

**INTERACTION LECTINES-OSES
ET DERIVES D'OSES SUBSTITUES :**
**étude par inhibition quantitative
de l'hémagglutination et
par modélisation et ancrage moléculaire**

T H E S E

POUR LE

**DIPLOME D'ETAT
DE DOCTEUR EN PHARMACIE**

présentée et soutenue publiquement le lundi 11 décembre 1995

par

Franck DUMEIRAIN

né le 21 juillet 1968 à Limoges (Haute-Vienne)

EXAMINATEURS de la THESE

Monsieur le Professeur Michel BERNARD PRESIDENT
Monsieur le Professeur Christian MËSCH JUGE
Monsieur Jean-Luc DUROUX, Maître de Conférences JUGE
Monsieur le Docteur Jean DEBORD JUGE

UNIVERSITÉ DE LIMOGES

Faculté de Pharmacie

DOYEN DE LA FACULTÉ : Monsieur le Professeur RABY Claude.

ASSESSEURS : Monsieur le Professeur GHESTEM Axel
Monsieur DREYFUSS Gilles - Maître de Conférences.

PROFESSEURS :

BENEYTOU Jean-Louis	BIOCHIMIE
BERNARD Michel	PHYSIQUE - BIOPHYSIQUE
BOSGIRAUD Claudine	BACTÉRIOLOGIE - VIROLOGIE - PARASITOLOGIE
BROSSARD Claude	PHARMACOTECHNIE
CARDOT Philippe	CHIMIE ANALYTIQUE
BUXERAUD Jacques	CHIMIE ORGANIQUE - CHIMIE THÉRAPEUTIQUE
CHULIA Albert	PHARMACOGNOSIE
CHULIA Dominique	PHARMACOTECHNIE
DELAGE Christiane	CHIMIE GÉNÉRALE ET MINÉRALE
GHESTEM Axel	BOTANIQUE ET CRYPTOLOGIE
HABRIOUX Gérard	BIOCHIMIE FONDAMENTALE
LACHATRE Gérard	TOXICOLOGIE
MESCH Christian	HYGIÈNE
LOUDART Nicole	PHARMACODYNAMIE
RABY Claude	PHARMACIE CHIMIQUE ET CHIMIE ORGANIQUE.

SECRÉTAIRE GÉNÉRAL DE LA FACULTÉ - CHEF DES SERVICES ADMINISTRATIFS :

POMMARET Maryse.

A notre Président de Jury,
Monsieur le Professeur Michel BERNARD,
Professeur des Universités en Physique et Biophysique.

Nous sommes très honorés que vous ayez accepté de présider ce Jury de Thèse.
La haute qualité de vos enseignements nous a permis d'acquérir d'excellentes bases
en Biophysique, indispensables pour appréhender une discipline comme la
Modélisation Moléculaire.
Nous vous exprimons notre respectueuse reconnaissance.

A notre Directeur de Thèse,
Monsieur le Professeur Pierre ROUGÉ,
Professeur à la Faculté des Sciences Pharmaceutiques de Toulouse.

Nous vous remercions de nous avoir permis d'intégrer votre équipe de recherche.
Vous avez su nous guider efficacement dans ce travail, tout au long duquel nous
avons pu apprécier à la fois vos hautes compétences, votre disponibilité et votre
gentillesse.
Soyez assurés de notre respectueuse considération.

A nos Membres du Jury,

A Monsieur le Professeur Christian MÆSCH,
Professeur des Universités en Hygiène, Environnement et Santé Publique.

Vous avez su nous faire pleinement partager, lors du premier certificat de Maîtrise de Sciences Biologiques et Médicales, votre passion pour la recherche dans le domaine de la Chimie et éveiller en nous une vocation.

Il était donc tout naturel de vous solliciter pour juger ce dernier acte du cursus des études pharmaceutiques.

Qu'il nous soit permis de vous témoigner ici notre gratitude.

A Monsieur Jean-Luc DUROUX,
Maître de Conférences.

Nous avons apprécié l'entière disponibilité dont vous avez toujours fait preuve au cours de nos études.

Nous vous adressons nos sincères remerciements.

A Monsieur le Docteur Jean DEBORD,
Assistant à la Faculté de Médecine de Limoges.

Nous sommes honorés de la présence d'un chercheur de votre niveau dans ce Jury.
Veuillez recevoir nos plus vifs remerciements.

A Mademoiselle Annick BARRE, Assistant à la Faculté des Sciences Pharmaceutiques de Toulouse, ainsi qu'à toute l'équipe du Département de Biologie Structurale et Ingénierie des Protéines.

Nous vous adressons tous nos remerciements.

A Monsieur le Professeur Gérard GRASSY,
Professeur à la Faculté de Pharmacie de Montpellier.

Nous vous remercions de vos précieux conseils pour l'élaboration du chapitre théorique consacré à la Modélisation Moléculaire, discipline dans laquelle vous êtes à tous notre Maître.

A Monsieur le Professeur Michel PAYARD,
Professeur à la Faculté des Sciences Pharmaceutiques de Toulouse.

Nous vous remercions de votre apport de compétences dans le domaine chimique et de votre soutien logistique.

A Monsieur Jean-Luc STIGLIANI,
Maître de Conférences à la Faculté des Sciences Pharmaceutiques de Toulouse.

Nous vous remercions de nous avoir guidé, avec beaucoup de sympathie, dans l'application des méthodes de la mécanique quantique à la modélisation moléculaire.

A la mémoire de mon Grand-père.

A ma Grand-mère.

A ma Mère.

A ma Marraine, à mon Parrain.

A toute ma Famille.

A notre chère Corpo.

A tous mes Amis.

Nos travaux de recherche ont été réalisés au sein du Département IV- Biologie Structurale et Ingénierie des Protéines, dirigé par Monsieur le Professeur ROUGÉ, du Laboratoire de Pharmacologie et Toxicologie Fondamentales (U.P.R. C.N.R.S. 8221), à la Faculté des Sciences Pharmaceutiques de Toulouse - Université Paul Sabatier (Toulouse III).

SOMMAIRE.

INTRODUCTION	2
PARTIE THÉORIQUE.....	4
I- Les acteurs de la symbiose Légumineuse-<i>Rhizobium</i>	5
1- Les facteurs Nod	5
2- Les lectines	11
3- La symbiose Légumineuse- <i>Rhizobium</i>	19
4- Les gènes de nodulation ou gènes <i>nod</i>	22
II- La modélisation moléculaire et ses méthodes	29
1- Bases de la modélisation moléculaire	29
2- Propriétés moléculaires.....	31
3- Représentation de l'énergie interne moléculaire	34
4- Minimisation de l'énergie interne moléculaire	42
5- Méthodes de la dynamique moléculaire	50
6- Principaux logiciels de modélisation moléculaire : caractéristiques et applications	53
PARTIE EXPÉRIMENTALE.....	54
1- Oses et dérivés d'oses substitués utilisés	55
2- Isolement des lectines et des isolectines	55
3- Techniques d'hémagglutination	58
4- Modélisation moléculaire	59
RÉSULTATS.....	65
1- Affinité des oses et dérivés d'oses substitués pour les lectines.....	66
2- Modélisation moléculaire	68
DISCUSSION ET CONCLUSION	81
BIBLIOGRAPHIE.....	83
TABLE DES MATIÈRES.....	93

INTRODUCTION.

Plusieurs observations récentes montrent que la spécificité de reconnaissance de glycoconjugués (oligosaccharides, glycolipides, glycoprotéines) à rôles biologiques dépend en particulier de la sulfatation de certains de leurs oses ou dérivés d'oses constitutifs.

Ainsi, chez les animaux, la reconnaissance par la L-sélectine LECAM-1 du Rat de certains de ses ligands (sulfatides, céramides) est conditionnée par la présence d'un résidu terminal non réducteur de galactose sulfaté en C3 : $-3\text{Gal}\beta 1\text{-SO}_3^-$ (1). De même, la reconnaissance par la L-sélectine de Souris d'un de ses deux ligands naturels, GlyCAM-1, qui est une glycoprotéine d'environ 50 kDa, nécessite la sulfatation de cette glycoprotéine (2). Ces phénomènes de reconnaissance ont un intérêt particulier car ils interviennent dans l'hébergement (ou « homing ») et la maturation des lymphocytes T dans le thymus ou dans les ganglions lymphatiques.

L'existence, dans des glycoprotéines d'origines variées, d'oses ou de dérivés d'oses sulfatés a été rapportée à différentes reprises, notamment par Lis et Sharon (3), qui soulignent l'existence de galactosamine-3sulfate ($\text{galactosamine-3SO}_3^-$) dans la thyroglobuline et dans les mucines, de *N*-acétylgalactosamine-4 SO_3^- dans des hormones pituitaires, de *N*-acétylglucosamine 6 SO_3^- dans la thyroglobuline, de mannose-4 SO_3^- et de mannose-6 SO_3^- dans l'ovalbumine.

C'est cependant chez les végétaux, avec la découverte récente de signaux, les facteurs de nodulation ou facteurs Nod, d'origine bactérienne (*Rhizobium*), qui sont des glycolipides responsables de l'établissement d'une symbiose spécifique et fonctionnelle avec la plante-hôte (Légumineuse), que l'importance de la sulfatation dans les phénomènes de reconnaissance a été clairement démontrée et est activement étudiée.

Nous avons ainsi étudié l'influence de la substitution, essentiellement par des groupes sulfate ou phosphate, des hydroxyles portés en différentes positions (C1, C2, C3 et C6) du cycle pyranique de divers oses (mannose et glucose) et dérivés d'oses (mannosamine, glucosamine et *N*-acétylglucosamine), dont certains entrent dans la partie oligosaccharidique de facteurs Nod, sur la reconnaissance de lectines de Légumineuses (LoLI et LoLII de *Lathyrus ochrus*, PsA du Pois, LcA de la Lentille et la ConA de *Canavalia ensiformis*) glucose/mannose-spécifiques.

Cette étude a été réalisée par deux méthodes complémentaires : d'une part, une expérimentation classique à l'aide de la technique quantitative d'inhibition de l'hémagglutination qui a servi à mesurer les affinités des oses et dérivés d'oses substitués pour ces lectines, et, d'autre part, une approche de modélisation et d'ancrage moléculaire (ou « docking ») qui a permis d'expliquer, par le calcul et la visualisation de différents types d'interactions (Van der Waals, liaisons hydrogène, et électrostatique) entre l'ose ou le dérivé d'ose et le site de la lectine (de structures tridimensionnelles connues ou modélisées) et par l'observation d'encombrements stériques, la spécificité d'action de certains dérivés substitués.

PARTIE THÉORIQUE

I- Les acteurs de la symbiose Légumineuse-*Rhizobium*.

II- La modélisation moléculaire et ses méthodes.

I- Les acteurs de la symbiose Légumineuse-*Rhizobium*.

1- Les facteurs Nod.

Initialement découverts chez *Rhizobium meliloti* (4) qui colonise spécifiquement la Luzerne (*Medicago sativa*), les facteurs Nod sont des lipo-oligosaccharides, à caractère fortement amphiphile, qui correspondent à des tétra- (chitotétraoses) ou pentasaccharides (chitopentaoses) de *N*-acétylglucosamine liées en β 1-4, dont le résidu de glucosamine non réducteur se lie par son groupe amine porté en C2 du cycle pyranique à un acide gras mono- ou polyinsaturé.

Les facteurs Nod de *Rh. meliloti* (5, 6, 7) (**Figure 1A**) sont constitués d'un enchaînement de quatre (NodRm-IV) ou de cinq (NodRm-V) unités de glucosamine liées en β 1-4. Les trois ou quatre unités de glucosamine en partant de l'extrémité réductrice sont acétylées (*N*-acétylglucosamine). L'unité de glucosamine située à l'extrémité non réductrice est, d'une part, liée à un acide gras insaturé (C16:1 ou acide palmitoléique, C16:2 ou acide hexadécadiénoïque, ou C16:3 ou acide hexadécatriénoïque) au niveau de son groupe amine et, d'autre part, peut être acétylée (Ac) en C6. De même, l'unité de *N*-acétylglucosamine de l'extrémité réductrice peut être sulfatée (S) en C6.

Selon la nomenclature proposée par Roche *et al.* (1991) (6), les structures suivantes ont jusqu'à présent été caractérisées :

Structures	Auteurs
NodRm-IV(S)(C16:2)	Lerouge <i>et al.</i> (1990) (4) Roche <i>et al.</i> (1991) (5) Schultze <i>et al.</i> (1992) (7)
NodRm-IV(Ac,S)(C16:2)	Roche <i>et al.</i> (1991) (5)
NodRm-V(S)(C16:2)	Roche <i>et al.</i> (1991) (6) Schultze <i>et al.</i> (1992) (7)
NodRm-V(Ac,S)(C16:2)	Roche <i>et al.</i> (1991) (6) Schultze <i>et al.</i> (1992) (7).

Différents tests, notamment le test Had (abréviation de « Hair deformation ») (9) qui consiste à mesurer l'influence de l'inoculation des racines d'une Légumineuse par une souche de *Rhizobium* homologue (ou, comme dans le cas précis, par les facteurs Nod correspondants) sur la déformation de ses poils absorbants, réalisés sur la Luzerne, ont servi à déterminer l'activité des facteurs NodRm.

Les facteurs sulfatés et en particulier NodRm-IV(S)(C16:2) se sont avérés les plus actifs, à des concentrations de 10^{-10} à 10^{-12} M (4, 5). A des concentrations supérieures (10^{-7} M), ils induisent dans le cortex racinaire la formation de véritables nodosités en l'absence de bactéries (10). Les autres facteurs NodRm sulfatés ont des activités beaucoup plus faibles (100 fois moins actifs) et sont moins spécifiques de la Luzerne.

D'autres facteurs Nod ont été isolés d'autres souches de *Rhizobium* et caractérisés :

- chez *Rh. leguminosarum* bv. *viciae*, dont la plante-hôte est une Vesce (Pois, Lentille, Fève...), les facteurs Nod ont la même structure générale (chitotétraoses ou chitopentaoses) (**Figure 1B**), mais différent des facteurs de *Rh. meliloti* par quelques particularités (11) : 1) l'unité de *N*-acétylglucosamine située à l'extrémité réductrice n'est pas sulfatée en C6 ; 2) l'unité de glucosamine de l'extrémité non réductrice est généralement *O*-acétylée en C6 ; 3) les acides gras portés par ce même résidu sont en C18 et monoinsaturé $\Delta 11$ (C18:1 ou acide vaccénique) ou tétra-insaturé avec trois doubles liaisons conjuguées $\Delta 2,4,6,11$ (C18:4 ou acide octadécatétraène-2,4,6,11 oïque).

Figure 1.**A-** Structure des facteurs Nod isolés de *Rhizobium meliloti* (8).

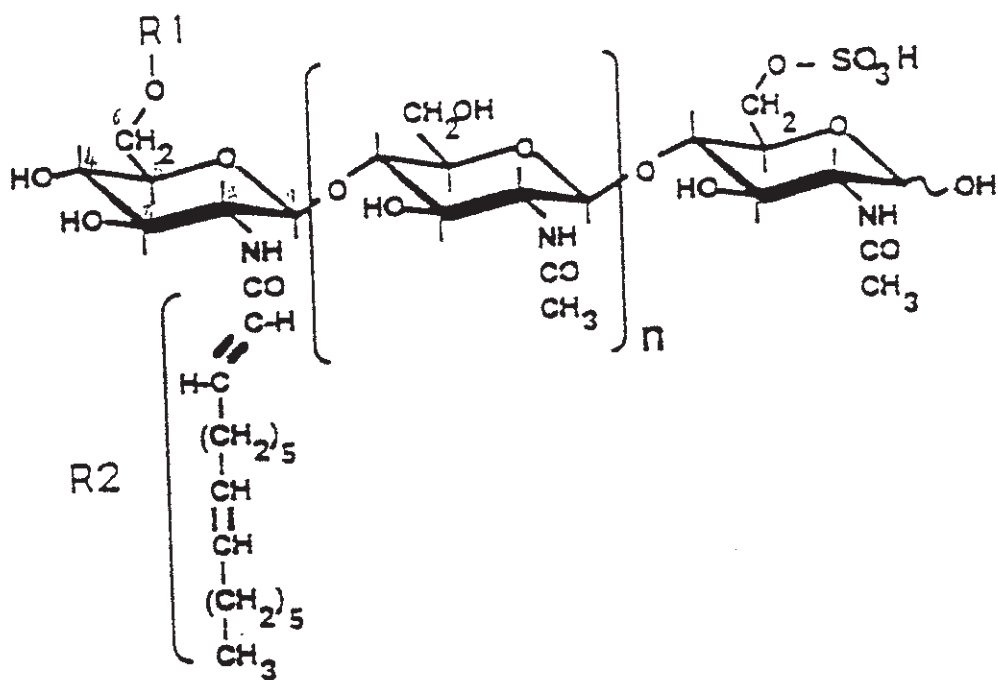
Ils correspondent à des chitotétraoses ($n = 2$) ou à des chitopentaoses ($n = 3$) dont le résidu de glucosamine situé à l'extrémité non réductrice est, d'une part, lié à un acide gras insaturé ($R_2 = C_{16}:1, C_{16}:2$ comme dans le cas présent ou $C_{16}:3$), et peut, d'autre part, être acétylé en C6 ($R_1 = COCH_3$ ou H).

On note la présence d'un groupe O6-sulfaté sur l'unité de N-acétylglucosamine de l'extrémité réductrice.

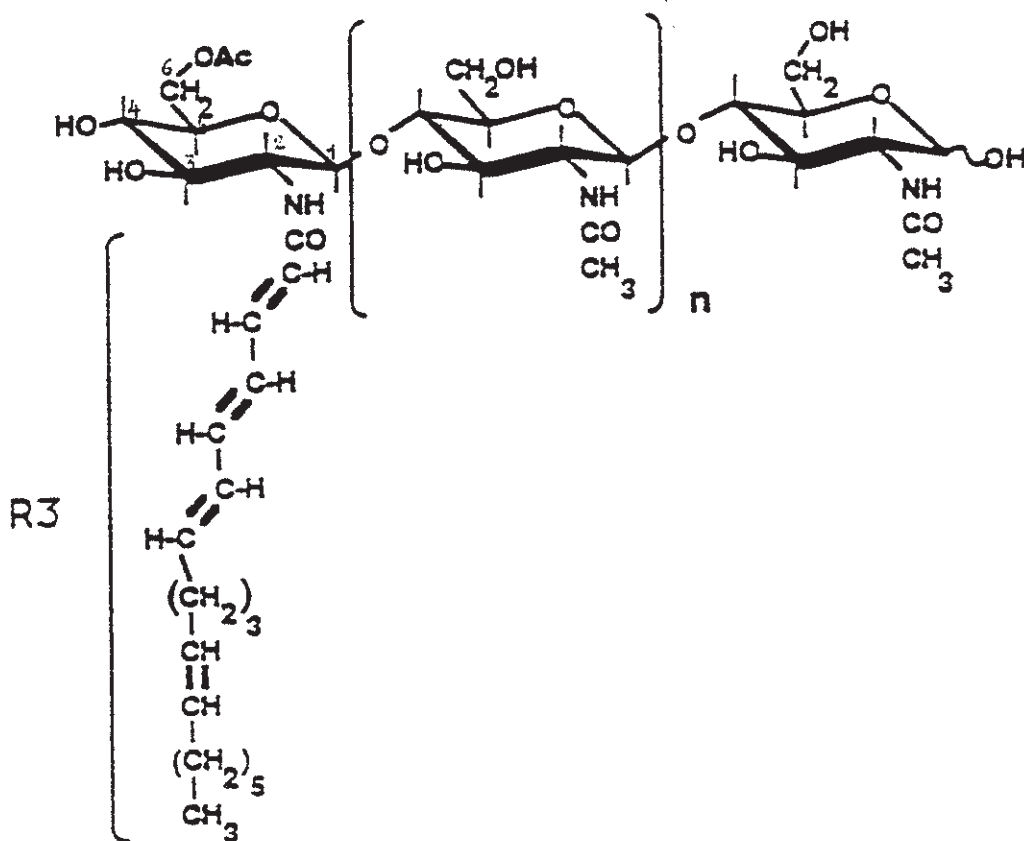
B- Structure des facteurs Nod isolés de *Rhizobium leguminosarum* bv. *viciae* (8).

Il s'agit de chitotétraoses ($n = 2$) ou de chitopentaoses ($n = 3$) dont le résidu de glucosamine O6-acétylé situé à l'extrémité non réductrice porte un acide gras insaturé ($R_3 = C_{18}:1$ ou $C_{18}:4$ comme dans le cas présent).

A



B



Selon la nomenclature de Roche *et al.* (6), les structures décrites correspondent à :

- NodRlv-IV(C18:1) ;
- NodRlv-IV(C18:4) ;
- NodRlv-V(C18:1) ;
- NodRlv-V(C18:4).

- chez *Rhizobium sp.* souche NGR234, dont la spécificité d'hôte est extrêmement lâche (plus de 70 genres de Légumineuses peuvent être nodulées par cette souche) et s'étend même à des familles autres que les Légumineuses, les facteurs Nod correspondent à un ensemble de chitopentaoses diversement substitués (12) :

- 1) par un groupe carbamyle en C3 ou en C4 de l'unité de glucosamine située à l'extrémité non réductrice ;
- 2) par un groupe méthyle sur le groupe amine de ce même résidu ;
- 3) par du 2-O-méthyl-fucose en C6 de l'unité de N-acétylglucosamine de l'extrémité réductrice, pouvant lui-même être substitué par un groupe sulfate en C3 ou par un groupe acétate en C4 ;
- 4) par un acide gras insaturé Δ^{11} (C18:1) ou par un acide gras saturé (C16 ou acide palmitique) sur le groupe amine de l'unité de glucosamine de l'extrémité non réductrice.

D'où les structures suivantes :

- NodNGR-V(Carb1,NMe,C18:1)(MeFuc) ;
- NodNGR-V(Carb1,NMe,C18:1)(MeFuc,S) ;
- NodNGR-V(Carb2,NMe,C18:1)(MeFuc,Ac) ;
- NodNGR-V(Carb2,NMe,C18:1)(MeFuc) ;
- NodNGR-V(Carb2,NMe,C18:1)(MeFuc,S) ;
- NodNGR-V(Carb2,NMe,C18:1)(MeFuc,Ac) ;

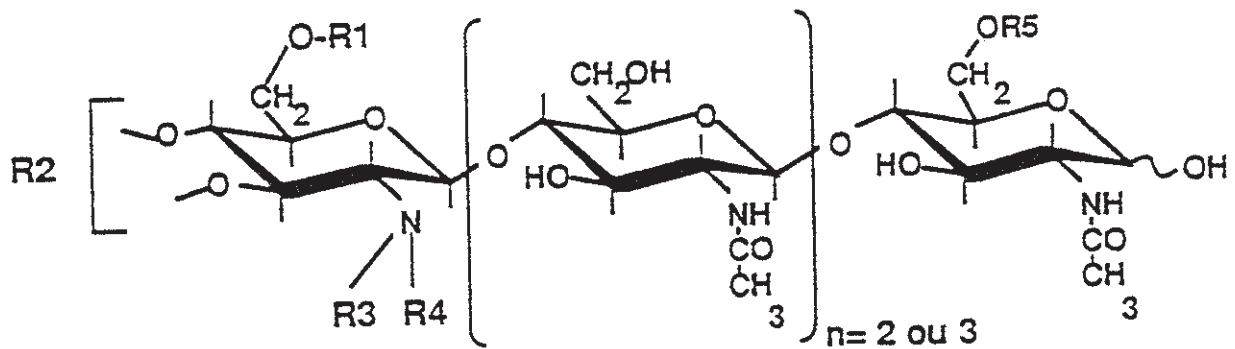
selon la nomenclature de Roche *et al.* (6).

- d'autres facteurs Nod ont été isolés de *Bradyrhizobium japonicum* (13), de *Rhizobium fredii* (14), d'*Azorhizobium caulinodans* (15) et de *Rh. tropici* (16).

Les particularités structurales de tous ces facteurs Nod sont indiquées sur la **Figure 2**.

Figure 2.

Structure des facteurs Nod de différentes espèces des genres *Rhizobium* et apparentés (*Azorhizobium* et *Bradyrhizobium*) (8).



