

**République Algérienne Démocratique et Populaire**  
**Ministère de L'enseignement Supérieur et de la Recherche Scientifique**



UNIVERSITE ABOU-BEKR BELKAID DE TLEMCCEN

Faculté des sciences – Département de chimie

**Laboratoire des Substances Naturelles et BIOactives**  
(LASNABIO)

**THESE**

Présentée par : **M<sup>me</sup> SAIL née KHEBICHAT Naouel**

En vue de l'obtention du diplôme de doctorat

**Spécialité : Chimie Physique : Option Chimie Théorique et Modélisation Moléculaire**

**Thème :**

Analyse conformationnelle des macromolécules biologiques

Devant le jury :

Prof. Houcine ALLALI	Université de Tlemcen	Président
Prof Mohamed KOTBI	Université de Tlemcen	Examinateur
Prof Mohamed KAMAL	Université Roi Abdulaziz-Jeddah	Examinateur
Prof Aissa BELOUATEK	Centre Universitaire-Relizane	Examinateur
Prof Tewfik BOUHAOUR	Université de Tlemcen	Examinateur
Prof Said GHALEM	Université de Tlemcen	Directeur

*A la mémoire de mon Père*

*A mes parents*

*A mon mari NOUREDDINE*

*A mon trésor mes fils Ibrahim Modjib Errahmene et Ahmed Zakarya*

*A mes sœurs SAMRA et HAMIDA*

*A mes neveux Mohammed Rayane, Amine Noor Essalam, et Mohammed Elkhalil*

*A ceux qui m'aiment malgré tous mes défauts*

# Remerciements

*Le présent travail a été réalisé au sein du laboratoire LASNABIO (Laboratoire des Substances Naturelles et BIOactives) dans l'équipe de modélisation moléculaire dirigée par monsieur Said GHALEM professeur à l'université de Tlemcen et directeur du laboratoire.*

*Je suis honorée et il m'est agréable d'adresser mes remerciements à Monsieur le Professeur S.Ghalem, de qui j'ai tiré un immense profit de son savoir ; par ses conseils éclairés et sa vision objective et rationnelle de la problématique ainsi que sa bienveillante disponibilité.*

*Mes remerciements vont également au Professeur R. Borsali, directeur du Centre de Recherche sur les Macromolécules végétales (CERMAV-CNRS) Grenoble-France, pour m'avoir accueilli chaleureusement au centre de recherche, et avoir mis à ma disposition les programmes et les moyens informatiques qui ont été essentiels dans l'aboutissement de ce travail.*

*J'exprime ma profonde et respectueuse gratitude à Monsieur Allali. H. Professeur à l'université de Tlemcen pour m'avoir fait l'honneur de présider le jury.*

*J'adresse mes vifs et sincères remerciements à Messieurs Bouchaour T. et Kotbi M., professeurs à l'université de Tlemcen pour l'honneur qu'ils m'ont fait d'accepté d'être les examinateurs de ce travail. Mes vifs et sincères remerciements vont à Monsieur M.A. Kamal, Professeur à l'université Roi Abdulaziz-Jeddah (Arabie saoudite), et Monsieur A. Belouatek, Professeur au centre universitaire-Relizane pour l'honneur qu'ils nous ont fait en acceptant de faire le déplacement à Tlemcen pour participer au jury.*

*Je voudrais adresser mes vifs et sincères remerciements à mon mari Noureddine pour son encouragement, son aide et sa patience durant toutes ces années.*

*J'exprime ma profonde reconnaissance à M<sup>elle</sup> H.Chemouri et K. Nekkaz pour leur disponibilité, leurs conseils et leurs discussions qui m'ont aidé tout au long de ces années.*

*Je remercie enfin tous les membres du laboratoire LASNABIO (équipe de modélisation moléculaire) ainsi que tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.*

# Table des matières

	Page
Table des matières.....	i
Liste des abréviations.....	iv
<b>INTRODUCTION GENERALE.....</b>	<b>1</b>
<b>Référence.....</b>	<b>7</b>
<b>CHAPITRE I : LES BIOMOLECULES.....</b>	<b>9</b>
A. Les protéines.....	10
1. Structure des protéines.....	11
1.1. Structure primaire.....	11
1.2. Structure secondaire.....	12
1.2.a. Les helices $\alpha$ .....	15
1.2.b. Les feuillet $\beta$ .....	15
1.2.c. Les coudes.....	15
1.2.d. Les boucles.....	16
1.3. Structure tertiaire.....	16
1.4. structure quaternaire .....	16
2. Les protéines et les ions métalliques.....	17
3. Le cuivre dans l'organisme.....	18
B. Les glucides.....	21
1. Relations structure-fonction.....	22
2. Glycobiologie: une science récente.....	23
2.1. Les rôles de reconnaissance moléculaire.....	24
2.2. Antigènes membranaires.....	26
2.3. Les systèmes Lewis.....	27

3. Analyse conformationnelle des glucides: facteurs déterminant la structure tridimensionnelle.....	28
3.1. Monosaccharide.....	29
3.2. Disaccharide.....	31
3.3. Oligosaccharide.....	33
<b>Références.....</b>	<b>35</b>
<b>CHAPITRE II : MODELISATION MOLECULAIRE.....</b>	<b>37</b>
1. Mécanique quantique.....	38
2. Mécanique moléculaire et champ de force.....	40
1.1. Méthodes de minimisation d'énergie.....	43
a- La méthode Steepest Descent.....	43
b- La méthode du Gradient conjugué.....	44
3. Dynamique moléculaire.....	44
1.2. Principe de base.....	44
1.3. Algorithme SHAKE.....	45
1.4. Procédure d'une dynamique moléculaire.....	46
1.5. Pas du temps d'intégration .....	48
1.6. Hydratation du système.....	48
1.7. Conditions aux limites périodiques, PBC.....	49
1.8. Traitement des interactions non liées.....	51
a- Cut-off .....	51
b- Sommation d'Ewald (Particule Mesh Ewald, PME).....	51
2. Analyse conformationnelle.....	52
<b>Références.....</b>	<b>56</b>
<b>CHAPITRE III : ANALYSE CONFORMATIONNELLE DES GLUCIDES.....</b>	<b>59</b>
I. Modélisation des glucides.....	60
1. Champs de force pour les glucides.....	60
1.1. Glycam06.....	61
1.2. Autres champs de force pour les carbohydrates.....	63
2. Etudes conformationnelles in silico des carbohydrates.....	66
2.1. Procédures de recherche systématiques.....	66
2.2. Rôle du solvant dans les études conformationnelle des carbohydrates.....	68
II. Analyse d'une simulation de dynamique moléculaire.....	69
1. Validation des structures.....	69
a- Diagramme de Ramachandran.....	69

b- Convergence des structures.....	71
<b>Référence</b> .....	72
<b>CHAPITRE IV : RESULTATS ET DISCUSSIONS</b> .....	76
A. Etude théorique des complexes di-peptidiques de cuivre (II).....	78
1. introduction.....	78
2. approche théorique.....	80
3. résultat.....	82
4. discussion.....	84
5. conclusion.....	86
B. Recherche conformationnelle sur les structure de Lewis X par dynamique moléculaire : étude du tri- et pentasaccharide.....	92
1. introduction.....	92
2. approche théorique.....	94
3. résultat.....	96
3.1 la racine carrée moyenne de déplacement du système.....	99
3.2 Calcul de Phi / Psi des oligosaccharides.....	100
4. discussion.....	101
5. conclusion.....	101
C. Recherche conformationnelle sur la structure des tripeptides par dynamique moléculaire: étude de L,L-Gly-Phe-Phe, L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly.....	110
1. Introduction.....	110
2. Méthodes.....	112
3. résultat.....	114
3.1. la racine carrée moyenne de déplacement du système.....	116
3.2. Calcul de Phi / Psi pour oligopeptides.....	118
4. discussion.....	119
5. conclusion.....	123
<b>Références</b> .....	124
<b>CONCLUSION GENERALE</b> .....	128

## Liste des Abréviations

ABNR	Adopted Basis Newton-Raphson
AMBER	Assisted Model Building with Energy Refinement
CHARMM	Chemistry at HARvard of Macromolecular Mechanics
Cu(II)	Cuivre (II)
DFT	Density Functional Theory
DM	Dynamique Moleculaire
fs	femto seconde
Fuc	fucose
Gal	D-galactose
GalNAc	<i>N</i> -acétylgalactosamine
GC	Gradient Conjugue
Glc	D-glucose
GlcNAc	<i>N</i> -acétylglucosamine
Gly	Glycine
GTA	Glycosyltransferase A
GTB	Glycosyltransferase B
Le	Lewis
Leu	Leucine
M	masse molaire (en mol/l)
Me	methyl
Met	Methionine
MM	mécanique moléculaire
MQ	mécanique quantique
NeuAc	Acide <i>N</i> -acétylneuraminique
ns	nanoseconde
PBC	Periodic Boundary Conditions
PDB	Protein Data Bank
PE	potential Energy
PES	Potential Energy Surface
Phe	Phenylalanine
PME	Particule Mesh Ewald



ps	pico seconde
RMN	Résonance magnétique nucléaire
RMSD	Root Mean Square Deviation
RPE	Résonance Paramagnétique Electronique
RX	Rayons X
SD	Steepest Descent
TE	Total Energy



Introduction générale

# Introduction générale

---

La glycobiologie est la science qui étudie les structures, les propriétés et les fonctions des glucides dans la matière vivante [1]. Ces études ne concernent pas seulement les glucides mais également toutes les autres classes de molécules biologiques (protéines, lipides, nucléotides) qui interagissent avec eux. Le but de la glycobiologie structurale est bien évidemment de déterminer la structure 3D et d'établir des modèles pour les interactions.

Les propriétés physique, chimique, et biologique d'une molécule sont intimement liées à la structure tridimensionnelle qu'elle adopte, les rôles structuraux et/ou fonctionnels des biomolécule s'expliquent par l'architecture qu'elle adopte [2].

Les glycanes sont des enchaînements linéaires ou branchés de monosaccharides. Ils se présentent sous des formes très diverses et sont indispensables à la vie. Liés aux protéines ou aux lipides, ils sont sécrétés dans le milieu externe ou associés à la membrane plasmique des cellules. Ils servent à protéger le squelette peptidique et la surface cellulaire. Ils sont aussi reconnus de manière très spécifique par des récepteurs physiologiques et se trouvent impliqués dans un grand nombre de processus biologiques comme la reconnaissance et la communication intercellulaire. Le développement des méthodes d'analyse des glycoconjugués a permis de dévoiler progressivement l'extraordinaire diversité des glycanes et a révélé le remarquable potentiel d'informations que représentent ces molécules. La connaissance approfondie des structures glycaniques, de leurs fonctions biologiques et de leurs voies de biosynthèse permet aujourd'hui de trouver de nombreuses applications aux glycoconjugués tant dans le domaine thérapeutique que dans l'alimentaire et la cosmétique. Notamment, grâce aux molécules glucidiques, nous sommes capables de déceler certaines maladies graves telles que le cancer en réalisant un diagnostic plus spécifique et plus précoce, et les thérapies basées sur les molécules portant les glycanes sont plus efficaces et moins lourdes que d'autres utilisées actuellement. Néanmoins, nos connaissances sur l'activité biologique des

## Introduction générale

---

glycoconjugués sont freinées par un manque d'outils appropriés. Les glycopeptides et les oligosaccharides synthétiques représentent de bons modèles pour mener à bien une étude fonctionnelle. Cependant, la complexité des structures glucidiques ne facilite pas la production de tels mimes, et la plupart du temps, les problèmes de purification et les rendements faibles de synthèse permettent difficilement de réaliser une évaluation biologique du produit synthétisé [3] .

Cette thèse est consacrée à l'étude conformationnelle des macromolécules biologiques et plus particulièrement des oligosaccharides (les structures de Lewis X : étude du tri-et pentasaccharides) et des oligopeptides (des tripeptides : L,L-Gly-Phe-Phe , L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly).

Le terme conformation qualifie tout la géométrie tridimensionnelle d'une molécule, ce terme ne doit pas être confondu avec celui de la configuration qui précise une possibilité géométrique pour un ensemble d'atomes pour passer d'une configuration à une autre, il faut nécessairement rompre des liaisons covalentes et les réorganiser ; par contre les conformations d'une molécule peuvent être obtenues sans aucune rupture de la liaison covalente.

La rotation autour des liaisons covalentes simples peut modifier la chaînes glycanique dans l'espace, la définition de la conformation peut être étendue sachant que de petits distorsions dans les angles de valences, et les longueurs de liaison s'accompagnent de changement conformationnel, et ces rotations peuvent se produire dans des systèmes conjugués, ces possibilités créent de nombreuses orientations possibles pour la chaîne polysaccharidique, orientation désignée par l'expression possibilités conformationnelle [4] .

La recherche conformationnelle est la composante chef de l'analyse conformationnelle, dont l'objectif est d'identifier les conformations préférées d'une molécule, ces derniers vont

## Introduction générale

---

déterminer le comportement de ces molécules, cela exige l'aptitude que nous localisons les conformations qui sont en points minimaux sur la surface d'énergie potentielle.

Les méthodes de minimisations jouent un rôle crucial dans l'analyse conformationnelle car elles donnent la rapidité et la précision des résultats, donc il faut bien choisir entre ces méthodes [5].

La minimisation de l'énergie potentielle d'une molécule consiste à résoudre un problème d'optimisation géométrique, presque toutes les méthodes de minimisation ont au moins un point commun: le commencement à un endroit approximatif de l'hyper surface énergie potentielle qui correspond à une géométrie très approximative [6].

Cette hyper surface de l'énergie potentielle est l'ensemble de points générés par la fonction de l'énergie potentielle, cette dernière qui est une fonction de variables géométriques du système moléculaire étudié, elle sera représentée par une courbe de potentiel, l'ensemble de points générées par la fonction est plan [7].

Il est important de reconnaître la différence entre une recherche conformationnelle et une dynamique moléculaire ; la recherche conformationnelle concerne seulement le déplacement de structure d'énergie minimale, tandis que la simulation produit un ensemble des états qui inclut des structures dans les minimums d'énergie, cependant, comme nous savons que les méthodes de dynamique moléculaires peuvent être utilisées comme la partie d'une stratégie de la recherche conformationnelle [8].

Si c'est possible, il est souhaitable d'identifier toutes les conformations d'énergie minimales sur la surface d'énergie potentielle, cependant, le nombre de minimums, peut être si grand qu'il est peu pratique d'envisager de les trouver tous, dans de telle circonstances il est habituellement envisageable d'essayer de trouver les minimums accessibles.

## Introduction générale

---

Les populations relatives des conformations d'une molécule peuvent être calculées en utilisant la statistique via la distribution de Boltzmann, elle nous renseigne sur la population des niveaux occupés à une température donnée.

Les méthodes d'analyse conformationnelle ont été utilisées pour trouver le minimum énergétique des molécules les plus complexes [9].

Ces méthodes incluent les procédures qui impliquent la variation des angles de torsion, la variation des angles dièdres [10-11], la variation des distances internucléaires [12-13], et les méthodes de dynamique moléculaire [14-15].

Les effets de solvant sont aussi importants; actuellement on peut estimer l'énergie du solvant de différentes façons, et on peut l'inclure dans l'expression de l'énergie intramoléculaire. Les résultats de l'analyse conformationnelle montrent que pour les molécules en solution les conformations stables se trouvent généralement dans le jeu de conformation stable 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 minimas dans l'espace conformationnelle, on s'accorde donc pour considérer que les résultats d'un calcul conformationnel sur la molécule isolée donnent une bonne idée des possibilités conformationnelle de cette molécule [16].

Pour quelques molécules il y a tant de minimums sur la surface d'énergie qu'il est peu pratique d'essayer de les trouver tous, dans de telles circonstances il est souvent constaté que la conformation est seule avec la valeur la plus basse de la fonction d'énergie, cette conformation d'énergie est d'habitude mentionnée comme la conformation d'énergie minimale globale.

## Introduction générale

---

La conformation d'énergie minimale globale ne peut pas être la structure active. En effet, dans quelques cas, il est possible que la conformation active ne fasse pas de transition à aucun minimum sur la surface d'énergie de la molécule isolée. Cela peut être nécessaire pour une molécule d'adopter plus qu'une conformation par exemple pourrait lier dans une conformation à l'enzyme et adopter ensuite une conformation différente avant la réaction. Les méthodes de recherche conformationnelle peuvent être convenablement classées dans les catégories suivantes:

- ❖ Algorithmes de recherches systématiques
- ❖ Méthodes approches aléatoires
- ❖ Géométrie
- ❖ Distance et dynamique moléculaire

Le but du présent travail consiste à faire une analyse conformationnelle des oligosaccharides (un trisaccharide et un pentasaccharide) de Lewis X et de trois tripeptides L,L-Gly-Phe-Phe (commercial), L,L-Phe-Gly-Phe, and L,L-Phe-Phe-Gly (synthesized by Pr S. GHALEM) [17]. Les différents outils de la modélisation sont utilisés pour mener à bien le travail de ce mémoire (mécanique moléculaire et dynamique moléculaire). les paramètres géométriques et énergétiques de ces structures différentes vont nous aider à discuter les différents réarrangements spatiaux pour cibler les conformations activement biologiques.

## Introduction générale

---

### Références :

- [1] M.Taylor, and K. Drickamer, Introduction to Glycobiology. Oxford University Press. (2003).
- [2] Bartron, “The Conformation Of The Steroid Nucleus”, *Experientia*, (1950), chapitre 6, 316.
- [3] V. Bourgeaux, Thèse de doctorat , l’universite d’Orleans, (2006).
- [4] Reginald H. Garrett, Charles M. Grisham, Biochimie, De Boeck university, (2000), chapitre 4, 81.
- [5] Hermann Dugas, « Principes de base en modélisation moléculaire, Aspects théoriques et pratiques », Chapitre 3 introduction aux méthodes de minimisation d’énergie, Librairie de l’Université de Montréal, (1996).
- [6] J. S. Lomas, la mécanique moléculaire, une méthode non quantique pour le calcul de la structure et de l’énergie d’entité moléculaire, l’actualité chimique, (1986), chapitre7.
- [7] S. Djennane Bousmaha, Thèse de doctorat, USTHB Alger, (1999).
- [8] Andrew R. Leach, *Molecular Modelling, principles and application*, Pearson Education Limited, (1996), chapitre 8, 413.
- [9] Istavan Kolossvary, Waynec Guida, *J.Am.Soc*, 118: (1996), 5011.
- [10] Fine. R. M, wang. H, Shenkin. P. S, Yarmush. D. L, Levinthal. C, “Proteins: Struct. Funct. Gene”t, (1986), chapitre 1, 342.
- [11] Weinberg. N, Wolfe. S, *J. Am. Chem. Soc*, 116: (1994), 9860.



## Introduction générale

---

[12] Weiner. P. K, Profeta. S, Wipff. G, Havel. T, Kuntz. I. D, Lanngridge. R, Kollman. P. A, *Tetraheron*, 39: **(1983)**, 1113.

[13] G. M. Crippen, F. T. Havel, “Distance Geomerty And Molecular Confotmation”, John Wiley, **(1988)**.

[14] Y. Sun, P. A. Kollman, *J. Comput. Chem*, 13: **(1992)**, 33.

[15] D. Byrne, J. li, E. Patt, B. Robson, P. Weiner, J, *Comput-Aided Mol. Des*, 8 : **(1994)**, 67.

[16] Y. Fukasawa, and al, *Tetrahedron Lett*, 27 :(1986), 1825.

[17] S. Ghalem, Thèse de doctorat, Université Paris 7, **(1996)**.



**CHAPITRE I**  
**LES BIOMOLECULES**

# Chapitre I : les biomolécules

---

Les êtres vivants sont principalement constitués de quatre classes de biomolécules. Les acides nucléiques, tout d'abord, contiennent l'information génétique des organismes permettant leur reproduction. Les lipides sont également des réserves d'énergie mais forment aussi la majorité des constituants des membranes biologiques. Les protéines, constituées d'acides aminés, jouent des rôles cruciaux dans pratiquement tous les processus biologiques (la catalyse enzymatique, le transport et la mise en réserve des petites molécules et d'ions, la protection immune, le contrôle de la croissance et de la différenciation des cellules et peuvent également être un support mécanique dans la peau et les os. Enfin, les glucides, servent de réservoir d'énergie et d'intermédiaire métaboliques. [1]

L'espace qui entoure les cellules, appelé espace extracellulaire, contient un ensemble de macromolécules, polysaccharides ou glycosaminoglycanes, protéines, sels et eau, que l'on désigne sous le nom de matrice extracellulaire. Cette matrice extracellulaire joue un rôle essentiel dans la constitution, le maintien et le remodelage de l'architecture tissulaire.

## **A. LES PROTEINES**

Les protéines occupent une place centrale dans le fonctionnement des cellules des êtres vivants. Elles y jouent en effet tous les rôles imaginables, depuis la catalyse des réactions chimiques jusqu'à la défense contre les infections. Notre capacité à nous déplacer dépend de protéines présentes sous la forme de muscles; les os qui constituent l'architecture interne de l'organisme doivent leur résistance à une matrice de protéines collagènes; qu'une cellule se divise ou qu'elle devienne cancéreuse dépend de l'activité régulatrice de protéines, etc. Le secret de cette remarquable variété de fonctions tient à la très large gamme de conformations que les protéines peuvent adopter et aux propriétés chimiques très différentes qui en résultent.

Les protéines sont constituées d'une succession d'acides aminés (vingt types différents chez l'homme); la chaîne ainsi formée se plie pour aboutir à la forme active de la protéine.

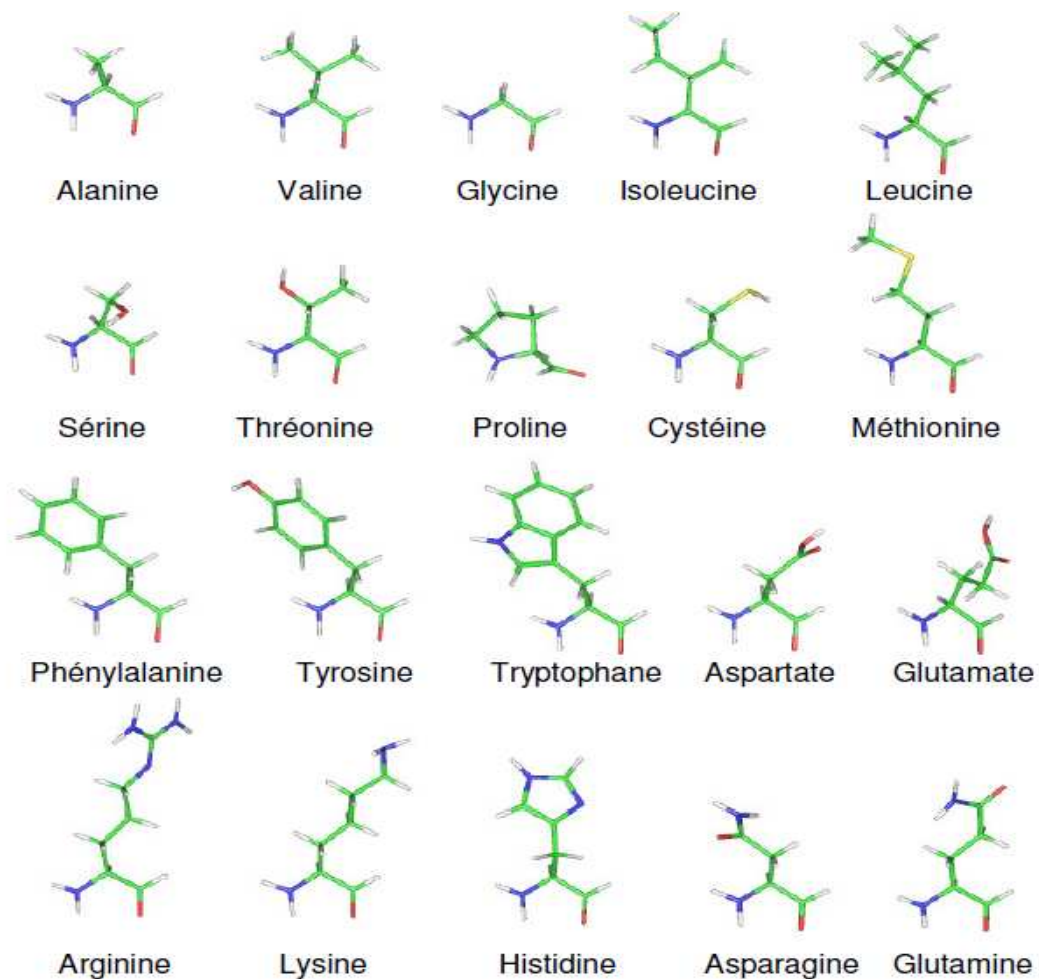
# Chapitre I : les biomolécules

Cette conformation totalement repliée correspond à la forme la plus stable de la protéine dans des conditions normales. On distingue quatre niveaux d'organisation structurale pour une protéine, respectivement appelés structures primaire, secondaire, tertiaire et quaternaire [2].

## 1. Structure des protéines

### 1.1. Structure primaire

La structure primaire correspond à la séquence proprement dite de la protéine, c'est-à-dire à l'enchaînement linéaire des acides aminés par la liaison peptidique. On ne dénombre pas moins de 20 acides aminés qui ne diffèrent les uns des autres que par leur chaîne latérale (groupement R) portée par le  $C_{\alpha}$  (Figure 01). On parle de peptide lorsque le nombre de résidus formant la molécule est inférieur à 50 et de protéine au-delà de 50 résidus.



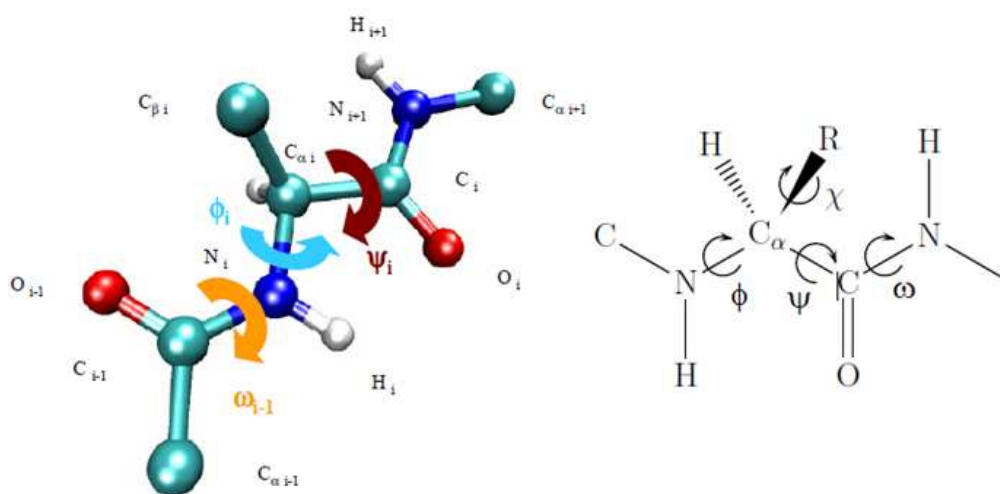
**Figure 01** : Structure des 20 acides aminés composant les protéines.

## 1.2. Structure secondaire

Elle résulte d'un repliement local de la protéine créé par des interactions stériques et électrostatiques et stabilisé par des liaisons hydrogène. On distingue deux angles de torsion principaux dans un résidu (figure 02) :

–  $\phi$ , angle C-N-C $_{\alpha}$ -C;

–  $\psi$ , angle N-C $_{\alpha}$ -C-N.



**Figure 02** : Définition des angles dièdres  $\omega$ ,  $\phi$  et  $\psi$ .

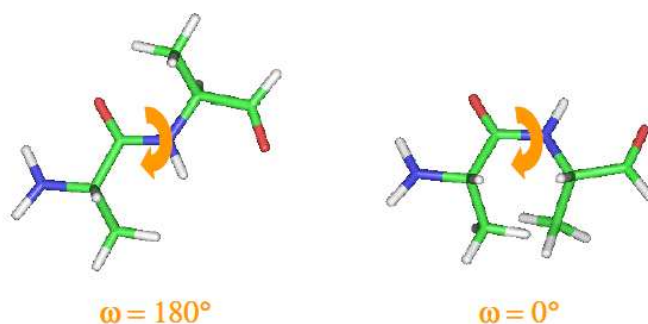
La liaison peptidique étant plane, l'angle (C $_{\alpha}$ -C-N-C $_{\alpha}$ ) est bloqué la plupart du temps sur une valeur proche de 180°, ce qui maintient la liaison peptidique dans la conformation *trans*. L'angle  $\chi$  définit la rotation de la chaîne latérale du résidu.

Dans les protéines, les acides aminés sont reliés entre eux par une liaison peptidique **O=C-N-H** caractérisée par 2 formes imposant ainsi à l'angle de torsion de la liaison C-N (angle dièdre  $\omega$ ) deux valeurs particulières (Figure 03):

0° : liaison peptidique *cis*: son taux d'occurrence est très faible. C'est avec la proline que l'on a le plus de chance d'observer une liaison peptidique *cis*: jusqu'à 10 % de l'ensemble des prolines [3].

## Chapitre I : les biomolécules

180° : liaison peptidique *trans*: c'est la forme la plus stable correspondant à la forme quasiment exclusivement observée dans les protéines.

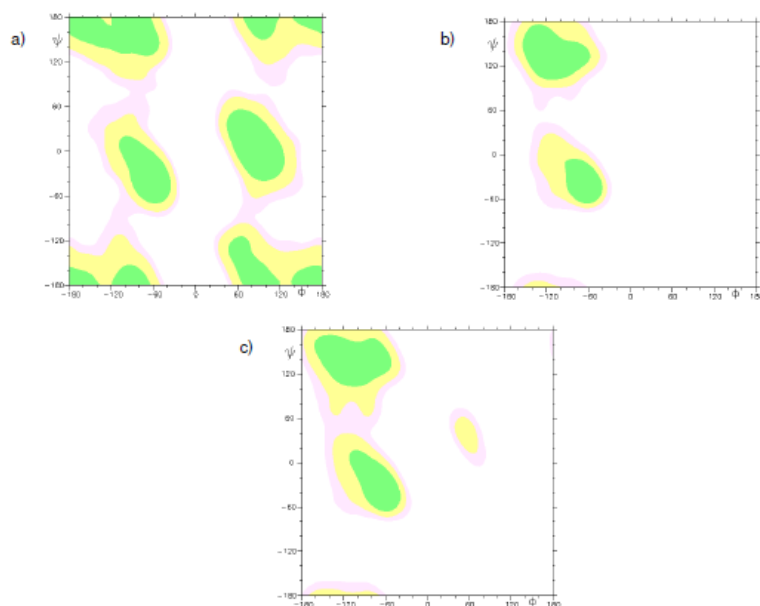


**Figure 03** : Liaisons peptidiques *trans* ( $\omega = 180^\circ$ ) et *cis* ( $\omega = 0^\circ$ ).

Puisque la grande majorité des liaisons peptidiques sont en conformation *trans* ( $\omega = 180^\circ$ ), il est donc possible de décrire la structure d'une protéine à partir de ses seuls angles dièdres  $\phi$  et  $\psi$ . L'angle dièdre  $\phi$  correspond à l'angle de torsion de la liaison  $\text{N-C}^\alpha$  et  $\psi$ , à l'angle de torsion de la liaison  $\text{C}^\alpha\text{-C}$  (Figure 02). Ramachandran [4] a été le premier à proposer d'utiliser la valeur de  $\phi$  en abscisse et celle de  $\psi$  en ordonnée pour tracer une carte 2D, depuis appelée carte de Ramachandran (Figure 04), sur laquelle chacun des résidus de la protéine est représenté par un point.

Ce type de représentation permet ainsi de voir le comportement différent que peut adopter un résidu. La glycine dont la chaîne latérale est un simple atome d'hydrogène, ne connaît pratiquement pas de restriction de valeurs d'angles dièdres : une très grande partie de la carte de Ramachandran est explorée (Figure 04a). A l'opposé, la proline, seul acide aminé cyclique, ne possède que deux régions véritablement favorables du point de vue énergétique.

Ainsi, les liaisons  $\text{N-C}_\alpha$  et  $\text{C}_\alpha\text{-C}$  effectuent librement des mouvements de rotation, mais seules quelques conformations parviennent à minimiser la gêne stérique. On peut repartir ces structures secondaires en catégories représentées dans le tableau 1.



**Figure 04** : Cartes de Ramachandran de la glycine (a), de la valine, de l'isoleucine et de la thréonine (b) et des autres résidus (c), d'après le logiciel MOLMOL: un programme pour visualiser et analyser des structures macromoléculaires [5].

**Tableau 1.** : Paramètres angulaires des différentes structures secondaires [6]. La 4<sup>ème</sup> colonne représente le nombre de résidus nécessaires pour faire un tour, autrement dit l'inclinaison d'un résidu. La dernière colonne permet de connaître la longueur du peptide suivant la structure secondaire considérée et le nombre de résidus.

Conformation	$\Phi$ [°]	$\Psi$ [°]	Résidus/tour	Translation/résidu [°A]
Hélice $\alpha$ droite	-57	-47	3.6	1.50
Hélice $\alpha$ gauche	+57	+47	3.6	1.50
Hélice $3_{10}$ droite	-49	-26	3.0	2.50
Hélice $\pi$ droite	-57	-70	4.4	1.15
Hélice gauche du collagène	-51	+153	3.0	3.13
Feuillet plissé $\beta$ antiparallèle	-139	+135	2.0	3.40
Feuillet plissé $\beta$ parallèle	-119	+113	2.0	3.20
Chaîne étirée	$\pm 180$	$\pm 180$		

# Chapitre I : les biomolécules

---

Donc les structures secondaires ordonnées des protéines peuvent être classées en trois catégories: les hélices, les brins et les coudes. Lorsqu'une portion de protéine ne satisfait à aucun des critères caractérisant les structures précédentes, on est en présence de structure secondaire non ordonnée appelée boucle ou pelote statistique (random coil).

## 1.2.a. Les hélices

Une structure en hélice est définie comme étant une portion de la chaîne peptidique s'enroulant en spirale. Il en existe plusieurs types selon le nombre de résidus par tour d'hélice (voir tableau 1).

L'hélice  $\alpha$  est l'élément de structure secondaire le plus courant. C'est une hélice enroulée à droite constituée de 3,6 résidus par tour. C'est une structure très stable grâce à la présence de liaisons hydrogène entre le C=O du résidu  $i$  et le N-H du résidu  $i+4$ .

Ses valeurs d'angles dièdres sont :  $\phi = -57^\circ$  et  $\psi = -47^\circ$ .

## 1.2.b. Les feuillets $\beta$

Le brin  $\beta$  est le deuxième type de structure secondaire ordonnée retrouvée dans les protéines. C'est une structure proche des structures étendues avec des valeurs standards d'angles dièdres  $\phi = -120^\circ$  et  $\psi = +120^\circ$ .

Deux ou plusieurs brins  $\beta$  peuvent s'associer en feuillet  $\beta$  grâce à des liaisons hydrogènes caractéristiques ; il en existe, entre autres, deux types caractéristiques:

- **le feuillet  $\beta$  anti-parallèle** dont l'orientation des brins successifs est inversée avec des angles dièdres  $\phi = -139^\circ$  et  $\psi = +135^\circ$ .

- **le feuillet  $\beta$  parallèle** dont l'orientation des brins successifs est la même, avec des angles dièdres  $\phi = -119^\circ$  et  $\psi = +113^\circ$ .

## 1.2.c. Les coudes

Les coudes permettent à la chaîne polypeptidique d'effectuer un changement de direction à  $180^\circ$ . Ils sont classés selon le nombre de résidus impliqués dans ce changement de direction :



## Chapitre I : les biomolécules

---

on parle de coude  $\delta$  pour 2 résidus, de coude  $\gamma$  pour 3 résidus (le coude  $\beta_{3_{10}}$  est un cas très particulier de coude à 3 résidus, dont la répétition peut engendrer une hélice  $\alpha_{3_{10}}$ ), de coude  $\beta$  pour 4 résidus, de coude  $\alpha$  pour 5 résidus et de coude  $\pi$  pour 6 résidus [7].

Les coudes  $\gamma$  et  $\beta$  sont les plus couramment mis en évidence dans les structures des protéines.

Ils sont en particulier souvent observés dans la portion peptidique reliant deux brins  $\beta$  formant un feuillet  $\beta$  antiparallèle.

Différents types de coudes ont été définis à partir de valeurs précises des angles dièdres  $\phi$  et

$\psi$ . Ainsi un coude  $g$  est dit de type classique pour des valeurs d'angles dièdres du résidu

Central  $\phi = +75^\circ$  et  $\psi = -64^\circ$  et de type inverse pour  $\phi = -79^\circ$  et  $\psi = +69^\circ$  [8].

### 1.2.d. Les boucles

Le terme boucle ou pelote statistique (random coil) est utilisé lorsque aucune autre structure secondaire ordonnée ne peut être affectée à une conformation locale. Mais il faut cependant bien garder à l'esprit que cela ne signifie pas l'absence d'une structure stable ou d'une structure particulière.

### 1.3. Structure tertiaire

La structure tertiaire des protéines correspond au repliement et à l'assemblage des différents éléments de structures secondaires. La structure tertiaire correspond en fait à la structure tridimensionnelle (structure 3D) de la protéine. Ce sont des interactions non liantes de type électrostatique et de van der Waals, ainsi que les ponts salins, les liaisons hydrogènes et les ponts disulfures qui permettent de stabiliser ce type de structure.

### 1.4 Structure quaternaire

La structure quaternaire d'une protéine résulte de l'association de plusieurs chaînes polypeptidiques ou sous-unités. Ces sous-unités peuvent être identiques comme c'est le cas pour l'hémoglobine ou bien radicalement différente, comme pour le récepteur de l'élastine.

# Chapitre I : les biomolécules

---

L'association peut être nécessaire à l'activité biologique qui, dans ce cas, est le plus souvent extrêmement diminuée, voire inexistante au niveau tertiaire.

## 2. Les protéines et les ions métalliques

La chimie des protéines est un vaste domaine. Elle décrit diverses catégories de macromolécules ; les molécules fibreuses et globulaires, les enzymes, les phosphoprotéines, les hormones, les précurseurs d'hormones en considérant leurs aspects physiques tels que la masse moléculaire et la structure des sous unités, elle constitue ainsi un large volet de la biochimie.

Par ailleurs et en raison du faible poids moléculaire des peptides par rapport aux protéines, la chimie des peptides s'associe beaucoup à la chimie organique. Ainsi la synthèse des peptides est possible avec les moyens qu'offre la chimie organique. Les peptides synthétisés constituent **des modèles** de complexes protéiques [9].

Le corps humain contient environ 3% en masse d'ions métalliques. Malgré cette faible teneur, ces ions jouent des rôles extrêmement importants dans les divers mécanismes du métabolisme des organismes vivants [10] pour lesquels les métaux sont indispensables [11].

La solution à beaucoup de problèmes biochimiques exige une connaissance de la nature des interactions des protéines avec les ions métalliques. Le développement des techniques d'isolement des protéines et l'application des concepts théoriques modernes ont permis de beaucoup apprendre au sujet des propriétés énergétiques et moléculaires des complexes métal-protéine [12].

Parmi les ions métalliques nécessaires aux organismes vivants, certains possèdent des propriétés chimiques particulières qui leur confèrent des rôles uniques. De plus, les sites biologiques par lesquels s'effectuent les interactions entre les métaux et les organes présentent souvent une sélectivité particulièrement élevée pour certains ions. Ceci nous amène à constater que ces sites peuvent réagir avec certains métaux, mais non avec d'autres. Par

exemple, les métaux de transition apparaissent généralement dans les centres actifs des enzymes, bien que certains d'entre eux soient présents seulement sous forme de traces [11].

### 3. Le cuivre dans l'organisme

Le cuivre est présent dans tous les tissus et la plupart des fluides organiques (voir tableau I [13, 14]). L'organisme d'un adulte en contient entre 1,4 et 2,1 mg/kg et celui d'un nourrisson trois fois plus. Ingré par l'alimentation, le cuivre est absorbé par l'estomac et l'intestin grêle. Il est stocké par le foie et normalement excrété par la bile. On le retrouve en quantité importante dans le cerveau, le cœur et le système rénal [11]. En fait la concentration du cuivre reste très faible ( $1\mu\text{M}$ ) dans le plasma et les tissus par rapport aux autres ions métalliques ( $\text{Ca}^{+2}$   $2,5\mu\text{M}$ ,  $\text{Mg}^{+2}$   $1,0\mu\text{M}$ ,  $\text{Zn}^{+2}$   $20\mu\text{M}$  etc...) [15].

Alors la question qui s'impose d'elle-même est la suivante : Comment le cuivre, avec une si faible concentration peut-il entrer en concurrence avec les autres ions métalliques ? Les peptides étant constitués d'acides aminés, leur importance et leur aptitude à la complexation avec l'ion cuivrique suivent celles des acides aminés, lesquels d'ailleurs forment les complexes les plus stables avec le cuivre en comparaison avec les autres ions [16].

## Chapitre I : les biomolécules

**Tableau I:** Concentration moyenne du cuivre dans quelques tissus et fluides biologiques humains [13, 14].

Tissus	( $\mu\text{g/g}$ de cendre du tissu)	Fluides	( $\mu\text{g/ml}$ )
Vésicule biliaire	750	Bile	547
Foie	680	Suc pancréatique	105
Cerveau	370	Sang (total)	1,01
Cœur	350	Erythrocytes	0,98
Rein	270.	Plasma	1,12
Intestins	180 à 300	Sérum	1,19
Estomac	230	Sueur	0,55
Glandes surrénales	210	Lait	0,2 à 0,5
Pancréas	150	Salive	0,08
Œsophage	140	Urine	0,04
Ovaire, Poumon	130		
Peau, Vessie	120		
Prostate, Utérus	110		
Thyroïde	100		
Aorte	97		
Testicules	95		
Rate	93		
Muscle	85		
Os	25		
Ongles	23		
Cheveux	19		

Par exemple, la complexation, de l'ion Nickel (II) par les peptides a été moins étudiée que celle de l'ion cuivrique. Ceci est dû à la plus grande facilité de chélation de l'ion cuivrique ainsi qu'à son rôle biologique [17].

## Chapitre I : les biomolécules

---

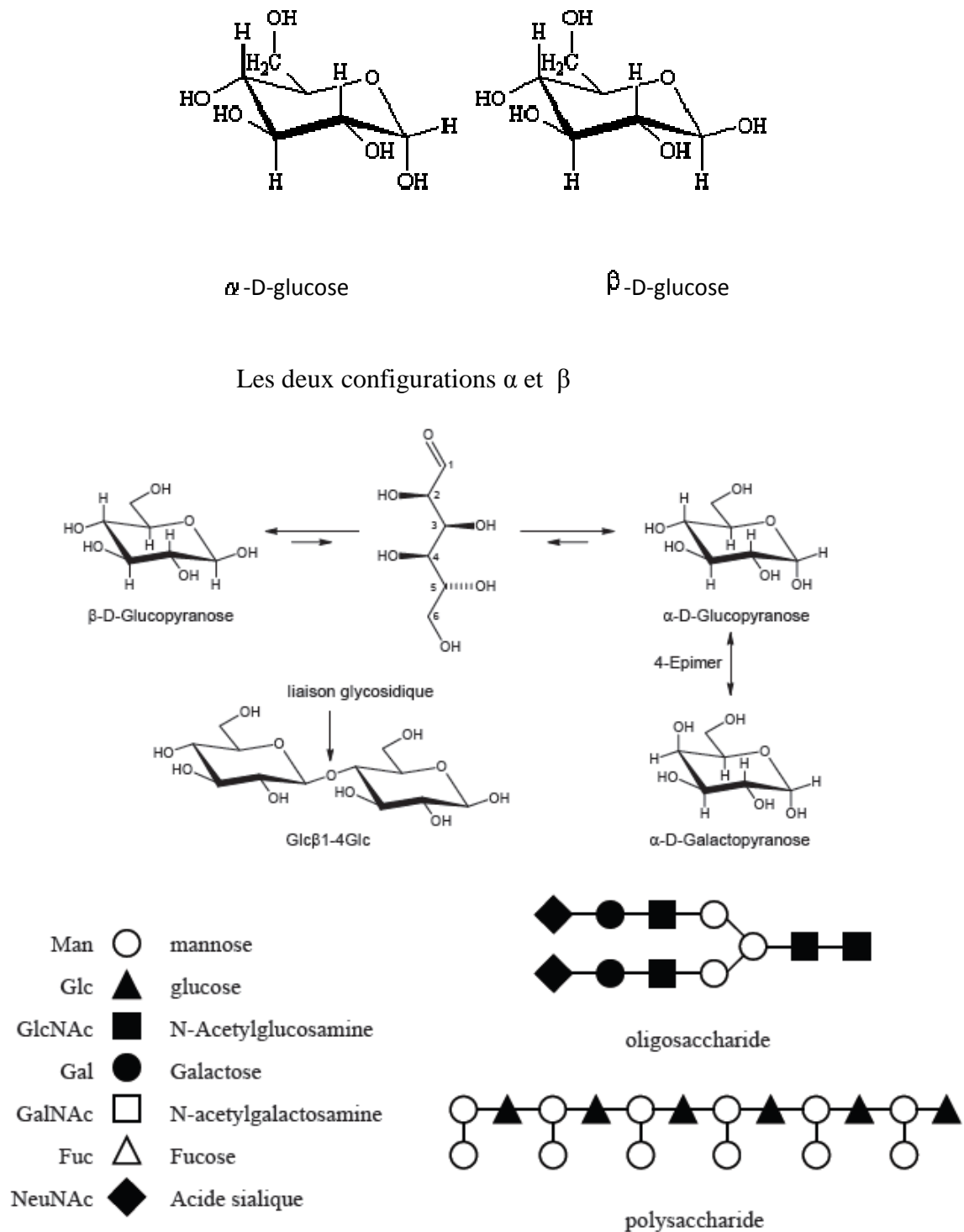
De plus la grande capacité de l'ion cuivrique à déprotoner l'azote peptidique qui constitue un site de chélation potentiel, en plus des sites spécifiques des acides aminés, rend les complexes peptidiques du Cu (II) très stables [11].

