

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

Discipline : SCIENCES DE LA VIE ET DE LA SANTE

Spécialité : Bio-informatique

Présentée et soutenue publiquement par

CAMILLE BESANÇON

Le 9 décembre 2019

**DOGME - Développement d'un Outil de visualisation moléculaire et application à l'étude
in silico des Glycosylations de protéines de la Matrice Extracellulaire.**

Thèse dirigée par **STEPHANIE BAUD, JESSICA JONQUET ET NICOLAS BELLOY**

JURY

M. Marc BAADEN,	Directeur de Recherche,	Institut de Biologie Physicochimique (IBPC),	Président
Mme Stéphanie BAUD,	Maître de Conférences HDR,	Université de Reims Champagne-Ardenne,	Directrice de thèse
Mme Jessica JONQUET,	Maître de Conférences,	Université de Reims Champagne-Ardenne,	Co-Directrice de thèse
M. Nicolas BELLOY,	Ingénieur de Recherches,	Université de Reims Champagne-Ardenne,	Co-Directeur de thèse
Mme Annick DEJAEGERE,	Professeur,	Université de Strasbourg,	Rapporteur
M. Philippe HÜNENBERGER,	Professeur,	Zurich ETH, Suisse,	Rapporteur
Mme Elisa FADDA,	Docteur,	Université de Maynooth, Irlande,	Examinatrice

Remerciements

Ma thèse de doctorat s’est déroulée au sein de l’unité UMR URCA/CNRS 7369 « Matrice Extracellulaire et Dynamique Cellulaire » (MEDyC) et je souhaite, dans un premier temps, remercier son directeur, le Professeur Laurent MARTINY.

Je présente tous mes remerciements aux professeurs Annick DEJAEGERE et Philippe HÜNENBERGER de me faire l’honneur d’évaluer ce travail de thèse en tant que rapporteurs.

J’adresse tous mes remerciements au docteur Elisa FADDA et au professeur Marc BAADEN pour avoir accepté de siéger à mon jury de thèse en tant qu’examinateurs de ce travail de thèse.

Enfin, je tiens à exprimer ma profonde gratitude au professeur Stéphanie BAUD et aux docteurs Jessica JONQUET et Nicolas BELLOY pour leur encadrement au cours de cette thèse.

Un travail de thèse ne peut être mené à bien sans le soutien financier de nombreux organismes. Ainsi, je tiens à adresser mes remerciements à la région Champagne-Ardenne / région Grand-Est, au Centre National de la Recherche Scientifique et à l’Université de Reims Champagne-Ardenne. Je remercie également les centres de calculs ROMEO et CINES pour l’accès aux supercalculateurs m’ayant permis d’obtenir mes données.

Enfin, il est sûrement temps de remercier ceux qui m’ont entouré pendant ces trois ans. Au-delà des remerciements formels, je tiens à remercier une fois de plus Nanie, Jess et Nico pour leur encadrement, leurs conseils et leur soutien tout au long de cette thèse. J’ai beaucoup appris grâce à vous, tant scientifiquement qu’humainement et vous avez été de ceux qui m’ont aidée un peu à sortir de ma coquille. Je tiens également à remercier le professeur classe ex’ Manu Dauchez pour ses encouragements et sa rengaine d’entraîneur sportif : « On lâche rien, on reste fort sur les appuis ! » Je garde tous vos conseils en tête. Je pense que vous savez tous que j’ai cette tendance à me sous-estimer, mais même si je n’en ai pas parlé, c’est un peu aussi grâce à vous que je me sens un peu plus à l’aise dans mes bottes de scientifique.

Bien sûr, je me dois de remercier mes camarades de bureau, Jean-Marc, Hua, Aïcha, Camille et Camille. Derrière le calme tranquille qui règne pendant la journée, je sais que la promesse d’une petite binouze – ou d’un jus de fruit – suffit à révéler vos vraies personnalités... ;) Même si je vais partir, j’aurais Hua et Aïcha à l’œil grâce à ce formidable outil qu’est WhatsApp et je compte sur le premier pour prendre soin de la seconde en mon absence... Sache, mon cher Hua, que Meiko continuera toujours de te juger et que je suis toujours prête à t’affronter dans un combat Pokémon sans merci ! Aïcha, toi qui entres dans seconde année de thèse, c’est là que les choses sérieuses commencent, les plus flippantes mais aussi les plus excitantes ! Fais attention à toi, et ne lâche rien. Ensuite, je tiens à léguer solennellement ma place de bureau à Camille D. Profite de cette place, c’est sans doute la meilleure. A toi les courants d’air en été et le radiateur en hiver ! Il me faudrait sans doute quatre pages pour remercier l’intégralité de l’équipe MIME comme il se doit, mais sachez que ça aura été un réel plaisir de travailler (et plaisanter) avec vous tous. Je vous souhaite de garder la même énergie pour la suite. Ah, et je dois avouer une chose

à tous les rémois qui me lisent : Je n'ai jamais été très fan de champagne, et je ne le suis toujours pas. . .

