	8,8	55,6	92,2	34,2
Galega officinalis	Galegoide	3,7	6,7	9,0	5,1	7,8	62,1	94,4	28,6
Glycyrrhiza glabra	Galegoide	13,9	5,4	10,1	5,7	3,9	53,2	92,2	25,1
Hippocrepis emerus	Galegoide	30,1	5,0	6,5	3,7	4,3	40,4	90,0	19,5
Lathyrus latifolius	Galegoide	2,6	1,5	6,8	4,0	12,3	58,2	85,4	24,5
Lens esculenta	Galegoide	4,1	3,6	4,5	4,5	10,3	58,6	85,6	23,0
Lotus corniculatus	Galegoide	9,1	5,2	8,5	5,7	4,5	57,0	89,8	23,8
Lotus tetragonolobus	Galegoide	14,5	4,0	4,9	6,0	8,3	53,2	90,9	23,1
Medicago lupulina	Galegoide	6,2	4,3	8,8	9,3	11,3	45,5	85,5	33,7
Medicago truncatula	Galegoide		4,0	5,6	9,3	4,5	47,7	71,2	23,4
Melilotus albus	Galegoide	6,3	8,1	7,2	5,2	6,4	55,9	89,2	26,9
Onobrychis viciifolia	Galegoide	5,8	6,1	9,1	5,2	10,9	52,6	89,7	31,3
Pisum arvense	Galegoide	2,9	3,0	6,3	4,7	11,9	54,3	83,1	25,9
Pisum sativum	Galegoide	3	2,1	7,5	4,2	9,6	61,7	88,0	23,3
Securigera varia	Galegoide	9,3	8,8	9,2	4,9	12,4	46,6	91,2	35,3
Trifolium fragiferum	Galegoide	9,7	5,8	8,5	4,2	6,1	58,0	92,2	24,5
Trigonella foenum-graecum	Galegoide	4,6	8,9	10,5	9,8	7,4	40,8	82,1	36,7
Vicia hirsuta	Galegoide	1,9	5,4	6,2	3,4	12,7	61,2	90,8	27,7
Baptisia australis	Genistoide	10,5	8,6	9,0	5,1	12,4	42,5	88,0	35,0
Bolusanthus speciosus	Genistoide	11,5	7,9	9,0	5,3	9,3	48,9	92,0	31,6
Calpurnia aureus	Genistoide	10,9	10,3	5,5	4,1	5,8	55,3	91,9	25,7
Crotalaria eremaea	Genistoide	5,7	7,3	8,4	4,3	12,0	55,8	93,5	32,0
Cytisus sessifolius	Genistoide	11,6	9,0	10,9	4,3	5,6	50,3	91,7	29,9
Daviesia corymbosum	Genistoide	1,3	5,8	10,8	4,6	11,9	55,8	90,2	33,1
Diplotropis purpurea	Genistoide	47,2	12,0	5,9	2,4	6,9	18,8	93,2	27,2
Gompholobium scabrum	Genistoide	19,7	5,1	11,3	5,4	8,1	39,3	89,0	30,0
Goodia latifolia	Genistoide	14,7	6,6	6,5	4,1	5,8	54,8	92,6	23,1
Laburnum anagyroides	Genistoide	11,3	7,1	8,3	5,1	14,0	45,5	91,3	34,5
Lupinus albus	Genistoide	9,3	8,1	8,7	10,7	12,8	45,9	95,5	40,3
Lupinus angustifolius	Genistoide	6,3	6,8	4,6	2,6	11,9	54,6	86,8	25,9
Podalyria biflora	Genistoide	14,8	8,6	8,0	4,3	16,8	29,9	82,3	37,7
Spartium junceum	Genistoide	16,7	7,7	10,5	4,0	3,0	50,0	91,9	25,2
Templetonia retusa	Genistoide	23,5	6,3	9,6	4,7	10,3	39,6	94,1	30,9
Ulex europaeus	Genistoide	10,7	7,4	8,0	4,4	11,9	46,0	88,5	31,7
Viminaria juncea	Genistoide	6,2	5,0	6,3	8,6	10,6	55,7	92,3	30,4
Sophora arizonica	Genistoide	21,4	11,0	9,9	4,2	8,4	34,7	89,5	33,4