Bactérie	R1	R2	R3	R4	R5
<i>Rh. meliloti</i>	Ac ou H	H	C16:2	H	SO ₃ H
<i>Rh. leguminosarum</i>	Ac	H	C18:4 ou C18:1	H	H
<i>Rh. NGR234</i>	H	NH ₂ CO	C18:1	Me	2-O-MeFuc
<i>B. japonicum</i>	H	H	C18:1	H	2-O-MeFuc
<i>A. caulonidans</i>	NH ₂ CO	H	C18:1 ou C18	Me	D-Ara ou H

A quelques exceptions près, qui demandent peut-être à être confirmées, la plupart des facteurs Nod actuellement caractérisés sont sulfatés et la présence d'un groupe sulfate paraît être un élément important assurant leur reconnaissance par la plante-hôte et leur spécificité d'action (établissement d'une symbiose spécifique et fonctionnelle entre l'hôte et la bactérie).

2- Les lectines.

Parmi les récepteurs végétaux pouvant reconnaître spécifiquement les facteurs Nod, les lectines, protéines capables de reconnaître et de fixer sélectivement les glucides simples et complexes (17), apparaissent jusqu'à présent être les plus probables pour jouer ce rôle. Plusieurs observations, concernant la spécificité de la symbiose et la localisation des lectines dans la plante, vont en ce sens :

1) Le Pois (*Pisum sativum*) et le Trèfle blanc (*Trifolium repens*) sont deux Légumineuses appartenant à des tribus différentes (*Vicieæ* pour le Pois et *Trifolieæ* pour le Trèfle blanc), qui sont infectées spécifiquement par *Rhizobium leguminosarum* bv. *vicieæ* pour la première et par *Rh. leguminosarum* bv. *trifolii* pour la seconde (18). Cette spécificité d'inoculation est très stricte et, dans des conditions naturelles, n'est jamais transgressée.

Diaz *et al.* (19) ont montré que l'introduction, à l'aide d'*Agrobacterium tumefaciens* utilisé comme vecteur, du gène (*Psl*) codant pour la lectine de Pois (PsA) chez le Trèfle blanc, entraînait une rupture de cette spécificité d'hôte : le Trèfle blanc pouvait, tout comme le Pois, être infecté par *Rh. leguminosarum* bv. *vicieæ*. Cette expérience démontrait sans ambiguïté que les lectines intervenaient dans la spécificité d'hôte lors de l'interaction Légumineuse-*Rhizobium*.

2) Chez de nombreuses Légumineuses, une localisation racinaire des lectines a été rapportée (20) et, plus récemment, Diaz *et al.* (21) ont mis en évidence, par des réactions immunocytochimiques, que chez le Pois, la lectine se trouve à l'extrémité

des poils absorbants de la zone pilifère des jeunes racines, par lesquels s'effectue l'infection par les bactéries libres (*Rh. leguminosarum* bv. *viciae*) du sol.

Les structures tridimensionnelles de plusieurs lectines et isolectines de Légumineuses ont été déterminées par radiocristallographie (**Tableau 1**). Leur étude montre une forte identité structurale basée sur la présence constante d'éléments caractéristiques.

Les lectines/isolectines sont constituées d'un monomère de base d'environ 26 kDa, formé d'une seule (lectines dites "à une chaîne") ou de deux (lectines dites "à deux chaînes") sous-unités appelées sous-unité lourde ou β , de 20 kDa, et sous-unité légère ou α , de 6 kDa. Ces monomères s'associent par des liaisons non covalentes en dimères de 52 kDa (PsA ou *Pisum sativum* Agglutinine du Pois, LcA ou *Lens culinaris* Agglutinine de la Lentille, Favin de la Fève, LoLI de *Lathyrus ochrus* (Gesse)) (**Figure 3**) ou en tétramères d'environ 100 kDa (Con A ou Concanavaleine A de *Canavalia ensiformis*, PNA ou Peanut Agglutinine de l'Arachide).

Chaque monomère possède un site de fixation des glucides ou « monosaccharide-binding site » : les lectines dimériques sont bivalentes (deux sites de fixation), tandis que les lectines tétramériques sont tétravalentes (quatre sites de fixation).

Les lectines à deux chaînes sont bivalentes et ne s'observent que dans la tribu des *Viciae* (Pois, Lentille, Fève, Gesses). Elles reconnaissent spécifiquement le glucose, le mannose, et leurs dérivés, et sont dites glucose/mannose-spécifiques (Glc/Man-spécifiques).

Les lectines à une chaîne sont tétravalentes et s'observent dans toutes les autres tribus. Elles reconnaissent le plus souvent le galactose, la N-acétylgalactosamine, et leurs dérivés (lectines Gal/GalNAc-spécifiques), ou le fucose et ses dérivés (lectines Fuc-spécifiques).

Il existe toutefois des exceptions : la lectine de Sainfoin (*Onobrychis viciifolia*) et la Con A sont toutes les deux des lectines à une chaîne Glc/Man-spécifiques. De même, la lectine de *Lathyrus sphaericus* (24) et les isolectines de *L. nissolia* (25) sont des lectines à une chaîne Glc/Man-spécifiques, bien qu'appartenant à la tribu des *Viciae*.

Figure 3.

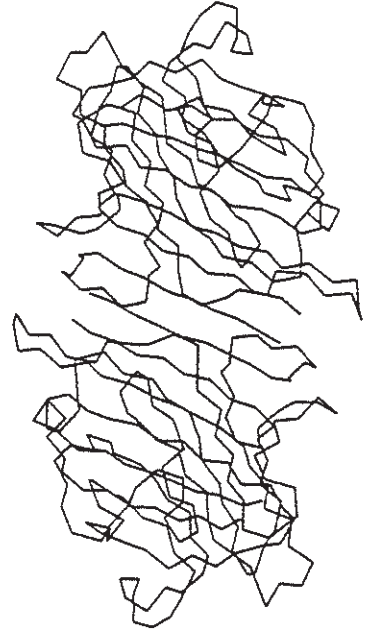
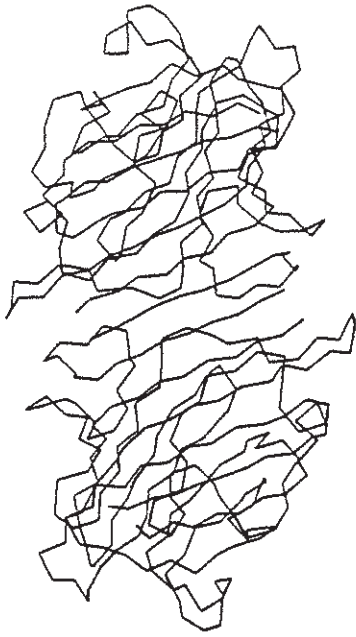
Structure tridimensionnelle de la lectine de Pois ou PsA (A) et de l'isolectine I de *Lathyrus ochrus* ou LoLI (B), modélisées à partir de leurs coordonnées cristallographiques (22, 23) par les logiciels Insight et Discover (Biosym Technologies, San Diego CA, USA) par Rougé *et al.* et visualisées à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

Ces deux lectines sont des dimères (52 kDa) résultant de l'association de deux monomères composés d'une chaîne lourde ou β (20 kDa) et d'une chaîne légère ou α (6 kDa).

Les éléments de structure secondaire de ces lectines correspondent exclusivement à des feuillets β à disposition antiparallèle reliés par des boucles. Cette disposition antiparallèle des feuillets est nettement apparente sur la face dorsale plane (« back face », modèles de gauche) qui est formée de six feuillets. Elle l'est par contre beaucoup moins sur la face frontale (« front face », modèles de droite) formée de sept feuillets à disposition enroulée.

Chaque monomère présente la forme d'un dôme aplati et possède un site de reconnaissance des monosaccharides situé à l'extrémité du dôme et constitué de sept résidus d'acides aminés portés par quatre boucles convergentes.

A



B

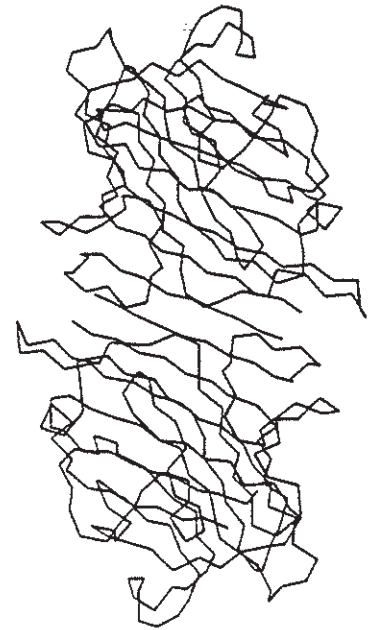
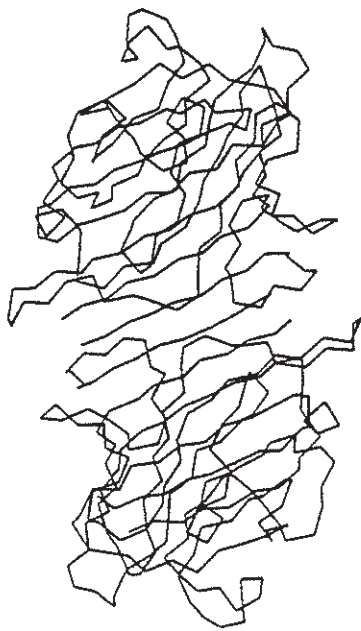


Tableau 1. Structures tridimensionnelles des lectines/isolectines de Légumineuses obtenues par radiocristallographie.

Lectines	Légumineuses	Spécificité	Résolution(Å)	Auteurs
Con A	<i>Canavalia ensiformis</i>	Glc/Man	1,75	Hardman <i>et al.</i> (1982) (26)
PsA	Pois (<i>Pisum sativum</i>)	Glc/Man	3,0	Einspahr <i>et al.</i> (1986) (23)
Favin	Fève	Glc/Man	2,8	Reeke & Becker (1986) (27)
LoLI	<i>Lathyrus ochrus</i>	Glc/Man	1,9	Bourne <i>et al.</i> (1990) (22)
GsIV	Gesses	Fuc	2,8	Delbaere <i>et al.</i> (1990) (28)
EcoL	<i>Erythrina corallo dendron</i>	Gal/GalNAc	2,0	Shaanan <i>et al.</i> (1991) (29)
LcA	Lentille (<i>Lens culinaris</i>)	Glc/Man	2,3	Loris <i>et al.</i> (1993) (30)
PNA	Arachide (<i>Arachis</i>)	Gal/GalNAc	2,95	Banerjee <i>et al.</i> (1994) (31)

Depuis quelques années, la résolution par diffraction aux rayons X de complexes obtenus en co-cristallisant diverses lectines avec des oses, des dérivés d'oses, des glycanes ou des glycoconjugués complexes (**Tableau 2**) permet de comprendre, en terme d'interactions moléculaires, le mécanisme de reconnaissance de glucides simples ou complexes par les lectines.

Tableau 2. Complexes lectines-ligands analysés par radiocristallographie à haute résolution.

Lectines	Ligands	Auteurs
Con A	Man, Glc	Derewenda <i>et al.</i> (1989) (32)
LoLI	MeMan, Man, Glc	Bourne <i>et al.</i> (1990) (33)
LoLI	Trisaccharide	Bourne <i>et al.</i> (1990) (34)
GsIV	Lewis	Delbaere <i>et al.</i> (1990) (28)
EcoL	Lactose	Shaanan <i>et al.</i> (1991) (29)
LoLI	Octasaccharide	Bourne <i>et al.</i> (1992) (35)
PsA	Trisaccharide	Rini <i>et al.</i> (1993) (36)
LoLI	Muramyl-dipeptides	Bourne <i>et al.</i> (1994) (37)
LoLII	Glycoprotéine	Bourne <i>et al.</i> (1994) (38)

L'étude des complexes obtenus avec LoLI et la Con A permet d'expliquer l'interaction lectine-ose comme suit :

- l'ose, qui possède une conformation de type chaise, s'insère dans une cavité en forme de fente constituée de sept résidus d'acides aminés (Asp⁸¹, Gly⁹⁹, Asn¹²⁵, Gly²⁰⁹, Ala²¹⁰, Glu²¹¹ et Phe¹²³) dont les six premiers appartiennent à quatre boucles différentes mais convergentes : c'est le site de reconnaissance des monosaccharides (un sur chaque monomère) ;
- le complexe est ensuite stabilisé par un réseau de liaisons hydrogène qui s'établissent entre les atomes d'oxygène O3, O4, O5 et O6 des groupes hydroxyles portés en C3, C4, C5 et C6 sur le cycle pyranique de l'ose et tous les résidus du site, à l'exception de Phe¹²³ qui est responsable d'une interaction hydrophobe avec le cycle pyranique (**Figure 4**).

Ce schéma se retrouve indépendamment de la complexité du glucide qui se lie à la lectine.

Figure 4.

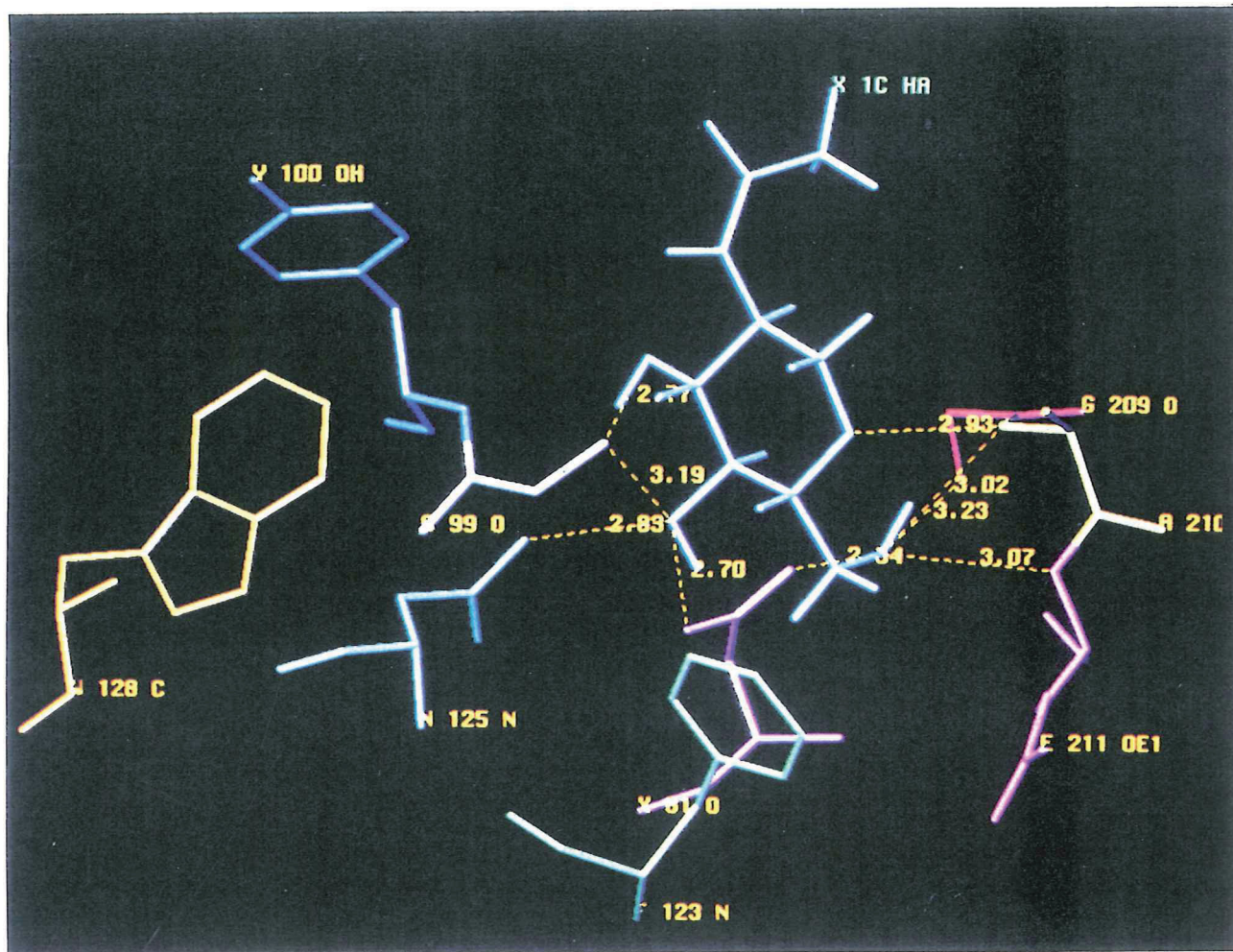
Photographie. Interaction de la *N*-acétylglucosamine avec le site de reconnaissance de LoLI, réalisée par Rougé *et al.* par ancrage moléculaire à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

Le site est constitué de sept résidus (Asp⁸¹, Gly⁹⁹, Phe¹²³, Asn¹²⁵, Gly²⁰⁹, Ala²¹⁰ et Glu²¹¹) portés par quatre boucles différentes mais convergentes. Deux résidus aromatiques (Tyr¹⁰⁰ et Trp¹²⁸) situés au voisinage immédiat du site sont également représentés. Le réseau de liaisons hydrogène s'établissant entre ce dérivé d'ose et les résidus du site est symbolisé par des pointillés avec indication des distances interatomiques. Il se produit en plus une interaction de type hydrophobe entre le cycle pyranique de l'ose et le résidu Phe¹²³.

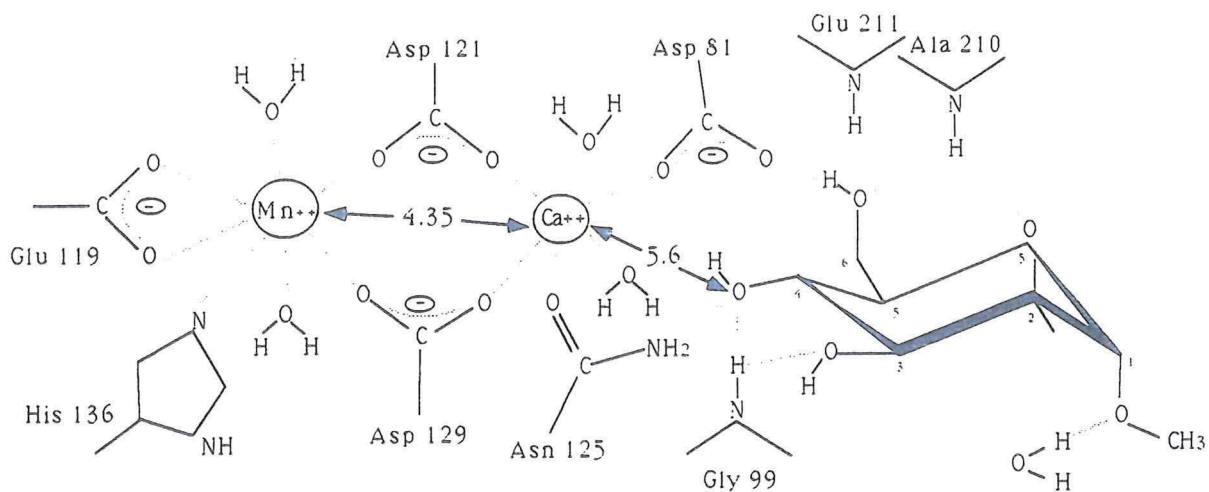
Schéma. Interaction entre l' α -méthyl-mannopyranoside (conformation de type chaise) et le site de reconnaissance de LoLI (26).

Cette représentation montre que deux résidus du site, Asp⁸¹ et Asn¹²⁵, interviennent également dans la fixation du calcium (les lectines sont des métalloprotéines à calcium et à manganèse). La coordinance des deux ions divalents est satisfaite grâce à des molécules d'eau. Ce réseau de liaisons explique l'absence d'interaction des lectines préalablement démétallisées avec les oses et leurs dérivés.

A



B



Deux autres résidus aromatiques, Tyr¹⁰⁰ et Trp¹²⁸, situés au voisinage immédiat du site, peuvent également donner des interactions hydrophobes avec certains ligands, notamment avec le groupe lactyl de l'acide *N*-acétylmuramique (MurNAc) ou avec des muramyl-dipeptides (diastéréoisomères MDP D-D et MDP L-D) bactériens (37).

Lorsque des glycannes plus complexes (trisaccharides, octa- et dodécasaccharides biantennés de type *N*-acétyllactosaminique issus de la lactotransferrine humaine) interagissent avec la lectine, l'unité de mannose de l'antenne A, liée en $\alpha(1\rightarrow3)$ au manche de la fourche, s'insère dans le site, tandis que les deux antennes se plaquent plus ou moins étroitement sur la surface de la lectine, à proximité du site. De nombreuses interactions hydrophobes et des liaisons hydrogène médiées par des molécules d'eau de surface assurent la liaison des deux partenaires (34, 37, 38).

3- La symbiose Légumineuse-*Rhizobium*.

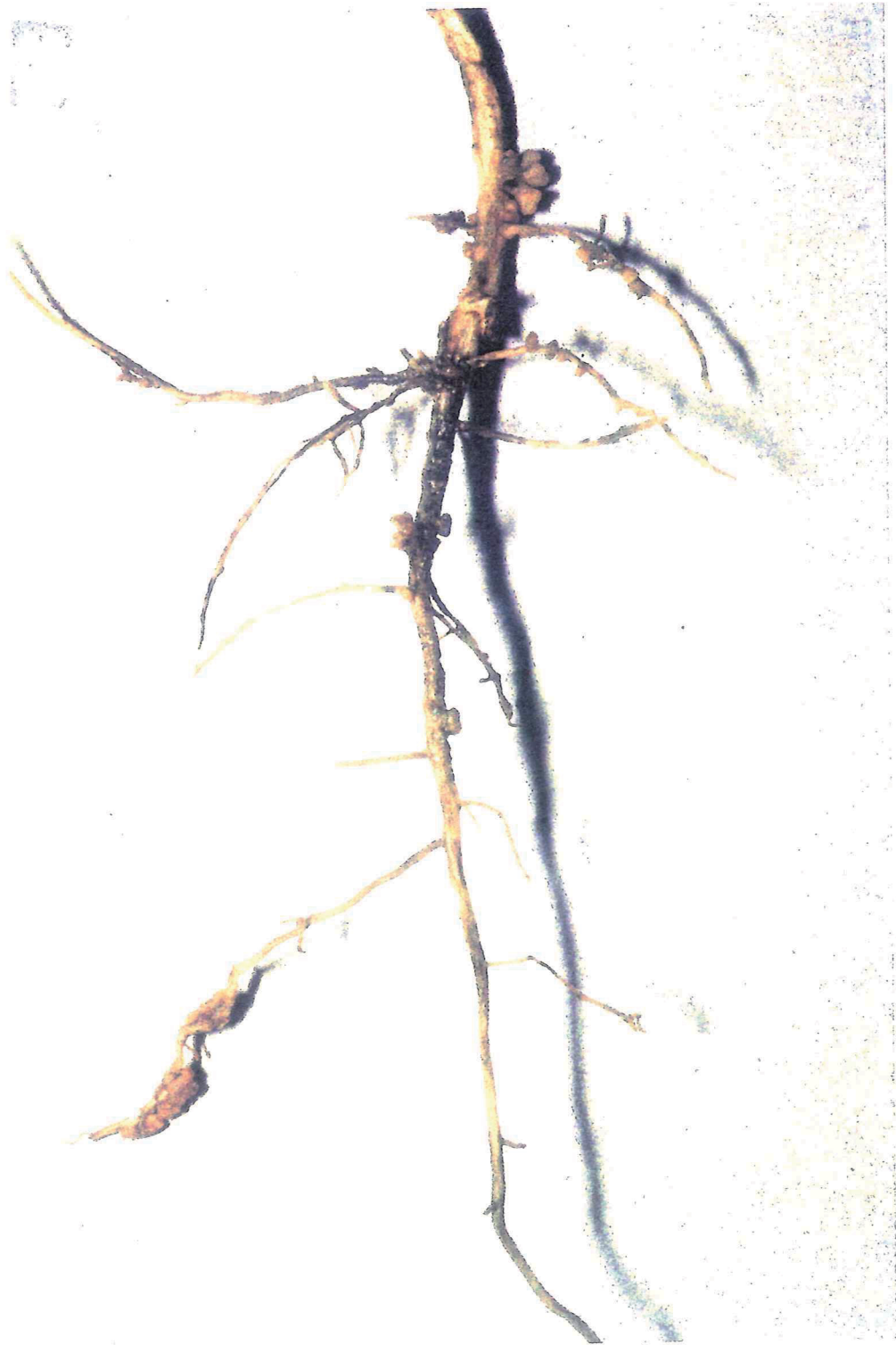
L'établissement d'une symbiose spécifique et fonctionnelle entre une Légumineuse et une bactérie fixatrice d'azote du genre *Rhizobium* revêt une importance économique considérable. Elle permet à la plante-hôte d'utiliser, en plus de l'azote minéral (nitrates NO₃⁻) qu'elle puise dans le sol, l'azote de l'air (N₂) pour ses propres synthèses protéiques. Il s'ensuit des rendements en matière sèche et en azote protéique plus importants que ceux des autres plantes, qui ont fait très tôt se servir des Légumineuses comme engrais verts (Trèfle, Luzerne) dans la pratique de l'assolement. D'ailleurs, la plupart des protéagineux (Soja, Pois, Fèverole) appartiennent à cette famille.

