



ÉTUDE STRUCTURALE ET DYNAMIQUE PAR RÉSONANCE MAGNÉTIQUE NUCLÉAIRE, D'UNE PROTÉINE DE TRANSFERT DE LIPIDES ISOLÉE DANS LE TABAC

Pedro da Silva

► To cite this version:

Pedro da Silva. ÉTUDE STRUCTURALE ET DYNAMIQUE PAR RÉSONANCE MAGNÉTIQUE NUCLÉAIRE, D'UNE PROTÉINE DE TRANSFERT DE LIPIDES ISOLÉE DANS LE TABAC. domain_other. Université d'Orléans, 2004. Français. NNT : . tel-00009320v2

HAL Id: tel-00009320

<https://theses.hal.science/tel-00009320v2>

Submitted on 28 Jul 2005 (v2), last revised 11 Aug 2005 (v3)

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.



THESE

PRESENTEE

A L'UNIVERSITE D'ORLEANS

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITE D'ORLEANS

Discipline : Biophysique Moléculaire

Par

Pedro Da Silva

**Étude structurale et dynamique
par résonance magnétique nucléaire,
d'une protéine de transfert de lipides isolée dans le tabac**

Soutenue le 10 décembre 2004

MEMBRES DU JURY :

- | | |
|--|--------------------|
| • Jean Pierre Blein ; Directeur de recherche, INRA, Dijon | Président de jury |
| • Françoise Vovelle ; Professeur, Université d'Orléans | Directeur de thèse |
| • Muriel Delepierre ; Directeur de recherche, CNRS, Paris | Rapporteur |
| • Jean Pierre Simorre ; Directeur de recherche, CNRS, Grenoble | Rapporteur |
| • Céline Landon ; Chargée de recherche, CNRS, Orléans | Examineur |
| • Luc Morin-Allory ; Professeur, Université d'Orléans | Examineur |



THESE

PRESENTEE

A L'UNIVERSITE D'ORLEANS

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITE D'ORLEANS

Discipline : Biophysique Moléculaire

par
Pedro Da Silva

**Étude structurale et dynamique
par résonance magnétique nucléaire,
d'une protéine de transfert de lipides isolée dans le tabac**

Soutenue le 10 décembre 2004

MEMBRES DU JURY :

- Jean Pierre Blein ; Directeur de recherche, INRA, Dijon
- Françoise Vovelle ; Professeur, Université d'Orléans
- Muriel Delepierre ; Directeur de recherche, CNRS, Paris
- Jean Pierre Simorre ; Directeur de recherche, CNRS, Grenoble
- Céline Landon ; Chargée de recherche, CNRS, Orléans
- Luc Morin-Allory ; Professeur, Université d'Orléans

Président de jury
Directeur de thèse
Rapporteur
Rapporteur
Examineur
Examineur

Remerciements

Ce travail de thèse a été effectué au Centre de Biophysique Moléculaire du CNRS à Orléans dirigé par Monsieur Jean-Claude Beloeil.

Je tiens à remercier chaleureusement toutes les personnes qui ont partagé mon quotidien et qui ont rendu possible ce travail.

Je remercie tout particulièrement tous les membres de l'équipe de RMN et de modélisation moléculaire de protéines et de peptides que dirige avec autorité le Professeur Françoise Vovelle et dont les membres sont ou ont été : Henri Iabbé, Anita Caille, Denise Sy, Florent Barbault, Romain Beltoise, Loïc Burnel, Gilles Rautureau, Hervé Meudal, Laurence Jouvensal et Céline Landon. J'ai eu la chance au cours de ma thèse et de mon DEA de travailler avec tous les membres de l'équipe et je tiens à les remercier de m'avoir fait partager leurs savoir-faire, leurs connaissances et leur bonne humeur.

D'autre part, ma thèse m'a permis de collaborer avec de nombreuses équipes et je voudrais ici exprimer mes sincères remerciements à Michel Ponchet et Benoît Industri de l'INRA d'Antibes à Lionel Dubois, Hervé Desvaux, Patrick Berthault et Gaspard Huber du CEA de Saclay, à Stéphanie Duclos, Véronique et Friedrich Piller du CBM d'Orléans, à toutes les personnes impliquées dans le réseau LTP et enfin à Norbert Garnier du CBM d'Orléans.

Je remercie également toutes les personnes qui ont participé aux moments de détente, pauses café, pots ... et qui ont su créer un climat agréable pendant mes quatre années passées au CBM ; encore MERCI.

Et enfin, je voudrais dédier ma thèse à Claire, ma femme et à ma famille qui m'ont entouré et supporté dans tous les sens du terme.

Sommaire

Introduction	11
I) Les protéines de transfert de lipides d'origine végétale	15
1. LTPs : famille multigénique	17
2. LTPs : propriétés structurales	21
2.1 LTP1 : structures tridimensionnelles sous forme libre.....	21
2.2 LTP1 : structures de complexes	24
2.3 LTP2 : structures tridimensionnelles	26
2.4 Protéines de même repliement	27
2.4.1 Allergie alimentaire.....	28
3. Activités biologiques.....	29
3.1 La cuticule	29
3.2 La subérine.....	31
3.3 LTPs : activité antimicrobienne.....	32
3.4 LTPs : signalisation cellulaire	33
3.4.1 Les élicitines	33
3.4.2 Dialogue entre LTPs et élicitines.....	37
4. Ligands naturels	38
4.1 Ligand de la LTP1 d'orge.....	38
4.2 Ligand de la LTP1 de blé.....	39
4.3 Nouvel éclairage sur les structures des ligands des LTP1 d'orge et de blé.....	40
II) Matériels et Méthodes	41
1. Détermination de la structure tridimensionnelle d'une protéine (par RMN et modélisation moléculaire).....	43
1.1 Méthodes d'attribution RMN des protéines	44
1.1.1 Méthode d'attribution séquentielle.....	44
1.1.2 Attribution facilitée par un simple marquage ¹⁵ N ou ¹³ C.....	45
1.1.3 L'expérience 3D ¹ H- ¹⁵ N NOESY-HSQC	46
1.1.4 Expériences tridimensionnelles triple résonance	47
a) HNCA et HN(CO)CA	48
b) HN(CA)CO et HNCO	49
c) CBCANH et CBCA(CO)NH	50
1.1.5 Attribution des chaînes latérales.....	50
1.2 Les contraintes structurales utilisées en modélisation moléculaire	51
1.2.1 Les contraintes de distance.....	51
1.2.2 Les contraintes d'angles dièdres	51
1.2.3 Données qualitatives.....	54
1.3 Modélisation moléculaire.....	54
1.3.1 Le logiciel CNS	54
a) Fonction empirique de l'énergie d'une molécule	55
b) Recuit simulé	55
1.3.2 Le logiciel ARIA	56
a) Contraintes de distance ambiguë.....	56
b) Principe de fonctionnement du logiciel ARIA.....	56
c) Calibration des pics NOE	58
d) Attribution des pics NOE ambigus	60
1.3.3 Sélection et analyse des structures	61
a) Nombre de violations et géométrie covalente.....	61

b) Les logiciels d'analyse PROCHECK et PROMOTIF	61
2. Etude dynamique interne des protéines par relaxation nucléaire	63
2.1 Mouvements moléculaires des protéines	63
2.2 La fonction de densité spectrale	64
2.3 La relaxation nucléaire	64
2.3.1 La relaxation longitudinale	65
2.3.2 La relaxation transversale	66
2.3.3 La relaxation croisée ; l'effet NOE η	66
2.3.4 Paramètres expérimentaux mesurés	67
2.3.5 Le phénomène d'échange	68
2.4 Détermination expérimentale des paramètres de relaxation du noyau ^{15}N	68
2.4.1 Mesure de la relaxation longitudinale R_1 et transversale R_2	68
2.4.2 Mesure de l'effet NOE	69
2.5 Interprétation des paramètres de relaxation en terme de mouvement	70
2.5.1 Le modèle de Lipari et Szabo	70
2.5.2 Modèle de Lipari Szabo étendu	72
2.5.3 Détermination du temps de corrélation globale de la protéine τ_c	73
2.5.4 Modelfree et Fast-Modelfree	73
2.5.5 Choix des modèles	74
3. Simulations de dynamique moléculaire	76
3.1 La fonction d'énergie potentielle	76
3.2 Les simulations de dynamiques moléculaires	78
3.2.1 Résolution des équations du mouvement	79
3.2.2 Les conditions initiales	79
3.2.3 Dynamique moléculaire à température constante	80
3.2.4 Dynamique moléculaire à pression constante	81
3.3 Utilisation du logiciel CHARMM	81
3.3.1 Mise en œuvre des simulations	81
4. Principe de la modélisation comparative	83
4.1 Recherche et choix des modèles	84
4.2 Alignement des séquences	85
4.3 Construction du modèle	87
4.4 Evaluation du modèle	88
III) Etude structurale des LTP1 de tabac	89
1. La famille des LTP1 de <i>Nicotiana tabacum</i>	91
2. Etude structurale de la LTP1_1 de tabac	92
2.1 Purification de la LTP1_1 de tabac	93
2.2 Article sur la LTP1_1 de tabac : structure 3D et études d'interactions	94
2.3 Conclusions et perspectives	112
3. Modélisation comparative des LTP1 identifiées dans le tabac	114
3.1 Modèle structural	114
3.2 Alignement multiple	114
3.3 Les structures 3D modèles	115
3.4 Conclusion et perspectives	118
IV) Dynamique interne de la LTP1_1 de tabac	121
1. Etude par RMN de la dynamique interne de la LTP1_1 de tabac	123
1.1 Détermination des paramètres de relaxation du noyau ^{15}N	123
1.1.1 Mesures des vitesses de relaxation longitudinale R_1 et transversale R_2	123
1.1.2 Mesure de l'effet NOE hétéronucléaire ^1H - ^{15}N	126
1.1.3 Détermination du temps de corrélation τ_c	127

1.2 Interprétation des paramètres de relaxation par le formalisme de Lipari-Szabo.	128
1.3 Discussion	131
2. Simulation de dynamique moléculaire à 293K	132
2.1 Stabilité de la LTP1_1 de tabac	133
3. Simulations de dynamique moléculaire à 500K	136
3.1 Comportement de la LTP1_1	137
3.2 Discussion	141
4. Conclusion et perspectives	143
V) Cartographie de la cavité hydrophobe	145
1. Xénon polarisé par laser	148
2. Le SPINOE	150
3. Etude de la cavité hydrophobe de la LTP1_1 de tabac	150
4. Conclusions et perspectives	167
Conclusions et perspectives	169
Bibliographie	175
Annexe	193

Abréviations

2D : bidimensionnel(le)
3D : tridimensionnel(le)
Ace-AMP1 : *Allium cepa* Antimicrobial Protein
ADNc : Acide DésoxyriboNucléique complémentaire
ADR : Ambiguous Distance Restraint
ARIA : Ambiguous distance Restraints for Iterative Assignments
ARNm : Acide RiboNucléique messenger
BLAST : Basic Alignment Search Tool
BLOSUM : BLOck Substitution Matrix
CHARMM : Chemistry at HARvard Macromolecular Mechanics
CNS : Crystallography NMR System
COSY : COrrrelation SpectroscopY
DMPG : DiMyristoylPhosphatidylGlycérol
HCA : Hydropobic Cluster Analysis
HOHAHA : HOmonuclear HArtman HAhn spectroscopy
HPLC : High Performance Liquid Chromatography
HPS : Hydrophobic Protein of Soybean
HSQC : Heteronuclear Single Quantum Coherence
INEPT : Insensitive Nuclei Enhanced Polarisation Transfer
LLPG : LysoLauroylPhosphatidylGlycérol
LMPC : LysoMyristoylPhosphatidylCholine
LOX : Lipoxygénase
LTPs : Lipid Transfer Proteins
NMR : Nuclear Magnetic Resonance
NOE : Nuclear Overhauser Effect
NOESY : Nuclear Overhauser Effect SpectroscopY
PAM : Percent Accepted Mutation
PDB : Protein Data Bank
pI : point Isoléctrique
POX : Péroxygénase
PRP : Pathogenesis Related Protein
PSI-BLAST : Position Specific Iterated BLAST
RH : Réponse Hypersensible
RMN : Résonance Magnétique Nucléaire
RMSD : Root Mean Square Deviation
RSA : Réponse Systémique Acquise
RX : Rayons X
SDS-PAGE : Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis
TOCSY : Total Correlation SpectroscopY
TROSY : Transverse Relaxation Optimized SpectroscopY

Introduction

Les lipides sont des molécules biologiques qui jouent un rôle clef dans la physiologie des organismes vivants. Sous la forme de phospholipides et de glycolipides associés à des protéines, ils sont les constituants principaux des membranes biologiques qui assurent la compartimentation cellulaire. Chez les végétaux supérieurs, ils constituent les briques élémentaires des couches de protection et d'imperméabilisation des organes de la plante (la cutine et la subérine) qui constituent la première barrière physique aux agressions environnementales (stress hydrique, bioagresseurs...). Toutes ces fonctions nécessitent un trafic intra et extracellulaire de ces molécules hydrophobes, qui doivent donc être transportées par des molécules hydrophiles. C'est dans ce contexte qu'une famille de petites protéines hydrophiles a été identifiée : les LTPs pour Lipid Transfer Proteins.

Depuis la montée en puissance des outils de génomique et de transcriptomique, de nombreux ARNm (ADNc) ou gènes codant des séquences homologues aux protéines de transfert de lipides ont été isolés dans des organes et contextes physiologiques variés. Les LTPs d'origine végétale sont de petites protéines basiques possédant 8 résidus cystéine qui sont regroupées en 2 familles : les LTP1 (9kDa) et les LTP2 (7kDa). Ces 2 familles sont multigéniques et plus de 150 séquences sont présentes dans les bases de données. Bien que la fonction biologique des LTPs d'origine végétale ne soit pas connue, une des hypothèses les plus étayées par de nombreux travaux, est celle d'un rôle dans les mécanismes de défense des plantes. En effet, il a été montré que ces protéines participeraient à la synthèse de la cutine ou de la subérine ; certaines présentent une activité antifongique et/ou seraient impliquées dans la signalisation cellulaire.

Un réseau pluridisciplinaire français impliquant des équipes de l'INRA, du CNRS et de l'Université a été mis en place dans le but de préciser la ou les fonction(s) biologique(s) des LTPs en s'appuyant sur des études de relations structure-activité. Depuis plusieurs années, l'équipe de RMN et modélisation des peptides et protéines du Centre de Biophysique Moléculaire d'Orléans participe activement à ce réseau pluridisciplinaire par la détermination des structures en solution de LTPs libres ou sous forme de complexes. Au début de ma thèse les études structurales étaient limitées aux LTPs extraites de graines de céréales (Monocotylédones). Nous disposions donc de données structurales extrêmement réduites et probablement incomplètes si l'on considère que les LTPs sont décrites chez tous les végétaux supérieurs et qu'elles sont exprimées de manière différentielle selon les organes et les stades de développement de la plante. C'est pourquoi nous avons entrepris d'étudier une LTP1 identifiée chez une plante modèle : le tabac *Nicotiana tabacum*, en collaboration avec l'INRA d'Antibes.

L'objectif de mon travail de thèse a donc été dans un premier temps d'élargir la base de données structurales des LTPs, en déterminant la structure 3D par RMN de la première LTP1 constitutive d'organes végétatifs. Cette protéine, nommée LTP1_1, a été surexprimée dans la levure *Pichia pastoris*, permettant ainsi son marquage isotopique ^{15}N et ^{13}C . Nous avons déterminé sa structure tridimensionnelle par RMN 2D, 3D et modélisation moléculaire et avons réalisé différentes études d'interactions et de dynamique qui nous ont permis de décrire les propriétés biophysiques particulières de cette nouvelle LTP1.

Ce mémoire s'articule en cinq parties. La première présente la bibliographie des LTPs d'origine végétale. La seconde est consacrée à l'exposé des techniques utilisées pour réaliser ce travail. Les trois chapitres suivants sont dédiés à la présentation des résultats obtenus : le premier décrit la structure tridimensionnelle obtenue par RMN 2D, 3D et modélisation moléculaire et les études d'interaction de cette nouvelle LTP1_1 avec des molécules lipidiques. Ce chapitre présente également des résultats de modélisation comparative que nous avons obtenus pour les séquences de LTP1 identifiées chez *N. tabacum*. Le second chapitre détaille les résultats des études de dynamique réalisées par RMN et *in silico* de la LTP1_1. Enfin, le troisième expose les résultats de l'exploration de la cavité hydrophobe de la LTP1_1 en utilisant la RMN du xénon hyperpolarisé. Par ailleurs, en marge de mes principaux travaux de thèse, j'ai réalisé un travail de modélisation comparative en collaboration avec l'équipe de biologie cellulaire et moléculaire de la glycosylation du Centre de Biophysique Moléculaire d'Orléans. Ces résultats sont présentés en annexe de mon manuscrit sous la forme d'un article qui a été accepté dans le journal *Protein Engineering Design and Selection*

I) Les protéines de transfert de lipides d'origine végétale

Les protéines de transfert de lipides d'origine végétales (LTPs pour Lipid Transfer Proteins) forment une famille ubiquitaire chez les plantes supérieures. Elles sont présentes dans tous les organes végétaux et sont particulièrement abondantes dans les graines où elles peuvent représenter jusqu'à 5% des protéines solubles (Kader 1997). Découvertes en 1975, lors de travaux sur le trafic intracellulaire des lipides (Kader 1975), on a longtemps supposé qu'elles participaient au transfert de lipides d'une membrane à une autre, en raison de leur capacité de fixation et de transfert, *in vitro*, de molécules aussi bien amphiphiles qu'hydrophobes (Kader 1997). Mais en 1988, (Tchang et al. 1988) le premier ARN messenger (*via* son ADNc ou ADN complémentaire) codant une LTP a été identifié, révélant la présence d'un peptide signal. Cela signifie que les protéines matures correspondantes sont adressées vers le compartiment extracellulaire (Bernhard et al. 1991; Sterk et al. 1991). L'hypothèse de la participation des LTPs au trafic intracellulaire des lipides était ainsi sérieusement mise à mal et a été par la suite définitivement exclue. Plus de 150 séquences d'ADNc ou de gènes codant des LTPs ont été identifiées dans des organes et des contextes physiologiques variés. Pour une plante donnée, la diversité tant génétique que spatiotemporelle des LTPs suggère que ces protéines jouent plusieurs rôles au sein de la plante. Toutefois ces fonctions restent encore inconnues. Dans la suite de ce chapitre, je vais présenter la famille des LTPs, leurs structures tridimensionnelles, ainsi que leurs fonctions biologiques supposées.

1. LTPs : famille multigénique

Les LTPs forment une famille multigénique chez les plantes supérieures : environ 170 séquences appartenant à 60 espèces différentes sont répertoriées dans la banque de données de séquences protéiques UniProt, (Apweiler et al. 2004). Par exemple, pour la plante modèle *Arabidopsis thaliana* dont le génome a été totalement séquencé (Nature, 2000) plus de trente gènes putatifs codant des LTPs ont été identifiés. De plus, le profil d'expression a été déterminé pour six d'entre elles et il a été montré qu'au niveau de la feuille et du fruit d'*Arabidopsis thaliana* (les siliques) cinq LTPs s'expriment en même temps (Arondel et al. 2000).

En fait deux groupes de LTPs sont connus à ce jour : les LTP1 et les LTP2, ayant respectivement des masses moléculaires d'environ 9kDa et 7kDa. La majorité (environ 80 %) des séquences protéiques de LTPs déposées dans les banques de données correspondent à la famille des LTP1.

Code d'accès UniProt	Séquences de protéines matures (sans peptide signal)	origine	organe	pI
	10 20 30 40 50			
NLT1_ARATH	-ALS C GSVNSNLAA C IGYVLQ-GGV I PPA C SGVKNLSIAKTTP DR QQA	<i>A.thaliana</i>	fleur	9.4
NLTA_BRAOL	-ALT C GTVNSNVAP C IGYITQ-GGT L PGA C CTGVSKLNSMARTTP DR QQA	choux	feuille	9.1
NLTP_MALDO	-A I T C GQVTS L AP C IGYVRS-GGAVPPA C CNG I RTINGLARTTA DR QTA	pommier	fruit	9.3
NLT1_PRUDO	-- I T C GQVSSNLAP C IN Y VKG-GGAVPPA C CNG I RVNNLARTTA DR RAA	prunier	fruit	9.4
NLT4_ARATH	-A I T C GTVASSLSP C LGYLSK-GGVVPPP C CAGVKKLNGMAQTTP DR QQA	<i>A.thaliana</i>	fleur	9.2
NLT1_WHEAT	-- I D C GHVDSLVR P C LSYVQG-GPG S SGQ C CDGVKNLHNQARSQS DR QSA	blé	graine	8.2
NLT1_HORVU	-- L N C GQVDSK M K P C LT Y VQG-GPG S SGE C CNGVRDLHNQAQSSG DR QTV	orge	graine	8.2
NLTP_MAIZE	-A I S C GQVASAIAP C IS Y ARGQSG S SAG C CSGVRS L NNAARTTA DR RAA	maïs	graine	9.1
NLT1_ORYSA	-- I T C GQVNSAVGP C LT Y ARG-GAG S SA C CSGVRS L FAAASTTA DR RTA	riz	graine	9.4
NLTP_BETVU	-A I T C GLVASKLAP C IGYLQG-APG S SA C CGG I KSLSNAAASP DR KTA	betterave	graine	9.2
NLTP_PRUDU	A I T C GQVSSNLAP C IPYVRG-GGAVPPA C CNG I RVNNLARTTP DR QAA-	amandier	fleur	9.1
NLT6_AMBAR	AS P T C DTVQNI L AP C AG F LTG--QE S KA C CTGVNNLNSRKT K A DR VAV	ambroisie	fleur	8.9
Consensus	* C * * C * + * * CC . * * DR			
	60 70 80 90			
NLT1_ARATH	C NC I QGAARALGSGLNAGRAAG I PKA C GVNIP Y -KISTSTNC K TVR- 93	<i>A.thaliana</i>	fleur	9.4
NLTA_BRAOL	C RC L ETAARALGPNLNAGRAAG I PKA C GVSV P F-P I STNTNC N V K - 93	choux	feuille	9.1
NLTP_MALDO	C NC L KNLAGS I SG-VNPNNAA L PGK C GVNP Y -KISTSTNC A TV K - 92	pommier	fruit	9.3
NLT1_PRUDO	C NC L KQLSGS I PG-VNPNNAA L PGK C GVNP Y -K I ASTNC A TV K - 91	prunier	fruit	9.4
NLT4_ARATH	C RC L QSA A K----GVNPSLASGLPGK C GV S IP Y -P I STSTNC A TI K - 89	<i>A.thaliana</i>	fleur	9.2
NLT1_WHEAT	C NC L KGIARG I HN-LNEDNARS I PPK C GVNL P Y -T I SLNID C SR V -- 90	blé	graine	8.2
NLT1_HORVU	C NC L KGIARG I HN-LNLNNAAS I PSK C VNP Y -T I SPDID C SR I Y - 91	orge	graine	8.2
NLTP_MAIZE	C NC L KNAAG V SG-LNAGNAAS I PSK C GV S IP Y -T I STSTD C SR V N- 93	maïs	graine	9.1
NLT1_ORYSA	C NC L KNAARG I KG-LNAGNAAS I PSK C GV S IP Y -T I SASID C SR V S- 91	riz	graine	9.4
NLTP_BETVU	C T C LKSAAT S IKG-INYGKAAS L PRQ C GV S IP Y -A I SPNTNC A TH- 92	betterave	graine	9.2
NLTP_PRUDU	C NC L KQLSAS V PG-VNPNNAA L PGK C GVNP Y -Q I SPSTNC A NV K - 91	amandier	fleurs	9.1
NLT6_AMBAR	C NC I KE L TKS I AY--DPKRM P LL S TK C GV K PD F PAVDKNLD C SK L PV 93	ambroisie	fleur	8.9
Consensus	C C * * * * C * + * * C : . *			
	10 20 30			
NL21_WHEAT	---A C QASQ--LAV C ASAI L S--GAKPSGE C C G N L R-----AQQG	blé	graine	8.2
NLT2_HORVU	---A C EPAQ--LAV C ASAI L G--GTKPSGE C C G N L R-----AQQG	orge	graine	7.0
NLTX_ORYSA	--AG C NAGQ--L T V C TGA I AG--GARPTA C CS L R-----AQQG	riz	graine	9.1
NLT2_MAIZE	-AN P C NPAQ--L T P C AGP A L F --GGAVPPA C CA L R-----AQQG	maïs	graine	9.1
NLTP_VIGUN	--VT C NPTE--L S S C VPA I TG--GSKPSST C CS K L K -----VQEP	haricot	racine	8.9
Consensus	. C * C * CC L			
	40 50 60 70			
NL21_WHEAT	C F C QYAKDPTYG Q Y IRSPHARDTL T S C GLAVP----- H C ----- 67	blé	graine	8.2
NLT2_HORVU	C L C QYVKDPNYGH Y VSSPHARDTL N L C GIPV----- H C ----- 67	orge	graine	7.0
NLTX_ORYSA	C F C QFAKDPRYGR Y VNSPNARKAVSS C GIALP----- T C H----- 69	riz	graine	9.1
NLT2_MAIZE	C L C G Y ARSPNYGS Y IRSPNAAR L FA I C N L P M----- R C R----- 70	maïs	graine	9.1
NLTP_VIGUN	C L C N Y IKNPS L K Q Y VNSPGAK V LS N C GV T YP----- N C ----- 68	haricot	racine	8.9
Consensus	C * C + * * C C			

Figure 1.2 : Alignement multiple de quelques séquences de LTP1 et de LTP2 présentes dans la banque de séquences protéiques UniProt. * et + sont respectivement les positions hydrophobes et aromatiques conservées parmi les séquences de LTPs répertoriées.

Les séquences protéiques des LTPs sont pour la plupart déduites de leur ARNm (ADNc) et seulement quelques protéines (environ une vingtaine) ont été séquencées et caractérisées. Enfin, la propriété de transfert de lipides n'a été étudiée réellement que sur un échantillon de LTPs encore plus réduit (environ une dizaine). Par conséquent, parmi les 170 séquences référencées comme étant des LTPs dans les banques de données, on ne connaît pas le vrai nombre de « LTP » possédant une activité de transfert de lipides et celui de protéines apparentées qui ne fixent pas de lipides.

Les LTP1 et LTP2 ne présentent pas ou peu d'homologie de séquence à l'exception d'un même motif cystéine (figure 1.1). Ces cystéines sont toutes impliquées dans la formation de ponts disulfure dont les connexions sont similaires, sauf un « mésappariement » au niveau du motif consensus CXC entre LTP1 et LTP2 (Douliez et al. 2001b) (figure 1.1). On peut noter qu'au niveau de ce motif CXC le résidu central X est hydrophobe pour les LTP2 et hydrophile pour les LTP1 (figure 1.2). Cette différence suggère que la nature du résidu X est importante, voire déterminante, pour l'organisation du réseau de ponts disulfure et peut-être même pour la structure tridimensionnelle des LTPs.

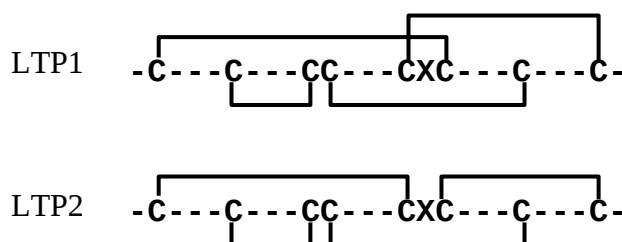


Figure 1.1 : Comparaison du réseau de ponts disulfure des familles LTP1 et LTP2

Au niveau de leur séquence, les LTPs sont dépourvues de résidus tryptophane. Les LTP1 possèdent généralement deux résidus tyrosine qui sont respectivement dans les parties N- et C-terminale. De plus, parmi toutes les séquences de LTP1 répertoriées dans les banques de données de nombreux résidus hydrophobes sont très bien conservés, ainsi qu'un motif hydrophile « DR » (figure 1.2). Ceci indique que ces résidus sont probablement importants aussi bien pour le rôle biologique des LTP1 que pour leur repliement. Les LTP2 possèdent elles aussi des positions hydrophobes et hydrophiles très bien conservées mais différentes de celles des LTP1, suggérant peut-être une fonction différente au sein de la plante. Enfin, les LTP1 et LTP2 sont des protéines généralement basiques avec un pI (point isoélectrique) supérieur ou égal à 8 (figure 1.2). Cependant, deux LTPs acides ont été identifiées dans le milieu extracellulaire de cellules embryogéniques de *Cichorium* en culture (Blanckaert et al. 2002). Chez *A. thaliana* plusieurs gènes codant des LTPs acides ont également été mis en évidence (Arondel et al. 2000).

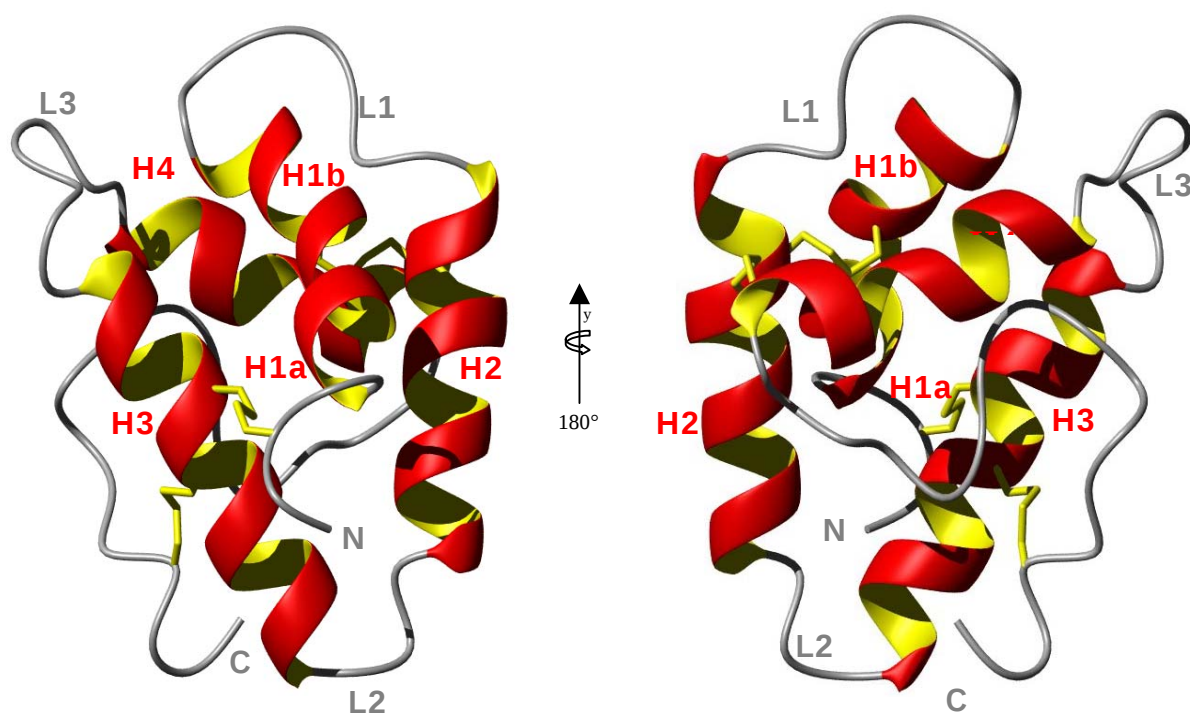


Figure 1.3 : Illustration de la structure tridimensionnelle de la LTP1 de blé représentative de la famille LTP1. Les ponts disulfure sont représentés en jaune.

2. LTPs : propriétés structurales

Les LTPs sont des protéines très stables. Par exemple une LTP1 d'orge résiste bien à des traitements par des dénaturants chimiques (chlorure de guanidinium), par des températures élevées ($> 95\text{ }^{\circ}\text{C}$) et par des protéases (Lindorff-Larsen and Winther 2001). On observe la même stabilité pour une LTP2 de riz (Samuel et al. 2002). Cette remarquable stabilité est certainement importante pour les fonctions biologiques des LTPs et est sûrement due à leur structure tridimensionnelle (3D) particulière très compacte. Actuellement, seules les structures tridimensionnelles de LTPs extraites de graines de céréales sont connues : blé, maïs, orge, riz.

2.1 LTP1 : structures tridimensionnelles sous forme libre

Historiquement, pour les LTP1, il a tout d'abord été proposé un modèle de structure 3D essentiellement composée de feuillets β . Cette hypothèse était basée sur l'analyse des séquences de LTP1 et sur des prédictions de structures secondaires (Tchang et al. 1988). Ces prédictions ont été invalidées par des études spectroscopiques RMN, infrarouge et Raman qui ont révélé que les LTP1 étaient principalement formées d'hélices α (Simorre et al. 1991; Desormeaux et al. 1992). Notre équipe a confirmé ces résultats préliminaires grâce à la détermination de la première structure 3D de la LTP1 de blé (Gincel et al. 1994). Depuis, trois autres structures de LTP1 ont été obtenues par RMN ou radiocristallographie : les LTP1 de graines de maïs (Shin et al. 1995; Gomar et al. 1996), d'orge (Heinemann et al. 1996), et de riz (Lee et al. 1998; Poznanski et al. 1999) (tableau 1.1).

La structure des LTP1 est formée de 4 hélices α (H1 à H4) reliées entre elles par des boucles (L1 à L3). Ces quatre hélices sont enroulées en spirale droite et suivies par un long segment C-terminal qui ne présente aucune structure secondaire canonique. La première hélice H1 présente une courbure, dont la valeur varie d'une LTP1 à l'autre, et se divise ainsi en deux hélices H1a et H1b. Ce repliement est stabilisé par quatre ponts disulfure répartis deux par deux de part et d'autre de la structure. Ils relient les hélices entres elles et les extrémités N- et C-terminales aux hélices (figure 1.3). L'existence d'une cavité tapissée d'acides aminés hydrophobes constitue l'originalité du repliement des LTP1. Les quatre LTP1 présentent toute cette cavité dont la taille varie d'une protéine à l'autre (tableau 1.1). Cette cavité hydrophobe confère aux LTP1 la remarquable propriété de fixer toutes sortes de molécules hydrophobes, à l'exception des stérols (Kader 1996; Douliez et al. 2000; Douliez et al. 2001a; Pato et al. 2001).

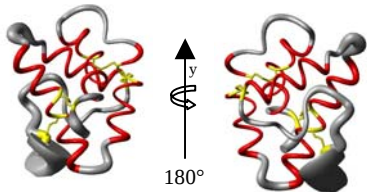
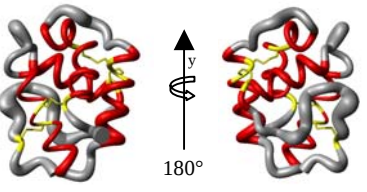
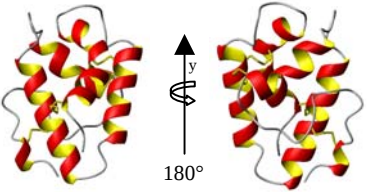
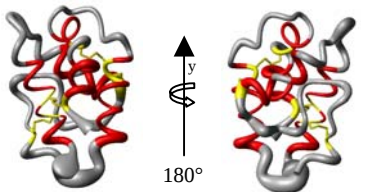
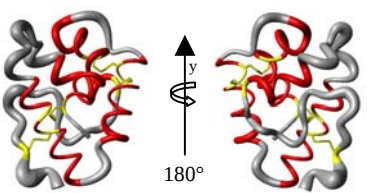
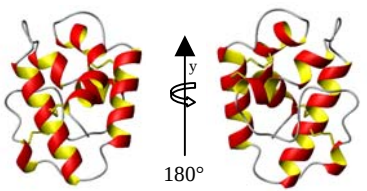
Source	Code PDB Technique	Structures 3D des LTP1	Volume de la Protéine (Å ³)	Volume de la cavité (Å ³)	Références
Graine de blé	1GH1 RMN 15 modèles		7780	330 ± 50	(Gincel et al. 1994)
Graine de maïs	1AFH RMN 15 modèles		7300	290 ± 100	(Gomar et al. 1996)
	1MZL RX		7380	420	(Shin et al. 1995)
Graine d'orge	1LIP RMN 4 modèles		7750	160 ± 30	(Heinemann et al. 1996)
Graine de riz	1BV2 RMN 14 modèles		7260	240 ± 70	(Poznanski et al. 1999)
	1RZL RX		7320	100	(Lee et al. 1998)

Tableau 1.1 : Tableau récapitulatif des structures 3D de LTP1 : les structures 3D des LTP1 déterminées par RMN ont été représentées sous forme de « boudin » ; plus le diamètre de celui-ci est petit, meilleure est la définition structurale. Les tailles des cavités ont été obtenues à l'aide du programme « cavité » (Gomar et al. 1998)

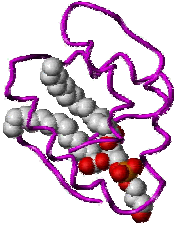
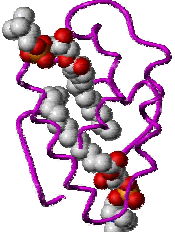
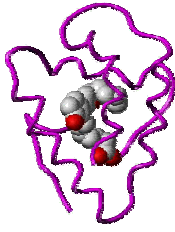
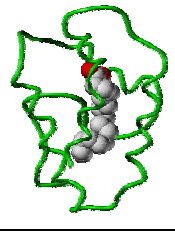
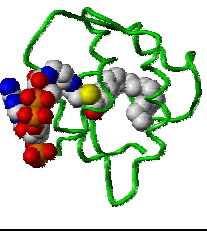
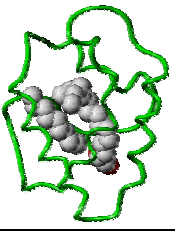
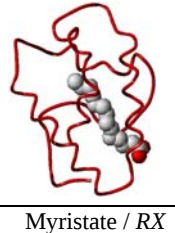
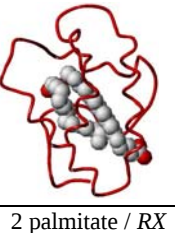
A) Complexes avec la LTP1 de blé			
Ligand / Technique Code PDB / Cavité (Å ³) Référence	DMPG / RMN 750±250 (Sodano et al. 1997)	2 LMPC / RX 1BW0 / 1180 (Charvolin et al. 1999)	Prostaglandine / RMN 1CZ2 / 700±120 (Tassin-Moindrot et al. 2000)
B) Complexes avec la LTP1 d'orge			
Ligand / Technique Code PDB / Cavité (Å ³) Référence	Palmitate / RMN 1BE2 / 490±30 (Lerche and Poulsen 1998)	Coenzyme A / RX 1JTB / 1180 (Lerche et al. 1997)	LLPG / RMN 1MID / 770 (Henriksen 2004)
C) Complexes avec la LTP1 de riz			
Ligand / Technique Code PDB / Cavité (Å ³) Référence	Myristate / RX 1UVA / 1080 (Cheng et al. 2004b)	2 palmitate / RX 1UVB / 1350 (Cheng et al. 2004b)	Stéarate / RX 1UVC / 780 (Cheng et al. 2004b)

Tableau 1.2 : Représentation des différents complexes de LTP1 réalisés expérimentalement. A) Trois complexes avec la LTP1 de blé : un avec le DiMyristoylPhosphatidylGlycérol (DMPG), un autre avec 2 molécules de LysoMyristoylPhosphatidylCholine (LMPC) et un dernier avec la prostaglandine. B) Trois complexes avec la LTP1 d'orge : le premier avec le palmitate, le deuxième avec le coenzyme A et le plus récent avec le LysoLauroylPhosphatidylGlycérol (LLPG). C) Trois complexes avec la LTP1 de riz : le premier avec du myristate, le deuxième avec deux palmitate et le dernier avec du stéarate.

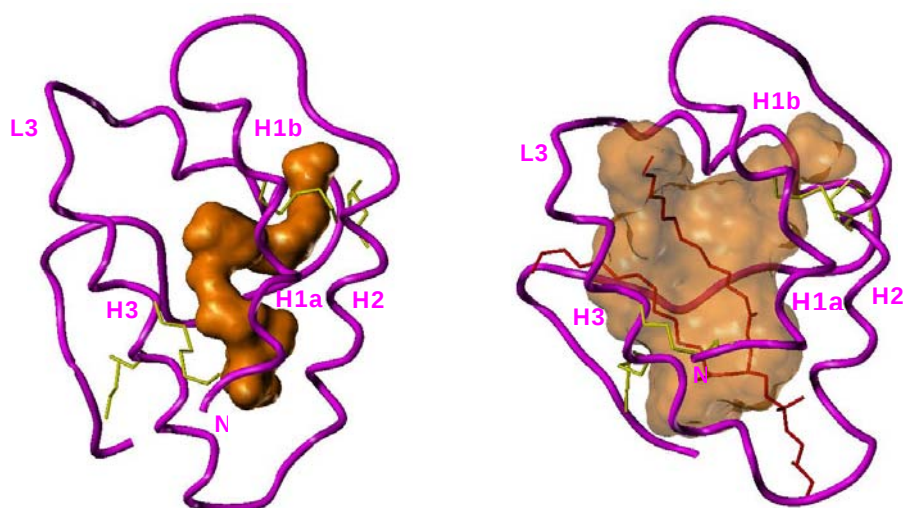


Figure 1.4 : Comparaison de la structure 3D de la LTP1 de blé sous forme libre (à gauche) et de la même LTP1 (à droite) fixant un phospholipide à deux chaînes grasses le DMPG, en rouge. La cavité est représentée en orange. Les ponts disulfure sont également illustrés en jaune.

2.2 LTP1 : structures de complexes

Jusqu'à présent de nombreux complexes de LTP1 ont été réalisés avec des ligands de nature et de taille variées. Les tableaux 1.2 et 1.3 récapitulent tous les complexes réalisés expérimentalement. On peut remarquer que les LTP1 adaptent la taille de leur cavité à celle de leur ligand sans déformation majeure de leur structure tridimensionnelle, simplement par un écartement des hélices et un repositionnement des chaînes latérales (figure 1.4). Par exemple la taille de la cavité de la LTP1 de blé passe de 300 Å³ sans ligand à 750 Å³ après fixation d'un phospholipide à deux chaînes aliphatiques. Toutefois on peut noter que lorsque la taille du ligand est trop importante, la structure 3D de la LTP1 se déforme. Ainsi, quand la LTP1 d'orge fixe le palmitoyl-coenzyme A (ester biologiquement actif de l'acide palmitique, activation des fonctions carboxyliques, activation des fonctions carboxyliques qui leur permette d'être des substrats de nombreux systèmes enzymatiques de type transférase qui donnent des conjugués type esters, amides...), elle présente une déformation anormale comparativement aux autres complexes.

La partie C-terminale apparaît très importante pour la fixation des ligands. Elle se comporte comme un couvercle et son soulèvement permet l'augmentation de la taille de la cavité (figure 1.4). De plus, bien qu'il n'existe aucune preuve expérimentale, elle pourrait constituer la porte d'entrée par laquelle le ligand arrive. La question du mode de fixation des ligands aux LTP1 n'a pas encore trouvé de réponse définitive et fait l'objet d'investigations.

Dans tous les complexes réalisés, on observe que la fixation des ligands aux LTP1 est non covalente. En effet, la stabilité des complexes est assurée par un certain nombre d'interactions hydrophobes, de liaisons hydrogène et d'interactions électrostatiques. Ces liaisons faibles sont assurées par des résidus conservés parmi les séquences des LTP1. On retrouve par exemple la tyrosine C-terminale, Y79 (numérotation de la LTP1 de blé). Cette tyrosine, qui obstrue l'entrée possible de la cavité dans la forme libre, pivote d'un angle de 100° pour former dans la plupart des cas une liaison hydrogène avec le ligand à l'aide de sa fonction hydroxyle (phénolique). De plus, une LTP1 de blé dans laquelle la Tyr 79 est mutée en alanine (Lullien-Pellerin et al. 1999) ne présente plus d'activité de transfert, confirmant l'importance de ce résidu dans la fixation du ligand. Enfin, le motif d'acides aminés chargés « DR », également très bien conservé parmi les LTP1, est impliqué souvent dans des liaisons électrostatiques et/ou hydrogène avec le ligand et constitue ainsi un point d'arrimage possible pour les ligands.

La différence majeure existant entre tous les complexes de LTP1 est l'orientation du ligand au sein de la cavité. L'orientation prédominante du ligand dans de nombreux complexes est la suivante : sa tête polaire est localisée vers la boucle L2 tandis que la partie C-terminale et sa queue hydrophobe sont enfouies dans la cavité. Cette orientation est

retrouvée dans tous les complexes avec la LTP1 de maïs, dans le complexe blé/DMPG et les complexes blé/LMPC et riz/palmitate où les LTP1 de blé et de riz fixent respectivement deux molécules de LMPC et de palmitate. Ces deux derniers complexes présentent une deuxième orientation possible du ligand. En effet, une des molécules de LMPC ou de palmitate a sa tête polaire située entre la boucle L1 et L3. Cette deuxième orientation est également observée dans le complexe orge/palmitate et dans le complexe maïs/caprato où il a été observé expérimentalement que l'acide gras pouvait adopter deux orientations différentes au sein de la cavité (Han et al. 2001). Enfin, une troisième orientation est observée dans le complexe orge/coenzymeA. Dans ce complexe la tête polaire du ligand est orientée entre le milieu de l'hélice H1 et H3. Ces différences d'orientation du ligand suggèrent donc trois entrées et/ou sorties possibles de la cavité hydrophobe. Deux d'entre elles, illustrées par le complexe blé/LMPC, ont été observées récemment par dynamique moléculaire à haute température (Sy et al. 2003).

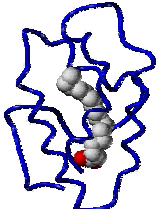
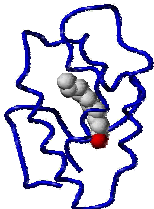
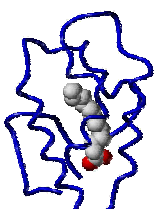
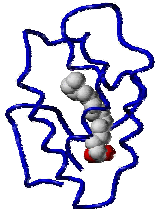
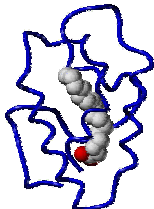
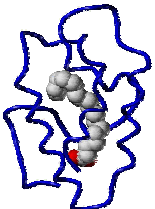
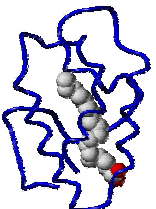
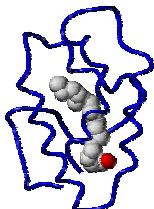
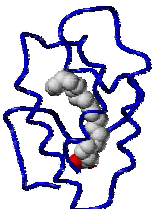
D) Complexes avec la LTP1 de maïs			
Ligand / carbone : sat Code PDB / Cavité (Å ³) Référence	Palmitate / C16 : 0 1MZM / 740 (Shin et al. 1995)	Caprate / C10 : 0 1FKO / 560 (Han et al. 2001)	Laurate / C12 : 0 1FK1 / 550 (Han et al. 2001)
Complexes avec la LTP1 de maïs			
Ligand / carbone : sat Code PDB / Cavité (Å ³) Référence	Myristate / C14 : 0 1FK2 / 570 (Han et al. 2001)	Palmitoléate / C16 : 1 1FK3 / 570 (Han et al. 2001)	Stéarate / C18 : 0 1FK4 / 560 (Han et al. 2001)
Complexes avec la LTP1 de maïs			
Ligand / carbone : sat Code PDB / Cavité (Å ³) Référence	Oléate / C18 : 1 1FK5 / 560 (Han et al. 2001)	Linoléate / C18 : 3 1FK6 / 610 (Han et al. 2001)	Ricinélate / C18 : 1 1FK7 / 620 (Han et al. 2001)

Tableau 1.3 : Représentation de neuf complexes avec la LTP1 de maïs obtenues par diffraction X. Les ligands sont tous des acides gras possédant des longueurs de chaînes aliphatiques variables allant de C10 à C18 ainsi que des degrés divers de saturation (sat : nombre de doubles liaisons)

2.3 LTP2 : structures tridimensionnelles

Plus récemment, les structures 3D d'une LTP2 de riz sous forme libre (Samuel et al. 2002) et de blé sous forme complexée (Pons et al. 2003) ont été déterminées par RMN et modélisation moléculaire. La structure tridimensionnelle des LTP2 présente un repliement plus compact que celui des LTP1 (figure 1.5). Elles possèdent cinq hélices (H1 à H5). Les hélices H1, H2 et H5 trouvent respectivement leur correspondance avec les hélices H1, H2 et H4 des LTP1 (figure 1.5). Les plus grandes différences structurales résident d'une part au niveau des deux courtes hélices H3 et H4 qui sont remplacées par une longue hélice H3 dans les LTP1, et d'autre part dans l'absence de la longue partie C-terminale présente dans les LTP1. Comme les LTP1, les LTP2 possèdent également une cavité qui est douée d'une certaine plasticité. Ainsi la cavité de la LTP2 de riz sous forme libre mesure environ $150 \text{ \AA}^3$ contre $340 \text{ \AA}^3$ pour la cavité de la LTP2 de blé liée à une molécule de LPG (LysoPalmitoylphosphatidylGlycérol). D'autre part, contrairement aux LTP1, les LTP2 sont capables de fixer des molécules de stérols. En effet, la LTP2 de riz peut fixer du cholestérol (Cheng et al. 2004a).

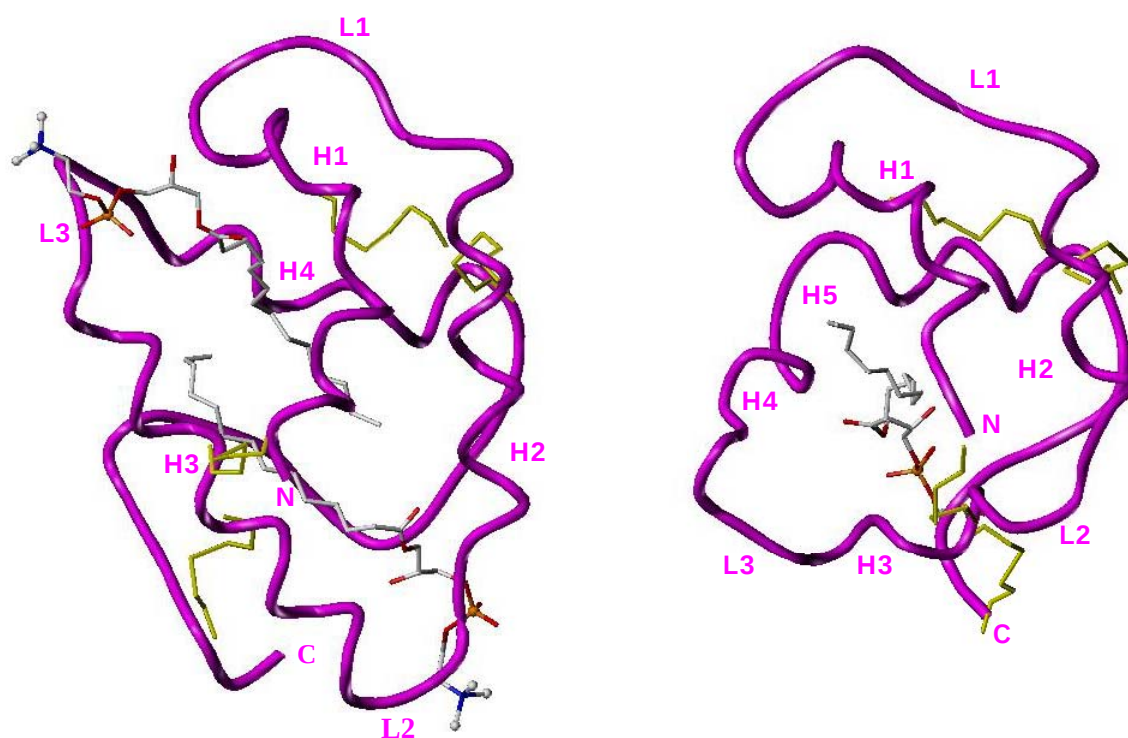


Figure 1.5 : Comparaison des structures 3D de la LTP1 de blé fixant 2 LMPC (à gauche) et de la LTP2 liant un LPG (à droite).

2.4 Protéines de même repliement

L'étude des structures de LTPs sous forme libre ou liée a permis d'établir une relation directe entre leur structure et leur capacité de transfert *in vitro*. Par ces propriétés de fixation de lipides, les LTPs constituent donc une famille de transporteurs à part entière. L'explosion du nombre de structures protéiques dans la Protein Data Bank de Brookhaven a permis de mettre en évidence que le repliement des LTPs se retrouve dans d'autres protéines végétales comme les inhibiteurs d' α amylases des céréales (Oda et al. 1997), la protéine hydrophobe du soja (HPS pour Hydrophobic Protein of Soybean) (Baud et al. 1993; Sodano and Ptak 1995), les protéines de réserve du type 2S albumine des graines (Rico et al. 1996) et enfin une protéine antifongique extraite des graines d'oignon, Ace-AMP1 (*Allium cepa* Antimicrobial Protein) (Tassin et al. 1998). Toutefois, aucune de ces protéines ne possède la capacité de transfert de lipide même si elles partagent le même repliement que les LTPs. Cette similarité structurale est évidente pour la protéine Ace-AMP1 qui est totalement superposable aux structures des LTP1 (figure 1.6). Il apparaît donc que le repliement des LTPs constitue une plate forme structurale permettant l'expression de fonctions biologiques très diverses. De plus il est à noter que, parmi ces protéines, certaines ont été identifiées comme des agents allergènes.

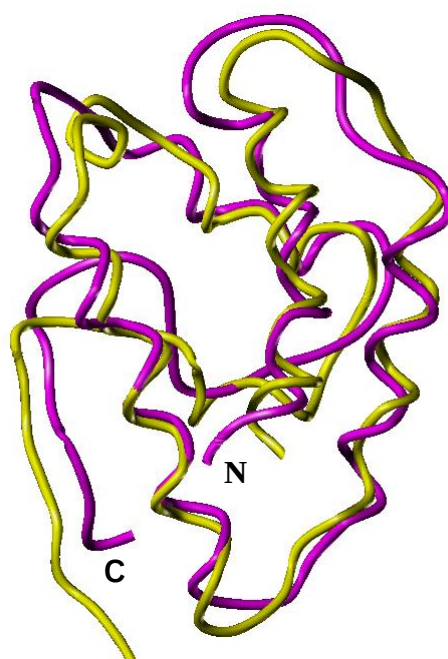


Figure 1.6 : Superposition des structures 3D de la LTP1 de blé (en gris) et de celle de Ace-AMP1 (en blanc).

2.4.1 Allergie alimentaire

A la fin du 20^e siècle, on a remarqué que certaines personnes dans le bassin méditerranéen développaient une allergie alimentaire aux *Rosaceae*. Parmi les membres de cette famille végétale, on retrouve de nombreux fruits comme les pommes, les poires, les cerises ou encore les pêches. Les allergènes impliqués ont été identifiés comme étant des LTP1 situées principalement dans la peau des fruits (Marion et al. 2003). D'autres LTP1 allergènes ont été isolées de plantes n'appartenant pas à la famille des *Rosaceae* comme la noisette, l'olive ou encore certaines céréales tels que le maïs ou l'orge (Marion et al. 2003). De plus, il s'avère que les protéines présentant un repliement de « type LTP » sont également des allergènes alimentaires, comme la protéine HPS (Gonzalez et al. 1995), les 2S albumines (Pastorello et al. 2001) et les inhibiteurs d' α -amylases (Armentia et al. 1993; Amano et al. 1998) (figure 1.7). Ces protéines ont aussi en commun le motif cystéine retrouvé dans les LTPs. Ces particularités structurales confèrent à ces protéines une stabilité remarquable. Les LTPs comme les 2S albumines sont ainsi très résistantes à l'hydrolyse pepsique (enzyme du suc digestif) et à la dénaturation thermique (Lindorff-Larsen and Winther 2001; Murtagh et al. 2002). Cette stabilité est à rapprocher de la propriété allergène de toutes ces protéines. Le motif cystéine des LTPs apparaît comme « une signature de protéine allergène ». Sa recherche dans les banques de données de séquences de protéines pourrait permettre de détecter des allergènes potentiels dans les produits végétaux.

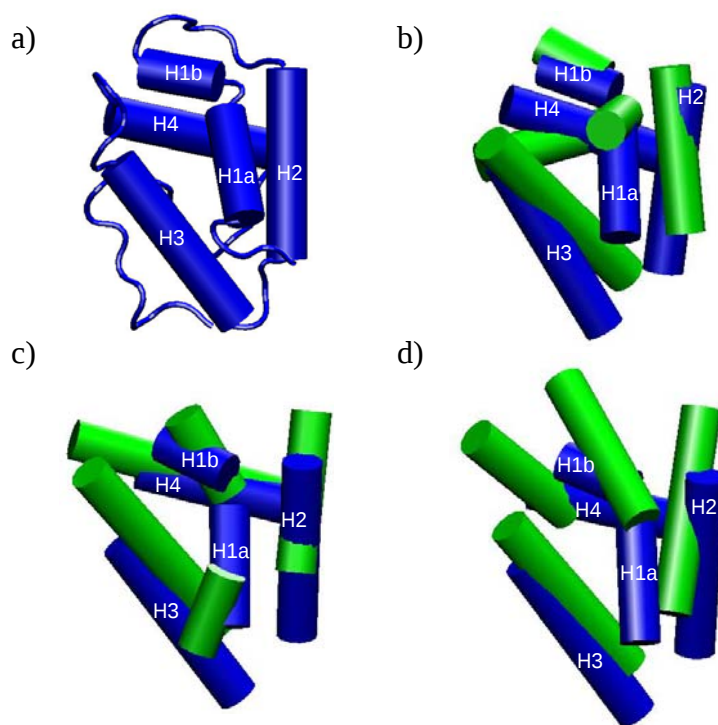


Figure 1.7 : Représentation schématique du domaine structural de protéine « type LTP ». (a) Structure 3D de la LTP1 de maïs, (b) (c) et (d) représentent respectivement la superposition schématique du domaine structural des LTP1 en noir, et de celui de la protéine HPS, des 2S albumines et des inhibiteurs d' α -amylase, en gris.

3. Activités biologiques

In vitro, les LTPs sont capables de complexer et de transférer des molécules amphiphiles ou hydrophobes d'une membrane à l'autre alors que leur fonction biologique au sein de la plante reste encore inconnue. Ces propriétés de complexation sont liées à la présence d'une cavité hydrophobe au cœur de la protéine.

Bien que la (les) fonction(s) biologique(s) des LTPs végétales ne soi(en)t pas connue(s), leur rôle dans les mécanismes de défense des plantes semble clair, que ce soit en participant à la synthèse de la cutine ou de la subérine et/ou comme agents antimicrobiens (Kader 1996). De plus, selon des travaux récents (Buhot et al. 2001; Maldonado et al. 2002), les LTPs seraient impliquées dans la Réponse Systémique Acquise (RSA) (protection de l'ensemble de la plante) souvent consécutive d'une Réponse Hypersensible (RH) (nécrose des tissus suite à une interaction incompatible avec un agent pathogène).

3.1 La cuticule

Les parties aériennes des plantes sont recouvertes d'une couche extracellulaire continue : la cuticule, qui joue un rôle important dans la protection de la plante contre la perte d'eau, l'irradiation des rayons UV, les dommages du gel, mais aussi contre l'attaque des agents pathogènes. L'élément structural principal de la cuticule est la cutine, un polymère de dérivés d'acide palmitique C16 ou oléique C18 (figures 1.8 et 1.9). La cutine est recouverte par la cire, un mélange complexe de composés aliphatiques à très longues chaînes (C26-C36). La combinaison de la cutine et des cires forme la cuticule. Bien que la voie générale de biosynthèse de la cutine soit bien décrite depuis de nombreuses d'années (Kolattukudy 1981), on connaît très peu choses sur le transport des monomères de cutine de leur lieu de synthèse vers la cuticule. Les LTPs pourraient constituer des candidates potentielles pour ce transport. En effet, les LTPs sont des protéines sécrétées et leur localisation est extracellulaire (Bernhard et al. 1991; Sterk et al. 1991). Certaines se situent au niveau de la paroi cellulaire (Thoma et al. 1993), d'autres sont sécrétées dans le milieu de culture de cellules embryonnaires (Sterk et al. 1991; Coutos-Thevenot et al. 1993). Cette localisation extracellulaire a conduit Sterk et ses collaborateurs (Sterk et al. 1991) à formuler pour la première fois l'hypothèse du rôle possible des LTPs dans le transport de monomères de cutine. Cette hypothèse est confortée par la capacité des LTPs à fixer *in vitro* des ligands aussi

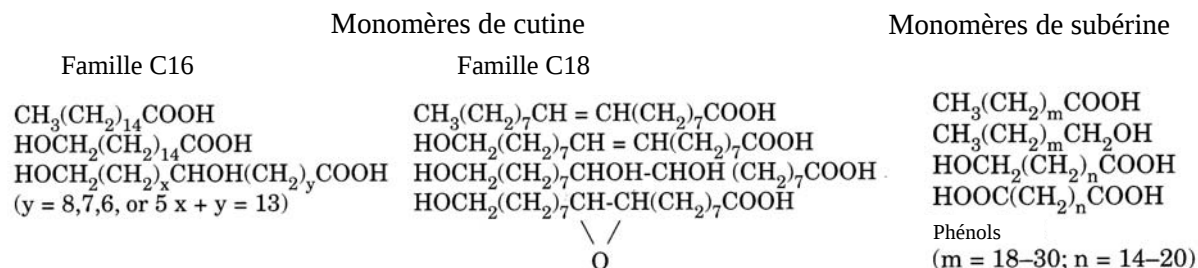


Figure 1.8 : Composants majeurs de la cutine et de la subérine.

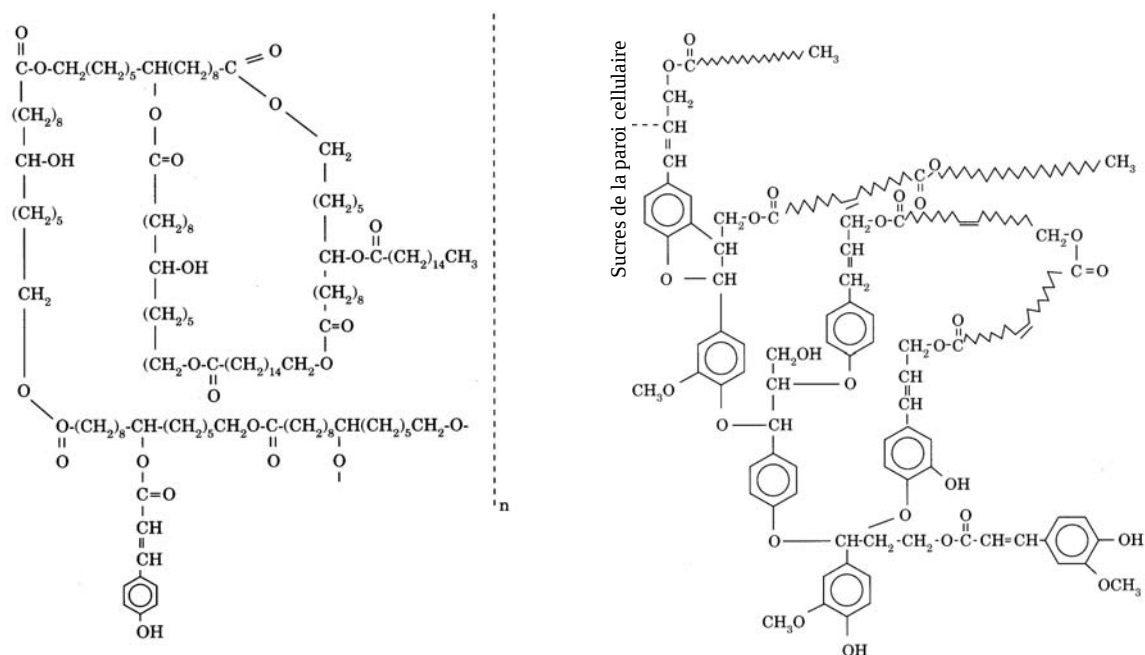


Figure 1.9 : Structures modèles de polymères de cutine (à gauche) et de subérine (à droite).

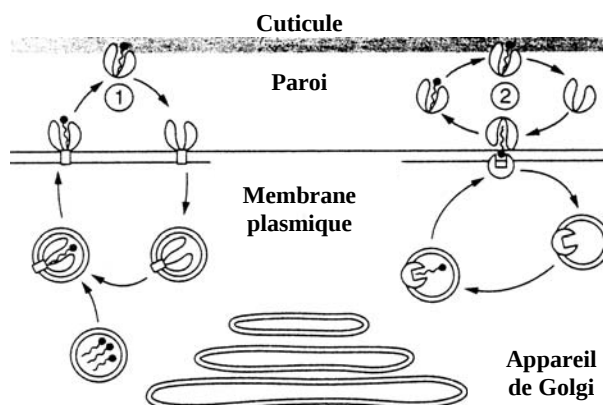


Figure 1.10 : Modèle du rôle des LTPs dans la formation de la cuticule d'après (Hendriks 1994).

Les LTPs agissent comme des transporteurs de monomères de cutine, de leur lieu de synthèse vers la cuticule. Deux modèles d'interactions sont présentés. Le premier (à gauche) : après fusion de deux vésicules l'une contenant les LTPs, l'autre les monomères de cutine provenant de l'appareil de Golgi, les LTPs chargent les monomères de cutine dans la cellule. Le deuxième (à droite) : les LTPs chargent les monomères de cutine au niveau de la paroi cellulaire. Ces derniers se présentent à la surface de la membrane plasmique après la fusion d'une vésicule contenant une protéine membranaire qui lie les monomères de cutine.

bien amphiphiles qu'hydrophobes, de taille et de nature variées (cf. paragraphe 2). D'autre part, cette fonction est également appuyée par le fait que de nombreuses LTP1 sont exprimées dans les tissus épidermiques des plantes (Sossountzov et al. 1991; Kotilainen et al. 1994). Une LTP1 a même été identifiée comme étant la protéine majoritaire de la surface cireuse des feuilles de brocoli (Pyee et al. 1994). De plus, on a remarqué une corrélation entre l'augmentation de l'épaisseur de la cuticule et la surexpression d'une LTP1 dans la feuille d'orge (Hollenbach et al. 1997). Tous ces résultats suggèrent fortement l'implication des LTPs dans la formation de la cuticule. Hendricks et ses collaborateurs (Hendriks 1994) ont même proposé deux modèles moléculaires de transport de monomères de cutine de leur lieu de synthèse vers la cuticule (figure 1.10). On retrouve des LTPs dans des organes végétaux dépourvus de cuticule comme les racines mais qui contiennent un autre polymère hydrophobe, la subérine (figures 1.8 et 1.9).

3.2 La subérine

La subérine est un polymère hydrophobe complexe composé de molécules phénoliques (dérivés cinnamiques) et d'acides gras dicarboxylés et hydroxylés (figures 1.8 et 1.9) qui recouvre les organes souterrains des plantes, comme les racines. Il a été montré que les LTPs qui s'expriment dans les racines appartiennent à la famille des LTP2, alors que les LTP1 y sont faiblement ou pas du tout exprimées (Sossountzov et al. 1991; Sterk et al. 1991; Canevascini et al. 1996; Vignols et al. 1997). De plus les LTP1 et LTP2 sont synthétisées dans les graines où la cutine et la subérine sont simultanément présentes. Enfin au cours de la symbiose entre une plante de la famille des *Fabaceae* (exemple : Légumineuses) (Smith et al. 1986), et la bactérie du genre *Rhizobium*, les nodosités racinaires (organes siège de la symbiose) qui se forment sont constituées de composés de type subérine ; et au cours du développement de la nodosité, on observe l'expression d'une LTP2 (Krause et al. 1994).

Par conséquent, tous ces résultats suggèrent que les LTP1 participent au transport des monomères de cutine et les LTP2 à celui de subérine. Toutefois, jusqu'à présent, ces observations ne constituent que des faisceaux de présomptions qui convergent vers le rôle possible des LTPs dans la formation de la cuticule et de la subérine. Mais aucune preuve moléculaire, comme par exemple l'isolement d'une LTP liée avec un monomère de cutine ou de subérine, n'a été faite.

En plus de ce rôle probable dans la mise en place de polymères hydrophobes, il a été montré que certaines LTPs possédaient des propriétés antimicrobiennes.

3.3 LTPs : activité antimicrobienne

L'activité antimicrobienne *in vitro* de certaines LTPs a été découverte par deux groupes différents : celui de Broekaert en Belgique et l'équipe espagnole de Garcia-Olmedo. Ils ont identifié respectivement l'activité antifongique d'une LTP extraite de graines de radis (Terras et al. 1992) et l'activité antimicrobienne de LTPs extraites de feuilles d'orge et de maïs (Molina et al. 1993). Depuis, de nombreuses LTPs possédant une activité antimicrobienne ont été isolées de feuilles d'*Arabidopsis thaliana* et d'épinard (Segura et al. 1993) et de feuilles de betterave sucrière (Nielsen et al. 1996). Ceci a amené certains auteurs à classer les LTPs parmi les « PR (Pathogenesis Related)-proteins », protéines induites lors de l'attaque d'agents pathogènes (Van Loon and Van Strien 1999). Plus récemment, de nouvelles LTPs antifongiques ont été caractérisées : une dans les feuilles de betterave sucrière (Kristensen et al. 2000) et deux autres dans les graines de riz (Ge et al. 2003).

De plus, l'attaque d'agents pathogènes induit *in planta* la surexpression de gènes codant pour des LTPs (Garcia-Olmedo et al. 1995; Broekaert et al. 1997). Par exemple lors de la réplication du virus de la mosaïque du tabac chez le poivron (*Capsicum annum*), on observe la surexpression d'une LTP (Park et al. 2002). On peut également noter que des plants de tabac transgénique surexprimant une LTP antimicrobienne d'orge sont plus résistants à l'attaque d'agents pathogènes (Molina and Garcia-Olmedo 1997). Toutefois certaines LTPs, comme la LTP1 de la graine de blé ne possèdent pas d'activité antimicrobienne mais leur présence augmente l'activité antimicrobienne des thionines, peptides de 4 à 5 kDa impliqués dans la défense des plantes (Dubreil et al. 1998). Elles agiraient donc en synergie avec les thionines pour lutter efficacement contre l'attaque pathogène.

Actuellement, on ne connaît pas le mode d'action des LTPs vis-à-vis des agents pathogènes. On suppose qu'elles pourraient agir en perméabilisant la membrane des agents pathogènes, du fait de leur point isoélectrique élevé (Kader 1996) ce qui a été partiellement démontré en comparant l'activité antifongique des LTPs et des histones (Gomes et al. 2003).

La découverte d'une protéine Ace-AMP1 (*Allium cepa* AntiMicrobial Protein), isolée des graines d'oignon, homologue des LTP1 et possédant une puissante activité antimicrobienne renforce l'idée que les LTPs participent à la défense des plantes contre les agents pathogènes (Cammue et al. 1995). Toutefois, Ace-AMP1 présente peu d'identité de séquence avec les autres LTP1, à l'exception des huit cystéines. Elle possède 2 résidus

tryptophane alors que celui-ci est absent des séquences des LTP1. De plus, *Ace*-AMP1 contient 19 résidus arginine contre, en général, quatre fois moins pour les LTP1. Enfin sa structure tridimensionnelle est parfaitement superposable à celles des LTP1 mais est dépourvue de cavité hydrophobe (Tassin et al. 1998) et est donc incapable de fixer des lipides. *Ace*-AMP1 pourrait représenter le premier maillon d'une divergence fonctionnelle des LTP1. On suppose donc que la fonction principale des LTPs est de participer à la synthèse de cuticule et de subérine et que secondairement, grâce à leur pI basique, elles possèdent une activité antimicrobienne. Certaines LTPs ou protéines de type LTPs comme *Ace*-AMP1 ont peut-être développé cette dernière activité au détriment de l'activité de transport de molécules lipidiques.

Des travaux récents ont suggéré un nouveau rôle possible des LTPs dans la signalisation cellulaire, rôle qui serait complémentaire des deux fonctions précédemment évoquées.

3.4 LTPs : signalisation cellulaire

Il a été montré que les LTPs possédaient un récepteur spécifique qui se trouve être en fait celui des élicitines (Buhot et al. 2001). En recherchant le mode d'action des élicitines, il a donc été découvert une fonction jusqu'alors inconnue des LTPs, qui serait d'induire la mise en place de mécanisme de défense.

3.4.1 Les élicitines

Les élicitines sont des protéines de petite masse moléculaire (10 kDa) sécrétées par les *Phytophthora* et les *Pythium*, oomycètes connus pour les dégâts qu'ils causent aux cultures de par le monde. Ces protéines, décrites pour la première fois à Antibes (Bonnet et al. 1986), sont connues depuis une quinzaine d'années pour leurs propriétés inductrices de Réaction Hypersensible (RH, nécrose) et de Résistance Systémique Acquise (RSA, protection vis-à-vis d'autres agents pathogènes). Les élicitines possèdent la propriété de se charger en lipides (Osman et al. 2001a), avec une forte affinité pour les stérols, et de transférer ces petites molécules entre membranes biologiques (Mikes et al. 1998). Cette capacité à se complexer avec un stérol est très importante, car elle permet d'expliquer leur activité élicitrice chez le tabac.