Évidemment, je me dois aussi de remercier mes parents, qui ont toujours été là pour m'aider (moralement) et refaire le monde pendant deux heures au téléphone, manger les riches et adopter tous les chats du voisinage. . . Je sais quelle chance j'ai de m'entendre aussi bien avec vous, même si j'ai été profondément vexée (c'est faux) que vous osiez me demander si j'allais vous inviter à la soutenance. Comme si j'allais vous oublier.

A ma sœur, j'adresse un non-remerciement tout spécial : La gueuse, pas-merci pour rien. Pas-merci d'être là, pas-merci de m'envoyer des photos trop mignonnes de ton chat, pas-merci d'être la pire des frangines. ;)

Finalement, je n'oublie pas la petite bande d'amis qui m'ont supportée dans tous les sens du terme. J'ai eu la chance d'être bien entourée, et il faut bien le dire : ça aide ! Raphaël et Yannis, on se connaît depuis combien de temps déjà ? Quinze ans ? Pourquoi vous me parlez encore ?? :O On ne se voit peut et je ne suis pas la meilleure pour donner des nouvelles, mais vous êtes des amis comme on en a peu. Agathe et Max, merci encore pour les week-ends et les petits moments de détente pour faire passer mes coups de déprime. Et bientôt, ce sera à moi de vous dire de faire un jet d'initiative. . . :3

Un gros gros merci également à tous ceux qui sont loin des yeux, mais pas du cœur. Aurélie, Micha et Pascalou, le trio de choc tellement fluffly qu'on se noie dans la choupitude. Mon Poussaing de Marseille, sache que nos échanges de photos d'animaux mignons étaient un p'tit rayon de soleil dans mes journées. Merci d'être the best Poussaing et prends soin de toi. Et pour conclure, un petit salut à JB et Nitix, les oiseaux de nuit du serveur Discord que j'ai pas mal écouté blablater pendant que je tentais de bosser. . .

Glossaire

A

AIMD *Ab Initio* Molecular Dynamics.

C

CARP *C*Arbohydrate *R*amachandran *P*lot.

CRE Complexe Récepteur à l'Elastine.

CSS *C*arbohydrate *S*tructure *S*uite.

E

EBP *E*lastin *B*inding *P*rotein.

G

GAG GlycosAminoGlycane.

L

LINUCS *L*inear *N*otation *f*or *U*nique *d*escription *o*f *C*arbohydrate *S*equences.

LRR *L*eucine *R*ich *R*epet.

M

MEC Matrice ExtraCellulaire.

N

Neu-1 Neuraminidase-1.

P

PDB *P*rotein *D*ata *B*ank.

R

RI Récepteur à l'Insuline.

RMN Résonnance Magnétique Nucléaire.

S

SLRP *Small Leucine Rich Protein.*

Table des matières

1	Introduction	1
1.1	Impact de la désialylation du récepteur à l'insuline	1
1.1.1	Le récepteur à l'insuline : structure et fonction	1
1.1.2	Désialylation du récepteur à l'insuline par le complexe récepteur à l'élastine	3
1.2	Acides sialiques et flexibilité des glycanes	4
1.2.1	Caractérisation des conformations principales : méthodes de clustering	4
1.2.2	Développement de l' <i>Umbrella Visualization</i> et problèmes soulevés pendant l'étude	6
1.2.3	Vers une nouvelle approche de l'étude des glycosylations dans le cadre du Projet DOGME	7
2	Contexte scientifique	9
2.1	Biologie de la matrice extracellulaire	9
2.1.1	Organisation générale de la MEC	9
2.1.2	Les protéoglycanes	10
2.1.3	Les glycoprotéines	11
2.1.4	Les collagènes	11
2.1.5	L'élastine	11
2.2	Glycosylations	12
2.2.1	Liaison glycosidique et diversité structurale des glycanes	12
2.2.2	Représentation schématique et nomenclature SNFG	13
2.2.3	Protéoglycanes et glycoprotéines	13
2.2.4	N-glycosylations	13
2.3	Glycosylations et modélisation moléculaire	16
2.3.1	Identification des sites de glycosylation	16
2.3.2	Analyse des structures	17
2.3.3	Construction des systèmes Protéine/Glycanes	18
2.3.4	Travaux connexes	19