Dans la plante-hôte, les bactéries, modifiées en bactéroïdes plus volumineux, sont hébergées dans des petites nodosités racinaires ou nodules, colorées en rose, lorsqu'elles sont fonctionnelles (Figure 5), par une hémoglobine (leghémoglobine) qui régule l'apport d'oxygène à la bactérie. Grâce à un complexe enzymatique où intervient une nitrogénase, l'azote moléculaire atmosphérique N₂ est réduit en ions ammonium NH₄⁺ qui serviront à une synthèse d'acides aminés supplémentaires.

Figure 5.

Nodosités d'une racine de *Lathyrus ochrus*.

Elles se présentent sous forme de petites protubérances, souvent associées en grappes, colorées en rose lorsqu'elles sont fonctionnelles par une leghémoglobine qui régule l'apport d'oxygène aux bactéroïdes (*Rhizobium leguminosarum* bv. *viciae* dans ce cas) fixateurs d'azote.



La bactérie utilisera ces acides aminés pour ses propres synthèses protéiques, mais le surplus ainsi produit sera rétrocédé à la plante qui, en contrepartie, fournira à la bactérie des glucides issus de la photosynthèse (Figure 6).

Il existe une étroite spécificité de reconnaissance entre ces bactéries telluriques et le système racinaire des plantes-hôtes. Différents groupes de bactéries de genre *Rhizobium* ou apparenté (*Azorhizobium* et *Bradyrhizobium*), dits d'inoculation croisée, peuvent suivant le cas infecter une ou plusieurs espèces de Légumineuses (Tableau 3).

La nature des molécules responsables de cette reconnaissance reste mystérieuse, mais la découverte récente des facteurs Nod permet de supposer qu'ils pourraient être les signaux bactériens reconnus par la plante-hôte. L'activation et l'expression de gènes de *Rhizobium* impliqués dans la synthèse des facteurs Nod par des composés excrétés par la plante-hôte laisse sous-entendre l'existence d'un véritable dialogue chimique dont l'interaction lectine-facteur Nod ne serait qu'un des éléments. En outre, la localisation de lectines dans des nodules fonctionnels, rapportée par Kishinevsky *et al.* (39), suggère qu'elles pourraient aussi jouer un rôle morphogène indispensable à la différenciation des tissus du nodule.

4- Les gènes de nodulation ou gènes *nod* (8).

Ces gènes, gènes *nod* ou *nol*, sont impliqués dans la reconnaissance spécifique de l'hôte, son infection et l'induction de l'organogénèse nodulaire. Ils sont situés sur le mégaplasmide symbiotique pSym qui porte également les gènes de fixation de l'azote, à l'exception des genres *Bradyrhizobium* et *Azorhizobium* où ils sont chromosomiques.

Figure 6.

Schéma descriptif du métabolisme nodulaire (8).

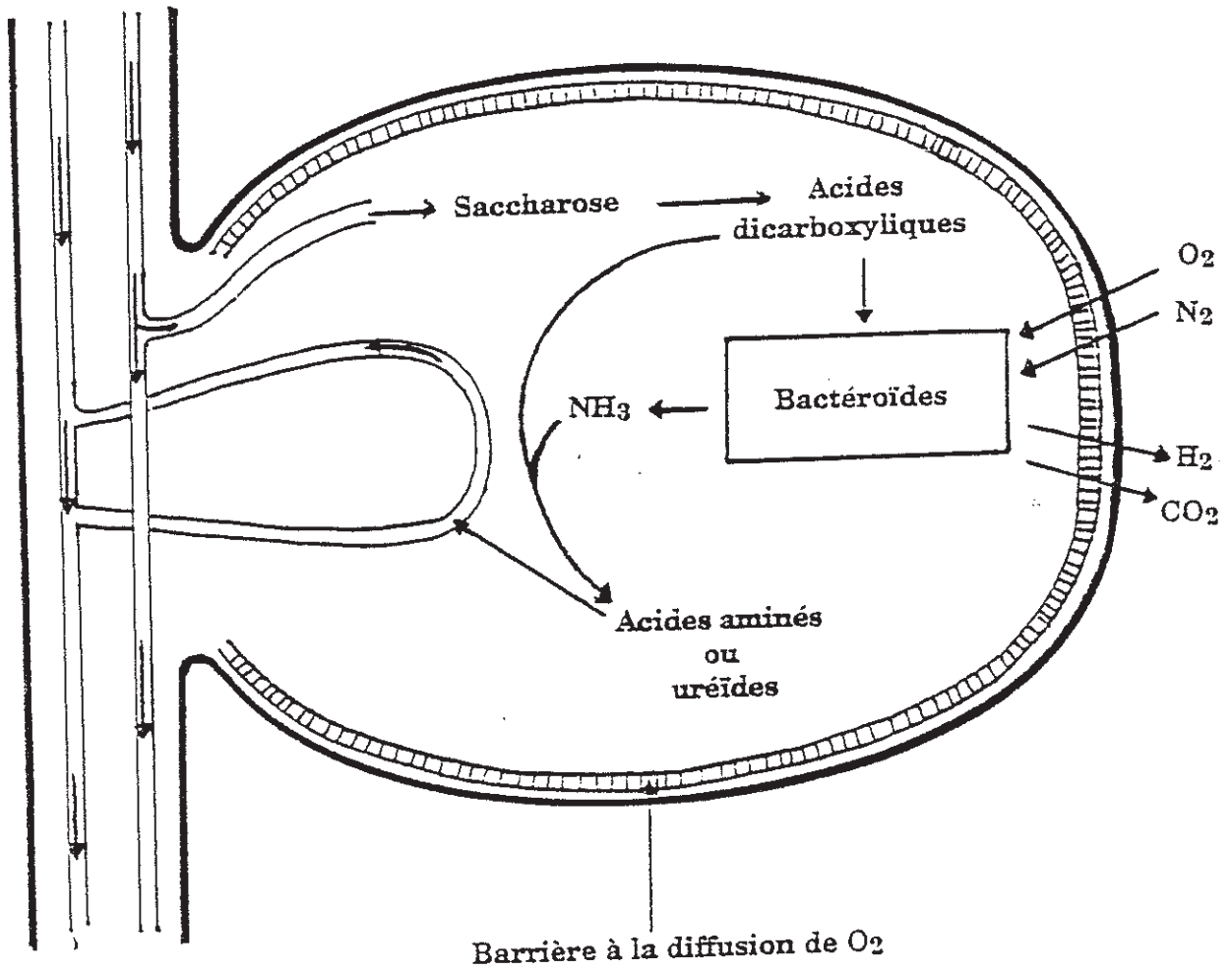


Tableau 3. Spécificité de nodulation des bactéries des genres *Rhizobium* et apparentés à l'égard des Légumineuses-hôtes (18).

Espèces ou biovars	Plantes-hôtes
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i>	<i>Lathyrus</i> (Gesses) <i>Lens</i> (Lentille) <i>Pisum</i> (Pois) <i>Vicia</i> (Vesses)
bv. <i>trifolii</i>	<i>Trifolium</i> (Trèfle)
bv. <i>phaseoli</i>	<i>Phaseolus</i> (Haricot)
<i>Rhizobium meliloti</i>	<i>Melilotus</i> (Mélilot) <i>Medicago</i> (Luzerne) <i>Trigonella</i>
<i>Rhizobium loti</i>	<i>Lotus</i> (Lotier) <i>Anthyllis</i>
<i>Rhizobium ciceri</i>	<i>Cicera</i> (Pois chiche)
<i>Rhizobium galegae</i>	<i>Galega</i>
<i>Rhizobium fredii</i>	<i>Glycine</i> (Soja)
<i>Rhizobium</i> sp. NGR234	<i>Macroptilium</i>
<i>Bradyrhizobium japonicum</i>	<i>Glycine</i> <i>Vigna</i> (Niébé)
<i>Bradyrhizobium lupini</i>	<i>Lupinus</i> (Lupin)
<i>Bradyrhizobium parasponiae</i>	<i>Parasponia</i> (non Légumineuse)
<i>Bradyrhizobium</i> sp. (<i>Arachis</i>)	<i>Arachis</i> (Arachide)
<i>Azorhizobium caulonidans</i>	<i>Sesbania</i>
« Cowpea » <i>Rhizobium</i>	Nombreuses Légumineuses tropicales

Les gènes *nod* sont subdivisés en deux grands groupes : les gènes *nod* communs ou *nodABC*, identifiés chez toutes les Rhizobiacées, et les gènes *nod* spécifiques, propres à une ou à plusieurs espèces. Les gènes *nodABC* sont structurellement voisins et organisés en général en un seul opéron polaire de *nodA* vers *nodC*.

Les gènes *nod* ne sont pas exprimés de façon constitutive et sont régulés (**Figure 7A**) par les gènes de régulation *nodD* (41) selon un mécanisme commun à tous les *Rhizobium* : les gènes *nodD* codent pour des protéines NodD qui sont des activateurs des gènes *nod*. Le nombre de copies du gène *nodD* varie selon les espèces : de un chez *Rh. leguminosarum* bv. *viciae* à trois chez *Rh. meliloti*.

Les protéines NodD, d'une part, interagiraient directement avec des composés sécrétés au niveau des exsudats racinaires de la plante-hôte, essentiellement des flavonoïdes (42), et, d'autre part, se lieraient au DNA par leur extrémité N-terminale très conservée (43, 44) au niveau de séquences d'environ quarante paires de bases de la région promoteur des divers opérons *nod*, les « *nod*-box » (45, 46), jouant ainsi le rôle d'activateurs transcriptionnels. Les flavonoïdes sont des molécules signal qui peuvent agir soit comme inducteur, soit comme inhibiteur des gènes *nodD*, selon la spécificité de la protéine NodD (47, 48).

L'induction est réprimée par des concentrations élevées en ions ammonium (49), donc il semble que les gènes *nod* aient également pour rôle de réguler les différents opérons *nod* en fonction de la concentration en azote combiné.

Un répresseur de l'induction des gènes *nod*, NolR, a été identifié chez *Rh. meliloti* (50) et plus récemment chez d'autres souches des genres *Rhizobium* et *Bradyrhizobium* (51, 52). Il est capable de se lier au promoteur des gènes *nodD1* et *nodD2*, diminuant ainsi leur transcription.

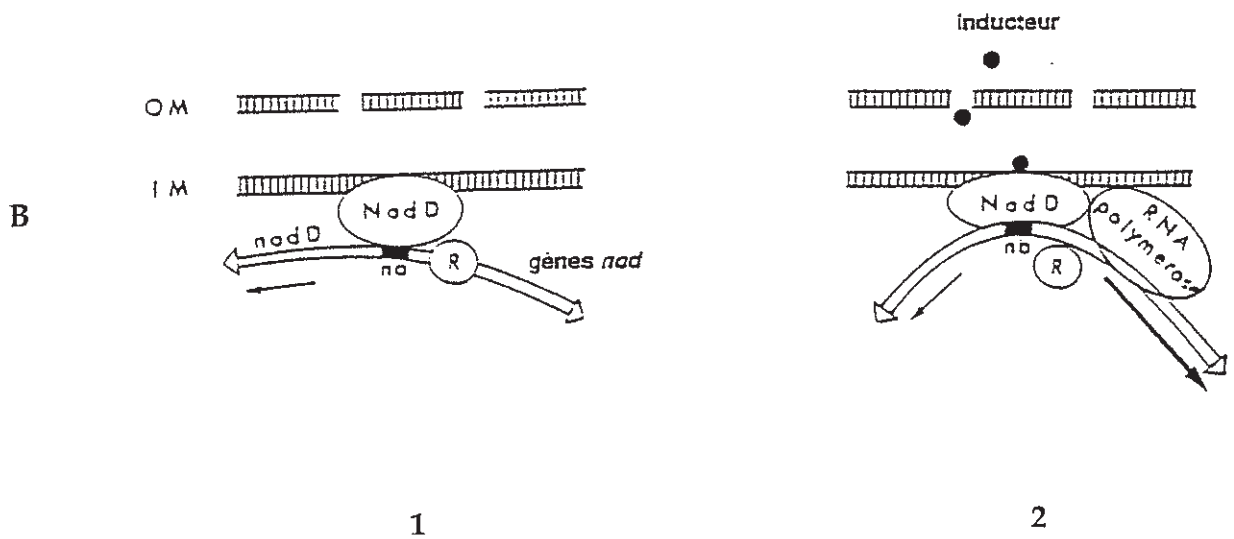
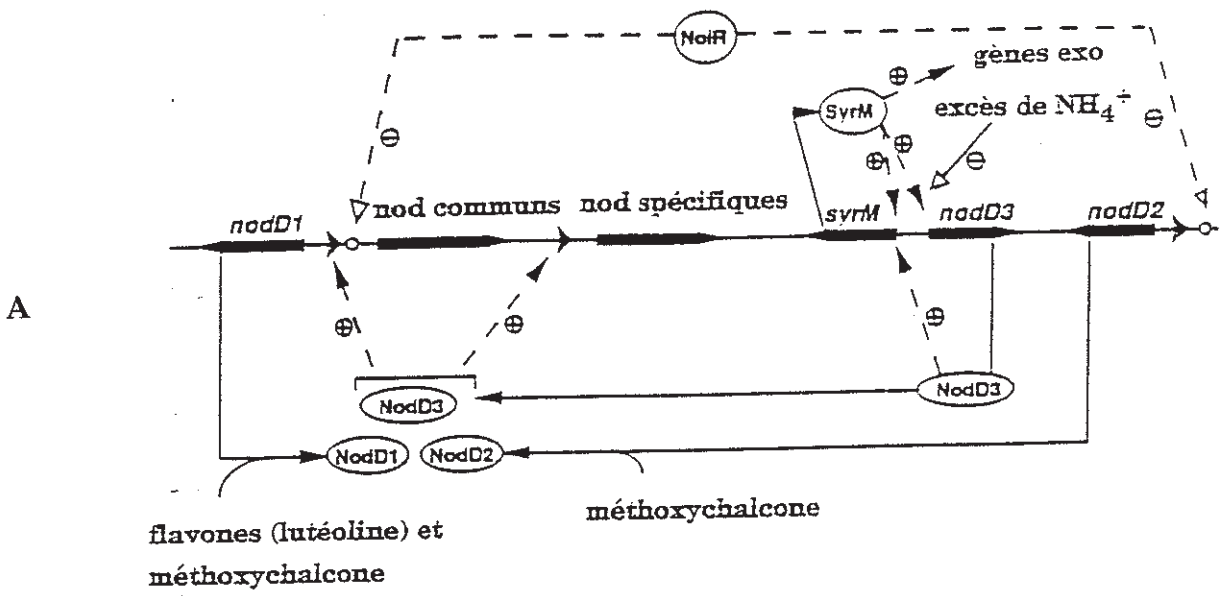
Un modèle moléculaire de la régulation des gènes *nod* a été proposé chez *Rh. meliloti* (**Figure 7B**) (53).

Figure 7.

A- Schéma descriptif de la régulation des gènes de nodulation de *Rhizobium meliloti* (40) (avec ► : *nod*-box et ○ : sites de liaison des répresseurs).

Il convient de noter la présence de gènes régulateurs *syrm*, codant pour des protéines *syrm*, qui ont un rôle analogue à celui des gènes *nodD1*, *nodD2* et *nodD3*, mais ce sont en plus des activateurs transcriptionnels des gènes *exo* codant pour les exopolysaccharides EPS1 et EPS2, situés à la surface de la capsule bactérienne, qui sont impliqués dans le processus d'infection.

B- Modèle moléculaire de la régulation des gènes *nod* de *Rhizobium meliloti* (53) (avec IM membrane interne, OM membrane externe, nb *nod*-box et R répresseur).



En absence de répresseur, la protéine NodD, localisée dans la membrane interne des *Rhizobium*, se lie dans sa conformation inactive au DNA au niveau des *nod*-box, mais ne peut provoquer ni la fixation de la polymérase, ni l'initialisation de la transcription. Dans les souches contenant un répresseur, celui-ci se lie au niveau des sites promoteurs des gènes *nodD*, diminuant fortement leur transcription. Lorsque l'inducteur est sécrété par les plantes, il s'accumule au niveau de la membrane interne et interagit alors avec la protéine NodD. La reconnaissance spécifique inducteur-protéine NodD induit le changement conformationnel de la protéine et permet la fixation de la RNA polymérase qui, en compétition avec le répresseur, conduira à la transcription des opérons *nod*.

Il nous a semblé indispensable de consacrer un chapitre théorique important sur la modélisation moléculaire, discipline à l'heure actuelle encore ésotérique, pour la bonne compréhension des travaux s'y afférant relatés dans cette thèse.

II- La modélisation moléculaire et ses méthodes (54).

1- Bases de la modélisation moléculaire (55 à 62).

La modélisation moléculaire est une récente approche multidisciplinaire en recherche utilisant l'informatique et recouvrant différents domaines de la biologie, la chimie et la physique.

Elle est devenue un complément utile à l'expérimentation classique.

1.1- Définition.

La modélisation moléculaire est une application de méthodes théoriques et de méthodes de calcul pour résoudre des problèmes impliquant la structure moléculaire et la réactivité chimique.

Elle fait appel à la simulation moléculaire qui consiste en la détermination numérique de propriétés structurales, énergétiques ou dynamiques, et de grandeurs de la thermodynamique statistique.

A l'origine, elle se limitait à des représentations manuelles de type DREIDING ou CPK (Corey, Pauling, Koltun).

Elle requiert désormais l'emploi d'ordinateurs alliant puissance de calcul et infographie sophistiquée qui permettent d'appliquer cette technique à des molécules de plus grande taille et aux bio-macromolécules.

1.2- Fondements physiques.

1.2.1- Mécaniques.

Des méthodes théoriques comme la mécanique moléculaire, la mécanique quantique semi-empirique ou la mécanique quantique *ab initio* permettent d'évaluer les propriétés physiques d'une molécule.

1.2.2- Thermodynamiques.

Les molécules de taille intermédiaire, plus flexibles que les petites molécules, possèdent plusieurs états conformationnels de population variable à la température ambiante ou physiologique pour les molécules d'intérêt biologique.

1.3- Méthodologie informatique.

1.3.1- Obtention d'un modèle initial.

Ces modèles peuvent être :

- construits à partir d'éléments structuraux (atomes, groupements, résidus nucléotidiques ou peptidiques) sélectionnés dans une bibliothèque de programme ;
- importés intégralement ou partiellement de banques de données (Protein Data Bank (PDB), Cambridge Structural Data (CSD)...) dans lesquelles sont stockées des structures établies par radiocristallographie ;
- établis à partir de données expérimentales, c'est-à-dire de résultats de radiocristallographie ou de résonance magnétique nucléaire (RMN).

1.3.2- Optimisation de la géométrie.

Cette opération revient à minimiser l'énergie interne du modèle par deux grands types de méthodes :

- soit de mécanique moléculaire, en appliquant sur le modèle initial un champ de force, et en se servant d'un ou de plusieurs algorithmes de minimisation ;

- soit de dynamique moléculaire, en simulant en sus les mouvements internes sur des temps très courts.

Elle conduit à une ou à un ensemble de conformations.

1.3.3- Techniques de graphisme moléculaire.

Elles permettent de représenter sur un écran la structure en 2 ou en 3 dimensions (2D ou 3D) d'une molécule, de la manipuler (rotations, translations, changement de conformation, superposition...) de façon interactive et de l'analyser (calculs de paramètres géométriques tels que distances, angles, surfaces accessibles...).

1.4- Configurations informatiques requises.

Les différents types de matériel habituellement utilisés sont des stations Silicon Graphics, Sun, Apollo, Evans & Sutherland et Vax, ainsi que des PC et Macintosh haut de gamme, disposant d'un processeur rapide et d'importantes ressources mémoire. Le système d'exploitation dominant est Unix avec une interface graphique gérée par le protocole de fenêtrage X-Windows.

Nos travaux de modélisation ont été réalisés sur deux stations Silicon Graphics - Iris 4D25G et Indigo R3000 - équipées d'écrans de 20 pouces, montées en réseau à partir d'une unité centrale comprenant un processeur RISC cadencé à 100 MHz, un disque dur de 1,2 Go et une mémoire vive de 16 Mo. L'interface graphique est ici Open GL.

2- Propriétés moléculaires.

2.1- Structure moléculaire.

2.1.1- Définition.

Une structure moléculaire est un enchaînement atomique caractérisé par la nature des atomes constitutifs, leur ordre et leur mode de liaison, par sa géométrie interne

(longueurs et angles de liaison, configuration) et par sa conformation en structure 3D.

2.1.2- Aspect énergétique.

En considérant une molécule dans son environnement (solvant, membrane, macromolécule...), l'état d'équilibre du système correspond à un minimum d'enthalpie libre ou énergie libre de Gibbs G :

$$\Delta G = \Delta H - T\Delta S \quad \text{avec H enthalpie, S entropie et T température absolue.}$$

Cette énergie contient trois termes :

- l'énergie interne de la molécule traduisant des interactions interatomiques intramoléculaires ;
- l'énergie d'interaction molécule-solvant ;
- l'énergie du solvant en dehors des zones d'interaction avec le soluté (qui ne sera ici pas développée).

L'énergie des liaisons simples est élevée, de l'ordre de 400 à 600 kJ.mole⁻¹.

Pour deux atomes séparés par deux liaisons, l'énergie d'interaction ne dépend pas de la rotation autour des liaisons. Par contre, pour deux atomes séparés par au moins trois liaisons, elle dépend de la rotation autour des liaisons entre les atomes N - k et N - (k + 1), pour k allant de 1 à N - 3 et N étant le nombre d'atomes. Les effets de ces interactions s'exercent dans un domaine de distances de l'ordre de 2 à 10 Å.

L'énergie interne moléculaire comporte par conséquent des termes fixes et des termes variables qui sont fonction des angles de rotation essentiellement autour des liaisons simples, mais également, dans des cas particuliers, autour de liaisons doubles (changements de configuration).

Une molécule isolée présente donc généralement plusieurs minima d'énergie correspondant à plusieurs conformations stables ou conformères, le minimum le plus bas définissant la forme la plus stable.

Les résultats de l'analyse conformationnelle montrent que pour les molécules en solution la conformation stable se trouve souvent dans le jeu des conformations

stables de la molécule isolée. Les interactions avec le solvant ont surtout pour effet de modifier les valeurs relatives des énergies minimales ou de déplacer légèrement les positions des minima dans l'espace conformationnel.

On admet donc que les résultats d'un calcul conformationnel sur la molécule isolée donnent une bonne idée des possibilités conformationnelles de cette molécule et peuvent conduire dans de nombreux cas à la détermination de la forme la plus stable ou d'une forme prédominante.

On peut en outre, par une distribution de Boltzmann ou par la méthode de Monte Carlo, explorer complètement l'espace conformationnel d'une molécule en examinant les possibilités d'équilibre conformationnel ou faire une analyse statistique des conformères. On est également capable de simuler, surtout par dynamique moléculaire, des effets de solvant à condition de disposer de logiciels adéquats et d'ordinateurs très puissants.

Il existe de nombreuses méthodes (quantiques, semi-empiriques, empiriques) de calcul de l'énergie interne moléculaire qui seront traitées plus loin (cf. II-3).

2.2- Mouvements moléculaires.

Une structure moléculaire est animée de différents types de mouvements internes d'amplitude, de direction et de fréquence très variables.

2.2.1- Mouvements de vibration (allongement ou « stretching », et plissement ou « bending »).

Ils sont d'amplitude limitée ($\approx 0,1 \text{ \AA}$, et/ou quelques degrés) et de fréquence élevée ($10^{13} - 10^{14} \text{ Hz}$).

2.2.2- Mouvements de rotation.

Ils peuvent avoir des amplitudes et des fréquences très variables (de 10^{12} - 10^{11} Hz à quelques Hz).

Une rotation peut être partielle ou limitée à quelques degrés (autour d'une liaison double) ou à quelques dizaines de degrés (autour d'une liaison simple).

Elle peut être complète, le fragment oscillant alors entre deux et trois positions. Les conformations correspondantes sont des isomères de rotation ou rotamères.

2.2.3- Agitation thermique.