Cependant, l'ion Nickel (II) présente un intérêt biologique certain. En effet certains auteurs rapportent qu'il est parmi les métaux les plus cancérigènes pour l'homme [18]. Son taux de présence est plus élevé chez les individus atteints de leucémie que chez les individus sains [19]. D'autant plus que l'hydroxyde de Nickel (II) précipite à pH=7 [20], ce qui est particulièrement intéressant, puisque ce pH est proche du pH physiologique (6,8-7,2). L'étude potentiométrique des systèmes Nickel (II)-peptide montre une plus faible stabilité des complexes du Nickel (II) par rapport aux complexes du Cuivre (II) de même formule brute, ce qui signifie probablement une plus grande facilité pour l'ion Nickel (II) d'être labile et d'être échangé plus facilement que l'ion cuivrique dans les milieux biologiques [17].

La majeure partie des fonctions biologiques passe par une réaction de complexation. En effet l'excès ou le défaut d'un métal peut engendrer des pathologies, signalons que chaque métal est lié à une maladie. Comprendre les mécanismes régissant les interactions entre ions métalliques et systèmes moléculaires (biologiques) rend plus facile l'apport de solutions dans des cas de carence ou d'intoxication. Le but est donc de mettre en évidence tous les paramètres impliqués dans la stabilité des complexes pour pouvoir sélectionner des structures détoxifiantes.

### B. LES GLUCIDES

Les glucides, encore appelés sucres (goût) ou hydrates de carbone  $(C.H_2O)_x$  représentent la plus importante source de la biomasse sur la planète terre et leur complexité, surtout en terme de variations structurales, n'a pas d'égale dans les autres classes de polymères biologiques. Du point de vue chimique, les sucres sont des polyhydroxy-cétones ou polyhydroxy-aldéhydes. Les monosaccharides ont une formule générale du type  $C_xH_{2x}O_x$  avec un nombre d'atomes de carbone de la chaîne principale qui peut varier de trois à dix unités. Une caractéristique générale des sucres est qu'un des atomes de carbone porte une fonction de type céto- ou aldéhyde et pour les structures à cinq (*pentoses*) et six atomes de carbone (*hexoses*), la condensation intramoléculaire est à l'origine de la structure cyclique (*furanose* ou *pyranose*) qui est la forme prédominante en solution. La cyclisation donne origine à un nouveau centre chiral dans la molécule, le carbone anomère C1, qui peut avoir deux configurations appelées  $\alpha$  ou  $\beta$ . Ce carbone anomère peut condenser avec un des hydroxyles d'un deuxième monosaccharide pour donner un disaccharide. Ainsi, l'addition de monomères résulte en la formation de structures dont la complexité augmente considérablement avec le nombre de résidus. Les structures les plus simples, linéaires ou branchées, sont formées par l'addition de deux à cinq monomères et sont nommées en utilisant les préfixes di-, tri-, tétra- et pentasaccharides. Le terme oligosaccharides, concerne des molécules qui contiennent jusqu'à une vingtaine de monosaccharides. Quand la taille et la complexité de la structure deviennent très importantes (plus de 20 résidus) on parle d'une manière générale des polysaccharides.



**Figure05** : Quelques concepts centraux de la structure de glucides

## 1. Relations structure-fonction :

L'énorme potentiel biologique codé par les glucides résulte de leurs capacités à créer des structures de grandes complexités, incluant des points de ramification, à partir des simples

## Chapitre I : les biomolécules

---

monomères (exemple : la laine). Cette caractéristique des glucides est unique parmi tous les polymères biologiques. Par exemple, deux acides aminés différents peuvent se lier seulement de deux façons différentes. Par contre, un simple monosaccharide comme le glucose peut se lier par son carbone anomère (de façon  $\alpha$  ou  $\beta$ ) avec les cinq hydroxyles d'un deuxième glucose, donnant naissance à onze isomères structuraux différents. On peut donc comprendre que le nombre de variations structurales assurées par un tel type de molécules est énorme. La structure des glycannes (oligosaccharides, glycoconjugués, polysaccharides) dépend donc fortement de leur composition et de plus, la présence des liaisons glycosidiques introduit une flexibilité structurale et conformationnelle considérable. La conséquence directe de cette variabilité structurale est une grande diversité dans le rôle et dans la localisation des glucides dans les organismes vivants, tels que représentés ci-dessous :

- ❖ Renforcement structural (polysaccharides)
- ❖ Accumulation d'énergie (polysaccharides)
- ❖ Reconnaissance moléculaire (oligosaccharides/glycoconjugués)
- ❖ Croissance et développement (oligosaccharides/glycoconjugués)
- ❖ Défense (oligosaccharides/glycoconjugués)
- ❖ Ancrage (ancres GPI)
- ❖ Ciblage (oligosaccharides/glycoconjugués)

### 2. Glycobiologie : une science récente

La glycobiologie est la science qui étudie les structures, les propriétés et les fonctions des glucides dans la matière vivante [21]. Ces études ne concernent pas seulement les glucides mais également toutes les autres classes de molécules biologiques (protéines, lipides, nucléotides) qui interagissent avec eux. Le but de la glycobiologie structurale est bien



# Chapitre I : les biomolécules

---

évidemment de déterminer la *structure 3D* et d'établir des modèles pour les *interactions* impliquant ces molécules.

C'est seulement à partir de la moitié du vingtième siècle que les glucides ont été reconnus comme des molécules biologiques extrêmement importantes. Il a longtemps été admis que toute l'information biologique était portée par les deux grandes classes de molécules que sont les protéines et les acides nucléiques. Durant les années 1900, on s'intéressait surtout aux polysaccharides des plantes car ils étaient largement utilisés dans de nombreuses applications industrielles. Les connaissances dans la biologie des glucides étaient donc limitées aux rôles structuraux (cellulose, chitine, etc....) et à l'accumulation et au transport d'énergie (amidon, glucose). La prise de conscience que les glycannes, libres ou sous la forme de glycoconjugués, pouvaient jouer des rôles extrêmement divers et de première importance fut la motivation principale qui a donné naissance à la Glycobiologie moderne. Des découvertes majeures ont marqué l'élucidation du rôle des oligosaccharides :

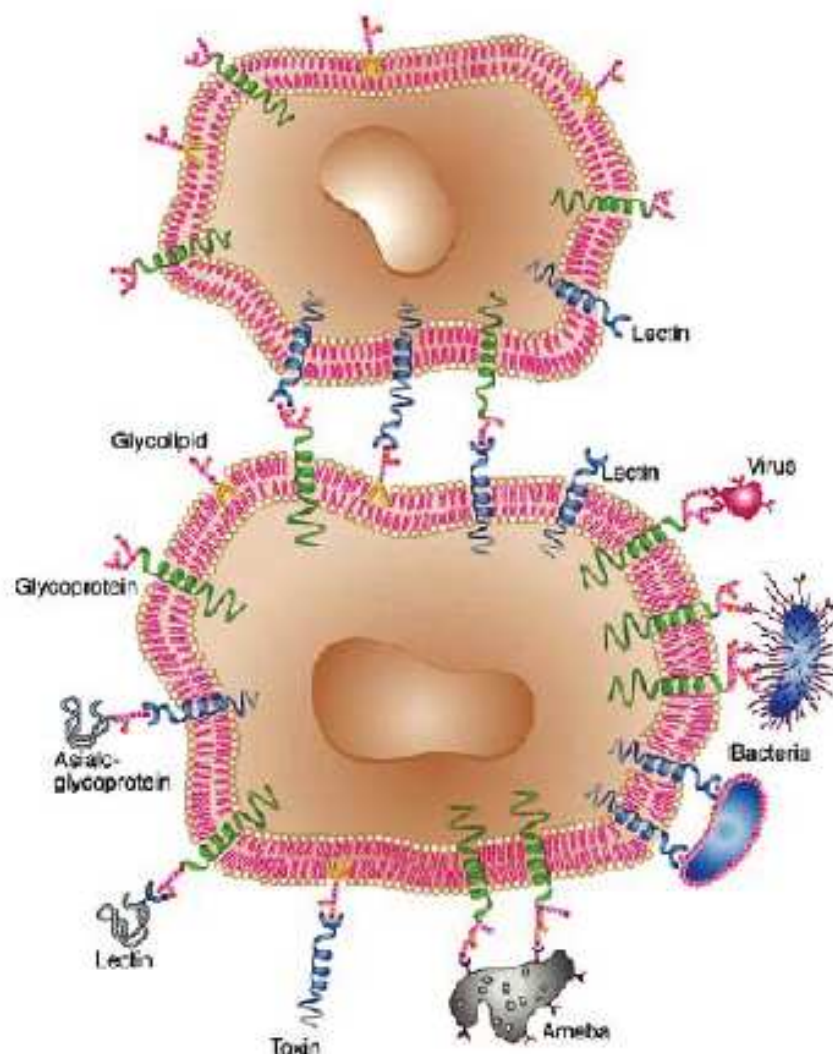
- ❖ La structure glucidique des déterminantes du système ABH sanguin a été établie.
- ❖ L'interaction entre une hémagglutinine et l'acide sialique a été démontrée lors de l'agglutination des érythrocytes par le virus de la grippe.
- ❖ il a été observé que la défucosilation des lymphocytes empêche leur migration dans les tissus. En 1968, Ashwell et Morell identifièrent la première lectine animale, une protéine impliquée dans la clairance des glycoprotéines circulant dans le sang [22]. Cette découverte fût à l'origine des nombreuses recherches qui ont depuis été consacrées au rôle joué par les glucides dans les processus de communication et de reconnaissance moléculaire.

## 2.1. Les rôles de reconnaissance moléculaire

Les oligosaccharides participent à de nombreux phénomènes de reconnaissance moléculaire de type récepteur-ligand impliquant des lectines et de nombreuses études confirment leur rôles dans la vie sociale des cellules [23]. Les lectines reconnaissent généralement les

## Chapitre I : les biomolécules

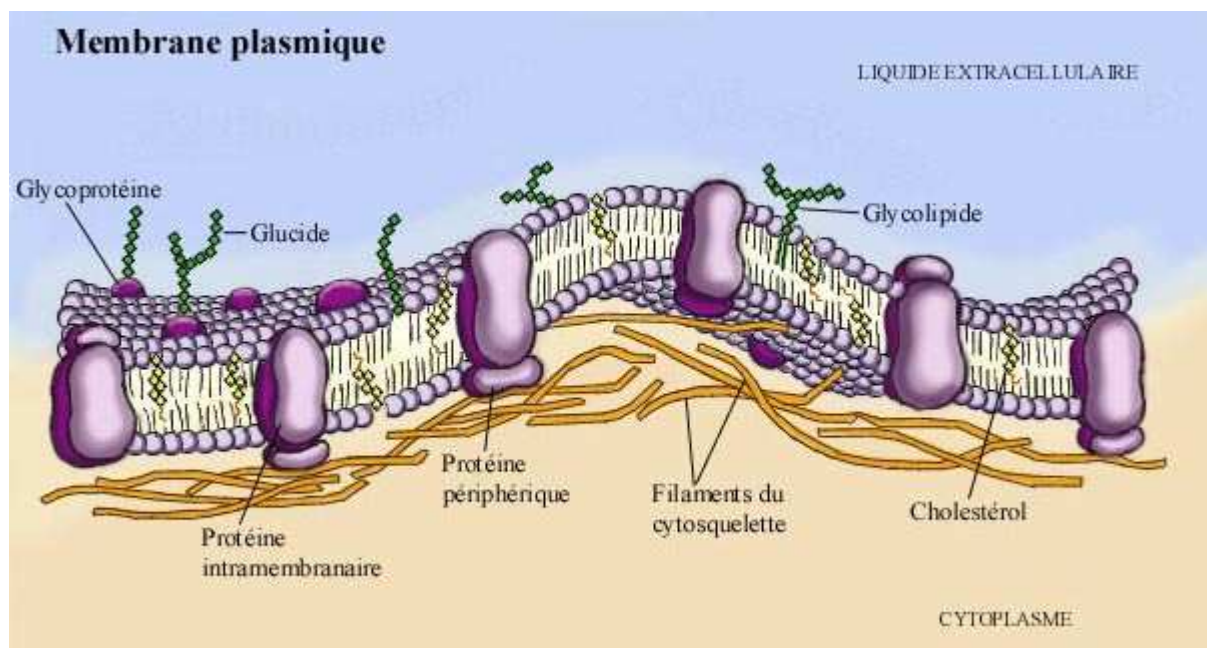
monomères terminaux des oligosaccharides. Au niveau des structures périphériques des glycoconjugués (surtout Oglycannes et glycolipides), on trouve souvent des épitopes antigéniques comme les déterminants des groupes sanguins ABO (antigènes A, B, H) et Lewis (Lewis a, b, X et Y) ainsi que leurs dérivés sialylés. Par ailleurs, il a été montré que des anomalies de glycosylation accompagnent fréquemment les états physiologiques anormaux ou pathologiques, comme le cancer. En conséquence, ce sujet reçoit maintenant une grande attention et de nombreuses études ont été effectuées dans le but d'élucider les structures complexes et les rôles de ces glycannes.



**Figure 06 :** Exemple de rôles de glycoconjugués situés sur la surface cellulaire

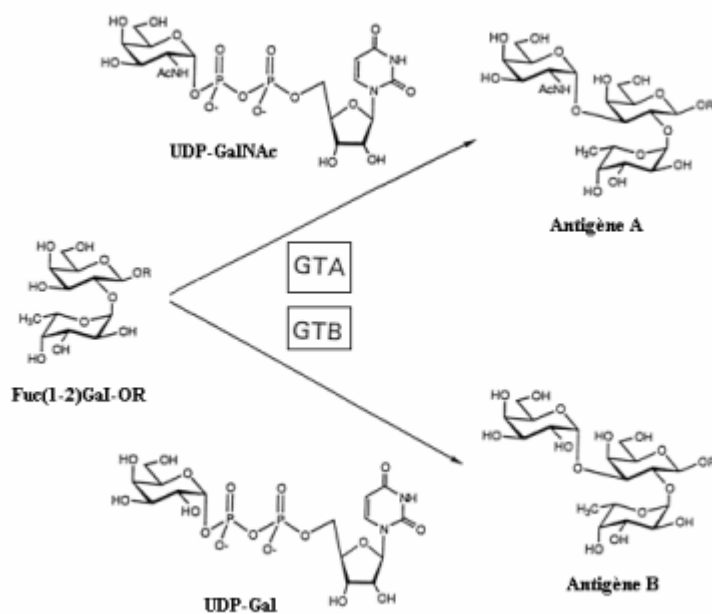
## 2.2. Antigènes membranaires

On trouve à la surface des globules rouges(ou érythrocytes) des molécules capables d'être reconnues par le système immunitaire et de déclencher une réponse immunitaire. Ce sont les antigènes membranaires érythrocytaires. Leur nature chimique est variable protéine et glycoprotéine ou glycolipide (figure 1). Il s'agit de transporteurs et canaux membranaires (protéines assurant les transports de molécules à travers la membrane), d'enzyme, de protéines structurales de la membrane (charpente du globule), de molécules d'adhésions ou de récepteurs membranaires (protéines capables de lier une molécule signal ou informative) [24].



**Figure 07 :** Schéma d'une membrane cellulaire avec les différentes structures glycaniques [25].

La découverte du système ABO en 1901 par Landsteiner a permis d'expliquer pourquoi certaines transfusions sanguines étaient couronnées de succès alors que d'autres se terminaient en tragiques accidents (hémolyse des érythrocytes transfusés[26]).



**Figure 08 :** Biosynthèse des antigènes du groupe sanguin A et B.

L'enzyme Glycosyltransferase A (GTA) humaine catalyse le transfert de GalNAc sur le processur Fuc  $\alpha(1-2)$ Gal  $\beta$ -OR pour former l'antigène A Fuc  $\alpha(1-2)$ [GalNAc  $\alpha(1-3)$ ]Gal  $\beta$ -OR. La Glycosyltransferase B (GTB) catalyse le transfert du Gal pour générer l'antigène B Fuc  $\alpha(1-2)$ [Gal  $\alpha(1-3)$ ]Gal  $\beta$ -OR. R= résidu glucidique attaché à une glycoprotéine ou un glycolipide [27].

Ce système se définit par la présence ou non d'antigènes A, B ou O à la surface des globules rouges des individus et la présence ou non des anticorps associés dans le sérum. Les antigènes des groupes sanguins A et B représentés dans la figure 7 sont des oligosaccharides, présents sur la membrane des globules rouges, et également sécrétés sous forme soluble dans la salive.

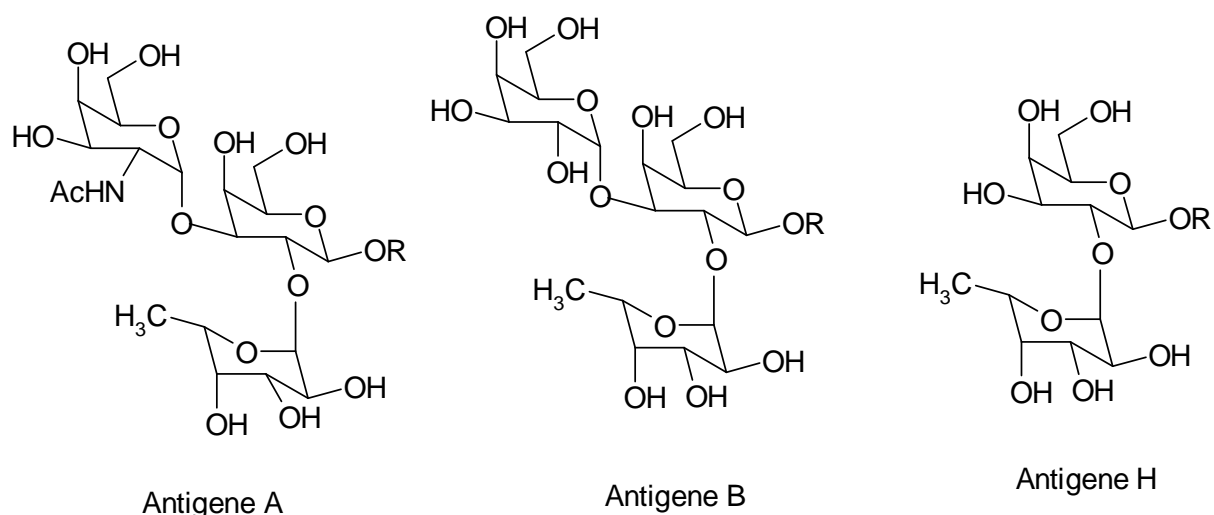
### 2.3. Les systèmes Lewis :

L'antigène Lewis a été identifié la première fois par un anticorps dans un échantillon de sérum provenant de Mme H.G.D.Lewis. Les antigènes de système Lewis sont d'un grand

## Chapitre I : les biomolécules

intérêt dans l'immunohistologie du fait qu'ils sont les seuls antigènes qui ne sont pas produits par les globules rouges. Ce sont des glycoprotéines synthétisés par des glandes salivaires [28] et les cellules de tube digestif [29-31], ces glycoprotéines sont des substances hydrosolubles présentes éventuellement dans les sécrétions et le plasma et qui se fixent secondairement sur les chaînes glycosidiques qui portent les antigènes A, B et H en leur ajoutant de nouvelles spécificités Le(a) et Le(b).

Les antigènes A, B et H sont des oligosaccharides portés par des glycolipides membranaires des hématies, des cellules épithéliales et endothéliales. L'antigène A et l'antigène B sont des trisaccharides, et l'antigène H est un disaccharide qui est l'épine dorsale des trisaccharides A et B.



**Figure 09** : les structures des déterminants antigéniques A, B et H

### 3. Analyse conformationnelle des glucides: facteurs déterminant la structure tridimensionnelle

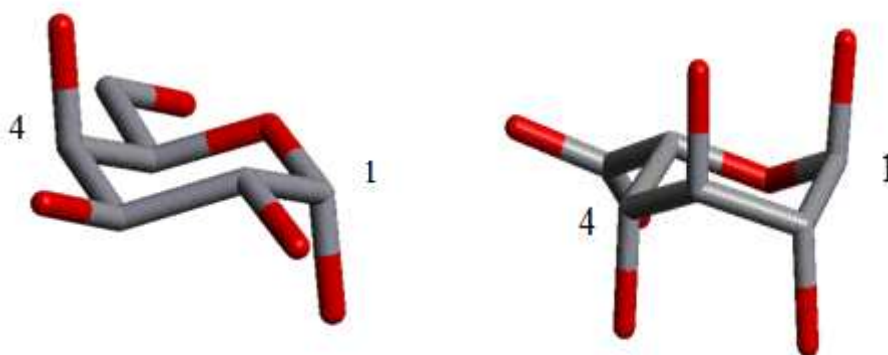
Les méthodes de la chimie computationnelle sont devenues extrêmement importantes dans l'analyse conformationnelle des molécules organiques et notamment des glucides complexes. L'augmentation constante de la puissance de calcul des ordinateurs a permis de

relever de nouveaux défis, notamment l'exploration des surfaces multidimensionnelles, qui étaient techniquement irréalisables auparavant.

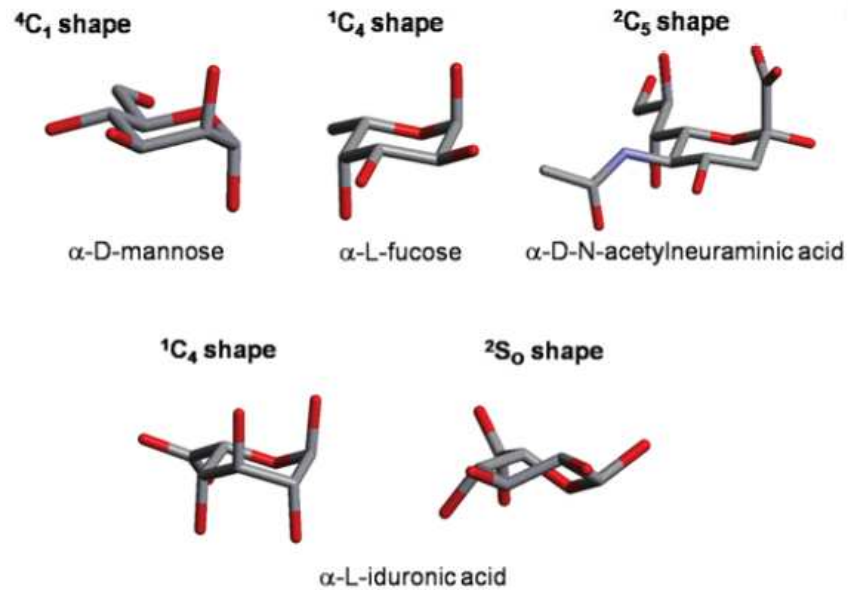
Je voudrais ici attirer l'attention sur les facteurs qui déterminent la structure tridimensionnelle des glucides mais aussi sur les difficultés rencontrées dans l'analyse conformationnelle des structures complexes. Les méthodes de calcul seront décrites dans le troisième chapitre.

### 3.1. Monosaccharides

Dans l'analyse conformationnelle des oligosaccharides, la chaîne principale est généralement considérée comme formée de monomères rigides même si les pyranoses sont dotés d'une certaine flexibilité au niveau du cycle. Pour la plupart des monosaccharides, la conformation de plus basse énergie est la forme *chaise* et plus précisément  ${}^4C_1$  et  ${}^1C_4$  (figure 10) pour les glucides de série D ou L respectivement [32]. Certains monosaccharides tels que les cycles idoses peuvent de plus adopter des conformations de type sofa ( ${}^2S_0$ ).

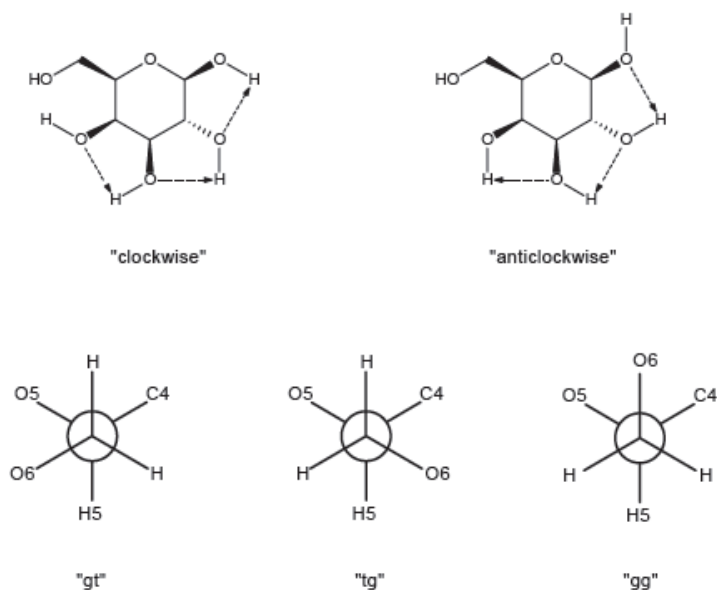


**Figure 10** : Les deux conformations chaise  ${}^4C_1$  et  ${}^1C_4$



**Figure 11 :** Exemples de formes préférées possibles des cycles pyranose

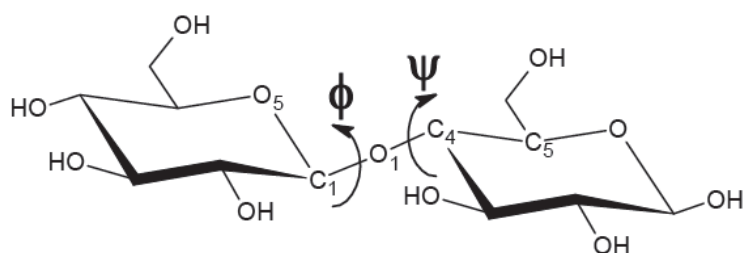
Le groupement hydroxyméthyle exocyclique peut tourner autour de la liaison C5-C6 mais l'analyse cristallographique a montré qu'il adopte trois conformations de basse énergie, nommées *gt*, *gg* et *tg*, pour limiter les interactions avec le O4 et le O5 (Figure. 12) [33,34]. Par contre, les hydroxyles secondaires sont libres de tourner presque librement même si deux orientations relatives (*sens des aiguilles d'une montre ou sens inverse*), montrées dans la Figure 12, favorisent la formation des réseaux de liaisons hydrogène intramoléculaires [35]. L'ensemble de ces paramètres doit être pris en compte dans l'analyse structurale des structures complexes.



**Figure 12 :** Réseaux des liaisons hydrogène pour le galactose et conformation du O6 en projection de Newman

## 3.2. Disaccharides

Nous avons indiqué plus haut que deux pyranoses peuvent se combiner pour donner onze disaccharides différents. La nature de la liaison glycosidique ( $\alpha$  ou  $\beta$ , 1 $\rightarrow$ 1, 1 $\rightarrow$ 2, 1 $\rightarrow$ 3, 1 $\rightarrow$ 4 ou 1 $\rightarrow$ 6) est le paramètre clé qui détermine la structure tridimensionnelle d'un disaccharide. En fait, la molécule présente une certaine flexibilité autour des deux angles de torsion ( $\Phi$  et  $\Psi$ ) de la liaison glycosidique, mais il a été montré que toutes les conformations ne peuvent pas être atteintes avec la même probabilité. Les deux angles de torsion glycosidiques qui définissent la géométrie de la liaison sont illustrés dans la Figure 13.



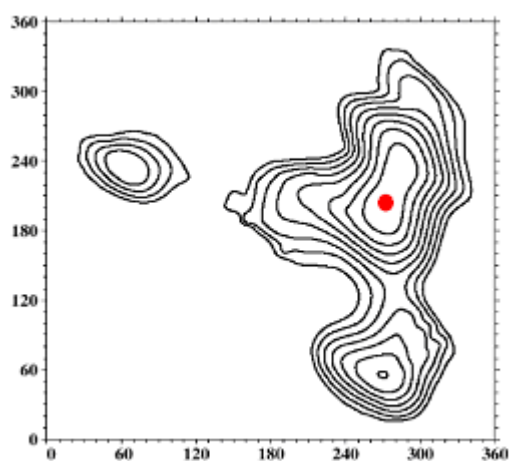
**Figure 13:** Les angles de torsion glycosidique pour Glc $\beta$ 1  $\rightarrow$ 4Glc



## Chapitre I : les biomolécules

L'angle  $\varphi$  (O5—C1—O1—Cx) est influencé en premier lieu par l'effet *exoanomère*, qui est un effet stéréo-électronique résultant des interactions entre les orbitales des atomes O5, C1, O1 et Cx, impliqués dans la liaison glycosidique [36]. L'angle  $\psi$  (C1—O1—Cx—C(x+1)) montre une variabilité plus grande et est plutôt influencé par les interactions non liées (liaisons hydrogène, interactions de van der Waals) entre les deux monosaccharides. La structure globale d'un disaccharide est donc déterminée par la nature de la liaison et par l'ensemble de ces interactions.

La probabilité d'existence pour une conformation donnée d'un disaccharide peut être représentée sur un diagramme bidimensionnel (Figure 14) qui représente la valeur de l'énergie potentielle de la molécule pour chaque combinaison des valeurs de  $\varphi$  et  $\psi$ .



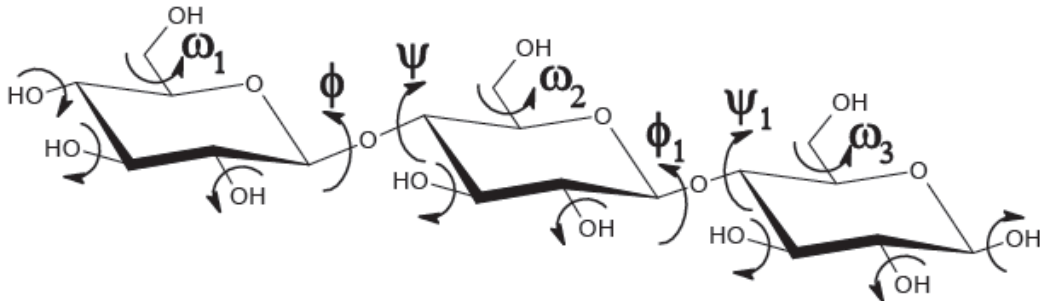
**Figure 14** : Carte adiabatique d'énergie potentielle pour le disaccharide Glcβ1→4Glc. Chaque isocontour représente un niveau d'énergie de 1kcal. Les conformations de plus basse énergie sont indiquées par des points rouges.

On peut voir dans la figure 14 que les conformations de plus basse énergie pour la liaison  $\text{Glc}\beta 1 \rightarrow 4\text{Glc}$  sont confinées dans des régions relativement limitées de l'espace conformationnel.

### 3.3. Oligosaccharides

Les oligosaccharides sont des structures généralement flexibles qui peuvent adopter différentes conformations (ou plus souvent familles de conformations) de basse énergie. La liberté conformationnelle peut être limitée par la présence des interactions (liaisons hydrogène, interactions de van der Waals) entre deux monomères qui ne sont pas forcément liés par une liaison glycosidique. Par exemple, dans le cas particulier des oligosaccharides du type Lewis X, la forte interaction hydrophobe de *stacking* entre le Gal et le Fuc rigidifie la structure de l'oligosaccharide qui n'adopte en fait qu'une seule conformation en solution [37].

Dans l'analyse conformationnelle l'oligosaccharide peut être considéré comme constitué d'un ensemble de disaccharides. Les valeurs de  $\phi$  et de  $\psi$  pour chaque disaccharide seront donc, en première approximation, considérées comme indépendantes des résidus voisins. Ceci est une simplification importante mais qui permet de dire que la structure globale d'un oligosaccharide dépend en premier lieu de la nature de chaque liaison entre ses disaccharides constituants. Malgré ça, l'analyse conformationnelle des oligosaccharides reste un travail long et difficile car le nombre des paramètres mis en jeu est généralement très important. Pour un simple trisaccharide, comme celui montré en Figure 15, on doit prendre en compte six conformations différentes pour chaque monomère (trois pour le C6 et deux pour les hydroxyles) et  $24^4$  combinaisons pour les angles de torsion  $\phi$ ,  $\psi$ ,  $\phi_1$ ,  $\psi_1$  pour un totale de 71663616 conformations possibles !



**Figure 15 :** Espace conformationnel pour un trisaccharide  $\text{Glc}\beta 1 \rightarrow 4\text{Glc}\beta 1 \rightarrow 4\text{Glc}$

Bien qu'il s'agisse d'un simple trisaccharide, un tel nombre de minimisations d'énergie nécessite déjà une puissance de calcul non négligeable. On peut donc comprendre que pour l'analyse conformationnelle des glycanes encore plus complexes (plus de 10 monomères) on a forcément besoin d'une stratégie d'exploration de l'espace conformationnel très puissante et intelligente.

## Chapitre I : les biomolécules

---

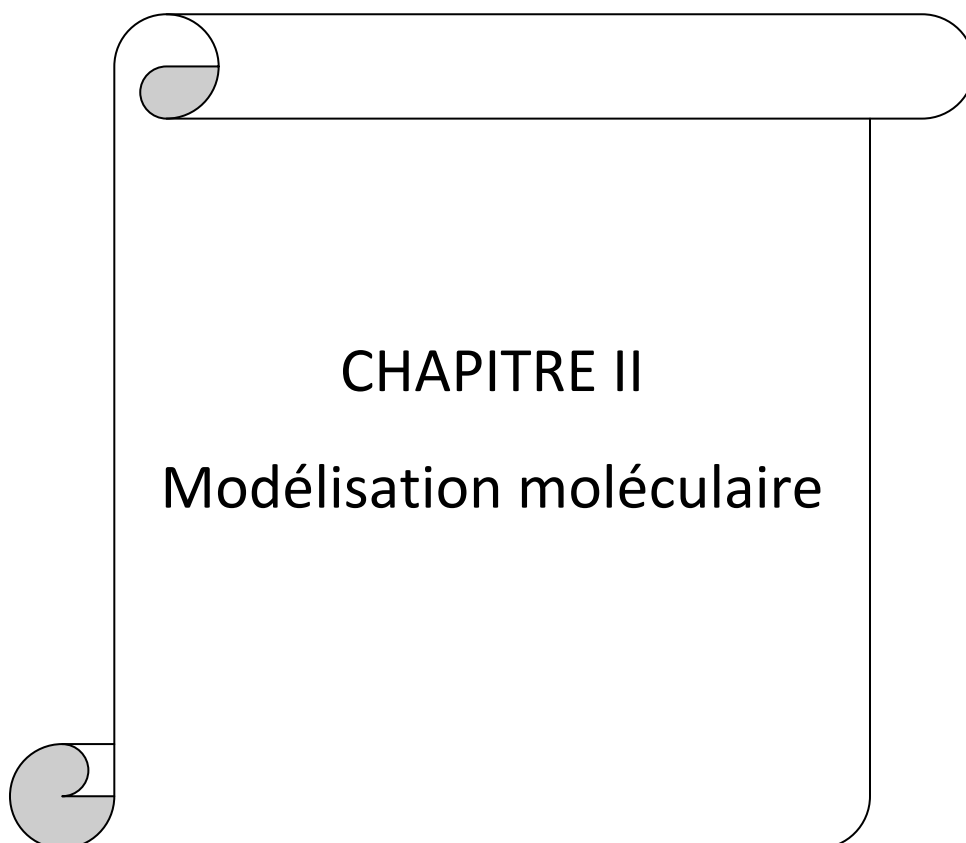
### Références

- [1] G. Moroy, *Thèse de doctorat*. L'université de Reims Champagne-Ardenne. (2005)
- [2] Adrien MELQUIOND, *Thèse de doctorat*. L'université Paris 7 - Denis Diderot. (2007)
- [3] S. Fischer, R. L. J. Dunbrack, and M. Karplus, *J Am Chem Soc*, 116: (1994), 11931-11937
- [4] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J Mol Biol*, 7: (1963), 95-99.
- [5] R. Koradi, M. Billeter, and K. Wuthrich, *J Mol Graph*, 14(1): (1996), 51-5, 29-32.
- [6] S. Weinman, P. Méhul. *Biochimie-Structure et fonction des protéines*. Dunod, Paris, (2000).
- [7] K. C. Chou, *Anal Biochem*, 286(1) : (2000), 1-16.
- [8] E. Milner-White, B.M. Ross, R. Ismail, K. Belhadj-Mostefa, R. Poet, *J Mol Biol*, 204(3) : (1988), 777-82.
- [9] Miklos Bodanszky, *Peptide Chemistry, A Practical Textbook* 2<sup>nd</sup> Edition, Springer-Verlag Berlin Heidelberg ,(1993).
- [10] D. R Willams. *Chem. Rev.*, 72 (3): (1972), 203-213.
- [11] S. Ghalem, *thèse de doctorat*, Université Denis Diderot- Paris 7- France, (1996).
- [12] Thomas R. Hughes, Irving M. Klotz, *Analysis of Metal-Protein Complexes, Methods of Biochemical Analysis*, Volume III, Northwestern University.
- [13] G. V. Iyengar, W. E. Kollmer, H. J. M. Bowen, “*The elemental composition of human tissue and body fluids*” Springer, New-York, (1978).
- [14] I. H. Tipton, M. J. Cook, *Health Phys.*, 9: (1963), 105.
- [15] R. Österberg; *Coord. Chem. Rev.*, 12: (1974), 309.
- [16] L. G. Sillen, A. E. Martell ; “*Stability Constants*“, 2<sup>nd</sup> Edn., *Chem. Soc. Spec. Publ.* N° 17, (1967); Supplement N° 1, *Chem. Soc. Spec. Pub.*, N° 25, (1971).
- [17] B. Mrabet, et al., *J. Chim. Phys.*, 91: (1994), 1537-1551.

## Chapitre I : les biomolécules

---

- [18] F. W. Sundermann Jr., *Ann. Clin. Lab. Sci.*, 7: (1977), 377.
- [19] E. L. Andronikashvili, L. Mosulishvili, "Metal ions in Biological systems", H. Sigel ed., Marcel Dekker Inc. N. Y.; 10: (1980), 185.
- [20] S. Chaberek, R.C. Courtney, A.E. Martel, *J. Am. Chem. Soc.*; 74: (1952), 5057.
- [21] M. Taylor, K. Drickamer, "Introduction to Glycobiology". *Oxford University Press* (2003).
- [22] G.Ashwell, A.G. Morell, *Adv. Enzymol.*, 41: (1974), 99-128.
- [23] A. Varki, *Glycobiology*, 3 : (1993), 97-130.
- [26] V. Bourgeaux, thèse de doctorat. Université d'Orléans, (2006).
- [25] P. H. Seeberger, D. B. Werz, *Nat Rev Drug Discov*, 4 (9): (2005), 751-63.
- [24] C. Gourmala, *thèse de doctorat*, université de Tlemcen, Algérie, (2005).
- [27] N. O. Seto, , M. M. Palcic, C. A. Compston, H. Li, D. R. Bundle, S. A. Narang, *J Biol Chem.*, 272 (22): (1997), 14133-8.
- [28] U.Mandel, et al., *Vox Sang*, 61: (1991), 205-14.
- [29] C.Cordon-Cardo, et al., *Int.J.cancer*, 37: (1986), 667-76.
- [30] O. Seitz, *Chembiochem.*, 1 (4): (2000), 214-46.
- [31] J. Sakomoto, et al., *cancer Res*, 46 : (1986), 1553-61.
- [32] V.S.R. Rao, P.K. Qasba, P.V. Balaji, R. Chandrasekaran, *Conformation of Carbohydrates*. (1988) Harwood academic publishers.
- [33] R.H. Marchessault, S. Pérez, *Biopolymers*, 18: (1979), 2369-2374.
- [34] S.Pérez, C. Gautier, and A. Imberty, *Wiley*, (2000) 969-1001.
- [35] J.W. Brady, R. K. Schmidt, *J. Phys. Chem.*, (1993), 958-966.
- [36] R.U. Lemieux, S. Koto, D. Voisin,. The exo-anomeric effect. In Szarek, W.A. and Horton, D. (eds), *Anomeric effect: origin and consequence*, ACS symposium Series No.87. *Amer. Chem. Soc.*, Washington DC, (1979), 17-29.
- [37] A.Imberty, S. Pérez, *Chem. Rev.*, 100: (2000), 4567-4588.



**CHAPITRE II**  
**Modélisation moléculaire**

La recherche et la synthèse de nouveaux composés chimiques et biochimiques sont aujourd'hui souvent associées à une étude par modélisation moléculaire. La modélisation moléculaire est le terme générique regroupant les méthodes qui permettent de simuler le comportement d'un système de particules. La taille du système étudié peut aller d'une simple molécule diatomique jusqu'à des macromolécules biologiques de plusieurs dizaines de milliers d'atomes.

La modélisation moléculaire peut ainsi permettre de décrire le comportement électronique des atomes et des molécules pour expliquer leurs réactivités, comprendre les processus de repliement d'une protéine ou encore expliquer l'importance de certains acides aminés d'un site catalytique enzymatique.

Pour réaliser ce type d'étude, il est nécessaire de déterminer une expression de l'énergie d'interaction des atomes du système moléculaire en fonction de leurs positions relatives. Pour ce faire, les méthodes de modélisation moléculaire reposent entre autres sur deux approches différentes appelées Mécanique Quantique (MQ) et Mécanique Moléculaire (MM), pour évaluer cette énergie d'interaction.

### 1. Mécanique quantique

La mécanique quantique (MQ) repose sur la résolution de l'équation de Schrödinger pour décrire les interactions au niveau atomique en tenant compte des électrons et des noyaux de chaque atome du système :

$$H \Psi = E \Psi$$

Où  $\Psi$  décrit la fonction d'onde du système,  $H$  est l'opérateur Hamiltonien (contenant les énergies cinétique et potentielle du système) s'appliquant à  $\Psi$  et  $E$  l'énergie du système.

Cependant, la résolution analytique exacte de cette équation n'est possible que pour les atomes mono électroniques tels que les hydrogénoïdes. Pour des systèmes de taille plus

## Chapitre II : Modélisation moléculaire

---

importante, on a recours en particulier à l'approximation de Born-Oppenheimer qui permet de découpler les termes de l'opérateur Hamiltonien. Comme la masse du noyau atomique est significativement plus grande que la masse des électrons, le mouvement des électrons peut ainsi être considéré comme découplé du mouvement des noyaux, ce qui signifie que les électrons s'adaptent instantanément aux changements de positions des noyaux. La fonction d'onde totale de la molécule peut désormais s'écrire :

$$\Psi_{\text{tot}}(\text{noyaux, électrons}) = \Psi(\text{électrons}) \Psi(\text{noyaux})$$

L'énergie totale d'une molécule est égale à la somme de l'énergie nucléaire (la répulsion électrostatique entre les noyaux) et de l'énergie électronique. L'énergie électronique comprend l'énergie cinétique et potentielle des électrons se déplaçant dans le champ électrostatique des noyaux.

$$E_{\text{tot}} = E(\text{électrons}) + E(\text{noyaux})$$

Quand l'approximation de Born-Oppenheimer est utilisée, les noyaux sont supposés fixes quand ce sont les mouvements des électrons qui sont en jeu. Pour chaque réarrangement de la position des noyaux, l'équation de Schrödinger est résolue seulement pour les électrons.

C'est ce type de procédure qui est la base sur laquelle repose les méthodes de modélisation moléculaire quantique ab-initio. Le terme ab-initio provient du fait que les calculs sont effectués indépendamment de toute donnée expérimentale, mais ils permettent d'avoir accès à l'ensemble des propriétés qui découlent des distributions électroniques, en particulier pour les réactions chimiques dans lesquelles les liaisons sont brisées et/ou formées.

Cependant, les calculs qui en résultent sont encore extrêmement longs et limités aux petits systèmes de quelques dizaines d'atomes à cause de l'ensemble des électrons pris en compte et de la complexité des équations à résoudre pour l'évaluation de l'énergie d'interaction. C'est



## Chapitre II : Modélisation moléculaire

---

pourquoi ce type d'évaluation de l'énergie d'interaction est incompatible avec nos systèmes protéiques qui peuvent comprendre plusieurs dizaines de milliers d'atomes.

Nous avons donc eu recours à des calculs de Mécanique Moléculaire qui reposent sur les principes de la mécanique classique. Dans ce type de méthode, les électrons ne sont pas pris en compte, mais seulement les noyaux, ce qui permet de diminuer grandement les temps de calculs.

### 2. Mécanique Moléculaire et champs de force

Les méthodes de mécanique moléculaire (MM) utilisent les lois de la physique classique pour la prédiction des structures et de leurs propriétés, comme l'énergie. Au même titre que les méthodes de MQ, les méthodes de MM utilisent une alternative à l'approximation de Born-Oppenheimer, mais les électrons ne sont pas explicitement inclus dans le calcul. Cette simplification assume que les électrons dans la molécule sont répartis dans leur distribution optimale, et seules les positions des noyaux sont prises en considération pour la résolution du problème chimique particulier. La molécule est alors mécaniquement simplifiée à un ensemble de masses auxquelles des charges ont été attribuées. Ces masses interagissent entre eux par des forces classiques, décrites par des fonctions d'énergie potentielle. Les effets électroniques sont omis ou simplifiés, et inclus dans le **champ de forces** pour certains atomes spécifiques.