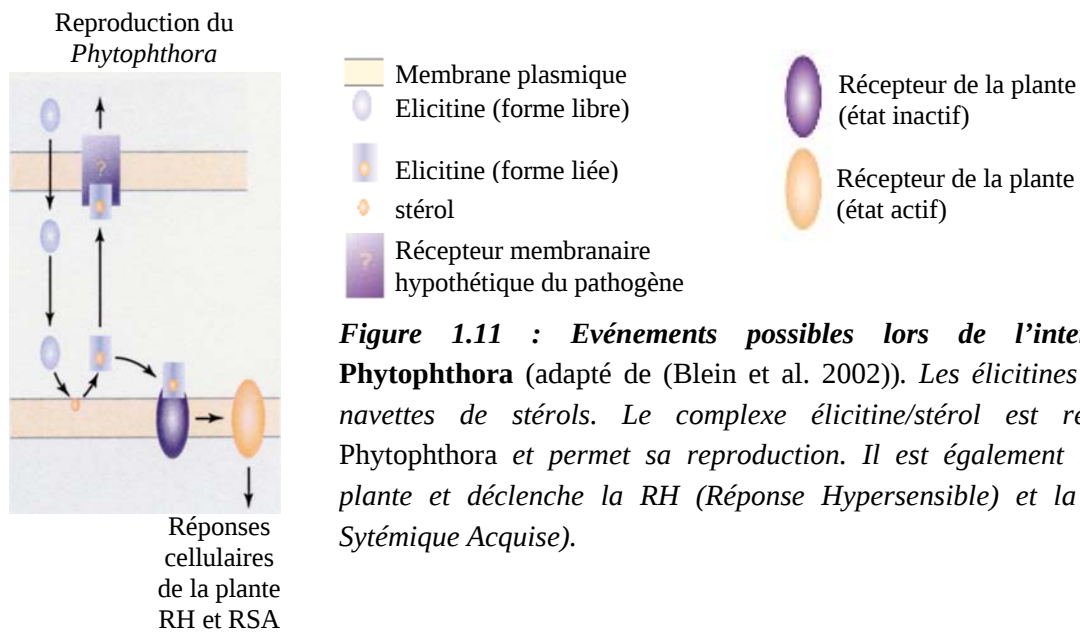


Figure 1.11 : Evénements possibles lors de l'interaction plante *Phytophthora* (adapté de (Blein et al. 2002)). Les élicitines constituent des navettes de stérols. Le complexe élicitrine/stérol est reconnu par le *Phytophthora* et permet sa reproduction. Il est également reconnu par la plante et déclenche la RH (Réponse Hypersensible) et la RSA (Réponse Systémique Acquise).

En effet, seule une élicitrine chargée peut se fixer sur un récepteur spécifique localisé sur la membrane plasmique, puis déclencher une cascade d'événements conduisant à la résistance (Osman et al. 2001b). Les Oomycètes qui sécrètent ces protéines sont connus pour leur incapacité à fabriquer des stérols qui, pourtant, sont importants pour leur croissance et leur reproduction (Hendrix and Guttman 1970). On suppose donc que les élicitines sont des navettes qui permettraient à ces agents pathogènes de pouvoir récupérer ces précieuses molécules au détriment de la plante qu'ils envahissent (figure 1.11).

Jusqu'à présent seules les structures tridimensionnelles de deux élicitines ont été déterminées par RMN ou radiocristallographie : la cryptogéine (Boissy et al. 1996; Fefeu et al. 1997) et la cinnamomine (Rodrigues et al. 2002) extraites respectivement de *Phytophthora cryptogea* et de *Phytophthora cinnamomi*. La structure 3D de ces deux élicitines est similaire. Elles présentent 5 hélices ($\alpha 1$ - $\alpha 5$), un feuillet β antiparallèle à deux brins et une boucle Ω . L'ensemble de la structure est stabilisé par 3 ponts disulfure (figure 1.12). Les élicitines possèdent également une cavité hydrophobe dans laquelle se fixent les stérols. De plus, deux structures 3D de la cryptogéine liée à l'ergostérol (Boissy et al. 1999) ou au cholestérol (Lascombe et al. 2002) obtenues par radiocristallographie ont révélé que cette fixation n'entraînait pas de modifications structurales importantes à l'instar des complexes avec les LTP1 (figure 1.12).

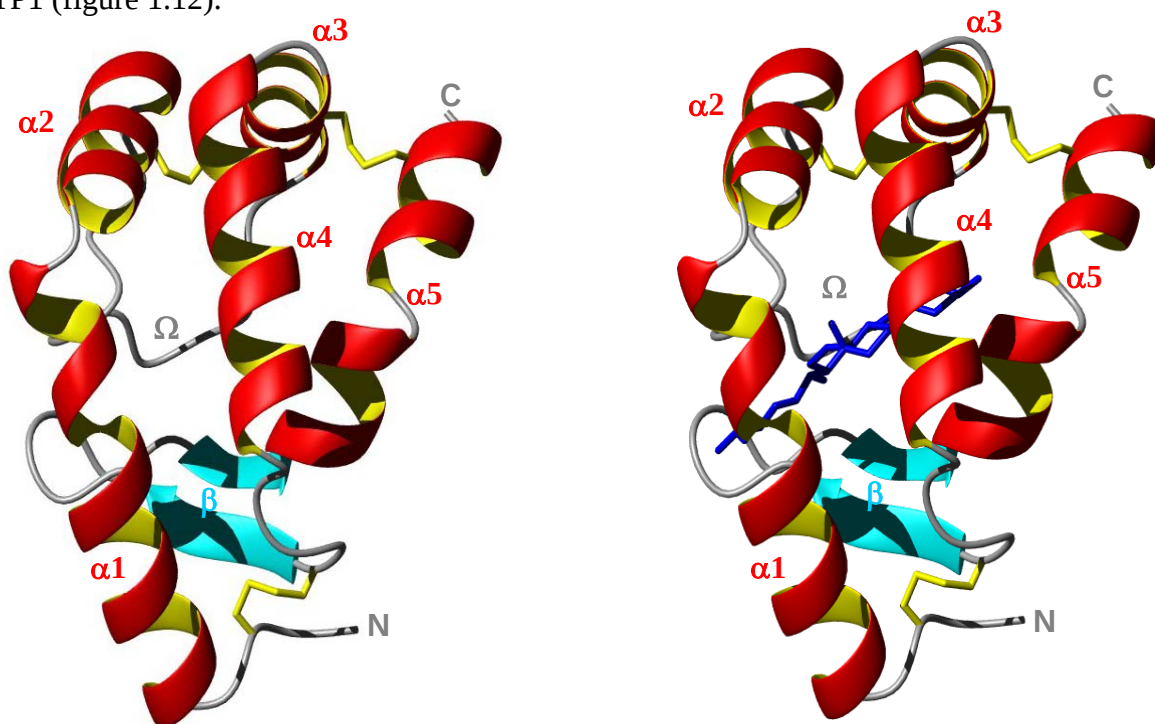


Figure 1.12 : Comparaison des structures 3D de la cryptogéine sous forme libre (à gauche) et sous forme liée avec une molécule de cholestérol (en bleu) (à droite). Les ponts disulfure sont représentés en jaune.

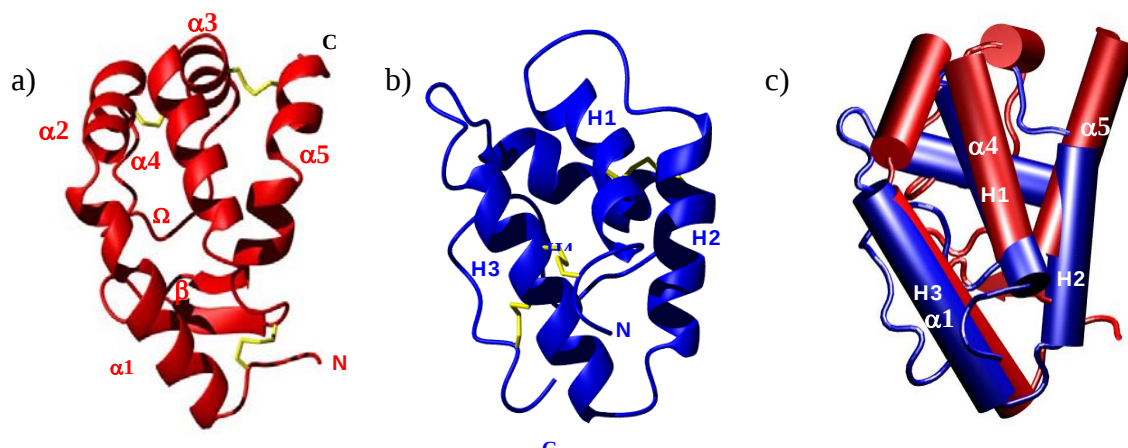


Figure 1.13 : Comparaison des structures 3D des élicitines et des LTP1. a) Structure 3D de la cryptogéine, b) Structure 3D de la LTP1 de blé, c) Superposition des structures 3D de la cryptogéine et de la LTP1 de blé.

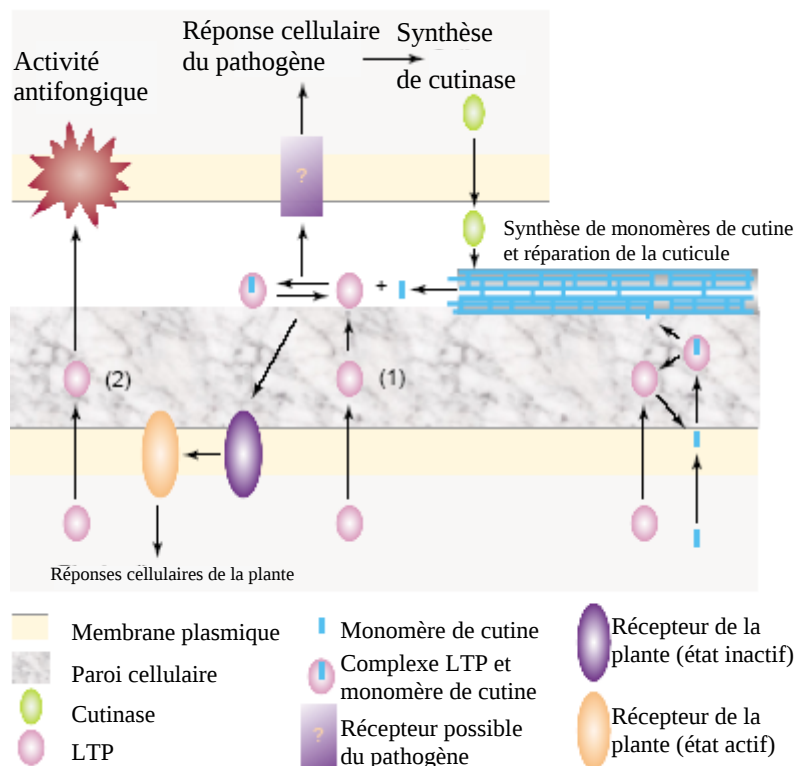


Figure 1.14 : Actions possibles des LTPs au cours de l'attaque d'un pathogène (cas général), adaptée de (Blein et al. 2002). L'action de la cutinase constitutive produite par le pathogène entraîne la libération de monomères de cutine qui se fixent aux LTPs sentinelles de la paroi cellulaire (1). Ce complexe LTP/monomère de cutine peut se fixer sur son récepteur et déclencher alors une cascade d'événements (2) aboutissant à la synthèse de protéines comme des LTPs pour réparer la cuticule et lutter contre le développement du pathogène. Ce complexe peut également se fixer sur le récepteur possible du pathogène et induire la synthèse de cutinase.

3.4.2 Dialogue entre LTPs et élicitines

Récemment il a été montré que les LTP1 étaient capables de déplacer les élicitines de leur récepteur et d'entrer en compétition avec les activités biologiques induites par ces protéines. En fait, une LTP1 de graine de blé est capable d'entrer en compétition avec la cryptogéine tant pour sa fixation à son récepteur situé sur la membrane plasmique de tabac que pour les réponses qu'elle induit sur cultures cellulaires (Buhot et al. 2001). Bien que les LTP1 et les élicitines ne possèdent pas d'identité de séquence ni d'homologies structurales évidentes, il est néanmoins possible de superposer des hélices des LTP1 avec celles des élicitines (Buhot et al. 2001). On peut ainsi noter que les hélices $\alpha 1$, $\alpha 4$, $\alpha 5$ des élicitines trouvent leurs homologues respectifs dans les hélices H3, H1, H2 des LTP1 (figure 1.13). Cette analogie structurale pourrait expliquer en partie l'affinité des LTP1 pour les récepteurs des élicitines.

Cependant la fonction des LTPs ne se limiterait pas à inhiber l'action des élicitines. En effet, au cours de l'attaque d'un champignon pathogène, la cuticule est considérée comme une zone d'échange de signaux entre la plante et les bio-agresseurs. Dans plusieurs systèmes plante/pathogène, la cutinase, une enzyme produite par différents agents pathogènes fongiques, est impliquée dans la pénétration de la cuticule. Les monomères de cutine produits par l'action de la cutinase se sont avérés importants en tant que molécules signal, d'une part dans la plante pour activer potentiellement des mécanismes de défense (Fauth et al. 1998), d'autre part chez les microorganismes pathogènes en induisant l'expression de cutinase (Kolattukudy et al. 1995). Dans ce contexte, les LTPs participeraient au transport de ces molécules signal. Le complexe LTP/monomère de cutine pourrait alors se fixer sur son récepteur et déclencher une cascade d'événements aboutissant à la synthèse de protéines comme des LTPs pour réparer la cuticule et empêcher la pénétration de l'agent pathogène. Mais l'agresseur pourrait aussi détourner le complexe LTP/lipide à son avantage. Ce complexe se fixerait alors sur un possible récepteur membranaire et déclencherait la synthèse de cutinase. La régulation de ce complexe apparaît donc essentielle pour la survie de la plante ou du bio-agresseur. La figure 1.14 résume le double rôle possible du complexe LTP/lipides, ainsi que les différentes fonctions supposées des LTPs.

D'autre part, le rôle de signalisation cellulaire des LTPs a été conforté par une découverte récente faite par une équipe en Grande Bretagne (Maldonado et al. 2002). Cette équipe a montré qu'un mutant d'*A. thaliana* incapable de développer de la RSA était affecté au niveau de l'expression d'un gène codant une protéine de type LTP2. De plus notre équipe,

en collaboration avec l'INRA de Nantes, essaie d'identifier les ligands « naturels » des LTP1. Ceci permettrait d'apporter une preuve moléculaire de la fonction biologique supposée de signalisation cellulaire des LTPs.

Il ressort de l'ensemble de ces études que les LTPs seraient des actrices importantes dans la formation des couches de protection des organes végétaux (cuticule, subérine) et dans la défense des plantes contre leurs agresseurs microbiens. Comme nous l'avons vu, ces deux fonctions ne sont pas indépendantes, dans la mesure où la cuticule constitue une barrière physique aux bio-agresseurs et que ses composants participent à la signalisation de l'agression.

4. Ligands naturels

Jusqu'à présent, les structures de complexe de LTPs ont été déterminées avec des ligands modèles qui correspondent à des molécules qui se rencontrent rarement sous forme libre au sein de la plante. Cependant, lorsque les LTP1 des graines d'orge et de blé sont extraites, un adduit d'une masse moléculaire de 294 Da est observé. Cet adduit est fixé de manière covalente et ne s'hydrolyse qu'en milieu alcalin (pH 8).

4.1 Ligand de la LTP1 d'orge

Le groupe danois de Poulsen (Lindorff-Larsen et al. 2001) a identifié l'adduit de la LTP1 d'orge comme un composé lipidique ajouté lors d'une modification post-traductionnelle dont la formule serait $C_{17}O_5H_{28}$ et dont la structure supposée est présentée sur la figure 1.15.

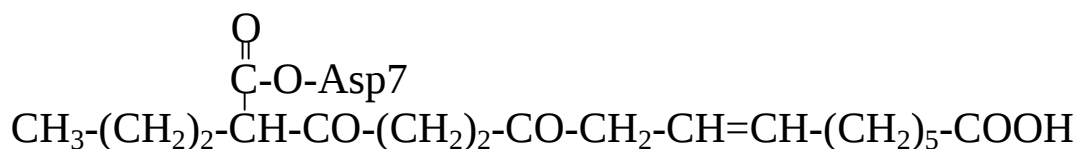


Figure 1.15: Structure semi-développée proposée par (Lindorff-Larsen et al. 2001) de la modification post-traductionnelle de la LTP1 d'orge. Le ligand identifié formerait une liaison ester avec la chaîne latérale du résidu Asp7. La double liaison du ligand est en configuration *cis*.

Cependant la nature de cet adduit post-traductionnel constitue seulement une proposition. En effet, pour l'identifier, les auteurs de l'article ont eu recours tout d'abord à une protéolyse de la LTP1 d'orge et ont ensuite étudié par RMN la structure du ligand sur le fragment protéique possédant la modification. La structure chimique qu'ils proposent pour le ligand, avec un nombre impair de carbone, en C_{17} , est très rare dans la nature mais, d'après les auteurs, aucune formule en C_{2n} , plus compatible avec le métabolisme des lipides ne pouvait correspondre à leurs résultats.

4.2 Ligand de la LTP1 de blé

De même, nous avons essayé d'identifier l'adduit de 294 Da de la LTP1 de blé. D'après nos expériences RMN et de spectrométrie de masse, nous avons obtenu un faisceau de présomptions qui nous conduit à proposer une formule de l'adduit en $C_{18}O_4H_{32}$ (figure 1.16) sans pouvoir identifier sur quel acide aminé est fixé ce ligand.

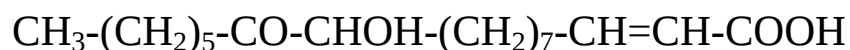


Figure 1.16: Proposition de structure pour l'adduit de la LTP1 de blé. D'après les résultats RMN, la double liaison est en configuration trans.

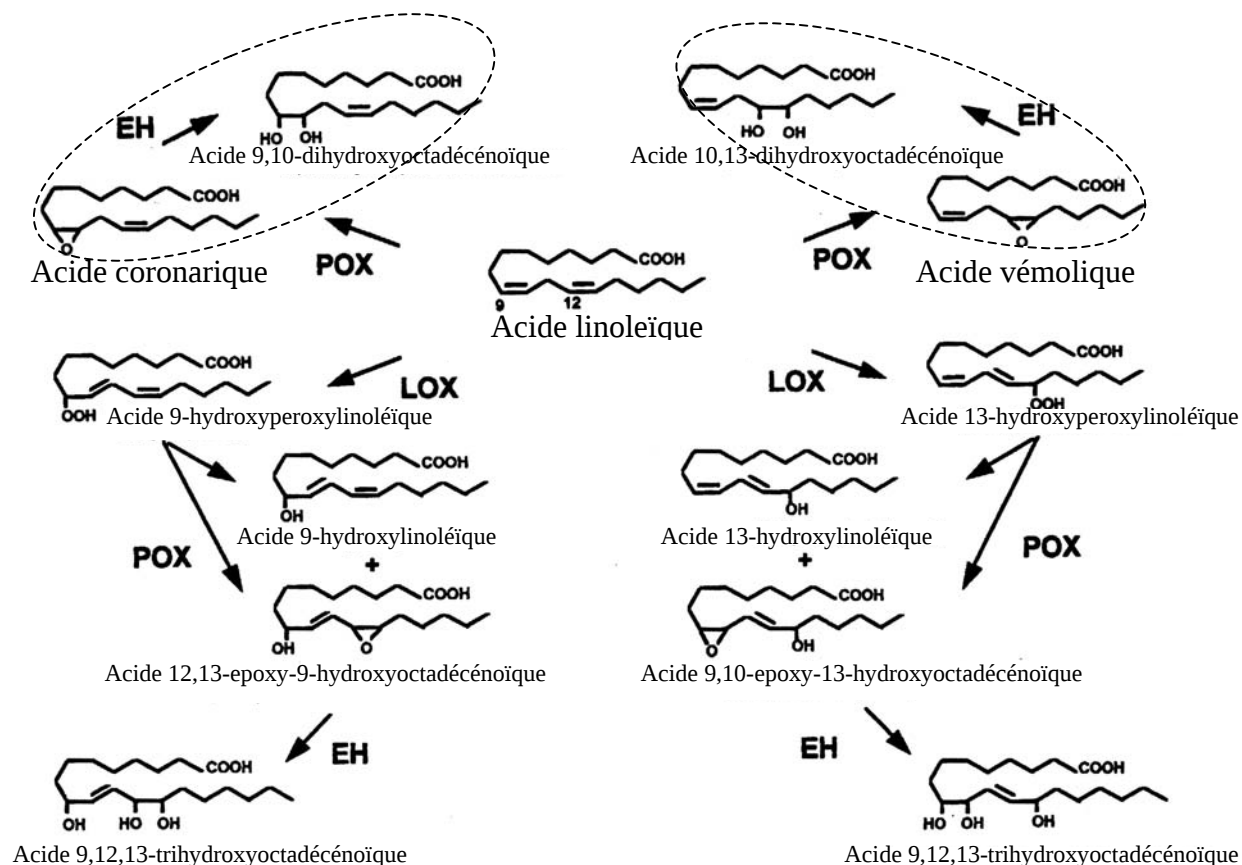


Figure 1.17: Différentes oxilipines issues des voies LOX POX. LOX : lipoxigénase, POX : peroxygénase, EH, époxyde hydrolase.

4.3 Nouvel éclairage sur les structures des ligands des LTP1 d'orge et de blé

Récemment, nos collaborateurs de l'INRA de Nantes ont découvert que l'adduit n'est pas observé quand les graines d'orge ou de blé sont délipidées avant extraction des LTPs. Cette découverte révèle que l'adduit serait issu du pool des lipides. En effet, si on observe attentivement les deux structures hypothétiques des ligands des LTP1 de blé et d'orge, on remarque qu'ils présentent des analogies structurales avec des oxylipines (acides gras oxygénés issus de la voie des lipoxygénases (LOX) et des peroxygénases (POX)). Les LOX et les POX sont des enzymes impliquées en général dans la biosynthèse des acides gras précurseurs de la cutine et en particulier dans la génération d'acides gras peroxydés cytotoxiques pouvant expliquer les nécroses induites au cours d'une RH (Rusterucci et al. 1999 ; Montillet et al. 2002). De par sa structure en C18, la formule que nous avons proposée ressemble aux produits de la voie de peroxidation des acides gras : acides coronarique et vémolique et leurs dérivés (Blee 1998) (figure 1.17). Cependant les produits des voies de peroxidation des acides gras sont dans la cellule alors que les LTPs sont sécrétées, il n'y donc pas à priori de raison pour que les LTPs fixent les oxilipines, sauf si la compartimentation cellulaire disparaît. C'est le cas lors de l'extraction et la purification des LTPs, et également lors d'une attaque par un bioagresseur qui en détruisant quelques cellules pour se nourrir pourrait induire des réactions en cascade conduisant à une RH (mort cellulaire programmée et nécrose des tissus).

En conclusion, les LTP1 d'orge et de blé fixeraient des oxilipines issues de la voie des LOX et POX, précurseurs de la cutine et impliquées dans la signalisation de défense des plantes. Cependant on peut se poser la question de l'origine des formes complexées des LTP1 d'orge et de blé. S'agit-il d'une modification post-traductionnelle ou d'un artefact de purification ? Pour vérifier cela, il serait nécessaire d'inhiber la voie des LOX et POX au cours de l'extraction des LTPs sans réaliser de délipidation. Si la forme complexée des LTP1 d'orge et de blé n'est plus observée, on concluerait donc à une origine artéfactuelle de l'adduit. Cet artéfact de purification pourraient par contre refléter une réalité biologique, puisque la décompartimentation cellulaire a bien lieu lors d'une attaque pathogène de la plante. Il est donc nécessaire, d'une part de déterminer la structure exacte du complexe, et d'autre part d'étudier les effets biologiques de ce complexe sur la plante. La formation d'un adduit covalent pourrait constituer un messenger responsable du développement de la réponse systémique acquise (RSA), qui intervient dans la protection de l'ensemble de la plante contre une agression pathogène. Nous aurions alors apporté la preuve moléculaire de l'implication des LTPs dans la signalisation cellulaire.

II) Matériels et Méthodes

Dans ce chapitre je présenterai les techniques utilisées pour réaliser l'étude structurale et dynamique sur la LTP1_1 de tabac : dans une première partie j'exposerai la méthodologie employée pour déterminer une structure tridimensionnelle de protéine par RMN et modélisation moléculaire ; la seconde partie sera consacrée aux méthodes d'études de la dynamique interne des protéines par relaxation nucléaire ; la troisième partie traitera des simulations de dynamique moléculaire et enfin la dernière partie exposera la méthodologie de la modélisation comparative.

1. Détermination de la structure tridimensionnelle d'une protéine (par RMN et modélisation moléculaire)

Depuis la publication de la première structure 3D de protéine par RMN en 1985 (Williamson et al. 1985), d'énormes progrès technologiques et méthodologiques ont été réalisés pour rendre la spectroscopie RMN performante dans la détermination de structures tridimensionnelles de macromolécules (protéines, acides nucléique, sucres, complexes protéine/protéine ou protéine/acide nucléique). Ces progrès ont permis d'augmenter la résolution et la sensibilité de la RMN, les points faibles de la technique. Ils ont consisté dans le développement de spectromètres à haut champ (25 T à 2K actuellement) et de sondes cryogéniques. Ces progrès technologiques ont été accompagnés de développements méthodologiques : avec l'apparition du marquage isotopique (^{15}N , ^{13}C et ^2H), le développement de séquences RMN 3D, et 4D (Cavanagh et al. 1996; Sattler et al. 1999; Kanelis et al. 2001), la mise au point de la séquence TROSY (pour Transverse Relaxation Optimized Spectroscopy) (Pervushin et al. 1997) et dernièrement, l'avènement de l'utilisation des constantes de couplages dipolaires (Tolman et al. 1995; Tjandra and Bax 1997). Ces développements ont permis de réaliser l'étude structurale de biomolécules de taille de plus en plus importante (Tugarinov et al. 2004). Par exemple, l'enzyme malate synthase de *E. coli* de 723 résidus (Tugarinov et al. 2004) est actuellement la plus grosse protéine dont les déplacements chimiques sont connus (^1H , ^{13}C et ^{15}N).

Dans la suite de cette partie, je vais présenter tout d'abord les méthodes d'attribution RMN des protéines puis les contraintes structurales déduites de données RMN et utilisées en modélisation moléculaire et enfin les logiciels de modélisation moléculaire utilisés pour déterminer la structure tridimensionnelle de la LTP1_1 de tabac

1.1 Méthodes d'attribution RMN des protéines

Quelle que soit sa taille, l'étude RMN d'une protéine ne commence réellement qu'après l'attribution de ses déplacements chimiques. La structure 3D de l'inhibiteur de trypsine de taureau de 53 résidus, publiée en 1985 (Williamson et al. 1985) par l'équipe de Kurt Wüthrich, prix Nobel de chimie 2002, a été la première structure 3D de protéine en solution. L'attribution des déplacements chimiques de ses protons, seuls noyaux alors observables en abondance naturelle de façon exploitable à l'époque, a été obtenue par la méthode d'attribution séquentielle (Wüthrich 1986) basée sur l'effet NOE (Nuclear Overhauser Effect).

1.1.1 Méthode d'attribution séquentielle

Cette méthode exploite tout d'abord les informations de corrélations scalaires révélées par les expériences de type COSY (CORRélation SpectroscopY) et TOCSY (TOtal CORrelation SpectroscopY, encore appelées HOHAHA (HOmonuclear HArtman HAhN spectroscopy)). Ces expériences qui mettent en évidence des corrélations entre protons au travers des liaisons permettent d'obtenir les signatures caractéristiques des systèmes de spins de chaque acide aminé. L'étape suivante consiste à déterminer la position des acides aminés ainsi identifiés dans la séquence de la protéine. C'est ce que l'on appelle l'attribution séquentielle. Celle-ci se base sur des expériences NOESY (Nuclear Overhauser Effect SpectroscopY) qui permettent d'observer des protons distants de moins de 5 Å. Pour l'attribution séquentielle, on cherche les corrélations dipolaires entre le résidu i et le résidu $i+1$ et plus particulièrement, les corrélations $H\alpha_{(i)}-H_{N(i+1)}$ qui permettent de reconstituer l'enchaînement des systèmes de spins. Cette méthode d'attribution séquentielle est largement utilisée et applicable aux protéines de petite taille, généralement inférieure à une centaine de résidus. Toutefois, pour les protéines de plus de 80 résidus que sont les LTP1, l'attribution des protons à partir des expériences 2D est très difficile à réaliser en raison de nombreuses superpositions de signaux. Pour la LTP1_1 de tabac, un marquage au ^{15}N a donc été effectué. Ce simple marquage a été suffisant pour réaliser l'attribution des fréquences de résonance des noyaux ^1H et ^{15}N et déterminer par la suite la structure 3D de la protéine.

1.1.2 Attribution facilitée par un simple marquage ^{15}N ou ^{13}C

La méthode d'attribution séquentielle présentée précédemment est facilitée par un simple marquage ^{15}N (ou ^{13}C). L'attribution se réalise alors à partir de spectres 3D hétéronucléaires comme par exemple les expériences 3D TOCSY-ou NOESY-HSQC éditées ^{15}N ou ^{13}C . Ces expériences 3D sont simplement la combinaison de deux expériences 2D : la première TOCSY ou NOESY ^1H , et la deuxième HSQC (pour Heteronuclear Single Quantum Coherence) qui permet d'étaler l'information selon la fréquence de l'hétéronoyau auquel sont liés les protons (figure 2.1). L'expérience HSQC peut en effet être considérée comme une expérience « COSY hétéronucléaire » de type $^{15}\text{N}/\text{H}_\text{N}$ (ou $^{13}\text{C}/\text{H}_\text{C}$) qui se compose de deux séries d'impulsions : une séquence INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) et une séquence INEPT-inverse qui assurent respectivement le transfert de polarisation du proton vers le noyau ^{15}N (ou ^{13}C) puis de ce dernier à nouveau vers le proton, qui est finalement détecté.

Le gain en résolution de ces expériences 3D est très important grâce à la dispersion des gammes de déplacements chimiques ^{15}N ou ^{13}C . C'est cette dispersion dans la dimension ^{15}N ou ^{13}C qui facilite l'attribution des déplacements chimiques ^1H .

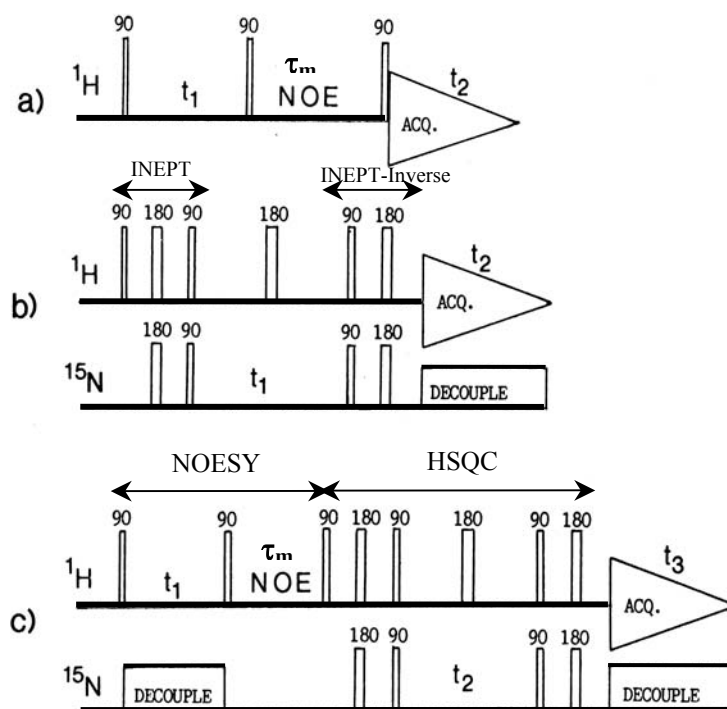


Figure 2.1 : Exemples de schémas simplifiés de séquences d'impulsions pour des expériences 2D et 3D. (a) Séquence d'impulsions pour l'expérience 2D NOESY ; (b) Séquence d'impulsions pour l'expérience 2D ^1H - ^{15}N HSQC ; (c) Séquence d'impulsions pour l'expérience 3D ^1H - ^{15}N NOESY-HSQC, obtenue en combinant celles de (a) et (b). Les impulsions radiofréquences sont indiquées par des barres verticales étroites et larges qui correspondent respectivement à des impulsions de 90 et 180°.

1.1.3 L'expérience 3D ^1H - ^{15}N NOESY-HSQC

Le schéma de base de la séquence d'impulsions de l'expérience 3D ^1H - ^{15}N NOESY-HSQC est illustré dans la figure 2.1c. Schématiquement l'aimantation originelle du spin K_i (^1H) est transférée sur le spin I_j (^1H) pendant le temps de mélange τ_m de la séquence NOESY. L'expérience HSQC transfère ensuite l'aimantation du spin I_j vers le spin S_j (^{15}N) via le couplage scalaire $^1J_{\text{IS}}$. Pendant le temps t_1 l'aimantation évolue sous l'effet du seul déplacement chimique de S ; enfin on observe le retour de l'aimantation sur le spin I_j via le couplage scalaire $^1J_{\text{IS}}$ pour la détection (figure 2.2). Le résultat de cette expérience est schématisé sur la figure 2.3.

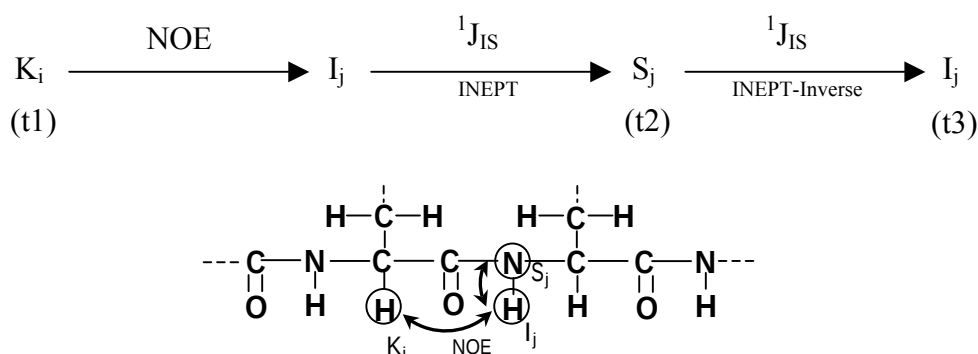


Figure 2.2: transfert d'aimantation dans une expérience 3D ^1H - ^{15}N NOESY-HSQC.

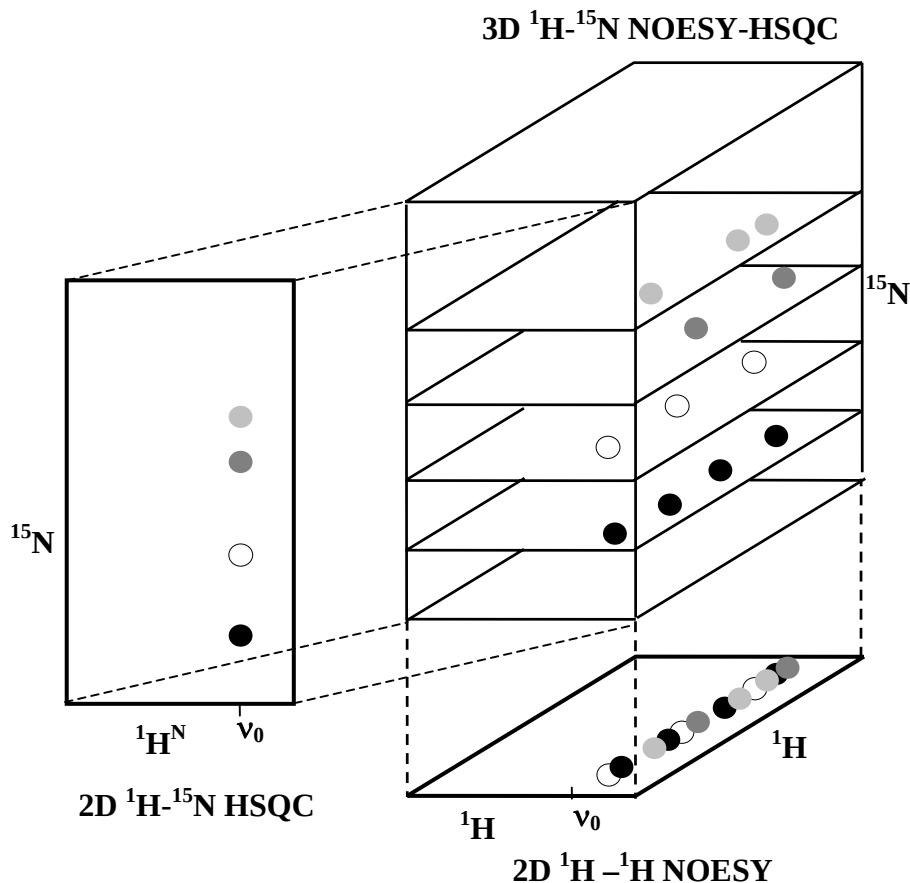


Figure 2.3: Schéma de la relation entre une expérience 3D ^1H - ^{15}N NOESY-HSQC et deux expériences 2D ^1H - ^{15}N HSQC et ^1H - ^1H NOESY. Les systèmes de spins dont les protons ^1H résonnent à la fréquence ν_0 sont superposés dans l'expérience ^1H - ^1H NOESY. Cette dégénérescence est levée grâce à l'expérience 2D ^1H - ^{15}N HSQC. L'expérience 3D permet de disperser l'information contenue dans l'expérience NOESY avec la résolution des déplacements chimiques ^{15}N obtenue avec l'expérience HSQC.

1.1.4 Expériences tridimensionnelles triple résonance

La LTP1_1 a également été doublement marquée en ^{13}C et ^{15}N . Ce double marquage isotopique nous a permis de réaliser l'attribution des déplacements chimiques ^{13}C utiles pour la suite de notre étude structurale de la protéine (cf. chapitre xénon). D'autre part, l'avantage de posséder une protéine doublement marquée réside dans la possibilité de réaliser l'attribution des déplacements chimiques sans avoir recours à la méthode d'attribution séquentielle classique. En effet, comme nous l'avons vu précédemment, cette stratégie d'attribution est basée sur l'existence et l'identification des NOEs. Il existe pour les protéines doublement marquées une attribution séquentielle alternative basée sur les transferts de cohérence à travers les principaux couplages scalaires homonucléaires et hétéronucléaires suivants : $^1J_{\text{C-C}}$, $^1J_{\text{C-N}}$, $^1J_{\text{H-N}}$, $^1J_{\text{H-C}}$, $^2J_{\text{H-H}}$, $^2J_{\text{C-N}}$, et enfin $^2J_{\text{C-C}}$. La figure 2.4 résume les constantes de couplages scalaires 1J et 2J existant au sein d'une protéine. L'attribution est facilitée et réalisée à partir d'une série d'expériences 3D triple résonance qui permet, d'une part de réaliser l'attribution des résonances des noyaux du squelette peptidique, et d'autre part celle des noyaux des chaînes latérales des acides aminés. Je me contenterai de décrire brièvement le résultat des expériences 3D triple résonance sans détailler leurs séquences d'impulsions.

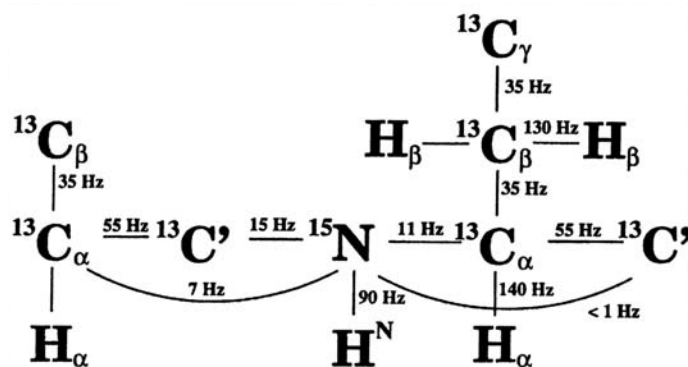


Figure 2.4 : Valeurs caractéristiques des constantes de couplage scalaire 1J et 2J pour une protéine

Les expériences 3D triple résonances HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCANH, CBCA(CO)NH (Cavanagh et al. 1996; Sattler et al. 1999; Kanelis et al. 2001) permettent l'attribution complète des noyaux du squelette peptidique ($^1\text{H}_\text{N}$, ^{15}N , ^{13}CO , $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$), ainsi que celles des protons et des carbones β des chaînes latérales ($^1\text{H}_\beta$ et $^{13}\text{C}_\beta$).

Les noms de ces expériences proviennent des noyaux qui sont impliqués dans les chemins de cohérences suivis. Les noyaux qui sont entre parenthèses ne sont pas édités au cours de l'expérience. Par exemple, l'expérience HNCA corrèle les résonances $^1\text{H}_\text{N}$ et ^{15}N à la

résonance $^{13}\text{C}_\alpha$ du même résidu i et à la résonance $^{13}\text{C}_\alpha$ du résidu $i-1$. En revanche l'expérience $\text{HN}(\text{CO})\text{CA}$ relie les résonances $^1\text{H}_\text{N}$ et ^{15}N d'un même résidu avec seulement la résonance $^{13}\text{C}_\alpha$ du résidu $i-1$ via le carbonyle CO, ce dernier n'étant pas édité au cours de l'expérience. Les expériences HNCA et $\text{HN}(\text{CO})\text{CA}$, HNCO et $\text{HN}(\text{CA})\text{CO}$, CBCANH et $\text{CBCA}(\text{CO})\text{NH}$ s'interprètent donc par paires. De plus, chacun de ces couples fournit un chemin d'attribution séquentielle possible.

a) HNCA et $\text{HN}(\text{CO})\text{CA}$

Ce couple d'expériences permet d'obtenir un chemin d'attribution séquentielle qui relie les $^{13}\text{C}_\alpha$ à la fois aux résonances $^1\text{H}_\text{N}$, ^{15}N du même résidu et à celles du résidu précédent (figure 2.5).

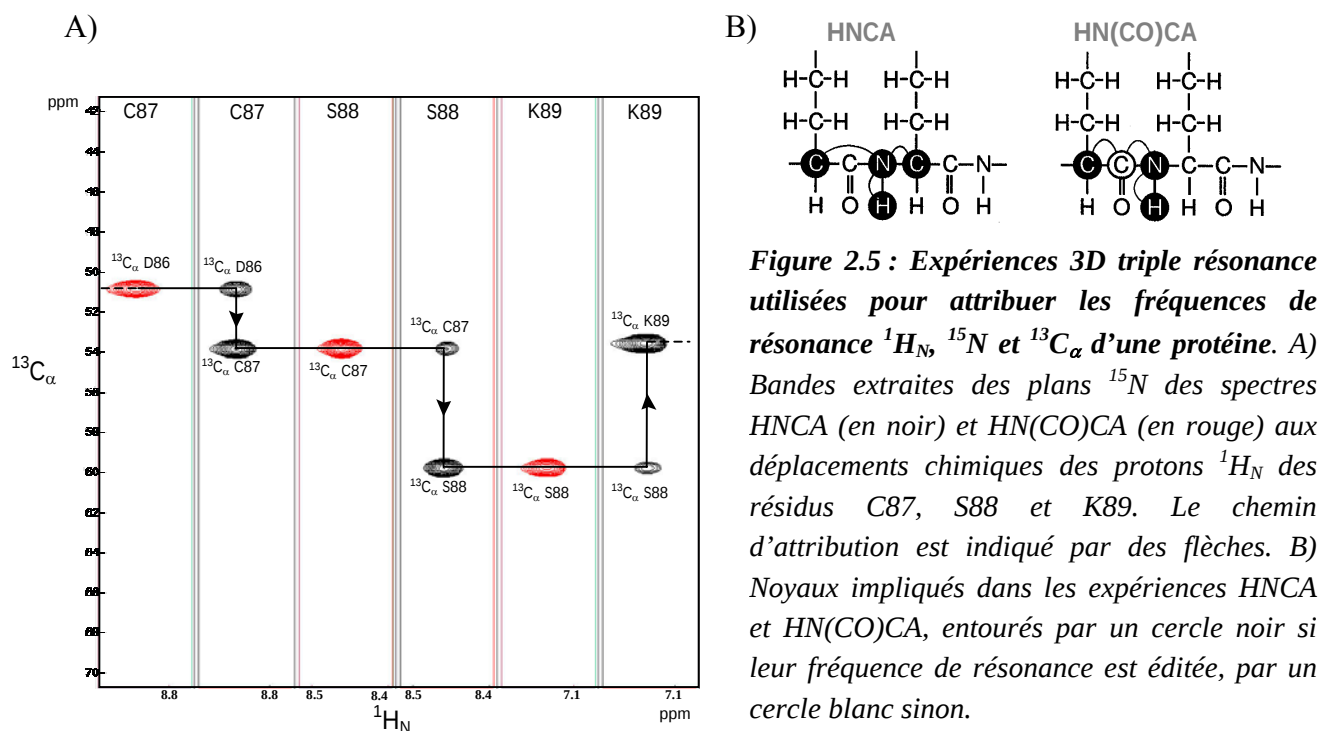
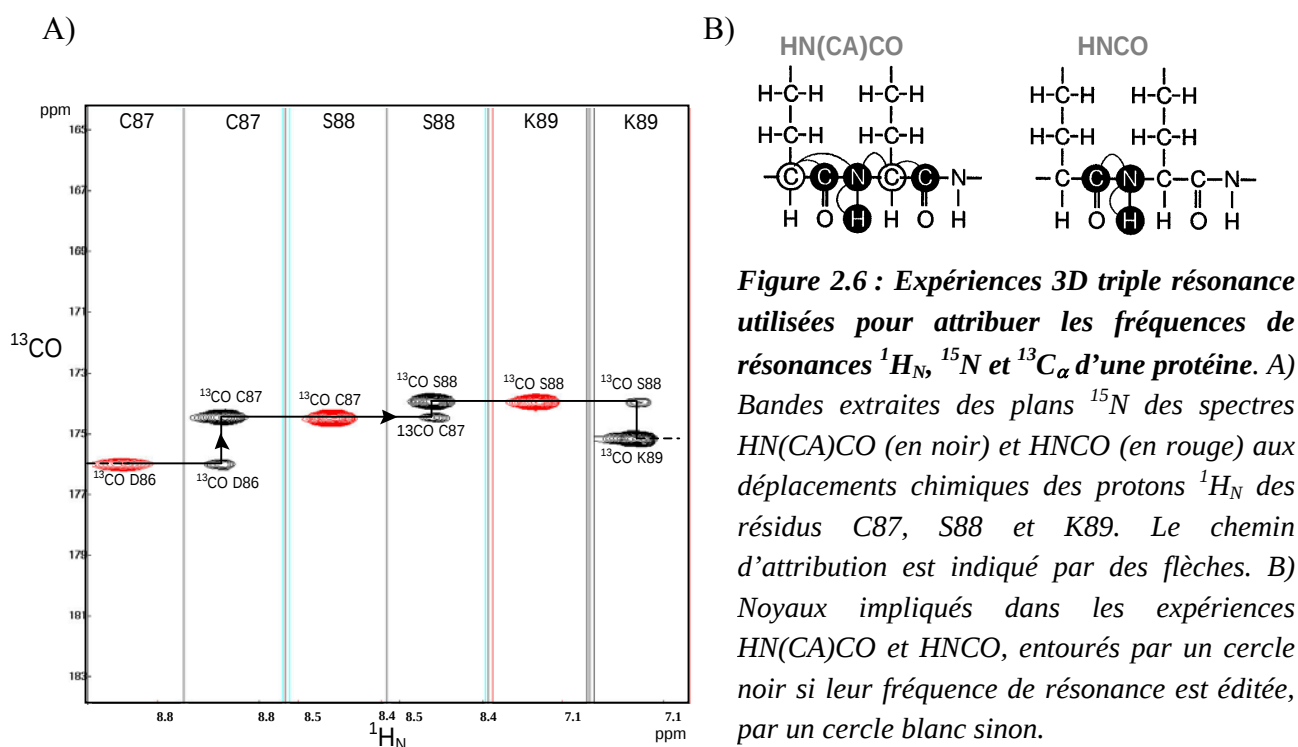


Figure 2.5 : Expériences 3D triple résonance utilisées pour attribuer les fréquences de résonance $^1\text{H}_\text{N}$, ^{15}N et $^{13}\text{C}_\alpha$ d'une protéine. A) Bandes extraites des plans ^{15}N des spectres HNCA (en noir) et $\text{HN}(\text{CO})\text{CA}$ (en rouge) aux déplacements chimiques des protons $^1\text{H}_\text{N}$ des résidus C87, S88 et K89. Le chemin d'attribution est indiqué par des flèches. B) Noyaux impliqués dans les expériences HNCA et $\text{HN}(\text{CO})\text{CA}$, entourés par un cercle noir si leur fréquence de résonance est éditée, par un cercle blanc sinon.

Ces deux expériences suffiraient donc à identifier la position des acides aminés au sein de la séquence. Cependant il peut y avoir des superpositions de corrélations ne permettant pas d'identifier tous les résidus. Ces ambiguïtés sont alors levées par l'interprétation d'un autre couple d'expériences qui fournit des corrélations alternatives : le couple $\text{HN}(\text{CA})\text{CO}$ et HNCO .

b) HN(CA)CO et HNCO

Selon le même principe, les expériences HN(CA)CO et HNCO procurent les corrélations entre les résonances $^1\text{H}_\text{N}$, ^{15}N et les CO intra et interrésiduels ; HNCO fournit uniquement les corrélations interrésiduelles tandis que HN(CA)CO, en plus de ces corrélations, donne celles qui sont intrarésiduelles (figure 2.6).

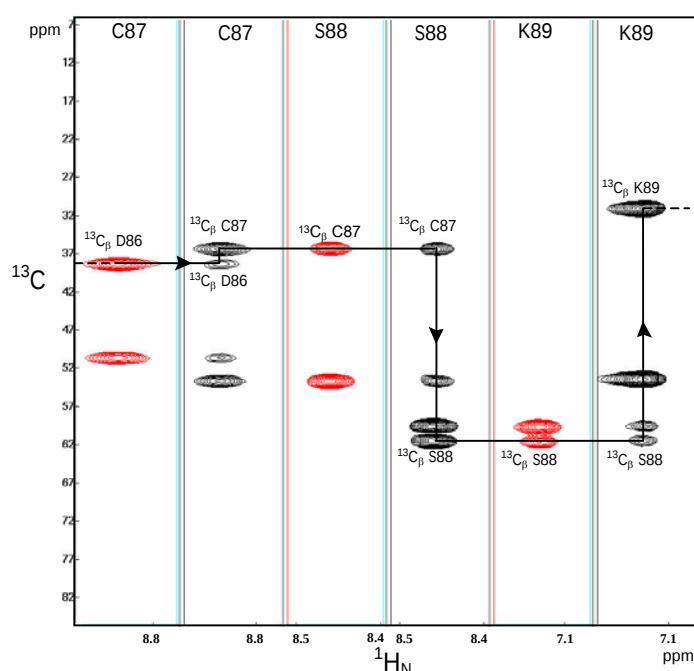


L'analyse combinée de ces deux couples d'expériences offre donc systématiquement au moins un chemin d'attribution séquentielle. Il faut noter toutefois que ce dernier couple d'expériences, même s'il constitue un chemin d'attribution séquentielle alternatif, est rarement utilisé. En effet, la dispersion des déplacements chimiques des carbonyle (CO) est trois fois moins grande (environ 10 ppm) que celle des C_α (environ 30 ppm), donc le risque de superpositions et d'ambiguïtés éventuelles est plus important. Pour lever d'éventuelles ambiguïtés persistantes, on a recours au dernier couple d'expériences : CBCANH et CBCA(CO)NH.

c) CBCANH et CBCA(CO)NH

Ce dernier couple d'expériences fournit les mêmes informations que le couple HNCA et HN(CO)CA, de plus met en évidence les corrélations entre les résonances des noyaux $^1\text{H}_\text{N}$, ^{15}N et les $^{13}\text{C}_\beta$ intra et interrésiduels, constituant un troisième et dernier chemin d'attribution séquentielle possible (figure 2.7).

A)



B)

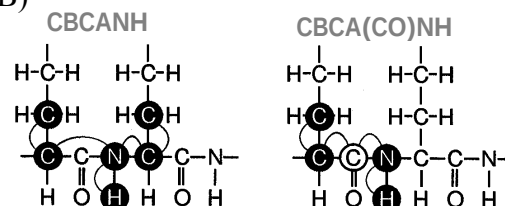


Figure 2.7 : Expériences 3D triple résonance utilisées pour attribuer les fréquences de résonances $^1\text{H}_\text{N}$, ^{15}N et $^{13}\text{C}_\alpha$ et $^{13}\text{C}_\beta$ d'une protéine. A) Bandes extraites des plans ^{15}N des spectres CBCANH (en noir) et CBCA(CO)NH (en rouge) aux déplacements chimiques des protons $^1\text{H}_\text{N}$ des résidus C87, S88 et K89. Le chemin d'attribution est indiqué par des flèches. Seules les connectivités relatives aux $^{13}\text{C}_\beta$ sont annotées. B) Noyaux impliqués dans les expériences CBCANH et CBCA(CO)NH, entourés par un cercle noir si leur fréquence de résonance est éditée, par un cercle blanc sinon.

De toutes ces expériences, on tire donc des données complémentaires qui permettent de réaliser l'attribution séquentielle plus facilement, en limitant les risques d'erreurs. Jusqu'à présent, seules les résonances des protons et des carbones β des chaînes latérales sont attribuées. Il reste à réaliser l'attribution complète des chaînes latérales pour aboutir par la suite à la structure 3D de la protéine.

1.1.5 Attribution des chaînes latérales

L'expérience HCCH-TOCSY permet d'identifier tous les protons et les carbones aliphatiques des chaînes latérales. Le résultat de cette expérience 3D contient la même information qu'une expérience 3D TOCSY-HSQC éditée ^{13}C . Toutefois, cette expérience est bien plus sensible que l'expérience 3D TOCSY-HSQC éditée ^{13}C car le transfert HCCH-TOCSY est bien plus efficace qu'un transfert par les hydrogènes. En effet, les

constantes de couplage 2J ou 3J entre les hydrogènes sont de l'ordre de quelques hertz (4-10 Hz) alors que celles 1J entre les carbones et les hydrogènes sont beaucoup plus élevées (de l'ordre de centaine d'Hertz 140 Hz).

Après avoir réalisé l'attribution de tous les déplacements chimiques, il est nécessaire pour obtenir la structure 3D d'une protéine d'extraire les contraintes structurales utilisées dans la modélisation moléculaire.

1.2 Les contraintes structurales utilisées en modélisation moléculaire

1.2.1 Les contraintes de distance

Les paramètres les plus importants pour la modélisation moléculaire sont les contraintes de distance. Ces dernières sont extraites des expériences de type NOESY : 2D homonucléaire proton, des expériences 3D NOESY-HSQC édité ^{15}N ou ^{13}C . Les volumes des corrélations dipolaires obtenues dans ces expériences peuvent être en première approximation convertis en contraintes de distance grâce à la relation suivante :

$$d_{ij} \propto I_{ij}^{1/6} \quad (2.1)$$

où d_{ij} et I_{ij} désignent respectivement la distance et le volume du pic de corrélation dipolaire entre les atomes i et j .

1.2.2 Les contraintes d'angles dièdres

Les angles dièdres qui décrivent le squelette peptidique d'une protéine sont au nombre de trois : ϕ , ψ , ω (figure 2.8). Les angles dièdres ϕ et ψ , pour un résidu i , sont définis respectivement autour des liaisons $N_i-C_{\alpha i}$ et $C_{\alpha i}-C'_i$. L'angle dièdre ω décrit la liaison peptidique qui est généralement plane.

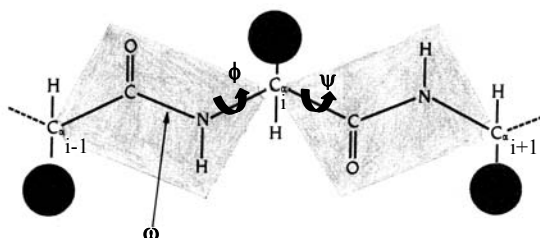


Figure 2.8 : Représentation schématique du squelette peptidique d'une protéine. Les cercles noirs représentent les chaînes latérales. Les angles dièdres ϕ , ψ , ω qui définissent le squelette peptidique sont indiqués.

Les contraintes d'angles dièdres sont extraites des constantes de couplage scalaire 3J entre protons séparées par trois liaisons. Ces constantes de couplage sont reliées à l'angle de torsion θ autour de la liaison centrale par l'intermédiaire de la formule empirique suivante, appelée relation de Karplus (Karplus 1959):

$$^3J(\theta) = A \cos^2 \theta + B \cos \theta + C \quad (2.2)$$

où A , B , C sont des coefficients semi-empiriques

La constante de couplage $^3J_{\text{HN-H}\alpha}$ est la plus communément mesurée. Elle est reliée à l'angle dièdre ϕ par la relation de Karplus avec les coefficients semi-empiriques A , B , C définis par Pardi en 1984 (Pardi et al. 1984) (figure 2.9) :

$$^3J_{\text{HN-H}\alpha}(\theta) = 6.4 \cos^2 \theta - 1.4 \cos \theta + 1.9 \text{ avec } \theta = |\phi - 60^\circ| \quad (2.3)$$

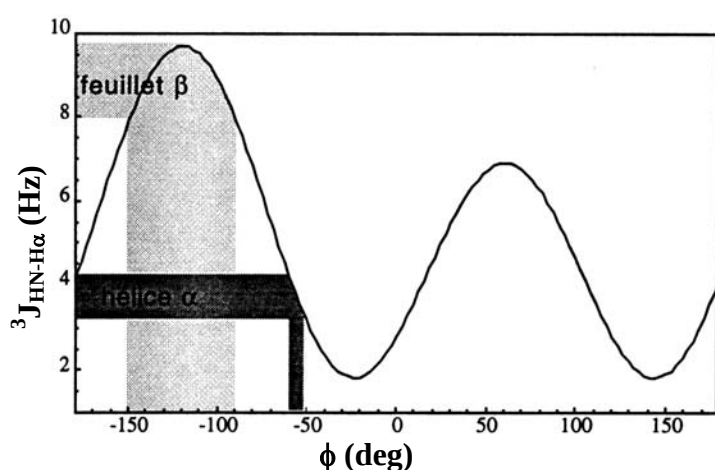


Figure 2.9 : Courbe de variation de la constante de couplage $^3J_{\text{HN-H}\alpha}$ en fonction de l'angle ϕ . Les coefficients semi-empiriques A , B , C sont ceux définis par Pardi en 1984 rappelés dans l'équation (2.3) (Pardi et al. 1984). Les plages d'intervalles pour la constante de couplage $^3J_{\text{HN-H}\alpha}$ des hélices α droites et des feuillets β canoniques sont indiquées en fonction des valeurs angulaires permises.

Différentes méthodes (Cavanagh et al. 1996) permettent de mesurer les constantes de couplage $^3J_{\text{HN-H}\alpha}$. Pour l'étude de la LTP1_1 nous avons utilisé une méthode quantitative (Vuister and Bax 1993) reposant sur l'expérience HNHA.

Cette expérience est très largement utilisée lors de l'étude par RMN de la structure 3D de protéines simplement marquées ^{15}N . Elle met en corrélation les résonances $^1\text{H}_\text{N}$, ^{15}N et $^1\text{H}_\alpha$ d'un même résidu. Cette expérience permet par ailleurs d'identifier les résonances $^1\text{H}_\alpha$ des résidus. Le résultat de l'expérience HNHA se traduit par la présence de deux pics en antiphase aux déplacements chimiques de ^{15}N et du $^1\text{H}_\text{N}$ du résidu i . Dans le plan ^{15}N , nous avons un pic positif sur la diagonale et une corrélation croisée négative.

Le rapport d'intensité des deux pics est relié à la constante de couplage $^3J_{\text{HN-H}\alpha}$ par la relation suivante :

$$I_c/I_d = -\tan^2(2 \pi ^3J_{\text{HN-H}\alpha} \zeta) \quad (2.4)$$

où I_c et I_d correspondent respectivement aux intensités des pics de corrélations croisées et de la diagonale. ζ est un délai de 13.5 ms défini dans la séquence d'impulsions de l'expérience 3D HNHA. La courbe $^3J_{\text{HN-H}\alpha}$ en fonction de $|I_c/I_d|$ déduite de l'expression (2.4) permet d'obtenir les constantes de couplages $^3J_{\text{HN-H}\alpha}$ et d'identifier rapidement les résidus impliqués dans les hélices et dans les structures étendues comme les feuillets (figure 2.10).

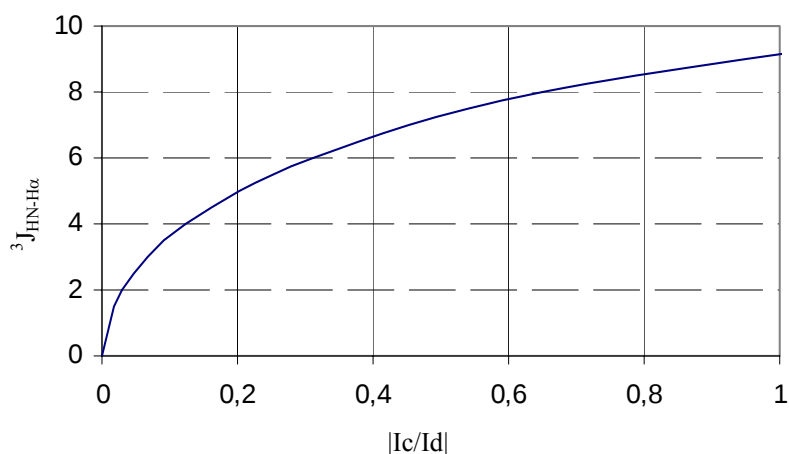


Figure 2.10 : Evolution de $^3J_{\text{HN-H}\alpha}$ en fonction du rapport $|I_c/I_d|$ (d'après (Vuister and Bax 1993)). Les résidus impliqués dans une hélice possèdent une valeur de constante de couplage $^3J_{\text{HN-H}\alpha}$ faible ($\leq 5\text{Hz}$) alors que ceux participant à un feuillet en ont une élevée ($\geq 8\text{Hz}$).

Remarque : la méthode de détermination quantitative HNHA entraîne une sous estimation de la constante de couplage $^3J_{\text{HN-H}\alpha}$. Dans le cas de la nucléase du staphylocoque, protéine de 136 résidus utilisée par Vuister et ses collaborateurs (Vuister and Bax 1993) pour valider la méthode HNHA, une sous estimation de 11% a été mesurée. D'autre part, le cas des résidus glycine doit être considéré séparément. En effet, ces résidus possèdent deux protons $\text{H}\alpha$ et le couplage géminale tend à moyenner les deux constantes de couplages mesurées. De plus, l'absence d'attribution stéréospécifique des deux protons $\text{H}\alpha$ ne permet pas de conclure sur la définition de l'angle dièdre ϕ .

1.2.3 Données qualitatives

Des informations qualitatives sur l'existence de liaisons hydrogène peuvent être obtenues à partir d'expériences d'échange des protons après dissolution dans un solvant deutéré. Expérimentalement, l'échange des protons amide de la LTP1_1 avec ceux du solvant deutéré a été suivi en réalisant des spectres ^1H - ^{15}N HSQC. Les protons amide de la protéine exposés au solvant s'échangent rapidement et leurs signaux disparaissent immédiatement alors que les protons engagés dans une liaison hydrogène ou inaccessibles au solvant s'échangent lentement et leurs signaux persistent plus ou moins longtemps.

Les paramètres structuraux obtenus à partir des expériences RMN sont utilisés ensuite en modélisation moléculaire pour construire le modèle 3D de la protéine.

1.3 Modélisation moléculaire

Déterminer la structure tridimensionnelle d'une protéine par RMN et modélisation moléculaire consiste à trouver un ensemble de conformations qui correspondent à des minima énergétiques et qui sont en accord avec les données expérimentales. En 1985, la structure 3D de l'inhibiteur de trypsine de taureau (Williamson et al. 1985) a été obtenue par la méthode de géométrie des distances développée par W. Braun (Braun and Go 1985) dans l'équipe de Kurt Wüthrich. Cette méthode de géométrie des distance est de moins en moins utilisée, au bénéfice des méthodes de calcul par recuit simulé ; elle ne sera donc pas exposée ici. Nous avons choisi pour déterminer la structure 3D de la LTP1_1 le logiciel CNS (Crystallography & NMR System) <http://cns.csb.yale.edu/v1.1/> (Brünger et al. 1998) qui utilise ces méthodes de calcul de recuit simulé.

1.3.1 Le logiciel CNS

Ce programme est un logiciel de modélisation moléculaire utilisé pour déterminer les structures 3D de macromolécules en utilisant les données expérimentales obtenues par radiocristallographie ou RMN. Avant d'expliquer la méthode de recuit simulé, je vais introduire brièvement la notion d'énergie potentielle d'une molécule qui sera expliquée plus en détail avec la technique de dynamique moléculaire dans le paragraphe 3 de simulations de dynamique moléculaire.

a) Fonction empirique de l'énergie d'une molécule

On peut considérer l'énergie potentielle d'une molécule comme la somme de deux termes : un terme représentant l'ensemble des interactions entre atomes non liés et un terme lié à la géométrie covalente de la molécule. Ces termes seront détaillés dans la troisième partie de ce chapitre (p 63).

A cette énergie potentielle de la molécule, on peut ajouter un troisième terme énergétique représentant les contraintes expérimentales de distance déduites de l'effet NOE et d'angle dièdre résultant de la mesure des constantes de couplages. Ce dernier terme énergétique est nécessaire pour déterminer la structure 3D de la protéine.

b) Recuit simulé

Les méthodes de recuit simulé, dynamique moléculaire à très haute température et à température constante permettent en un temps raisonnable d'explorer convenablement l'espace conformationnel et de générer des structures compatibles avec les données expérimentales. Le recuit simulé peut se réaliser en utilisant les coordonnées cartésiennes ou internes.

Utilisation des coordonnées cartésiennes

Partant de structures aléatoires, le recuit simulé s'effectue en plusieurs étapes : une première étape de chauffage permet d'atteindre une température élevée (1000 ou 2000 K), une seconde période permet l'évolution du système à haute température pour l'exploration de l'espace conformationnel et la recherche d'un puits de potentiel; enfin une dernière étape de refroidissement permet de stabiliser le système dans le puits de potentiel ainsi déterminé. Pour diminuer les temps de calcul, le champ de force utilisé pour le recuit simulé est simplifié (pas de contribution des termes électrostatiques et termes de van der Waals simplifiés). Les constantes de force des termes d'énergie interne sont augmentées afin de maintenir une bonne géométrie covalente.

Utilisation des coordonnées internes

Le logiciel CNS offre également la possibilité d'effectuer les calculs dans l'espace des angles dièdres. Les structures initiales sont des structures aléatoires et le protocole est similaire à celui décrit précédemment, mais les températures de chauffage sont beaucoup plus élevées 10000K.

La réalisation d'un recuit simulé en coordonnées internes permet une exploration de l'espace conformationnel beaucoup plus rapide par rapport un recuit simulé dans les coordonnées cartésiennes.

Dans ce travail nous avons également utilisé ARIA (Ambiguous Restraints for Iterative Assignment) <http://www.pasteur.fr/recherche/unites/Binfo/aria/> (Nilges et al. 1997; Nilges and O'Donoghue 1998; Linge et al. 2001; Linge et al. 2003a), un logiciel d'attribution semi-automatique des pics NOE qui fonctionne en interface avec le logiciel CNS.

1.3.2 Le logiciel ARIA

ARIA exploite de nombreux concepts nouveaux. Le plus important d'entre eux est la notion d'ADR (Ambiguous Distance Restraint) ou contrainte de distance ambiguë qui a été introduite par M. Nilges en 1995 (Nilges 1995). Cette notion permet d'utiliser les contraintes de distance ambiguës dans la modélisation moléculaire alors qu'auparavant seulement les contraintes de distance non ambiguës étaient utilisées.

a) Contraintes de distance ambiguës

La notion d'ADR permet de considérer un pic NOE non plus comme le résultat de l'interaction dipolaire d'une unique paire de protons mais comme la somme des contributions dipolaires entre plusieurs paires de protons dont les fréquences de résonance sont proches. Un pic NOE avec une attribution multiple ($n > 1$, n : nombre d'attribution) peut être considéré comme la superposition de n pics NOE. Sachant que la distance entre 2 protons est reliée au volume du pic NOE par la relation (2.1), la contrainte sera appliquée sur une distance moyenne $\bar{d}$ calculée par la relation suivante :

$$\bar{d} = \left(\sum_{i=1}^n d_i^{-6} \right)^{-1/6} \quad (2.5)$$

où chaque distance d_i correspond à la distance entre une paire d'atomes d'hydrogène p_i et q_i et n représente le nombre d'attributions possibles.

b) Principe de fonctionnement du logiciel ARIA

L'utilisation simultanée des logiciels ARIA et CNS permet de réaliser l'attribution des pics NOE et le calcul des structures 3D de manière itérative (figure 2.11).

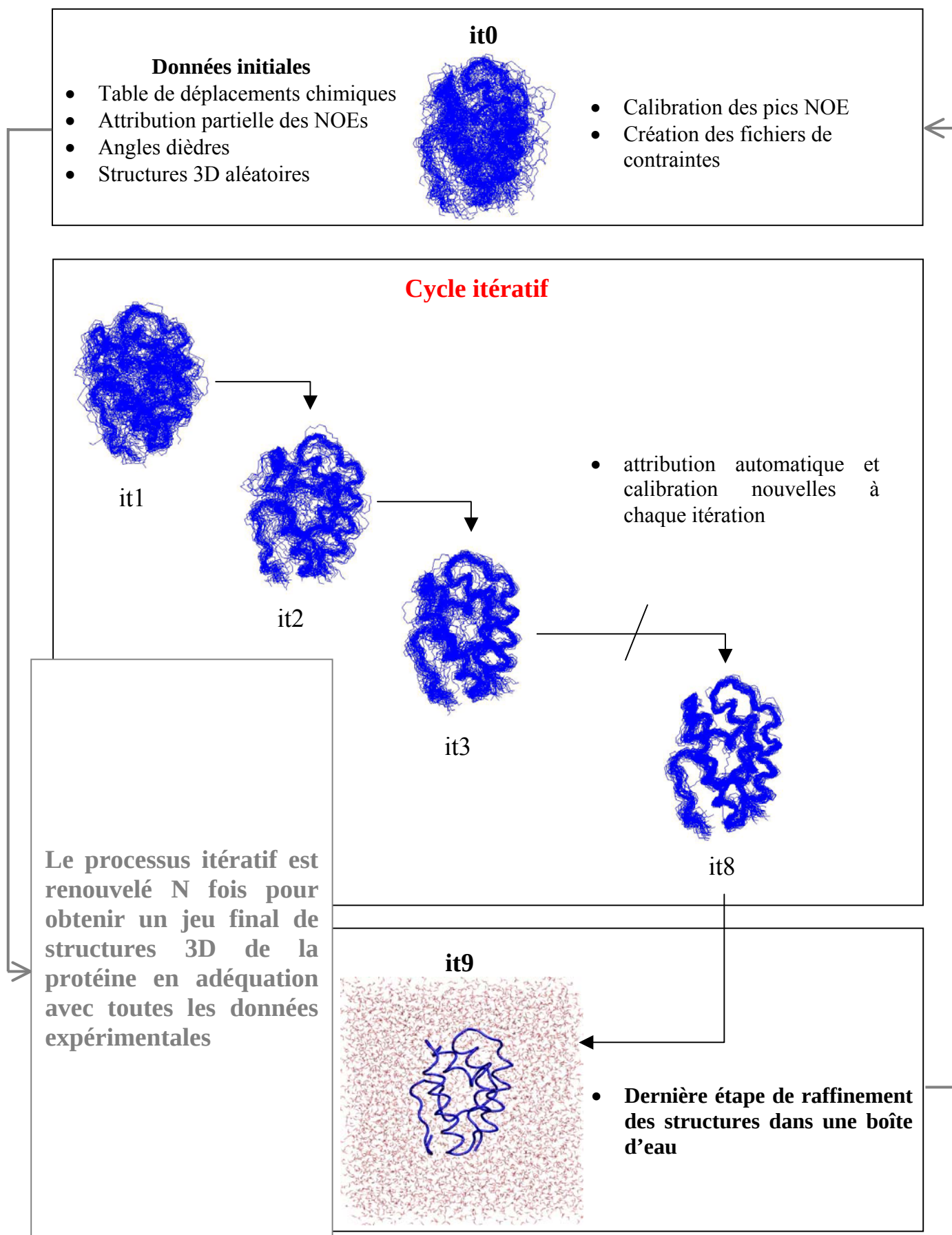


Figure 2.11 : description schématique simplifiée du processus itératif ARIA/CNS utilisé pour la modélisation moléculaire de la LTP1_1 de tabac.