Les déplacements des atomes d'une molécule présentant une conformation stable unique autour de leur position d'équilibre sont caractérisés par le facteur d'agitation thermique. L'échelle de temps correspondante est celle des mouvements rapides : de 10^{-13} à 10^{-10} s.

En solution, les mouvements intramoléculaires ont une amplitude plus grande (au moins dans les parties flexibles de la molécule) et couvrent une échelle de temps beaucoup plus large. Dans certains cas, l'équilibre conformationnel peut impliquer deux conformères nettement différents qui s'échangent avec une vitesse suffisamment lente pour qu'une méthode spectroscopique (ultraviolets, fluorescence, RMN...) permette de les distinguer.

3- Représentation de l'énergie interne moléculaire.

Les méthodes quantiques ne sont appropriées que pour des petites molécules ou des fragments moléculaires d'au plus quelques dizaines d'atomes.

Les méthodes semi-empiriques peuvent être appliquées à des structures plus grandes, mais elles sont difficiles d'emploi.

On utilise donc généralement des méthodes empiriques de représentation de l'énergie interne moléculaire qui sont à la base des calculs de mécanique moléculaire et de dynamique moléculaire.

3.1- Méthodes quantiques et semi-empiriques.

L'objectif de la mécanique quantique est principalement de déterminer l'énergie et la distribution électronique sur les orbitales.

L'énergie moléculaire est calculée par l'équation de Schrödinger. Son expression est au préalable séparée en deux termes par l'approximation de Born-Oppenheimer qui suppose que, les noyaux bougeant beaucoup plus lentement que les électrons, ils peuvent alors être assimilés à des masses fixes.

Cette équation peut être résolue :

- telle quelle (méthodes *ab initio*) : on se sert de toutes les orbitales atomiques occupées à l'état fondamental ;
- en introduisant des approximations (méthodes semi-empiriques) : on prend en compte uniquement les orbitales de valence.

Pour mieux analyser les interactions entre charges, on peut placer la molécule dans une "boîte" (espace virtuel de forme cubique) et en fixant une sonde (proton) sur chaque sommet on obtient, par calcul de densité électronique, un potentiel moléculaire électrostatique.

Certains logiciels permettent aussi de manipuler une courbe d'isopotential en examinant les zones chargées et leur complément moléculaire dans une interaction récepteur-ligand.

3.2- Méthodes empiriques.

En mécanique moléculaire, on considère une structure moléculaire comme étant composée de billes et de ressorts (forces harmoniques) associés à une série de fonctions de potentiel.

3.2.1- Définition du champ de force ou « force field » moléculaire.

Le champ de force est un modèle mathématique qui, par un choix approprié des fonctions de potentiel et des paramètres qu'elles contiennent, doit représenter aussi bien que possible les variations de l'énergie potentielle totale avec la géométrie moléculaire. Il est exprimé sous la forme d'une somme de fonctions de potentiel appelée fonctionnelle d'énergie conformationnelle.

$$E_{\text{totale}} = E_{\text{liaison}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{Van der Waals}} + E_{\text{électrostatique}} + E_{\text{pont H}}$$

ou

$$E_{\text{totale}} = E_b + (E_\theta + E_\zeta) + E_\varphi + E_{vw} + E_{el} + E_h$$

A chacun de ces termes correspond une position d'équilibre préférentielle ; la recherche de l'énergie minimale se fait donc par optimisation de la géométrie en réduisant au maximum les écarts structuraux par rapport à des valeurs de référence (cf. 3.2.3.1).

Il n'existe cependant pas encore de modèle unique permettant de simuler tous les aspects du comportement moléculaire, mais un ensemble de modèles.

Le développement de l'expression analytique à 3N variables des champs de force vers des termes plus complexes et plus précis est limité par la puissance de calcul des ordinateurs.

3.2.2- Formule empirique du champ de force (63).

On distingue les interactions intramoléculaires entre atomes liés et non-liés.

$$E_{\text{totale}} = E_{\text{liés}} + E_{\text{non-liés}}$$

3.2.2.1- Energie d'interaction entre atomes liés.

3.2.2.1.1- Energie de déformation du squelette.

a) élongation des liaisons.

Elles sont faibles, de l'ordre de 0,05 Å.

$$E_b = \frac{1}{2} \sum_{\text{liaisons}} K_b (b - b_0)^2$$

avec b longueur de la liaison et $K_b \cong 400 \text{ kJ.mol}^{-1}.\text{Å}^{-2}$.

b) distorsion des angles de valence.

Elles sont de l'ordre de quelques degrés.

$$E_\theta = \frac{1}{2} \sum_{\substack{\text{angle} \\ \text{valence}}} K_\theta (\theta - \theta_0)^2$$

avec θ angle de valence et $K_\theta \cong 40 \text{ kJ.mol}^{-1}.\text{deg}^{-2}$.

c) déformation des angles dièdres impropres.

Un angle dièdre impropre est défini par trois atomes liés à un quatrième atome qui est soit de type sp^2 , les quatre atomes étant maintenus dans une configuration plane, soit de type sp^3 , les quatre atomes étant maintenus dans une configuration tétraédrique.

Cette déformation décrit donc les sorties de plans de certains atomes, non interprétables par des angles dièdres classiques.

$$E_\zeta = \frac{1}{2} \sum_{\substack{\text{dièdres} \\ \text{impropres}}} K_\zeta (\zeta - \zeta_0)^2$$

avec ζ angle dièdre impropre.

3.2.2.1.2- Energie de torsion.

Elle est associée à la rotation autour d'une liaison charnière entre deux plans formés par quatre atomes consécutifs définissant ainsi un angle dièdre.

$$E_{\varphi} = \frac{1}{2} \sum_{\text{dièdres}} K_{\varphi} [1 + \cos(n\varphi - \delta)]$$

avec φ angle dièdre, K_{φ} constante de torsion ou barrière de rotation de l'ordre de 40 à 70 kJ.mol⁻¹, n périodicité de la rotation ($n = 2$ si sp² ou $n = 3$ si sp³) et δ angle de phase.

La combinaison des fonctions de liaison, d'angle et de torsion d'angle représente le champ de force de valence.

3.2.2.2- Energie d'interaction entre atomes non-liés.

3.2.2.2.1- Interactions de Van der Waals.

L'énergie correspondante ou fonction de Lennard-Jones est la somme d'un terme attractif due aux interactions entre dipôles fluctuants appelé attraction de London et d'un terme répulsif traduisant le recouvrement des nuages électroniques à courte distance appelé répulsion de Lennard-Jones.

$$E_{vw} = \frac{1}{2} \sum_{\text{paires}} K(D_{ij}/r_{ij}^{12} - C_{ij}/r_{ij}^6)$$

avec r distance interatomique, i et j couple d'atomes et, C et D constantes.

Cette fonction est la plus souvent employée dans les programmes de modélisation moléculaire. On peut cependant rencontrer quelques variantes comme les fonctions de Buckingham ou de Morse où :

$$E_{vw} = \frac{1}{2} k(r - r_0)^2.$$

3.2.2.2.2- Interactions électrostatiques.

La molécule est décrite par des charges ponctuelles centrées sur les atomes. L'énergie électrostatique suit la loi de Coulomb :

$$E_{el} = \sum_{\text{paires}} (1/4\pi\epsilon \cdot q_i q_j / r_{ij}) \quad \text{avec } \epsilon \text{ permittivité et } q \text{ charges ponctuelles.}$$

Les programmes incluent une fonction de coupure ou « switch function » pour limiter la portée des potentiels de Van der Waals et électrostatique à une certaine distance appelée r_{cut} comprise entre 8 à 10 Å afin de diminuer le temps de calcul pour de grandes molécules.

3.2.2.2.3- Liaisons ou ponts hydrogène.

Ce sont des interactions d'importance intermédiaire (8 à 20 kJ.mol⁻¹) entre un atome d'hydrogène déficient en électron et un atome de forte densité électronique porteur d'un doublet libre. Deux fonctions tenant compte de la directivité de la liaison hydrogène sont utilisées :

$$\begin{aligned}
 & \text{- la fonction "10-12"} & E_h &= A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{10} \\
 & \text{- la fonction "6-12"} & E_h &= A'_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6
 \end{aligned}$$

avec A, A', B et B' constantes.

3.2.3 - Paramétrisation.

3.2.3.1- Paramètres de référence.

Les fonctions de potentiel présentées précédemment contiennent de nombreux paramètres qui sont optimisés à partir de données expérimentales déduites de l'analyse de composés types (spectroscopie vibrationnelle, cristallographie, thermodynamique...) ou dans quelques cas déduits de calculs de mécanique quantique.

3.2.3.2- Paramètres de substitution.

La plupart des banques de données sont très restreintes.

L'absence de paramètres de référence, en particulier pour les systèmes hétérocycliques, nécessite un choix de paramètres de substitution ou *ad hoc*.

3.2.4- Un exemple de champ de force, celui utilisé par le logiciel Discover : Consistent Valence Force Field ou CVFF (64) (Figure 8).

L'expression de l'énergie potentielle totale comprend différentes fonctions dites de "termes croisés" qui traduisent les phénomènes de couplage des mouvements de déformation de liaison et des angles de valence, dièdres et dièdres impropres.

Figure 8.

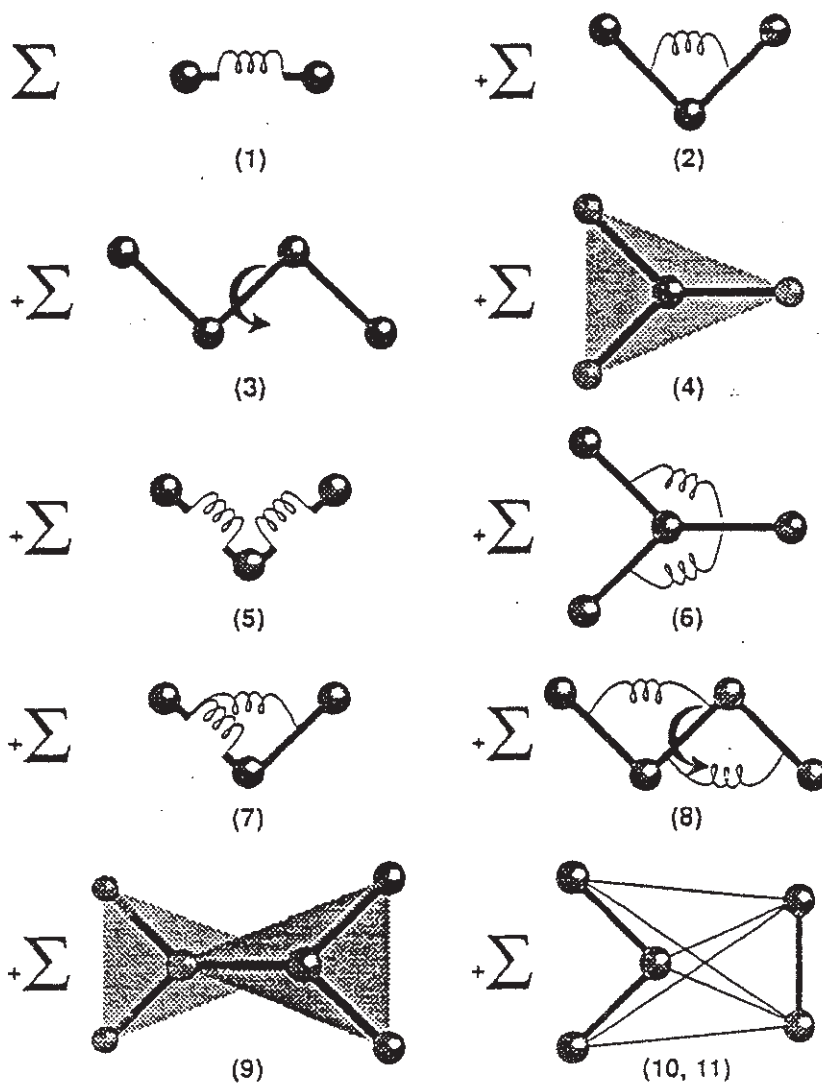
Expression mathématique (A) du champ de force CVFF ou « Consistent Valence Force Field » de Discover, avec illustration des interactions associées aux différentes fonctions de potentiel (B) (64).

Les termes (1) à (9) représentent le champ de force de valence, dont (5) à (9) sont des "termes croisés".

- (1) élongation des liaisons ou « stretching ».
- (2) distorsion des angles de valence ou « bending ».
- (3) torsion (déformation des angles dièdres).
- (4) déformation des angles dièdres impropres pour des atomes de type sp^3 .
- (5) influence de l'élongation d'une liaison sur la suivante.
- (6) influence de la distorsion d'un angle de valence sur le suivant dans le cas de trois atomes situés dans un même plan liés à un atome commun.
- (7) influence réciproque de l'élongation d'une liaison sur la distorsion de l'angle de valence associé ou « stretch-bending ».
- (8) influence de la distorsion de deux angles de valence sur leur angle dièdre.
- (9) influence de la déformation d'un angle dièdre impropre sur le suivant dans le cas d'une liaison entre deux atomes de type sp^3 .
- (10) liaisons hydrogène (fonction "6-12").
- (11) interactions électrostatiques.

$$\begin{aligned}
 E_{\text{pot}} = & \sum_b D_b \left[1 - e^{-\alpha(b-b_0)^2} \right] + \frac{1}{2} \sum_{\theta} H_{\theta} (\theta - \theta_0)^2 + \frac{1}{2} \sum_{\phi} H_{\phi} [1 + s \cos(n\phi)] \\
 & + \frac{1}{2} \sum_{\chi} H_{\chi} \chi^2 + \sum_b \sum_{b'} F_{bb'} (b - b_0) (b' - b'_0) + \sum_{\theta} \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0) (\theta' - \theta'_0) \\
 & + \sum_b \sum_{\theta} F_{b\theta} (b - b_0) (\theta - \theta_0) + \sum_{\phi} F_{\phi\theta\theta'} \cos \phi (\theta - \theta_0) (\theta' - \theta'_0) + \sum_{\chi} \sum_{\chi'} F_{\chi\chi'} \chi\chi' \\
 & + \sum \varepsilon [(r^*/r)^{12} - 2(r^*/r)^6] + \sum q_i q_j' \varepsilon r_{ij}
 \end{aligned}
 \tag{1} \tag{2} \tag{3} \tag{4} \tag{5} \tag{6} \tag{7} \tag{8} \tag{9} \tag{10} \tag{11}$$

A



3.2.5- Faiblesses des champs de force.

Le principal défaut des champs de force vient de la description des charges ponctuelles fixes. De plus, le traitement de l'énergie électrostatique est discutable car on néglige les effets dus aux charges distribuées en volume et l'évolution de cette distribution en fonction de la conformation. De même, la polarisation et les effets à longue portée sont ignorés.

Les autres termes sont d'une approximation satisfaisante.

Il n'y a pas de règles absolues concernant le genre et le nombre de fonctions de potentiel utilisables, c'est pourquoi on rencontre un grand nombre de champs de force différents en mécanique moléculaire.

4- Minimisation de l'énergie interne moléculaire (65-69).

4.1- Stratégie générale.

L'énergie interne est une fonction de $3N$ variables. Il faut donc minimiser cette fonction cible $F(x)$ décrivant la surface de l'énergie potentielle, où x représente l'ensemble des coordonnées internes de la molécule, par un ou plusieurs algorithmes ou minimiseurs.

Ceux-ci supposent que cette surface est approximativement harmonique. Même pour les surfaces non-harmoniques, l'allure de la surface devient harmonique s'il y a convergence vers un minimum.

On doit au préalable définir si le système étudié prend en compte ou non les atomes d'hydrogène et la sphère de solvatation. On peut en outre introduire des contraintes et des restrictions pour contrôler et diriger la minimisation, comme dans le cas d'un ancrage moléculaire enzyme-substrat.

Une telle fonction mène, en général, à un minimum absolu ou global et à un grand nombre de minima relatifs ou locaux. Il n'existe pas de méthodes analytiques ou numériques permettant de déterminer le minimum absolu : les algorithmes

aboutissent dans le meilleur des cas à un minimum relatif de faible énergie et l'on ne peut qu'espérer que ce minimum ne soit pas trop éloigné en énergie du minimum absolu.

On distingue les méthodes qui utilisent :

- uniquement la pente de la surface (dérivée première) : algorithmes d'ordre 1 ;
- à la fois cette pente et la courbure de la surface (dérivées première et secondes) : algorithmes d'ordre 2.

4.2- Principaux algorithmes.

4.2.1- Algorithmes d'ordre 1.

On définit une fonction cible à deux variables $F(x,y)$ en prenant comme valeurs initiales de x et y le jeu de coordonnées cartésiennes brutes x^0 et y^0 .

On recherche ensuite son minimum par ajustement itératif des coordonnées le long d'un axe particulier ; un tel algorithme est dit de recherche de ligne ou « line search » (Figure 9A). Celle-ci correspond à une minimisation unidimensionnelle dans une direction donnée.

La première direction (pente) s^0 est calculée à partir des valeurs respectives des dérivées partielles $\partial F/\partial x$ et $\partial F/\partial y$ en x^0 et y^0 . Elles sont proportionnelles aux coordonnées de sorte que plus on est loin du minimum, plus leurs valeurs sont grandes. Un minimiseur détermine par conséquent à la fois la direction du minimum et la distance à parcourir dans cette direction pour atteindre le puits de valeur minimale.

La minimisation converge quand le gradient g^0 est nul. La nouvelle direction s^1 est alors orthogonale à la direction précédente ($s^1 = -s^0$), et ne dépend pas de l'algorithme. La recherche de ligne permet donc d'extraire toute l'information énergétique dans une direction avant de passer à la suivante.

On continue ainsi jusqu'à ce que l'énergie de deux itérations successives soit constante, le minimum relatif étant alors atteint.

4.2.1.1- Méthode "de la plus grande pente" ou « steepest descent » (Figure 9A).

La direction suivie est indiquée par l'opposé du gradient d'énergie au point considéré :

$$s^i = -g^i.$$

C'est la direction dans laquelle l'énergie diminue le plus vite, du moins localement.

On peut en sus définir un pas α (Figure 9C), c'est-à-dire un changement de variable ou de coordonnées ou bien un angle de rotation autour d'une liaison, et examiner son influence sur l'énergie.

Puisque cette méthode permet de décroître rapidement la valeur de la fonction, elle est habituellement très efficace pour supprimer les mauvais contacts ou les principaux problèmes stéréochimiques qui existent dans les coordonnées brutes de structures cristallines ou modélisées, tout en les perturbant très peu. Mais la convergence devient faible et très lente en s'approchant du minimum (phénomènes oscillants, remontées d'énergie), particulièrement si la surface énergétique possède des vallées étroites.

4.2.1.2- Gradient conjugué ou « conjugate gradient » (Figure 9B).

La nouvelle direction tient compte du gradient au point considéré mais est également conjuguée à toutes les directions précédentes :

$$s^{i+1} = -g^{i+1} + \gamma s^i \quad \text{avec } \gamma = (g^{i+1})^2 / (g^i)^2.$$

Il n'y a ainsi pas de perte d'information énergétique.

Le pas est ajusté plus finement à chaque cycle pour obtenir une meilleure diminution d'énergie.

L'intérêt de cet algorithme est d'éviter un comportement oscillatoire autour du minimum et d'accélérer la convergence. Il se révèle cependant moins efficace que steepest descent ou même inutilisable (pas de convergence) pour des structures qui présentent beaucoup de mauvais contacts, telles que les structures moyennées sur la trajectoire d'une dynamique moléculaire.

Figure 9.

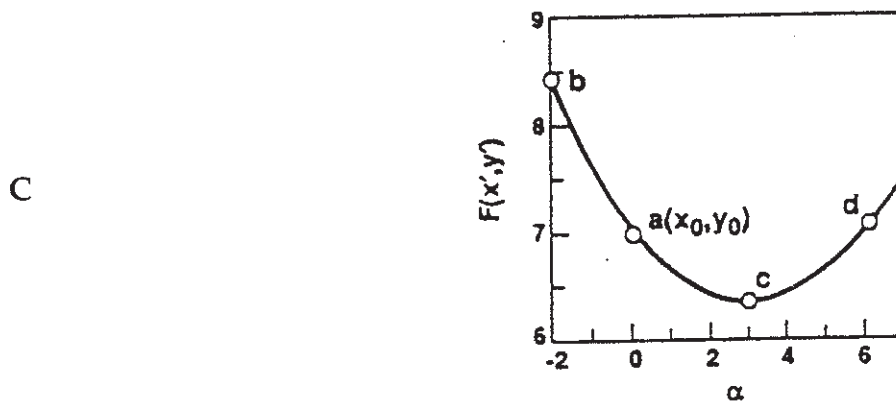
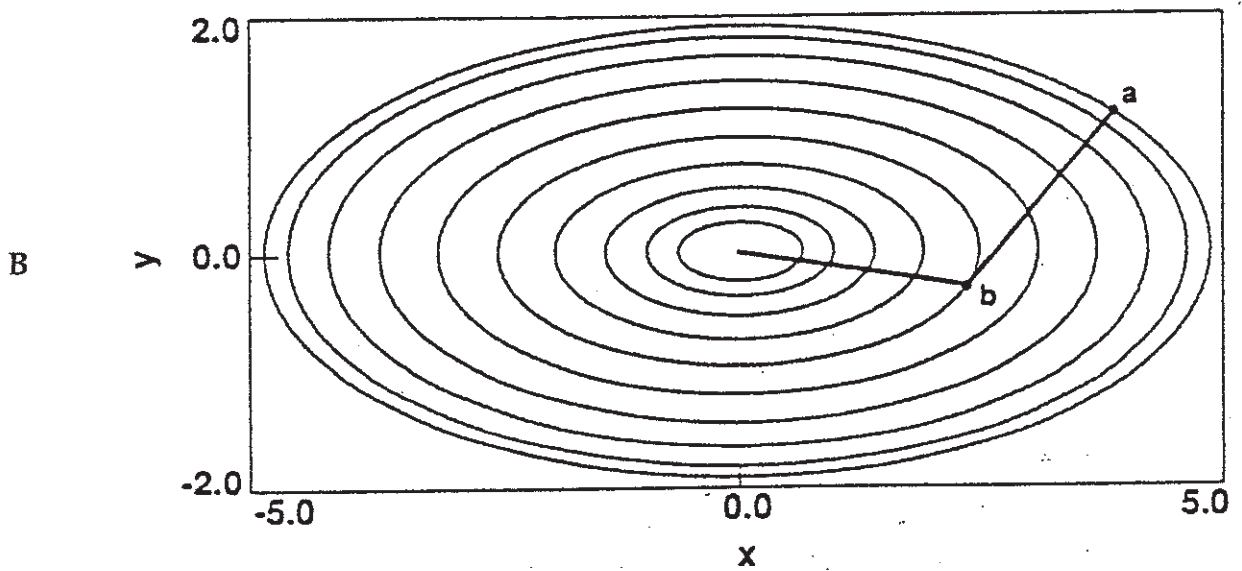
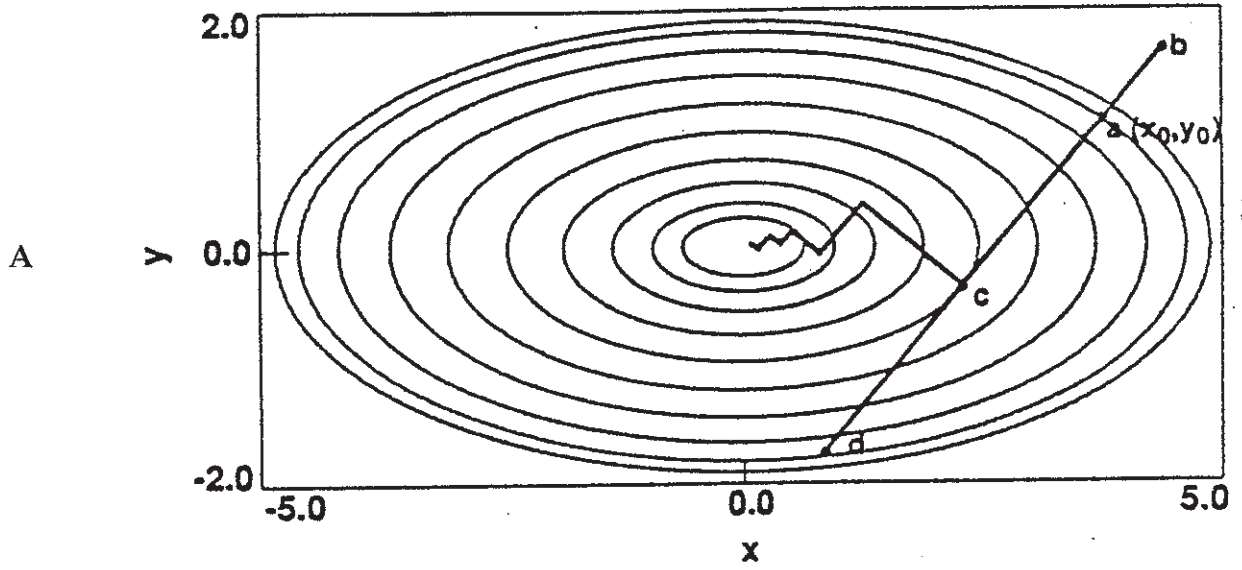
Principe de la minimisation de l'énergie interne moléculaire à l'aide de la recherche de ligne ou « search-line » utilisée par les algorithmes d'ordre 1.

