# Virus et sucres : une approche par modélisation moléculaire

Solène Grosdidier

▶ **To cite this version:**

Solène Grosdidier. Virus et sucres : une approche par modélisation moléculaire . Sciences pharmaceutiques. 2005. dumas-01312828

**HAL Id: dumas-01312828**

**https://dumas.ccsd.cnrs.fr/dumas-01312828**

Submitted on 9 May 2016

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

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

UFR de Pharmacie
UNIVERSITÉ Grenoble Alpes

SICD1 Bibliothèques de l'UJF et de Grenoble INP

# AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il n'a pas été réévalué depuis la date de soutenance.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact au SID de Grenoble : **thesebum@ujf-grenoble.fr**

# LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/juridique/droit-auteur
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

B.U.M GRENOBLE
D 115 087138 9

1er exemplaire

UNIVERSITE JOSEPH FOURIER
FACULTE DE PHARMACIE DE GRENOBLE

n° 7034

Année 2005

# VIRUS ET SUCRES : UNE APPROCHE PAR MODELISATION MOLECULAIRE

THESE
PRESENTEE POUR L'OBTENTION DU TITRE DE DOCTEUR EN PHARMACIE

DIPLÔME D'ETAT

SOLENE GROSDIDIER

*[Données à caractère personnel]*

THESE SOUTENUE PUBLIQUEMENT A LA FACULTE DE PHARMACIE DE GRENOBLE

Le 21/10/05

DEVANT LE JURY COMPOSE DE

Président de jury : Pr. Emmanuel DROUET

Membres :

Dr. Anne IMBERTY
Dr. Hugues LORTAT-JACOB
Dr. Wilhelm BURMEISTER

*La faculté de Pharmacie de Grenoble n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions sont considérées comme propres aux auteurs.*

BIBLIOTHEQUE UNIVERSITAIRE
GRENOBLE
MÉDECINE - PHARMACIE

Université Jospeh Fourier
**FACULTE DE PHARMACIE DE GRENOBLE**
Domaine de la Merci 38700 LA TRONCHE

Doyen de la Faculté : M. Le Professeur **P. DEMENGE**
Vice-Doyenne : Mme **A. VILLET**

## PROFESSEURS DE PHARMACIE

| | | |
|---|---|---|
| **BAKRI** | Abdelaziz | Pharmacie Galénique |
| **BENOIT-GUYOD** | Jean-Louis | (Emérite) |
| **BURMEISTER** | Wilhelm | Biophysique |
| **CALOP** | Jean | Pharmacie Clinique et Bio-Technique |
| **DANEL** | Vincent | Toxicologie |
| **DECOUT** | Jean-Luc | Chimie Bio-Inorganique |
| **DEMENGE** | Pierre | Physiologie / Pharmacologie |
| **DROUET** | Emmanuel | Immunologie / Microbiologie / Biotechnologie |
| **FAVIER** | Alain | Biochimie / Biote |
| **GOULON** | Chantal | Biophysique |
| **GRILLOT** | Renée | Parasitologie |
| **LECLERC** | Gérard | Chimie Organique |
| **MARIOTTE** | Anne-Marie | Pharmacognosie |
| **PEYRIN** | Eric | Chimie Analytique |
| **RIBUOT** | Christophe | Physiologie / Pharmacologie |
| **ROUSSEL** | Anne-Marie | Biochimie |
| **SEIGLE-MURANDI** | Françoise | Botanique / Cryptogamie |
| **STEIMAN** | Régine | Biologie Cellulaire |
| **WOUESSIDJEWE** | Denis | Pharmacie Galénique |

## PROFESSEUR ASSOCIE (PAST)

| | | |
|---|---|---|
| **CHAMPON** | Bernard | Pharmacie Clinique |

## MAITRES DE CONFERENCES DE PHARMACIE

| | | |
|---|---|---|
| **ALDEBERT** | Delphine | Parasitologie |
| **ALLENET** | Benoit | Pharmacie Clinique |
| **BARTOLI** | Marie-Hélène | Pharmacie Clinique et Biotechnologie |
| **BOUMENDJEL** | Ahcène | Pharmacognosie |
| **CARON** | Cécile | Biologie Moléculaire |
| **CHOISNARD** | Luc | Pharmacotechnie et Génie de la formulation |
| **DELETRAZ** | Martine | Droit Pharmaceutique / Economie |
| **DEMEILLERS** | Christine | Biochimie |
| **DESIRE** | Jérôme | Chimie Bioorganique |
| **DIJOUX-FRANCA** | Marie-Geneviève | Pharmacognosie |
| **DURMORT-MEUNIER** | Claire | Virologie |
| **ESNAULT** | Danielle | Chimie Analytique |
| **FAURE** | Patrice | Biochimie |
| **FAURE-JOYEUX** | Marie | Physiologie / Pharmacognosie |
| **FOUCAUD-GAMEN** | Jacqueline | Immunologie |
| **GEZE** | Annabelle | Pharmacotechnie Galénique |
| **GERMI** | Raphaële | Bactériologie |
| **GILLY** | Catherine | Chimie Analytique |
| **GODIN-RIBUOT** | Diane | Physiologie / Pharmacologie |
| **GROSSET** | Catherine | Chimie Analytique |
| **GUIRAUD** | Pascale | Biologie Cellulaire / Génétique |
| **HININGER-FAVIER** | Isabelle | Biochimie |
| **KRIVOBOK** | Serge | Botanique / Cryptogamie |
| **MORAND** | Jean-Marc | Chimie Thérapeutique |
| **NICOLLE** | Edwige | Chimie Organique / Chimie Thérapeutique |
| **PINEL** | Claudine | Parasitologie |
| **RAVEL** | Anne | Chimie Analytique |
| **RAVELET** | Corinne | Chimie Analytique |
| **RICHARD** | Jean-Michel | |
| **RIONDEL** | Jacqueline | Physiologie / Pharmacologie |
| **SEVE** | Michel | Physique / Biochimie |
| **VANHAVERBEKE** | Cécile | Chimie Bio-Inorganique |
| **VILLEMAIN** | Danielle | Mathématiques / Informatique |
| **VILLET** | Annick | Chimie Analytique |

## PROFESSEUR AGREGE (PRAG)

| | | |
|---|---|---|
| **ROUTABOUL** | Christel | Chimie Générale |

Remerciements

Je souhaite tout d'abord adresser mes plus sincères remerciements à Anne Imberty, mon Directeur de thèse, sans qui je n'aurais pu réaliser ce travail. J'ai tout particulièrement apprécié sa présence et tous ses conseils qui m'ont beaucoup aidée, son énergie incroyable, sa joie de vivre et son humour permanent. Merci de m'avoir tant appris.

Un grand merci à Valérie et Catherine pour m'avoir accueillie dans leur bureau, pour leur complicité de tous les jours et leur bonne humeur.
Merci à Franck et Emilie de m'avoir tant fait rire durant ces six mois et d'avoir toujours été présents à mes côtés. Merci à Charles pour ces grandes discussions philosophiques sur lesquelles nous n'avons pas toujours été d'accord !

Merci à tous pour votre accueil si chaleureux qui me laissera un souvenir impérissable du C.E.R.M.A.V. où je reviendrai avec plaisir.

Et pour finir, je souhaite également remercier toutes les personnes qui me sont proches : merci à ma famille pour son soutien, merci à mon frère, Aurélien, qui est toujours là quand il le faut et qui continuera je l'espère ! Enfin, merci à tous mes amis qui ont pensé à prendre des nouvelles régulièrement pendant la rédaction de cette thèse.

# Table des matières


| | *page* |
|---|---|
| Liste actualisée des enseignants de l'UFR de Pharmacie de Grenoble | 1 |
| Remerciements | 3 |
| Table des matières | 4 |
| **INTRODUCTION** | 6 |
| **RAPPELS HISTORIQUE ET BIBLIOGRAPHIQUE** | 7 |
| 1- LES VIRUS AIMENT LES SUCRES | 7 |
| 1.1- Les virus | 7 |
| 1.2- Reconnaissance des glycoconjugués | 8 |
| 1.3- Définition des lectines | 9 |
| 1.4- Rôle des lectines | 9 |
| 1.5- Classification des lectines | 10 |
| 1.6- Banque de données des lectines | 11 |
| 2- L'HEMAGGLUTININE DU VIRUS DE LA GRIPPE | 13 |
| 2.1- Virus de la grippe | 13 |
| 2.2- Stratégies antivirales | 17 |
| 3- CAPSIDE DU VIRUS DE NORWALK | 23 |
| 3.1- Classification/phylogénie | 23 |
| 3.2- Epidémiologie | 24 |
| 3.3- Diagnostic virologique | 27 |
| 3.4- Mesures préventives | 28 |
| 3.5- Le virus de Norwalk | 29 |
| 3.6- Les récepteurs du virus | 34 |
| **MATERIEL ET METHODES** | 37 |
| 1- Alignement de séquences | 37 |
| 2- Modélisation des carbohydrates | 37 |
| 3- Amarrage moléculaire | 41 |
| 4- Algorithme génétique | 42 |
| 5- Algorithme génétique lamarckien | 44 |
| **RESULTATS** | 46 |
| 1- HEMAGGLUTININE DU VIRUS DE LA GRIPPE | 46 |
| 1.1- Structures | 46 |
| 1.2- Modélisation d'un pentasaccharide Lstc | 48 |
| 2- MODELISATION DE LA CAPSIDE DU VIRUS DE NORWALK AVEC DEUX OLIGOSACCHARIDES | 52 |
| 2.1- Alignement de séquences avec GeneDoc, ClustalX et analyse des familles hydrophobes | 52 |
| 2.2- Amarrage moléculaire | 55 |
| 2.2.1- Analyse spatiale des solutions trouvées par amarrage moléculaire | 57 |

*page*

2.2.2- Familles déterminées par AutoDock 59

2.3- Minimisation dans le champ de force TRIPOS 64

2.3.1- Analyse conformationnelle du trisaccharide H type 1 67

2.3.2- Analyse conformationnelle du trisaccharide représentant l'antigène de groupe A 69

2.4- Evaluation de l'énergie d'interaction 70

2.5- Analyse de la distribution des positions obtenues après l'amarrage moléculaire : représentation du paysage énergétique en coordonnées sphériques 72

2.6- Sélection des meilleures poches de reconnaissance 74

**DISCUSSION** 85

MODELISATION DE LA CAPSIDE DU VIRUS DE NORWALK AVEC DEUX OLIGOSACCHARIDES : POCHE DE FIXATION 85

1.1- Données expérimentales publiées 85

1.2- Prédiction par amarrage pour l'antigène de groupe sanguin A 88

1.3- Prédiction par amarrage pour l'antigène de groupe sanguin O 90

**CONCLUSION** 96

**LISTE DES ABREVIATIONS** 97

**BIBLIOGRAPHIE** 98

**ANNEXES** 102

**Serment de Galien** 114

# INTRODUCTION

Les premières étapes dans la découverte d'un nouveau médicament correspondent à l'identification et à la validation d'une cible thérapeutique, ces phases sont suivies par l'identification de molécules d'intérêt par criblage haut débit qui sont ensuite optimisées avant leur étude pré-clinique et clinique. Le nombre de structures tridimensionnelles protéiques disponibles aujourd'hui est en perpétuelle augmentation grâce aux améliorations dans le domaine de la cristallographie et ces structures résolues constituent autant de nouvelles cibles potentielles pour le domaine médical. La bioinformatique qui est devenue un outil incontournable dans la recherche de nouveaux médicaments intervient entre autres dans l'identification de la cible et les méthodes de modélisation moléculaire sont utilisées pour l'optimisation de petites molécules d'intérêt thérapeutique [17].

L'objectif de cette thèse a été de modéliser le comportement d'oligosaccharides impliqués dans la reconnaissance de nos cellules par deux virus : le virus de la grippe et le virus de Norwalk. Dans un premier temps, un pentasaccharide sous contrainte a été créé par mécanique moléculaire afin d'améliorer sa reconnaissance par l'hémagglutinine du virus de la grippe en vue de réaliser des inhibiteurs compétitifs de cette lectine. Puis, une étude d'interaction par amarrage moléculaire entre les lectines d'un Norovirus et de deux oligosaccharides a été réalisée afin d'identifier des poches potentielles de fixation et d'établir des hypothèses pouvant expliquer le profil d'infection particulier de ce virus.

# RAPPELS HISTORIQUE et BIBLIOGRAPHIQUE

## 1- LES VIRUS AIMENT LES SUCRES

### 1.1- Les virus

Les virus contiennent une information génétique sous la forme d' ARN ou ADN et possèdent souvent une structure de protection protéique et compacte appelée capside afin de le protéger. Ils présentent un mode particulier de multiplication, différent de la mitose cellulaire ou de la multiplication bactérienne. En effet, ce sont des parasites intracellulaires obligatoires ne pouvant se multiplier qu'à l'intérieur d'une cellule hôte en détournant sa machinerie cellulaire. Dans un premier temps, le virus doit se fixer sur les cellules qu'il infecte par le biais de récepteurs spécifiques. Il existe ensuite deux modes différents de pénétration virale. Les virus nus sont phagocytés puis transférés vers d'autres compartiments cellulaires ou libérés dans le cytoplasme. Les virus enveloppés pénètrent, eux, par fusion de leur enveloppe avec la membrane cellulaire ou la vacuole de phagocytose ce qui aboutit à la libération de la nucléocapside dans le cytoplasme. Une fois entrés, les virus peuvent prendre le pouvoir : tous les systèmes de production de la cellule sont alors détournés afin de produire les éléments nécessaires à l'élaboration des virions qui, relâchés, pourront infecter d'autres cellules. Les virus de notre environnement utilisent souvent les sucres exposés à la surface de nos cellules comme récepteurs spécifiques qu'ils reconnaissent par l'intermédiaire de leurs lectines et cette première étape est indispensable à l'infection.

### 1.2- Reconnaissance des glycoconjugués

Les glycoconjugués exprimés à la surface de nos cellules sont des médiateurs très importants de l'interaction cellulaire comme l'adhésion cellulaire, l'adhésion des leucocytes ou encore la diapédèse. Ces glycoconjugués sont également impliqués dans l'adhésion de nombreux pathogènes tels que les bactéries ou les virus. *Campylobacter jejuni* est une des causes les plus courantes de diarrhées dans le monde. Cette bactérie doit effectuer toute une série d'étapes complexes avant de provoquer ce symptôme. Mais c'est l'adhésion, nécessaire à l'invasion des cellules intestinales, qui constitue la première étape obligatoire pour l'infection. Et celle-ci se fait par l'intermédiaire de la reconnaissance de l'épitope fucosyl $\alpha$1,2 de l'antigène de groupe sanguin O dans le mucus intestinal [25]. Les virus sont également nombreux à utiliser les glycoconjugués comme récepteurs à la surface des cellules à infecter. Le virus de la grippe, affection respiratoire touchant les voies aériennes supérieures accompagnée de fièvre, en est l'exemple le plus connu. Il reconnaît l'acide sialique en $\alpha$2,6 situé à l'extrémité des chaînes oligosaccharidiques des cellules épithéliales de notre tractus respiratoire. La reconnaissance des glycoconjugués s'effectue, aussi bien pour les cellules de notre propre organisme que pour les pathogènes de notre environnement, par des protéines bien particulières capables de fixer spécifiquement les carbohydrates connus sous le nom de lectines.

D'autres protéines, comme la gp120 du virus de l'immunodéficience humaine, ne reconnaissent que les sucres quand ils sont associés à la matrice extracellulaire sous forme de protéoglycane et ne sont alors pas appelées lectines.

### 1.3- Définition des lectines

Les lectines sont des protéines fixant les sucres qui diffèrent des anticorps par leur origine non immune et des enzymes car elles ne modifient pas le sucre auquel elles se lient. Elles existent chez la plupart des organismes vivants tels que les virus, les bactéries, les animaux ou les plantes. Les lectines interagissent avec les carbohydrates de manière non-covalente et réversible. Malgré la faible affinité intrinsèque entre la protéine et son sucre, avec une constante de dissociation de l'ordre du millimolaire pour les monosaccharides, les lectines sont hautement spécifiques. Elles ont des structures très diversifiées et sont en général des protéines oligomériques composées de sous-unités dont une ou plus possède un site de fixation pour les sucres. Elles varient en termes de taille, de composition en acides aminés, de nombre de sous-unités, d'organisation (présence ou non de métaux) ainsi que dans leurs structures tridimensionnelles. Dans leur diversité structurale, les lectines peuvent d'ailleurs être comparées aux enzymes bien qu'elles ne montrent pas d'activité catalytique [1].

### 1.4- Rôle des lectines

Les interactions lectine-sucre jouent un rôle clé dans l'organisation de nombreux systèmes biologiques. Elles sont, par exemple, impliquées dans la clairance des glycoprotéines du système circulatoire, le recrutement des leucocytes aux sites inflammatoires ainsi que dans les interactions cellulaires du système immunitaire et la formation de métastases. Elles sont également importantes dans l'embryogénèse, le développement neuronal, l'activité hormonale, la prolifération de certaines cellules ainsi que leur organisation en tissus spécifiques [26].

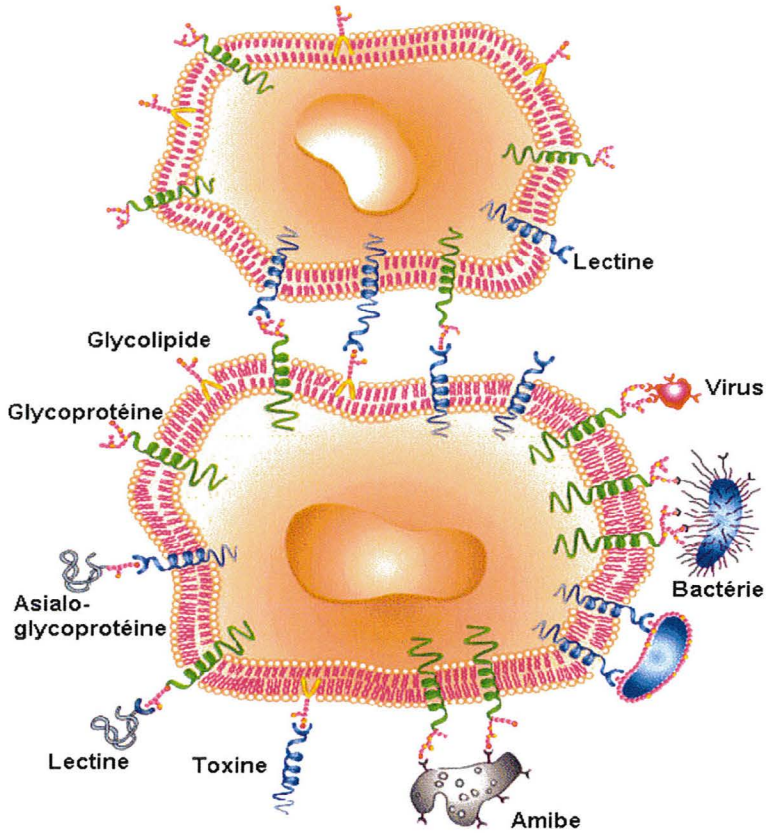



***Figure n°1***
*Selon Sharon et al. (2004)*
*Schéma représentant les différents rôles des lectines*

Ces interactions jouent également un rôle primordial dans l'adhésion d'agents infectieux tels que les virus ou les bactéries et constituent la première étape obligatoire pour l'infection.

### 1.5- Classification des lectines

La grande majorité des lectines ne reconnaissent que quelques monosaccharides sur la centaine existant dans la nature, principalement : le mannose, le glucose, le galactose, la N-acétylglucosamine, le fucose et l'acide N-acétylneuraminique (aussi connu sous le nom d'acide sialique). Par contre, elles peuvent reconnaître une grande quantité d'oligosaccharides lorsque ceux-ci sont composés de ces monosaccharides. Elles sont capables de traduire les

informations codées par les carbohydrates et possèdent un rôle de reconnaissance déterminant dans de nombreux processus biologiques. Pour cette raison, elles sont aujourd'hui considérées comme les "déchiffreuses du glycocode".

La classification des lectines est souvent basée sur leur spécificité vis-à-vis du saccharide qu'elles reconnaissent le mieux. Cependant, une lectine ne montre pas d'affinité pour un saccharide unique mais a la capacité d'en reconnaître plusieurs. Les lectines peuvent être regroupées en familles caractérisées par leur similarité de séquences et de structure tridimensionnelle. La résolution de plusieurs centaines de structures tridimensionnelles a permis d'établir cette nouvelle classification basée sur l'existence de traits structuraux communs (http://www.cermav.cnrs.fr/lectines/). La plupart des lectines appartiennent ainsi à trois classes distinctes : il existe des lectines simples, mosaïques (ou multidomaines) et des assemblages macromoléculaires. Les lectines simples sont composées d'un petit nombre de sous-unités dont chaque monomère possède un site de fixation pour le sucre. Cette classe est composée en majeure partie par les lectines de plantes. Les lectines mosaïques sont composées de plusieurs molécules possédant différents types de domaines protéiques dont un seul possède un site de fixation aux carbohydrates. Ces lectines mosaïques sont surtout rencontrées chez les virus et les animaux tandis que les assemblages macromoléculaires formant des organelles filamenteuses sont retrouvés chez les bactéries.

### 1.6- Banque de données des lectines

Le nombre de structures résolues ayant augmenté de manière exponentielle avec les avancées techniques en cristallographie et résonance magnétique nucléaire (RMN), l'enjeu actuel est d'arriver à classer ces structures dans des banques de données en vue de les exploiter et de faire de la conception rationnelle de molécules à fort potentiel thérapeutique. Le

C.E.R.M.A.V. possède une banque de données rassemblant plus de 150 structures différentes de lectines d'origine animale, végétale, fongique, bactérienne et virale. Les lectines bactériennes et virales sont particulièrement intéressantes puisqu'elles présentent un intérêt thérapeutique direct.

***Tableau I.*** *Lectines virales dans la banque de données du C.E.R.M.A.V.*

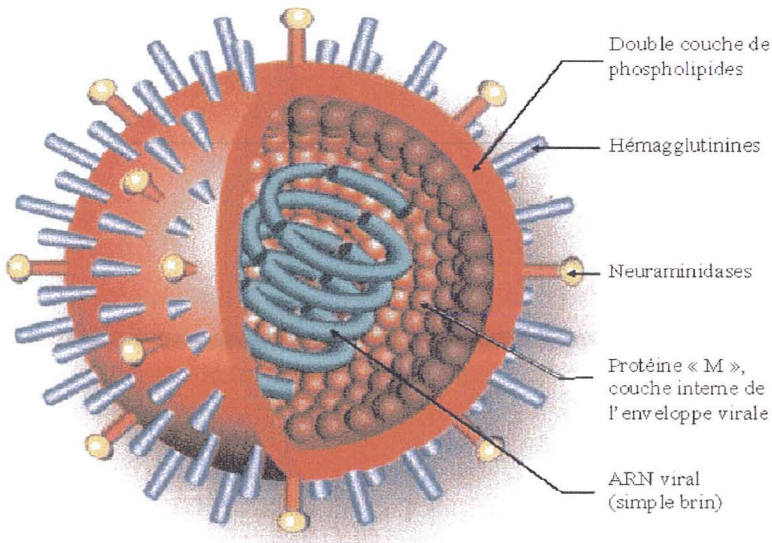
| ***Virus*** | ***Protéine*** | ***Code PDB*** | ***Résolution (Å)*** | ***Année*** |
|---|---|---|---|---|
| **Rotavirus** | Protéine de capside | 1KRI | 0 | 2002 |
| | | 1KQR | 1,4 | 2002 |
| **Virus de la fièvre aphteuse** | Protéine d'enveloppe | 1QQP | 1,9 | 1999 |
| **Polyomavirus de souris** | Protéine d'enveloppe | 1VPN | 2 | 1997 |
| | | 1VPS | 1,9 | 1997 |
| | | 1SID | 3,65 | 1997 |
| | | 1SIE | 3,65 | 1996 |
| **Adénovirus** | Protéine fibre | 1UXA | 1,5 | 2004 |
| | | 1UXB | 1,75 | 2004 |
| | | 1UXE | 2 | 2004 |
| **Phage P22** | Protéine de spicule de queue | 1TSP | 2 | 1994 |
| | | 1TYU | 1,8 | 1996 |
| | | 1TYV | 1,8 | 1996 |
| | | 1TYW | 1,8 | 1996 |
| | | 1TYX | 1,8 | 1996 |
| | | 1CLW | 2 | 1999 |
| | | 1QA1 | 2 | 1999 |
| | | 1QA2 | 2 | 1999 |
| | | 1QA3 | 2 | 1999 |

## 2- L'HEMAGGLUTININE DU VIRUS DE LA GRIPPE

### 2.1- virus de la grippe

Le virus de la grippe cause une affection respiratoire touchant les voies aériennes supérieures : le nez, la gorge et les bronches. L'infection se manifeste par l'apparition brutale d'une forte fièvre accompagnée de myalgies, de céphalées et d'une importante asthénie. La toux est sèche, et le malade présente également une rhinite ainsi qu'une pharyngite. La guérison survient une à deux semaines plus tard sans traitement particulier dans la plupart des cas. Cependant, la grippe peut causer d'importantes complications chez les sujets fragiles comme les jeunes enfants et les personnes âgées. Les personnes souffrant de pneumopathie, de diabète, de problèmes cardiaques ou rénaux constituent aussi des populations à risque par complications de pathologies concomitantes pouvant entraîner la mort.

La grippe est due à trois groupes de virus *Influenzae* : A, B et C.




**Structure simplifiée du virus Influenza**

***Figure n°2***
*(source : internet)*

Ce virus contient huit fragments d'ARN simples brins de ribonucléoprotéines associés à une nucléoprotéine. Ce génome segmenté est contenu dans une enveloppe lipoprotéique et une protéine antigénique appelée protéine de matrice (MP1), à l'intérieur de l'enveloppe, se lie aux fragments d'ARNs. L'enveloppe porte deux types de protéines différentes. La première appelée neuraminidase (NA) appartenant à 9 groupes antigéniques différents, possède une activité enzymatique permettant la libération des virions après le bourgeonnement de la cellule. L'autre protéine correspond à l'hémagglutinine (HA) et présente 15 types antigéniques distincts.

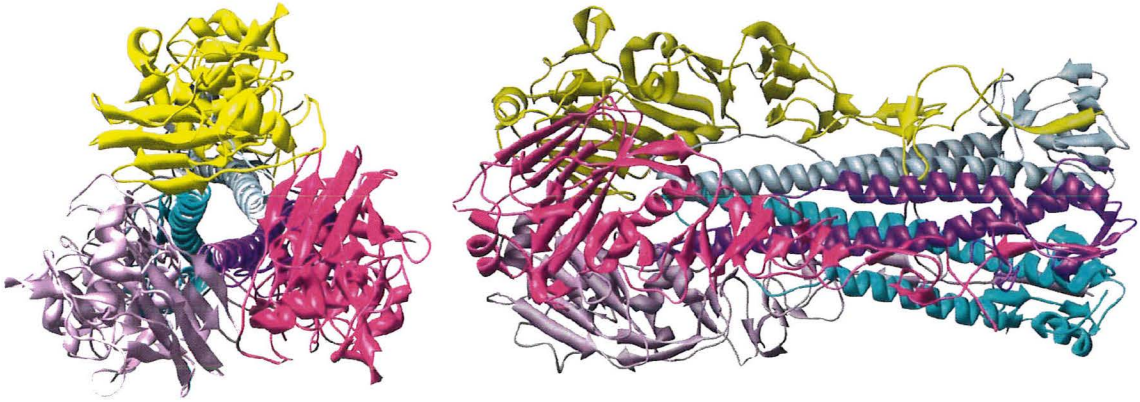


***Figure n°3***
*Selon Gamblin et al. (2004)*
*Structure de l'hémagglutinine du virus de la grippe obtenue par cristallographie (code PDB 1RVT): à gauche, vue de face; à droite, vue de profil.*

Dans un premier temps, l'hémagglutinine va reconnaître puis se fixer spécifiquement à l'acide sialique, carbohydrate présent à la surface de nos cellules puis intervient dans la fusion des membranes virale et cellulaire. L'acide sialique est en général lié en α2,3 ou α2,6 à du galactose, sucre prédominant dans les carbohydrates N-liés de nos chaînes latérales. La préférence de fixation d'une hémagglutinine donnée pour l'une ou l'autre des ces liaisons glycosidiques détermine la capacité d'infection de la souche virale. Cette première étape est indispensable à l'internalisation du virus par endocytose. Le cycle viral s'effectue alors, puis

les virions se forment, bourgeonnent et sont libérés : c'est la neuraminidase qui clive l'acide sialique de la surface cellulaire des virions afin de permettre leur libération et l'infection de nouvelles cellules.

Les virus grippaux B et C sont relativement stables tandis que la souche A évolue sans cesse selon deux mécanismes principaux. Le premier, appelé glissement antigénique, est dû à des mutations au niveau des gènes codant les protéines de surface (hémagglutinine et neuraminidase) et ne provoque que des modifications mineures du virus. Le deuxième phénomène de variation est appelé cassure et peut être plus grave. Des changements radicaux des protéines antigéniques du virus, avec le remplacement d'une protéine par une autre, donnent naissance à un virus totalement différent de celui qui circulait jusque-là. Ce nouveau virus apparaît brutalement et peut ensuite gagner tous les continents. Comme l'immunité pré-existante ne protège pas et que les vaccins préparés avec les souches précédentes sont inefficaces : c'est la pandémie. C'est ainsi que de nouveaux virus sont apparus, causant des pandémies dramatiques. Trois fois au cours du siècle dernier, les virus grippaux A ont subi des modifications génétiques majeures sur l'hémagglutinine provoquant des pandémies avec une forte morbidité et mortalité. La pandémie la plus tristement célèbre est la « grippe espagnole » qui, en 1918 et 1919, a touché une grande partie de la population mondiale et a tué au moins 40 millions de personnes selon les estimations. Plus récemment, deux autres pandémies se sont produites : en 1957 (« grippe asiatique ») et en 1968 (« grippe de Hong Kong »). En se basant sur les tendances historiques, on peut s'attendre en moyenne à trois à quatre pandémies par siècle, avec l'émergence de nouveaux sous-types viraux se transmettant facilement d'une personne à l'autre. Mais il est impossible de prévoir le moment exact où elles surgissent. Aujourd'hui, le virus de la grippe fait donc l'objet d'une surveillance particulière de la part de l'Organisation Mondiale de la Santé car l'émergence d'une nouvelle pandémie est redoutée. Lors des épidémies annuelles, 5 à 15 % de la population souffre

d'infections des voies respiratoires supérieures. Même si ces chiffres sont difficiles à évaluer, on pense que ces épidémies entraînent entre 3 et 5 millions de cas graves et 250 000 à 500 000 décès par an dans le monde. Dans les pays industrialisés, la plupart des décès dus à la grippe surviennent chez les personnes de plus de 65 ans. Bien que les vaccins anti-grippaux existent, la mise au point d'un antiviral efficace contre l'influenza aurait des retombées économiques considérables, ce virus intéresse donc particulièrement l'industrie pharmaceutique. Le virus de la grippe suscite aujourd'hui un vif intérêt et se trouve ainsi au coeur d'une recherche intense.

Tous les virus grippaux de type A qui sont spécifiques de différentes espèces sont génétiquement instables et parfaitement adaptés pour échapper aux défenses immunitaires de l'hôte. Ainsi, les virus aviaires sont des virus grippaux qui n'infectent que les oiseaux et les porcs. Ces virus ne disposent pas de mécanismes de réparation des erreurs qui se produisent pendant la réplication. Bien que ces virus aviaires n'infectent normalement pas l'espèce humaine, un premier cas documenté d'infection s'est produit à Hong Kong en 1997, lorsqu'une souche aviaire (H5N1), causant une épidémie hautement pathogène chez la volaille, a provoqué une affection respiratoire sévère chez 18 personnes dont 6 sont décédées. Le 5 août 2005, le ministère de la santé viêtnamien a confirmé trois nouveaux cas par le virus H5N1, dont deux des patients sont décédés. D'après le rapport de l'OMS du 5 août, le nombre cumulé de cas humains confirmés depuis décembre 2003 est de 112 infections et de 57 décès.

C'est l'hémagglutinine qui va déterminer la capacité d'infection de la souche virale de par ses préférences de fixation pour les liaisons glycosidiques de l'acide sialique. Ainsi, cette protéine est responsable de la barrière d'espèces. Les hémagglutinines aviaires se fixent préférentiellement à l'acide sialique lié en α-2,3 tandis que celles infectant les cellules humaines reconnaissent l'acide sialique en α-2,6, conformation la plus rencontrée dans notre tractus respiratoire. Un changement dans la reconnaissance du virus a été requis pour le transfert de l'oiseau à l'homme [11].