ARIA utilise comme données initiales une table de déplacements chimiques, une liste de pics NOE avec une attribution partielle réalisée manuellement (corrélations intra-résiduelles, séquentielles, et certaines longues distances non ambiguës). Il peut prendre également en compte des contraintes d'angles dièdres, de constantes de couplage dipolaire résiduelle et de liaisons hydrogène. ARIA réalise d'abord la calibration des pics NOE et crée un fichier de contraintes de distance qui est utilisé par le logiciel CNS pour générer un premier jeu de structures 3D de la protéine (par défaut 20 structures) : c'est l'itération 0. Les structures 3D de départ sont des structures 3D générées aléatoirement. On entre ensuite dans un processus itératif où à chaque itération ARIA réalise des attributions automatiques à partir d'une fenêtre de déplacement chimique fixée par l'utilisateur. ARIA affine également la calibration des pics NOE et crée un fichier de contraintes que CNS utilise pour générer un nouveau jeu de structures. Ce sont les structures 3D de plus basse énergie (par défaut 7 structures 3D sur les 20 générées) de l'itération $i-1$ qui sont exploitées par ARIA pour réaliser l'attribution automatique et la calibration des pics NOE à l'itération i . Le nombre d'itérations réalisée par ARIA est fixé à 8. La neuvième itération correspond à une étape de raffinement et d'analyse des structures 3D. En effet, ARIA offre la possibilité de raffiner les structures 3D en réalisant une courte trajectoire de dynamique moléculaire en solvant explicite dans un champ de force réaliste (Linge et al. 2003b). Les attributions des NOE réalisées automatiquement par ARIA au cours du cycle itératif sont vérifiées manuellement et sont introduites, ou non au cycle itératif suivant. Ce processus est renouvelé N fois jusqu'à introduction de toutes les données expérimentales qui doivent être en accord avec les structures 3D obtenues.

c) Calibration des pics NOE

L'effet NOE est lié aux interactions dipôle-dipôle entre les moments magnétiques de spin de deux protons voisins dans l'espace. Dans le cas d'une paire de spins isolés, si on travaille à temps de mélange suffisamment court (temps pendant lequel on laisse évoluer le système), l'intensité NOE I_{ij} est proportionnelle à la vitesse de relaxation croisée σ_{ij} .

Cette vitesse se met alors sous la forme :

$$\sigma_{ij} = f(\tau_c) / d_{ij}^6 \quad (2.6)$$

où τ_c est le temps de corrélation rotationnel du vecteur d_{ij} . (τ_c = temps moyen nécessaire au vecteur d_{ij} pour tourner de 1 radian).

On voit donc que dans ce cas I_{ij} est inversement proportionnel à d_{ij}^6 et ceci permettra une première étape de « calibration » des NOE où l'intensité des pics NOE est convertie en contraintes de distance.

Dans ARIA, la calibration est obtenue à partir d'un jeu de structures 3D initiales. Pour chaque paire de protons correspondant à un pic NOE de volume I_{ij} on calcule une distance moyenne $\langle d_{ij} \rangle$ sur les structures. On en déduit un facteur de calibration global (C) de la manière suivante :

$$C = \sum_{NOE} \frac{1}{\langle d_{ij} \rangle^6 I_{ij}} \quad (2.7)$$

Les contraintes de distance observées (d_{ij}^{obs}) sont alors calculées de la façon suivante :

$$d_{ij}^{obs} = \frac{1}{(C \times I_{ij})^{1/6}} \quad (2.8)$$

Une des difficultés majeures de la modélisation des structures 3D provient de la dynamique interne des molécules. Les mouvements intramoléculaires des protéines qui couvrent une large échelle de temps (10^{-15} à 1s) rendent difficile la quantification précise des distances inter protons. La flexibilité de la molécule dans certaines régions telles que les boucles, les parties terminales etc. peut entraîner une diminution de certaines intensités NOE. Pour tenir compte de ce phénomène, on utilise la notion d'intervalle de valeurs permises pour les distances inter protons : $d_{min} < d_{ij}^{obs} < d_{max}$. Dans ARIA, l'erreur sur l'évaluation des contraintes de distance (Δ) est estimée à 12,5 % du carré de la distance observée d_{ij}^{obs} . Les valeurs d_{min} et d_{max} sont alors égale à :

$$d_{min} = d_{ij}^{obs} - \Delta \quad (2.9)$$

$$d_{max} = d_{ij}^{obs} + \Delta \quad (2.10)$$

$$\text{avec } \Delta = 0,125 (d_{ij}^{obs})^2. \quad (2.11)$$

D'autre part, il faut noter qu'ARIA offre la possibilité de réaliser une calibration plus élaborée en tenant compte de la matrice de relaxation pour limiter les effets de la diffusion de spin. Pour ce faire, il est nécessaire de connaître le temps de corrélation τ_c de la protéine (Linge et al. 2001; Linge et al. 2003a).

d) Attribution des pics NOE ambigus

Pour chaque pic NOE ambigu, c'est-à-dire dont une coordonnée de déplacement chimique correspond à la fréquence de résonance de plusieurs protons, une contrainte de distance ambiguë est créée. Le pic NOE est donc la somme de plusieurs contributions C_k correspondant aux différentes attributions possibles. Chaque contribution C_k est calculé de la manière suivante :

$$C_k = \frac{d_k^{-6}}{\sum_{i=1}^n d_i^{-6}} \quad (2.12)$$

où d_k est la distance moyenne de la paire de proton k calculée sur les 7 structures de plus basse énergie de l'itération précédente. On classe ensuite les contributions C_k par ordre décroissant et on exclut les contributions C_k de faible poids dès lors que :

$$\sum_{i=1}^{N_p} C_k > p \quad (2.13)$$

où p constitue la valeur du seuil d'attribution et N_p est le nombre de contributions C_k participant majoritairement à la définition du pic NOE. La valeur du seuil d'attribution est proche de 1 à la première itération et décroît progressivement au fur à mesure des itérations pour atteindre la valeur de 0,8 à l'itération 8. Pour essayer de comprendre intuitivement la signification de la valeur seuil d'attribution p (Nilges and O'Donoghue 1998), on considère un pic NOE qui possède deux attributions possibles donnant lieu à deux distances distinctes (figure 2.12). Si la plus courte des deux distances est de 2,5 Å, une valeur seuil $p = 0,999$ exclura une seconde distance supérieure à 7,9 Å, une valeur seuil de $p = 0,95$ exclura une distance supérieure à 4,1 Å et enfin une valeur seuil de $p = 0,8$ exclura une distance supérieure à 3,3 Å. Si la distance la plus courte est de 4 Å, les distances minimales exclues pour les valeurs seuil de p précédemment évoquées sont respectivement 12,6, 6,6 et 5,2 Å. Ceci signifie qu'au fur et à mesure des itérations, le nombre d'attributions des pics NOE ambigus est réduit en faveur des attributions les plus probables d'un point de vue de la structure 3D. Cette réduction du nombre d'attributions facilite le calcul et la convergence des structures 3D.

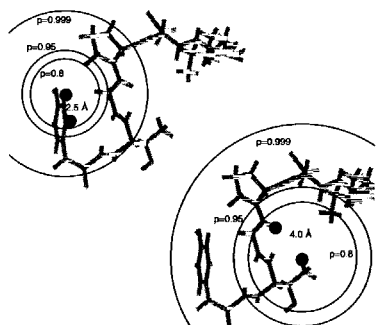


Figure 2.12 : Effet de différentes valeurs du seuil d'attribution p sur les contributions rejetées. Plus la valeur de p est petite, plus les faibles contributions correspondant à des atomes éloignés sont rejetées. La figure illustre le cas de deux distance : 2,5 Å et 4,0 Å avec différentes valeur de p .

De plus, au fur à mesure des itérations le traitement ambiguë des pics NOE permet d'exclure de la modélisation des pics dont les distances correspondantes sont systématiquement violées sur plus de la moitié des structures analysées. Le seuil d'exclusion correspond à une violation de plus de 50% de l'intensité du pic NOE considéré. Ces pics exclus sont vérifiés manuellement après chaque cycle itératif et peuvent correspondre à des pics qui résultent de situations diverses : problèmes de correction de ligne de base, bruit t1, mauvaise suppression du solvant, diffusion de spin.

Pour traiter correctement les pics NOE ambigus, les paramètres importants sont la valeur du seuil d'attribution p et celle de la fenêtre de déplacement chimique. Ce sont ces deux paramètres qui ont été optimisés pour obtenir la structure 3D de la LTP1_1 de tabac.

1.3.3 Sélection et analyse des structures

La stratégie ARIA/CNS permet de générer un grand nombre de structures. Une sélection s'impose pour ne conserver que les meilleures, du point de vue de la géométrie covalente et du respect des contraintes expérimentales. Les structures retenues seront considérées comme représentatives de la molécule en solution. Dans ce travail, une dizaine de modèles ont été retenus pour la LTP1_1 de tabac. Ce choix s'est effectué en utilisant divers critères :

a) Nombre de violations et géométrie covalente.

Après le raffinement des structures dans la boîte d'eau, ne sont conservées que les structures possédant un nombre minimal de violations de distance excédant 0,3 Å. Les structures retenues doivent également posséder une bonne géométrie covalente.

b) Les logiciels d'analyse PROCHECK et PROMOTIF

Les programmes PROCHECK (Laskowski et al. 1996) et PROMOTIF (Hutchinson and Thornton 1996) ont été utilisés pour l'analyse des structures obtenues de la LTP1_1. Ils facilitent l'identification des structures secondaires. Le logiciel PROCHECK fournit de nombreux indicateurs sur la qualité des structures, notamment la carte de Ramachandran qui définit les régions de l'espace conformationnel (ϕ , ψ) stériquement et énergétiquement favorables (Ramachandran et al. 1971).

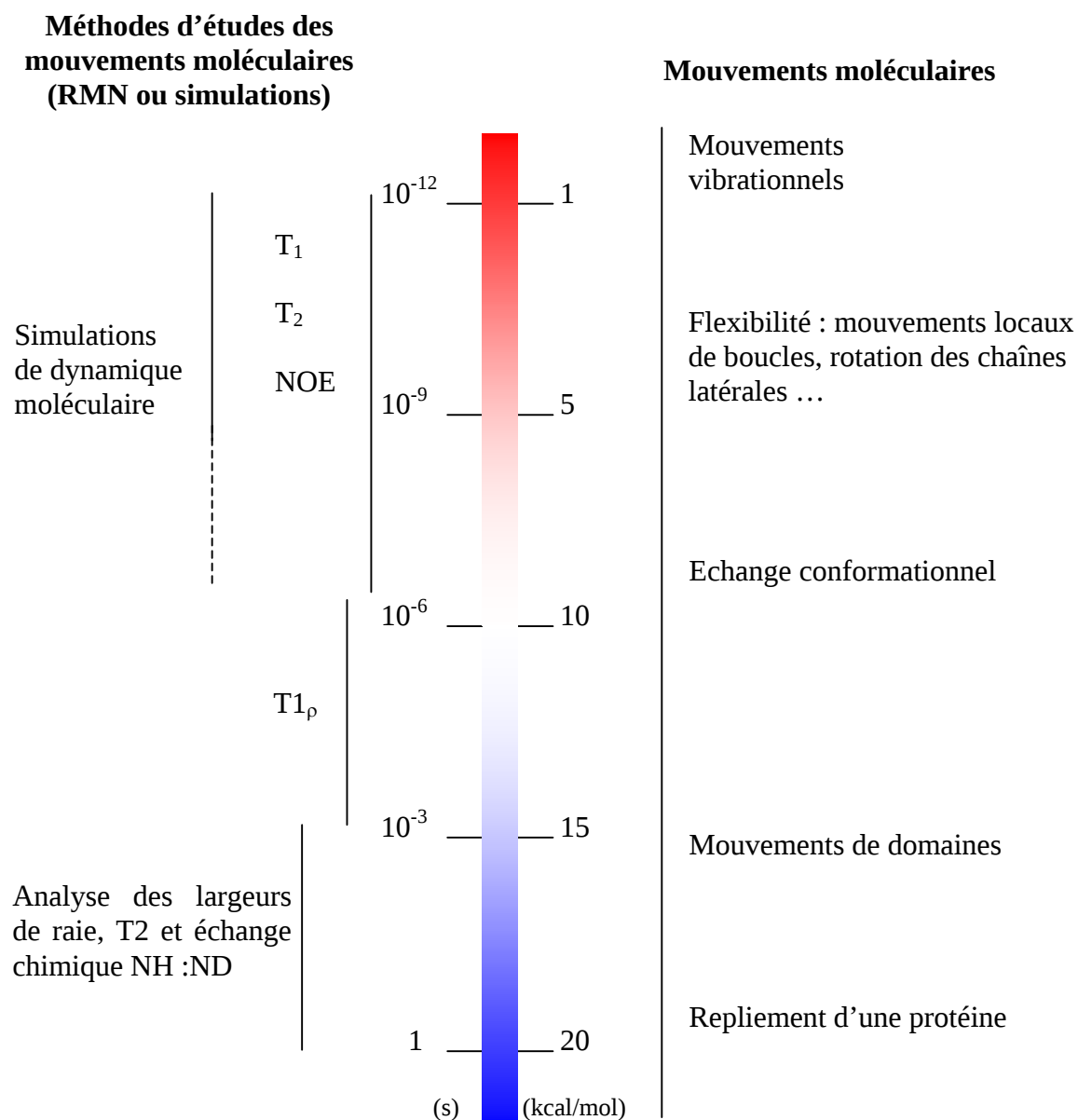


Figure 2.13 : échelle de temps (à gauche) et barrière énergétique (à droite) des mouvements moléculaires observables par RMN et simulations de dynamique moléculaire

2. Etude de la dynamique interne des protéines par relaxation nucléaire

Une des possibilités les plus intéressantes offertes par la RMN est sa capacité à sonder les mouvements moléculaires, sur une gamme de temps très large, localement et au niveau atomique. En effet, les spins nucléaires ont des propriétés remarquables pour analyser la dynamique interne des molécules. Ils sont sensibles à leur environnement proche et leur relaxation s'étale sur une gamme de fréquences très large qui traduit celle des mouvements des macromolécules. Depuis le développement du marquage isotopique des protéines avec les noyaux ^{15}N et ^{13}C , les études de dynamique par relaxation nucléaire se sont multipliées afin de comprendre la relation entre la structure, la dynamique et la fonction d'une protéine. La dynamique interne des protéines est étudiée essentiellement à partir de la relaxation des spins nucléaires ^{15}N ou ^{13}C qui portent un seul proton. L'analyse des mouvements internes de la LTP1_1 de tabac a été centrée sur l'étude de la relaxation du noyau ^{15}N sur un échantillon uniformément marqué en ^{15}N . Ce sont les mouvements du vecteur $^{15}\text{N-H}$ et par conséquent ceux du squelette peptidique qui ont été explorés. Dans la suite de ce chapitre, je décrirai brièvement les principes théoriques qui sont sous-jacents à l'étude de la dynamique interne d'une protéine par relaxation nucléaire. Cette description est inspirée de revues scientifiques qui exposent en détails les théories de l'analyse des mouvements internes des protéines par RMN (Fischer et al. 1998; Korzhnev et al. 2001; Palmer III 2001).

2.1 Mouvements moléculaires des protéines

Les protéines en solution ne sont pas rigides, mais présentent des mouvements moléculaires dus à l'agitation thermique qui les anime. La cinétique de ces mouvements moléculaires s'étale sur une large gamme de temps allant de la femtoseconde à la minute. Les mouvements de faible amplitude nécessitant peu d'énergie sont extrêmement rapides alors qu'inversement, ceux de grandes amplitudes sont plus lents et plus coûteux en énergie. Ce sont les mouvements rapides (de la pico à la nanoseconde) qui peuvent être sondés par l'étude de la relaxation nucléaire. La figure 2.13 illustre l'échelle de temps et la barrière d'énergie des mouvements moléculaires observables par RMN.

2.2 La fonction de densité spectrale

Les mouvements moléculaires provoquent des réorientations du vecteur X-H (X correspondant à ^{15}N ou ^{13}C). Les variations angulaires de ce vecteur sont décrites mathématiquement par une fonction d'autocorrélation $C(t)$ qui représente la probabilité du vecteur X-H de retrouver son orientation originelle après un intervalle de temps donné. Si on considère simplement une protéine comme une sphère rigide adoptant un mouvement brownien isotrope, la fonction d'autocorrélation $C(t)$ peut être associée à une fonction monoexponentielle décroissante (Abragam 1961) :

$$C(t) = \frac{1}{5} \left[e^{\left(\frac{-t}{\tau_c} \right)} \right] \quad (2.14)$$

où τ_c représente le temps de corrélation d'un vecteur X-H. Ce temps correspond à la durée nécessaire au vecteur pour tourner d'un angle de 1 radian.

La transformée de Fourier de la fonction d'autocorrélation $C(t)$ est la fonction de densité spectrale $J(\omega)$ qui fournit une analyse en fréquence des mouvements des vecteurs X-H. Elle caractérise ainsi la « quantité » de mouvements à une fréquence définie pour un vecteur donné. Dans le cas d'une sphère rigide en mouvement brownien, la densité spectrale est décrite par une fonction Lorentzienne :

$$J(\omega) = \frac{2}{5} \left[\frac{\tau_c}{1 + (\omega\tau_c)^2} \right] \quad (2.16)$$

où ω correspond à la fréquence de Larmor du noyau considéré ^{15}N ou ^{13}C .

2.3 La relaxation nucléaire

Lorsque les spins nucléaires sont soumis à un champ magnétique B_0 , ils se répartissent sur différents niveaux d'énergie, le système est alors à l'équilibre thermodynamique. Après une excitation provoquée par un champ magnétique B_1 , le système est déplacé hors de sa position d'équilibre et le retour à cette position se réalise grâce à la relaxation nucléaire. De nombreux mécanismes sont à l'origine de la relaxation nucléaire, les deux plus importants étant l'interaction dipôle-dipôle (DD) et l'anisotropie de déplacement chimique (CSA pour Chemical Shift Anisotropy). Les deux phénomènes se traduisent par l'existence de champs magnétiques locaux autour des noyaux de la protéine. Ces champs magnétiques fluctuent à cause des mouvements globaux et internes de la protéine. Parmi tous ces mouvements de la

protéine certains présentent des fréquences qui correspondent à celles des transitions des spins nucléaires. Plus ils sont nombreux, plus la relaxation est rapide. Les mouvements rapides avec une fréquence moyenne de 100MHz sont à l'origine de l'efficacité de la relaxation du système. Le système retourne à son état d'équilibre rapidement en empruntant les transitions décrites par la figure 2.14

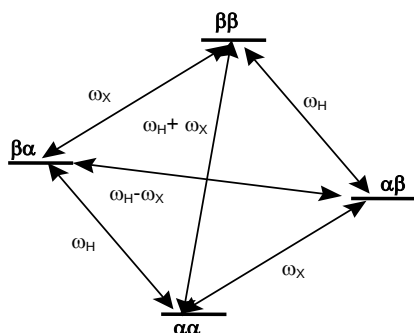


Figure 2.14 : Niveaux d'énergie et transitions correspondantes pour un couple de spins XH. Les transitions simple quantum aux fréquences ω_X et ω_H , zéro quantum à $\omega_H - \omega_X$ et double quanta à $\omega_H + \omega_X$ sont représentées par des doubles flèches.

Les vitesses de relaxation d'un système de spins couplés dépendent donc des valeurs adoptées par la fonction de densité spectrale $J(\omega)$ aux fréquences de transition décrites dans figure 2.14 c'est à dire : $J(\omega_X)$, $J(\omega_H + \omega_X)$, $J(\omega_H)$, $J(\omega_H - \omega_X)$.

2.3.1 La relaxation longitudinale

Liée aux échanges d'énergie entre les spins et leur environnement (appelé réseau : on parle parfois de relaxation spin/réseau), elle décrit le retour à l'équilibre de l'aimantation M_z le long de l'axe z, associé au passage des spins de l'état β à l'état α (et réciproquement). La vitesse de relaxation sera notée $R_X(X_z)$ où X_z désigne l'aimantation longitudinale du noyau X. Cette relaxation est favorisée par des fluctuations de champ magnétique à haute fréquence ($\omega \approx \omega_0$). Seules celles qui fluctuent à des fréquences de l'ordre de ω_0 sont notables dans le repère tournant. Le temps de relaxation longitudinal T_1 est donc minimum pour $\tau_c = 1/\omega_0$ (figure 2.15).

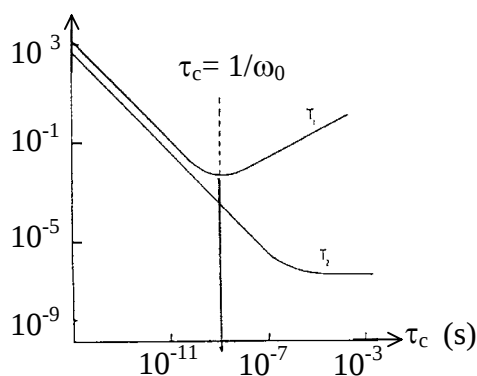


Figure 2.15 : Variations des temps de relaxation T_1 et T_2 en fonction du temps de corrélation τ_c .

2.3.2 La relaxation transversale

Produite par échange d'énergie entre les spins sans interaction avec le réseau, elle décrit le retour à zéro de l'aimantation transversale M_t située dans le plan xy perpendiculairement au champ magnétique B_0 . La vitesse de relaxation est notée $R_X(X_{x,y})$. Elle correspond au déphasage des aimantations nucléaires sous l'effet des fluctuations du champ local à toutes les fréquences : les mouvements « lents » (devant les fréquences de transition RMN) y contribuent cette fois aussi bien que les mouvements haute fréquence. T_2 est inférieur à T_1 puisqu'un plus grand nombre de phénomènes contribuent au T_2 . La courbe décrivant les variations de T_2 en fonction de τ_c est décroissante, indiquant que la relaxation est d'autant plus efficace que les mouvements sont lents (figure 2.15).

2.3.3 La relaxation croisée ; l'effet NOE η

Comme nous l'avons déjà vu, il s'agit d'une interaction dipôle-dipôle transmise à travers l'espace. La relaxation croisée résulte en un gain ou une perte de polarisation du noyau ^{15}N ou ^{13}C suite à la modification de polarisation du proton avec lequel il intéragit. L'expérience donne accès au gain ou à la perte d'aimantation par mesure de l'effet NOE qui dépend de $R_X(X_Z)$ et de la vitesse de relaxation croisée $R_X(H_Z-X_Z)$. Le facteur NOE η s'écrit de la manière suivante :

$$\eta = \frac{\gamma_H}{\gamma_X} \times \frac{R_X(H_Z \leftrightarrow X_Z)}{R_X(X_Z)} \quad (2.16)$$

où γ_H et γ_X sont les rapports gyromagnétiques respectifs du proton et de hétéronoyau ^{15}N ou ^{13}C .

2.3.4 Paramètres expérimentaux mesurés

Dans notre étude, nous avons mesuré pour le spin nucléaire ^{15}N les paramètres de relaxation suivants : la relaxation longitudinale ($R_1=1/T_1$), la relaxation transversale ($R_2=1/T_2$) et le NOE hétéronucléaire ($\text{NOE} = 1+\eta$) à l'état stationnaire. Ces trois paramètres dépendent de la fonction de densité spectrale $J(\omega)$ de la façon suivante (Abragam 1961):

$$R_1 = R_x(X_z) = d^2 \{J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\} + c^2 J(\omega_N) \quad (2.17)$$

$$R_2 = R_x(X_{x,y}) = \frac{d^2}{2} \{4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\} + \frac{c^2}{6} \{3J(\omega_N) + 4J(0)\} + R_{ex} \quad (2.18)$$

$$\text{NOE} = 1 + \left(\frac{\gamma_H}{\gamma_N} \right) \times d^2 \{6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\} \times \frac{1}{R_1} \quad (2.19)$$

Les constantes d^2 et c^2 sont définies de la façon suivante :

$$d^2 = \left(\frac{1}{10} \right) \left(\frac{h\gamma_H\gamma_N}{2\pi r_{NH}^3} \right)^2 \quad (2.20)$$

$$c^2 = \frac{2}{15} \omega_N^2 (\sigma_{//} - \sigma_{\perp})^2 \quad (2.21)$$

Le terme d^2 représente la constante d'interaction dipolaire et c^2 la constante d'anisotropie de déplacement chimique. Les fréquences ω_H , ω_N représentent respectivement les fréquences de résonances des noyaux ^1H et ^{15}N , γ_H et γ_N sont leurs rapports gyromagnétiques avec $\gamma_H = 26,756 \times 10^7 \text{ rad.s}^{-1}.\text{T}^{-1}$ et $\gamma_N = -2,711 \times 10^7 \text{ rad.s}^{-1}.\text{T}^{-1}$, h est la constante de Planck dont la valeur est $6,626 \times 10^{-34} \text{ J.s}$, r_{NH} est la distance interatomique de $1,02 \text{ \AA}$ entre ^1H et ^{15}N et enfin $\sigma_{//}$ et $\sigma_{\perp}$ sont les composantes parallèles et perpendiculaires du tenseur de déplacement chimique de ^{15}N . La quantité $\sigma_{//} - \sigma_{\perp}$ qui constitue l'anisotropie de déplacement chimique de ^{15}N est proche de -170 ppm . Le terme R_{ex} est inclus dans l'équation (2.19) pour tenir compte d'éventuels phénomènes d'échange au sein de la protéine.

2.3.5 Le phénomène d'échange

La présence d'échange chimique pour un résidu particulier se traduit expérimentalement par une valeur de R_2 élevée par rapport à la valeur moyenne mesurée. Le paramètre de relaxation transversale mesuré expérimentalement R_2^{exp} est en fait la somme de la vitesse de relaxation transversale réelle $R_2^{\text{réel}}$ et du terme d'échange R_{ex} :

$$R_2^{\text{exp}} = R_2^{\text{réel}} + R_{\text{ex}} \quad (2.22)$$

$$\text{avec } R_{\text{ex}} = \frac{4\pi^2 \Delta\nu^2 P_A P_B \tau_{\text{ex}}}{1 + \omega^2 \tau_{\text{ex}}^2} \quad (2.23)$$

dans le cas d'un échange chimique entre deux états A et B de populations P_A et P_B dont les fréquences sont ν_A et ν_B . $\Delta\nu$ correspond à la différence de déplacement chimique ($\Delta\nu = \nu_A - \nu_B$). τ_{ex} est égale à l'inverse de la somme des constantes de vitesse d'échange entre ces deux états A et B , k_1 et k_{-1} , $\tau_{\text{ex}} = (k_1 + k_{-1})^{-1}$:



2.4 Détermination expérimentale des paramètres de relaxation du noyau ^{15}N

Les séquences d'impulsions qui ont été utilisées pour mesurer les paramètres de relaxation de la LTP1_1 de tabac sont celles décrites par Farrow et ses collaborateurs en 1994 (Farrow et al. 1994) pour étudier la dynamique interne de l'ubiquitine, protéine modèle des spectroscopistes RMN.

2.4.1 Mesure de la relaxation longitudinale R_1 et transversale R_2

Les séquences d'impulsions pour mesurer les vitesses de relaxation longitudinale R_1 et transversale R_2 sont basées sur la séquence d'impulsions HSQC (figure 2.1b). Cette séquence commence par la série INEPT pour transférer l'aimantation du proton vers le noyau ^{15}N . Ensuite une série d'impulsions amène le système hors de son état d'équilibre, dans l'un des états dont on veut mesurer la relaxation. Suit un délai tr de relaxation, avec une série d'impulsions qui permettent de conserver le simple processus de relaxation en supprimant les effets de relaxation croisée dipôle-dipôle et d'anisotropie de déplacement chimique. Enfin l'aimantation du noyau ^{15}N est traitée afin de transférer sa polarisation vers le proton à travers

une série INEPT-inverse. Une série d'expériences est enregistrée en faisant varier le délai tr afin d'échantillonner au mieux la courbe de relaxation.

On obtient donc une série de spectres de type ^1H - ^{15}N HSQC pour les différentes valeurs de tr . L'intensité des pics croisés décroît de manière exponentielle et tend vers zéro quand les temps de relaxation tendent vers l'infini. Les valeurs de R_1 et R_2 sont déterminées en réalisant un ajustement de l'intensité des pics croisés à l'aide d'une fonction monoexponentielle décroissante :

$$I(tr) = I_0 e^{(-R \times tr)} \quad (2.25)$$

où $I(tr)$ est l'intensité du pic croisé au délai de relaxation tr et I_0 l'intensité au temps $tr = 0$ et R correspond à la constante de vitesse de relaxation R_1 ou R_2 .

L'optimisation des valeurs de I_0 et R est obtenue en minimisant la fonction cible χ^2 dont l'expression est la suivante :

$$\chi^2 = \sum_{i=1}^N \frac{(I(t_R)_{calc} - I(tr)_{exp})^2}{\sigma_I^2} \quad (2.26)$$

où $I(tr)_{calc}$ et $I(tr)_{exp}$ sont respectivement les intensités des pics calculées et mesurées expérimentalement, et σ_I correspond à l'incertitude de mesure sur l'intensité du pic. Plus la valeur du χ^2 est faible plus la correspondance entre le modèle et les valeurs expérimentales est correcte.

Les incertitudes sur les valeurs des paramètres de relaxation R_1 et R_2 sont évaluées par la méthode de Monte Carlo (Press et al. 1986). Les intensités calculées grâce à l'équation 2.25 sont perturbées aléatoirement selon une distribution normale avec un écart type correspondant à celui de l'erreur sur la mesure. Un échantillonnage statistique de 200 vitesses de relaxation par acide aminé a été obtenu et a permis d'aboutir à une vitesse de relaxation moyenne avec un écart type.

2.4.2 Mesure de l'effet NOE

La mesure de l'effet NOE nécessite la réalisation de deux expériences : une avec irradiation des ^1H et l'autre sans. Cette irradiation est assurée par un train d'impulsions qui permet de saturer les protons. Lorsque les protons sont saturés, le système se trouve à l'état stationnaire. Il relaxe alors vers l'équilibre thermique aux transitions zéro quantum et double quanta avec un transfert de polarisation du proton vers le noyau ^{15}N . On obtient deux spectres de type ^1H - ^{15}N HSQC correspondant aux deux expériences avec et sans irradiation. Il faut noter que la sensibilité de mesure de l'effet NOE est moins importante que celle des

paramètres de relaxation R_1 et R_2 car on ne peut bénéficier du premier transfert de polarisation du ^1H vers ^{15}N réalisé par la première série INEPT.

Les valeurs du NOE hétéronucléaire $^1\text{H} - ^{15}\text{N}$ sont déterminées par le rapport suivant :

$$NOE = \frac{I_{irr}}{I_{eq}} \quad (2.27)$$

où I_{irr} et I_{eq} correspondent respectivement aux intensités des pics croisés dans l'expérience avec et sans irradiation du proton. L'incertitude sur la mesure du NOE est estimée à $\pm 10\%$ de l'intensité du pic mesuré.

2.5 Interprétation des paramètres de relaxation en terme de mouvement

De nombreuses approches théoriques ont été réalisées pour modéliser la fonction de densité spectrale à partir des constantes de relaxation R_1 et R_2 et du NOE . Je me contenterai de décrire dans cette partie le modèle le plus couramment utilisé dans l'interprétation des paramètres de relaxation en terme de mouvements : l'approche « model-free » introduite par Lipari et Szabo en 1982 (Lipari and Szabo 1982a; b).

2.5.1 Le modèle de Lipari et Szabo

L'approche de ce modèle dits « à deux paliers » consiste à réduire les mouvements moléculaires d'un vecteur N-H d'une protéine à deux types de mouvements: le mouvement de rotation globale de la molécule et les mouvements internes locaux. On considère que ces deux types de mouvements présentent des échelles de temps très différentes et sont par conséquent indépendants. De plus ce formalisme considère la protéine comme une sphère rigide présentant un mouvement global caractérisé par le temps de corrélation isotrope τ_c . Alors, la fonction d'autocorrélation $C(t)$ qui décrit le mouvement de l'ensemble de la molécule est le résultat de la multiplication de deux fonctions, l'une caractérisant le mouvement global de la protéine $C_c(t)$, et l'autre les mouvements internes $C_i(t)$. Son expression est la suivante :

$$C(t) = C_c(t) \times C_i(t) \quad (2.28)$$

$$\text{où : } C_c(t) = \frac{1}{5} \times e^{\left(\frac{-t}{\tau_c}\right)} \text{ et } C_i(t) = S^2 + (1 - S^2) \times e^{\left(\frac{-t}{\tau_e}\right)} \quad (2.29)$$

où S^2 est le paramètre d'ordre généralisé et représente la valeur limite de $C_i(t)$ quant t tend vers l'infini. Il est interprété comme décrivant la restriction spatiale des mouvements internes

du vecteur N-H (figure 2.16). Il est égal à zéro quand il n'y a pas de restriction de mouvement et prend la valeur de 1 quand au contraire les mouvements internes sont restreints et n'ont pas d'amplitude (alors $C(t) = 1$). Enfin, τ_c et τ_e caractérisent respectivement les temps de corrélation de la molécule et des mouvements internes.

La transformée de Fourier de cette fonction d'autocorrélation est la fonction de densité spectrale $J(\omega)$ suivante. Elle est le résultat de la somme de deux lorentziennes :

$$J(\omega) = \frac{2}{5} \left\{ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S^2) \tau_e}{1 + (\omega \tau_e)^2} \right\} \quad (2.30)$$

$$\text{avec } \tau = \frac{\tau_c \tau_e}{\tau_c + \tau_e} \quad (2.31)$$

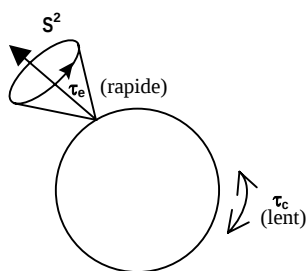


Figure 2.16 : Illustration schématique des paramètres du model-free défini par Lipari-Szabo. Le mouvement global de la protéine est caractérisé par un temps de corrélation τ_c alors que le mouvement interne du vecteur N-H est défini par un temps de corrélation τ_e et par le paramètre S^2 ($\tau_c \gg \tau_e$).

Le modèle de Lipari et Szabo n'est valable que si l'échelle de temps des mouvements internes τ_e (de l'ordre de la picoseconde) est largement inférieure à celle du mouvement global de la protéine τ_c (de l'ordre de la nanoseconde).

La figure 2.17 représente la fonction d'autocorrélation $C(t)$ en considérant τ_c largement supérieure à τ_e .

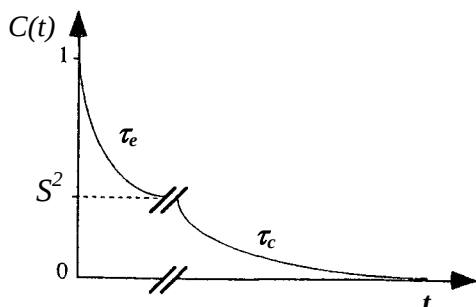


Figure 2.17 : Représentation de la fonction d'autocorrélation $C(t)$ avec τ_c largement supérieure à τ_e

D'autre part, le modèle de Lipari Szabo se simplifie si on considère que les mouvements internes sont très rapides avec τ_e inférieur à quelques picosecondes. L'expression de la fonction de densité spectrale $J(\omega)$ (2.30) devient alors égale à :

$$J(\omega) = \frac{2}{5} \left\{ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} \right\} \quad (2.32)$$

Les mouvements des vecteurs N-H ne sont plus dépendants que du mouvement global de la protéine caractérisé par le temps de corrélation τ_c

En 1990, Clore et ses collaborateurs ont remarqué que le modèle proposé par Lipari et Szabo ne reproduisait pas le comportement dynamique de certains acides aminés de la nucléase de staphylocoque et de l'interleukine-1 β (Clore et al. 1990a; Clore et al. 1990b). C'est alors qu'ils ont proposé un modèle étendu du « model-free » de Lipari Szabo.

2.5.2 Modèle de Lipari Szabo étendu

Ce modèle étendu introduit un deuxième temps de corrélation pour décrire les mouvements internes. On considère donc avec ce nouveau modèle que les mouvements internes sont décrits sur deux échelles de temps distinctes, avec deux temps de corrélation τ_f et τ_s et deux paramètres d'ordre S_f^2 et S_s^2 .

La fonction de densité spectrale $J(\omega)$ caractérisant les mouvements moléculaires de l'ensemble de la molécule devient alors la somme de trois lorentziennes :

$$J(\omega) = \frac{2}{5} \left\{ \frac{S_f^2 S_s^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{S_f^2 (1 - S_s^2) \tau_s'}{1 + (\omega \tau_s')^2} + \frac{(1 - S_f^2) \tau_f'}{1 + (\omega \tau_f')^2} \right\} \quad (2.33)$$

$$\text{avec : } \tau_f' = \frac{\tau_c \tau_f}{\tau_c + \tau_f} \text{ et } \tau_s' = \frac{\tau_c \tau_s}{\tau_c + \tau_s} \quad (2.34)$$

Le paramètre d'ordre S^2 représentant l'amplitude de l'ensemble des mouvements internes devient égal à :

$$S^2 = S_f^2 \times S_s^2 \quad (2.35)$$

Ce modèle étendu de Lipari Szabo décrit donc trois types de mouvements caractérisés par trois temps de corrélation τ_f , τ_s , τ_c avec $\tau_c \gg \tau_s > \tau_f$ et τ_s et τ_f différent d'au moins un

ordre de grandeur. Quand les mouvements internes rapides sont très rapides, on néglige le terme τ_f' . L'expression (2.34) de la fonction de densité spectrale devient alors égale à :

$$J(\omega) = \frac{2}{5} S_f^2 \left\{ \frac{S_s^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S_s^2) \tau_s'}{1 + (\omega \tau_s')^2} \right\} \quad (2.36)$$

où il ne reste plus que quatre paramètres à ajuster : S_f^2 , S_s^2 , τ_s' et τ_c . De plus si les mouvements internes rapides définis par S_f^2 sont restreints, S_f^2 est égale à 1 et S^2 devient alors égal à S_s^2 : on retrouve le modèle de Lipari Szabo et l'expression (2.36) de la fonction de densité spectrale $J(\omega)$ devient égale à celle numérotée (2.32).

Les paramètres S^2 , τ_e et τ_c du modèle de Lipari Szabo ou ceux du modèle étendu S_f^2 , S_s^2 , τ_s et τ_c sont ajustés à l'aide des mesures des données de relaxation R_1 , R_2 , et NOE.

Il est nécessaire de déterminer le temps de corrélation global de la protéine τ_c afin de faciliter l'ajustement des paramètres de relaxation.

2.5.3 Détermination du temps de corrélation globale de la protéine τ_c

Le temps de corrélation globale de la protéine τ_c est classiquement accessible à partir des rapports R_2/R_1 des résidus dont les vecteur N-H présentent une mobilité interne réduite et pas de phénomène d'échange qui conduirait à surestimer la valeur de R_2 (Clore et al. 1990a). Les résidus qui réunissent ces deux conditions sont généralement impliqués dans des structures secondaires. Pour ces résidus, les mouvements internes sont négligeables et le temps de corrélation qui leur est associé τ_e est quasiment nul. Le rapport R_2/R_1 devient indépendant de τ_e et de S^2 et on en déduit le temps de corrélation τ_c (Kay et al. 1989; Clore et al. 1990a).

Le logiciel Modelfree (Palmer III et al. 1991; Mandel et al. 1995) <http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software/modelfree.html>, en interface graphique avec le programme Fast-Modelfree (Cole and Loria 2003) <http://xbeams.chem.yale.edu/~loria/software.htm>, nous a permis de modéliser les mouvements caractéristiques des vecteurs N-H de la LTP1_1 de tabac.

2.5.4 Modelfree et Fast-Modelfree

Le logiciel Modelfree permet de modéliser les mouvements caractéristiques des vecteurs N-H à partir des données de relaxation R_1 , R_2 , et NOE . Ce logiciel utilise cinq modèles qui sont basés sur le formalisme de Lipari Szabo. Ces cinq modèles sont représentés dans le tableau 2.1 et contiennent un nombre de paramètres ajustables différents.

Modèles	Hypothèses	Paramètres ajustables	Degré de liberté
1. Lipari-Szabo simplifié	τ_e très court < 20ps est négligé	S^2	2
2. Lipari-Szabo	τ_e est pris en compte dans le modèle	S^2, τ_e	1
3. Lipari-Szabo simplifié avec terme d'échange	modèle 1 avec terme d'échange	S^2, R_{ex}	1
4. Lipari-Szabo, avec terme d'échange	modèle 2 avec terme d'échange	S^2, τ_e et R_{ex}	0
5. Lipari-Szabo étendu	les mouvements internes sont de deux types et on néglige les temps de corrélations les plus courts avec $\tau_f = 0$	S_f^2, S_s^2 et $\tau_s = \tau_e$	0

Tableau 2.1 : Présentation des cinq modèles utilisés dans Modelfree pour modéliser les mouvements caractéristiques des vecteurs N-H.

2.5.5 Choix des modèles

Le choix des modèles utilisés pour modéliser les mouvements caractéristiques des vecteurs N-H de la LTP1_1 de tabac a été réalisé à l'aide d'une évaluation statistique utilisée par Mandel et ses collaborateurs en 1995 pour décrire la dynamique interne de la ribonuclease HI d'*E. coli* (Mandel et al. 1995). L'objectif de cette analyse statistique est de faire correspondre au mieux à chaque vecteur N-H l'un des cinq modèles précédemment présentés, afin de décrire au mieux sa dynamique.

Cette procédure statistique consiste à tester successivement chacun des modèles pour chaque vecteur N-H en ajustant leurs paramètres de relaxation. Pour tester la validité du modèle à reproduire les paramètres de relaxation, on détermine la fonction cible χ^2 suivante :

$$\chi^2 = \frac{(R_1^{mes} - R_1^{calc})^2}{\sigma_{R_1}^2} + \frac{(R_2^{mes} - R_2^{calc})^2}{\sigma_{R_2}^2} + \frac{(NOE^{mes} - NOE^{calc})^2}{\sigma_{NOE}^2} \quad (2.38)$$

où les σ correspondent aux incertitudes sur les mesures des paramètres de relaxation.

Les modèles sélectionnés doivent reproduire les données expérimentales à $\pm 5\%$, et posséder un faible χ^2 traduisant ainsi leur validité. Comme nous l'avons vu, les modèles possèdent un nombre de paramètres ajustables différents. Pour trancher entre deux modèles i

et j et déterminer si l'ajustement avec un paramètre supplémentaire est statistiquement justifié, on utilise le test de Fischer-Snedecor (test - F). La valeur du test-F est calculée de la manière suivante :

$$test - F = \frac{\nu_i (\chi_j^2 - \chi_i^2)}{(\nu_j - \nu_i) \chi_i^2} \quad (2.39)$$

où χ_j^2, χ_i^2 sont respectivement les valeurs de la fonction cible χ^2 pour les modèles j et i et ν_j et ν_i correspondent respectivement aux degrés de liberté des modèles j et i. Le nombre de degré de liberté ν de chacun des modèles est égal à la différence du nombre de données expérimentales mesurées (N=3) et du nombre de paramètres à ajuster pour chaque modèle (n). Par exemple pour le modèle 1, le nombre de paramètre à ajuster est égal à 1 donc nous avons deux degrés de liberté pour ce modèle (tableau2.1).

D'après les valeurs χ^2 calculées pour chaque modèle, on en déduit une valeur test-F que l'on compare à la valeur théorique test-F_{th} donnée par la table de Fisher-Snedecor. Si test-F est supérieur à test-F_{th}, le modèle possédant un degré de liberté plus petit est choisi. L'utilisation du logiciel Fast-Modelfree permet l'automatisation du choix des modèles.

Remarque : les modèles 4 et 5 possèdent le même nombre de paramètres expérimentaux et ajustables, par conséquent leur nombre de degrés de liberté est nul. On ne peut donc pas calculer pour ces modèles la valeur test-F. Ces modèles 4 et 5 sont alors acceptés si leur fonction cible χ^2 est voisine de zéro.

3. Simulations de dynamique moléculaire

Pendant les trente dernières années, les expériences numériques de simulations de dynamique moléculaire par ordinateur ont remarquablement contribué à la compréhension des propriétés de la matière dans l'état liquide. Plus récemment, l'évolution des moyens de calcul a permis l'étude de systèmes plus complexes, notamment les macromolécules biologiques. La méthode de simulation numérique dite de dynamique moléculaire permet de simuler le comportement macroscopique de la matière dans l'état solide, liquide ou gazeux à partir de la description microscopique d'un système de particules soumises à un champ de force classique bien défini.

3.1 La fonction d'énergie potentielle

Les informations sur les interactions entre atomes sont contenues dans la fonction d'énergie potentielle, qui définit donc le modèle étudié. Les variables de l'énergie potentielle sont les coordonnées nucléaires. La forme fonctionnelle de l'énergie potentielle est représentée *a priori* de façon empirique. Ses variables dépendent de paramètres qui doivent être ajustés de manière à reproduire les propriétés énergétiques d'un système moléculaire particulier, ce qui constitue le champ de force.

Pour un système atomique, l'énergie potentielle est souvent divisée en termes qui caractérisent les atomes individuellement, les interactions entre les paires d'atomes, triplets, etc. Dans la simulation des liquides on considère d'habitude seulement les termes à deux corps en incluant dans ceux-ci des corrections pour tenir compte en moyenne des termes d'ordre supérieur.

Une expression classique de l'énergie potentielle d'une protéine isolée est celle proposée par Brooks et ses collaborateurs en 1983 (Brooks 1983) et utilisée dans le programme de modélisation CHARMM (Chemistry at HARvard Macromolecular Mechanics).

Celle-ci a la forme suivante :

$$\begin{aligned}
 Ep(r) = & \sum_{\text{liaisons}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{impropres}} k_\omega (\omega - \omega_0)^2 \\
 & + \sum_{\text{dièdres}} \sum_{n=1}^{\infty} k_{\phi,n} [1 + \cos(n\phi - \delta)] + \sum_{ij} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q_i q_j}{\varepsilon_{ij}} \quad (2.39)
 \end{aligned}$$

où r désigne l'ensemble des positions des noyaux atomiques. On distingue dans cette fonction les quatre premiers termes, appelés *termes liés*, qui sont directement en relation avec les liaisons covalentes entre les atomes, et les deux autres termes, appelés *termes non-liés*, qui concernent les interactions entre les atomes non-liés de manière covalente.

Les trois premiers termes rendent compte du coût en énergie de l'élongation des liaisons covalentes (r), des fluctuations de l'angle formé par deux liaisons covalentes d'un atome (θ) ou encore de l'angle dièdre impropre formé par trois liaisons (ω) et maintient la planarité de certains groupements ou la chiralité des C_α . Les valeurs de référence dérivent de géométries moléculaires expérimentales obtenues par diffraction de rayons X ou de neutrons, ou de calculs de mécanique quantique *ab-initio*. Les constantes de forces k_r , k_θ et k_ω sont ajustées pour reproduire les fréquences de vibrations de molécules en phase gazeuse. Le quatrième terme est un potentiel de torsion sur les angles dièdres (ϕ).

Le cinquième terme de la fonction d'énergie est un terme de van der Waals. Ce potentiel a été introduit historiquement, par London pour expliquer la condensation des gaz rares. L'énergie d'attraction de deux atomes identiques de gaz rare s'explique par une fluctuation de leurs nuages électroniques respectifs, qui conduit à une interaction du type dipôle induit/ dipôle induit. Il en résulte un terme attractif de la forme $-A/r^6$ appelé interaction de dispersion, où r est la distance séparant les deux noyaux et A est une constante qui dépend de la nature chimique des atomes. La partie répulsive du potentiel traduit le principe de Pauli. On exprime souvent le terme qui en résulte sous la forme B/r^{12} pour représenter la répulsion des nuages électroniques à courte distance.

L'énergie de van der Waals s'exprime alors par le potentiel de Lennard-Jones qui a pour expression :

$$\begin{aligned}
 V_{LJ}(r) &= \frac{B}{r^{12}} - \frac{A}{r^6} \\
 &= 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (2.40)
 \end{aligned}$$

Il comporte un puits de potentiel dont la profondeur est donnée par ε et dont le minimum se trouve en $r_0 = 2^{1/6}\sigma$. La valeur $r_0/2$ est le rayon de van der Waals de l'espèce chimique considérée.

Jusqu'ici nous avons considéré des interactions entre atomes de même nature ; lorsque les atomes sont de nature différente on utilise les règles de combinaison suivantes $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ et $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$, où les indices i et j font référence aux espèces chimiques i et j .

Le dernier terme de la fonction d'énergie (2.39) est le terme d'interaction électrostatique. Dans cette fonction, le potentiel électrostatique est généré dans l'approximation de charges ponctuelles portées par les atomes. Les termes de polarisation sont négligés.

Les termes électrostatiques et de van der Waals comportent chacun une sommation sur $\frac{1}{2} N (N-1)$ paires d'atomes. S'il y a plusieurs milliers d'atomes, le calcul peut devenir très long. On a alors recours à une approximation qui permet de diminuer notablement le temps de calcul. Cette approximation consiste à ne tenir compte que des interactions non-liées d'un atome donné avec ses plus proches voisins situés dans une sphère, dont le rayon r_c est appelé rayon de coupure. Pour éviter les instabilités, on multiplie la fonction d'énergie potentielle par une fonction "switch" permettant de faire tendre lentement la fonction énergie potentielle vers 0 entre r_{on} et r_c .

$$sw(r) = \begin{cases} 1 & \text{pour } r < r_{on} \\ \frac{(r_c - r)^2(r_c + 2r - 3r_{on})}{(r_c - r_{on})^3} & \text{pour } r_{on} \leq r < r_c \\ 0 & \text{pour } r \geq r_c \end{cases} \quad (2.41)$$

La plupart des programmes de simulation (CHARMM, AMBER, GROMOS etc...) utilisent une fonction d'énergie potentielle semblable, mais pas toujours rigoureusement identique à celle donnée en (2.39). Elles diffèrent principalement par les paramètres du champ de force.

3.2 Les simulations de dynamiques moléculaires

La dynamique moléculaire permet de simuler l'évolution du système à une température donnée sur un intervalle de quelques nanosecondes. Elle permet le franchissement de barrières énergétiques grâce au terme d'énergie cinétique. Les moyens techniques actuellement

disponibles permettent d'étudier des systèmes où le nombre N d'atomes (ou centres d'interaction) est de l'ordre de 20000. Le système est généralement confiné dans une région définie de l'espace, de volume V . Des conditions périodiques imposées au bord de ce volume, ou cellule de simulation, permettent de minimiser les effets de bord.

3.2.1 Résolution des équations du mouvement

La dynamique moléculaire standard est basée sur la résolution de l'équation fondamentale de la dynamique (2.42). Les forces étant reliées à la fonction d'énergie potentielle (2.43) connaissant la position à l'instant t , il est possible, par application de développements de Taylor, d'obtenir les positions $\vec{r}(t + \delta t)$ et les vitesses $\vec{v}(t + \delta t)$ des différents atomes.

$$\vec{F} = m \vec{a} \quad (2.42)$$

$$\vec{F} = -\vec{\nabla} E_p \quad \Rightarrow \quad \vec{a} = -\frac{\vec{\nabla} E_p}{m} \quad (2.43)$$

$$\vec{r}(t + \delta t) = 2\vec{r}(t) - \vec{r}(t - \delta t) + \delta t^2 \vec{a}(t) \quad (2.44)$$

$$\vec{v}(t) = \left[\vec{r}(t + \delta t) - \vec{r}(t - \delta t) \right] / 2\delta t \quad (2.45)$$

Cet algorithme introduit par Verlet (Verlet 1967) est aussi simple que stable et précis. Le pas d'intégration δt doit être petit devant le temps caractéristique le plus court dans un système moléculaire. Dans le cas d'une protéine, ce dernier correspond à la période de vibration des liaisons, soit 10 fs. Cela rend nécessaire l'utilisation d'un pas d'intégration δt de l'ordre de 0,5 fs (1fs = 10^{-15} s). On peut passer à 2 fs si on s'affranchit des vibrations de liaisons peu intéressantes en utilisant l'option SHAKE (Ryckaert et al. 1977).

3.2.2 Les conditions initiales

Une simulation de dynamique moléculaire s'effectue à partir d'une structure initiale, en attribuant aux atomes des vitesses initiales. Les vitesses sont généralement échantillonnées à partir d'une distribution de Maxwell dont la variance est liée à la température à laquelle on veut effectuer la simulation.

Le choix des positions initiales est très important dans la simulation des protéines. En raison de la complexité de l'espace conformationnel accessible à une chaîne polypeptidique et de l'échelle de temps très longue impliquée dans la relaxation conformationnelle, en initiant la

simulation à partir d'une conformation aléatoire et dépliée, on a peu d'espoir d'arriver à une conformation d'équilibre.

Les simulations des protéines ne peuvent pas être effectuées sans la connaissance des positions des atomes obtenues expérimentalement par diffraction de rayons X et de neutrons, ou par résonance magnétique nucléaire (RMN). Pour la LTP1_1 de tabac, les structures 3D initiales sont celles que nous avons déterminées expérimentalement par RMN et modélisation moléculaire.

3.2.3 Dynamique moléculaire à température constante

Lors de la simulation de systèmes d'intérêt biologique tels que les protéines, suivant l'ensemble thermodynamique dans lequel on se place, il est parfois nécessaire de travailler à température constante. Plusieurs méthodes existent pour maintenir la température, le principe étant de réajuster les vitesses des différents atomes. Cela peut se faire grâce à la relation (2.46) qui relie la vitesse à la température par l'intermédiaire de l'énergie cinétique :

$$Ec = \sum \frac{1}{2}mv^2 = \frac{3}{2}k_bNT \quad (2.46)$$

où k_b est la constante de Boltzman et N le nombre de degrés de liberté du système.

La méthode la plus couramment utilisée, est celle introduite par Berendsen et ses collaborateurs (Berendsen et al. 1984). Dans cette méthode, la température est maintenue à une valeur déterminée, en couplant le système étudié avec un bain externe maintenu à la température désirée (T_{bain}). A chaque étape les vitesses sont réajustées suivant la relation suivante :

$$v_{corrigée} = \lambda v \quad \text{avec} \quad \lambda^2 = 1 + \frac{\delta t}{\tau} \left(\frac{T_{bain}}{T(t)} - 1 \right) \quad (2.47)$$

où τ est le paramètre de couplage, si τ est grand, le couplage est considéré comme faible entre le bain et le système. Dans le cas où τ est petit, le couplage est fort. L'avantage de cette approche est qu'elle permet à la température du système de fluctuer au voisinage de la température désirée.

3.2.4 Dynamique moléculaire à pression constante

Selon un principe tout à fait similaire, il est également possible de réaliser des simulations de dynamique moléculaire à pression constante. En effet, comme de nombreuses mesures expérimentales sont réalisées dans des conditions de température et de pression constantes, il est préférable de réaliser les simulations dans les mêmes conditions isothermes et isobares. Afin de réguler la pression, on simule l'existence d'un piston couplé au système, qui autorise des variations des dimensions du volume d'étude suivant une ou plusieurs directions.

3.3 Utilisation du logiciel CHARMM

CHARMM (Chemistry at HARvard Macromolecular Mechanics) est un logiciel qui utilise des fonctions d'énergie empiriques définies précédemment afin de modéliser des systèmes macromoléculaires. Le programme peut lire et construire des modèles de structures, minimiser leur énergie, accomplir des simulations de dynamique moléculaire.

CHARMM permet également d'analyser la structure et de déterminer les propriétés dynamiques des systèmes étudiés.

3.3.1 Mise en œuvre des simulations

Nous avons utilisé comme structure 3D de départ de la LTP1_1 de tabac celle qui présentait l'énergie de contrainte la plus faible. Nous avons procédé alors à une étude de dynamique moléculaire. Nous avons tout d'abord réalisé une simulation à 293K sur une période de 5 ns afin de vérifier la stabilité de la structure expérimentale dans le champ de force de CHARMM, parfaitement adapté au cas des protéines. Le processus de dépliement est relativement lent, mais on peut artificiellement accélérer le système en travaillant à haute température dans une limite raisonnable pour les champs de force. Ceci permet de simuler par exemple le dépliement d'une protéine. Nous avons donc ensuite réalisé deux trajectoires de simulation à 500K sur une période de 3 ns. Les simulations ont été effectuées dans l'ensemble NPT : nombre de particules, pression et température constantes.

Afin de modéliser au mieux la protéine dans de bonnes conditions, la protéine est introduite dans une boîte d'eau de dimensions 55x55x60 (Å), générée en utilisant le modèle TIP3 pour l'eau. Une étape de minimisation énergétique a été réalisée avant l'introduction de la protéine dans la boîte contenant environ 6000 molécules d'eau. Les dimensions de la boîte

ont été déterminées de façon à ce que la protéine se trouve au centre et à une distance convenable des parois de la boîte.

Le protocole de dynamique moléculaire à 293K a consisté à simuler les fluctuations des coordonnées de la protéine pendant une période de 5ns, avec un pas d'intégration fixé à 2 fs. Les coordonnées des atomes ont été collectées après chaque picoseconde de dynamique moléculaire.

La température du système a ensuite été élevée à 500K par paliers successifs de 50K sur des périodes de 10 ps. Une fois la température de 500K atteinte, deux trajectoires de simulation de dynamique moléculaire à haute température ont été réalisées durant une période de 3 ns, à partir de deux structures initiales différentes, selon le même principe que la dynamique à 293K.

Nous avons réalisé l'étude structurale, par RMN et modélisation de la LTP1_1 de tabac, qui s'exprime dans l'épiderme des jeunes feuilles mais quatre autres LTP1 ont été isolées dans d'autres organes de la plante. Afin d'identifier les caractéristiques structurales de ces quatre LTP1 nous avons entrepris de déterminer leurs structures 3D par modélisation comparative. Dans la suite de cette partie, je vais donc introduire le principe de modélisation comparative et présenter le logiciel Modeller <http://salilab.org/modeller/modeller.html> (Sali and Blundell 1993) qui a été utilisé pour réaliser les modèles 3D de la famille des LTP1 du tabac *Nicotiana tabacum*.

4. La modélisation comparative

Plus d'un million de séquences protéiques uniques sont actuellement annotées dans les banques de protéines SWISSPROT et TrEMBL (Boeckmann et al. 2003) alors que seulement 25000 structures 3D de protéines sont déposées dans la PDB (pour Protein Data Bank) (Berman et al. 2000). On estime que plus d'un tiers des séquences de protéines connues possèdent un repliement déjà déterminé expérimentalement (Marti-Renom et al. 2000). En effet, le nombre de repliements existants dans la nature est limité par rapport au nombre de séquences apparues au cours de l'évolution (Chothia 1992) et deux protéines ayant un taux d'identité de séquence suffisant (supérieur à 30 %) sont susceptibles d'adopter le même repliement. On observe parfois même des similarités de structures importantes entre deux protéines qui possèdent un taux d'identité de séquence très faible (inférieur à 10%). Il est donc envisageable de construire par modélisation comparative la structure d'une protéine en utilisant celle d'une protéine de séquence voisine comme modèle et ainsi de combler (en partie) l'écart entre le nombre de structures et de séquences connues.

Les modèles ainsi obtenus apportent des informations globales sur la structure des protéines, servant soit de travail préliminaire à une étude structurale expérimentales soit de support pour l'expérimentation biologique comme nous l'avons décrit dans l'article concernant l' α -GalNac transférase où des mutations suggérées à partir du modèle ont été validées par l'expérience (cf annexe).

Le but de la modélisation comparative est de déterminer la structure tridimensionnelle d'une protéine (la cible) à partir de l'alignement de sa séquence avec celle d'une ou plusieurs protéines de structure 3D connues (le(s) modèle(s)) (figure 2.18).

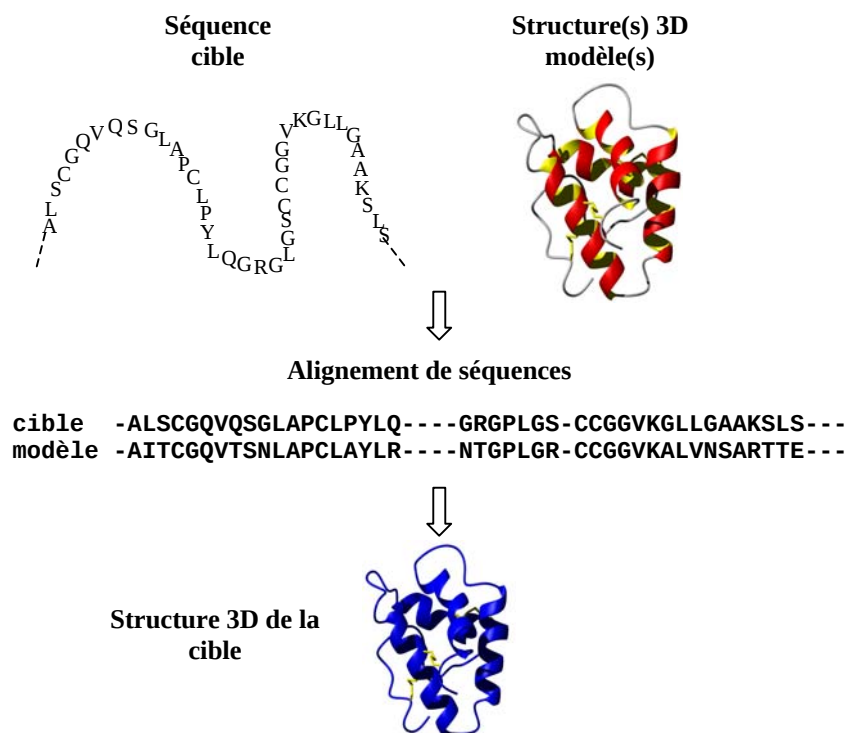


Figure 2.18 : Principe de la modélisation comparative.

La modélisation comparative se déroule en cinq étapes successives : recherche des modèles, choix des modèles, alignement des séquences, construction des structures 3D et évaluation des structures.

4.1 Recherche et choix des modèles

La modélisation comparative débute par la recherche des structures 3D modèles dans les banques de données en utilisant la séquence de la protéine cible. Pour ce faire, on recherche des alignements locaux entre la séquence cible et les banques de données. Les algorithmes de comparaison de séquences les plus couramment utilisés, BLAST (pour Basic Local Alignment Search Tool) (Altschul et al. 1990) et FASTA (Pearson and Lipman 1988), permettent d'interroger les banques de données comme la PDB et d'identifier des séquences de protéines qui sont similaires à celle de la protéine cible. Pour trouver les séquences les plus semblables à celle de la protéine cible, ces algorithmes se réfèrent à des matrices de score. Ces dernières servent à quantifier le degré de similitude entre deux séquences et permettent d'évaluer à quel point l'alignement est significatif. Les matrices les plus utilisées sont les matrices de type BLOSUM (pour BLOck SUBstitution Matrix) (Henikoff and Henikoff 1992) et PAM (pour Percent Accepted Mutation) (Schwartz and Dayhoff 1978).

D'autres algorithmes permettent d'identifier des séquences similaires dans les banques de données, comme par exemple la version améliorée de BLAST, PSI-BLAST (pour Position-Specific-Iterated BLAST) (Altschul et al. 1997). Cet algorithme fonctionne de la manière suivante (figure 2.19) : à partir de la séquence cible et des homologues proches trouvés par l'algorithme BLAST, on établit un alignement multiple d'où l'on extrait un profil, c'est-à-dire que l'on identifie les acides aminés les plus conservés parmi toutes les séquences. On recherche alors ce profil dans les banques de données pour identifier une famille de séquences qui le contiennent. On affine ensuite l'alignement multiple et on extrait un nouveau profil ; ce processus itératif profil/alignement multiple est poursuivi pendant quelques cycles jusqu'à ne plus faire apparaître de nouvelle séquence et avoir ainsi identifié tous les membres de la famille.

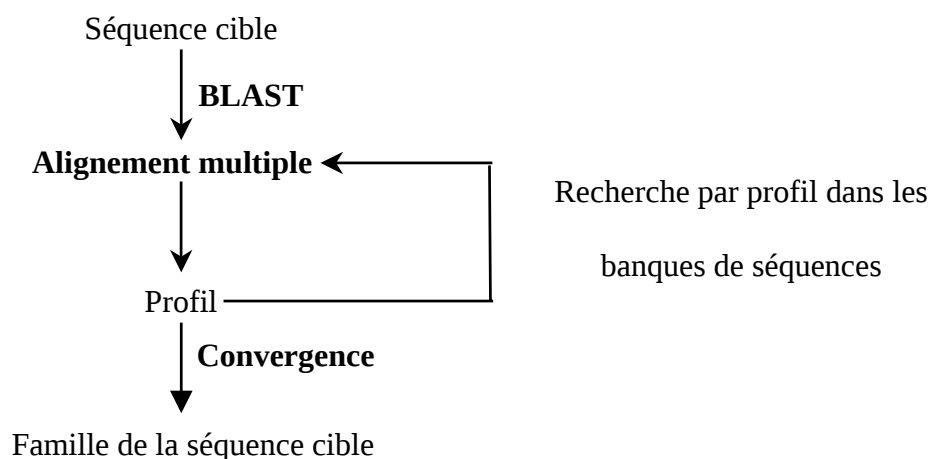


Figure 2.19 : Principe de fonctionnement de l'algorithme PSI-BLAST

Parmi ceux-ci, les protéines de structure 3D connue et possédant la similarité de séquence la plus importante avec la cible sont sélectionnées comme modèles pour prédire la structure 3D de la cible. D'autres critères peuvent guider le choix du modèle comme, par exemple, sa fonction biologique (protéines de même famille que la cible) ou encore la résolution et la qualité de sa structure 3D.

4.2 Alignement des séquences

Cette étape est déterminante dans le déroulement de la modélisation comparative. En effet, un mauvais alignement entraîne systématiquement un modèle 3D erroné pour la cible. Comme nous l'avons précédemment vu, la recherche des séquences similaires à celle de la séquence cible engendre un alignement de séquences. Toutefois, cet alignement n'est que

local et ne concerne que les acides aminés similaires de la cible et du modèle, puisque les algorithmes de recherche de séquences sont optimisés dans la recherche de similarités. Il reste donc à réaliser l'alignement complet.

En général, il est préférable de réaliser des alignements multiples afin d'éviter des erreurs éventuelles. Par exemple, si les séquences de la cible et du modèle ne partagent que 30 % d'identité, il est très difficile de réaliser directement l'alignement complet des deux séquences. On utilise donc si elles existent, des séquences intermédiaires possédant des identités de séquences supérieures à 30 % avec la cible et le modèle (mais de structure 3D inconnue), qui permettent d'optimiser l'alignement. (Figure 2.20).

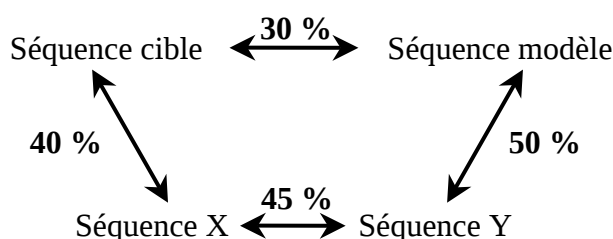


Figure 2.20 : Principe de l'alignement multiple. Les identités de séquences sont indiquées en (%)

Le programme qui est généralement utilisé pour réaliser des alignements multiples est le logiciel CLUSTALW (Thompson et al. 1994). Ce dernier réalise l'alignement multiple de séquences en commençant par les paires les plus identiques. Il faut néanmoins vérifier manuellement l'alignement multiple car une erreur d'alignement sur une paire peut se répercuter sur l'ensemble de l'alignement. L'alignement initial de la séquence de la cible et du modèle a été obtenu à l'aide du logiciel CLUSTALW pour la modélisation des 4 LTP1 du tabac. Un autre logiciel permet également de réaliser l'alignement de la séquence de la cible avec le modèle, il s'agit de la méthode HCA (pour Hydrophobic Cluster Analysis) développée par Jean Paul Mornon et son équipe (Gaboriaud et al. 1987). Cette méthode permet d'aligner des séquences de protéines pour de très basses identités (en dessous de 25-30 %) là où les méthodes d'alignement multiple comme CLUSTALW sont inefficaces. Cette méthode permet d'identifier les structures secondaires présentes au sein de la séquence cible. En effet, les amas hydrophobes des tracés bidimensionnels qu'elle utilise pour représenter la séquence d'une protéine correspondent généralement aux faces internes de ses structures secondaires (Woodcock et al. 1992). C'est le logiciel que nous avons employé au cours du travail de l' α -GalNac transférase (cf. Annexe).

4.3 Construction du modèle

Une fois l'étape cruciale d'alignement de séquence de la cible et du modèle réalisée, il existe de nombreuses méthodes pour construire le modèle tridimensionnel de la cible (Martini-Renom et al. 2000). Je vais vous présenter seulement l'une d'entre elle, celle que nous avons utilisée dans notre étude : la modélisation à partir de contraintes géométriques avec le logiciel MODELLER. Cette technique s'inspire largement de la modélisation moléculaire à partir de contraintes RMN. La modélisation comporte trois étapes : l'alignement de la séquence cible avec celles des modèles (l'alignement initial est celui obtenu à partir des logiciels CLUSTALW ou HCA), la génération des fichiers de contraintes et enfin le calcul sous contraintes du modèle tridimensionnel de la cible par des méthodes recuit simulé et de minimisation énergétique (Figure 2.21).

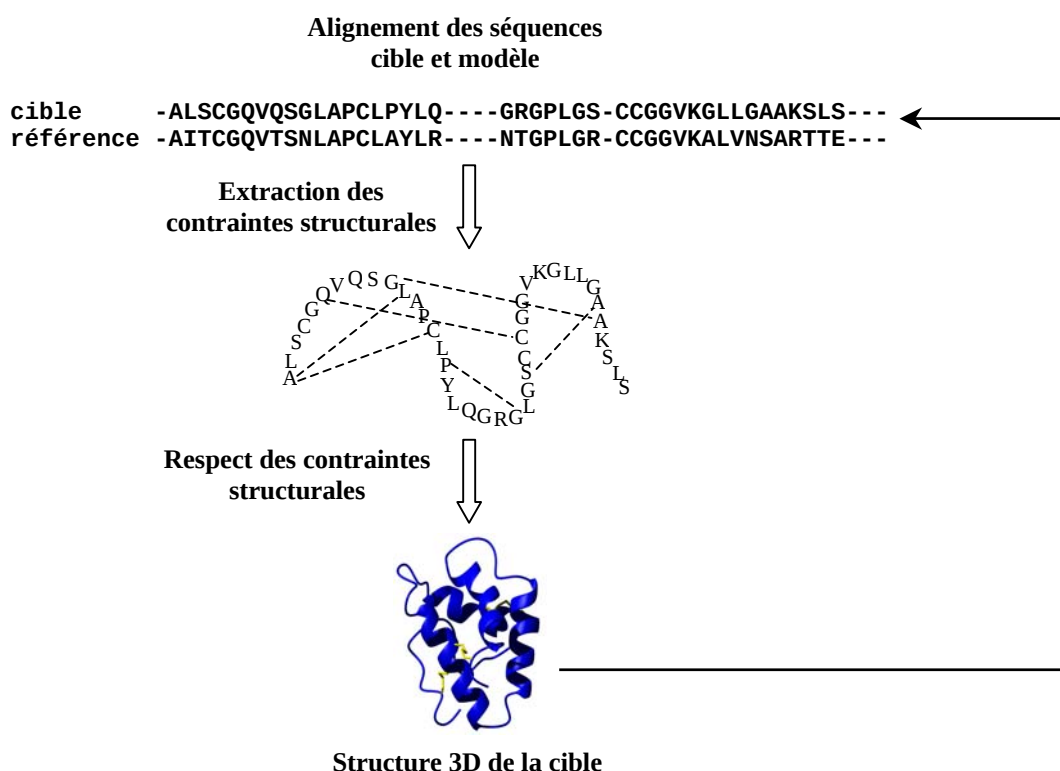


Figure 2.21 : Principe de fonctionnement du logiciel MODELLER

MODELLER utilise des contraintes de distance et d'angle dièdre générées à partir de l'alignement des séquences et d'une base de données structurales incluant 78495 résidus de 405 protéines appartenant à 17 familles. En effet, on considère que les valeurs des distances et des angles dièdres mesurés sur les résidus alignés de la séquence cible et du modèle sont du même ordre. Ces contraintes structurales, ainsi que des contraintes stéréochimiques sont utilisées pour construire le modèle de la cible, à l'image de la modélisation moléculaire à partir des données RMN que nous avons décrite précédemment.

La mise en œuvre du programme suit un processus itératif. En effet, les structures 3D prédites pour la cible sont utilisées pour faire un alignement structural avec le(s) modèle(s) qui permet d'améliorer l'alignement de séquence. De nouvelles contraintes spatiales sont alors générées et on réalise de nouveaux calculs jusqu'à ce que l'alignement ne soit plus modifié.

4.4 Evaluation du modèle

Après avoir déterminé le modèle 3D de la cible, il est nécessaire d'évaluer la qualité de son repliement. Nous avons utilisé les logiciels VERIFY 3D (http://shannon.mbi.ucla.edu/DOE/Services/Verify_3D/) (Bowie et al. 1991) et PROSAIL (Sippl 1993). Ces derniers vérifient si l'environnement d'un type d'acide aminé donné est conforme à celui prédit par une analyse de base de données structurales.

Nous avons également vérifié la stéréochimie des modèles à l'aide du logiciel PROCHECK (Laskowski et al. 1996).

D'autre part, nous avons visualisé chacun des modèles pour détecter d'éventuelles erreurs telles que des défauts d'empilement ou une mauvaise orientation des chaînes hydrophobes et hydrophiles.

En conclusion, la réalisation d'un bon modèle structural dépend essentiellement de la qualité de l'alignement et celui-ci est facilité si le taux d'identité de séquence est importante (supérieure à 50 %). En dessous de ce pourcentage d'identité il faut toujours vérifier manuellement l'alignement proposé par les différents logiciels. Au cours de cette thèse j'ai été confronté aux deux cas de figures. Pour la modélisation des LTP1 de tabac, nous étions dans la gamme d'identité de séquence permettant d'obtenir un modèle fiable (entre 36 et 72 %). Pour l'étude de l' α -GalNac transférase nous n'étions plus dans la gamme « de sécurité » d'identité de la modélisation comparative (inférieure à 30 %). Néanmoins malgré la faible identité de séquence entre le modèle et la cible nous avons réussi à prédire un modèle structural qui a été confirmé par l'expérience. En effet, certaines zones de la protéine sont mieux conservées que d'autres au cours de l'évolution, comme par exemple son cœur hydrophobe et, dans notre cas le site actif de l'enzyme.