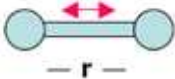
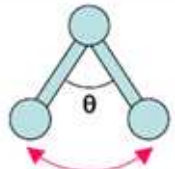
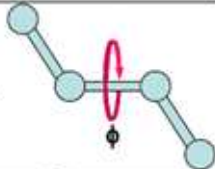
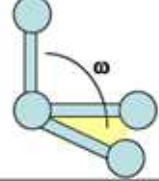
Un champ de forces utilise généralement un ensemble d'équations pour définir comment l'énergie potentielle du système varie en fonction de la position des atomes. Le champ définit aussi les propriétés "chimiques" pour chaque atome (en prenant en compte son environnement chimique), telles que la charge partielle, l'hybridation ou le rayon de Van der Waals. Enfin, le champ de force définit les paramètres spécifiques, telles que les longueurs, les angles et les torsions de liaison, et les barrières de rotation des angles définies par les constantes de force.

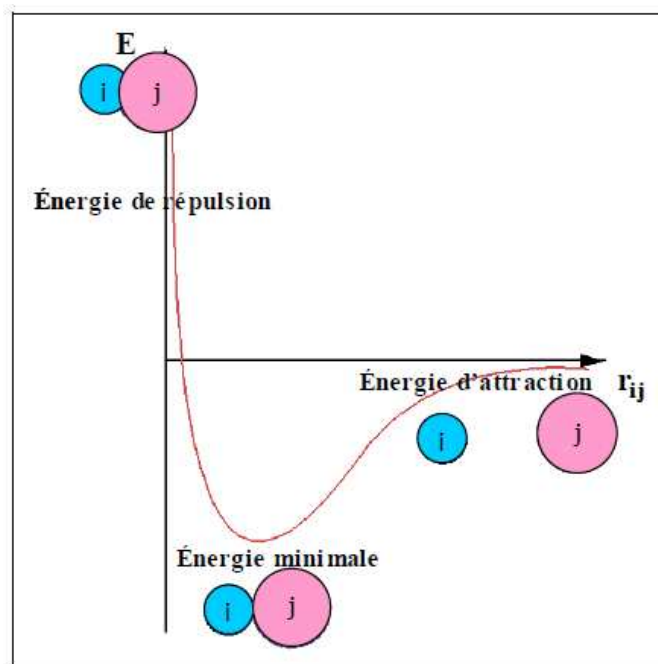
## Chapitre II : Modélisation moléculaire

Ces valeurs sont généralement importées à partir de données expérimentales ou de calculs *ab initio* de haut niveau.

Le théorème fondamental de la MM est que l'énergie potentielle totale ( $E_{tot}$ ) peut être considérée comme la somme de plusieurs contributions:

$$E_{tot} = E_{liaisons} + E_{angles} + E_{dièdres} + E_{impropres} + E_{vdW} + E_{elec} + E_{liaison\ H}$$

Terme du champ de force	Expression générale	Description	Illustration
Energie de liaisons	$\sum k_r (r - r_0)^2$ $r_0$ : longueur de liaison à l'équilibre $k_r$ : constante de force	Définit l'énergie mise en jeu par les variations de la longueur $r$ des liaisons après extension et compression à partir de leurs valeurs d'équilibre	
Energie de déformation des angles de valence	$\sum k_\theta (\theta - \theta_0)^2$ $\theta_0$ : angle à l'équilibre $k_\theta$ : constante de force	Définit l'énergie impliquée dans les variations d'angle $\theta$ entre deux liaisons contiguës, autour d'une valeur d'équilibre $\theta_0$	
Energie de torsion	$\sum k_\phi (1 + \cos n\phi)$ $k_\phi$ : constante de force $n$ : périodicité ( $n = 1, 2, 3, 4, 6$ )	Définit l'énergie impliquée dans la rotation d'une partie d'une molécule autour d'une liaison	
Energie des angles impropres	$E_{impropres} = \sum k_\omega (\omega - \omega_0)^2$ $k_\omega$ : constante de force $\omega_0$ : valeur idéale définie dans le champ de force	Assure le maintien de la planéité ou de la chiralité d'un groupe d'atomes déterminés	
Energie de van der Waals	$E_{van\ der\ Waals} = \sum_{i,j} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$ $A_{ij}, B_{ij}$ : paramètres dépendant de la nature des atomes $i$ et $j$ , distants de $r_{ij}$ .		
Energie des interactions électrostatiques	$E_{électrostatique} = \sum_{i,j} \frac{q_i q_j}{\epsilon_{ij} r_{ij}}$ $q_i, q_j$ : charges partielles des atomes $i$ et $j$ , distants de $r_{ij}$ . $\epsilon_{ij}$ : constante diélectrique rendant compte de l'atténuation de l'interaction électrostatique par l'environnement (solvant ou molécule elle-même).		
Energie des liaisons hydrogène	$E_{liaisons\ hydrogène} = \sum \left[ \frac{A'}{r_{AD}^{12}} - \frac{B'}{r_{AD}^{10}} \right] \cos^m \theta_{A-H-D} \cos^n \theta_{AA-A-H}$ $A', B'$ : paramètres dépendant de la nature du donneur et de l'accepteur de liaisons H, distants de $r_{AD}$ . $\theta_{A-H-D}$ : angle formé par l'accepteur (A), l'hydrogène (H) et le donneur (D). $\theta_{AA-A-H}$ : angle formé par l'antécédent de l'accepteur (AA), A et H. $m, n$ : exposants donnés par le type de D et A. $m = 0, 2, 4$ ; $n = 0, 2$		



**Figure 1** : Courbe  $E = f(r_{ij})$  représentative de la fonction d'énergie de van der Waals entre deux atomes  $i$  et  $j$  (potentiel de Lennard-Jones[1]).

De nombreux champs de force sont couramment utilisés tels que AMBER, CHARMM, GROMOS, MM1 à MM4, OPLS ou TRIPOS. Une revue générale des champs de forces disponibles est donnée dans le livre de Young [2]. Nous avons, durant cette étude, utilisé les champs de forces (GLYCAM 06 pour oligosaccharides et FF99SB pour les oligopeptides) du programme de modélisations moléculaire AMBER 10 (Assisted Model Building with Energy Refinement) [3].

Du point de vue du temps de calcul, l'ensemble des approximations utilisées rendent les calculs de MM beaucoup moins exigeants que les calculs de MQ. On peut donc appliquer ces méthodes à un plus grand nombre de systèmes, y compris les biomolécules (protéines, acides nucléiques). A l'heure actuelle, les calculs sont souvent effectués en prenant en compte le solvant, ce qui représente mieux les conditions naturelles du système. Malheureusement, les méthodes de MM comportent aussi quelques limitations:

(1) le nombre de systèmes paramétrés, pour lesquelles la méthode MM donne les résultats validés,

(2) l'omission des électrons pendant le calcul, ce qui signifie que la MM en principe n'est pas capable à résoudre les problèmes où dominant les effets électroniques (réactions chimiques, transferts d'électrons ou de protons etc.).

Différentes stratégies de calcul, utilisés dans ce travail pour étudier les propriétés des systèmes biologiques, vont être décrites. Ces méthodes reposent sur la description énergétique des systèmes.

### 2.1 Méthodes de minimisation d'énergie

Une fois le champ de force choisi, il faut rechercher la conformation possédant l'énergie la plus basse. Une minimisation d'énergie en mécanique moléculaire implique une succession de calculs itératifs de détermination de la structure d'une molécule à partir d'une conformation initiale, laquelle est soumise à une optimisation géométrique complète. Tous les paramètres définissant la géométrie du système sont systématiquement modifiés par petits incréments jusqu'à ce que l'énergie atteigne un minimum. Généralement, cette méthode permet d'atteindre un minimum local le plus proche qui est donc rarement le minimum global correspondant à l'énergie la plus basse. Parmi les algorithmes de minimisation, deux sont très souvent utilisés :

- a. **La méthode de la plus grande pente SD (Steepest Descent) :** dont le principe est le suivant : après avoir calculé l'énergie correspondant à une géométrie initiale, chaque atome est déplacé individuellement selon ses trois coordonnées cartésiennes et l'énergie est recalculée après chaque déplacement. La dérivée première ( $dE/dr_i$ ) correspondant à la pente est calculée, puis les atomes sont déplacés sur une distance de telle sorte que  $dE/dr_i$  diminue le plus vite, c'est à dire dans le sens de la plus grande pente. Cette méthode, principalement efficace pour diminuer les contacts

## Chapitre II : Modélisation moléculaire

---

stériques, est utilisée sur les structures initiales dans les premières étapes de minimisation.

- b. Gradient conjugué CG (Conjugate gradient) :** repose sur le même principe mais permet un ajustement du pas et de la direction de la pente en prenant en compte pour chaque calcul le résultat dérivant du précédent. Cet algorithme évite un comportement oscillatoire autour du minimum, ce qui permet une convergence plus rapide, mais se révèle moins efficace si la structure présente de nombreux contacts stériques.

### 3. Dynamique moléculaire

La dynamique moléculaire (DM) est la méthode la plus fréquemment utilisée pour la simulation des systèmes réels. La DM est capable de simuler un comportement structural en fonction du temps [4, 5, 6, 7, 8] et dans des conditions spécifiées (pression, volume, température). Le principe de base de la DM est l'échantillonnage de l'espace conformationnelle du système étudié par l'intégration des équations du mouvement de Newton pour tous les atomes présents dans le système.

Les simulations de DM sont très importantes pour la recherche du comportement structural des biomolécules en fonction du temps. En utilisant la DM, on peut étudier la flexibilité ou la rigidité des biomolécules, mesurer les interactions intermoléculaires entre la protéine et les ligands ou d'autres biomolécules, calculer l'énergie libre ou bien étudier l'effet du solvant sur la structure des biomolécules. Malheureusement, la DM classique, utilisant des champs de force empiriques, est limitée de la même manière que la MM et donc n'est pas applicable, par exemple, à l'étude des réactions enzymatiques.

#### 3.1 Principe de base

Lorsque l'on ignore les effets quantiques, chaque atome d'une molécule est considéré comme une masse ponctuelle dont le mouvement est déterminé par l'ensemble des forces exercées par tous les autres atomes en fonction du temps. Les intervalles de temps de chaque cycle de

## Chapitre II : Modélisation moléculaire

---

dynamique sont suffisamment courts, environ 1 fs ( $10^{-15}$  s), pour que l'on puisse considérer les forces interatomiques comme constantes. Il devient alors possible d'intégrer numériquement les équations de mouvement et de déterminer ainsi les positions des atomes de la molécule étudiée dans le temps. Cette évolution temporelle –**trajectoire**– d'une molécule est décrite par la résolution des équations de mouvement de Newton. La molécule est donc perçue comme une entité dynamique dont les atomes modifient leurs positions spatiales au cours du temps. Au temps  $t$ , chaque atome  $i$ , de vecteur position  $r_i$  et de masse  $m_i$  subira une accélération  $a_i$  telle que selon la loi de Newton:

$$\vec{F}_i = m_i \vec{a}_i = - \frac{dV}{d\vec{r}_i} = m \frac{d^2 \vec{r}_i}{dt^2}$$

Où  $m_i$  est la masse d'atome  $i$ ,  $a_i$  est son accélération,  $t$  est le temps,  $r_i$  sont les coordonnées cartésiennes de l'atome, et  $V$  est l'énergie potentielle.

On peut alors intégrer les équations de mouvement et obtenir une trajectoire pour chaque atome en fonction du temps. Pour cela on utilise l'algorithme de précision de deuxième ordre « **l'algorithme de Verlet** [9] » (en troncature de la série de Taylor). Sa validité est aisément testée à l'aide de la troisième loi fondamentale de la thermodynamique qui implique que l'énergie totale d'un système doit être conservée au cours d'une simulation. En général le potentiel d'interaction  $V_{total}$  utilisé est un potentiel empirique. Ce qui veut dire qu'il contient les contributions des énergies de liaisons, des angles de valence, des angles dièdres et des interactions de van der Waals et électrostatiques.

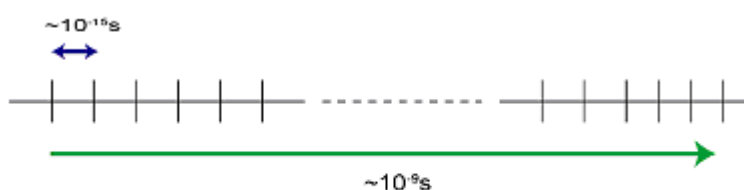
### 3.2 Algorithme SHAKE

Afin de pouvoir tenir compte toutes les interactions entre atomes, le pas d'intégration  $\Delta t$  (algorithme de Verlet) doit être au moins 10 à 20 fois inférieur aux fréquences de vibration les plus élevées. Ce qui est généralement de l'ordre de 1 femto seconde (1 fs =  $10^{-15}$  s). En effet, un des problèmes de fond de la dynamique moléculaire est l'inadéquation entre l'échelle du

## Chapitre II : Modélisation moléculaire

---

temps des phénomènes physiques observés et celle des simulations accessibles par la technologie actuelle. Une simulation courante dure plusieurs nanosecondes (Figure 2), tandis que les échelles de temps des processus dynamiques observés varient du femto seconde (e.g. l'élongation des liaisons covalentes) à la seconde (e.g. la vibration des cycles aromatiques des résidus d'acides aminés). La **procédure SHAKE** [10], qui néglige les élongations des liaisons, permet d'allonger le pas d'intégration jusqu'à 2, voire 4 femto secondes. En effet, cet algorithme permet de fixer la longueur de certaines liaisons durant une dynamique moléculaire. Le plus souvent, SHAKE n'est appliqué que sur des liaisons impliquant un atome d'hydrogène dont les vibrations sont les plus rapides. Le choix du pas d'intégration dépend également de la température du système. Dans les simulations à haute température, il faut éviter des pas trop grands, qui entraînent un éclatement du système.



**Figure 2** : Un pas d'intégration  $\Delta t$  est d'ordre de femto seconds ( $10^{-15}$  s), alors que la trajectoire dure généralement quelques nano seconds ( $10^{-9}$  s).

### 3.3 Procédure d'une dynamique moléculaire

Les simulations de dynamique moléculaire sont généralement réalisées en quatre étapes (figure 3) sous des conditions isotherme-isobare.

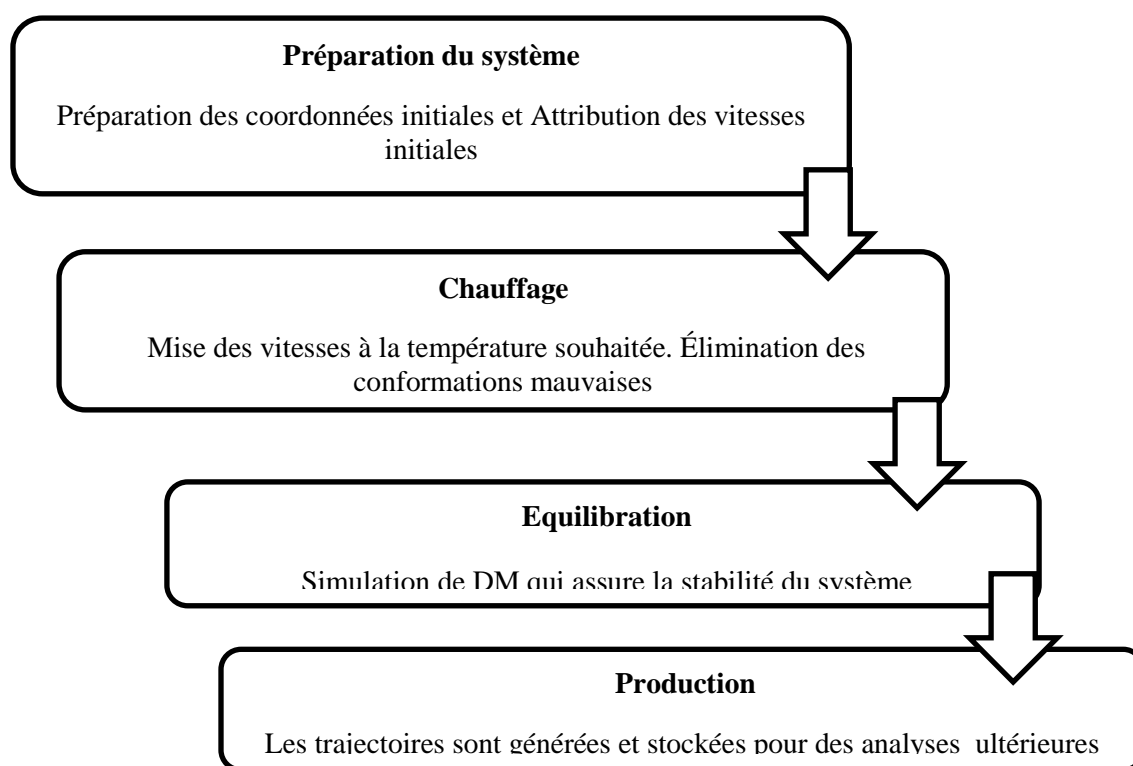
Dans la première étape, le système, provenant des données RMN, ou RX est préparée, en ajoutant des atomes manquants, et soumis à des cycles de minimisation. La deuxième étape consiste à chauffer le système afin d'éliminer les mauvais contacts, ce qui augmente affecte de nouvelles vitesses aux atomes. Les vitesses initiales sont généralement déterminées à partir de

## Chapitre II : Modélisation moléculaire

la distribution de Maxwell-Boltzmann. La troisième étape est appelée équilibration, dans laquelle l'énergie, la température et le RMSD d'un système convergent vers des valeurs stables ce qui permet la collection des trajectoires (étape de production) qui seront analysées successivement. L'écart quadratique moyen RMSD (Root Mean Square Deviation) entre les atomes est défini comme suit:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atomes}} d_i^2}{N_{atomes}}}$$

$N_{atomes}$  est le nombre d'atomes prises en compte pour le calcul RMSD, et  $d_i$  est la distance entre les coordonnées de l'atome  $i$  lorsque deux structures du même système sont superposées. Afin de réaliser une simulation de dynamique moléculaire (DM), plusieurs choix devrait être faite avant de commencer les calculs.



**Figure 3 :** Protocole général pour des simulations DM



### **3.4 Pas du temps d'intégration**

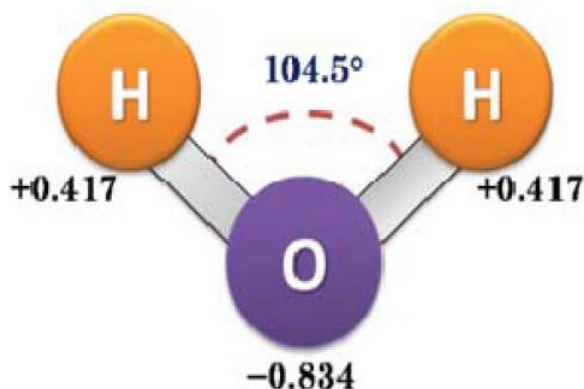
Le pas du temps  $\Delta t$  doit être petit par rapport à la période du mouvement le plus rapide du système simulé. Pour les protéines, les mouvements les plus rapides sont les vibrations d'élongation des liaisons entre les hydrogènes et les atomes lourds ( $\tau = 10\text{fs}$ ). Une règle générale pour choisir un pas de temps est :  $\tau / \Delta t \sim 20$

Ainsi, pour les protéines, un pas de temps de 0,5 fs serait couteux en ressources informatiques le long d'une simulation MD. L'algorithme SHAKE a été développé pour garder les longueurs des liaisons (X-H) constantes, permettant davantage des pas de temps plus longs [11].

### **3.5 Hydratation du système :**

Dans les simulations de dynamique moléculaire, le solvant peut être traité de différentes manières. L'effet de solvant peut être obtenu par l'utilisation d'un constant diélectrique dépendant de la distance (calculs *in vacuo*). Des approches alternatives ont été développées où le solvant est considéré comme un milieu continu entourant le soluté. Elles peuvent simuler les effets de solvant en utilisant des formules pour calculer les champs électrostatiques et des conditions de surface afin d'évaluer les interactions hydrophobes.

Le modèle du solvant continu le plus utilisé, PB-SA, est basée sur des algorithmes capables de résoudre l'équation de Poisson-Boltzmann pour tenir compte des interactions électrostatique, y compris les termes de surface [12]. Un autre modèle de solvant continu, plus simple et avec une précision similaire, utilise l'approximation de Born généralisée pour limiter les effets de solvation [13]. Cependant, le but principal de ces simulations est de reproduire le comportement des systèmes biologiques qui existent dans la nature, dans un environnement hydraté. Ainsi, un moyen plus fiable pour modéliser un environnement aqueux est de créer une boîte de solvant explicite autour de la molécule.



**Figure 4 :** modèle d'eau TIP3P

Plusieurs modèles de l'eau ont été développés. Ils ont été classés par le nombre des points utilisés pour définir le modèle, si la structure est rigide ou flexible, et si le modèle inclut ou non les effets de polarisation. Le modèle le plus simple et le plus utilisé est le modèle TIP3P [14]. Il est caractérisé par trois sites d'interaction, correspondant aux trois atomes de la molécule d'eau. Une charge ponctuelle est attribuée à chaque atome pour obtenir les paramètres de Lennard-Jones (figure 4).

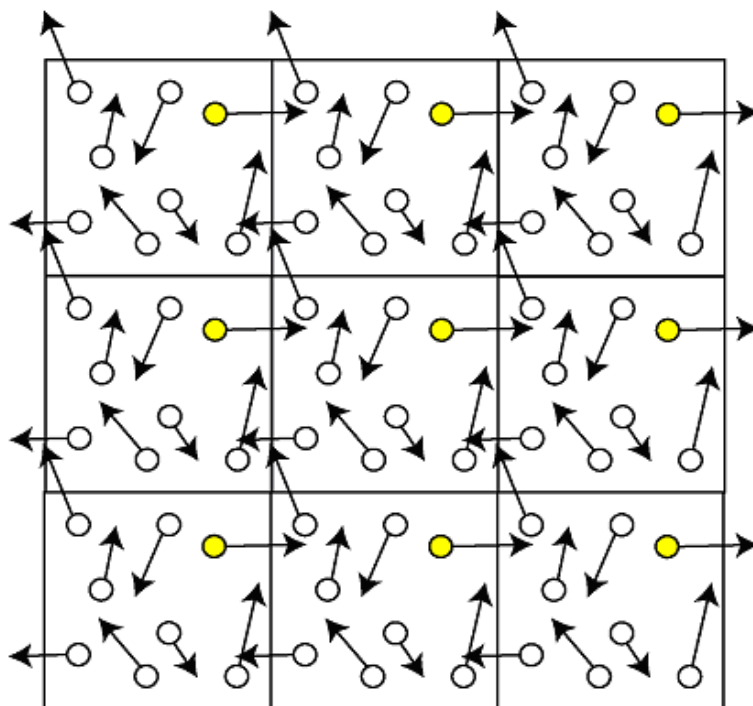
### **3.6 Conditions aux limites périodiques, PBC (Periodic Boundary Conditions)**

Dans un modèle biologique réaliste, les molécules d'eau devraient être situées autour du soluté. Si le modèle biologique est placé dans une boîte de molécules d'eau, une partie des atomes de soluté / solvant, pendant la simulation, se trouvent au bord de la boîte, en contact avec le vide qui l'entoure. Pour éviter cette image artificielle du volume du liquide et pour assurer une immersion complète du soluté au cours des simulations, des conditions aux limites périodiques sont employées.

Le solvant est introduit dans le modèle sous forme d'une boîte au sein de laquelle, on place le soluté étudié. En pratique, on utilise un artifice de calcul appelé **conditions périodiques** qui permet de simuler des systèmes infinis. Celui-ci implique que la boîte de solvant est entourée de 26 copies identiques à elle-même (Figure 5). Quand une molécule de la boîte centrale sort par une face, elle est immédiatement remplacée par une copie qui apparaît par la face

## Chapitre II : Modélisation moléculaire

opposée. Pour reproduire des propriétés d'un grand système, une molécule dans la boîte centrale ne peut interagir qu'avec une copie de chaque particule (« minimum image convention »). En conséquence, la troncature de potentiel (cut-off) doit être inférieure à la demi-longueur des trois dimensions de boîte.



**Figure 5:** Conditions périodiques dans l'espace bidimensionnel.

Ainsi les conditions périodiques rendent possibles des simulations à pression et température constantes tout en maintenant fixe le nombre de molécules de l'ensemble moléculaire. Elles permettent donc de simuler des processus chimiques dans l'ensemble thermodynamique isotherme-isobare ( $N, T, P$ ) qui correspond aux conditions généralement rencontrées au laboratoire. La méthode de Berendsen [15] sert à réguler la température. Le système est lié avec un bain externe et les vitesses de chaque pas sont multipliées par un facteur qui dirige la température vers la valeur désirée. De la même manière, la pression peut être contrôlée en ajustant le volume du système.

### *3.7 Traitement des interactions non liées*

Les interactions non-liées, composées par des interactions électrostatiques et de van der Waals, sont considérées comme un problème pour la simulation de dynamique moléculaire en termes de coût de calcul. Ainsi, des approximations sont nécessaires.

#### *a. Cut-off*

Comme la force des interactions de van der Waals entre les atomes diminue avec la distance interatomique, on peut tronquer les interactions de Lennard Jones à une distance cut-off spécifique. Une distance cut-off optimale est déterminée en évaluant l'effet de la variation de l'énergie totale du système. Lorsque l'énergie totale converge rapidement à un cut-off particulier, le gain en allant aux grands cut-off est faible et ne justifie plus le coût supplémentaire de calcul. Si les conditions aux limites périodiques sont appliquées, le cut-off doit être inférieure ou égale à la moitié de la taille de la boîte de solvant. Cela empêche la possibilité pour une particule d'interagir avec l'un de ses images de la même particule au même temps.

#### *b. Sommation d'Ewald (Particle Mesh Ewald, PME)*

Le traitement des interactions électrostatiques à longues distances présente toujours un problème, notamment dans les simulations de dynamique moléculaire de molécules chargées. Une des solutions possibles, l'utilisation d'une grande distance de troncature, augmente le temps de calcul en fonction du cube de la distance. Comme cela a été démontré dans plusieurs simulations de molécules d'ADN [16-17], l'introduction de la **sommation d'Ewald** [18] semble plus efficace et plus précise. Cette méthode est basée sur un algorithme rapide qui calcule l'énergie électrostatique complète d'une boîte périodique, dans un système macroscopique infini d'images répétées. Elle évite ainsi l'évaluation standard des interactions non-liées et utilise sa propre fonction composée d'une sommation dans l'espace direct et

d'une autre dans l'espace réciproque pour calculer des interactions de type Lennard-Jones et les interactions électrostatiques.

#### 4. Analyse conformationnelle :

L'analyse conformationnelle est une méthode qui permet de calculer l'énergie relative associée à la conformation d'une molécule [19]. Son but est donc de trouver le plus de minima possible, et de calculer la population de Boltzmann [20].

$$\left. \begin{array}{l} N_i = N_0 \exp(-E_i / KT) \\ N_j = N_0 \exp(-E_j / KT) \end{array} \right\} \Rightarrow N_i / N_j = \exp(-\Delta E_{ij} / KT)$$

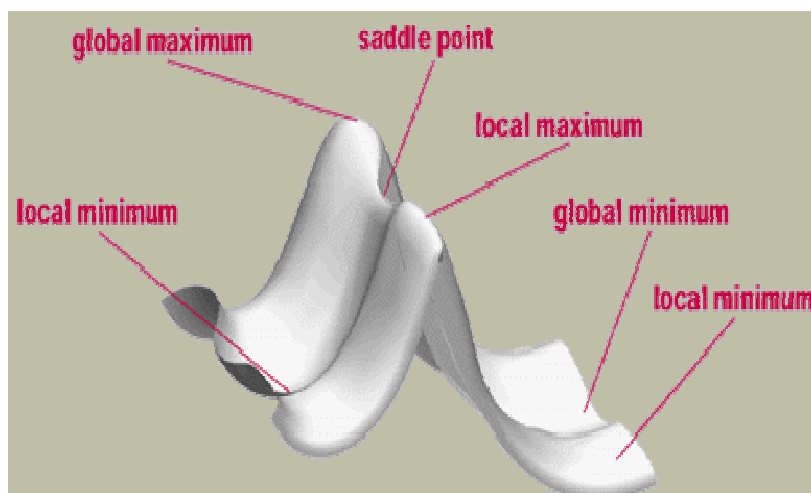
$E_i$  : Energie de la molécule.

$K$  : Constante de Boltzmann.

$T$  : Température (Kelvin).

Les molécules flexibles présentent un nombre astronomique de conformations différentes, mais les plus importantes sont celles ayant une énergie libre suffisamment basse pour qu'elles puissent être peuplées à des températures raisonnables [21].

Un des modèles les plus réussis utilisés en chimie computationnelle est la surface d'énergie potentielle (Potential Energy Surface « PES ») (figure 6). C'est une représentation remarquable des molécules individuelles, des complexes moléculaires et des interconversions conformationnelles. Elle permet aussi de faire le lien entre les concepts chimiques clés et les caractéristiques géométriques de base des surfaces. Dans le cas de l'analyse conformationnelle, les minima et les points de selle (saddle points) sur la PES sont les caractéristiques les plus importantes. Les minima correspondent à des conformations d'énergie minimale et les points de selle à des états de transition associés à des interconversions conformationnelles [21].



**Figure 6** : surface d'énergie potentielle

En analyse conformationnelle, un grand nombre de géométries de départ de haute énergie sont générées, minimisées et stockées. Pour que la recherche soit minutieuse, les géométries de départ brutes doivent couvrir toute la surface d'énergie potentielle. Dans le cas contraire, on n'est pas sûr de trouver tous les minima [20].

Certaines méthodes de recherche conformationnelle sont destinées au traitement des petits cycles, des peptides acycliques et des molécules organiques [22, 23, 24], d'autres pour les petites protéines [25], des polymères [26] et les hydrates de carbone [27].

Dans le paragraphe précédent, on suppose une surface d'énergie potentielle parfaite. Toutefois, en pratique, la fonction d'énergie potentielle et la surface d'énergie sont sujettes à beaucoup d'approximations. En effet, il y a de nombreuses fonctions d'énergie potentielle utilisées basées soit sur la mécanique quantique soit sur la mécanique moléculaire. Malgré la précision des potentiels de la mécanique quantique, ceux de la mécanique moléculaire sont exclusivement utilisés en analyse conformationnelle puisqu'elle permet de traiter de grosses molécules. Il est à noter cependant, que des champs de force différents génèrent des surfaces de potentiel différentes, ainsi la localisation des minima locaux dépend du champ de force

## Chapitre II : Modélisation moléculaire

---

utilisé. Les différents champs de force donnent en général les mêmes résultats pour de petites molécules, mais présentent des divergences dans les conformations des molécules plus complexes [21]. Par exemple, il apparaît 262 conformations d'énergie minimale pour le cycloheptadecane, la plus basse étant à 3 Kcal/mol avec MM2, et seulement 134 conformations avec MM3 [23].

L'analyse conformationnelle permet de trouver les minima et les points de selle, et d'identifier les interconversions impliquant des petites barrières énergétiques. Pour cela il faut générer un ensemble de structures d'essai, et passer à la minimisation ensuite. Dans un système moléculaire sans contrainte et avec un nombre d'atomes  $n > 2$ , le nombre de degrés de liberté est égal au nombre de coordonnées internes c'est à dire  $3n - 6$ . Lors de la minimisation de l'énergie, l'ensemble des structures d'essai est généré en faisant varier les degrés de liberté les plus flexibles à savoir les angles dièdres [21].

De façon générale, pour les liaisons simples, la modification d'un angle dièdre coûte moins cher en Kcal/mol que la déformation d'un angle de valence qui, à son tour, est plus facile que l'élongation d'une liaison. On dit que les angles dièdres sont plus mous que les autres coordonnées internes [28]. Pour pouvoir faire varier tous les angles dièdres d'une molécule, on dresse une grille pour obtenir toutes les possibilités combinatoires. Ainsi, une molécule avec trois angles dièdres va générer 216 conformations si l'incrément d'angle est de  $60^\circ$ . Pour une molécule ayant six angles dièdres on atteindra déjà 46656 conformères, tandis qu'avec neuf angles on arrive à plus de 10 millions de conformations possibles [19].

La recherche et la localisation des structures des états d'équilibre et de transition sur l'hypersurface d'énergie potentielle d'un système moléculaire sont nécessaires pour la détermination des propriétés thermodynamiques et cinétiques, de même que pour

## Chapitre II : Modélisation moléculaire

---

l'interprétation de spectres de vibration et de rotation. Des techniques spéciales d'optimisation de géométrie ont été élaborées pour cela [29].

L'optimisation de la géométrie d'un système moléculaire consiste à trouver la géométrie de ce système qui correspond à l'énergie la plus basse du système. C'est pourquoi ces méthodes sont dites de minimisation. Il est important de préciser que ces dernières s'appliquent au voisinage des points stationnaires, donc pour des petites variations de la géométrie. Ceci est dû au fait que ces méthodes partent toutes du développement de l'énergie en série de Taylor qui n'est valable que pour un petit écart de la position d'équilibre recherchée.

Elles peuvent être en défaut et déceler un minimum global englobant ainsi plusieurs minimums locaux, si la géométrie de départ est trop éloignée de celle de l'état d'équilibre [30].

Donc, l'analyse conformationnelle reste un moyen efficace de prédiction des conformations stables d'une structure moléculaire correspondant aux minima de son énergie intramoléculaire. L'énergie calculée mesure la différence entre l'énergie de la structure moléculaire considérée et celle d'une structure hypothétique dont toutes les coordonnées prendraient leurs valeurs de référence. Sa valeur absolue dépend du champ de force considéré, c'est-à-dire des fonctions d'énergie potentielle et des paramètres utilisés. Les valeurs de cette énergie sont toutefois très utiles lorsqu'il s'agit de comparer la stabilité relative des conformations ou des stéréo-isomères d'une même molécule [19].



### Références

- [1] J. E. Lennard-Jones, *Proc.Camb.Phil.Soc.* 27: (1931), 469.
- [2] D. C. Young, "Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems". John Wiley and Sons, New York. (2001).
- [3] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. J. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J Am Chem Soc*, 117: (1995), 5179-5197.
- [4] O. M. Becker, A. MacKerell, B. Roux, M. Watanabe, "Computational biochemistry and biophysics". Dekker M: New York, (2001).
- [5] R. Eklund, "Computational analysis of carbohydrates: dynamical properties and interactions". Stockholm University, Stockholm (Sweden), (2005).
- [6] R. C. Walker, "The development of a QM/MM based linear response method and its application to proteins". Imperial College London, London (U.K.), (2003).
- [7] A. Kranjc, "Predicting structural determinants and ligand poses in proteins involved in neurological diseases: bioinformatics and molecular simulation studies International School for Advanced Studies (SISSA/ISAS)", Trieste (Italy), (2009).
- [8] C. J. Cramer, "Essentials of computational chemistry: theories and models". John Wiley & Sons: New York, (2004).
- [9] L. Verlet, *Mol. Phys. Rev*, 159: (1967), 98-103.
- [10] W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed. Eng*, 29: (1990), 992-1023.
- [11] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *Journal of Computational Physics*, 23 (3): (1977), 327-341.

## Chapitre II : Modélisation moléculaire

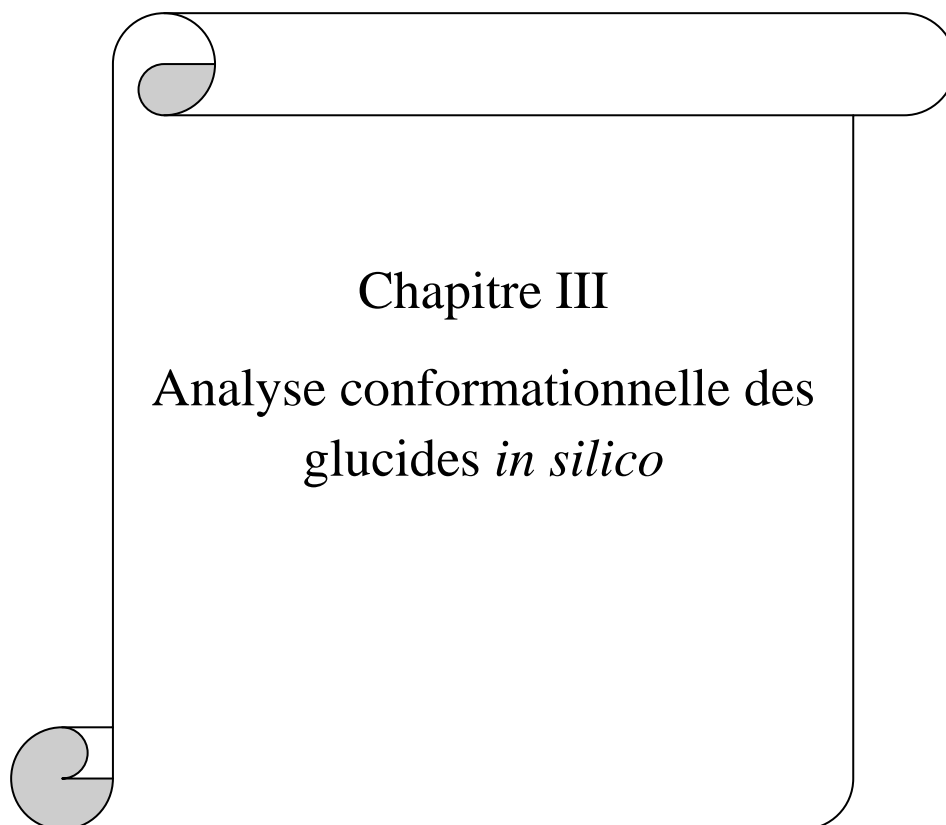
---

- [12] M. K. Gilson, K. A. Sharp, B. H. Honig, *Journal of Computational Chemistry*, 9 (4): (1988), 327-335
- [13] M. Wojciechowski, B. Lesyng, *Journal of physical chemistry. B, Condensed matter, materials, surfaces, interfaces, & biophysical chemistry*, 108 (47): (2004), 18368-18376
- [14] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *The Journal of Chemical Physics*, 79: (1983), 926-935.
- [15] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, R. J. Haak, *J. Phys. Chem.*, 81: (1984), 3684-3690.
- [16] T. E. Cheatham III, J. L. Miller, T. Fox, T. A. Darden, P. A. Kollman, *J. Am. Chem. Soc.* 117: (1995), 4193-4194.
- [17] D. M. York, W. Yang, H. Lee, T. A. Darden, L. G. Pedersen, *J. Am. Chem. Soc.*, 117: (1995), 5001-5002.
- [18] U. Essmann, L. Perera, M. L. Berkowitz, *Journal of Chemical Physics*, 103: (1995), 8577-8593.
- [19] H. Dugas, « Principes de base en modélisation moléculaire, Aspects théoriques et pratiques », quatrième édition, Librairie de l'Université de Montréal, (1996).
- [20] J. Bartol, P. Comba, M. Melter, M. Zimmer, *J. Comput. Chem.*, 20: (1999), 1549.
- [21] G. Keserü, I. Kolossvary, "Molecular Mechanics and Conformational Analysis in Drug Design", Department of Chemical Information Technology, Technical University of Budapest, Blackwell Science Ltd, (1999).
- [22] J. Thomas Ngo and M. Karplus, *J. Am. Chem. Soc.*, 119: (1997), 5657.
- [23] I. Kolossvary and W. C. Guida, *J. Am. Chem. Soc.*, 118: (1996), 5011.
- [24] H. Senderowitz, F. Guarieri, W. C. Still., *J. Am. Chem. Soc.*, 117: (1995), 8211.

## Chapitre II : Modélisation moléculaire

---

- [25] M. J. Dudeck, H. A. Scheraga, *J. Comput.Chem*, 11: (1990), 121.
- [26] E. Leontidis, J. J. de Pablo, M. Laso, U. W. Suter, *Adv. Polym. Sci*, 116: (1994), 283.
- [27] C. W. Von Der Lieth, T. Kozar, W. E. Hull, *Theochem. J. Mol. Struct*, 395: (1997), 225.
- [28] J. S. Lomas, « la mécanique moléculaire, une méthode non quantique pour le calcul de la structure et de l'énergie d'entité moléculaire », l'actualité chimique, (1986), chapitre7.
- [29] A. Hinchliffe, *Modelling Molecular structures*, John Wiley and Sons, New York, (1994).
- [30] S. Djennane – Bousmaha, *Thèse de doctorat*, USTHB Alger, Algérie, (1999).



### **I. Modélisation des glucides**

On a déjà vu dans le chapitre précédent les facteurs qui déterminent la structure tridimensionnelle des glucides et plus particulièrement des oligosaccharides. L'analyse conformationnelle d'un oligosaccharide s'effectue par étapes. Il faut d'abord identifier les liaisons entre les monosaccharides qui composent la structure et analyser la surface d'énergie potentielle de chaque liaison glycosidique. La connaissance des conformations de plus basse énergie pour chaque fragment disaccharidique permet de construire un modèle (ou plusieurs) de la structure globale de la molécule dans une conformation de basse énergie. Ces modèles serviront de structures de départ pour l'analyse de la surface d'énergie potentielle à  $n$ -dimensions. La dernière étape est la génération des familles des conformations de basse énergie et l'analyse des résultats.

La complexité de la détermination des structures tridimensionnelles des glucides provient de la grande mobilité de ces molécules et de la présence des conformations différentes qui coexistent en équilibre [1].

Les méthodes computationnelles ont été développées, en fournissant des outils complémentaires aux informations structurales RX et RMN des glucides [2].

La variabilité de l'état anomérique, la position de liaison de, taille du cycle, la forme et la topologie représentent un problème pour les études des conformations des hydrates de carbone in silico [3]. Pour ces raisons, des champs de force spécifiques ont été développés avec succès.

#### **1. Champs de force pour les glucides**

Pour étudier les structures et les propriétés des glucides en utilisant des techniques de modélisation moléculaire, les fonctions d'énergie potentielle et des paramètres spécifiques pour cette catégorie de molécules sont nécessaires. Des champs de force appropriés pour les

hydrates de carbone ont été développés, dont le but de reproduire les effets particuliers qui influent sur leurs propriétés structurales en solution [4].

Le comportement exocyclique du groupement hydroxyméthyl définie par l'angle  $\omega$  (O5-C5-C6-O6) et sa préférence pour les conformations *gauche* peuvent être reproduit par l'introduction des facteurs qui modifient légèrement les interactions 1-4 non liées [5].

Les interactions 1-4 non liées définissent l'influence, en termes des potentiels électrostatiques et de Van der Waals, des deux atomes situés à chaque extrémité d'un angle de torsion. Les interactions 1-4 non liées ne sont pas traitées de la même manière dans tous les champs de force et cela pourrait être un problème dans les simulations des systèmes complexes dans lesquels deux champs de force différents doivent être utilisée. Dans ces cas, le traitement séparé des interactions non liées 1-4 peut assurer une compatibilité entre les champs de force.

Dans la littérature, plusieurs revues décrivent et comparent la performance des champs de force pour les glucides [6, 7]. Le champ de force le plus utilisé, GLYCAM06, est décrit dans ce chapitre en détail.

### 1.1 Glycam06

Le champ de force Glycam06 est très cohérent pour la modélisation des carbohydrates, des glycoprotéines et des glycolipides [8, 9]. Il peut être utilisé pour décrire les propriétés physico-chimiques des dérivés de carbohydrate complexes et il est entièrement compatible avec le champ de force AMBER [10]. Glycam06 peut être utilisé dans des package de simulation autre que AMBER à travers l'utilisation des outils de conversion de fichier appropriés [11]. Les paramètres sont développés en tenant compte un ensemble d'essai de 100 molécules à partir des familles chimiques des hydrocarbures, des alcools, des éthers, des

amides, des esters, des carboxylates, des molécules avec des groupements fonctionnels mixtes et aller aux données de chimie quantique.

Pour faciliter la transférabilité des paramètres, toutes les séquences atomiques ont un ensemble de termes de torsion défini explicitement, sans aucun terme générique, et les mêmes paramètres PARM94 utilisés dans AMBER, sont utilisés pour modéliser les termes de van der Waals [10]. Aucun facteur d'échelle pour le traitement des interactions 1-4 n'est introduit pour reproduire l'effet gauche sur l'angle  $\omega$  [12].

Dans Glycam06, les effets stéréo-électroniques qui influencent les variations de liaison et d'angle de l'atome de carbone anomérique sont inclus dans un type unique d'atome anomère. Cette fonctionnalité permet d'imiter la bague retournement observé en monomères glycosidiques qui se produisent, par exemple, lors d'événements catalytiques [13].

Une comparaison avec les données expérimentales a confirmé que le champ de force Glycam06 est capable de bien reproduire les énergies de rotation et les caractéristiques des carbohydrates s'il est combiné avec un ensemble de charge approprié, pour les molécules fortement polaires pour lesquels des conditions empiriques ont été introduites pour corriger les erreurs des énergies de torsion [8]. Dans le champ de force Glycam06, les charges atomiques partielles sont calculées résidu par résidu. Pour chaque résidu, une simulation de dynamique moléculaire de 50-100ns est effectuée, 100-200 snapshots sont extraites et les charges atomiques sont ajustées en utilisant le potentiel électrostatique (ESP) donné par les méthodes de mécanique quantique. Cette stratégie est adoptée pour intégrer la dépendance de conformations moléculaires sur des charges partielles. Puisque les charges portées par les atomes d'hydrogène aliphatiques C-H ne sont pas significatives pour la reproduire les moments dipolaires [14-15], des contraintes(RESP) sont employées dans la procédure d'ajustement ESP afin de s'assurer que les charges portées par tous les atomes d'hydrogène

aliphatiques sont nuls. Une contrainte de charge optimale RESP égale à 0,01 est appliquée, en se basant sur des simulations des réseaux cristallins des glucides [16].