2.4	Visualisation scientifique	20
2.4.1	Visualisation scientifique : définition et enjeux principaux	20
2.4.2	Logiciels de visualisation moléculaire : historique et évolutions	22
2.4.3	Modes de représentations classiques des objets biologiques	23
3	Matériel et méthodes	29
3.1	Dynamique moléculaire : principes et usages	29
3.1.1	Principes de la dynamique moléculaire	29
3.1.2	Mise en place d’une expérience de dynamique moléculaire	35
3.1.3	Préparation et paramètres utilisés lors des simulations	39
3.1.4	Analyses des trajectoires	45
3.2	Méthodes de visualisation appliquées aux sucres	47
3.2.1	<i>Umbrella Visualization</i> : une approche 2D	47
3.2.2	UnityMol : un logiciel pour une implémentation en 3D	49
3.2.3	Moteur de jeux vidéos Unity	51
4	Caractérisation du comportement dynamique de chaînes sucrées : de la chaîne isolée aux glycosaminoglycanes	53
4.1	Constitution d’une librairie de glycanes	53
4.2	Analyses d’expériences de dynamique moléculaire portant sur des glycanes en chaînes isolées	55
4.2.1	Caractérisation de glycanes bi-antennés bisectants	55
4.2.2	Caractérisation de glycanes tétra-antennés	58
4.3	Fibromoduline et chaînes de keratan sulfate	66
5	Développement d’un nouvel outil de visualisation intégré à UnityMol	73
5.1	Problématique liée à la représentation des chaînes de glycosylation	73
5.2	Adaptation de l’ <i>Umbrella Visualization</i> à UnityMol	74
5.2.1	Présentation des éléments clés utilisés lors de l’implémentation	74
5.2.2	Implémentation de la méthodologie <i>Umbrella Visualization</i> au logiciel UnityMol	77
5.2.3	Intégration et rendu au sein de l’interface de UnityMol	84
5.3	Application et utilisation de l’ <i>Umbrella Visualization</i> pour décrypter l’impact de la désialylation	89
5.3.1	Comparaison de l’ <i>Umbrella Visualization</i> avec d’autres méthodes	89
5.3.2	Comportement des glycanes bi-antennés	91
5.3.3	Comportement des glycanes tri-antennés	94

6	Discussion et conclusion	101
6.1	Caractérisation de la flexibilité et de la dynamique de chaînes sucrées	101
6.1.1	Propriétés intrinsèques de glycanes isolés	101
6.1.2	Étude de l'interaction réciproque protéine / glycane	103
6.2	Apports et perspectives offerts grâce à l'implémentation de l' <i>Umbrella Visualization</i> sous UnityMol	104
6.3	Désialylation du Récepteur à l'insuline : pertinence de la visualisation scientifique dans la compréhension des conséquences moléculaires associées	107
6.4	Conclusion générale	109
	Annexes	111

Table des figures

1.1	Structure et organisation du récepteur à l'insuline.	2
1.2	Représentation schématique du complexe récepteur à l'élastine.	3
1.3	Familles conformationnelles des glycanes bi-antennés.	5
1.4	Structure des liaisons α 1-6 et α 1-3	6
2.1	Organisation globale de la matrice extracellulaire.	10
2.2	Synthèse des N-glycanes	15
2.3	Premières représentations scientifiques historiques.	21
2.4	Représentations en fil de fer et en bâtons sur la fibromoduline.	24
2.5	Représentations appliquées à la fibromoduline.	25
2.6	Principe de la représentation en surface.	26
2.7	Exemple de coloration de la fibromoduline en fonction de son facteur thermique.	27
3.1	Potentiel d'énergie en fonction de la longueur de liaison et de l'angle entre les atomes.	32
3.2	Fonction périodique décrivant l'énergie en fonction de la position des atomes d'un angle dièdre considéré.	33
3.3	Fonction d'énergie des interactions non-liées.	34
3.4	Processus d'une étape de dynamique moléculaire.	35
3.5	Exemple de système de grande taille étudié par des trajectoires de dynamique moléculaire.	36
3.6	Exemple de fichier au format pdb sur une alanine.	37
3.7	Principe de la minimisation d'énergie.	38
3.8	Illustration des conditions périodiques aux limites.	38
3.9	Structures des glycanes étudiés par les simulations de dynamique moléculaire.	40
3.10	Glycane utilisé sur la fibromoduline.	43
3.11	Structure de l'asparagine.	44
3.12	Structure du récepteur à l'insuline et glycanes utilisés pour les trajectoires.	45

3.13	Principe de l' <i>Umbrella Visualization</i>	48
3.14	Vecteurs entre les centres de masse des glycanes.	48
4.1	Exemples de glycanes intégrés à la librairie.	54
4.2	Résultat du clustering sur les glycanes bisectants.	56
4.3	Résultat de l' <i>Umbrella Visualization</i> sur les glycanes bisectants.	57
4.4	Résultat de l' <i>Umbrella Visualization</i> sur les glycanes tétra-antennés, sur 30 000 structures.	59
4.5	Mesure des valeurs d'angles dièdres au cours du temps pour la liaison mannose α 1-6.	61
4.6	Distribution des valeurs d'angles adoptées par la liaison mannose α 1-6.	62
4.7	Résultat de l' <i>Umbrella Visualization</i> sur les glycanes tétra-antennés sans acides sialiques et avec acides sialiques liés en α 2-3, sur 20 000 structures.	63
4.8	Conformation majoritaire issue du clustering pour les trois types de glycanes. . .	64
4.9	Comparaison des conformations principales du glycanes tétra-antenné avec acides sialiques liés en α 2-3 obtenues sur la trajectoire 1 d'une part et sur l'ensemble des trois trajectoires d'autre part.	65
4.10	Repliement des chaînes de glycosaminoglycanes autour de la fibromoduline. . . .	67
4.11	Détail des points de contacts des glycosaminoglycanes repliés sur la fibromoduline.	68
4.12	Distance au cours du temps entre les centres de masse des groupements sulfate et des résidus des points de contact.	69
4.13	Comparaison entre la position du premier glycanes replié et le LRR7 de la fibromoduline.	70
5.1	Rotation et déplacement du plan par rapport à l'asparagine glycosylée	79
5.2	Représentation des atomes pour l' <i>Umbrella Visualization</i>	80
5.3	<i>Umbrella Visualization</i> appliquée à une seule structure.	80
5.4	Méthodes de projection en fonction des données à analyser.	82
5.5	Méthode de compilation des ombres.	83
5.6	Atténuation de l'ombre en fonction de l'orientation de la surface et de la distance.	84
5.7	Interface utilisateur de UnityMol.	86
5.8	Méthodes de représentation du résultat final.	87
5.9	Résultats de l' <i>Umbrella Visualization</i> en fonction du nombre d'images combinées.	88
5.10	Comparaison entre les résultats de clustering et l' <i>Umbrella Visualization</i>	90
5.11	Résultats du clustering des glycanes Gb1 et Gt1.	91
5.12	Conformations majoritaires des glycanes bi-antennés.	92