III) Etude structurale des LTP1 de tabac

Comme nous l'avons vu précédemment, les LTPs constituent une famille de protéines multigénique dont la fonction n'a pas encore été clairement identifiée et notre base de données structurales est restreinte à la seule connaissance de structures 3D de LTPs extraites de graines de céréales. Pour essayer de comprendre la fonction de ces protéines, nous nous sommes donc intéressés à des LTPs d'origine différente et plus particulièrement aux 5 LTP1 identifiées dans le tabac *Nicotiana tabacum* (figure 3.1). Dans la suite de ce chapitre, je vais présenter brièvement ces LTP1 de tabac, l'étude structurale qui a été réalisée par RMN sur l'une d'entre elles, la LTP1_1, et enfin la modélisation comparative qui a été effectuée sur les 4 autres LTP1 de la famille.



Figure 3.1 : La plante de tabac, *Nicotiana tabacum*.
Ses feuilles sont utilisées pour fabriquer le tabac qui est communément fumé.

1. La famille des LTP1 de *Nicotiana tabacum*

	10	20	30	40	50	60	
LTP1_1	-AITC	GGVTSNLA	PCLAYLR---	NTGPLGR-	CCGGVKALVNSARTTEDRQIA	CTCLKSAAGAI	SGINLG
LTP1_2	-ALSC	GGVQSGLA	PCLPYLQ---	GRGPLGS-	CCGGVKGLLGA	AKSPDRKTA	CTCLKSAANA
LTP1_3	APPSC	PVTTLQLA	PCLSYIQ---	GGGDP	SVPCCTGINNI	YELAKTKEDRVAI	CNCLKTA
LTP1_4	APPSC	QVTTLQLA	PCLSYIQNRVKGGN	PSVPCCTGINNI	YELAKTKEDRVAI	CNCLKNA	FIHAGNVNPT
LTP1_5	-AISL	CGVVTRLSP	CINYVR---	QGGAL	PAPCCSGIKALNNQATTT	PDRQAAC	CNCIKSAAASISGINFN
	70	80	90		pI		organes
LTP1_1	KAAGLP	STCGVNI	PYK-IS	PSTD	CSKVQ--	91	organes végétatifs
LTP1_2	KAAGLP	GACGVNI	PCK-IS	PSTD	CSKVQ--	91	fleurs
LTP1_3	LVAQL	PKKCGIS	FNMPPIDKN	YDCNTISM	Y	96	anthères
LTP1_4	LVAEL	PKKCGIS	FNMPPIDKN	YDCNTISM	Y	100	anthères
LTP1_5	LAGSL	PSKCGVNL	PYK-IS	PSID	CSTVQ--	92	organes végétatifs

Figure 3.2 : Séquences des 5 LTP1 isolées dans le tabac dans différents organes. Les cystéines sont indiquées en rouge, les aromatiques et les proline en orange.

Actuellement, 5 LTPs ont été isolées dans le tabac dans des organes différents (figure 3.2). Elles présentent les caractéristiques de la famille des LTP1, c'est-à-dire qu'elles possèdent le réseau de cystéines qui est la signature avec un résidu de nature hydrophile au sein du motif CXC. Comme la plupart des protéines de cette famille, elles sont basiques, avec

un point isoélectrique compris entre 7,7 et 9,1 (figure 3.2). On peut remarquer que les deux LTP1 (3 et 4) exprimées dans les anthères (organes de reproduction) présentent une extension C-terminale de deux résidus (Met, Tyr). En outre, ces deux protéines comportent dans leur séquence un nombre important de résidus aromatiques (6 au lieu de 2 dans la plupart des LTP1) et un réseau de prolines notablement différent de celui observé pour les autres LTP1 de tabac, en particulier dans le segment N-terminal.

2. Etude structurale de la LTP1_1 de tabac

La LTP1_1 est exprimée de manière constitutive dans les organes végétatifs de la plante *Nicotiana tabacum*. Elle est plus particulièrement présente dans l'épiderme des jeunes feuilles (Fleming et al. 1992). Des études préliminaires sur son rôle biologique ont révélé une activité élicitrice, c'est à dire qu'elle est capable d'induire une nécrose sur la plante de tabac, à l'instar des élicitines, d'où notre intérêt pour son étude structurale par RMN et modélisation moléculaire.

C'est une protéine de 91 résidus, de masse moléculaire 9164 Da, dont les 8 cystéines sont impliquées dans le réseau de ponts disulfure indiqué figure 3.3. La présence de 12 résidus chargés, dont neuf arginines et lysines, rend cette protéine basique avec un pI théorique de 9,1. La séquence de la LTP1_1 présente de fortes similarités de séquence avec celles des LTP1 de graines de céréales (§ 2.2 figure 3 de l'article).

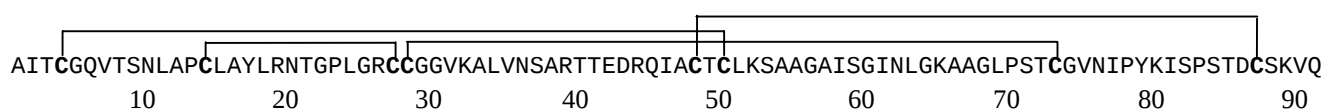


Figure 3.3 : Séquence de la LTP1_1 de tabac. Les ponts disulfure sont indiqués.

Afin de produire une quantité importante de protéine pour les études RMN (environ une dizaine de milligrammes), la LTP1_1 de tabac a été surexprimée dans la levure *Pichia pastoris*. En effet, cet hôte est reconnu très efficace pour produire des protéines riches en ponts disulfure alors que des essais de production de LTPs dans la bactérie *Escherichia coli* conduisait à la formation de corps d'inclusion, nécessitant ensuite des méthodes de purification dans des conditions de dénaturation très drastiques. Il faut néanmoins noter que tout récemment l'équipe de D. Marion de l'INRA de Nantes (Elmorjani et al. 2004) a réussi

grâce à une souche mutante de *E. coli* à produire la LTP1 de blé sans formation de corps d'inclusion dans le cytoplasme de l'hôte.

Le travail de production de la LTP1_1 de tabac dans *Pichia pastoris* et de sa purification a été réalisé par nos collaborateurs de l'INRA d'Antibes, Michel Ponchet, Antoine Marais et Benoit Industri. J'ai néanmoins (modestement) participé aux étapes de purification que je vais décrire succinctement. La stratégie de production de la LTP1_1 de tabac est expliquée en détail dans l'article qui est présenté dans le paragraphe 2.2 de ce chapitre.

2.1 Purification de la LTP1_1 de tabac

Pour réaliser l'étude structurale RMN de la LTP1_1, des échantillons de protéine non marquée, simplement marquée ^{15}N et doublement marquée ^{13}C et ^{15}N ont été produits et purifiés. Les échantillons marqués ^{15}N et ^{13}C ont été obtenus en ajoutant dans le milieu nutritif de culture de *Pichia Pastoris*, du sulfate d'ammonium ($^{15}\text{NH}_4$) $_2\text{SO}_4$ enrichi à plus de 98% en ^{15}N et du méthanol ($^{13}\text{CH}_3\text{OH}$) enrichi à plus de 99% en ^{13}C . Le protocole de purification est le même quelle que soit la nature du marquage

Les levures recombinantes qui ont intégré les plasmides codant pour la LTP1_1, expriment et sécrètent la protéine dans le milieu de culture. Après élimination du milieu de culture, la LTP1_1 est purifiée par chromatographies successives : une première échangeuse d'ions et une seconde en phase inverse. La LTP1_1 est alors dialysée contre de l'eau ultra pure et enfin lyophilisée. A chaque étape de purification, la présence et la pureté de la LTP1_1 sont vérifiées par HPLC (pour High Performance Liquid Chromatography) et SDS-PAGE (pour Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis). La figure 3.4 illustre le profil d'élution de la LTP1_1 de tabac après la chromatographie en phase inverse.

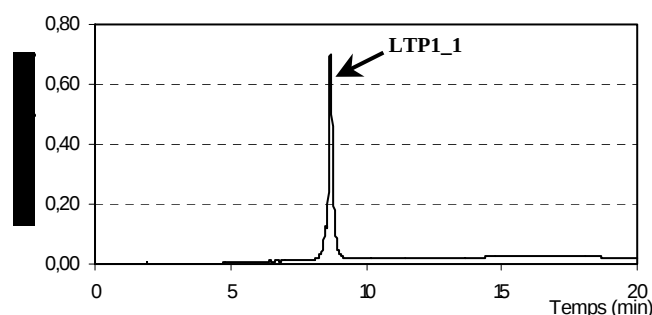


Figure 3.4 : Profil d'élution de la LTP1_1 de tabac en HPLC après l'étape de purification par chromatographie en phase inverse.

L'intégrité de la protéine est vérifiée par spectrométrie de masse. La figure 3.5 représente le spectre de masse obtenu pour l'échantillon de LTP1_1 non marqué.

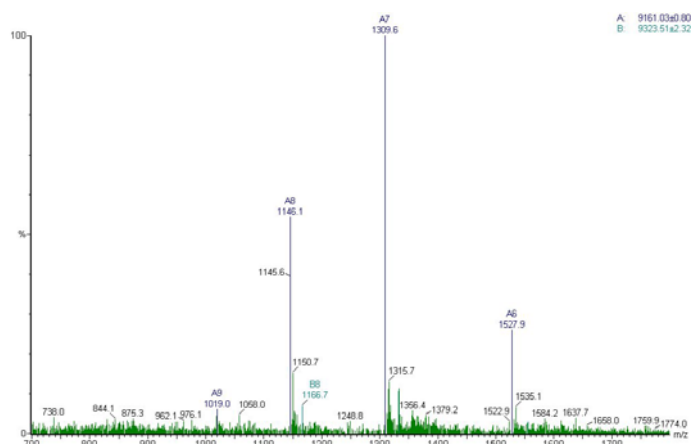


Figure 3.5 : Spectre de masse de la LTP1_1 de tabac non marquée.

L'analyse du spectre de masse de la LTP1_1 de tabac non marqué révèle deux masses : une première de 9161 Da et une deuxième de 9323 Da. Ces deux masses correspondent respectivement à la LTP1_1 réduite (avec ses ponts disulfure formés) et à la LTP1_1 avec un adduit glycosydique de 162 Da de type mannose ou glucose (seulement 8% de l'échantillon est glycosylé).

D'autre part, l'analyse des spectres de masse des échantillons marqués de la LTP1_1 montre que ces derniers sont enrichis à plus de 96 %, que ce soit en ^{15}N ou ^{13}C .

Le rendement de production de la LTP1_1 de tabac après purification est de l'ordre de 15 à 20 mg par litre de culture de levure, quelles que soient les conditions de marquage isotopique. Ces quantités de protéine sont largement suffisantes pour envisager l'étude structurale de la LTP1_1 par RMN.

2.2 Article sur la LTP1_1 de tabac : structure 3D et études d'interactions

Ce paragraphe présente la structure 3D de la LTP1_1 de tabac, ainsi que les études d'interactions de la LTP1_1 avec des lipides (phospholipides, acides gras). Il prend la forme d'un article intitulé : « Solution structure of a tobacco Lipid Transfer Protein exhibiting new biophysical and biological features », qui est sous presse dans le journal, *Proteins: Structure, Function, and Bioinformatics*

Solution Structure of a Tobacco Lipid Transfer Protein Exhibiting New Biophysical and Biological Features

Pedro Da Silva,¹ Céline Landon,¹ Benoit Industri,² Antoine Marais,² Didier Marion,³ Michel Ponchet,² and Françoise Vovelle^{1*}

¹Centre de Biophysique Moléculaire, UPR 4301 CNRS affiliated with Orléans University, rue Charles Sadron, 45071 Orléans cedex 2, France

²UMR Interactions Plantes-Microorganismes et Santé Végétale, INRA, 400 Route des Chappes, BP 167, 06903 Sophia Antipolis cedex, France

³Unité de Recherche sur les Protéines Végétales et leurs Interactions, INRA, La Géraudière, BP71627, 44316 Nantes cedex 3, France

ABSTRACT Plant lipid transfer proteins are small soluble extracellular proteins that are able to bind and transfer a variety of lipids *in vitro*. Recently, it has been proposed that lipid transfer proteins may play a key role in plant defence mechanisms, especially during the induction of systemic acquired resistance. However, very little is known about the proteins expressed in developing plants and tissues, since almost all the biophysical and structural data available to date on lipid transfer proteins originate from proteins present in storage tissues of monocot cereal seeds. In this paper, we report the structural and functional characteristics of a lipid transfer protein (named LTP1_1) constitutively expressed in young aerial organs of *Nicotiana tabacum* (common tobacco). The unlabelled and uniformly labelled proteins were produced in the yeast *Pichia pastoris*, and we determined the three-dimensional (3D) structure of LTP1_1 using nuclear magnetic resonance (NMR) spectroscopy and molecular modeling techniques. The global fold of LTP1_1 is very close to the previously published structures of LTP1 extracted from cereal seeds, including an internal cavity. However, the chemical shift variations of several NMR signals upon lipid binding show that tobacco LTP1_1 is able to bind only one LysoMyristoylPhosphatidylCholine (LMPC), while wheat and maize LTPs can bind either one or two. Titration experiments using intrinsic tyrosine fluorescence confirm this result not only with LMPC but also with two fatty acids. These differences can be explained by the presence in tobacco LTP1_1 of a hydrophobic cluster closing the second possible access to the protein cavity. This result suggests that LTP1 lipid binding properties could be modulated by subtle changes in a conserved global structure. The biological significance of this finding is discussed in the light of the signalling properties of the tobacco LTP1_1-jasmonate complex described elsewhere.¹ Proteins 2005;59:000–000. © 2005 Wiley-Liss, Inc.

Key words: LTPs; cysteine-rich protein; hydrophobic cavity; tobacco; orthogonal bundle

INTRODUCTION

Plant lipid transfer proteins (LTPs) form a very homogeneous class of proteins. These small, abundant and mostly basic proteins contain eight cysteine residues implicated in four conserved disulfide bridges.^{2,3} They are capable of transferring lipids between membranes *in vitro* but their biological function *in vivo* remains unclear. LTPs form a multigenic family of proteins that can be divided into two subfamilies, LTP1 [molecular weight (MW) ≈9 kDa] and LTP2 (MW ≈7 kDa). About 30 genes of putative LTPs have been identified in the model plant *Arabidopsis thaliana*,⁴ and their expression pattern is complex. It is characterized by strong developmental and tissue specificity and modulated by environmental and stress conditions.³ These data, together with physiological and structural studies, suggest various possible roles for plant LTPs involving cutin biosynthesis,⁵ surface wax formation,⁶ pathogen defense⁷ and adaptation of the plant to environmental changes.^{8–10} Recently, we proposed that LTPs could act as systemic signalling molecules,¹¹ since a wheat LTP1 from seed was shown to be able to compete with elicitors on their tobacco plasma membrane receptor.¹² Elicitors are small cysteine-rich holoproteins secreted by the phytopathogenic oomycetes *Phytophthora* and *Pythium* that trigger hypersensitive response (HR) and systemic acquired resistance (SAR) in tobacco and some *Brassicaceae* species.¹³ Although no elicitor activity of wheat LTP1 could be observed, *in vivo* competition between this LTP1 and the elicitor cryptogein led to a loss of cryptogein activity on tobacco.¹² The results

Abbreviations: 3D, three-dimensional; ¹H, proton; 2D NMR and 3D NMR, two- and three-dimensional nuclear magnetic resonance; LTP, lipid transfer protein; LMPC, LysoMyristoylPhosphatidylCholine; SAR, systemic acquired resistance; HR, hypersensitive response; YNB, yeast nitrogen base; RMSD, root mean square deviation; ARIA, Ambiguous Restraints for Iterative Assignment

SWISS-PROT accession number: LTP1_1 (Q42952).

PDB accession number: 1T12.

*Correspondence to: Professor Françoise Vovelle, Centre de Biophysique Moléculaire, CNRS UPR 4301, Rue Charles Sadron, 45071 Orléans Cedex 2, France. E-mail: vovelle@cnrs-orleans.fr.

Received 9 June 2004; Revised 22 October 2004; Accepted 26 October 2004

Published online 00 Month 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20405

with wheat LTP1 were obtained in a "heterologous" biological system (wheat/tobacco), so that any attempt to characterize LTP involvement in plant defence mechanisms requires homologous proteins in the tobacco-elicitor biological system. To our knowledge, no native protein and only four nucleotidic sequences encoding tobacco LTP1 have so far been described in the literature and sequence data banks. Our strategy was first to select a cDNA encoding a canonical LTP1 in terms of sequence homology. The cDNA encoding the NLT1_TOBAC, accession Q42952 (http://www.expasy.org/sprot/), fit our criteria, and the corresponding mature protein will be termed LTP1_1 throughout this work. The *Ltp1* gene is constitutively expressed in the whole plant except for the roots. The highest levels of transcripts were observed in the upper young aerial leaves with a localization in the epidermis,¹⁴ which was confirmed by promoter analysis of the *ltp1* gene with GUS fusion and expression pattern.¹⁵ The second step was to obtain the corresponding mature protein through heterologous overexpression in the yeast *Pichia pastoris* in order to determine its structural, biophysical and biological properties. This constitutes the first integrative approach concerning an LTP1 expressed in the developing tissues of plants. So far, the only known three-dimensional (3D) structures have been established for LTP1 extracted from cereal seeds^{16–21} by X-ray or nuclear magnetic resonance (NMR) spectroscopy techniques. They are all characterized by the presence of a four α -helix orthogonal bundle and a hydrophobic cavity capable of accommodating a variety of lipidic molecules. In order to establish the 3D structure of the tobacco LTP1_1 using 2D and 3D NMR, the protein was expressed unlabeled and uniformly labeled. Its binding properties were analyzed by following the chemical shift variations of several NMR signals upon lipid binding and fluorescence techniques. This study reveals that, despite a global fold very similar to that of cereal seed LTP1 and the presence of a large cavity, the tobacco protein presents dissimilar binding abilities.

MATERIALS AND METHODS

Over-Expression of Tobacco LTP1 and ¹⁵N Labeling

The tobacco LTP1_1 cDNA was first amplified by polymerase chain reaction (PCR) with two oligonucleotides, oligo1 (5'-CTAGCTAGATACTTCATT-3') and 2 (5'-ATAGACGAACACATCATATT-3') using the total cDNA of tobacco leaves as a template. The PCR fragment was cloned into pGEM-T easy (Promega) and sequenced to verify its complete identity to Q42952. The encoding sequence was reamplified using two oligonucleotides, oligo3 (5'-GTATCTCTCGAGAAAAGAGCCATAACCTGTGGC-CAG-3') and 4 (5'-ATAATACCTAGGTTACTGGACCTTG-GAGCA-3') as forward and reverse primers, respectively. The oligonucleotide 3 was designed to introduce a XhoI site and to eliminate the end of the yeast signal peptide corresponding to the EAEA amino acid sequence. The oligonucleotide 4 introduced an AvrII site to ensure that plasmid integration was in the right direction. The amplified fragment was purified and cloned into pPIC9. The correct phas-

ing of the reading frame was checked by PCR screening and confirmed by DNA sequencing. The yeast transformation (strain GS115, obtained from Invitrogen) was performed as previously described.²²

At least 15 *P. pastoris* transformant strains were cultivated on a derived YNB medium containing (NH₄)₂SO₄ (10 g/L), KH₂PO₄ (2 g/L), MgSO₄ · 7H₂O (2.046 g/L), NaCl (0.2 g/L), potassium phosphate buffer (100 mM, pH 6), H₃BO₃ (1 mg/L), CuSO₄ · 5H₂O (0.05 mg/L), KI (0.1 mg/L), MnSO₄ · H₂O (5 mg/L), Na₂MoO₄ · 2H₂O (0.2 mg/L), ZnSO₄ · 7H₂O (10 mg/L), CoCl₂ · 6H₂O (0.05 mg/L), calcium pantothenate (1 mg/L), nicotinic acid (1 mg/L), pyridoxin-HCl (1 mg/L), thiamine-HCl (1 mg/L), biotin (1 mg/L), folic acid (0.01 mg/L), inositol (5 mg/L), p-amino benzoic acid (0.5 mg/L), riboflavin (0.5 mg/L), CaCl₂ · 2H₂O (0.264 g/L) and FeSO₄ · 7H₂O (0.6 mg/L). Methanol 0.5% (v/v, final concentration) containing 100 mg/L of biotin was added daily as the sole carbon source. All cultures were grown at 29°C, in the dark, on rotatory shakers (250 rpm) holding 1 L Erlenmeyer flasks containing 100 mL of the medium. A preculture was carried out for 24 h and was used to initiate the culture after a four-fold concentration of the yeast suspension by centrifugation (5,000g, 10 min). After 5 days of culture, the medium was recovered after centrifugation (10,000 g, 10 min, 4°C). The strain producing the highest amount of protein was used for production and labeling experiments.

For ¹⁵N-LTP1_1 labeling, cultures were grown in the previously described medium containing (¹⁵NH₄)₂SO₄ (>98%, Spectra Stable Isotope, USA) at 1 g/L instead of 10 g/L. For ¹⁵N and ¹³C labeling, the previous medium for ¹⁵N feeding was used and methanol was replaced by ¹³C methanol (99%, Euriso-Top, France).

Purification of Recombinant Labeled and Unlabeled Proteins

After culture centrifugation, the pH of the supernatant was adjusted to 2.0 by the addition of TFA and was loaded onto Amberlite XAD7 (Fluka), pre-equilibrated with 0.1% aqueous TFA. After washing with 0.1% aqueous TFA and 10% CH₃CN, 0.1% TFA, the tobacco LTP1_1 was eluted by 40% CH₃CN, 0.1% TFA. CH₃CN was evaporated under vacuum, and the fraction was adjusted to pH 7 with 1M Tris and loaded onto a Trisacryl SP ion exchange column (M grade, Biosepra), pre-equilibrated with Tris-HCl 20 mM, pH 7.0 buffer. A step gradient of NaCl in the same buffer was applied to the column: 0.1, 0.25, 0.5M. The protein was recovered in the 0.25M fraction. This fraction was adjusted to pH 2.0 with TFA, and CH₃CN was added to reach a 10% (v/v) final concentration; then it was settled on a reversed phase column (Uptisphere C4, 300 Å, 20 µm, Interchim). After washing with CH₃CN 10%, 0.1% TFA, the protein was eluted with 20% CH₃CN, 0.1% TFA. The fraction containing the tobacco LTP1_1 was dialyzed extensively against ultra pure water and freeze-dried. The protein content and purity was checked throughout the purification by sodium-dodecyl-sulfate-polyacrylamide-gel electrophoresis (SDS-PAGE) and high-performance liquid chromatography (HPLC). HPLC was performed on a Zor-

AQ: 6

AQ: 5

AQ: 5

AQ: 7

AQ: 5

AQ: 5

bax 300 SB-CN (Agilent) column (250 × 4.6 id, 5 μm, 300 Å) at a flow rate of 1 mL/min. A gradient of 7% to 49% CH₃CN in 0.1% TFA (16 min, linear) was applied to the column, and detection was secured by a photodiode array to check peak spectral purity (M625 solvent delivery, PDA 996, Millennium software, Waters).

Mass Spectrometry (MS)

The recombinant tobacco LTP1_1, both unlabeled and labeled, was analyzed by MS. Analyses were performed on a triple quadrupole mass spectrometer (Quattro II, Micromass, Manchester, UK) equipped with a nebulizer-assisted electrospray source. Instrument control and data analysis were accomplished using Masslynx application software, version 4.0, from Micromass (Manchester, UK). Calibration of the mass spectrometer was performed using myoglobin. The sample cone voltage was maintained at 45 V.

Nuclear Magnetic Resonance (NMR) Experiments Sample Preparation

The sample used for NMR spectroscopy was prepared by dissolving 10 mg of recombinant tobacco LTP1_1 powder in 600 μL of 90% H₂O:10% ²H₂O, resulting in a final concentration of 1.8 mM. The pH of the protein solution was adjusted to 5.2 with minute increments of HCl 1N in order to observe the amide protons. For ¹H/²H exchange kinetics, the sample was lyophilized and quickly redissolved in 99.9% ²H₂O. The exchange was then monitored from short ¹H-¹⁵N HSQC spectra.

NMR Spectroscopy

Two-dimensional and 3D NMR spectra were recorded on a VARIAN INOVA NMR spectrometer equipped with a z-axis field-gradient unit and operating at a proton frequency of 600 MHz. In order to solve assignment ambiguities, a series of homonuclear and heteronuclear 2D experiments were performed at two temperatures (293K and 303K): ¹H-¹⁵N-HSQC,²³ DQF-COSY,²⁴ clean-TOCSY with a spin lock of 80 ms produced by the MLEV-17 sequence^{25,26} and NOESY spectra collected at mixing times

AQ: 8 of 100 and 200 ms.²⁷ ¹H and ¹⁵N assignments were achieved using standard procedure through the analysis of a combination of 3D ¹⁵N-edited NOESY-HSQC²⁸ at mixing times of 100 and 200 ms and TOCSY-HSQC²⁸ spectra with a spin lock of 80 ms. For the TOCSY- and NOESY-HSQC experiments, 1024 (¹H) × 140 (¹H) × 32 (¹⁵N) complex points were recorded using eight scans, for a total measuring time of ≈36 h and ≈60 h, respectively. A AQ: 9 HNHA²⁹ spectrum was also acquired to determine ³J_{HN-Hα} coupling constants from which restraints for the ϕ torsion angle were extracted by means of the appropriate Karplus equations.³⁰ ¹³C sequence-specific backbone assignments were performed using a suite of triple-resonance NMR experiments in the Varian Biopack, including HNCA,³¹ HN(CO)CA,³² HNCO,³¹ HN(CA)CO,³² HNCACB,³³ AQ: 10 CBCA(CO)NH.³⁴ Additional assignments and chemical shift data were obtained from a 2D ¹³C HSQC³⁵ and 3D ¹³C HCCH-TOCSY³⁶ experiments (aromatic and ali-

phatic). All NMR datasets were processed using the NMRpipe/NMRdraw software package.³⁷

Structure Calculations and Analysis

Spectra analysis, volume integrations and ³J_{HN-Hα} measurement were achieved with NMRVIEW software.³⁸ Combined NOE cross-peak assignment and 3D protein structure calculations were performed using ARIA 1.1³⁹ implemented in the CNS 1.1 software.⁴⁰ The input consisted of the chemical shift list obtained from the resonance assignment, the list of cross-peak positions and volumes with or without assignment and the dihedral angles obtained from measured ³J_{HN-Hα} coupling constants. These ϕ angles were restrained to -60 ± 40° for ³J_{HN-Hα} < 6 Hz; and -120 ± 30° for ³J_{HN-Hα} > 8 Hz. AQ: 11 Moreover, since the disulfide pairing of tobacco LTP1_1 has been determined unequivocally on the basis of inter-residual NOE connectivities for the four disulfide bridges, the following distance restraints between two cysteine residues involved in the linkage were added for Cys4-Cys50, Cys14-Cys27, Cys28-Cys73 and Cys48-Cys87 disulfide pairs: 2.0 Å < d(Sγ',Sγ') < 2.1 Å; 3.0 Å < d(Cβ',Sγ'); d(Sγ',Cβ') < 3.1 Å. During the ARIA run, an automated calibration was performed, unassigned NOE peaks were assigned, distance restraints were generated and the 3D protein structures were calculated. At the end of each run, rejected restraints and residual NOE violations were analyzed. The new assignments were checked and introduced or excluded the subsequent run. This procedure of assignment/refinement was repeated iteratively until complete assignment of NOESY spectra. In the final iteration, the 50 structures with the lowest restraint energy values were further refined by molecular dynamics simulation in water to remove artefacts.⁴¹ The 10 lowest restraint energy structures were used for statistical analysis. AQ: 12

The structures were displayed and analyzed using the MOLMOL,⁴² SYBYL (TRIPOS Inc., St. Louis, MO) and PROMOTIF⁴³ programs. Their stereochemical quality was evaluated using PROCHECK.⁴⁴ The volume of the LTP1 internal cavity was calculated by 'CAVITE' software.⁴⁵

Lipid Binding Studies

The binding of LysoMyristoylPhosphatidylCholine (LMPC) to wheat, maize and tobacco LTP1 was probed by comparing a series of one-dimensional proton spectra at different lipid-to-protein molar ratios (*R_i*). Moreover, to AQ: 3 investigate the precise molar ratio of the LTP1_1-LMPC complex and to survey regions of LTP1_1 to interact with LMPC, a series of ¹H-¹⁵N HSQC and ¹⁵N-filtered 1D ¹H-NMR spectra was performed using ¹⁵N labeled LTP1_1. To quantitatively evaluate the shift of the ¹H-¹⁵N amide group cross-peaks in the ¹H-¹⁵N HSQC spectra, a "total LMPC-induced shift," Δ, was defined as

$$\Delta = [(w\Delta\delta_H)^2 + (\Delta\delta_N)^2]^{1/2}.$$

Here, Δδ_H and Δδ_N denote the induced chemical shift AQ: 12 variations, measured in parts per million, of the ¹H and AQ: 3 ¹⁵N dimension, respectively. To account for the difference

The ^1H - ^{15}N -HSQC of LTP1_1 (Fig. 1) displays an excellent dispersion for the proton and nitrogen resonances and is characteristic of a protein with an ordered structure. Sequence-specific assignments for the ^1H and ^{15}N resonances were performed using standard 3D heteronuclear experiments (TOCSY- and NOESY-HSQC) and 2D homonuclear COSY, TOCSY, NOESY spectra. From these well-resolved spectra, a complete assignment was obtained for all backbone nitrogens and protons and for all non-labile side-chain hydrogens. The labile side-chain protons and nitrogens were completely assigned, with the exception of ηN and ηNH_2 of arginine residues (at positions 19, 26, 39 and 44), ζN of lysine residues (at positions 32, 52, 65, 80, 89) and ζNH_2 of lysine 89. All ^{13}C chemical shifts were assigned except for Ala1 and Cys4 carbonyl (see supplement-

tary material for the chemical shift table). Strong $H\alpha(i)$ – $H\delta(i+1)$ and $NH(i)$ – $H\delta(i+1)$ NOE crosspeaks for Ala12–Pro13, Gly22–Pro23, Leu69–Pro70, Ile77–Pro78 and Ser82–Pro83 indicate the presence of *trans* peptide bonds for the five prolines. Moreover, when the freeze-dried sample of tobacco LTP1_1 was dissolved in 2H_2O , the signals of 18 amide protons, namely Cys14, Leu15, Leu18, Arg19, Val31, Lys32, Ala33, Leu34, Val35, Arg44, Gln45, Ile46, Ala47, Cys48, Thr49, Cys50, Leu51 and Lys52, were still observed after 24 h (see supplementary material), strongly suggesting that they are involved in hydrogen bonds.

Structure Calculation

More than fifteen hundred distance restraints were derived from the volume integration of cross peaks observed in the 2D NOESY and 3D HSQC–NOESY spectra recorded at 303K with a mixing time of 100 ms. Forty-seven ϕ dihedral angles were deduced from $^3J_{HN-H\alpha}$ coupling constants measured on the HNHA spectrum. Six characteristic NOEs [Cys4(H β 1)–Cys50(H α), Cys14(H α)–Cys27(H β 2), Cys14(H β 2)–Cys27(H β 2), Cys28(H β 1)–Cys73(H α), Cys28(H β 2)–Cys73(H α) and Cys48(HN)–Cys87(H β 1)] between cysteine protons indicate the presence of four disulfide bridges Cys4–Cys50, Cys14–Cys27, Cys28–Cys73 and Cys48–Cys87, in agreement with the conserved disulfide bridge array of LTP1. The corresponding disulfide distance restraints (see Materials and Methods) were also introduced in the restraint list.

After nine ARIA iterations, all of the NOE peaks were assigned. The final NOE data file comprised 557.2 intra-residue, 385.5 sequential, 360.7 medium-range ($2 < |i-j| < 5$), 222.6 long-range ($|i-j| \geq 5$) restraints (an average of 17 restraints per residue) among which 1258 were non-ambiguous. A family of ten structures with the lowest restraint energies was selected from the 50 structures refined by molecular dynamics simulation. These structures are in very good agreement with the standard covalent geometry and do not show experimental violations greater than 0.3 Å (Table I). Both negative van der Waals and electrostatic energy terms are indicative of favorable non-bonded interactions. Moreover, the Ramachandran plot exhibits 97% of the (ϕ, ψ) angles of the ten selected structures in the most favored and additional allowed regions according to the PROCHECK software nomenclature.

Structure Description

A sausage representation of the backbone atoms of the ten best structures of tobacco LTP1_1 is shown in Figure 2(a). The overall fold of tobacco LTP1_1 includes four α -helices: H1 (Cys4–Leu18), H2 (Cys27–Val37), H3 (Glu42–Ala55) and H4 (Leu63–Thr72), followed by a C-terminal segment involving three γ -turns (Fig. 3). Due to the presence of a proline (Pro13) in its sequence, helix H1 is kinked at residues Leu11, the pattern of hydrogen bonds is disrupted and the helix is divided into two sub-helices H1a, H1b [Fig. 2(a)] with an angle of roughly 75° between H1a and H1b. Pro70 in helix H4 does not produce a

TABLE I. Structural Statistics for 10 Models of Tobacco LTP1_1

Conformational Restraints		
NOE-based Distance Restraints		
Total		1526
Intraresidue ($ i-j = 0$)		557.2
Sequential ($ i-j = 1$)		385.5
Medium range ($2 \leq i-j < 5$)		360.7
Long-range ($ i-j \geq 5$)		222.6
Unambiguous restraints		1258
Ambiguous restraints		268
Dihedral angle restraints		47
Residual Violations		
	Average Number Per Structure	
Distance restraints $> 0.3 \text{ \AA}$	0	
Dihedral angle restraints $> 5^\circ$	0	
RMS deviation from Standard Geometry		
Bond lengths ($\text{\AA}$)	0.0035 ± 0.0002	
Bond angles ($^\circ$)	0.49 ± 0.02	
Energies (kcal/mol)		
NOE	17 ± 2	
Van der Waals	-252 ± 16	
Electrostatic	-3152 ± 92	
Ramachandran Maps ^a		
Most favoured regions	79.6%	
Additional allowed regions	17.4%	
Generously allowed regions	1.6%	
Disallowed regions	1.4%	
Average RMSD (pairwise, $\text{\AA}$)		
	Backbone N, C $_{\alpha}$ and C'	All heavy atoms
Whole	1.03 ± 0.15	1.55 ± 0.15
Secondary structures	0.69 ± 0.12	1.20 ± 0.12
Helix (4–18)	0.42 ± 0.10	0.92 ± 0.14
Helix (27–37)	0.38 ± 0.11	0.88 ± 0.17
Helix (42–55)	0.32 ± 0.10	1.23 ± 0.25
Helix (63–72)	0.36 ± 0.10	0.84 ± 0.17

^aRamachandran plots were calculated with PROCHECK.

marked kink but rather a curvature centered at residue Gly68. The C-terminal segment involves three γ -turns: the first is formed by Val75, Asn76 and Ile77; the second by Ile77, Pro78 and Tyr79; the third by Ser82, Pro83 and Ser84. They are all stabilized by a canonical C = O(i)–N–H($i+2$) hydrogen bond. The overall structure is stabilized by the four disulfide bridges distributed in pairs on opposite sides of the molecule. On one side, the 4–50 disulfide bridge links the N-terminus to the H3 helix, and the 48–87 bridge links H3 to the C-terminus. On the other side, the 14–27 bridge connects H1 to the beginning of H2 and the 28–73 bridge links H2 to the C-terminal segment. As evidenced by the structural statistics (Table I), the structure of tobacco LTP1_1 is particularly well defined. The pairwise RMS deviations on the N, C α , C' backbone atoms of residues 1 to 91 are 1.03 Å (1.55 Å on all heavy atoms) and drop to 0.69 Å (1.20 Å for heavy atoms) when calculated on the secondary structure elements. The presence of hydrogen bonds typical of the secondary structures

T1

F2

F3

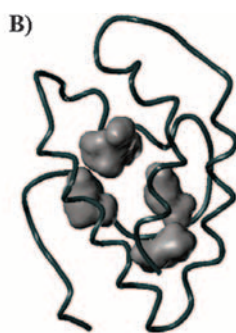
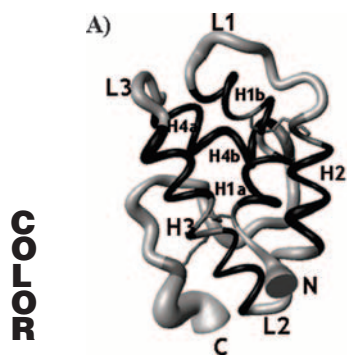


Fig. 2. Sausage representation of the backbone and of the cavity of tobacco LTP1_1: (a) sausage representation of the 10 selected structures of tobacco LTP1_1. The radius of the sausage is proportional to the RMSD on the coordinates of the backbone atoms of the 10 structures. The four helices (H1, H2, H3, H4) are in black, the loops and turns in light gray and the four disulfide bonds in grey; (b) cavity of tobacco LTP1_1; the backbone of the molecule is represented as a tube.

is observed on the ten models and is confirmed by the ^1H - ^2H exchange experiments.

The main characteristic of this structure is the presence of an internal cavity [Fig. 2(b)] delineated by the hydrophobic faces of the amphipathic helices facing each other and by the hydrophobic residues of the C-terminal segment, namely Ile77, Tyr79, Ile81. Even if the side chains of the hydrophobic residues that point in towards the core of the protein are relatively well defined (small χ_1 and χ_2 circular variances), the shape and size of the cavity fluctuate from one model to another. This internal cavity very often appears to be divided into several smaller cavities [Fig. 2(b)].

Lipid Binding

The binding characteristics of tobacco LTP1_1 were studied by monitoring the chemical shift variations on ^1H - ^{15}N -HSQC, ^{15}N -filtered 1D ^1H -NMR and 2D ^1H -TOCSY spectra with increasing concentrations of LMPC. The spectral changes induced by the interaction of LMPC with LTP1_1 are typical of a slow exchange on the NMR time scale. Several peaks appear or disappear upon titration, indicating the coexistence of free and bound forms of the protein. Figure 4 depicts the variation in the tobacco LTP1_1 1D NMR spectrum when the lipid/protein molar ratio, R_i , varies from 0 to 2. The overall aspect of the spectrum changed up to an R_i value of 1.0, but no further modifications were observed for larger R_i values, demonstrating that saturation is reached at a ratio of one lipid per protein. The modifications of the NMR spectra of tobacco LTP1_1 in the 6.6-7.1 ppm region [Fig. 4(a)] evidence that Tyr79 signals are clearly affected by lipid binding, while Tyr17 ones are unaltered. In the 0.0-0.7 ppm region [Fig. 4(b)], the methyl group signals of Leu24, Ile77 and Val90 exhibit different behavior. The Ile77 signals of the free protein diminish for increasing R_i values, up to $R_i = 1.0$, whereas, concomitantly, the Val90 signals of the bound form begin to appear. In contrast, like Tyr17, Leu24 is not affected by lipid binding.

Moreover, Δ values (see Mat. and Meth.) were evaluated from the shifts of the ^1H and ^{15}N amide group cross peaks

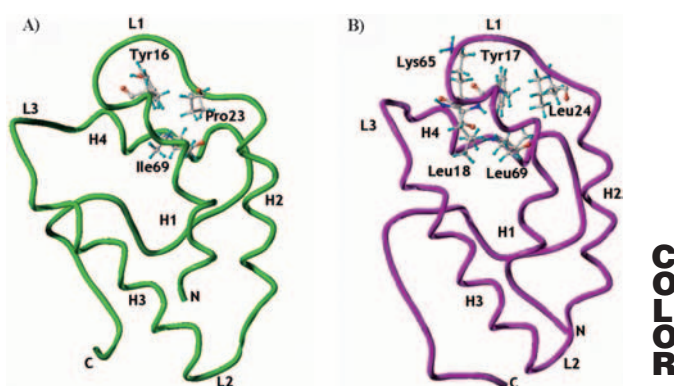


Fig. 6. Representation of hydrophobic cluster of (a) wheat and (b) tobacco LTP1_1. The backbone of each protein is represented as a tube.

in the ^1H - ^{15}N -HSQC spectra of LTP1_1 in complex with one molecule of LMPC and of the free protein (Fig. 5). The highest Δ values (2.0–7.0 ppm) occur for amino acid groups located within three different regions. The first is delimited by the C-terminal region (Val75, Ile77, Tyr79, Ile81, Thr85, Asp86, Ser88, Gln91); the second is located in the L2 loop (Ala38, Arg39); and the third involves Asn10, Leu11 and Ala12 residues close to the kink of H1 helix. The largest shift is observed for Arg39, located in the second region.

In order to determine the residues most perturbed by lipid binding, we have monitored the ^1H chemical shift variations of aliphatic protons. The chemical shift differences larger than 0.12 ppm are listed in Table III. They correspond to 20 amino acids, Leu11, Cys27, Leu34, Val35, Cys48, Leu51, Ala54, Ala55, Ile58, Ser59, Ile61, Ile77, Tyr79, Lys80, Ile81, Ser82, Ser84, Thr85, Asp86 and Val90. Among them, 14 are located on the inner surface of tobacco LTP1_1 facing the cavity. The six other amino acids are closed to the preceding one, Ser59, Lys80, Ser82, Ser84, Asp86 and Val90. No significant variation of the proton chemical shifts was detected for side chains on the outer surface of the protein.

The variation in NMR signals of maize and wheat LTP1 upon titration with LMPC were followed by monitoring the chemical shift variations on 1D ^1H -NMR, using the Tyr81e signals and Val75 γ_2 methyl group, respectively. As for tobacco LTP1_1, changes in the proton signals were observed for increasing values of R_i , but the saturation occurred at 1.4, indicating that these two types of LTP1 are able to bind either one or two LMPC molecules in their cavity (see supplementary material).

In order to obtain supplementary acyl binding characteristics for tobacco LTP1_1, titration experiments were carried out by following the tyrosine fluorescence intensity variations upon binding of LMPC (as for NMR) and of two fatty acids, namely palmitic acid (C16:0) and oleic acid (*cis* C18:1). This technique allows determination of both the affinity constant (K_d) and the number of binding sites (n). The results presented in Table II indicate that tobacco LTP1_1 possesses a unique binding site not only for LMPC but also for fatty acids, which are classical constituents of plant membranes or lipidic vesicles. Moreover, the affinity

AQ: 3

F4

AQ: 3

AQ: 3

AQ: 3

AQ: 3

F5

T3,
AQ: 16

AQ: 14

AQ: 3

SOLUTION STRUCTURE OF A TOBACCO LTP1

	10	20	30	40	50	60		
tobacco	A ITCGQV TSN LAPCLAYLRN - -TGPLGR CCGGVKALVNSARTTEDRQI ACTCLKS AA GAI 58							
rice	-ITCGQVNSAVGPCLTY ARG -GAGPSA ACCSG VRS LKAAASTTADRR TACNCLKN AA ARGI 58							
maize	A ISCGQV AS AIAPCISY AR QGSGPS AG CCSG VRS LNNAARTTADRR AA CN CLKN AA AGV 60							
barley	-L NC CGQVDS KMKPCLTYVQ G-GPGPS GECCNG V RD LHNQAQSSGDRQ TV CNCLKG IA ARGI 58							
wheat	-IDCGH VDSLVRPCLSYVQ G-GPGPS SGCCD GV KNLHNQARSQ SDRQ SACN CLKG IA ARGI 58							
	70	80	90	Identity(%)		Similarity(%)	RMSD(Å)	pI
tobacco	SGINLGK AAAGLPST CGVN IPYKIS PSTDCSKVQ			91				9.1
rice	KGLNAGNAAS IPSK CGV SV PTTISASIDCSRVS			91	53	63	1.13	9.4
maize	SGLNAGNAAS IPSK CGV SI PTTISTSTDCSRVN			93	57	77	1.36	9.1
barley	HNLN LNNAASIP SKCN VN VPYTTISPDIDCSRIY			91	46	59	1.49	8.2
wheat	HNLNEDNARS IPPK CGVN LP TTISLNIDCSRVS			90	48	61	1.51	8.2

Fig. 3. Sequence alignment of tobacco, rice, maize, barley and wheat LTP1. The residues delineating the cavity of each LTP1 are in bold. The residues involved in α -helices are boxed in gray. The total number of residues, the percentages of identity and the similarity of each protein with tobacco LTP1_1 are indicated. The RMSD between the coordinates of the backbone atoms of the secondary structure elements of rice, maize, wheat, and barley LTP1 with the corresponding residues of tobacco LTP1_1, and the theoretical isoelectric point (pI) of each LTP1 are also indicated.

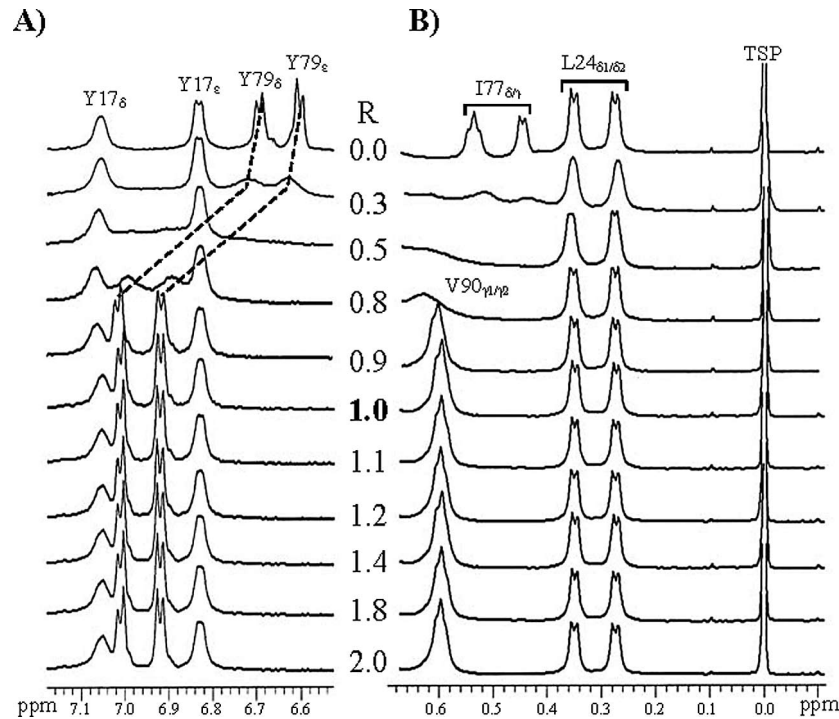


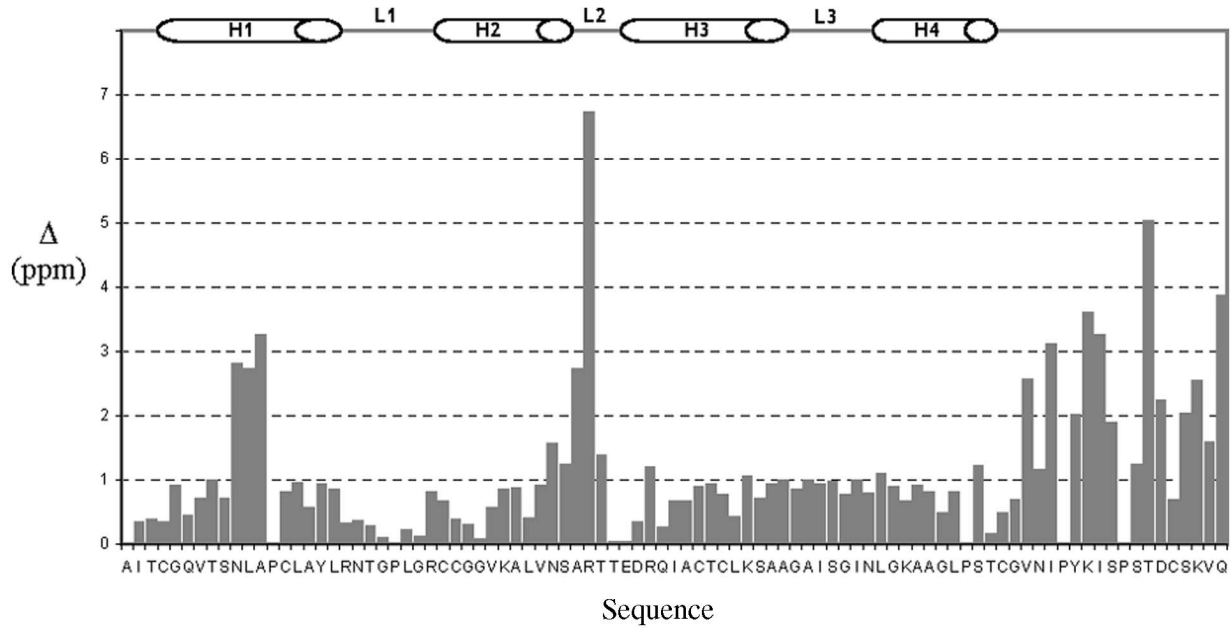
Fig. 4. Modification of the ^{15}N -filtered 1D proton NMR spectra of tobacco LTP1_1 upon binding of LMPC: (a) 6.6 to 7.1 ppm region, (b) 0.0 to 0.6 ppm region. The lipid/protein molar ratio R_l varies from 0 to 2 and is annotated on each spectrum. Several signals are indicated for Tyr17, Tyr79, Leu24, Ile77 and Val90 residues.

constant increases by a factor of 5 to 10 from LMPC to fatty acids.

Discussion

We have described an efficient method to produce LTP1_1 in yeast, both unlabeled and labeled. The major interest of the yeast heterologous system lies in the fact that the protein enters the secretory pathway and thus has a high probability of correct folding to be processed.⁴⁷ Moreover, the protein is easily purified from supernatant and does not require fusion with tags or easy-to-purify foreign

proteins that are difficult to remove in order to recover the recombinant protein. In the present work, the construct allowed us to obtain the whole predicted tobacco LTP1_1 without any additional residue, as confirmed by MS. A major concern of the yeast overexpression technique is a high level of random glycosylation upon the protein of interest. In the case of tobacco LTP1_1, this phenomenon is present but remains very low (about 8% of the protein molecules contain one sugar) and does not interfere with NMR measurements. The protein recovery is not very high (15–20 mg/L) when compared to examples given in the



AQ: 15 Fig. 5. “Total LMPC-induced shift,” Δ , observed for LTP1_1/LMPC complex at $R_i = 1$ for each amino acid of the sequence. The secondary structure elements are indicated on the top of the picture.

TABLE II. Thermodynamic Properties of the Tobacco LTP1 as Measured by Tyrosine Fluorescence (K_d : dissociation constant, n : number of binding sites)

	K_d (μM)	n
LMPC	0.5 ± 0.2	1.0 ± 0.1
Palmitic acid	5.6 ± 3.1	1.1 ± 0.1
Oleic acid	3.9 ± 1.2	1.1 ± 0.1

TABLE III. List of the Aliphatic Protons Experiencing Specific Chemical Shift Variations Upon LMPC Binding

Residue	Atom	$ \Delta\delta_{complex} $
Leu11	H_α	0.26
Leu11	$H_{\delta 1}$	0.12
Cys27	$H_{\beta 1}$	0.14
Leu34	$H_{\beta 2}$	0.13
Leu34	H_γ	0.13
Val35	H_α	0.16
Cys48	$H_{\beta 1}$	0.48
Leu51	H_γ	0.14
Ala54	H_β	0.22
Ala55	H_β	0.18
Ile58	H_α	0.22
Ile58	H_β	0.14
Ile58	$H_{\gamma 1}$	0.12
Ser59	H_α	0.23
Ile61	H_β	0.12
Ile77	H_α	0.20
Ile77	$H_{\gamma 1}$	0.32
Ile77	$H_{\delta 1}$	0.36
Tyr79	H_α	0.12
Tyr79	$H_{\beta 2}$	0.57
Tyr79	$H_{\delta 1,2}$	0.33
Tyr79	$H_{\epsilon 1,2}$	0.32
Lys80	H_α	0.23
Ile81	H_β	0.18
Ile81	$H_{\gamma 1}$	0.48
Ile81	$H_{\gamma 2}$	0.18
Ser82	H_α	0.13
Ser84	H_α	0.18
Thr85	H_γ	0.19
Asp86	H_α	0.15
Val90	H_β	0.13
Val90	$H_{\gamma 1}$	0.41
Val90	$H_{\gamma 2}$	0.45

literature (up to 1 g/L); nevertheless it is sufficient to ensure structural and biological experiments, all the more so as labeling yields are higher than 95% without affecting the protein production.

AQ: 3 At the beginning of this work, our knowledge of the structure and lipid binding abilities of LTP1 suffered from two biological biases: (i) LTP1 originates from very special plant tissues (seeds) and (ii) the studied plant species always belong to monocot cereals. Meanwhile, LTP1 proteins are encoded by a multigenic family and are widespread in all plant organs and tissues. For example, a search for the PFAM PF00234 signature (LTP, amylase inhibitor, seed storage protein...) in the *Arabidopsis* genome (<http://www.arabidopsis.org/>) with MW cut-offs 8–12 kDa returns 44 matches (126 matches without the MW filter). A careful analysis of these hits indicates that more than 35 sequences encode LTP1 or LTP2 putative proteins. Thus, a clear contrast exists between “bias-limited” structural data and LTP1 diversity in the plant kingdom. For this reason, we have determined the 3D solution structure of the recombinant LTP1_1 from the shoot apex of *Nicotiana tabacum* and compared it to cereal LTP structures. Tobacco LTP1_1 adopts the characteristic fold of LTPs: a four-helix orthogonal bundle⁴⁸ followed by a long C-terminal tail with no canonical secondary struc-

AQ: 1

ture. The structure is stabilized by the four disulfide bridges common to all LTP1 proteins and presents the characteristic hydrophobic cavity in the core of the protein, which constitutes the binding site for various hydrophobic ligands.

Four structures of cereal seed LTP1 have been determined by NMR spectroscopy: wheat LTP1,¹⁶ maize LTP1,¹⁹ barley LTP1¹⁸ and rice LTP1.²¹ Maize and rice LTP1 structures have also been obtained by X-ray crystallography^{17,20} and are very close to the NMR structures. Because NMR provides us with an ensemble of models that may reflect the particular plasticity of LTP1, we have restrained our comparisons to the four NMR structures of cereal seed LTP1. Globally, when considering for each LTP1 the average structures of the ensemble of NMR models, the RMS deviation between the backbone heavy atom coordinates of tobacco LTP1_1 and of cereal seed LTP1 are rather close regardless of the molecule (Fig. 3). The value ranges from 1.13 Å for rice LTP1 to 1.51 Å for wheat LTP1.

AQ: 3 The positions and lengths of the helices in the different types of LTP1 are very similar (Fig. 3). In rice LTP1, the H1 helix presents a large kink of 70° (±5), comparable to the value for tobacco LTP1_1 (74° ± 4). In wheat, barley and maize LTP1, the kinks are smaller: 45° (±5), 54° (±3) and 54° (±9), respectively. The kink occurs at position (*i* - 2) with respect to the conserved proline of LTP1_1 in the tobacco protein and at position (*i* - 1) in all the other LTP1. In all considered structures except for maize LTP1, where it is rectilinear, the H4 helix presents a more-or-less marked curvature at position (*i* - 2) with respect to the conserved proline (Pro70, tobacco numbering, see Fig. 3), and it always exposes its concave face to the solvent. The variability of the H1 kink and H4 curvature are the factors most probably involved in the adaptability of the size of the internal cavity to the dimensions of the ligand and in its accessibility. For instance, in barley LTP1 in complex with Acyl-CoA,⁴⁹ the H1 kink angle increases significantly (70° ± 4), and the curvature of H4 is lost. In wheat LTP1, in complex with Prostaglandin B2 (PGB2)⁵⁰ or DiMyristoylPhosphatidylGlycerol (DMPG),⁵¹ helix H4 also becomes rectilinear, but the kink in H1 helix does not change much upon lipid binding. When bounded by two LMPC, the kink increases significantly (75°), probably to allow entrance of the second lipid. In the case of maize LTP1, the structures of H1 and H4 helices are almost identical in the structures of the free protein and bounded proteins.^{17,52} H1 helix, which is already highly kinked, keeps the same kink and H4 remains rectilinear.

The main feature of all LTP1 structures is an elongated cavity running throughout the molecule, which may accommodate one or two hydrophobic ligands. Depending on the origin of the protein and on the NMR model considered, the shape and volume of the cavity can vary noticeably. In particular, for the same LTP1, the cavity can be split into several tiny cavities or can form a continuous single cavity. Tobacco LTP1_1 cavity is fragmented into several parts. It has been investigated in detail by NMR using xenon atoms as a spy.⁵³ Tobacco, wheat and maize LTP1 display the

largest cavities, with average volumes of 318 ± 36 Å³, 326 ± 51 Å³ and 290 ± 102 Å³, respectively. Rice LTP1 presents an intermediate cavity with an average volume of 236 ± 68 Å³ and barley LTP1 has the smallest cavity, with an average volume of 160 ± 33 Å³. However, the binding ability of LTP1 does not appear to be correlated with the cavity size of the molecules, since barley LTP1, with the smallest cavity, is able to accommodate a ligand as large as Acyl-CoA. In fact, binding ability depends on the plasticity of the cavity, which can expand considerably upon lipid binding, and on its accessibility.

AQ: 3

Dissimilarities appear between the various LTP1 types in their binding abilities. In NMR titration experiments of tobacco LTP1_1 with LMPC, saturation was reached at a lipid/protein ratio of 1.0, indicating that the protein can bind only one lipid (Fig. 4), in agreement with fluorescence data leading to saturation at a ligand/protein ratio of 1. LTP1_1 contains only two tyrosine residues at positions 17 and 79. Addition of LMPC induces a noticeable shift of Tyr79 resonances while Tyr17 is unaffected [Fig. 4(a)]. This result clearly indicates that the increase of tyrosine fluorescence observed during titration experiments with different acyl residues is due to Tyr79 environment changes. For maize and wheat LTP1, this saturation occurs at a lipid/protein ratio of 1.4, suggesting that the two proteins can bind either one or two lipids (see supplementary figures). Measurements on wheat LTP1 intrinsic tyrosine fluorescence have shown that saturation is reached at a lipid/protein ratio of 1.6, suggesting that two lipids can fit within the protein.⁴⁶ This is consistent with the X-ray structure of wheat LTP1, revealing the presence of two LMPC molecules in the core of the structure in a head-to-tail orientation.⁵⁴ Similarly, the recently determined X-ray structure of rice LTP1⁵⁵ reveals the presence of two palmitic acid molecules in a head-to-tail orientation. Moreover, it has been shown that maize LTP1 is able to bind two fluorescent-labeled fatty acids in its cavity with an affinity for the second ligand noticeably weaker than the affinity for the first one.⁵⁶ Fluorescence data also suggest that tobacco LTP1_1 shows a differential affinity towards lipidic ligands (Table II). LTP1 from monocot seeds, especially from wheat,⁵⁷ never differ in their affinity for lysophospholipids or fatty acids, with average *K_d* values ranging from 0.5 to 1 μM. In the case of tobacco LTP1, the same *K_d* was retrieved for LMPC whereas there is a five- to 10-fold increase for fatty acids.

**FS1,
FS2**
AQ: 3
AQ: 3
AQ: 3

Since the cavity of the tobacco protein is large enough to accommodate two mono-acylated lipids, the difference in the number of binding sites is presumably due to differences in the accessibility of the cavity. Examination of NMR and X-ray structures of LTP1 in complex with various lipids allows delineation of two possible entrances. The first is located between loop L2 and the Val75-Gln91 segment at the C-terminus, which acts as a cover for the protein cavity. In this case, the lipid insertion involves a movement of this segment toward the solvent and a rotation of the side chain of the conserved Tyr79 (tobacco numbering), which interacts with the lipid through the hydrogen bond via its OH group.⁵⁴ The second entrance,

which presumably prevails for the insertion of Acyl-CoA in the Acyl-CoA/barley LTP1 complex and for the insertion of the second LMPC molecule in wheat LTP1 in complex with two LMPC molecules, necessitates an opening of the cavity in the neighborhood of loops L1 and L3, and of the end of helices H1 and H3. In this case, the lipid insertion requires the disruption of the hydrophobic interactions that maintain the packing of helices H1 and H3. However, this organization differs from the Ace-AMP1 LTP1-like protein from onion seeds, in which the hydrophobic cavity is clearly obstructed by bulky residues,⁵⁸ leading to a molecule without any lipid-binding properties.

These hypotheses are supported by high temperature MD simulations on wheat LTP1, showing that very rapidly, the Leu77–Asn84 segment of the C-terminus deviates significantly from its experimental position, providing an obvious pathway for ligand entry.⁵⁹ Another significant conformational alteration occurs later in the dynamics trajectory (1 ns approximately), involving a large movement of loops L1 and L3 opening this second cavity entrance. The packing interactions in tobacco LTP1_1 do not differ drastically from those found in the other LTP1 structures, but differences appear in the organization of the side chains of the end of H1, of loop L1 and of the beginning of helix H4. In rice, maize, barley and wheat LTP1, there is a small hydrophobic patch including Tyr16, Pro23 and Ile69 (wheat numbering), with Pro23 pointing inward. In tobacco LTP1_1, there is a reorganization of the side chains due to the replacement of a serine and an asparagine, conserved over the cereal seeds LTP1, by a leucine (Leu24) and a lysine (Lys65) respectively. The L1 loop differs in the tobacco structure, with Leu24 pointing inwards and Pro23 partially exposed to the solvent. The ring of Tyr17 is enclosed between the side chains of Leu18, Leu24, the hydrophobic part of Lys65 (the ammonium group of Lys65 being exposed) and Leu69 (Fig. 6), leading to high field shifts of the Leu24 (δCH_3) and Lys65 (δCH_2) signals (see supplementary material). In fact, a continuous network of hydrophobic interactions among H1, L1, H3, L3 and the N-terminus of H4 (involving Leu11, Leu15, Tyr17, Leu18, Leu51, Ala54, Ile58, Ile61, Lys65, Ala66 and Leu69) is formed. Such strong hydrophobic interactions prevent the opening of the second pathway for lipid entrance and release in the case of tobacco LTP1_1. It should also be noted that the chemical shifts of protons in this region are not affected by lipid binding, as exemplified in Figure 4(b) by the Tyr17 and Leu24 signals. In contrast, the resonances of residues bordering the first entrance are much altered. For example, the drastic shift of the Arg39 signal (Fig. 5) is probably due to the position of this exposed amino acid directly in the middle of the assumed first entrance. This result reinforces our hypothesis of a pathway for lipid entrance between L2 and the coil around Ile81, which is further confirmed by preliminary high temperature molecular dynamics simulations on tobacco LTP1_1 (data not shown).

The comparison of the various cereal seed LTP1–lipid complexes show large differences in lipid orientation^{17,19,49–51,54,55,60} and in the position of the aliphatic

tails in the cavity.^{52,55} Several LTP1 complexes have already been studied by NMR^{19,49–51,60} and it appears that the chemical shift variations depend on the size of the lipid and also on the size of the cavity. In our case, analysis of chemical shift variations of the aliphatic protons indicates that the residues most perturbed by the lipid binding are residues bordering the cavity or neighboring these residues. They are located principally in the C-terminal segment but also in H3 and some in H1 and H2. This indicates that the lipid is inserted into the cavity with its large polar head protruding outside the first lipid entrance, accounting for the large variations of the induced shift Δ of Arg 39 and of the C-terminal residues. These NMR results, favoring an orientation identical to that found for the LPPC–maize LTP1 or palmitate–maize LTP1 complex,^{17,19} are fully compatible with our hypothesis on the lipid entrance.

CONCLUSIONS

The global fold of tobacco LTP1_1 is very close to the structure previously published for cereal seeds. However, local differences in the packing of hydrophobic residues may explain the differences in the binding properties of tobacco LTP1_1 from those of cereal seed LTP1. This is the first demonstration that LTP1 can bind only one monoacylated lipid and exhibit different binding properties. These considerations reveal that different types of LTP1 probably display different lipid binding abilities (number of sites, affinity and/or specificity), reflecting their biological function. In order to answer the interesting question of structure–function relationships and to understand how small changes in the amino acid composition and subtle structure modifications affect the activity of LTP1, more investigations are required. These include (i) further comparative experiments with different ligands varying in chain length, level of saturation, nature of hydrophilic head (net charge, space filling, ability to interact with exposed residues...); (ii) site-directed mutagenesis on residues involved in the second entrance and lipid titration experiments on the mutated recombinant proteins, (iii) 3D structures and binding properties of other LTPs, such as another tobacco LTP1⁶¹ or tomato LTP1,⁹ showing the same residues at the level of the second entrance. From this intricate puzzle, further description of ‘the right protein in the right complexation state able to trigger a given cell response’ constitutes a very attractive challenge. In this work, we present the first analysis of LTP1_1 ¹H–¹⁵N chemical shift variations upon lipid binding. This will constitute a very helpful tool to investigate the mechanism of lipid binding by these plant extracellular proteins. For example, comparing the modifications induced by linolenic and jasmonic acids should be of particular interest, since it was recently demonstrated that the jasmonate complex (and not the linolenate) is able to activate plant protection against pathogens.¹

One of the research avenues of interest is that elicitor high-affinity sites located on the plasma membrane,⁶² which have been shown to trigger HR and SAR,²² can also bind wheat LTP1 and tobacco LTP1_1.^{1,12} The elicitor

F6

AQ: 15

AQ: 3

AQ: 3 receptor should be studied as an LTP receptor,¹¹ suggesting these proteins' abilities for cell-to-cell (local) and tissue-to-tissue (systemic) signaling.⁶³ This function fits well with the ubiquity of LTPs in the plant kingdom, as with the description of elicitor receptors in all plant plasma membranes tested up to now.¹³

ACKNOWLEDGMENTS

We wish to thank Corinne Buré for her technical support in Mass Spectroscopy.

REFERENCES

- Buhot N, Gomes E, Milat ML, Ponchet M, Lequeu J, Delrot S, Coutos-Thevenot P, Blein JP. A jamaonic acid-lipid transfer protein complex binds to a plasma membrane receptor and specifically induces protection against *Phytophthora parasitica* in tobacco. *Mol Biol Cell* 2004; In press.
- Kader JC. Lipid-transfer proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* 1996;47:627–654.
- Kader JC. Lipid-transfer proteins: a puzzling family of plant proteins. *Trends in plant science* 1997;2(2):66–70.
- Aronel V, Vergnolle C, Cantrel C, Kader JC. Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. *Plant Science* 2000(157):1–12.
- Sterk P, Booi H, Schellekens GA, Van Kammen A, De Vries SC. Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* 1991;3(9):907–921.
- Pyee J, Yu H, Kolattukudy PE. Identification of a lipid transfer protein as the major protein in the surface wax of broccoli (*Brassica oleracea*) leaves. *Arch Biochem Biophys* 1994;311(2):460–468.
- Garcia-Olmedo F, Molina A, Segura A, Moreno M. The defensive role of nonspecific lipid-transfer proteins in plants. *Trends Microbiol* 1995;3(2):72–74.
- Keresztessy Z, Hughes MA. Homology modeling and molecular dynamics aided analysis of ligand complexes demonstrates functional properties of lipid-transfer proteins encoded by the barley low-temperature-inducible gene family, blt4. *Plant J* 1998;14(5):523–533.
- Torres-Schumann S, Godoy JA, Pintor-Toro JA. A probable lipid transfer protein gene is induced by NaCl in stems of tomato plants. *Plant Mol Biol* 1992;18(4):749–757.
- Souferi IA, Vergnolle C, Miginiac E, Kader JC. Germination-specific lipid transfer protein cDNAs in *Brassica napus* L. *Planta* 1996;199:229–237.
- Blein JP, Coutos-Thevenot P, Marion D, Ponchet M. From elicitors to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. *Trends Plant Sci* 2002;7(7):293–296.
- Buhot N, Douliet JP, Jacquemard A, Marion D, Tran V, Maume BF, Milat ML, Ponchet M, Mikes V, Kader JC, Blein JP. A lipid transfer protein binds to a receptor involved in the control of plant defence responses. *FEBS Lett* 2001;509(1):27–30.
- Ponchet M, Panabieres F, Milat ML, Mikes V, Montillet JL, Suty L, Triantaphylides C, Tirilly Y, Blein JP. Are elicitors cryptograms in plant-Oomycete communication? *Cell Mol Life Sci* 1999;56:1020–1047.
- Fleming AJ, Mandel T, Hofmann S, Sterk P, de Vries SC, Kuhlemeier C. Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. *Plant J* 1992;2(6):855–862.
- Canevascini S, Caderas D, Mandel T, Fleming AJ, Dupuis I, Kuhlemeier C. Tissue-specific expression and promoter analysis of the tobacco Ltp1 gene. *Plant Physiol* 1996;112:513–524.
- Petit MC, Sodano P, Marion D, Ptak M. Two-dimensional 1H-NMR studies of maize lipid-transfer protein. Sequence-specific assignment and secondary structure. *Eur J Biochem* 1994;222(3):1047–1054.
- Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* 1995;3(2):189–199.
- Heinemann B, Andersen KV, Nielsen PR, Bech LM, Poulsen FM. Structure in solution of a four-helix lipid binding protein. *Protein Sci* 1996;5(1):13–23.
- Gomar J, Petit MC, Sodano P, Sy D, Marion D, Kader JC, Vovelle F, Ptak M. Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. *Protein Sci* 1996;5(4):565–577.
- Lee JY, Min K, Cha H, Shin DH, Hwang KY, Suh SW. Rice non-specific lipid transfer protein: the 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity. *J Mol Biol* 1998;276(2):437–448.
- Poznanski J, Sodano P, Suh SW, Lee JY, Ptak M, Vovelle F. Solution structure of a lipid transfer protein extracted from rice seeds. Comparison with homologous proteins. *Eur J Biochem* 1999;259(3):692–708.
- Osman H, Vauthrin S, Mikes V, Milat ML, Panabieres F, Marais A, Brunie S, Maume B, Ponchet M, Blein JP. Mediation of elicitor activity on tobacco is assumed by elicitor-sterol complexes. *Mol Biol Cell* 2001;12(9):2825–2834.
- Kay LE, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 1992;114:10663–10665.
- Piantini U, Sorensen OW, Ernst RR. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc* 1982;104:6800–6801.
- Bax A, Davis DG. MLEV-17-Based two-dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson* 1985;65:355–360.
- Griesinger C, Otting G, Wüthrich K, and Ernst R. R. Clean TOCSY for 1H spin system identification in macromolecules. *J Am Chem Soc* 1988;110:7870–7872.
- Jeener J, Meier B. H., Bachmann P., and Ernst R. R. Investigation of exchange processes by two dimensional NMR spectroscopy. *J Chem Phys* 1979;71:4546–4553.
- Zhang O, Kay LE, Olivier JP, Forman-Kay JD. Backbone ¹H and ¹⁵N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J Biomol NMR* 1994;4(6):845–858.
- Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(H^NH^α) coupling constants in ¹⁵N-enriched proteins 1993;115:7772–7777.
- Karplus M. Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys* 1959;30:11–15.
- Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of ¹H, ¹³C, and ¹⁵N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 1990;29(19):4659–4667.
- Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. A suite of triple resonance NMR experiments for backbone assignment of ¹⁵N, ¹³C, ²H labeled proteins with high sensitivity. *J Am Chem Soc* 1994;116:11655–11666.
- Wittekind M, Mueller L. HNCACB, a high-sensitivity 3D NMR experiments to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson* 1993;B101:201–205.
- Grzesiek S, Bax A. Correlating backbone amide and side chain resonance in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc* 1992;114:6291–6293.
- John BK, Plant D, Hurd RE. Improved proton-detected heteronuclear correlation using gradient-enhanced z and zz filters. *J Magn Reson* 1992;A101:113–117.
- Bax A, Clore GM, Gronenborn AM. ¹H-¹H correlation via isotropic mixing of ¹³C magnetisation, a new three-tridimensional approach for assigning ¹H and ¹³C spectra of ¹³C-enriched proteins. *J Magn Reson* 1990;88:425–431.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6(3):277–293.
- Johnson BA, Blevins RA. NMRview: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 1994;4:603–614.
- Linge JP, O'Donoghue SI, Nilges M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 2001;339:71–90.
- Brünger AT, Adams PD, Clore GM, De Lano WL, Gros P, Grosse-Kuntzele RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular struc-

- ture determination. *Acta Crystallogr D Biol Crystallogr* 1998; 54(Pt 5):905–921.
41. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. *Proteins* 2003;50(3):496–506.
 42. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996;14(1):51–55.
 43. Hutchinson EG, Thornton JM. PROMOTIF a program to identify and analyze structural motifs in proteins. *Protein Sci* 1996;5(2): 212–220.
 44. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8(4):477–486.
 45. Gomar J, Sodano P, Sy D, Shin DH, Lee JY, Suh SW, Marion D, Vovelle F, Ptak M. Comparison of solution and crystal structures of maize nonspecific lipid transfer protein: a model for a potential in vivo lipid carrier protein. *Proteins* 1998;31(2):160–171.
 46. Douliez JP, Jegou S, Pato C, Molle D, Tran V, Marion D. Binding of two mono-acylated lipid monomers by the barley lipid transfer protein, LTP1, as viewed by fluorescence, isothermal titration calorimetry and molecular modelling. *Eur J Biochem* 2001;268(2): 384–388.
 47. Guo RT, Chou LJ, Chen YC, Chen CY, Pari K, Jen CJ, Lo SJ, Huang SL, Lee CY, Chang TW, Chaung WJ. Expression in *Pichia pastoris* and characterization by circular dichroism and NMR of rhodostomin. *Proteins* 2001;43(4):499–508.
 48. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH—a hierarchic classification of protein domain structures. *Structure* 1997;5(8):1093–1108.
 49. Lerche MH, Kragelund BB, Bech LM, Poulsen FM. Barley lipid-transfer protein complexed with palmitoyl CoA: the structure reveals a hydrophobic binding site that can expand to fit both large and small lipid-like ligands. *Structure* 1997;5(2):291–306.
 50. Tassin-Moindrot S, Caille A, Douliez JP, Marion D, Vovelle F. The wide binding properties of a wheat nonspecific lipid transfer protein. Solution structure of a complex with prostaglandin B2. *Eur J Biochem* 2000;267(4):1117–1124.
 51. Sodano P, Caille A, Sy D, de Person G, Marion D, Ptak M. ¹H NMR and fluorescence studies of the complexation of DMPG by wheat non-specific lipid transfer protein. Global fold of the complex. *FEBS Lett* 1997;416(2):130–134.
 52. Han GW, Lee JY, Song HK, Chang C, Min K, Moon J, Shin DH, Kopka ML, Sawaya MR, Yuan HS, Kim TD, Choe J, Lim D, Moon HJ, Suh SW. Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution x-ray crystallography. *J Mol Biol* 2001;308:263–278.
 53. Dubois L, Da Silva P, Landon C, J. HG, Ponchet M, Vovelle F, Berthault P, Desvaux H. Probing hydrophobic cavities of Lipid Transfer Protein from *Nicotiana tabacum* through Xenon-based NMR. submitted 2004.
 54. Charvolin D, Douliez JP, Marion D, Cohen-Addad C, Pebay-Peyroula E. The crystal structure of a wheat nonspecific lipid transfer protein (ns-LTP1) complexed with two molecules of phospholipid at 2.1 Å resolution. *Eur J Biochem* 1999;264(2):562–568.
 55. Cheng HC, Cheng PT, Peng P, Lyu PC, Sun YJ. Lipid binding in rice nonspecific lipid transfer protein-1 complexes from *Oryza sativa*. *Protein Sci* 2004;13(9):2304–2315.
 56. Zachowski A, Guerbette F, Grosbois M, Jolliot-Croquin A, Kader JC. Characterisation of acyl binding by a plant lipid-transfer protein. *Eur J Biochem* 1998;257(2):443–448.
 57. Douliez JP, Michon T, Marion D. Steady-state tyrosine fluorescence to study the lipid-binding properties of a wheat non-specific lipid-transfer protein (nsLTP1). *Biochim Biophys Acta* 2000; 1467(1):65–72.
 58. Tassin S, Broekaert WF, Marion D, Acland DP, Ptak M, Vovelle F, 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;37(11): 3623–3637.
 59. Sy D, Le Gravier Y, Goodfellow J, Vovelle F. Protein stability and plasticity of the hydrophobic cavity in wheat ns-LTP. *J Biomol Struct Dyn* 2003;21(1):15–29.
 60. Lerche MH, Poulsen FM. Solution structure of barley lipid transfer protein complexed with palmitate. Two different binding modes of palmitate in the homologous maize and barley nonspecific lipid transfer proteins. *Protein Sci* 1998;7(12):2490–2498.
 61. Masuta C, Furuno M, Tanaka H, Yamada M, Koiwai A. Molecular cloning of a cDNA clone for tobacco lipid transfer protein and expression of the functional protein in *Escherichia coli*. *FEBS Lett* 1992;311(2):119–123.
 62. Bourque S, Binet MN, Ponchet M, Pugin A, Lebrun-Garcia A. Characterization of the cryptogin binding sites on plant plasma membranes. *J Biol Chem* 1999;274(49):34699–34705.
 63. Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK. A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. *Nature* 2002;419(6905):399–403.