Un outil de calcul simple a été développé pour faciliter la préparation des fichiers nécessaires pour exécuter des simulations de dynamique moléculaire en utilisant Glycam implanté dans AMBER package ([www.glycam.com](http://www.glycam.com)).

### 1.2 Autres champs de force pour les carbohydrates

GROMOS-45A4, CHARMM et OPLS-AA sont d'autres champs de force utilisés avec Glycam06, pour décrire les propriétés conformationnelles des hydrates de carbone dans les calculs [7] (figure 1).

Le champ de force GROMOS a été développé au début pour les simulations de dynamique moléculaire des protéines, des nucléotides, ou des sucres dans des solutions aqueuses ou apolaires ou sous forme cristalline, mais récemment, il a été modifié pour inclure les effets anomériques des mono et oligopyranoses [17]. Comme dans Glycam06, les méthodes de mécanique quantique sont utilisées pour le calcul des constantes de force de liaison et d'angle de valence, tandis que les paramètres d'angle dièdre et le terme de Van der Waals sont directement pris des versions précédente de GROMOS [18, 19]. Une procédure d'ajustement des potentiels électrostatiques ESP, avec des contraintes sur les atomes d'hydrogène aliphatique, est choisie pour reproduire le potentiel électrostatique, en utilisant un trisaccharide comme un modèle pour le développement de charges [17]. Aucune distinction n'est faite entre les monomères  $\alpha$  et  $\beta$  en termes de charges et de type d'atome anomérique et de facteurs électrostatique-Van der Waals 1-4 ne sont pas introduites pour reproduire correctement les effets *gauche* sur les angles  $\omega$ . De longues simulations de dynamique moléculaire de 20 ns dans les eaux explicites (SCP water models [20]) ont été utilisées pour valider le champ de force, montrant la capacité pour prédire correctement les effets stéréo-



électroniques et les conformations cycliques les plus stables, mais parfois il présente un défaut dans la reproduction de leurs énergies [17]. Récemment, GROMOS a été proposé comme le champ de force le plus adaptée pour simuler le transit des conformations chaise  ${}^4C_1$  au bateau oblique (skew boat) des résidus d'acide iduronique dans les simulations de dynamique moléculaire de l'héparine [21].

Le champ de force CHARMM a été récemment développé pour le glucopyranose et ses diastéréoisomères [22]. Plusieurs modifications pour les glucides ont été proposées afin d'élargir l'utilisation de ce champ de force pour des sucres de cycle à cinq et des oligosaccharides [23-25]. Les mêmes procédures de paramétrisation et du traitement des interactions 1-4 non liées sont utilisées pour assurer une compatibilité totale avec les autres champs de force biomoléculaires CHARMM [26-28]. Des ensembles de paramètres préliminaires sont créés en utilisant des modèles de petites molécules correspondant à des fragments des cycles pyranose puis successivement appliquée pour compléter les monosaccharides pyranose. Les paramètres des angles dièdres manquants sont développés en ajustant plus de 1800 énergies conformationnelles de l'hexopyranose obtenues par les méthodes de mécanique quantique. Les charges atomiques partielles et les valeurs des paramètres de Lennard-Jones, prises à partir des versions précédentes de CHARMM, sont ajustées afin de reproduire les énergies et les distances d'interaction carbohydrate-eau des calculs de mécanique quantique, et améliorés davantage pour reproduire les chaleurs de vaporisation et les volumes moléculaires expérimentaux des liquides. En utilisant des simulations de dynamique moléculaire dans le modèle TIP3P des molécules-d'eau, le champ de force, avec différent type d'atome pour les anomères  $\alpha$  et  $\beta$ , a été validé puisqu'il reproduit des propriétés expérimentales calculées par la mécanique quantique.

Le champ de force OPLS a été étendu récemment pour inclure les carbohydrates [29]. Le champ de force OPLS-AA pour carbohydrates est modifié en appliquant un facteur d'échelle sur les interactions électrostatiques intramoléculaire et en optimisant les paramètres de torsions. Ce nouveau modèle de "scaling" permet une meilleure reproduction des énergies relatives des conformations en phase gazeuse de l'éthanoléol et de plusieurs hexopyranoses. Les simulations par dynamique moléculaire d'hexopyranoses en solution qui sont conduites avec le nouveau champ de force, nommé "OPLS-AA-SEI" (Scaling Electrostatic Interactions), présentent un meilleur échantillonnage des groupements hydroxyles. Dans le cas des disaccharides, un excellent accord est obtenu entre les simulations par dynamique moléculaire et les données expérimentales. Le schizophyllane est un polysaccharide pour lequel une transition conformationnelle a été observée expérimentalement dans l'eau mais dont le mécanisme moléculaire n'a pas été élucidé. Afin d'élucider plus avant le mécanisme de cette transition, des simulations par dynamique moléculaire sont entreprises avec le champ de force OPLS-AA-SEI [29].

Autres champs de force sont utilisés pour comprendre les propriétés des glucides *in silico*. En particulier, le champ de force MM3, destiné initialement pour les hydrocarbures [30], mais maintenant applicable à une large série de composés [31, 32], est largement utilisé pour la construction des cartes adiabatiques des disaccharides. Le champ de force TRIPOS de mécanique moléculaire est conçu pour simuler les peptides et les petites molécules organiques [33], mais une extension de paramètre pour les oligosaccharides comporte des fragments de glycosaminoglycanes sulfatés et des glycopeptides [33, 34]. Le champ de force TRIPOS implémenté dans SYBYL package (Tripos Associate, St. Louis, MO) est utilisé pour l'optimisation des géométries.

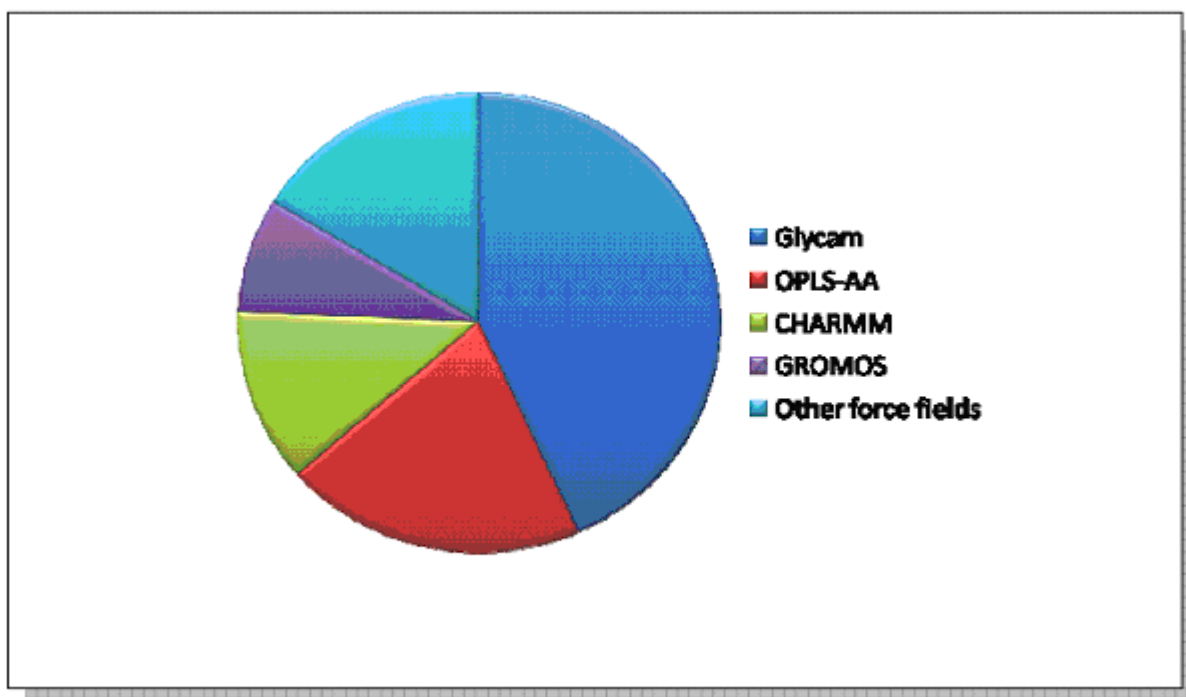


Figure 1 : usage approximative des champs de force (GLYCAM, CHARMM, OPLS-AA, GROMOS) pour les carbohydrates dans les cinq ans passés selon le site ISI — Web of Science (<http://scientific.thomsonreuters.com/products/wos/>) [7].

### 2. Etudes conformationnelles *in silico* des carbohydrates

La recherche des conformations bioactives est un élément important pour comprendre la relation entre la structure et l'activité biologique d'une molécule. Les Propriétés conformationnelles des hydrates de carbone et les surfaces d'énergie potentielles respectives peuvent être évaluées par des recherches conformationnelles *in silico* [36,3].

#### 2.1 Procédures de recherche systématiques

La recherche systématique est l'une des méthodes les plus utilisées dans l'analyse conformationnelle en chimie computationnelle. Elle est réalisée en variant systématiquement les angles de torsion d'une molécule pour générer toutes les conformations possibles et évaluer l'énergie potentielle associée à chaque conformation.

Les angles de torsion  $\phi$  et  $\psi$  définissent la liaison glycosidique des carbohydrates. Ils influencent la conformation globale des oligosaccharides. Les angles glycosidiques peuvent être tournés par des petits incréments et l'énergie potentielle correspondante peut être calculée selon le champ de force appliqué. Toutes les informations de l'énergie en fonction des changements  $\phi / \psi$  tirés de la recherche conformationnelle peuvent être tracées pour donner des cartes adiabatiques dans lesquelles la position des minima et les barrières de transition peuvent être visualisées.

Des cartes adiabatiques rigides sont calculées uniquement en tenant compte la flexibilité des angles  $\phi$  et  $\psi$ . Au cours de la rotation  $\phi / \psi$ , les mouvements des monomères sont instantanés, ce qui va provoquer des gênes stériques et, par conséquent, des conformations de haute énergies. Pour décrire correctement le comportement des carbohydrates, plusieurs de degrés de liberté peuvent être attribués à l'ensemble des structures glycosidiques, bien sûr, ne pas ignorer les effets stéréo-électroniques.

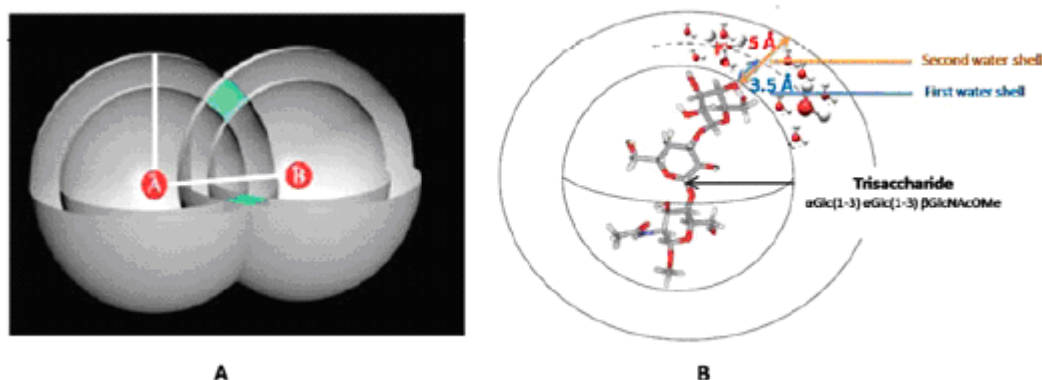
Des cartes adiabatiques sont construites en appliquant la méthode décrite précédemment. Une minimisation complète après chaque étape, à l'exception des deux angles de torsion glycosidiques, permet l'ajustement des longueurs de liaisons et des angles, et la diminution des énergies conformationnelles [37]. Cependant, la minimisation commence à partir d'une conformation particulière du sucre due à la gêne stérique, peut ne pas atteindre des minima stables. A chaque étape de la recherche conformationnelle, toutes les combinaisons des orientations possibles du fragment disaccharidique, doivent être prises en compte.

Les Cartes adiabatiques sont largement utilisées dans la chimie computationnelle pour comprendre les propriétés énergétiques des glucides [38-40].

### 2.2 Rôle du solvant dans les études conformationnelles des carbohydrates

Les carbohydrates, comme les autres systèmes, sont fortement influencés par le solvant. Ainsi, lorsque les études conformationnelles sont réalisées, l'effet du solvant doit être pris en considération [41,42]. Le modèle TIP3P de l'eau a été couramment utilisé pour décrire le comportement des interactions carbohydrates-eau [43]. Lors d'une solvatisation explicite d'eau, les interactions eau-soluté peuvent être caractérisées par l'évaluation de la fonction de la distribution radiale, de la première couche d'hydratation et du temps moyen de repos.

La fonction de la distribution radiale est une méthode qui décrit statistiquement la variation de la densité de l'eau en fonction de la distance d'un atome particulier. Le calcul de la fonction de la distribution radiale 2D, permet l'identification d'intersection formé par deux couches sphériques entourant deux atomes d'oxygène du soluté dans lequel un atome d'oxygène des molécules d'eau peut être trouvé [44]. (**Figure 2 A**). La première couche d'hydratation donne le nombre de molécules d'eau à moins de  $3.5\text{\AA}$  à partir des atomes d'oxygène du soluté. (**Figure 2 B**).



**Figure 2 :** Illustration du volume d'intersection formé par deux couches sphériques entourant deux atomes d'oxygène du soluté, A et B, dans lequel un atome d'oxygène des molécules d'eau peut être trouvé (A) [3]. Représentation de la première et la seconde couche d'eau à partir de l'atome O4 de la partie terminale non réduite d'un trisaccharide (B).

Le temps moyen de repos exprime le temps qu'occupent les molécules d'eau dans une région spécifique de l'espace. Le temps moyen de repos des molécules d'eau en contact avec l'oxygène du groupement hydroxyle d'un hydrate de carbone est environ de 0,6-0,7 ps avec quelques exceptions pour des molécules comme le saccharose et le tréhalose, où les temps de repos de 2,7ps et 3,0 ps sont calculés respectivement pour certains groupements hydroxyle.

### ***II. Analyse d'une simulation de dynamique moléculaire***

Une fois la simulation réalisée, l'ensemble des données enregistrées permet d'analyser le comportement dynamique de la structure étudiée sous plusieurs formes. L'analyse la plus courante et la plus simple d'une trajectoire de dynamique consiste à tracer une composante géométrique en fonction du temps.

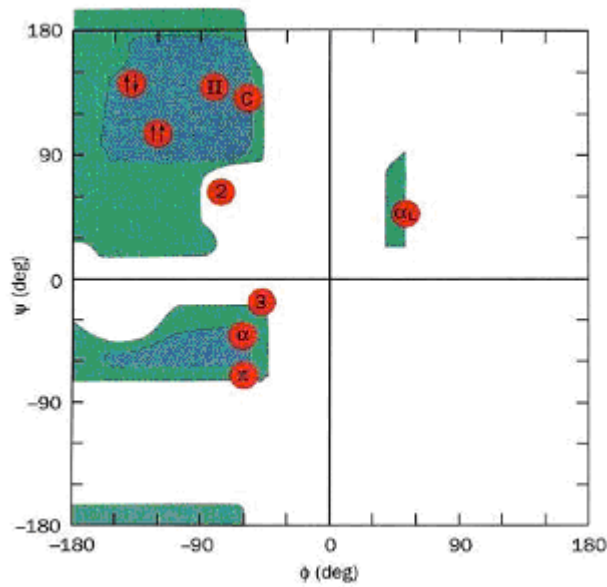
#### **1. Validation des structures**

Différents critères sont utilisés pour évaluer la qualité des structures générées.

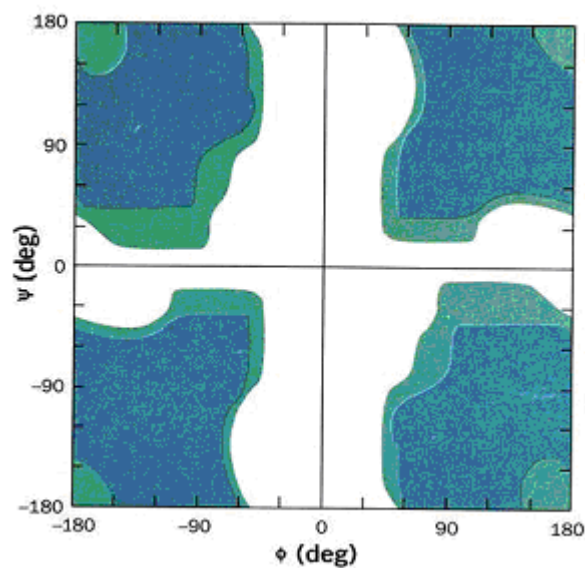
##### **a. Diagramme de Ramachandran**

Lorsqu'on porte sur un graphe des deux angles dièdres  $\phi$  et  $\psi$ , on construit un espace à deux dimensions appelé carte de Ramachandran, chercheur ayant beaucoup travaillé sur les peptides [45]. Ramachandran a eu l'idée de construire un modèle réduit de peptide où il pouvait faire pivoter les angles  $\phi$  et  $\psi$  de  $360^\circ$  [46]. Il a alors démontré que la plupart des zones de la carte étaient interdites à cause de l'encombrement stérique des chaînes latérales (figure 3). Des statistiques effectuées sur les protéines issues de la Protein Data Bank [47] ont permis plus tard de confirmer que seules certaines parties de la carte étaient représentées et que la proline et la glycine avaient des cartes différentes des autres acides aminés [48]. La glycine est en effet le seul résidu sans carbone  $\beta$ . L'hydrogène étant beaucoup moins encombrant qu'une chaîne latérale carbonée, la carte de Ramachandran pour le résidu glycine montre plus de conformations accessibles que pour les résidus plus gros (figure 4). De plus, la

non chiralité de ce résidu rend la carte symétrique par rapport à l'origine. Dans le cas des prolines, les variations de  $\phi$  sont limitées par les tensions du cycle.



**Figure 3** : Carte de Ramachandran d'une dialanine. En bleu sont les conformations les plus stables, en vert les conformations possibles et en blanc les conformations non atteintes. Les structures secondaires (hélice  $\alpha$  droite ( $\alpha$ ) et gauche ( $\alpha L$ ), feuillets  $\beta$  parallèles ( $\uparrow\uparrow$ ) et antiparallèles ( $\uparrow\downarrow$ ) sont indiquées.



**Figure 4** : Carte de Ramachandran des glycines.

### **b. Convergence des structures**

La convergence des structures obtenues est aussi un critère de sélection et elle est déterminée à partir de l'écart quadratique moyen (RMSD pour "root mean square deviation") de chaque structure par rapport à une structure moyenne calculée. Les calculs de RMSD sont réalisés sur le squelette et sur les chaînes latérales. Des valeurs élevées de RMSD indiquent des régions flexibles ou mal définies, alors que des valeurs faibles indiquent une bonne structuration du squelette et une orientation privilégiée pour ce qui concerne les chaînes latérales. En principe, des valeurs de RMSD inférieures à 1 Å reflètent une bonne convergence des structures obtenues.



### Références

- [1] M. L. DeMarco, R. J. Woods, *Glycobiology*, 18 (6): (2008), 426-440.
- [2] R. J. Woods, *Glycoconjugate journal*, 15 (3): (1998), 209-216.
- [3] N. Takahashi, H. Yagi, K. Kato, *Comprehensive Glycoscience in Elsevier*, 2: (2007).
- [4] E. J. Sorin, V. S. Pande, *Journal of computational chemistry*, 26 (7): (2005), 682-690.
- [5] K. N. Kirschner, R. J. Woods, *Proceedings of the National Academy of Sciences of the United States of America*, 98 (19): (2001), 10541-10545.
- [6] A. Imberty, S. Perez, *Chem. Rev*, 100 (12): (2000), 4567-4588
- [7] E. Fadda, R. J. Woods, *Drug Discovery Today*, 15(15-16): (2010), 596-609.
- [8] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, *Journal of Computational Chemistry*, 29 (4): (2008), 622-655.
- [9] M. B. Tessier, M. L. DeMarco, A. B. Yongye, R. J. Woods, *Molecular Simulation*, 34 (4): (2008), 349-364.
- [10] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. K. M. Gould, D. M. Merz, Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *Journal of the American Chemical Society*, 117 (19): (1995), 5179-5197.
- [11] A. Dejoux, P. Cieplak, N. Hannick, G. Moyna, F. Y. Dupradeau, *Journal of Molecular Modeling*, 7 (11): (2001), 422-432.
- [12] Kirschner, K. N.; Woods, R. J., *Proceedings of the National Academy of Sciences of the United States of America*, 98 (19): (2001), 10541-10545.
- [13] X. Biarnes, J. Nieto, A. Planas, C. Rovira, *Journal of Biological Chemistry*, 281 (3): (2006), 1432-1441.
- [14] M. Basma, S. Sundara, D. Calgan, T. Vernali, R. J. Woods, *Journal of computational chemistry*, 22 (11): (2001), 1125-1137.

- [15] R. J. Woods, M. Khalil, W. Pell, S. H. Moffat, V. H. Smith, *Journal of Computational Chemistry*, 11 (3): (1990), 297-310.
- [16] R. J. Woods, R. Chappelle, *Journal of Molecular Structure:THEOCHEM*, 527(1-3): (2000), 149-156.
- [17] R. D. Lins, P. H. Hunenberger, *Journal of computational chemistry*, 26 (13): (2005), 1400-1412.
- [18] L. D. Schuler, X. Daura, W. F. Van Gunsteren, *Journal of Computational Chemistry* 22 (11): (2001), 1205-1218.
- [19] L. D. Schuler, W. F. Van Gunsteren, *Molecular Simulation*, 25 (5): (2000), 301-319.
- [20] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, J. Hermans, *Intermolecular forces*, (1981), 331-333.
- [21] N. S. Gandhi, R. L. Mancera, *Glycobiology*, 19 (10): (2009), 1103-1115.
- [22] O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor, A. D. Mackerell Jr, *Journal of Computational Chemistry*, 29 (15): (2008), 2543-2564.
- [23] O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor, A. D. MacKerell Jr, *Journal of Chemical Theory and Computation*, 5 (9): (2009), 2353-237.
- [24] E. R. Hatcher, O. Guvench, A. D. MacKerell Jr, *J. Chem. Theory Comput*, 5 (5): (2009), 1315-1327.
- [25] E. Hatcher, O. Guvench, A. D. MacKerell Jr, *The Journal of Physical Chemistry B*, 113(37): (2009), 12466-12476.
- [26] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *Journal of Physical Chemistry B-Condensed Phase*, 102 (18): (1998), 3586-3616.

- [27] A. D. Mackerell, M. Feig, C. L. Brooks, *Journal of computational chemistry*, 25(11): (2004), 1400-1415.
- [28] A. D. MacKerell Jr, N. Banavali, N. Foloppe, *Biopolymers*, 56 (4): (2000-2001), 257-265.
- [29] D. Kony, W. Damm, S. Stoll, W. F. Van Gunsteren, *Journal of Computational Chemistry*, 23 (15): (2002), 1416-1429.
- [30] N. L. Allinger, Y. H. Yuh, J. H. Lii, *Journal of the American Chemical Society*, 111 (23): (1989), 8551-8566.
- [31] J. H. Lii, N. L. Allinger, *Journal of Computational Chemistry*, 12, (2): (1991), 186-199.
- [32] N. L. Allinger, F. Li, L. Yan, J. C, Tai, *Journal of Computational Chemistry*, 11 (7): (1990), 868-895.
- [33] M. Clark, R. D. Cramer, N. Van Opdenbosch, *Journal of Computational Chemistry*, 10 (8): (1989), 982-1012.
- [34] A. Imberty, K. D. Hardman, J. P. Carver, S. Perez, *Glycobiology*, 1 (6): (1991), 631-642.
- [35] S. Perez, C. Meyer, A. Imberty, *Modeling of Biomolecular Structures and Mechanisms*, A. Pullman, J. Jortner, and B. Pullman, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), (1995), 425-454.
- [36] H. D. Höltje, W. Sippl, D. Rognan, G. Folkers, "Molecular modeling: basic principles and applications". *Wiley-Vch*, (2008).
- [37] A. D. French, *Biopolymers*, 27 (9): (1988), 1519-1525.
- [38] E. Mikros, G. Labrinidis, S. Pérez, *Journal of Carbohydrate Chemistry*, 19 (9): (2000), 1319-1349.
- [39] I. Braccini, R. P. Grasso, S. Pérez, *Carbohydrate research*, 317 (1-4): (1999), 119-130.
- [40] P. G. Nyholm, L. A. Mulard, C. E. Miller, T. Lew, R. Olin, C. P. J. Glaudemans,

*Glycobiology*, 11 (11): (2001), 945-955.

[41] T. Kozar, C. W. von der Lieth, *Glycoconjugate journal*, 14 (8): (1997), 925-933.

[42] M. Orozco, C. Alhambra, X. Barril, J. M. López, M. A. Busquets, F. J. Luque, *Journal of Molecular Modeling*, 2 (1): (1996), 1-15.

[43] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *The Journal of Chemical Physics*, 79: (1983), 926-935.

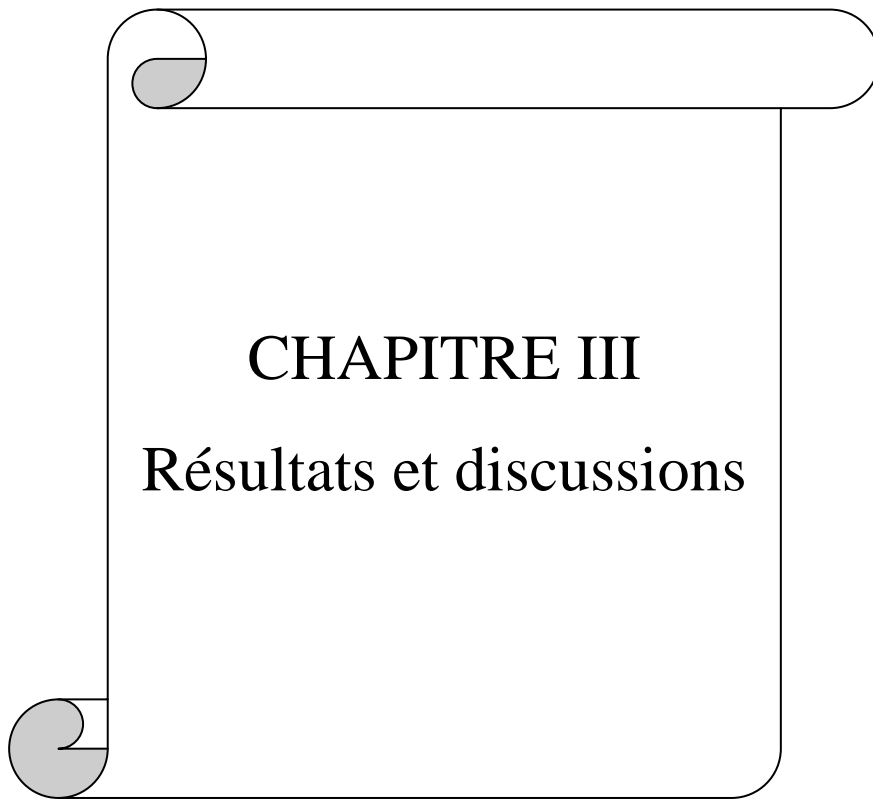
[44] C. Andersson, S. Balling Engelsen, *Journal of Molecular Graphics and Modelling*, 17 (2): (1999), 101-105.

[45] C. Ramakrishnan, *Protein Sci.*, 10(8): (2001), 1689-91.

[46] G. N. Ramachandran, C. Ramakrishnan, *J Mol Biol.*, 7: (1963), 95-99.

[47] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* 28(1): (2000), 235-42.

[48] G. N. Ramachandran, V. Sasisekharan, *Adv Protein Chem.*, 23: (1968), 283-438.



## Chapitre IV : résultats et discussions

---

Dans ce travail, des méthodes classiques de modélisation moléculaire (mécanique moléculaire MM et dynamique moléculaire DM) ont été utilisées pour élucider les caractéristiques structurales et dynamiques des biomolécules. Deux sujets principaux ont été abordés:

- i) Etude théorique de la stabilité des complexes dipeptidiques de cuivre (II).
- ii) Etude conformationnelle de quelques biomolécules en particulier des oligosaccharides : tri- et pentasaccharide lewis X, et des tripeptides: L,L-Phe-Phe-Gly, L,L-Phe-Gly -Phe, et L,L-Gly-Phe-Phe.

Dans ce chapitre nous avons englobé tous les résultats de calcul effectués dans cette étude.

La discussion que nous allons engager est basée sur les énergies stérique, et totale ainsi que les valeurs des angles dièdre ( $\phi$ ,  $\psi$  et  $\omega$ ) et RMSD des différents systèmes étudiés.

L'étude de la stabilité des complexes dipeptidique de cuivre (II) a été réalisée par le programme EMO (Energie d'une Molécule) [13]. L'étude conformationnelle des deux oligosaccharides et les trois tripeptides a été réalisée par le programme AMBER 10 [39], et les courbes de RMSD en fonction du temps et les cartes adiabatiques des différents systèmes étudiés ont été générées en utilisant les logiciels Origin 6.0 et Surfer [44].

### A. Etude théorique des complexes di-peptidique de cuivre (II) [47]

#### 1. Introduction :

Le cuivre est un élément essentiel largement distribué dans tout le corps des mammifères. Parmi les principaux ions métalliques de transition rencontrés en biologie, Cu(II) est probablement le plus efficace en chélation avec les peptides au pH physiologiques [1]. Dans le corps, l'essentiel du cuivre est présent sous une forme non-labile (par exemple : céruloplasmine). Le cuivre labile est présent sous forme de complexes avec des acides aminés ou des peptides [2].

Le Cu(II) forme des complexes très stables avec des oligopeptides simples. Les modes principaux de coordination des ions cuivriques avec des dipeptides simples sont bien établis maintenant [3].

A pH faible, l'espèce  $[\text{CuA}]^+$  se forme avec le dipeptide réagissant comme un ligand bidenté (I) [voir figure]. Vers pH=5, l'hydrogène peptidique peut être déprotéger en présence de Cu(II), permettant de réarranger les centres donneurs pour former la structure (II) [voir figure1], qui est l'espèce  $\text{CuH}_1\text{A}$ . Dans la zone de haut pH (>9), une des molécules  $\text{H}_2\text{O}$  se dissocie pour former un complexe mono-hydroxylique dont la formule est  $[\text{CuH}_2\text{A}]^-$  ou  $[[\text{CuH}_1(\text{OH})\text{A}]^-$ , qui possède la structure (III) montrée dans figure. Outre les mono-ligands, les structures des complexes biligands sont également bien établies. Ces complexes comprennent  $\text{CuA}_2$  (IV).

Sigel et al [1] indiquent qu'il existe de grandes différences entre les complexes cuivriques de la diglycine et les complexes avec des dipeptides où le résidu glycyll est monosubstitué par des chaînes latérales coordinantes ou non. Une diminution significative de stabilité a été observée dans les complexes cuivriques avec les dipeptides où résidu glycyll a été substitué. En accord avec Rabin [4], Sigel et al. concluent que la stabilité augmentent avec l'allongement des chaînes latérales. Ce phénomène a été expliqué par les interactions entre les

## Chapitre IV : résultats et discussions

chaînes latérales non covalentes [5,6]. De plus certains auteurs proposent l'existence d'interaction  $\pi$ -d entre les noyaux aromatiques des chaînes latérales et les ions métalliques dans certains complexes [7,8].

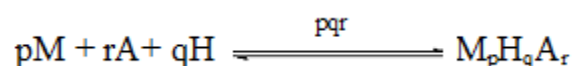
Des études potentiométriques [11] (tableau 1) sur les complexes mixtes de Cu(II) en présence de plusieurs acides aminés [9,10] ont montré que les espèces les plus stables sont formées lorsque un des ligands est aliphatique et l'autre aromatique (par exemple : phénylalanine ou tyrosine) [11,12].

**Tab 1:** Longueur de liaison de référence  $l_0$  pour le Cu-ligand (Cambridge Structural Database 1995).

Atom chaining	$l_0$ (angstroms Tot.Obs).	
Cu-O ( $sp^3$ ) axial	2.39 (2.38)	1117
Cu-O ( $sp^3$ ) equatorial	1.96 (1.94)	1162
Cu-O ( $sp^3$ ) equatorial	(2.03)	
Cu-O ( $sp^3$ ) axial	(2.15)	
Cu-O ( $sp^3$ ) equatorial	2.00 (2.04)	2279

Les valeurs de longueur de liaison de référence entre parenthèses sont tirées de la littérature [25].

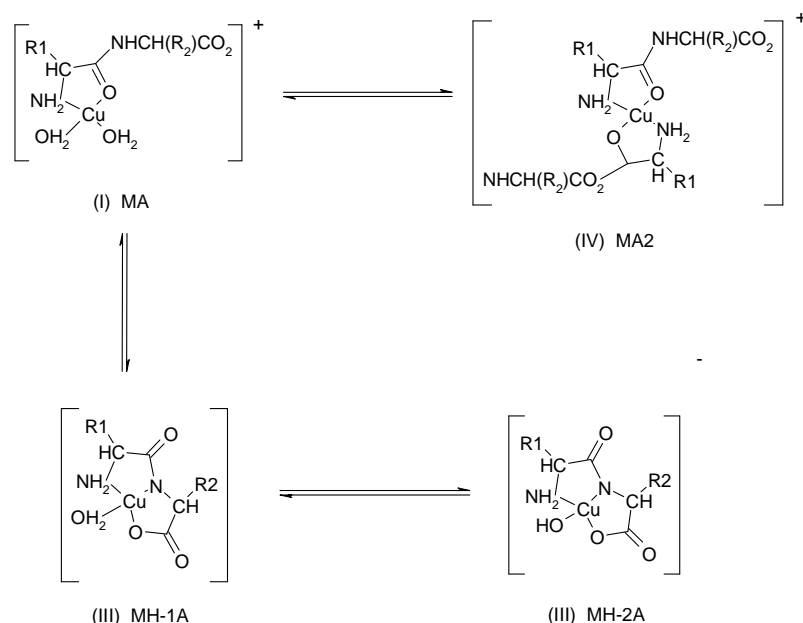
Tous les dipeptides utilisés dans cette étude sont de conformation L, L. Les équilibres dans les systèmes ion métallique-dipeptide et les constantes de stabilité des complexes sont définis par les équations suivantes (les charges sont ignorées pour plus de simplicité) :



$$\beta_{pqr} = [M_p H_q A_r] / [M]^p [H]^q [A]^r$$

Où p, q, r sont les coefficients stœchiométriques. Une valeur négative de q désigne une forme déprotonée ou hydroxylées de l'espèce considérée. La base conjuguée du ligand est noté A. HA est la forme zwitterions, et M l'ion métallique





**Fig1** : modèles structuraux des complexes cuivriques avec des dipeptides simples.

### 2. Approche théorique :

La modélisation moléculaire est utilisée pour étudier l'influence des chaînes latérales dans les dipeptides, et surtout ceux qui contiennent un résidu possédant un noyau aromatique.

Nous nous sommes intéressé à étudier la complexation de l'ion  $\text{Cu}^{2+}$  au voisinage du pH physiologique (l'espèce (II)  $\text{MH}_1\text{A}$  est la plus prédominante) avec les quatres dipeptides suivant : L,L-Phe-Leu, L,L-Phe-Met, L,L-Leu\_Phe, et L,L-Met-Phe (chaque dipeptide contient une chaîne latérale aromatique et une alkyl).

Ce travail consiste à étudier l'espèce  $\text{MH}_1\text{A}$  dans les deux cas Phe en C-terminal et Phe en N-terminal avec le programme EMO (Energie d'une MOlecule) suivi par des calculs semi empiriques par la méthode SAM1/d.

EMO a été développé par B. Blaive en 1993 [13] pour traiter les molécules organiques, il comprend trois menus :

- Menu1 : entrée de la molécule à l'aide du clavier ; chaque atome est codifié selon son hybridation. Le nombre maximal d'atomes est de 135, les hydrogènes compris.

## Chapitre IV : résultats et discussions

---

- Menu2 : manipulations géométriques sur la molécule.
- Menu3 : minimisation de l'énergie par mécanique moléculaire ; elle se fait sans dérivation de la fonction énergie. La version de ce programme mise à notre disposition est conçue pour fonctionner sur PC [14-17].

Ce programme utilise le champ de force MM2, qui est le champ de force d'Allinger [18], qui a été conçu au début pour les molécules simples (alcane, alcène et alcyne non conjugués, les composés carbonylés, les sulfures, les amines...). Il est utilisé pour traiter des molécules de plus en plus complexes.

Les paramètres retenus, concernant les structures des ligands de notre complexe, sont ceux proposés par le professeur Blaive complétée par l'étude statistique effectuée au niveau du CSD (Cambridge Structural Database, 1995).

Plusieurs effets sont souvent énoncés dans l'interprétation des résultats expérimentaux [19]: (a) l'effet donneur de la chaîne latérale, (b) l'interaction hydrophobique entre chaîne latérale-chaîne latérale; (c) les effets stériques des chaînes latérales (d) les effets de l'environnement sphère du solvant. Un autre effet a été proposé par les auteurs pour le cuivre (II)-dipeptide, le nickel (II)-dipeptide, les complexes contenant des chaînes latérales aromatiques [20,21]. Il a été démontré que la présence d'un résidu aromatique en position C-terminale augmente la stabilité des complexes.

- **rayon van der Waals du cuivre (II):** Les paramètres utilisés dans ce programme sont inspirés du travail de Benmenni [22] qui a pris la corrélation linéaire entre le rayon van der Waals publié par Allinger, et ceux déterminés par Bondi [23] établis par Bouraoui [15] qui a calculé les constantes dans le cas du fer.

$r^*(Allinger, MM2) = 0.856r (Bondi) + 0,058$  ce terme d'Allinger nous informe que les rayons van der Waals utilisés dans la mécanique moléculaire peuvent prendre des valeurs différentes sur la base des résultats cristallographiques. On a pris les valeurs du cuivre (II): (r

## Chapitre IV : résultats et discussions

---

\*  $\text{Cu}^{2+} = 1,90$  et  $\epsilon\text{Cu}^{2+} = 0,06$ , calculé par Bouraoui, une autre valeur de  $r^* = 1,70$  qui provient des données du programme MOLCAD développés par Brickman [24].

- **Modélisation des interactions Cu-Ligand:** Pour référence longueur de liaison, les liaisons Cu-N et Cu-O diffèrent selon le type d'atome d'oxygène ou d'azote lié au cuivre. Nous avons interrogé la CSD. Nous avons fait une étude statistique sur plus de 1000 cas ayant la même configuration de chromophore que nos complexes. Nous avons choisi Les valeurs en prenant les valeurs moyennes de celles qui sont en accord avec cette étude statistique (voir tableau 1).
- **Calculs de constante k:** Certaines mesures exécutées sur les complexes de cuivre, ont permis d'identifier la fréquence de vibration  $\gamma(\text{Cu-O})$  entre 200 et 500  $\text{cm}^{-1}$ . Cela a permis d'évaluer par calcul la constante de force k par la relation:  $\gamma = 1/2\pi (k/m)^{1/2}$  et  $k = 4\pi^2 m \gamma^2$  avec m (la masse réduite):  $1/m = 1/m_1 + 1/m_2$ .  $m_1$  et  $m_2$  sont les masses atomiques de l'oxygène et du cuivre, pour la vibration de Cu-O, on peut considérer ce métal est fixé par le ligand qui le porte, et la masse réduite prendra la valeur de l'oxygène:  $1/m = 1/m_1 = 1/16$

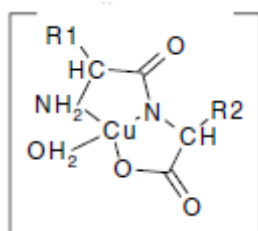
Après avoir minimisé l'énergie des différents complexes peptidiques du Cu(II), une optimisation d'énergie a été réalisée par la méthode SAM1/d.

SAM1/d est une méthode semi empirique qui tient compte les orbitales d des métaux de transition.

### 3. Résultats :

Nous nous sommes intéressés à étudier l'espèce  $\text{MH}_1\text{A}$  car c'est le complexe majoritaire des systèmes Cu(II)-dipeptide entre pH=4.5 et pH=8 [11] i.e au voisinage du pH physiologique. Tous les dipeptides utilisés dans cette étude sont de conformation L, L.

## Chapitre IV : résultats et discussions



Structure du  $\text{CuH}_1\text{A}$

Toutes les structures étudiées dans notre travail ont été construites à l'aide du programme EMO (version 2001) en introduisant le code d'Allinger des atomes des molécules étudiées, par le clavier de l'ordinateur ensuite l'énergie est minimisée en utilisant les paramètres semi-empiriques. La conformation la plus stable est obtenue à partir de diverses géométries de départ, après optimisation.

Afin d'éviter les minima locaux correspondants à des conformères instables, nous l'avons exécuté avec l'option « **SCAN** » qui permet de balayer la surface d'énergie potentielle (**PES**). Ceci, nous a permis d'éliminer les géométries ayant peu de chance d'engendrer un ensemble de conformères probablement stables. Les énergies des conformères trouvés sont optimisées par la méthode semi empirique **SAM1/d**. Les conformations les plus stables ont les énergies inférieures. Pour chaque structure, on donne les énergies:  $E_{\text{EMO}}$ ,  $E_{\text{liaison}}$ ,  $E_{\text{valence}}$ ,  $E_{\text{torsion}}$ ,  $E_{\text{van der Waals}}$ , et  $E_{\text{SAM1/d}}$ . Les résultats (les énergies) obtenus sont rassemblés dans le tableau 3.

**Tab 3 :** énergies des complexes calculées à l'aide d'EMO et SAM1/d.

	L,L-Leu-Phe-Cu(II)	L,L-Phe-Leu-Cu(II)	L,L-Met-Phe-Cu(II)	L,L-Phe-Met-Cu(II)
$E_{\text{EMO}}$ (kJ/mol)	66.711	67.506	60.011	63.445
$E_{\text{liaison}}$ (kJ/mol)	4.77	5.05	3.88	4.01
$E_{\text{valence}}$ (kJ/mol)	29.19	25.29	29.72	29.62
$E_{\text{torsion}}$ (kJ/mol)	-11.26	-9.78	-11.97	-10.23
$E_{\text{vdw}}$ (kJ/mol)	44.02	46.95	38.39	40.05
$E_{\text{SAM1/d}}$ (eV)	-5153.092	-5153.081	-5191.404	-5191.336

## Chapitre IV : résultats et discussions

### 4. Discussion :

Les constantes de formation des différentes espèces peut être déterminée par les mesures potentiométriques [11], voir le tableau 2.

**Tab 2:** Constantes de formation des complexes cuivre(II)-dipeptide à 25 °C, I=0,2 mol dm<sup>-3</sup> KNO<sub>3</sub> [11]

Dipeptide	$\log\beta_{1-11}$	$\Delta\log\beta_{1-11}$
L,L-Phe-Leu	$0.93 \pm 0.01$	0.96
L,L-Leu-Phe	$1.89 \pm 0.01$	
L,L-Phe-Met	$1.61 \pm 0.01$	0.09
L,L-Met-Phe	$1.70 \pm 0.01$	

En comparant chaque paire de ligands, où deux résidus aminés sont les mêmes, mais différent dans la séquence, on remarque une augmentation significative de la stabilité des complexes lorsque le résidu C-terminal contient un noyau aromatique. Les différences de constantes de formation sont rassemblées dans le Tableau 2 ( $\log\beta_{L,L\text{-Leu-Phe}} > \log\beta_{L,L\text{-Phe-Leu}}$  et  $\log\beta_{L,L\text{-Met-Phe}} > \log\beta_{L,L\text{-Phe-Met}}$ ).

Il est connu que les influences des chaînes latérales non-coordinantes sont différentes selon les espèces. Si nous examinons la structure de l'espèce  $\text{CuH}_1\text{A}$  déterminé par les mesures e.s.r. [26,27], nous voyons que les deux chaînes latérales sont toutes les deux impliquées dans la stabilité du complexe. Généralement, pour les L, L-dipeptides contenant deux groupes de chaîne latérale non-covalentes, la stabilité des  $\text{MH}_1\text{A}$  espèce est plus grande que celle des L,D- ou D,L-dipeptides correspondant à cause d'une interaction hydrophobe entre les groupes des deux chaînes latérales qui sont situés sur le même côté [5]. Cependant, cette interaction hydrophobe est moins évidente lorsque les dipeptides contiennent un groupe benzyle, parce que l'interaction entre le groupe alkyle et le cycle aromatique est moins efficace [5]. Cette augmentation de stabilité pour les L, L-Leu-Phe et L, L-Met-Phe peut être

## Chapitre IV : résultats et discussions

---

proposée comme le résultat de l'interaction  $\pi$ -d entre le noyau aromatique et les ions de cuivre [28-30].

Pour chaque paire de dipeptides qui contiennent les mêmes résidus d'acides aminés, on pourrait faire la comparaison suivante [12]:

(1) *Effet électronique.* Pour les espèces  $MH_{1A}$  Cet effet devrait être similaire pour les deux dipeptides, car il ne dépend que de la distance. Se référant à la représentation structurale, R1 et R2 ont presque la même distance de l'ion métallique central.

(2). *Effet stérique* Cet effet déstabilise le complexe. Le modèle tridimensionnel montre que le résidu R2 met au-dessus de l'ion métallique. Dans le cas de L, L-Leu-Phe et L, L-Phe-Leu, les deux chaînes latérales  $(CH_3)_2CHCH_2$  et  $C_6H_5CH_2$  sont différentes. Mais leurs effets stériques peuvent être comparables [5].