Acacia dealbata	Mimosoideae	8,9	10,2	14,7	5,8	8,2	40,5	88,3	38,9
Albizia julibrissin	Mimosoideae	14	12,3	13,6	18,3	6,2	24,3	88,7	50,4
Albizia lebbeck	Mimosoideae	7,9	10,3	14,5	11,3	10,0	33,1	87,2	46,2
Desmanthus illinoense	Mimosoideae	3,8	5,3	6,8	9,4	6,6	57,5	89,5	28,2
Entada phaseoloides	Mimosoideae	6,7	20,3	16,7	7,7	9,5	29,5	90,5	54,3
Leucaena leucocephala	Mimosoideae	11,8	6,3	8,0	7,8	8,1	48,3	90,3	30,2
Parkia biglobosa	Mimosoideae	22,7	10,9	10,7	4,1	10,9	29,7	89,0	36,7
Prosopis africana	Mimosoideae	5	2,0	3,9	12,6	16,3	51,9	91,9	35,0
Alysicarpus ovalifolius	Phaseoloide	5,7	7,1	8,4	5,2	9,4	54,5	90,3	30,1
Bituminaria bituminosa	Phaseoloide	11,3	12,4	7,3	2,9	7,6	49,2	90,8	30,3
Canavalia brasiliensis	Phaseoloide	2,8	5,8	9,8	7,2	16,6	52,1	94,4	39,4
Desmodium canadense	Phaseoloide	9,7	8,2	7,8	7,6	9,6	44,7	87,6	33,2
Dioclea grandifolia	Phaseoloide	2,6	5,4	9,9	6,4	4,9	54,9	84,1	26,6
Erythrina crusgalli	Phaseoloide	19,9	14,0	7,7	3,2	6,8	37,7	89,3	31,7
Glycine max	Phaseoloide	20,7	9,0	3,1	5,9	7,8	38,1	84,6	25,8
Hardenbergia comptoniana	Phaseoloide	11,1	5,4	9,4	6,3	12,0	47,9	92,1	33,1
Indigofera astragalina	Phaseoloide	3,8	5,1	5,9	3,6	8,1	65,1	91,7	22,8
Kennedia nigricans	Phaseoloide	11,4	6,2	10,9	7,8	10,7	47,2	94,2	35,6
Lonchocarpus capassa	Phaseoloide	16,6	5,1	15,0	7,3	8,5	37,4	90,0	35,9
Mundulea sericea	Phaseoloide	9,2	8,0	9,9	6,2	15,8	45,1	94,2	39,9
Phaseolus vulgaris	Phaseoloide	3	4,5	2,3	4,9	10,6	60,7	86,1	22,3
Rhynchosia pentheri	Phaseoloide	5,8	4,6	8,3	3,1	14,2	56,5	92,5	30,2
Robinia pseudoacacia	Phaseoloide	14,1	4,2	8,3	7,2	12,0	46,2	92,0	31,7
Sesbania sesban	Phaseoloide	6	5,7	8,4	5,4	5,7	57,0	88,1	25,1
Vigna radiata	Phaseoloide	2,8	3,3	6,2	3,4	11,2	57,8	84,6	24,0
Vigna subterranea	Phaseoloide	5,1	5,3	0,0	7,0	12,6	57,7	87,7	24,9
Vigna unguiculata	Phaseoloide	1,2	4,6	3,9	5,5	10,1	59,9	85,3	24,2
Wisteria sinensis	Phaseoloide	10,6	11,0	16,0	7,4	7,2	37,1	89,3	41,6
Amorpha fruticosa	Sophoroide	10,7	11,0	4,8	5,6	10,1	44,9	87,0	31,4
Baphia massaiensis	Sophoroide	15,6	7,1	5,6	2,7	4,7	54,9	90,7	20,2
Pterocarpus rotundifolia	Sophoroide	21,9	10,1	4,7	4,8	21,4	26,4	89,4	41,1
Styphnolobium japonicum	Sophoroide	9,2	5,6	7,6	6,8	8,8	53,8	91,8	28,8
Swartzia madagascarensis	Sophoroide	5,6	6,3	6,3	12,6	15,8	41,8	88,4	41,1
Polygala myrtifolia	outgroup	66,7	4,8	2,6	1,4	1,7	17,3	94,5	10,6
moyennes		11,3	6,9	7,8	6,0	9,4	48,4	89,7	30,1