5.13	Présence de résidus acides accepteurs de liaisons hydrogène à proximité du site de glycosylation.	93
5.14	Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l' <i>Umbrella Visualization</i> pour les deux glycanes bi-antennés fucosylés sans et avec acides sialiques.	95
5.15	Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l' <i>Umbrella Visualization</i> pour les deux glycanes bi-antennés non-fucosylés sans et avec acides sialiques.	96
5.16	Conformations majoritaires des glycanes tri-antennés.	97
5.17	Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l' <i>Umbrella Visualization</i> pour les deux glycanes tri-antennés fucosylés sans et avec acides sialiques.	99
5.18	Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l' <i>Umbrella Visualization</i> pour les deux glycanes tri-antennés non-fucosylés sans et avec acides sialiques.	100

Liste des tableaux

3.1	Détails des systèmes de glycanes étudiés par les trajectoires de dynamique moléculaire des glycanes tétra-antennés.	41
3.2	Détails des systèmes de glycanes utilisés pour les trajectoires de dynamique moléculaire des glycanes bi-antennés bisectant.	41
3.3	Modifications effectuées sur les charges de l'asparagine glycosylée afin de neutraliser le système.	44
3.4	Composition des systèmes de fibromoduline utilisés pour les trajectoires de dynamique moléculaire.	44
3.5	Cut-off des clusters sur les trajectoires du récepteur à l'insuline (RI), en nm. . . .	46
3.6	Cut-off des clusters sur les trajectoires des glycanes isolés, en nm.	46
3.7	Logiciels considérés pour l'implémentation de l' <i>Umbrella Visualization</i>	50
5.1	Tableau récapitulant les intervalles de valeurs r, g, b des pixels pour la représentation en courbes de niveaux.	86

Chapitre 1

Introduction

1.1 Impact de la désialylation du récepteur à l'insuline

1.1.1 Le récepteur à l'insuline : structure et fonction

Le récepteur à l'insuline est une protéine transmembranaire composée de deux sous-unités α et β (130 et 95 kDa, respectivement). La sous-unité α est exclusivement extracellulaire. La sous-unité β comporte environ 190 acides aminés extracellulaires, une hélice transmembranaire et environ 400 acides aminés intracellulaires [1]. Ces deux sous-unités sont reliées par des ponts disulfures entre les cystéines $\alpha 647$ et $\beta 872$, formant ainsi un monomère $\alpha\beta$. Ces deux sous-unités assemblées vont ensuite former un dimère fonctionnel $\alpha 2\beta 2$ maintenu par deux ponts disulfures supplémentaires [2, 3].

La chaîne α comporte deux domaines riches en leucine L1 et L2 séparés par un domaine riche en cystéine C. Elle comporte également un premier domaine fibronectine complet. Un second domaine fibronectine et un domaine d'insertion sont répartis entre les chaînes α et β . La chaîne β porte un troisième domaine fibronectine complet, un domaine transmembranaire/juxtamembranaire et un domaine Tyrosine Kinase (domaine TK) suivi d'une queue C-terminale [4]. Ce dernier domaine comporte de nombreuses tyrosines qui sont phosphorylées à l'activation du récepteur [5]. L'ensemble de ces domaines et leur organisation sur la structure tridimensionnelle sont représentés sur la figure 1.1 A.

Deux isoformes du récepteur à l'insuline existent, appelées A et B. L'isoforme A est formée grâce à l'exclusion de l'exon 11 par épissage alternatif, contrairement à l'isoforme B qui contient toujours cet exon. Ces deux isoformes peuvent dimériser de façon indifférente et former des dimères A-A, B-A ou B-B [6]. Les différences entre les isomères sont mal caractérisées mais celles-ci présentent des affinités légèrement différentes pour l'insuline [7].