Author Proof

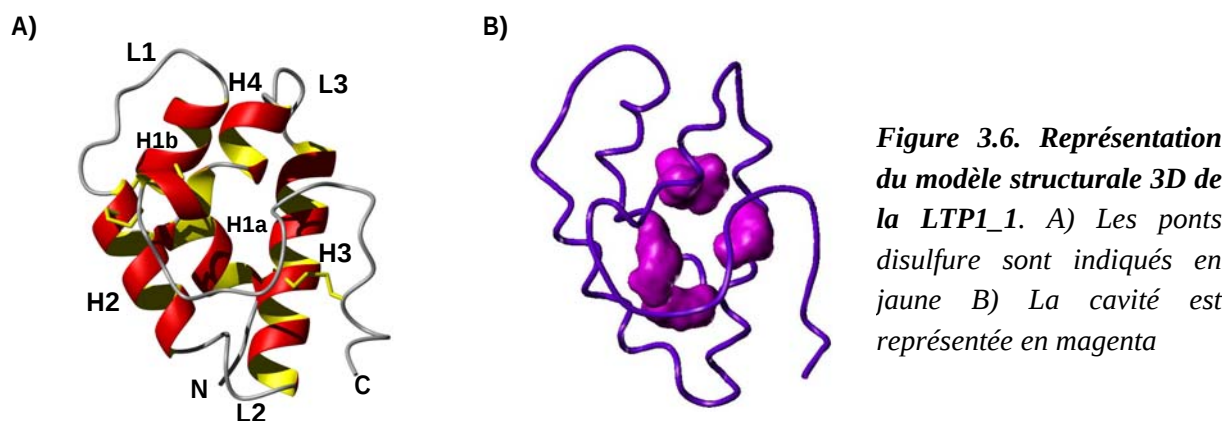
Cette première étude structurale d'une LTP1_1 exprimée dans les feuilles de tabac nous éclaire sur les différences de propriétés de fixation existant entre les LTPs d'origines diverses qui sont probablement liées à des fonctions différentes au sein de la plante. L'étude de LTP1, présentant les mêmes mutations que la LTP1_1 de tabac confirmerait l'importance des résidus que nous avons déterminés comme responsables des propriétés de complexation originale de la LTP1_1 de tabac. Des expériences de mutagenèse dirigées de ces résidus prouveraient également leur importance. En effet, comme nous l'avons vu pour la LTP1_1 de tabac, de petites modifications dans la composition en acides aminés entraînent une modification des propriétés de fixation et peut-être de fonction biologique. D'autre part, une étude récente de différents complexes entre la LTP1_1 de tabac et différents ligands de nature et de taille variées (Buhot et al. 2004) a révélé qu'un complexe entre la LTP1_1 de tabac et une molécule de jasmonate provoque une réponse biologique dans le tabac alors qu'un complexe avec une molécule de linoléate n'entraîne aucune réponse. Une étude structurale comparative de ces deux complexes fournirait peut-être des éléments de réponse dans notre quête d'identification des déterminants structuraux qui confèrent à une LTP donnée une fonction donnée.

3. Modélisation comparative des LTP1 identifiées dans le tabac

Afin d'identifier les caractéristiques structurales des quatre autres LTP1 identifiées dans le tabac, nous avons entrepris de déterminer leurs structures 3D par modélisation comparative.

3.1 Modèle structural

Nous avons déterminé par RMN et modélisation moléculaire la structure 3D de la LTP1_1 de tabac. Nous avons donc la chance de disposer, d'un modèle structural qui va être utilisé pour la détermination des structures 3D des quatre autres LTP1 de tabac. Parmi l'ensemble des structures 3D représentatives de la structure de la LTP1_1 en solution, nous avons choisi comme modèle une structure possédant une cavité de taille moyenne (environ 320 Å³) (figure 3.6).



3.2 Alignement multiple

L'alignement de la séquence modèle de la LTP1_1 avec celles des quatre autres LTP1 a été obtenu en réalisant un alignement multiple de séquences à l'aide du logiciel CLUSTALW (figure 3.7).

	10	20	30	40	50	60		
LTP1_1	-AITCGQVTSN	LAPCLAYLR---	NTGPLGR-	CCGGVKALVNS	ARTTEDRQIA	CTCLKS	AAG	ISGINLG
LTP1_2	-ALSCGQVQSG	LAPCLPYLQ---	GRGPLGS-	CCGGVKGLLGA	AKSPDRKTAC	TLKSA	ANA	IKGIDMG
LTP1_3	APPSCPTVTTQ	LAPCLSYIQ---	GGDPSVP	CCTGINNIYEL	AKTKEDRVAI	CNCLK	TAF	HAGNVNPT
LTP1_4	APPSCQTVTTQ	LAPCLSYIQNRV	KGGNPSVP	CCTGINNIYEL	AKTKEDRVAI	CNCLK	NAF	IAHAGNVNPT
LTP1_5	-AISCGQVVTRL	SPCINIVR---	QGALPAP	CCSGIKALNNQ	ATTTEDRQAAC	NCIKS	AAS	ISGINFN
Cons	C V L PC Y	G	CC G	A	DR	C C K A		
	70	80	90				% identité	similarité
LTP1_1	KAAGLPSTCGVNI	IPYK-ISPSTDC	SKVQ--	91				
LTP1_2	KAAGLPGACGVNI	IPCK-ISPSTDC	SKVQ--	91			72	83
LTP1_3	LVAQLPKKCGISF	NMPPIDKNYDC	NTISMY	96			38	54
LTP1_4	LVAELPKKCGISF	NMPPIDKNYDC	NTISMY	100			36	52
LTP1_5	LAGSLPSKCGVNL	IPYK-ISPSIDC	STVQ--	92			63	72
Cons	LP CG	I DC						

Figure 3.7: Alignement multiple des séquences des 5 LTP1 réalisé avec le logiciel CLUSTALW. Les acides aminés strictement conservés entre les séquences sont indiqués dans la séquence consensus (Cons) et coloriés: les cystéines en rouge, les glycines en gris, les acides aminés hydrophobes en vert, ceux chargés en bleu.

Cet alignement multiple révèle une séquence consensus de résidus strictement conservés qui est proche de celle obtenue pour l'ensemble des LTP1, avec quelques mutations conservatives de résidus hydrophobes (§ 2.2 figure 3 article). On remarque également dans la famille des LTP1 de tabac, l'existence de deux sous-familles composées, pour la première, des LTP1_1, LTP1_2 et LTP1_5, et pour la deuxième, des LTP1_3 et LTP1_4. En effet, la LTP1_1 possède un taux d'identité de séquence élevé avec les LTP1_2 et LTP1_5 : 72 et 63 % respectivement, alors que pour les LTP1_3 et LTP1_4 le taux d'identité n'excède pas 40 % (38 et 36 % respectivement). Si on s'intéresse maintenant uniquement au taux de similarité de la LTP1_1 avec les autres LTP1, celui-ci est supérieur à 50 % quelle que soit la LTP1 considérée et permet d'envisager avec confiance l'utilisation de la modélisation comparative.

3.3 Les structures 3D modèles

Les structures 3D modèles des LTP1 de tabac ont été construites par modélisation comparative sous contraintes géométriques à l'aide du logiciel MODELLER et évaluées comme décrit dans le chapitre Matériels et Méthodes.

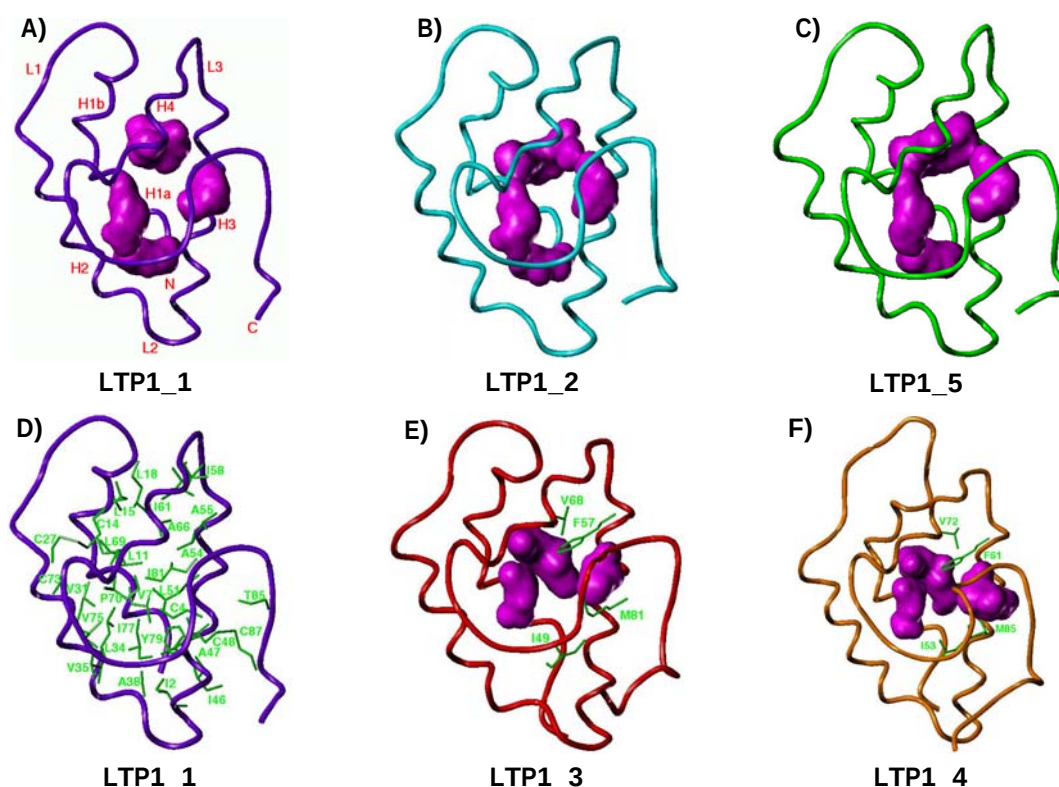


Figure 3.8 : Représentation des structures 3D de la famille des LTP1 de tabac. A, B, C) structures 3D des LTP1_1, 2 et 5 avec la cavité représentée en magenta. D) structure 3D de la LTP1_1 avec les acides aminés qui bordent la cavité (en vert). E et F) structures 3D des LTP1_3 et 4 avec leurs cavités en magenta et en vert les acides aminés qui modifient la taille et l'accessibilité de la cavité comparativement à celles des LTP1_1, 2 et 5

Les quatre LTP1 modélisées possèdent le même repliement que celui de la LTP1_1 (figure 3.8) : 4 hélices α enroulées en spirale droite et suivies par un long segment C-terminal. La différence structurale la plus notable s'observe au niveau de la boucle L1 de la structure 3D de la LTP1_4. En effet, l'insertion de 4 acides aminés à ce niveau (figure 3.2) provoque un allongement de cette boucle sans pour autant que le reste de la structure soit affecté.

Les résidus qui bordent la cavité hydrophobe de la LTP1_1 sont pour la plupart conservés dans les quatre LTP1 (figure 3.9). Les LTP1_2 et 5 possèdent une cavité de taille comparable à celle de la LTP1_1 (de l'ordre de 300 Å³). Par contre, la cavité des LTP1_3 et 4 est deux fois plus petite car des résidus hydrophobes encombrants (I49, F57 et V68, numérotation de la LTP1_3) remplacent des alanines, réduisant ainsi le volume de la cavité (figure 3.9).

On peut également noter que la tyrosine 79 connue pour intervenir dans la fixation et l'insertion des lipides à une entrée de la cavité des LTPs, est remplacée dans la LTP1_2 par une cystéine et par une méthionine dans les LTP1_3 et 4. Toutefois une phénylalanine (Phe79 et Phe84 dans les LTP1_3 et 4 respectivement) vient, comme la tyrosine 79 des autres LTP1 de structure connue, obstruer l'entrée de la cavité. Alors que pour la LTP1_2 la cystéine qui remplace la tyrosine 79 est un résidu moins encombrant qui libère l'entrée de la cavité.

	10	20	30	40	50	60	
LTP1_1	-AITCGQVTSN	LAPCLAYLR---	NTGPLGR-	CCGGVKALVNS	ARTTEDRQIA	CTCLKSAAGA	ISGINLG
LTP1_2	-ALSCGQVQSG	LAPCLPYLQ---	GRGPLGS-	CCGGVKGLLGA	AKSPDRKTA	CTCLKSAANA	IKGIDMG
LTP1_3	APPSCTVT	TQLAPCLSYIQ---	GGGDPSVP	CCTGINNIYEL	AKTKEDRVA	ICNCLKTAFT	HAGNVNPT
LTP1_4	APPSCTVT	TQLAPCLSYIQNRV	KGGGNPSVP	CCTGINNIYEL	AKTKEDRVA	ICNCLKNAFI	HAGNVNPT
LTP1_5	-AISCQVVTRL	LSPCINIVR---	QGGALPAP	CCSGIKALNNQ	ATTTEDRQA	ACNCIKSAA	ASISGINFN
	70	80	90				
LTP1_1	KAAGLPSTCGVNI	PK-ISPSTDCSKVQ	--	91			
LTP1_2	KAAGLPGACGVNI	PK-ISPSTDCSKVQ	--	91			
LTP1_3	LVAQLPKKCGISF	NMPPIDKNYDCNT	ISMY	96			
LTP1_4	LVAELPKKCGISF	NMPPIDKNYDCNT	ISMY	100			
LTP1_5	LAGSLPSKCGVNL	PK-ISPSTDCSTVQ	--	92			

Figure 3.9 : Alignement multiple de séquences des 5 LTP1 de tabac. Les acides aminés qui bordent la cavité sont indiqués en vert, ceux qui modifient la taille et l'accessibilité des cavités des LTP1_3 et 4 sont encadrés en vert. Ceux encadrés en bleu et en orange sont les résidus contribuant au pôle hydrophile mutés en résidus hydrophobes dans les LTP1_3 et 4.

L'étude des propriétés de complexation de la LTP1_1 a montré qu'elle n'était capable de fixer qu'un seul phospholipide alors que les LTP1 de graines de céréales en fixent deux. L'analyse des structures a révélé (§ 2.2 article) que cette différence provenait de l'obstruction par un amas hydrophobe de la seconde entrée de la cavité observée dans les LTP1 de graines

de céréales. Ce cluster hydrophobe propre à la LTP1_1 est constitué des résidus Tyr17, Leu18, Leu24, Leu69 et de la partie hydrophobe de la chaîne latérale de Lys65. Dans toutes les LTP1 de tabac, ces résidus sont strictement conservés ou mutés par des résidus homologues. On peut même noter que la Lys65 est remplacée par une leucine dans les LTP1_3 et 4, conduisant à l'existence d'un cluster hydrophobe plus important pour ces deux dernières LTP1.

Nous avons déterminé les potentiels de lipophilie des 5 LTP1 de tabac (figure 3.10). Les LTP1_1, 2 et 5 présentent un pôle hydrophile définis par les résidus Glu42, Asp43, Arg44, Gln45, Asp86 et Lys89 (résidus et numérotation pour la LTP1_1) alors que les LTP1_3 et 4 n'en possèdent pas. En effet, dans les LTP1_3 et 4 certains résidus hydrophiles présents dans les LTP1_1, 2 et 5 sont remplacés par des résidus hydrophobes (figure 3.9). De plus, ces deux protéines possèdent une extension C-terminale de deux résidus hydrophobes Met95 et Tyr96 (numérotation de la LTP1_3). Tous ces facteurs contribuent à une diminution importante de l'hydrophilie de ce pôle dans ces deux LTP1 (figures 3.9, 3.10). Cette absence de pôle hydrophile des LTP1_3 et 4 suggère un mode de fixation de lipides différent des autres LTP1. En effet, ce pôle hydrophile se situe au niveau de la boucle L2 et de la partie C-terminale, précisément dans la région identifiée comme l'entrée présumée des phospholipides dans la cavité de la LTP1_1 de tabac (§ 2.2 article).

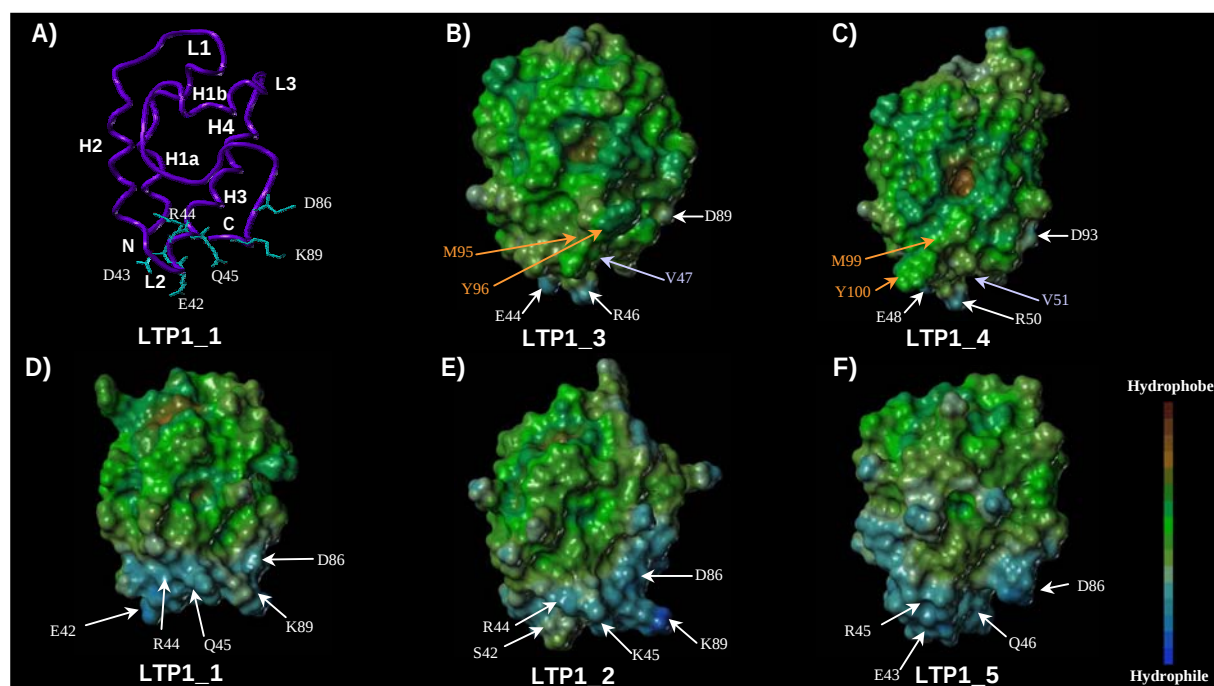


Figure 3.10 : représentation des potentiels hydrophobes des LTP1 de tabac. A, orientation et structure 3D de la LTP1_1 avec, annotés, les acides aminés définissant son pôle hydrophile. B, C, D, E, et F, illustration des potentiels hydrophobes des LTP1_2, 5, 1, 3 et 4 avec annotés les acides aminés intervenant dans le pôle hydrophile de LTP1_1 ou leur équivalent dans les autres séquences.

3.4 Conclusion et perspectives

Comme on pouvait s'y attendre, compte tenu du fort taux de similarité de séquence entre le modèle et les protéines cibles, le repliement global des 4 autres LTP1 de tabac est voisin de celui du modèle LTP1_1. Nous avons remarqué pour les quatre LTP1 modélisées que les résidus hydrophobes tapissant la cavité sont conservés. Cependant, dans les LTP1_3 et 4 certains résidus hydrophobes encombrants (isoleucine, phénylalanine, valine) remplacent des alanines, modifiant ainsi la taille de la cavité et son accessibilité. De plus, la tyrosine 79, qui intervient dans la complexation dans la LTP1_1 est remplacée par une méthionine dans les LTP1_3 et 4 et par une cystéine dans la LTP1_2, traduisant peut-être un mode de complexation différent. Enfin, nous avons observé une répartition des résidus hydrophiles et hydrophobes différente dans les LTP1_3 et 4 où le pôle hydrophile qui est présent dans les LTP1_1, 2 et 5 a disparu, suggérant peut-être une fonction différente.

Dans ce cas particulier des LTP1_3 et 4, bien que le taux de similarité de séquence soit important avec la LTP1_1 et les autres LTP1, on peut se demander si on ne se trouve pas dans un cas limite de l'utilisation de la modélisation comparative. En particulier, dans la région C-terminale où la présence d'un doublet de prolines spécifique des LTP1_3 et 4 peut considérablement modifier la structure de la boucle. La présence de nombreux résidus aromatiques (4 Tyr et 2 Phe) peut aussi fortement affecter la taille de la cavité. Dans nos modèles, seule la Phe79 vient bloquer l'entrée de la cavité, sans pour autant y pénétrer, mais il n'est pas impossible que d'autres résidus aromatiques viennent en réalité s'y insérer. La modélisation comparative ne permet pas encore de définir avec précision la conformation des chaînes latérales. Dans ce cas, on pourrait imaginer que la cavité est complètement obstruée, comme dans la protéine de graines d'oignon *Ace-AMP1* qui ne fixe pas les lipides.

Pour aller plus loin dans ce travail et afin de confirmer ou infirmer les prédictions structurales que nous avons obtenues par modélisation comparative, il serait intéressant d'étudier plus en détails ces quatre LTP1 de tabac.

D'après nos résultats, il apparaît, d'une part, que les LTP1_2 et 5 sont proches sur le plan structural de la LTP1_1. L'étude de leur mode de complexation pourrait donc confirmer l'existence d'une sous famille de LTP1 capable de ne lier qu'un seul lipide. Nous pourrions ainsi identifier plus en détail les déterminants structuraux participant à la fixation des lipides. D'autre part, les LTP1_3 et 4 apparaissent comme une nouvelle sous-famille structurale de LTP1. En effet, d'après nos résultats, ces LTP1 possèdent une cavité de taille réduite et pas

de pôle hydrophile, suggérant un mode de fixation différent, et peut-être même une impossibilité de complexation. Une étude structurale couplée à des études des propriétés de complexation de l'une de ces protéines pourrait nous aider dans la compréhension de la fonction de cette famille de protéines.

IV) Dynamique interne de la LTP1_1 de tabac

Le marquage en ^{15}N de la LTP1_1 de tabac nous a permis de réaliser, pour la première fois, l'étude par RMN de la relaxation des noyaux ^{15}N d'une protéine de la famille des LTP1 nous donnant ainsi accès aux propriétés dynamiques de son squelette polypeptidique. Afin de compléter cette étude et comparer nos données expérimentales à celles obtenues par le calcul, nous avons effectué des simulations de dynamique moléculaire sur plusieurs nanosecondes. Les résultats de ces études réalisées expérimentalement et « *in silico* » sur la LTP1_1 de tabac sont exposés successivement dans ce chapitre. Ils nous ont éclairés sur les régions de la LTP1_1 vraisemblablement impliquées dans la fixation des molécules lipidiques.

1. Etude par RMN de la dynamique interne de la LTP1_1 de tabac

1.1 Détermination des paramètres de relaxation du noyau ^{15}N

L'étude par RMN de la dynamique de la LTP1_1 a été réalisée sur un échantillon uniformément marqué ^{15}N présentant une concentration de 1,8 mM et un pH de 5,2. Les expériences RMN ont été enregistrées sur un spectromètre Inova Varian à 600 MHz, à une température de 293 K. Le traitement et l'analyse des données ont été effectués à l'aide des logiciels NMRPipe (Delaglio et al. 1995) et NMRView (Johnson and Blevins 1994).

1.1.1 Mesures des vitesses de relaxation longitudinale R_1 et transversale R_2

Afin de déterminer les valeurs des vitesses de relaxation longitudinale R_1 et transversale R_2 , une série d'expériences de type HSQC a été enregistrée en faisant varier le délai tr (cf. Matériels et Méthodes). Douze valeurs de ce délai ont été choisies afin d'échantillonner au mieux les courbes de relaxation correspondant à R_1 et à R_2 . Pour R_1 , il s'agit de $tr = 10, 60, 120, 180, 240, 300, 360, 480, 600, 900, 1200$ et 1500 ms. Pour R_2 plus rapide, $tr = 10, 30, 50, 70, 90, 110, 130, 150, 170, 190, 210$ et 230 ms.

Afin d'estimer l'erreur sur la mesure de l'intensité des pics, le spectre avec un délai tr égal à 240 ms a été réalisé deux fois pour R_1 et les spectres acquis avec un délai tr de 10 et 50 ms ont également été doublés pour R_2 . Les figures 4.1 a, b et c représentent quelques-uns des spectres réalisés pour déterminer les valeurs des vitesses de relaxation R_1 . Celles-ci, comme les vitesses de relaxation R_2 , ont été extraites en réalisant un ajustement de l'intensité des pics croisés à l'aide d'une fonction monoexponentielle décroissante, comme nous l'avons décrit dans le chapitre Matériels et Méthodes. Les figures 4.2 a et b illustrent respectivement

l'ajustement des intensités des pics croisés extraites des spectres de mesure de R_1 et de R_2 pour les résidus Ala16 et Ile81.

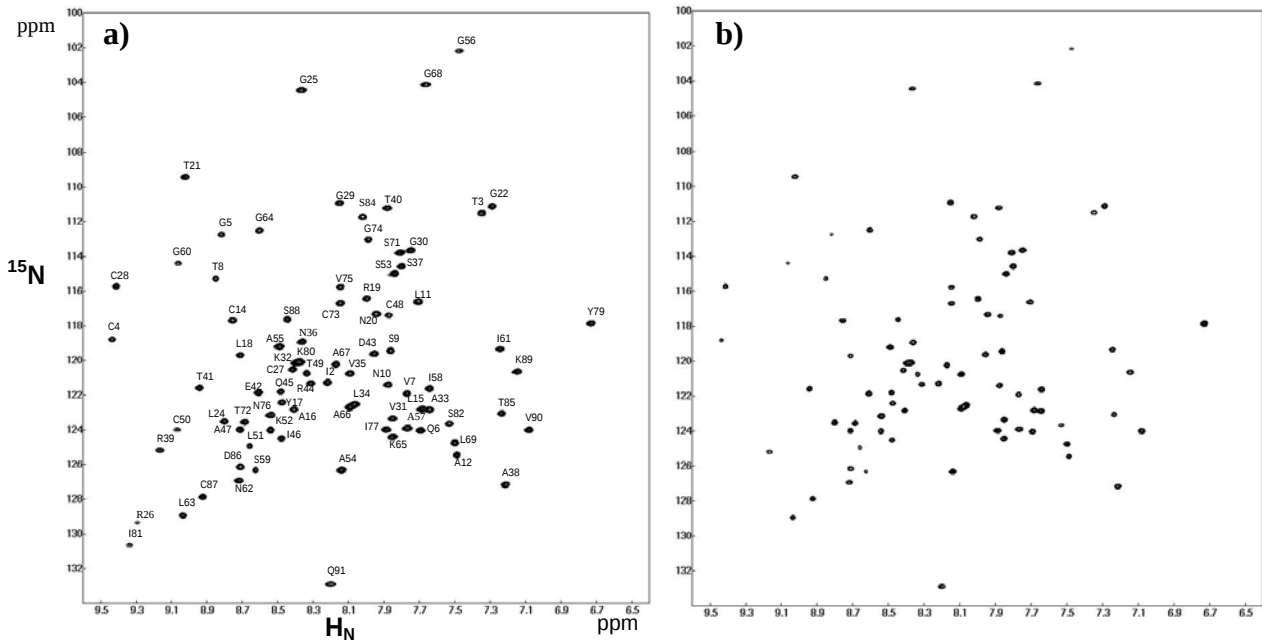


Figure 4.1 : Exemples de spectres réalisés pour déterminer les vitesses de relaxation R_1 . Les temps de relaxation tr correspondant aux spectres a, b, c ont pour valeurs respectives 10, 240, et 600 ms. L'intensité des pics croisés décroît au fur et à mesure que le temps de relaxation tr augmente.

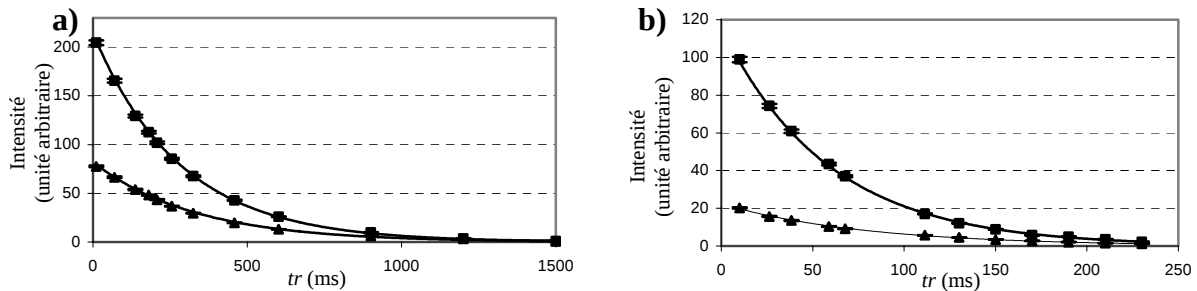


Figure 4.2 : Ajustements monoexponentiels des intensités des pics croisés extraites de la série de spectres de mesure de R_1 (a) ou de R_2 (b). Cas des résidus Ala16 (■) et Ile81 (▲). Les barres d'erreurs sont indiquées et leurs tailles sont inférieures à celles des motifs ■ et ▲.

Les valeurs des vitesses de relaxation longitudinale R_1 et transversale R_2 ont été mesurées pour 83 résidus et sont réunies dans la figure 4.3. Pour des raisons de superpositions de signaux, ces paramètres de relaxation n'ont pas pu être obtenus pour les lysines 32 et 80.

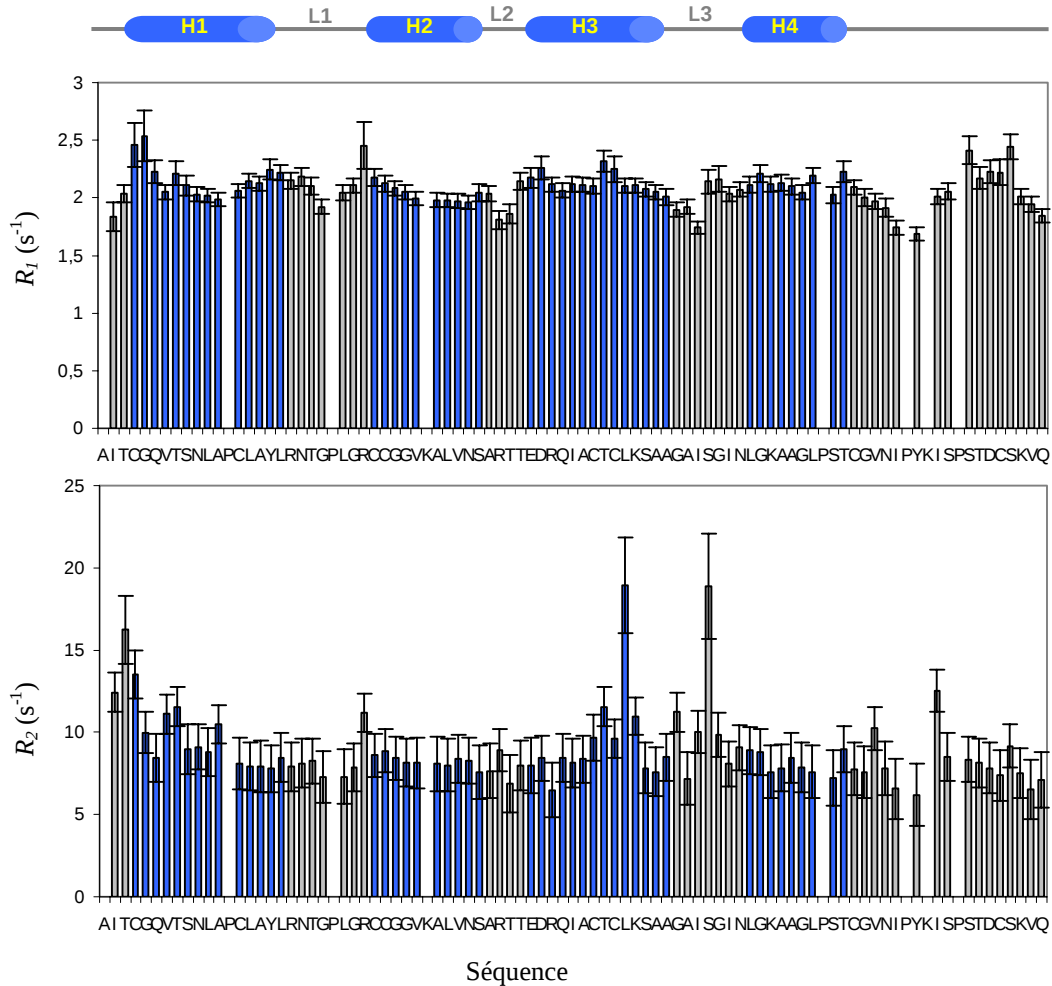


Figure 4.3 : Valeurs des vitesses de relaxation longitudinale R_1 et transversale R_2 des noyaux ^{15}N du squelette le long de la séquence de la LTP1_1. Les éléments de structures secondaires sont schématisés en haut de la figure. Les hélices sont représentées par des cylindres. Les paramètres de relaxation concernant les résidus impliqués dans ces structures secondaires sont coloriés en gris foncé.

La valeur moyenne de la vitesse de relaxation R_1 des noyaux ^{15}N de la LTP1_1 de tabac est de $2,09 \pm 0,15 \text{ s}^{-1}$. La valeur de R_1 est homogène tout au long de la séquence ; en effet, très peu de résidus présentent une valeur supérieure ou inférieure de plus d'un écart type à celle de la moyenne. On peut toutefois remarquer que les résidus Cys4, Gly5 et Arg26, Ser84 et Ser88 présentent une valeur de R_1 importante ($> 2,4 \text{ Hz}$). Ces résidus se situent en positions N- et C- terminales ou au niveau de la boucle L1 pour l'arginine 26, c'est-à-dire

dans des régions de la protéine très exposées au solvant, ce qui explique des valeurs de R_1 élevées.

La valeur moyenne des vitesses de relaxation transversale R_2 est de $7,76 \pm 2,26 \text{ s}^{-1}$. Tout au long de la séquence, les valeurs de R_2 oscillent faiblement autour de la valeur moyenne. En effet, la majorité des résidus (70 sur 83) présentent des valeurs de R_2 comprises entre 5 et 10 Hz. Ces dernières ne varient donc pas de plus d'un écart type par rapport à la moyenne. Six résidus (Ile2, Thr3, Cys4, Leu51, Ser59, Ile81) possèdent cependant une valeur de R_2 significativement élevée ($> 12 \text{ Hz}$). Ces résidus pourraient être impliqués dans un phénomène d'échange chimique et avoir de ce fait un terme de relaxation supplémentaire.

1.1.2 Mesure de l'effet NOE hétéronucléaire ^1H - ^{15}N .

Pour la mesure de l'effet NOE hétéronucléaire ^1H - ^{15}N , le délai d'irradiation des ^1H a été fixé à 5 s, soit environ 10 fois la valeur moyenne du T_1 des noyaux ^{15}N mesuré pour la LTP1_1 de tabac. Les deux spectres correspondant aux deux expériences avec et sans irradiation des protons sont illustrés par la figures 4.4. Comme pour R_1 et R_2 , les mesures ont été faites sur 83 résidus et sont récapitulées dans la figure 4.5.

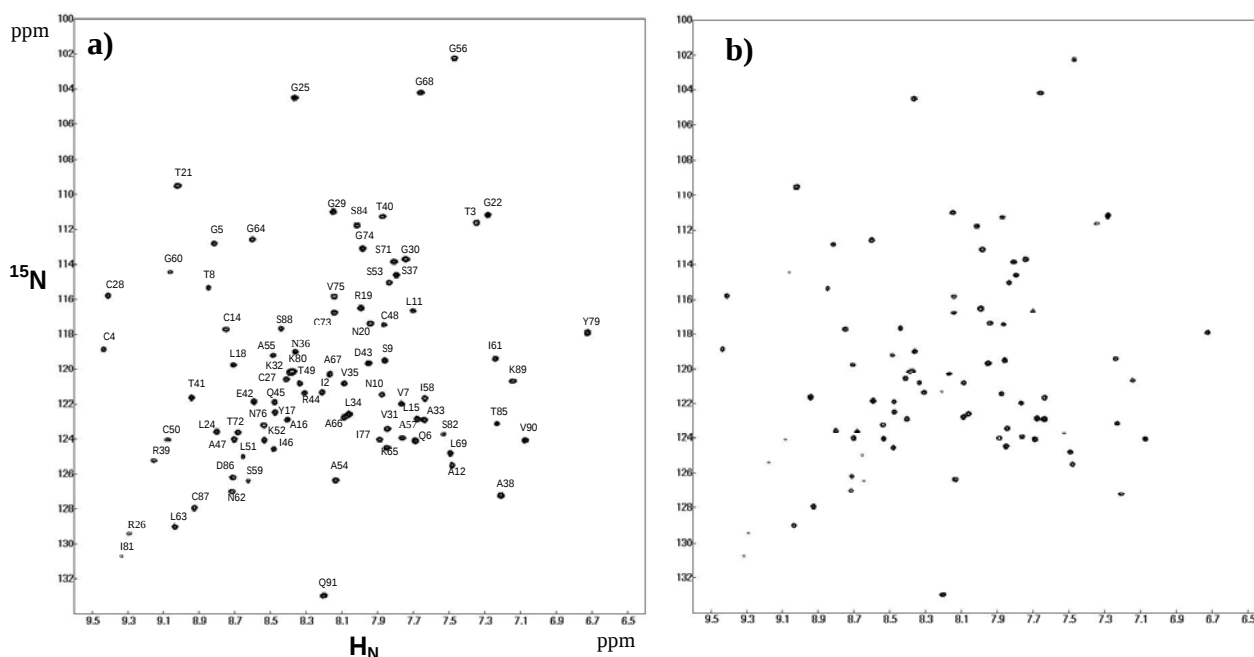


Figure 4.4 : Spectres réalisés pour mesurer le NOE hétéronucléaire ^1H - ^{15}N de la LTP1_1 a) sans irradiation des ^1H et b) avec irradiation des ^1H .

Les résidus de la LTP1_1 présentent une valeur moyenne de NOE hétéronucléaires ^1H - ^{15}N de $0,72 \pm 0,07$. Huit résidus (Ile2, Arg39, Ile77, Tyr79, Ile81, Asp86, Lys89 et Val90) présentent des valeurs de NOE inférieures à 0,65 suggérant que ces derniers présentent une

dynamique interne importante. Ces résidus sont situés dans trois régions différentes de la protéine : les parties N- et C- terminales et la boucle L2 (Arg39).

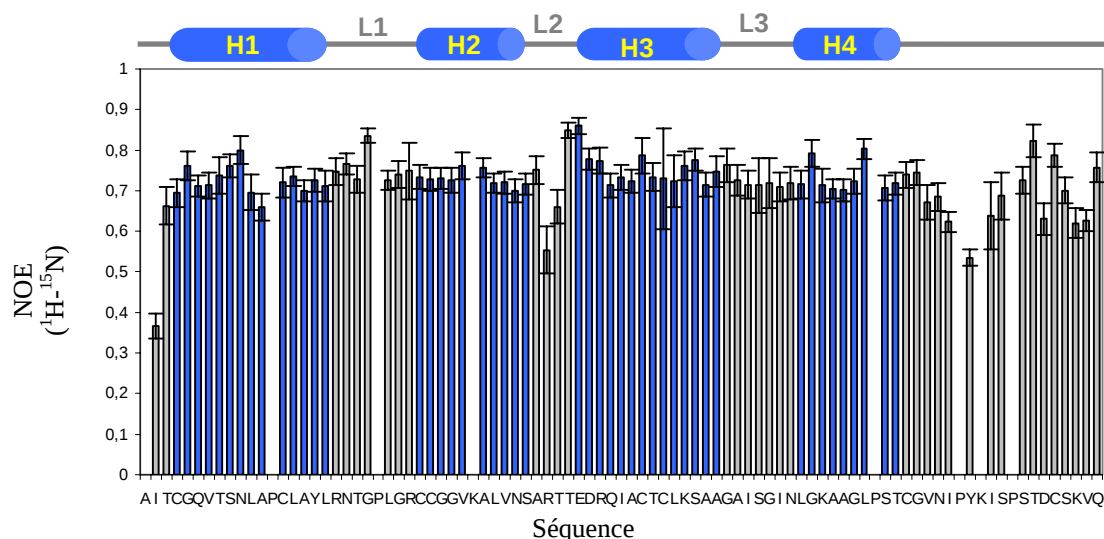


Figure 4.5 : Valeurs du NOE hétéronucléaire ^1H - ^{15}N le long de la séquence de la LTP1_1

1.1.3 Détermination du temps de corrélation τ_c

Pour déterminer le temps de corrélation de la LTP1_1 nous avons calculé la valeur moyenne des rapports R_2/R_1 des résidus impliqués dans les hélices (à l'exception de Leu51 qui présente un R_2 élevé) en considérant un modèle de diffusion isotrope qui suffit pour satisfaire les données expérimentales de relaxation (figure 4.6). La valeur obtenue, $5,3 \pm 0,1$ ns, correspond à la valeur attendue pour une protéine globulaire de cette taille.

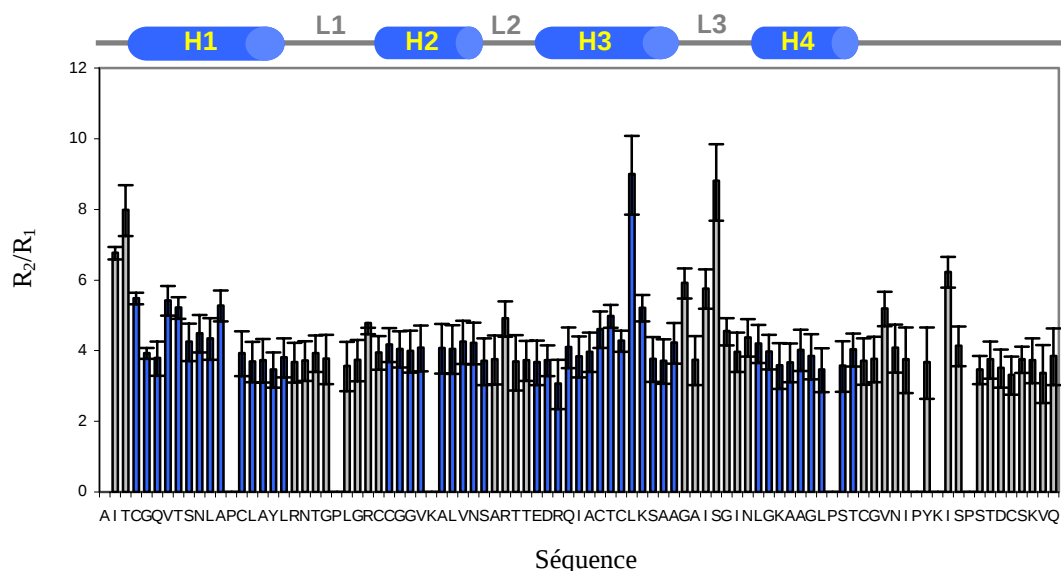


Figure 4.6 : Valeur du rapport R_2/R_1 en fonction de la séquence de la LTP1_1.

1.2 Interprétation des paramètres de relaxation par le formalisme de Lipari-Szabo.

Les données de relaxation obtenues expérimentalement ont été interprétées par le formalisme de Lipari Szabo à l'aide du logiciel Fast-Modelfree, comme décrit dans le chapitre Matériels et Méthodes. Les résultats obtenus sont rassemblés dans la figure 4.7 :

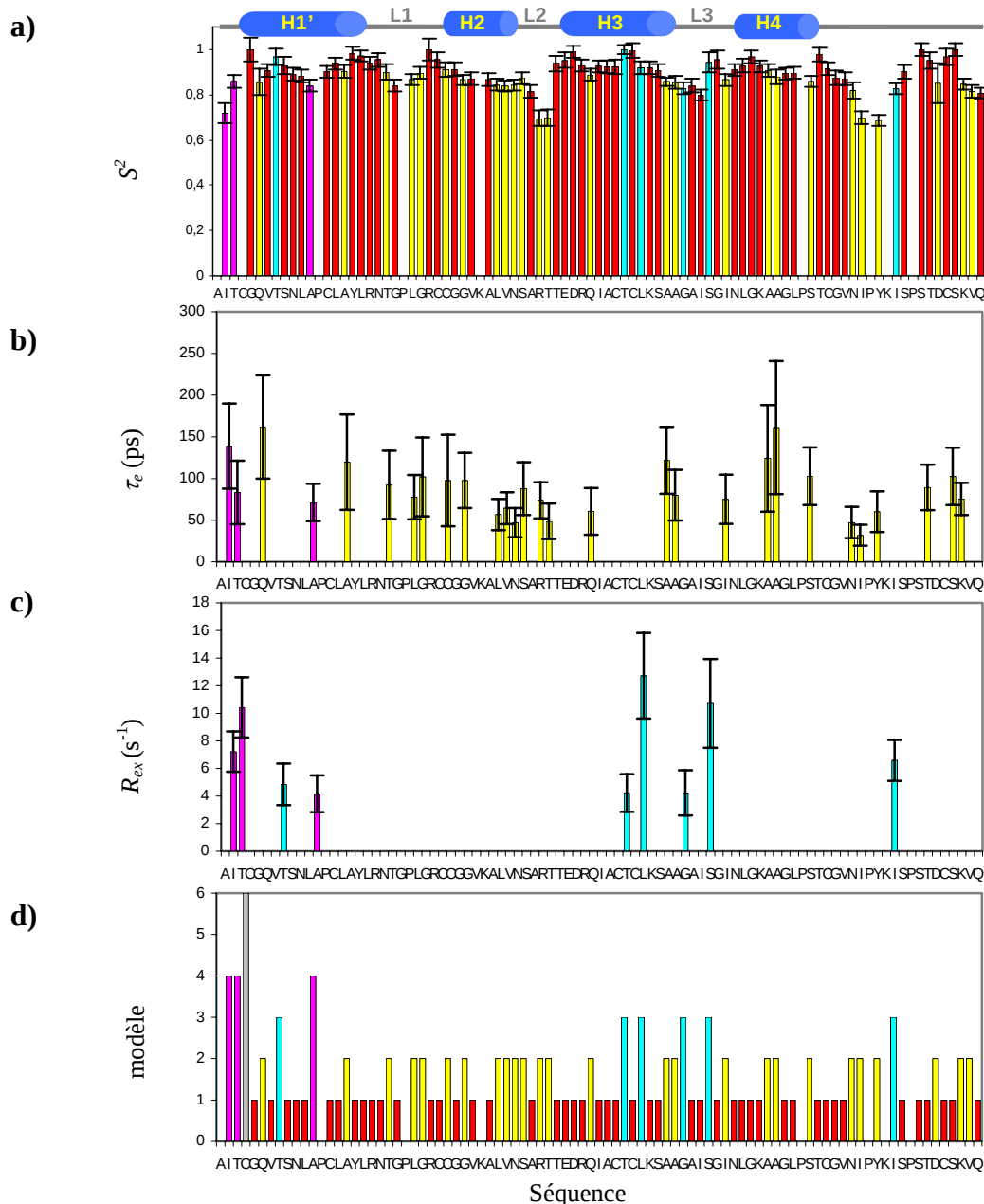


Figure 4.7 : Valeurs des différents paramètres ajustables des modèles utilisés par le logiciel Fast-Modelfree en fonction de la séquence. a) le paramètre d'ordre S^2 , b) le temps de corrélation des mouvements internes τ_e , c) la vitesse d'échange chimique R_{ex} et d) le numéro du modèle utilisé pour l'ajustement. Les modèles sont représentés par un code couleur : rouge pour le modèle 1, jaune pour 2, cyan pour 3 et magenta pour 4. Le numéro 6 coloré en gris signifie qu'aucun des 5 modèles de Fast-Modelfree n'a pu satisfaire les données expérimentales.

L'analyse effectuée à l'aide du logiciel Fast-Modelfree a permis d'ajuster les paramètres de relaxation de tous les résidus à l'exception de la cystéine 4 pour lequel aucun modèle n'était satisfaisant et des lysines 32 et 80 pour lesquelles les paramètres n'étaient pas mesurables. La majorité de ces paramètres (47 résidus sur 83) a été décrit par le modèle 1 (S^2). Il s'agit de 47 résidus répartis uniformément tout au long de la séquence (figures 4.7 a et d), confirmant que la LTP1_1 de tabac se meut en solution de manière isotrope. En effet, le modèle 1 ne considère qu'un seul type de mouvement, avec un temps de corrélation τ_c unique correspondant à celui de la protéine. Les mouvements des vecteurs N-H de ces résidus sont donc uniquement dépendants du mouvement global de la protéine.

Environ 1/3 des résidus (26 sur 83) ont été interprétés par le modèle 2 correspondant au formalisme de Lipari-Szabo originel « à deux paliers » qui ne considère que deux types de mouvements se déroulant sur deux échelles de temps distinctes. Le premier correspond au mouvement global de la protéine, avec un temps de corrélation isotrope τ_c qui est égal pour la LTP1_1 à 5,3 ns. Le second caractérise les mouvements internes des vecteurs N-H de ces résidus qui présentent un temps de corrélation interne τ_e de l'ordre de la picoseconde. La valeur moyenne de τ_e pour les résidus considérés est égale à une centaine de picosecondes environ. Ces résidus présentent donc une dynamique interne simple, sans intervention d'un terme d'échange R_{ex} , et sont répartis tout au long de la séquence (figures 4.7 a, b et d).

Les modèles 3 et 4 correspondant respectivement aux modèles 1 et 2 avec un terme d'échange chimique supplémentaire (R_{ex}) ont permis de modéliser les mouvements internes des vecteurs N-H de 9 résidus : 6 résidus par le modèle 3 et 3 résidus par le modèle 4 (figures 4.7). Ces résidus sont situés dans les parties N- et C-terminales de la LTP1_1 (Ile2, Thr3 et Ile81), les hélices H1 et H3 (Thr8, Ala12, Thr49, L51) et la boucle L3 (Gly56 et Ser59). De plus, ces résidus se trouvent essentiellement à proximité des ponts disulfure (Cys4-Cys50, Cys14-Cys27, Cys28-Cys73, Cys48-Cys87) ou des tyrosines (Tyr17 et Tyr79), ce qui peut expliquer l'origine des termes supplémentaires d'échange chimique. En effet, les courants de cycle des chaînes latérales aromatiques créent de forts gradients de déplacement chimique provoquant des chemins de relaxation efficaces et conduisant par conséquent à des vitesses de relaxation plus élevées pour les noyaux proches de ces cycles. De la même façon, l'isomérisation conformationnelle des ponts disulfure peut être à l'origine de l'augmentation des vitesses de relaxation des noyaux qui sont à leur voisinage.

Les paramètres d'ordre S^2 calculés pour les vecteurs N-H des résidus de la LTP1_1 (figure 4.7 a et tableau 4.1) montrent que cette dernière est structurée en solution et que ses

mouvements internes sont très restreints. En effet, la valeur moyenne des S^2 tout au long de la séquence est élevée et égale à 0,89. Elle augmente jusqu'à 0,93 pour les résidus impliqués dans les hélices. Même les résidus des régions N-et C-terminales présentent des S^2 élevés. Par exemple, la moyenne est de 0,87 pour l'ensemble du segment C-terminal (tableau 4.1), alors que pour la plupart des protéines les valeurs n'excèdent pas 0,6 dans ces régions. Les seuls résidus qui montrent des valeurs de S^2 relativement faibles ($<0,7$) sont situés au niveau de la boucle L2 (Arg39 et Thr40) et dans la partie C-terminale pour les résidus Ile77 et Tyr79, suggérant que ces régions présentent des mouvements internes d'amplitude importante. Les paramètres d'ordre S^2 sont illustrés sur la structure tridimensionnelle de la LTP1_1 dans la figure 4.8.

Régions de la LTP1_1	S^2
Résidus (2-91)	0.89 ± 0.03
Hélices α (H1 à H4)	0.91 ± 0.03
Hélice H1 (4-18)	0.92 ± 0.03
Hélice H2 (27-37)	0.88 ± 0.03
Hélice H3 (42-55)	0.93 ± 0.03
Hélice H4 (63-72)	0.92 ± 0.03
Segment C-terminal (73-91)	0.87 ± 0.03

Tableau 4.1 : Valeurs moyennes des paramètres d'ordre S^2 dans différentes régions de la LTP1_1.

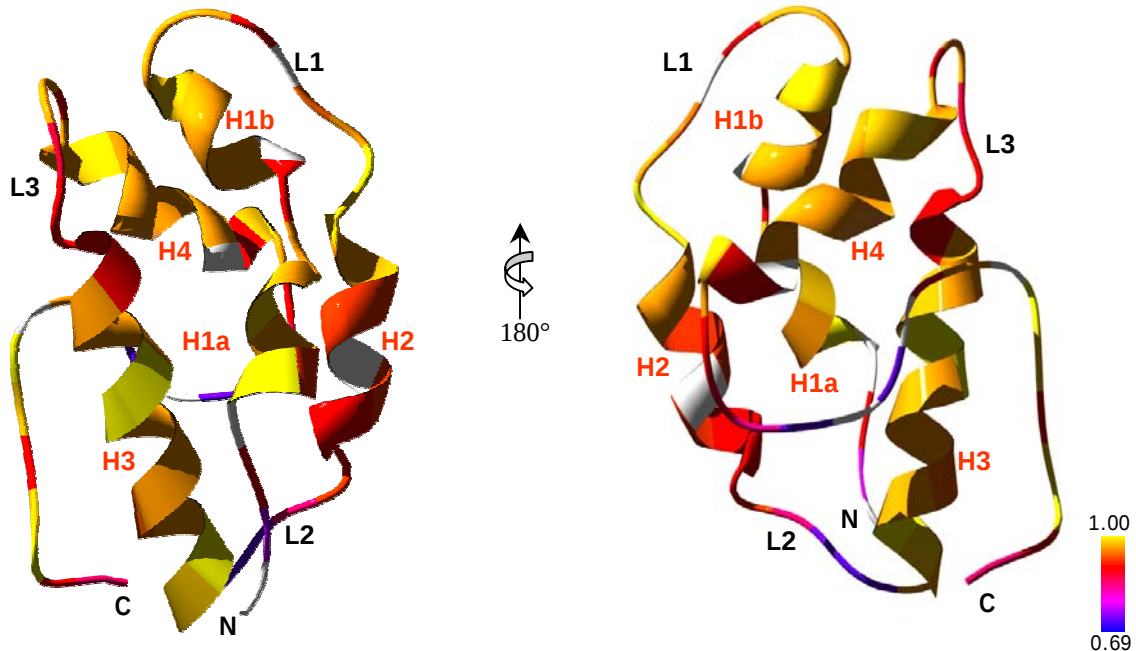


Figure 4.8 : Représentation des valeurs du paramètre d'ordre S^2 le long de la structure tridimensionnelle de la LTP1_1. L'échelle des couleurs est indiquée en bas à droite. Les résidus pour lesquels le S^2 n'a pu être déterminé sont en blanc.

1.3 Discussion

La mesure des paramètres de relaxation des noyaux ^{15}N de la LTP1_1 de tabac a mis en lumière les propriétés de dynamique particulières de son squelette peptidique.

Tout d'abord, la valeur du temps de corrélation global de la LTP1_1, $\tau_c = 5,3 \pm 0,1$ ns, montre que la protéine est sous forme monomérique en solution. En effet, cette valeur est comparable à celle mesurée dans les mêmes conditions pour des protéines de même taille. Par exemple, le temps de corrélation global de la partie C-terminale de la calmoduline, qui comporte 90 résidus, a été estimé à 5,3 ns à une température de 291 K et à pH 6.0 (Evenas et al. 1999).

L'analyse des valeurs des paramètres d'ordre S^2 de la LTP1_1 a montré qu'elle est très bien structurée en solution et que ses mouvements internes sont très restreints. Ces résultats s'expliquent d'une part par le fait que plus de la moitié des résidus sont impliqués dans les hélices (50 résidus sur 91) et que d'autre part la LTP1_1 possède quatre ponts disulfure. Ces résultats sont cohérents avec les propriétés biophysiques des protéines de la famille des LTP1. En effet, comme nous l'avons vu dans le premier chapitre de rappels bibliographiques, les LTP1 sont des protéines extrêmement stables : elles résistent à la dénaturation chimique, à des températures élevées et à certaines protéases (Lindorff-Larsen and Winther 2001).

Cette étude a également révélé une dynamique particulière de la LTP1_1 au niveau de la région définissant l'entrée de la cavité (figure 4.9). L'analyse des mouvements des vecteurs N-H montre que les résidus de la boucle L2 (Arg39 et Thr40) et de la partie C-terminale (Ile77 et Tyr79) présente une dynamique importante (faibles valeurs des NOE hétéronucléaires et des S^2). De plus, l'ajustement des paramètres de relaxation de l'isoleucine 81 fait intervenir un terme d'échange chimique qui peut être dû à la proximité de la tyrosine 79 ou à différentes orientations de la chaîne latérale de cette isoleucine. Ces résultats, ajoutés à ceux des études d'interactions entre le LMPC et la LTP1_1, fournissent un faisceau de présomptions sur l'implication des résidus Arg39, Thr40, Ile77, Tyr79 et Ile81 dans la définition de l'entrée de la cavité de la LTP1_1 et sur leur rôle dans la fixation des molécules lipidiques (figure 4.9). En effet, ces résidus présentent une forte variation de leurs déplacements chimiques lors de la complexation et sont ceux dont la dynamique est la plus importante au sein de la protéine.

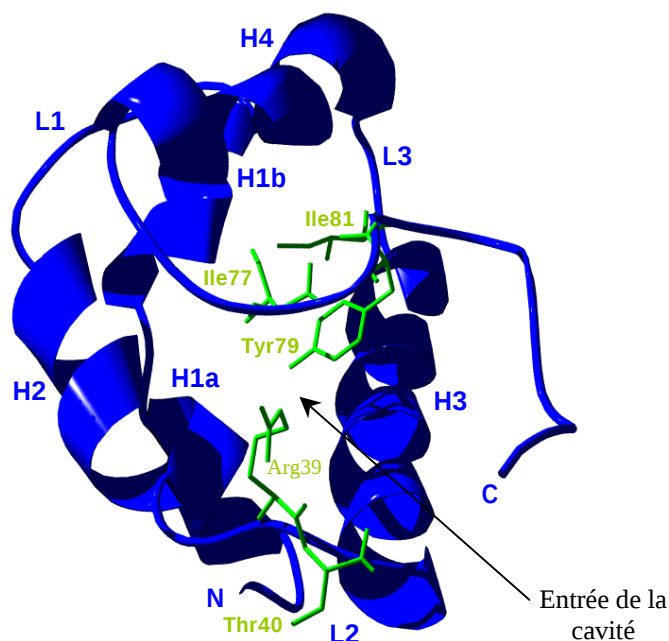


Figure 4.9 : Illustration des résidus impliqués dans la définition de l'entrée de la cavité. Les résidus sont indiqués en vert : Arg39, Thr40, Ile77, Tyr79 et Ile81.

Afin de réaliser une description tridimensionnelle des mouvements atomiques de la LTP1_1 et de comprendre le mode d'insertion des molécules lipidiques, nous avons entrepris d'effectuer des simulations de dynamique moléculaire. Celles-ci présentent l'avantage de procurer des informations à la fois quantitatives et qualitatives sur les mouvements des macromolécules et constituent une approche complémentaire de l'approche expérimentale.

Nous avons donc réalisé, dans un premier temps, une simulation à température ambiante (293K) afin de vérifier la stabilité de notre système dans les conditions de simulation utilisées, puis deux simulations à haute température (500K) afin d'étudier les possibilités de dépliement de la LTP1_1.

2. Simulation de dynamique moléculaire à 293K

La première étape de notre étude de dynamique moléculaire *in silico* consacrée à la LTP1_1 a été effectuée à température ambiante en utilisant le protocole présenté dans le chapitre Matériels et Méthodes. Cette dynamique moléculaire a été réalisée pour vérifier la stabilité de la structure de la LTP1_1 dans le champ de force de CHARMM, à partir d'une structure RMN présentant à la fois une énergie NOE faible (environ 15 kcal/mol) et une taille de cavité moyenne (environ 320 Å³). Cette simulation a été effectuée sur une période de 5 ns.

2.1 Stabilité de la LTP1_1 de tabac

Le paramètre le plus simple et le plus informatif pour quantifier les résultats de notre dynamique moléculaire est le RMSD (Root Mean Square Deviation) qui permet d'accéder aux fluctuations des positions atomiques par rapport à une structure de référence. Le calcul des RMSD a été réalisé sur les atomes C_α de la totalité du squelette peptidique, de chacune des 4 hélices α et de la partie C-terminale (figure 4.10 et tableau 4.2). L'analyse du RMSD sur le squelette peptidique en fonction du temps, par rapport à la structure RMN de départ minimisée, révèle des variations significatives sur les 500 premières ps puis une stabilisation avec l'apparition d'un plateau autour d'une valeur moyenne de $1,94 \pm 0,18 \text{ \AA}$. Les variations observées sur les 500 premiers ps correspondent à une « équilibration du système » dans le champ de force de CHARMM, entraînant un réajustement des coordonnées de la protéine par rapport à celles de la structure expérimentale initiale. Les valeurs moyennes de RMSD calculées sur les 4 hélices α sont plus faibles ($\sim 0,5 \text{ \AA}$), à l'exception de l'hélice H1 qui présente un RMSD deux fois plus important que les autres, $1,03 \pm 0,22 \text{ \AA}$, résultant d'un réarrangement conformationnel notamment au niveau du coude de l'hélice. Les atomes C_α du segment C-terminal présentent les plus grandes fluctuations, $1,82 \pm 0,25 \text{ \AA}$, suggérant une flexibilité plus importante de cette région de la LTP1_1 par rapport aux structures secondaires (figure 4.10 et tableau 4.2). De plus, on observe que ces fluctuations sont corrélées à celles de l'ensemble du squelette peptidique ($1,94 \pm 0,18 \text{ \AA}$).

Les faibles valeurs des écarts types montrent la remarquable stabilité de la structure 3D de la LTP1_1 après que l'équilibre du système (protéine/molécules d'eau) a été atteint. L'analyse des valeurs moyennes des RMSD calculés, par rapport à la structure 3D de la LTP1_1 à l'équilibre confirme la stabilité de cette structure. En effet, ces RMSD sont plus faibles que ceux calculés par rapport à la structure expérimentale, comme par exemple le RMSD sur les atomes C_α du squelette peptidique qui est réduit d'environ 1/3, passant de $1,94 \pm 0,18 \text{ \AA}$ à $1,39 \pm 0,25 \text{ \AA}$ (tableau 4.2). On remarque une flexibilité réduite des hélices dont les fluctuations sont faibles après 500 ps, notamment pour l'hélice H1 qui, après sa déviation par rapport à la structure initiale, présente un RMSD similaire à ceux des 3 autres hélices restées plus proches de la structure 3D expérimentale. Toutefois, malgré ses réarrangements conformationnels, l'hélice H1 conserve sa particularité structurale : un coude marqué au niveau de la proline 13.

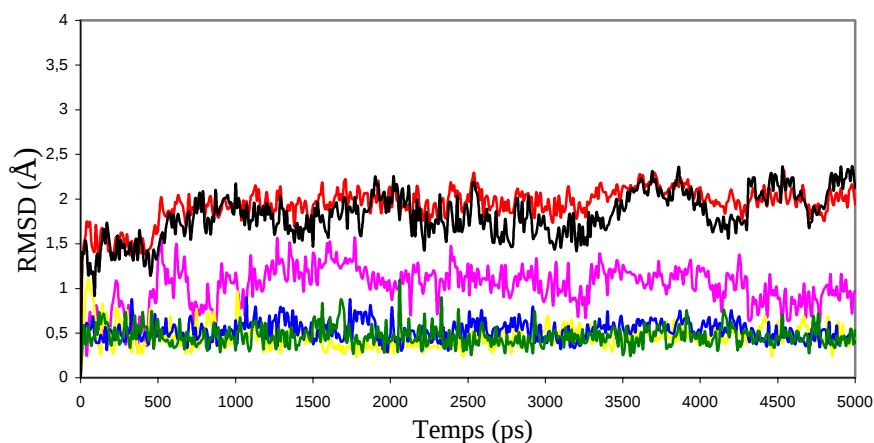


Figure 4.10 : RMSD calculé sur les C_{α} des différents segments de la protéine : en rouge le squelette, en rose l'hélice H1, en jaune H2, en bleu H3, en vert H4, en noir la partie C terminale.

Résidus	Atomes C_{α} 1-91	H1 4-18	H2 27-37	H3 42-55	H4 63-72	C-terminal 73-91
MD (293K) ¹	$1,94 \pm 0,18 \text{ \AA}$	$1,03 \pm 0,22 \text{ \AA}$	$0,46 \pm 0,12 \text{ \AA}$	$0,53 \pm 0,10 \text{ \AA}$	$0,46 \pm 0,11 \text{ \AA}$	$1,82 \pm 0,25 \text{ \AA}$
MD (293K) ²	$1,39 \pm 0,25 \text{ \AA}$	$0,55 \pm 0,15 \text{ \AA}$	$0,43 \pm 0,20 \text{ \AA}$	$0,44 \pm 0,11 \text{ \AA}$	$0,43 \pm 0,14 \text{ \AA}$	$1,50 \pm 0,24 \text{ \AA}$

¹ Calcul du RMSD par rapport de la structure RMN minimisée

² Calcul du RMSD par rapport de la structure après équilibration (500 ps)

Tableau 4.2 : RMSD sur différents segments de la LTP1_1.

Cette grande stabilité de la structure globale de la LTP1_1 est confirmée par d'autres analyses : par exemple l'évolution du rayon de giration montre que le rayon moyen de la protéine ne varie pas au cours de la simulation à 293K, avec une valeur moyenne de 12,3 Å et un écart type inférieur à 0,2 Å (figure 4.11).

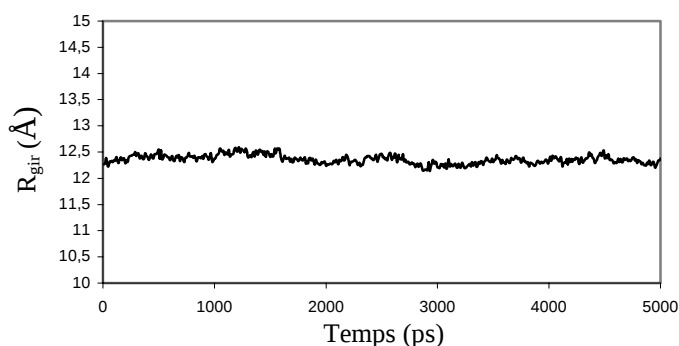


Figure 4.11 : Rayon de giration de la LTP1_1 tout au long de la simulation.

De même, l'analyse des distances entre les C_α des cystéines engagées dans les ponts disulfure (Figure 4.12) montre qu'elles restent quasiment constantes au cours de la simulation avec des fluctuations voisines de 0,2 Å.

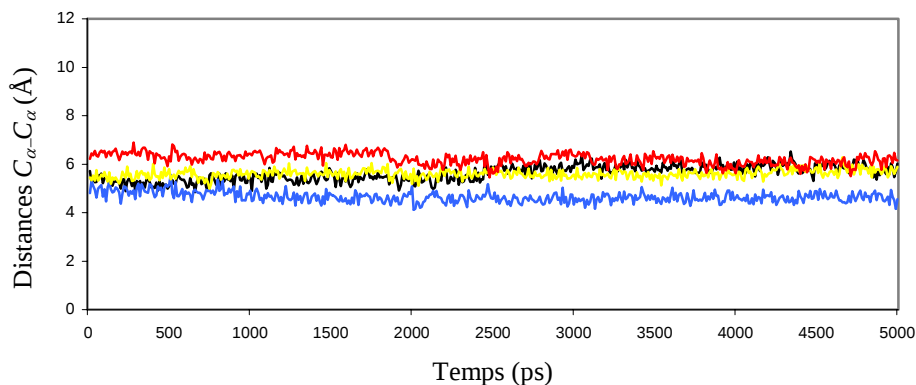


Figure 4.12 : Distances $C_\alpha-C_\alpha$ dans les ponts disulfure : noir 4-50, bleu 14-27, jaune 28-73, rouge 48-87.

Enfin, la détermination du pourcentage de liaisons hydrogène $i, i+4$ caractéristiques des hélices α , sur un ensemble de structures séparées de 50 ps, montre la persistance et la stabilité des structures secondaires de la LTP1_1 tout au long de la simulation (figure 4.13).

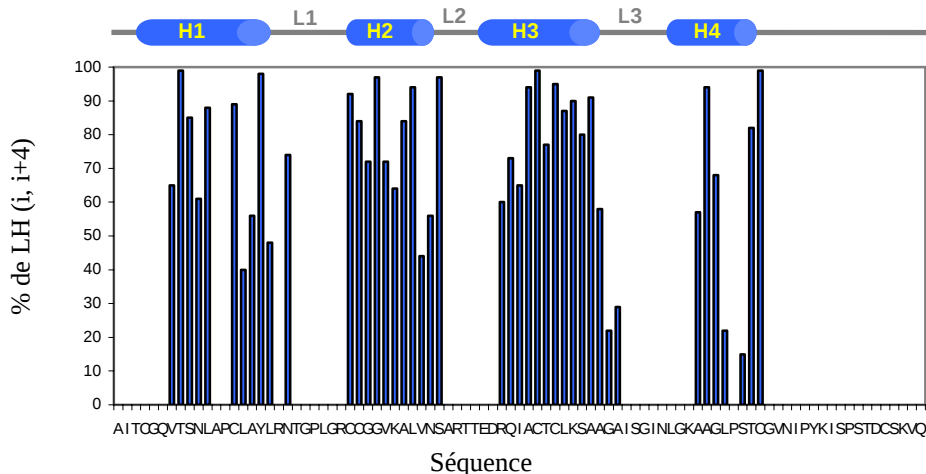


Figure 4.13 : Pourcentage de liaisons hydrogène (LH) $i, i+4$ déterminé sur le squelette peptidique de la LTP1_1 tout au long de la simulation de dynamique moléculaire de 5ns.