(3) *Interaction hydrophobe entre les deux chaînes latérales.* La structure établie pour les espèces  $MH_{1A}$  En fonction des études e.s.r montre un  $CuNO_2$  dans le plan chromophore, ce qui nécessite deux chaînes latérales situées sur le même côté, à l'égard de ce plan de base lorsque L, L-dipeptides sont considérés. Un tel arrangement des résultats dans une interaction donne lieu à des micelles hydrophobes, diminuant ainsi l'enthalpie de solvation.

La discussion ci-dessus exclut la possibilité que l'augmentation significative de la stabilité, pour les dipeptides contenant un résidu aromatique en C-terminal par rapport à la situation N-terminale, soit le résultat des différents autres effets que ceux énumérés en (1) à (3). Cette augmentation est alors très probablement due à l'interaction  $\pi$ -d

A la lumière des résultats obtenus on remarque que l'ordre de stabilité des systèmes étudiés est le suivant : L, L-Met-Phe-Cu(II) > L, L-Phe-Met-Cu(II) > L, L-Leu-Phe-Cu(II) > L, L-Phe-Leu-Cu(II).

Nous remarquons une augmentation significative de la stabilité des complexes

## Chapitre IV : résultats et discussions

---

cuivriques, lorsque le groupement aromatique est en position C-terminale (L, L-Leu-Phe et L, L-Met-Phe) par rapport au dipeptide contenant les mêmes acides aminés, mais le résidu phényl est en N-terminal (L, L-Phe-Leu et L, L-Phe-Met), ( $E_{L, L-Leu-Phe} < E_{L, L-Phe-Leu}$  et  $E_{L, L-Met-Phe} < E_{L, L-Phe-Met}$ ). Ce phénomène est attribué à l'interaction entre les électrons p du cycle aromatique et les orbitales d du Cu (II). Les résultats obtenus par le calcul sont en accord avec ceux obtenus par l'étude potentiométrique ( $\log\beta_{L, L-Leu-Phe} > \log\beta_{L, L-Phe-Leu}$  et  $\log\beta_{L, L-Met-Phe} > \log\beta_{L, L-Phe-Met}$ ).

### 5. Conclusion :

La présence d'un cycle aromatique dans le résidu C-terminal donne généralement des espèces [MH-1A] plus stables. Cet effet de stabilisation a été attribué à une interaction entre l'ion métallique et le cycle aromatique. Cette interaction se trouve que lorsque le cycle aromatique est situé dans le résidu C-terminal, parce que la liaison Cu-O est faible et facile à déformer. Cependant, la liaison rigide Cu-N est formée lorsque le cycle aromatique est dans l'extrémité N-terminal.

# Theoretical study of dipeptide complexes of copper(II)

N. Khebichat, S. Ghalem \*

*Abou-Bakr Belkaid University, Faculty of Sciences, Department of Chemistry, Laboratory of Organics Chemistry, Analyses and Natural Substances (COSNA), BP 119, Tlemcen 13000, Algeria*

Received 17 July 2006; accepted 1 August 2006

Available online 9 August 2006

## Abstract

The L,L-Phe-Leu-Cu(II), L,L-Leu-Phe-Cu(II), L,L-Phe-Met-Cu(II), and L,L-Met-Phe-Cu(II) systems were studied using molecular modelling. The results obtained, which are in good agreement with results obtained by potentiometric studies show a significant increase in stability of copper complexes, when an aromatic residue is located in C-terminal (compared to the L,L-dipeptides containing the same amino acid residues), this phenomenon is attributed to the interaction between the d-orbital of copper and the  $\pi$ -electrons of the aromatic ring.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Molecular modelling; Force field; Peptide; Copper(II)

## 1. Introduction

Copper is an essential element widely distributed in all the body of the mammals. Among the principal ions of transition metal in biology, Cu (II) is probably most effective in chelation with peptides at physiological pH [1]. In the body, the essence of copper is present under a non-labile form (for example: céruloplasmine). Labile copper is present in the form of complexes with amino acids or peptides [2].

Cu(II) forms very stable complexes with simple oligopeptides. The principal modes of coordination of the cupric ions with simple dipeptides are well established now [3].

At low pH, the species  $[\text{CuA}]^+$  is formed with the dipeptide reacting like a bidentate ligand (I) (see Fig. 1). Towards pH 5, peptidic hydrogens can be deprotected in the presence of Cu(II), making it possible to rearrange the donors centres to form the structure (II) (see Fig. 1), which is the  $\text{CuH-1A}$  species. In the zone of high pH ( $>9$ ), one of the molecules  $\text{H}_2\text{O}$  dissociates to form a mono-hydroxylic complex whose formula is  $[\text{CuH}_2\text{A}]^-$

or  $[\text{CuH}_1(\text{OH})\text{A}]^-$ , which has the structure (III) shown in Fig. 1. In addition to the monoligands, the structures of the complexes biligands are also well established. These complexes include  $\text{CuA}_2$  (IV) in Fig. 1.

Sigel et al. [1] indicate that there are great differences between the cupric complexes of the diglycine and the complexes with dipeptides where the residue glycyl is monosubstituted by coordinating side chains or not. A significant reduction in stability was observed in the cupric complexes with the dipeptides where the residue glycyl was substituted.

In agreement with Rabin [4], Sigel et al. conclude that stability increases with the lengthening of the side chains. This phenomenon was explained by the interactions between the non-covalent side chains [5,6]. Moreover certain authors propose the existence of interaction  $\pi$ -d between the aromatic ring of the side chains and the metal ions in certain complexes [7,8].

Potentiometric studies [11] (see Table 1) on the mixed complexes of Cu(II) in the presence of several amino acids [9,10] showed that the most stable species are formed when one of the ligands is aliphatic and the other one is aromatic (for example: phenylalanine or tyrosine) [11,12].

All the dipeptides used in this study are of L,L conformation. The equilibria in metal ion–dipeptide systems and the

\* Corresponding author.

E-mail address: s\_ghalem2002@yahoo.fr (S. Ghalem).



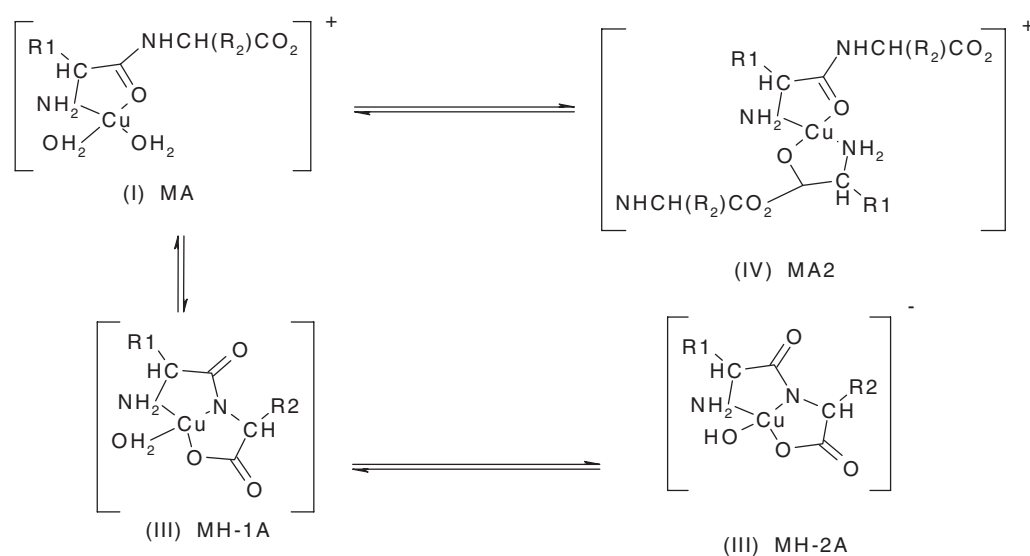


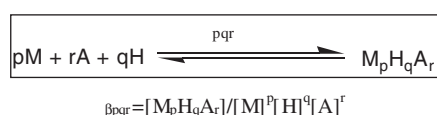
Fig. 1. Structural models of the cupric complexes with simple dipeptides.

Table 1  
Reference bond length  $l_0$  for Cu–ligand (Cambridge Structural Database 1995)

Atoms chaining	$l_0$ (angstroms Tot. Obs).	
Cu–O ( $sp^3$ ) axial	2.39 (2.38)	1117
Cu–O ( $sp^2$ ) equatorial	1.96 (1.94)	1162
Cu–O ( $sp^3$ ) equatorial	(2.03)	
Cu–O ( $sp^2$ ) axial	(2.15)	
Cu–N ( $sp^3$ ) equatorial	2.00 (2.04)	2279

Values of reference bond length in brackets are taken from literature [25].

stability constants of complexes are defined by the following equations (charges are omitted for simplicity):



where  $p$ ,  $q$ , and  $r$  are the stoichiometric numbers. A negative value of  $q$  denotes a deprotonated or hydroxylated form of the species considered. The conjugated base of the ligand is noted  $A$ .  $HA$  is the zwitterions form, and  $M$  the metal ion.

## 2. Theoretical approach

Molecular modelling is used to study the influence of the side chains in the dipeptides, and especially those which contain a residue having an aromatic ring. We were interested to study the complexation of the  $Cu^{2+}$  close to the physiological pH, where the species (II)  $MH_{-1}A$  is most predominant, with the four dipeptides according to: L,L-phenylalanyl-leucine (L,L-Phe-Leu), L,L-leucyl-phenylalanine (L,L-Leu-Phe), L,L-phenylalanyl-methionine (L, L-

Phe-Met), and L,L-methionyl-phenylalanine (L,L-Met-Phe). Each dipeptide contains an aromatic side chain and an alkyl.

The  $MH_{-1}A$  species is studied with the program EMO (Energy of MOlecule) follow-up by semi empirical calculations by the SAM1/d method, in both two cases:

- The phenyl group is located in C-terminal.
- The phenyl group is located in N-terminal.

EMO was developed by Blaive in 1993 [13] to treat the organic molecules, it includes three menus:

- Menu1: entry of the molecule using the keyboard; each atom is codified according to its hybridization. The maximum number of atoms is 420, hydrogen included.
- Menu2: geometrical handling on the molecule.
- Menu3: minimization of energy by molecular mechanics; it is done without derivation of the function energy. The version of this program placed at our disposal is designed to function on PC [14–17].

This program uses the force field MM2, which is the force field of Allinger [18], which was conceived at the beginning for the simple molecules (alkanes, carbonyl compounds, sulphides, amines, etc.). It is used to treat increasingly complex molecules.

The retained parameters, concerning structures of our complex ligands, are those proposed by the professor Blaive completed by the statistical study done at the level of the CSD (Cambridge Structural Database 1995).

Several effects are often set forth in the interpretation of experimental results [19]: (a) side-chain donor effect; (b) hydrophobic side chain–side chain interaction; (c) steric effects between the side chains; (d) effects from the surrounding solvent sphere. Another effect was proposed by

authors for copper(II)–dipeptide, nickel(II)–dipeptide, complexes containing aromatic side-chains [20,21]. It has been shown that the presence of an aromatic residue in a C-terminal site increases the stability of complexes.

- **van der Waals radius of copper(II):** Parameters used in this program are inspired from the work of Benmenni [22] who took the linear correlation between van der Waals radius published by Allinger, and those determined by Bondi [23] established by Bouraoui [15] who calculated constants in the case of iron.  $r^*(\text{Allinger,MM2}) = 0.856r(\text{Bondi}) + 0.058$  this term of Allinger informs us that van der Waals radius used in molecular mechanics can take different values on the basis of crystallographic results. One took values of the copper(II): ( $r^*\text{Cu}^{2+} = 1.90$  and  $\epsilon\text{Cu}^{2+} = 0.06$ , calculated by Bouraoui, another value of  $r^* = 1.70$  that comes from data of the MOLCAD program developed by Brickman [24].
- **Modelling of the Cu–Ligand interactions:** For reference bond length, the Cu–N and Cu–O bonds differ according to the type of atom of oxygen or nitrogen linked to the copper. We interrogated the CSD. We did a statistical study on more than 1000 cases having the same configuration of chromophore that our complexes. We chose values by taking the average values of those met at the time of this statistical study (see Table 1).
- **Calculations of constant  $k$ :** Some measurements executed on the copper complexes, permitted to identify the frequency of the vibration  $\gamma$  (Cu–O) between 200 and 500  $\text{cm}^{-1}$ . That permitted to evaluate by calculation the force constant  $k$  by the relation:  $\gamma = 1/2\pi(k/m)^{1/2}$  and  $k = 4\pi^2 m\gamma^2$  with  $m$  (reduced mass):  $1/m = 1/m_1 + 1/m_2$ ,  $m_1$  and  $m_2$  atomic masses of the oxygen and the copper, for the vibration Cu–O, we can consider that metal is fixed by the ligand that it carries, and the reduced mass will take the value of the oxygen:

Table 2  
Formation constants of copper(II)–dipeptide complexes at 25 °C,  $I = 0.2 \text{ mol dm}^{-3} \text{ KNO}_3$  [11]

Dipeptide	$\log \beta_{1-11}$	$\Delta \log \beta_{1-11}$
L,L-Phe-Leu	$0.93 \pm 0.01$	0.96
L,L-Leu-Phe	$1.89 \pm 0.01$	
L,L-Phe-Met	$1.61 \pm 0.01$	0.09
L,L-Met-Phe	$1.70 \pm 0.01$	

$\Delta \log \beta_{\text{pqr}} = \log \beta_{\text{pqr}}(\Phi \text{ C-terminal}) - \log \beta_{\text{pqr}}(\Phi \text{ N-terminal})$ .

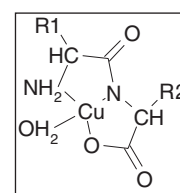
Table 3  
Energies of the complexes calculated using EMO and SAM1/d

	L,L-Leu-Phe-Cu(II)	L,L-Phe-Leu-Cu(II)	L,L-Met-Phe-Cu(II)	L,L-Phe-Met-Cu(II)
$E_{\text{EMO}}$ (kJ/mol)	66.711	67.506	60.011	63.445
$E_{\text{stretch}}$ (kJ/mol)	4.77	5.05	3.88	4.01
$E_{\text{binding}}$ (kJ/mol)	29.19	25.29	29.72	29.62
$E_{\text{torsion}}$ (kJ/mol)	–11.26	–9.78	–11.97	–10.23
$E_{\text{vdW}}$ (kJ/mol)	44.02	46.95	38.39	40.05
$E_{\text{SAM1/d(AMPAC)}}$ (eV)	–5153.092	–5153.081	–5191.404	–5191.336

$1/m = 1/m^1 = 1/16$ . After having minimized the energy of the various peptidic complexes of Cu(II), an optimisation of energy is carried out by the SAM1/d method. SAM1/d is a semiempirical method which takes in consideration d orbitals of transition metals.

### 3. Results

The study of the  $\text{MH}_{-1}\text{A}$  species is interesting because it is the majority complex of the Cu(II)–dipeptide systems between pH 4.5 and 8 [11] i.e. close to the physiological pH. All the dipeptides used in this study are of L,L conformation.



Structure of  $\text{CuH}_{-1}\text{A}$

All the structures studied in our work were built using the program EMO (version 2001) by introducing the Allinger code of the atoms of the molecules studied, by the keyboard of the computer then the energy is minimized by using the semi-empirical parameters. The most stable conformation is obtained from various starting geometries, after optimisation.

In order to avoid the local minima corresponding to unstable conformers, we carried out it with the option “SCAN” which makes it possible to sweep the surface of potential energy (PES). This enabled us to eliminate the geometries having little chance to generate the most stable conformers. Energies of the found conformers are optimised by the semi empirical method SAM1/d. The most stable conformations have the lower energies. For each structure we done the energies:  $E_{\text{EMO}}$ ,  $E_{\text{stretching}}$ ,  $E_{\text{bending}}$ ,  $E_{\text{torsion}}$ ,  $E_{\text{Van Der Waals}}$ , and  $E_{\text{SAM1/d}}$ .

The results (energies) obtained are gathered in Table 3.

### 4. Discussion

The formation constants of different species can be determined by the potentiometric measurements [11], see Table 2.

Comparing each pair of ligands, where two amino residues are the same, but differing in sequence, we remark a significant increase in stability of complexes when the C-terminal residue contains an aromatic ring. The differences of formation constants are gathered in Table 2 ( $\log \beta_{L,L\text{-Leu-Phe}} > \log \beta_{L,L\text{-Phe-Leu}}$  and  $\log \beta_{L,L\text{-Met-Phe}} > \log \beta_{L,L\text{-Phe-Met}}$ ).

It is known that the influences of non-coordinating side chains are different according to the species. If we examine the structure of the  $\text{CuH}_{-1}\text{A}$  species determined by e.s.r. measurements [26,27], we see that the two side chains are both involved in the stability of complex. Generally for L,L-dipeptides containing two large non-covalent side chain groups, the stability of  $\text{MH}_{-1}\text{A}$  species is greater than of the corresponding L,D- or D,L dipeptides due to a hydrophobic interaction between two side chain groups which are located on the same side [5]. However, this hydrophobic interaction is less evident when the dipeptides contain a benzyl group, because the interaction between the alkyl group and the aromatic ring is less effective [5]. This increase in stability for the L,L-Leu-Phe and L,L-Met-Phe can be proposed as the result of a  $\pi$ -d interaction between the aromatic ring and the copper ion [28–30].

For every pair of dipeptides which contain the same amino acid residues, we could make the following comparison [12]:

- (1) *Electronic effect.* For  $\text{MH}_{-1}\text{A}$  species, this effect should be similar for the two dipeptides, because it is only dependent upon the distance. Referring the structural representation, R1 and R2 have almost the same distance from the center metal ion.
- (2) *Steric effect.* This effect destabilizes the complex. The tridimensional model shows that the R2 residue sets above the metal ion. In the case of L,L-Leu-Phe and L,L-Phe-Leu, the two side chain  $(\text{CH}_3)_2\text{CHCH}_2$  and  $\text{C}_6\text{H}_5\text{CH}_2$  are different. But their steric effect may be comparable [5].
- (3) *Hydrophobic interaction between two side chains.* The structure established for the  $\text{MH}_{-1}\text{A}$  species according to the e.s.r. studies shows a  $\text{CuNO}_2$  in-plane chromophore, which requires two side chains located on the same side, with respect to this basal plane when L,L-dipeptides are considered. Such an arrangement results in an interaction giving rise to hydrophobic micelles, thus decreasing the solvation enthalpy.

The above discussion excludes the possibility that the significant increase in stability, for the dipeptides containing an aromatic residue in C-terminal with respect to the N-terminal location, is the result of the other different effects than those enumerated in (1) to (3). This increase is then very probably due to  $\pi$ -d interaction.

In the light of the results obtained by calculation, we notice that the order of stability of the studied systems is as follows: L,L-Met-Phe-Cu(II) > L,L-Phe-Met-Cu(II) and L,L-Leu-Phe-Cu(II) > L,L-Phe-Leu-Cu(II).

We notice a significant increase in the stability of the cupric complexes, when the aromatic group is in C-terminal position (L,L-Leu-Phe and L,L-Met-Phe) compared to the dipeptide containing the same amino acids, but the residue phenyl is into N-terminal (L,L-Phe-Leu and L,L-Phe-Met), ( $E_{L,L\text{-Leu-Phe}} < E_{L,L\text{-Phe-Leu}}$  and  $E_{L,L\text{-Met-Phe}} < E_{L,L\text{-Phe-Met}}$ ). This phenomenon is attributed to the interaction between the  $\pi$  electrons of the aromatic ring and the d orbitals of the Cu(II). Results obtained by the calculation are in agreement with those obtained by the potentiometric study ( $\log \beta_{L,L\text{-Leu-Phe}} > \log \beta_{L,L\text{-Phe-Leu}}$  and  $\log \beta_{L,L\text{-Met-Phe}} > \log \beta_{L,L\text{-Phe-Met}}$ ).

## 5. Conclusion

The presence of an aromatic ring in the C-terminal residue gives generally a more stable  $[\text{MH}_{-1}\text{A}]$  species. This stabilising effect was attributed to an interaction between the metal ion and the aromatic ring. This interaction is found only when the aromatic ring is located in the C-terminal residue, because the Cu—O bond is weak and easily distorted. However, the rigid Cu—N bond formed when the aromatic ring is in the N-terminal residue does not favour this interaction [11].

## Acknowledgement

The authors thank particularly Dr. B. Blaive Professor at the University of Marseille, France for sending the molecular mechanics program EMO.

## References

- [1] H. Sigel, R.B. Martin, Chem. Rev. 82 (1982) 385–426.
- [2] J.W.H. Robert, L.D. Pettit, J.C.S. Dalton. (1981) 13331.
- [3] A. Kaneda, A.E. Martell, J. Am. Chem. Soc. 99 (1977) 1586.
- [4] B.R. Rabin, Trans. Faraday. Soc. 52 (1956) 1130.
- [5] R.P. Bnomo, R. Cali, V. cucinotta, Inorg. Chem. 25 (1986) 1641–1646.
- [6] T. Kiss, Z. Szucs, J. Chem. Soc. Dalton Trans. (1986) 2443.
- [7] R.M. Izatt, J.W. Wrathall, K.P. Anderson, J. Phys. Chem. 65 (1961) 1914.
- [8] A. Wahlborg, E. Friedon, Arch. Biochem. Biophys. 111 (1965) 672.
- [9] A. Gergerly, I. Savago, I. Nagypal, R. Kiraly, Inorg. Chim. Acta 6 (1972) 435.
- [10] A. Gergerly, I. Savago, J. Inorg. Nucl. Chem. 35 (1973) 4355.
- [11] L. Xiao, M. Jouini, B.T. Fan, G. Lapluye, J. Huet, J. Chem. Soc. Dalton Trans. 286 (1989) 1–10.
- [12] S. Ghalem, B.T. Fan, L. Xiao, G. lapluye, J. Chim. Phys. 91 (1994) 113–125.
- [13] S. Ghalem, Thèse de Doctorat, Université Denis Diderot- Paris VII, France (1996).
- [14] A. Bouraoui, M. Fathallah, B. Blaive, R. Gallo, J. Chem. Soc. Perkin Trans. 2 (1990) 1211–1214.
- [15] A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive, R. Gallo, in J.L. Rivail (Eds.), Proceedings of an International Meeting, Studies in Physical and Theoretical Chemistry, vol. 71 Nancy, France, 11–15, Septembre, 1989, pp. 381–393.
- [16] A. Zinelabidine, A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive, R. Gallo, J. Mol. Struct. Theochem. 286 (1993) 267–274.
- [17] B. Blaive, G. Legsaï, R. Laï, J. Mol. Struct. 354 (1995) 245–250.

- [18] Z. Li, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 84 (1987) 6611.
- [19] L.D. Petit, R.J.W. Hefford, in: H. Sigel (Ed.), Metal Ions in Biological Systems, vol. 9, 1979, p.173.
- [20] R.P. Bonomo, R. Cali, V.G.I. Cucinotti, E. Rizzarelli, Inorg.Chem. 25 (1986) 1641.
- [21] S.H. Kim, R.B. Martin, J.Am.Chem. Soc. 106 (1984) 1707.
- [22] L. Benmenni, Thèse de doctorat en sciences, Fac. de Sciences de St-Jérôme, Université d'Aix-Marseille III (1994).
- [23] A. Bondi, J. Phys. Chem. 68 (1964) 441.
- [24] J. Brickmann, MOLCAD Molecular Modelling Package, Institut für physikalische Chemie, Technische Hochschule, 6100 Darmstad, Germany (1989).
- [25] L. Benmenni, E.H. Allilou, M. Giorgi, M. Pierrot, M. Regulier, J. Chem. Crystallogr. 24 (1994) 5.
- [26] J. Peisach, W.E. Blumberg, Arch. Biochem. Biophys. 165 (1974) 691.
- [27] D.M. Dooley, J. Rawlings, J.H. Dawson, P.J. Stephenes, L.E. Andresson, B.G. Malmstrom, H.B. Grey, J. Am. Chem. Soc. 101 (1979) 5038.
- [28] O. Yamauchi, K. Tsujide, A. Odani, J. Am. Chem. Soc. 107 (1985) 659.
- [29] W.L. Kwik, K.P. Ang, G.I. Chen, J. Inorg. Nucl. Chem. 42 (1980) 303.
- [30] P.I. Vestues, R.B. Martin, J. Am. Chem. Soc. 102 (1980) 7906.

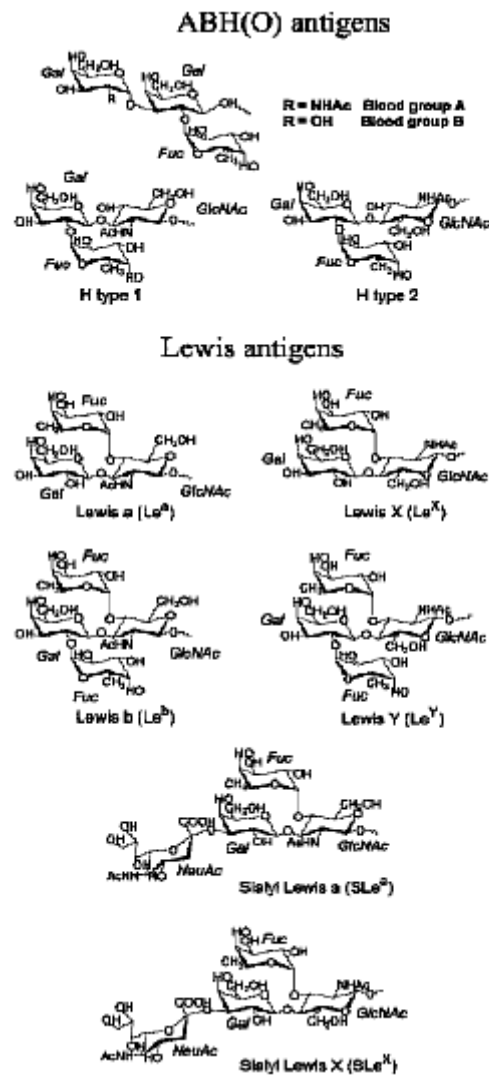
### **A. Recherche conformationnelle sur les structures de Lewis X par dynamique moléculaire: étude du tri-et pentasaccharide [48]**

#### **1. Introduction**

Dans la nature, les hydrates de carbone forment une famille importante de biomolécules, comme les hydrates de carbone simples ou complexes, seuls ou liés d'une façon covalente aux protéines ou lipides [31]. Ils jouent des rôles vitaux dans beaucoup de processus biologiques, tel que la reconnaissance, l'adhésion et la communication entre les cellules. Les hydrates de carbone qui sont liés à une partie non-sucre (protéines, peptides ou lipides) sont les plus proéminents structures de la surface-exposées cellulaire. La variabilité structurelle très diverse d'hydrates de carbone les fait des bons candidats pour les récepteurs cellulaires et les molécules de reconnaissance [32].

La détermination de la structure à trois dimensions des oligosaccharides et comprendre la base moléculaire de leur reconnaissance par les récepteurs représente les défis principaux de glycobiologie structurelle. L'élucidation des structures à trois dimensions et les propriétés dynamiques des oligosaccharides est une condition préalable pour mieux comprendre de la biochimie des processus de reconnaissance et pour le design rationnel des médicaments dérivés des carbohydrates [31].

Les systèmes histo-blood-group ABH (O) humains étaient les premiers allo-antigènes humains à être identifiés. La nature d'hydrate de carbone des antigènes A, B, H, et Lewis ont été établis en 1950 [33-34]. La Fig. 1: montre une représentation schématique des antigènes histo sang groupe courants bien que les types supplémentaires rares qui existent aussi [35]. L'épitope oligosaccharidique des antigènes A, B, H (O) et les systèmes Lewis ont été le sujet de beaucoup d'investigations structurelles. Les cartes d'énergie de tous les fragments de disaccharides ont été établies par les méthodes des mécaniques moléculaires [36-37].



**Fig 1:** Représentation schématique des oligosaccharides histo-blood-group

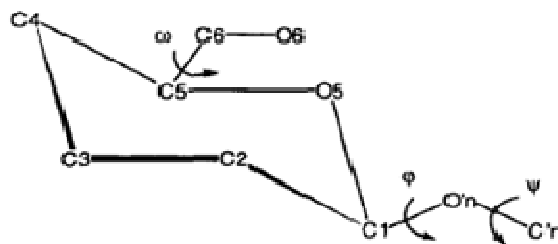
La détermination des préférences conformationnelles des oligosaccharides est mieux approchée en décrivant leurs conformations préférées sur les surfaces de l'énergie potentielle comme une fonction des angles dièdres de la liaison osidique  $\varphi$ ,  $\psi$ . Aujourd'hui, des techniques complexes de simulation telles que la dynamique moléculaire sont utilisées de plus en plus. Plusieurs exemples peuvent être trouvés dans la littérature des stratégies de recherches conformationnelles pour simplifier la complexité du problème de minimums multiples [38].

## Chapitre IV : résultats et discussions

Dans ce travail, nous avons réalisé une étude de simulation DM (solvant explicite) pendant 10 ns pour les tri- et pentasaccharides Lewis X. Chaque unité de monosaccharide dans les molécules Lewis X adopte la conformation chaise  ${}^4C_1$ , sans déviation significative de la forme pyranose classique. L'objectif de cette étude était d'explorer le comportement conformationnelle de ces oligosaccharides dans le modèle d'eau explicite à 300 K pendant 10 ns en utilisant le champ de force GLYCAM 06.

### 2. Approche théorique:

Le marquage atomique et les angles de torsions sont indiqués dans la Fig. 2. Les angles de torsion  $\varphi$ ,  $\psi$  sont définis  $\varphi = \theta$  (O5-C1-O'2-C'2),  $\psi = \theta$  (C1-O'2-C'2-C'3), et  $\omega = \theta$  (O5-C5-C6-O6). Les conformations avec angle de torsion spécifique  $\omega$  sont considérés comme gauche-trans (gt;  $\omega = 60^\circ$ ), gauche-gauche (gg;  $\omega = -60^\circ$ ) et trans-gauche (tg;  $\omega = 180^\circ$ ).



**Fig. 2** : marquage atomique et les angles de torsions de l'intérêt

Les simulations de DM ont été réalisées en utilisant AMBER10 [39] à une température de 300 K. Cette température a été choisie parce que notre objectif principal était d'étudier la dynamique des oligosaccharides à des conditions physiologiquement pertinentes. Les coordonnées de départ et les paramètres des monomères de tri- et penta-saccharide Lewis X ont été construits avec le "Glycam Biomolecule Builder", disponible en ligne sur le site Web de R.J. Wood Group [40].

## Chapitre IV : résultats et discussions

---

Étant donné que ces coordonnées ne sont pas optimisées, une première minimisation par mécanique moléculaire (1000 pas de steepest descent suivie par 1000 pas du conjugate gradient avec une tolérance de gradient de 0,05 kcal mol<sup>-1</sup>) a été faite avec le programme AMBER 10 [39]. Le champ de force GLYCAM 06 a été utilisé pour les oligosaccharides (il est utilisé pour simuler les interactions interatomiques responsables de la dynamique des oligosaccharides et glycoprotéines et il est actuellement adapté à tous les liaisons N- et O-biologiquement pertinentes dans des oligo- et polysaccharides et des glycoprotéines [41]). Les simulations ont été réalisées en utilisant un solvant explicite avec le programme AMBER 10. Les systèmes ont été préparés en utilisant une boîte d'eau explicite de 8 Å dans le module Xleap d'AMBER10. Le nombre de molécules d'eau était de 569 pour le trisaccharide, et de 1072 pour le pentasaccharide. Toutes les simulations ont été effectuées avec le module SANDER d'AMBER avec l'algorithme SHAKE [42] (tolérance = 0.0005 Å<sup>o</sup>) pour contraindre les liaisons covalentes impliquant les hydrogènes, en utilisant des conditions aux limites périodiques, un pas de temps de 2 fs, un couplage de température [43], un cut-off de 8 Å a été appliqué à l'interaction de Lennard-Jones, et une pression constante de 1 atm. La liste non liée a été actualisée toutes les 10 étapes. Cette étude préliminaire avait pour but pour trouver la structure la plus favorable énergétiquement qui a été trouvée à 4856,9 ps pour le trisaccharide, et à 2411,8 ps pour le pentasaccharide. Ces conformations ont été trouvés avec un script Perl "home-made" qui se contentent de vérifier toutes les valeurs des énergies, trouver la plus basse et en donnant la structure associée [32].

Le module PTRAJ d'AMBER10 a été utilisé pour analyser et traiter les fichiers de la trajectoire et des coordonnées à partir de SANDER, où les analyses pourraient inclure la réalisation de superposition, l'analyse des liaisons hydrogènes, le calcul des fluctuations dans des liaisons, des angles ou des dièdres, des fonctions de corrélation etc. et la racine carrée moyenne des déviations (RMSDs).



## Chapitre IV : résultats et discussions

Enfin, les cartes de contour iso-énergie ont été générés par les logiciels ORIGINE et SURFER 9 (Golden Software, Inc, Golden Co) [44]. Les positions des minima dans ces cartes ont été ensuite calculées en minimisant l'énergie sans retenue ces points indiquant les minima sur les cartes.

### 3. Résultats :

Les surfaces d'énergie conformationnelles adiabatiques ( $\phi$ ,  $\psi$ ) calculées pour le tri- et pentasaccharide Lewis X sont présentées comme des courbes d'énergies potentielles dans les Fig. 3 et Fig. 4 respectivement. Les contours sont séparés par 1 kcal/mol et allant de 1 à 10 kcal/mol. Les énergies sont données par rapport au minimum le plus bas. La géométrie et les énergies potentielles relatives des minimums prédis sont énumérées dans la Table 1 et la Table 2 et les conformères de plus basse énergie pour chaque composé sont indiqués dans la Fig. 3 et Fig. 4.

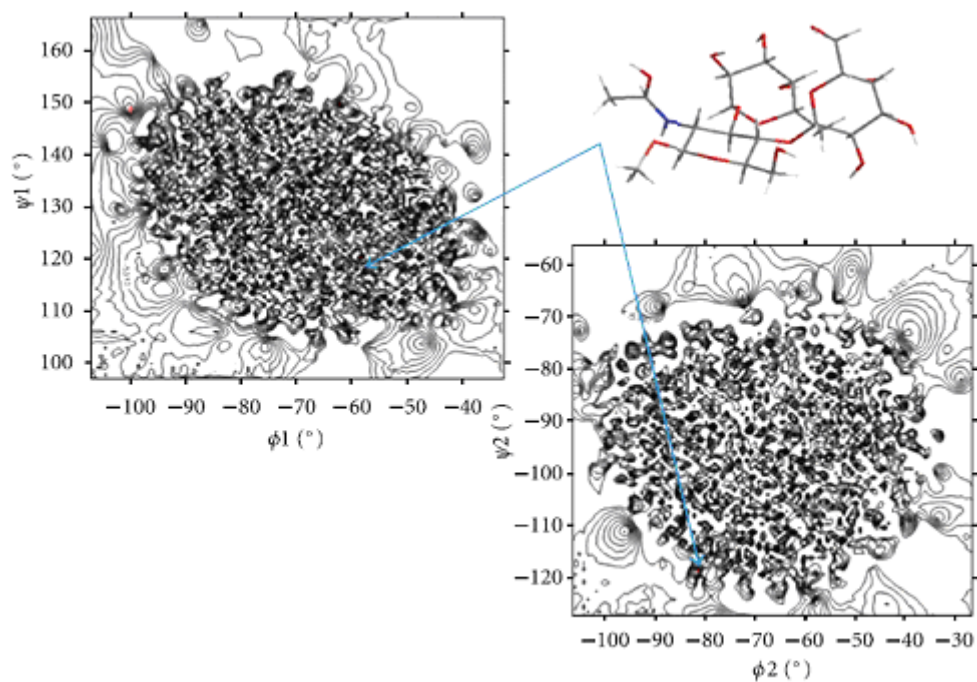
**Table1:** Energie potentielle, angles de torsion des conformères de plus basses énergies du trisacchride Lewis X.

	PE(kcal/mol)	$\phi_1(^{\circ})$	$\psi_1(^{\circ})$	$\phi_2(^{\circ})$	$\psi_2(^{\circ})$
Conformer1	-5446.95	-61.84	124.7	-61.54	-104.33
Conformer2	-5442.50	-84.91	155.63	-69.95	-92.95
Conformer3	-5440.08	-80.60	139.25	-55.46	-98.51
Conformer4	-5437.12	-56.58	130.28	-59.61	-80.28

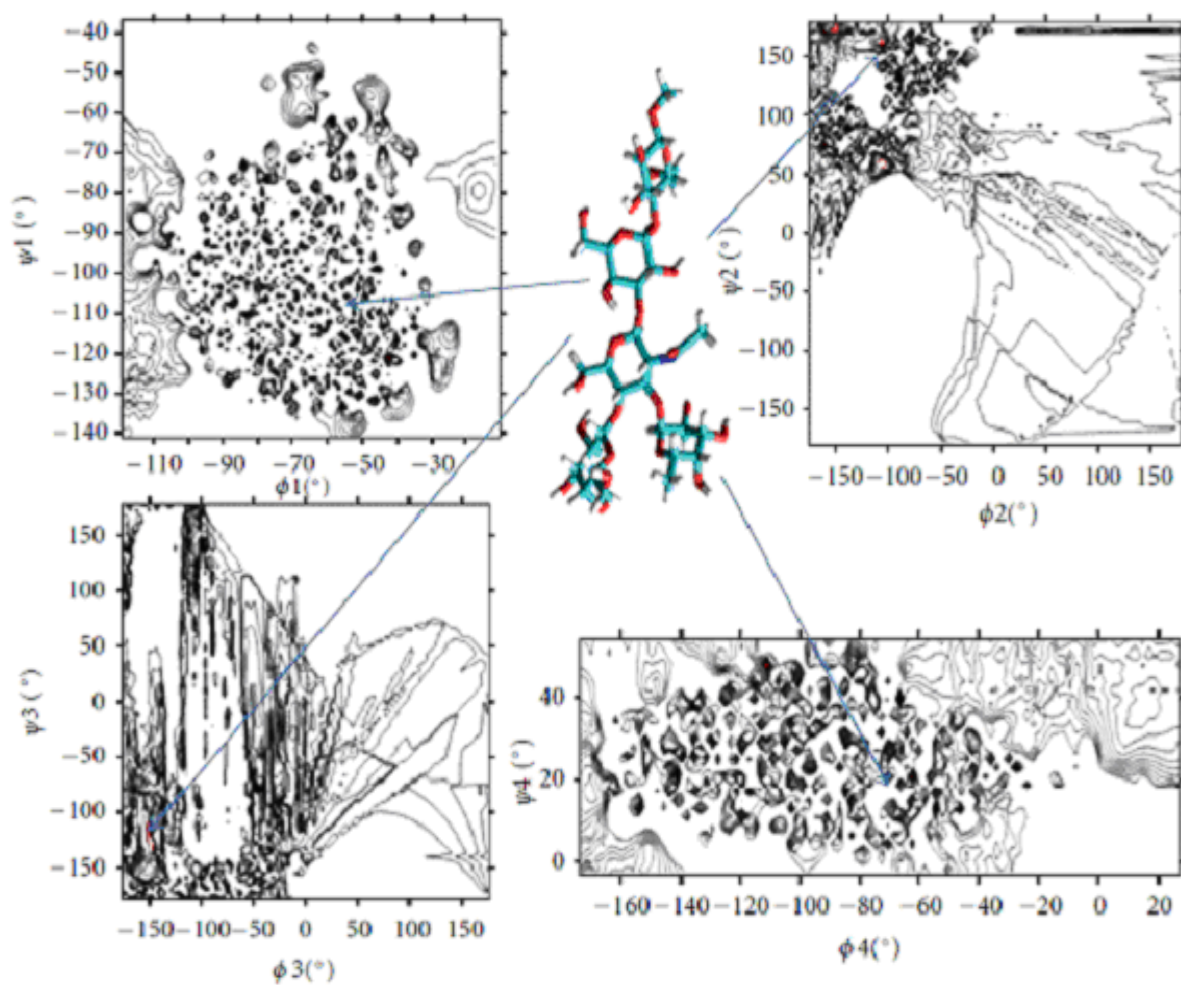
## Chapitre IV : résultats et discussions

**Table 2:** Energie potentielle, angles de torsion des conformères de plus basses énergies du pentasacchride Lewis X.

	PE(kcal/mol)	$\phi_1$ (°)	$\psi_1$ (°)	$\phi_2$ (°)	$\psi_2$ (°)	$\phi_3$ (°)	$\psi_4$ (°)	$\phi_4$ (°)	$\psi_4$ (°)
Conformer1	-10191.15	-78.30	-97.88	-124.4	64.97	-108.27	-173.37	-57.54	15.18
Conformer2	-10162.71	-65.65	-106.09	-73.57	136.95	-52.97	-155.29	-71.26	22.90
Conformer3	-10162.25	-56.49	-112.26	-78.58	144.67	-55.95	-157.92	-69.85	23.71
Conformer4	-10152.28	-71.67	-108.12	-75.73	152.13	-73.72	-163.60	-74.94	18.04



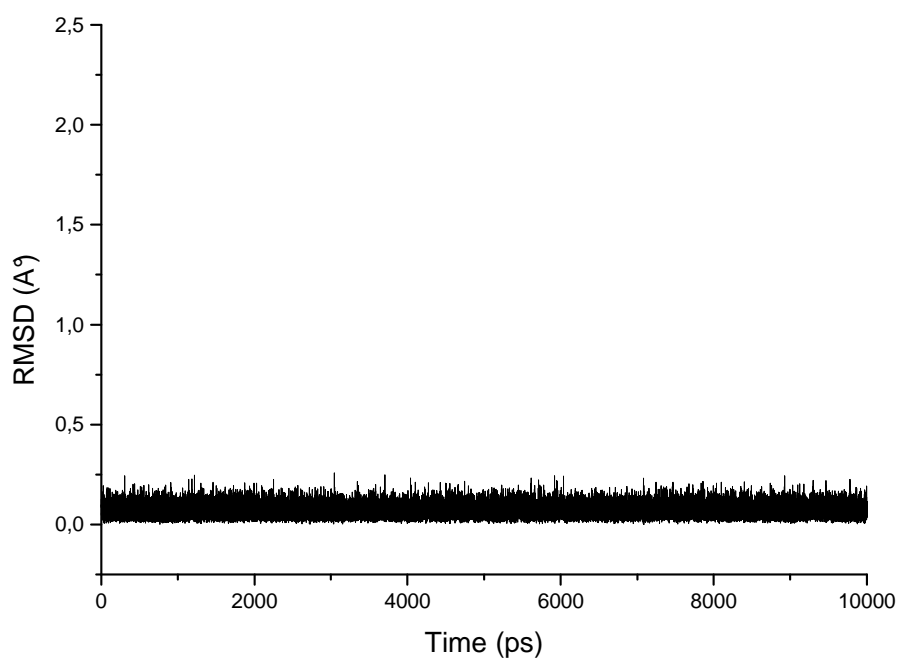
**Figure 3 :** Carte adiabatique de RAMCHANDRAN du trisaccharide Lewis X calculée en utilisant le champ de force GLYCAM 06.



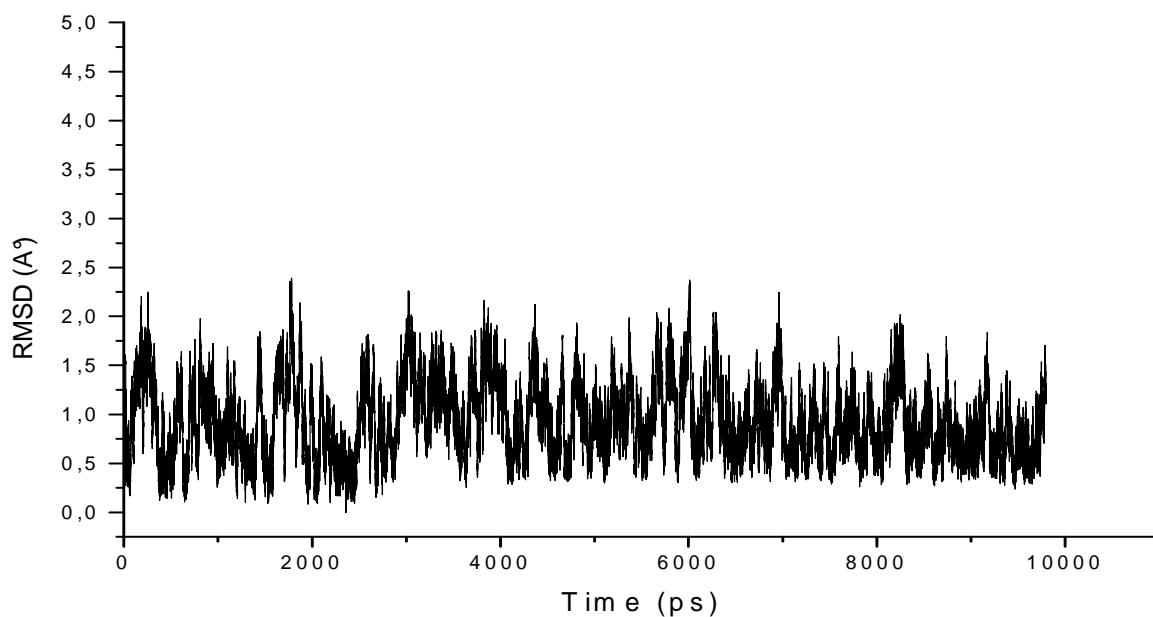
**Figure 4 :** Carte adiabatique de RAMCHANDRAN du pentasaccharide Lewis X calculée eu utilisant le champ de force GLYCAM 06.

### 3.1. La racine carrée moyenne de déplacement du système

RMSD est un paramètre important pour décrire la stabilité du système. Fig. 5 et Fig. 6 représentent les courbes RMSD en fonction du temps du tri-et pentasaccharide Lewis X respectivement durant le processus de simulation.