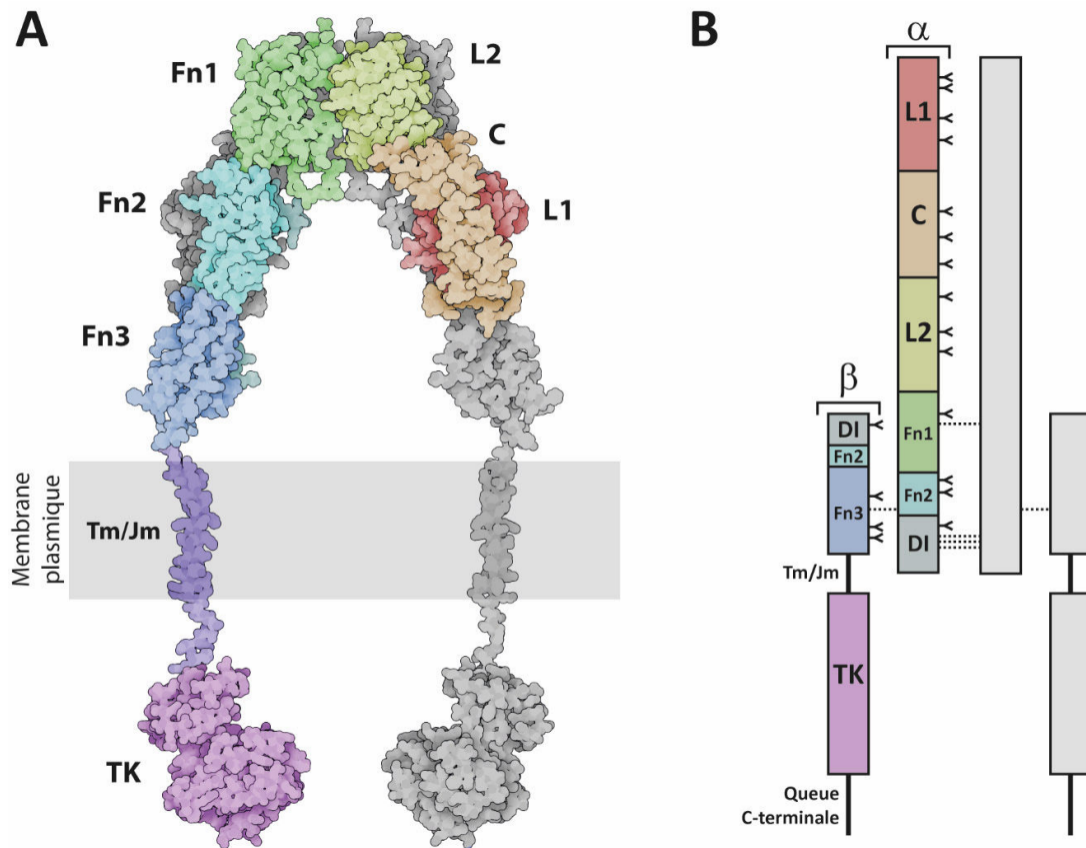


FIGURE 1.1 – **Structure et organisation du récepteur à l'insuline.** Les différents domaines du récepteur sont mis en évidence par couleur : les domaines fibronectines (Fn1, Fn2 et Fn3), les domaines riches en leucine L1 et L2, le domaine riche en cystéines C, le domaine transmembranaire/juxtamembranaire Tm/Jm et le domaine tyrosine kinase (TK). **A** : Structure tridimensionnelle du récepteur à l'insuline. Les domaines sont colorés pour une seule sous-unité (chaîne α et β). **B** : Représentation schématique du récepteur à l'insuline et de ses domaines. Les ponts disulfures sont en pointillés et les points potentiels de N-glycosylation sont indiqués par les symboles "Y". Image par A. Guillot [8]

Le récepteur à l'insuline possède une activité tyrosine kinase grâce à sa partie intracellulaire. Après la fixation du ligand, les deux domaines TK du dimère du récepteur se rapprochent et déclenchent une cascade d'autophosphorylation ainsi que l'activation du récepteur, entraînant ainsi le recrutement et l'activation de substrats et le déclenchement de voies de signalisations [9, 10]. Parmi les fonctions régulées par le récepteur à l'insuline, on retrouve notamment le stockage du glucose sous forme de glycogène (polymère de glucose lié en α 1-4). Ce récepteur a également un rôle dans le cycle de vie cellulaire et est impliqué dans la mitose, l'apoptose ou la différenciation cellulaire [11].

Le récepteur à l'insuline est une protéine fortement glycosylée qui, en plus de nombreuses

O-glycosylations, comporte 18 sites potentiels de N-glycosylation (14 sur la sous-unité α et 4 sur la sous unité β , figure 1.1 B) [12]. Ces glycosylations sont majoritairement situées sur la région extracellulaire du récepteur. L'impact des glycosylations sur le récepteur à l'insuline est très variable et concerne aussi bien le repliement de la protéine que sa dimérisation ou son affinité pour le ligand [13].

Le récepteur à l'insuline et son ligand sont très étudiés dans le cadre de pathologies, notamment les deux types de diabète. De par sa nature de protéine transmembranaire, ce récepteur est exposé à la matrice extracellulaire et interagit avec ses éléments. Au sein de notre laboratoire, il a été notamment montré que l'activation du complexe récepteur à l'élastine (CRE) par les produits de dégradation de l'élastine affectait les fonctions du récepteur à l'insuline [14].