Une représentation sous forme de « boudin » de cet ensemble de structure (figure 4.14), avec un rayon de cylindre proportionnel au RMSD par rapport à la structure moyenne, illustre la faible variabilité des structures secondaires. Les zones où la dynamique est la plus marquée sont le C-terminus et les boucles.

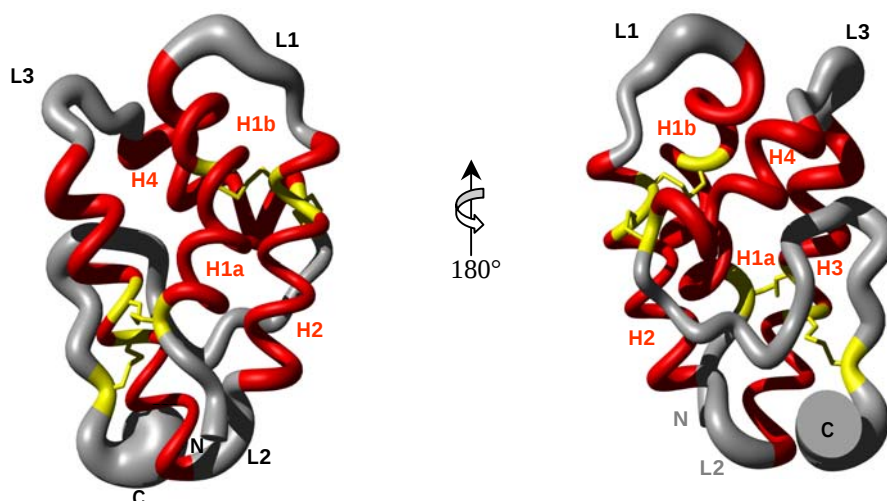


Figure 4.14 : Représentation « en forme de boudin » de 50 structures 3D de la LTP1_1 correspondant à un ensemble de conformations obtenues toutes les 100 ps° tout au long de la simulation de dynamique moléculaire. Le rayon du cylindre est proportionnel au RMSD par rapport à la structure moyenne

La caractéristique principale des protéines de transfert de lipides est la présence d'une cavité hydrophobe au cœur de la protéine. La taille de celle de la LTP1_1 varie tout au long de la dynamique, essentiellement en raison des fluctuations de la position du segment C-terminal et des chaînes hydrophobes des résidus tapissant la cavité. Son volume moyen est de $210 \text{ \AA}^3$ avec une fluctuation importante de $100 \text{ \AA}^3$ environ.

3. Simulations de dynamique moléculaire à 500K

Afin de poursuivre notre étude de la dynamique de la LTP1_1, nous avons effectué une étude à haute température (500K). La structure initiale utilisée est celle obtenue en fin de dynamique à température ambiante, sur laquelle nous avons appliqué la procédure de chauffage décrite dans le chapitre Matériels et Méthodes. Deux simulations à haute température ont été réalisées en modifiant la valeur des vitesses initiales, afin de s'assurer de la reproductibilité des résultats observés. Ces deux simulations à haute température ont été effectuées sur une période de 3 ns. L'objectif d'une simulation à 500K est d'utiliser une température élevée pour accélérer la dynamique de la protéine, afin d'observer un éventuel mécanisme d'ouverture de la cavité hydrophobe. De tels mouvements ne pourraient être observés à 293K dans des limites de temps de calcul raisonnables.

3.1 Comportement de la LTP1_1

Les deux simulations de dynamique moléculaire à haute température ont révélé des comportements identiques de la protéine, comme le traduisent les valeurs moyennes des RMSD calculés sur différentes régions de la LTP1_1 (tableau 4.3). Par conséquent, par souci de clarté, je ne vais présenter que les analyses réalisées sur la première trajectoire de dynamique moléculaire à haute température. Ces analyses sont pour la plupart identiques à celles effectuées pour la simulation à température ambiante, nous permettant ainsi d'observer de nettes différences au niveau de la stabilité de la protéine. Elles ont portées sur les calcul du RMSD sur différentes régions de la protéine, celui du rayon giration et des distances C_{α} - C_{α} des cystéines impliquées dans les ponts disulfure. Ces résultats sont exposés respectivement dans les figures 4.15 a, b et c.

Résidus	Atomes C_{α} 1-91	H1 4-18	H2 27-37	H3 42-55	H4 63-72	C-terminal 73-91
MD (500K) ¹	$5,32 \pm 1,66 \text{ \AA}$	$1,79 \pm 0,98 \text{ \AA}$	$1,86 \pm 1,07 \text{ \AA}$	$1,34 \pm 0,64 \text{ \AA}$	$3,08 \pm 1,02 \text{ \AA}$	$4,33 \pm 1,27 \text{ \AA}$
MD (500K) ²	$5,03 \pm 1,69 \text{ \AA}$	$1,68 \pm 0,77 \text{ \AA}$	$1,88 \pm 1,04 \text{ \AA}$	$1,04 \pm 0,40 \text{ \AA}$	$2,76 \pm 1,22 \text{ \AA}$	$4,59 \pm 1,35 \text{ \AA}$

¹ et ² numéro de chaque simulation, l'analyse présentée est celle correspondant à la trajectoire numéro 1.

Tableau 4.3 : RMSD sur différents segments de la LTP1_1.

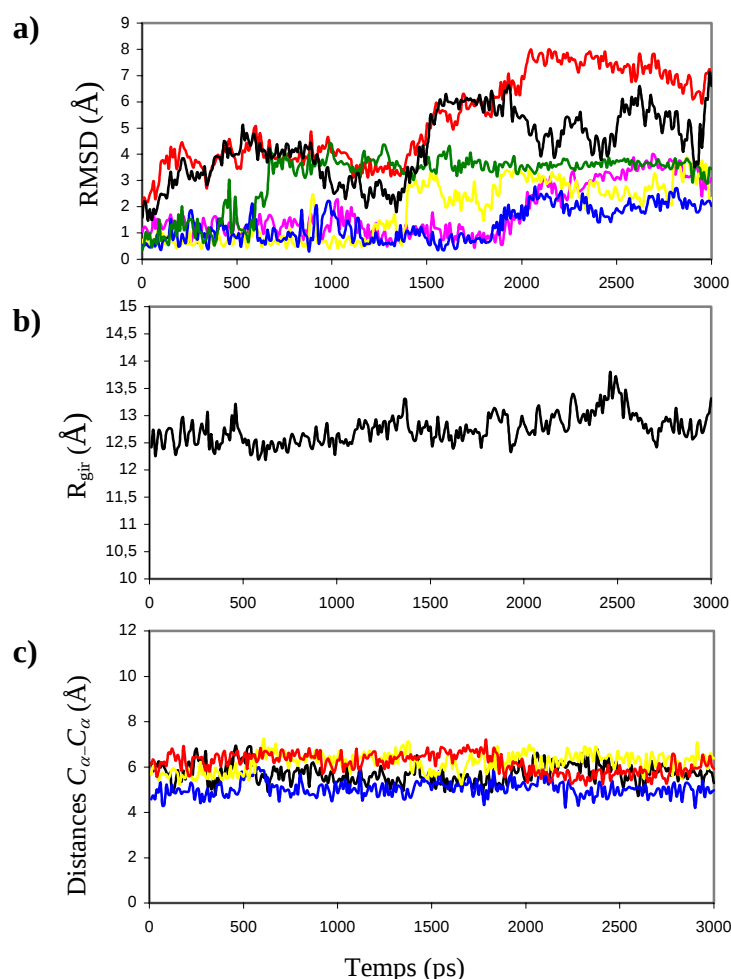


Figure 4.15 : Analyses effectuées lors d'une trajectoire sur une simulation à haute température. a) RMSD sur les C_{α} des différents segments de la protéine : en rouge le squelette, en rose l'hélice H1, en jaune H2, en bleu H3, en vert H4, en noir le segment C-terminal. b) Rayon de giration de la LTP1_1 tout au long de la simulation. c) Distances C_{α} - C_{α} des ponts disulfure : noir 4-50, bleu 14-27, jaune 28-73, rouge 48-87.

A haute température, on observe un dépliement partiel de la protéine. En effet, la valeur du RMSD sur les atomes C_{α} du squelette peptidique augmente rapidement, (figure 4.15 a), pour atteindre une valeur maximale de 8 Å après 2,1 ns de simulation. Cette valeur se stabilise autour de ce maximum jusqu'à la fin de la dynamique. De la même façon, on remarque une augmentation progressive du rayon de giration de la protéine qui atteint une valeur maximale proche de 14 Å aux alentours de 2,5 ns de dynamique (figure 4.15 b). Le rayon de giration diminue ensuite à une valeur proche de 13 Å à la fin de la dynamique. Néanmoins, on peut remarquer que le rayon de giration n'augmente pas de manière importante, ce qui traduit le maintien d'une structuration partielle de la protéine tout au long de la simulation. Cette persistance de structuration de la LTP1_1 est à corrélérer avec la présence des quatre ponts disulfure qui maintiennent la forme globulaire de la protéine. En effet, la valeur moyenne des distances C_{α} - C_{α} (figure 4.15 c) des cystéines impliquées dans les ponts disulfure est proche de celle obtenue lors de la simulation de dynamique à température ambiante, autour de 5,8 Å.

Si on s'intéresse plus particulièrement à certaines régions de la protéine, on observe des comportements différents de chacune des hélices ainsi que du segment C-terminal (figure 4.15 a). L'analyse des RMSD sur les atomes C_{α} révèle de très grandes fluctuations du segment C-terminal avec un RMSD variant entre 1,5 et 6,6 Å avec une valeur moyenne de 4,3 Å. Au niveau des hélices α , on remarque un comportement similaire des hélices H1 et H3 qui s'éloignent faiblement de leur conformation initiale. En effet, pendant les 2 premières ns de simulations, les hélices H1 et H3 sont proches de leur conformation initiale avec un RMSD voisin de 1,1 Å ; puis sur la fin de la simulation, on observe une augmentation du RMSD qui atteint un plateau à 2,1 Å pour l'hélice H3 et à 3,1 Å pour H1. Les hélices H2 et H4 présentent un comportement totalement différent. L'hélice H4 s'éloigne très rapidement de sa conformation initiale après 500 ps de simulation pour atteindre une valeur plateau de son RMSD voisine de 3,6 Å. L'hélice H2, quant à elle, s'éloigne de sa conformation initiale aux alentours de 1,2 ns de simulation et la valeur de son RMSD se stabilise ensuite autour de 2,6 Å. Une analyse des structures secondaires dans les conformations de la LTP1_1 séparées de 100 ps tout au long de la simulation révèle que certaines hélices α sont progressivement déstructurées (figure 4.16). Ces résultats sont en accord avec ceux observés pour les RMSD : on remarque ainsi que les hélices H4 puis H2 se déstructurent progressivement au cours de la simulation ; l'hélice H2 disparaît totalement alors que les hélices H1 et H3 présentent une stabilité plus importante.

LTP1_1	AITCGQVTSNLAPCLAYLRNTGPLGCCGGVKALVNSARTTEDRQIACTCLKSAAGAISGINLGKAAGLPSTCGVNIPYKISPSTDCKSVQ
100 ps	-- - HHHHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHH -- - HHHHHHHHHHHH -- - - HHHHH - HHHH -----
200	-- - HHHHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - HHHHH - HHHHH -----
300	-- - HHHHHHHH - HHHHHHHH -- - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - HHHHHHHHHH -----
400	- - - - HHHHHHHHHHHHHHHH -- - HHHHHHHH - HHHHHH -- - HHHHHHHHHHHHHHHH -- - - - HHHHHHHHHH -----
500	-- - HHHHHHHHHHHHHHHH -- - - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHH -- - - - HHHHH -----
600	- - - - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - - HHHHH -----
700	-- - HHHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHH -- - - - - HHHHHHHHHHHH -- - - - HHHHHH -----
800	-- - HHHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
900	- - - - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
1000	- - - - HHHHHHHHHHHHHHHH -- - HHHHHHHHH -- - - - - HHHHHHHHHHHH -- - - - - HHHHHH -----
1100	-- - HHHHHHHHHHHHHHHHH -- - - HHHH - HHHHHHHH -- - HHHHHHHHHH -- - HHHHH -- - - - HHHHHH -----
1200	-- - HHHHHHHHHHHHHHHHH -- - - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHH -- - HHHH -- - - - HHHHHH -----
1300	-- - HHHHHHHHHHHHHHHHH -- - - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
1400	- - - - HHHHHH - HHHHHHHH -- - - HHHHHHHHHH -- - - - - HHHHHHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
1500	-- - HHHHHHHH - HHHHHHHH -- - HHHH -- - - HHHHHH -- - - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
1600	-- - HHHHH -- - HHHHHHHHH -- - - - - - HHHH -- - - - - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
1700	-- - HHHHHHHH - HHHHHHHH -- - - - - - HHHHHHHH -- - HHHHHHHHHHHHHHHHHH -- - - - - HHHHH -----
1800	-- - HHHHHHHH - HHHHHH -- - - - - - HHHHHH -- - - - - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
1900	-- - HHHHHHHH - HHHHHHHH -- - - - HHHHHHHHHHHHHHHH -- - HHHHHHHHHHHHHHHH -- - - - - HHHHH -----
2000	- - - - HHHH -- - HHHHHHHHH -- - - - - - HHHH -- - - - - HHHHHHHHHHHHHHHH -- - - - - HHHHH -----
2100	- - - - - - - - HHHHHHHH -- - HHHH -- - - HHHHHH -- - - - - HHHHHHHHHH -- - - - - HHHHHH -----
2200	- - - - - - - - HHHHHHHHH -- - - - - - - - - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
2300	- - - - - - - - HHHHHHHHH -- - - - - - - - - HHHHHHHHHH -- - - - - HHHHH -----
2400	- - - - - - HHHHHHHHHH -- - - - - - - - - HHHHHHHHHHHHHHHHHH -- - - - - HHHHH -----
2500	- - - - - - - - HHHHHHHHH -- - - - - - - - - HHHHHHHHHHHH -- - - - - HHHHHH -----
2600	- - - - - - - - HHHHHHHHH -- - - - - - - - - HHHHHHHHHHHHHHHH -- - - - - HHHHHH -----
2700	- - - - - - - - HHHHHHHHH -- - - - - - - - - HHHHHHHHHH -- - - - - HHHHH -----
2800	- - - - - - - - HHHHHHHHHH -- - - - - HHHHHHHHH -- - - - - - - - - HHHHHH -----
2900	- - - - - - - - HHHHHHHHHH -- - - - - - - - - HHHHHHHHHH -- - - - - - - - - HHHHH -----
3000	- HHH -- - - - HHHHHHHHH -- - - - - - - - - HHHHHHHHHHHHHHHH -- - - - - - - - - HHHHH -----

Afin d'identifier les mouvements des différentes régions de la LTP1_1, nous avons réalisé en complément des analyses précédentes, un examen visuel des différentes conformations obtenues au cours de la simulation. Cette étude a révélé, qu'après 1,5 ns de simulation, la LTP1_1 adopte des structures intermédiaires entre l'état natif et celui dénaturé. Les hélices H2 et H4 sont presque entièrement déstructurées et les valeurs des RMSD dépassent 3 Å pour la plupart des régions de la LTP1_1, à l'exception des hélices H1 et H3 dont le RMSD est voisin de 1 Å (figures 4.15 a et 4.16).

simulation. Ensuite, la forme globulaire et la cavité de la LTP1_1 ne persistent que grâce à la présence des quatre ponts disulfure et le volume de la cavité est difficile à définir. Cette partie de la dynamique n'a donc pas été exploitée plus en avant.

L'examen des conformations sur les premières 1,5 ns montrent que très rapidement dès 400 ps, le segment C-terminal de la LTP1_1 se comporte comme un couvercle qui se soulève et qui s'éloigne fortement de sa position initiale. La figure 4.17 A illustre ce mouvement : la région Val75-Ser84 du segment C-terminal adopte une conformation étendue qui se positionne parallèlement à l'hélice H4, alors que la région Thr85-Gln91 devient parallèle à l'hélice H3.

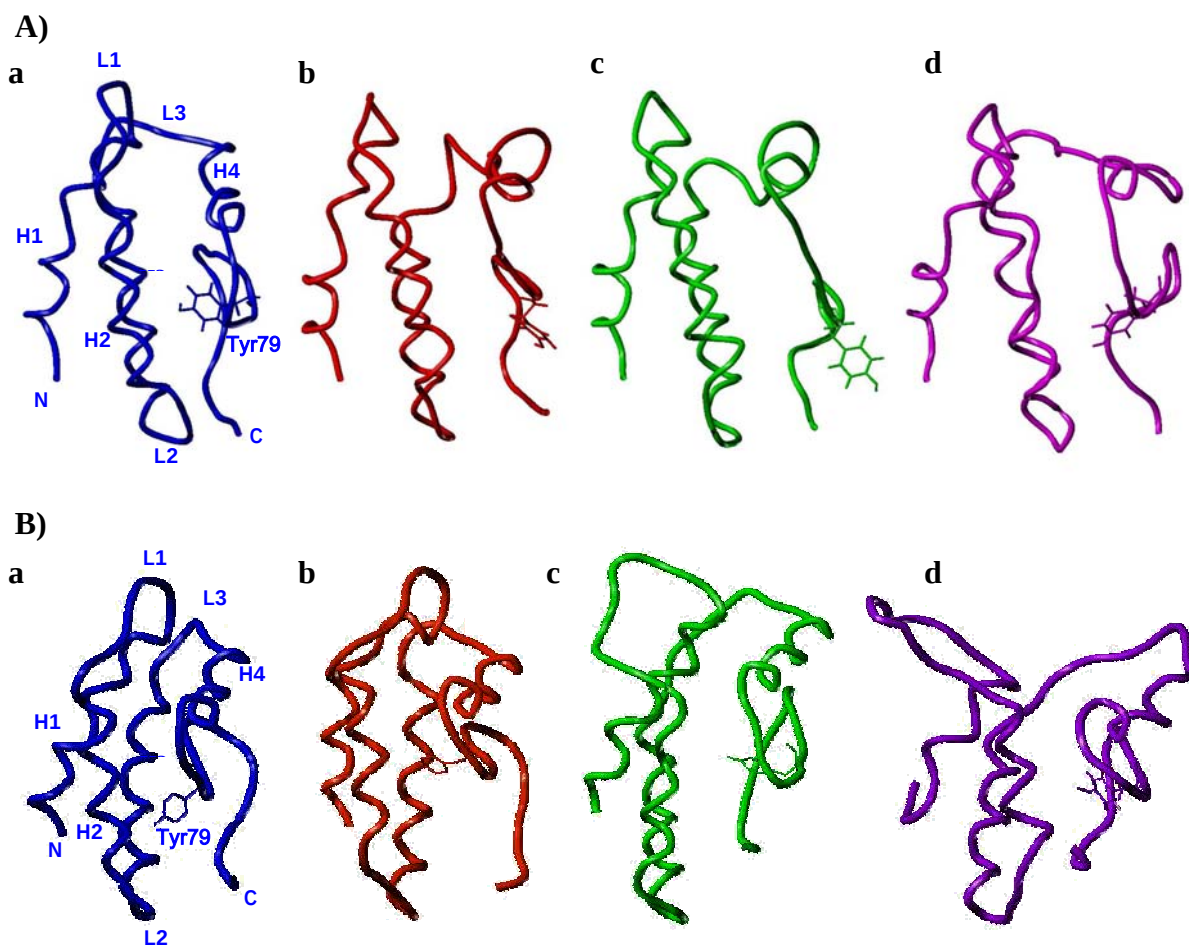


Figure 4.17 : Représentation de différentes conformations des LTP1 de tabac et de blé au cours de simulations de dynamique moléculaire à haute température. A) Conformations de la LTP1_1 de tabac : a) référence au début de la simulation ; b) 0,5 ns ; c) 1 ns et d) 1,4 ns. B) Conformation de la LTP1 de blé adaptée de (Sy et al. 2003): a) référence au début de la simulation ; b) 0,5 ns ; c) 1 ns et d) 1,35 ns.

Afin d'évaluer quantitativement l'amplitude des mouvements de la LTP1_1 au cours des premières 1,5 ns, plusieurs distances caractéristiques ont été mesurées sur des conformations séparées de 100 ps. Trois distances, $C_{\alpha}11$ (H1)- $C_{\alpha}31$ (H2), $C_{\alpha}11$ (H1)- $C_{\alpha}51$ (H3) et $C_{\alpha}31$ (H2)- $C_{\alpha}51$ (H3) permettent d'évaluer les mouvements relatifs des hélices H1, H2 et H3. Deux distances, $C_{\alpha}40$ (L2)- $C_{\alpha}78$ (C-term) et $C_{\alpha}20$ (L1)- $C_{\alpha}60$ (L3), quantifient respectivement les mouvements entre la boucle L2 et la partie C-terminale et entre les boucles L1 et L3, ces deux régions constituant des entrées possibles de la cavité (figure 4.18). Comme on le remarque sur la figure 4.18, les distances calculées entre les 3 hélices sont plutôt stables au cours de la dynamique, la plus grande variation est observée entre les hélices H2 et H3. La distance entre la boucle L2 et la partie C-terminale qui contrôle l'ouverture de la première entrée de la cavité de la LTP1_1 présente la plus grande variation. En effet, celle-ci augmente rapidement de 10 Å à 23 Å à 1000 ps de simulation. En revanche, la distance entre les boucles L1 et L3, permettant de suivre l'ouverture de l'autre entrée possible de la cavité, ne varie pas au cours de la simulation.

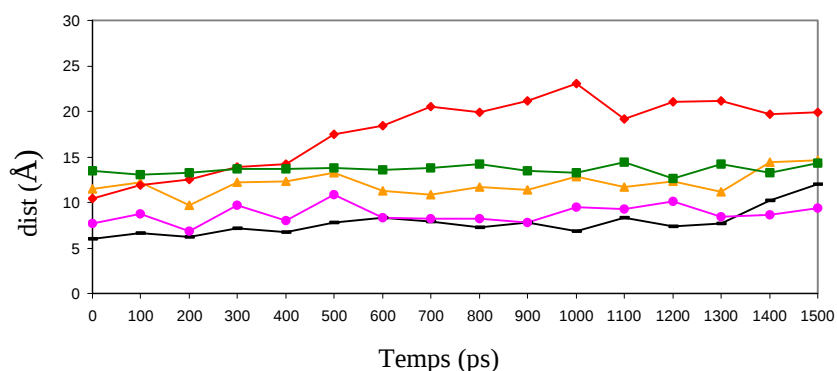


Figure 4.18 : Distances interatomiques entre différentes régions de la LTP1_1 sur une série de conformations séparées de 100 ps : en — distance H1-H2, $C_{\alpha}(11)$ - $C_{\alpha}(31)$; en • distance H1-H3, $C_{\alpha}(11)$ - $C_{\alpha}(51)$; en ▲ distance H2-H3, $C_{\alpha}(31)$ - $C_{\alpha}(51)$; en ◆ distance L2-C-term, $C_{\alpha}40$ (L2)- $C_{\alpha}78$ (C-term) et en ■ distance L1-L3, $C_{\alpha}(20)$ - $C_{\alpha}(60)$

3.2 Discussion

La simulation de la dynamique moléculaire à température ambiante de la LTP1_1 de tabac a mis en évidence la remarquable stabilité de sa structure tridimensionnelle. Ses structures secondaires, ainsi que sa cavité hydrophobe, sont présentes tout au long de la dynamique moléculaire. Les 4 hélices α constituent les régions de la protéine qui possèdent les valeurs de RMSD sur les atomes C_{α} les plus faibles, avec des fluctuations allant de 0,3 à

0,7 Å. Le segment C-terminal (Cys73-Gln91) montre une flexibilité importante avec un RMSD moyen de 1,5 Å en accord avec les résultats de notre étude RMN sur la dynamique interne du squelette de la LTP1_1. Cette flexibilité au niveau de la région C-terminale entraîne une variation du volume de la cavité entre 100 et 300 Å³. A titre comparatif, les cavités calculées pour l'ensemble des conformations de la LTP1_1 déduites des données RMN (cf. Chapitre 3) possèdent une taille moyenne de 320 ± 40 Å³. Ces variations traduisent une « respiration » de la protéine entraînant par conséquent une augmentation/diminution de la taille de la cavité.

Les simulations de la dynamique de la LTP1_1 effectuées à haute température nous ont permis de localiser les régions de la LTP1_1 qui sont susceptibles de subir des mouvements de large amplitude. Le cœur hydrophobe de la protéine reste très stable durant les premières 1,5 ns de la simulation, les distances caractéristiques mesurées entre les hélices H1, H2 et H3, (figure 4.18) ne variant que faiblement. En revanche, le segment C-terminal s'écarte très rapidement de sa position initiale avec un soulèvement de la région comportant les résidus Val75 à Ser84. La distance C_α40 (L2)-C_α78 (C-term) qui quantifie ce mouvement augmente brusquement à partir de 500 ps de simulation pour atteindre une valeur de 23 Å à 1 ns de simulation et se stabiliser ensuite autour d'une valeur de 20 Å. On observe donc une ouverture de la cavité entre la boucle L2 et la région C-terminale incluant plus particulièrement les résidus Val75 à Ser84. On observe notamment que le résidu Tyr79 obstruant initialement l'entrée de la cavité se déplace très fortement. Ils se trouve alors totalement exposé et libère l'ouverture de la cavité. Une étude de dynamique moléculaire à haute température réalisée sur la LTP1 de blé (Sy et al. 2003) avait déjà mis en évidence l'importance de ce résidu tyrosine dans l'ouverture de la cavité (figure 4.17 B). Mais cette étude avait révélé l'ouverture d'une deuxième entrée aux alentours de 1 ns de simulation (figure 4.17 B) mettant en jeu l'écartement des boucles L1 et L3 et conduisant à une augmentation considérable de la distance C_α20 (L1)-C_α60 (L3) passant de 12 Å à plus de 30 Å après 1 ns de simulations (Sy et al. 2003). Dans la simulation de la LTP1_1 de tabac une telle ouverture n'est pas observée et la distance C_α20 (L1)-C_α60 (L3) reste stable (~ 13 Å) (figures 4.17 A et 4.18). Les résultats de la simulation de la dynamique de la LTP1 de blé sont corrélés aux études expérimentales qui ont montré la capacité des LTP1 de graines de céréales à fixer 2 molécules amphiphiles. En particulier, la structure cristallographique (Charvolin et al. 1999) de la LTP1 de blé révèle que 2 phospholipides (LMPC) s'insèrent tête bêche dans la cavité. De même, la structure cristallographique de la LTP1 de riz en complexe

avec l'acide palmitique établie récemment par l'équipe de Y.J. Sun (Cheng et al. 2004b) met en évidence une organisation identique des deux lipides dans la cavité. Contrairement aux LTP1 de graines de céréales, les résultats de simulation à haute température pour la LTP1_1 de tabac montrent l'existence d'une unique entrée pour la cavité. Ces résultats confirment ceux observés lors de l'étude de la dynamique interne par RMN du squelette peptidique de la LTP1_1 qui montrent que les régions présentant une flexibilité notable sont la boucle L2 (Arg39, Thr40) et le segment C-terminal avec les résidus Ile77, Tyr79 et Ile81. De plus ils permettent de valider l'hypothèse que nous avons émise de l'existence d'une seule entrée de la cavité de la LTP1_1 de tabac pour expliquer ses propriétés particulières de complexation.

4. Conclusion et perspectives

Dans ce travail, nous avons étudié la dynamique interne de la LTP1_1 par RMN et par simulation de dynamique moléculaire à température ambiante et à haute température. Notre objectif était de déterminer les propriétés dynamiques du squelette peptidique de la LTP1_1 et d'identifier les régions de la protéine impliquées dans la fixation de molécules amphiphiles telles que les phospholipides.

Dans un premier temps, les études de dynamique interne par RMN du squelette peptidique de la LTP1_1 et de simulation de dynamique moléculaire à température ambiante ont montré une remarquable stabilité de la structure 3D de la protéine. En effet, en dehors de quelques modifications mineures de la structure 3D, la seule région présentant une certaine flexibilité est le segment C-terminal (73-91). L'analyse de la dynamique interne par RMN du squelette peptidique de la LTP1_1 a révélé des mouvements moléculaires internes de grande amplitude dans la gamme de la picoseconde à la nanoseconde pour les résidus impliqués dans la définition de l'entrée de la cavité (boucle L2 et segment C-terminal). Toutefois, l'analyse dynamique que nous avons réalisée à un seul champ magnétique et à une seule température n'est pas suffisante pour déterminer d'éventuels mouvements du squelette dans la gamme de temps allant de la microseconde à la milliseconde. Une étude plus approfondie de la dynamique interne de la LTP1_1 à différents champs magnétiques, mais également à différentes températures, pourraient révéler ce type de mouvement.

Les simulations de dynamique moléculaire à haute température ont également mis en évidence les régions de la protéine qui sont les plus stables et celles qui présentent des mouvements de plus grande amplitude. On a remarqué une forte déstructuration des structures

secondaires, notamment pour les hélices H2 et H4 qui se déplient très rapidement. Au niveau des entrées possibles de la cavité, nous avons montré que la LTP1_1 de tabac possède une seule et unique ouverture alors que celles des graines de céréale en possèdent deux. Cependant, quelle que soit la LTP1 considérée, on remarque le rôle essentiel de la tyrosine 79, très conservée parmi les LTP1, dans l'ouverture de l'entrée commune des cavités de ces LTP1 de graine de céréale.

Afin de compléter les résultats que nous avons obtenus par RMN et *in silico* sur la dynamique de la LTP1_1 de tabac, des études semblables à celles que nous avons effectuées pourraient être menées sur des complexes de la LTP1_1 impliquant des molécules lipidiques comme par exemple des acides gras ou, dans l'idéal, un ligand naturel que nous aurions identifié. Nous pourrions ainsi déterminer les résidus responsables de la stabilisation du complexe et des différentes affinités observées pour cette LTP1.

V) Cartographie de la cavité hydrophobe de la LTP1_1

Comme les LTP1 extraites de graines de céréales, la LTP1_1 de tabac possède une cavité hydrophobe qui lui permet de complexer des lipides. Cependant, comme nous l'avons montré, ses propriétés de fixation diffèrent de celles des LTP1 de graines de céréales. Il nous a donc semblé important d'étudier sa cavité hydrophobe afin de cartographier l'accessibilité des résidus qui la tapissent.

Actuellement, peu de méthodes ont été développées pour explorer les cavités hydrophobes des protéines. Parmi celles-ci, on compte l'étude par diffraction des RX d'un monocristal de protéine sous pression de xénon ou de krypton (Prangé et al. 1998) et la mesure par RMN des vitesses de relaxation croisée entre des protons de la protéine et ceux de molécules organiques dissoutes, comme par exemple le méthane ou le cyclopropane (Otting et al. 1997). Dans la méthode de diffraction des RX, nous sommes confrontés au problème de l'adéquation de la taille de la sonde à celle de la cavité hydrophobe. Par exemple une sonde de taille trop petite par rapport à celle de la cavité diffusera trop rapidement et ne sera pas observable. D'autre part, si la taille de la sonde est adéquate, avec cette technique nous avons une vision très statique de l'interaction sonde/cavité. La deuxième méthode nécessite l'emploi de pressions notables afin de solubiliser ces petits composés hydrophobes et on risque ainsi de modifier les dimensions et l'accessibilité de la cavité. Dans ce travail nous avons choisi d'explorer la cavité de la LTP1_1 en utilisant la RMN du xénon polarisé par laser, puisque cette méthode permet l'étude de la protéine à température et pression ambiantes et de la dynamique du xénon au sein de la cavité.

Le xénon, grâce à son caractère hydrophobe, apparaît comme une sonde idéale des cavités hydrophobes des protéines (Schiltz et al. 2003). L'interaction du xénon avec les sites hydrophobes peut être révélée par RMN en analysant les variations des déplacements chimiques ^1H , ^{15}N et ^{13}C et/ou la dynamique interne de la protéine, mais également le déplacement chimique et les vitesses de relaxation du xénon. Ces deux derniers paramètres sont facilement mesurables grâce à l'emploi de xénon hyperpolarisé (cf § 1). De plus, cette technique permet de déterminer les distances $^{129}\text{Xe}-^1\text{H}$ entre les xénons et les protons de la protéine grâce à l'effet SPINOE (pour Spin Polarisation Induced NOE) (Navon et al. 1996).

Dans la suite de ce chapitre, je vais présenter le principe de la RMN du xénon polarisé par laser et l'étude que nous avons réalisée sur la LTP1_1. Ces travaux ont été effectués en collaboration avec Hervé Desvaux, Patrick Bertault, Gaspard Huber et Lionel Dubois du CEA de Saclay.

1. Xénon polarisé par laser

L'un des inconvénients de la RMN est son manque de sensibilité (figure 5.3). Celle-ci est proportionnelle à la polarisation de spin nucléaire P donnée par :

$$P = \frac{n_\alpha - n_\beta}{n_\alpha + n_\beta} \approx \frac{\gamma \hbar B_0}{2kT} \quad (5.1)$$

où n_α et n_β correspondent aux populations des états de spins α et β , γ est le rapport gyromagnétique du noyau considéré, B_0 le champ magnétique statique, T la température, k la constante de Boltzmann et $\hbar = \frac{h}{2\pi}$ avec h constante de Planck.

On remarque que la polarisation nucléaire P est proportionnelle au champ magnétique appliqué B_0 et à l'inverse de la température. Pour augmenter la polarisation nucléaire P , on ne peut intervenir que sur ces deux paramètres puisque le rapport gyromagnétique γ est une constante propre à chaque noyau. Pour le champ statique B_0 , nous sommes limités par les valeurs des champs critiques des alliages supraconducteurs et actuellement ces valeurs plafonnent vers 25 T à la température de 2 K. Au niveau de l'échantillon, nous sommes restreints à une gamme de température compatible avec les processus biologiques. Nous ne pouvons donc pas, à l'aide de ces deux paramètres, augmenter sensiblement P par une approche thermodynamique.

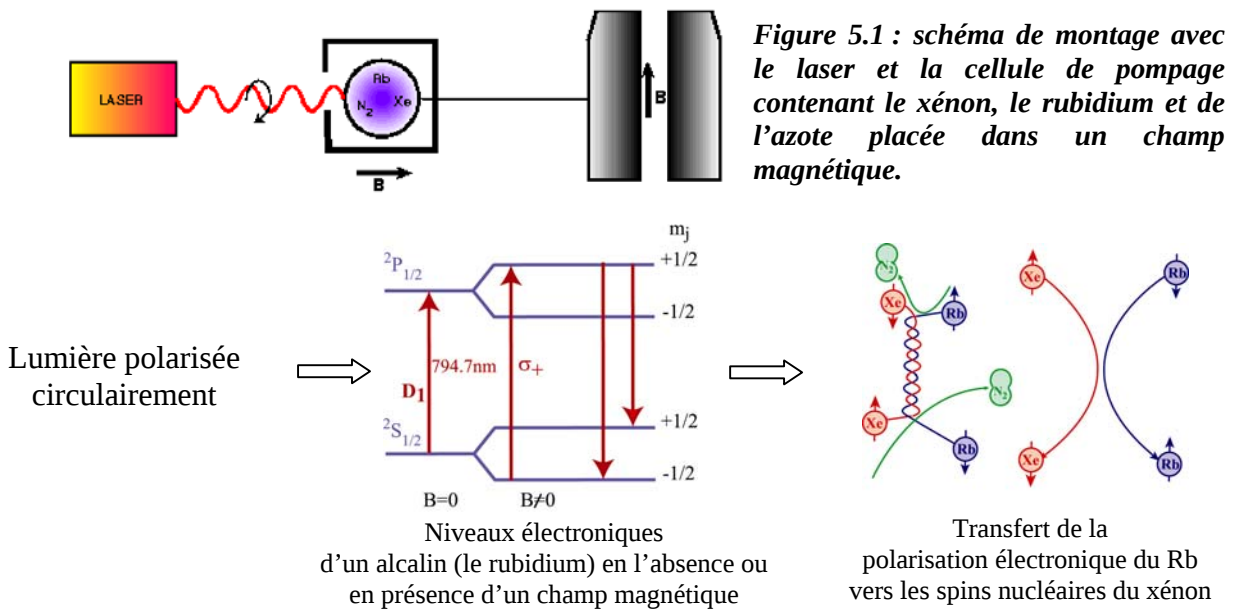


Figure 5.2 : Principe du pompage optique pour augmenter la polarisation nucléaire P du ^{129}Xe . La lumière polarisée circulairement est issue d'un laser titane-saphir et l'alcalin utilisé est le rubidium.

L'expérience de pompage optique (Walker and Happer 1997) permet d'augmenter de plusieurs ordres de grandeur la polarisation nucléaire P du xénon ^{129}Xe . Le principe de cette expérience, illustré figures 5.1 et 5.2, consiste à créer une différence de population entre les états magnétiques d'un alcalin (le rubidium, Rb) et à transférer cette polarisation vers le xénon. En pratique, en présence d'un champ magnétique $\vec{B}$, les deux niveaux d'énergie les plus bas $^2S_{1/2}$ et $^2P_{1/2}$ du rubidium se séparent en deux sous niveaux. Sous l'effet d'un rayonnement laser polarisé circulairement (σ^+) de longueur d'onde adaptée, on induit uniquement des transitions entre le sous niveau $m_j = -1/2$ de $^2S_{1/2}$ et le sous niveau $m_j = +1/2$ de $^2P_{1/2}$. Le système se désexcite très rapidement et retourne à son état fondamental (transitions de $m_j = +1/2$ de $^2P_{1/2}$ vers $m_j = +1/2$ et $-1/2$ de $^2S_{1/2}$ (cf. figure 5.2)). Comme on pompe en permanence le niveau d'énergie $m_j = -1/2$ de $^2S_{1/2}$, l'état $+1/2$ devient totalement peuplé. Les spins électroniques du Rb sont ainsi totalement polarisés et, par relaxation croisée, transfèrent leur polarisation vers les niveaux d'énergie voisins des spins nucléaires du xénon contenu dans la cellule (figure 5.1). Comme les spins électroniques du Rb sont polarisés en permanence par pompage optique, on augmente de fait la polarisation du xénon. La limite de cette méthode est donc la vitesse de pompage optique par rapport à celle de relaxation de l'électron (équilibre entre les états $m_j = +1/2$ et $-1/2$ de $^2S_{1/2}$).

Par ce principe, nos collaborateurs ont atteint une polarisation du xénon P égale à 0,30, à comparer avec la valeur à l'équilibre thermodynamique à 11,7 T qui est de $1,2 \cdot 10^{-5}$, soit une augmentation de P d'un facteur 25000 (figure 5.3).

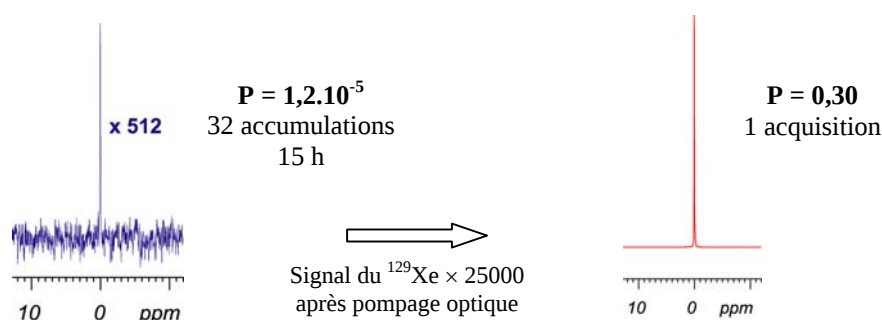


Figure 5.3 : **Signaux du ^{129}Xe polarisé par laser (à droite) ou non (à gauche).** Le signal est augmenté d'un facteur 25000.

Cette amélioration de sensibilité du ^{129}Xe obtenue permet de mesurer facilement son déplacement chimique et donc ses paramètres de relaxation, très sensibles à l'environnement chimique (Cherubini and Bifone 2003). Elle permet aussi d'envisager la mesure des NOEs

hétéronucléaires ^{129}Xe - ^1H à l'aide de la méthode du SPINOE et d'explorer au niveau atomique les cavités hydrophobes.

2. Le SPINOE

Le SPINOE a été observé pour la première fois par l'équipe de A. Pines en 1996 (Navon et al. 1996) qui a remarqué un transfert de polarisation du ^{129}Xe hyperpolarisé aux spins voisins, par relaxation croisée, sans irradiation préalable des spins du ^{129}Xe . Pour essayer de comprendre intuitivement cet effet, on peut considérer le xénon polarisé par laser comme un système possédant des spins nucléaires à très basse température (Dubois et al. 2004). En effet, d'après l'équation (5.1) la polarisation nucléaire P est inversement proportionnelle à T . Par exemple dans notre cas, la polarisation nucléaire P de 30 % de notre échantillon de xénon, à un champ magnétique de 11,7 T, correspond à une température des spins nucléaires du xénon de 11×10^{-3} K. Les spins nucléaires voisins de ceux du xénon qui sont initialement à température ambiante sont alors refroidis par relaxation dipolaire croisée et par conséquent, on observe une variation de leur polarisation (dans un sens qui dépend du signe de la polarisation du xénon). Le principe de l'expérience SPINOE est de mesurer la variation de température de spins des protons.

Cette découverte a ouvert de nombreuses perspectives, dont la caractérisation de cavités hydrophobes de protéines à pression et température ambiantes. Le principe consiste à réunir les conditions de mise en œuvre du transfert de polarisation du ^{129}Xe vers des protons proches afin de localiser les acides aminés tapissant la cavité. La première observation de transfert de polarisation du ^{129}Xe hyperpolarisé au sein d'une cavité hydrophobe a été obtenue par Hervé Desvaux et Patrick Bertault en collaboration avec notre équipe dans la LTP1 de blé non marquée (Landon et al. 2001). Depuis, les conditions de mise en œuvre de l'expérience SPINOE ont été optimisées et le marquage isotopique ^{13}C et ^{15}N de la LTP1_1 de tabac nous a permis de réaliser une étude bien plus complète.

3. Etude de la cavité hydrophobe de la LTP1_1 de tabac

L'étude des propriétés structurales, thermodynamiques et dynamiques de la cavité hydrophobe de la LTP1_1 du tabac *Nicotiana tabacum* a été réalisée sur un échantillon de LTP1_1 100 % marquée ^{15}N et sur un échantillon marqué 10 % ^{13}C et 100 % ^{15}N . Cette étude

a donc nécessité non seulement l'attribution préalable des déplacements chimiques ^1H et ^{15}N utilisés pour la détermination de la structure 3D de la LTP1_1 mais aussi l'attribution ^{13}C . L'essentiel de ces résultats est présenté sous la forme d'un article intitulé « Probing hydrophobic cavities of Lipid Transfer Protein from *Nicotiana tabacum* through Xenon-based NMR » qui a été accepté pour publication dans la revue : *Journal of the American Chemical Society*.

L'expérience SPINOE a conduit à l'identification des sites de fixation du xénon et des résidus tapissant la cavité hydrophobe. La réalisation de cette expérience pour la première fois dans l'eau légère a permis de montrer que seuls les protons des chaînes hydrophobes sont impliqués dans l'interaction directe avec les atomes de ^{129}Xe . De plus, L'analyse des déplacements chimiques ^1H , ^{13}C et ^{15}N de la LTP1_1 en fonction de différentes concentrations de xénon a permis de déterminer les constantes d'affinité du xénon pour ses différents sites au sein de la cavité hydrophobe. Enfin, la détermination des paramètres de relaxation (relaxation longitudinale du xénon et transversale des noyaux ^{15}N de la LTP1_1) a permis une meilleure compréhension de la dynamique du xénon au sein de la cavité de la LTP1_1.

Probing the Hydrophobic Cavity of Lipid Transfer Protein from *Nicotiana tabacum* through Xenon-Based NMR Spectroscopy

Lionel Dubois,[†] Pedro Da Silva,[‡] Céline Landon,[‡] J. Gaspard Huber,[†]
Michel Ponchet,[§] Françoise Vovelle,[‡] Patrick Berthault,[†] and Hervé Desvaux^{*,†}

Contribution from DSM/DRECAM/Service de Chimie Moléculaire, URA CEA/CNRS 331 Claude Fréjacques, C.E.A./Saclay, F-91191 Gif sur Yvette, France, Centre de Biophysique Moléculaire, UPR 4301 CNRS, affiliated with Orléans University, Rue Charles Sadron, 45071 Orléans Cedex 2, France, and UMR Interactions Plantes-Microorganismes et Santé Végétale, INRA, 400 route des Chappes, BP 167, 06903 Sophia Antipolis Cedex, France

Received June 28, 2004; E-mail: hdesvaux@cea.fr

Abstract: The hydrophobic cavity of Lipid Transfer Protein 1 from *Nicotiana tabacum* is investigated in detail by NMR using xenon as a spy. The analysis of the ^{129}Xe chemical shifts and self-relaxation times gives evidence of protein–xenon interaction. Thermodynamics of the binding is characterized through the study of aliphatic ^1H and ^{13}C chemical shift variation as a function of xenon pressure. The binding constant is evaluated to $75.5 \pm 1.0 \text{ M}^{-1}$ at 293 K. The location of xenon inside the cavity is deduced from SPINOE experiments. The noble gas appears to occupy four sites, and xenon self-relaxation experiments indicate that it quickly jumps between different sites. The chemical shifts of amide protons and nitrogens also depend on the xenon concentration, either specifically or nonspecifically for atoms at the external surface of the protein. Yet, contrary to aliphatic atoms, they do not correspond to short-range interactions as confirmed by magnetization transfer experiments between laser-polarized xenon and protons in H_2O . These ^{15}N chemical shift variations, used in combination with ^{15}N transverse self-relaxation rates to determine the lower limit of the binding rate, consequently reveal subtle changes in the structure of the protein upon binding.

I. Introduction

In the past, only a few experimental methods have been proposed to characterize and investigate hydrophobic sites or surfaces of proteins at the atomic scale. The use of small probes (“spy” molecules) designed to interact with these sites to describe them is an elegant approach. However, the results are strongly dependent on the size of the spy and on its affinity for the sites, which is related to the van der Waals energy, the variation of the electrostatic contribution (water desorption, for example), but also of the entropy contribution due to degeneracy of the location of the spy.

Actually, the approaches for detecting the interaction between the spy and the protein atoms can be separated into two classes, direct methods or indirect ones, according to their capability to localize the spy or not. In the first case, the observed effect is directly dependent on the distance between the probe and the protein atoms. Methods such as X-ray diffraction on protein mono-crystals under high pressure of noble gas,^{1–4} NMR measurements of homonuclear dipolar cross-relaxation between

small organic compounds and protein protons,⁵ or NMR methods exploiting the relaxation enhancement resulting from paramagnetic probes⁶ belong to this class. Variations of the fluorescence properties or observation of the spy chemical shifts are indirect processes. This second class of approaches is usually adequate to extract data on the thermodynamics of the interaction through the determination of binding constants. The monitoring of the protein ^1H , ^{13}C , or ^{15}N chemical shifts as a function of the probe and protein concentrations belong to both classes. They are direct parameters if they can be easily transformed into distances to the spy atoms (in the case of chemical shift reagents, for instance⁷). On the other hand, binding can cause structural changes distant from the interaction site, rendering the mapping only via chemical shifts difficult.⁸

Xenon appears as a promising spy of hydrophobic regions.⁴ It has been shown, by X-ray diffraction on protein crystals

[†] DSM/DRECAM/Service de Chimie Moléculaire.

[‡] Centre de Biophysique Moléculaire.

[§] INRA.

(1) Schoenborn, B. *Nature* **1965**, 208, 760–762.

(2) Prangé, T.; Schiltz, M.; Pernot, L.; Colloc'h, N.; Longhi, S.; Bourguet, W.; Fourme, R. *Proteins: Struct., Funct., Genet.* **1998**, 30, 61–73.

(3) Quillin, M. L.; Breyer, W. A.; Grisworld, I. J.; Matthews, B. W. *J. Mol. Biol.* **2000**, 302, 955–977.

(4) Schiltz, M.; Fourme, R.; Prangé, T. *Methods Enzymol.* **2003**, 374, 83–119.

(5) Otting, G.; Liepinsh, E.; Halle, B.; Frey, U. *Nat. Struct. Biol.* **1997**, 4, 396–404.

(6) Pintacuda, G.; Otting, G. *J. Am. Chem. Soc.* **2002**, 124, 372–373.

(7) Banci, L.; Bertini, I.; Luchinat, C. *Nuclear and Electron Relaxation*; VCH: Weinheim, 1991.

(8) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy*; Academic Press: San Diego, CA, 1996.

pressurized by xenon, that the polarizability of its electronic cloud makes it a selective probe of hydrophobic cavities, fitting its van der Waals diameter (4.3 Å). This observation is reinforced by molecular dynamics simulations, although they reveal a large amplitude of xenon dynamics inside cavities.⁹ The richness of NMR to tackle molecular dynamics certainly explains why many procedures using this spectroscopy are employed to characterize interaction sites between protein and xenon. The existence of hydrophobic sites can be detected through the influence of xenon concentration on the protein ¹H and ¹⁵N chemical shifts,^{10,11} on the dynamics of the protein,¹² or on the ¹²⁹Xe chemical shift^{13–15} or longitudinal self-relaxation time.^{16,17} The availability of laser-polarized xenon makes the last two parameters easy to monitor, but the possible contribution from nonspecific interactions limits their interpretation.¹³ Nevertheless, the use of laser-polarized xenon also enables determination of ¹²⁹Xe—¹H distances through magnetization transfer and then leads to the location of the binding sites. This so-called SPINOE was initially observed on partially deuterated benzene.¹⁸ Its first observation on a protein dissolved in D₂O^{19,20} has led to the determination of three binding sites of xenon in the wheat Lipid Transfer Protein 1.

In this paper, a combined use of NMR techniques has been undertaken to derive the structural, thermodynamical, and dynamical properties of the hydrophobic cavity of a Lipid Transfer Protein, from *Nicotiana tabacum* (tobacco LTP1•1). ¹H, ¹³C, and ¹⁵N NMR experiments relate protein chemical shift variations to the xenon concentration, allowing one to determine the affinity of xenon to the different sites. Dipolar heteronuclear cross-relaxation experiments, enhanced by the high magnetization of laser-polarized xenon, allow the observation of polarization transfer to nearby protons to locate the noble gas atoms inside the cavity. For the first time, it has been possible to perform indifferently these experiments in D₂O or in H₂O. These experiments focus on direct effects and therefore helpfully complement the chemical shift mapping. Finally, information about dynamics extracted from xenon self-relaxation times and ¹⁵N transverse relaxation times allows a better understanding of the complete phenomenon of xenon binding inside tobacco LTP1•1.

Lipid Transfer Proteins constitute an exciting class of proteins for such a study, as their structures contain hydrophobic cavities that can accommodate various lipidic molecules. The 3D structure of the recombinant LTP1•1 from tobacco shoot apex has been recently determined using 2D and 3D NMR experi-

ments.²¹ Despite a very similar global fold, the wheat LTP1 and tobacco LTP1•1 present different lipid binding properties which can be related to subtle structural changes. Wheat LTP1 is able to bind two LMPC (lysomyristoylphosphatidylcholine) molecules, whereas tobacco LTP1•1 binds only one such lipid.

II. Materials and Methods

A. Protein Samples Preparation. Lipid Transfer Protein from *Nicotiana tabacum* was produced and purified as described in ref 21. Different samples were available: a classical nonenriched protein for NMR experiments where isotope edition was useless, a 100% ¹⁵N-enriched sample for the study of amide ¹H and ¹⁵N chemical shift variation under xenon pressure, and a 10% ¹³C-, 100% ¹⁵N-enriched protein for ¹H and ¹³C chemical shift variation under xenon pressure.

B. Protein NMR Experiments. 1. NMR Spectrometers and Peak Assignment. All NMR experiments were performed at 293 K on Bruker DRX500 and DRX600 spectrometers running at 500 MHz (11.7 T) and 600 MHz (14 T) proton resonance frequencies, respectively. The protein ¹H, ¹³C, and ¹⁵N chemical shift assignment was deduced from ref 21 using HSQC and HSQC-TOCSY experiments. For the NMR experiments, the protein was dissolved in 99.97% D₂O (Eurisotop) except when amide protons were monitored (solvent H₂O/D₂O 90:10). The NMR tubes were 5 mm o.d. medium-size walled tubes from Wilmad and Cortec, closed with J. Young valves. After introduction of the protein solution, the tubes were carefully degassed by at least three freeze—pump—thaw cycles.

2. Variation of Protein Atom Chemical Shifts under Xenon Pressure. The protein chemical shift variation under xenon pressure was monitored using HSQC experiments with sensitivity improvement by gradients.²² The experiments were run on the DRX600 spectrometer equipped with a HCN cryoprobehead (¹H—¹⁵N HSQC, protein concentrations: 0.75, 1.0, and 1.5 mM) or on the DRX500 spectrometer (¹H—¹³C HSQC, protein concentration 3.2 mM) using a HCN inverse probehead equipped with xyz gradients. These last experiments were optimized (resonance frequencies and ¹J_{CH} scalar couplings) for the aliphatic region.

Xenon from Air Liquide was added to the NMR tube containing the protein using two different protocols. For 15 out of the 19 ¹⁵N experiments (pressure ranging between 0 and 6.9 bar), a known amount of xenon was added to the tube by condensation above the frozen solution. For the four last experiments (between 0 and 1.5 bar), a xenon static pressure was directly applied to the solution. The same technique, more reliable and precise in particular for low pressures, was used for the ¹³C experiments (12 pressures ranging between 0 and 5.9 bar).

Peak picking and assignment of the 2D-HSQC experiments were performed using SPARKY.²³ The extracted frequency lists were further processed to determine the binding constant using a home-written C program with nonlinear least-squares fitting based on the Levenberg Marquardt algorithm.²⁴ The initial conditions of the fitting procedure were optimized by a grid search. Estimation of the errors on the fitted parameters was determined by Monte Carlo simulations using the experimental uncertainties on the frequency of the peak maximum.

C. Laser-Polarized Xenon Experiments. 1. Preparation of Laser-Polarized Xenon. Xenon was polarized by the spin-exchange method²⁵ with the apparatus described in detail in ref 26. In comparison to this paper, the volume of the pumping cell was increased (inner diameter

- (9) Mann, G.; Hermans, J. *J. Mol. Biol.* **2000**, *302*, 979–989.
- (10) Rubin, S. M.; Lee, S.-Y.; Ruiz, E. J.; Pines, A.; Wemmer, D. E. *J. Mol. Biol.* **2002**, *322*, 425–440.
- (11) Gröger, C.; Möglich, A.; Pons, M.; Koch, B.; Hengstenberg, W.; Kalbitzer, H. R.; Brunner, E. *J. Am. Chem. Soc.* **2003**, *125*, 8726–8727.
- (12) Mulder, F. A. A.; Hon, B.; Muhandiram, D. R.; Dahlquist, F. W.; Kay, L. E. *Biochemistry* **2000**, *39*, 12614–12622.
- (13) Rubin, S. M.; Spence, M. M.; Goodson, B. M.; Wemmer, D. E.; Pines, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9472–9475.
- (14) Rubin, S. M.; Spence, M. M.; Pines, A.; Wemmer, D. E. *J. Magn. Reson.* **2001**, *152*, 79–86.
- (15) Locci, E.; Dehouck, Y.; Casu, M.; Saba, G.; Lai, A.; Lühmer, M.; Reisse, J.; Bartik, K. *J. Magn. Reson.* **2001**, *150*, 167–174.
- (16) Locci, E.; Casu, M.; Saba, G.; Lai, A.; Reisse, J.; Bartik, K. *ChemPhysChem* **2002**, *9*, 812–814.
- (17) Dubois, L.; Parrès, S.; Huber, J. G.; Berthault, P.; Desvaux, H. *J. Phys. Chem. B* **2004**, *108*, 767–773 *ibid* 15952.
- (18) Navon, G.; Song, Y.-Q.; Rööm, T.; Appelt, S.; Taylor, R. E.; Pines, A. *Science* **1996**, *271*, 1848–1851.
- (19) Berthault, P.; Landon, C.; Vovelle, F.; Desvaux, H. *C. R. Acad. Sci. Ser. IV* **2001**, *2*, 327–332.
- (20) Landon, C.; Berthault, P.; Vovelle, F.; Desvaux, H. *Protein Sci.* **2001**, *10*, 762–770.

- (21) Da Silva, P.; Landon, C.; Industri, B.; Marais, A.; Ponchet, M.; Vovelle, F. *Proteins: Struct., Funct., Bioinf.* **2005**, in press.
- (22) Cavanagh, J.; Rance, M. *Annu. Rep. NMR Spectrosc.* **1993**, *27*, 1–58.
- (23) Goddard, T. D.; Kneller, D. G. *SPARKY 3*; University of California, San Francisco.
- (24) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C. The art of scientific programming*; Cambridge University Press: Cambridge, 1988.
- (25) Walker, T. G.; Happer, W. *Rev. Mod. Phys.* **1997**, *69*, 629–642.
- (26) Desvaux, H.; Gautier, T.; Le Goff, G.; Péro, M.; Berthault, P. *Eur. Phys. J. D* **2000**, *12*, 289–296.

22 mm, length 130 mm), and a gas mixture composed of 10–40 Torr of xenon and 200 Torr of nitrogen was used. After about 10 min of optical pumping at 368 K using a 4.5 W titanium:sapphire laser, polarized xenon was condensed and thus separated from nitrogen in a solenoid immersed in liquid nitrogen. The gas was then transferred in the fringe field of the NMR magnet to the previously degassed NMR tube. Using this protocol, with 15 Torr of xenon inside the pumping cell, we obtained more than 1 bar of xenon inside the NMR tube with a useful polarization ranging between 20% and 30%.

2. ^{129}Xe Chemical Shift and Self-Relaxation. The study of ^{129}Xe chemical shift and self-relaxation as a function of the xenon pressure and protein concentration was performed at 500 MHz using a Nalorac direct broadband probehead. Laser-polarized noble gas was employed for these experiments. The ^{129}Xe chemical shift was referenced to the NMR signal of 1 bar of xenon gas. The longitudinal self-relaxation time T_1 was determined by a nonlinear least-squares fit of xenon signal decay as a function of time in repeated acquisitions using small-flip angle rf pulses (ca. 2.5°). For each protein concentration and xenon pressure, three different decay curves composed of 32 points were obtained, and the average self-relaxation time was used for further exploitation. After the NMR experiments, the pressure in the NMR tube was evaluated by expansion into a calibrated volume equipped with a 0–1000 Torr Ceramic capacitance manometer from Varian. The measured pressures were compared to the amount of xenon initially present in the optical pumping cell. Four different protein concentrations were used. The relative values of the two mere solutions (0.47 and 2.2 mM) were checked by UV–visible spectroscopy, and the two others were obtained by dilution by a factor of 2.

3. SPINOE Experiments. The detection of magnetization transfer between laser-polarized xenon and protein protons was obtained by SPINOE^{26,27} and SPIROE²⁸ pulse sequences. For these experiments, we took benefit from the last improvements in their implementation.²⁹ The initial conditions of the proton spin system corresponded to complete saturation obtained by a series of 32 proton 90° hard pulses, each followed by a pulsed field gradient of random amplitude. Very short recycling delays (100 ms + 220 ms of acquisition) were used. Each SPINOE spectrum, acquired in about 2 min, was constituted of 32–48 spectra depending on the mixing time values. The spectra were added in the processing step, thus enabling optimization of the number of accumulations. Before each SPINOE or SPIROE experiment, the NMR tube was vigorously shaken to refresh dissolved laser-polarized xenon by exchange from the gas phase. After reinsertion of the tube into the magnet, the xenon magnetization was monitored by a small read pulse (about 2°), and the SPINOE (or SPIROE) experiment, which started by four dummy scans, was acquired.

In polarization transfer experiments, peak assignment was obtained by comparison of the 1D SPINOE spectra to 2D TOCSY experiments acquired under exactly the same conditions (temperature, sample, spectrometer, and xenon pressure). This assignment was obviously difficult due to spectral overlap and possible artifacts. However, the recording of a large number of spectra at different mixing times (four mixing times for SPINOE ranging from 200 to 500 ms and four mixing times for SPIROE ranging from 100 to 300 ms) reduced the risk of a misinterpretation.

For the SPINOE and SPIROE experiments, the protein concentration was equal to 0.85 mM in D_2O and 0.75 mM in H_2O , and xenon enriched in isotope 129 (96%) from Chemgas was used. The typical pressure inside the NMR tube ranged between 1.5 and 3 bar. A volume of 450 μL of solution was prepared to avoid homogeneity degradation due to the foam appearing after the sample has been shaken. These experimental conditions were chosen to optimize the xenon longitudinal self-

Table 1. Structural Characterization of the Four Hydrophobic Alveoli of Tobacco LTP1•1 As Determined by Molecular Modeling^a

site	amino acid side chains bordering the alveoli
A	L11, C14, L15, L18, L51, A54, A55, I58, I61, A66, L69, I81
B	C48, K52, Y79, I81, T85, C87
C	L11, V31, L34, V35, A38, L51, L69, P70, V75, I77, Y79, I81
D	I2, V7, L34, S37, A38, I46, I47, C50

^a The lists of amino acids bordering each site are given. It is based on the 10 structural models deposited in the PDB.

relaxation times. In our present conditions, the measured T_1 was about 150 s (about 72 s for experiments in H_2O). Ten runs of optical pumping were performed (eight for D_2O and two for H_2O). The SPINOE and SPIROE spectra were acquired at 500 and 600 MHz using Bruker broadband inverse and ^{15}N -broadband inverse TBI probeheads, respectively; both are equipped with xyz gradients.

D. Molecular Mechanics. 1. Solvation Simulations. Molecular modeling was conducted with SYBYL 6.9 (TRIPOS Associate Inc.) under Linux, using the 10 refined solution structures of tobacco LTP1•1 deposited at the PDB (code 1T12). The solvation simulation was performed using the xenon parameters defined in SYBYL, considering two layers of xenon solvent around the protein, kept rigid during all of the procedure.

2. Molecular Modeling. To locate the xenon interaction sites from SPINOE data, we started from a solution structure, keeping the protein rigid except for the 24 residues experiencing SPINOE (residues I2, V7, L11, L15, L18, V31, L34, V35, I46, A47, L51, K52, A54, A55, I58, I61, A66, L69, V75, I77, Y79, I81, S82, and T85). The xenon atoms were then added. The 58 SPINOE cross-peaks were converted into distance restraints ($3.0 < r < 5.5 \text{ \AA}$) using double weighting for the 26 unambiguous restraints. When methyl groups were involved, pseudo-atoms located at the carbon coordinates were used and the range of distance restraints was enlarged ($4.0\text{--}7.5 \text{ \AA}$). The xenon binding sites were finally obtained by energy minimization under distance restraints in the TRIPOS force field.

III. Results and Discussion

A. Simulation of Solvation of *Nicotiana tabacum* LTP1•1 by Xenon. LTP1•1 from *Nicotiana tabacum* presents a large hydrophobic cavity defined by four α -helices linked together by four disulfide bridges.²¹ The volume and shape of the hydrophobic cavity obviously depend on the exact location of the amino acid side chains. Using a spherical probe of 1 $\text{\AA}$ radius in the program CAVITE,³⁰ its overall volume varies between 240 and 440 $\text{\AA}^3$ for the 10 structures deposited in the PDB. This hydrophobic volume is composed of different alveoli separated by amino acid side chains. Table 1 gives the names of the residues bordering the four main alveoli.

Based on all structural models, the four different alveoli (A, B, C, and D) are able to bind xenon (Figure 1). The first three have an equivalent in wheat LTP1, while D is specific to tobacco LTP1•1. Site A has the highest occupancy, because xenon is found in 9 out of the 10 models, and on average 1.1 xenon atoms occupy this site. Sites B and C are occupied in 5 and 7 out of the 10 models, respectively, while site D is occupied only three times. Only sites A and C can accommodate more than one xenon atom. On average, the whole cavity volume of tobacco LTP1•1 allows the binding of 2.7 xenon atoms.

B. Laser-Polarized Xenon NMR Spectroscopy. The simplest procedure to explore the interaction between xenon and tobacco LTP1•1 when laser-polarized xenon is available consists

- (27) Song, Y.-Q.; Goodson, B. M.; Taylor, R. E.; Laws, D. D.; Navon, G.; Pines, A. *Angew. Chem.* **1997**, *36*, 2368–2370.
 (28) Desvaux, H.; Huber, J. G.; Brotin, T.; Dutasta, J.-P.; Berthault, P. *ChemPhysChem* **2003**, *4*, 384–387.
 (29) Dubois, L.; Berthault, P.; Huber, J. G.; Desvaux, H. C. R. *Physique* **2004**, *5*, 305–313.

- (30) Gomar, J.; Sodano, P.; Sy, D.; Shin, D. H.; Lee, J. Y.; Suh, S. W.; Marion, D.; Vovelle, F.; Ptak, M. *Proteins* **1998**, *31*, 160–171.

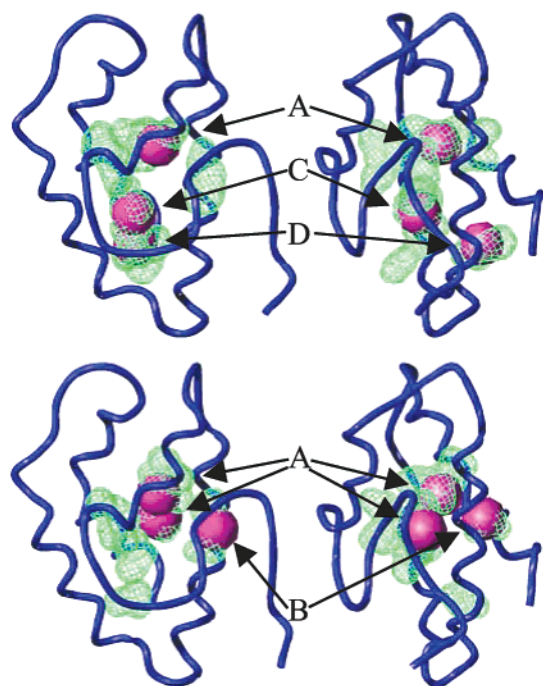


Figure 1. Examples of solvation simulation of tobacco LTP1•1 by xenon. The protein backbone is shown as a blue tube, the cavity calculated using the home-written program CAVITE is represented as green volumes, and xenon is represented by its van der Waals sphere. In model 2 (top), xenon occupies the three sites A, C, and D, and in model 3 (bottom), only sites A (doubly occupied) and B are filled with xenon.

of monitoring the xenon chemical shift and longitudinal self-relaxation time T_1 as a function of the protein concentration and xenon pressure.³¹ The use of thermally polarized xenon renders these experiments difficult to setup, as the low polarization requires long acquisition times or high xenon concentrations. In this latter case, nonspecific contributions (directly proportional to xenon concentration) may hide the specific contribution (formation of a complex).^{13–15}

In the case of tobacco LTP1•1, whatever the xenon pressure and the protein concentration, only one ^{129}Xe peak is observed, the chemical shift of which clearly depends not only on protein concentration but also on xenon pressure (Figure 2A). This proves that xenon is in fast exchange between different sites. However, no simple behavior is observed. Increasing the amount of xenon at constant protein concentration induces a downfield shift, and increasing the protein concentration at constant xenon concentration leads also to a downfield shift. Such a behavior certainly results from the superposition of nonspecific interactions which tend to increase the xenon chemical shift as observed for other proteins^{10,14,15} and specific binding with a chemical shift of bound xenon higher than that of xenon dissolved in water.

In contrast to xenon chemical shift, xenon longitudinal self-relaxation is much more dependent on specific interactions¹⁷ because the associated xenon–proton correlation times acting on this phenomenon are much longer than those associated with nonspecific interactions. Variation of xenon T_1 , being essentially due to dipolar contributions, is a function of the “residence” time (Xe–H dipolar correlation time) and scales in r^{-6} , where r is the xenon–proton distance. Xenon T_1 thus represents

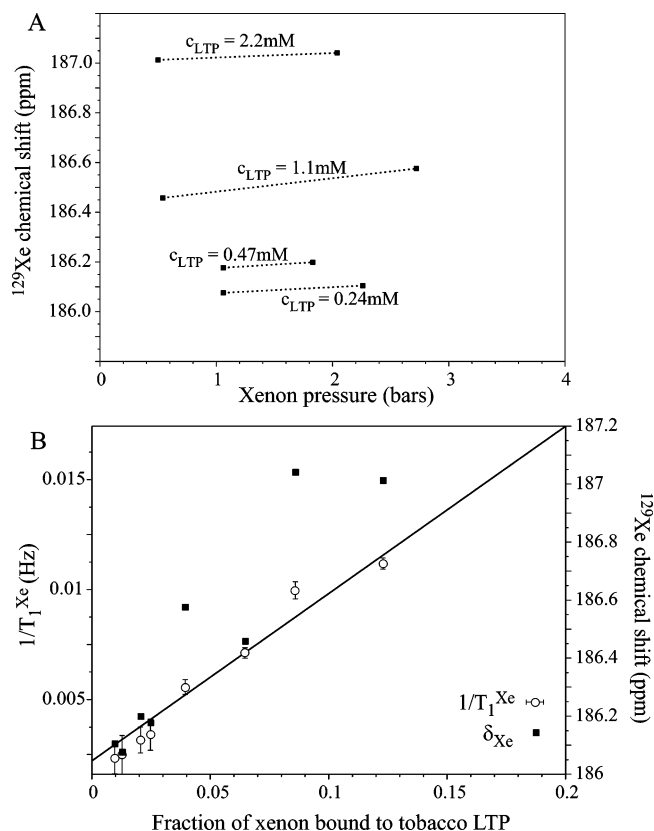


Figure 2. (A) Xenon chemical shift as a function of the protein concentration and xenon pressure. (B) Xenon T_1 and chemical shift for these protein concentrations and xenon pressures represented as a function of the fraction of xenon bound to the protein. These values are computed using $K = 75.5 \text{ M}^{-1}$, a result found in paragraph III.C. The best-fit theoretical curve to eq 1 is superimposed.

specific short-range interactions. Figure 2 shows the dependence of $1/T_1$ as a function of the fraction of bound xenon f , which depends on the xenon pressure and protein concentration:

$$\frac{1}{T_{1\text{measured}}} = f \frac{1}{T_{1\text{complex}}} + (1 - f) \frac{1}{T_{1\text{free}}} \quad (1)$$

Using a binding constant of 75.5 M^{-1} (vide infra), the extracted $T_{1\text{complex}}$ of xenon bound to tobacco LTP1•1 is $13.1 \pm 1.0 \text{ s}$, a value in good agreement with that obtained for wheat LTP1 at the same magnetic field.¹⁷ This would imply similar dipolar contribution and dynamics of xenon inside the cavity.

This analysis of xenon NMR parameters is made easily accessible thanks to laser-polarized xenon. In the present case, it reveals the existence of an interaction between the noble gas and tobacco LTP1•1. Nevertheless, even if the binding constant can be estimated from xenon chemical shift¹³ and T_1 measurements,¹⁷ these simple ^{129}Xe NMR experiments suffer from the inability to characterize the structural elements of the protein involved in the binding of xenon.

C. Protein ^1H and ^{13}C Chemical Shifts under Variable Xenon Pressure. The interaction between xenon and tobacco LTP1•1 has been explored by monitoring the protein nuclei. Twelve ^1H – ^{13}C HSQC spectra under xenon pressures varying between 0 and 5.9 bar have been acquired in D_2O . As illustrated in Figure 3, the resonance frequencies of some peaks are affected by the presence of xenon, but no new signal appears. This confirms a fast exchange situation for xenon binding, with a

(31) Cherubini, A.; Bifone, A. *Prog. NMR Spectrosc.* **2003**, *42*, 1–30.

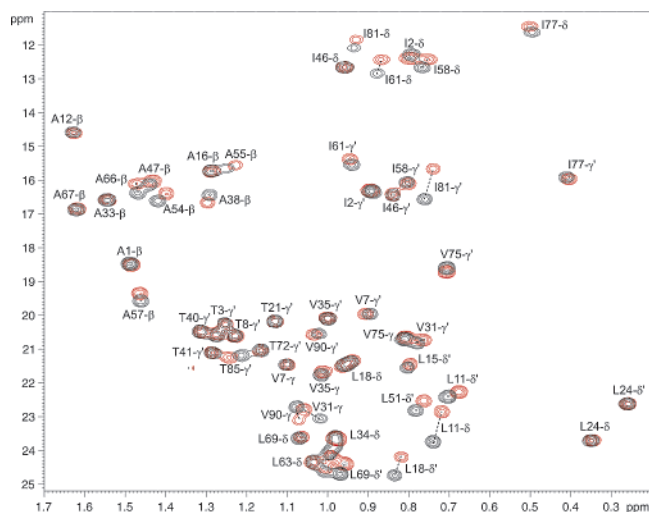


Figure 3. Comparison of ^{13}C - ^1H HSQC contour plots of tobacco LTP1•1 in the presence of 0.5 bar of xenon (red lines) and of 6 bar of xenon (black lines). Only an expansion of the aliphatic region is displayed.

residence time typically shorter than 1 ms, based on the maximum chemical shift variation observed (1.17 ppm in ^{13}C at 11.7 T).