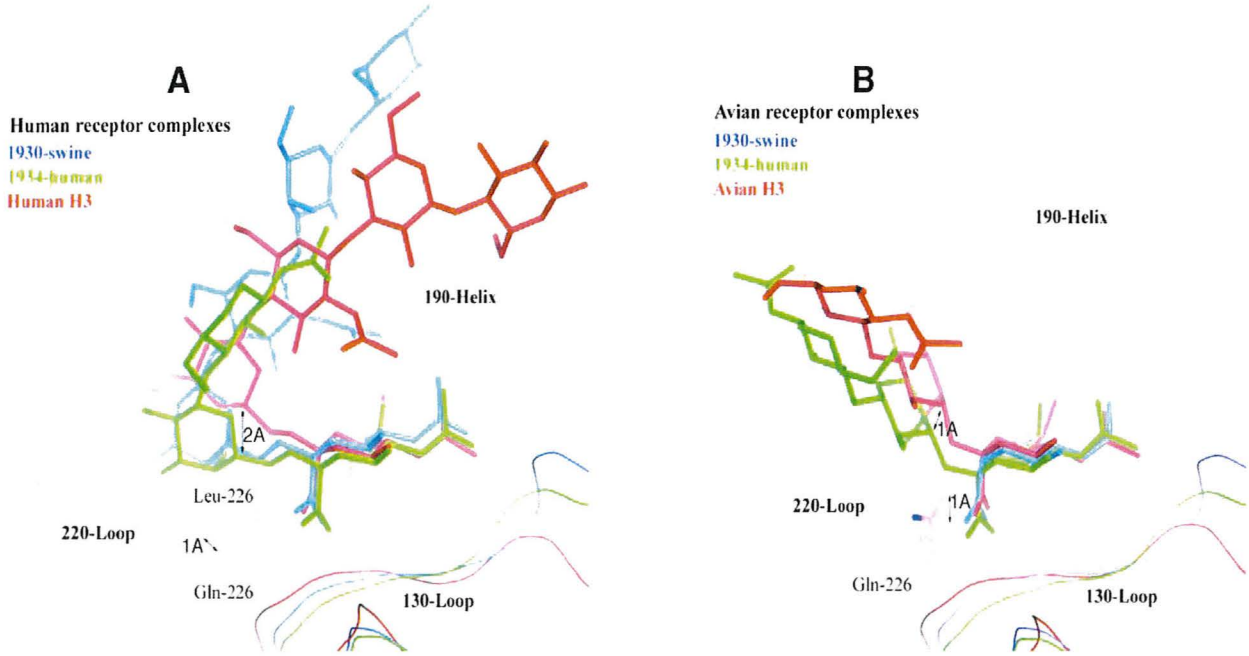



***Figure n°4***
*Selon Gamblin et al. (2004)*
*Différences d'orientation des récepteurs humains et aviaires dans le site actif de trois souches différentes d'hémagglutinines d'origine porcine, aviaire et humaine. Les sialopentasaccharides sont colorés en fonction de l'hémagglutinine à laquelle ils sont liés.*
**A.** *modèle du récepteur humain : Lstc,* **B.** *modèle du récepteur aviaire : Lsta*

### 2.2- Stratégies antivirales

Il existe actuellement peu d'antiviraux pour soigner la grippe. Certains sont utilisés depuis plusieurs années comme l'amantadine et la rimantadine (qui n'est aujourd'hui plus commercialisée en France). Ces deux médicaments ne présentent cependant qu'une efficacité limitée sur les souches du virus de type A seulement. Le mode d'action de l'amantadine reste encore aujourd'hui peu connu. Cette molécule préviendrait et ralentirait la libération de l'acide nucléique de la cellule hôte, réduisant ainsi la formation de nouvelles particules virales et l'infection de nouvelles cellules. La rimantadine agit, elle, sur la pénétration et la décapsidation du virus. Le coût et la fréquence des effets secondaires (tels que des délires et

des convulsions) ne permettent pas un usage de ces deux médicaments en routine et ils ne sont que très peu utilisés en Europe.

La vaccination constitue la meilleure arme en terme de santé publique contre le virus de la grippe. La composition du vaccin à base de particules virales inactivées varie d'une année à l'autre selon le contexte épidémiologique puisque la souche change tous les ans. Cette vaccination, efficace dans 80% des cas, est très fortement recommandée chez les sujets à risque comme les personnes de plus de 70 ans, les personnes immunodéficientes ou souffrant de maladies chroniques ainsi que le personnel soignant.

Aujourd'hui, de nouvelles stratégies anti-virales sont développées afin de trouver des molécules ciblant les deux protéines structurales du virus : l'hémagglutinine et la neuraminidase. La neuraminidase, dont la structure tridimensionnelle a été découverte en 1983, consiste en quatre monomères protéiques identiques [7].

Chaque monomère possède une profonde cavité où se situe le site de reconnaissance de l'acide sialique. Bien que les variations génétiques normales produisent des changements dans la séquence de l'enzyme, la cavité demeure identique et sa nature hautement conservée a rendu cette poche particulièrement intéressante pour le développement d'antiviraux [8].

Deux inhibiteurs oligosaccharidiques de la neuraminidase : le zanamivir (Relenza® ou la 4-guanidino-Neu5Ac2ène) et l'oseltamivir (Tamiflu®) ont ainsi récemment vu le jour pour le traitement des grippes A et B.

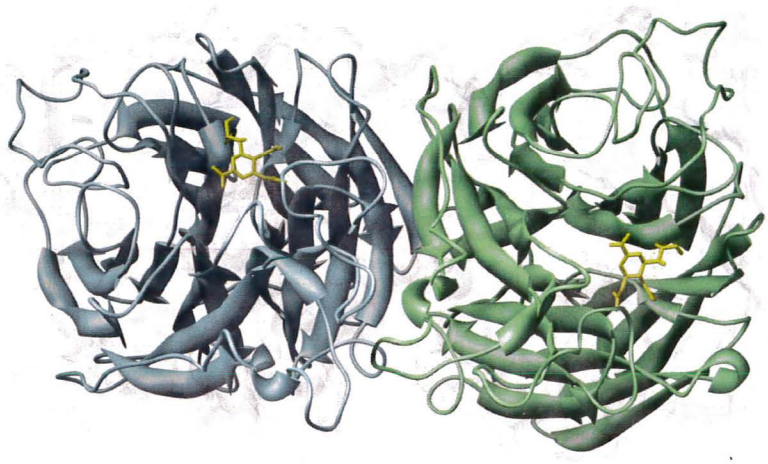


***Figure n°5***
*Selon Cleasby et al. (1998)*
*Deux monomères de neuraminidases en complexe avec deux analogues carboxamides du Zanamivir (code PDB 1A4G)*

La neuraminidase clive les liaisons formées entre les hémagglutinines et les acides sialiques de la cellule hôte afin de libérer les nouveaux virions en fin de cycle viral. Ces deux molécules, analogues de l'acide sialique, inhibent cette enzyme par blocage de ses sites actifs et empêchent les virions de se libérer des acides sialiques présents sur nos cellules aboutissant à leur agrégation au niveau de la surface cellulaire.

La prise de ces deux médicaments dans les premières 48 heures suivant l'apparition des symptômes permet d'en réduire la durée en limitant leur sévérité.

La mise au point d'inhibiteurs de l'hémagglutinine, contrairement à ceux de la neuraminidase, permettrait de cibler le virus pendant son étape de reconnaissance et de prévenir son entrée dans la cellule. Le principe général de la thérapie anti-adhésive correspond à l'inhibition de la reconnaissance de nos cellules par les micro-organismes en utilisant des analogues de leurs

récepteurs sous forme soluble. Malgré l'attractivité évidente de cette technique, cette thérapie anti-adhésive semble encore loin d'être applicable pour l'hémagglutinine. Les oligosaccharides monomériques sont en effet incapables d'entrer en compétition avec les mêmes oligosaccharides à la surface des cellules à cause de leur trop faible affinité d'interaction, la reconnaissance entre le virus et la cellule s'effectuant en plusieurs endroits différents et de manière coopérative le plus souvent. De plus, la reconnaissance du carbohydrate est rapidement suivie par une cascade d'événements rendant l'entrée du virus inévitable. Une manière d'augmenter l'affinité des oligosaccharides sous forme soluble consiste à créer des formes multivalentes, or pour pouvoir être utilisée, une molécule thérapeutique doit être de petite taille. Afin de surmonter ces difficultés, une équipe de chercheurs russes travaille, en collaboration avec le C.E.R.M.A.V., sur l'obtention de petits assemblages glycopeptidiques. La partie carbohydrate du glycopeptide est responsable de la fixation à la lectine virale qu'elle inhibe par compétition tandis que la région protéique intervient dans la formation de la structure appelée tectomère. Suivant leur organisation, l'assemblage des tectons (unité glycopeptidique) en tectomères peut être spontané ou dépendant de la présence du virus ciblé; ainsi, des sialopeptides ont pu être obtenus en présence du virus de la grippe. Ces sialopeptides ou tetrahédrons (tectomère tétravalent) ont déjà montré une bonne efficacité chez la souris [5][6].

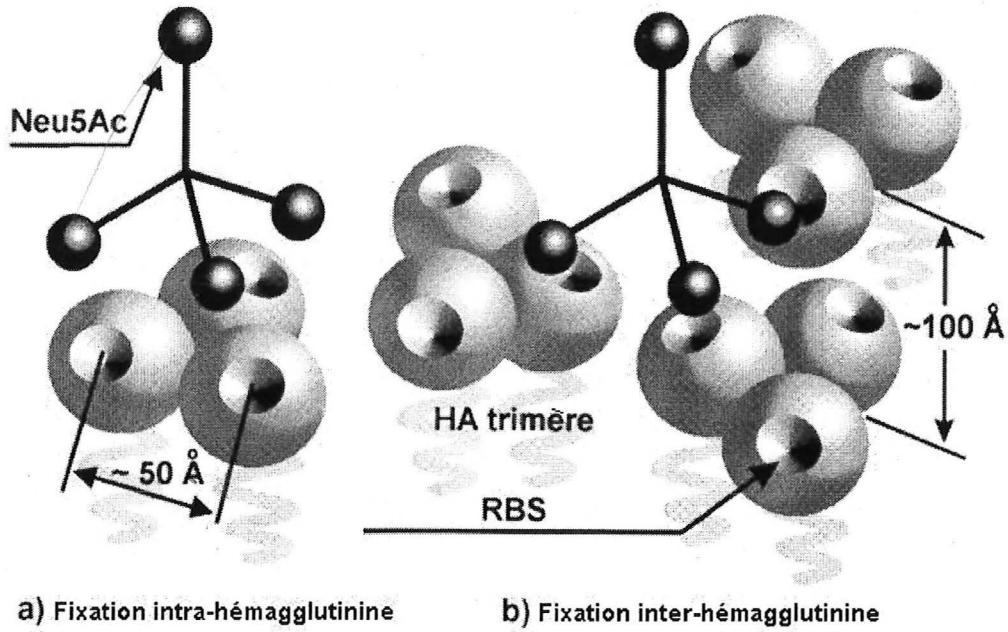



***Figure n°6***
*Selon Chinarev et al. (1999)*
*Les tétrahédrons peuvent inhiber les hémagglutinines du virus de la grippe de deux manières différentes, soit par interaction intra-hémagglutinine en fixant les sites de reconnaissance de l'acide sialique au niveau de chacun des trimères, soit par interaction inter-hémagglutinine impliquant plusieurs trimères différents.*

Le récepteur humain pour l'hémagglutinine est modélisé par un pentasaccharide appelé Lstc correspondant à l'acide sialique lié à deux galactoses, à une N-acétyl glucosamine et une glucosamine (αNeuAc-2-6-βGal-1-4-βGlcNAc-1-3-βGal-1-4-Glc) [11]. Après étude, le groupe N-acétyl appartenant au troisième résidu de ce fragment Lstc semble pouvoir former

une liaison hydrogène avec un des groupements hydroxyls de l'acide sialique. Et il semblerait aujourd'hui que seule cette conformation repliée interagisse avec l'hémagglutinine.

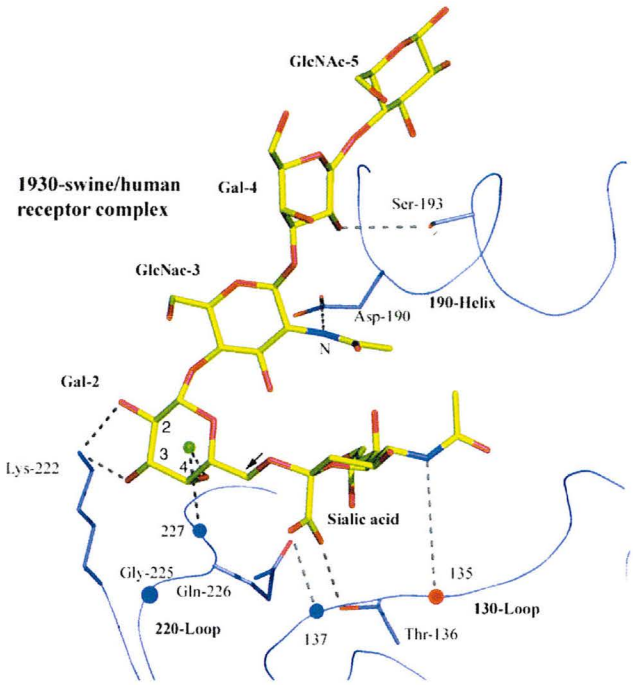



***Figure n°7***
*Selon Gamblin et al.*
*Structure du Lstc, modèle du récepteur humain pour l'hémagglutinine, dans le site de reconnaissance d'une hémagglutinine de souche porcine.*
*(code d'accès PDB : 1RVT)*

Afin d'optimiser l'inhibition de l'hémagglutinine par les sialopeptides, la conception d'une conformation sous contrainte du Lstc assistée par ordinateur a été réalisée. Cela, afin de forcer le pentasaccharide à adopter la conformation repliée quelle que soit la structure du tecton. En effet, il faut faire une conformation bloquée pour limiter la flexibilité du sucre et ainsi diminuer l'entropie [22].

## 3- CAPSIDE DU VIRUS DE NORWALK

La gastroentérite est une affection inflammatoire touchant l'estomac, l'intestin grêle et le gros intestin caractérisée par l'émission brutale et fréquente de selles liquides et abondantes. Souvent appelée à tort « grippe intestinale » , la gastroentérite virale peut être causée par une grande variété de virus incluant les rotavirus, les adénovirus, les astrovirus et les calicivirus.

### 3.1- Classification/phylogénie

Les *Caliciviridae* doivent leur nom aux dépressions en forme de calice observées en microscopie électronique à la surface de certains de leurs virus. Cette famille est divisée en quatre genres basés sur des différences morphologiques et génomiques : les Vesivirus, les Lagovirus, les Sapporovirus comme le Sapporo virus aussi appelé SLV (« Sapporo-like virus ») et enfin les Norovirus aussi appelés NVL pour « Norwalk-like virus » [12].

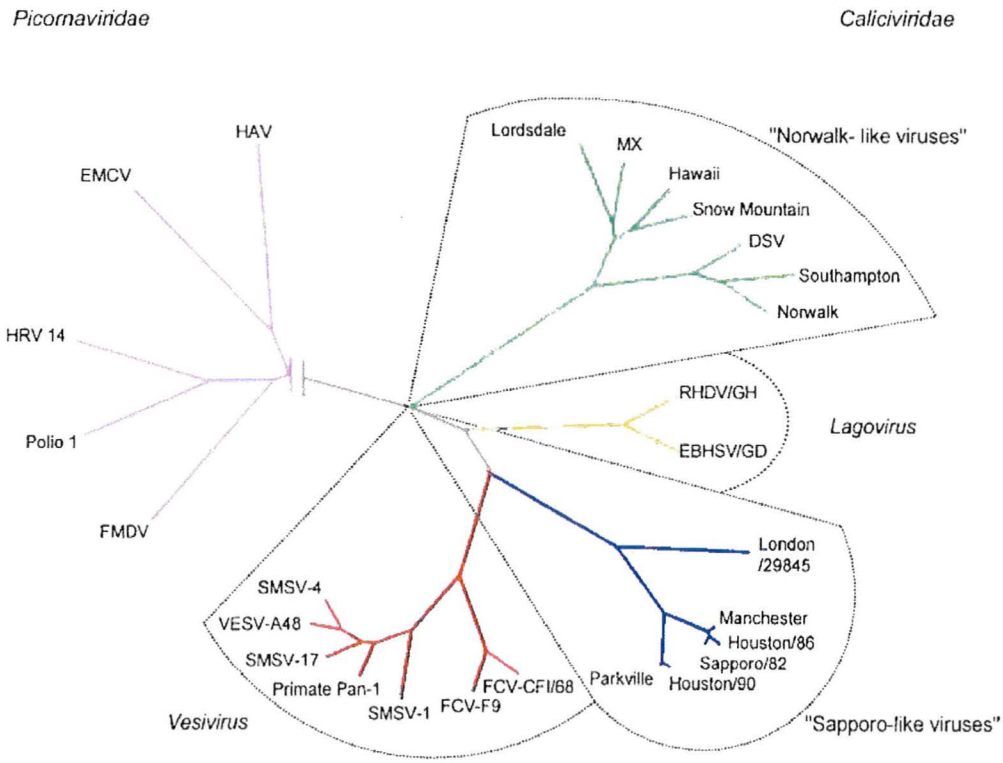



***Figure n°8***
*Selon Green et al. (2000)*
*Taxonomie des Calicivirus*

Les Norovirus représentent plus de 90% des gastroentérites non bactériennes dans le monde et sont génétiquement très divers avec de nombreux génotypes classés en 3 génogroupes différents : GGI, GGII et GGIII.

### 3.2- Epidémiologie

Le virus de Norwalk (NV) est la souche la plus représentative du groupe des *Caliciviridae* et appartient au génogroupe I. Aux Etats-Unis, le nombre d'infections est estimé à 23 millions par an, dont 50.000 aboutissent à une hospitalisation et 300 décès sont dénombrés chaque année. La contamination se fait principalement par voie oro-fécale d'individu à individu, par ingestion d'eau (de boisson, de piscine...) ou d'aliments contaminés [8]. Le virus de Norwalk est un virus très résistant notamment au traitement chloré utilisé pour décontaminer l'eau ce qui explique que la contamination résulte d'un défaut d'hygiène ou d'assainissement. En théorie, tous les aliments sont susceptibles de contenir des virus par contamination fécale, cependant certains sont plus souvent impliqués que d'autres dans l'apparition d'épidémies comme par exemple les coquillages. Les huîtres ou les clams ont en effet tendance à concentrer les NLVs présents dans leur environnement par filtration d'eau contaminée. De plus les virus ne sont pas toujours inactivés par la cuisson puisqu'ils sont capables de résister à un chauffage jusqu'à 60°c.

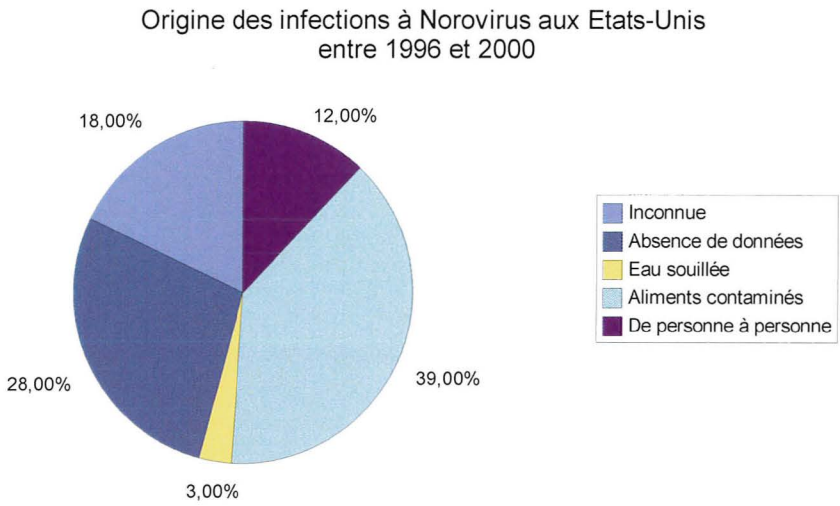



***Figure n°9***
*Selon Fankhauser et al. (1998)*

La transmission du virus par aérosol a été évoquée et demeure une hypothèse forte qui expliquerait la large diffusion virale lors des grandes épidémies hivernales de Novembre à Mars dans les pays tempérés. Ces différents modes de transmission ont un rôle indéniable dans la propagation du virus favorisée par la promiscuité et explique l'apparition d'épidémies fréquentes dans les institutions telles que les maisons de retraite, les crèches, les hôpitaux ou même les bateaux de croisière.

Une fois ingéré, le virus infecte les entérocytes matures de l'intestin grêle qui prennent alors une apparence anormale avec une diminution des microvillosités, une dilatation du réticulum endoplasmique. De grosses mitochondries et des oedèmes intracellulaires sont également observés. Les symptômes : diarrhées, crampes abdominales, nausées, vomissements peuvent être accompagnés de malaises, maux de tête, myalgies et de faible fièvre. Ils peuvent persister jusqu'à 72 heures. L'enfant semble plus touché par les vomissements tandis que l'adulte souffre majoritairement de diarrhées aqueuses. L'évolution est le plus souvent favorable et il est rare que les patients fassent appel à leur médecin, cependant la déshydratation importante

peut entraîner un déséquilibre hydroélectrolytique à l'origine de complications graves pouvant être létales. La période d'incubation peut varier de 12 à 48 h et les personnes atteintes sont contagieuses pendant la phase aiguë de la maladie et jusqu'à deux à trois jours suivant la disparition de tous signes cliniques. Ces caractéristiques, combinées avec l'incapacité de développer une immunité à long terme, font de cette famille de virus un important problème de santé publique en Europe et dans le monde.

***Tableau II.*** *Principales caractéristiques des Norovirus facilitant la dissémination virale lors des épidémies. Selon Parashar et al. (2001)*

| **Caractéristiques** | **Conséquences** |
|---|---|
| Faible dose infectieuse (<100 particules) | Permet la **dissémination virale** de personne à personne |
| Dissémination post-symptomatique prolongée jusqu'à 2 semaines | Risque augmenté de **transmission secondaire** mais surtout du début des symptômes jusqu'à 2-3 jours après leur disparition |
| Stabilité dans l'environnement : résistance à la congélation, au chauffage à 60°c ou à un traitement <10 mg/L de chore | **Elimination** du virus **difficile** dans l'eau contaminée, les glaçons ou les huîtres mêmes cuites peuvent causer la maladie |
| Grandes diversités de souches | **Diagnostic difficile. Infections à répétition** possibles par des souches différentes compliquant les estimations de prévalence |
| Absence d'immunité durable | Exposition virale durant l'enfance n'empêche pas les infections à l'âge adulte. **Développement d'un vaccin** permettant la protection à vie **difficile** |

### 3.3- Diagnostic virologique

Le diagnostic repose sur les signes cliniques du patient. Cependant, ces dix dernières années, les avancées dans les méthodes de détection des NLVs ont changé la compréhension de l'épidémiologie de ces virus. Trois techniques sont aujourd'hui utilisées sans toutefois être disponibles en routine : la microscopie électronique avec marquage immunologique, le dosage immunoenzymatique (Enzyme Linked Immuno Sorbent Assay) et l'amplification génomique du virus (Reverse Transcriptase Polymerase Chain Reaction) à partir de prélèvement de fèces ou de vomissures [2].

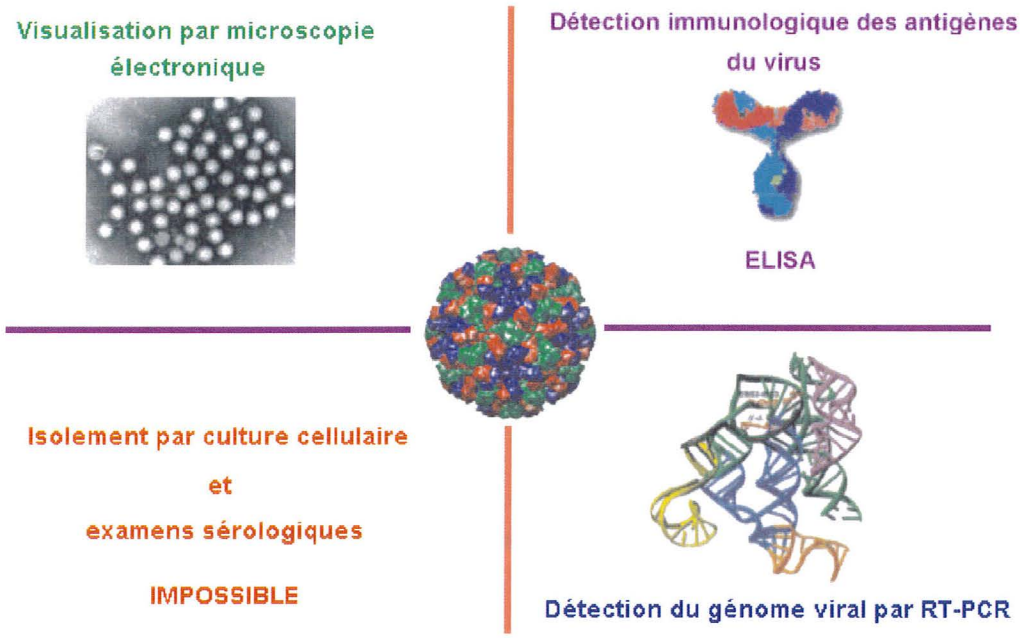



***Figure n°10***
*Moyens diagnostics actuellement disponibles pour le virus de Norwalk (source des images : internet)*

***Tableau III.*** *Diagnostic différentiel avec les autres souches virales à l'origine de gastroentérites.*

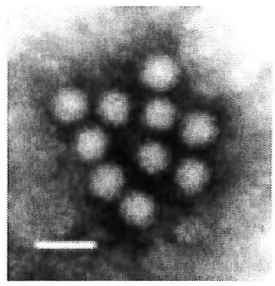
| | ***Caliciviridae*** | ***Rotaviridae*** | ***Astroviridae*** | ***Adenoviridae*** |
|---|---|---|---|---|
| **Signes Cliniques** | Diarrhées +++<br>Vomissements ++<br>Fièvre | Diarrhées<br>Fièvre<br>Déshydratation +<br>Vomissements +/- | Diarrhées modérées<br>Vomissements +/-<br>Fièvre +/- | |
| **Variation saisonnière** | Surtout l'hiver | Pic hivernal | Hiver/printemps | Toute l'année |
| **Caractéristique épidémiologique** | Majoritairement épidémique | Sporadique et saisonnière | Épidémique ou sporadique | |
| **Population cible** | Tout âge | Jeunes enfants<br>Personnes âgées | Enfants<br>Personnes âgées<br>Immunodéprimés | Enfants de moins de 2 ans |

### 3.4- Mesures préventives

Bien que la transmission de personne à personne semble difficilement contrôlable, des mesures faciles à mettre en oeuvre peuvent tout de même limiter la dissémination virale. Le lavage fréquent et énergique des mains pendant plus de 10 secondes au savon suivi d'un long rinçage à l'eau est un moyen simple et efficace de prévention. Une hygiène générale est primordiale : les vêtements doivent être lavés en machine sur le cycle le plus long avec séchage et le nettoyage des surfaces à l'aide de produits détergents appropriés est également important. Pour finir, les camps de vacances ou les croisières au cours desquels l'épidémie est susceptible d'être propagée, par renouvellement de la population exposée, devraient être fermés afin de permettre une décontamination soigneuse des locaux.

### 3.5- Le virus de Norwalk

Le premier calicivirus fut découvert en 1972 dans des échantillons cliniques prélevés lors d'une épidémie hivernale survenue quatre ans plus tôt dans une école de Norwalk dans l'Ohio (USA) par microscopie électronique. La microscopie électronique avec marquage immunologique a permis une visualisation directe par interaction antigène-anticorps de la particule virale en faible quantité.



Celle-ci ne présentant pas de caractéristiques dictinctives fut tout d'abord désignée dans la littérature par l'expression « small round structured virus » ou SRSVs. Sa caractérisation épidémiologique et biologique n'a débuté que récemment. Son étude est en effet rendue difficile par plusieurs facteurs. Il existe tout d'abord peu de systèmes de culture simples et appropriés pour les plupart des Calicivirus. Seuls les Vesivirus peuvent se multiplier dans des cellules en culture. Les virions natifs infectieux sont également difficiles à caractériser étant donné leur faible quantité d'excrétion dans les selles et les vomissures et la microscopie électronique n'est souvent pas assez sensible pour les détecter. Des avancées récentes dans diverses techniques de biologie moléculaire et l'apparition d'un système d'expression ont mis en lumière l'incidence des *Caliciviridae* dans toutes les classes d'âge ainsi que leur capacité à causer l'infection par diverses voies de transmission aussi bien que leur incroyable diversité génétique. L'analyse de la séquence des NLVs a confirmé cette extraordinaire diversité pouvant refléter un nombre aussi important de sérotypes que pour les rhinovirus, ce qui suggère que des antiviraux puissent être nécessaires pour traiter les infections dues aux *Caliciviridae*.

Le virus de Norwalk possède un ARN positif simple brin de 7.7kb et son génome est organisé en 3 cadres de lecture distincts.

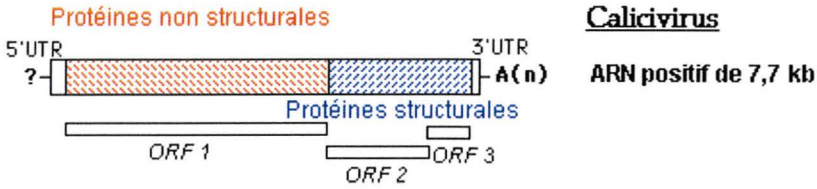



***Figure n°11***
*Organisation du génome des Calicivirus dont fait partie le virus de Norwalk (source : www.allthevirologyonthewww.com )*

Le premier ORF débute à l'extrémité 5' du génome après une courte séquence non codante et son expression produit une polyprotéine non structurale de plus de 190 kDa. Elle contient entre autres une hélicase, la Vpg, une protéase 3C-like et une ARN polymérase ARN dépendante. Les deux ORFs suivants codent pour les protéines structurales, le second cadre de lecture permet d'obtenir la protéine de capside appellée VP1 et le troisième code la protéine structurale mineure : VP2 ayant pour rôle de stabiliser l'ensemble de la capside du virus [3]. (voir ANNEXE A/ p102)

Les *Caliciviridae* constituent une classe à part parmi les virus. Certains possèdent en effet une capside composée d'une seule protéine. En conséquence, toutes les entités fonctionnelles nécessaires à ces virus afin d'assurer leur intégrité structurale, leur immunogénicité et leur infectivité sont codées dans une structure protéique unique. La compréhension de la structure et des fonctions de cette protéine de capside constituent donc les points de départ, clés pour le développement de stratégies antivirales contre les Norovirus. La protéine de capside du virus de Norwalk a pu être produite en cellules d'insectes après clonage de l'ORF2 en baculovirus et son expression dans ces cellules résulte en l'autoassemblage de virus recombinants vides

appelés VLPs pour « Virus-Like Particles ». Ces particules sont morphologiquement identiques et possèdent les mêmes caractéristiques antigéniques que le virus natif.

La structure cristallographique de la capside du virus de Norwalk, résolue par rayon X en 1999 à 3.4 Å et enregistrée dans la PDB sous le code 1IHM, permet aujourd'hui l'étude des propriétés antigéniques du virus ainsi que l'élaboration d'hypothèses quant à l'assemblage de la capside [21]. Chaque particule virale est composée de 180 protéines VP1, organisées en 90 dimères ou capsomères nécessaires à la formation de la capside de 38 nm de diamètre. Afin de former une structure icosaédrique T=3, la protéine de capside doit s'adapter à trois positions quasi équivalentes, et les sous-unités à ces positions sont conventionnellement appelées : A, B et C. Les protéines de capside A s'assemblent avec les protéines de capside B pour former des dimères tandis que les protéines de capside C s'associent entre elles.