1.1.2 Désialylation du récepteur à l'insuline par le complexe récepteur à l'élastine

Le complexe récepteur à l'élastine est formé d'une sous-unité *Elastin Binding Protein* (EBP), une protéine protectrice (Cathepsin A), et d'une Neuraminidase 1 (Neu-1) [15] (figure 1.2). La sous-unité Neu-1 de ce complexe possède une activité sialidase et hydrolyse la liaison entre les acides sialiques terminaux et le reste de la chaîne de glycosylation [16]. Dans le cadre de l'étude de l'influence de la dégradation de l'élastine, il a ainsi été montré que, après activation du CRE, la sous-unité Neu-1 de ce complexe interagit avec le récepteur à l'insuline [14].

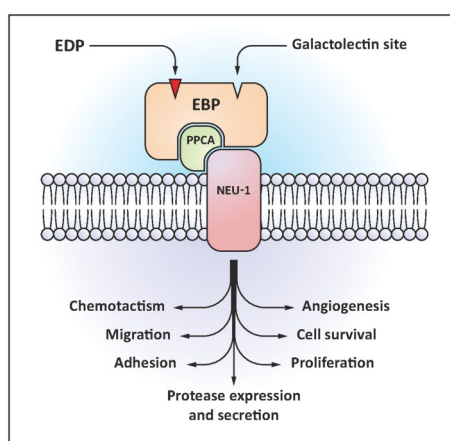


FIGURE 1.2 – **Représentation schématique du complexe récepteur à l'élastine.** Le complexe est composé de trois sous-unités : une *Elastin Binding Protein* (EBP), une protéine protectrice / cathepsin A (PPCA) et une Neuraminidase-1 (Neu-1). Cette protéine possède notamment une activité sialidase et est impliquée dans de nombreuses voies de signalisation ou de régulation. Image tirée de Scandolera et al. [15]

Une expérience de docking moléculaire a confirmé la possibilité structurale de cette interac-

tion. De plus, la mesure des niveaux d'acides sialiques sur le récepteur a permis de montrer que l'action de Neu-1 réduit le taux de sialylation des glycanes portés par le récepteur à l'insuline. A l'inverse, l'inhibition de Neu-1 restaure l'état de sialylation du récepteur. L'étude détaillée de l'impact de l'hydrolyse des acides sialiques menée pendant la thèse d'Alexandre Guillot a montré que la désialylation n'affecte pas l'internalisation du récepteur à l'insuline par la cellule, ni le *shedding* (clivage de la partie extracellulaire) du récepteur, son niveau d'expression et son affinité à l'insuline. Il est donc probable que l'hydrolyse des acides sialiques ait un effet plus fin au niveau de la structure des récepteurs [8, 14].

Ainsi, il a été décidé d'étudier l'influence des acides sialiques sur la dynamique et la flexibilité des glycanes afin de comprendre comment la désialylation impacte la structure et la dynamique des glycanes.

1.2 Acides sialiques et flexibilité des glycanes

L'objectif des travaux suivants a donc été d'élucider l'impact de la sialylation sur la dynamique et la flexibilité des glycanes. Cette étude a notamment permis le développement d'une méthode d'analyse dédiée aux chaînes de glycosylation [17].

1.2.1 Caractérisation des conformations principales : méthodes de clustering

Dans un premier temps, une étude de dynamique moléculaire sur des glycanes bi- et tri-antennés en chaînes isolées a débuté. Huit glycanes ont été étudiés au total. L'utilisation de méthodes de clustering a permis de classer les conformations principales obtenues en fonction des grandes familles conformationnelles définies par Mazurier et al [18] et présentées sur la figure 1.3. Sur des glycanes bi-antennés : la conformation *Bird* correspond à un glycanes aux deux branches étendues vers le solvant. Les conformations *Broken Wing* et *Backfolded* correspondent à des conformations où l'une des branches du glycanes est repliée le long du cœur du glycanes [18].

Ainsi, ces travaux réalisés sur 1 500 ns de trajectoire pour chaque glycanes ont montré que les conformations principales adoptées par le glycanes au cours de la trajectoire varient en fonction de l'état de sialylation : les glycanes bi-antennés passent d'une préférence pour la conformation *Broken Wing* en présence d'acides sialiques à une préférence pour la conformation *Bird* (extension des branches vers le solvant). Les glycanes tri-antennés, à l'inverse, montrent que la conformation principale similaire à la conformation *Broken Wing* est stabilisée en l'absence d'acides sialiques : la proportion de cette conformation passe en effet de 68% à 87% de la trajectoire. Le RMSF (*Root Mean Square Fluctuations*, fluctuations autour de la structure moyenne) diminue ainsi au