La somme des fractions est inférieure à 100% à cause des pertes à chaque étape.
non réserve : somme des fractions hors lipides et résidu.

v. Annexe 5 : Temps létaux 50

TL50 des différentes fractions de graines des espèces végétales sélectionnées sur S. oryzae de souche sensible (S) et résistante (R).

Les toxicités différentielles R/S significatives (p_{BGW}<0,0010) sont soulignées, et indiquées en gras dans le cas de différentiel total (pas (>20) ou très peu (>15) de mortalité).

Sous-famille ou clade (Papilionoideae)	***Espèce***	TL50 pour chaque combinaison fraction/souche de charançons (Erreur standard)									
		MeOH		**MeOH60**		**H_2O5**		**H_2O8**		**Résidu**	
		S	R	S	R	S	R	S	R	S	R
Caesalpinioideae	*Cassia occidentalis*	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
	Cercis siliquastrum	>15	>15	**5,7** (0,6)	**>15**	>20	>20	>20	>20	4,2 (0,3)	4,8 (0,4)
	Ceratonia siliqua	>20	>20	>20	>20	>15	>15	>20	>20	5,3 (0,6)	5,4 (0,3)
	Delonix regia	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
	Gleditsia tricanthos	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
	Intsia bijuga	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
	Schotia afra	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
	Tamarindus indica	6,1 (0,4)	11,7 (1,2)	>15	>20	>20	>20	>20	>20	>15	>20
Mimosoideae	*Acacia dealbata*	**10,6** (0,7)	**>20**	13,0 (3,9)	>15	**12,0** (0,8)	**>20**	13,0 (1,0)	>15	6,9 (0,2)	8,4 (0,3)
	Albizia lebbeck	4,1 (0,5)	4,3 (0,4)	4,7 (0,3)	4,7 (0,4)	4,9 (0,3)	6,8 (0,2)	7,1 (1,0)	11,8 (0,7)	6,1 (0,3)	8,4 (0,3)
	Albizia julibrissin	9,3 (0,6)	12,9 (0,4)	6,1 (0,5)	8,1 (0,2)	7,0 (1,0)	9,0 (0,5)	>15	>20	**10,0** (1,8)	**>20**
	Parkia biglobosa	**7,0** (0,8)	**>20**	>20	>20	>20	>20	>20	>20	>20	>20
Sophoroïdes	*Amorpha fruticosa*	**5,9** (0,3)	**>15**	**10,5** (1,9)	**>20**	>20	>20	>20	>20	>20	>20
	Bobgunnia madagascarensis	>20	>20	5,2 (0,4)	4,8 (0,3)	8,7 (0,6)	8,3 (0,5)	6,0 (0,6)	5,7 (0,4)	6,2 (0,4)	6,1 (0,4)
	Pterocarpus rotundifolius	13,8 (0,7)	>20	4,2 (0,3)	4,5 (0,2)	3,8 (0,2)	4,4 (0,4)	3,6 (0,2)	5,1 (0,3)	3,9 (0,2)	5,2 (0,2)
	Styphnolobium japonicum	**3,9** (0,2)	**>20**	**5,2** (0,5)	**>20**	**5,7** (0,5)	**>20**	**5,4** (0,7)	**>20**	4,9 (0,5)	6,9 (0,4)
Génistoïdes	*Baptisia australis*	**7,2** (0,7)	**>20**	**9,0** (1,9)	**>20**	>15	>20	>20	>20	>20	>20
	Bolusanthus speciosus	4,6 (0,3)	4,1 (0,3)	4,9 (0,5)	4,0 (0,2)	6,5 (0,4)	5,7 (0,6)	9,1 (0,5)	10,2 (0,5)	5,7 (0,3)	5,4 (0,4)
	Calpurnia aurea	3,9 (0,3)	4,5 (0,2)	3,8 (0,2)	4,9 (0,2)	4,6 (0,4)	8,4 (0,7)	**6,0** (0,4)	**>15**	4,6 (0,4)	5,0 (0,5)
	Crotalaria eremaea	6,3 (0,4)	5,9 (0,2)	*5,6 (0,3)*	*4,3 (0,4)*	6,4 (0,3)	6,0 (1,9)	9,7 (0,7)	11,5 (0,3)	5,7 (0,4)	5,0 (0,2)
	Laburnum anagyroides	**9,0** (1,9)	**>20**	**5,9** (0,5)	**>15**	>20	>20	>20	>20	8,0 (0,4)	13,0 (1,9)
	Lupinus albus	>20	>20	>20	>20	16,7 (0,7)	19,2 (1,2)	15,4 (0,7)	17,5 (1,0)	9,9 (0,6)	15,4 (0,5)
	Lupinus angustifolius	11,3 (0,6)	>15	**11,4** (0,4)	**>20**	9,1 (0,5)	12,2 (0,9)	6,8 (0,6)	4,9 (0,3)	6,9 (0,5)	8,2 (0,5)
	Sophora arizonica	5,6 (0,5)	4,6 (0,2)	5,4 (0,3)	5,2 (0,5)	5,3 (0,3)	8,8 (0,7)	7,4 (0,4)	11,7 (0,5)	6,8 (0,3)	9,0 (0,6)
	Templetonia retusa	**5,5** (0,4)	**>20**	5,4 (0,2)	4,6 (0,2)	11,2 (0,9)	>15	8,2 (0,5)	9,8 (0,7)	6,9 (0,5)	7,7 (0,4)
	Ulex europaeus	>15	>20	>20	>20	>20	>20	>20	>20	>20	>20