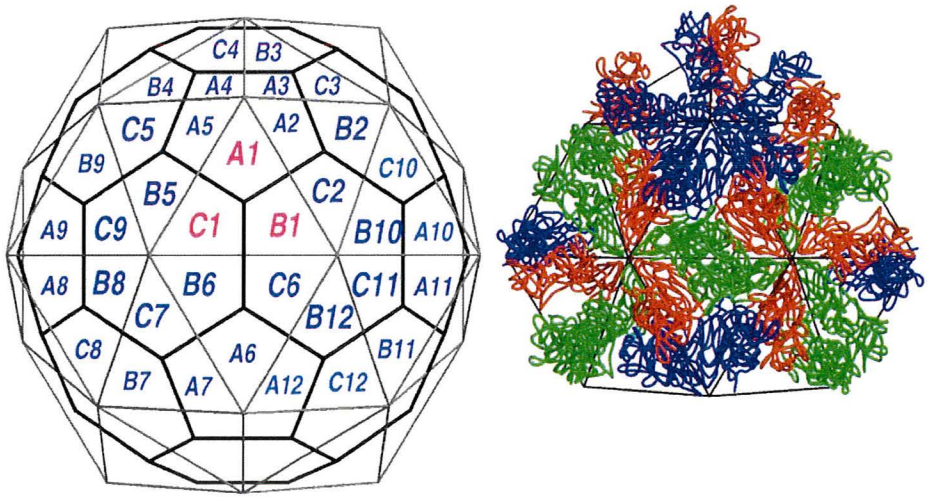



***Figure n°12***
*Capsomère : Cette figure montre l'assemblage des squelettes carbonés des différentes sous-unités superposées à la structure icosaédrique afin de souligner la relation entre la chimie et la géométrie de ces assemblements remarquables. (source : VIPER)*

La protéine de capside, d'environ 60 kDa, est organisée en deux domaines principaux appelés S et P. Le domaine S (pour « Shell » ou coque) est impliqué dans la formation de la coque icosaédrique, il est constitué des 225 premiers acides aminés de l'extrémité N-terminale. Les résidus 50 à 225 forment le classique « ß-sandwich » composé de huit feuillets ß retrouvé dans beaucoup d'autres capsides virales. Le domaine P (pour « Protrusion ») forme une protrusion proéminente émanant de la coque et s'étendant du résidu 226 jusqu'à l'extrémité C-terminale. Ce dernier ne ressemble à aucune autre protéine virale. Il est divisé en deux sous-domaines distincts, le premier, P1, est composé des résidus 226 à 278 et 406 à 520 et le deuxième, P2, est représenté par les résidus 279 à 405.

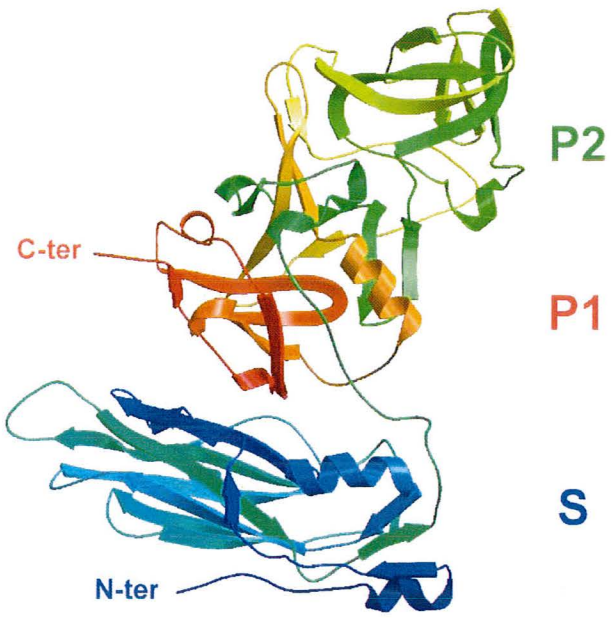
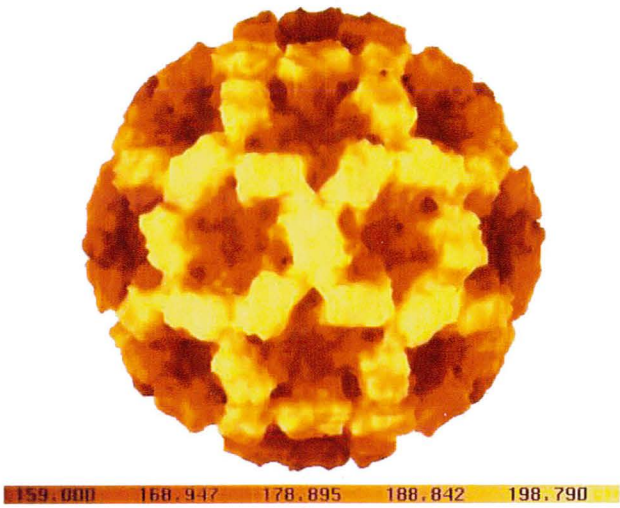



***Figure n°13***
*Vue de la protéine de capside avec le domaine S en bleu, le domaine P1 en rouge et P2 en vert. Ce sont les domaines P1 et P2 qui s'assemblent avec ceux d'une autre protéine de capside et forment une protrusion émanant de la coque virale. (source: VIPER)*

Dans la structure modulaire de la capside, le domaine S est responsable des contacts formant l'icosaèdre tandis que le domaine P intervient exclusivement dans les contacts dimériques. Les dimères AB et CC sont majoritairement stabilisés par les interactions entre chaînes latérales des deux domaines P des sous-unités formant le dimère.

***Figure n°14***
*Surface virale d'une VLP du virus de Norwalk*




*L'image a été calculée à partir des coordonnées atomiques de la base de données VIPER. Les couleurs reflètent la profondeur de la capside selon une gamme monochromatique du marron clair pour les diamètres les plus importants au marron plus foncé au fur et à mesure que le diamètre diminue. La surface est construite de manière à avoir un rendu proche de l'image obtenue par reconstruction après cryo-microscopie électronique à 20 Å de résolution. La topologie de la surface est caractéristique : des «canyons» et des «sommets» : les calices.*
*(source VIPER)*

Afin d'étudier les différents domaines fonctionnels de la protéine de capside impliquée dans l'assemblage du virus, des mutants de cette protéine ont été produits et leur capacité à s'autoassembler en VLPs a été analysée notamment par microscopie électronique. Ainsi, les 20 premiers résidus de l'extrémité N-terminale se sont montrés indispensables pour l'assemblage T=3 de la particule virale. La délétion de 3 des 8 feuillets ß a pour conséquence une perte totale de la capacité d'autoassemblage, le domaine S ayant perdu son intégrité.

Le domaine P, connu pour intervenir dans les contacts dimériques et conférer une stabilisation à l'ensemble de la structure icosaédrique, est également impliqué dans le contrôle du diamètre de la capside. En effet, quelle que soit l'importance des délétions dans le domaine P, elles ont toutes pour conséquence la formation de particules virales anormales se traduisant par une

variation de son diamètre. La délétion des 230 derniers résidus correspondant à la quasi totalité du domaine P conduit à la formation de particules entièrement lisses de 30 nm avec une structure icosaédrique conservée. Pour tous les mutants à l'extrémité C-terminale, l'absence d'interactions entre les domaines P et S semble donc compromettre le diamètre de la capside, or il a été trouvé par cristallographie qu'une liaison hydrogène existait entre l'extrémité C-terminale et le domaine S, il se pourrait donc que celle-ci soit à l'origine du contrôle de la taille de la capside [4].

La comparaison de diverses protéines capsidiques de souches distinctes de *Caliciviridae* a montré que le domaine S est la région la plus conservée. Le sous-domaine P1 est plus variable mais c'est la séquence entre les acides aminés 280 et 400 dans le sous-domaine P2 qui est hypervariable. P2 étant également la portion la plus exposée à la surface de l'enveloppe virale, il intervient très certainement dans la reconnaissance et l'attachement à la cellule intestinale de l'hôte.

### 3.6- Les récepteurs du virus

Ce virus a un pouvoir hautement pathogène : moins de 100 particules virales sont requises pour l'infection. En dépit de cela, des études auprès de volontaires ont montré que certains individus étaient résistants à toute infection malgré l'ingestion de charges virales importantes. La preuve d'une immunité protectrice est controversée, une immunité à court terme a pu être observée mais les anticorps ne semblent pas fournir de protection contre l'infection virale. De plus, malgré la forte prévalence de ces anticorps au sein de la population, le virus de Norwalk infecte des individus de tous âges. L'absence d'anticorps anti-NVs dans un groupe de patients volontaires résistants suggère une non-susceptibilité d'origine génétique par absence de récepteurs pour le virus [13]. Des études récentes ont montré que des VLPs de différents

génogroupes : GGI et GGII, révèlent des profils de fixation différents selon les antigènes de groupe sanguin (HBGAs) des volontaires [24]. Ces antigènes, présentés à l'extrémité terminale de chaînes oligosaccharidiques liées à des protéines ou des lipides, sont très exprimés dans les tissus humains. Ainsi, les antigènes de groupe sanguin ABH pourraient influencer la susceptibilité face au virus de Norwalk. Il existe dans la population deux groupes de personnes répondant de façon différente à l'ingestion du virus. Le premier groupe, composé d'individus dits « non sécréteurs » (Se-), n'exprime pas leurs antigènes de groupes sanguins dans divers fluides comme la salive ainsi qu'à la surface de certains types de cellules épithéliales et se trouve être résistant au virus de Norwalk par absence de récepteurs à la surface de leurs cellules épithéliales intestinales. L'autre groupe composé de « sécréteurs » (Se+) est susceptible d'être infecté, cependant le virus possède un profil de fixation spécifique. Les individus de groupe sanguin O (exprimant uniquement l'antigène H type 1) et ceux de groupe A présentent ainsi des risques d'infection tandis que l'antigène de groupe sanguin B non reconnu, confère aux individus de groupe B une résistance naturelle au virus.

Les données sur la fixation des VLPs suggèrent que ces 20% de personnes sont génétiquement résistantes au virus de Norwalk [18].

Actuellement pour le virus de Norwalk, le croisement des données avec d'autres souches virales a permis de découvrir des profils de fixation spécifiques selon les HBGAs. Des alignements de séquences ainsi que la réalisation de divers mutants de la protéine de capside ont permis de proposer une poche de fixation potentielle du virus. Cependant, des études d'interaction capside-antigènes de groupes sanguins assistées par ordinateur n'ont jamais été rapportées, ce virus constituant un challenge pour les méthodes d'amarrage moléculaire de par la taille de sa capside.

En considérant ces différents profils d'interaction, il se pourrait que la protéine de capside possède plusieurs poches de fixation préférentielles selon l'antigène de groupe sanguin d'un

individu qu'il serait actuellement particulièrement intéressant de découvrir. De plus, il serait très enrichissant d'établir des hypothèses quant à l'interaction sucres/virus au niveau moléculaire afin de mieux comprendre le déroulement de la reconnaissance virale. L'hypothèse selon laquelle le virus reconnaîtrait l'antigène de groupe A en grande partie par le biais du groupement N-acétyl semble par exemple très probable lorsque l'on considère que l'antigène B, qui ne possède pas ce groupe, n'est que très peu reconnu.

Afin de remplir ces deux objectifs, une étude des interactions moléculaires assistée par ordinateur a été réalisée en utilisant le programme AutoDock 3.0. Les données ont ensuite été analysées avec le logiciel SYBYL® (TRIPOS™).

Les antigènes de groupe sanguin : H type 1 d'une part et l'antigène de groupe A d'autre part sont les deux sucres les mieux reconnus par le virus de Norwalk, pour cette raison ils ont été utilisés comme ligand vis-à-vis du virus et respectivement représentés par les trisaccharides suivants : l'α-L-Fuc(1-2)ß-D-Gal(1-3)ß-D-GlcNAc et l'α-D-GalNAc(1-3)[α-L-Fuc(1-2)]ß-D-Gal.

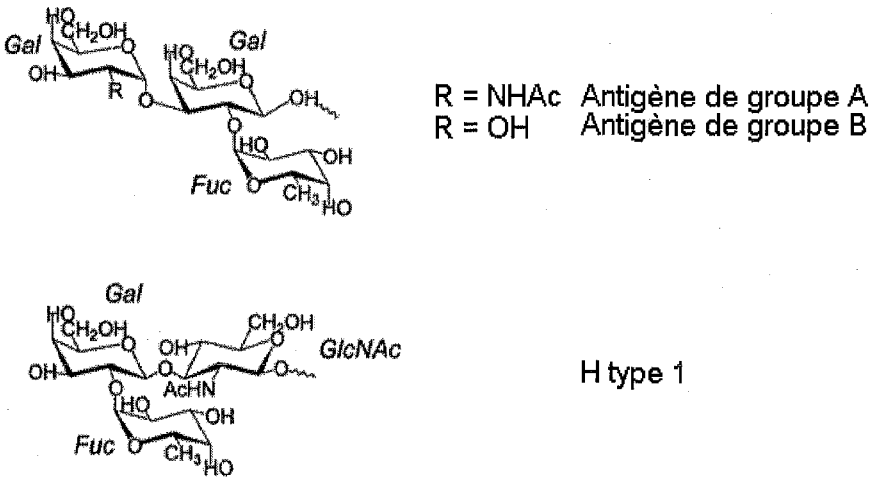



***Figure n°15***
*Selon Imberty et al. (2000)*
*Structure des deux trisaccharides utilisés pour l'amarrage moléculaire sur le virus de Norwalk*

# MATERIEL ET METHODES

## 1- Alignement de séquences

Un alignement de plusieurs souches virales a été réalisé afin de comparer celles ayant le même profil de fixation et d'essayer de visualiser des résidus conservés (ou proches sur le plan physicochimique) pouvant indiquer un site de reconnaissance du récepteur.

Les séquences ont été récupérées à partir du site d'Infobiogen (voir ANNEXE B/ p103). L'alignement a, dans un premier temps, été réalisé avec les logiciels ClustalX et GeneDoc. Il a ensuite été nécessaire de faire une analyse des familles hydrophobes (HCA) à cause de la difficulté d'alignement de certaines souches étant donné leur faible pourcentage d'identité.

Les méthodes d'alignements, comme ClustalX ou GeneDoc, sont basées sur la détection du pourcentage d'identité maximal entre les séquences étudiées. Mais même si la similarité est moins marquée, cela n'implique pas forcément d'importantes différences sur la structure tridimensionnelle et le repliement des protéines. L'analyse des familles hydrophobes utilise une représentation en hélice-α à deux dimensions de la séquence protéique. Cette technique repose sur l'hypothèse que le polypeptide naissant à partir du ribosome serait structuré en hélice-α avant son repliement. Bien que ce mécanisme ne soit pas prouvé, cette cartographie permet de faire une comparaison visuelle plus facile et plus efficace des déterminants structuraux dans la séquence par rapport aux méthodes linéaires [10].

## 2- Modélisation des carbohydrates

Il est indéniable que les protéines constituent un challenge pour la modélisation assistée par ordinateur. La description des mouvements des hélices et rubans en terme de repliement seulement peut déjà facilement dépasser les capacités d'un superordinateur. Cependant, la complexité des protéines n'est rien en comparaison de celle des carbohydrates. Bien que les protéines solubles soient souvent composées de milliers d'atomes formant une structure

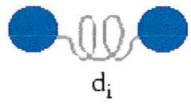
élaborée, elles tendent à conserver leur structure globulaire. Les carbohydrates, en contraste, sont flexibles et fréquemment branchés et la modélisation de leur comportement au cours de simulations par ordinateur est en conséquence particulièrement difficile.

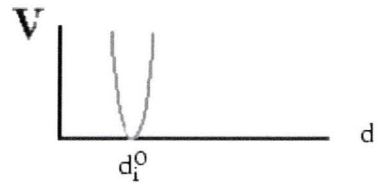
Le logiciel SYBYL de TRIPOS™, Inc a été utilisé pour faire la modélisation moléculaire. Ce logiciel possède de nombreux modules permettant, par exemple, de faire de la conception rationnelle de molécules, du criblage virtuel, de la chimie combinatoire et de la biologie structurale.

Afin d'évaluer la structure d'une molécule, des fonctions représentant l'énergie de celle-ci sont utilisées. Bien que tout indiquée pour cela, la mécanique quantique ne peut être utilisée pour modéliser de gros systèmes à cause des limites informatiques actuelles. Le compromis est alors d'utiliser la physique classique afin de déterminer les fonctions d'énergie ou champ de force. La mécanique moléculaire aussi désignée sous le nom de « méthode à champ de force » calcule l'énergie d'un système comme une fonction des positions des noyaux de chaque atome, les électrons étant implicites contrairement à la mécanique quantique. Un champ de force comprend plusieurs termes traduisant d'une part les interactions liantes entre atomes liés de manière covalente et d'autre part les interactions non-liantes et permet de déterminer quelles sont les conformations préférentielles d'une molécule. SYBYL est le seul logiciel à posséder le champ de force TRIPOS développé en majeure partie pour la modélisation de petites molécules organiques.

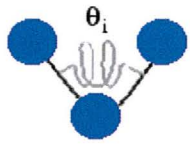
## Champ de force TRIPOS

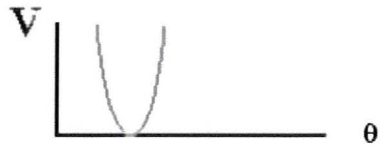
$$E_{str} = \sum_{i=1}^{Nbonds} \frac{1}{2} k_i^d (d_i - d_i^0)^2$$



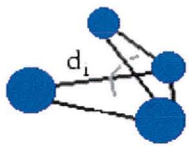



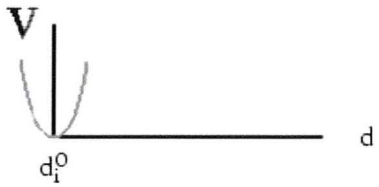

$$E_{bend} = \sum_{i=1}^{Nangles} \frac{1}{2} k_i^\theta (\theta_i - \theta_i^0)^2$$



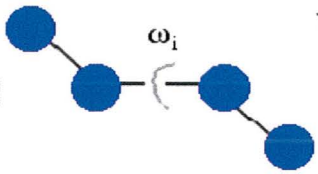



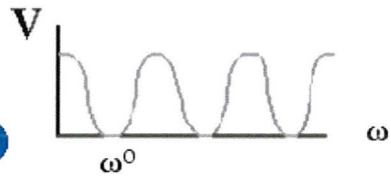

$$E_{oop} = \sum_{i=1}^{Noop} \frac{1}{2} k_i^{oop} d_i^2$$



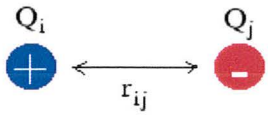




$$E_{tor} = \sum_{i=1}^{Ntor} \frac{1}{2} V_i^\omega [1 + S_i \cos(|n_i| \cdot \omega_i)]$$




$$E_{ele} = 332.17 \times \sum_{i=1}^{Natoms} \sum_{j>1}^{\text{non bonded}} \frac{Q_i Q_j}{D_{ij} r_{ij}}$$



$$E_{vdw} = \sum_{i=1}^{Natoms} \sum_{j>1}^{\text{non bonded}} E_{ij} \left[ \frac{1.0}{a_{ij}^{12}} - \frac{2.0}{a_{ij}^{6}} \right]$$

***Figure n°16***
*Présentation des différents termes intervenant dans le calcul du champ de force TRIPOS (voir la partie Abréviations)*

$$E = \sum E_{str} + \sum E_{bend} + \sum E_{oop} + \sum E_{tors} + \sum E_{vdw}$$
$$+[\sum E_{ele} + \sum E_{dist_c} + \sum E_{ang_c} + \sum E_{tors_c} + \sum E_{range_c}$$
$$+\sum E_{multi} + \sum E_{fieldfit}]$$

***Figure n°17***
*Equation permettant d'évaluer l'énergie totale de la molécule.*
*(voir la partie Abréviations)*

Le champ de force TRIPOS a été adapté pour les carbohydrates au sein du laboratoire. Il peut maintenant être utilisé pour les sucres lorsque les charges PIM sont attribuées à chacun de ses atomes.

Chaque minimisation ou recherche du minimum énergétique pour déterminer la conformation la plus probable d'une molécule ou d'un complexe a été réalisée avec la méthode POWELL.

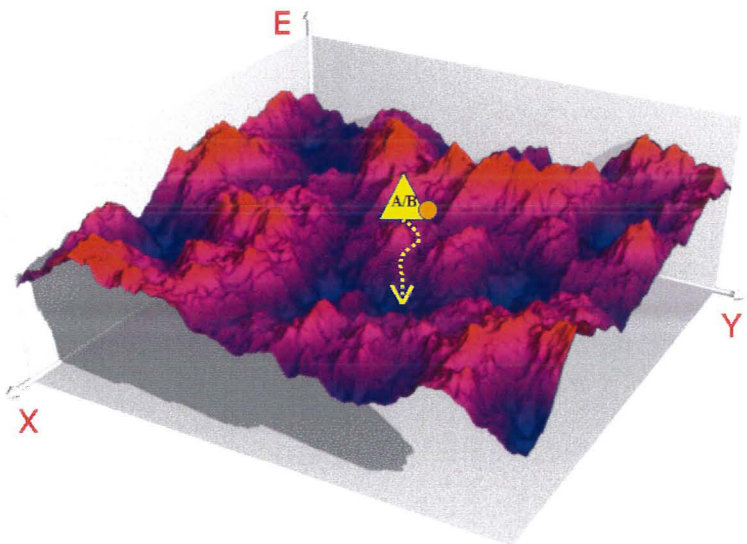



***Figure n°18***
*Paysage énergétique : représentation tridimensionnelle de l'énergie d'un complexe*

Cette méthode appartient à la famille des gradients conjugués. Elle utilise cependant des règles plus avancées pour déterminer la direction de la descente. Elle est également plus tolérante et trois fois plus rapide que la méthode des gradients conjugués. Cette procédure consiste en une série de recherches linéaires à partir de la position de la protéine dans son paysage énergétique. Une suite de pas est effectuée dans une direction choisie de l'espace tridimensionnel [23].

## 3- Amarrage moléculaire

Au cours de ces dernières années, des efforts considérables ont été faits afin de développer des méthodes automatisées pour comprendre et prédire la façon dont les protéines interagissent avec leur ligand. Cet emballement est en grande partie dû à l'intérêt indéniable pour la recherche et l'industrie pharmaceutique. Comprendre et prédire la façon dont les lectines interagissent avec leur ligands oligosaccharidiques devrait permettre, à terme, d'interférer avec le processus biologique dans lequel ces protéines sont impliquées.

L'amarrage moléculaire consiste à analyser les interactions entre une protéine et son ligand. Cet amarrage se déroule en trois étapes : la première consiste à placer le ligand vis-à-vis de son récepteur, elle est suivie par une exploration spatiale des degrés de liberté du ligand afin d'identifier les conformations qui s'adaptent le mieux au récepteur : c'est l'échantillonnage de l'espace des configurations du complexe ligand-récepteur. Plusieurs méthodes existent pour effectuer cette deuxième phase : des méthodes géométriques de reconnaissance de motifs pour apparier des descripteurs du ligand et du récepteur basées sur des propriétés physico-chimiques, pharmacophoriques... La dynamique moléculaire peut aussi être utilisée tout comme les algorithmes génétiques. La troisième phase permet d'attribuer un score à chaque conformation trouvée en fonction de son énergie d'interaction avec la protéine et d'établir un classement des meilleures solutions au problème posé.

## 4- Algorithme génétique

L'algorithme génétique (GA) se définit comme un programme informatique qui exploite les principes de l'évolution de la biologie tels que la transmission des gènes, les mutations et la recombinaison génétique pour proposer des solutions à des problèmes de grande complexité. Ce problème à résoudre est décrit comme une liste de paramètres, appelés chromosomes ou génome, qui peuvent être utilisés en vue d'effectuer une procédure d'évaluation.

Dans le cas de l'amarrage moléculaire, le ligand, appelé individu, prend une position particulière par rapport à la protéine à laquelle il est lié. Cette position peut être définie comme un ensemble de valeurs comme son orientation et sa conformation qui représentent, dans l'algorithme génétique, des variables appelées gènes. L'état du ligand correspond alors au génotype et ses coordonnées atomiques au phénotype de l'individu. Pour AutoDock 3.0, il existe 3 gènes de translation : x, y et z, coordonnés dans l'espace, et 4 gènes définissant l'orientation du ligand représentés par un quaternion, un angle de rotation et deux angles de torsion [19].

Au départ, un grand nombre de chromosomes est créé de façon randomisée pour constituer une population d'individus. Une étape de sélection va permettre de déterminer quels individus vont se reproduire en fonction de leur forme physique. Les individus possédant une très bonne santé ont proportionnellement plus de chance d'être parents. A chaque génération, la forme physique de chaque organisme est ainsi évaluée comme l'énergie d'interaction entre le ligand et la protéine. Seuls les individus ayant la meilleure forme physique constituant les meilleures solutions au problème posé vont pouvoir devenir parents. Autodock génère de la diversité par cette repoduction en respectant les règles de la génétique classique. Le programme utilise également des opérateurs tels que les crossovers et les mutations.

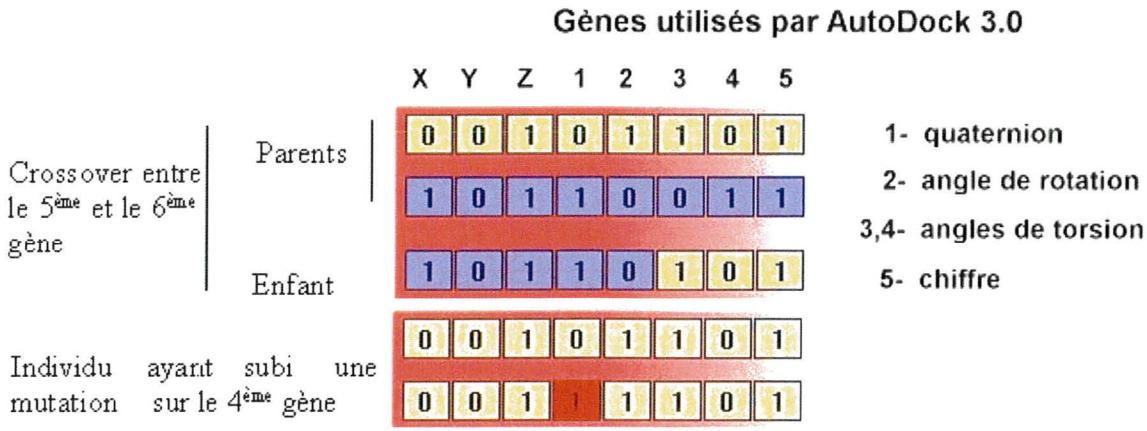



*Figure n°19*

Les crossovers et mutations sont appliqués à un certain nombre d'individus choisis au hasard au sein de la population selon un taux défini dès le départ. Les crossovers sont créés en cassant le génome entre deux gènes et non dans un gène afin d'éviter les changements erratiques dans leur valeur. Ainsi, chaque parent verra son génome cassé en 3 morceaux aux mêmes positions avant la recombinaison. Le crossover est l'élément clé qui distingue l'algorithme génétique d'autres méthodes comme le « passage de barrières » ou le « recuit-simulé » (SA). Il permet en effet à chaque individu d'explorer l'espace tridimensionnel à proximité immédiate avec référence à ce que d'autres individus ont découvert. Les enfants remplacent ensuite leurs parents dans la population de façon à maintenir le nombre d'individus constant et seul l'individu possédant la meilleure santé peut survivre à la génération suivante. Le processus de reproduction se poursuit jusqu'à ce qu'une des deux conditions d'arrêt de la recherche soit remplie, la première correspond au nombre d'évaluation d'énergie et la deuxième : au nombre de générations.

## 5- Algorithme génétique lamarckien

Les GAs miment, pour la plupart, les principales caractéristiques de l'évolution décrite par Darwin et appliquent la génétique selon Mendel : le génome détermine le phénotype. Cependant, il est parfois possible d'utiliser des fonctions mathématiques permettant de déduire le génotype d'un individu à partir de son phénotype : et ainsi de réaliser des recherches locales pour ensuite remplacer un individu par le résultat de cette recherche. Cette étape de recherche locale est caractéristique des GAs dit lamarckiens, dont AutoDock fait partie. Ils sont en effet basés sur la théorie de Jean-Baptiste Lamarck, célèbre naturaliste français, selon laquelle les caractéristiques phénotypiques acquises pendant la vie d'un individu peuvent être transmises à sa descendance.

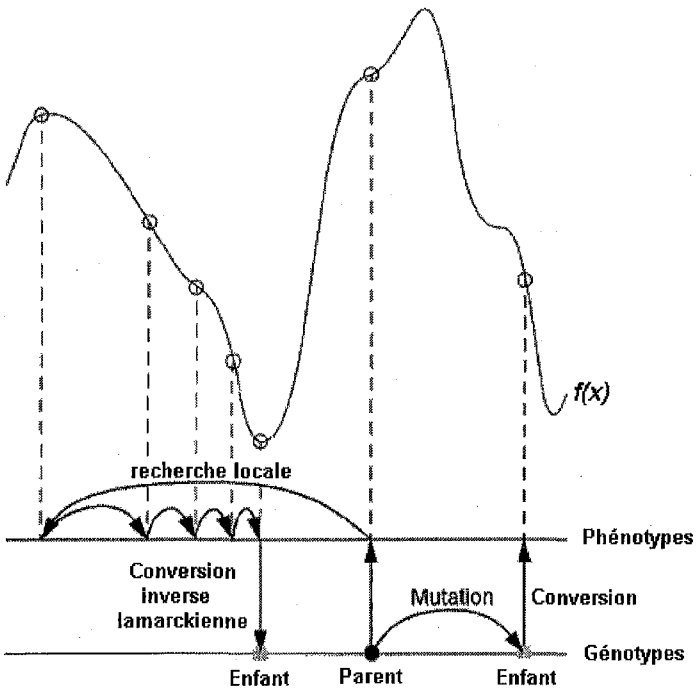



***Figure n°20***
*Selon Morris et al. (1998)*

La santé de chaque individu soumis au champ de force implémenté dans le programme est évaluée grâce à une fonction de score. Dans AutoDock, cette fonction empirique est un modèle ayant pour but d'apprécier l'énergie libre de liaison du ligand à son récepteur. Afin

d'estimer cette énergie libre de Gibbs, de nombreux facteurs doivent être considérés. AutoDock prend en compte la dispersion/répulsion des deux molécules, la formation des liaisons hydrogènes, les interactions électrostatiques aussi bien que les changements dans les longueurs et les angles des liaisons. La rotation globale, les translations, la désolvatation et l'effet hydrophobe sont également examinés.

$$\Delta G = \Delta G_{\text{vdw}} + \Delta G_{\text{hbond}} + \Delta G_{\text{elec}} + \Delta G_{\text{conform}} + \Delta G_{\text{tor}} + \Delta G_{\text{sol}}$$

$$\Delta G = \Delta G_{\text{vdW}} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{\text{hbond}} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \Delta G_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}$$

$$+ \Delta G_{\text{tor}} N_{\text{tor}} + \Delta G_{\text{sol}} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2 / 2\sigma^2)}$$

***Figure n°21***
*Selon Morris et al. (1998)*

# RESULTATS

## 1- HEMAGGLUTININE DU VIRUS DE LA GRIPPE

### 1.1- Structures

La plus connue des lectines virales correspond à l'hémagglutinine, lectine du virus de la grippe. La première partie de mon travail de thèse a été de mettre à jour la famille des hémagglutinines dans la banque de données des lectines. Aujourd'hui, 18 nouvelles structures différentes d'hémagglutinines sont accessibles depuis le site Internet du C.E.R.M.A.V. en plus des 11 qui étaient déjà présentes.