**Fig. 5:** RMSD du trisaccharide Lewis X en fonction du temps, par rapport à la conformation de plus basse énergie.



**Fig. 6:** RMSD du pentasaccharide Lewis X en fonction du temps, par rapport à la conformation de plus basse énergie.

### 3.2. Calcul de Phi / Psi des oligosaccharides:

Dans la présente étude, la prédiction des structures des oligosaccharides a été réalisée par les angles de torsion phi ( $\phi$ ) / psi ( $\psi$ ). La combinaison des angles  $\phi$  et  $\psi$  détermine parfaitement la conformation d'un oligosaccharide. Fig. 3 et Fig. 4 montrent la répartition des conformations permis stériquement d'énergie minimisée dans l'espace glycosidique  $\phi / \psi$  obtenu à partir de notre simulations de DM (solvant explicite) pendant 10 ns dans le cas du tri-et pentasaccharide Lewis X, respectivement. Ces figures montrent également les meilleurs conformères (qui ont des énergies faibles). Ces conformères et leurs valeurs angulaires  $\phi/\psi$  correspondants sont donnés dans les tableaux 1 et 2 pour le tri-et pentasaccharide Lewis X, respectivement.

## Chapitre IV : résultats et discussions

---

### 4. Discussion:

Le suivi des énergies cinétique, potentielle et totale le long de la trajectoire, ainsi que la pression et la température, indiquent que le système périodique global est stable. L'énergie cinétique a augmenté lentement, alors que l'énergie potentielle a diminué progressivement au cours des simulations. L'énergie totale d'un système à simuler doit être conservée, tandis que la fluctuation de la température est causée par les composantes de l'énergie cinétique et potentielle [45].

Les variations RMSD représentées en fonction du temps (Fig. 5 et Fig. 6 pour le tri - et pentasaccharide Lewis X respectivement) indiquent que les deux systèmes sont équilibrés après 500 ps. L'examen visuel des deux trajectoires de 10 ns montre que les deux oligosaccharides sont stables. Une analyse RMSD a été emportée pour mesurer les différences structurelles. Le RMSD entre les structures PDB initiales (obtenues à partir de Glycam biomolecule builder) et les conformères de basse énergie en tenant compte seulement les atomes lourds sont de  $0.03\text{\AA}$  et  $0.20\text{\AA}$  pour le tri - et le pentasaccharide respectivement. Cela montre que les deux oligosaccharides sont stables.

Pendant la simulation de ces oligosaccharides, nous remarquons que pour le pentasaccharide le conformère le plus stable est obtenu à 2411.8 ps, et que celui pour le trisaccharide a été obtenu à 4856.9 ps. Cela signifie que le pentasaccharide est moins flexible que le trisaccharide à cause de l'encombrement stérique.

### 5. Conclusion

Une compréhension complète du rôle des glucides dans les systèmes biologiques dépend de l'information disponible sur le mélange à l'équilibre et sur la conformation préférée des glucides en solution. L'analyse conformationnelle offre un outil qui permet de déterminer toutes les conformations possibles qui influencent le comportement des glucides en solution.

## Chapitre IV : résultats et discussions

---

Ce travail représente une étude théorique des deux oligosaccharides tri-et pentasaccharide Lewis X en utilisant une simulation de dynamique moléculaire dans l'eau explicite avec le programme AMBER10. Les mesures RMSD des glucides autour de leurs conformations moyennes jouent un rôle d'indicateur important de nombreux processus biologiques tels que la reconnaissance macromoléculaire, et des formations complexes [46].

## Research Article

# Conformational Search on the Lewis X Structure by Molecular Dynamic: Study of Tri- and Pentasaccharide

N. Khebichat, K. Nekkaz, and S. Ghalem

Laboratory of Natural Products and Bio Actives, Department of Chemistry, Faculty of Sciences, Aboubakr Belkaid University, Tlemcen 13000, Algeria

Correspondence should be addressed to S. Ghalem, s\_ghalem2002@yahoo.fr

Received 4 July 2011; Revised 7 October 2011; Accepted 12 October 2011

Academic Editor: Yuriy A. Knirel

Copyright © 2012 N. Khebichat et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Carbohydrates play vital roles in many biological processes, such as recognition, adhesion, and signalling between cells. The Lewis X determinant is a trisaccharide fragment implicated as a specific differentiation antigen, tumor antigen, and key component of the ligand for the endothelial leukocyte adhesion molecule, so it is necessary or essential to determine and to know their conformational and structural properties. In this work, conformational analysis was performed using molecular dynamics (MD) simulation with the AMBER10 program package in order to study the dynamic behavior of the Lewis X trisaccharide ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-OMe) and the Lewis X pentasaccharide ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-(1,3)- $\beta$ -D-Gal-(1,4)- $\beta$ -D-Glu-OMe) in explicit water model at 300 K for 10 ns using the GLYCAM 06 force field.

## 1. Introduction

In nature, carbohydrates form an important family of biomolecules, as simple or complex carbohydrates, either alone or covalently linked to proteins or lipids [1]. They play vital roles in many biological processes, such as recognition, adhesion, and communication between cells. Carbohydrates that are covalently linked to a nonsugar moiety (proteins, peptides, or lipids) are the most prominent cell-surface-exposed structures. The highly diverse structural variability of carbohydrates makes them good candidates for cell receptors and recognition molecules [2].

Determination of the three-dimensional structure of oligosaccharides and understanding the molecular basis of their recognition by receptors represent the main challenges of structural glycobiology. Elucidation of the three-dimensional structures and the dynamical properties of oligosaccharides is a prerequisite for a better understanding of the biochemistry of recognition processes and for the rational design of carbohydrate-derived drugs [1].

The human histo-blood group ABH (O) systems were the first major human alloantigens to be identified. The carbohydrate nature of the A, B, H, and Lewis antigens was

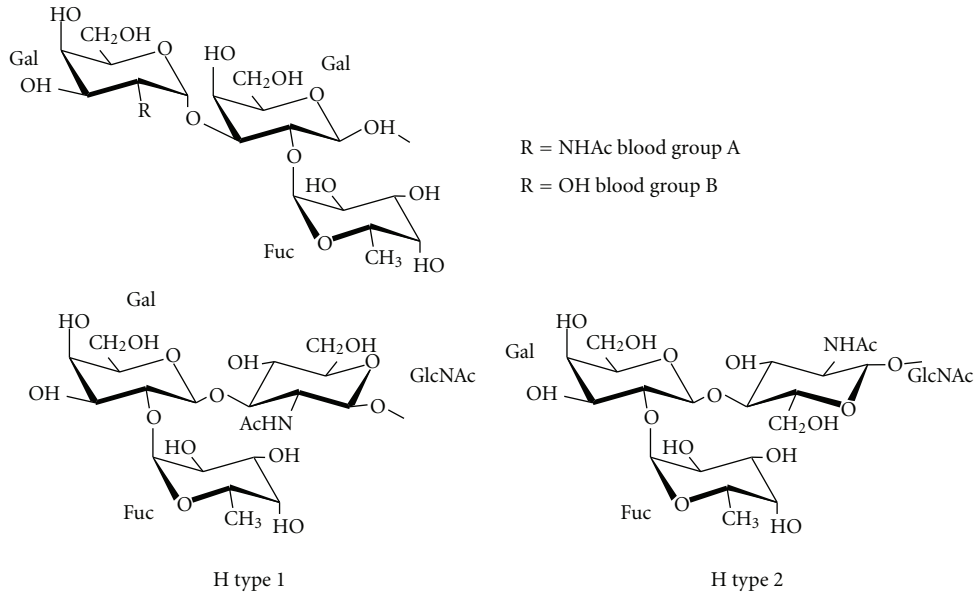
established in 1950s [3, 4]. Figure 1 shows a schematic representation of current histo-blood group antigens although additional rare types also exist [5]. The oligosaccharidic epitope of ABH (O) and Lewis histo-blood groups have been the subject of many structural investigations. The energy maps of all of the disaccharides have been established by molecular mechanics methods [6, 7].

The determination of conformational preferences of oligosaccharides is best approached by describing their preferred conformations on potential energy surfaces as a function of the glycosidic linkage  $\varphi$ ,  $\psi$  torsional angles. In the early days of conformational analysis of oligosaccharides, the calculations of either rigid or, later, relaxed conformational maps predominated. Today, complex simulation techniques such as molecular dynamics are increasingly being used. Several examples can be found in the literature of “state-of-the-art” conformational search strategies in order to simplify the complexity of the multiple-minima problem [8].

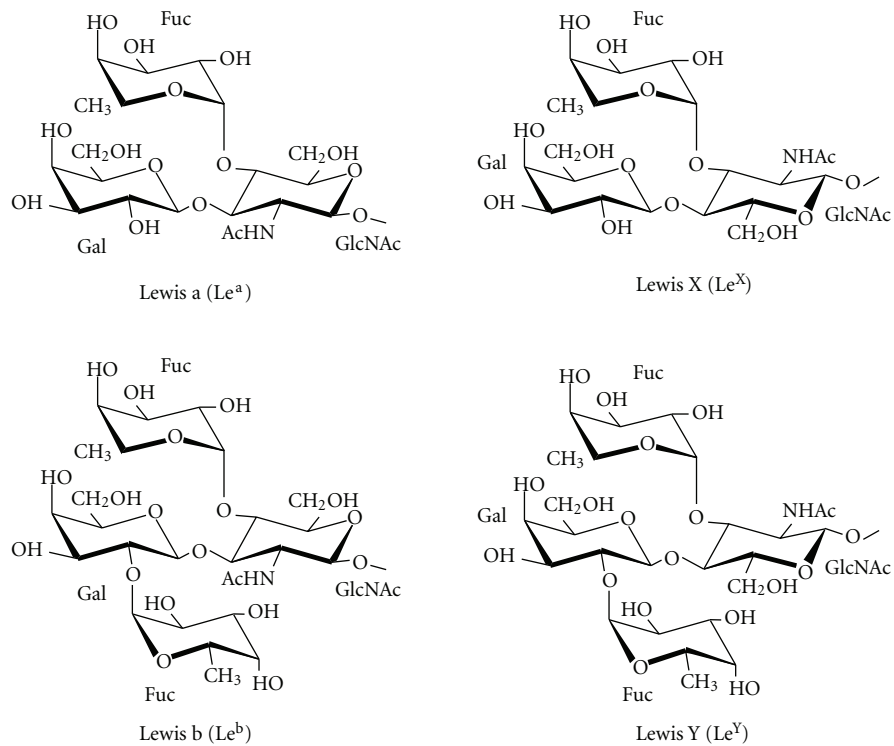
In this work, we carried out extensive, explicit solvent MD simulation studies for tri- and pentasaccharide Lewis X for 10 ns. Each monosaccharide unit in the Lewis X molecules adopts the typical  ${}^4C_1$  chair conformation, with no significant deviation from the classical pyranose ring shape.



## ABH(O) antigens



## Lewis antigens



(a)

FIGURE 1: Continued.

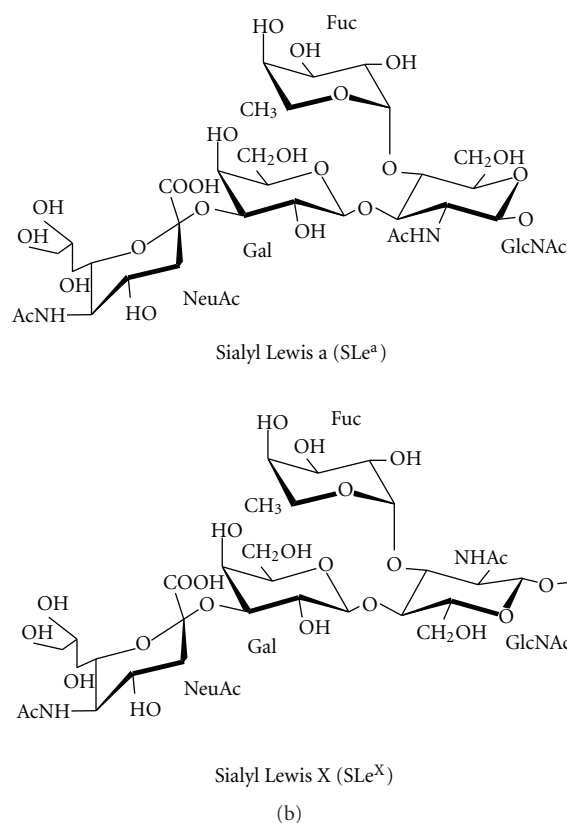


FIGURE 1: Schematic representation of histo-blood group oligosaccharides.

The objective of this study was to explore the conformational behaviour of those oligosaccharides in explicit water model at 300 K for 10 ns using the GLYCAM 06 force field.

## 2. Theoretical Approach

The atomic labelling and the torsion angles of interest are given in Figure 2. The torsion angles  $\varphi$ ,  $\psi$  are defined  $\varphi = \vartheta$  (O5-C1-O'2-C'2),  $\psi = \vartheta$  (C1-O'2-C'2-C'3), and  $\omega = \vartheta$  (O5-C5-C6-O6). The conformations with specific torsion angles  $\omega$  are referred to as gauche-trans (gt;  $\omega = 60^\circ$ ), gauche-gauche (gg;  $\omega = -60^\circ$ ) and trans-gauche (tg;  $\omega = 180^\circ$ ).

The MD simulations were carried out using AMBER10 [9] at a temperature of 300 K. This temperature was chosen because our primary objective was to study the dynamics of the oligosaccharides at physiologically relevant conditions. The starting Lewis X pentasaccharide and trisaccharide monomer coordinates and parameters were constructed with the "Glycam Biomolecule Builder" available online from the website of Woods group [10] (see Schemes 1 and 2).

Since these coordinates are not optimized, a first minimization by molecular mechanics (1,000 steps of steepest descent followed by 1,000 steps of conjugate gradient minimization with a gradient tolerance of  $0.05 \text{ kcal mol}^{-1}$ ) was made with AMBER package version 10 [9]. Force field GLYCAM 06 was used for oligosaccharides (it used to simulate the interatomic interactions responsible for oligosaccharide

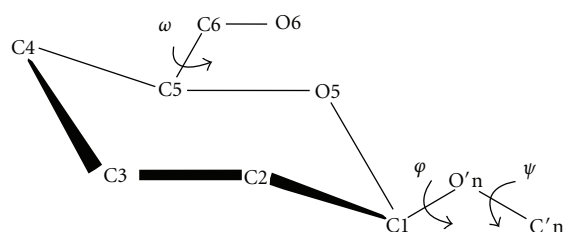


FIGURE 2: Atomic labeling and torsion angles of interest.

and glycoprotein dynamics and is currently suitable for all biologically relevant N- and O-linkages in oligo- and polysaccharides and glycoproteins [11]). The simulations were carried out using an explicit solvent environment with AMBER package version 10. The solvated system for the oligosaccharides molecules was prepared using an explicit water box of  $8 \text{ \AA}$  in the xleap module of AMBER10. The number of water molecules was 569 for the trisaccharide and 1072 for the pentasaccharide. All simulations were run with the SANDER module of AMBER with SHAKE algorithm [12] (tolerance =  $0.0005 \text{ \AA}$ ) to constrain covalent bonds involving hydrogens, using periodic boundary conditions, a 2 fs time step, a temperature coupling [13], a  $8 \text{ \AA}$  cutoff was applied to the Lennard-Jones interaction, and constant pressure of 1 atm. The nonbonded list was updated every

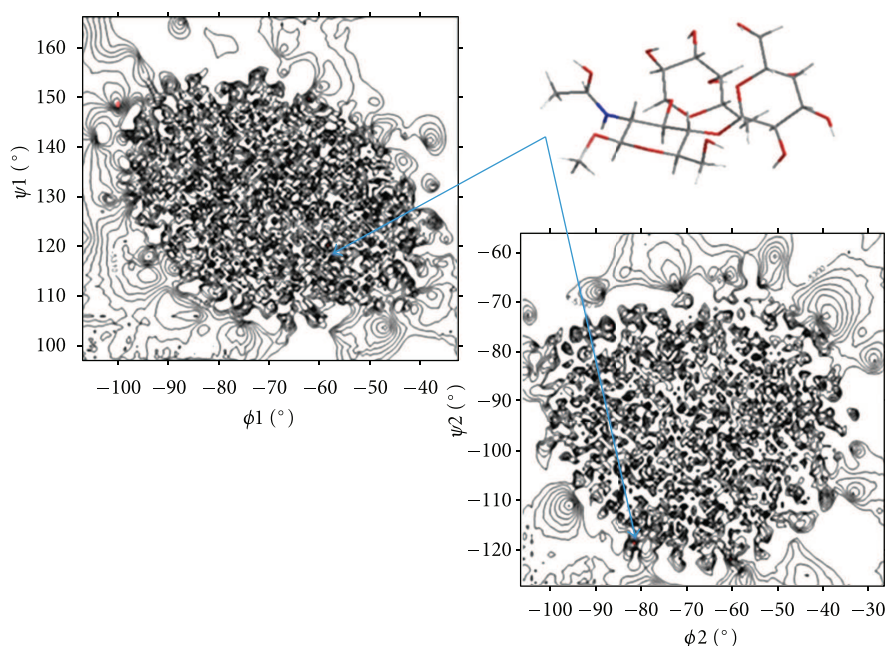


FIGURE 3: Relaxed conformational energy maps for the trisaccharide Lewis X calculated using the GLYCAM 06 force field.

10 steps. This preliminary study was aimed at finding the most energetically favorable structure which was found at 4856.9 ps for the trisaccharide and at 2411.8 ps for the pentasaccharide. Those conformations were found with a “home-made” Perl script that merely checked all energies values, finding the lowest one and giving the associated structure [2].

The PTRAJ module of AMBER10 was used to analyze and process the trajectory and coordinate files from SANDER, where the analyses might include carrying out superimposition, clustering analysis of hydrogen bonds, calculating fluctuations in bonds, angles, or dihedrals, correlation functions, and so forth, and the root mean-square deviations (RMSDs).

Finally, isoenergy contour maps were generated with ORIGIN and SURFER 9 (Golden Software, Inc., Golden Co) [14]. The positions of the minima in these maps were subsequently calculated by minimizing the energy without restraints of those grid points indicating minima on the maps.

### 3. Results

The calculated adiabatic conformational energy surfaces ( $\phi$ ,  $\psi$ ) for the tri- and pentasaccharide Lewis X are presented as potential energy contour maps in Figures 3 and 4, respectively. The contours are separated by 1 kcal/mol and range from 1 to 10 kcal/mol. The energies are given relative to the lowest minimum. Geometries and relative potential energies of the predicted minima are listed in Tables 1 and 2, and the lowest energy conformers for each compound are shown in Figures 3 and 4.

TABLE 1: Potential Energy, torsional angles for the conformers with low energies of the trisaccharide Lewis X.

	PE (kcal/mol)	$\phi_1$ (°)	$\psi_1$ (°)	$\phi_2$ (°)	$\psi_2$ (°)
Conformer 1	-5446.95	-61.84	124.7	-61.54	-104.33
Conformer 2	-5442.50	-84.91	155.63	-69.95	-92.95
Conformer 3	-5440.08	-80.60	139.25	-55.46	-98.51
Conformer 4	-5437.12	-56.58	130.28	-59.61	-80.28

**3.1. System Root-Mean-Square Displacement.** RMSD is an important parameter for describing the system stability. Figures 5 and 6 show the function curves between mass weighted RMSD fit and the time of the tri- and pentasaccharide Lewis X during dynamics simulation process, respectively.

**3.2. Computation of Phi/Psi for Modeled Oligosaccharide.** In the present study, structure prediction of oligosaccharides has been made through the torsion angles phi ( $\phi$ )/psi ( $\psi$ ). The combination of  $\phi$  and  $\psi$  angles fully determine the conformation of an oligosaccharide. Figures 3 and 4 show the distribution of sterically allowed and energy minimized conformations in  $\phi/\psi$  glycosidic space obtained from our explicit solvent MD simulations for 10 ns in the case of the tri- and pentasaccharide Lewis X, respectively. These figures also show the best conformers (which have the low energies). These conformers and their corresponding  $\phi/\psi$  angular values are reported in Tables 1 and 2 for the tri- and pentasaccharide Lewis X, respectively.

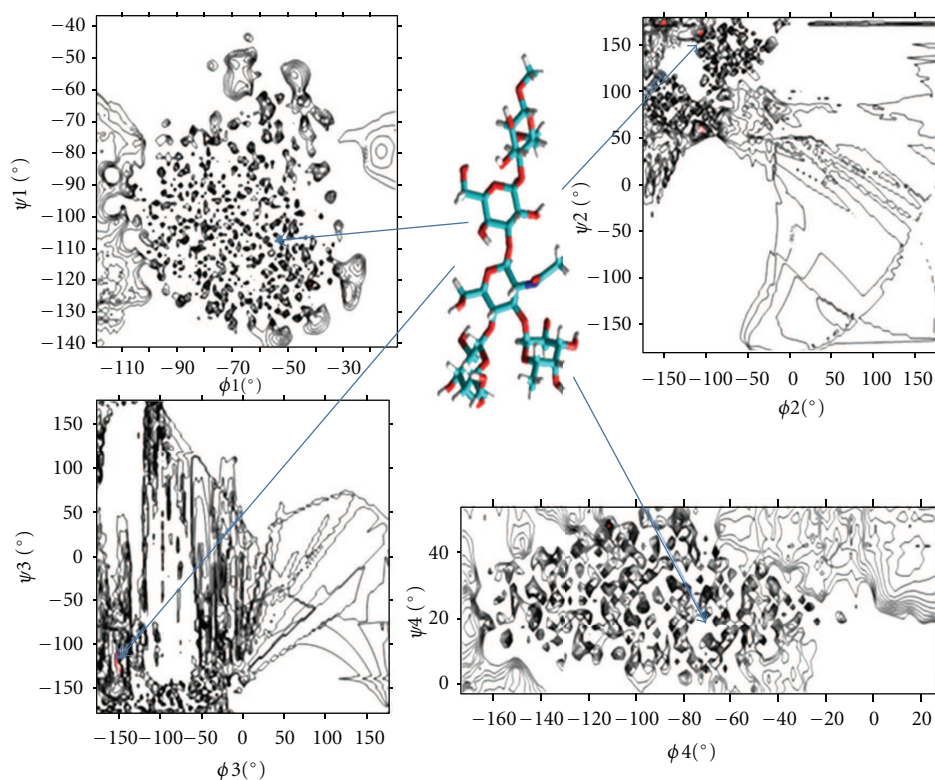


FIGURE 4: Relaxed conformational energy maps for the pentasaccharide Lewis X calculated using the GLYCAM 06 force field.

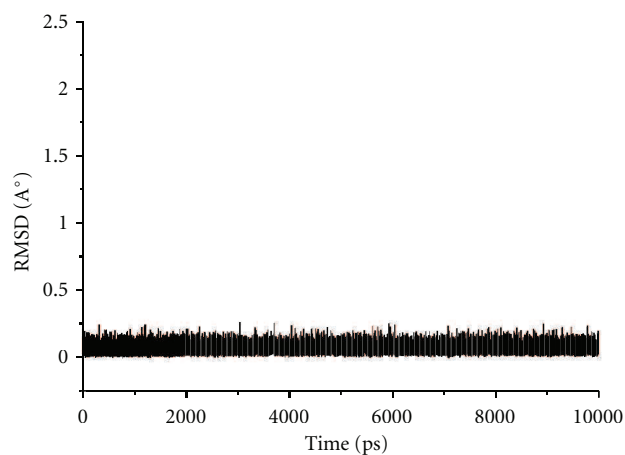


FIGURE 5: RMSD of the trisaccharide Lewis X as a function of time during MD, compared to the lowest-energy conformation.

TABLE 2: Potential energy, torsional angles for the conformers with low energies of the pentasaccharide Lewis X.

	PE (kcal/mol)	$\varphi_1$ (°)	$\psi_1$ (°)	$\varphi_2$ (°)	$\psi_2$ (°)	$\varphi_3$ (°)	$\psi_4$ (°)	$\varphi_4$ (°)	$\psi_4$ (°)
Conformer 1	-10191.15	-78.30	-97.88	-124.4	64.97	-108.27	-173.37	-57.54	15.18
Conformer 2	-10162.71	-65.65	-106.09	-73.57	136.95	-52.97	-155.29	-71.26	22.90
Conformer 3	-10162.25	-56.49	-112.26	-78.58	144.67	-55.95	-157.92	-69.85	23.71
Conformer 4	-10152.28	-71.67	-108.12	-75.73	152.13	-73.72	-163.60	-74.94	18.04

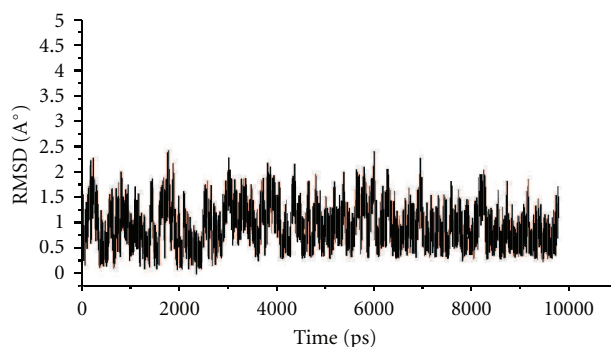
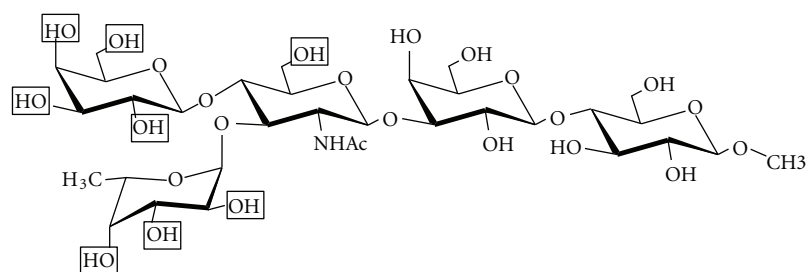
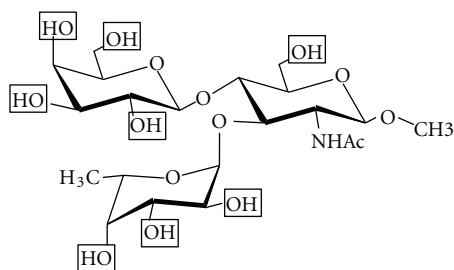


FIGURE 6: RMSD of the pentasaccharide Lewis X as a function of time during MD, compared to the lowest-energy conformation.



Lewis X pentasaccharide

SCHEME 1: Lewis X trisaccharide.



Lewis X trisaccharide

SCHEME 2: Lewis X pentasaccharide.

#### 4. Discussion

The monitoring of kinetic, global, and potential energies along the trajectory, as well as the pressure and temperature, indicate that the global periodic system is stable and does not present simulation artefacts. Kinetic energy increased slowly, whereas potential energy decreased gradually during the simulations. The total energy of a system to be simulated should be conserved, whereas the fluctuation of a temperature is caused by the interconversion of the kinetic and potential energy components [15].

The RMS variations as a function of time are displayed on Figures 5 and 6 for the tri- and penta-saccharide Lewis X

respectively, and indicate that both systems are equilibrated after 500 ps. The visual examination of the two 10 ns trajectories shows that both oligosaccharides are stable. An RMSD analysis was carried out to quantify the structural differences. The RMSD between the PDB starting structure of the simulations (obtained from Glycam biomolecule builder), and the low-energy conformers structure considering only the heavy atoms is 0.03 Å and 0.20 Å for the tri- and penta-saccharide respectively. This shows that both oligosaccharides are stable.

During the simulation of those oligosaccharides, we notice that for the pentasaccharide the most stable conformer is obtained at 2411.8 ps and that for the trisaccharide was obtained at 4856.9 ps. This means that the pentasaccharide is less flexible than the trisaccharide due to the steric effect.

#### 5. Conclusion

A complete understanding of the role of carbohydrates in biological systems is to a large extent dependent on the information available about the equilibrium mixture and about the preferred conformation of the carbohydrate molecules in solution. The conformational analysis offers a tool which can determine all possible conformations which influence the solution behavior of carbohydrates. This work represents a theoretical study of two oligosaccharides tri- and pentasaccharide Lewis X using molecular dynamic simulations in explicit water with the AMBER10 program

package. RMSD measures of the accuracy of carbohydrates around their average conformations play an important indicator of many biological processes such as macromolecular recognition and complex formations [16].

## Acknowledgments

The authors are grateful to Dr. R. Borsali, Dr. K. Mazeau, Dr. A. Rivet, and Dr. A. Nurisso (CERMAV) France, and they thank particularly Mr. Guendouzi A for his help.

## References

- [1] A. Imberty and S. Perez, "Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations," *Chemical Reviews*, vol. 100, no. 12, pp. 4567–4588, 2000.
- [2] Y. Luo, F. Barbault, C. Gourmala et al., "Cellular interaction through LewisX cluster: theoretical studies," *Journal of Molecular Modeling*, vol. 14, no. 10, pp. 901–910, 2008.
- [3] E. A. Kabat, "Immunochemical studies on the carbohydrate moiety of water-soluble blood group A, B, H, Le<sup>a</sup>, and Le<sup>b</sup> substances and their precursor I antigens," in *Carbohydrates in Solution*, H. Isbell, Ed., vol. 117, pp. 334–355, American Chemical Society, Washington, DC, USA, 1973.
- [4] W. M. Watkins, "Biochemistry and genetics of the ABO, Lewis, and P blood group systems," in *Advances in Human Genetics*, H. Harris and K. Hirsch-Horn, Eds., vol. 10, pp. 1–136, Plenum Press, New York, NY, USA, 10th edition, 1980.
- [5] S. I. Hakomori, "Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer," *Biochimica et Biophysica Acta*, vol. 1473, no. 1, pp. 247–266, 1999.
- [6] F. Bizik and I. Tvaroska, "Conformational analysis of disaccharide fragments of blood group determinants in solution by molecular modelling," *Chemical Papers*, vol. 49, no. 4, pp. 202–214, 1995.
- [7] A. Imberty, E. Mikros, J. Koca, R. Mollicone, R. Oriol, and S. Perez, "Computer simulation of histo-blood group oligosaccharides: energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens," *Glycoconjugate Journal*, vol. 12, no. 3, pp. 331–349, 1995.
- [8] T. Kozar and C.-W. von der Lieth, "Efficient modelling protocols for oligosaccharides: from vacuum to solvent," *Glycoconjugate Journal*, vol. 14, no. 8, pp. 925–933, 1997.
- [9] D. Case, T. Darden, T. Cheatham et al., "AMBER," University of California, CA, USA, 10th edition, 2006.
- [10] R. J. Woods, R. A. Dwek, C. J. Edge, and B. Fraser-Reid, "Glycam biomolecule builder," [http://www.glycam.com/CCRC/biombuilder/biomb\\_index.jsp](http://www.glycam.com/CCRC/biombuilder/biomb_index.jsp).
- [11] C. Gourmala, Y. Luo, F. Barbault et al., "Elucidation of the LewisX-LewisX carbohydrate interaction with molecular dynamics simulations: a glycosynapse model," *Journal of Molecular Structure*, vol. 821, no. 1–3, pp. 22–29, 2007.
- [12] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.
- [13] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, article 3684, 7 pages, 1984.
- [14] E. Y. Yoon, "A conformational study of linkage positions in oligosaccharides investigated by 2-D NMR spectroscopy and molecular modeling," *Bulletin of the Korean Chemical Society*, vol. 24, no. 3, pp. 339–344, 2003.
- [15] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1987.
- [16] E. B. Philip and W. Helge, *Structural Bioinformatics*, 1st edition, 2003.

### **A. Recherche conformationnelle sur la structure des tripeptides par dynamique moléculaire: étude de L,L-Gly-Phe-Phe, L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly**

#### **1. Introduction**

Les oligopeptides naturels constitués de 2 à 50 résidus d'acides aminés [49,50] se trouvent dans pratiquement tous les organismes vivants. Leurs fonctions sont si diverses qu'il est difficile d'indiquer un processus d'une importance vitale qui se produit sans leur participation. Ces oligopeptides sont généralement appelés régulateurs, parce que la majorité d'entre eux sont impliqués dans la régulation des processus physiologiques dans tous les systèmes de régulation: systèmes nerveux, endocrinien et immunitaire [51]. En outre leurs fonctions de régulation dépassent les limites d'un organisme ou espèce biologique. Ainsi, les toxines oligopeptidiques des eucaryotes jouent un rôle important dans la régulation des relations interspécifiques, et les oligopeptides antimicrobiens des procaryotes régulent la concurrence dans l'occupation de niches écologiques et agissent comme des molécules de signalisation dans les communications intercellulaires [52].

La connaissance complète de la structure chimique d'un oligopeptide est basée sur la détermination de sa séquence d'acides aminés, la détection des modifications de ses N et C terminaux, et d'autres modifications post-traductionnelles, ainsi que de la liaison S-S et les autres liaisons chimiques intermoléculaires entre les résidus d'acides aminés éloignés, y compris la liaison peptidique supplémentaire produite lors de la cyclisation. La détermination directe de la structure chimique exacte des oligopeptides est associée, dans de nombreux cas, à de grandes difficultés en raison de leur teneur naturelle est souvent  $10^{-15}$ - $10^{-12}$  M seulement [53]. L'emplacement des oligopeptides dans la cellule est également associé à des problèmes.

## Chapitre IV : résultats et discussions

---

Par conséquent, des techniques biochimiques et immunohistochimiques très sensibles sont utilisés pour les analyses.

Les approches computationnelles sont utilisées de plus en plus pour explorer la structure et les interactions de différentes macromolécules biologiques. Une telle approche computationnelle est la dynamique moléculaire (DM), qui permet l'exploration des systèmes biologiques à l'échelle atomique et peut fournir de précieuses indications sur les propriétés dynamiques de ces systèmes.

Dans ce travail, nous avons effectué une étude de simulation DM pour trois oligopeptides: L, L-Gly-Phe-Phe (commercial), L, L-Phe-Gly-Phe, et L, L-Phe-Phe-Gly (synthétisé par le Pr S. Ghalem) [13]. Ces oligopeptides ont été simulés pendant 10 ns dans le modèle d'eau explicite à 300 K en utilisant le champ de force ff99SB. L'objectif de cette étude était de comprendre la stabilité structurale de ces oligopeptides contenant des chaînes latérales aromatiques non-coordinantes. Nous savons que dans les systèmes protéiques, les acides aminés portant une chaîne latérale aromatique, telle que la tyrosine, la phénylalanine, etc ... jouent souvent des rôles importants dans la complexation avec les ions métalliques. Souvent ces chaînes latérales ne constituent pas directement comme un site de ligand. Mais leurs présence influence largement la complexation par différents effets, par exemple, l'effet stérique, l'effet  $\pi$ -d, l'effet hydrophobe, etc... [54,12].

En effet, un pentapeptide tel que l'enképhaline (Tyr-Gly-Gly-Phe-Met ou Tyr-Gly-Gly-Phe-Leu) de faible masse moléculaire [55] est décrit par certains auteurs [56,57,58] comme étant une substance endogène existant dans le cerveau et réagit comme antagoniste aux sites récepteurs d'opiacés. Toute modification des résidus récepteurs tyrosyl en position N-terminal pouvant conduire à une forme protonée dans cette position, au pH physiologique, inhibe l'activité biologique [59]. La disposition spatiale relative des chaînes latérales aromatiques de

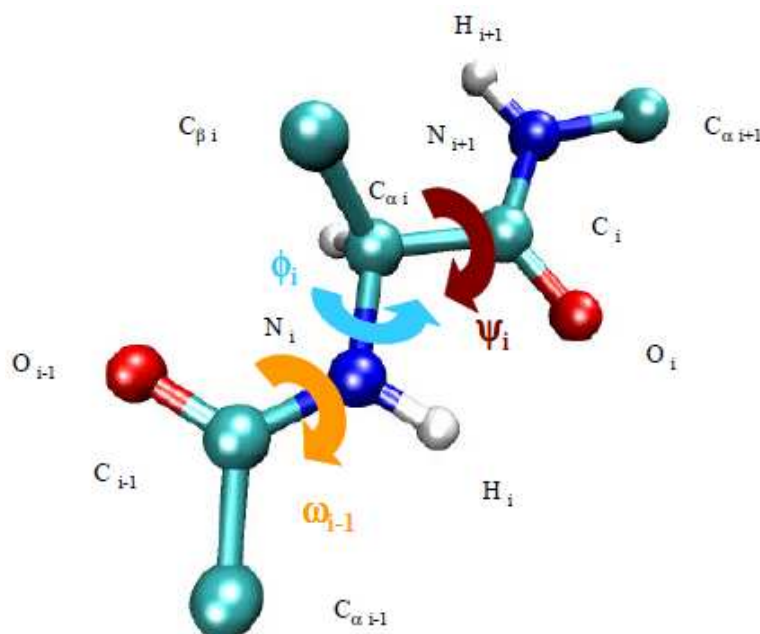


## Chapitre IV : résultats et discussions

la phénylalanine est aussi intéressante en raison d'une ressemblance possible avec l'enképhaline. La complexation du cuivre (II) par [5-leucine] - et la [5-méthionine]-enképhaline a fait l'objet d'études dans les conditions physiologiques [60]. Afin de préciser la conformation biologiquement active de l'enképhaline [61,62], il a été remarqué que, dans la séquence de l'enképhaline, la modification des résidus L-Tyr<sup>1</sup>, Gly<sup>2</sup>, et L-Phe<sup>4</sup> entraîne la diminution significative de l'activité. Au contraire, le résidu Gly<sup>2</sup> pourrait être remplacé par un autre acide de la conformation D [63], sans modification sensible d'activité.

### 2. Méthodes

Le marquage atomique et les angles de torsion sont donnés dans la figure. 1. Les angles dièdres du squelette des protéines sont appelés  $\phi$  (phi, impliquant les atomes C'-N-C <sup>$\alpha$</sup> -C'), ( $\psi$  (psi, impliquant les atomes N-C <sup>$\alpha$</sup> -C'-N) et  $\omega$  oméga, impliquant les atomes C <sup>$\alpha$</sup> -C'-N-C <sup>$\alpha$</sup> ). Ainsi, la torsion  $\phi$  contrôle la distance C'-C',  $\psi$  contrôle la distance N-N et  $\omega$  contrôle la distance C <sup>$\alpha$</sup> -C <sup>$\alpha$</sup> .



**Figure 1:** marquage atomique et les angles de torsion d'intérêt

## Chapitre IV : résultats et discussions

---

Les simulations de dynamique moléculaire ont été réalisées à l'aide AMBER10 [39] à une température de 300 K. Cette température a été choisie parce que notre objectif principal était d'étudier la dynamique des oligopeptides dans des conditions physiologiquement pertinents. Les coordonnées de départ des oligopeptides sont construites en utilisant la banque de données des acides aminés inclus dans Hyperchem 7,5. Les simulations ont été réalisées en utilisant un solvant (eau) explicite avec le programme AMBER 10. Les systèmes ont été préparés en utilisant une boîte d'eau explicite de 8 Å (TIP3PBOX 8) dans le module Xleap d'AMBER10. Le nombre de molécules d'eau était de 906 molécules, identique pour les trois tripeptides.

Étant donné que ces coordonnées ne sont pas optimisées, une première minimisation par mécanique moléculaire (1000 pas de steepest descent suivie par 1000 pas du conjugate gradient) a été faite avec AMBER10 pour éliminer tous les contacts défavorables et ensuite équilibrée pendant 120 ps pour amener la température de  $\sim 0$  K à 300 K, ensuite la phase de production a été effectuée pendant 10 ns. Le champ de force ff99SB a été utilisée pour ces oligopeptides. Toutes les simulations ont été effectuées avec le module SANDER d'AMBER avec l'algorithme SHAKE [42] (tolérance =  $0.0005\text{Å}$ ) pour contraindre les liaisons covalentes impliquant les hydrogènes, en utilisant des conditions aux limites périodiques, un pas de temps de 2 fs, un couplage de température [43], un cut-off de 8 Å a été appliqué à l'interaction de Lennard-Jones, et une pression constante de 1 atm. Les conditions aux limites périodiques ont été utilisées pour annuler les effets de surface à des limites de la boîte. La méthode de sommation d'Ewald (particle mesh Ewald summation) [64] a été utilisée pour calculer le potentiel électrostatique. Les liaisons contenant des hydrogènes ont été contraintes à l'aide SHAKE [42], et ainsi les interactions impliquant des liaisons hydrogènes n'ont pas été calculées de façon concomitante. Cette étude préliminaire visait à trouver la structure la plus favorable énergétiquement qui a été trouvée à 3696,2 ps pour L, L-Gly-Phe-Phe, à 3855,6 ps

## Chapitre IV : résultats et discussions

---

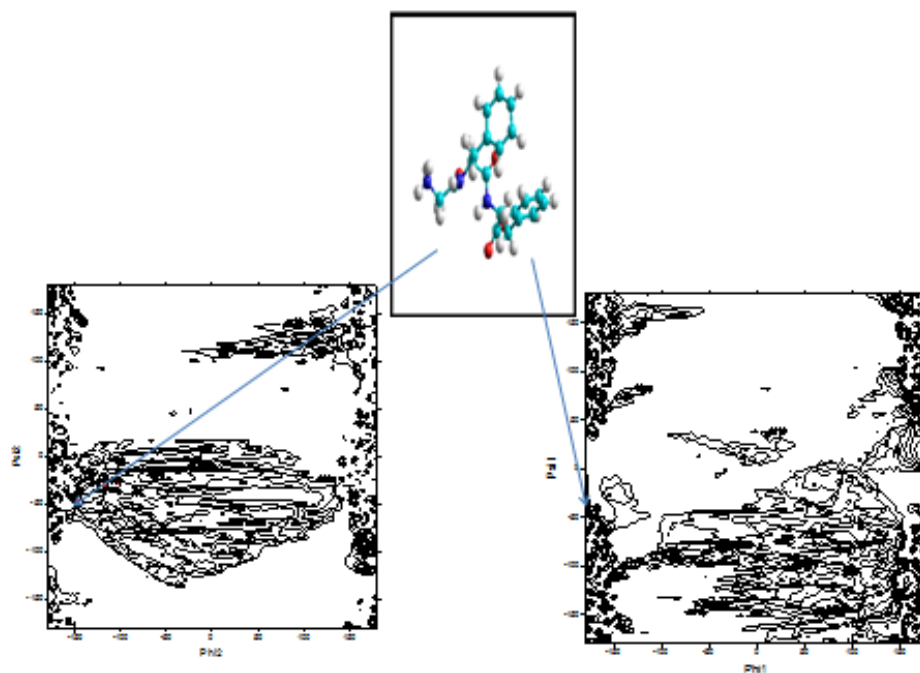
pour L, L-Phe-Gly-Phe, et à 5634,8 ps pour L, L-Phe-Phe-Gly. Ces conformations ont été trouvées avec un script Perl "home-made" qui se contentent de vérifier toutes les valeurs des énergies, trouver la plus basse et en donnant la structure associée [32].

Le module de PTRAJ du programme AMBER10 a été utilisé pour analyser et traiter les fichiers de la trajectoire et des coordonnées à partir SANDER, où les analyses pourraient inclure la réalisation de superposition, l'analyse des liaisons hydrogènes, le calcul des fluctuations dans des liaisons, des angles ou des dièdres, des fonctions de corrélation, etc et la racine quadratique moyenne des écarts (RMSDs).

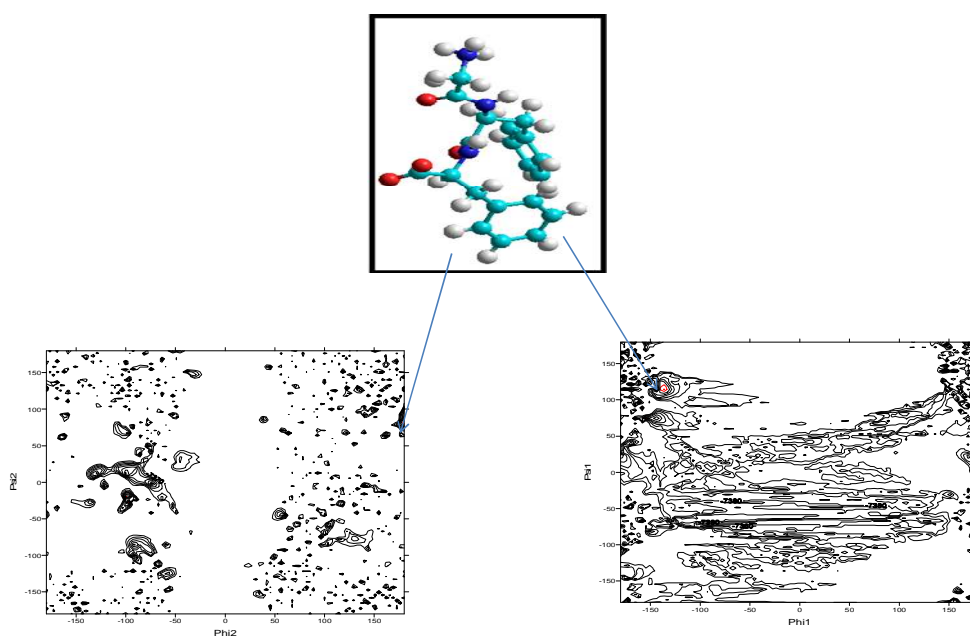
Enfin, des cartes de contours iso-énergie ont été générés avec les logiciels ORIGIN et SURFER 9 [44]. Les positions exactes des minima dans ces cartes ont été ensuite calculées en minimisant l'énergie sans retenue ces points indiquant les minima sur les cartes.