Galégoïdes	*Anthyllis barba-jovis*	>20	>20	3,7 (0,4)	4,9 (0,3)	3,7 (0,4)	4,4 (0,2)	3,9 (0,2)	5,6 (0,7)	4,1 (0,3)	4,5 (0,3)
	Astragalus monspessulanus	>20	>20	6,6 (0,4)	13,0 (0,4)	9,5 (1,8)	11,5 (3,6)	>20	>20	9,6 (0,6)	>15
	Caragana arborescens	>20	>20	5,0 (0,4)	6,3 (0,4)	12,7 (0,2)	12,4 (0,2)	>20	>20	>20	>20
	Carmichaelia stevensoni	**7,7** (0,6)	**>20**	5,3 (0,4)	6,4 (0,3)	4,7 (0,1)	6,0 (0,2)	4,8 (0,2)	7,5 (0,3)	6,1 (0,4)	8,6 (0,3)
	Cicer arietinum	>20	>20	13,0 (3,9)	>20	>20	>20	>20	>20	>20	>20
	Colutea arborescens	**6,6** (0,5)	**>20**	5,0 (0,6)	6,2 (0,4)	5,0 (0,5)	6,2 (0,2)	5,0 (0,5)	7,6 (0,2)	6,3 (0,4)	8,3 (0,4)
	Hippocrepis emerus	>20	>20	5,0 (0,4)	8,2 (0,2)	6,6 (0,3)	8,2 (0,2)	>20	>20	6,9 (0,3)	9,6 (0,3)
	Lathyrus latifolius	>20	>20	**14,7** (1,9)	**>20**	>15	>15	**3,9** (0,2)	**>20**	**6,7** (0,4)	**>20**
	Lens culinaris	>20	>20	**4,6** (0,2)	**>20**	**5,7** (0,4)	**>20**	**4,6** (0,2)	**>20**	**5,4** (0,7)	**>20**
	Lotus corniculatus	>20	>20	7,3 (0,6)	9,1 (0,2)	5,0 (0,7)	7,4 (0,2)	>20	>20	7,7 (0,4)	7,2 (0,5)
	Lotus tetragonolobus	>20	>20	>15	>20	>20	>20	>20	>20	**10,6** (0,3)	**>20**
	Medicago truncatula	3,9 (0,4)	8,8 (0,6)	3,5 (0,3)	3,6 (0,2)	4,5 (0,2)	6,4 (0,3)	4,6 (0,3)	5,2 (0,2)	4,3 (0,4)	5,2 (0,2)
	Melilotus albus	**5,9** (0,5)	**>20**	4,7 (0,4)	9,0 (1,9)	>20	>20	>20	>20	**5,2** (0,2)	**>17**
	Onobrychis viciifolia	>20	>20	**5,8** (0,2)	**>20**	**5,6** (0,3)	**>20**	**6,5** (0,5)	**>20**	5,3 (0,3)	8,2 (0,6)
	Pisum sativum	>20	>20	4,5 (0,2)	>20	5,7 (0,4)	>20	5,0 (0,4)	>20	5,3 (0,3)	>20
	Trigonella foenum-graecum	4,8 (0,3)	6,2 (0,5)	4,5 (0,3)	5,2 (0,5)	9,9 (0,9)	12,4 (0,7)	13,7 (0,6)	15,1 (0,3)	6,3 (0,4)	6,7 (0,3)
	Vicia hirsuta	>15	>15	**5,0** (1,3)	**>20**	**6,1** (0,4)	**>20**	**5,0** (0,4)	**>20**	**6,2** (0,7)	**>20**
Aeschynoménoïdes	*Abrus precatorius*	>20	>20	**6,6** (0,9)	**>20**	>20	>20	**5,5** (0,1)	**>20**	5,5 (0,2)	7,8 (0,3)
	Adesmia bicolor	>20	>20	**10,0** (1,3)	**>20**	>20	>20	>20	>20	8,9 (0,7)	13,6 (0,9)
	Arachis hypogaea	>20	>20	>20	>20	>20	>20	>20	>20	9,0	>20
	Dalbergia purpurescens	4,7 (0,4)	5,3 (0,4)	**7,7** (0,7)	**>15**	>15	>15	>20	>20	7,5 (0,7)	7,9 (0,9)
Phaséoloïdes	*Alysicarpus ovalifolius*	>20	>20	**6,7** (0,7)	**>15**	>20	>20	>20	>20	7,3 (0,4)	9,1 (1,1)
	Bituminaria bituminifolia	5,6 (0,4)	6,0 (0,4)	4,7 (0,3)	7,7 (0,2)	**11,8** (0,3)	**>20**	**8,3** (0,4)	**>20**	5,2 (0,4)	8,2 (0,3)
	Canavalia brasiliensis	>15	>20	3,8 (0,2)	3,8 (0,2)	4,8 (0,4)	5,3 (0,2)	4,4 (0,3)	5,7 (0,2)	>20	>20
	Desmodium canadense	**5,5** (0,4)	**>20**	4,9 (0,2)	7,2 (0,5)	5,4 (0,5)	8,7 (0,4)	>15	>20	**6,7** (0,6)	**>20**
	Dioclea grandiflora	>20	>20	**5,1** (0,3)	**>20**	4,7 (0,3)	4,8 (0,2)	**10,7** (0,9)	**>20**	>20	>20
	Glycine max	**7,2** (0,9)	**>20**	**6,5** (0,9)	**>20**	>20	>20	**10,8** (0,6)	**>20**	**8,0** (1,3)	**>20**
	Hardenbergia comptoniana	>20	>20	4,6 (0,2)	5,7 (0,3)	4,5 (0,2)	6,7 (0,1)	4,7 (0,2)	8,4 (0,5)	4,3 (0,2)	7,3 (0,3)