La fonction décrivant la surface d'énergie potentielle choisie comme exemple, $F(x,y) = x^2 + 5y^2$, est représentée par des abaques (A et B) (64).

Le gradient de F en un point initial $a(x_0, y_0)$ donne les composantes du vecteur directeur de la première ligne de recherche (droite b-a-c-d) (A). Les intersections de cette droite avec les ellipses successivement rencontrées définissent un pas α que l'on va suivre ($F(x',y') = f(\alpha)$) (C) (64) pour déterminer la valeur minimale de F dans cette direction, qui est ici atteinte au point c pour $\alpha = 3$. Le gradient s'y annule et va par conséquent orienter la recherche de ligne dans une seconde direction orthogonale à la première, ce qui est le cas pour la méthode de la plus grande pente ou « steepest descent » (A).

Dans la méthode du gradient conjugué ou « conjugate gradient » (B), la nouvelle direction va être en sus conjuguée à la précédente, ce qui va permettre de l'affiner et ainsi de réduire le nombre d'étapes conduisant au minimum.

Dans ces deux cas, l'algorithme va réitérer le processus de convergence jusqu'à ce que l'énergie de deux étapes successives soit identique, un minimum, le plus souvent relatif, étant alors atteint.



C'est une méthode de choix pour les structures qui sont trop importantes pour pouvoir stocker et manipuler une matrice de dérivées secondes.

4.2.2- Algorithmes d'ordre 2 : Newton-Raphson (et méthodes apparentées).

En plus du gradient qui contribue à indiquer une direction, on utilise la courbure de la fonction pour prédire où, le long du gradient, la fonction va changer de direction et passer par un minimum (Figure 10).

Pour cela, on a recours à une approximation quadratique de $F(x)$ par un développement en série de Taylor limité à l'ordre 2.

$$1) \text{ Au point origine } G, \text{ on a : } F(x) = F(G) + \sum_i \frac{\partial F}{\partial x_i} x_i + \frac{1}{2} \sum_{ij} \frac{\partial^2 F}{\partial x_i \partial x_j} x_i x_j + \dots$$

$$2) \text{ On pose ensuite : } c = F(G) ; b = -\nabla F ; A_{ij} = \frac{\partial^2 F}{\partial x_i \partial x_j}.$$

L'expression apparaît alors sous la forme : $F(x) = c - bx - \frac{1}{2}x^tAx$, qui dérivée donne : $dF = -b + Ax$.

$$3) \text{ Au minimum } dF = 0, \text{ d'où } x_m = A^{-1}.b.$$

$$4) \text{ Par conséquent, au point } x_i, \text{ on a : } x_m - x_i = A^{-1}.(-dF(x_i))$$

avec A^{-1} inverse de la matrice des dérivées secondes ou matrice hessienne.

Puisque la matrice complète des dérivées secondes définit la courbure de la surface dans la direction de chaque gradient, la matrice hessienne peut être multipliée par le gradient pour obtenir les composantes du vecteur qui mène directement au minimum le plus proche. Théoriquement, ce minimum peut donc être atteint en une étape.

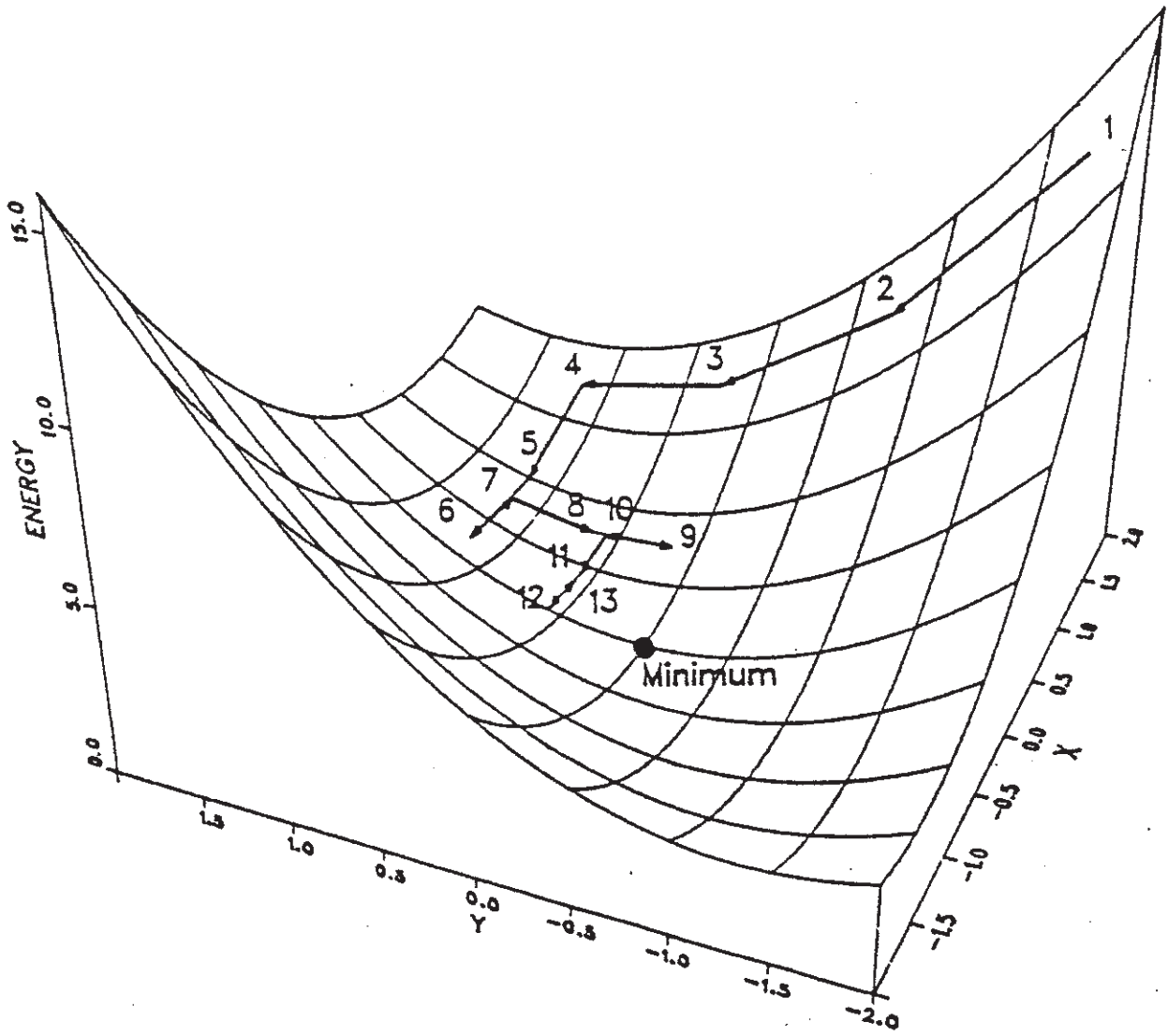
Le gros défaut de la méthode Newton-Raphson est que le résultat global peut être aussi bien un minimum qu'un col ou "selle de cheval" de la surface énergétique. Pour y remédier, certains programmes font appel à quelques étapes de steepest descent afin de revenir dans une région où la dérivée seconde de l'énergie par rapport au vecteur propre est positive. On converge alors vers un vrai minimum.

Figure 10.

Principe de la minimisation de l'énergie interne moléculaire à l'aide des algorithmes d'ordre 2 : Newton-Raphson et méthodes apparentées (64).

La fonction F décrivant l'hypersurface énergétique est dans ce cas représentée en 3 dimensions pour visualiser sa courbure.

Le gradient de F au point 1 va indiquer une direction de recherche (1->4) parallèle à un axe de coordonnées le long de la courbure de la surface que l'on va suivre à pas successifs jusqu'à la rencontre de la valeur minimale de F , atteinte ici au point 4. Les dérivées secondes s'y annulent et changent de signe, et le gradient de E en 4 va générer une seconde direction de recherche (4->7) orthogonale à la précédente. Il en va ainsi jusqu'à ce que l'énergie de deux étapes successives soit identique, un minimum, le plus souvent relatif, étant alors atteint.



Les désavantages de cette méthode résident dans le fait qu'une matrice hessienne demande un temps de calcul très long. Si la structure est loin du minimum, la minimisation peut devenir instable et conduire à un point encore plus loin du minimum. C'est le cas quand les forces sont grandes et la courbure de la surface petite.

Donc Newton-Raphson et les méthodes apparentées sont réservées pour les cas de convergence rapide avec un minimum bien défini et un écart moyen ou « root mean square » (rms) inférieur à $0,1 \text{ Kcal.mol}^{-1}.\text{Å}^{-1}$.

4.1.3- Un exemple d'algorithme : celui du logiciel Discover (64).

Il fait intervenir successivement steepest descent, conjugate gradient, et quasi Newton-Raphson va09.

Ce dernier est une variante de Newton-Raphson qui consiste à calculer la matrice hessienne numériquement en utilisant seulement l'information des dérivées premières plutôt que l'analyse par les dérivées secondes. L'évaluation de la fonction d'énergie est plus rapide, mais puisque l'on garde une matrice hessienne la taille mémoire demeure une limitation.

5- Méthodes de la dynamique moléculaire (70 à 73).

5.1- Objectifs.

Ces méthodes, par la simulation des mouvements intramoléculaires sur une échelle de temps allant de 10^{-15} à 10^{-10} s, permettent :

- d'optimiser les structures de façon plus efficace qu'en mécanique moléculaire afin d'éviter les minima multiples ;
- de décrire thermodynamiquement le système étudié (entropie, énergie libre, chaleur spécifique...).

5.2- Principes.

En mécanique classique, chaque atome est assimilé à une masse ponctuelle. Il est soumis à des forces dues aux interactions avec les atomes environnants (énergie interne totale) et à l'agitation thermique ($E_c = \frac{1}{2} k_B T_0$ par degré de liberté, avec k_B constante de Boltzmann et T_0 température absolue) qui vont engendrer un mouvement.

L'évolution spatio-temporelle ou trajectoire des atomes va être décrite par la résolution des équations de mouvement de Newton.

N.B. : seul le cas d'une dynamique moléculaire à température constante est ici envisagé.

1) Selon la loi de Newton, au temps t , on a : $\mathbf{F}_t = m\mathbf{a}_t = -(\mathbf{dE})_t$ (1) ($a = dv/dt = d^2r/dt^2$) avec F force, m masse, a accélération, v vitesse, r coordonnées et E énergie potentielle totale.

2) De (1), on extrait : $\mathbf{r}_t = -1/m \iint (\mathbf{dE})_t dt^2$ (2).

3) L'intégration de (2) se fait par approximation en se servant de l'algorithme de Verlet qui revient à un développement en série de Taylor limité à l'ordre 2.

Pour simuler le mouvement, on évalue les nouvelles coordonnées des atomes après un court intervalle de temps Δt .

a) Soit v la vitesse moyenne pendant Δt , on a alors : $\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\Delta t$ (3).

v peut être approximée par la vitesse exacte au milieu de l'intervalle : $\mathbf{v} \cong \mathbf{r}'(t + \Delta t/2)$.

b) Soit a l'accélération moyenne pendant l'intervalle $-\Delta t/2$ à $+\Delta t/2$, on a donc :

$$\mathbf{v} = \mathbf{r}'(t + \Delta t/2) = \mathbf{r}'(t - \Delta t/2) + \mathbf{a}\Delta t \quad (4).$$

a peut être approximée comme étant l'accélération exacte au milieu de l'intervalle $-\Delta t/2$ à $+\Delta t/2$: $\mathbf{a} \cong \mathbf{r}''(t)$.

c) On obtient donc finalement : $\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{r}'(t - \Delta t/2).\Delta t + \mathbf{r}''(t).\Delta t^2$.

A chaque pas d'une simulation de dynamique moléculaire, on effectue par conséquent successivement les opérations suivantes :

- calcul de l'accélération au temps t par (1) avec une itération sur tous les termes de la fonctionnelle d'énergie conformationnelle ;
- calcul de la vitesse au temps $t + \Delta t/2$ à partir de sa valeur au temps $t - \Delta t/2$ par (4) ;
- mise à jour des coordonnées au temps $t + \Delta t$ à partir de leurs valeurs au temps t par (3).

Les vitesses initiales sont assignées au hasard par une distribution de Maxwell-Boltzmann.

Le pas Δt doit être au moins 10 fois inférieur aux fréquences de vibration les plus élevées (pour des molécules organiques, ce sont les vibrations de valence C-H à 3600 cm^{-1}) pour que l'on puisse considérer les forces interatomiques comme constantes sur l'intervalle d'intégration. Il est de l'ordre de la femtoseconde (10^{-15} s).

On limite habituellement les temps de simulation à 20-30 picosecondes, soit 20000 à 30000 pas, pour les protéines en raison d'un temps de calcul élevé.

L'énergie totale du système doit être conservée : $E_{\text{totale}} = E_{\text{potentielle}} + E_{\text{cinétique}} = \text{Constante}$. Si l'énergie potentielle diminue, alors l'énergie cinétique et la température devraient augmenter. Par conséquent, pour maintenir la température constante, de l'énergie cinétique (déterminée par simulation à 300°K ou à 600°K) est redistribuée en énergie potentielle par des artifices mathématiques qui introduisent un facteur de relaxation τ .

En dynamique moléculaire, on va générer un grand nombre de conformations, environ 10^4 , au voisinage d'un minimum. Il faut alors minimiser par Newton-Raphson ou une méthode apparentée celles dont les énergies sont les plus basses.

5.3- Avantages.

La dynamique moléculaire permet :

- d'obtenir des conformations qui dépendent moins de la structure initiale que par mécanique moléculaire ;

- de franchir des maxima d'énergie et d'explorer d'autres conformations stables, mais cette recherche est limitée par le temps de simulation ;
- d'étudier des propriétés dépendantes du temps comme la diffusion et le repliement moléculaire.

6- Principaux logiciels de modélisation moléculaire : caractéristiques et applications (Tableau 4).

MECANIQUE MOLECULAIRE ET DYNAMIQUE MOLECULAIRE		
Logiciels	Champs de force (auteurs et dates)	Applications
BIOGRAF	DREIDING-II (Goddard, 1990)	Peptides, protéines et acides nucléiques
INSIGHT/ DISCOVER	CVFF (Hagler, 1985)	Peptides, protéines et acides nucléiques
MACROMODEL	AMBER (Kollman, 1981)	Peptides, protéines et acides nucléiques
MODEL	MM2 (Allinger, 1976)	Hydrocarbures Fonctions simples avec hétéroatomes
QUANTA	CHARMM (Karplus, 1983)	Produits organiques et biologiques, macromolécules Interactions entre petites molécules et macromolécules (ancrage moléculaire) Polymères et matériaux
SYBYL	MAXIMIN2 (Clark, 1988)	Produits organiques et biologiques
MECANIQUE QUANTIQUE SEMI-EMPIRIQUE		
Logiciels	Programmes	Applications
AMPAC	AM1 MNDO MINDO/3	Réactions chimiques et états de transition Paramètres de champs de force : charge atomique et barrière de rotation
MOPAC	AM1 MNDO MINDO	Réactions chimiques impliquant des molécules, des ions et des polymères linéaires

PARTIE EXPÉRIMENTALE

1- Oses et dérivés d'oses substitués utilisés.

Vingt-huit oses et dérivés d'oses diversement substitués - le glucose et ses dérivés, la glucosamine et la *N*-acétylglucosamine, et le mannose et son dérivé, la mannosamine, mono- ou di- phosphorylés ou sulfatés, ou *O*-méthylés - de provenance commerciale (Sigma Chemical Co., USA), ont été utilisés. Le galactose et ses dérivés, la galactosamine et la *N*-acétylgalactosamine, pour lesquels les lectines étudiées n'ont aucune affinité, ont été choisis comme témoins négatifs (Tableau 5).

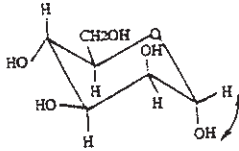
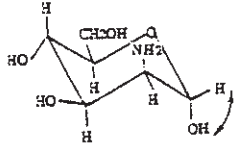
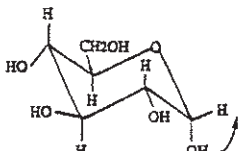
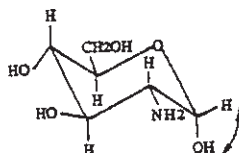
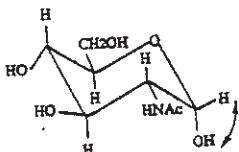
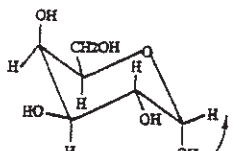
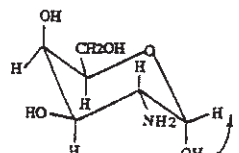
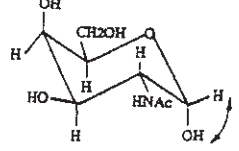
2- Isolement des lectines et des isolectines.

Les cinq lectines et isolectines de Légumineuses utilisées - les isolectines LoLI et LoLII de *Lathyrus ochrus*, PsA du Pois, LcA de la Lentille et la Con A de *Canavalia ensiformis* - sont toutes Glc/Man-spécifiques.

Elles ont été isolées et purifiées par chromatographie d'affinité sur gel de Sephadex G100 (Pharmacia), à l'exception de la Con A qui est ici de provenance commerciale (Sigma Chemical Co., Paris). Ce support, constitué de dextran - polymère de glucose liés en $\alpha(1\rightarrow6)$, fixe par ses résidus terminaux les lectines Glc/Man-spécifiques, excluant donc les autres protéines. Il suffit d'ajouter un excès de glucose au tampon d'élution pour recueillir les lectines préalablement retenues (74). Le protocole d'isolement est le même pour toutes ces lectines.

Deux cent cinquante grammes (250 g) de farine de graines sont mises à macérer dans un excès (1 litre) de tampon Tris-HCl 50 mM, NaCl 150 mM (pH 7,6) (TBS) pendant une nuit à 4°C. Le surnageant est récupéré par centrifugation (5000 g, 20 min), les débris insolubles étant éliminés. Il est ensuite amené à 90% de saturation en $(\text{NH}_4)_2\text{SO}_4$, puis maintenu sous agitation à 4°C durant 4 h. Les protéines précipitées, récupérées par centrifugation (20000 g, 30 min), sont reprises par un minimum de tampon TBS, puis dialysées contre du TBS durant 48 h. Après centrifugation, le surnageant est filtré sur une colonne (70 x 5 cm) de Sephadex G100 équilibré en tampon TBS. Les protéines non retenues sont éluées, puis les lectines retenues sont

Tableau 5. Oses et dérivés d'oses substitués utilisés.

Oses et dérivés d'oses	Formule (forme chaise)	Oses et dérivés d'oses substitués
Mannose		α -méthyl-mannopyranoside Mannose-1PO ₃ ⁻ Mannose-6PO ₃ ⁻ Mannose-6SO ₃ ⁻
Mannosamine		
Glucose		α -méthyl-glucopyranoside Glucose-1PO ₃ ⁻ Glucose-6PO ₃ ⁻ Glucose-1,6diPO ₃ ⁻ Glucose-6SO ₃ ⁻
Glucosamine		Glucosamine-1PO ₃ ⁻ Glucosamine-6PO ₃ ⁻ Glucosamine-2SO ₃ ⁻ Glucosamine-3SO ₃ ⁻ Glucosamine-6SO ₃ ⁻ Glucosamine-2,3diSO ₃ ⁻ Glucosamine-2,6diSO ₃ ⁻
N-acétylglucosamine		N-acétylglucosamine-1PO ₃ ⁻ N-acétylglucosamine-6PO ₃ ⁻ N-acétylglucosamine-3SO ₃ ⁻ N-acétylglucosamine-6SO ₃ ⁻
Galactose		
Galactosamine		
N-acétylgalactosamine		

précipitées par ajout de $(\text{NH}_4)_2\text{SO}_4$ à 90% de saturation et le précipité, récupéré par centrifugation (20000 g, 30 min), est repris par un minimum d'eau distillée, puis dialysé contre de l'eau distillée durant 48 heures et ensuite lyophilisé.

Les rendements sont de l'ordre de 300 à 400 mg de lectine purifiée par kg de graines.

Les deux isolectines LoLI et LoLII sont séparées par chromatofocalisation sur une colonne de PBE 94 (Pharmacia), en utilisant un gradient de pH compris entre 8,4 et 5,0.

Trente milligrammes (30 mg) de lectine totale en solution dans 2 ml d'eau distillée sont filtrés sur une colonne (30 x 1 cm) de PBE 96 (Pharmacia) équilibré à pH 8,9 par un tampon Tris-HCl de pH 8,4. L'élution est assurée par un mélange (30/70, v/v) de Polybuffers 96 et 94 dilué au 1/10^{ème} et ajusté à pH 5,0 avec de l'acide acétique 1,0 M, à un débit de 30 ml.min⁻¹. On réalise ainsi un gradient de pH compris entre 8,4 et 5,0 par l'intermédiaire duquel les isolectines LoLI (1^{er} pic d'élution, pH 7,2) et LoLII (2^{ème} pic d'élution, pH 6,0) sont parfaitement séparées. Les fractions correspondant aux deux pics d'élution sont amenées à 90% de saturation en $(\text{NH}_4)_2\text{SO}_4$, puis agitées pendant une nuit à 4°C. Les isolectines précipitées, récupérées par centrifugation (20000 g, 30 min), sont redissoutes dans 5 ml d'eau distillée, puis dialysées contre de l'eau distillée durant 48 h. Après centrifugation (20000 g, 30 min, 4°C), les surnageants recueillis sont lyophilisés.

La pureté des préparations de lectines et d'isolectines a été vérifiée par électrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) à l'aide du PhastSystem (Pharmacia), en utilisant des gels à 20% d'acrylamide. Les lectines/isolectines à deux chaînes (LoLI, LoLII, PsA et LcA) montrent les deux types de sous-unités (chaînes lourdes ou β et chaînes légères ou α) associées à de petites quantités de précurseur (chaînes $\alpha + \beta$ non dissociées).

3- Techniques d'hémagglutination.

La technique quantitative d'inhibition de l'hémagglutination d'hématies humaines a été utilisée pour étudier l'affinité des oses et de leurs dérivés, substitués ou non, vis-à-vis du site de reconnaissance des lectines et isolectines.

3.1- Hémagglutination quantitative.

L'activité hémagglutinante des solutions de lectines a été mesurée par la technique d'hémagglutination en godets (micro-plaques OMS de 96 godets à fond en U, Flow Laboratories).

La solution à titrer (lectine à 1 mg.ml^{-1} dans du tampon phosphate 50 mM (pH 7,4) (PBS)) est diluée selon une progression géométrique de raison 2 dans du PBS et distribuée sous un volume de 25 μl dans chaque godet. On y ajoute 25 μl d'une suspension d'hématies humaines (préalablement lavées trois fois dans du PBS) O Rh+ à 1% (v/v) dans du tampon PBS et, après agitation, la micro-plaque est abandonnée à température ambiante. L'hémagglutination est lue (hémagglutination en voile franc) au bout de 2 h puis, pour contrôle, 12 h plus tard.

L'activité hémagglutinante est exprimée par le titre d'hémagglutination, en unités hémagglutinantes (UH), qui est l'inverse de la dilution la plus élevée donnant encore une hémagglutination en voile franc. Les résultats correspondent à la moyenne de deux titrages indépendants, effectués dans les mêmes conditions et avec les mêmes hématies.

3.2- Inhibition de l'hémagglutination.

L'activité inhibitrice des oses et des dérivés d'oses à l'égard des lectines a été mesurée par la technique de l'inhibition de l'hémagglutination ou « haptén-inhibition technic ».