### 3. Résultats:

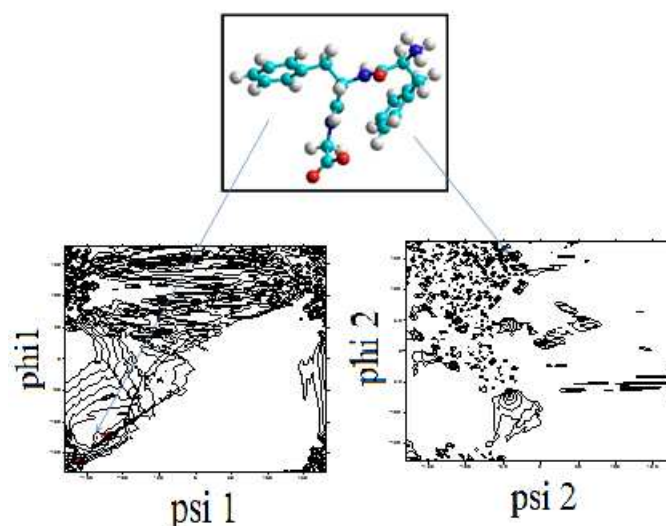
Les surfaces d'énergie conformationnelle adiabatiques ( $\phi$ ,  $\psi$ ) calculées pour les trois tripeptides sont présentées comme des courbes d'énergie potentielle dans les Figure2, figure 3 et figure 4 (les conformères d'énergie plus basse pour chaque composé sont indiqués). Les contours sont séparées par 1 kcal/mol et allant de 1 à 10kcal/mol. Les énergies sont données par rapport au minimum le plus bas. Les géométries et les énergies potentielles relatives des minimums prédits sont énumérées dans le tableau 1, tableau 2 et tableau 3, et les conformères de plus basse énergie pour chaque composé sont indiqués dans les figures 2, 3 et 4.



**Figure2:** Carte adiabatique de RAMCHANDRAN du tripeptide L,L-Gly-Phe-Phe calculée eu utilisant le champ de force ff99SB.

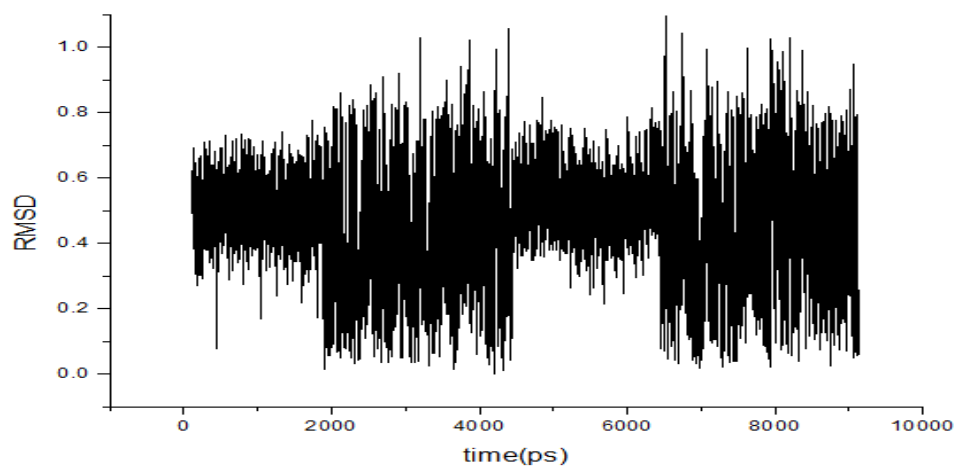


**Figure3 :** Carte adiabatique de RAMCHANDRAN du tripeptide L,L-Phe-Gly-Phe calculée eu utilisant le champ de force ff99SB.

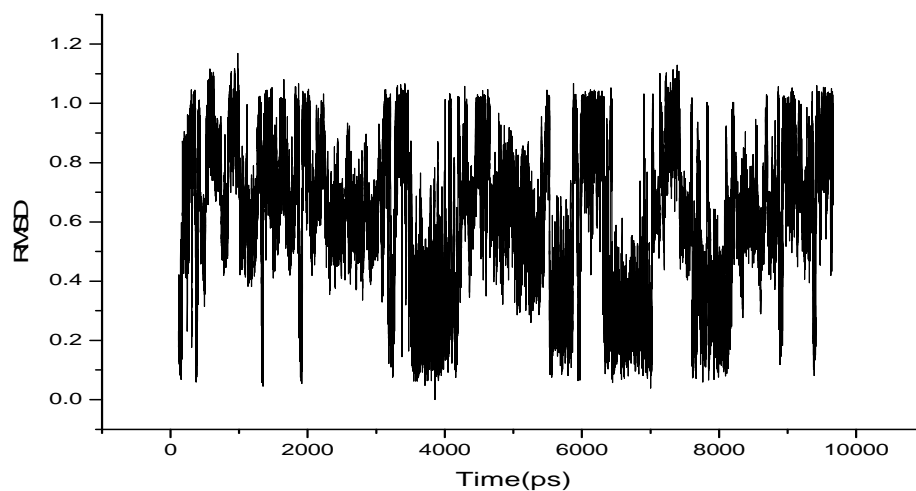


**Figure 4:** Carte adiabatique de RAMCHANDRAN du tripeptide L,L-Phe-Phe-Gly calculée en utilisant le champ de force ff99SB.

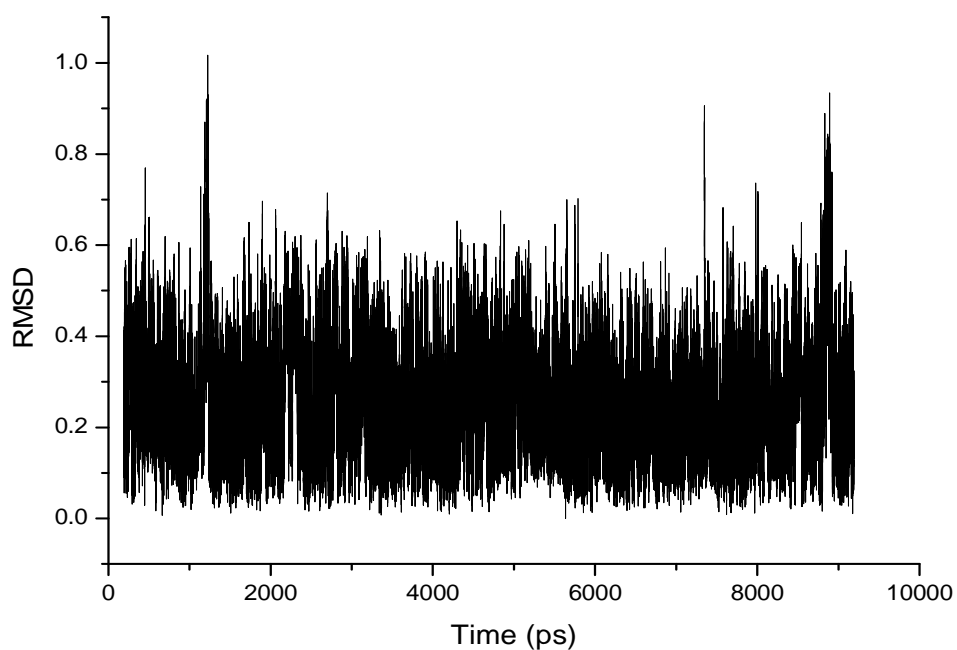
**3.1. La racine quadratique moyenne du déplacement.** RMSD est un paramètre important pour décrire la stabilité du système. Les figures 5, 6 et 7 montrent les courbes de fonction entre la masse ajustement RMSD pondérée et le temps de L, L-Gly-Phe-Phe, L, L-Phe-Gly-Phe, et L, L-Phe-Phe-Gly, respectivement pendant le processus de simulation.



**Figure 5 :** RMSD du tripeptide L,L-Gly-Phe-Phe en fonction du temps, par rapport à la conformation de plus basse énergie.



**Figure 6 :** RMSD du tripeptide L,L-Phe-Gly-Phe en fonction du temps, par rapport à la conformation de plus basse énergie.



**Figure7:** RMSD du tripeptide L,L-Phe-Phe-Gly en fonction du temps, par rapport à la conformation de plus basse énergie.

## Chapitre IV : résultats et discussions

**3.2. Calcul de Phi / Psi pour Oligopeptide:** Dans la présente étude, la prédiction des structures des oligopeptides a été réalisée par les angles de torsion phi ( $\phi$ ) / psi ( $\psi$ ). La combinaison des angles  $\phi$  et  $\psi$  détermine entièrement la conformation d'un oligopeptide. Les figures 2, 3 et 4 montrent la répartition des conformations permises stériquement d'énergie minimisée dans l'espace peptidique  $\phi / \psi$  obtenus à partir de nos simulations de DM pendant 10 ns dans le cas des trois tripeptides, L,L-Gly-Phe-Phe, L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly, respectivement. Ces figures montrent également les meilleurs conformères (qui ont des basses énergies). Ces conformères et de leurs valeurs angulaires  $\phi/\psi$  correspondants sont donnés dans les tableaux 1, 2 et 3 pour le L, L-Gly-Phe-Phe, L, L-Phe-Gly-Phe, et L, L-Phe-Phe-Gly, respectivement.

**Table1:** Energie potentielle, angles de torsion des conformères de plus basses énergies du tripeptide L,L-Gly-Phe-Phe.

	<b>TE</b>	<b><math>\phi</math>1</b>	<b><math>\phi</math>2</b>	<b><math>\psi</math>1</b>	<b><math>\psi</math>2</b>	<b>Omega1</b>	<b>Omega2</b>
<b>Conformer1</b>	<b>-4975.023</b>	<b>-175.34</b>	<b>174.31</b>	<b>115.98</b>	<b>-10.54</b>	<b>-102.95</b>	<b>88.56</b>
<b>Conformer2</b>	<b>-4974.554</b>	<b>-161.67</b>	<b>160.58</b>	<b>147.10</b>	<b>4.089</b>	<b>129.69</b>	<b>-91.56</b>
<b>Conformer3</b>	<b>-4969.982</b>	<b>179.44</b>	<b>177.60</b>	<b>126.45</b>	<b>160.75</b>	<b>-104.77</b>	<b>149.71</b>

## Chapitre IV : résultats et discussions

**Table2:** Energie potentielle, angles de torsion des conformères de plus basses énergies du tripeptide L,L-Phe-Gly-Phe.

	TE	$\phi_1$	$\phi_2$	$\psi_1$	$\psi_2$	Omega1	Omega2
Conformer1	-6168.924	-177.17	-109.38	138.19	164.62	-109.38	145.41
Conformer2	-6162.356	-168.48	-73.69	120.77	158.49	-73.69	132.42
Conformer3	-6159.94	-169.81	-86.44	126.26	161.015	-86.44	-134.4

**Table3:** Energie potentielle, angles de torsion des conformères de plus basses énergies du tripeptide L,L-Phe-Phe-Gly.

	TE	$\phi_1$	$\phi_2$	$\psi_1$	$\psi_2$	Omega1	Omega2
Conformer1	-4405.32	179.31	-122.99	157.36	124.24	-122.99	166.16
Conformer2	-4397.93	166.41	-92.50	129.74	162.51	-92.50	68.31
Conformer3	-4386.12	-170.36	-68.01	155.03	154.73	-68.01	74.37

### 4. discussion

Le suivi des énergies cinétiques, globales, et potentielles le long de la trajectoire, ainsi que la pression et la température, indiquent que le système périodique global est stable et ne présente pas de simulation artéfact. L'énergie cinétique a augmenté lentement, alors que l'énergie potentielle a diminué progressivement au cours des simulations. L'énergie totale d'un système à simuler doit être conservée, tandis que la fluctuation de la température est provoquée par les composantes de l'énergie cinétique et le potentiel [45].

Les variations RMSD en fonction du temps (les figures 5, 6 et 7 pour la L, L-Gly-Phe-Phe, L, L-Phe-Gly-Phe, et L, L-Phe-Phe-Gly, respectivement) indiquent que les trois



## Chapitre IV : résultats et discussions

---

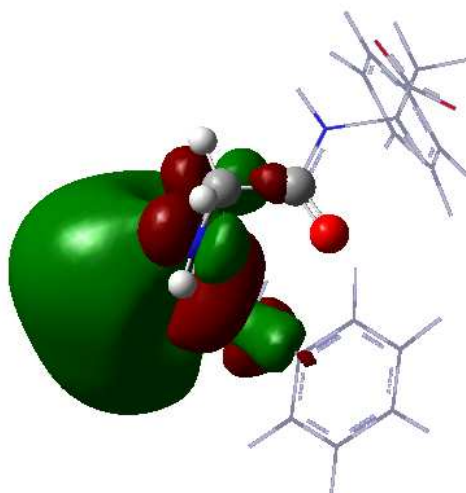
systèmes sont équilibrés après 500 ps. L'examen visuel des trois trajectoires de 10 ns montre que les trois oligopeptides sont stables. Une analyse RMSD a été réalisée pour quantifier les différences structurales. Le RMSD entre les structures PDB de départ (construits avec la banque de données des acides aminés incluse dans Hyperchem 7,5), et les conformères de faibles énergies en considérant seulement les atomes lourds est de 0,19 Å, 0,41 Å et 0,1 Å pour la L,L-Gly-Phe-Phe, L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly, respectivement. Cela montre que les trois oligopeptides sont stables.

Au cours de la simulation de ces oligopeptides, nous remarquons que pour L,L-Gly-Phe-Phe le conformère le plus stable est obtenue à 3696,2 ps, pour L,L-Phe-Gly-Phe est obtenu à 3855,6 ps, et pour L,L-Phe-Phe-Gly a été obtenu à 5634,8 ps. Cela signifie que L, L-Gly-Phe-Phe est moins flexible que L,L-Phe-Gly-Phe, qui est moins flexible que L,L-Phe-Phe-Gly à cause de la gêne stérique.

Pour ces trois tripeptides contenant les mêmes résidus d'acides aminés mais différent par la position du résidu glycine, et à la lumière des résultats obtenus par le calcul:  $E_{L,L-Phe-Gly-Phe} < E_{L,L-Gly-Phe-Phe} < E_{L,L-Phe-Phe-Gly}$ , nous remarquons que l'ordre de stabilité des systèmes étudiés est le suivant: L,L-Phe-Gly-Phe > L,L-Gly-Phe-Phe > L,L-Phe-Phe-Gly. Cet ordre de stabilité est dû aux effets stérique (le volume et la position du groupe phényle) et électroniques (distribution électronique).

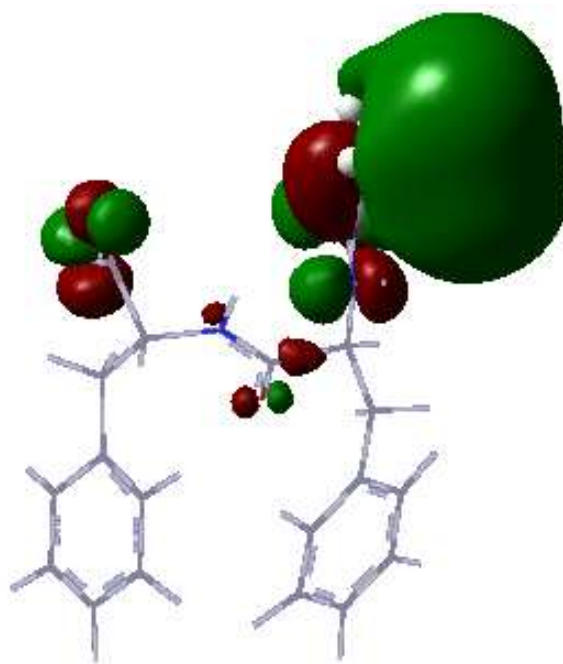
Pour confirmer cela, j'ai fait un calcul de population électronique au niveau DFT avec la base 6-31G\* à l'aide du programme Gaussian 09 [65].

Pour le L,L-Phe-Gly-Phe tripeptide, nous notons que la position des groupements phényle est homogène (figure 8) ce qui permettra de créer un équilibre stérique et électronique.

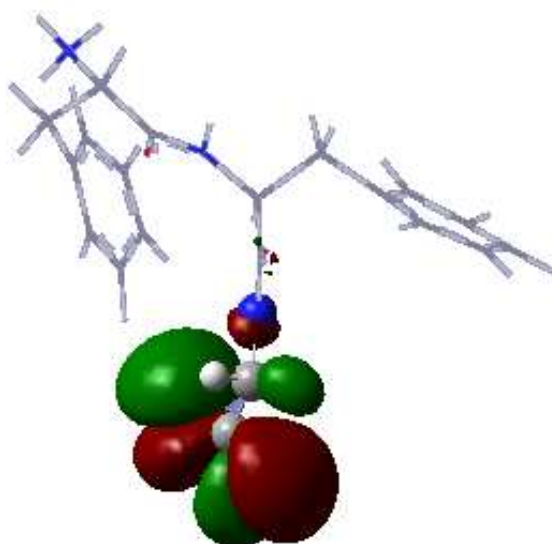


**Figure 8:** La distribution électronique donnée par le programme Gaussian 09 du conformère le plus stable du tripeptide L,L-Phe-Gly-Phe trouvé par simulation moléculaire dynamique (AMBER10).

Alors que pour les autres tripeptides L, L-Gly-Phe-Phe et L, L-Phe-Phe-Gly, la géométrie obtenue montre que la position des résidus phényle n'est pas homogène, plus nous constatons une augmentation significative de la stabilité lorsque le groupement aromatique est en position C-terminal (L,L-Gly-Phe-Phe) par rapport au tripeptide contenant les mêmes acides aminés, mais le phényle est en résidu N-terminal (L,L-Phe-Phe-Gly),  $E_{L,L-Gly-Phe-Phe} < E_{L,L-Phe-Phe-Gly}$ . Ainsi, pour le tripeptide L,L-Gly-Phe-Phe, les deux groupements phényle sont situés dans le même côté (figure 9), tandis que pour le tripeptide L,L-Phe-Phe-Gly, le groupement phényle du troisième résidu d'acide aminé se situe dans le même côté que N-terminal (figure 10), ce qui provoque une répulsion électronique, d'où l'instabilité de ce tripeptide par rapport aux autres tripeptides.



**Figure 9:** La distribution électronique donnée par le programme Gaussian 09 du conformère le plus stable du tripeptide L,L-Gly-Phe- Phe trouvé par simulation moléculaire dynamique (AMBER10).



**Figure 10:** La distribution électronique donnée par le programme Gaussian 09 du conformère le plus stable du tripeptide L,L-Phe-Phe-Gly trouvé par simulation moléculaire dynamique (AMBER10).

### 5. Conclusion

Une compréhension complète du rôle des peptides dans les systèmes biologiques dépend de l'information disponible sur la conformation préférée des peptides en solution. L'analyse conformationnelle offre un outil qui permet de déterminer toutes les conformations possibles qui influencent le comportement des peptides en solution. Ce travail représente une étude théorique de trois oligopeptides: L, L-Gly-Phe-Phe, L, L-Phe-Gly-Phe, et L, L-Phe-Phe-Gly par dynamique moléculaire dans l'eau explicite avec le programme AMBER10. Les mesures RMSD des oligopeptides autour de leurs conformations moyennes jouent un rôle d'indicateur important de nombreux processus biologiques tels que la formation des complexes. Ainsi la présence d'un cycle aromatique dans le résidu C-terminal donne généralement un tripeptide plus stable par rapport au tripeptide contenant les mêmes acides aminés, mais le phényle est en résidu N-terminal.

## Chapitre IV : résultats et discussions

---

### Reference:

- [1] H. Sigel et R.B. Martin, *Chem.Rev.*, 82: (1982), 385-426.
- [2] J.W.H. Robert et L.D. Pettit, *J.C.S. Dalton.*, (1981), 13331.
- [3] A. Kaneda et A.E. Martell, *J. Am. Chem. Soc.*, 99: (1977), 1586.
- [4] B.R. Rabin, *Trans. Faraday. Soc.*, 52: (1956), 1130.
- [5] R. P. Bnomo, R. Cali et V. cucinotta, *Inorg. Chem.*, 25: (1986), 1641-1646.
- [6] T. Kiss, Z. Szocs, *J. Chem. Soc. Dalton. Trans.*, (1986), 2444.
- [7] R.M. Izatt, J.W. Wrathall et K. P. Anderson, *J. Phys. Chem.*, 65: (1961), 1914.
- [8] A. Wahlborg et E. Friedon, *Arch. Biochem. Biophys.*, 111: (1965), 672.
- [9] A. Gergerly, I. Savago, I. Nagypal et R. Kiraly, *Inorg. Chim. Acta.*, 6: (1972), 435.
- [10] A. Gergerly, I. Savago, *J. Inorg. Nucl. Chem.*, 35: (1973), 4355.
- [11] L. Xiao, M. Jouini, B. T. Fan, G. Lapluye, J.Huet, *J. Chem. Soc. Dalton Trans.*, (1989)
- [12] S. Ghalem, BT Fan, L. Xiao, G. lapluye, *J. Chim. Phys.*, 1994, 91, 113-125.
- [13] S. Ghalem, *Thèse de Doctorat*, Université Denis Diderot- Paris VII- France, (1996).
- [14] A. Bouraoui, M. Fathallah, B. Blaive, R. Gallo, *J. Chem. Soc. Perkin Trans. , 2:* (1990), 1211-1214.
- [15] A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive, R. Gallo, *Proceeding of an International Meeting*, Nancy, France, 11-15, Septembre, ed. J.L. Rivail., *Studies in Physical and Theoretical Chemistry*, 71: (1989), 381-393.
- [16] A. Zinelabidine, A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive, R. Gallo, *J. Mol. Struct. Theochem.*, 286: (1993), 267-274.
- [17] B. Blaive, G. Legsaï, R. Lai, *J. Mol. Struct.*, 354: (1995), 245-250.
- [18] Z. Li, H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, 84: (1987), 6611.
- [19] L.D. Petit, R.J.W. Hefford, in: H. Sigel (Ed.), *Metal Ions in Biological Systems*, 9: (1979), 173.
- [20] R.P. Bonomo, R. Cali, V.G.I. Cucinotti, E. Rizzarelli, *Inorg.Chem.*, 25: (1986), 1641.
- [21] S.H. Kim, R.B. Martin, *J.Am.Chem. Soc.*, 106: (1984), 1707.

## Chapitre IV : résultats et discussions

---

- [22] L. Benmenni, Thèse de doctorat en sciences, Fac. des Sciences de St-Jérôme, Université d'Aix-Marseille III (1994).
- [23] A. Bondi, *J. Phys. Chem.* 68: (1964), 441.
- [24] J. Brickmann, MOLCAD Molecular Modelling Package, Institut für physikalische Chemie, Technische Hochschule, 6100 Darmstadt, Germany (1989).
- [25] L. Benmenni, E.H. Allilou, M. Giorgi, M. Pierrot, M. Regulier, *J. Chem. Crystallogr.*, 24: (1994), 5.
- [26] J. Peisach, W.E. Blumberg, *Arch. Biochem. Biophys.*, 165: (1974), 691.
- [27] D.M. Dooley, J. Rawlings, J.H. Dawson, P.J. Stephenes, L.E. Andresson, B.G. Malmstrom, H.B. Grey, *J. Am. Chem. Soc.* , 101: (1979), 5038.
- [28] O. Yamauchi, K. Tsujide, A. Odani, *J. Am. Chem. Soc.*, 107: (1985), 659.
- [29] W.L. Kwik, K.P. Ang, G.I. Chen, *J. Inorg. Nucl. Chem.*, 42: (1980), 303.
- [30] P.I. Vestues, R.B. Martin, *J. Am. Chem. Soc.*, 102: (1980), 7906.
- [31] A. Imberty, S. Perez, *Chem. Rev.*, 100: (2000), 4567–4588
- [32] Y. Luo, F. Barbault, C. Gourmala, Y. Zhang, F. Maurel, Yongzhou Hu, Bo Tao Fan *J Mol Model.*, 14: (2008), 901–910.
- [33] E. A. Kabat, In Carbohydrates in Solution; Isbell, H., Ed.; *American Chemical Society: Washington, DC*, 117: (1973), 334-355.
- [34] W. M. Watkins, “In Advances in human genetics”; Harris, H., Hirsch-horn, K., Eds.; *Plenum Press: New York*, 10: (1980), 1-136.
- [35] S. I. Hakomori, *Biophys. Biochim. Acta.*, 1473: (1999), 247.
- [36] F. Bizik, Tvaroska, *I. Chem. Pap.*, 49: (1995), 202.
- [37] A. Imberty, E. Mikros, J. Koca, R. Mollicone, R. Oriol, S. Perez, *Glycoconjugate J.*, 12: (1995), 331.

## Chapitre IV : résultats et discussions

---

- [38] Tibor Kozar, Claus-Wilhelm von der Lieth, *Glycoconjugate Journal.*, 14: (1997), 925-933.
- [39] D. Case, T. Darden, T. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Merz, D. Pearlman, M. Crowley, R. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. Mathews, C. Schafmeister, W. Ross, P. Kollman (2006) AMBER University of California.
- [40] RJ. Woods, RA. Dwek, CJ Edge, B. Fraser-Reid, Glycam Biomolecule Builder [http://www.glycam.com/CCRC/biombuilder/biomb\\_index.jsp](http://www.glycam.com/CCRC/biombuilder/biomb_index.jsp)
- [41] Chafika Gourmala, Yun Luo, Florent Barbault, Yongmin Zhang, Said Ghalem, François Maurel, BoTao Fan, *Journal of Molecular Structure: THEOCHEM.*, 821: (2007), 22-29.
- [42] JP. Ryckaert, G. Ciccotti, HJC. Berendsen, *J Comput Phys.*, 23: (1977), 327–341.
- [43] HJC. Berendsen, JPM. Postma, van WF. Gusteren, *J Chem Phys.*, 81: (1984), 3684–3690.
- [44] Eunsun Yoo Yoon, *Bull. Korean Chem. Soc.*, 24:3 (2003), 339-344.
- [45] J.A McCammon, and S.C.Harvey, “Dynamics of proteins and Nucleic acids”, *Cambridge University press, Cambridge*, (1987), 60-66.
- [46] E. B. Philip and W. Helge, *Structural Bioinformatics, Ist editions*, (2003), 509-523.
- [47] N. Khebichat, S. Ghalem, *Journal of Molecular Structure: THEOCHEM*, 777: (2006) 107–111.
- [48] N. Khebichat, K. Nekkaz, S. Ghalem, *International Journal of Carbohydrate Chemistry*

## Chapitre IV : résultats et discussions

---

Volume (2012), Article ID 725271, 7 pages doi:10.1155/2012/725271.

[491] A. A. Zamyatnin, *Neirokhimiya*, 9: (1990), 71-82.

[50] A. A. Zamyatnin, *Prot. Seq. Data Anal.*, 4: (1991), 49-52.

[51] N.Sewald, and H. D.Jakubke, "Peptides: Chemistry and Biology", *Wiley VCH Verlag GmbH*, Weinheim, (2002).

[52] P. C.Woo, A. P.To, S. K.Lau, and K. Y. Yuen, *Med. Hypotheses*, 61:(2003), 503 508.

[53] I. P. Ashmarin, and M. F. Obukhova, *Zh. Vyssh. Nervn. Deyat.*, 35: (1985), 211 221.

[54] L. Xiao, M. Jouini, B.T. Fan, J. Huet, G. Lapuye., *J. Chem. Soc. Dalton Trans.*, (1991), 1137.

[55] J. Hughes, T. Smieh, B. Morgan, L. Fothergill, *Life Sci.*, 16: (1975), 1753-1758.

[56] L. Ternius, A. Wahlström, *acta. Pharmac. Tox.*, 35 Suppl1: (1974), 55.

[57] L. Ternius, A. Wahlström, *acta. Physiol. scand.*, 94: (1975), 74-81.

[58] J. Hughes, *Brain Res.*, 88: (1975), 295-308.

[59] B. P. Roques, C. Garbay-Jaureguirberry, R. oberlin, *Nature*, 262: (1976), 778-779.

[60] J. E. Gairin, H. Mazarguil, P. Sharbrock, *Inorg. Chem.*, 21: (1982), 1846.

[61] R. Haran, P. Sharrock, J. Gairin, *Int. J. peptide Protein Res.*, 20: (1982), 380.

[62] H. Mazarguil, R. Haran, J. P .Laussac, *Biochim. Biophys. acta.*, 717: (1982), 465.

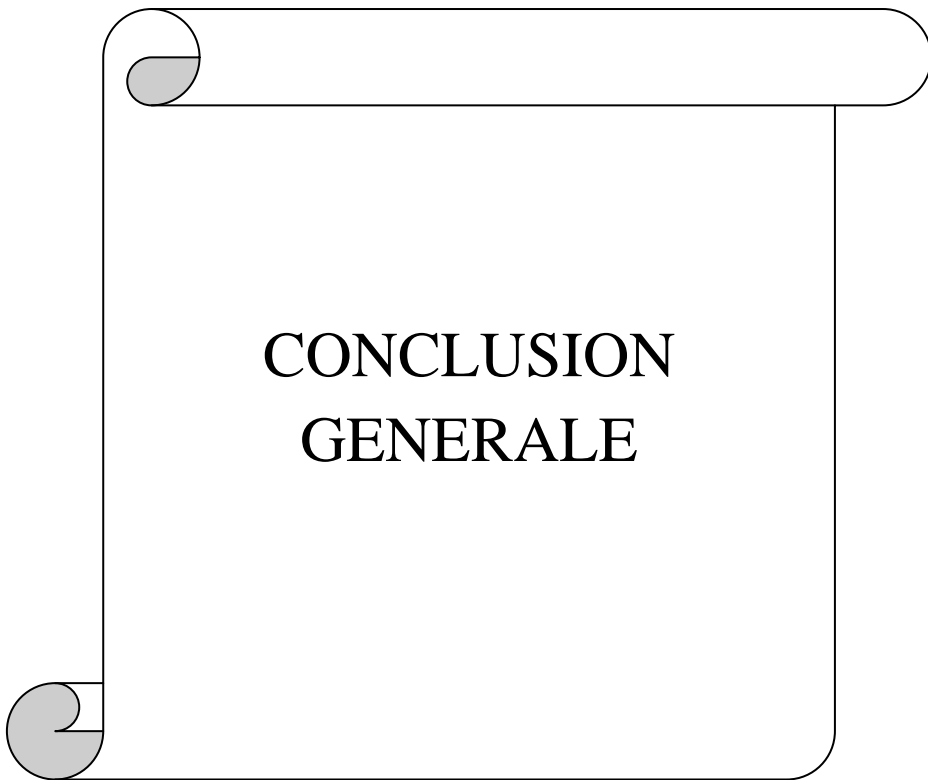
[63] T. Ishida, M. Kenmotsu, Y. mino, *Biochim. J.*, 218: (1984), 677-689.

[64] T. Darden, D. Yorkand, et L. Pedersen. 1993. Particle mesh Ewald: an N. log(N) method for Ewald sums in large systems., *J. Chem. Phys.*, (1998), 10089–10092.

[65] M. J. Frisch, et al, *Gaussian, Inc.*, Wallingford CT, (2009).







## Conclusion générale

Ce travail a été consacré à l'étude de quelques biomolécules par analyse conformationnelle. Notre intérêt a été porté sur les oligosaccharides: le tri- et le pentasaccharide Lewis X, les oligopeptides (tripeptides) : L,L-Gly-Phe-Phe, L,L-Phe-Gly-Phe, et L,L -Phe-Phe-Gly,. Ainsi pour l'étude de quelques complexes peptidiques du cuivre (II).

Notre thèse est structurée en quatre chapitres:

Le premier chapitre de cette thèse présente une description détaillée des glucides, leurs propriétés structurales et leur importance biologique. Les protéines sont également présentées, en insistant sur les différents niveaux d'énergie structurale (primaire, secondaire, tertiaire et quaternaire) et les différents types d'interactions qui sont mise en jeu pour la stabilité des édifices moléculaire, en particulier les complexes peptidiques du cuivre (II).

Dans le deuxième chapitre les techniques de modélisation moléculaire utilisées dans ce travail sont présentées et amplement décrites. À la fin de cette partie, le lecteur non spécialisé atteint une vision globale des techniques utilisées.

Dans le 3<sup>ème</sup> chapitre nous avons tenu à mettre l'accent sur le conflit existant entre la structure moléculaire et l'analyse conformationnelle des glucides. Les différents champs de force pour les carbohydrates, en particulier le champ de force Glycam06, sont défini dans ce chapitre. L'importance de paramétrisation justifie le choix du champ de force.

Le quatrième chapitre présente les résultats obtenus suite à l'application des approches computationnelles sur les systèmes biologiques. Tous les travaux sont présentés sous forme d'articles scientifiques en incluant une introduction complémentaire.

- Le premier article (publié dans *Journal of Molecular Structure: THEOCHEM*) est consacré à une étude théorique des complexes dipeptidiques de cuivre (II)

au voisinage du pH physiologique (l'espèce (II) MH<sub>1</sub>A est la plus prédominante) avec les quatre dipeptides suivant : L,L-Phe-Leu, L,L-Phe-Met, L,L-Leu\_Phe, et L,L-Met-Phe (chaque dipeptide contient une chaîne latérale aromatique et une alkyl) pour voir l'influence des chaînes latérales dans les dipeptides, et surtout ceux qui contiennent un résidu possédant un noyau aromatique. ce travail nous a permis de conclure que, lorsque le groupement aromatique est en position C-terminale par rapport au dipeptide contenant les mêmes acides aminés, mais le résidu possédant un noyau aromatique est en N-terminal donne généralement des espèces [MH-1A] plus stables. Ce phénomène est attribué à l'interaction entre les électrons  $\pi$  du cycle aromatique et les orbitales d du Cu (II), parce que la liaison Cu-O est faible et facile à déformer, cependant, la liaison rigide Cu-N est formée lorsque le cycle aromatique est dans l'extrémité N-terminal.

- Le deuxième article (publié dans *International Journal of Carbohydrate Chemistry*) montre un exemple de recherche conformationnelle sur les structures de Lewis X par dynamique moléculaire : étude du trisaccharide LewisX ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-OMe) et pentasaccharide LewisX ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-(1,3)- $\beta$ -D-Gal-(1,4)- $\beta$ -D-Glu-OMe). L'objectif de cette étude était d'explorer le comportement conformationnelle de ces oligosaccharides dans le modèle d'eau explicite à 300 K pendant 10 ns en utilisant le champ de force GLYCAM 06.

Durant la simulation les deux oligosaccharides montrent une flexibilité importante, mais le pentasaccharide est moins flexible que le trisaccharide à cause de l'encombrement stérique. Un large intervalle d'angle de torsion a été exploré. L'examen visuel des deux RMSD au

cours du temps (10 ns) montre que l'équilibration des deux systèmes est atteinte, et les deux oligosaccharides sont stables.

Le comportement conformationnelle des différents oligosaccharides montre l'existence des conformations privilégiées et des rotations chimiques entre les différents résidus.

- Une troisième étude porte sur une recherche conformationnelle sur la structure des tripeptides par dynamique moléculaire: étude de L,L-Gly-Phe-Phe (commercial), L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly (synthétisé par le Pr S. Ghalem). L'objectif de cette étude était de comprendre la stabilité structurale de ces oligopeptides contenant des chaînes latérales aromatiques non coordinantes, dans le modèle d'eau explicite à 300 K pendant 10 ns en utilisant le champ de force ff99SB.

Durant la simulation, ces trois tripeptides contenant les mêmes résidus d'acides aminés mais diffèrent par la position du résidu glycine montrent une flexibilité importante, mais L, L-Gly-Phe-Phe est moins flexible que L,L-Phe-Gly-Phe, qui est moins flexible que L,L-Phe-Phe-Gly à cause de la gêne stérique, et la distribution électronique.

Un large intervalle d'angle de torsion a été exploré. L'examen visuel des trois RMSD au cours du temps (10 ns) montre que l'équilibration des trois systèmes est atteinte, et les trois oligopeptides sont stables. La présence d'un cycle aromatique dans le résidu C-terminal donne généralement un tripeptide plus stable par rapport au tripeptide contenant les mêmes acides aminés, mais le phényle est en résidu N-terminal.

# Theoretical study of dipeptide complexes of copper(II)

N. Khebichat, S. Ghalem \*

*Abou-Bakr Belkaid University, Faculty of Sciences, Department of Chemistry, Laboratory of Organics Chemistry, Analyses and Natural Substances (COSNA), BP 119, Tlemcen 13000, Algeria*

Received 17 July 2006; accepted 1 August 2006

Available online 9 August 2006

## Abstract

The L,L-Phe-Leu-Cu(II), L,L-Leu-Phe-Cu(II), L,L-Phe-Met-Cu(II), and L,L-Met-Phe-Cu(II) systems were studied using molecular modelling. The results obtained, which are in good agreement with results obtained by potentiometric studies show a significant increase in stability of copper complexes, when an aromatic residue is located in C-terminal (compared to the L,L-dipeptides containing the same amino acid residues), this phenomenon is attributed to the interaction between the d-orbital of copper and the  $\pi$ -electrons of the aromatic ring.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Molecular modelling; Force field; Peptide; Copper(II)

## 1. Introduction

Copper is an essential element widely distributed in all the body of the mammals. Among the principal ions of transition metal in biology, Cu (II) is probably most effective in chelation with peptides at physiological pH [1]. In the body, the essence of copper is present under a non-labile form (for example: céruloplasmine). Labile copper is present in the form of complexes with amino acids or peptides [2].

Cu(II) forms very stable complexes with simple oligopeptides. The principal modes of coordination of the cupric ions with simple dipeptides are well established now [3].

At low pH, the species  $[\text{CuA}]^+$  is formed with the dipeptide reacting like a bidentate ligand (I) (see Fig. 1). Towards pH 5, peptidic hydrogens can be deprotected in the presence of Cu(II), making it possible to rearrange the donors centres to form the structure (II) (see Fig. 1), which is the  $\text{CuH-1A}$  species. In the zone of high pH ( $>9$ ), one of the molecules  $\text{H}_2\text{O}$  dissociates to form a mono-hydroxylic complex whose formula is  $[\text{CuH}_2\text{A}]^-$

or  $[\text{CuH}_1(\text{OH})\text{A}]^-$ , which has the structure (III) shown in Fig. 1. In addition to the monoligands, the structures of the complexes biligands are also well established. These complexes include  $\text{CuA}_2$  (IV) in Fig. 1.

Sigel et al. [1] indicate that there are great differences between the cupric complexes of the diglycine and the complexes with dipeptides where the residue glycyl is monosubstituted by coordinating side chains or not. A significant reduction in stability was observed in the cupric complexes with the dipeptides where the residue glycyl was substituted.

In agreement with Rabin [4], Sigel et al. conclude that stability increases with the lengthening of the side chains. This phenomenon was explained by the interactions between the non-covalent side chains [5,6]. Moreover certain authors propose the existence of interaction  $\pi$ -d between the aromatic ring of the side chains and the metal ions in certain complexes [7,8].

Potentiometric studies [11] (see Table 1) on the mixed complexes of Cu(II) in the presence of several amino acids [9,10] showed that the most stable species are formed when one of the ligands is aliphatic and the other one is aromatic (for example: phenylalanine or tyrosine) [11,12].

All the dipeptides used in this study are of L,L conformation. The equilibria in metal ion–dipeptide systems and the

\* Corresponding author.

E-mail address: s\_ghalem2002@yahoo.fr (S. Ghalem).

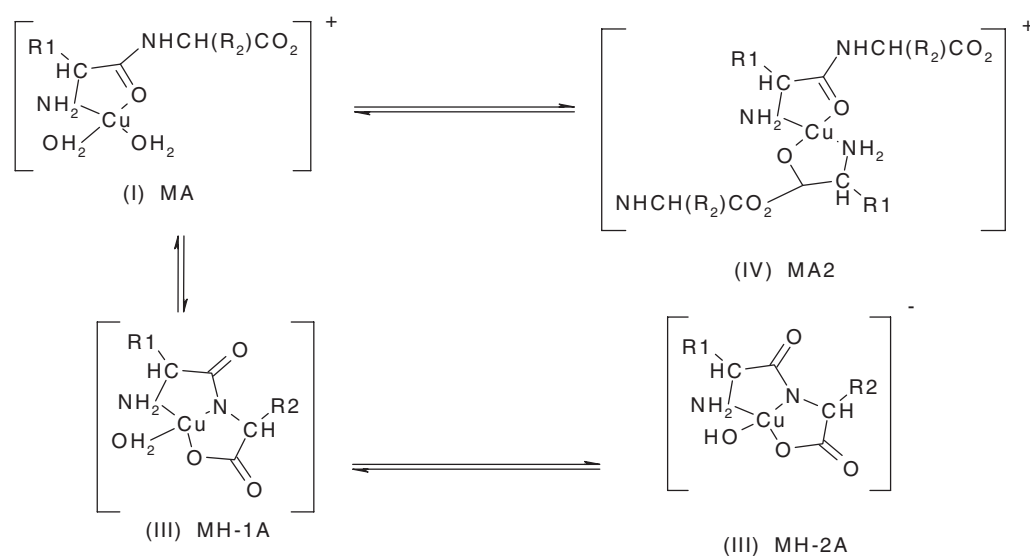


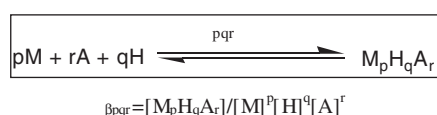
Fig. 1. Structural models of the cupric complexes with simple dipeptides.

Table 1  
Reference bond length  $l_0$  for Cu–ligand (Cambridge Structural Database 1995)

Atoms chaining	$l_0$ (angstroms Tot. Obs).	
Cu–O ( $sp^3$ ) axial	2.39 (2.38)	1117
Cu–O ( $sp^2$ ) equatorial	1.96 (1.94)	1162
Cu–O ( $sp^3$ ) equatorial	(2.03)	
Cu–O ( $sp^2$ ) axial	(2.15)	
Cu–N ( $sp^3$ ) equatorial	2.00 (2.04)	2279

Values of reference bond length in brackets are taken from literature [25].

stability constants of complexes are defined by the following equations (charges are omitted for simplicity):



where  $p$ ,  $q$ , and  $r$  are the stoichiometric numbers. A negative value of  $q$  denotes a deprotonated or hydroxylated form of the species considered. The conjugated base of the ligand is noted  $A$ .  $HA$  is the zwitterions form, and  $M$  the metal ion.

## 2. Theoretical approach

Molecular modelling is used to study the influence of the side chains in the dipeptides, and especially those which contain a residue having an aromatic ring. We were interested to study the complexation of the  $Cu^{2+}$  close to the physiological pH, where the species (II)  $MH_{-1}A$  is most predominant, with the four dipeptides according to: L,L-phenylalanyl-leucine (L,L-Phe-Leu), L,L-leucyl-phenylalanyl (L,L-Leu-Phe), L,L-phenylalanyl-methionine (L, L-

Phe-Met), and L,L-methionyl- phenylalanine (L,L-Met-Phe). Each dipeptide contains an aromatic side chain and an alkyl.

The  $MH_{-1}A$  species is studied with the program EMO (Energy of MOlecule) follow-up by semi empirical calculations by the SAM1/d method, in both two cases:

- The phenyl group is located in C-terminal.
- The phenyl group is located in N-terminal.

EMO was developed by Blaive in 1993 [13] to treat the organic molecules, it includes three menus:

- Menu1: entry of the molecule using the keyboard; each atom is codified according to its hybridization. The maximum number of atoms is 420, hydrogen included.
- Menu2: geometrical handling on the molecule.
- Menu3: minimization of energy by molecular mechanics; it is done without derivation of the function energy. The version of this program placed at our disposal is designed to function on PC [14–17].

This program uses the force field MM2, which is the force field of Allinger [18], which was conceived at the beginning for the simple molecules (alkanes, carbonyl compounds, sulphides, amines, etc.). It is used to treat increasingly complex molecules.

The retained parameters, concerning structures of our complex ligands, are those proposed by the professor Blaive completed by the statistical study done at the level of the CSD (Cambridge Structural Database 1995).

Several effects are often set forth in the interpretation of experimental results [19]: (a) side-chain donor effect; (b) hydrophobic side chain–side chain interaction; (c) steric effects between the side chains; (d) effects from the surrounding solvent sphere. Another effect was proposed by

authors for copper(II)–dipeptide, nickel(II)–dipeptide, complexes containing aromatic side-chains [20,21]. It has been shown that the presence of an aromatic residue in a C-terminal site increases the stability of complexes.

- **van der Waals radius of copper(II):** Parameters used in this program are inspired from the work of Benmenni [22] who took the linear correlation between van der Waals radius published by Allinger, and those determined by Bondi [23] established by Bouraoui [15] who calculated constants in the case of iron.  $r^*(\text{Allinger,MM2}) = 0.856r(\text{Bondi}) + 0.058$  this term of Allinger informs us that van der Waals radius used in molecular mechanics can take different values on the basis of crystallographic results. One took values of the copper(II): ( $r^*\text{Cu}^{2+} = 1.90$  and  $\epsilon\text{Cu}^{2+} = 0.06$ , calculated by Bouraoui, another value of  $r^* = 1.70$  that comes from data of the MOLCAD program developed by Brickman [24].
- **Modelling of the Cu–Ligand interactions:** For reference bond length, the Cu–N and Cu–O bonds differ according to the type of atom of oxygen or nitrogen linked to the copper. We interrogated the CSD. We did a statistical study on more than 1000 cases having the same configuration of chromophore that our complexes. We chose values by taking the average values of those met at the time of this statistical study (see Table 1).
- **Calculations of constant k:** Some measurements executed on the copper complexes, permitted to identify the frequency of the vibration  $\gamma$  (Cu–O) between 200 and 500  $\text{cm}^{-1}$ . That permitted to evaluate by calculation the force constant  $k$  by the relation:  $\gamma = 1/2\pi(k/m)^{1/2}$  and  $k = 4\pi^2 m\gamma^2$  with  $m$  (reduced mass):  $1/m = 1/m_1 + 1/m_2$ ,  $m_1$  and  $m_2$  atomic masses of the oxygen and the copper, for the vibration Cu–O, we can consider that metal is fixed by the ligand that it carries, and the reduced mass will take the value of the oxygen:

Table 2  
Formation constants of copper(II)–dipeptide complexes at 25 °C,  $I = 0.2 \text{ mol dm}^{-3} \text{ KNO}_3$  [11]

Dipeptide	$\log \beta_{1-11}$	$\Delta \log \beta_{1-11}$
L,L-Phe-Leu	$0.93 \pm 0.01$	0.96
L,L-Leu-Phe	$1.89 \pm 0.01$	
L,L-Phe-Met	$1.61 \pm 0.01$	0.09
L,L-Met-Phe	$1.70 \pm 0.01$	

$\Delta \log \beta_{\text{pqr}} = \log \beta_{\text{pqr}}(\Phi \text{ C-terminal}) - \log \beta_{\text{pqr}}(\Phi \text{ N-terminal})$ .