Phaséolöides	*Indigofera astragalina*	>20	>20	>20	>20	>20	>20	>20	>20	6,3 (0,3)	7,0 (0,6)
	Lonchocarpus capassa	>20	>20	3,2 (0,3)	4,3 (0,2)	3,4 (0,2)	5,0 (0,2)	3,9 (0,2)	7,6 (0,3)	4,4 (0,2)	6,4 (0,2)
	Mundulea sericea	6,8 (0,3)	6,0 (0,5)	6,2 (0,6)	6,0 (1,3)	>20	>20	**11,0** (0,8)	**>20**	5,8 (0,4)	14,0 (1,8)
	Phaseolus vulgaris	>20	>20	**6,3** (0,3)	**>20**	>20	>20	**8,7** (0,9)	**>20**	**10,6** (0,2)	**>20**
	Rhynchosia pentheri	>20	>20	**5,1** (0,3)	**>20**	**8,5** (0,7)	**>20**	5,2 (0,3)	6,5 (0,1)	>15	>20
	Robinia pseudoacacia	3,7 (0,3)	4,9 (0,2)	3,7 (0,2)	6,5 (0,3)	5,8 (0,2)	9,0 (1,9)	3,7 (0,4)	4,9 (0,2)	4,0 (0,8)	5,5 (0,3)
	Sesbania sesban	**14,2** (0,9)	**>20**	3,9 (0,2)	6,2 (0,4)	5,4 (0,7)	6,6 (0,2)	5,9 (0,4)	7,2 (0,3)	4,8 (0,3)	7,2 (0,5)
	Vigna unguiculata	>20	>20	**5,5** (0,5)	**>20**	**4,9** (0,2)	**>20**	**4,6** (0,3)	**>20**	**13,4** (0,7)	**>20**
	Vigna subterranea	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
	Wisteria sinensis	>20	>20	4,5 (0,7)	6,4 (0,3)	4,6 (0,4)	6,6 (0,3)	5,8 (0,4)	7,7 (0,2)	>20	>20