***Tableau IV.*** *Hémagglutinines accessibles depuis le site internet du C.E.R.M.A.V.*

| ***Code PDB*** | ***Résolution (Å)*** | ***Complexe*** | ***Année*** |
|---|---|---|---|
| 1EO8 | 2,8 | Anticorps neutralisant | 2000 |
| 1HGD | 2,7 | / | 1992 |
| 1HGE | 2,6 | α–Me-NeuAc | 1992 |
| 1HGF | 3 | / | 1992 |
| 1HGG | 2,9 | α-2,3-sialyllactose | 1992 |
| 1HGH | 2,7 | α-2-O-(4'-benzylamidocarboxybutyl)-5-N-AcNeuAc | 1992 |
| 1HGI | 2,7 | 4-O-Ac-α-MeNeuAc | 1992 |
| 1HGJ | 2,7 | 9-amino-9-deoxy-α-Me-NeuAc | 1992 |
| 2HMG | 3 | / | 1990 |
| 3HMG | 2,9 | / | 1990 |
| 4HMG | 3 | NeuAc | 1990 |
| 5HMG | 3,2 | NeuAc | 1990 |
| 1HTM | 2,5 | / | 1994 |
| 1KEN | 3,5 | Anticorps | 2002 |
| 1MQM | 2,6 | Lstc[1] (αNeuAc26βGal14βGlcNAc13βGαl14Glc) | 2003 |

| *Code PDB* | *Résolution (Å)* | *Complexe* | *Année* |
|---|---|---|---|
| 1MQN | 3,2 | Lstc | 2003 |
| 1QFU | 2,8 | Anticorps neutralisant | 1999 |
| 1RD8 | 3 | / | 2004 |
| 1RU7 | 2,3 | / | 2004 |
| 1RUZ | 2,9 | / | 2004 |
| 1RV0 | 2,5 | Lsta[2] (αNeuAc23βGal13βGlcNac13βGal14Glc) | 2004 |
| 1RVT | 2,5 | Lstc | 2004 |
| 1RVX | 2,2 | Lsta | 2004 |
| 1RVZ | 2,25 | Lstc | 2004 |
| 2VIR | 3,25 | Anticorps neutralisant | 1998 |
| 2VIS | 3,25 | Anticorps neutralisant | 1998 |
| 2VIT | 3,25 | Anticorps neutralisant | 1998 |
| 2VIU | 2,5 | / | 1998 |

[1] Lstc : oligosaccharide modèle du récepteur humain pour le virus de la grippe
[2] Lsta : oligosaccharide modèle du récepteur aviaire pour le virus de la grippe

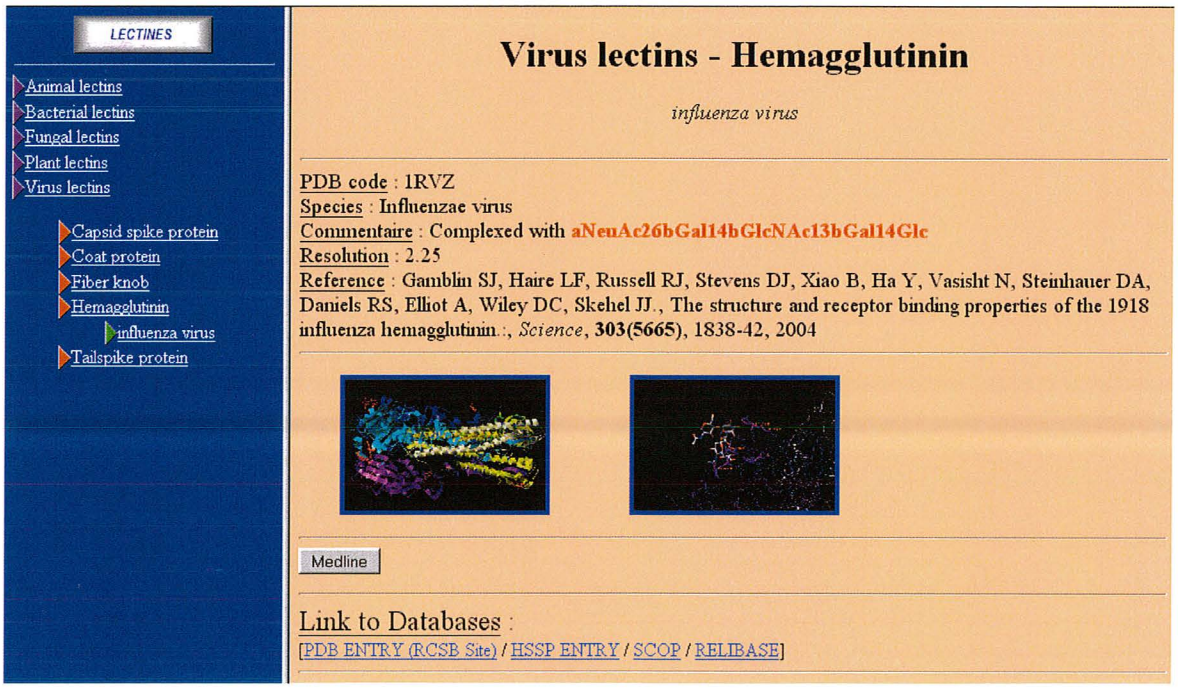



***Figure n°22***
*Capture d'écran de la banque de données des lectines accessible depuis le site internet du C.E.R.M.A.V.*

Ces lectines ont été cristallisées seules ou co-cristallisées avec des anticorps ou des oligosaccharides et leur structure a été résolue par cristallographie. Le site propose aussi des liens vers Medline ou la Protein Data Bank (PDB) correspondant à la structure considérée. Des images sont également disponibles afin de visualiser rapidement la structure de la lectine et de ses ligands éventuels.

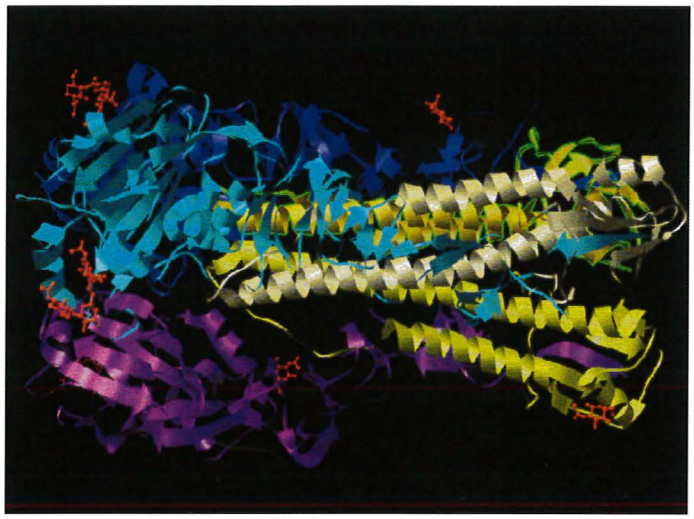


***Figure n°23***
*Exemple de structure d'hémagglutinine disponible depuis la banque.*

### 1.2- Modélisation d'un pentasaccharide Lstc

La structure de départ 1RVT (PDB) correspond à la structure co-cristallisée de l'hémagglutinine de porc (1930-swine) avec le fragment Lstc en conformation replié comme modèle du récepteur humain.

Afin d'optimiser l'inhibition de l'hémagglutinine par les sialopeptides sous forme de tétrahédrons, la conception d'une conformation sous contrainte du Lstc assistée par ordinateur a été réalisée afin de forcer le pentasaccharide à adopter une conformation repliée quelle que soit la structure du tecton. Le logiciel SYBYL a permis de construire deux modèles du pentasaccharide pontés au niveau des atomes O3 et O7 des résidus 3-GlcNAc et 1Neu5Ac

respectivement. Le premier modèle correspond à l'introduction d'un pont à 1 carbone (Lstc b1C) tandis que le deuxième possède une structure pontée à 2 atomes de carbone (Lstc b2C). A partir du pentasaccharide de la structure 1RVT, plusieurs angles ont été modifiés afin de rendre la conformation de Lstc plus probable. La fonction amide de l'acide sialique a été mise en position trans et les angles de torsion C4-C3-O3-H des résidus 3 et 5-GlcNAC ont été ramenés à 0°.

Pour chacun des modèles, une rapide minimisation de la nouvelle chaîne carbonée uniquement a été réalisée afin de permettre le réajustement des angles et des longueurs de liaisons. Une minimisation a ensuite été réalisée avec les atomes d'hydrogène du sucre libre. Enfin une dernière minimisation est effectuée avec le sucre et les atomes d'hydrogène de l'hémagglutinine libres.

Les paramètres des minimisations effectuées sont toujours identiques : le champ de force TRIPOS est appliqué, on utilise la méthode POWELL avec un gradient de 0,1 kcal.mol$^{-1}$.A$^{-1}$), la constante diélectrique est fixée à 4.

***Tableau V.*** *Energies des différents modèles de Lstc.*

| | **Agrégat** | **Lstc** | **Lstc b1C** | **Lstc b2C** |
|---|---|---|---|---|
| **Energie d'élongation de liaison** | 407.021 | 41.760 | 41.789 | 42.258 |
| **Energie de déformation d'angle** | 672.144 | 438.059 | 442.859 | 447.389 |
| **Energie de torsion** | 339.320 | 577.136 | 585.721 | 581.899 |
| **Energie de déformation de plan** | 8.956 | 82.562 | 82.586 | 82.542 |
| **Energie de Van Der Waals 1-4** | -117.346 | 814.603 | 815.050 | 812.864 |
| **Energie de Van der Waals** | -1160.881 | -331.019 | -329.998 | -330.311 |
| **Energie électrostatique 1-4** | 779.435 | -7.638 | -8.819 | -8.027 |
| **Energie électrostatique** | 2032.810 | -3012.010 | -3007.335 | -3008.805 |
| **Energie totale** | 2961.459 kcal.mol$^{-1}$ | 1564.913 kcal.mol$^{-1}$ | 1583.312 kcal.mol$^{-1}$ | 1581.269 kcal.mol$^{-1}$ |

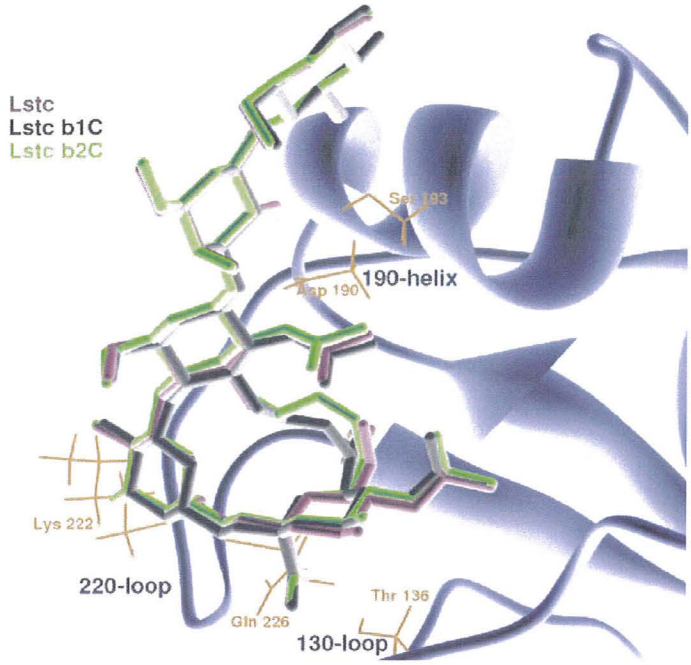



***Figure n° 24***
*Vue du Lstc natif superposé avec les deux Lstc construits, le premier avec un pont à un atome de carbone et le second à deux atomes de carbone.*

Au vu des résultats sur le plan énergétique comme sur le plan conformationnel, il semble tout à fait possible de créer un Lstc sous contrainte par introduction d'un pont carboné qui permettrait peut-être par la suite d'augmenter l'affinité du tectomère pour l'acide sialique et ainsi bloquer les hémagglutinines présentes à la surface du virus de la grippe. D'après les énergies des deux conformations sous contrainte, les deux structures sont moins stables que la structure native ce qui est normal étant donné que ces deux modèles subissent des contraintes. Cependant, le Lstc possédant un pont à deux atomes de carbone présente une meilleure énergie, sa conformation est donc préférentielle en comparaison avec celle avec un pont à un atome de carbone. En superposant les deux modèles avec la forme native, les deux structures adoptent des conformations très proches de celle du Lstc natif. Comme l'acide sialique est considéré comme étant l'unique carbohydrate par lequel se fait la reconnaissance de nos cellules par l'hémagglutinine, il est particulièrement important que sa structure change le moins possible. Les acides sialiques des deux constructions n'ont presque pas été modifiés par

rapport à la structure de départ ce qui laisse à penser que ces Lstc pontés peuvent encore être reconnus par l'hémagglutinine. Comme les deux modèles adoptent des structures quasiment identiques dans l'espace et semblent en conséquence équivalents, la meilleure construction n'est déterminée que sur des critères énergétiques. Ainsi, le Lstc ponté à deux atomes de carbone est plus stable, il doit être synthétisé et faire l'objet d'analyses plus poussées.

# 2- MODELISATION DE LA CAPSIDE DU VIRUS DE NORWALK AVEC DEUX OLIGOSACCHARIDES

## 2.1- Alignement de séquences : avec GeneDoc, ClustalX et analyse des familles hydrophobes

Les quatorze protéines de capsides qui ont été alignées appartiennent toutes à des virus de la classe des Norovirus comme le virus de Norwalk, cependant toutes ces souches ne présentent pas le même profil de fixation aux antigènes de groupe sanguin. Ainsi l'alignement a été effectué afin de déterminer des régions non conservées qui pourraient être à l'origine de ces différences de reconnaissance. (références des souches en ANNEXE B/ p103)

***Tableau VI.*** *Souches virales utilisées pour l'alignement de séquences avec leurs caractéristiques de fixation aux antigènes de groupes sanguins.*

| ***Souches virales*** | ***Résidus identiques par rapport au NV*** | ***Non sécréteurs*** | ***O sécréteurs*** | ***A sécréteurs*** | ***B sécréteurs*** |
|---|---|---|---|---|---|
| VA387 | 42,00% | - | +++ | ++++ | ++++ |
| GrV | 42,00% | - | +++ | ++++ | ++++ |
| Norwalk | / | - | +++ | ++++ | + |
| Dijon | 42,00% | - | +++ | ++++ | ++++ |
| C59 | 67,00% | - | + | ++ | - |
| MxV | 43,00% | - | + | +++ | +++ |
| HV | 45,00% | - | - | ++ | - |
| MOH | 45,00% | - | - | ++++ | ++++ |
| SMV | 45,00% | - | - | - | +++ |
| VA207 | 43,00% | +++ | ++ | + | +/- |
| Boxer | 61,00% | ++ | +++ | + | +/- |

| *Souches virales* | *Résidus identiques par rapport au NV* | *Non sécréteurs* | *O sécréteurs* | *A sécréteurs* | *B sécréteurs* |
|---|---|---|---|---|---|
| DSV | 66,00% | - | - | - | - |
| VA115 | 65,00% | - | - | - | - |
| RHDV[1] | 14,00% | - | - | - | - |

[1] virus hémorragique du lapin (Rabbit Hemorragic Desease Virus)

Après alignement des séquences capsidiques des quatorze souches de Norovirus, il a été possible de faire une cartographie de chaque acide aminé du virus de Norwalk selon sa conservation. Le virus hémorragique du lapin n'a pas été pris en compte étant donné l'existence d'insertions trop importantes par rapport au NV. Le domaine P2 est bien constitué d'une série d'acides aminés très peu conservés d'une souche à l'autre.

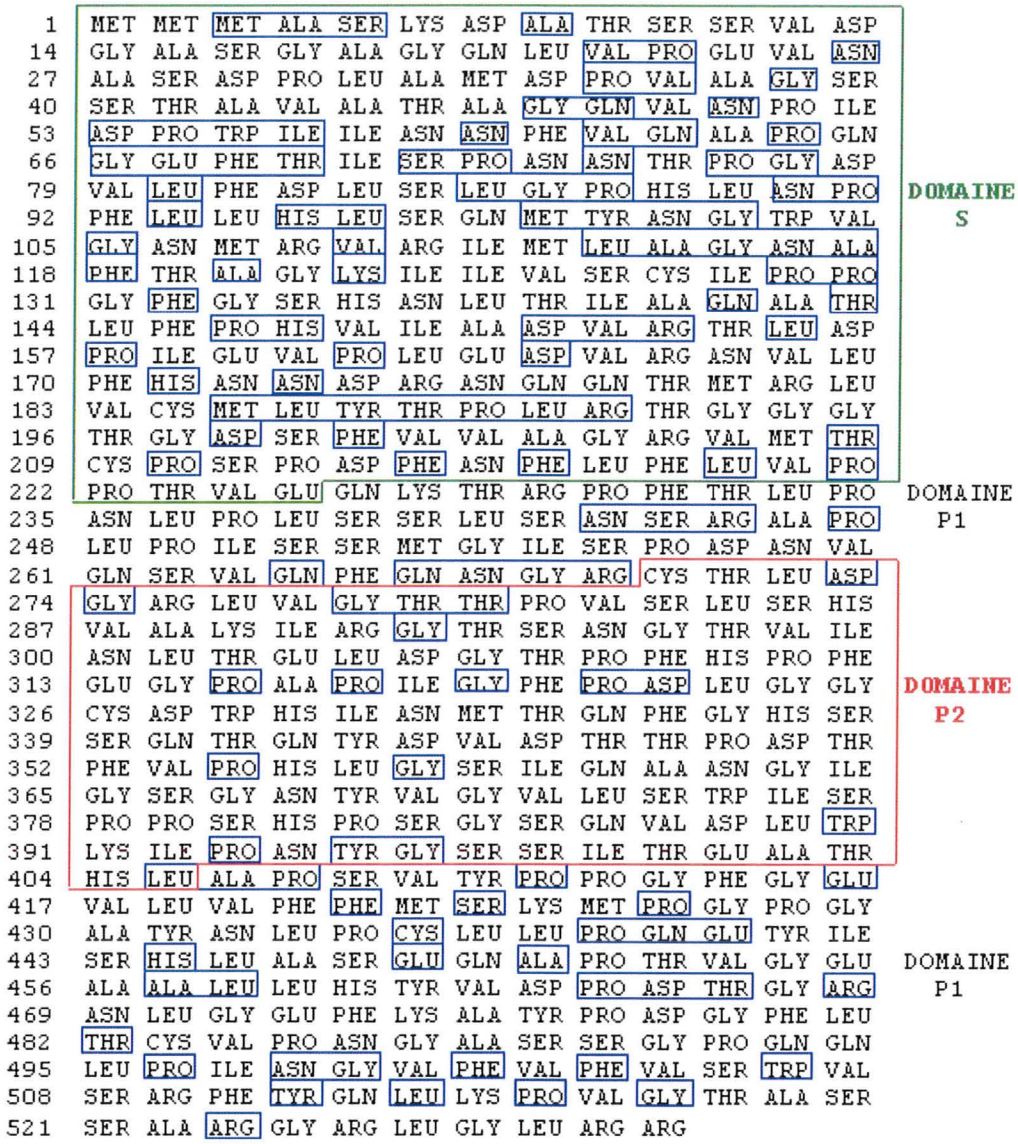



***Figure n°25***

*Séquence protéique de la capside virale du virus de Norwalk avec les domaines S (en vert), P1 et P2 (en rouge). Les acides aminés conservés à 100% dans toutes les souches sont encadrés en bleu.*

Une analyse des familles hydrophobiques a été réalisée pour aligner les séquences du virus de Norwalk et celles du virus hémorragique de lapin qui sont très dissemblables, ce dernier présentant d'importantes insertions dans son génome. La souche VA207 a été utilisée afin de faciliter l'alignement car elle semble être intermédiaire.

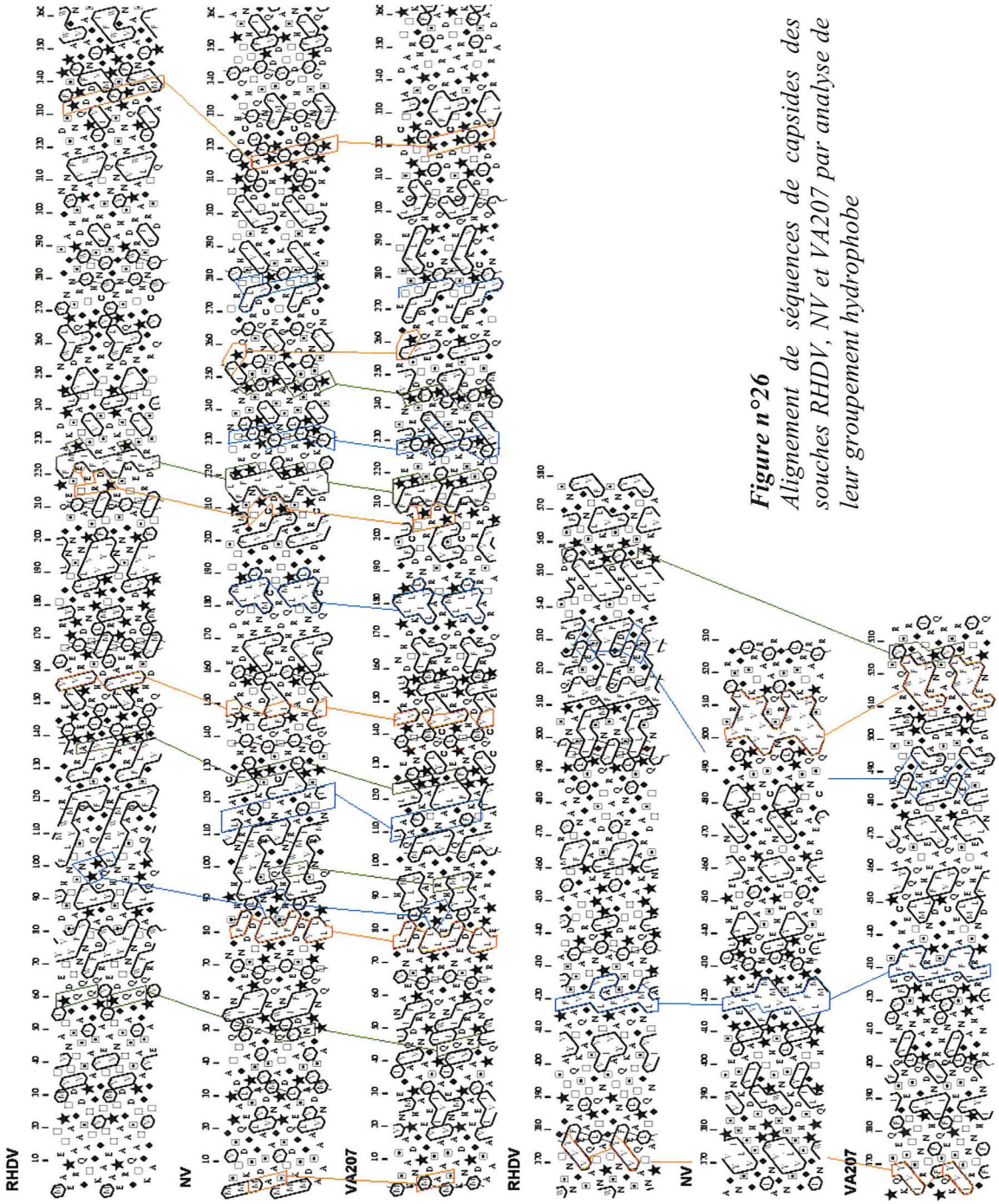



***Figure n°26***
*Alignement de séquences de capsides des souches RHDV, NV et VA207 par analyse de leur groupement hydrophobe*

## 2.2- Amarrage moléculaire

La capside complète du virus de Norwalk est très importante, bien trop grosse pour pouvoir réaliser les calculs d'amarrage moléculaire avec AutoDock sur son ensemble. Il a ainsi été décidé d'utiliser la plus petite entité structurale de la capside pour réaliser l'amarrage. Les

protéines de capside A, B et C étant identiques dans leur séquence, les dimères AB et CC ont été considérés comme quasi équivalents. Afin de diminuer les temps de calculs et d'interprétation des résultats, seul le dimère AB a été retenu pour l'étude.

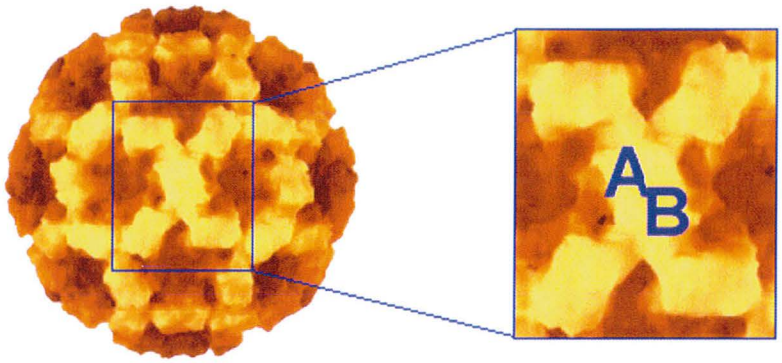



***Figure n°27***
*A gauche : vue de la totalité de la capside virale; à droite : grossissement sur un dimère AB formant une protrusion à la surface du virus.*

Les coordonnées du complexe AB ont été obtenues en utilisant le générateur d'oligomères sur le site de VIPER qui les a calculées à partir de la structure cristalline du virus (code PDB 1IHM). Dans un premier temps, les étapes de préparation des protéines et sucres ont été réalisées avec SYBYL : les atomes d'hydrogène ont été rajoutés, puis les charges de type Pullman ont été attribuées à tous les atomes du dimère. Les charges partielles PIM ont été assignées à chacun des deux trisaccharides. Pour l'étape d'amarrage moléculaire, les grilles d'énergies d'interaction et potentiels électrostatiques pour chaque atome sonde (ici C, O, H et N; soit chacun des types atomiques composant le ligand) ont été calculés en utilisant le programme Autogrid d'AutoDock dans une boîte cubique de 60 Å de côté, taille maximale pour une grille définie par des points avec un intervalle régulier de 0,5 Å. L'emplacement de cette boîte sur la protéine est particulièrement important puisqu'elle définit la zone de l'étude. Ici, la boîte a été centrée sur la protéine de capside B pour limiter les temps de calcul, A et B étant quasiment identiques. La zone de chevauchement entre les deux protéines a

soigneusement été incluse à l'intérieur de la boîte qui recouvre les domaines P1 et P2. Le domaine S étant en grande partie rendu inaccessible par les protrusions, il est peu probable qu'il possède la région de reconnaissance responsable de la fixation [27].

Pour chaque calcul, la population de départ de 100 individus a évolué sur 3000 générations avec un seul survivant d'une génération à l'autre. Le taux de crossover est de 0,8 tandis que le taux de mutation est à 0,02. La limite des calculs d'énergie a été fixée à $20.10^6$ évaluations afin d'assurer 3000 générations de façon certaine ; le nombre de générations et d'évaluations étant les deux conditions d'arrêt de la recherche. (Voir fichier de paramétrage de l'amarrage moléculaire dans l'ANNEXE C/ p104)

### 2.2.1- Analyse spatiale des solutions trouvées par amarrage moléculaire

Pour l'amarrage moléculaire, AutoDock effectue une exploration des conformations que le ligand peut adopter en fonction de la surface de la protéine avec laquelle il interagit : c'est l'échantillonnage. Une visualisation de la conformation ainsi que de la dispersion des solutions sur le récepteur permet d'évaluer rapidement cette deuxième étape de l'amarrage.

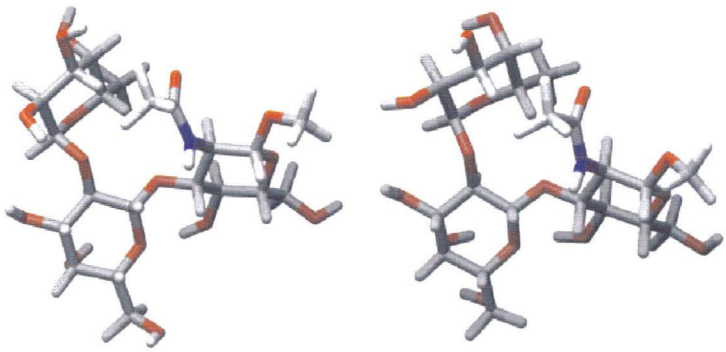


***Figure n°28***
*Exemple d'exploration des angles de torsion de deux solutions trouvées par AutoDock pour l'antigène H type 1. La partie N-acétyl glucosamine a été orientée de la même manière afin de faciliter la comparaison des deux conformations : les groupements N-acétyl et méthoxy étaient fixes pendant l'amarrage moléculaire et leur orientation n'a donc pas été modifiée. Les liaisons glycosidiques présentent des angles presque identiques, en revanche l'orientation des groupements hydroxyls est très variable.*

En analysant rapidement la conformation adoptée par les solutions proposées par AutoDock, l'exploration des différents angles de torsion qui ont été laissés libres de tourner pendant l'étape d'amarrage moléculaire s'est déroulée correctement. En effet, les angles de torsion des liaisons glycosidiques sont modifiés d'une solution à une autre et les angles de torsion des groupements hydroxyls montrent des variations très importantes.

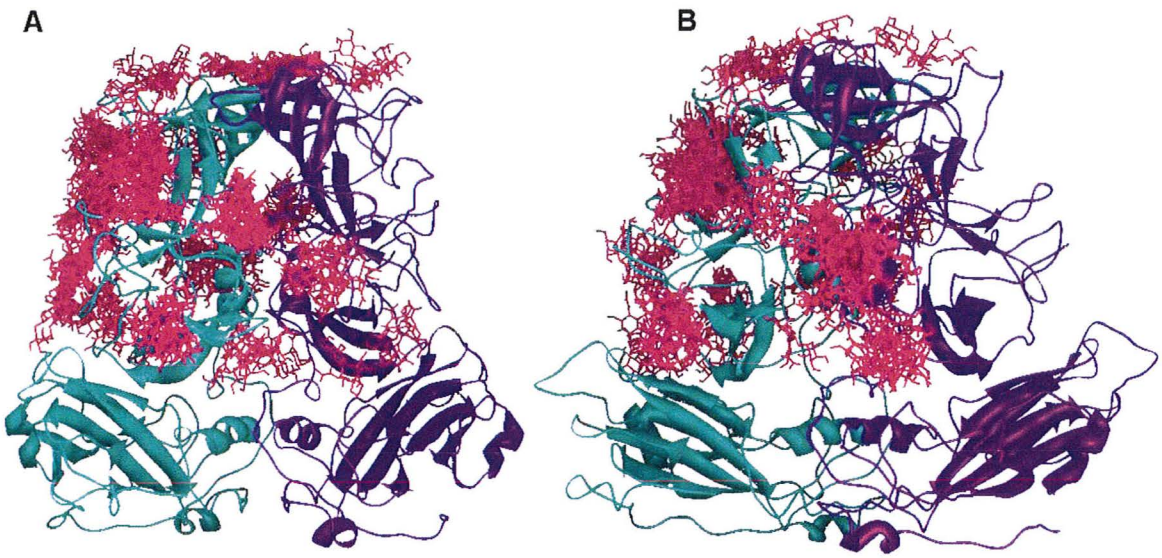



***Figure n°29***
*Représentation des solutions trouvées par AutoDock sur le dimère AB.* **A.** *Solutions pour le trisaccharide représentant l'antigène de groupe sanguin A,* **B.** *Solutions pour le trisaccharide représentant l'antigène de groupe sanguin O (H type 1).*

Deux cents solutions ont été obtenues pour chacun des deux types de trisaccharides. Une première analyse dans l'espace permet de dire que les solutions se sont bien réparties sur la surface du dimère dans les limites de la boîte d'amarrage moléculaire. L'échantillonnage est correct et AutoDock a bien exploré la surface. Les antigènes de groupe A et de groupe O adoptent une dispersion similaire par rapport au dimère bien que certaines zones soient peuplées différemment.

### 2.2.2- Familles déterminées par AutoDock

AutoDock classe les solutions d'amarrage moléculaire dans des familles en fonction de leur « Root Mean Square Deviation » (R.M.S.D.). Le R.M.S.D. est une déviation moyenne ou un écart moyen entre deux structures A et B d'une même molécule qui permet de comparer très facilement les deux conformations entre elles. Plus le R.M.S.D. est faible et plus les deux structures adoptent des conformations voisines dans l'espace. L'étude des familles déterminées par AutoDock permet donc une première analyse des poches potentielles de fixation des antigènes de groupes sanguins. Ces familles ayant été définies par rapport à un R.M.S.D. de 2.5Å, seules les solutions possédant des coordonnées tridimensionnelles très proches appartiennent à un même groupe. Ce R.M.S.D. est en revanche assez important pour tenir compte de la liberté de rotation des liaisons glycosidiques.

Pour des raisons de temps, les 200 calculs d'amarrage moléculaire ont été réalisés en deux fois pour chaque carbohydrate. En effet, pour obtenir 100 solutions avec les paramètres définis précédemment, l'étape d'amarrage moléculaire seule prend plus de 55h. Pour cela, les familles présentées ici correspondent à 100 solutions et non 200.