For each peak of the ^1H - ^{13}C HSQC spectrum, the ^1H and ^{13}C chemical shift variations on the xenon concentration are fitted by a specific model of interaction and a nonspecific one. In the specific model, the formation of a local complex of stoichiometry 1:1 without cooperative effect is assumed and the chemical shift variation $\Delta\delta$ is given by

$$\Delta\delta = \frac{K[\text{Xe}]}{1 + K[\text{Xe}]}\Delta\delta_{\text{complex}} \quad (2)$$

with K as the binding constant, and $\Delta\delta_{\text{complex}}$ as the chemical shift difference between the complexed and free protein (degassed solution). The xenon concentration $[\text{Xe}]$ is related to the pressure p_{Xe} by Henry's law, the solubility of xenon in water being taken as equal to 5 mM/bar.³²

In the nonspecific model of interactions, the variation of the ^1H or ^{13}C chemical shift is considered to only result from nonspecific interactions or identically from weak binding effects. It is consequently proportional to the xenon concentration. For convenience in the units, we choose

$$\Delta\delta = \frac{[\text{Xe}]}{[\text{Xe}]_{\text{max}}} \Delta\delta_{\text{max}} \quad (3)$$

with $[\text{Xe}]_{\text{max}}$ as the maximal concentration of xenon used and $\Delta\delta_{\text{max}}$ as a value to be fitted.

The experimental variation of the proton and carbon chemical shifts as a function of xenon concentration is fitted by both models (eqs 2 and 3), and the adjustments are compared through the F-test. The specific model is retained if the probability of improvement by chance is lower than 1%. Furthermore, only peaks for which at least one chemical shift variation $\Delta\delta$ is larger than 10 times the uncertainty on the frequency determination are considered. Among the 337 ^1H and 337 ^{13}C chemical shift variations, 21 and 21, respectively, are best represented by the specific model (Figure 4). When variations of the chemical shifts

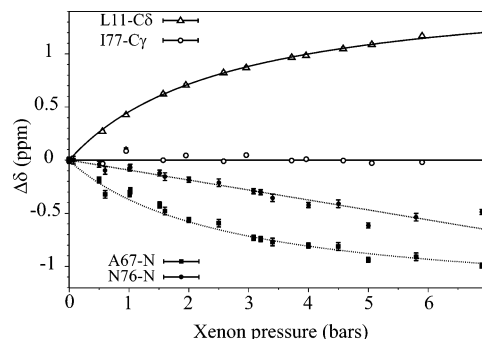


Figure 4. Variation of the ^{13}C (open symbols) and ^{15}N (close symbols) chemical shifts as a function of xenon pressure. The best-fit theoretical curves to the specific model (eq 2) considering a binding constant of 75.5 M^{-1} (L11 and A67) or to the nonspecific model (eq 3) (N76 and 177) are displayed. The same models can be used for the aliphatic or amide protons, but the variation expressed in ppm is smaller (maximum of 0.11 ppm for the aliphatic protons and 0.17 ppm for the amide protons).

of a heteronucleus and its bonded proton are both significant, the same model is the best for both partners. Finally, in all cases, an excellent fit is obtained (reduced χ^2 lower than 1), which proves that considering more sophisticated models such as binding of more than one xenon atom or superimposition of specific and nonspecific interactions is not relevant.

The observed chemical shift variations correspond to 22 different amino acids (see Supporting Information for a structural representation). Among them, 18 residues associated to 36 chemical shift variations correspond to protons or carbons located on the inner surface of tobacco LTP1•1, facing the cavity (Table 2). The four other amino acids (A57, P78, K80, and S82) are neighbors of the preceding ones and exhibit reduced chemical shift variation $\Delta\delta_{\text{complex}}$ (the maximal value of $\Delta\delta_{\text{complex}}$ is 0.42 ppm for ^{13}C and 0.029 ppm for ^1H). No significant variation of the ^1H or ^{13}C chemical shifts is detected for side chains on the outer surface of the protein. Finally, using this method, only side chains bordering the four alveoli have been detected (Table 2).

The binding constants K derived for all of these peaks range between 50 and 111 M⁻¹, most values being found between 60 and 80 M⁻¹. The determination of each K value is precise, as evidenced by Monte Carlo simulations (creation of new data sets based on experimental errors followed by the same fitting procedure), giving hope to cluster the K values and then to distinguish between sites of different affinities. However, the range of K values for which a good agreement, as defined by the χ^2 test between the experimental points and the specific model curve, is obtained is much larger (about 30 M⁻¹). This is particularly true for small $\Delta\delta_{\text{complex}}$ values. As a consequence, this direct discrimination between sites is, in fact, impossible. We have consequently sorted the nuclei experiencing a chemical shift variation in the presence of xenon according to their location relative to the four alveoli of tobacco LTP1•1. This reduces the hit list to 36 chemical shift variations (19 protons and 17 carbons). The frequency variations belonging to one site are then fitted altogether. Table 2 gives the name of the residue, the number of nuclei of which chemical shift values can be fitted by the specific model, the extracted binding constant, and the maximum of the chemical shift variation values $\max(\Delta\delta_{\text{complex}})$.

Whether finding different binding constants for the different alveoli is relevant or not has been explored through the F-test,

(32) Clever, H. L. *IUPAC Solubility Data Series*; Pergamon Press: Oxford, 1979.

Table 2. Fit Parameters of the Specific Model (Eq 2) for Each Hydrophobic Site of Tobacco LTP1•1

	site A	site B	site C	site D
amino acid detected	L11,L18,L51,A54,A55,I58,I61,A6, I81	K52,I81,T85	L11,L34,V35,A38,L51,P70,V75,I77,I81	L34,A38,I46
number of ^{13}C	14	3	6	1
number of ^1H	11	4	10	2
K	$76.1 \pm 0.6 \text{ M}^{-1}$	$76.5 \pm 1.0 \text{ M}^{-1}$	$72.5 \pm 0.7 \text{ M}^{-1}$	$110 \pm 3 \text{ M}^{-1}$
reduced χ^2	0.85	0.75	0.79	0.28
^{13}C max ($\Delta\delta_{\text{complex}}$)	1.69 ppm	1.69 ppm	1.73 ppm	0.25 ppm
^1H max ($\Delta\delta_{\text{complex}}$)	0.104 ppm	0.046 ppm	0.104 ppm	0.032 ppm

comparing the previous model using one binding constant K for two sites with the model of one binding constant per site. Performing these tests successively, it appears that the simplest model (the same binding constant for all sites) is the most relevant one for sites A, B, and C, a result in agreement with the binding constant values K reported in Table 2. The case of site D is troublesome: fitting with the second model represents an improvement with a probability of significance $P_F > 99.98\%$. A careful analysis reveals that this distinction is only based on one peak: the γ carbon of I46 (see Supporting Information for a graphical comparison of the best-fit theoretical curves using the second model). The small value of $\Delta\delta_{\text{complex}}$ (0.25 ppm) and the small number of involved peaks lead us to simplify the description of the binding of xenon inside the tobacco LTP1•1 by considering the same binding constant for all sites. Fitting the 36 chemical shift variations altogether leads to a binding constant K equal to 75.5 M^{-1} with an uncertainty of 1.0 M^{-1} as determined by Monte Carlo simulation (reduced $\chi^2 = 0.697$). Based on a selection criterion on the χ^2 values (reduced χ^2 lower than 1.0), the range of K values is between 63 and 93 M^{-1} .

D. Amide ^1H and ^{15}N Chemical Shifts under Variable Xenon Pressure. As for ^1H and ^{13}C nuclei, the peak frequencies in ^1H – ^{15}N HSQC spectra vary significantly according to xenon pressure (Figure 4). However, in contrast to ^1H – ^{13}C , the best model to describe the significant chemical shift variations is not always the specific one. For example, for the amide nitrogen of N76, represented in Figure 4, the nonspecific model can suitably describe the variation. This amide group is located at the surface of the protein (surface accessibility factor of 42%). This chemical shift variation could thus be interpreted as a change in the protein hydration induced by the amount of dissolved xenon, maybe associated with a nearby surface hydrophobic patch. This example, however, illustrates the usefulness of monitoring the complete chemical shift variations under xenon pressure instead of a few points, to validate a specific interaction.

When the specific model of binding is adequate to represent the variation of nitrogen chemical shifts under xenon pressure, the extracted binding constants range between 50 and 100 M^{-1} and are compatible with the values determined in the previous section. However, conversely to the ^1H – ^{13}C study, the detected amide groups are now located all along the protein backbone and not only in the vicinity of the hydrophobic cavity (see Supporting Information). It is the case, for instance, for A67 represented in Figure 4, which experiences large chemical shift variations. A67 does not really border an alveolus, because its side chain points outward from the protein, but is close in the sequence from residues involved in specific interaction with xenon. The detected effects can consequently be remote effects. They certainly result from subtle changes of the protein structure favoring the binding of xenon or in a modification of structural

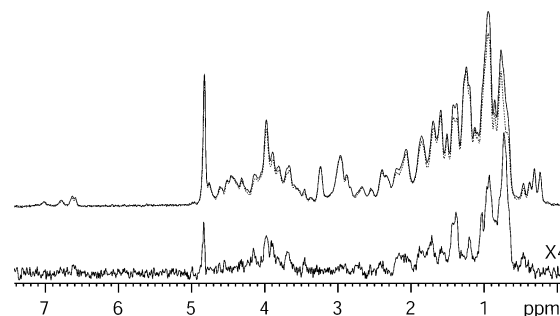


Figure 5. Example of SPINOE spectrum acquired on tobacco LTP1•1 in D_2O . The upper display corresponds to the proton spectra acquired in the presence of positively polarized xenon (solid line) and negatively polarized xenon (dotted line). Variations in the peak intensity can be observed in particular in the aliphatic region. The difference between these two spectra gives the bottom spectrum. Experimental conditions: 500 MHz spectrometer, NMR tube of reduced length, 3 bar of 96% enriched xenon polarized at 30%, tobacco LTP1•1 concentration of 0.85 mM, mixing time for proton magnetization build-up of 500 ms, 64 accumulations of each proton spectrum (positively and negatively polarized xenon), total acquisition time 2 min. In these experimental conditions, the xenon T_1 is equal to 150 s.

equilibria induced by the presence of xenon inside the cavity. Similar observations were also reported for the I14A mutant of histidine-containing phosphocarrier protein.¹¹

E. SPINOE Experiments. To locate xenon inside the hydrophobic cavity of tobacco LTP1•1 while avoiding biases resulting from the analysis of chemical shift variations, an estimation of the distances between the protein nuclei and the noble gas is required. The low solubility of xenon, its large van der Waals radius, and its small gyromagnetic ratio prevent the observation of dipolar cross-relaxation between protons and thermally polarized xenon. The enhancement of xenon magnetization by a factor of $\sim 25\,000$ in our experiments, afforded by optical pumping, allows us to circumvent the low polarization transfer rate. The SPINOE is a difference spectroscopy between proton signals in the presence of positively and negatively polarized xenon magnetization. Inversion of the xenon magnetization is performed via rf 180° pulses of the CHIRP type in the presence of field gradients.³³ Thanks to heteronuclear Xe–H dipolar cross-relaxation, the difference of even and odd ^1H subspectra enables selective detection of protons close to xenon.

Figure 5 illustrates such an experiment acquired on tobacco LTP1•1 in D_2O . Taking benefit from our last developments in the preparation of polarized xenon and in the implementation of SPINOE and SPIROE experiments,²⁹ significant signal enhancements can now be reached for a protein. For instance, for tobacco LTP1•1, the ratio of the ^1H peak intensity at 0.75 ppm with positively polarized xenon to that with negatively polarized xenon is 1.24. The signal enhancement per accumulation as defined in ref 29 is consequently equal to about 11%.

(33) Berthault, P.; Desvaux, H.; Le Goff, G.; P  tro, M. *Chem. Phys. Lett.* **1999**, 314, 52–56.

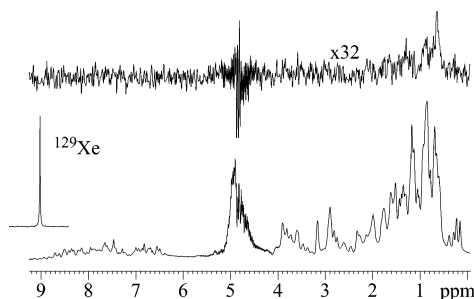


Figure 6. Example of a SPINOE spectrum acquired on tobacco LTP1•1 in H₂O. Bottom: ¹H spectrum. Top: SPINOE subspectrum. Experimental conditions: 600 MHz spectrometer, 1.8 bar of 96% enriched ¹²⁹Xe polarized at 30%, tobacco LTP1•1 concentration of 0.75 mM, mixing time for proton magnetization build-up of 500 ms, 64 accumulations of each proton spectrum (xenon spins up and down), total acquisition time 2 min.

Using this convention for defining the signal enhancement, this represents a factor of improvement bigger than 7 relative to our previous SPINOE on wheat LTP1.²⁰ However, we have checked that the xenon–proton cross-relaxation rates are on the same order of magnitude for both proteins.

To detect labile protons in SPINOE experiments, we have modified the pulse sequence by adding water saturation through low power rf irradiation during the mixing time and WATERGATE³⁴ before acquisition. The motivation of such a development lies in the previous question of whether the amide chemical shift variations under xenon pressure are induced by structural rearrangement. The spectrum of Figure 6 is obtained using this sequence. SPINOE signals are detected only in the aliphatic region. The signal-to-noise ratio lower than in Figure 5 arises from the shorter xenon *T*₁ due to relaxation with the solvent protons but also from the receiver gain which cannot be simultaneously adapted to the protein signals and the water signal.

To circumvent this last limit and to focus on the amide protons, we have taken profit of the digital filtering, to acquire the SPINOE signal in the range 5.5–10 ppm, leading to a good signal-to-noise ratio for the amide protons in each subspectrum. However, no SPINOE is detected (data not shown). This is a clear indication that the chemical shift variations experienced by the amide protons and nitrogens result from structural modifications or change in chemical exchange equilibria during xenon binding and not directly from the proximity of the noble gas atom. This fully agrees with the analysis of the tobacco LTP1•1 structure, which reveals that amide protons close to the cavity are included in the hydrogen network defining the four α helices. The xenon–amide proton distances are consequently big.

F. Location of Xenon in Tobacco LTP1•1 by SPINOE-Type Experiments. Based on the SPINOE and SPIROE spectra acquired for different mixing times, 26 peaks are unambiguously assigned. They correspond to 15 amino acid side chains, all facing the interior of the tobacco LTP1•1 cavity. Thirty-two additional peaks are also detected, but they are either less intense or nonambiguously assigned. Considering them, 9 other amino acids pointing toward the interior of the cavity are detected by SPINOE, leading to a total of 24 involved residues (see Supporting Information for a complete list).

Among these 24 amino acids, 17 were already detected by the ¹³C–¹H chemical shift variations under xenon pressure.

However, the comparison of the protons detected by SPINOE and the protons or carbons detected by the analysis of chemical shift variation under xenon pressure reveals that some of them, such as C _{γ} of I77 (Figure 4), can exhibit strong SPINOE signal but no chemical shift variation. This is not a surprise because detection of binding through chemical shifts assumes net variation between resonances of the free and bound forms ($\Delta\delta_{\text{complex}} \neq 0$). This shows again that the structural characterization of binding only from chemical shifts is far from being straightforward.

The protons detected by SPINOE belong to residues facing the four different alveoli of tobacco LTP1•1. To refine the location of xenon inside tobacco LTP1•1, all peaks in SPINOE are consequently used as restraints, with different weighting for ambiguous and unambiguous peaks. Because SPINOE to protons of amino acid side chains of the four alveoli are detected, four noble gas atoms are used in the molecular simulation. Applying the xenon–proton restraints and allowing mobility of the amino acid side chains experiencing SPINOE leads to a structural model (Figure 7) without a distance violation for the 26 unambiguous SPINOE peaks and with only three violations (0.2 Å) for the ambiguous peaks. They correspond to H _{α} of L11 and H _{α} of V75, for which proton–proton spin diffusion is more likely to explain their detection by SPINOE and H _{γ} of L51, but this amino acid is close to sites A and C, preventing specific assignment. Based on SPINOE, xenon atoms occupy the four different alveoli.

G. Dynamical Aspects of the Binding of Xenon inside Tobacco LTP1•1. Results of the analysis of the protein chemical shift variation or xenon self-relaxation times give only one xenon atom bound to the protein (even with a noble gas pressure leading to 72% of the protein molecules filled by a xenon atom, Figure 4). SPINOE results find xenon atoms in the four sites. This could appear incompatible. However, the similar affinity of the sites for xenon can be seen as a fast mobility of xenon between these sites: the noble gas atom jumps from site to site. This has been previously observed through numerical simulations of xenon dynamics inside a mutant of lysozyme T4.⁹ Such a mobility between the different sites was already noticed for wheat LTP1¹⁷ and is comforted by the finding of similar xenon *T*₁ and xenon–proton cross-relaxation rate values for both xenon–wheat LTP1 and xenon–tobacco LTP1•1 complexes. For the first complex, a xenon–proton dipolar correlation time value on the order of 0.4 ns has been found. Finally, the higher apparent affinity of xenon for D can be interpreted in this model of xenon intersite jumps by the fact that the noble gas spends more time in site D. Indeed, site D is the most embedded one, the only allowed jump being between D and C, while A, B, and C are all interconnected. Moreover, there is only one entrance to the hydrophobic cavity in tobacco LTP1•1. This entrance is close to site B and thus distant from site D.

Due to the diffusion limit, the process of xenon binding inside the protein should be slower than the previously discussed jump from one site to the next one inside tobacco LTP1•1. The study of the protein line widths in HSQC spectra (Figure 3) or of xenon line widths does not reveal a large broadening arising from the exchange between the two states. This proves that the exchange rate is typically faster than $\max(\Delta\delta_{\text{complex}}) = 1$ kHz.

The ¹⁵N relaxation study allows one to go further. Indeed, although the xenon–amide protons distances are too big to

(34) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661–665.

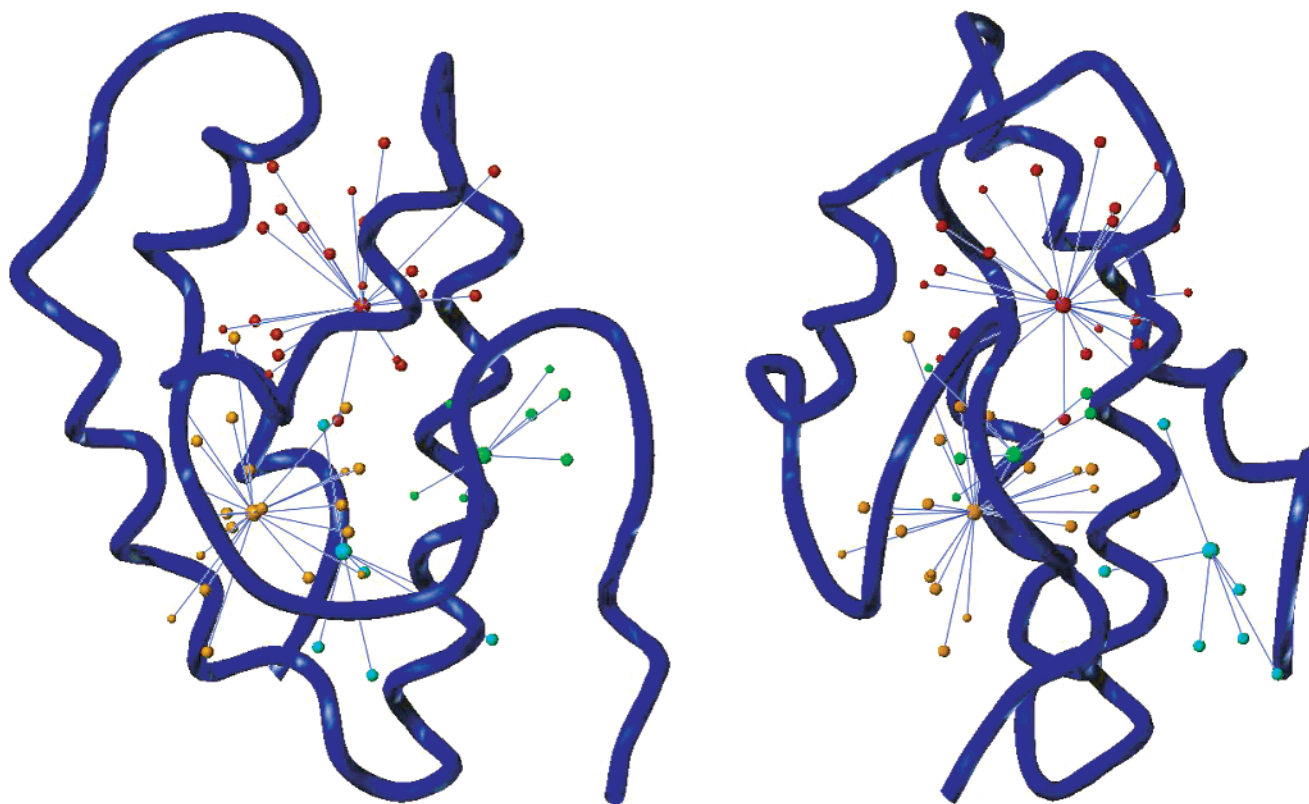


Figure 7. Location of xenon inside the solution structure of tobacco LTP1•1 as determined by SPINOE spectra. The four sites (A in red, B in green, C in orange, and D in cyan) are filled. The lines represent the distance restraints between xenon atoms and protein protons.

enable polarization transfer, the variation of nitrogen chemical shifts between the free and bound forms is a clear indication of the protein structure adaptation during the binding. The rate of this phenomenon can a priori be explored through ^{15}N transverse self-relaxation of the nitrogen amide as performed for a mutant of lysozyme T4.¹² In the case of tobacco LTP1•1, the comparison of nitrogen transverse self-relaxation times, measured by the CPMG method³⁵ at 11.7 T on the same solution degassed or under 2.9 bar of xenon (fraction of xenon–protein complex equal to 0.52), does not reveal significant variation. This puts an upper limit on the exchange rate. Considering (i) the largest chemical shift variation for an amide nitrogen equal to -1.3 ppm for A67 (Figure 4) and (ii) an average error of 0.8 Hz on the transverse relaxation rates, the exchange rate between free and bound xenon inside tobacco LTP1•1 is deduced to be faster than about $1/20 \mu\text{s} = 50$ kHz. This exchange is 1 order of magnitude faster than that of a mutant of lysozyme T4,¹² but in this latter case the xenon binding is associated with large amplitude motions of protein domains.

IV. Conclusions

This paper describes the combination of different xenon-based NMR methods for characterizing hydrophobic sites of a protein, tobacco LTP1•1. Xenon has been chosen as a “spy” of hydrophobic sites for several reasons favoring its binding. (i) It is rather soluble in water.³² (ii) Due to the high polarizability of its electron cloud, this noble gas binds specifically hydrophobic regions.⁴ (iii) The spherical shape of the single atom leads to the absence of rotational entropy.²⁰ (iv) Its adequation

to many sites allows large degeneracy in locations and then an entropy-driven binding.¹⁷ Moreover, the NMR-based approach benefits from (i) the capability of NMR to study dynamic processes,⁸ (ii) the wide chemical shift range of ^{129}Xe NMR,³¹ and (iii) the capability to enhance the xenon magnetization by several orders of magnitude through optical pumping.²⁵

The study of ^{129}Xe chemical shifts and relaxation times as a function of the noble gas or protein concentrations gives a fast insight into the affinity of xenon to the protein. Nevertheless, the superposition of specific and nonspecific interactions renders, as observed here, the exploitation of ^{129}Xe chemical shifts not straightforward, even if the availability of laser-polarized gas allows measurements with low amounts of dissolved gas which may privilege specific interactions.

The analysis of the protein atom chemical shifts under xenon pressure gives a fine insight into the thermodynamical parameters of the binding. This study benefits from the dependence of the binding only on the xenon pressure, and from the capability to acquire a large number of measurements allowing the distinction between specific and nonspecific interactions. Moreover, as observed for site D of tobacco LTP1•1, various affinities to different binding sites can be detected. Nevertheless, chemical shift variations of protein atoms under xenon pressure are an indirect approach, and their exploitation to locate the noble gas is very difficult. Indeed, the variations can result from specific contacts between xenon and protein atoms, as observed for many aliphatic groups of amino acid side chains, but also from small structural rearrangements, particularly sensitive for the protein backbone atoms such as the amide protons and nitrogens. The SPINOE approach allows one to circumvent the incapability of the precedent methods to locate xenon atoms.

(35) Farrow, N. A.; Zhang, O. W.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* 1995, 34, 868–878.

Indeed, SPINOE allows the extraction of xenon–proton proximities, even now in H₂O. With this approach, the interaction sites can be mapped at the atomic level. The understanding of the xenon binding inside the protein can finally be refined by xenon and protein relaxation studies. To end, xenon-based NMR experiments provide a composite picture, including structure, thermodynamics of binding, kinetics of binding, and dynamics of xenon inside the cavities.

All of these experiments converge to a tobacco LTP1•1/xenon binding constant value of the order of 75 M⁻¹, while the use of other spies for the related protein, wheat LTP1, did not reveal such strong interactions.³⁶ Xenon seems to jump at the nanosecond time between the four hydrophobic sites A, B, C, and D, with D specific to tobacco LTP1•1 as compared to wheat LTP1. The exchange rate between free- and bound-xenon tobacco LTP1•1 is faster than 50 kHz, indicating the absence of large amplitude motion of the protein during xenon binding. Nevertheless, LTP1 proteins have a very particular hydrophobic cavity. Thus, it is only through the analysis of protein

hydrophobic cavities of different geometries that a better characterization of the affinity of xenon to these volumes can be obtained. Such a task is already in development in our laboratory. This would open the way to a thermodynamics- and kinetics-based description of the biological role of protein hydrophobic cavities, which would complement the X-ray crystallography structural approach.^{3,4}

Acknowledgment. Benoît Industri and Antoine Marais are acknowledged for their involvements in the production of the tobacco LTP1•1. The French Ministry of Research is acknowledged for financial support (ACI no. 4103 to H.D.).

Supporting Information Available: Two tables containing the list of atoms experiencing chemical shift variations under xenon pressure; two tables giving the list of atoms detected in SPINOE; figure showing the location on the protein structure of ¹H and ¹³C atoms experiencing chemical shift variation under xenon pressure; figure displaying the fit of C_γ of I46 with K = 75.5 M⁻¹ and K = 110 M⁻¹. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA046195I

(36) Liepinsh, E.; Sodano, P.; Tassin, S.; Marion, D.; Vovelle, F.; Otting, G. *J. Biomol. NMR* **1999**, *15*, 213–225.

4. Conclusions et perspectives

La cavité hydrophobe de la LTP1_1 a été explorée en détail par RMN à l'aide de la sonde hydrophobe que constitue le xénon. L'analyse du déplacement chimique et du temps de relaxation longitudinale T_1 du ^{129}Xe a permis de mettre en évidence une interaction entre la LTP1_1 et le xénon. De plus, cette analyse a révélé un échange chimique rapide à l'échelle de temps de la RMN entre la forme libre du xénon et sa forme liée avec la LTP1_1. Les paramètres thermodynamiques de fixation du xénon ont été obtenus en analysant les variations de déplacement chimique ^1H et ^{13}C de la LTP1_1 en fonction de la concentration de xénon. Leur interprétation, ainsi que des simulations de solvation de la cavité de la LTP1_1 par du xénon, ont montré l'existence de quatre sites de fixation du xénon (A, B, C, D) ayant des constantes d'affinité différentes (cf. figure 1 et table II de l'article) variant de 72,5 à 76,5 M^{-1} pour les sites A, B et C et égale à 110 M^{-1} pour le site D. Les expériences SPINOE ont permis la localisation du xénon et l'identification des acides aminés formant les quatre sites privilégiés de son interaction. Les mesures des temps de relaxation longitudinale T_1 du ^{129}Xe indiquent que la LTP1_1 pourrait fixer un atome de xénon et que ce dernier passerait d'un site à l'autre très rapidement avec des temps de résidence inférieurs à une nanoseconde. D'autre part, nous avons montré que les déplacements chimiques des protons et azotes amides ($^1\text{H}_\text{N}$ et ^{15}N) variaient en fonction de la concentration de xénon. Mais, contrairement aux variations des déplacements chimiques ^1H et ^{13}C des chaînes aliphatiques, ces variations ne correspondent pas uniquement à des interactions spécifiques. En effet, elles ne sont pas corrélées aux NOEs hétéronucléaires ^{129}Xe - ^1H que nous avons obtenus par l'expérience SPINOE réalisée dans H_2O . Elles traduisent une « respiration » de la LTP1_1, c'est-à-dire que la protéine subit de petites modifications structurales après fixation du xénon. Enfin, l'analyse combinée des déplacements chimiques $^1\text{H}_\text{N}$ et ^{15}N et des temps de relaxation transversale T_2 du noyau ^{15}N permet d'appréhender la vitesse d'échange entre la forme libre de la LTP1_1 et sa forme liée au xénon. Celle-ci est supérieure à 50 kHz, indiquant une absence de mouvements de large amplitude pendant la fixation du xénon.

L'ensemble de ces résultats permet de définir une constante d'association globale LTP1_1/xénon de l'ordre de 75,5 M^{-1} pour l'ensemble des quatre sites A, B, C et D. C'est la première fois que l'on observe une constante d'association aussi élevée pour une sonde hydrophobe interagissant avec une protéine de la famille des LTP1. Jusqu'à présent, aucune

des sondes hydrophobes utilisées pour explorer la cavité de la LTP1 de blé (Liepinsh et al. 1999) n'a permis de mettre en évidence une interaction aussi importante. D'autre part, la LTP1_1 de tabac possède un site de fixation du xénon spécifique supplémentaire par rapport à ceux identifiés dans la LTP1 de blé. C'est le site D qui est défini par les 3 résidus L34, A38 et I46 et présente la constante d'affinité la plus élevée 110 M^{-1} . Parmi ces trois résidus, seul le résidu I46 est spécifique à la séquence de la LTP1_1 de tabac. En effet, dans les LTP1 de graines de céréales, cet acide aminé est muté en résidu hydrophile (Thr ou Ser) ou en alanine. Ce résidu apparaît donc très important pour la définition du site D de fixation du ligand et pourrait expliquer la différence d'affinité de ce site comparé aux sites A, B et C. Il pourrait même nous permettre d'expliquer l'origine des différences d'affinité de la LTP1_1 pour un phospholipide (le LMPC) et pour des acides gras (acides oléique et palmitique) qui ont été observées par fluorescence (cf. chapitre 3). La constante d'affinité de la LTP1_1 est en effet augmentée d'un facteur variant de 5 à 10 entre le LMPC et les acides gras.

En conclusion, la RMN du xénon polarisé par laser se révèle être un excellent outil pour l'étude des cavités hydrophobes de protéines. En effet, grâce à ses propriétés particulières, hydrophobie, sensibilité à son environnement, la sonde xénon a permis une cartographie complète de la cavité hydrophobe de la LTP1_1 de tabac et l'identification des résidus impliqués dans sa fixation. La détermination de ces résidus a peut-être permis de découvrir le résidu responsable de la spécificité de la LTP1_1 (Ile 46). Afin de confirmer cette hypothèse, il serait nécessaire de muter spécifiquement cette isoleucine par une alanine ou les résidus hydrophiles (Thr ou Ser) observés dans les LTP1 de graines de céréales et de réaliser les mêmes expériences de fixation de molécules lipidiques (LMPC et acides gras).

Conclusions et perspectives

L'objectif principal de mon travail de thèse a été de réaliser l'étude structurale de protéines appartenant à la famille des LTPs. Je me suis plus particulièrement intéressé à la famille des LTP1 identifiées chez le tabac *Nicotiana tabacum*, afin d'établir des relations structure/fonction et d'élucider à terme la ou les fonctions de ces protéines.

Avant la réalisation de mes travaux de thèse, seules les structures tridimensionnelles de LTPs extraites de graines de céréales étaient connues, c'est pourquoi nous avons entrepris d'étudier les LTP1 présentes dans la plante entière. La première LTP1 exprimée dans les organes végétatifs de tabac a été surexprimée dans *Pichia pastoris*, permettant ainsi son marquage isotopique ^{15}N et ^{13}C . La production de la première LTP1 recombinante s'est révélée très efficace, avec des rendements de production de l'ordre d'une dizaine de mg par litre de milieu de culture. Nous avons ensuite pu déterminer sa structure tridimensionnelle par RMN 2D, 3D et modélisation moléculaire et réaliser des études d'interaction de cette nouvelle LTP1 avec un phospholipide LMPC (LysoMyristoylPhosphocholine) et des acides gras (acides palmitique et oléique).

La structure tridimensionnelle de cette première LTP1 présente dans les organes végétatifs de tabac ne diffère pas globalement de celle des LTP1 isolées dans les graines de céréales. Elle possède quatre hélices α surenroulées suivies d'un long segment C-terminal qui ne présente pas de structure secondaire. De plus, comme les autres LTP1, elle présente une cavité hydrophobe qui lui permet de lier des molécules amphiphiles. En revanche, ses propriétés de complexation sont différentes. En effet, contrairement aux LTPs de graine de céréales qui complexent indifféremment des phospholipides et des acides gras avec la même affinité, la LTP1_1 de tabac ne peut fixer qu'une seule molécule amphiphile, un LMPC ou un acide gras, et avec une affinité plus faible pour ces derniers.

Ces différences de complexation ne résultent pas de l'absence de cavité hydrophobe mais de petites différences structurales dans l'organisation des résidus hydrophobes, qui entraînent une accessibilité différente de la cavité. Les LTP1 de graine de céréales possèdent deux entrées pour leur cavité hydrophobe alors que la LTP1_1 ne présente qu'une seule entrée située entre la boucle L2 et le segment C-terminal, la deuxième étant obstruée par des résidus hydrophobes propres à cette protéine, en particulier la leucine 24 et la chaîne latérale de la lysine 65. Des expériences de mutagénèse dirigée ainsi que l'étude de LTP1 présentant les mêmes caractéristiques que la LTP1_1 de tabac devraient permettre de démontrer

l'importance de ces résidus.

D'autre part, la mesure des vitesses de relaxation hétéronucléaire R1, R2, et des NOEs (^{15}N ^1H) nous a permis de décrire les mouvements caractéristiques des vecteurs NH le long du squelette peptidique. Cette étude par RMN de la dynamique du squelette polypeptidique de la LTP1_1 de tabac a été complétée par des simulations de dynamique moléculaire à température ambiante et à haute température. Nous avons ainsi montré que les propriétés remarquables de complexation de la LTP1_1 sont liées non seulement à la présence de sa cavité hydrophobe mais également à la dynamique particulière de certains acides aminés situés au niveau de l'entrée de la cavité qui permettraient la fixation du lipide. Il s'agit des acides aminés de la boucle L2 (Arg39 et Thr 40) et ceux du segment C-terminal Val75-Ser84 impliqués d'une part dans la définition de l'entrée de la cavité de la LTP1_1 de tabac et d'autre part dans son ouverture. De plus, contrairement à la LTP1 de blé, les résultats de simulation à haute température pour la LTP1_1 de tabac montrent l'existence d'une entrée unique pour la cavité. Afin de compléter les résultats que nous avons obtenus par RMN et *in silico* sur la dynamique de la LTP1_1 de tabac, des études semblables à celles que nous avons effectuées pourraient être menées sur des complexes lipide-protéine. Ces derniers impliqueraient des molécules lipidiques comme par exemple des acides gras ou, dans l'idéal, un ligand naturel que nous aurions identifié, afin de déterminer les résidus responsables de la stabilisation du complexe et des différentes affinités observées pour cette LTP1.

Par ailleurs, nous avons étudié la cavité hydrophobe de la LTP1_1 en utilisant la RMN du xénon polarisé par laser qui se révèle être un excellent outil pour l'étude des cavités hydrophobes de protéines. En effet, grâce à ses propriétés particulières : hydrophobie, sensibilité à son environnement, la sonde xénon a permis une cartographie complète de la cavité hydrophobe de la LTP1_1 de tabac et l'identification des résidus impliqués dans sa fixation. La détermination de ces résidus a peut-être permis de découvrir le résidu responsable de la spécificité de la LTP1_1 (Ile 46). Afin de confirmer cette hypothèse, il serait nécessaire de muter spécifiquement cette isoleucine par une alanine ou par les résidus hydrophiles (Thr ou Ser) correspondants dans les LTP1 de graine de céréales et de réaliser les mêmes expériences de fixation de molécules lipidiques (LMPC et acides gras).

Nous avons ainsi montré avec la LTP1_1 de tabac que certaines LTPs peuvent présenter des caractéristiques de complexation et d'affinité différentes les unes des autres sans modification majeure de leur structure tridimensionnelle. Ces différences de propriétés ont

elles une signification biologique ? La multigénicité des LTPs au sein d'une même espèce végétale suggère probablement qu'il n'y a pas redondance mais multiplicité des fonctions. Ainsi, le fait que la LTP1_1 de tabac ne fixe qu'une seule molécule lipidique et que selon le type de ligand l'affinité n'est pas la même, indique que si nous pouvions demain comparer la structure et l'activité d'un ensemble de LTPs d'un même organisme, nous aurions probablement des surprises quand à la diversité des ligands qu'elles peuvent fixer et par conséquent de leurs fonctions. Nous avons déjà commencé cette étude par la modélisation comparative. Jusqu'à présent, seulement 5 LTP1 ont été identifiées chez le tabac. Nos résultats de modélisation suggèrent que deux d'entre elles (la 3 et la 4) formeraient une nouvelle sous-famille structurale de LTP1 avec une cavité de taille réduite et une absence de pôle hydrophile. Ces résultats suggèrent un mode de fixation différent, voire même une impossibilité de complexation. Une étude structurale de cette sous-famille couplée à des études des propriétés, pourrait nous aider à définir leur rôle particulier, et à progresser encore dans les relations structure/fonction de la famille des LTPs.

Nos résultats peuvent aussi signifier que les interactions entre les LTP1 et les macrostructures amphiphiles (membranes, domaines lipophiles) sont de nature différente. Par exemple, il pourrait exister des spécificités d'interaction qui adresserait des protéines vers la construction de subérine/cutine ou vers les interactions protéine-protéine (récepteur membranaire) en fonction des modifications fines de la structure globale conservée au sein de ces protéines.

Bibliographie

2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- Abraham, A. 1961. *Principles of nuclear magnetism*. Clarendon, Oxford.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
- Amano, M., Ogawa, H., Kojima, K., Kamidaira, T., Suetsugu, S., Yoshihama, M., Satoh, T., Samejima, T., and Matsumoto, I. 1998. Identification of the major allergens in wheat flour responsible for baker's asthma. *Biochem J* 330 (Pt 3): 1229-1234.
- Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 32 Database issue: D115-119.
- Armentia, A., Sanchez-Monge, R., Gomez, L., Barber, D., and Salcedo, G. 1993. In vivo allergenic activities of eleven purified members of a major allergen family from wheat and barley flour. *Clin Exp Allergy* 23: 410-415.
- Arondel, V., Vergnolle, C., Cantrel, C., and Kader, J.C. 2000. Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. *Plant Science*: 1-12.
- Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S., and Lehmann, M.S. 1993. Crystal structure of hydrophobic protein from soybean; a member of a new cysteine-rich family. *J Mol Biol* 231: 877-887.
- Berendsen, H., Postma, J., Van Gunstesen, W., Di Nola, A., and Haak, J. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys* 81: 3684-3690.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* 28: 235-242.
- Bernhard, W.R., Thoma, S., Botella, J., and Somerville, C.R. 1991. Isolation of a cDNA clone for spinach lipid transfer protein and evidence that the protein is synthesized by the secretory pathway. *Plant Physiol.* 95: 164-170.
- Blanckaert, A., Belingheri, L., Sautiere, P.E., Vasseur, J., and Hilbert, J.L. 2002. 9-kDa acidic and basic nsLTP-like proteins are secreted in the culture-medium conditioned by somatic embryogenesis in *Cichorium*. *Plant physiology and biochemistry* 40: 339-345.
- Blee, E. 1998. Phytooxylipins and plant defense reactions. *Prog Lipid Res* 37: 33-72.

- Blein, J.P., Coutos-Thevenot, P., Marion, D., and Ponchet, M. 2002. From elicitors to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. *Trends Plant Sci* 7: 293-296.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31: 365-370.
- Boissy, G., de La Fortelle, E., Kahn, R., Huet, J.C., Bricogne, G., Pernellet, J.C., and Brunie, S. 1996. Crystal structure of a fungal elicitor secreted by *Phytophthora cryptogea*, a member of a novel class of plant necrotic proteins. *Structure* 4: 1429-1439.
- Boissy, G., O'Donohue, M., Gaudemer, O., Perez, V., Pernellet, J.C., and Brunie, S. 1999. The 2.1 Å structure of an elicitor-ergosterol complex: a recent addition to the Sterol Carrier Protein family. *Protein Sci* 8: 1191-1199.
- Bonnet, P., Poupet, A., Abad, P., Venard, P., and Cardin, L. 1986. Induction de nécroses foliaires, de protéines b et de résistance dans les interactions tabac-*Phytophthora*. *Agronomie* 6: 829-837.
- Bowie, J.U., Luthy, R., and Eisenberg, D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253: 164-170.
- Braun, W., and Go, N. 1985. Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. *J Mol Biol* 186: 611-626.
- Broekaert, W.F., Cammue, B.P.A., De Bolle, M.F.C., Thevissen, K., De Samblanx, G.W., and Osborn, R.W. 1997. Antimicrobial peptides from plants. *Critical Reviews in Plant Sciences* 16: 297-323.
- Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. 1983. CHARMM: A program for macromolecular energy minimization, and dynamics calculations. *J. Comp. Chem.* 4: 187-217.
- Brünger, A.T., Adams, P.D., Clore, G.M., De Lano, W.L., Gros, P., Grosse-Kuntze, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system : A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54: 905-921.
- 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. 2001. A lipid transfer protein binds to a receptor involved in the control of plant defence responses. *FEBS Lett* 509: 27-30.

- Buhot, N., Gomes, E., Milat, M.L., Ponchet, M., Lequeu, J., Delrot, S., Coutos-Thevenot, P., and Blein, J.P. 2004. A jamonic acid-lipid transfer protein complex binds to a plasma membrane receptor and specifically induces protection against *Phytophthora parasitica* in tobacco. *Mol Biol Cell* 15: 5047-5052.
- Cammue, B.P., Thevissen, K., Hendriks, M., Eggermont, K., Goderis, I.J., Proost, P., Van Damme, J., Osborn, R.W., Guerbet, F., Kader, J.C., et al. 1995. A potent antimicrobial protein from onion seeds showing sequence homology to plant lipid transfer proteins. *Plant Physiol* 109: 445-455.
- Canevascini, S., Caderas, D., Mandel, T., Fleming, A.J., Dupuis, I., and Kuhlemeier, C. 1996. Tissue-specific expression and promoter analysis of the tobacco Ltp1 gene. *Plant Physiol.* 112: 513-524.
- Cavanagh, J., Fairbrother, W.J., Palmer III, A.G., and Skelton, N.J. 1996. *Protein NMR spectroscopy - Principles and Practice*, Academic Press Inc ed.
- Charvolin, D., Douliez, J.P., Marion, D., Cohen-Addad, C., and Pebay-Peyroula, E. 1999. The crystal structure of a wheat nonspecific lipid transfer protein (ns-LTP1) complexed with two molecules of phospholipid at 2.1 Å resolution. *Eur J Biochem* 264: 562-568.
- Cheng, C.S., Samuel, D., Liu, Y.J., Shyu, J.C., Lai, S.M., Lin, K.F., and Lyu, P.C. 2004a. Binding mechanism of nonspecific lipid transfer proteins and their role in plant defence. *Biochemistry* in press.
- Cheng, H.C., Cheng, P.T., Peng, P., Lyu, P.C., and Sun, Y.J. 2004b. Lipid binding in rice nonspecific lipid transfer protein-1 complexes from *Oryza sativa*. *Protein Sci* 13: 2304-2315.
- Cherubini, A., and Bifone, A. 2003. Hyperpolarised xenon in biology. *Progress in NMR spectroscopy* 42: 1-30.
- Chothia, C. 1992. Proteins. One thousand families for the molecular biologist. *Nature* 357: 543-544.
- Clore, G.M., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. 1990a. Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear ¹⁵N-¹H NMR spectroscopy. *Biochemistry* 29: 7387-7401.
- Clore, G.M., Szabo, A., Bax, A., Lewis, E.K., Driscoll, P.C., and Gronenborn, A.M. 1990b. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc* 112: 4989-4991.
- Cole, R., and Loria, J.P. 2003. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J Biomol NMR* 26: 203-213.

- Coutos-Thevenot, P., Jouenne, T., Maes, O., Guerbette, F., Grosbois, M., Le Caer, J.P., Boulay, M., Deloire, A., Kader, J.C., and Guern, J. 1993. Four 9-kDa proteins excreted by somatic embryos of grapevine are isoforms of lipid-transfer proteins. *Eur J Biochem* 217: 885-889.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277-293.
- Desormeaux, A., Blochet, J.E., Pezolet, M., and Marion, D. 1992. Amino acid sequence of a non-specific wheat phospholipid transfer protein and its conformation as revealed by infrared and Raman spectroscopy. Role of disulfide bridges and phospholipids in the stabilization of the alpha-helix structure. *Biochim Biophys Acta* 1121: 137-152.
- Douliez, J.P., Jegou, S., Pato, C., Molle, D., Tran, V., and Marion, D. 2001a. Binding of two mono-acylated lipid monomers by the barley lipid transfer protein, LTP1, as viewed by fluorescence, isothermal titration calorimetry and molecular modelling. *Eur J Biochem* 268: 384-388.
- Douliez, J.P., Michon, T., and Marion, D. 2000. Steady-state tyrosine fluorescence to study the lipid-binding properties of a wheat non-specific lipid-transfer protein (nsLTP1). *Biochim Biophys Acta* 1467: 65-72.
- Douliez, J.P., Pato, C., Rabesona, H., Molle, D., and Marion, D. 2001b. Disulfide bond assignment, lipid transfer activity and secondary structure of a 7-kDa plant lipid transfer protein, LTP2. *Eur J Biochem* 268: 1400-1403.
- Dubois, L., Berthault, P., Huber, J.G., and Desvaux, H. 2004. mapping hydrophobic molecular regions using dissolved laser-polarised xenon NMR. *C. R. Acad. Sci. Physique*.
- Dubreil, L., Gaborit, T., Bouchet, B., Gallant, D.J., Broekaert, W., Quillien, L., Marion, D. 1998. Saptial and temporal distribution of the major isoforms of puroindolines (puroindoline-a and puroindoline-b) and non specific lipid transfer protein (ns-LTPe1) of *Triticum aestivum* seeds. Relationships with their *in vitro* antifungal properties. *Plant Science* 138. 121-135.
- Elmorjani, K., Lurquin, V., Lelion, A., Rogniaux, H., and Marion, D. 2004. A bacterial expression system revisited for the recombinant production of cystein-rich plant lipid transfer proteins. *Biochem Biophys Res Commun* 316: 1202-1209.

- Evenas, J., Forsen, S., Malmendal, A., and Akke, M. 1999. Backbone dynamics and energetics of a calmodulin domain mutant exchanging between closed and open conformations. *J Mol Biol* 289: 603-617.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33: 5984-6003.
- Fauth, M., Schweizer, P., Buchala, A., Markstadter, C., Riederer, M., Kato, T., and Kauss, H. 1998. Cutin monomers and surface wax constituents elicit H₂O₂ in conditioned cucumber hypocotyl segments and enhance the activity of other H₂O₂ elicitors. *Plant Physiol* 117: 1373-1380.
- Fefeue, S., Bouaziz, S., Huet, J.C., Pernollet, J.C., and Guittet, E. 1997. Three-dimensional solution structure of beta cryptogein, a beta elicitin secreted by a phytopathogenic fungus *Phytophthora cryptogea*. *Protein Sci* 6: 2279-2284.
- Fischer, M.W.F., Majumdar, A., and Zuiderweg, E.R.P. 1998. Protein NMR relaxation: theory, applications and outlook. *Prog Nucl Magn Reson Spectrosc* 33: 207-272.
- Fleming, A.J., Mandel, T., Hofmann, S., Sterk, P., de Vries, S.C., and Kuhlemeier, C. 1992. Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. *Plant J* 2: 855-862.
- Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J.P. 1987. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett* 224: 149-155.
- Garcia-Olmedo, F., Molina, A., Segura, A., and Moreno, M. 1995. The defensive role of nonspecific lipid-transfer proteins in plants. *Trends Microbiol* 3: 72-74.
- Ge, X., Chen, J., Sun, C., and Cao, K. 2003. Preliminary study on the structural basis of the antifungal activity of a rice lipid transfer protein. *Protein Eng* 16: 387-390.
- Gincel, E., Simorre, J.P., Caille, A., Marion, D., Ptak, M., and Vovelle, F. 1994. 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* 226: 413-422.
- Gomar, J., Petit, M.C., Sodano, P., Sy, D., Marion, D., Kader, J.C., Vovelle, F., and Ptak, M. 1996. Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. *Protein Sci* 5: 565-577.

- Gomar, J., Sodano, P., Sy, D., Shin, D.H., Lee, J.Y., Suh, S.W., Marion, D., Vovelle, F., and Ptak, M. 1998. Comparison of solution and crystal structures of maize nonspecific lipid transfer protein: a model for a potential in vivo lipid carrier protein. *Proteins* 31: 160-171.
- Gomes E., Sagot E., Gaillard C., Laquitaine L., Poinssot B., Sanejouand Y.H., Delrot S., Coutos-Thevenot P. 2003. Nonspecific lipid-transfer protein genes expression in grape (*Vitis* sp.) cells in response to fungal elicitor treatments. *Mol Plant Microbe Interact* 16(5):456-64.
- Gonzalez, R., Varela, J., Carreira, J., and Polo, F. 1995. Soybean hydrophobic protein and soybean hull allergy. *Lancet* 346: 48-49.
- Han, G.W., Lee, J.Y., Song, H.K., Chang, C., Min, K., Moon, J., Shin, D.H., Kopka, M.L., Sawaya, M.R., Yuan, H.S., et al. 2001. Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution x-ray crystallography. *J. Mol. Biol.* 308: 263-278.
- Heinemann, B., Andersen, K.V., Nielsen, P.R., Bech, L.M., and Poulsen, F.M. 1996. Structure in solution of a four-helix lipid binding protein. *Protein Sci* 5: 13-23.
- Hendriks, T. 1994. *The carrot extracellular lipid transfer protein EP2: quantitative aspects with respect to its putative role in cutin synthesis in Plant Molecular Biology*, pp. 85-94.
- Hendrix, J.W., and Guttman, S.M. 1970. Sterol or calcium requirement by *Phytophthora parasitica* var. *nicotianae* for growth on nitrate nitrogen. *Mycologia* 62: 195-198.
- Henikoff, S., and Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89: 10915-10919.
- Henriksen, A. 2004. Binding of L-Alpha-Lysophosphatidylcholine, Lauroyl in the hydrophobic channel of Barley Lipid Transfer Protein. *To be published*.
- Hollenbach, B., Schreiber, L., Hartung, W., and Dietz, K.J. 1997. Cadmium leads to stimulated expression of the lipid transfer protein genes in barley: implications for the involvement of lipid transfer proteins in wax assembly. *Planta* 203: 9-19.
- Hutchinson, E.G., and Thornton, J.M. 1996. PROMOTIF a program to identify and analyze structural motifs in proteins. *Protein Sci* 5: 212-220.
- Johnson, B.A., and Blevins, R.A. 1994. NMRview: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4: 603-614.

- Kader, J.C. 1975. Proteins and the intracellular exchange of lipids. I. Stimulation of phospholipid exchange between mitochondria and microsomal fractions by proteins isolated from potato tuber. *Biochim Biophys Acta* 380: 31-44.
- Kader, J.C. 1996. Lipid-transfer proteins in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47: 627-654.
- Kader, J.C. 1997. Lipid-transfer proteins: a puzzling family of plant proteins. *Trends in plant science* 2: 66-70.
- Kanelis, V., Forman-Kay, J.D., and Kay, L.E. 2001. Multidimensional NMR methods for protein structure determination. *IUBMB Life* 52: 291-302.
- Karplus, M. 1959. Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* 30: 11-15.
- Kay, L.E., Torchia, D.A., and Bax, A. 1989. Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28: 8972-8979.
- Kolattukudy, P. 1981. Structure, biosynthesis and biodegradation of cutin and suberin. *Annual Review of plant physiology*: 539-567.
- Kolattukudy, P.E. 2001. Polyesters in higher plants. *Adv Biochem Eng Biotechnol* 71: 1-49.
- Kolattukudy, P.E., Rogers, L.M., Li, D., Hwang, C.S., and Flaishman, M.A. 1995. Surface signaling in pathogenesis. *Proc Natl Acad Sci U S A* 92: 4080-4087.
- Korzhnev, D.M., Billeter, M., Arseniev, A.S., and Orekhov, V.Y. 2001. NMR studies of tumbling and internal motions in proteins. *Prog Nucl Magn Reson Spectrosc* 38: 197-266.
- Kotilainen, M., Helariutta, Y., Elomaa, P., Paulin, L., and Teeri, T.H. 1994. A corolla- and carpel-abundant, non-specific lipid transfer protein gene is expressed in the epidermis and parenchyma of *Gerbera hybrida* var. Regina (Compositae). *Plant Mol Biol* 26: 971-978.
- Krause, A., Sigrist, C.J., Dehning, I., Sommer, H., and Broughton, W.J. 1994. Accumulation of transcripts encoding a lipid transfer-like protein during deformation of nodulation-competent *Vigna unguiculata* root hairs. *Mol Plant Microbe Interact* 7: 411-418.
- Kristensen, A.K., Brunstedt, J., Nielsen, K.K., Roepstorff, P., and Mikkelsen, J.D. 2000. Characterization of a new antifungal non-specific lipid transfer protein (nsLTP) from sugar beet leaves. *Plant Science* 155: 31-40.

- Landon, C., Berthault, P., Vovelle, F., and Desvaux, H. 2001. Magnetization transfer from laser-polarized xenon to protons located in the hydrophobic cavity of the wheat nonspecific lipid transfer protein. *Protein Sci* 10: 762-770.
- Lascombe, M.B., Ponchet, M., Venard, P., Milat, M.L., Blein, J.P., and Prange, T. 2002. The 1.45 Å resolution structure of the cryptogein-cholesterol complex: a close-up view of a sterol carrier protein (SCP) active site. *Acta Crystallogr D Biol Crystallogr* 58: 1442-1447.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477-486.
- Lee, J.Y., Min, K., Cha, H., Shin, D.H., Hwang, K.Y., and Suh, S.W. 1998. Rice non-specific lipid transfer protein: the 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity. *J Mol Biol* 276: 437-448.
- Lerche, M.H., Kragelund, B.B., Bech, L.M., and Poulsen, F.M. 1997. Barley lipid-transfer protein complexed with palmitoyl CoA: the structure reveals a hydrophobic binding site that can expand to fit both large and small lipid-like ligands. *Structure* 5: 291-306.
- Lerche, M.H., and Poulsen, F.M. 1998. Solution structure of barley lipid transfer protein complexed with palmitate. Two different binding modes of palmitate in the homologous maize and barley nonspecific lipid transfer proteins. *Protein Sci* 7: 2490-2498.
- Liepinsh, E., Sodano, P., Tassin, S., Marion, D., Vovelle, F., and Otting, G. 1999. Solvation study of the non-specific lipid transfer protein from wheat by intermolecular NOEs with water and small organic molecules. *J Biomol NMR* 15: 213-225.
- Lindorff-Larsen, K., Lerche, M.H., Poulsen, F.M., Roepstorff, P., and Winther, J.R. 2001. Barley lipid transfer protein, LTP1, contains a new type of lipid-like post-translational modification. *J Biol Chem* 276: 33547-33553.
- Lindorff-Larsen, K., and Winther, J.R. 2001. Surprisingly high stability of barley lipid transfer protein, LTP1, towards denaturant, heat and proteases. *FEBS Lett* 488: 145-148.
- Linge, J.P., Habeck, M., Rieping, W., and Nilges, M. 2003a. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19: 315-316.
- Linge, J.P., O'Donoghue, S.I., and Nilges, M. 2001. Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 339: 71-90.

- Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M., and Nilges, M. 2003b. Refinement of protein structures in explicit solvent. *Proteins* 50: 496-506.
- Lipari, G., and Szabo, A. 1982a. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. Theory and range of validity. *J. Am. Chem. Soc* 104: 4546.
- Lipari, G., and Szabo, A. 1982b. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. Analysis of experimental result. *J.Am.Chem.Soc.* 104: 4559.
- Lullien-Pellerin, V., Devaux, C., Ihorai, T., Marion, D., Pahin, V., Joudrier, P., and Gautier, M.F. 1999. Production in *Escherichia coli* and site-directed mutagenesis of a 9-kDa nonspecific lipid transfer protein from wheat. *Eur J Biochem* 260: 861-868.
- Maldonado, A.M., Doerner, P., Dixon, R.A., Lamb, C.J., and Cameron, R.K. 2002. A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* 419: 399-403.
- Mandel, A.M., Akke, M., and Palmer, A.G., 3rd. 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* 246: 144-163.
- Marion, D., Douliez, J.P., Gautier, M.F., and Elmorjani, K. 2003. *Plant lipid Transfer Proteins: relationship between allergenicity and structural, biological and technological properties*, Mills, C Shewry, P. R ed. Blackwell publishing, pp. 59-69.
- Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325.
- Mikes, V., Milat, M.L., Ponchet, M., Panabieres, F., Ricci, P., and Blein, J.P. 1998. Elicitins, proteinaceous elicitors of plant defense, are a new class of sterol carrier proteins. *Biochem Biophys Res Commun* 245: 133-139.
- Molina, A., and Garcia-Olmedo, F. 1997. Enhanced tolerance to bacterial pathogens caused by the transgenic expression of barley lipid transfer protein LTP2. *Plant J* 12: 669-675.
- Molina, A., Segura, A., and Garcia-Olmedo, F. 1993. Lipid transfer proteins (nsLTPs) from barley and maize leaves are potent inhibitors of bacterial and fungal plant pathogens. *FEBS Lett* 316: 119-122.

- Montillet J.C., Agnel J.P., Ponchet M., Vailleau F., Roby D. and Triantaphylidès C. 2002. Lipoxygenase-mediated production of fatty acid hydroperoxides is a specific signature of the hypersensitive reaction in plants. *Plant Physiol Biochem* 40: 633-639.
- Murtagh, G.J., Dumoulin, M., Archer, D.B., and Alcocer, M.J. 2002. Stability of recombinant 2 S albumin allergens in vitro. *Biochem Soc Trans* 30: 913-915.
- Navon, G., Song, Y.Q., Room, T., Appelt, S., Taylor, R.E., and Pines, A. 1996. Enhancement of solution NMR and MRI with laser-polarized xenon. *Science* 271: 1848-1851.
- Nielsen, K.K., Nielsen, J.E., Madrid, S.M., and Mikkelsen, J.D. 1996. New antifungal proteins from sugar beet (*Beta vulgaris* L.) showing homology to non-specific lipid transfer proteins. *Plant Mol Biol* 31: 539-552.
- Nilges, M. 1995. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J Mol Biol* 245: 645-660.
- Nilges, M., Macias, M.J., O'Donoghue, S.I., and Oschkinat, H. 1997. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 269: 408-422.
- Nilges, M., and O'Donoghue, S.I. 1998. Ambiguous NOEs and automated NOE assignment. *Prog NMR Spectrosc* 32: 107-139.
- Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T., and Morimoto, T. 1997. Tertiary and quaternary structures of 0.19 alpha-amylase inhibitor from wheat kernel determined by X-ray analysis at 2.06 Å resolution. *Biochemistry* 36: 13503-13511.
- Osman, H., Mikes, V., Milat, M.L., Ponchet, M., Marion, D., Prange, T., Maume, B.F., Vauthrin, S., and Blein, J.P. 2001a. Fatty acids bind to the fungal elicitor cryptogein and compete with sterols. *FEBS Lett* 489: 55-58.
- Osman, H., Vauthrin, S., Mikes, V., Milat, M.L., Panabieres, F., Marais, A., Brunie, S., Maume, B., Ponchet, M., and Blein, J.P. 2001b. Mediation of elicitin activity on tobacco is assumed by elicitin-sterol complexes. *Mol Biol Cell* 12: 2825-2834.
- Otting, G., Liepinsh, E., Halle, B., and Frey, U. 1997. NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. *Nat Struct Biol* 4: 396-404.
- Palmer III, A.G. 2001. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* 30: 129-155.

- Palmer III, A.G., Rance, M., and Wright, P. 1991. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterised by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J.Am.Chem.Soc.* 113.
- Pardi, A., Billeter, M., and Wuthrich, K. 1984. Calibration of the angular dependence of the amide proton-C alpha proton coupling constants, 3JHN alpha, in a globular protein. Use of 3JHN alpha for identification of helical secondary structure. *J Mol Biol* 180: 741-751.
- Park, C.J., Shin, R., Park, J.M., Lee, G.J., You, J.S., and Paek, K.H. 2002. Induction of pepper cDNA encoding a lipid transfer protein during the resistance response to tobacco mosaic virus. *Plant Mol Biol* 48: 243-254.
- Pastorello, E.A., Pompei, C., Pravettoni, V., Brenna, O., Farioli, L., Trambaioli, C., and Conti, A. 2001. Lipid transfer proteins and 2S albumins as allergens. *Allergy* 56 Suppl 67: 45-47.
- Pato, C., Le Borgne, M., Le Baut, G., Le Pape, P., Marion, D., and Douliez, J.P. 2001. Potential application of plant lipid transfer proteins for drug delivery. *Biochem Pharmacol* 62: 555-560.
- Pearson, W.R., and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444-2448.
- Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* 94: 12366-12371.
- Pons, J.L., de Lamotte, F., Gautier, M.F., and Delsuc, M.A. 2003. Refined solution structure of a liganded type 2 wheat nonspecific lipid transfer protein. *J Biol Chem* 278: 14249-14256.
- Poznanski, J., Sodano, P., Suh, S.W., Lee, J.Y., Ptak, M., and Vovelle, F. 1999. Solution structure of a lipid transfer protein extracted from rice seeds. Comparison with homologous proteins. *Eur J Biochem* 259: 692-708.
- Prange, T., Schiltz, M., Pernot, L., Colloc'h, N., Longhi, S., Bourguet, W., and Fourme, R. 1998. Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins* 30: 61-73.
- Press, W.H., Flannery, B.P., Teukolsy, S.A., and Vetterling, W.T. 1986. *Numerical recipes*, Cambridge.

- Pyee, J., Yu, H., and Kolattukudy, P.E. 1994. Identification of a lipid transfer protein as the major protein in the surface wax of broccoli (*Brassica oleracea*) leaves. *Arch Biochem Biophys* 311: 460-468.
- Ramachandran, G.N., Chandrasekaran, R., and Kopple, K.D. 1971. Variation of the NH-C α -H coupling constant with dihedral angle in the NMR spectra of peptides. *Biopolymers* 10: 2113-2131.
- Rico, M., Bruix, M., Gonzalez, C., Monsalve, R.I., and Rodriguez, R. 1996. ¹H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. *Biochemistry* 35: 15672-15682.
- Rodrigues, M.L., Archer, M., Martel, P., Jacquet, A., Cravador, A., and Carrondo, M.A. 2002. Structure of beta-cinnamomin, a protein toxic to some plant species. *Acta Crystallogr D Biol Crystallogr* 58: 1314-1321.
- Rusterucci C., Montillet J.L., Agnel J.P., Battesti C., Alonso B., Knoll A., Bessoule J.J., Etienne P., Suty L., Blein J.P., Triantaphylides C. 1999. Involvement of lipoxygenase-dependent production of fatty acid hydroperoxides in the development of the hypersensitive cell death induced by cryptogein on tobacco leaves. *J Biol Chem.* 274(51):36446-55.
- Ryckaert, J.P., Ciccotti, G., and H.J.C., B. 1977. Numerical integration of cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327.
- Sali, A., and Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815.
- Samuel, D., Liu, Y.J., Cheng, C.S., and Lyu, P.C. 2002. Solution Structure of Plant Nonspecific Lipid Transfer Protein-2 from Rice (*Oryza sativa*). *J Biol Chem* 277: 35267-35273.
- Sattler, M., Schleucher, J., and Griesinger, C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradient. *Progress in NMR spectroscopy* 34: 93-158.
- Schiltz, M., Fourme, R., and Prange, T. 2003. Use of noble gases xenon and krypton as heavy atoms in protein structure determination. *Methods Enzymol* 374: 83-119.
- Schwartz, R.M., and Dayhoff, M.O. 1978. "Matrices for detecting distant relationships." In "Atlas of Protein Sequence and Structure,". Natl. Biomed. Res. Found, Washington, DC., pp. 353-358.

- Segura, A., Moreno, M., and Garcia-Olmedo, F. 1993. Purification and antipathogenic activity of lipid transfer proteins (LTPs) from the leaves of *Arabidopsis* and spinach. *FEBS Lett* 332: 243-246.
- Shin, D.H., Lee, J.Y., Hwang, K.Y., Kim, K.K., and Suh, S.W. 1995. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* 3: 189-199.
- Simorre, J.P., Caille, A., Marion, D., and Ptak, M. 1991. Two- and three-dimensional ¹H NMR studies of a wheat phospholipid transfer protein: sequential resonance assignments and secondary structure. *Biochemistry* 30: 11600-11608.
- Sippl, M.J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins* 17: 355-362.
- Smith, C.A., Skvirsky, R.C., and Hirsch, A.M. 1986. Histochemical evidence for the presence of a suberin-like compound in *Rhizobium*-induced nodules of the non-legume *Parasponia rigida*. *Canadian journal of botany*: 1474-1483.
- Sodano, P., Caille, A., Sy, D., de Person, G., Marion, D., and Ptak, M. 1997. ¹H NMR and fluorescence studies of the complexation of DMPG by wheat non-specific lipid transfer protein. Global fold of the complex. *FEBS Lett* 416: 130-134.
- Sodano, P., and Ptak, M. 1995. Secondary structure in solution of the hydrophobic protein of soybean (HPS) as revealed by ¹H NMR. *J Biomol Struct Dyn* 12: 1009-1022.
- Sossountzov, L., Ruiz-Avila, L., Vignols, F., Jolliot, A., Arondel, V., Tchang, F., Grosbois, M., Guerbette, F., Miginiac, E., Delseny, M., et al. 1991. Spatial and temporal expression of a maize lipid transfer protein gene. *Plant Cell* 3: 923-933.
- Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A., and De Vries, S.C. 1991. Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* 3: 907-921.
- Sy, D., Le Gravier, Y., Goodfellow, J., and Vovelle, F. 2003. Protein stability and plasticity of the hydrophobic cavity in wheat ns-LTP. *J Biomol Struct Dyn* 21: 15-29.
- Tassin, S., Broekaert, W.F., Marion, D., Acland, D.P., Ptak, M., Vovelle, F., and Sodano, P. 1998. Solution structure of Ace-AMP1, a potent antimicrobial protein extracted from onion seeds. Structural analogies with plant nonspecific lipid transfer proteins. *Biochemistry* 37: 3623-3637.
- Tassin-Moindrot, S., Caille, A., Douliez, J.P., Marion, D., and Vovelle, F. 2000. The wide binding properties of a wheat nonspecific lipid transfer protein. Solution structure of a complex with prostaglandin B2. *Eur J Biochem* 267: 1117-1124.

- Tchang, F., This, P., Stiefel, V., Arondel, V., Morch, M.D., Pages, M., Puigdomenech, P., Grellet, F., Delseny, M., Bouillon, P., et al. 1988. Phospholipid transfer protein: full-length cDNA and amino acid sequence in maize. Amino acid sequence homologies between plant phospholipid transfer proteins. *J Biol Chem* 263: 16849-16855.
- Terras, F.R.G., Goderis, I.J., van Leuven, F., Vanderleyden, J., Cammue, B.P.A., and Broekaert, W.F. 1992. In vitro antifungal activity of a radish (*Raphanus sativus* L.) seed protein homologous to nonspecific lipid transfer protein. *Plant Physiol.* 100: 1055-1058.
- Thoma, S., Kaneko, Y., and Somerville, C. 1993. A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein. *Plant J* 3: 427-436.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
- Tjandra, N., and Bax, A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278: 1111-1114.
- Tolman, J.R., Flanagan, J.M., Kennedy, M.A., and Prestegard, J.H. 1995. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A* 92: 9279-9283.
- Tugarinov, V., Hwang, P.M., and Kay, L.E. 2004. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* 73: 107-146.
- Van Loon, L.C., and Van Strien, E.A. 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis PR-1 type proteins. *Physiological and Molecular Plant Pathology* 55: 85-97.
- Verlet, L. 1967. "Computer experiment" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review* 159: 98.
- Vignols, F., Wigger, M., Garcia-Garrido, J.M., Grellet, F., Kader, J.C., and Delseny, M. 1997. Rice lipid transfer protein (LTP) genes belong to a complex multigene family and are differentially regulated. *Gene* 195: 177-186.
- Vuister, G.W., and Bax, A. 1993. Quantitative J correlation: a new approach for measuring homonuclear three-bond $J(\text{H}^{\text{N}}\text{H}^{\text{a}})$ coupling constants in ^{15}N -enriched proteins. 1993 115: 7772-7777.
- Walker, T.G., and Happer, W. 1997. Spin-exchange optical pumping of noble-gas nuclei. *Reviews of Modern Physics* 69: 629-642.

- Williamson, M.P., Havel, T.F., and Wuthrich, K. 1985. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. *J Mol Biol* 182: 295-315.
- Woodcock, S., Mornon, J.P., and Henrissat, B. 1992. Detection of secondary structure elements in proteins by hydrophobic cluster analysis. *Protein Eng* 5: 629-635.
- Wüthrich, K. 1986. *NMR of proteins and nucleic acids.*, New York.

Annexe

En marge de mes principaux travaux de thèse, j'ai réalisé un travail de modélisation comparative en collaboration avec l'équipe de biologie cellulaire et moléculaire de la glycosylation de notre centre, afin d'obtenir un modèle tridimensionnel d'une α -GalNac transférase, enzyme de la O-glycosylation. Ce modèle a nécessité la construction du ligand de l'enzyme, l'UDP-*N*-acetylgalactosamine et son positionnement au sein du site catalytique par des méthodes de docking. Les résultats obtenus ont permis de suggérer des mutations qui ont abouti à l'identification de résidus importants pour l'activité enzymatique de l' α -GalNac transférase. Ces résultats sont présentés sous la forme d'un article qui est sous presse dans le journal *Protein Engineering Design and Selection*.

Characterization of the UDP-*N*-acetylgalactosamine binding domain of bovine polypeptide α *N*-acetylgalactosaminyltransferase T1

Stéphanie Duclos, Pedro Da Silva, Françoise Vovelle, Friedrich Piller and Véronique Piller¹

Centre de Biophysique Moléculaire, CNRS UPR 4301, affiliated with INSERM and the University of Orléans, F 45071 Orléans Cédex 2, France

¹To whom correspondence should be addressed at: Centre de Biophysique Moléculaire, Bât. B, rue Charles Sadron, F45071 Orléans Cédex 2, France. E-mail: pillere@cns-orleans.fr

UDP-GalNAc:polypeptide α *N*-acetylgalactosaminyltransferases (ppGaNTases) transfer GalNAc from UDP-GalNAc to Ser or Thr. Structural features underlying their enzymatic activity and their specificity are still unidentified. In order to get some insight into the donor substrate recognition, we used a molecular modelling approach on a portion of the catalytic site of the bovine ppGaNTase-T1. Fold recognition methods identified as appropriate templates the bovine α 1,3galactosyltransferase and the human α 1,3*N*-acetylgalactosaminyltransferase. A model of the ppGaNTase-T1 nucleotide-sugar binding site was built into which the UDP-GalNAc and the Mn²⁺ cation were docked. UDP-GalNAc fits best in a conformation where the GalNAc is folded back under the phosphates and is maintained in that special conformation through hydrogen bonds with R193. The ribose is found in van der Waals contacts with F124 and L189. The uracil is involved in a stacking interaction with W129 and forms a hydrogen bond with N126. The Mn²⁺ is found in coordination both with the phosphates of UDP and the DXH motif of the enzyme. Amino acids in contact with UDP-GalNAc in the model have been mutated and the corresponding soluble forms of the enzyme expressed in yeast. Their kinetic constants confirm the importance of these amino acids in donor substrate interactions.

Keywords: comparative molecular modelling/glycosyltransferases/mutagenesis/polypeptide α *N*-acetylgalactosaminyltransferases/UDP-GalNAc

Introduction

UDP-GalNAc:polypeptide α *N*-acetylgalactosaminyl-transferases (ppGaNTases EC 2.4.1.41) belong to a large family of enzymes (for review, see Ten Hagen *et al.*, 2003), 14 members of which have been cloned and 13 functionally characterized to date. Although they have been first described in mammals, large families of these enzymes have also been found in nematodes and in insects. They transfer, in the presence of Mn²⁺, GalNAc from the sugar donor UDP- α GalNAc to serine and threonine residues of the acceptor polypeptide in order to form: GalNAc α 1-*O*-Ser/Thr. This constitutes the first step of sugar addition during *O*-glycosylation of mucins, a process encountered in all animals. Acceptor substrates of ppGaNTases constitute a huge variety of peptide sequences that obviously require many different enzymes of this family in order to

glycosylate all the potential mucin-type *O*-glycan attachment sites. Besides, the different members of this family show specific spatiotemporal expression patterns during development that are probably related to the unique specificity of each member of the family. However, despite intense work aimed at finding the recognition criteria of acceptor substrates, we still do not know precisely the fine specificity of each enzyme. This is also due to the fact that no X-ray structure has been as yet reported for any member of the family and that no specific inhibitors have ever been designed for any enzyme isoform. Moreover, although the donor substrate (UDP-GalNAc) is identical for all the family members, very few data are available on the structural features that underlie the enzymatic transfer reaction and the specificity of recognition of both substrates.