VI. Annexe 6 : Plasmide et séquences

A : carte du plasmide pET-32a et séquence de son multi-site de clonage.

B : séquence de la protéine de fusion TRX-DP-M-*PA1b*.

S D K I I H L T D D S F D T V D L K A D G A I L V D F W A E W C G P C K M I A P I L D E I A D E Y Q G K L T V A K L N I D Q N P G T A P K Y G I R G I P T L L L F K N G E V A A T K V G A L S K G Q L K E F L D A N L A G S G S G H M H H H H H H S S G L V P R G S G M K E T A A A K F E R Q H M D S P D L G T D D D D P M *A S C N G V C S P F E M P P C G T S A C R C I P V G L V I G Y C R N P S G*

La partie encadrée a été modifiée par rapport au plasmide pET32 par insertion d'oligonucléotides, en conservant les sites *Bgl*II et *Nco*I (ElMorjani et al., 1997). La séquence correspondante de TRX-DDDDK-M-PA1b est :

-PDLGTDDDDKAM*ASC*-

FOLIO ADMINISTRATIF

THESE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

NOM : **LOUIS** DATE de SOUTENANCE : 20 février 2004
(avec précision du nom de jeune fille, le cas échéant)

Prénoms : **Sandrine**

TITRE :

Diversité structurale et d'activité biologique des albumines entomotoxiques de type A1b des graines de Légumineuses.