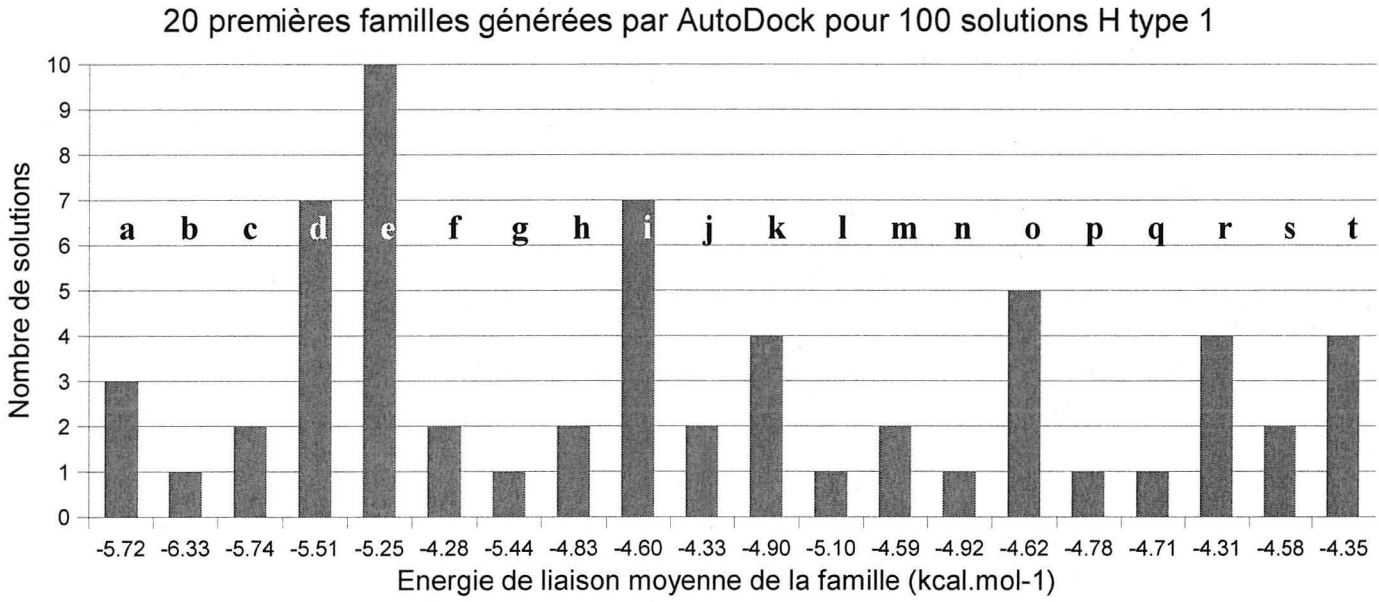



***Figure n°30***
*20 premières familles nommées de a à t pour le trisaccharide représentant l'antigène de groupe O classées dans l'ordre d'apparition des meilleures énergies d'interaction trouvées par AutoDock.*

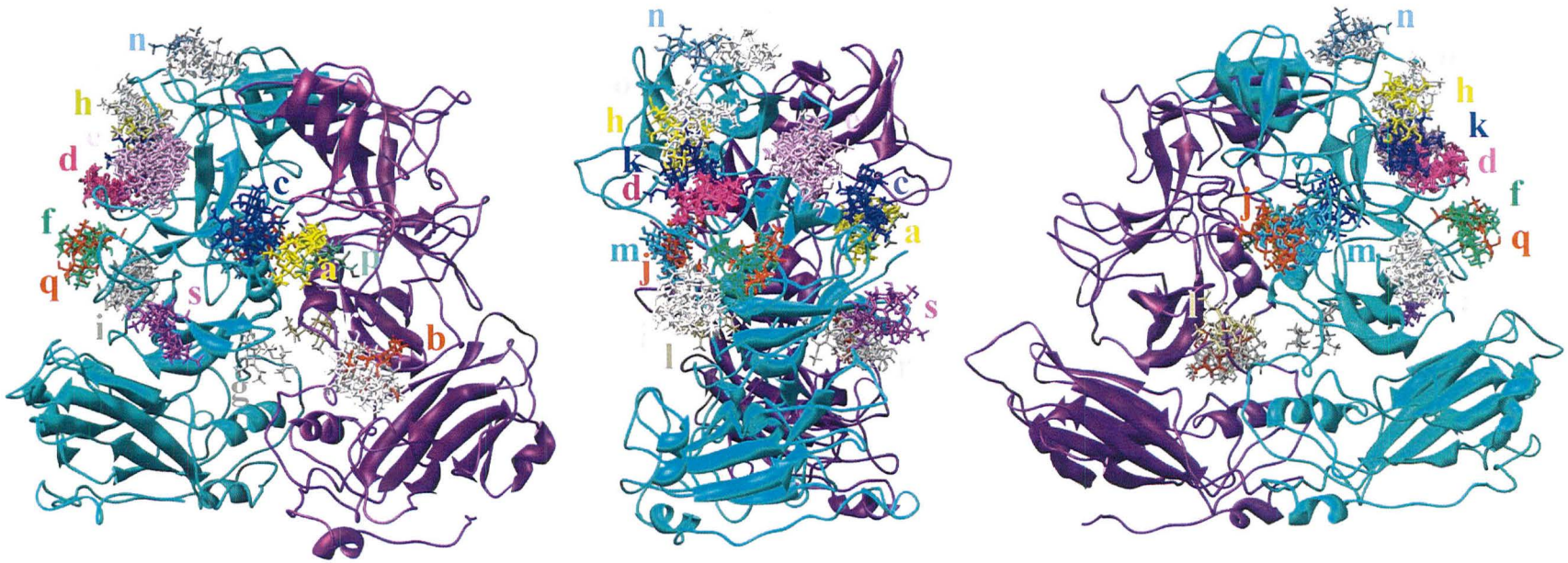



***Figure n°31***

*Vue de face , de profil et d'arrière du dimère avec les 20 premières familles issues de 100 calculs pour l'antigène de groupe sanguin O.*

*La protéine de capside B est représentée en vert tandis que celle de capside A est en violet.*

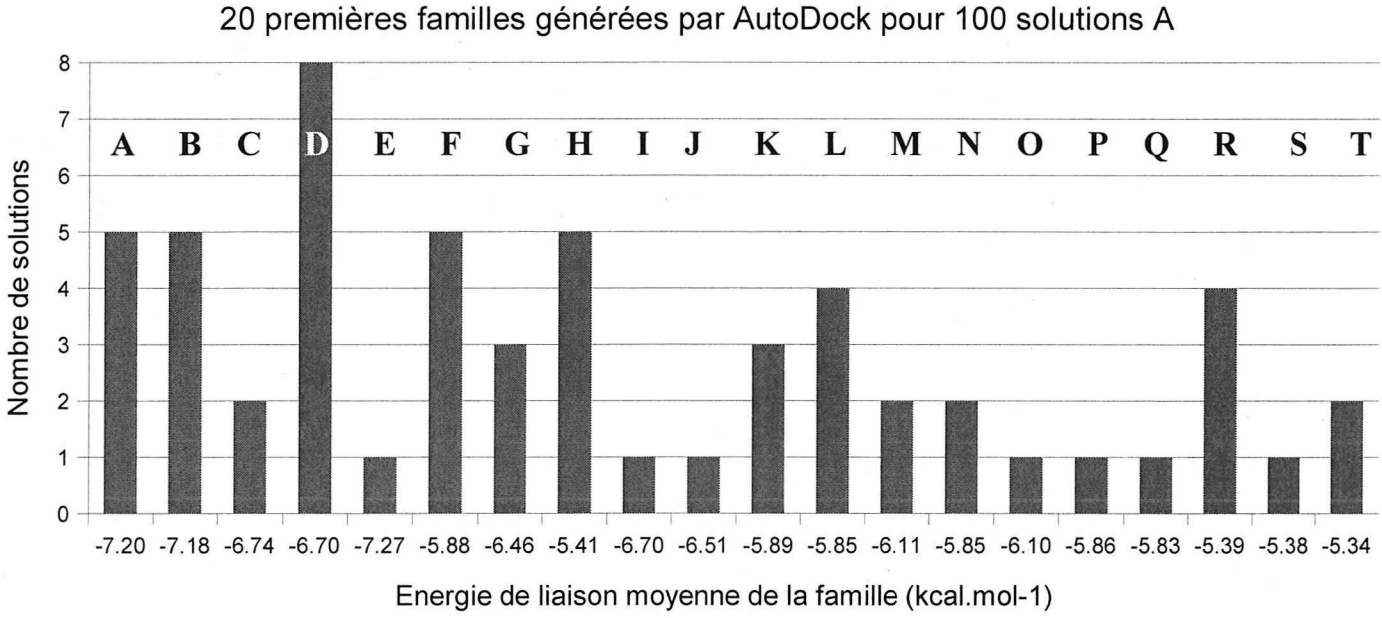
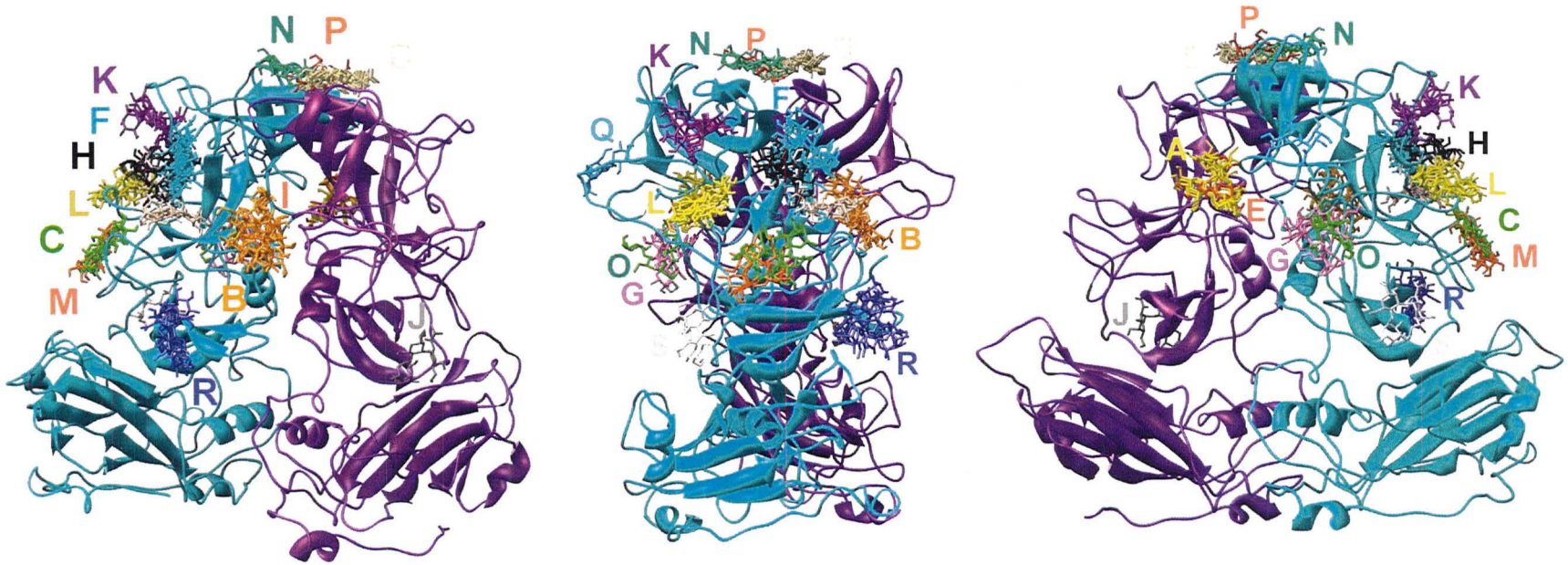



***Figure n°32***
*20 premières familles nommées de A à T pour le trisaccharide représentant l'antigène de groupe A classées dans l'ordre d'apparition des meilleures énergies d'interaction trouvées par AutoDock.*




***Fighure n°33***

*Vue de face , de profil et d'arrière du dimère avec les 20 premières familles issues de 100 calculs pour l'antigène de groupe sanguin A.*

*La protéine de capside B est représentée en vert tandis que celle de capside A est en violet.*

Ce sont les 20 premières familles qui rassemblent le plus de solutions et, quel que soit le trisaccharide considéré, le nombre de sucres par famille est d'autant plus important que les énergies de liaisons sont basses. Plus une famille contient de solutions, plus la zone qu'elle couvre est intéressante car elle a tendance à fixer les trisaccharides. Ainsi, cette première analyse permet de déterminer facilement plusieurs zones de fixation pouvant être intéressantes pour l'étude des interactions entre les antigènes de groupe sanguin A et O et les protéines de capside virale. Ainsi les familles **d** et **e** de l'antigène de groupe O qui sont particulièrement peuplées forment deux poches à étudier avec d'autant plus de soin qu'elles sont situées dans une même région. De même, les poches **A**, **B**, **D**, **F** et **G** sont à analyser pour l'antigène de groupe A.

### 2.3- Minimisation dans le champ de force TRIPOS

AutoDock permet de faire un amarrage moléculaire avec un ligand flexible, les angles de torsion définis au départ sont donc explorés pendant l'analyse de l'interaction avec le récepteur. Par contre, le récepteur est considéré comme rigide et les chaînes latérales à proximité du ligand n'ont pas été modifiées. Or, bien que les sites de fixation des lectines soient pré-organisés, plusieurs chaînes latérales peuvent changer considérablement pour favoriser la formation de liaisons hydrogènes et de contacts de Van Der Waals avec le ligand. Afin d'optimiser l'interaction entre les solutions obtenues et le dimère AB, des minimisations ont été effectuées dans le champ de force TRIPOS. De plus, ce champ de force possède des paramètres qui ont été développés spécialement pour les carbohydrates au sein du laboratoire contrairement à celui d'AutoDock, ainsi la conformation des solutions peut être améliorée par une telle minimisation.

Pour chacun des sucres, tous les angles de torsion susceptibles de tourner ont été laissés libres pendant l'amarrage moléculaire (liaisons glycosidiques, liaisons aux groupements hydroxyls, et groupements hydroxyméthyls) exceptées les liaisons aux groupements N-acétyl des GalNAc et GlcNAc et les liaisons oxyméthyls (groupement non présent dans les antigènes de groupe sanguin et rajouté pour mimer la liaison glycosidique aux autres carbohydrates formant les antigènes). Cependant les différents monomères d'un sucre sont connus pour adopter des conformations préférentielles [16].

Afin d'évaluer si les solutions une fois minimisées n'adoptent pas de structure qui serait peu probable, il a été nécessaire d'analyser les angles de torsion de chaque liaison glycosidique pour chacune des solutions trouvées pour ensuite les comparer avec leur cartes adiabatiques relaxées.

Les angles de torsion décrivant la liaison glycosidique se définissent comme suit :

$\Phi = O_5\text{-}C_1\text{-}O\text{-}C'_X$

$\Psi = C1\text{-}O\text{-}C'_X\text{-}C'_{X+1}$

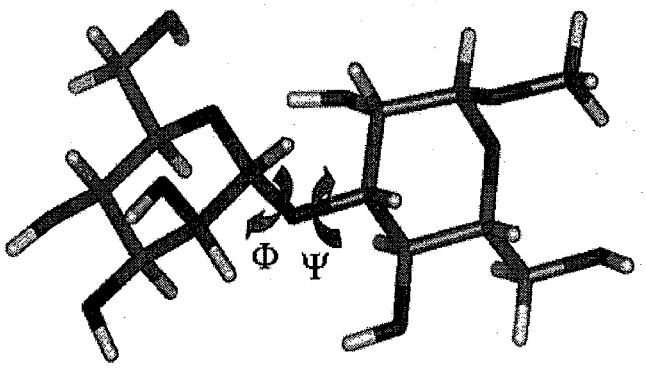



***Figure n°34***
*Angles de torsion Φ et Ψ*

avec les premiers atomes appartenant à la chaîne réductrice et la numérotation étant en accord avec les règles de nomenclature de l'I.U.P.A.C.

Afin d'évaluer rapidement l'effet des contraintes subies par le carbohydrate quand il est en interaction avec le dimère, une mesure des angles de torsion de chaque solution a été réalisée en utilisant SYBYL. (Voir fichier d'évaluation des angles de torsion dans l'ANNEXE F/ p108)

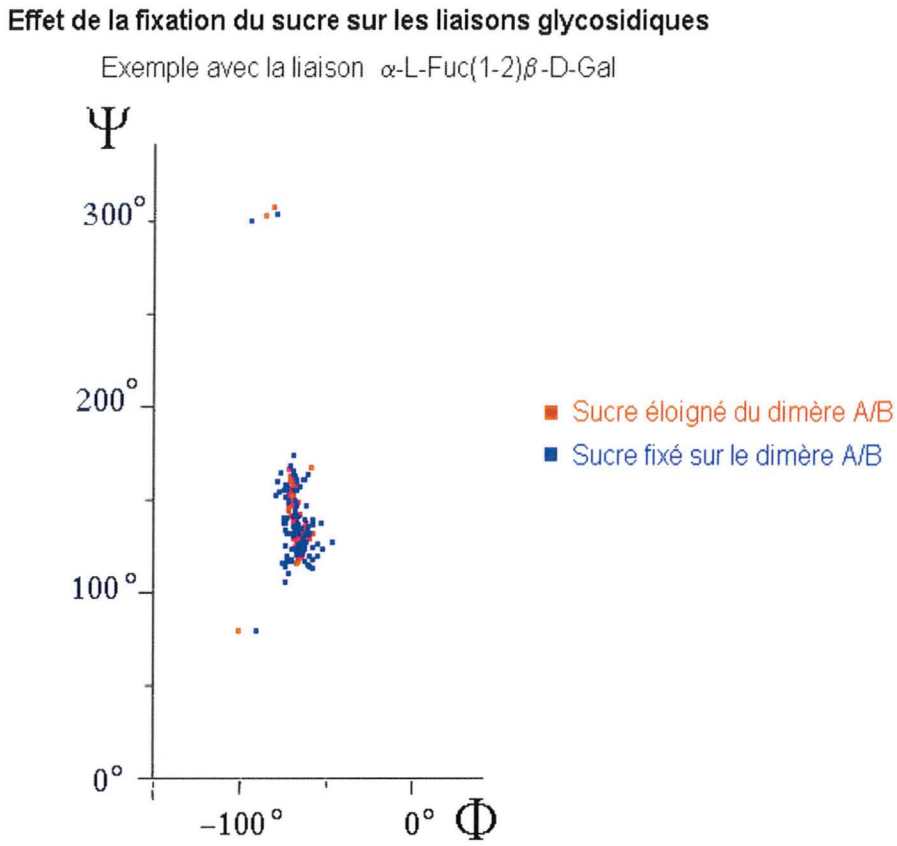



***Figure n°35***

Une dispersion du nuage de points est visible pour les angles de torsion correspondant au sucre en complexe avec AB en comparaison avec ceux qui en sont éloignés sur chacune des liaisons glycosidiques de deux trisacchacarides (données non montrées).

L'oligosaccharide, une fois fixé, subit des contraintes dues aux deux protéines. Pour cette raison, les angles de torsion Φ et Ψ peuvent varier de quelques degrés.

### 2.3.1- Analyse conformationnelle du trisaccharide H type 1 :

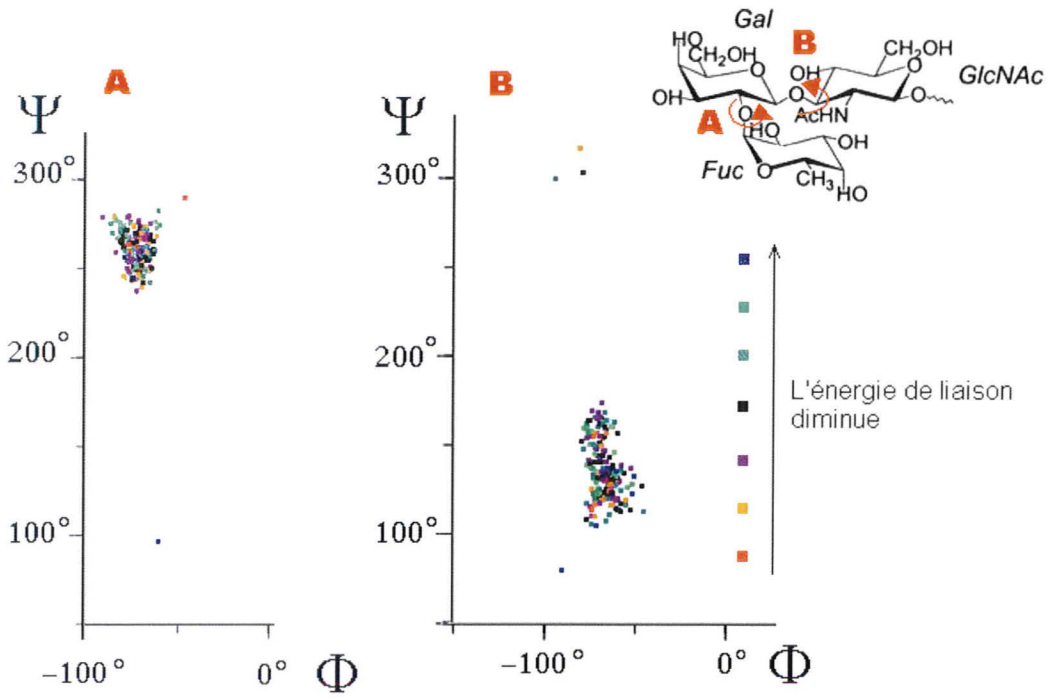
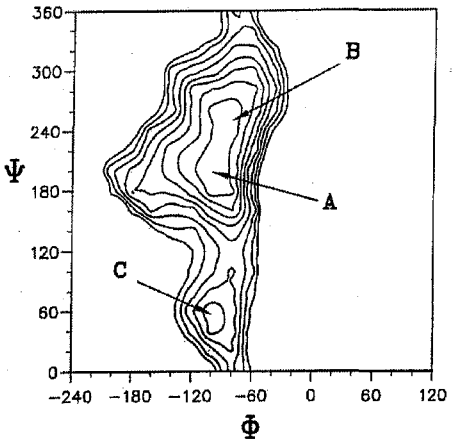



***Figure n°36***

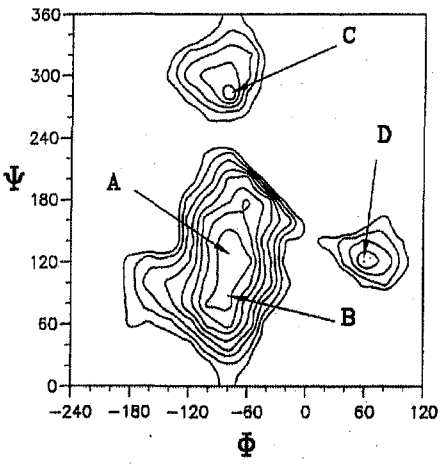
*Angles de torsion pour le trisaccharide H type 1 :* **A**, *angles de torsion de la liaison* α-*L-Fuc (1-2)*β-*D-Gal,* **B**, *angles de torsion de la liaison* β-*D-Gal(1-3)*β-*D-GlcNAc*

La plupart des trisaccharides représentant H type 1 possèdent des angles de torsion pour leur liaison glycosidique très similaires. Cependant, que ce soit pour la liaison α-L-Fuc(1-2)β-D-Gal ou β-D-Gal(1-3)β-D-GlcNAc, il existe quelques solutions isolées. Il existe notamment trois sucres pour la liaison galactose/N-acétylglucosamine qui présentent des angles de torsion très différents des autres et forment une petite population. La répartition des oligosaccharides en fonction de leur $\Delta E_{liaison}$ reste homogène vis-à-vis de leurs angles de torsion.

α-L-Fuc(1-2)β-D-Gal




β-D-Gal(1-3)β-D-GlcNAc




***Figures n° 37** (à gauche) **et 38** (à droite)*
*Selon Imberty et al. (1995)*
*Les isocontours de ces cartes ont été définis comme une augmentation de 1 kcal.mol$^{-1}$ à partir du minimum absolu.*

D'après les cartes adiabatiques de la littérature, pour la liaison α-L-Fuc(1-2)β-D-Gal de l'antigène H type 1, 199 solutions représentent la population B tandis que la population A n'existe pas dans l'échantillonnage obtenu. L'unique solution isolée appartient à la population C moins stable de 4 kcal.mol$^{-1}$ . Pour la liaison glycosidique β-D-Gal(1-3)β-D-GlcNAc, les 3 populations A, B et C sont représentées.

### 2.3.2- Analyse conformationnelle du trisaccharide représentant l'antigène de groupe A :

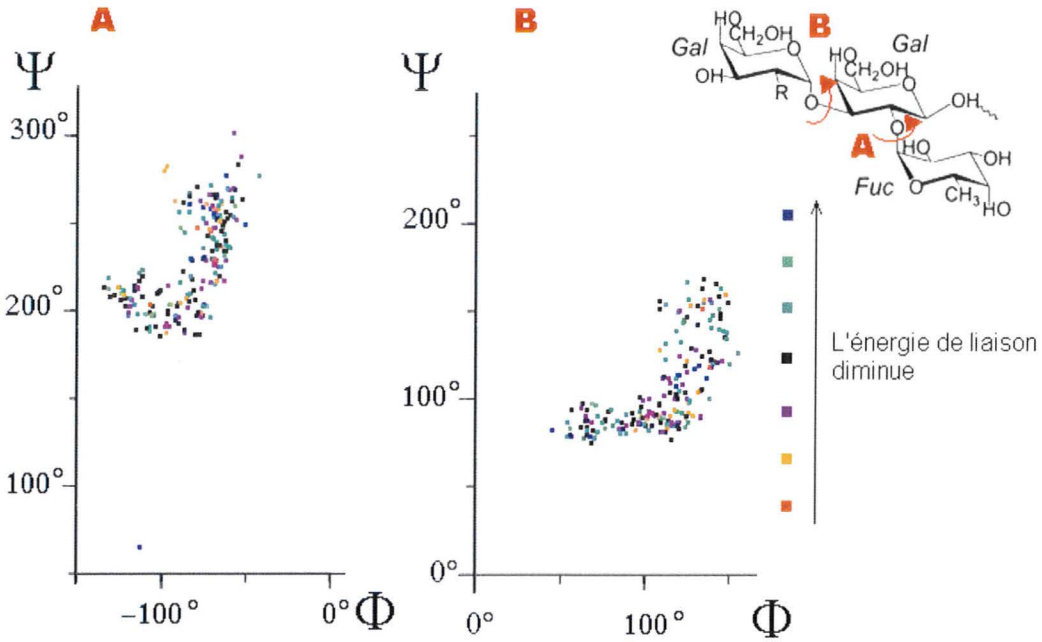



***Figure n°39 :*** *Angles de torsion pour l'antigène de groupe sanguin A :* **A**, *angles de torsion de la liaison glycosidique* α-*L-Fuc(1-2)*β-*D-Gal,* **B**, *angles de torsion de la liaison glycosidique* α-*D-GalNAc(1-3)*β-*D-Gal.*

De nouveau, comme pour l'antigène de groupe O, les angles de torsion de l'antigène de groupe sanguin A adoptent tous des valeurs presque similaires. Pour la liaison glycosidique α-L-Fuc (1-2)β-D-Gal, une solution est cependant isolée. En comparant ce profil avec celui pour la même liaison sur l'antigène H type 1, il semble que deux solutions aient des angles de torsion assez proches et pourraient former une population très minoritaire. Bien que cette population ne soit composée que de deux sucres sur 400, ils affichent des énergies d'interaction avec le dimère parmi les plus basses. Ainsi, une solution appartient à la population C ***(Figure n°37)***. La population A est cependant mieux représentée que pour l'antigène H type 1 ***(Figure n°37)***. La répartition des trisaccharides en fonction de leur $\Delta E_{liaison}$ est également très homogène pour l'antigène de groupe sanguin A.

α-D-GalNAc(1-3)β-D-Gal

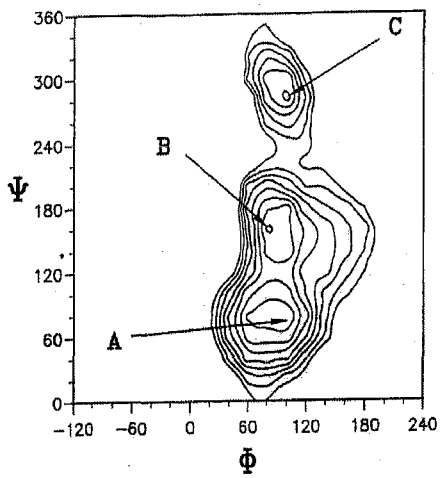



***Figure n°40***
*Selon Imberty et al. (1995)*
*Les isocontours de ces cartes ont été définis comme une augmentation de 1 kcal.mol$^{-1}$ à partir du minimum absolu.*

Pour la liaison α-D-GalNAc(1-3)β-D-Gal, les populations A et B sont représentées, aucune solution ne correspond à la population C.

Tous les angles Φ et Ψ calculés ne rentrent pas strictement dans ces cartes adiabatiques, cela s'explique facilement par les contraintes imposées aux carbohydrates par les protéines avec lesquelles ils sont en complexe. De plus ces cartes trouvées dans la littérature ont été réalisées dans le champ de force MM3, ce qui explique de légères variations puisque les trisaccharides ont ici été soumis au champ de force TRIPOS. L'obtention de quelques représentants de plusieurs familles moins stables en terme d'énergie d'après les cartes adiabatiques, montre que la recherche conformationnelle des deux ligands a été complète. Ces solutions moins stables ont tout de même été retenues car elles sont intéressantes en terme d'énergie d'interaction dans le champ de force d'AutoDock. Aucune solution n'a montré d'angles de torsion abérrants ainsi aucune n'a été retirée de l'étude.

### 2.4- Evaluation de l'énergie d'interaction

Comme par principe l'énergie se conserve au sein d'un système, une approximation de l'énergie libre de liaison s'obtient par soustraction de l'énergie totale du système quand le

ligand est lié à son récepteur avec l'énergie totale de ce même système lorsque le ligand et le récepteur sont hors d'influence l'un de l'autre. L'évaluation de l'énergie d'interaction du complexe ($\square E_{liaison}$) a été réalisée avec le champ de force TRIPOS. Ainsi pour chaque solution obtenue après amarrage moléculaire, deux minimisations des mêmes régions ont été effectuées : la première sur le complexe lectine/sucre et la deuxième avec le sucre éloigné de 200Å de la protéine afin de pouvoir par la suite évaluer le $\Delta E_{liaison}$ de chaque solution. Les paramètres de ces deux minimisations ont été strictement identiques. Les régions pour les calculs de minimisations et de champ de force ont été définies afin de permettre une diminution des temps de calcul. Le nombre d'évaluation maximum a été fixé à 100 000 itérations et la minimisation a été réalisée avec la méthode POWELL. (Voir Fichiers de minimisation dans les ANNEXES D/ et E/ p106 et 107)

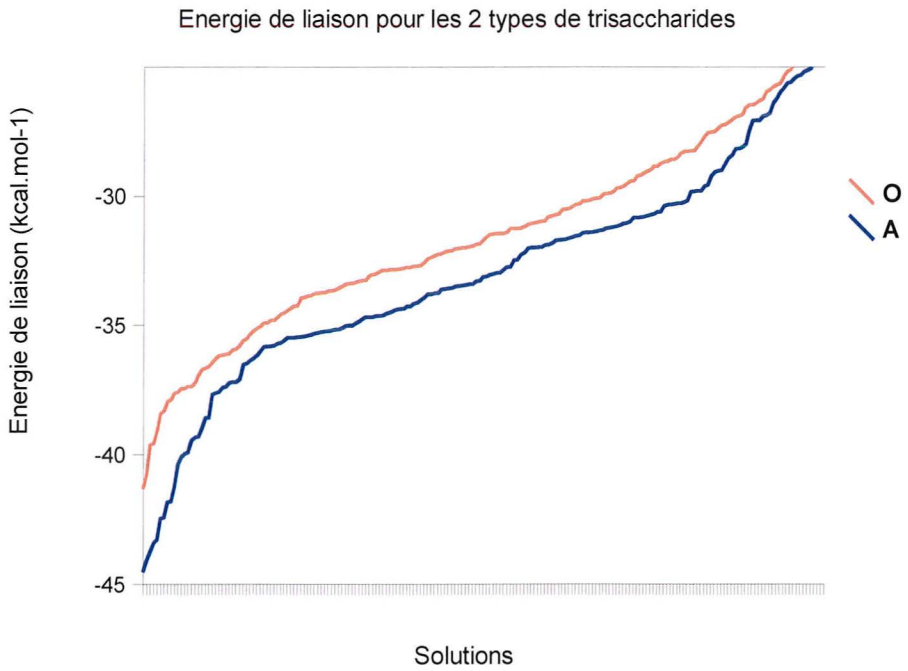



***Figure n°41***

Les énergies de liaison de chaque solution pour l'antigène de groupe A sont toutes inférieures aux énergies de liaison de l'antigène H type 1. L'énergie de liaison étant un critère important pour déterminer quelles sont les solutions les plus probables, elle a été surtout utilisée pour l'analyse de la qualité des poches de fixation en terme d'énergie.