Sequence analysis of the cDNAs of the different isoforms of the family and of the corresponding proteins have shown that these enzymes are type II membrane proteins (Figure 1A). They are anchored in the Golgi and contain a short (6–19 amino acids) N-terminal cytosolic domain, a unique transmembrane anchor (15–27 amino acids) followed by a luminal region which can be subdivided into three parts: a stem region of variable length (55–416 amino acids) which projects the catalytic entity (about 350 amino acids) fully accessible into the lumen of the Golgi; this catalytic domain is followed by a lectin-like domain of 130–150 amino acids at the C-terminus of all functional ppGaNTases. The comparison of the entire sequence of ppGaNTases shows that cytosolic, transmembrane, stem and lectin-like regions present very little sequence similarity, the regions of higher amino acid identity being situated in the catalytic domains (Figure 1B). The lectin-like domains show only structural similarities: molecular modelling studies have led to the conclusion that this part of the enzyme could adopt the same fold as the lectin domain of ricin (Imberty *et al.*, 1997; Hagen *et al.*, 1999). On the other hand, a careful examination of the catalytic domains of mammalian ppGaNTases emphasizes numerous sequence identities in all isoforms since a number of residues are perfectly conserved (Figure 1B). A majority of them is also found at the same position in the ppGaNTase isoforms of *Drosophila melanogaster* (Schwientek *et al.*, 2002) or of *Caenorhabditis elegans* (Hagen and Nehrke, 1998), which strongly suggests that they may be important for enzyme activity. A comparison of the peptide sequence of the catalytic domain among the different members of the family allow one to distinguish two highly conserved regions: a glycosyltransferase 1 or GT1 pattern and a Gal/GalNAcT pattern. The GT1 pattern comprises about 110 amino acids in the N-terminal half of the catalytic domain and contains a DXH motif, thought to correspond to the DXD motif present in many glycosyltransferases of different specificities (Breton *et al.*, 1998; Busch *et al.*, 1998; Kapitonov and Yu, 1999). In all glycosyltransferases of known three-dimensional (3D) structure which present this motif, it has been shown to bind the

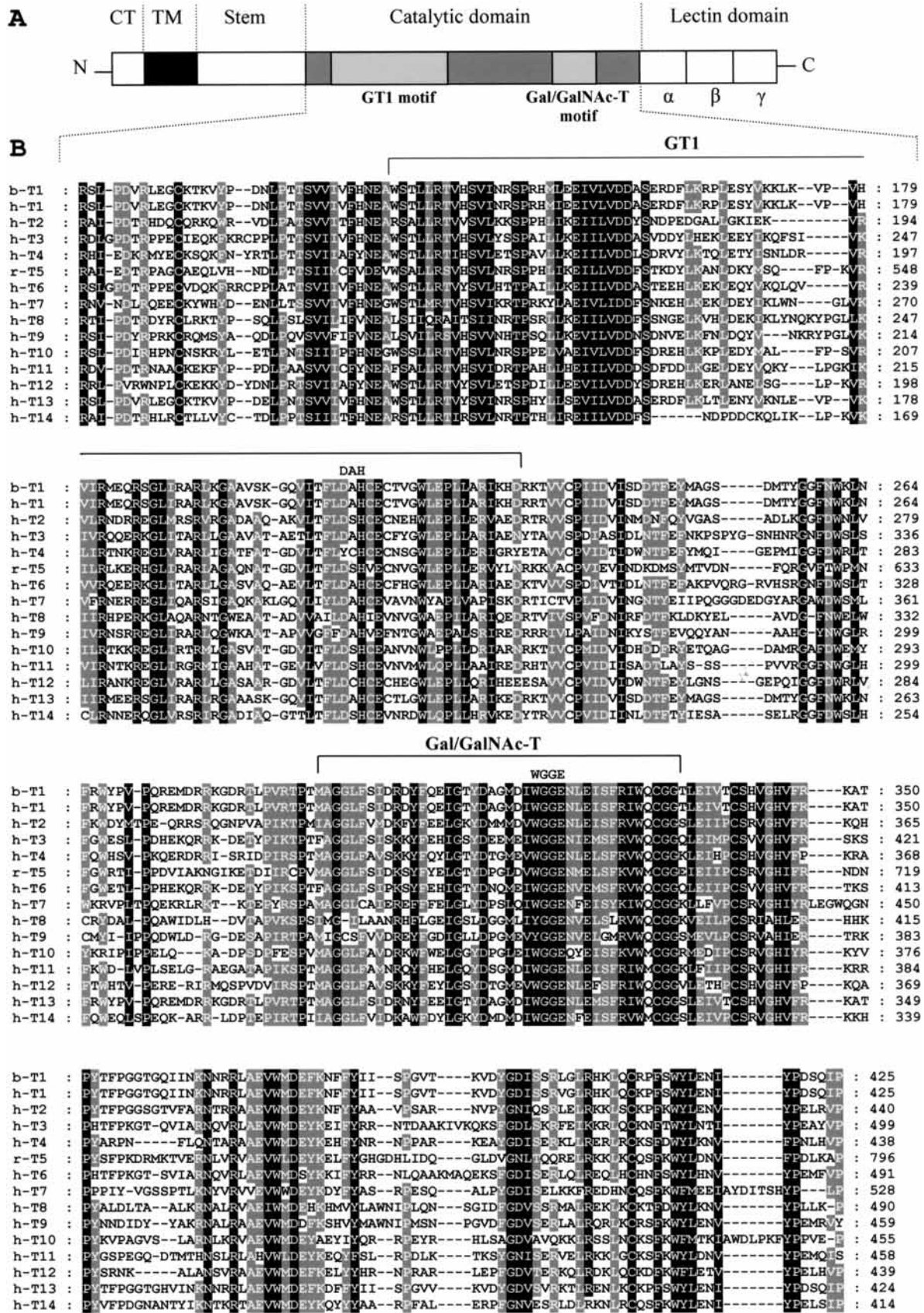


Fig. 1. Domain organization of ppGaNTases and multiple alignment of their catalytic domain. (A) Domain organization: CT, cytoplasmic tail; TM, trans-membrane region. α , β and γ point out the three different (QxW) units of the lectin domain. (B) Multiple alignment of the catalytic domains including the GT1 motif and Gal/GalNAc-T motif, of bovine ppGaNTase-T1 (b-T1), 13 human ppGaNTases (h-T1 to h-T4 and h-T6 to h-T14) one of which is an inactive form (h-T8) and ppGaNTase-T5 from rat (r-T5). Hyphens indicate introduced gaps and conserved residues are boxed (black for all 14 sequences, dark grey for 12–14 sequences and light grey for 9–11 sequences). The DAH and WGGE motifs are indicated above the aligned sequences.

sugar-nucleotide donor via an associated divalent metal ion. The Gal/GalNAcT pattern comprises about 40 amino acids in the C-terminal portion of the catalytic domain of ppGaNTases and contains a motif WXXE which has been suggested, in four of the glycosyltransferases of known 3D structure, to participate in the transfer of Gal or GalNAc (Gastinel *et al.*, 1999; Boix *et al.*, 2002; Patenaude *et al.*, 2002), although the available data do not allow firm conclusions. Until now, the atomic coordinates of only 18 crystallized glycosyltransferases have been published (Vrielink *et al.*, 1994; Charnock and Davies, 1999; Gastinel *et al.*, 1999, 2001; Ha *et al.*, 2000; Pedersen *et al.*, 2000, 2003; Unligil *et al.*, 2000; Mulichak *et al.*, 2001, 2003, 2004; Persson *et al.*, 2001; Boix *et al.*, 2002; Gibbons *et al.*, 2002; Gibson *et al.*, 2002; Patenaude *et al.*, 2002; Chiu *et al.*, 2004; Lobsanov *et al.*, 2004). Most of these enzymes present no or very little amino acid sequence similarity among each other. However, they all share some similitude in their overall protein fold. Bioinformatic analysis and 3D structure comparison have led to the proposal that glycosyltransferases adopt principally two different topologies: the nucleotide-diphospho-sugar transferase from *Bacillus subtilis* or SpsA fold and the DNA-modifying β -glucosyltransferase from bacteriophage T4 or BGT fold (Unligil and Rini, 2000; Zhang *et al.*, 2003 and references cited therein). The SpsA fold consists of a $\beta/\alpha/\beta$ Rossmann-like fold (Rossmann *et al.*, 1974) organized in such a way that central parallel β -sheets are flanked by α -helices and arranged in two tightly associated subdomains, one involved in the nucleotide binding and one in the acceptor binding. The BGT fold is characterized by two similar $\beta/\alpha/\beta$ fold subdomains less tightly associated than in the SpsA fold and with the binding pocket of the donor and acceptor substrates situated in the cleft between the two subdomains. Until now, all enzymes of characterized 3D structure with a SpsA fold have shown a DXD motif involved in the binding of a divalent cation itself coordinated to the dinucleotide sugar. Among the 18 crystallized enzymes of known 3D structure, half are retaining glycosyltransferases (same anomeric linkage of donor substrate and product) while the other half are inverting (different anomeric linkage of donor substrate and product). Six of them adopt the BGT fold while 12 adopt the SpsA fold that is the type of fold predicted for ppGaNTases (Hagen *et al.*, 1999; Breton *et al.*, 2002). The overall fold of the six enzymes which function with a retention of the transferred sugar configuration and present a SpsA topology is similar especially in the active-site pocket where the UDP-sugar is bound. However, although great efforts have been made to understand the enzymatic mechanism involved in the sugar transfer reaction for inverting as well as retaining glycosyltransferases (Unligil and Rini, 2000; Ly *et al.*, 2002) there are no data yet available on the precise enzymatic mechanism involved for any of the enzymes of known 3D structure.

Molecular modelling has been used successfully in order to predict the conformation of proteins of unknown 3D structure (for review, see Schonbrun *et al.*, 2002 and references therein). The models are built from templates provided by the X-ray structure of proteins that present a maximum of sequence similarities with the candidate enzyme. Analogous methods have been used for the prediction of all or part of the catalytic domain structure of glycosyltransferases (Imberty *et al.*, 1997, 1999; Tsai *et al.*, 2000; Unligil *et al.*, 2000; Rao and Tvaroska, 2001; de Vries *et al.*, 2001; Lazarus *et al.*, 2002; Ouzzine *et al.*, 2002; Gulberti *et al.*, 2003; Heissigerova *et al.*, 2003).

However, in most cases the predictions were not supported by biological or biochemical data. Since we are currently lacking structural information on ppGaNTases as well as on their enzymatic mechanism, we used molecular modelling to predict some features of the catalytic domain. For that approach we restricted our modelling to the GT1 subdomain that is the larger conserved motif of all ppGaNTases and used as templates the X-ray data of retaining glycosyltransferases which adopt a SpsA fold. The amino acids found to be in direct contact and/or important for the UDP-GalNAc binding were mutated in a secreted form of the ppGaNTase-T1; mutants were expressed in yeast, purified and their kinetic properties analysed.

Materials and methods

Sequence alignment and secondary structure prediction

The amino acid sequences of the bovine ppGaNTase-T1 and of 13 human ppGaNTases plus the ppGaNTase-T5 cloned from rat, were aligned using the ClustalW method (Thompson *et al.*, 1994) and refined manually using GeneDoc (Nicholas *et al.*, 1997). This multiple alignment of ppGaNTases was employed to carry out secondary structure predictions of the bovine ppGaNTase-T1 with two programs available on WWW servers: the PHDsec methods (Rost and Sander, 1993, 1994; <http://www.EMBL-Heidelberg.DE/services/index.html>), using a neural network algorithm based on position-specific conservation weight and the quadratic-logistic method (Di Francesco *et al.*, 1995; <http://abs.cit.nih.gov/ql2/>), operating according to a logistic discriminating function methodology. Other programs, which only work with unique sequences were run with the ppGaNTase-T1 sequence: the NPS method (http://pbil.ibcp.fr/NPSA/npsa_prediction.html) which provides a consensus secondary prediction using different algorithms and PREDATOR (Frishman and Argos, 1997; <http://bioweb.pasteur.fr/seqanal/interfaces/predator.html>) which uses information from related proteins in order to predict a secondary structure.

Fold prediction

Fold recognition was used to propose a structural model for the catalytic domain of the ppGaNTase-T1. Several fold recognition programs were run using the putative catalytic domain of the bovine ppGaNTase-T1 (amino acids 84–426) as query sequence: Superfamily (Gough *et al.*, 2001; <http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/>), mGenTHREADER (Jones, 1999; McGuffin *et al.*, 2000; <http://bioinf.cs.ucl.ac.uk/psipred/>), Hybrid Fold Recognition (Fischer, 2000; <http://www.cs.il/~bioinbgu/>) and LIBELLULA (Juan *et al.*, 2003; <http://www.pdg.cnb.uam.es/servers/libellula.html>), that uses and improves the results of two other methods (SAMT99 and 3DPSSM). These 'threading' processes (Lemer *et al.*, 1995) allow one to align one sequence with one structure.

Comparative modelling

The region comprised between amino acids 96 and 227 of the bovine ppGaNTase-T1 was aligned on the nucleotide-sugar binding segment of bovine α 1,3galactosyltransferases (α 3GalT) and the corresponding region of human blood group A transferase or α 1,3N-acetylgalactosaminyltransferase (GTA). Because of the very low sequence identity between ppGaNTase-T1 and the other transferases α 3GalT and GTA, the alignment was carried out according to the hydrophobic

cluster analysis (HCA) program (Gaboriaud *et al.*, 1987; <http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html>). This graphical method allows to easily visualize α -helices and β -strands and to align the conserved hydrophobic clusters.

The models of the [96–227] fragment of ppGaNTase-T1 were built using MODELLER 6.2 software (Sali and Blundell, 1993). MODELLER is a program based on the satisfaction of spatial restraints generated on the target sequence from its alignment with the 3D structure of the templates [α 3GalT in complex with UDP (1K4V; Boix *et al.*, 2001) and GTA in complex with UDP and the H antigen (1LZI; Patenaude *et al.*, 2002)] and also from the statistical analysis of a data base including 416 protein structures. The PROCHECK program (Laskowski *et al.*, 1993) was used to assess the stereochemical quality of the structures. Additional evaluations were completed by various profile programs: PROSA II (Sippl, 1993) and VERIFY3D (Luthy *et al.*, 1992; http://www.doe-mbi.ucla.edu/Services/Verify_3D/) and also by examination of the structure on a graphic display. Using these assessments, a representative model was selected and energy minimised with the SYBYL software (TRIPOS Inc., St Louis, MO). The secondary structures of the model were analysed with the PROMOTIF program (Hutchinson and Thornton, 1996).

Docking of UDP-GalNAc

UDP-GalNAc was built using the geometry of UDP-2FGal-Mn²⁺ in complex with the α 1,4 galactosyltransferase from *Neisseria meningitidis* or LgtC (Persson *et al.*, 2001). The fluorine atom was changed into an acetyl group and the coordinates of the molecule were refined by energy minimization within SYBYL. The ligand was then docked into the active site of ppGaNTase-T1. In a first step, the UDP moiety of UDP-GalNAc was placed in a position similar to the position of UDP ligands in α 3GalT and GTA X-ray structures. In a second step, the geometry of the whole ligand and of the protein side chains in the vicinity of the active site were energy minimized within SYBYL.

Site-directed mutagenesis

Site-directed mutagenesis was performed by PCR using the Quik-Change[®] site-directed mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. The plasmid pcDNA3 (Invitrogen, San Diego, CA) containing an insert encoding a soluble form of the bovine ppGaNTase-T1 (S.Duclos, N.Bureau, F.Piller and V.Piller, manuscript in preparation) was used as template with the following couples of primers: W129A, 5'-GTTTTCCACAATGAGGCT-GCCAGCACACTTCTGCGAAC-3' upstream and 5'-GTT-CGCAGAAGTGTGCTGGCAGCCTCATTGTGGAAAAC-3' downstream; W129F: 5'-GATTGTTTTCCACAATGAGGCTTTTCAGCACACTTCTGCGAACTGTC-3' upstream and 5'-GACAGTTCGACAGAAGTGTGCTGAAAGCCTCATTGTGGAAAACAATC-3' downstream; W129R: 5'-GGTGATTGTTTTCCACAATGAGGCTCGGAGCACACTTCTGCGA-ACTGTCCATAGC-3' upstream and 5'-GCTATGGACAGTTCGACAGAAGTGTGCTCCGAGCCTCATTGTGGAAAAC-AATCACC-3' downstream; N126A: 5'-CCTACAACCAGTGTGGTGATTGTTTTCCACGCTGAGGCTTGGAGCACACTTCTGCGAAC-3' upstream and 5'-GTTTCGACAGAAGTGTGCTCCAAGCCTCAGCGTGGAACAATCACCACACTGTTGTAGG-3' downstream; F124A: 5'-CCTTCCTACAACCAGTGTGGTGATTGTTGCCCAATGAGGCT-

TGGAGCACACTTCTGTC-3' upstream and 5'-GCAGAAGTGTGCTCCAAGCCTCATTGTGGGCAACAATCACCACACTGTTGTAGGAAGG-3' downstream; L189A: 5'-CGAA-CGAATGGAGCAGCGTTCTGGAGCCGATCAGAGCTAGGTTAAAAGG-3' upstream and 5'-CCTTTTAACCTAGCTCTGATCGCTCCAGAACGCTGCTCCATTTCG-3' downstream; R193A: 5'-CGTTCTGGATTGATCAGAGCTGCTTTAAAAGGTGCTGCTGTGTCTAAAGG-3' upstream and 5'-CCTTTAGACACAGCAGCACCTTTTAAAGCAGCTCTGATCAATCCAGAACG-3' downstream; D209A: 5'-AAGTGATCACCTTTTTAGCCGCGCACTGTGAG-3' upstream and 5'-CTCACAGTGCAGCGGCTAAAAGGTGATCACTT-3' downstream. Nucleotides mutated to convert one amino acid into another residue are underlined and in bold characters. All mutations were confirmed by sequencing (MWG Biotech, Germany). Each mutated cDNA was excised from the pcDNA3 vector by *Bam*HI and *Not*I and cloned in frame with the alpha-factor signal peptide into the YE_{FL}AGTM-1 yeast expression vector (Sigma, St Louis, MO) opened by the same enzymes.

Expression of wild type and mutant ppGaNTase-T1 in *Saccharomyces cerevisiae*

Saccharomyces cerevisiae (strain BJ5465) cells were transformed with the YE_{FL}AGTM-1 vector containing either the wild type cDNA or a mutated cDNA, using the lithium acetate transformation method (Ito *et al.*, 1983). The selection method is based on the Trp auxotrophy of the BJ5465 strain. Transformants were selected on Trp-deficient synthetic complete agar (0.67% yeast nitrogen base without amino acids, 0.074% complete supplement mixture minus Trp, 2% dextrose and 2% agar) since the TRP1 gene of the vector YE_{FL}AGTM-1 complements the mutation of the BJ5465 strain and restores the Trp-prototrophy. Selected clones were grown at 30°C in SC minus Trp liquid medium for 48 h (OD_{600nm} = 8–10). Starter cultures were diluted to OD_{600nm} = 0.4 in the expression medium (1% yeast extract, 8% peptone, 1% glucose, 3% glycerol and 1 mM CaCl₂) for induction of protein expression. The expression cultures were grown at 30°C for 72 h.

Purification of recombinant ppGaNTase-T1 and mutants

Ten millilitres of expression culture were centrifuged for 5 min at 10 000 g and 4°C, and the supernatant was dialysed against 50 mM sodium phosphate buffer, pH 8, at 4°C. The supernatant was adjusted to 300 mM NaCl and 5 mM imidazole and then applied onto a 1 ml nickel-agarose column (Qiagen, Germany) equilibrated with the 50 mM sodium phosphate buffer, pH 8, 300 mM NaCl and 5 mM imidazole. The column was washed once with 5 ml of the same buffer and once with 5 ml of the same buffer but 10 mM in imidazole. The bound enzyme was eluted in 500 μ l fractions of 250 mM imidazole, pH 7 and analysed by SDS-PAGE, western blot and enzymatic activity.

SDS-PAGE and immunoblotting

Eluted enzymes (5 μ l) were analysed by SDS-PAGE (Laemmli, 1970) using an 8% (w/v) polyacrylamide separation gel. Proteins were transferred onto nitrocellulose membrane (Schleicher & Schuell, Germany) and blocking was carried out for 30 min in Tris-buffered saline (TBS) containing 0.5% (v/v) Tween-20 (TBS-T). The membrane was first incubated for 30 min with a mouse anti-FLAGTM antibody (Sigma)

at 10 µg/ml TBS. After three washings in TBS, the second antibody, a goat anti-mouse antibody coupled to alkaline phosphatase (Bio-Rad, Hercules, CA) was used at a dilution of 1:2500 in TBS-T. The coloured reaction was developed with the Western BlueTM detection reagent (Promega, Madison, WI) and the amount of ppGaNTase-T1 or mutant enzymes was determined by densitometry analysis (Image Quant; Molecular Dynamics, Sunnyvale, CA) by comparison to 50 and 100 ng of FLAGTM-tagged inactive bacterial alkaline phosphatase (BAP-FLAGTM; Sigma).

Determination of ppGaNTase activity and kinetic parameters

PpGaNTase-T1 activity was assayed by measuring the synthesis of [³H]GalNAc-substituted peptide in 25 µl final volume reactions. Standard reaction mixtures contained 50 mM MES, pH 6.5, 15 mM MnCl₂, 1 mg/ml BSA, 20 µM UDP-[³H]GalNAc (American Radiolabeled Chemicals, St Louis, MO) with a specific activity of 50 000 c.p.m./nmol, 0.7 mM acceptor peptide STPSTPSTPSTP (STP₅, Sigma) and an appropriate amount of enzyme such as <10% of UDP-[³H]GalNAc were consumed at the end of the reaction (about 0.3 µU enzyme per test). After 30 min incubation at 37°C and separation by Dowex-1X8 (Serva, Germany) chromatography, the products were quantified by scintillation counting. All reactions were done in triplicates.

To determine the apparent K_m for UDP-GalNAc, the concentration of UDP-[³H]GalNAc was varied from 3.5 to 100 µM (and up to 250 µM for mutants N126A and W129A) in the presence of 0.7 mM STP₅. The apparent K_m for the acceptor peptide was obtained by varying the concentration of STP₅ from 0.15 to 1.5 mM in the presence of 100 µM UDP-[³H]GalNAc. Use of double reciprocal plots (1/V versus 1/[S]) allowed the calculation of kinetic parameters.

Results

Since the sequence identities between the protein to be modelled (ppGaNTase-T1) and the templates (retaining glycosyltransferases of known 3D structure) are low, the first step was to perform a secondary structure prediction for the ppGaNTase catalytic domain. Fold recognition programs were also run to search in the Protein Data Bank (PDB) for proteins adopting the same fold as predicted for the ppGaNTase catalytic domain. Once these proteins were identified, the corresponding sequences were aligned with those of the ppGaNTase-T1 and the alignment was then refined using the HCA analysis as well as the known secondary structures of the templates and the predicted structure of the ppGaNTase-T1. From this alignment a model of a subdomain of the bovine ppGaNTase-T1 could be built with the MODELLER program.

Sequence alignment, secondary structure prediction and fold recognition of ppGaNTases

Most ppGaNTase isoforms that have been cloned in mammals are of human origin; however, whenever a corresponding enzyme of another species has been cloned (bovine, mouse, rat or porcine) and functionally expressed, it showed only differences in a few amino acids with the human enzymes (90–99% identity among orthologues). For instance, among the 559 amino acids that constitute the entire protein sequence of the bovine ppGaNTase-T1 used for this study, only four

residues differ from the human ppGaNTase-T1. Thus, 13 human isoform sequences (ppGaNTases-T1–T4 and -T6–T14) and the primary structure of ppGaNTase-T5 from rat (because there is no human equivalent cloned to date) were compared in this study. The sequence of the bovine ppGaNTase-T1 is reported above the corresponding human sequence in Figure 1B. Due to low sequence identity in the cytosolic, transmembrane, stem and lectin domains, only the putative catalytic ppGaNTase domains (corresponding to residues 84–426 of the bovine ppGaNTase-T1) are aligned. The multiple alignment thus obtained and the sequence of the bovine ppGaNTase-T1 were submitted to several secondary structure prediction programs. The resulting predicted topology found for the ppGaNTase-T1 fold is characteristic of a Rossmann fold with alternating α -helices and β -strands. In the N-terminal half of the considered sequence (corresponding to residues 117–242 of the bovine ppGaNTase-T1) the different methods agree perfectly, suggesting a high reliability of the predictions. On the other hand, the predictions are more variable in the second half (residues 243–426 in the bovine ppGaNTase-T1) of the catalytic domain.

The 84–426 amino acid sequence of ppGaNTase-T1 was also submitted to a fold recognition analysis in order to search for proteins that share the same fold with the bovine ppGaNTase-T1. The secondary structures given by fold recognition programs were compared to those obtained by the secondary structure prediction programs and this allowed one to select the templates according to the highest scores of the structural prediction. Threading calculations aligned the ppGaNTase-T1 sequence with X-ray structures of glycosyltransferases which adopt a SpsA fold: SpsA (Charnock and Davies, 1999), LgtC (Persson *et al.*, 2001), GTA and human blood group B transferase (GTB) (Patenaude *et al.*, 2002) and α 3GalT (Gastinel *et al.*, 2001; Boix *et al.*, 2002). The first two are bacterial enzymes. SpsA is believed to be an inverting glycosyltransferase, a member of the family 2 of glycosyltransferases (GT-2 family, CAZY database), able to bind UDP as well as TDP (Tarbouriech *et al.*, 2001), although its specificity for both donor and acceptor substrates is still unknown. LgtC is a retaining glycosyltransferase which transfers Gal from UDP-Gal in an α -linkage to a terminal Gal of lipopolysaccharides, forming a Gal α 1,4 Gal linkage, and is assigned on the basis of sequence similarity to the GT-8 family (CAZY database). Human ABO blood group transferases, GTA and GTB are enzymes showing high sequence identity (only four different residues on a total of 354 amino acids); GTA transfers GalNAc from UDP-GalNAc and GTB transfers Gal from UDP-Gal to the same acceptor [Fuc α 1,2] Gal, both in α configuration, forming the blood group A and B antigens: GalNAc α [Fuc α 1,2] Gal and Gala [Fuc α 1,2] Gal, respectively. Based on sequence similarity, they are classified in the retaining GT-6 family as well as α 3GalT which is a bovine enzyme transferring Gal from UDP-Gal in α to Gal β 1,4GlcNAc structures, forming the xenoantigen Gal α 1,3Gal β 1,4GlcNAc responsible for the hyperacute rejection of xenografts in humans. GTA, GTB and α 3GalT are all involved in the biosynthesis of Gal(Nac) α 1,3Gal-containing antigens; they display many amino acid similarities and exhibit roughly the same fold: two subdomains separated by the catalytic pocket, one N-terminal binding UDP-Gal or UDP-GalNAc, adopting the Rossmann fold characteristic of nucleotide binding regions and one C-terminal involved in the acceptor substrate recognition.

They are enzymes of mammalian origin, resident in the Golgi apparatus and present a type II membrane protein orientation. We finally retained as templates only the GTA and the $\alpha 3\text{GalT}$ and discarded all the other potential candidates, either because their primary structure or their substrate specificities were too far from those of ppGaNTases (SpsA and LgtC) or because, in the case of GTB, the identity with GTA is so high that including both GTA and GTB would strongly bias the modelling procedure.

Alignment of $\alpha 3\text{GalT}$, GTA and bovine ppGaNTase-T1 and molecular modelling of the UDP-GalNAc binding domain

An alignment of the putative catalytic domain of ppGaNTase-T1 and of the identified catalytic domains of $\alpha 3\text{GalT}$ and GTA was undertaken using the hydrophobic cluster analysis (HCA). The HCA plot obtained for the first N-terminal part of the putative catalytic domain of ppGaNTase-T1 (amino acids 96–227) appeared closely related to the HCA plots of the N-terminal region of the catalytic domains of both $\alpha 3\text{GalT}$ and GTA corresponding to the UDP binding sites in the two enzymes. An alignment of the three HCA plots

was thus possible in this region and the hydrophobic clusters were adjusted manually in order to take into account the secondary structure data which coincide particularly well in this region showing for the templates as well as for ppGaNTase-T1 a regular alternation of α -helices and β -strands (Figure 2). The resulting sequence alignment is given in Figure 3. On the other hand, the C-terminal part of the putative catalytic domain of ppGaNTase-T1 is longer than the corresponding C-terminal catalytic domains of $\alpha 3\text{GalT}$ and GTA. The large loop (amino acids 242–274) separating the aligned stretches (GT1 and Gal/GalNAc-T, Figure 1) hinders the modelling process, therefore only the sequence comprised between amino acids 96 and 227 of bovine ppGaNTase-T1 was modelled from the corresponding sequences (amino acids 93–227 for GTA and 116–241 for $\alpha 3\text{GalT}$) of the templates. A best-fitted superimposition of the two templates in blue and green (Figure 4A) shows that the structures are very similar, although the coordinates of the 177–195 segment of GTA are missing in the PDB file. The r.m.s.d. between the C α coordinates of the aligned residues (107 in total) is 2.6 Å.

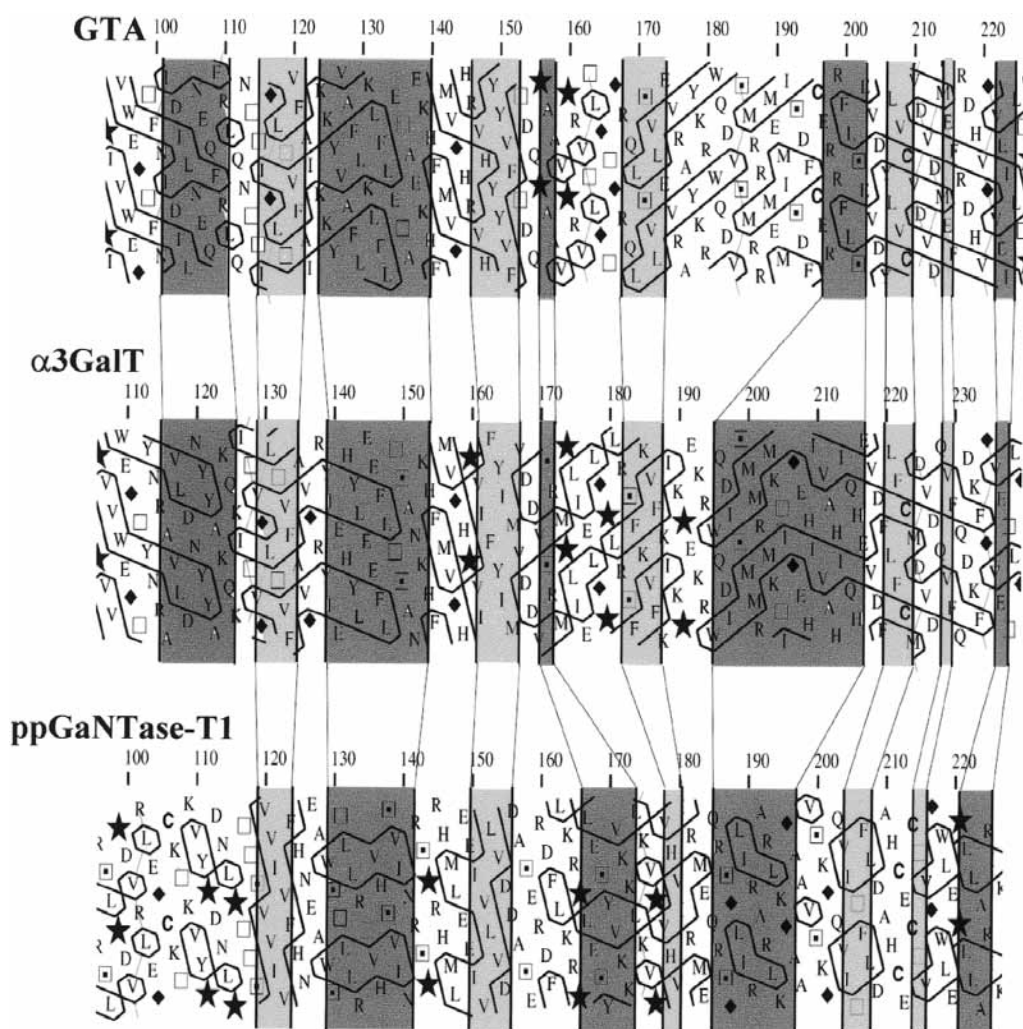


Fig. 2. HCA alignment of the N-terminal moiety of the catalytic domains of bovine $\alpha 3\text{GalT}$, human GTA and bovine ppGaNTase-T1. Diamond, star, square and square with a dot represent, respectively, G, P, T and S. Secondary structures are delimited by vertical lines and differentiated by the background colour; dark grey for α -helices and light grey for β -strands. Predicted secondary structures of bovine ppGaNTase-T1 follow the same representation. The DXD motif, known to coordinate the Mn^{2+} in GTA and $\alpha 3\text{GalT}$ is given as white text on black background, as well as the DAH motif of ppGaNTase-T1.

```

GTA:   PIVWEGTFNIDILNEQFRONTTIGLTVFAIKKYVA-FLK :131
α3GalT:PVVWEGTYNRAVLNDNYAKQKITVGLTVFAVGRYIEHYLE :145
b-T1:  RSLPDVRLEGCKTKVPYDNLPTTSVVIVFHNEAWST-LLR :134
      * * *

GTA:   LFLTEAEKHFMVGHVRVHYVFTDQPAVPRVT-----LG :165
α3GalT:EFLLTSANKHFMVGHVPVIFYIMVDDVSRMPLIE-----LG :179
b-T1:  TVHVSVINR--SPRHMLEEIVLVDDASERDFLKRPLESYVK :172

GTA:   TG-RQLSVLEVVRAYKRWQDVSMRRMEMISDFCERRFLSEV :204
α3GalT:PL-RSFKVFKIKPEKRWQDISMMRMKTIGEHIVAHQHEV :218
b-T1:  KIKVPVHVIRM---EORSGLIRARLKGA-----AVSKG :204
      * *

GTA:   DYLVCDVD--MEFRDHVGVEILTTP :227
α3GalT:DFLFCMDVD--QVFQDKFGVETLGE :241
b-T1:  QVITFLDAHCECTVGWLEPLLARIK :227

```

Fig. 3. Amino acid sequence alignment of bovine α3GalT, human GTA and bovine ppGaNTase-T1. b-T1, bovine ppGaNTase. Secondary structure elements are indicated by dark grey (helices) or light grey (β-sheets) boxes. Asterisks indicate amino acids of the bovine ppGaNTase-T1 interacting with UDP-GalNAc according to the model. The DXH motif is highlighted in black.

Overall structure of the model

Ten models of the segment 96–227 of bovine ppGaNTase-T1 were built according to the procedure described in the Materials and methods. After evaluation with PROCHECK, VERIFY-3D and PROSA II, one model was selected and considered as representative of the 3D structure of this part of the ppGaNTase-T1. As expected, the modelled enzyme subdomain adopts the α/β fold characteristic of nucleotide binding regions (Figure 4A). The structure starts at R96, with the α2 helix (α3GalT nomenclature) from E104 to Y111 followed by the β2 strand (V121–H125). This strand is connected to the β3 strand (E149–D155) by the long α3 helix (S130–S143), which is preceded by a 3_{10} helical turn from N126 to W129. The two chain reversals between the β3 and the β4 strands are ensured by a type I β turn between A157 and R160 which replaces the 3_{10} helical turn found in α3GalT and by a long loop between L167 and K175 corresponding to a large insertion compared to the sequences of the two templates. This β4 strand is very short (I181–M183), but residues 172–185 are found in an extended conformation. The α4 helix which is missing in the density map of GTA runs from R186 to V204; the β5 and β6 short strands which frame the DXD motif in α3GalT structures are not found although residues T206–A210 and G217–K223 are in an extended conformation and the 206–210 segment is parallel to the three parallel strands β2, β3 and β4 while the fragment 217–223 runs in an antiparallel direction. The presence of an insertion after the DXH motif in the ppGaNTase-T1 sequence leads to the formation of a six-residue turn between these two ‘pseudo’ strands. This turn is much longer than that observed for the two templates but after G127, the backbones of ppGaNTase-T1, α3GalT and GTA are well superimposed. Globally, the proposed 3D structure for ppGaNTase-T1 is rather close to the structure of the templates in its N-terminal part. The r.m.s.d. between the coordinates of the Cα atoms of the secondary structure elements of residues 120–186 of α3GalT and of the modelled structure is only 1.89 Å. However, the modelled structure differs more from that of the templates

in the helix α4 position and in the loops and turns which link the secondary structure elements in the C-terminal region.

UDP-GalNAc was docked into the active site which appears as a deep cleft located between the parallel β-sheets and helix α4 (Figure 4B). The catalytic pocket includes the DXH motif that replaces the DXD motif common to a wide range of glycosyltransferases. This motif was shown to be involved in the coordination of the Mn^{2+} divalent cation that is assumed to stabilize the UDP leaving group during bond cleavage (Pedersen *et al.*, 2000; Tarbouriech *et al.*, 2001). In our model, the Mn^{2+} ion shows seven interactions of coordination. It involves the two Od oxygens of D209 and the Ne of H211 (Figure 4C and Table I) but also the four phosphate oxygens of UDP. The characteristic distances of coordination range between 2.1 and 2.35 Å. The binding of UDP-GalNAc is stabilized by a number of specific interactions (Figure 4C and Table I). The uracil base is well stacked on the ring of W129, and its O2 carbonyl is involved in an hydrogen bond with the NH group of the N126 backbone, while the ribose is in van der Waals contacts with F124 and L189 and the NεH2 of N126 forms a H-bond with the O2 of the ribose. The galactosyl ring adopts a 4C_1 chair conformation similar to that found for other UDP-Gal(NAc) molecules (Persson *et al.*, 2001; Pedersen *et al.*, 2003). The O3' atom of the sugar is hydrogen bonded to the NHe atom of the guanidinium group of R193. A hydrogen bond is also found between the NηH2 of R193 and the oxygen atom of the acetyl group. Moreover, the GalNAc moiety makes favourable van der Waals contacts with L194 and L208.

Analysis of mutant forms of the ppGaNTase-T1

Molecular modelling of the most N-terminal part of the putative ppGaNTase-T1 catalytic domain points out several possible interactions between the donor substrate UDP-GalNAc and amino acids conserved in this subdomain for all ppGaNTases. In order to test if these interactions are really important for the expression and/or activity of the ppGaNTase-T1, we carried out a series of mutagenesis experiments on several amino acids present in this region and expressed the corresponding mutant enzymes. For that purpose, cDNA constructs have been designed for the expression of a soluble and tagged form of the ppGaNTase-T1 in the culture medium of the yeast *S.cerevisiae*, allowing one to readily obtain rather large amounts of pure enzyme after a fast and simple purification step (Figure 5). All the different forms of the enzyme were not grossly misfolded since they were expressed and secreted into the medium. The wild-type and mutant forms obtained therein were used for K_m and k_{cat} determinations with both the donor and the acceptor substrates. The results are reported in Table II. For several mutants, the low activity of the enzyme did not allow kinetic constant determination (F124A, L189A, R193A and D209A).

Mutagenesis of the DXH motif of the ppGaNTase-T1 (D209, A210, H211) (Hagen *et al.*, 1999) has already shown how this short sequence is important for the catalytic activity: when D209 was changed to A or N, the mutant had no activity and even when it was replaced by another acidic amino acid (D209E) the enzyme lost its function, showing how crucial is the positioning of this residue in the catalytic pocket. Mutation of the H residue of the same DXH sequence either strongly reduced (H211A) or completely abolished (H211D) the enzyme activity. In the model presented here, the DXH motif of the ppGaNTase-T1 is used for the docking of UDP-GalNAc and finally appears, after energy minimization, to

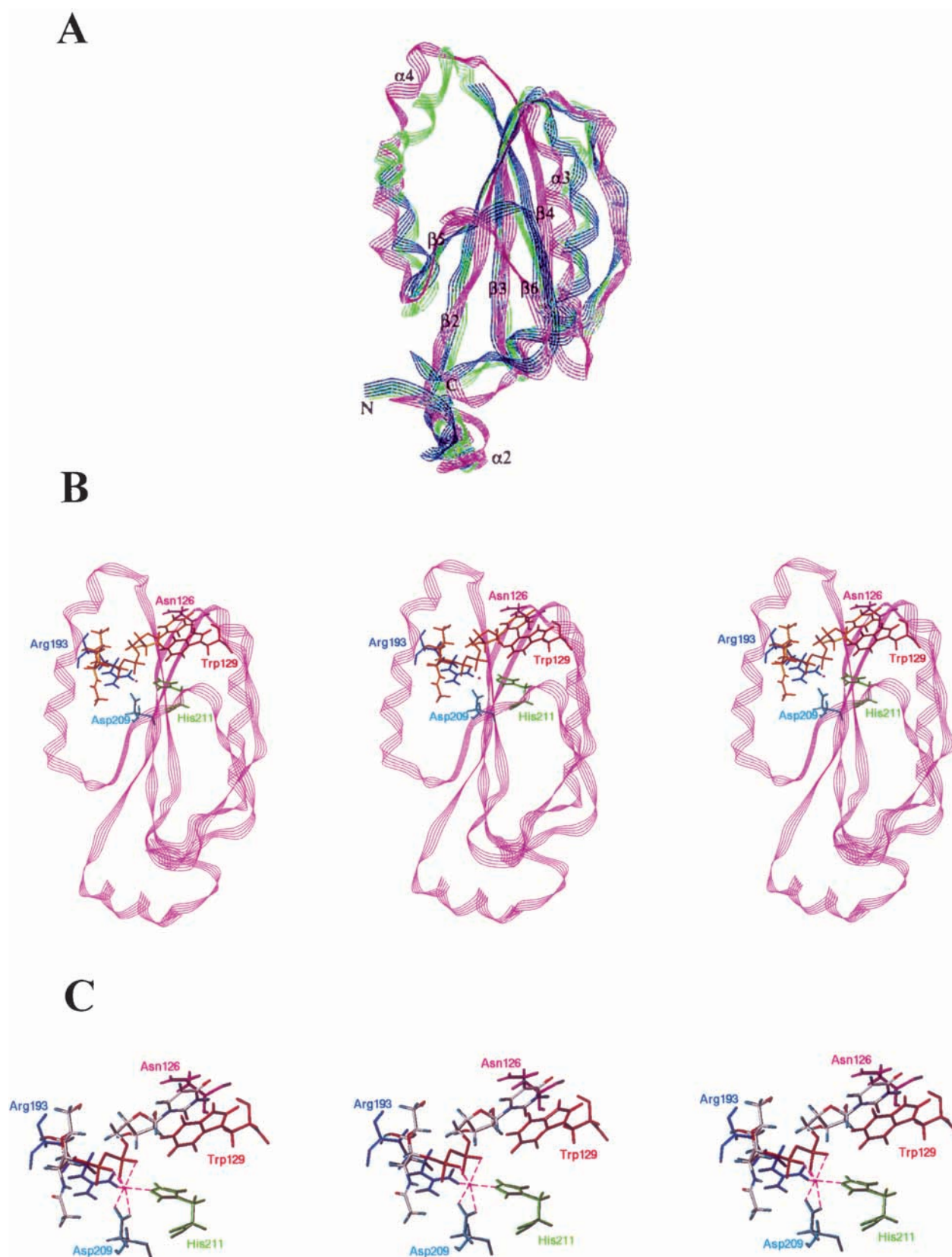


Fig. 4. Superimposed ribbon diagrams of the modelled region of bovine ppGaNTase-T1 with the structure of the two templates, stereo view of the ribbon diagram of the modelled region of bovine ppGaNTase-T1 with UDP-GalNAc and Mn^{2+} and close-up stereo view of the UDP-GalNAc binding site of the modelled region of bovine ppGaNTase-T1. The ribbon diagrams were drawn with SYBYL. (A) The modelled region of bovine ppGaNTase-T1 is in pink and the structures of the two templates are in blue (GTA) and in green ($\alpha 3$ GalT). (B) In the stereo view of the ribbon diagram of the modelled region of bovine ppGaNTase-T1 with UDP-GalNAc and Mn^{2+} , UDP-GalNAc is in orange, Mn^{2+} in magenta and for more clarity only the side chains of some of the amino acids involved in donor substrate binding are shown. (C) The ribbon diagrams from Figure 4A and B have been deleted and again, only some of the interacting aminoacids are shown. Ionic interactions are represented as purple dotted lines.

Table I. Interactions between nucleotide sugars and amino acids in the catalytic site for the crystallised enzymes α 3GalT (Boix *et al.*, 2002; Gastinel *et al.*, 2001) and GTA (Patenaude *et al.*, 2002) used as templates and for the ppGaNTase-T1 model

Ligand atom	Enzymes		
	α 3GalT	GTA	ppGaNTase-T1
Hydrogen bonds	Val136 (NH)		
Uracil (O2)	Val136 (O)	Ile123 (NH)	Asn126 (NH)
Uracil (NH3)	Asp227 (O δ 1)	Ile123 (O)	Asn126 (NeH2)
Ribose (O2)	Arg202 (N η H2)	Asp213 (O δ 1)	Arg193 (NH ϵ)
Ribose (O3)		Arg188 ^a	Arg193 (N η H2)
Gal (O3')			
GalNAc (O3')			
GalNAc (CO)			
van der Waals contacts			
Uracil	Tyr139	Tyr126	Trp129
Ribose	Phe134 ^b	Phe121 ^b	Phe124
Ribose	Ile198		Leu189
GalNAc			Leu194
GalNAc			Leu208
Electrostatic interactions			
Mn ²⁺	D225 (O δ 1)	D211 (O δ 1)	D209 (O δ 1)
Mn ²⁺			D209 (O δ 2)
Mn ²⁺	D227 (O δ 1)	D213 (O δ 1)	His211 (Ne)
Mn ²⁺	D227 (O δ 2)	D213 (O δ 2)	

^aArg188 was not seen in the crystal of the wild-type enzyme but only in a mutant and with a donor substrate modelled in the binding site.

^bThe two Phe which also establish hydrogen bonds with the O2 of the ribose.

stabilize the manganese ion and to bind via this cation the diphosphate moiety of UDP. The bidentate interaction of D209 through O δ 1 and O δ 2 with the Mn²⁺ ion as well as the interaction of H211 through its Ne atom with the manganese ion is consistent with the mutagenesis data obtained by (Hagen *et al.*, 1999) on this motif. W129 appears in our model to make a stacking interaction with the uracil ring. Mutation of this amino acid to A reduces the transfer efficiency to less than one tenth of the wild-type value since it increases the K_m for UDP-GalNAc by a factor 3.7 and for the acceptor peptide by a factor 2.5, suggesting a role of this residue in the substrates binding. On the other hand, replacement of this W129 by an F or by a R residue has little effect on the K_m for both substrates, indicating that both phenylalanine or arginine can interact with uracil as well as a tryptophan. N126 is also seen in our model as blocking the uridine position through two hydrogen bonds. When we mutate this residue to A, we observe an increase (5.5-fold) of the K_m for UDP-GalNAc while the K_m for the peptide is only 1.6 times higher, confirming the possible interaction of the donor substrate with this asparagine. The model also shows R193 interacting through two hydrogen bonds with the GalNAc moiety of the donor substrate. When we change this residue to A, the protein is still expressed and secreted but loses all its activity. Finally, the model also displays van der Waals contacts between the ribose moiety of uridine and F124 and L189 of the enzyme. When both amino acids are independently mutated to A, this results in a strong reduction of the activity.

Discussion

In order to get new insights into the enzymatic mechanism of the ppGaNTase family, we modelled a part of the catalytic domain of one of these retaining glycosyltransferases. For that

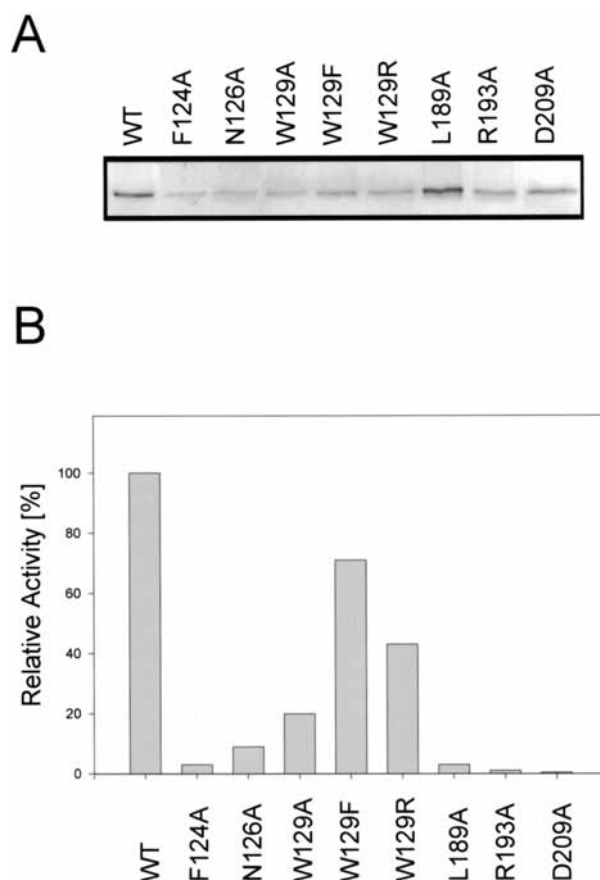


Fig. 5. Western blot of the mutants and wild-type soluble ppGaNTase-T1 and their relative enzyme activity. (A) Mutants and wild-type soluble ppGaNTase-T1 were expressed in yeast and purified by affinity chromatography. After western blotting the proteins were detected using an anti-FLAGTM monoclonal antibody followed by a goat anti-mouse antibody coupled to alkaline phosphatase. In order to compare the mutant protein expression in *S.cerevisiae*, the amount of mutant and wild-type proteins loaded onto the gel corresponded to the same volume of culture. The staining intensities were measured with the ImageQuant software and quantified using 50 and 100 ng of FLAGTM-tagged inactive bacterial alkaline phosphatase as standards. The values thus obtained were used for the relative activity determination shown in (B). The relative enzyme activity corresponds to the amount of GalNAc transferred by 1 μ g of enzyme protein in the standard enzyme assay (wild-type = 71 nmol/30 min) as described in Materials and methods. The values are the mean of three independent experiments.

Table II. Kinetic parameters for soluble recombinant wild-type and mutant ppGaNTase-T1

Enzyme	k_{cat} (s ⁻¹)	UDP-GalNAc		Acceptor peptide	
		K_m (μ M)	k_{cat}/K_m (μ M ⁻¹ s ⁻¹)	K_m (μ M)	k_{cat}/K_m (μ M ⁻¹ s ⁻¹)
Wild-type	3.47	8	0.433	250	0.014
F 124 A	–	–	–	–	–
N 126 A	2.53	44	0.057	410	0.006
W 129 A	0.8	29	0.027	650	0.001
W 129 F	2.43	14	0.173	205	0.012
W 129 R	2.79	17	0.164	200	0.014
L 189 A	–	–	–	–	–
R 193 A	–	–	–	–	–
D 209 A	–	–	–	–	–

The values were determined as indicated in the Materials and methods. –, activity not detectable.

purpose, we selected a sequence particularly well conserved among all ppGaNTases family members, established the most likely secondary structure for this sequence and used it to screen the PDB for proteins of similar fold. Using this approach two proteins with particular high structural similarity scores emerged from the search, GTA and $\alpha 3$ GalT, two recently crystallized retaining glycosyltransferases with known atomic coordinates. The structures of the regions homologous to ppGaNTase-T1 were used as templates to build a model of the ppGaNTase-T1 GT1 subdomain. The results are presented here as the model of the 96–227 segment of the bovine ppGaNTase-T1 in which the donor substrate UDP-GalNAc is docked in the conformation described in the crystal structure of the LgtC in complex with donor and acceptor sugar analogues (Persson *et al.*, 2001). A similar geometry has been found in a mouse $\alpha 1,4$ -N-acetylhexosaminyltransferase (EXTL2) co-crystallized with UDP-GalNAc (Pedersen *et al.*, 2003) as well as in the crystal structure of the human $\beta 1,3$ -glucuronyltransferase I (GlcAT-I) interacting with UDP-GlcUA (Pedersen *et al.*, 2002) and in the bacterial trehalose-6-phosphate synthase (OtsA) crystal in complex with UDP-Glc (Gibson *et al.*, 2004). For those enzymes the donor substrate has been found in a special conformation where the sugar is folded back under the diphosphates of the UDP. Although this peculiar curled conformation of the donor substrate could not be related to a precise mechanism of glycosyl transfer, it may be necessary for the catalysis (Negishi *et al.*, 2003; Gibson *et al.*, 2004). Our modelling shows that this donor substrate conformation may also exist in ppGaNTases. Besides, after energy minimization of the different models built in our study, the most favourable structures corresponded always to a folded back conformation of the UDP-GalNAc (data not shown).

UDP-GalNAc appears in the model described here to lie in a deep cleft containing the DXD-like motif and composed by a central sheet of β -strands surrounded by α -helices. Thus, the modelled catalytic subdomain of this enzyme adopts a typical SpsA fold as already predicted for this enzyme family (Breton *et al.*, 1998; Kikuchi *et al.*, 2003; Liu and Mushegian, 2003) and as expected from modelling carried out on the basis of two template sequences adopting a SpsA fold (GTA and $\alpha 3$ GalT).

Our model displays several well-defined interactions between the substrate and the enzyme, which can be directly compared to the same type of interactions found in both templates (see Table I). The amino acids suspected to be involved in those interactions have been mutated in the ppGaNTase-T1 and the results strongly support the model. First, the DXD sequence is a motif found in most prokaryotic and eukaryotic glycosyltransferases and in other nucleotide binding proteins as well. It has been shown to interact with the phosphate groups of the nucleotide through a divalent cation. If we consider the different glycosyltransferases of known 3D structure, the DXD sequence does not seem always to participate in the same manner in substrate binding. As a matter of fact the conservation of this short sequence is not absolute in glycosyltransferases since it can be either DXD, DDD, DXX or XDD. In the ppGaNTase family this sequence is a DXH motif, well conserved for all the members, except for the ppGaNTase-T4, for which the D209 residue is replaced by a Y. In the model, a change from D to Y is possible since the OH of the tyrosine can easily replace the COO^- of the aspartic acid in the interaction with the Mn^{2+} . However, Y being

much bulkier than D, this replacement probably alters the folding of the adjacent peptide chain which may in part account for the unique substrate specificity of the ppGaNTase-T4 as compared to the other members of the ppGaNTase family (Hassan *et al.*, 2000). The interaction of DXH with the Mn^{2+} ion involved in the donor substrate binding, as shown in the model, is characteristic of the type of interaction described for retaining glycosyltransferases, i.e. with both conserved residues of the DXH motif in direct contact with the cation, a feature which probably is of relevance for the catalytic action of these enzymes (Patenaude *et al.*, 2002). Compared to other glycosyltransferases, the presence of a histidine interacting with the Mn^{2+} ion has also been found in the LgtC binding site (H244) (Persson *et al.*, 2001) and in rabbit glycogenin (H211) (Gibbons *et al.*, 2002). On the other hand, replacement of H211 by a D in the ppGaNTase-T1 (Hagen *et al.*, 1999) abolished the enzyme activity, indicating that D could not functionally substitute for a H. However, in the model presented here, this substitution is possible without impairment of the donor substrate recognition (data not shown). In this context it should be emphasized that our model focuses only on a segment of the UDP-GalNAc binding domain and the H211D mutation could be of importance in interactions not taken into account by the present study.

Secondly, in our model the uracil base is stabilized in the binding pocket through a stacking interaction with W129. The equivalent is found not only in the templates GTA (Y126) and $\alpha 3$ GalT (Y139) but also in glycosyltransferases like LgtC with Y11 (Persson *et al.*, 2001), SpsA with Y11 (Charnock and Davies, 1999), the bovine $\beta 1,4$ galactosyltransferase 1 ($\beta 4$ Gal-T1) with F226 and R191 (Gastinel *et al.*, 1999), GlcAT-1 with Y84 (Pedersen *et al.*, 2002), glycogenin with Y14 (Gibbons *et al.*, 2002) and EXTL2 with Y74 (Pedersen *et al.*, 2003). W129 is not conserved in all ppGaNTases, but replaced by R (in ppGaNTases-T2 and -T4) or by L (in ppGaNTases-T8 and -T9), two amino acids which can be involved in N-H/p or C-H/p interaction, respectively, with an aromatic ring. Our mutation experiments show unambiguously that an R can replace W129 in the ppGaNTase-T1 without significant change in enzyme activity and even the W129A mutation conserves some activity. One can also predict that a leucine will make a hydrophobic interaction in this region as well, as it happens in glycogenin with a valine.

Thirdly, the N126 of the modelled ppGaNTase-T1 subdomain interacts via its NH and NH_2 groups through hydrogen bonds with the O2 of the uracil base and the O2 of the ribose, respectively. N126 is changed to D in the rat ppGaNTase-T5 isoform and otherwise conserved in all known mammalian members of the family. However, in the model an aspartic acid could interact with the uridine as well. On the other hand, it is most probable that a change from N to A does not allow any interaction with the ribose, which explains the low activity of the alanine mutant. In GTA (I123, F121, V212 and D213) (Patenaude *et al.*, 2002), $\alpha 3$ GalT (F134, V136 and D225) (Gastinel *et al.*, 2001), glycogenin (L8, N11 and D103) (Gibbons *et al.*, 2002), EXTL2 (Q72, N101, N130 and D152) (Pedersen *et al.*, 2003), GlcAT-1 (D113 and D195) (Pedersen *et al.*, 2002) and LgtC (A6, D8, N10 and I104) (Persson *et al.*, 2001), the same types of interaction have been observed in the crystals.

Fourthly, the model suggests the hydrogen bonding of the guanidinium group of R193 with the GalNAc of the donor

sugar. Again, the equivalent is seen in the templates used for the present modelling with R202 binding to the Gal moiety of UDP-Gal in the α 3GalT X-ray structure (Boix *et al.*, 2002) and most probably R188 binding to the GalNAc of UDP-GalNAc in the GTA binding site. This R188 is present in a disordered loop adjacent to the active site of the GTA but could be shown to form contacts with the donor substrate modelled into the binding site (Patenaude *et al.*, 2002). Interactions of a guanidinium group with the donor sugar were also found in LgtC where R86 binds to the Gal of UDP-2FGal (Persson *et al.*, 2001), in GlcAT-I where R161 forms a hydrogen bond with the GlcUA of the donor sugar (Pedersen *et al.*, 2002) and in EXTL2 where R135 is hydrogen bonded to the GalNAc of UDP-GalNAc (Pedersen *et al.*, 2003). It should be also mentioned that R193 is invariant in all the ppGaNTases cloned so far. Furthermore, a point mutation of R193 to W in a fruitfly ppGaNTase is lethal for the insect and, when expressed in COS cells, the resulting protein is inactive (Ten Hagen and Tran, 2002). Indeed in our model, the presence of a tryptophan instead of R193 would induce profound changes in the structure of the catalytic pocket. It is also noteworthy that R193 contributes notably to the folded back conformation of the UDP-sugar and may thus be important for the catalytic mechanism of the transferase.

Finally, the hydrophobic pocket formed by F124, L189, L194 and L208 in our model is found also in the GTA (F121 and V184) (Patenaude *et al.*, 2002), in the α 3GalT (F134 and I198) (Boix *et al.*, 2001), in LgtC (A6 and I104) (Persson *et al.*, 2001) and in glycogenin (L8 and V82) (Gibbons *et al.*, 2002). F124 and L189 are two strictly conserved amino acids in the ppGaNTases and mutagenesis confirmed their importance for enzyme activity. L208 is conserved in all ppGaNTases except in the ppGaNTase-T9 where it is conservatively replaced by a F. On the other hand, L194 is less well conserved since it is changed in V, I, M, N or S for six ppGaNTases of the 14 analysed. It is likely that in this region the contact with the substrate is looser.

Since we used GTA and α 3GalT as templates to model a subdomain of the ppGaNTase-T1, it may not be surprising, *a priori*, to find in the model the same type of interactions which are seen in the crystals between the enzymes and their substrates. However, these enzymes do not have the same function, even if they show common enzymatic features, and the primary sequence of the ppGaNTase-T1 is quite different from that of the GTA/GTB and of the α 3GalT. It is thus interesting to note that the global structure of the nucleotide binding site is similar to such an extent for the three enzymes. More restricted structural similarities could also be revealed between the model, the templates and three other retaining glycosyltransferases: LgtC, glycogenin and EXTL2. These observations strengthen our model and suggest that retaining glycosyltransferases may use very similar mechanisms.

Recently, derivatives of UDP-GalNAc O-methylated on C3, C4 or C6 of the GalNAc ring have been tested as substrates or inhibitors of the bovine ppGaNTase-T1 (Busca *et al.*, 2003). None of the UDP(3-OMe)GalNAc, UDP(4-OMe)GalNAc or UDP(6-OMe)GalNAc structures are substrates and they are only weak inhibitors with K_i s similar to those for UDP or UDP-GlcNAc. Furthermore, synthesis of UDP-GalNAc biotinylated on the C6 of the GalNAc moiety gave rise to a compound which was not transferred by the ppGaNTase-T1 (Bulter *et al.*, 2001). These results indicate that the GalNAc portion

of UDP-GalNAc fits tightly into the donor binding site which cannot accommodate OMe or biotinyl groups. This is in agreement with the model presented here, with the GalNAc bound in the site by hydrogen bonds with R193 as well as by hydrophobic interactions with L residues.

Although the model presented here has been tested by mutagenesis experiments, the final confirmation has to await X-ray crystallography of one of the members of the ppGaNTases family. The model, however, may contribute to our understanding of nucleotide sugar binding in glycosyltransferases and help with the design and development of new substrate analogues or inhibitors of the ppGaNTases.

Acknowledgements

The expert technical assistance of Nicole Bureau is gratefully acknowledged. Thanks are also due to Dr Anne Imberty and Professor Christelle Breton (CERMAV-CNRS, Grenoble, France) for their help in the initial phase of this study. This work was supported by grants from the Ligue Nationale contre le Cancer, comités départementaux du Loiret et du Loir et Cher, from the Centre National de la Recherche Scientifique: Protéomique et Génie des Protéines and by the Groupement de Recherche: Génomique et Génie des Glycosyltransférases. Stéphanie Duclos received a fellowship from the Région Centre Council and Pedro Da Silva received a fellowship from the MENESR.

References

- Boix,E., Swaminathan,G.J., Zhang,Y., Natesh,R., Brew,K. and Acharya,K.R. (2001) *J. Biol. Chem.*, **276**, 48608–48614.
- Boix,E., Zhang,Y., Swaminathan,G.J., Brew,K. and Acharya,K.R. (2002) *J. Biol. Chem.*, **277**, 28310–28318.
- Breton,C., Bettler,E., Joziassé,D.H., Geremia,R.A. and Imberty,A. (1998) *J. Biochem. (Tokyo)*, **123**, 1000–1009.
- Breton,C., Heissigerova,H., Jeanneau,C., Moravcova,J. and Imberty,A. (2002) *Biochem. Soc. Symp.*, **69**, 23–32.
- Bulter,T., Schumacher,T., Namdjou,D.J., Gutierrez Gallego,R., Clausen,H. and Elling,L. (2001) *ChemBiochem*, **2**, 884–894.
- Busca,P., Piller,V., Piller,F. and Martin,O.R. (2003) *Bioorg. Med. Chem. Lett.*, **13**, 1853–1856.
- Busch,C., Hofmann,F., Selzer,J., Munro,S., Jeckel,D. and Aktories,K. (1998) *J. Biol. Chem.*, **273**, 19566–19572.
- Charnock,S.J. and Davies,G.J. (1999) *Biochemistry*, **38**, 6380–6385.
- Chiu,C.P., Watts,A.G., Lairson,L.L., Gilbert,M., Lim,D., Wakarchuk,W.W., Withers,S.G. and Strynadka,N.C. (2004) *Nat. Struct. Mol. Biol.*, **11**, 163–170.
- de Vries,T., Knegt,R.M., Holmes,E.H. and Macher,B.A. (2001) *Glycobiology*, **11**, 119R–128R.
- Di Francesco,V., Munson,P.J. and Garnier,J. (1995) *28th Hawaii Int. Conf. Syst. Sci.*, **5**, 285–291.
- Fischer,D. (2000) *Pac. Symp. Biocomput.*, 119–130.
- Frishman,D. and Argos,P. (1997) *Proteins*, **27**, 329–335.
- Gaboriaud,C., Bissery,V., Benchetrit,T. and Mornon,J.P. (1987) *FEBS Lett.*, **224**, 149–155.
- Gastinel,L.N., Cambillau,C. and Bourne,Y. (1999) *EMBO J.*, **18**, 3546–3557.
- Gastinel,L.N., Bignon,C., Misra,A.K., Hinds-gaul,O., Shaper,J.H. and Joziassé,D.H. (2001) *EMBO J.*, **20**, 638–649.
- Gibbons,B.J., Roach,P.J. and Hurley,T.D. (2002) *J. Mol. Biol.*, **319**, 463–477.
- Gibson,R.P., Turkenburg,J.P., Charnock,S.J., Lloyd,R. and Davies,G.J. (2002) *Chem. Biol.*, **9**, 1337–1346.
- Gibson,R.P., Tarling,C.A., Roberts,S., Withers,S.G. and Davies,G.J. (2004) *J. Biol. Chem.*, **279**, 1950–1955.
- Gough,J., Karplus,K., Hughey,R. and Chothia,C. (2001) *J. Mol. Biol.*, **313**, 903–919.
- Gulberti,S., Fournel-Gigleux,S., Mulliert,G., Aubry,A., Netter,P., Magdalou,J. and Ouazzine,M. (2003) *J. Biol. Chem.*, **278**, 32219–32226.
- Ha,S., Walker,D., Shi,Y. and Walker,S. (2000) *Protein Sci.*, **9**, 1045–1052.
- Hagen,F.K. and Nehrkke,K. (1998) *J. Biol. Chem.*, **273**, 8268–8277.
- Hagen,F.K., Hazes,B., Raffo,R., deSa,D. and Tabak,L.A. (1999) *J. Biol. Chem.*, **274**, 6797–6803.
- Hassan,H., Benett,E.P., Mandel,U., Hollingworth,M.A. and Clausen,H. (2000) In: Ernst,B., Hart,G., and Sinaý,P. (eds), *Carbohydrates in Chemistry and Biology*. Wiley-VCH, pp. 273–292.
- Heissigerova,H., Breton,C., Moravcova,J. and Imberty,A. (2003) *Glycobiology*, **13**, 377–386.
- Hutchinson,E.G. and Thornton,J.M. (1996) *Protein Sci.*, **5**, 212–220.

- Imberty,A., Piller,V., Piller,F. and Breton,C. (1997) *Protein Eng.*, **10**, 1353–1356.
- Imberty,A., Monier,C., Bettler,E., Morera,S., Freemont,P., Sippl,M., Flockner,H., Ruger,W. and Breton,C. (1999) *Glycobiology*, **9**, 713–722.
- Ito,H., Fukuda,Y., Murata,K. and Kimura,A. (1983) *J. Bacteriol.*, **153**, 163–168.
- Jones,D.T. (1999) *J. Mol. Biol.*, **287**, 797–815.
- Juan,D., Grana,O., Pazos,F., Fariselli,P., Casadio,R. and Valencia,A. (2003) *Proteins*, **50**, 600–608.
- Kapitonov,D. and Yu,R.K. (1999) *Glycobiology*, **9**, 961–978.
- Kikuchi,N., Kwon,Y.D., Gotoh,M. and Narimatsu,H. (2003) *Biochem. Biophys. Res. Commun.*, **310**, 574–579.
- Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
- Laskowski,R.A., MacArthur,M.W., Moss D.S.A. and Thornton,J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.
- Lazarus,B.D., Milland,J., Ramsland,P.A., Mouhtouris,E. and Sandrin,M.S. (2002) *Glycobiology*, **12**, 793–802.
- Lemer,C.M., Rooman,M.J. and Wodak,S.J. (1995) *Proteins*, **23**, 337–355.
- Liu,J. and Mushegian,A. (2003) *Protein Sci.*, **12**, 1418–1431.
- Lobsanov,Y.D., Romero,P.A., Sleno,B., Yu,B., Yip,P., Herscovics,A. and Howell,P.L. (2004) *J. Biol. Chem.*, **279**, 17921–17931.
- Luthy,R., Bowie,J.U. and Eisenberg,D. (1992) *Nature*, **356**, 83–85.
- Ly,H.D., Loughheed,B., Wakarchuk,W.W. and Withers,S.G. (2002) *Biochemistry*, **41**, 5075–5085.
- McGuffin,L.J., Bryson,K. and Jones,D.T. (2000) *Bioinformatics*, **16**, 404–405.
- Mulichak,A.M., Losey,H.C., Walsh,C.T. and Garavito,R.M. (2001) *Structure (Camb.)*, **9**, 547–557.
- Mulichak,A.M., Losey,H.C., Lu,W., Wawrzak,Z., Walsh,C.T. and Garavito,R.M. (2003) *Proc. Natl Acad. Sci. USA*, **100**, 9238–9243.
- Mulichak,A.M., Lu,W., Losey,H.C., Walsh,C.T. and Garavito,R.M. (2004) *Biochemistry*, **43**, 5170–5180.
- Negishi,M., Dong,J., Darden,T.A., Pedersen,L.G. and Pedersen,L.C. (2003) *Biochem. Biophys. Res. Commun.*, **303**, 393–398.
- Nicholas,K.B., Nicholas,H.B. and Deerfield,D.W. (1997) *EMBNet News*, **4**, 14–17.
- Ouzzine,M., Gulberti,S., Levoine,N., Netter,P., Magdalou,J. and Fournel-Gigleux,S. (2002) *J. Biol. Chem.*, **277**, 25439–25445.
- Patenaude,S.I., Seto,N.O., Borisova,S.N., Szpacenko,A., Marcus,S.L., Palcic,M.M. and Evans,S.V. (2002) *Nat. Struct. Biol.*, **9**, 685–690.
- Pedersen,L.C., Tsuchida,K., Kitagawa,H., Sugahara,K., Darden,T.A. and Negishi,M. (2000) *J. Biol. Chem.*, **275**, 34580–34585.
- Pedersen,L.C., Darden,T.A. and Negishi,M. (2002) *J. Biol. Chem.*, **277**, 21869–21873.
- Pedersen,L.C., Dong,J., Taniguchi,F., Kitagawa,H., Krahn,J.M., Pedersen,L.G., Sugahara,K. and Negishi,M. (2003) *J. Biol. Chem.*, **278**, 14420–14428.
- Persson,K., Ly,H.D., Dieckelmann,M., Wakarchuk,W.W., Withers,S.G. and Strynadka,N.C. (2001) *Nat. Struct. Biol.*, **8**, 166–175.
- Rao,M. and Tvaroska,I. (2001) *Proteins*, **44**, 428–434.
- Rossmann,M.G., Moras,D. and Olsen,K.W. (1974) *Nature*, **250**, 194–199.
- Rost,B. and Sander,C. (1993) *Protein Eng.*, **6**, 831–836.
- Rost,B. and Sander,C. (1994) *Proteins*, **20**, 216–226.
- Sali,A. and Blundell,T.L. (1993) *J. Mol. Biol.*, **234**, 779–815.
- Schonbrun,J., Wedemeyer,W.J. and Baker,D. (2002) *Curr. Opin. Struct. Biol.*, **12**, 348–354.
- Schwientek,T. et al. (2002) *J. Biol. Chem.*, **277**, 22623–22638.
- Sippl,M.J. (1993) *Proteins*, **17**, 355–362.
- Tarbouriech,N., Charnock,S.J. and Davies,G.J. (2001) *J. Mol. Biol.*, **314**, 655–661.
- Ten Hagen,K.G. and Tran,D.T. (2002) *J. Biol. Chem.*, **277**, 22616–22622.
- Ten Hagen,K.G., Fritz,T.A. and Tabak,L.A. (2003) *Glycobiology*, **13**, 1–16.
- Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
- Tsai,Y.W., Chia,J.S., Shiau,Y.Y., Chou,H.C., Liaw,Y.C. and Lou,K.L. (2000) *FEMS Microbiol. Lett.*, **188**, 75–79.
- Unligil,U.M. and Rini,J.M. (2000) *Curr. Opin. Struct. Biol.*, **10**, 510–517.
- Unligil,U.M., Zhou,S., Yuwaraj,S., Sarkar,M., Schachter,H. and Rini,J.M. (2000) *EMBO J.*, **19**, 5269–5280.
- Vrielink,A., Ruger,W., Driessen,H.P. and Freemont,P.S. (1994) *EMBO J.*, **13**, 3413–3422.
- Zhang,Z., Kochhar,S. and Grigorov,M. (2003) *Protein Sci.*, **12**, 2291–2302.

Received July 13, 2004; revised September 6, 2004; accepted September 10, 2004

Edited by Stephen Withers