Les 12 godets d'une même rangée reçoivent 25 μl d'une solution d'ose ou de dérivé d'ose de concentration connue (mM) diluée suivant une progression géométrique de

raison 2 dans du tampon PBS. On y ajoute 25 μ l de la solution de lectine à analyser, diluée de façon à correspondre à 4 UH (1 UH correspond au titre de la solution de lectine déterminé par hémagglutination quantitative, 4 UH correspondent par conséquent à la dilution située 2 godets avant), puis on agite et on laisse reposer à température ambiante durant 30 min. On ajoute ensuite 25 μ l d'une suspension d'hématies humaines (préalablement lavées trois fois dans du PBS) O Rh+ à 1% (v/v) dans du tampon PBS et, après agitation, la micro-plaque est abandonnée à température ambiante. La lecture se fait au bout de 2 h et, pour contrôle, 12 h plus tard.

L'activité inhibitrice est exprimée par la concentration (mM) la plus basse d'ose ou de dérivé d'ose pour laquelle l'hémagglutination demeure complètement inhibée. Pour le calcul des concentrations, on tient compte de la dilution au 1/3 qui résulte des ajouts de la solution de lectine et de la suspension d'hématies. Tous les résultats correspondent à la moyenne de deux déterminations indépendantes effectuées dans des conditions identiques.

4- Modélisation moléculaire.

La modélisation moléculaire proprement dite de protéines de structure tridimensionnelle inconnue requiert généralement plusieurs opérations préliminaires.

4.1- Alignement et comparaison des séquences.

Ils ont été réalisés par l'intermédiaire du logiciel Ialign (PIR-NBRF, Washington, USA) sur une station MicroVax 3100.

Quand on compare deux séquences, afin d'obtenir le maximum d'identité, on est amené à décaler l'une des séquences par rapport à l'autre ; on introduit ainsi des blancs ou « gaps » qui correspondent souvent au niveau génétique à des insertions

ou des délétions d'acides aminés. Ils se traduisent par des pénalités dans le score d'identité (résidus identiques aux mêmes positions) ou d'homologie (résidus de structure ou de rôle voisins aux mêmes positions).

Pour y pallier, ce programme fait appel à la matrice de Dayhoff (75) qui prend en compte la fréquence statistique de remplacement d'un acide aminé par un autre dans un groupe de protéines apparentées au cours de l'évolution pour opérer une substitution. Il s'appuie pour cela sur des arbres phylogénétiques construits à partir de protéines fossiles, pour déduire de la distance génétique entre deux séquences (exprimée en unités PAM ou Point Accepted Mutation = nombre de mutations sur 100 résidus acceptées en 10^8 années) le temps évolutif.

Les valeurs standards (biais = 6 et pénalité = 6) ont été employées pour la création des gaps et leur extension, et la matrice de Dayhoff a été paramétrée à 500 PAMs.

4.2- Profils hydrophatiques.

Les méthodes utilisées pour prédire l'emplacement des régions exposées - correspondant à des épitopes ou déterminants antigéniques possibles - à la surface des chaînes polypeptidiques, sont basées sur le caractère hydrophile ou hydrophobe des acides aminés constitutifs.

Elles reposent sur des algorithmes qui opèrent de la manière suivante :

- 1) On affecte à chacun des 20 acides aminés protéiques un paramètre dont la valeur est directement ou indirectement liée à son caractère hydrophile ou hydrophobe.
- 2) Pour chaque acide aminé i d'une chaîne polypeptidique de longueur L , on calcule une valeur moyenne de l'hydrophilie sur n résidus soit :
 - à partir de celles des acides aminés $i - [(n - 1)/2]$, situés en amont - du côté N-terminal de la chaîne - à $i + [(n - 1)/2]$, situés en aval - du côté C-terminal - de l'acide aminé considéré ($n = 7, 9, 11$ ou 13). Le calcul débute par conséquent au niveau de l'acide aminé numéro $[(n - 1)/2]$ de la chaîne et s'arrête à l'acide aminé $L - (n - 1)/2$. Les valeurs des $(n - 1)/2$ premiers et des $(n - 1)/2$ derniers acides aminés

ne sont donc pas calculées. Des artifices mathématiques permettent toutefois de les déterminer. C'est la méthode de Kyte et Doolittle (76).

- sur celles des acides aminés i à $i + (n - 1)$ ($6 \leq n \leq 13$) ; c'est la méthode d'Hopp et de Woods (77).

Ces deux méthodes se distinguent également entre elles par l'origine et la valeur des paramètres d'hydrophilie/hydrophobie.

3) Chacune de ces valeurs moyennes est reportée sur un graphique où figurent en abscisses la séquence des acides aminés de la chaîne polypeptidique et en ordonnées les valeurs d'hydrophilie/hydrophobie correspondantes. On obtient ainsi un profil hydrophatique qui se traduit par des pics plus ou moins prononcés et étendus, situés au-dessus ou en-dessous d'une droite représentant la valeur moyenne calculée de l'hydrophilie, définissant ainsi respectivement les régions exposées (en surface) et incluses (enfouies) de la protéine.

Les profils hydrophatiques ont été établis par ces deux méthodes, en prenant $n = 11$, à l'aide du logiciel DNA Strider (Gif-sur-Yvette) sur Macintosh LC.

4.3- Profils HCA ou « Hydrophobic Cluster Analysis » (78).

Dans cette méthode, les séquences protéiques sont représentées en deux dimensions par projection de leur structure 3D sur un plan et on recherche les amas ou « clusters » constitués par des résidus hydrophobes contigus. En fonction de la taille et surtout de la forme de ces clusters, il est possible de prédire l'emplacement des principaux éléments de structure secondaire : les hélices α et les feuillets β (79). Les coudes et les boucles correspondent pour leur part à des zones où prédominent les résidus hydrophiles et flexibles.

Cette méthode permet également de reconnaître et de délimiter les régions structurellement conservées (RSCs) lorsqu'on compare deux séquences protéiques et a ainsi été utilisée pour préparer le travail de modélisation proprement dite de lectines de structures inconnues (LcicL de *Lathyrus cicera* et LniLII isolectine II de *L. nissolia*).

Les profils HCA ont été établis grâce au logiciel HCA-plot-V2 (Doriane, Paris) sur Macintosh LC.

4.4- Modélisation moléculaire des lectines.

Les lectines LcicL, LniLII, et LcA de la Lentille ont été modélisées à partir de leurs séquences primaires, en leur attribuant des coordonnées provenant de l'analyse par radiocristallographie de lectines apparentées : l'isolectine I (LoLI) de *L. ochrus* (22) et la lectine de Pois (PsA) (23).

Les modélisations ont été réalisées par l'intermédiaire des logiciels Insight II, Discover et Homology (Biosym Technologies, San Diego CA, USA) sur station Silicon Graphics.

Dans Homology, à partir de l'alignement des séquences primaires des lectines de structures connues et en s'aidant des résultats livrés par la méthode HCA, on définit des RSCs, qui correspondent ici surtout à des feuillets β . On aligne ensuite les séquences de la structure inconnue et de la structure modèle ainsi obtenue, et, en fonction de l'identité et de l'homologie, on repère les RSCs le long de cette dernière séquence. Il suffit, pour chaque RSC, d'assigner à la séquence de structure inconnue les coordonnées de la structure modèle pour construire de proche en proche la structure 3D de la protéine.

Pour certaines boucles ou certains coudes dont il n'existe pas d'équivalent structural dans la protéine modèle, il est nécessaire de rechercher dans une banque de données structurales (PDB, Brookhaven, New York, USA) puis d'importer et d'insérer l'élément le plus ressemblant, et le mieux adapté aux RSCs adjacentes.

On minimise finalement la structure par dynamique moléculaire (minimisation des boucles et des jonctions entre les RSCs), à température et volume constants, dans Discover, en appliquant le champ de force CVFF (cf. II-3.2.4) et en faisant appel aux algorithmes de minimisation d'énergie "de la plus grande pente", gradient conjugué et quasi Newton-Raphson va09 (cf. II-4.1.3).

Les calculs des charges atomiques nettes ont été effectués uniquement sur les résidus constituant le site lectinique par mécanique quantique en choisissant la méthode semi-empirique MNDO (Modified Neglect of Diatomic Overlap) dans le logiciel Pro-Quantum (Oxford Molecular, Palaiseau) sur station Silicon Graphics.

4.5- Construction des modèles d'oses substitués.

Les oses et dérivés d'oses substitués ont été modélisés à partir de la structure de l' α -D-glucopyranose (conformation de type chaise) et de groupements fonctionnels de la bibliothèque du module *Builder* du logiciel Insight II, et minimisés comme précédemment.

Les calculs des charges atomiques nettes ont été réalisés sur l'intégralité de la structure par MNDO.

4.6- Ancrage moléculaire ou « docking ».

L'étude de plusieurs complexes LoLI-glucides (cf. I-2) correspondant à des oses simples (glucose et mannose) (33), à des oligosaccharides (tri- et octasaccharide de type N-acétyllactosaminique) (34, 35), à des dérivés de l'acide muramique (37) et à un glycopeptide de 18 kDa dérivé de la lactotransferrine humaine (38), montre que, dans tous les cas, l'interaction repose sur l'insertion d'un ose ou d'un dérivé d'ose dans le site de reconnaissance. Dans cette interaction, le cycle pyranique de l'ose est toujours orienté de la même façon et le réseau de liaisons hydrogène qui le relie au site demeure inchangé.

Il est donc possible de mettre à profit ce mécanisme univoque d'insertion des oses et des dérivés d'oses dans le site lectinique pour réaliser l'ancrage moléculaire, à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille) sur station Silicon Graphics, des oses et dérivés d'oses substitués utilisés : il suffit de superposer graphiquement le cycle pyranique du dérivé à étudier et celui du modèle (mannose complexé à LoLI), d'effacer ensuite l'ose (mannose) et d'analyser les éventuelles

modifications introduites dans le réseau de liaisons hydrogène et la potentielle création d'encombrements stériques au niveau des substituants par la mesure de distances interatomiques et par la visualisation de surfaces de Van der Waals.

RÉSULTATS

Nous envisagerons successivement les résultats tirés des réactions d'inhibition de l'hémagglutination puis ceux obtenus par modélisation et ancrage moléculaire.

1- Affinité des oses et dérivés d'oses substitués pour les lectines.

Les réactions d'inhibition de l'hémagglutination des lectines/isolectines glucose/mannose-spécifiques utilisées (LoLI, LoLII, PsA, LcA et Con A) par les différents oses et dérivés d'oses substitués conduisent aux résultats suivants (Tableau 6) :

1) Toutes ces lectines interagissent fortement avec le glucose et le mannose. Les oses et dérivés d'oses aminés (glucosamine, mannosamine et *N*-acétylglucosamine) ont en quasi général des affinités plus faibles pour ces lectines. Par contre, le galactose et ses dérivés (galactosamine et *N*-acétylgalactosamine) n'ont aucune action inhibitrice sur ce type de lectines. Ces résultats sont en accord avec la spécificité glucose/mannose des lectines étudiées (80).

2) Les dérivés *O*-méthylés en C1 du mannose et du glucose se classent parmi les plus inhibiteurs pour ces lectines. La présence d'un groupe méthyle hydrophobe susceptible d'interagir avec des résidus hydrophobes situés au voisinage du site de fixation des glucides pourrait expliquer cette meilleure affinité (80).

3) Les dérivés phosphorylés et sulfatés du mannose, du glucose, de la glucosamine et de la *N*-acétylglucosamine ont un comportement différent selon la position des substituants :

- la phosphorylation ou la sulfatation de l'hydroxyle en C6 du cycle pyranique de l'ose ou du dérivé d'ose entraînent une perte totale de l'affinité : ainsi, le mannose-6PO₃⁻ et le mannose-6SO₃⁻, le glucose-6PO₃⁻ et le glucose-6SO₃⁻, la glucosamine-6PO₃⁻ et la glucosamine-6SO₃⁻, la *N*-acétylglucosamine-6PO₃⁻ et la *N*-acétylglucosamine-6SO₃⁻, sont dépourvus de toute activité inhibitrice

Tableau 6. Inhibition des lectines/isolectines de *Lathyrus ochrus* (LoLI, LoLII), de Pois (PsA), de Lentille (LcA) et de *Canavalia ensiformis* (Con A) par des oses et dérivés d'oses diversement substitués.

Les concentrations minimales inhibitrices correspondent aux plus faibles concentrations d'oses ou de dérivés d'oses substitués qui inhibent complètement 4 unités hémagglutinantes (4 UH) de lectine/isolectine. Les résultats correspondent à la moyenne de deux séries de déterminations indépendantes.

Oses et dérivés d'oses substitués	Concentration minimale (mM) inhibant 4 UH de :				
	LoLI	LoLII	PsA	LcA	ConA
Mannose	1,560	0,780	0,780	1,560	0,780
α -méthyl-mannopyranoside	0,780	0,780	0,195	0,390	0,098
Mannose-1PO ₃ ⁻	12,500	12,500	12,500	25,000	12,500
Mannose-6PO ₃ ⁻	>100	>100	>100	>100	>100
Mannose-6SO ₃ ⁻	>100	>100	>100	>100	>100
Mannosamine	6,250	12,500	12,500	6,250	12,500
Glucose	3,125	3,125	3,125	3,125	6,250
α -méthyl-glucopyranoside	0,780	1,560	0,780	0,780	0,780
Glucose-1PO ₃ ⁻	12,500	25,000	12,500	25,000	12,500
Glucose-6PO ₃ ⁻	>100	>100	>100	>100	>100
Glucose-1,6diPO ₃ ⁻	>100	>100	>100	>100	>100
Glucose-6SO ₃ ⁻	>100	>100	>100	>100	>100
Glucosamine	6,250	12,500	6,250	6,250	12,500
Glucosamine-1PO ₃ ⁻	6,250	6,250	6,250	6,250	3,125
Glucosamine-6PO ₃ ⁻	>100	>100	>100	>100	>100
Glucosamine-2SO ₃ ⁻	0,780	0,780	0,390	0,780	0,780
Glucosamine-3SO ₃ ⁻	1,560	1,560	1,560	1,560	3,125
Glucosamine-6SO ₃ ⁻	>100	>100	>100	>100	>100
Glucosamine-2,3diSO ₃ ⁻	1,560	1,560	1,560	1,560	1,560
Glucosamine-2,6diSO ₃ ⁻	>100	>100	>100	>100	>100
N-acétylglucosamine	3,125	3,125	3,125	3,125	3,125
N-acétylglucosamine-1PO ₃ ⁻	1,560	3,125	3,125	1,560	3,125
N-acétylglucosamine-6PO ₃ ⁻	>100	>100	>100	>100	>100
N-acétylglucosamine-3SO ₃ ⁻	0,390	0,780	0,390	0,390	0,390
N-acétylglucosamine-6SO ₃ ⁻	>100	>100	>100	>100	>100
Galactose	>100	>100	>100	>100	>100
Galactosamine	>100	>100	>100	>100	>100
N-acétylgalactosamine	>100	>100	>100	>100	>100

(pas d'inhibition à une concentration supérieure à 100 mM). Il en va de même pour une disubstitution avec un groupe phosphate ou sulfate en C6, ce qui est le cas pour le glucose-1,6diPO₃⁻ et la glucosamine-2,6diSO₃⁻. Ces résultats soulignent l'importance de l'hydroxyle en C6 dans l'interaction avec le site lectinique.

- la phosphorylation ou la sulfatation des autres hydroxyles peuvent avoir comme conséquence une augmentation d'affinité (glucosamine-1PO₃⁻, glucosamine-2SO₃⁻, glucosamine-3SO₃⁻, glucosamine-2,3diSO₃⁻, N-acétylglucosamine-1PO₃⁻, N-acétylglucosamine-3SO₃⁻) ou, plus rarement, une diminution d'affinité (mannose-1PO₃⁻, glucose-1PO₃⁻).

2- Modélisation moléculaire.

L'interprétation des résultats reposant en partie sur une approche de modélisation moléculaire, il nous était indispensable d'évaluer la validité des modèles de lectines avant de pouvoir les utiliser. En effet, si la modélisation de protéines par homologie donne en général des résultats satisfaisants concernant le squelette α -carboné, il n'en va pas toujours ainsi de l'orientation des chaînes latérales, particulièrement lorsqu'il s'agit de modéliser un site réactionnel (enzymatique ou lectinique).

2.1- Validité des modèles moléculaires.

Il nous a été possible de vérifier la validité des modèles (LcicL, LnLII et LcA) en comparant la structure de la LcA modélisée à partir des coordonnées de LoLI et de PsA avec celle (LAL) obtenue récemment par radiocristallographie (30).

La superposition des deux squelettes α -carbonés est totale, avec un rms voisin de 0,2 Å (**Figure 11A**). Lorsqu'elle porte sur les résidus qui constituent, par homologie avec LoLI ou la Con A, le site de reconnaissance des glucides (Asp^{81A}, Gly^{99A}, Phe^{123A}, Asn^{125A}, Gly^{29B}, Ala^{30B} et Glu^{31B}), nous avons alors des rms compris entre 0,1 (carbones α) et 0,4 Å (cycles aromatiques) et nous remarquons que les chaînes

latérales des résidus du site sont toujours orientées de façon sensiblement identique (**Figure 11B**). Ces observations montrent une excellente similitude entre les sites de reconnaissance du modèle et de la lectine cristallisée.

Nous avons abouti à des résultats analogues lorsque nous avons comparé les deux autres lectines modélisées, Lc1cL et LnLII, aux modèles cristallographiques (LoLI et PsA) ayant servi de références.

Ces comparaisons illustrent la fiabilité de la modélisation par homologie lorsqu'on considère des protéines appartenant à une famille homogène.

2.2- L'interaction moléculaire lectine-ose (cf. I-2).

Seuls les atomes d'oxygène O3, O4, O5 et O6 de l'ose participent au réseau de liaisons hydrogène qui assurent en partie la fixation de l'ose dans le site lectinique. Par contre, O1 et O2, dirigés constamment vers le solvant, n'interviennent apparemment pas dans le mécanisme de reconnaissance.

Nous pouvons donc supposer que toute substitution des groupes hydroxyles impliqués dans la liaison crée un encombrement stérique qui doit, *a priori*, diminuer ou supprimer l'interaction. C'est certainement le cas de l'hydroxyle en C6 qui se trouve profondément engagé dans le site lors de l'interaction.

Les résultats donnés par la technique d'inhibition de l'hémagglutination sont en parfait accord avec cette hypothèse : ainsi, toute substitution de l'hydroxyle en C6 par un groupe phosphate ou par un groupe sulfate supprime l'interaction (concentration inhibitrice supérieure à 100 mM), quels que soient l'ose ou le dérivé d'ose utilisés.

Les résultats obtenus lorsque la substitution porte sur les hydroxyles en C1, en C2 ou en C3 sont plus délicats à interpréter et nécessitent une approche de modélisation et d'ancrage moléculaire.

Figure 11.

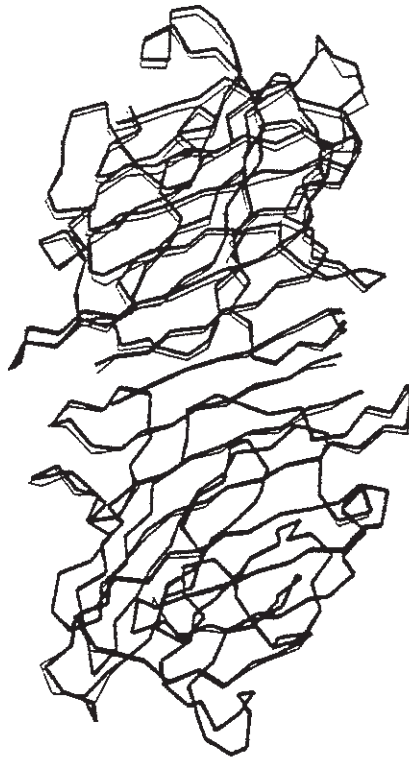
A- Superposition des modèles tridimensionnels de la lectine de Lentille LcA (trait fin) modélisée à partir de deux lectines voisines, PsA et LoLI, de coordonnées connues et de la lectine de Lentille LAL (trait épais) résolue par radiocristallographie à 2,3 Å par Loris *et al.* (30). Elle a été réalisée à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

L'écart moyen (rms) entre les deux chaînes α -carbonées est de 0,1 Å et nous n'observons aucune différence significative entre les deux modèles.

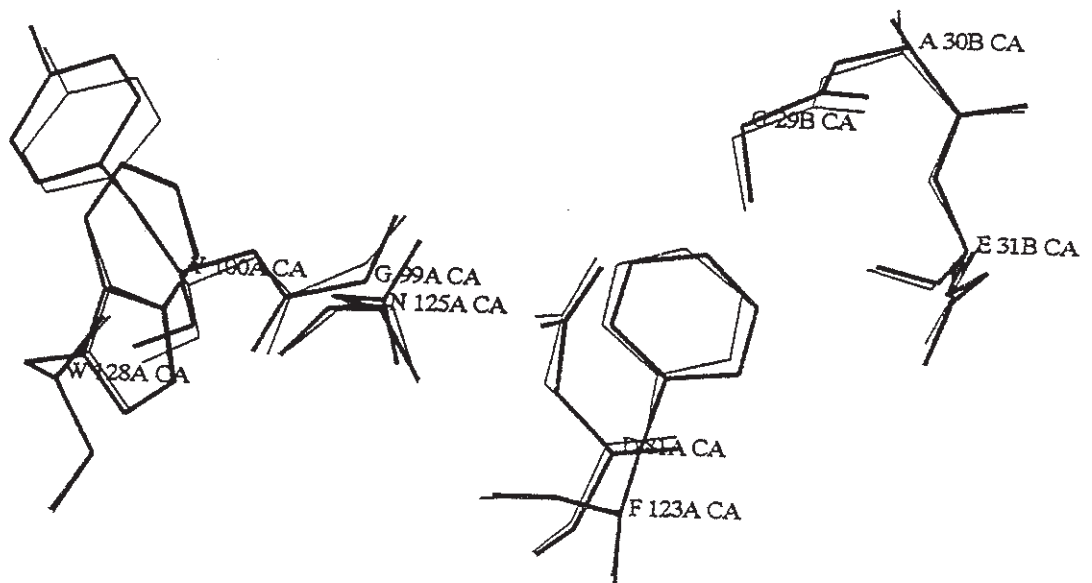
B- Superposition des sites de reconnaissance des monosaccharides de LcA (trait fin) et de LAL (trait épais), réalisée à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

La superposition a été effectuée uniquement sur neuf résidus : les sept résidus formant le site (Asp^{81A}, Gly^{99A}, Phe^{123A}, Asn^{125A}, Gly^{29B}, Ala^{30B} et Glu^{31B}) et deux résidus aromatiques (Tyr^{100A} et Trp^{128A}) situés au voisinage immédiat du site. Le rms sur ces neuf résidus est de 0,4 Å. L'orientation des chaînes latérales des neuf résidus est sensiblement identique, les variations les plus importantes s'observant avec les noyaux aromatiques des résidus Phe^{123A} et Tyr^{100A}.

A



B



2.3- Ancrage moléculaire.

2.3.1- Substituants en C6.

L'ancrage moléculaire des oses et dérivés d'oses O6- phosphatés ou sulfatés sur les sites des différentes lectines (LoLI, LciL, LnLII, LcA et PsA) nous permet d'observer une interpénétration des surfaces de Van der Waals du substituant et des résidus voisins du site (Asp⁸¹ et surtout les trois résidus Gly²⁰⁹, Ala²¹⁰ et Glu²¹¹ situés sur une même boucle) (**Figure 12**).

Les mesures de distances interatomiques nous indiquent l'existence de mauvais contacts sévères (jusqu'à 0,66 Å avec la glucosamine-6PO₃⁻) en particulier avec le résidu Gly²⁰⁹, mais également avec Asp⁸¹ et Phe¹²³ (**Tableaux 7 et 8**).

Cette gêne stérique, rencontrée également dans le cas d'une disubstitution (en 1,6 ou en 2,6), suffit à elle seule à expliquer l'absence d'affinité des oses et dérivés d'oses O6- phosphatés ou sulfatés pour les lectines (cf. les résultats des tests d'inhibition de l'héماغglutination, **Tableau 6**).

2.3.2- Substituants en C1 et en C2.

L'ancrage moléculaire ne nous montre dans ces cas aucun changement significatif dans le réseau de liaisons non covalentes qui relie l'ose ou le dérivé d'ose au site (**Figure 13A**), les hydroxyles en C1 et C2 étant dirigés vers le solvant.