### 2.5- Analyse de la distribution des positions obtenues après l'amarrage moléculaire : représentation du paysage énergétique en coordonnées sphériques

Pour délimiter plus facilement les poches préférentielles de fixation des sucres à la surface du dimère, une analyse graphique a été réalisée à partir des coordonnées sphériques des pseudo-centroïdes de chacune des solutions.

Le calcul des coordonnées s'est fait selon la méthode suivante : l'atome le plus proche du centre de masse de chaque sucre (**S**) a été représenté en coordonnées sphériques par rapport au centre de la zone d'amarrage moléculaire (**D**). (Voir fichier pour le calcul des coordonnées sphériques dans l'ANNEXE G/ p112)

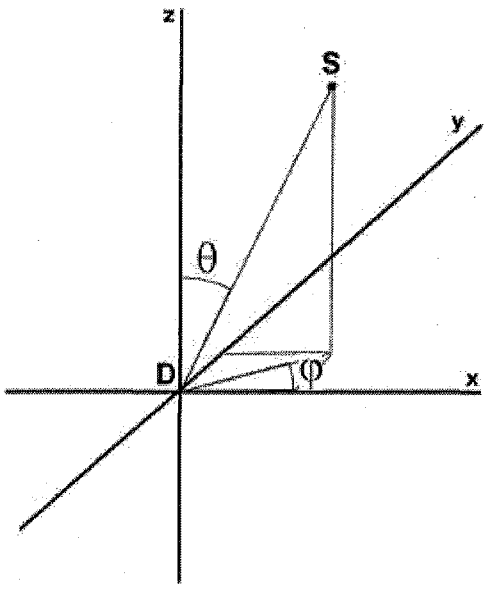



***Figure n°42***

**D** n'a pas été choisi comme le centre de la boîte définie par autodock puisque celle-ci ne recouvre qu'une partie de la protéine de capside A. Il a donc été fixé au centre du volume formé par les deux protéines à l'intérieur de la boîte.

L'angle φ a été calculé par rapport à l'axe des abscisses tandis que l'angle θ correspond à l'angle formé par la projection du pseudo-centre de masse **S** sur le plan yz. La représentation

graphique a été réalisée selon trois axes : l'axe des abscisses correspond à : φ sin θ, celui des ordonnées à : 90-θ tandis que la profondeur est déterminée selon la coordonnée x du pseudo-centroïde d'origine.

Chaque sucre représenté par son pseudo-centre de masse correspond à un point. Deux pseudo-centres de masse à proximité l'un de l'autre représentent deux sucres particulièrement proches dans l'espace tridimensionnel autour du dimère. Le nuage de points ainsi obtenu est légèrement déformé puisque le complexe formé par les deux protéines de capside n'est pas sphérique. Cette méthode permet cependant de recouper de manière simple et rapide les différentes bases de données des deux trisaccharides. Il est en effet difficile de regarder tous les sucres les uns après les autres. De plus, les différences conformationnelles des solutions peuvent rendre la délimitation précise des poches de fixation délicate. Cette projection permet aussi d'apprécier la densité des solutions dans l'environnement du dimère selon les énergies de liaison obtenues après minimisation. Dans l'espace, tous les points ont tendance à se concentrer sur un plan, la symétrie des protéines de capside est également visible.

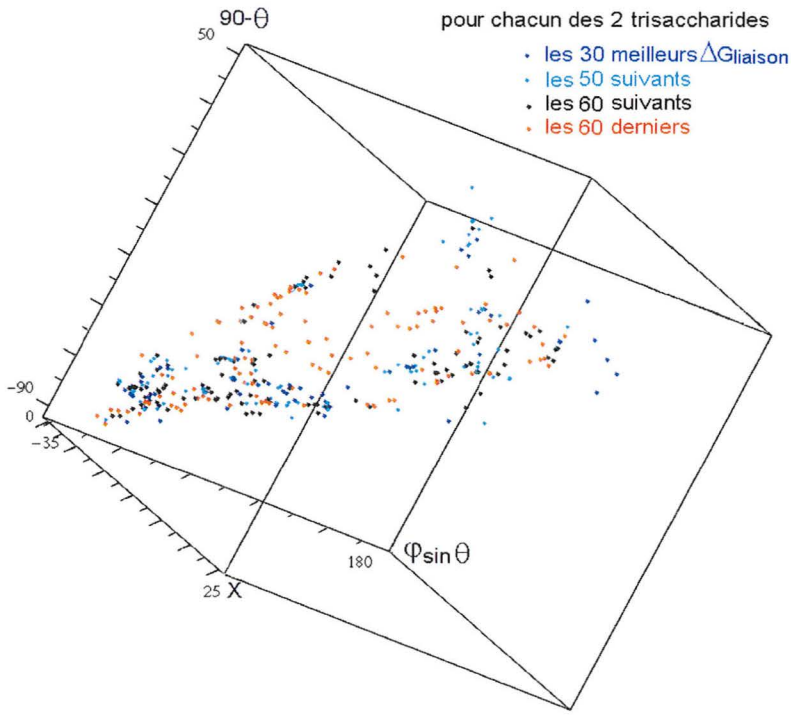



***Figure n°43***
*Pseudo-centres de masse de chaque solution colorée en fonction de son* $\Delta E_{liaison}$ *avec le dimère*

(Voir autre angle de vue dans l'ANNEXE H/ p113)

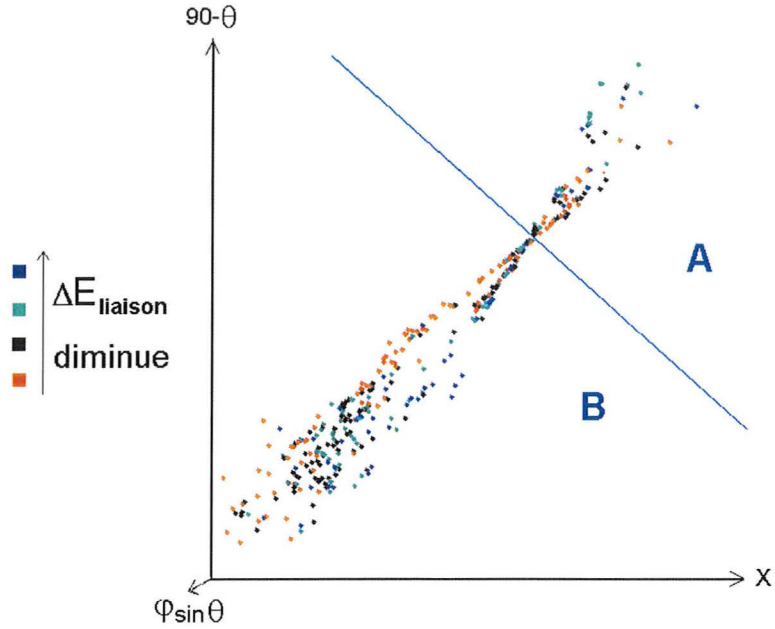



***Figure n°44***
*Vue de profil du nuage de points coplanaires. L'axe de symétrie est visible en bleu. La partie gauche inférieure correspond aux solutions localisées sur la protéine de capside B tandis que la partie supérieure droite n'est composée que des solutions en complexe avec la protéine de capside A.*

### 2.6- Sélection des meilleures poches de reconnaissance

Grâce à la représentation en coordonnées sphériques des pseudo-centroïdes de chaque sucre, des poches ont été facilement déterminées. Elles ont dans un premier temps été sélectionnées en fonction de la proximité et des énergies d'interaction des sucres après la minimisation des complexes dans TRIPOS. Puis le nombre de liaisons hydrogènes formées entre les trisaccharides et leur dimère a été évalué ainsi que la position des sucres vis-à-vis des différents domaines P1 et P2. Leur localisation par rapport aux acides aminés mutés connus pour diminuer ou inhiber la fixation du virus aux antigènes de groupes sanguins a également été étudiée. Pour finir, les groupements aminoéthyls ont été le centre d'une attention particulière et les liaisons hydrogènes impliquant leur atome d'azote, d'oxygène ou

d'hydrogène ont été soigneusement examinées. L'orientation des trisaccharides notamment celle de la liaison à laquelle est branché le reste de la chaîne glycosidique des deux antigènes a aussi été prise en compte.

***Tableau VII.*** *Poches de reconnaissance pour l'antigène de groupe sanguin O.*
* *classement dans l'ordre croissant des familles selon leur énergie de liaison moyenne dans le champ de force TRIPOS.*

| **Famille** | **Antigène de groupe O** | **Liaison hydrogène mettant en jeu le groupement aminoéthyl**<br><br>- O : liaison par le C=O<br>- H : liaison par le H N-H<br>- c : conservés à + de 80%<br>- m : mutés | **Nombre total de liaisons hydrogènes** | **Classement dans le champ de force TRIPOS** | **Energie de liaison dans le champ de force TRIPOS** |
|---|---|---|---|---|---|
| **c** | H1_38 | H SER251 | 7 | 2 | -39.12 (2) |
| | H1_8 | H SER251 | 7 | 7 | -37.46 (3) |
| | H2_1 | / | 7 | 10 | -41.34 (1) |
| | H2_11 | H SER251 | 7 | 5 | -34.59 (4) |
| | H2_33 | H ILE250 | 2 | 2 | -33.69 (5) |
| | | | | moy<br>**5.2** | moy<br>**-37.24**<br><br>**(1)*** |

| | | | | | |
|---|---|---|---|---|---|
| **z** | H1_50<br>H1_48<br>H1_77 | O LYS424<br>H ASP305 c<br>H ASP305 c | 8<br>10<br>5 | 11<br>37<br>44<br><br>moy<br>**30.6** | -36.46 (1)<br>-33.65 (2)<br>-32.39 (3)<br><br>moy<br>**-34.16**<br><br>**(6)*** |
| **a** | H1_4<br>H1_26<br>H1_28<br>H1_43<br>H1_87<br>H2_29<br>H2_52<br>H2_7<br>H2_93 | /<br>/<br>/<br>/<br>/<br>/<br>/<br>/<br>/ | 7<br>8<br>10<br>7<br>4<br>7<br>10<br>9<br>7 | 6<br>3<br>8<br>41<br>26<br>2<br>4<br>16<br>21<br><br>moy<br>**14.1** | -37.60 (4)<br>-37.97 (3)<br>-37.46 (5)<br>-32.77 (9)<br>-33.78 (8)<br>-39.63 (1)<br>-38.34 (2)<br>-34.93 (6)<br>-34.37 (7)<br><br>moy<br>**-36.31**<br><br>**(4)*** |

| | | | | | |
|---|---|---|---|---|---|
| **e** | H1_90<br>H1_7<br>H2_56<br>H2_90<br>H2_79<br>H2_28<br>H2_92<br>H1_89<br>H2_25<br>H2_57<br>H1_14 | /<br>/<br>O GLN386<br>/<br>/<br>/<br>O TRP390 c++<br>/<br>/<br>/<br>/ | 7<br>6<br>7<br>8<br>6<br>3<br>5<br>6<br>3<br>4<br>5 | 9<br>5<br>8<br>10<br>6<br>28<br>45<br>31<br>68<br>26<br>24<br><br>moy<br>**23.6** | -37.38 (2)<br>-37.90 (1)<br>-36.94 (4)<br>-36.61 (5)<br>-37.38 (2*)<br>-33.50 (9)<br>-31.86 (10)<br>-33.57 (8)<br>-30.13 (11)<br>-33.75 (7)<br>-34.27 (6)<br><br>moy<br>**-34.84**<br><br>**(5)*** |
| **b** | H1_54<br>H2_91 | /<br>/ | 11<br>9 | 15<br>7<br><br>moy<br>**11** | -36.10 (2)<br>-37.24 (1)<br><br>moy<br>**-36.67**<br><br>**(2)*** |
| **k** | H1_71<br>H1_57<br>H1_41<br>H2_46<br>H2_17 | OLYS391<br>OLYS391<br>OLYS391<br>OLYS391<br>OLYS391 | 3<br>3<br>1<br>2<br>3 | 14<br>10<br>18<br>5<br>12<br><br>moy<br>**11.8** | -36.14 (3)<br>-36.72 (2)<br>-35.53 (5)<br>-37.65 (1)<br>-35.97 (4)<br><br>moy<br>**-36.40**<br><br>**(3)*** |

| | | | | | |
|---|---|---|---|---|---|
| **i** | H1_99<br>H2_35<br>H2_98<br>H2_44<br>H1_29<br>H2_83<br>H2_16 | /<br>/<br>/<br>/<br>/<br>/<br>/ | 6<br>6<br>3<br>6<br>6<br>7<br>5 | 36<br>32<br>37<br>25<br>42<br>38<br>53<br><br>moy<br>**37.5** | -32.94 (3)<br>-33.28 (2)<br>-30.71 (7)<br>-33.75 (1)<br>-32.77 (4)<br>-32.73 (5)<br>-31.27 (6)<br><br>moy<br>**-32.49**<br><br>**(8)*** |
| **d** | H2_26<br>H1_2<br>H1_46<br>H1_20<br>H1_64<br>H2_73<br>H1_47<br>H2_31<br>H2_19<br>H1_80<br>H2_100<br>H2_50<br>H2_61 | HSER262<br>HSER262<br>HSER262<br>HSER262<br>HSER262<br>HSER262<br>/<br>/<br>/<br>/<br>/<br>/<br>/ | 6<br>6<br>5<br>6<br>6<br>5<br>5<br>8<br>7<br>7<br>7<br>5<br>6 | 29<br>33<br>54<br>29<br>27<br>36<br>60<br>20<br>14<br>22<br>37<br>39<br>50<br><br>moy<br>**34.6** | -33.40 (6)<br>-33.35 (7)<br>-31.95 (11)<br>-33.68 (5)<br>-33.77 (4)<br>-32.85 (8)<br>-30.99 (13)<br>-34.45 (3)<br>-35.14 (1)<br>-34.82 (2)<br>-32.83 (9)<br>-32.32 (10)<br>-31.46 (12)<br><br>moy<br>**-33.15**<br><br>**(7)*** |

***Tableau VIII.*** *Poches de reconnaissance pour l'antigène de groupe sanguin A.*
** classement dans l'ordre croissant des familles selon leur énergie de liaison moyenne dans le champ de force TRIPOS.*

| Famille | Antigène de groupe A | Liaison hydrogène mettant en jeu le groupement aminoéthyl<br>- O : liaison par le C=O<br>- H : liaison par le H N-H<br>- c : conservés à + de 80%<br>- m : mutés | Nombre total de liaisons hydrogènes | Classement dans le champ de force TRIPOS | Energie de liaison dans le champ de force TRIPOS |
|---|---|---|---|---|---|
| **B** | A_81 | / | 5 | 2 | -43.43 (2) |
| | A_48 | / | 4 | 4 | -41.85 (5) |
| | A_47 | / | 4 | 10 | -39.34 (7) |
| | A_73 | / | 5 | 13 | -38.58 (8) |
| | A_76 | O LYS424 | 6 | 23 | -35.84 (9) |
| | A_64 | H ASP305 c | 6 | 44 | -34.04 (11) |
| | A2_9 | / | 5 | 3 | -43.30 (3) |
| | A2_38 | / | 8 | 2 | -44.11 (1) |
| | A2_45 | / | 7 | 4 | -42.47 (4) |
| | A2_46 | O SER251 | 7 | 34 | -34.38 (10) |
| | A2_84 | / | 6 | 5 | -40.10 (6) |
| | | | | moy<br>**13.09** | moy<br>**-39.76**<br>**(1)*** |
| **A** | A_89 | / | 9 | 1 | -43.77 (1) |
| | A_78 | O ARG291 m | 8 | 9 | -39.92 (2) |
| | A_87 | / | 3 | 28 | -35.47 (4) |
| | A_49 | / | 5 | 36 | -35.02 (8) |
| | A_56 | O ARG291 m | 7 | 19 | -36.48 (3) |
| | A_75 | O ARG291 m | 7 | 31 | -35.33 (6) |
| | A2_13 | H ASP305 c | 5 | 32 | -34.55 (9) |
| | A2_23 | / | 4 | 35 | -34.37 (10) |
| | A2_39 | / | 4 | 21 | -35.30 (7) |
| | A2_76 | H ASN498 c++ | 5 | 19 | -35.46 (5) |
| | | | | moy<br>**23.1** | moy<br>**-36.58**<br>**(5)*** |

| | | | | | |
|---|---|---|---|---|---|
| **H** | A_83<br>A2_22<br>A2_60<br>A_9<br>A2_40 | /<br>/<br>/<br>H ASN432<br>/ | 7<br>7<br>2<br>6<br>7 | 3<br>40<br>29<br>12<br>15<br><br><br><br><br>moy<br>**19.8** | -42.44 (1)<br>-33.77 (5)<br>-34.65 (4)<br>-38.59 (2)<br>-35.82 (3)<br><br><br><br><br>moy<br>**-37.05**<br><br>**(4)*** |
| **K** | A_37<br>A2_47 | O CYS326<br>O CYS326 | 6<br>6 | 6<br>27<br><br><br><br><br>moy<br>**16.5** | -41.26 (1)<br>-34.75 (2)<br><br><br><br><br>moy<br>**-38.00**<br><br>**(3)*** |
| **G** | A_62<br>A_97<br>A_26<br>A2_74 | H LEU276 c<br>/<br>O GLY278 c++<br>/ | 5<br>5<br>5<br>4 | 30<br>25<br>40<br>22<br><br><br><br><br>moy<br>**29.25** | -35.38 (2)<br>-35.69 (1)<br>-34.69 (4)<br>-35.27 (3)<br><br><br><br><br>moy<br>**-35.26**<br><br>**(6)*** |
| **S** | A2_63<br>A2_97 | H GLU455<br>H GLU455 | 6<br>6 | 6<br>7<br><br><br><br><br>moy<br>**6.5** | -39.46 (1)<br>-39.31 (2)<br><br><br><br><br>moy<br>**-39.39**<br><br>**(2)*** |

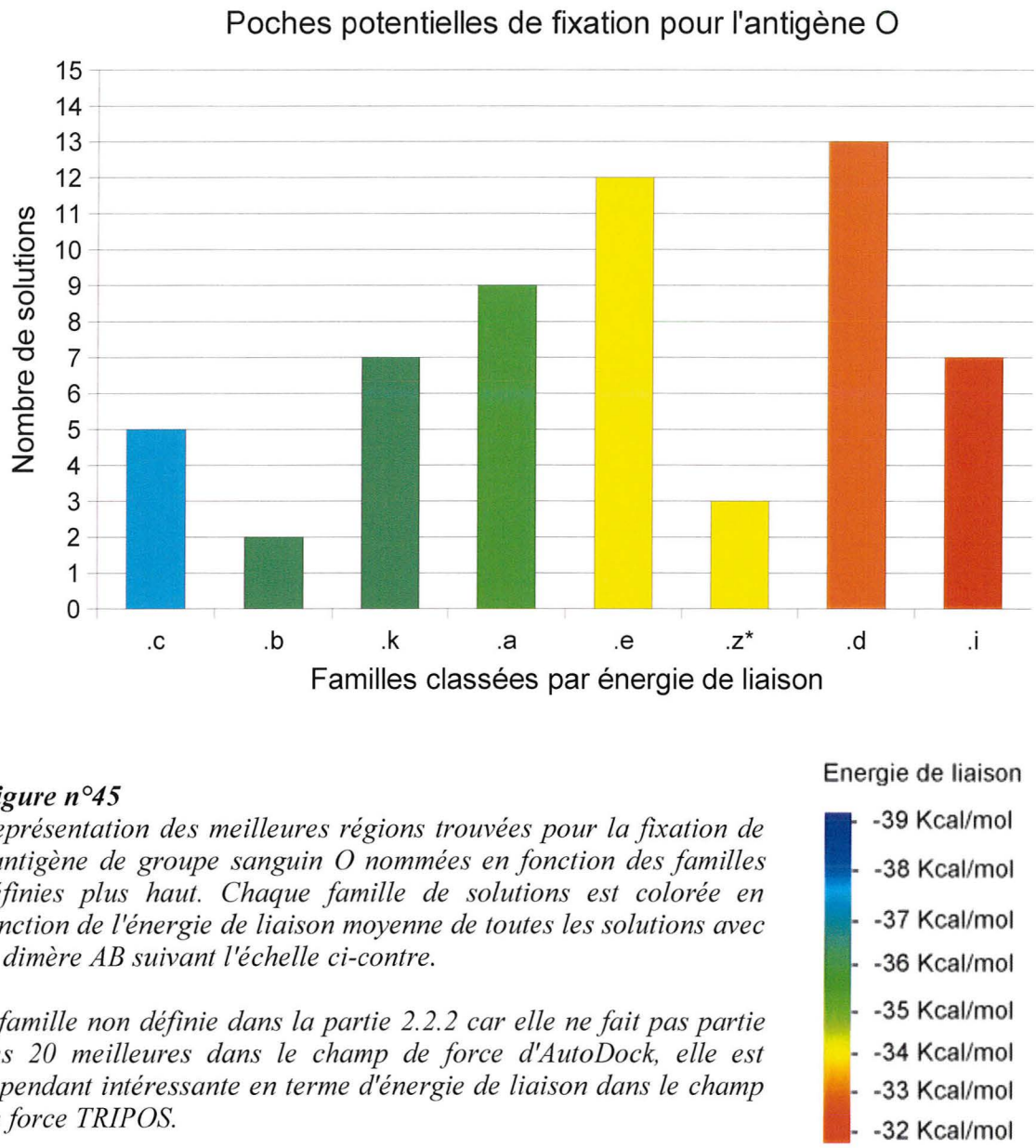



***Figure n°45***

*Représentation des meilleures régions trouvées pour la fixation de l'antigène de groupe sanguin O nommées en fonction des familles définies plus haut. Chaque famille de solutions est colorée en fonction de l'énergie de liaison moyenne de toutes les solutions avec le dimère AB suivant l'échelle ci-contre.*

** famille non définie dans la partie 2.2.2 car elle ne fait pas partie des 20 meilleures dans le champ de force d'AutoDock, elle est cependant intéressante en terme d'énergie de liaison dans le champ de force TRIPOS.*

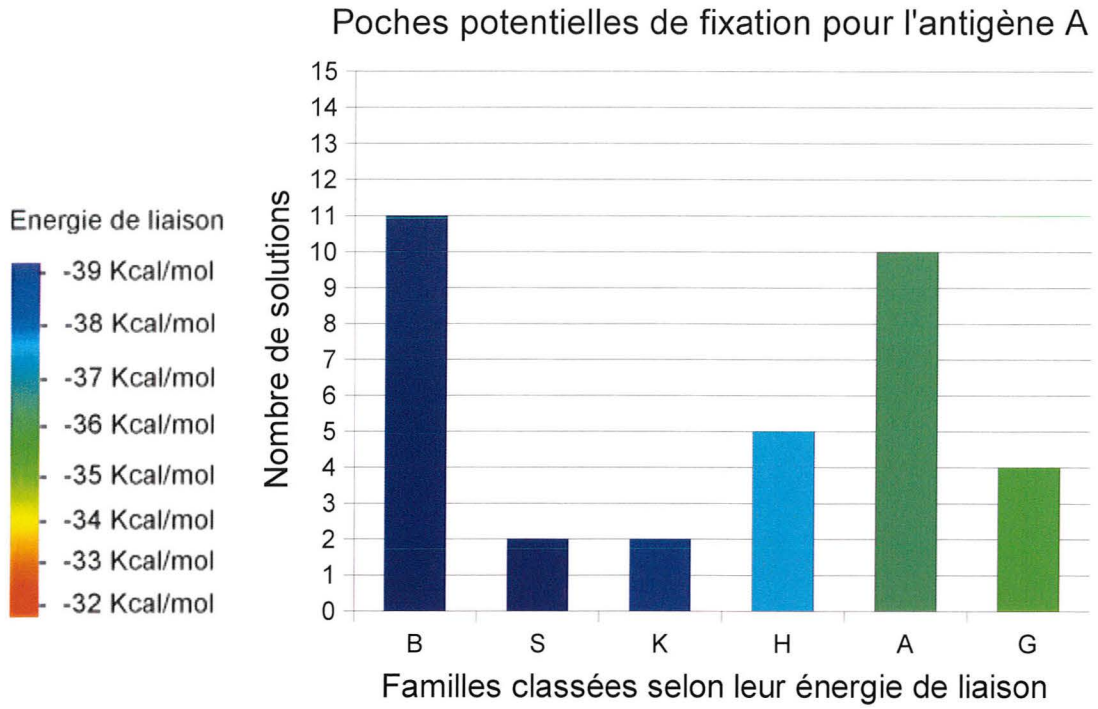



***Figure n°46***
*Représentation des meilleures régions trouvées pour la fixation de l'antigène de groupe sanguin A nommées en fonction des familles définies plus haut. Chaque famille de solutions est colorée en fonction de l'énergie de liaison moyenne de toutes les solutions avec le dimère AB suivant l'échelle ci-dessus.*

Une fois de plus, il est facile de voir que le trisaccharide représentant l'antigène de type A montre des énergies de liaison plus basses que celui représentant l'antigène de groupe O.

Chaque cavité contient plus ou moins de trisaccharides et certaines ne sont constituées que d'un seul des deux trisaccharides étudiés. Par exemple, seuls les antigènes H type 1 sont retrouvés dans les régions a, b et d tandis que la poche G n'est composée que d'antigènes de groupe A. Le nombre de solutions par niche est de 10 en moyenne mais varie de 17 à 2 au minimum. En effet, seuls deux trisaccharides remplissent la cavité b, cependant ils ont un RMSD très faible, donc présentent une conformation identique et sont, de plus, presque parfaitement superposés avec des énergies d'interaction dans le champ de force TRIPOS parmi les meilleures.

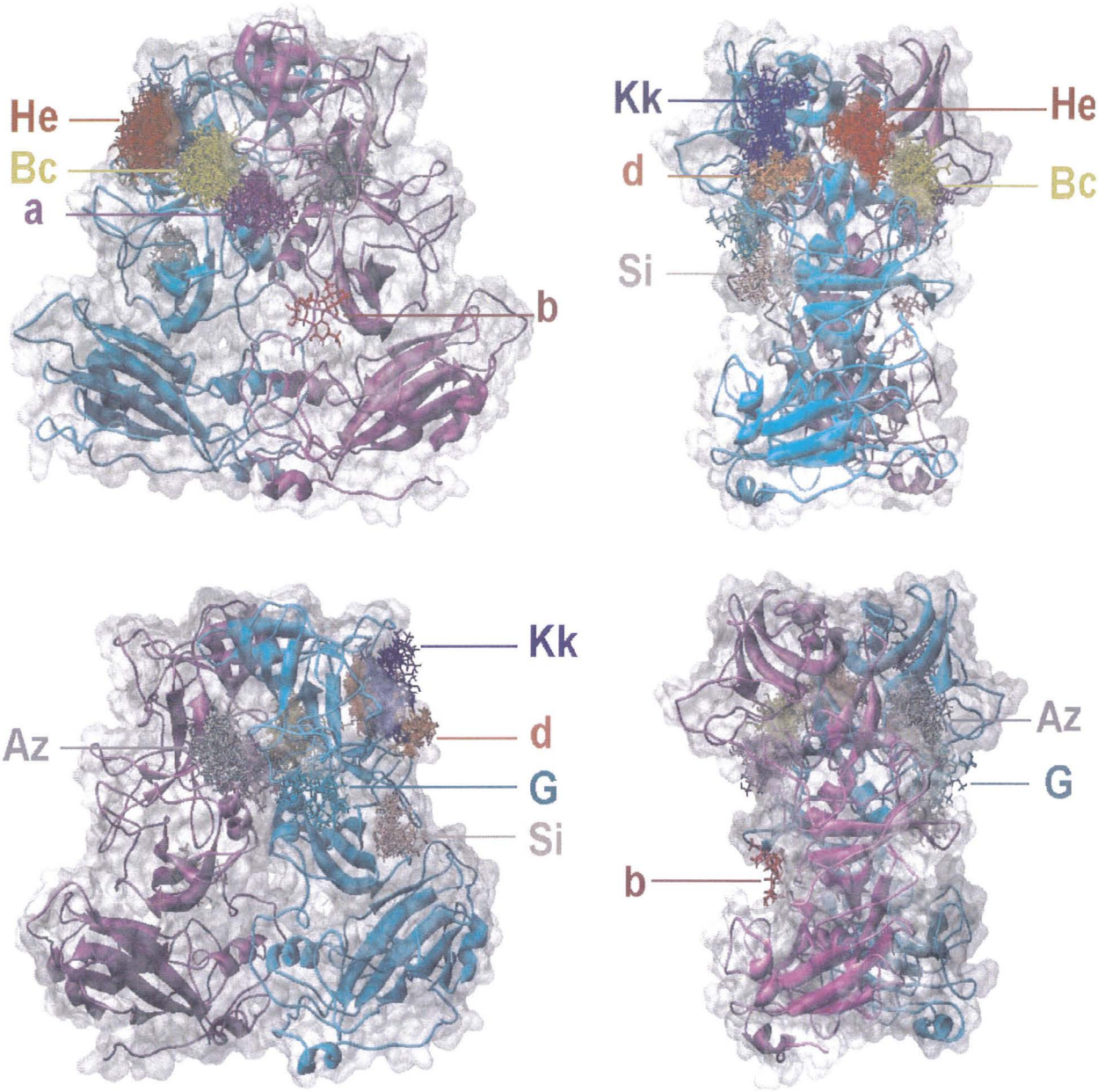



***Figure n°47***
*Vue des quatres faces du dimère AB avec les familles les plus intéressantes pour les antigènes A et O.*
*La protéine de capside B est représentée en vert tandis que celle de capside A est en violet.*

Afin de limiter les temps de calculs et en considérant la quasi équivalence des protéines A et B, la protéine de capside A n'a que partiellement été introduite dans la boîte. Deux poches ont pourtant été trouvées comme étant la symétrique l'une de l'autre. En effet, la région Bc est

symétrique à la poche Az. Ces deux poches sont aussi particulièrement intéressantes puisqu'elles ont des énergies d'interaction basses dans le champ de force TRIPOS.

Sur les 9 poches trouvées, 4 seulement possèdent des oligosaccharides situés à moins de 5 Å d'acides aminés très probablement impliqués dans un poche de fixation selon Tan *et al.*

# DISCUSSION

## MODELISATION DE LA CAPSIDE DU VIRUS DE NORWALK AVEC DEUX OLIGOSACCHARIDES : POCHE DE FIXATION

### 1.1- Données expérimentales publiées

Les Norovirus possèdent quatre profils de fixation différents aux antigènes de groupes sanguins. Pour les individus sécréteurs, le virus de Norwalk reconnaît les antigènes A et O, la souche VA387 fixe les antigènes A, B et O tandis que la souche MOH n'interagit qu'avec les antigènes A et B. Seule la souche VA207 reconnaît les individus lewis qu'ils soient sécréteurs ou non.

Des alignements multiples de séquences protéiques de capsides de plusieurs Norovirus possédant différents profils de fixation ont révélé un motif conservé RGD dans le domaine P2. Ce motif, connu comme étant un site de reconnaissance universel intervenant dans les interactions entre les cellules et leur environnement, est en effet impliqué dans les interactions récepteur-ligand, enzyme-substrat, hormone-cible ainsi que dans les systèmes de transduction du signal. Il est également responsable de la fixation aux cellules de virus tels que les Adénovirus ou les Coxsackievirus. A partir de ces alignements de séquences, il a été montré que le motif RGD est entouré de plusieurs résidus ou « points chauds » spécifiques à chaque souche de Norovirus. Une première prédiction d'un site de reconnaissance à la surface de la protéine de capside pour chacune des souches a ainsi pu être faite.