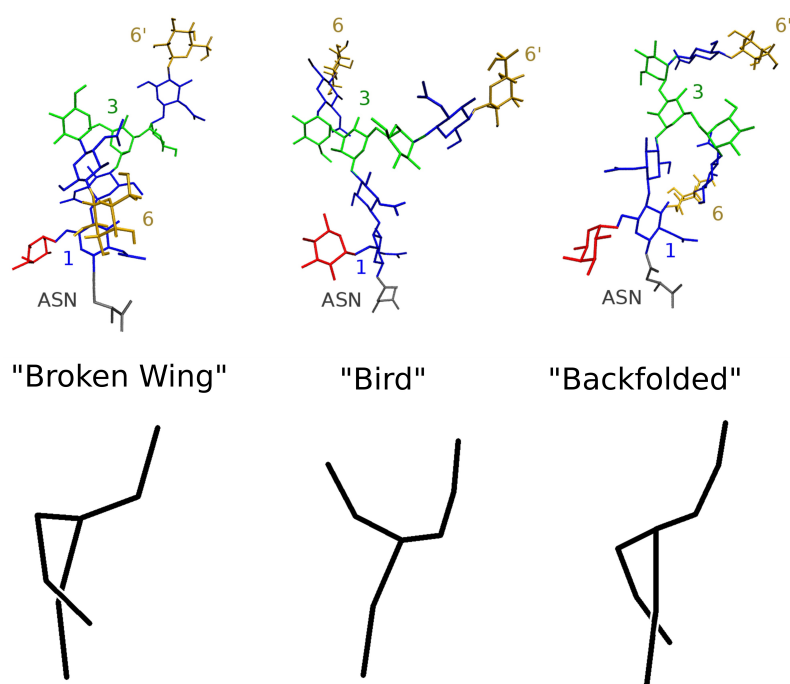


FIGURE 1.3 – **Familles conformationnelles des glycanes bi-antennés.** Les conformations *Broken Wing* et *Backfolded* ont une branche repliée le long du cœur du glycan tandis que la conformation *Bird* est étendue vers le solvant. Les N-acétyl-glucosamines sont en bleu, les mannoses en vert, les galactoses en jaune, le fucose en rouge et l'asparagine glycosylée en gris. Les structures en haut de la figure sont extraites de simulations de dynamique moléculaire et correspondent aux représentations schématiques en bas de la figure.

cours de la trajectoire.

Ces travaux ont également mis en avant le rôle de certains points de flexibilité bien particuliers, notamment la liaison glycosidique entre les mannoses du cœur du glycane. Les deux mannoses à la base des branches sont en effet reliés au mannose central par une liaison différente : la liaison $\alpha 1-3$ est formée de trois atomes et peut être caractérisée par deux angles dièdres Phi et Psi. La liaison $\alpha 1-6$ comporte un atome de plus et, par conséquent, un angle dièdre supplémentaire Omega est nécessaire pour décrire celle-ci (figure 1.4).

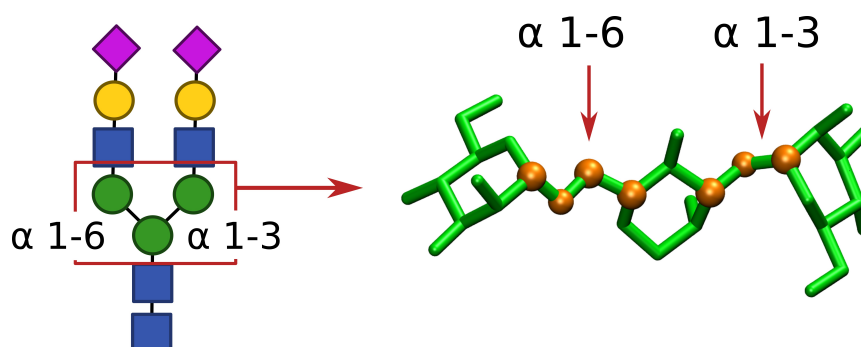


FIGURE 1.4 – **Structure des liaisons $\alpha 1-6$ et $\alpha 1-3$.** A gauche, la représentation d'un glycane selon la nomenclature SNFG présentée dans la section 2.2.2. L'emplacement des deux liaisons d'intérêt est encadré en rouge. A droite, la structure des trois mannoses impliqués est détaillée. Les atomes des liaisons $\alpha 1-3$ et $\alpha 1-6$ sont représentés par les sphères oranges. La liaison $\alpha 1-6$ comporte un atome de plus que la liaison $\alpha 1-3$. Les différents angles dièdres sont identifiés sur la structure.

Ces résultats sont également confirmés par les résultats obtenus grâce à une méthode développée afin d'évaluer la flexibilité et la zone explorée par les glycanes au cours de la trajectoire : l'*Umbrella Visualization*.

1.2.2 Développement de l'*Umbrella Visualization* et problèmes soulevés pendant l'étude

Cette méthode considère que le glycane agit comme un parapluie qui protège la surface de la protéine de sorte qu'une hypothétique molécule approchant cette région ne pourrait pas interagir directement avec la protéine. La section 3.2.1 décrit cette méthode et son utilisation.

L'application de cette méthode a ainsi permis de montrer la variation de flexibilité des glycanes en fonction de leur état de sialylation : les glycanes bi-antennés montrent en effet une flexibilité plus importante et explorent une plus grande région comparés aux glycanes bi-antennés

sialylés. La désialylation a cependant un impact différent sur les glycanes tri-antennés puisqu'une seule des trois branches montre un changement de flexibilité significatif [17].

La mise en place de cette méthode a ainsi permis de compléter les informations de clustering avec des informations de dynamique à l'aide de l'*Umbrella Visualization*.