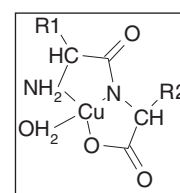
Table 3  
Energies of the complexes calculated using EMO and SAM1/d

	L,L-Leu-Phe-Cu(II)	L,L-Phe-Leu-Cu(II)	L,L-Met-Phe-Cu(II)	L,L-Phe-Met-Cu(II)
$E_{\text{EMO}}$ (kJ/mol)	66.711	67.506	60.011	63.445
$E_{\text{stretch}}$ (kJ/mol)	4.77	5.05	3.88	4.01
$E_{\text{binding}}$ (kJ/mol)	29.19	25.29	29.72	29.62
$E_{\text{torsion}}$ (kJ/mol)	–11.26	–9.78	–11.97	–10.23
$E_{\text{vdW}}$ (kJ/mol)	44.02	46.95	38.39	40.05
$E_{\text{SAM1/d(AMPAC)}}$ (eV)	–5153.092	–5153.081	–5191.404	–5191.336

$1/m = 1/m^1 = 1/16$ . After having minimized the energy of the various peptidic complexes of Cu(II), an optimisation of energy is carried out by the SAM1/d method. SAM1/d is a semiempirical method which takes in consideration d orbitals of transition metals.

### 3. Results

The study of the  $\text{MH}_{-1}\text{A}$  species is interesting because it is the majority complex of the Cu(II)–dipeptide systems between pH 4.5 and 8 [11] i.e. close to the physiological pH. All the dipeptides used in this study are of L,L conformation.



Structure of  $\text{CuH}_{-1}\text{A}$

All the structures studied in our work were built using the program EMO (version 2001) by introducing the Allinger code of the atoms of the molecules studied, by the keyboard of the computer then the energy is minimized by using the semi-empirical parameters. The most stable conformation is obtained from various starting geometries, after optimisation.

In order to avoid the local minima corresponding to unstable conformers, we carried out it with the option “SCAN” which makes it possible to sweep the surface of potential energy (PES). This enabled us to eliminate the geometries having little chance to generate the most stable conformers. Energies of the found conformers are optimised by the semi empirical method SAM1/d. The most stable conformations have the lower energies. For each structure we done the energies:  $E_{\text{EMO}}$ ,  $E_{\text{stretching}}$ ,  $E_{\text{bending}}$ ,  $E_{\text{torsion}}$ ,  $E_{\text{Van Der Waals}}$ , and  $E_{\text{SAM1/d}}$ .

The results (energies) obtained are gathered in Table 3.

### 4. Discussion

The formation constants of different species can be determined by the potentiometric measurements [11], see Table 2.



Comparing each pair of ligands, where two amino residues are the same, but differing in sequence, we remark a significant increase in stability of complexes when the C-terminal residue contains an aromatic ring. The differences of formation constants are gathered in Table 2 ( $\log \beta_{L,L\text{-Leu-Phe}} > \log \beta_{L,L\text{-Phe-Leu}}$  and  $\log \beta_{L,L\text{-Met-Phe}} > \log \beta_{L,L\text{-Phe-Met}}$ ).

It is known that the influences of non-coordinating side chains are different according to the species. If we examine the structure of the  $\text{CuH}_{-1}\text{A}$  species determined by e.s.r. measurements [26,27], we see that the two side chains are both involved in the stability of complex. Generally for L,L-dipeptides containing two large non-covalent side chain groups, the stability of  $\text{MH}_{-1}\text{A}$  species is greater than of the corresponding L,D- or D,L dipeptides due to a hydrophobic interaction between two side chain groups which are located on the same side [5]. However, this hydrophobic interaction is less evident when the dipeptides contain a benzyl group, because the interaction between the alkyl group and the aromatic ring is less effective [5]. This increase in stability for the L,L-Leu-Phe and L,L-Met-Phe can be proposed as the result of a  $\pi$ -d interaction between the aromatic ring and the copper ion [28–30].

For every pair of dipeptides which contain the same amino acid residues, we could make the following comparison [12]:

- (1) *Electronic effect.* For  $\text{MH}_{-1}\text{A}$  species, this effect should be similar for the two dipeptides, because it is only dependent upon the distance. Referring the structural representation, R1 and R2 have almost the same distance from the center metal ion.
- (2) *Steric effect.* This effect destabilizes the complex. The tridimensional model shows that the R2 residue sets above the metal ion. In the case of L,L-Leu-Phe and L,L-Phe-Leu, the two side chain  $(\text{CH}_3)_2\text{CHCH}_2$  and  $\text{C}_6\text{H}_5\text{CH}_2$  are different. But their steric effect may be comparable [5].
- (3) *Hydrophobic interaction between two side chains.* The structure established for the  $\text{MH}_{-1}\text{A}$  species according to the e.s.r. studies shows a  $\text{CuNO}_2$  in-plane chromophore, which requires two side chains located on the same side, with respect to this basal plane when L,L-dipeptides are considered. Such an arrangement results in an interaction giving rise to hydrophobic micelles, thus decreasing the solvation enthalpy.

The above discussion excludes the possibility that the significant increase in stability, for the dipeptides containing an aromatic residue in C-terminal with respect to the N-terminal location, is the result of the other different effects than those enumerated in (1) to (3). This increase is then very probably due to  $\pi$ -d interaction.

In the light of the results obtained by calculation, we notice that the order of stability of the studied systems is as follows: L,L-Met-Phe-Cu(II) > L,L-Phe-Met-Cu(II) and L,L-Leu-Phe-Cu(II) > L,L-Phe-Leu-Cu(II).

We notice a significant increase in the stability of the cupric complexes, when the aromatic group is in C-terminal position (L,L-Leu-Phe and L,L-Met-Phe) compared to the dipeptide containing the same amino acids, but the residue phenyl is into N-terminal (L,L-Phe-Leu and L,L-Phe-Met), ( $E_{L,L\text{-Leu-Phe}} < E_{L,L\text{-Phe-Leu}}$  and  $E_{L,L\text{-Met-Phe}} < E_{L,L\text{-Phe-Met}}$ ). This phenomenon is attributed to the interaction between the  $\pi$  electrons of the aromatic ring and the d orbitals of the Cu(II). Results obtained by the calculation are in agreement with those obtained by the potentiometric study ( $\log \beta_{L,L\text{-Leu-Phe}} > \log \beta_{L,L\text{-Phe-Leu}}$  and  $\log \beta_{L,L\text{-Met-Phe}} > \log \beta_{L,L\text{-Phe-Met}}$ ).

## 5. Conclusion

The presence of an aromatic ring in the C-terminal residue gives generally a more stable  $[\text{MH}_{-1}\text{A}]$  species. This stabilising effect was attributed to an interaction between the metal ion and the aromatic ring. This interaction is found only when the aromatic ring is located in the C-terminal residue, because the Cu—O bond is weak and easily distorted. However, the rigid Cu—N bond formed when the aromatic ring is in the N-terminal residue does not favour this interaction [11].

## Acknowledgement

The authors thank particularly Dr. B. Blaive Professor at the University of Marseille, France for sending the molecular mechanics program EMO.

## References

- [1] H. Sigel, R.B. Martin, Chem. Rev. 82 (1982) 385–426.
- [2] J.W.H. Robert, L.D. Pettit, J.C.S. Dalton. (1981) 13331.
- [3] A. Kaneda, A.E. Martell, J. Am. Chem. Soc. 99 (1977) 1586.
- [4] B.R. Rabin, Trans. Faraday. Soc. 52 (1956) 1130.
- [5] R.P. Bnomo, R. Cali, V. cucinotta, Inorg. Chem. 25 (1986) 1641–1646.
- [6] T. Kiss, Z. Szucs, J. Chem. Soc. Dalton Trans. (1986) 2443.
- [7] R.M. Izatt, J.W. Wrathall, K.P. Anderson, J. Phys. Chem. 65 (1961) 1914.
- [8] A. Wahlborg, E. Friedon, Arch. Biochem. Biophys. 111 (1965) 672.
- [9] A. Gergerly, I. Savago, I. Nagypal, R. Kiraly, Inorg. Chim. Acta 6 (1972) 435.
- [10] A. Gergerly, I. Savago, J. Inorg. Nucl. Chem. 35 (1973) 4355.
- [11] L. Xiao, M. Jouini, B.T. Fan, G. Lapluye, J. Huet, J. Chem. Soc. Dalton Trans. 286 (1989) 1–10.
- [12] S. Ghalem, B.T. Fan, L. Xiao, G. lapluye, J. Chim. Phys. 91 (1994) 113–125.
- [13] S. Ghalem, Thèse de Doctorat, Université Denis Diderot- Paris VII, France (1996).
- [14] A. Bouraoui, M. Fathallah, B. Blaive, R. Gallo, J. Chem. Soc. Perkin Trans. 2 (1990) 1211–1214.
- [15] A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive, R. Gallo, in J.L. Rivail (Eds.), Proceedings of an International Meeting, Studies in Physical and Theoretical Chemistry, vol. 71 Nancy, France, 11–15, Septembre, 1989, pp. 381–393.
- [16] A. Zinelabidine, A. Bouraoui, M. Fathallah, F. M'Henni, B. Blaive, R. Gallo, J. Mol. Struct. Theochem. 286 (1993) 267–274.
- [17] B. Blaive, G. Legsaï, R. Lai, J. Mol. Struct. 354 (1995) 245–250.

- [18] Z. Li, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 84 (1987) 6611.
- [19] L.D. Petit, R.J.W. Hefford, in: H. Sigel (Ed.), Metal Ions in Biological Systems, vol. 9, 1979, p.173.
- [20] R.P. Bonomo, R. Cali, V.G.I. Cucinotti, E. Rizzarelli, Inorg.Chem. 25 (1986) 1641.
- [21] S.H. Kim, R.B. Martin, J.Am.Chem. Soc. 106 (1984) 1707.
- [22] L. Benmenni, Thèse de doctorat en sciences, Fac. de Sciences de St-Jérôme, Université d'Aix-Marseille III (1994).
- [23] A. Bondi, J. Phys. Chem. 68 (1964) 441.
- [24] J. Brickmann, MOLCAD Molecular Modelling Package, Institut für physikalische Chemie, Technische Hochschule, 6100 Darmstad, Germany (1989).
- [25] L. Benmenni, E.H. Allilou, M. Giorgi, M. Pierrot, M. Regulier, J. Chem. Crystallogr. 24 (1994) 5.
- [26] J. Peisach, W.E. Blumberg, Arch. Biochem. Biophys. 165 (1974) 691.
- [27] D.M. Dooley, J. Rawlings, J.H. Dawson, P.J. Stephenes, L.E. Andresson, B.G. Malmstrom, H.B. Grey, J. Am. Chem. Soc. 101 (1979) 5038.
- [28] O. Yamauchi, K. Tsujide, A. Odani, J. Am. Chem. Soc. 107 (1985) 659.
- [29] W.L. Kwik, K.P. Ang, G.I. Chen, J. Inorg. Nucl. Chem. 42 (1980) 303.
- [30] P.I. Vestues, R.B. Martin, J. Am. Chem. Soc. 102 (1980) 7906.

## Research Article

# Conformational Search on the Lewis X Structure by Molecular Dynamic: Study of Tri- and Pentasaccharide

N. Khebichat, K. Nekkaz, and S. Ghalem

Laboratory of Natural Products and Bio Actives, Department of Chemistry, Faculty of Sciences, Aboubakr Belkaid University, Tlemcen 13000, Algeria

Correspondence should be addressed to S. Ghalem, s\_ghalem2002@yahoo.fr

Received 4 July 2011; Revised 7 October 2011; Accepted 12 October 2011

Academic Editor: Yuriy A. Knirel

Copyright © 2012 N. Khebichat et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Carbohydrates play vital roles in many biological processes, such as recognition, adhesion, and signalling between cells. The Lewis X determinant is a trisaccharide fragment implicated as a specific differentiation antigen, tumor antigen, and key component of the ligand for the endothelial leukocyte adhesion molecule, so it is necessary or essential to determine and to know their conformational and structural properties. In this work, conformational analysis was performed using molecular dynamics (MD) simulation with the AMBER10 program package in order to study the dynamic behavior of the Lewis X trisaccharide ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-OMe) and the Lewis X pentasaccharide ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-(1,3)- $\beta$ -D-Gal-(1,4)- $\beta$ -D-Glu-OMe) in explicit water model at 300 K for 10 ns using the GLYCAM 06 force field.

## 1. Introduction

In nature, carbohydrates form an important family of biomolecules, as simple or complex carbohydrates, either alone or covalently linked to proteins or lipids [1]. They play vital roles in many biological processes, such as recognition, adhesion, and communication between cells. Carbohydrates that are covalently linked to a nonsugar moiety (proteins, peptides, or lipids) are the most prominent cell-surface-exposed structures. The highly diverse structural variability of carbohydrates makes them good candidates for cell receptors and recognition molecules [2].

Determination of the three-dimensional structure of oligosaccharides and understanding the molecular basis of their recognition by receptors represent the main challenges of structural glycobiology. Elucidation of the three-dimensional structures and the dynamical properties of oligosaccharides is a prerequisite for a better understanding of the biochemistry of recognition processes and for the rational design of carbohydrate-derived drugs [1].

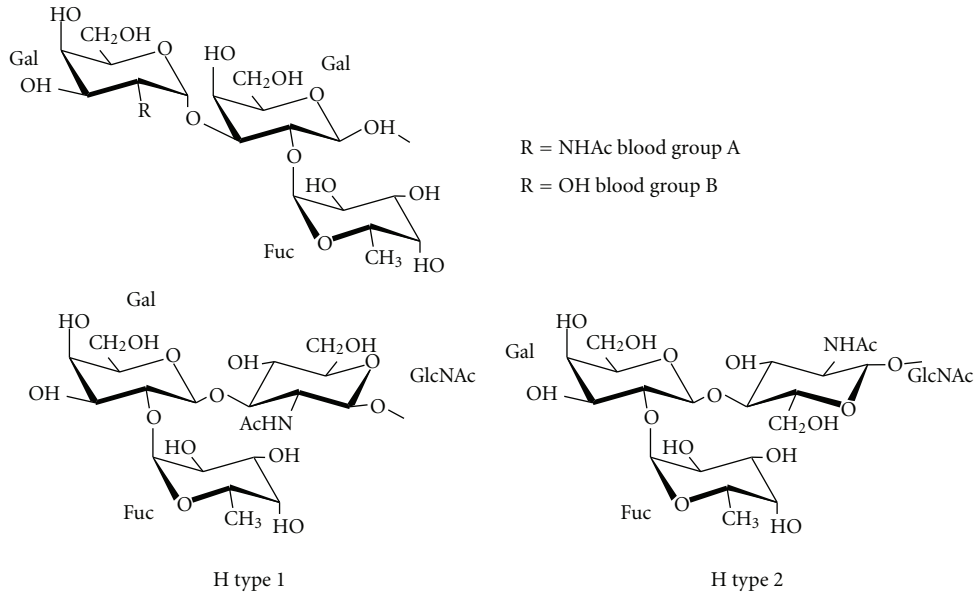
The human histo-blood group ABH (O) systems were the first major human alloantigens to be identified. The carbohydrate nature of the A, B, H, and Lewis antigens was

established in 1950s [3, 4]. Figure 1 shows a schematic representation of current histo-blood group antigens although additional rare types also exist [5]. The oligosaccharidic epitope of ABH (O) and Lewis histo-blood groups have been the subject of many structural investigations. The energy maps of all of the disaccharides have been established by molecular mechanics methods [6, 7].

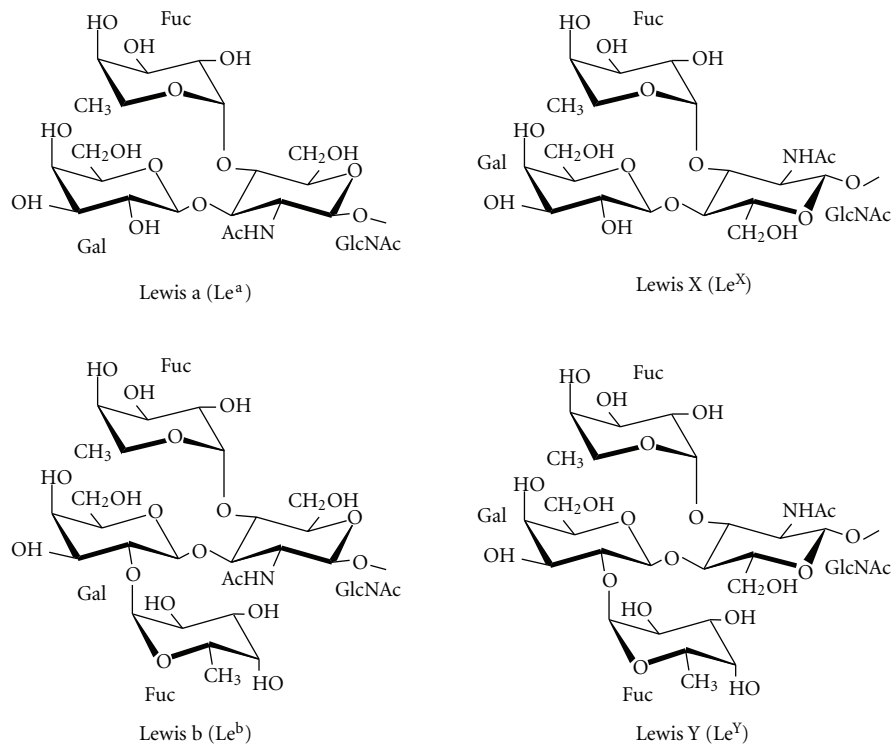
The determination of conformational preferences of oligosaccharides is best approached by describing their preferred conformations on potential energy surfaces as a function of the glycosidic linkage  $\varphi$ ,  $\psi$  torsional angles. In the early days of conformational analysis of oligosaccharides, the calculations of either rigid or, later, relaxed conformational maps predominated. Today, complex simulation techniques such as molecular dynamics are increasingly being used. Several examples can be found in the literature of “state-of-the-art” conformational search strategies in order to simplify the complexity of the multiple-minima problem [8].

In this work, we carried out extensive, explicit solvent MD simulation studies for tri- and pentasaccharide Lewis X for 10 ns. Each monosaccharide unit in the Lewis X molecules adopts the typical  ${}^4C_1$  chair conformation, with no significant deviation from the classical pyranose ring shape.

## ABH(O) antigens



## Lewis antigens



(a)

FIGURE 1: Continued.

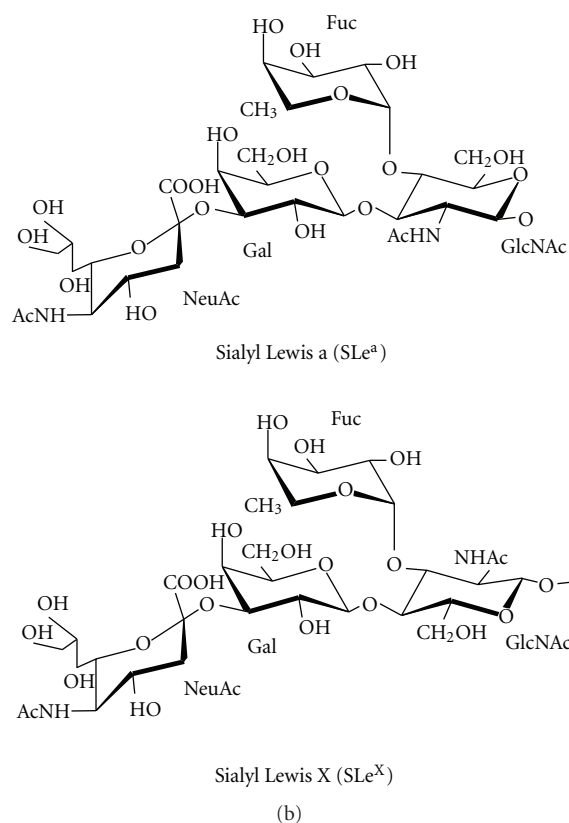


FIGURE 1: Schematic representation of histo-blood group oligosaccharides.

The objective of this study was to explore the conformational behaviour of those oligosaccharides in explicit water model at 300 K for 10 ns using the GLYCAM 06 force field.

## 2. Theoretical Approach

The atomic labelling and the torsion angles of interest are given in Figure 2. The torsion angles  $\varphi$ ,  $\psi$  are defined  $\varphi = \vartheta$  (O5-C1-O'2-C'2),  $\psi = \vartheta$  (C1-O'2-C'2-C'3), and  $\omega = \vartheta$  (O5-C5-C6-O6). The conformations with specific torsion angles  $\omega$  are referred to as gauche-trans (gt;  $\omega = 60^\circ$ ), gauche-gauche (gg;  $\omega = -60^\circ$ ) and trans-gauche (tg;  $\omega = 180^\circ$ ).

The MD simulations were carried out using AMBER10 [9] at a temperature of 300 K. This temperature was chosen because our primary objective was to study the dynamics of the oligosaccharides at physiologically relevant conditions. The starting Lewis X pentasaccharide and trisaccharide monomer coordinates and parameters were constructed with the "Glycam Biomolecule Builder" available online from the website of Woods group [10] (see Schemes 1 and 2).

Since these coordinates are not optimized, a first minimization by molecular mechanics (1,000 steps of steepest descent followed by 1,000 steps of conjugate gradient minimization with a gradient tolerance of  $0.05 \text{ kcal mol}^{-1}$ ) was made with AMBER package version 10 [9]. Force field GLYCAM 06 was used for oligosaccharides (it used to simulate the interatomic interactions responsible for oligosaccharide

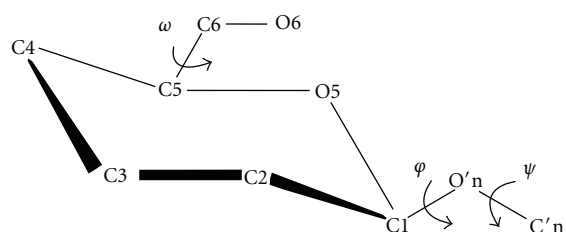


FIGURE 2: Atomic labeling and torsion angles of interest.

and glycoprotein dynamics and is currently suitable for all biologically relevant N- and O-linkages in oligo- and polysaccharides and glycoproteins [11]). The simulations were carried out using an explicit solvent environment with AMBER package version 10. The solvated system for the oligosaccharides molecules was prepared using an explicit water box of  $8 \text{ \AA}$  in the xleap module of AMBER10. The number of water molecules was 569 for the trisaccharide and 1072 for the pentasaccharide. All simulations were run with the SANDER module of AMBER with SHAKE algorithm [12] (tolerance =  $0.0005 \text{ \AA}$ ) to constrain covalent bonds involving hydrogens, using periodic boundary conditions, a 2 fs time step, a temperature coupling [13], a  $8 \text{ \AA}$  cutoff was applied to the Lennard-Jones interaction, and constant pressure of 1 atm. The nonbonded list was updated every

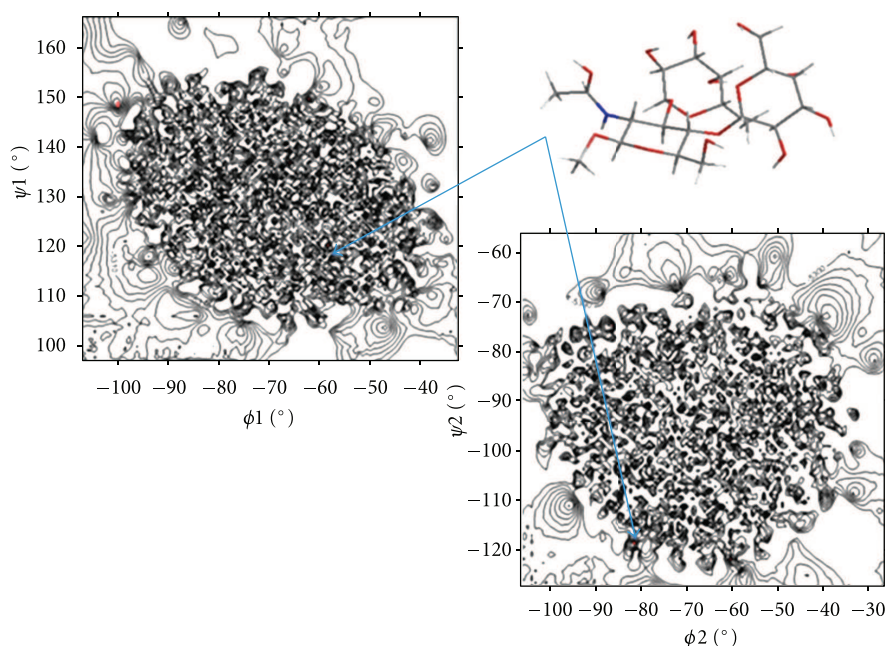


FIGURE 3: Relaxed conformational energy maps for the trisaccharide Lewis X calculated using the GLYCAM 06 force field.

10 steps. This preliminary study was aimed at finding the most energetically favorable structure which was found at 4856.9 ps for the trisaccharide and at 2411.8 ps for the pentasaccharide. Those conformations were found with a “home-made” Perl script that merely checked all energies values, finding the lowest one and giving the associated structure [2].

The PTRAJ module of AMBER10 was used to analyze and process the trajectory and coordinate files from SANDER, where the analyses might include carrying out superimposition, clustering analysis of hydrogen bonds, calculating fluctuations in bonds, angles, or dihedrals, correlation functions, and so forth, and the root mean-square deviations (RMSDs).

Finally, isoenergy contour maps were generated with ORIGIN and SURFER 9 (Golden Software, Inc., Golden Co) [14]. The positions of the minima in these maps were subsequently calculated by minimizing the energy without restraints of those grid points indicating minima on the maps.

### 3. Results

The calculated adiabatic conformational energy surfaces ( $\phi$ ,  $\psi$ ) for the tri- and pentasaccharide Lewis X are presented as potential energy contour maps in Figures 3 and 4, respectively. The contours are separated by 1 kcal/mol and range from 1 to 10 kcal/mol. The energies are given relative to the lowest minimum. Geometries and relative potential energies of the predicted minima are listed in Tables 1 and 2, and the lowest energy conformers for each compound are shown in Figures 3 and 4.

TABLE 1: Potential Energy, torsional angles for the conformers with low energies of the trisaccharide Lewis X.

	PE (kcal/mol)	$\phi_1$ ( $^\circ$ )	$\psi_1$ ( $^\circ$ )	$\phi_2$ ( $^\circ$ )	$\psi_2$ ( $^\circ$ )
Conformer 1	-5446.95	-61.84	124.7	-61.54	-104.33
Conformer 2	-5442.50	-84.91	155.63	-69.95	-92.95
Conformer 3	-5440.08	-80.60	139.25	-55.46	-98.51
Conformer 4	-5437.12	-56.58	130.28	-59.61	-80.28

**3.1. System Root-Mean-Square Displacement.** RMSD is an important parameter for describing the system stability. Figures 5 and 6 show the function curves between mass weighted RMSD fit and the time of the tri- and pentasaccharide Lewis X during dynamics simulation process, respectively.

**3.2. Computation of Phi/Psi for Modeled Oligosaccharide.** In the present study, structure prediction of oligosaccharides has been made through the torsion angles phi ( $\phi$ )/psi ( $\psi$ ). The combination of  $\phi$  and  $\psi$  angles fully determine the conformation of an oligosaccharide. Figures 3 and 4 show the distribution of sterically allowed and energy minimized conformations in  $\phi/\psi$  glycosidic space obtained from our explicit solvent MD simulations for 10 ns in the case of the tri- and pentasaccharide Lewis X, respectively. These figures also show the best conformers (which have the low energies). These conformers and their corresponding  $\phi/\psi$  angular values are reported in Tables 1 and 2 for the tri- and pentasaccharide Lewis X, respectively.

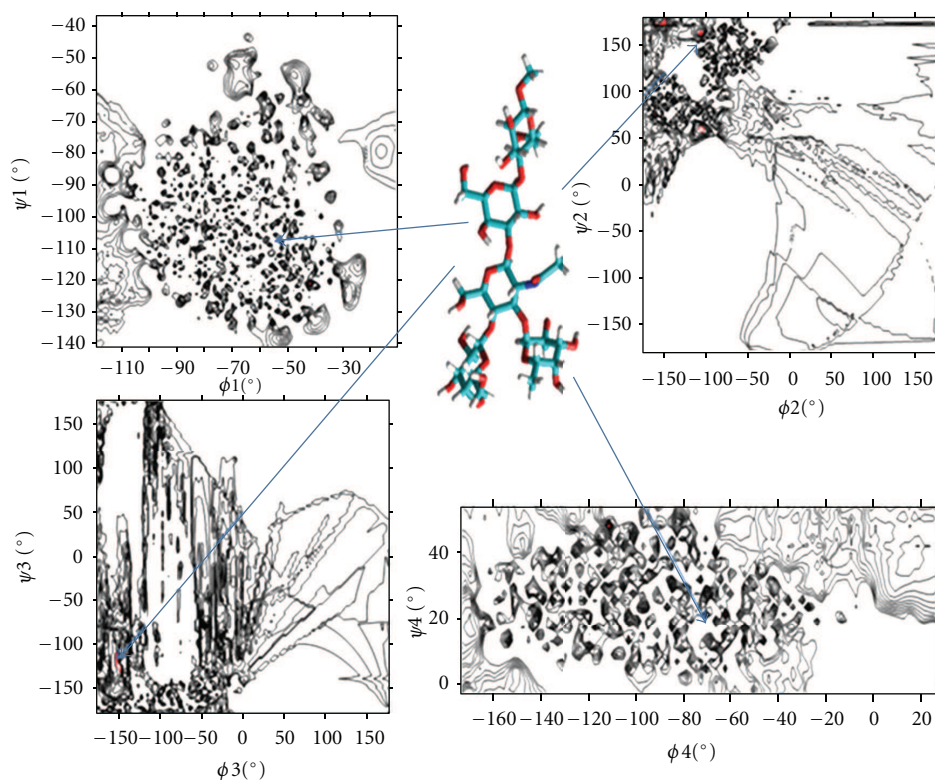


FIGURE 4: Relaxed conformational energy maps for the pentasaccharide Lewis X calculated using the GLYCAM 06 force field.

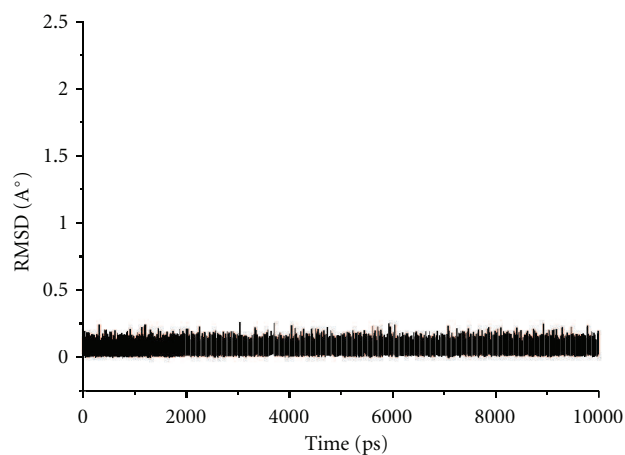


FIGURE 5: RMSD of the trisaccharide Lewis X as a function of time during MD, compared to the lowest-energy conformation.

TABLE 2: Potential energy, torsional angles for the conformers with low energies of the pentasaccharide Lewis X.

	PE (kcal/mol)	$\varphi_1$ (°)	$\psi_1$ (°)	$\varphi_2$ (°)	$\psi_2$ (°)	$\varphi_3$ (°)	$\psi_4$ (°)	$\varphi_4$ (°)	$\psi_4$ (°)
Conformer 1	-10191.15	-78.30	-97.88	-124.4	64.97	-108.27	-173.37	-57.54	15.18
Conformer 2	-10162.71	-65.65	-106.09	-73.57	136.95	-52.97	-155.29	-71.26	22.90
Conformer 3	-10162.25	-56.49	-112.26	-78.58	144.67	-55.95	-157.92	-69.85	23.71
Conformer 4	-10152.28	-71.67	-108.12	-75.73	152.13	-73.72	-163.60	-74.94	18.04

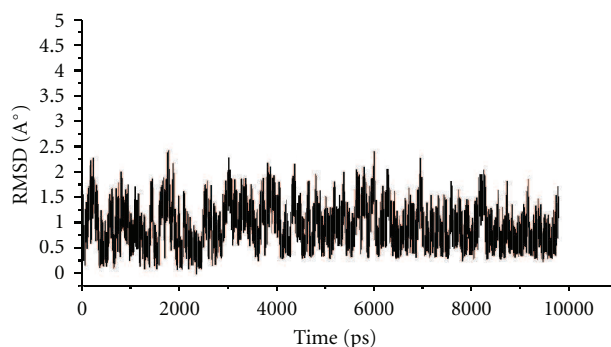
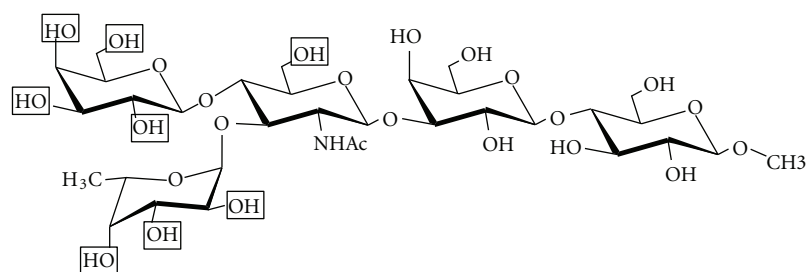
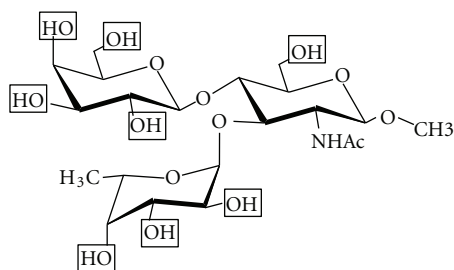


FIGURE 6: RMSD of the pentasaccharide Lewis X as a function of time during MD, compared to the lowest-energy conformation.



Lewis X pentasaccharide

SCHEME 1: Lewis X trisaccharide.



Lewis X trisaccharide

SCHEME 2: Lewis X pentasaccharide.

#### 4. Discussion

The monitoring of kinetic, global, and potential energies along the trajectory, as well as the pressure and temperature, indicate that the global periodic system is stable and does not present simulation artefacts. Kinetic energy increased slowly, whereas potential energy decreased gradually during the simulations. The total energy of a system to be simulated should be conserved, whereas the fluctuation of a temperature is caused by the interconversion of the kinetic and potential energy components [15].

The RMS variations as a function of time are displayed on Figures 5 and 6 for the tri- and penta-saccharide Lewis X

respectively, and indicate that both systems are equilibrated after 500 ps. The visual examination of the two 10 ns trajectories shows that both oligosaccharides are stable. An RMSD analysis was carried out to quantify the structural differences. The RMSD between the PDB starting structure of the simulations (obtained from Glycam biomolecule builder), and the low-energy conformers structure considering only the heavy atoms is 0.03 Å and 0.20 Å for the tri- and penta-saccharide respectively. This shows that both oligosaccharides are stable.

During the simulation of those oligosaccharides, we notice that for the pentasaccharide the most stable conformer is obtained at 2411.8 ps and that for the trisaccharide was obtained at 4856.9 ps. This means that the pentasaccharide is less flexible than the trisaccharide due to the steric effect.

#### 5. Conclusion

A complete understanding of the role of carbohydrates in biological systems is to a large extent dependent on the information available about the equilibrium mixture and about the preferred conformation of the carbohydrate molecules in solution. The conformational analysis offers a tool which can determine all possible conformations which influence the solution behavior of carbohydrates. This work represents a theoretical study of two oligosaccharides tri- and pentasaccharide Lewis X using molecular dynamic simulations in explicit water with the AMBER10 program



package. RMSD measures of the accuracy of carbohydrates around their average conformations play an important indicator of many biological processes such as macromolecular recognition and complex formations [16].

## Acknowledgments

The authors are grateful to Dr. R. Borsali, Dr. K. Mazeau, Dr. A. Rivet, and Dr. A. Nurisso (CERMAV) France, and they thank particularly Mr. Guendouzi A for his help.

## References

- [1] A. Imberty and S. Perez, "Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations," *Chemical Reviews*, vol. 100, no. 12, pp. 4567–4588, 2000.
- [2] Y. Luo, F. Barbault, C. Gourmala et al., "Cellular interaction through LewisX cluster: theoretical studies," *Journal of Molecular Modeling*, vol. 14, no. 10, pp. 901–910, 2008.
- [3] E. A. Kabat, "Immunochemical studies on the carbohydrate moiety of water-soluble blood group A, B, H, Le<sup>a</sup>, and Le<sup>b</sup> substances and their precursor I antigens," in *Carbohydrates in Solution*, H. Isbell, Ed., vol. 117, pp. 334–355, American Chemical Society, Washington, DC, USA, 1973.
- [4] W. M. Watkins, "Biochemistry and genetics of the ABO, Lewis, and P blood group systems," in *Advances in Human Genetics*, H. Harris and K. Hirsch-Horn, Eds., vol. 10, pp. 1–136, Plenum Press, New York, NY, USA, 10th edition, 1980.
- [5] S. I. Hakomori, "Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer," *Biochimica et Biophysica Acta*, vol. 1473, no. 1, pp. 247–266, 1999.
- [6] F. Bizik and I. Tvaroska, "Conformational analysis of disaccharide fragments of blood group determinants in solution by molecular modelling," *Chemical Papers*, vol. 49, no. 4, pp. 202–214, 1995.
- [7] A. Imberty, E. Mikros, J. Koca, R. Mollicone, R. Oriol, and S. Perez, "Computer simulation of histo-blood group oligosaccharides: energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens," *Glycoconjugate Journal*, vol. 12, no. 3, pp. 331–349, 1995.
- [8] T. Kozar and C.-W. von der Lieth, "Efficient modelling protocols for oligosaccharides: from vacuum to solvent," *Glycoconjugate Journal*, vol. 14, no. 8, pp. 925–933, 1997.
- [9] D. Case, T. Darden, T. Cheatham et al., "AMBER," University of California, CA, USA, 10th edition, 2006.
- [10] R. J. Woods, R. A. Dwek, C. J. Edge, and B. Fraser-Reid, "Glycam biomolecule builder," [http://www.glycam.com/CCRC/biombuilder/biomb\\_index.jsp](http://www.glycam.com/CCRC/biombuilder/biomb_index.jsp).
- [11] C. Gourmala, Y. Luo, F. Barbault et al., "Elucidation of the LewisX-LewisX carbohydrate interaction with molecular dynamics simulations: a glycosynapse model," *Journal of Molecular Structure*, vol. 821, no. 1–3, pp. 22–29, 2007.
- [12] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.
- [13] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, article 3684, 7 pages, 1984.
- [14] E. Y. Yoon, "A conformational study of linkage positions in oligosaccharides investigated by 2-D NMR spectroscopy and molecular modeling," *Bulletin of the Korean Chemical Society*, vol. 24, no. 3, pp. 339–344, 2003.
- [15] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1987.
- [16] E. B. Philip and W. Helge, *Structural Bioinformatics*, 1st edition, 2003.

**ملخص:** في هذه الأطروحة، تم استخدام الطرق التقليدية في النمذجة الجزيئية (الميكانيك وديناميك الجزيئية) لتوضيح الخصائص الهيكلية وديناميكية للجزيئات الحيوية: الكربوهيدرات (ثلاثي- و خماسي سكاريد لويس X )، أوليجوبيبتيد (ثلاثي بيببتيد)، و مركبات ثنائية نحاس (II). وتمت مناقشة موضوعين رئيسيين:

أ) تمت دراسة المركبات (L,L-Phe-Leu-Cu (II)، L,L-Leu-Phe-Cu (II)، L,L-Phe-Met-Cu (II)، و L,L-ehP-Met-Cu (II) -Cu (II) L,L-Met-Phe. النتائج المحصل عليها تتفق تماما مع نتائج الدراسة البتونسيومترية و التي تظهر زيادة كبيرة في استقرار هذه المركبات النحاسية عندما يكون Phe من جهة C-terminal بالنسبة الى ثنائي بيببتيد الذي يحتوي على نفس الأحماض الأمينية و لكن Phe يكون من جهة N-terminal، وهذا يرجع إلى التفاعل بين مدار d للنحاس (II) و الإلكترونات  $\pi$  ل Phe.

ب) تم إجراء تحليل بنيوي باستخدام الديناميك الجزيئية مع برنامج AMBER10 لدراسة السلوك الديناميكي لثلاثي السكاريد لويس X و خماسي السكاريد لويس X لمدة 10 نانوثانية على درجة الحرارة 300 K باستخدام مجال القوة GLYCAM06. تم استخدام نفس الإجراء للتحليل البنيوي لثلاثة ثلاثي بيببتيد (L,L-Phe-Phe-Gly، L,L-Phe-Gly-Phe، و L,L-Phe-Phe-Phe) باستخدام مجال القوة FF99SB لمدة 10 نانوثانية على درجة الحرارة 300 K باستخدام مجال القوة FF99SB.  
**الكلمات المفتاحية:** تحليل بنيوي، أوليجوساكاريد، لويس X، أوليجوبيبتيد، النمذجة الجزيئية

**Abstract:** In this thesis, classical molecular modeling methodologies (molecular mechanics and dynamics) have been used to elucidate structural and dynamic features of biomolecules: carbohydrates (Lewis X tri- and pentasaccharide), oligopeptides (tripeptides), and dipeptide complexes of copper (II). Two main topics were discussed:

i) Systems L,L-Phe-Leu-Cu (II), L,L-Leu-Phe-Cu (II), L,L-Phe-Met-Cu (II), and L,L-Met-Phe-Cu (II) were studied by molecular modeling. Results obtained by the calculation, which are in agreement with those obtained by the potentiometric study, show a significant increase in stability of complexes of copper, when an aromatic residue is located at the C-terminal (relative to L,L-dipeptides containing the same amino acid residues), this is due to the interaction between the d orbital of copper and  $\pi$  electrons of the aromatic ring.

ii) The conformational analysis was performed using the molecular dynamics (MD) simulation with the AMBER10 program to study the dynamic behavior of the Lewis X trisaccharide ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-OMe), and Lewis X pentasaccharide ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-(1,3)- $\beta$ -D-Gal-(1,4)- $\beta$ -D-Glu-OMe) in the explicit water model at 300 K for 10 ns using the GLYCAM06 force field. The same procedure is used for the conformational analysis of tripeptides: L,L-Gly-Phe-Phe (commercial), L,L-Phe-Gly-Phe, and L, L-Phe-Phe-Gly (synthesized by the Prof. S. Ghalem) using the force field FF99SB.

**Key words:** conformational analysis, oligosaccharide, Lewis X, oligopeptide, molecular modelling.

**Résumé:** Dans cette thèse, des méthodes classiques de modélisation moléculaire (mécanique et dynamique moléculaire) ont été utilisées pour élucider les caractéristiques structurales et dynamiques des biomolécules : des glucides (tri- et pentasaccharide Lewis X), des oligopeptides (des tripeptides) et des complexes di-peptidique de cuivre (II). Deux sujets principaux ont été abordés:

i) Les systèmes L,L-Phe-Leu-Cu (II), L,L-Leu-Phe-Cu (II), L,L-Phe-Met-Cu (II), et L,L-Met-Phe-Cu (II) ont été étudiés par modélisation moléculaire. Les résultats obtenus, qui sont en bon accord avec les résultats obtenus par les études potentiométriques montrent une augmentation significative de la stabilité des complexes de cuivre, quand un résidu aromatique est située en position C-terminal (par rapport au L,L-dipeptides contenant les mêmes résidus d'acides aminés), ce phénomène est attribuable à l'interaction entre l'orbitale d du cuivre et les électrons  $\pi$  du noyau aromatique.

ii) L'analyse conformationnelle a été réalisée en utilisant la simulation de dynamique moléculaire (DM) avec le programme AMBER10 afin d'étudier le comportement dynamique du trisaccharide Lewis X ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-OMe), et du pentasaccharide Lewis X ( $\beta$ -D-Gal-(1,4)-[ $\alpha$ -L-Fuc-(1,3)]- $\beta$ -D-GlcNAc-(1,3)- $\beta$ -D-Gal-(1,4)- $\beta$ -D-Glu-OMe) dans le modèle d'eau explicite à 300 K pendant 10 ns en utilisant le champ de force GLYCAM 06. La même procédure est utilisée pour l'analyse conformationnelle des tripeptides : L,L-Gly-Phe-Phe (commercial), L,L-Phe-Gly-Phe, et L,L-Phe-Phe-Gly (synthétisé par le Pr S. Ghalem) en utilisant le champ de force FF99SB.

**Mots-clés:** analyse conformationnelle, oligosaccharide, Lewis X, oligopeptide, modélisation moléculaire.