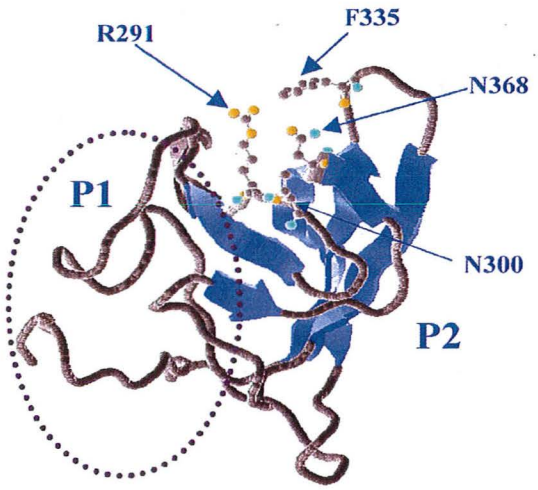



***Figure n°48***
*Selon Tan et al. (2003)*

*Vue du site actif prédit par alignement de séquences de plusieurs Norovirus pour le virus de Norwalk sur une protéine de capside.*

Pour le virus de Norwalk, le motif RGD se retrouve sous la forme RGT soit l'arginine 291, la glycine 292 et la thréonine 293. Selon Tan *et al.*, ces trois acides aminés constituent le fond de la poche de reconnaissance potentielle. Trois autres résidus sont apparus comme étant spécifiques du virus de Norwalk, ainsi la phénylalanine 335 et les asparagines 300 et 368 qui délimitent les contours de cette poche joueraient un rôle important. L'implication de ces divers acides aminés dans la reconnaissance des HBGAs a été étudiée par réalisation de différents mutants sur les souches virales VA387 et MOH [27].

Pour ces deux souches, le motif RGD joue un rôle dans la fixation aux antigènes de groupes sanguins. En effet, la disparition de celui-ci aboutit à une perte totale de reconnaissance des antigènes A, B et O ; cependant, une mutation sur l'arginine ou la glycine n'entraîne qu'une diminution de la fixation de l'antigène de groupe O [27]. Ces trois acides aminés sont donc nécessaires à la reconnaissance du virus pour ces deux souches, cependant leur rôle dans la spécificité de fixation reste difficile à déterminer.

Dans les souches VA387 et MOH, la phénylalanine 335 du virus de Norwalk est respectivement remplacée par une thréonine et une asparagine, lesquelles jouent un rôle primordial dans la fixation de la capside aux carbohydrates puisque la seule mutation de ces acides aminés entraîne la perte totale de reconnaissance des récepteurs [27].

Les acides aminés correspondant aux asparagines 300 et 368 du virus de Norwalk sembleraient être impliqués dans la spécificité de la reconnaissance des souches virales. Une mutation sur la souche VA387 au niveau de l'asparagine 302 (correspondant à l'asparagine 300 du NV selon l'alignement de séquences de *Tan et al.*) a montré une diminution de la reconnaissance de l'antigène de groupe sanguin O.

D'autres études ont été réalisées sur la protéine de capside du NV, les mutations de l'histidine 329 d'une part et du tryptophane 375 d'autre part ont montré une diminution de reconnaissance des antigènes de groupes sanguins dans les deux cas. Tandis que la mutation de l'asparagine 331 n'a aucun effet. Sur ces trois acides aminés, seul le tryptophane est présent à la surface de la protéine. Les mutations sur l'histidine 329 et l'asparagine 331 ont été réalisées afin d'évaluer leur rôle dans la structure de la poche de reconnaissance. L'asparagine 331 semble par sa position être plus impliquée dans le contact dimérique.

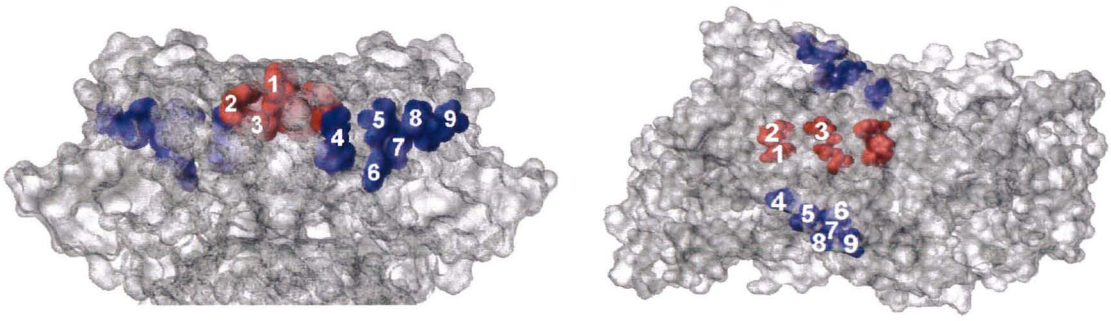



***Figure n°49***
*A gauche, vue de profil de la partie supérieure du dimère. A droite, vue de dessus. En bleu figure la poche de reconnaissance selon Tan et al. déterminée par alignement de séquences; en rouge, les 3 acides aminés mutés sur la capside du virus de Norwalk.*
***1.*** *TRP 375,* ***2.*** *HIS 329,* ***3.*** *ASN 331,* ***4.*** *PHE 335,* ***5.*** *ASN 368,* ***6.*** *ARG 291,* ***7.*** *GLY 392,* ***8.*** *THR 393,* ***9.*** *ASN 300*

## 1.2- Prédiction par amarrage moléculaire pour l'antigène de groupe sanguin A

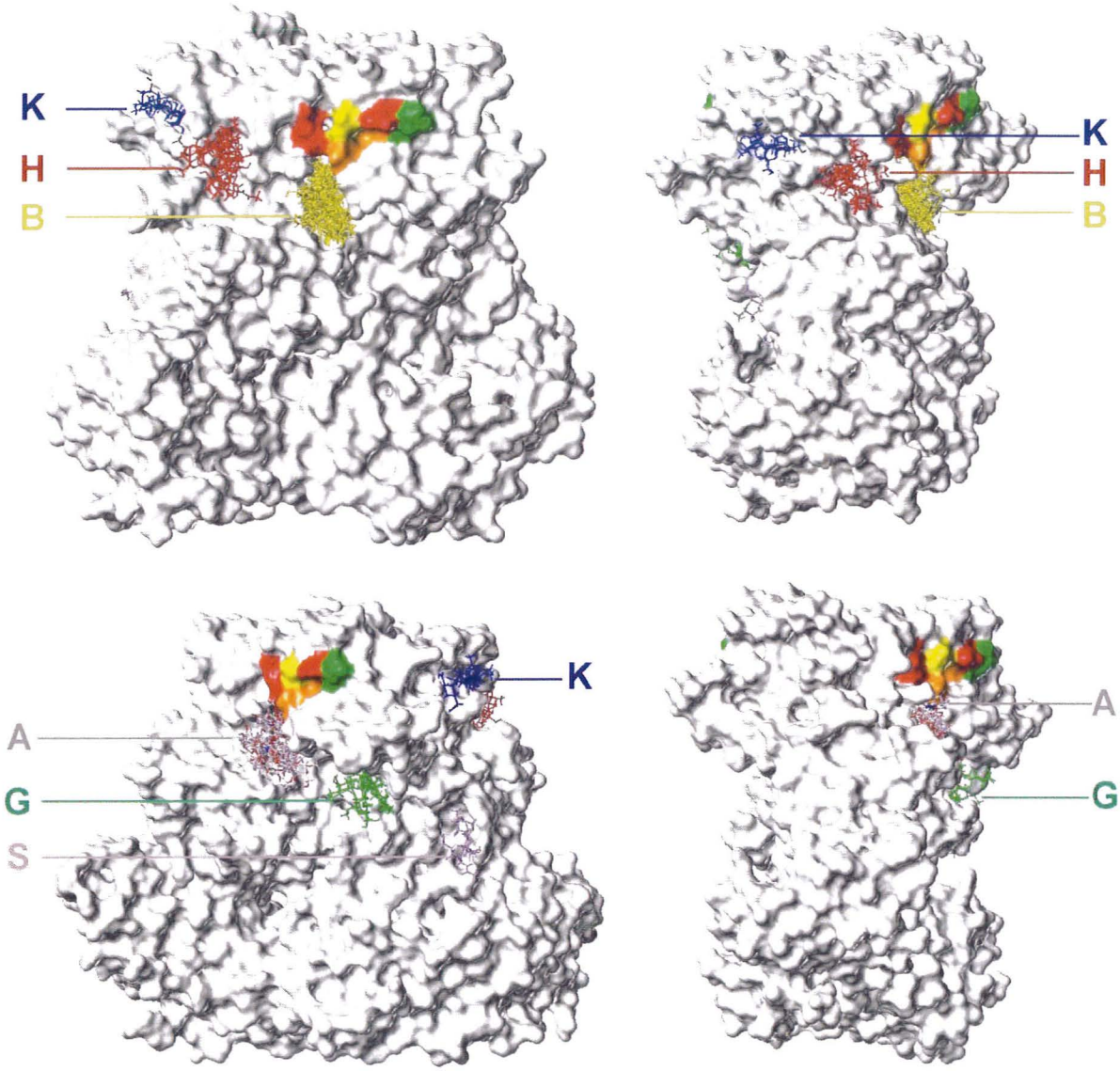



***Figure n°50***
*Surface du dimère du virus de Norwalk coloré en fonction de l'effet des acides aminés mutés sur la reconnaissance de l'antigène de groupe A pour la souche VA387.*
*En rouge apparaissent les acides aminés qui, mutés, diminuent la fixation de l'antigène, en vert, ceux qui l'augmentent. Enfin, en jaune figurent les acides aminés sans effet sur la fixation. L'orange désigne les acides aminés qui, mutés seuls, n'ont aucun effet mais qui, mutés avec d'autres résidus, diminuent la reconnaissance. Seules les solutions correspondant aux poches déterminées précédemment pour l'antigène de groupe A sont représentées sur cette figure.*

Après l'amarrage moléculaire, deux poches de fixation potentielles symétriques l'une de l'autre, **A** et **B**, dont les énergies de liaison avec les antigènes de groupe sanguin A sont parmi les plus basses, se sont révélées particulièrement intéressantes. Elles se situent en effet dans la région déterminée par Tan *et al.* comme étant impliquée dans la reconnaissance des HBGAs par le virus de Norwalk. Les solutions trouvées par AutoDock se positionnent ainsi à moins de 6 Å de plusieurs des acides aminés prédits comme intervenant dans la reconnaissance : l'arginine 291, la phénylalanine 335 et l'asparagine 368.

L'arginine 291 semblerait jouer un rôle très important dans l'interaction avec l'antigène de groupe sanguin A pour le virus de Norwalk. En effet, plusieurs solutions forment des liaisons hydrogènes par l'oxygène de leur groupement N-acétyl avec cet acide aminé. De plus, la reconnaissance par ce groupement N-acétyl semble d'autant plus probable lorsque l'on considère que l'antigène B, qui ne le possède pas, n'est que très peu reconnu.

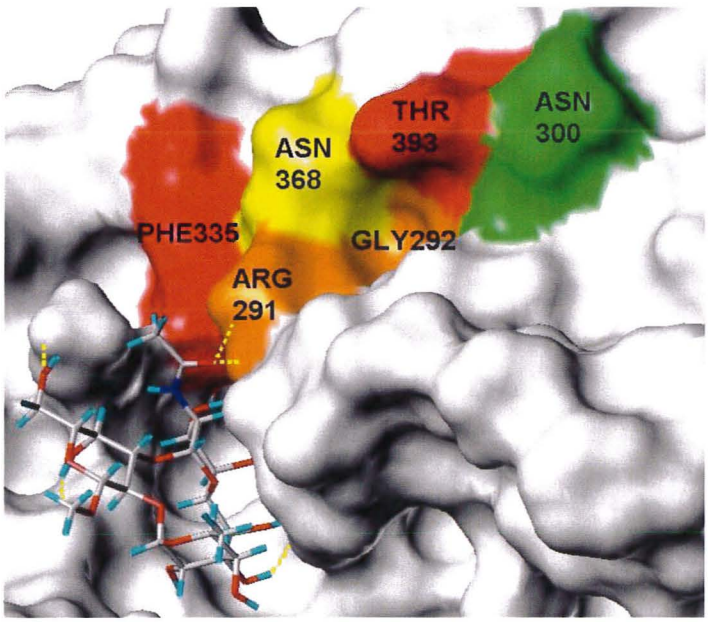



***Figure n°51***
*Trisaccharide représentant l'antigène de groupe A appartenant à la famille B proche des acides aminés mutés dans la souche VA387. Cette solution forme un total de huit liaisons hydrogènes avec la protéine dont deux avec l'ARG291 par l'intermédiaire de l'oxygène du groupement N-acétyl. Le groupement hydroxyméthyl du galactose représentant le point d'ancrage pour le reste de l'antigène de groupe sanguin est correctement orienté pour le branchement du reste de l'oligosaccharide.*

Cependant, selon Tan *et al.*, la mutation de l'acide aminé équivalent à cette arginine dans la souche VA387 n'affecte pas la reconnaissance de l'antigène de groupe sanguin A. Cette souche n'appartient néanmoins pas au même génogroupe que le virus de Norwalk, de plus, ces deux virus ne présentent pas le même profil de fixation, les données expérimentales obtenues par mutation de VA387 ne peuvent donc être étendues au virus de Norwalk qu'avec beaucoup de prudence. Il faut également prendre en compte le fait que le trisaccharide n'est pas non plus uniquement reconnu par l'intermédiaire d'une liaison hydrogène mais par l'ensemble de l'environnement de la protéine mettant en jeu des interactions électrostatiques, hydrophobes, des contacts de Van der Waals...

### 1.3- Prédiction par amarrage moléculaire pour l'antigène de groupe sanguin O

Les poches **c** et **z,** correspondant aux poches **A** et **B** pour l'antigène de groupe sanguin A, sont les deux poches les plus proches de celles prédites par Tan *et al.* pour l'antigène de groupe sanguin O.

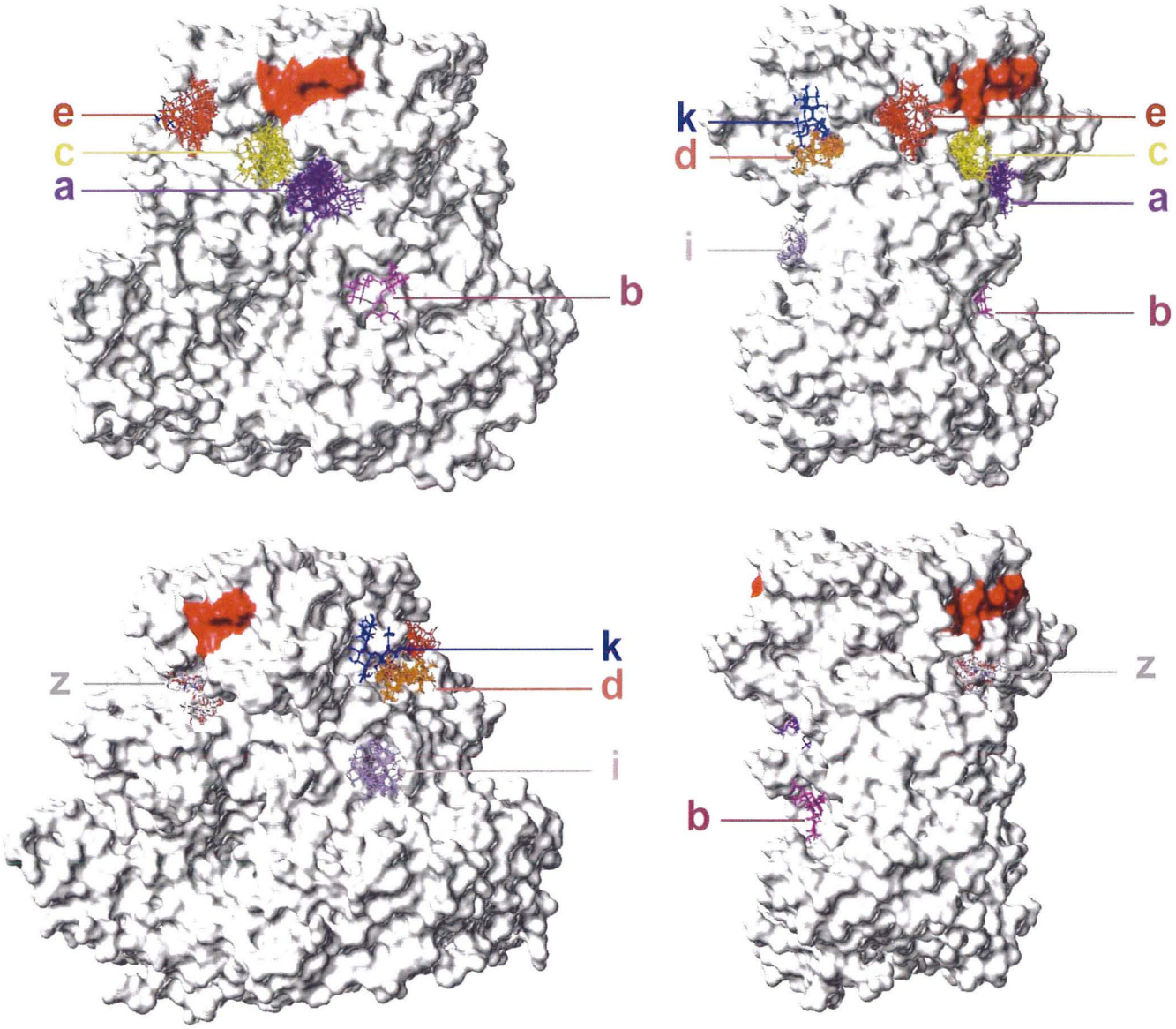



***Figure n°52***
*Surface du dimère du virus de Norwalk coloré en fonction de l'effet des acides aminés mutés sur la reconnaissance de l'antigène de groupe O pour la souche VA387. En rouge apparaissent les acides aminés qui, mutés, diminuent la fixation de l'antigène. Seules les solutions correspondant aux poches déterminées précédemment pour l'antigène de groupe O sont représentées sur cette figure.*

Le trisaccharide représentant H type 1 dans ces deux poches interagit respectivement avec la sérine 251 et l'acide aspartique 305 par formation d'une liaison hydrogène par l'intermédiaire du groupement N-acétyl. Si l'on considère les alignements de séquences effectués

précédemment sur les 12 souches de Norovirus (RHDV exclu), l'acide aspartique est conservé à plus de 80% contrairement à la sérine.

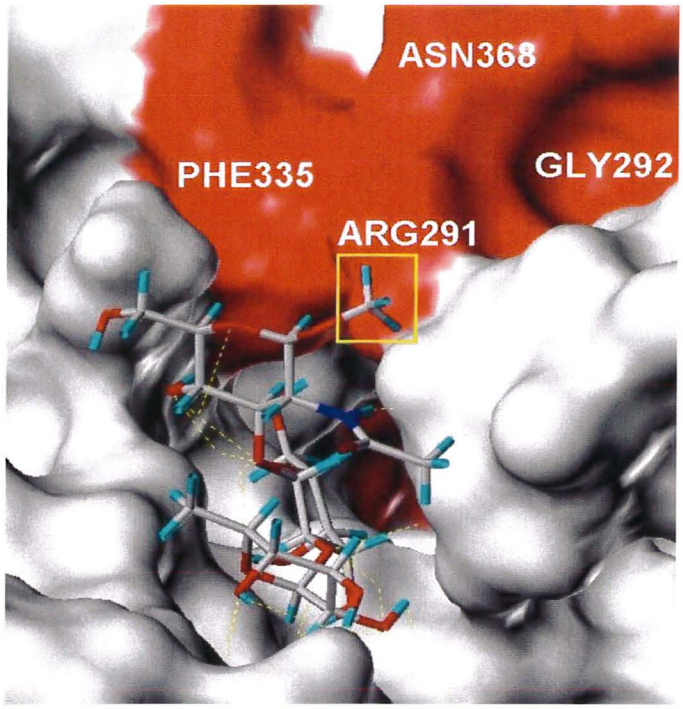



***Figure n°53***
*Trisaccharide représentant l'antigène de groupe O appartenant à la famille B proche des acides aminés mutés dans la souche VA387. Cette solution forme une liaison hydrogène avec l'acide aspartique 305 de la protéine par l'intermédiaire de l'hydrogène du groupement N-acétyl. Le groupement hydroxyméthane du GlcNac, encadré en jaune, représente le point d'ancrage pour le branchement du reste de l'antigène. Ce groupement est nettement orienté vers la région mutée.*

Selon Tan *et al.* les mutations des résidus correspondant à l'arginine 291, la glycine 292, la thréonine 293, l'asparagine 300 du NV diminuent la reconnaissance de l'antigène de groupe sanguin O sur les souches VA387 et MOH.

Après amarrage moléculaire, aucune solution pour l'antigène H type 1 n'a été trouvée interagissant directement avec les résidus 291, 292, 293 ou 300.

Cependant, ces résidus étant très proches de la cavité dans laquelle s'introduisent plusieurs solutions, il semble logique qu'un changement dans ces acides aminés entraîne un

remaniement de la structure tridimensionnelle de la poche qui ne serait alors plus capable de reconnaître le trisaccharide.

Il ne faut pas non plus perdre de vue que l'antigène de groupe O est une glycoprotéine ou un glycolipide fixé à la surface de cellules ou sécrété dans divers fluides. Les tests de reconnaissance des mutants VA387 et MOH ont d'ailleurs été réalisés avec la salive de volontaires contenant donc l'antigène dans sa totalité. Ici, le trisaccharide utilisé ne modélise que l'extrémité terminale de ce glycane et l'atome sur lequel se greffe le reste de la structure est orienté dans plusieurs cas vers la zone mutée. Si la mutation de ces acides aminés sur le virus de Norwalk aboutit comme dans la souche VA387 à une perte de reconnaissance de l'antigène de groupe O, cela pourrait aussi en partie être dû à une gêne du reste de la chaîne glycopeptidique ou glycolipidique dans cette région mutée.

Selon Huang *et al.*, les deux antigènes de groupe sanguin A et O partagent le même site de reconnaissance sur la protéine de capside. En effet, des tests d'inhibition à partir de différents oligosaccharides synthétiques ont montré que les VLPs du virus de Norwalk n'étaient plus capables de fixer un trisaccharide représentant l'antigène de groupe sanguin A lorsqu'ils avaient été préalablement incubés en présence d'oligosaccharides représentant l'antigène H type 1 [14]. Ainsi, les poches He, Kk et Si peuvent constituer des poches qu'il serait également intéressant d'étudier plus avant. La poche Kk présente de bonnes énergies de liaison puisqu'elle est classée troisième pour les deux types de trisaccharides. De plus, elle est très accessible à la surface du virus car elle est presque située au sommet de la protrusion contrairement à Si. He, elle, est située dans la même région que la poche Bc donc assez proche de la zone désignée par Tan *et al.*

Toutes ces poches représentent des régions à explorer plus en détails. L'étude des capacités de reconnaissance de VLPs du virus de Norwalk mutés au niveau de ces poches, notamment au

niveau de l'arginine 291, permettront de confirmer ou d'infirmer les hypothèses établies par amarrage moléculaire.

Les minimisations réalisées avec la méthode POWELL ont permis de trouver des minimums énergétiques pour les complexes. Cependant cette méthode, consistant en une suite de pas effectuée dans une direction choisie de l'espace tridimensionnel, favorise la découverte de minimums énergétiques locaux (voir ***Figure n°18*** p40). La méthode du « recuit-simulé » ou SA permettrait, par des chauffages et des refroidissements successifs du complexe, de surmonter les barrières énergétiques existantes pour trouver la conformation optimale correspondant au minimum global du paysage énergétique.

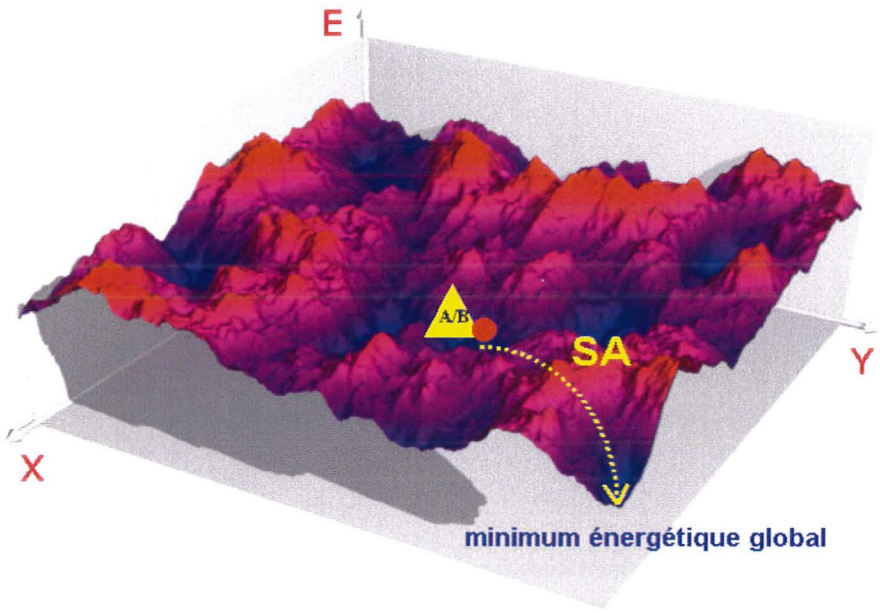



***Figure n°54***
*Le SA permet de trouver le minimum énergétique global dans le paysage du complexe.*

Il serait par la suite particulièrement important d'étudier la solvatation des meilleures solutions obtenues après le « recuit-simulé ». Il est en effet bien connu aujourd'hui que les molécules d'eau sont nettement impliquées dans les interactions protéines-sucres [1]. Watson *et al.* ont rapporté que, dans le cas de la complexation entre le glucose et la glycogène phosphorylase,

les liaisons hydrogènes mettant à contribution l'eau étaient aussi fortes que les liaisons directes sucre-lectine renforçant le concept de l'eau comme une extension de la surface protéique [1] [28]. Un modèle de solvatation des meilleures solutions devra être réalisé grâce à des simulations du complexe dans une boîte d'eau par dynamique moléculaire en utilisant AMBER, ceci afin de voir le comportement du complexe en milieu aqueux et de déterminer le rôle des molécules d'eau dans l'interaction.

Enfin, la souche virale VA207 reconnaît les antigènes de Lewis des individus qu'ils soient sécréteurs ou non contrairement au virus de Norwalk. Il serait intéressant, en mutant la protéine de caspide du NV avec des résidus de la souche VA207, d'étudier la fixation de ces antigènes par amarrage moléculaire pour essayer de déterminer les résidus clés intervenant dans cette reconnaissance.

THESE SOUTENUE PAR : SOLENE GROSDIDIER

TITRE : **VIRUS ET SUCRES : UNE APPROCHE PAR MODELISATION MOLECULAIRE**

# CONCLUSION

Les interactions lectine-sucre ont été négligées en faveur de l'étude des peptides, des protéines, de l'ADN et des interactions que ces différentes molécules peuvent avoir entre elles car on considérait que l'ensemble des informations d'un organisme était codé dans des acides aminés et des nucléotides. Aujourd'hui, on sait qu'en dépit de leur petite taille, les carbohydrates participent également au stockage d'information puis à la communication entre cellules et constituent un véritable glycocode. Les interactions lectine-sucre jouent un rôle primordial dans de nombreux phénomènes biologiques comme par exemple la reconnaissance et l'adhésion sur nos cellules de divers pathogènes tels que les bactéries ou les virus. Leur intérêt s'est donc considérablement développé ces dernières années au vu des applications thérapeutiques potentielles directes. L'élucidation de la structure tridimensionnelle et des propriétés dynamiques des oligosaccharides est nécessaire pour une meilleure compréhension du processus de reconnaissance des lectines. Cependant, les oligosaccharides, dans leur forme libre ou glycoconjuguée, restent difficiles à cristalliser. La modélisation moléculaire est alors devenue un important outil d'étude de ces composés. Les progrès réalisés dans les algorithmes de recherche et la puissance de calcul permettent maintenant de prédire de nouvelles structures oligosaccharidiques comme le modèle du Lstc sous contrainte pour inhiber par compétition l'hémagglutinine du virus de la grippe. La réalisation de simulations où les carbohydrates sont en interaction avec leur lectine est également devenue possible. Ainsi, l'amarrage moléculaire de deux trisaccharides sur la capside du virus de Norwalk aide à explorer la surface du virus pour découvrir des poches de reconnaissance des sucres afin de pouvoir développer ensuite des inhibiteurs de ces interactions. La modélisation moléculaire a donc trouvé aujourd'hui une place à part entière dans l'étude des carbohydrates pour la découverte de nouveaux médicaments.

VU ET PERMIS D'IMPRIMER
Grenoble, le 23/8/2005

LE DOYEN

UNIVERSITE JOSEPH FOURIER
U.F.R. DE PHARMACIE
GRENOBLE 1

P. DEMENGE

LABORATOIRE DE VIROLOGIE MOLÉCULAIRE ET STRUCTURALE

LE PRESIDENT DE THESE

Emmanuel DROUET
Professeur des Universités

E. DROUET

## LISTE DES ABREVIATIONS

1. C.E.R.M.A.V. : Centre d'Etude et de Recherche sur les Macromolécules Végétales
2. I.U.P.A.C. = International of Pure and Applied Chemistry
3. $E_{ang_c}$ = angle constraint energy : énergie associée avec les contraintes d'angles.
4. $E_{bend}$ = angle bending energy : énergie traduisant la variation de courbure des angles entre deux liaisons à partir de leur valeur normale, ou énergie de déformation d'angle.
5. $E_{dist_c}$ = angle constraint energy : énergie associée aux contraintes de distance.
6. $E_{ele}$ = electrostatic energy : énergie correpondant aux interactions électrostatiques.
7. $E_{multi}$ = énergies d'interaction multiple
8. $E_{oop}$ = out of plane energy : énergie correspondant à l'angle entre plusieurs atomes coplanaires et un atome hors de ce plan ou énergie de déformation de plan
9. $E_{range_c}$ = énergie associée aux contraintes
10. $E_{str}$ = bond stretching energy : correspond à l'énergie d'une liaison entre deux atomes lorsque sa longueur varie suite à un étirement ou une compression de celle-ci par rapport à sa distance idéale ou énergie d'élongation de la liaison
11. $E_{tor}$ = torsional energy : correspond à l'énergie de torsion due au vrillage des liaisons.
12. $E_{tors_c}$ = torsion constraint energy : énergie associée avec les contraintes au niveau des angles de torsion.
13. $E_{vdw}$ = van der waals energy : énergie correspondant aux interactions de van der Waals.
14. E.L.I.S.A : Ezyme Linked Immuno Sorbent Assay
15. IEM : Immune Electron Microscopy
16. GA : Genetic Algorithm
17. HBGAs : Histo-Blood Group Antigens
18. ORF : Open Reading Frame, cadre de lecture
19. PDB : Protein Data Bank
20. RMSD : Root Mean Square Deviation
21. RT-PCR : Reverse Transcriptase Polymerase Chain Reaction
22. SA : Simulated Annealing, méthode du « cuit et recuit »
23. Se+ : sécréteur
24. Se- : non sécréteur
25. $Se^{w}$ : sécréteur faible (weak secretor)
26. SRSV : Small Round Structured Virus
27. NLV : Norwalk-like virus
28. NV : Norwalk Virus
29. VLP : Virus-Like Particles

## BIBLIOGRAPHIE


1. Ambrosi M., Cameron NR., Davis BG. Lectins : tools for the molecular understanding of the glycocode.
   *Org. Biomol. Chem.*, 2005, **3**(9) : 1593-608.

2. Atmar RL., Estes MK., Diagnosis of noncultivable gastroenteritis viruses, the human caliciviruses.
   *Clin Microbiol. Rev.*, 2001, **14**(1) :15-37.

3. Bertolotti-Ciaret A., Crawford SE., Hutson AM., Estes MK., The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1 : a novel function for the VP2 protein.
   *J Virol.*, 2003, **77**(21) : 11603-11615.

4. Bertolotti-Ciarlet A., White LJ., Chen R., Prasad BV., Estes MK. Structural requirements for the assembly of norwalk-like particules.
   *J Virol.*, 2002, **76**(8) : 4044-55

5. Bovin NV., Tuzikov AB., Chinarev AA., Gambarayan AS. Multimeric glycotherapeutics: New paradigm.
   *Glycoconj J.*, 2004, **21**(8-9) : 471-478.

6. Chinarev NV., Alexander AB., Gambarayan AS., Mastrosovich MN., Imberty A., Bovin N. Tetravalent blockers for influenza virus hemagglutinin. "Sialobiology and Other Novel Forms of Glycosylation" (Y. Inoue, YC. Lee & FA. Troy, eds)
   *Gakushin Publishing Co*, 1999, 135-143.

7. Colman PM., Varghese JN., Laver WG. Structure of the catalytic and antigenic sites in influenza neuraminidase.
   *Nature*, 1983, **303**(5912):41-4.