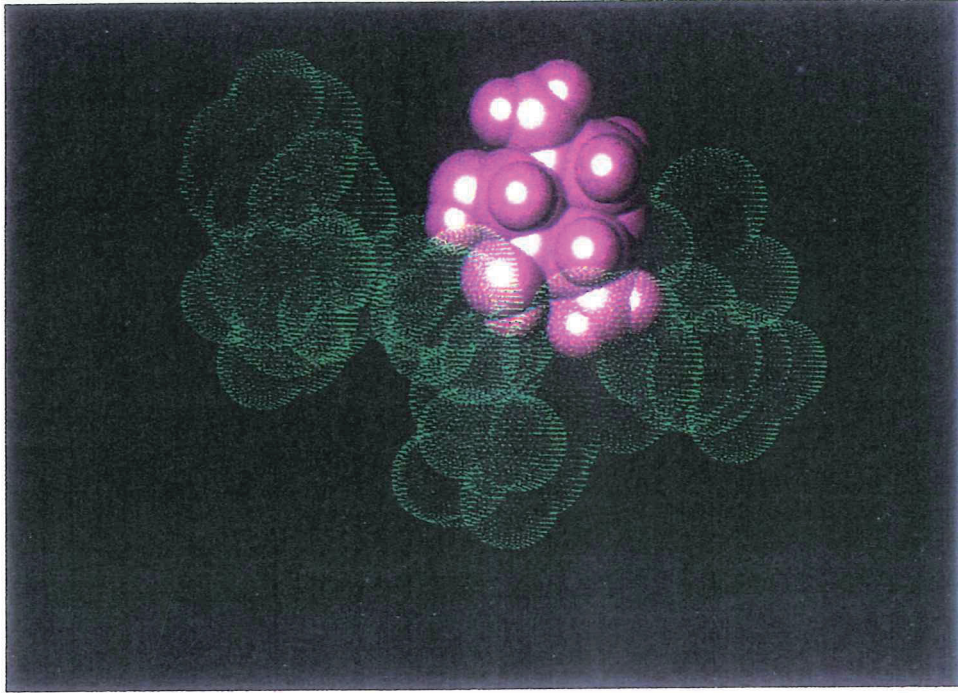
Les résultats des tests d'inhibition de l'héماغglutination nous indiquent que la substitution en C1 par un groupe phosphate (cas de la glucosamine-1PO₃⁻ et de la N-acétylglucosamine-1PO₃⁻) ne modifie quasiment pas l'affinité pour les lectines. Par contre, la même substitution du mannose (mannose-1PO₃⁻) et du glucose (glucose-1PO₃⁻) provoque une baisse importante de l'affinité : les concentrations minimales inhibitrices (CMI) sont en moyenne respectivement 14,2 et 5,2 fois moindres par rapport à celles des mêmes oses non substitués. Inversement, la substitution de la glucosamine en C2 par un groupe sulfate (glucosamine-2SO₃⁻) accroît grandement l'affinité : les CMI sont de 0,39/0,78 mM au lieu de 6,25/12,5 mM pour la glucosamine (facteur moyen d'accroissement = 13,2).

Figure 12.

Ancrage moléculaire (docking) de la glucosamine (A) et de la glucosamine-6SO₃⁻ (B) dans le site de reconnaissance des monosaccharides de LoLI, réalisé à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

La glucosamine et la glucosamine-6SO₃⁻ ont été représentées en CPK (en couleur mauve). Les résidus qui constituent le site lectinique (Asp⁸¹, Gly⁹⁹, Phe¹²³, Asn¹²⁵, Gly²⁰⁹, Ala²¹⁰ et Glu²¹¹) et deux résidus aromatiques (Tyr¹⁰⁰ et Trp¹²⁸, à gauche sur la photographie) situés au voisinage immédiat du site, ont été représentés par leurs surfaces de Van der Waals (en couleur verte). Ces deux derniers résidus n'interagissent pas avec la glucosamine et la glucosamine-6SO₃⁻, mais leur présence souligne la forme en cuvette du site.

A



B

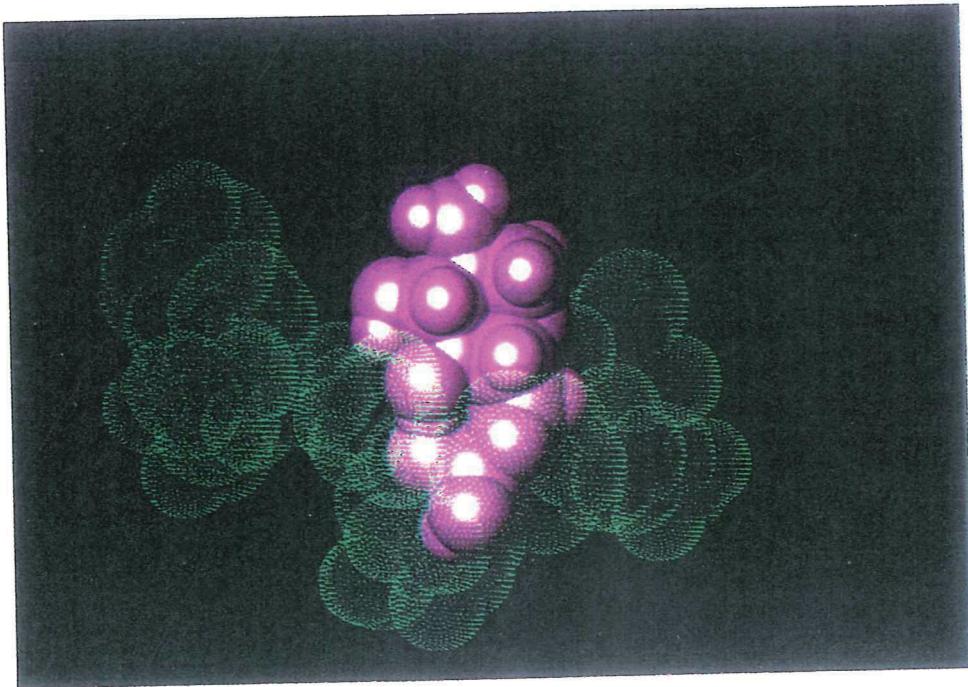


Tableau 7. Mauvais contacts entre le groupe sulfate de la glucosamine-6SO₃⁻ et les résidus du site de LoLI, déterminés par l'intermédiaire du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

Atomes du SO ₃ ⁻	Atomes du résidu	Résidu	Distance (Å)
H4	O	Glu ²¹¹	1,04
O3	N	Ala ²¹⁰	1,17
O4	O	Glu ²¹¹	1,58
O4	N	Gly ²⁰⁹	1,65
O2	N	Glu ²¹¹	1,68
O2	C α	Glu ²¹¹	1,70
O2	C β	Glu ²¹¹	1,76
O3	N	Glu ²¹¹	1,79
O3	C	Gly ²⁰⁹	1,84
H5	O δ 2	Asp ⁸¹	1,84
H4	N	Gly ²⁰⁹	1,96
O3	C α	Ala ²¹⁰	1,99
S1	N	Glu ²¹¹	2,04
O2	C	Glu ²¹¹	2,07
HO4	N	Gly ⁹⁹	2,07
O3	C	Ala ²¹⁰	2,10
O2	O	Glu ²¹¹	2,17
O1	O δ 1	Asp ⁸¹	2,52
O4	C α	Gly ²⁰⁹	2,56
S1	N	Ala ²¹⁰	2,62
O3	C α	Gly ²⁰⁹	2,68
S1	N	Gly ²⁰⁹	2,77
S1	C α	Glu ²¹¹	2,83
S1	C	Glu ²¹¹	2,87
S1	C α	Gly ²⁰⁹	2,99
S1	C	Ala ²¹⁰	3,06

Tableau 8. Mauvais contacts entre le groupe phosphate de la glucosamine-6PO₃⁻ et les résidus du site de LoLI, déterminés par l'intermédiaire du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

Atomes du PO ₃ ⁻	Atomes du résidu	Résidu	Distance (Å)
O3	Oδ2	Asp ⁸¹	0,66
O4	Cβ	Phe ¹²³	1,08
H4	Cβ	Phe ¹²³	1,52
O3	Cγ	Asp ⁸¹	1,57
O4	Cγ	Phe ¹²³	1,67
H5	Oδ2	Asp ⁸¹	1,83
HO4	N	Gly ⁹⁹	2,02
P1	Oδ2	Asp ⁸¹	2,04
O3	Oδ1	Asp ⁸¹	2,23
O2	Cγ	Asp ⁸¹	2,47
O4	Cδ1	Phe ¹²³	2,48
P1	Cγ	Asp ⁸¹	2,48
P1	Cγ	Phe ¹²³	2,50
O5	Cγ	Phe ¹²³	2,52
O5	Cδ1	Phe ¹²³	2,54
O2	Oδ1	Asp ⁸¹	2,56
P1	Cβ	Phe ¹²³	2,58
O4	Cα	Phe ¹²³	2,61
O4	Cδ2	Phe ¹²³	2,62
P1	Oδ1	Asp ⁸¹	2,84
P1	Cδ1	Phe ¹²³	2,86
P1	Cδ2	Phe ¹²³	3,18

Ces variations d'affinité sont donc difficilement explicables par ancrage moléculaire. Il n'existe aucune gêne stérique dans le cas d'oses et dérivés d'oses O1- ou O2-phosphatés ou sulfatés.

Les calculs des charges atomiques nettes révèlent des modifications notables, essentiellement au voisinage des substituants phosphate et sulfate, qui néanmoins ne nous permettent pas d'interpréter de manière satisfaisante ces variations d'affinité. De même, si l'augmentation de solubilité induite par des groupes sulfate ou phosphate fortement hydrophiles traduit assez bien l'accroissement d'affinité observée avec la glucosamine-2SO₃⁻, elle ne saurait expliquer les résultats contradictoires obtenus avec le mannose-1PO₃⁻ et le glucose-1PO₃⁻.

2.3.3- Substituants en C3.

La présence d'un groupe sulfate en C3 induit une augmentation sensible de l'affinité à l'égard des lectines qui est observée avec la glucosamine-3SO₃⁻, la glucosamine-2,3diSO₃⁻ et la N-acétylglucosamine-3SO₃⁻, leurs CMI étant en moyenne respectivement 4,8, 5,6 et 7,2 fois celles des mêmes dérivés d'oses non substitués (cf. **Tableau 6**).

Avec ces trois dérivés d'oses O3-sulfatés, l'ancrage moléculaire (**Figure 13B**) nous montre qu'une liaison hydrogène supplémentaire s'établit entre un atome d'oxygène du groupe sulfate et l'atome d'hydrogène de liaison peptidique -CO-NH- entre les résidus Gly⁹⁹ et Gly¹⁰⁰ du site, venant de ce fait renforcer l'interaction.

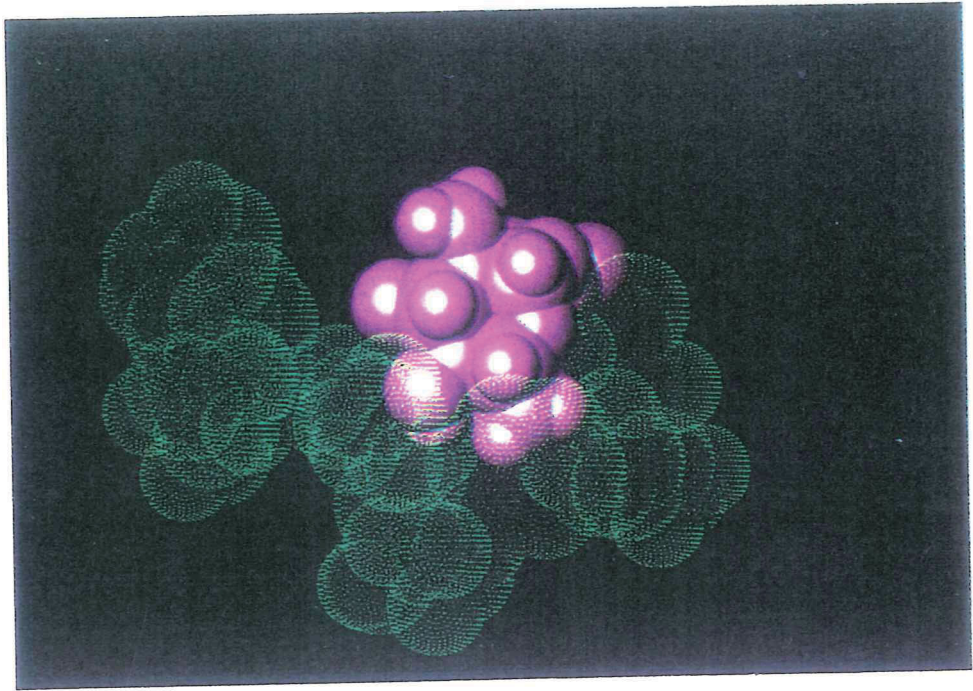
En outre, les calculs des charges atomiques nettes prouvent l'existence d'une forte charge positive sur l'atome de soufre du groupe sulfate de la glucosamine-3SO₃⁻ (+1,78 e), de la N-acétylglucosamine-3SO₃⁻ (+1,78 e), et de la glucosamine-2,3diSO₃⁻ (+1,82 e). Parallèlement, une charge négative se trouve sur les atomes d'azote des liaisons peptidiques unissant les résidus Gly⁹⁸ (-0,39 e), Gly⁹⁹ (-0,29 e) et Tyr¹⁰⁰ (-0,40 e). Les distances entre l'atome de soufre du groupe sulfate et ces atomes d'azote sont indiquées dans le **Tableau 9**.

Figure 13.

Ancrage moléculaire (docking) de la glucosamine-1SO₃⁻ (A) et de la glucosamine-3SO₃⁻ (B) dans le site de reconnaissance des monosaccharides de LoLI, réalisé à l'aide du logiciel TurboFrodo (Roussel et Cambillau, Marseille).

La glucosamine-1SO₃⁻ et la glucosamine-3SO₃⁻ ont été représentées en CPK (en couleur mauve). Les résidus qui constituent le site lectinique (Asp⁸¹, Gly⁹⁹, Phe¹²³, Asn¹²⁵, Gly²⁰⁹, Ala²¹⁰ et Glu²¹¹) et deux résidus aromatiques (Tyr¹⁰⁰ et Trp¹²⁸, à gauche sur la photographie) situés au voisinage immédiat du site, ont été représentés par leurs surfaces de Van der Waals (en couleur verte). Ces deux derniers résidus n'interagissent pas avec la glucosamine-1SO₃⁻ et la glucosamine-3SO₃⁻, mais leur présence souligne la forme en cuvette du site.

A



B

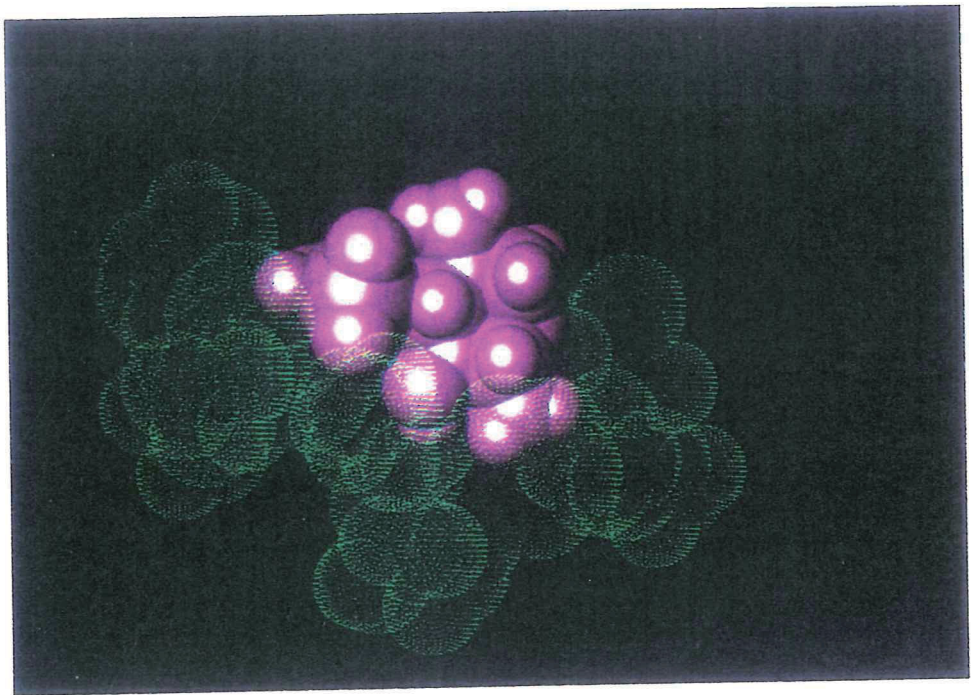


Tableau 9. Distances mesurées entre l'atome de soufre du groupe sulfate des dérivés d'oses O3-sulfatés utilisés et les atomes d'azote engagés dans les liaisons peptidiques entre les résidus Gly⁹⁸, Gly⁹⁹ et Tyr¹⁰⁰ du site de LoLI.

Dérivés d'oses substitués	Résidus	Distances S-NH-CO (Å)
Glucosamine-3SO ₃ ⁻	Gly ⁹⁸	5,49
	Gly ⁹⁹	3,05
	Tyr ¹⁰⁰	4,45
N-acétylglucosamine-3SO ₃ ⁻	Gly ⁹⁸	5,55
	Gly ⁹⁹	3,09
	Tyr ¹⁰⁰	4,49
Glucosamine-2,3diSO ₃ ⁻	Gly ⁹⁸	5,49
	Gly ⁹⁹	3,06
	Tyr ¹⁰⁰	4,46

Les distances entre l'atome de soufre du groupe sulfate et l'atome d'azote du résidu Gly⁹⁹, de l'ordre de 3 Å, sont tout à fait compatibles avec l'existence d'une interaction électrostatique entre ces deux atomes.

L'accroissement d'affinité des dérivés d'oses O3-sulfatés vis-à-vis des lectines pourrait donc être expliqué comme étant la résultante de trois phénomènes :

- 1) la création d'une liaison hydrogène supplémentaire, entre le groupe sulfate et le site lectinique ;
- 2) une forte attraction électrostatique entre l'atome de soufre du groupe sulfate et un atome d'azote impliqué dans une liaison peptidique entre deux résidus du site ;
- 3) une augmentation de l'hydrophilie due au groupe sulfate.

DISCUSSION ET CONCLUSION.

L'étude de l'interaction de divers oses (mannose et glucose) et dérivés d'oses (glucosamine et *N*-acétylglucosamine) substitués par des groupes phosphate ou sulfate avec des lectines de Légumineuses (LoLI et LoLII de *Lathyrus ochrus*, PsA du Pois, LcA de la Lentille et la ConA de *Canavalia ensiformis*) glucose/mannose-spécifiques, réalisée, d'une part, par inhibition quantitative de l'héماغglutination, et, d'autre part, par modélisation et ancrage moléculaire, montre une nette influence de la position du substituant sur la spécificité de reconnaissance :

- une substitution en C6 du cycle pyranique de l'ose entraîne toujours un encombrement stérique tellement important qu'il interdit toute insertion de ces dérivés dans le site lectinique. Leur affinité pour ces lectines devient quasi-nulle et comparable à celle d'oses et de dérivés d'oses non réactifs à l'égard de ce type de lectines (galactose, galactosamine et *N*-acétylgalactosamine). Dans ce cas, les résultats obtenus par modélisation et ancrage moléculaire corroborent et expliquent parfaitement les résultats expérimentaux tirés des tests d'inhibition de l'héماغglutination.
- lorsque la substitution concerne les atomes d'oxygène O1 et O2 des hydroxyles portés en C1 et C2, qui ne sont pas impliqués dans le réseau de liaisons hydrogène intervenant dans la fixation des oses et des dérivés d'oses au site, les résultats sont plus variables. Une substitution en C1 ne produit soit aucun changement, soit une diminution de l'affinité. Par contre, une substitution en C2 augmente l'affinité. Ces résultats ne sont pas interprétables en terme d'encombrement stérique. Bien que les calculs des charges atomiques nettes indiquent des différences marquées avec les mêmes oses et dérivés d'oses non substitués, leur corrélation avec les résultats observés est hypothétique.

- une substitution en C3 provoque un accroissement important de l'affinité, explicable par la conjonction de trois types d'interactions entre le dérivé d'ose substitué et le site : la création d'une liaison hydrogène supplémentaire, une attraction électrostatique et une augmentation de l'hydrophilie. Dans ce cas, les résultats obtenus par ancrage moléculaire et par l'application d'une méthode semi-empirique de calcul de charges rendent parfaitement compte des résultats expérimentaux.

Il n'a pas été possible d'étudier les éventuelles modifications d'affinité qu'induirait un substituant en C4, de tels dérivés n'étant pas disponibles dans le commerce.

L'ensemble de ces résultats suggère que dans l'hypothèse d'une reconnaissance *Rhizobium*-Légumineuse reposant sur une interaction facteurs Nod-lectines, la spécificité de reconnaissance ne serait pas dûe au groupe sulfate lié au C6 du résidu de *N*-acétylglucosamine situé à l'extrémité réductrice de ces lipo-oligosaccharides. Par contre, la présence d'un groupe sulfate en C2 ou en C3 pourrait jouer ce rôle, bien que de telles substitutions n'aient pas été rapportées dans les facteurs Nod isolés et caractérisés jusqu'à présent.

BIBLIOGRAPHIE

- (1) Suzuki, Y., Toda, Y., Tamatani, T., Watanabe, T., Suzuki, T., Nakao, T., Murase, K., Kiso, M., Hasegawa, A., Tadano-Aritomi, K., Ishizuka, I. and Miyasaka, M. (1993) *Biochem. Biophys. Res. Commun.* **190**, 426-434.
- (2) Imai, Y., Lasky, L.A. and Rosen, S.D. (1993) Sulphation requirement for GlyCAM-1, an endothelial ligand for L-selectin. *Nature* **361**, 555-557.
- (3) Lis, H. and Sharon, N. (1993) Protein glycosylation. *Eur. J. Biochem.* **218**, 1-27.
- (4) Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J.C. and Dénarié, J. (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781-784.
- (5) Roche, P., Lerouge, P., Ponthus, C. and Promé, J.C. (1991) Structural determination of bacterial nodulation factors involved in the *Rhizobium meliloti*-alfalfa symbiosis. *J. Biol. Chem.* **266**, 10933-10940.
- (6) Roche, P., Debellé, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Dénarié, J. and Promé, J.C. (1991) Molecular basis of symbiotic host specificity in *Rhizobium meliloti* : *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. *Cell* **67**, 1131-1143.
- (7) Schultze, M., Quiclet-Sire, B., Kondorosi, E., Virelizier, H., Glushka, J.N., Endre, G., Gero, S.D. and Kondorosi, A. (1992) *Rhizobium meliloti* produces a family of sulfated lipo-oligosaccharides exhibiting degrees of plant host specificity. *Proc. Natl. Acad. Sci. USA* **89**, 192-196.
- (8) Demont, N. (1993) Facteurs Nod de *Rhizobium meliloti* : rôle des gènes *nodFe* dans leur biosynthèse ; acides (ω -1) hydroxylés, nouveaux types de substituants, Thèse de Doctorat, Université de Toulouse III.

- (9) Zaat, S.A.J., Van Brussel, A.A.N., Tak, T., Pees, E. and Lugtenberg, B.J.J. (1987) Flavonoids induce *Rhizobium leguminosarum* to produce *nodABC* gene-related factors that cause thick, short roots and root hair responses on common vetch. *J. Bacteriol.* **139**, 3388-3391.
- (10) Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., De Billy, F., Promé, J.C. and Dénarié, J. (1991) Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. *Nature* **351**, 670-673.
- (11) Spaink, H.P., Sheeley, D.M., Van Brussel, A.A.N., Glushka, J., York, W.S., Tak, T., Geiger, O., Kennedy, E.P., Reinhold, V.N. and Lugtenberg, B.J.J. (1991) A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* **354**, 125-130.
- (12) Price, N.P.J., Relic, B., Talmont, F., Lewin, A., Promé, D., Pueppke, S.G., Maillet, F., Dénarié, J., Promé, J.C. and Broughton, W.J. (1992) Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are O-acetylated or sulphated. *Mol. Microbiol.* **6**, 3575-3584.
- (13) Bec-Ferté, M.P., Savagnac, A., Pueppke, S.G. and Promé, J.C. (1992) Nod factors from *Rhizobium fredii* USDA257. In *New Horizons in Nitrogen Fixation* (Palacios, R., Mora, J. and Newton, W.E., Eds.), pp. 157-159, Kluwer Academic Publishers, Dordrecht.
- (14) Sanjuan, J., Carlson, R.W., Spaink, H.P., Bhat, U.R., Barbour, W.M., Glushka, J. and Stacey, G. (1992) A 2-O-methylfucose moiety is present in the lipo-oligosaccharide modulation signal of *Bradyrhizobium japonicum*. *Proc. Natl. Acad. Sci. USA* **89**, 8789-8793.
- (15) Mergaert, P., Van Montagu, M., Promé, J.C. and Holsters, M. (1993) Three unusual modifications, a D-arabinosyl, an N-methyl and a carbamoyl group, are present on the Nod factors of *Azorhizobium caulonidans* strain ORS571. *Proc. Natl. Acad. Sci. USA* **90**, 1551-1555.