NATURE : Doctorat Numéro d'ordre : **04 ISAL 012**

Ecole doctorale : Evolution, Ecosystèmes, Microbiologie, Modélisation

Spécialité : Analyse et Modélisation des Systèmes Biologiques

Cote B.I.U. - Lyon : T 50/210/19 / et bis CLASSE :

RESUME :

PA1b (Pea Albumin 1 sous-unité b), une knottine toxique de 37 acides aminés, présente un grand intérêt dans la lutte contre les charançons des céréales (*Sitophilus spp*.), principaux ravageurs des céréales stockées.
Afin de mieux connaître la nouvelle famille peptidique de PA1b, sa var iabilité tant structurale que d'activité biologique a été étudiée au sein des Légumineuses. Après avoir validé notre approche sur 4 espèces végétales "test", nous avons caractérisé 24 gènes homologues chez 18 espèces de Papilionoideae. De plus, l'activité insecticide d'extraits de graines de 60 espèces des trois sous-familles de Légumineuses a été déterminée sur charançons de souche sensible et résistante à PA1b. Afin de relier variations de structure et d'activité, une approche par mutagenèse dirigée a été envisagée. Un système d'expression bactérienne et de purification de PA1b a été mis au point. Bien que de masse conforme (cystéines oxydées), le peptide recombinant ne présente pas d'activité biologique.

MOTS-CLES :

Albumine 1, A1b, Fabaceae, Curculionidae, phytotoxine, peptide de défense, résistance des plantes, expression hétérologue, cystine-knot

Laboratoire (s) de recherches :

UMR Biologie fonctionnelle, Insectes et Interactions (BF2I) INRA-INSA de Lyon
Unité de Recherche sur les Protéines Végétales et leurs Interactions (URPVI) INRA de Nantes

Directeur de thèse:

Yvan RAHBE

Président de jury :
J.M. LANCELIN

Composition du jury :

J.P. BLEIN, C. DOMONEY, J. GUEGUEN, E. GUIDERDONI, J.M. LANCELIN, Y. RAHBE

Diversité structurale et d'activité biologique des albumines entomotoxiques de type A1b des graines de Légumineuses.

Résumé :

PA1b (Pea Albumin 1 sous-unité b), une knottine toxique de 37 acides aminés, présente un grand intérêt dans la lutte contre les charançons des céréales (*Sitophilus spp.*), principaux ravageurs des céréales stockées.
Afin de mieux connaître la nouvelle famille peptidique de PA1b, sa variabilité tant structurale que d'activité biologique a été étudiée au sein des Légumineuses. Après avoir validé notre approche sur 4 espèces végétales "test", nous avons caractérisé 24 gènes homologues chez 18 espèces de Papilionoideae. De plus, l'activité insecticide d'extraits de graines de 60 espèces des trois sous-familles de Légumineuses a été déterminée sur charançons de souche sensible et résistante à PA1b. Afin de relier variations de structure et d'activité, une approche par mutagenèse dirigée a été envisagée. Un système d'expression bactérienne et de purification de PA1b a été mis au point. Bien que de masse conforme (cystéines oxydées), le peptide recombinant ne présente pas d'activité biologique.

Variability in structure and biological activity of A1b insecticidal albumins from legume seeds.

Résumé :

PA1b (Pea Albumin 1 subunit b), a 37 amino acid cystine-knot toxin, is of great interest for the control of cereal weevils (*Sitophilus spp.*), major pests of stored cereals.
To know more on the new A1b family, its variability in structure and biological activity was studied. After the validation of our approach on four "test" plant species, we characterised 24 homologous genes from 18 Papilionoideae species. Moreover, the insecticidal activity of seed extracts, from 60 species belonging to the three legumes subfamilies, was determined on weevil strains susceptible and resistant to PA1b. To link structure and activity variations, a mutagenesis approach was proposed. A bacterial expression and purification system for PA1b was determined. However the recombinant peptide was not biologically active, despite harbouring the expected mass (with three oxidised disulfide bridges).