8. Dreitlein WB., Maratos J., Brocavich J. Zanamivir and Oseltamivir: two new options for the treatment and prevention of influenza.
   *Clin. Ther.*, 2001 Mar, **23**(3) : 327-55.

9. Fankhauser RL., Noel JS., Monroe SS., Ando T., Glass RI., Molecular epidemiology of "Norwalk-like viruses" in outbreaks of gastroenteritis in the United States.
   *J Infect Dis.*, 1998, **178**(6) : 1571-8.

10. Gaboriaud C. , Bissery V., Benchetrit T., Mornon J.P., Hydrophobic cluster analysis : an efficient new way to compare and analyse amino acid sequences.
    *FEBBS letters*, 1987, **224**(1) : 149-155.

11. Gamblin SJ., Haire LF., Russell RJ., Stevens DJ., Xiao B., Ha Y., Vasisht N., Steinhauer DA., Skehel JJ. The structure and receptor binding properties of the 1918 influenza hemagglutinin.

*Science*, 2004, **303**(5665) : 1838-42.

12.Green KY, Ando T, Balayan MS, Berke T, Clarke IN, Estes MK, Matson DO, Nakata S, Neill JD, Studdert MJ, Thiel HJ. Taxonomy of the caliciviruses.
*J Infect Dis.*, 2000 May, **181** Suppl II : S322-30.

13.Huang P., Farkas T., Marionneau S., Zhong W., Ruvoën-Clouet N., Morrow AL., Altaye M., Pickering LK., Newburg DS., Le Pendu J., Jiang X. Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns.
*J Infect Dis.*, 2003, **188**(1) : 19-31.

14.Huang P., Farkas T., Zhong W., Tan M., Thornton S., Morrow AL., Jiang X. Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns.
*J Virol.*, June 2005, **79**(11) : 6714-6722.

15.Imberty A., Pérez S., Structure, conformation and dynamics of bioactive oligosaccharides : theoretical approaches and experimental validations.
*Chem Rev.*, 2000, **100** : 4567-4588

16.Imberty A., Mikros E., Koca J., Mollicone R., Oriol R., Pérez S. Computer simulation of histo-blood group ologosaccharides : energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens.
*Glycoconj J.*, 1995, **12**(3) : 331-349.

17.Kitchen DB., Decornez H., Furr JR., Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications.
*Nat Rev Drug Discov.*, 2004, **3**(11) : 935-949.

18.Lindesmith L., Moe C., Marionneau S., Ruvoen N., Jiang X., Lindblad L., Stewart P., LePendu J., Baric R. Human susceptibility and resistance to Norwalk virus infection.
*Nature Med.*, 2003, **9**(5) : 548-553.

19.Morris GM., Goodsell DS., Halliday RS., Huey R., Hart WE., Belew RK., Olson AJ. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function.
*J. Comput. Chem.*, 1998, **19**(14) : 1639-1662.

20.Parashar U., Quiroz ES., Mounts AW., Monroe SS., Fankhauser RL., Ando T., Noel JS., Bulens SN., Beard SR., Li JF., Bresee JS., Glass RI., "Norwalk-like viruses". Public health consequences and outbreak management.
*MMWR Recomm Rep.* 2001, **50**(RR-9) : 1-17.

21. Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G., Estes, M. K.: X-Ray Crystallographic Structure of the Norwalk Virus Capsid
*Science*, 1999, **286**(5438) : 287-290.



B. I. U. MED. GRENOBLE

22.Pochechueva TV., Galanina OE., Ushakova NA., Preobrazhenskaya ME., Sablina MA., Nifantiev NE., Tsvetkov YuV., Vozney YaV., Imberty A., Bovin NV. Uncharged P-selectin blockers.
*Glycoconj. J.*, 2004, **20**(2) : 91-97.

23.Powell MJD., Restart procedures for the conjugate gradient method.
*Math. Prog.*, 1977, **12 :** 241-254.

24.Rockx BH., Vennema H., Hoebe CJ., Duizer E., Koopmans MP. Association of histo-blood group antigens and susceptibility to Norovirus infection.
*J Infect Dis.*, 2005, **191**(5) : 749-754.

25.Ruiz-Palacios GM., Cervantes LE., Ramos P., Chavez-Munguia B., Newburg DS. Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection.
*J Biol Chem.*, 2003, **278**(16) : 14112-20.

26.Sharon N., Lis H. History of lectins : from hemagglutinins to biological recognition molecules.
*Glycobiology*, 2004, **14**(11) : 53-62.

27.Tan M., Huang P., Meller J., Zhong W., Farkas T., Jiang X. Mutations within the P2 domain of Norovirus capsid affect binding to human hisot-blood group antigens : evidence for a binding pocket.
*J Virol.*, 2003, **77**(23), 12562-12571.

28.Watson KA., Mitchell EP., Johnson LN., Son JC., Bichard CJ, Orchard MG., Fleet GW., Oikonomakos NG., Leonidas DD., Kontou M., Papageorgioui A. Design of inhibitors of glycogen phosphorylase: a study of alpha- and beta-C-glucosides and 1-thio-beta-D-glucose compounds.
*Biochemistry*, 1994, **33**(19) : 5745-5758.

**LIVRE :**

Sharon N., Lis H. Lectins, 2ème édition, Kluwer Academic Publishers, Dordrecht, 2003.

**SITES INTERNET :**

Center for disease control and prevention
http://www.cdc.gov/ avril 2005

Glycoword
http://www.gak.co.jp/FCCA/glycoword/wordE.html avril 2005

Influenza schéma
http://georges.dolisi.free.fr/Terminologie/Schemas/ août 2005

Infobiogen
http://www.infobiogen.fr mars 2005

Organisation mondiale de la santé
http://www.who.int/fr/ juillet 2005

Pasteur
http://www.pasteur.fr

Talk. Origins
http://www.talkorigins.org mai 2005

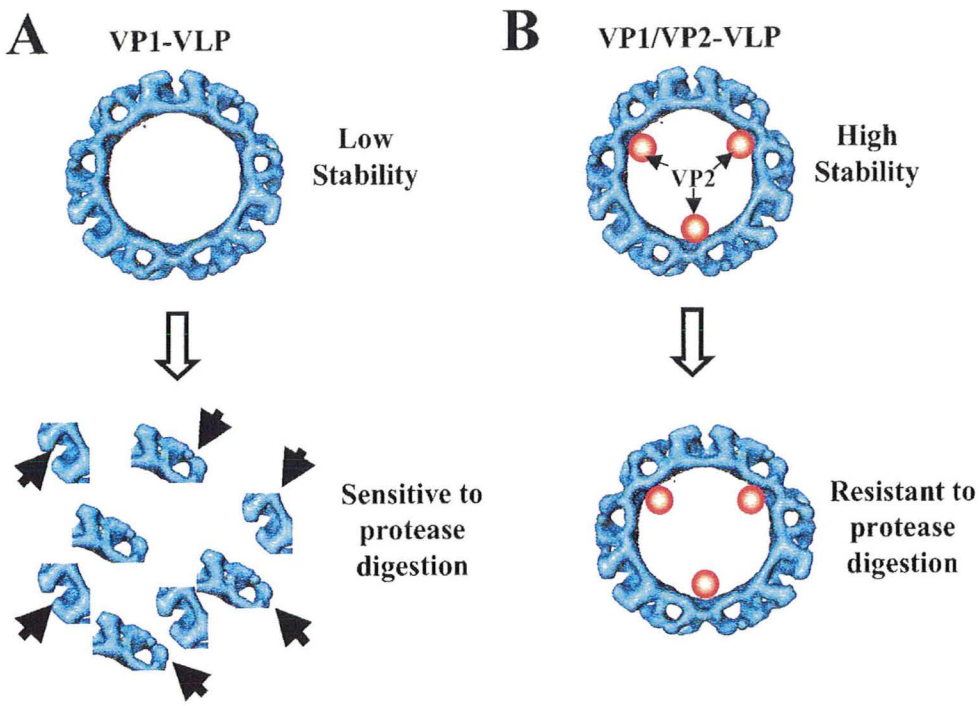
VIPER
http://mmtsb.scripps.edu/viper/ avril 2005

www.allthevirologyonthewww.com mai 2005

Wikipedia
http://en.wikipedia.org avril 2005

# ANNEXES

## ANNEXE A/




*Selon Bertolotti-Ciaret et al. (2003)*
*Rôle de VP2 dans la stabilisation de la capside virale.*

## ANNEXE B/

### Code d'accès aux différentes séquences pour l'alignement

| *Souches* | *# accès EMBL* | *# accès UniProt (SWISSPROT)* |
|---|---|---|
| VA387 | AY038600 | Q913Z3 |
| GrV | AJ004864 | Q9YQS0 |
| Norwalk | M87661 | Q83884 |
| Dijon | AF472623 | Q8QY55 |
| C59 | AF435807 | Q8VA27 |
| MxV | U22498 | Q68291 |
| HV[1] | U07611 | Q68104 |
| MOH | AF397156 | Q91I65 |
| SMV | U75682 | O12305 |
| VA207 | AY038599 | Q91H09 |
| Boxer | AF538679 | Q8BCA3 |
| DSV | U04469 | Q66418 |
| VA115 | AY038598 | Q91H12 |
| RHDV[2] | Z29514[3] | Q86117 |

[1] souche ne fixant pas la salive humaine mais capable de reconnaître les antigènes synthétiques A et B trisaccharides.
[2] virus hémorragique du lapin, ne fixant pas la salive humaine mais ayant la capacité d'agglutiner nos globules rouges.

ANNEXE C/

## AutoDock

## Fichier de paramétrage de l'amarrage moléculaire

```
seed    time pid    # for random number generator
types   CNOH        # atom type names
fld     ab2.maps.fld    # grid data file
map     ab2.C.map       # C-atomic affinity map file
map     ab2.N.map       # N-atomic affinity map file
map     ab2.O.map       # O-atomic affinity map file
map     ab2.H.map       # H-atomic affinity map file
map     ab2.e.map       # electrostatics map file

move    grA1.pdbq       # small molecule file
about   21.275 -3.371 -15.833     # small molecule center

# Initial Translation, Quaternion and Torsions
tran0   random          # initial coordinates/A or "random"
quat0   random          # initial quaternion or "random"
ndihe   14              # number of initial torsions
dihe0   random          # initial torsions

torsdof 0 0.3113  # num. non-Hydrogen torsional DOF & coeff.
#ligand_is_not_inhibitor  # uncomment if small molecule is substrate or T.S.

# Initial Translation, Quaternion and Torsion Step Sizes and Reduction Factors
tstep   2.0             # translation step/A
qstep   50.0            # quaternion step/deg
dstep   50.0            # torsion step/deg
trnrf   1.              # trans reduction factor/per cycle
quarf   1.              # quat reduction factor/per cycle
dihrf   1.              # tors reduction factor/per cycle

# Hard Torsion Constraints
#hardtorcon 1 -180. 30.   # constrain torsion, num., angle(deg), range(deg)

# Internal Non-Bonded Parameters
intnbp_r_eps  4.00 0.0222750  12  6       #C-C lj
intnbp_r_eps  3.75 0.0230026  12  6       #C-N lj
intnbp_r_eps  3.60 0.0257202  12  6       #C-O lj
intnbp_r_eps  3.00 0.0081378  12  6       #C-H lj
intnbp_r_eps  3.50 0.0237600  12  6       #N-N lj
intnbp_r_eps  3.35 0.0265667  12  6       #N-O lj
intnbp_r_eps  2.75 0.0084051  12  6       #N-H lj
intnbp_r_eps  3.20 0.0297000  12  6       #O-O lj
intnbp_r_eps  1.90 0.3280000  12 10       #O-H hb
intnbp_r_eps  2.00 0.0029700  12  6       #H-H lj

#intelec 0.1146          # calculate internal electrostatic energy

# Simulated Annealing Parameters
#rt0 616.                # SA: initial RT
#rtrf 0.95               # SA: RT reduction factor/per cycle
#linear_schedule # SA: do not use geometric cooling
#runs    10              # SA: number of runs
```

```
#cycles 50                  # SA: cycles
#accs   100                 # SA: steps accepted
#rejs   100                 # SA: steps rejected
#select m                   # SA: minimum or last

# Trajectory Parameters (Simulated Annealing Only)
#trjfrq  100                # trajectory frequency
#trjbeg  1                  # start trj output at cycle
#trjend  50                 # end trj output at cycle
#trjout  grA1.trj # trajectory file
#trjsel  E                  # A=acc only;E=either acc or rej

#watch  grA1.watch.pdb  # real-time monitoring file

outlev  1                   # diagnostic output level

# Docked Conformation Clustering Parameters for "analysis" command
rmstol  2.5                 # cluster tolerance (Angstroms)
rmsref  grA1.pdbq           # reference structure file for RMS calc.
#rmsnosym                   # do no symmetry checking in RMS calc.
write_all                   # write all conformations in a cluster

extnrg  1000.               # external grid energy
e0max   0. 10000            # max. allowable initial energy, max. num. retries


# Genetic Algorithm (GA) and Lamarckian Genetic Algorithm Parameters (LGA)
ga_pop_size 100             # number of individuals in population
ga_num_evals 20000000  # maximum number of energy evaluations
ga_num_generations 3000          # maximum number of generations
ga_elitism 1                # num. of top individuals that automatically survive
ga_mutation_rate 0.02       # rate of gene mutation
ga_crossover_rate 0.80      # rate of crossover
ga_window_size 10           # num. of generations for picking worst individual
ga_cauchy_alpha 0           # ~mean of Cauchy distribution for gene mutation
ga_cauchy_beta 1            # ~variance of Cauchy distribution for gene mutation
set_ga                      # set the above parameters for GA or LGA

# Local Search (Solis & Wets) Parameters (for LS alone and for LGA)
sw_max_its 300              # number of iterations of Solis & Wets local search
sw_max_succ 4               # number of consecutive successes before changing rho
sw_max_fail 4               # number of consecutive failures before changing rho
sw_rho 1.0                  # size of local search space to sample
sw_lb_rho 0.01              # lower bound on rho
ls_search_freq 0.06         # probability of performing local search on an indiv.
set_psw1                    # set the above pseudo-Solis & Wets parameters

# Perform Dockings
#do_local_only 50       # do only local search
#do_global_only 10 # do only global search (traditional GA)

#simanneal              # do as many SA runs as set by the "runs" command above

ga_run 100              # do this many GA or LGA runs

# Perform Cluster Analysis
analysis            # do cluster analysis on results
```

ANNEXE D/

## FICHIER SYBYL SPL

### Fichier pour la minimisation du sucre en complexe avec le dimère AB

```
@MACRO
cp_opt sybylbasic

####pour minimiser la proteine AVEC LE SUCRE dans un rayon de 10 A autour
####du sucre avec un FF prenant en compte 20 A autour du sucre

echo "\nATTENTION LES MOLECULES DE CETTE BASE DE DONNEES VONT ETRE MODIFIEES\n"

lookahead $*

setvar mdb %case("%arg(1 %dir(*.mdb))" "database name ? " "" %dir(*.mdb))
 set cgq 0
 database open $mdb update

MOL2 IN m2 AB_H_pullman.mol2


TAILOR SET FORCE_FIELD DIELECTRIC_FUNCTION DISTANCE | |
TAILOR SET MAXIMIN2 MINIMIZATION_METHOD POWELL  RMS_GRADIENT 0.05 | |
TAILOR SET MAXIMIN2 MAXIMUM_ITERATIONS 100000 | |

for i in %database(*)
   zap m1
   database get "$i" m1
   MERGE m2 m1

      REMOVE AGGREGATE m1(*)
      DEFINE AGGREGATE \
      M1((*)-(({Q_1_1}+{SUBST_SPHERE({Q_1_1},5.000000)})-{BACKBONE})) TOTO """"
      list agg m1(*) brief

      REMOVE LOCAL_SET m1({FF_INTERESTING})
      DEFINE STATIC_SET ATOM M1({Q_1_1}+{SUBST_SPHERE({Q_1_1},30.000000)}) \
      ff_interesting """"
      list loc m1(*) brief
      list loc m1(*) brief
      maximin2 m1 electro calc mark done interactive

data add m1 replace
endfor
```

ANNEXE E/

## Fichier pour la minimisation du sucre en présence du dimère mais éloigné de 200Å pour s'affranchir de toutes influences entre sucre et protéines.

```
@MACRO
opt_prot sybylbasic

####pour minimiser la proteine SANS LE SUCRE dans un rayon de 10 A autour
####du sucre avec un FF prenant en compte 20 A autour du sucre

echo "\nATTENTION LES MOLECULES DE CETTE BASE DE DONNEES VONT ETRE MODIFIEES\n"

lookahead $*

setvar mdb %case("%arg(1 %dir(*.mdb))" "database name ? " "" %dir(*.mdb))
 set cgq 0
 database open $mdb update

MOL2 IN m2 AB_centroid.mol2

####MOL2 IN m2 AB_H_pullman.mol2

TAILOR SET FORCE_FIELD DIELECTRIC_FUNCTION DISTANCE | |
TAILOR SET MAXIMIN2 MINIMIZATION_METHOD POWELL  RMS_GRADIENT 0.05 | |
TAILOR SET MAXIMIN2 MAXIMUM_ITERATIONS 100000 | |

for i in %database(*)
   zap m1
   database get "$i" m1
   MERGE m2 m1

      REMOVE AGGREGATE m1(*)
      DEFINE AGGREGATE \
      M1((*)-(({Q_1_1}+{SUBST_SPHERE({Q_1_1},5.000000)})-{BACKBONE})) TOTO """"
      list agg m1(*) brief
      REMOVE LOCAL_SET m1({FF_INTERESTING})
      DEFINE STATIC_SET ATOM M1({Q_1_1}+{SUBST_SPHERE({Q_1_1},30.000000)}) \
      ff_interesting """"
      list loc m1(*) brief

#### rajoute 1 liaison impossible (type du)
#### 15075 etant l atome ID d'un atome rajoute aux coordonnes du centroide
      ADD BOND M1(15075) M1(27) du
      MODIFY DISTANCE M1(15075) M1(27) 1000
      REMOVE BOND M1(15306)
      REMOVE ATOM M1(15075)

####ADD BOND M1(14632) M1(27) du
####MODIFY DISTANCE M1(14632) M1(27) 200
####REMOVE BOND M1(15306)

      maximin2 m1 electro calc mark done interactive

data add m1 replace
endfor
.
```

## ANNEXE F/

### Fichier pour le calcul des énergies après minimisation et évaluation des angles de torsion des liaisons glycosidiques puis l'affichage dans un tableau

```
@MACRO
mdb_E sybylbasic

####pour afficher les energies otpimisees d'une database de sucres dockes sur son recepteur ainsi que les angles de torsions

lookahead $*

if %not( %table_name() )
      echo "There are no spreadsheets currently open"
      return
endif

setvar MssName %case( %table_default() "Spreadsheet" \
      "Spreadsheet to optimise each conf " %table_name() )
  if $status
      return
  endif

set cgq 0

####creer 4 colonnes d'energie et 1 colonne de torsion
####energy column 1
setvar ColumnName1
while %not( $ColumnName1 )
      setvar ColumnName1 %prompt( STRING E_TOT "Name of Energy Column" \
            "Enter a column name" )
      if $status
            return
      endif
endwhile
####energy column 2
setvar ColumnName2
while %not( $ColumnName2 )
      setvar ColumnName2 %prompt( STRING E_SUCRE "Name of Energy Column" \
            "Enter a column name" )
      if $status
            return
      endif
endwhile
####energy column 3
setvar ColumnName3
while %not( $ColumnName3 )
      setvar ColumnName3 %prompt( STRING E_PROT "Name of Energy Column" \
            "Enter a column name" )
      if $status
            return
      endif
endwhile
####energy column 4
setvar ColumnName4
while %not( $ColumnName4 )
      setvar ColumnName4 %prompt( STRING D_E "Name of Energy Column" \
```

```
        "Enter a column name" )
    if $status
        return
    endif
endwhile
####torsion column 1
setvar ColumnNameTorPSIfuc
while %not( $ColumnNameTorPSIfuc )
    setvar ColumnNameTorPSIfuc %prompt( STRING FUC_PSITorsion "Name of Torsion Column PSI" \
        "Enter a column name" )
    if $status
        return
    endif
endwhile
####torsion column 2
setvar ColumnNameTorPHIfuc
while %not( $ColumnNameTorPHIfuc )
    setvar ColumnNameTorPHIfuc %prompt( STRING FUC_PHITorsion "Name of Torsion Column PHI" \
        "Enter a column name" )
    if $status
        return
    endif
endwhile

####torsion column 3 Htype 1
####setvar ColumnNameTorPSIgal
####while %not( $ColumnNameTorPSIgal )
####setvar ColumnNameTorPSIgal %prompt( STRING GAL_PSITorsion "Name of Torsion Column PSI" \
####"Enter a column name" )
####if $status
####return
####endif
####endwhile

####torsion column 3 Antigene A
setvar ColumnNameTorPSIgalNAc
while %not( $ColumnNameTorPSIgalNAc )
    setvar ColumnNameTorPSIgalNAc %prompt( STRING GALNAc_PSITorsion "Name of Torsion Column
PSI" \
        "Enter a column name" )
    if $status
        return
    endif
endwhile

####torsion column 4 Htype 1
####setvar ColumnNameTorPHIgal
####while %not( $ColumnNameTorPHIgal )
####setvar ColumnNameTorPHIgal %prompt( STRING GAL_PHITorsion "Name of Torsion Column PHI" \
####"Enter a column name" )
####if $status
####return
####endif
####endwhile

####torsion column 4 Antigene A
setvar ColumnNameTorPHIgalNAc
while %not( $ColumnNameTorPHIgalNAc )
    setvar ColumnNameTorPHIgalNAc %prompt( STRING GALNAc_PHITorsion "Name of Torsion Column
PHI" \
```

```
        "Enter a column name" )
    if $status
        return
    endif
endwhile

####remplie la colonne 1 energie totale

for mol in %mols(*)
    UNDISPLAY $mol
endfor

TAILOR SET FORCE_FIELD DIELECTRIC_FUNCTION DISTANCE | |

setvar old_area $DEFAULT_AREA
DISPLAY %table_attribute( mol_area ) >$NULLDEV
DEFAULT %table_attribute( mol_area ) >$NULLDEV

TABLE EXAMINE SHOW YES INPUT ROW_COLUMN 1  | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnName1 | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnName2 | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnName3 | >$NULLDEV
TABLE COLUMN_APPEND FUNCTIONAL_DATA $ColumnName4 "@1-@2-@3" | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnNameTorPSIfuc | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnNameTorPHIfuc | >$NULLDEV
#TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnNameTorPSIgal | >$NULLDEV
#TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnNameTorPHIgal | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnNameTorPSIgalNAc | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA $ColumnNameTorPHIgalNAc | >$NULLDEV

for row in %table( *  row number )
    TABLE EXAMINE SHOW YES INPUT ROW_COLUMN $row |    >$NULLDEV
    CONSTR REMOVE TORSION $DEFAULT_AREA
    REMOVE AGGREGATE $DEFAULT_AREA(*)
    DEFINE AGGREGATE $DEFAULT_AREA({Q_1_1}) sucre """"
    REMOVE LOCAL_SET $DEFAULT_AREA({FF_INTERESTING})
    ENERGY $DEFAULT_AREA DONE
    TABLE ENTER CELL $row $ColumnName1 $energy_total >$NULLDEV
    TABLE ENTER CELL $row $ColumnName2 $energy_agg_total >$NULLDEV
    REMOVE AGGREGATE $DEFAULT_AREA
    DEFINE AGGREGATE $DEFAULT_AREA((*)-{Q_1_1}) proteine """"
    LIST AGGREGATE $DEFAULT_AREA brief
    ENERGY $DEFAULT_AREA DONE
    TABLE ENTER CELL $row $ColumnName3 $energy_agg_total >$NULLDEV

    TABLE EVALUATE ALL_CELLS * 4 >$NULLDEV
    TABLE UPDATE

####mesure angle de torsion des liaisons glycosidiques de Htype1 a partir
####des atomes ID

####MEASURE TORSION  $DEFAULT_AREA(8) $DEFAULT_AREA(1) $DEFAULT_AREA(7) \
####$DEFAULT_AREA(15) |    >$NULLDEV
####TABLE ENTER CELL $row $ColumnNameTorPSIfuc $measure_torsion >$NULLDEV
####MEASURE TORSION  $DEFAULT_AREA(1) $DEFAULT_AREA(7) $DEFAULT_AREA(15) \
####$DEFAULT_AREA(16) |    >$NULLDEV
####TABLE ENTER CELL $row $ColumnNameTorPHIfuc $measure_torsion >$NULLDEV

####MEASURE TORSION  $DEFAULT_AREA(27) $DEFAULT_AREA(28) $DEFAULT_AREA(29) \
####$DEFAULT_AREA(44) |    >$NULLDEV
```

```
####TABLE ENTER CELL $row $ColumnNameTorPSIgal $measure_torsion >$NULLDEV
####MEASURE TORSION  $DEFAULT_AREA(26) $DEFAULT_AREA(27) $DEFAULT_AREA(28) \
####$DEFAULT_AREA(29) |    >$NULLDEV
####TABLE ENTER CELL $row $ColumnNameTorPHIgal $measure_torsion >$NULLDEV

####mesure angle de torsion des liaisons glycosidiques de l'antigene de type A
####a partir des atomes ID

MEASURE TORSION  $DEFAULT_AREA(7) $DEFAULT_AREA(1) $DEFAULT_AREA(8) \
$DEFAULT_AREA(9) |    >$NULLDEV
TABLE ENTER CELL $row $ColumnNameTorPHIfuc $measure_torsion >$NULLDEV
MEASURE TORSION  $DEFAULT_AREA(1) $DEFAULT_AREA(8) $DEFAULT_AREA(9) \
$DEFAULT_AREA(21) |    >$NULLDEV
TABLE ENTER CELL $row $ColumnNameTorPSIfuc $measure_torsion >$NULLDEV

MEASURE TORSION  $DEFAULT_AREA(24) $DEFAULT_AREA(23) $DEFAULT_AREA(22) \
$DEFAULT_AREA(21) |    >$NULLDEV
TABLE ENTER CELL $row $ColumnNameTorPHIgalNAc $measure_torsion >$NULLDEV
MEASURE TORSION  $DEFAULT_AREA(23) $DEFAULT_AREA(22) $DEFAULT_AREA(21) \
$DEFAULT_AREA(18) |    >$NULLDEV
TABLE ENTER CELL $row $ColumnNameTorPSIgalNAc $measure_torsion >$NULLDEV

endfor

TABLE UPDATE

DEFAULT $old_area >$NULLDEV
.
```

## ANNEXE G/

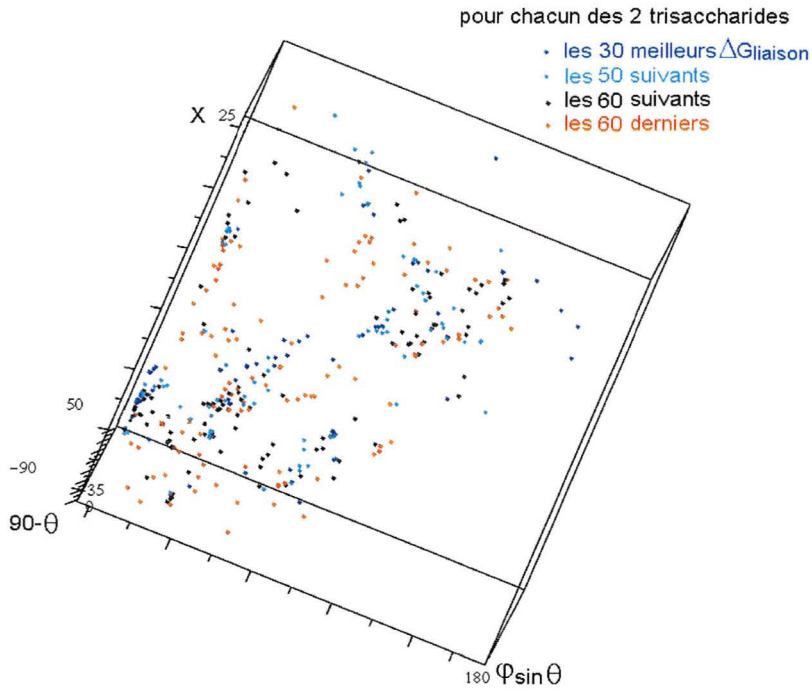
**Fichier pour le calcul des coordonnées sphériques du « pseudo-centroïde » de chaque sucre.**

```
@MACRO
2D sybylbasic

lookahead $*

setvar mdb %case("%arg(1 %dir(*.mdb))" "database name ? " "" %dir(*.mdb))
  set cgq 0
  database open $mdb readonly


zap(*)
MOL2 IN m2 box2.mol2


TABLE CREATE D2 DATABASE "2D" $mdb $DEFAULT_AREA  %DATABASE(*) | |

TABLE EXAMINE SHOW YES INPUT ROW_COLUMN 1  | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA TETA | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA PHI | >$NULLDEV
TABLE COLUMN_APPEND FUNCTIONAL_DATA PHIsinTETA "@2*SIND(@1)" | >$NULLDEV
TABLE COLUMN_APPEND FUNCTIONAL_DATA TETAordo "90-@1" | >$NULLDEV
TABLE COLUMN_APPEND EXPLICIT_DATA coordX | >$NULLDEV

for mol in %mols(*)
         DISPLAY $mol
endfor

for row in %table( *  row number )
    TABLE EXAMINE SHOW YES INPUT ROW_COLUMN $row |      >$NULLDEV
    DATABASE GET "$i" m1
    MERGE m2 m1

setvar X %ATOM_INFO((28) x)
setvar Y %ATOM_INFO((28) y)
setvar Z %ATOM_INFO((28) z)

####mesure l'angle theta
     MEASURE ANGLE M1(28) M1(15075) M1(15079)  | |
     TABLE ENTER CELL $row teta $measure_angle >$NULLDEV

####creer un rawatom ayant ID 15093 (projection du pseudo centre de masse
sur le plan YZ
     ADD RAWATOM M1 proj C.3 2.0186 $Y $Z
     MEASURE ANGLE M1(15083) M1(15075) M1(15076)  | |
     TABLE ENTER CELL $row PHI $measure_angle >$NULLDEV
     TABLE EVALUATE NEW_ONLY $row PHIsinTETA
     TABLE EVALUATE NEW_ONLY $row TETAordo
     TABLE ENTER CELL $row coordX $X

endfor
TABLE UPDATE
TABLE SAVE coord2_A2
```

```
DATABASE CLOSE

#zap(*)

#TAILOR SET GRAPHS

#TABLE GRAPH SCATTER D1 $row PHIsinTETA TETAordo coordX \

.
```

## ANNEXE H/

**Représentation en 3D coordonnées sphériques des pseudo-centroïdes de chaque solution**



Faculté de Pharmacie de Grenoble

**Serment de Galien**

Je jure, en présence des maîtres de la faculté, des conseillers de l'ordre des pharmaciens et de mes condisciples :

D'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement.

D'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement.

De ne jamais oublier ma responsabilité et mes devoirs envers le malade et sa dignité humaine ; en aucun cas je ne consentirai à utiliser mes connaissances et mon état pour corrompre les moeurs et favoriser des actes criminels.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.



B. I. U. MÉD. GRENOBLE

# VIRUS ET SUCRES :
# UNE APPROCHE PAR MODELISATION MOLECULAIRE


## Résumé de la thèse

Les bactéries et les virus sont nombreux à reconnaître nos cellules par l'intermédiaire des sucres présents à leur surface grâce à des protéines spécifiques des oligosaccharides apellées lectines. L'objectif de cette thèse a été de modéliser le comportement d'oligosaccharides impliqués dans la reconnaissance de nos cellules par deux virus : le virus de la grippe et le virus de Norwalk. Dans un premier temps, un pentasaccharide sous contrainte a été créé par mécanique moléculaire afin d'améliorer sa reconnaissance par l'hémagglutinine du virus de l'Influenza en vue de réaliser des inhibiteurs compétitifs de cette lectine. Puis, une étude d'interaction par amarrage moléculaire entre les lectines d'un Norovirus et de deux oligosaccharides a été réalisée afin d'identifier des poches potentielles de fixation et d'établir des hypothèses pouvant expliquer le profil d'infection particulier de ce virus.



**Mots clés : Virus, Sucres, Modélisation moléculaire, SYBYL, AutoDock**



Date de soutenance : 1/09/05
Auteur : Solène Grosdidier