- (16) Poupot, R., Martinez-Romero, E. and Promé, J.C. (1993) Nodulation factors from *Rhizobium tropici* are sulfated or nonsulfated chitopentasaccharides containing an N-methyl-N-acetylglucosaminyl terminus. *Biochemistry* **32**, 10430-10435.
- (17) Lis, H. and Sharon, N. (1991) Lectin-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 741-749.
- (18) Kijne, I., Diaz, C., de Pater, S. and Lugtenberg, B. (1992) Lectins in the symbiosis between *Rhizobia* and Leguminous plants. In *Advances in Lectin Research* (Franz, H., Ed.), Vol. 5, pp. 15-50, Ullstein Mosby GmbH & Co., Berlin.
- (19) Diaz, C.L., Melchers, L.S., Hooykaas, P.J.J., Lugtenberg, B.J.J. and Kijne, J.W. (1989) Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. *Nature* **338**, 579-581.
- (20) Van Driessche, E. (1988) Structure and function of Leguminosæ lectins. In *Advances in Lectin Research* (Franz, H., Ed.), Vol. 1, pp. 73-134, Springer-Verlag, Berlin.
- (21) Diaz, C.L., Hosselet, M., Logman, G.J.J., Van Driessche, E., Lugtenberg, B.J.J. and Kijne, J.W. (1990) Distribution of glucose/mannose-specific isolectins in pea (*Pisum sativum* L.) seedlings. *Planta* **181**, 451-461.
- (22) Bourne, Y., Abergel, C., Cambillau, C., Frey, M., Rougé, P. and Fontecilla-Camps, J.C. (1990) X-ray crystal structure determination and refinement at 1.9 Å resolution of isolectin I from the seeds of *Lathyrus ochrus*. *J. Mol. Biol.* **214**, 571-584.
- (23) Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. and Suddath, F.L. (1986) The crystal structure of pea lectin at 3.0-Å resolution. *J. Biol. Chem.* **261**, 16518-16527.
- (24) Richardson, M., Yarwood, A. and Rougé, P. (1987) The amino acid sequence of an atypical single-chain lectin from seeds of *Lathyrus sphaericus* (Retz.). *FEBS Lett.* **216**, 145-150.

- (25) Yarwood, A., Richardson, M., Morphet, B., Westby, M., Père, D. and Rougé, P. (1988) The amino acid sequences of two atypical single-chain *Viciae* isolectins from seeds of *Lathyrus nissolia* L. *Phytochemistry* **27**, 1719-1721.
- (26) Hardman, K.D., Agarwal, R.C. and Freiser, M.J. (1982) Manganese and calcium binding sites of concanavalin A. *J. Mol. Biol.* **157**, 69-86.
- (27) Reeke, Jr., G.N. and Becker, J.W. (1986) Three-dimensional structure of favin : saccharide binding-cyclic permutation in leguminous lectins. *Science* **234**, 1108-1111.
- (28) Delbaere, L.T.J., Vandonselaar, M., Prasad, L., Quail, J.W., Pearlstone, J.R., Carpenter, M.R., Smillie, L.B., Nikrad, P.V., Spohr, U. and Lemieux, R.U. (1990) Molecular recognition of a human blood group determinant by a plant lectin. *Can. J. Chem.* **68**, 1116-1121.
- (29) Shaanan, B., Lis, H. and Sharon, N. (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science* **254**, 862-866.
- (30) Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. and Wyns, L. (1993) Crystal structure determination and refinement at 2.3-Å resolution of the lentil lectin. *Biochemistry* **32**, 8772-8781.
- (31) Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A. and Vijayan, M. (1994) Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc. Natl. Acad. Sci. USA* **91**, 227-231.
- (32) Derewenda, Z., Yariv, J., Helliwell, J.R., Kalb (Gilboa), A.J., Dodson, E.J., Papiz, M.Z., Wan, T. and Campbell, J. (1989) The structure of the saccharide-binding site of concanavalin A. *EMBO J.* **8**, 2189-2193.
- (33) Bourne, Y., Roussel, A., Frey, M., Rougé, P., Fontecilla-Camps, J.C. and Cambillau, C. (1990) Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose : fine specificity of the monosaccharide-binding site. *Proteins* **8**, 365-376.

- (34) Bourne, Y., Rougé, P. and Cambillau, C. (1990) X-ray structure of a (α -Man(1->3) β -Man(1->4)GlcNAc)-lectin complex at 2.1 Å resolution. The role of water in sugar-lectin interaction. *J. Biol. Chem.* **265**, 18161-18165.
- (35) Bourne, Y., Rougé, P. and Cambillau, C. (1992) X-ray structure of a biantennary octasaccharide-lectin complex refined at 2.3 Å resolution. *J. Biol. Chem.* **267**, 197-203.
- (36) Rini, J.M., Hardman, K.D., Einspahr, H., Suddath, F.L. and Carver, J.P. (1993) X-ray crystal structure of pea lectin-trimannoside complex at 2.6 Å. *J. Biol. Chem.* **268**, 10126-10132.
- (37) Bourne, Y., Ayouba, A., Rougé, P. and Cambillau, C. (1994) Interaction of a legume lectin with two components of the bacteria cell wall. A crystallographic study. *J. Biol. Chem.* **269**, 9429-9435.
- (38) Bourne, Y., Mazurier, J., Legrand, D., Rougé, P., Montreuil, J., Spik, G. and Cambillau, C. (1994) Interaction of a legume lectin with the human lactotransferrin N2 fragment or with the isolated biantennary glycopeptide : role of the fucose moiety. *Structure* **2**, 209-219.
- (39) Kishinevsky, B.D., Law, I.J. and Strijdom, B.W. (1988) Detection of lectins in nodulated peanut and soybean plants. *Planta* **176**, 10-18.
- (40) Dénarié, J., Debelle, F. and Rosenberg, C. (1992) Signaling and host range variation in nodulation. *Annu. Rev. Microbiol.* **46**, 497-531.
- (41) Long, S.R. (1989) *Rhizobium*-legume nodulation : life together in the underground. *Cell* **56**, 203-214.
- (42) Maxwell, C.A., Hartwig, U.A., Joseph, C.M. and Phillips, D.A. (1989) A chalcone and two related flavonoids released from alfalfa roots induce *nod* genes of *Rhizobium meliloti*. *Plant Physiol.* **91**, 842-847.

- (43) Henikoff, S., Haughn, G.W., Calvo, J.M. and Wallace, J.C. (1988) A large family of bacterial activator proteins. *Proc. Natl. Sci. USA* **85**, 6602-6606.
- (44) Györgypal, Z., Kiss, G.B. and Kondorosi, A. (1991) Transduction of plant signal molecules by the *Rhizobium* NodD proteins. *BioEssays* **13**, 575-581.
- (45) Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A. and Kondorosi, A. (1986) Conservation of extended promoter regions of nodulation genes in *Rhizobium*. *Proc. Natl. Acad. Sci. USA* **83**, 1757-1761.
- (46) Debellé, F. and Sharma, S.B. (1986) Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity of nodulation. *Nucleic Acids Res.* **14**, 7453-7472.
- (47) Zaat, S.A.J., Schripsema, J., Wijffelman, C.A., Van Brussel, A.A.N. and Lugtenberg, B.J.J. (1989) Analysis of the major inducers of the *Rhizobium nodA* promoter from *Vicia sativa* root exudate and their activity with different *nodD* genes. *Plant Mol. Biol.* **13**, 175-188.
- (48) Györgypal, Z., Kondorosi, E. and Kondorosi, A. (1991) Diverse signal sensitivity of NodD protein homologs from narrow and broad host range *Rhizobia*. *Mol. Plant-Microbe Interact.* **4**, 356-364.
- (49) Dusha, L., Bakos, A., Kondorosi, A., De Bruijn, F.J. and Schell, J. (1989) The *Rhizobium meliloti* early nodulation genes (*nodABC*) are nitrogen-regulated : isolation of a mutant strain with efficient nodulation capacity on alfalfa in the presence of ammonium. *Mol. Gen. Genet.* **219**, 89-96.
- (50) Kondorosi, E., Gyuris, J., Schmidt, J., John, M., Duda, E., Hoffman, B. and Schell, J. (1989) Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. *EMBO J.* **8**, 1331-1340.

- (51) Kondorosi, A. (1989) *Rhizobium*-legume interactions : nodulation genes. In *Plant-Microbe Interactions* (Kosuge, T. and Nester, E., Eds.), Vol. 3, pp. 383-420, McGraw-Hill, New York.
- (52) Schlaman, H.R.M., Okker, R.J.H. and Lugtenberg, B.J.J. (1990) Subcellular localization of the *Rhizobium leguminosarum nodI* gene product. *J. Bacteriol.* 171, 4686-4693.
- (53) Kondorosi, A. (1992) Regulation of nodulation genes in *Rhizobia*. In *Molecular Signals in Plant-Microbe Communications* (Verna, Eds.), pp. 325-340, CRC Press Boca Raton, London.
- (54) Dugas, H. (1994) Modélisation moléculaire, stéréochimie et analyse conformationnelle, Librairie de l'Université de Montréal.
- (55) Liotta, D., Ed. (1988, 1990). *Adv. in Molecular Modeling* 1 et 2, JAI edition.
- (56) Clark, T. (1985) *A Handbook of Computational Chemistry*, Wiley, New York.
- (57) Venkataraghavan, B. and Feldman, R.J., Eds. (1985) Macromolecular Structure and Activity : Computer-assisting Modeling and Applications. *Ann. NY Acad. Sci.* 539.
- (58) Beveridge, D.L. and Jorgensen, W.L. (1986) Computer Simulation of Chemical and Biochemical Systems. *Ann. NY Acad. Sci.* 482.
- (59) Wilson, S. (1986) *Chemistry by Computer. An Overview of the Applications of Computers in Chemistry*, Plenum, New York.
- (60) Roberts, S.M., Ed. (1989) *Molecular Recognition*, Royal Soc. Chem, London.
- (61) Chemical Information System. Numéro spécial de *Chem. Britain* nov. 1989.
- (62) Richards, G., Ed. (1989) *Computer-Aided Molecular Design*, VCH Publishers.

- (63) Stawarz, B. (1990) Nouvelles méthodes pour la modélisation de structures tridimensionnelles en solution en relation avec les données de la RMN 2D, Thèse de Doctorat, Université d'Orléans.
- (64) Manuel d'utilisation du logiciel Discover version 2.1 (1992) Biosym Technologies, San Diego.
- (65) Engler, E.M., Andose, J.D. and von R. Schleyer, P. (1973) Critical evaluation of molecular mechanics. *J. Amer. Chem. Soc.* **95**, 8005-8025.
- (66) Still, W.C. (1986) *Molecular Mechanics*, Columbia Univ. Press.
- (67) Burkert, U.L. and Allinger, N. (1982) *Molecular Mechanics*. ACS **177**.
- (68) Boyd, P.B. and Lipkowitz, K.B. (1982) Molecular mechanics : the method and its underlying philosophy. *J. Chem. Educ.* **59**, 269-272.
- (69) Lomas, J.S. (1986) La mécanique moléculaire, une méthode non quantique pour le calcul de la structure et de l'énergie d'entités moléculaires. *L'actualité chimique* mai 1986, 7-21.
- (70) Bach, L.A. (1989) Modélisation moléculaire et RMN : recherche de structures et étude de la dynamique de caténates et caténands, Thèse de Doctorat, Université de Strasbourg I.
- (71) Clore, G.M., Nilges, M., Brünger, A.T., Karplus, M. and Gronenborn, A.M. (1987) A comparison of the restrained molecular dynamics and distance geometry methods for determining three-dimensionnal structures of proteins on the basis of interproton distances. *FEBS Lett.* **213**, 269-277.
- (72) Tirado-Rives, J. and Jorgensen, W.L. (1990) Molecular dynamics of proteins with the OPLS potential functions. Simulation of the third domain of silver pheasant ovomucoid in water. *J. Am. Chem. Soc.* **112**, 2773-2781.

- (73) Karplus, M. and McCammon, J.A. (1986) Dynamics of proteins. *Sci. Am.* avril 1986, 42-51.
- (74) Agrawal, B.B.L. and Goldstein, I.J. (1967) Protein-carbohydrate interaction. VI. Isolation of concanavalin A by specific adsorption on cross-linked dextran gels. *Biochim. Biophys. Acta* 147, 262-271.
- (75) Dayhoff, M.O., Eck, R.V. and Park, C.M. (1972) *Atlas of Protein Sequence and Structure*, Vol. 5, Natl. Biomed. Res. Foundation, Washington DC.
- (76) Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.
- (77) Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* 78, 3824-3828.
- (78) Gaboriaud, C., Bissery, V., Benchetrit, T. and Moron, J.P. (1987) Hydrophobic cluster analysis : an efficient new way to compare and analyse amino acid sequences. *FEBS Lett.* 224, 149-155.
- (79) Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Moron, J.P. (1990) Hydrophobic cluster analysis : procedure to derive structural and functional information from 2-D-representation of protein sequences. *Biochimie* 72, 555-574.
- (80) Goldstein, I.J. and Poretz, R.D. (1986). In *The lectins, properties, functions, and applications in biology and medicine* (Liener, I.E., Sharon, N. and Goldstein, I.J., Eds.), pp. 33-247, Academic Press Inc., New York.

TABLE DES MATIÈRES

INTRODUCTION.....	2
PARTIE THÉORIQUE	4
I- Les acteurs de la symbiose Légumineuse- <i>Rhizobium</i>	5
1- Les facteurs Nod	5
2- Les lectines	11
3- La symbiose Légumineuse- <i>Rhizobium</i>	19
4- Les gènes de nodulation ou gènes <i>nod</i>	22
II- La modélisation moléculaire et ses méthodes	29
1- Bases de la modélisation moléculaire.....	29
1.1- Définition.....	29
1.2- Fondements physiques.....	29
1.2.1- Mécaniques	30
1.2.2- Thermodynamiques	30
1.3- Méthodologie informatique.....	30
1.3.1- Obtention d'un modèle initial.....	30
1.3.2- Optimisation de la géométrie.....	30
1.3.3- Techniques de graphisme moléculaire	31
1.4- Configurations informatiques requises	31
2- Propriétés moléculaires	31
2.1- Structure moléculaire	31
2.1.1- Définition	31
2.1.2- Aspect énergétique	32
2.2- Mouvements moléculaires.....	33
2.2.1- Mouvements de vibration	33
2.2.2- Mouvements de rotation.....	34
2.2.3- Agitation thermique	34
3- Représentation de l'énergie interne moléculaire	34

3.1- Méthodes quantiques et semi-empiriques	35
3.2- Méthodes empiriques	35
3.2.1- Définition du champ de force ou « force field » moléculaire	35
3.2.2- Formule empirique du champ de force	36
3.2.2.1- Energie d'interaction entre atomes liés	36
3.2.2.1.1- Energie de déformation du squelette.....	37
a) élongation des liaisons	37
b) distorsion des angles de valence	37
c) déformation des angles dièdres impropres.....	37
3.2.2.1.2- Energie de torsion	37
3.2.2.2- Energie d'interaction entre atomes non-liés	38
3.2.2.2.1- Interactions de Van der Waals	38
3.2.2.2.2- Interactions électrostatiques.....	38
3.2.2.2.3- Liaisons ou ponts hydrogène	38
3.2.3 - Paramétrisation	39
3.2.3.1- Paramètres de référence.....	39
3.2.3.2- Paramètres de substitution.....	39
3.2.4- Un exemple de champ de force, celui utilisé par le logiciel Discover : Consistent Valence Force Field ou CVFF.....	39
3.2.5- Faiblesses des champs de force	39
4- Minimisation de l'énergie interne moléculaire	42
4.1- Stratégie générale	42
4.2- Principaux algorithmes	43
4.2.1- Algorithmes d'ordre 1	43
4.2.1.1- Méthode "de la plus grande pente" ou « steepest descent »	43
4.2.1.2- Gradient conjugué ou « conjugate gradient »	44
4.2.2- Algorithmes d'ordre 2 : Newton-Raphson (et méthodes apparentées).....	47
4.1.3- Un exemple d'algorithme : celui du logiciel Discover	50
5- Méthodes de la dynamique moléculaire	50
5.1- Objectifs	50
5.2- Principes	50
5.3- Avantages.....	52

6- Principaux logiciels de modélisation moléculaire : caractéristiques et applications	53
PARTIE EXPÉRIMENTALE.....	54
1- Oses et dérivés d'oses substitués utilisés	55
2- Isolement des lectines et des isolectines	55
3- Techniques d'hémagglutination	58
3.1- Hémagglutination quantitative.....	58
3.2- Inhibition de l'hémagglutination	58
4- Modélisation moléculaire	59
4.1- Alignement et comparaison des séquences	59
4.2- Profils hydrophobes.....	60
4.3- Profils HCA ou « Hydrophobic Cluster Analysis »	61
4.4- Modélisation moléculaire des lectines	62
4.5- Construction des modèles d'oses substitués.....	63
4.6- Ancrage moléculaire ou « docking »	63
RÉSULTATS.....	65
1- Affinité des oses et dérivés d'oses substitués pour les lectines.	66
2- Modélisation moléculaire	68
2.1- Validité des modèles moléculaires	68
2.2- L'interaction moléculaire lectine-ose	69
2.3- Ancrage moléculaire	72
2.3.1- Substituants en C6	72
2.3.2- Substituants en C1 et en C2	72
2.3.3- Substituants en C3	77
DISCUSSION ET CONCLUSION	81
BIBLIOGRAPHIE	83

ANNEXE 1 : TABLE DES ILLUSTRATIONS.

<p>Figure 1 : A- Structure des facteurs Nod isolés de <i>Rhizobium meliloti</i>. B- Structure des facteurs Nod isolés de <i>Rhizobium leguminosarum</i> bv. <i>viciae</i></p>	7
<p>Figure 2 : Structure des facteurs Nod de différentes espèces des genres <i>Rhizobium</i> et apparentés (<i>Azorhizobium</i> et <i>Bradyrhizobium</i>)</p>	10
<p>Figure 3 : A- Structure tridimensionnelle de la lectine de Pois ou PsA. B- Structure tridimensionnelle de l'isoelectine I de <i>Lathyrus ochrus</i> ou LoLI.</p>	13
<p>Figure 4 : Photographie. Interaction de la N-acétylglucosamine avec le site de reconnaissance de LoLI. Schéma. Interaction entre l'α-méthyl-mannopyranoside et le site de reconnaissance de LoLI</p>	17
<p>Figure 5 : Nodosités d'une racine de <i>Lathyrus ochrus</i></p>	20
<p>Figure 6 : Schéma descriptif du métabolisme nodulaire</p>	23
<p>Figure 7 : A- Schéma descriptif de la régulation des gènes de nodulation de <i>Rhizobium meliloti</i>. B- Modèle moléculaire de la régulation des gènes <i>nod</i> de <i>Rhizobium meliloti</i></p>	26
<p>Figure 8 : A- Expression mathématique du champ de force CVFF du logiciel Discover. B- Illustration des interactions associées aux différentes fonctions de potentiel de CVFF</p>	40

Figure 9 : Principe de la minimisation de l'énergie interne moléculaire à l'aide de la recherche de ligne ou « search-line » utilisée par les algorithmes d'ordre 1... 45

Figure 10 : Principe de la minimisation de l'énergie interne moléculaire à l'aide des algorithmes d'ordre 2 : Newton-Raphson et méthodes apparentées 48

Figure 11 : **A-** Superposition des modèles tridimensionnels de la lectine de Lentille LcA modélisée et de la lectine de Lentille LAL résolue par radiocristallographie.
B- Superposition des sites de reconnaissance des monosaccharides de LcA et de LAL..... 70

Figure 12 : **A-** Ancrage moléculaire de la glucosamine dans le site de reconnaissance des monosaccharides de LoLI.
B- Ancrage moléculaire de la glucosamine-6SO₃⁻ dans le site de reconnaissance des monosaccharides de LoLI..... 73

Figure 13 : **A-** Ancrage moléculaire de la glucosamine-1SO₃⁻ dans le site de reconnaissance des monosaccharides de LoLI.
B- Ancrage moléculaire de la glucosamine-3SO₃⁻ dans le site de reconnaissance des monosaccharides de LoLI..... 78

ANNEXE 2 : TABLE DES TABLEAUX.

Tableau 1 : Structures tridimensionnelles des lectines/isolectines de Légumineuses obtenues par radiocristallographie.....	15
Tableau 2 : Complexes lectines-ligands analysés par radiocristallographie à haute résolution.....	16
Tableau 3 : Spécificité de nodulation des bactéries des genres <i>Rhizobium</i> et apparentés à l'égard des Légumineuses-hôtes	22
Tableau 4 : Principaux logiciels de modélisation moléculaire : caractéristiques et applications	53
Tableau 5 : Oses et dérivés d'oses substitués utilisés.....	56
Tableau 6 : Inhibition des lectines/isolectines de <i>Lathyrus ochrus</i> (LoLI, LoLII), de Pois (PsA), de Lentille (LcA) et de <i>Canavalia ensiformis</i> (Con A) par des oses et dérivés d'oses diversement substitués	67
Tableau 7 : Mauvais contacts entre le groupe sulfate de la glucosamine-6SO ₃ ⁻ et les résidus du site de LoLI	75
Tableau 8 : Mauvais contacts entre le groupe phosphate de la glucosamine-6PO ₃ ⁻ et les résidus du site de LoLI	76
Tableau 9 : Distances mesurées entre l'atome de soufre du groupe sulfate des dérivés d'oses O3-sulfatés utilisés et les atomes d'azote engagés dans les liaisons peptidiques entre les résidus Gly ⁹⁸ , Gly ⁹⁹ et Tyr ¹⁰⁰ du site de LoLI	80

DUMEIRAIN (Franck). — Interaction lectines-oses et dérivés d'oses substitués : étude par inhibition quantitative de l'hémagglutination et par modélisation et ancrage moléculaire. — 99 f.; ill.; tabl.; 30 cm (Thèse : Pharm.; Limoges; 1995).

RESUME :

Chez certaines Légumineuses, il s'établit une symbiose spécifique et fonctionnelle entre la plante-hôte et une bactérie fixatrice d'azote du genre *Rhizobium*. De nombreux travaux récents montrent que cette reconnaissance reposerait, au niveau moléculaire, sur une interaction entre des protéines de localisation racinaire, les lectines, et des lipo-oligosaccharides bactériens diversément substitués, notamment par des groupes sulfate, les facteurs Nod.

Nous avons ainsi recherché l'influence de la substitution, essentiellement par des groupes sulfate ou phosphate, de divers oses (mannose et glucose) et dérivés d'oses (mannosamine, glucosamine et *N*-acétylglucosamine) sur leur affinité à l'égard de lectines (LoLI et LoLII de *Lathyrus ochrus*, PsA du Pois, LcA de la Lentille et la ConA de *Canavalia ensiformis*) glucose/mannose-spécifiques. Cette étude, réalisée par deux méthodes complémentaires, une expérimentation par inhibition quantitative de l'hémagglutination et une approche de modélisation et d'ancrage moléculaire, permet de supposer que la spécificité de reconnaissance ne serait pas due au groupe sulfate en C6 du résidu de *N*-acétylglucosamine de l'extrémité réductrice de certains facteurs Nod, mais pourrait faire intervenir un groupe sulfate en C2 ou C3 d'un dérivé d'ose bien que de telles substitutions n'aient pas été rapportées jusqu'à présent.

MOTS-CLES :

- Ancrage moléculaire.
- Facteurs Nod.
- Hémagglutination (inhibition de).
- Lectines.
- Légumineuses.
- Modélisation moléculaire.
- Ose.
- *Rhizobium*.

JURY : Président : M. le Professeur Michel BERNARD.
Juges : M. le Professeur Christian MCESCH.
M. Jean-Luc DUROUX, Maître de Conférences.
M. le Docteur Jean DEBORD.