1.2.3 Vers une nouvelle approche de l'étude des glycosylations dans le cadre du Projet DOGME

Le principal problème rencontré pendant les précédents travaux a été le manque d'outils adaptés à l'étude des glycosylations. En effet, bien que de nombreux logiciels et serveurs permettent la construction de systèmes de protéines glycosylées (voir section 2.3), les outils spécifiques dédiés à l'étude de l'impact des glycosylations sur les protéines sont peu nombreux. La première version de l'*Umbrella Visualization* a été développée pour pallier ce manque. Bien que cette méthode ait permis d'obtenir des résultats intéressants et décrivant la zone explorée par des glycanes en chaînes isolées, cette méthode ne permet pas de prendre les interactions avec la protéine ou la topologie de cette dernière.

Dans la continuité des travaux relatifs à l'*Umbrella Visualization* et engagés au sein de l'unité MEDyC nous avons choisi d'étendre cette dernière à une représentation tridimensionnelle qui permet de tenir compte des interactions des glycanes et la surface de la protéine. Ainsi, le projet DOGME (**D**éveloppements d'un **O**util de visualisation moléculaire et application à l'étude in silico de **G**lycosylations de protéines de la **M**atrice **E**xtracellulaire) dans lequel s'inscrivent les travaux de thèse présentés dans ce manuscrit a pour objectifs d'une part de développer un outil d'analyse visuel dédié aux sucres et d'autre part de tester ce dernier sur diverses trajectoires de dynamique moléculaire (glycanes isolés ou portés par une protéine).

Afin de présenter l'ensemble des travaux réalisés pendant les trois années de thèse ainsi que leur contexte, le manuscrit est organisé de la façon suivante : dans un premier temps le chapitre "Contexte scientifique" nous permet de présenter de manière plus approfondie l'objet d'étude de l'unité MEDyC, à savoir la matrice extracellulaire et de faire le lien avec les glycosylations, leur représentation schématique, et les différentes méthodes qui permettent d'appréhender leur étude. En clôture de ce chapitre, nous spécifions les enjeux de la visualisation scientifique associés à la représentation et l'étude des objets moléculaires. A travers le chapitre "Matériel et méthodes", nous présentons les outils spécifiques de dynamique moléculaire utilisés ainsi que le logiciel de visualisation avec lequel nous avons procédé à nos développements. L'ensemble des résultats et analyses des expériences de dynamique moléculaire réalisées sur des glycanes isolés

ou encore portés par une protéine (la fibromoduline) sont ensuite détaillés dans le chapitre 4. Nous y présentons notamment une étude comparative de deux types de liaisons des acides sialiques terminaux. Le chapitre intitulé "Développements d'un nouvel outil de visualisation intégré à UnityMol" revient sur les implémentations originales intégrées au logiciel UnityMol. Nous présentons ensuite les tests réalisés à l'aide de cet outil et validons son utilisation en application sur huit trajectoires de dynamique moléculaire décrivant le comportement du récepteur à l'insuline glycosylé. Enfin, dans le dernier chapitre, nous discutons et comparons l'ensemble des résultats obtenus au cours de ces travaux pour ensuite proposer de nouvelles pistes scientifiques à explorer.

Chapitre 2

Contexte scientifique

2.1 Biologie de la matrice extracellulaire

La matrice extracellulaire (MEC) est un réseau tridimensionnel de macromolécules impliqué dans de nombreux mécanismes et processus cellulaires. La MEC est présente dans tous les tissus et organes et leur procure un soutien physique grâce à ses propriétés mécaniques et un soutien biochimique grâce aux nombreuses molécules actives qu'elle contient. Ces molécules peuvent être des facteurs de croissance, des produits de dégradation ou des hormones et permettent ainsi les échanges entre la MEC et les cellules environnantes, activant différentes voies de signalisation et de régulation impliquées dans l'adhésion et la différenciation cellulaire, la migration, la prolifération [19].

La MEC est ainsi un milieu très dynamique grâce aux enzymes et protéases qu'elle contient. Synthétisées par les fibroblastes, ces enzymes dégradent et remodelent les composants de la MEC. On y trouve notamment les métalloprotéases matricielles (MMP), des protéases à Serine ou à Cystéine. Ces produits de dégradation, appelées matrikines, sont également impliqués dans diverses voies de signalisation [20, 21]. Cette section vise à présenter la matrice extracellulaire et ses caractéristiques principales : son organisation en membranes basales et matrices interstitielles et les grandes familles de molécules qui la composent.

2.1.1 Organisation générale de la MEC

La MEC est composée de quatre grandes familles de macromolécules : les protéoglycanes, les glycoprotéines de structure, les fibres de collagène et l'élastine. Les fibres de collagène et l'élastine forment un réseau dense dont les propriétés mécaniques permettent aux tissus de résister aux forces de tension. Les glycoprotéines et les protéoglycanes forment également un réseau, moins

dense, pour résister aux forces de compression [22]. La composition de la MEC, ainsi que ses propriétés mécaniques associées, varie en fonction des organes et des tissus. Ainsi, les tissus les plus élastiques, comme la peau, sont plus riches en élastine alors que les tissus plus riches en collagène, comme les os, sont plus rigides. Cette organisation en réseaux permet de différencier deux sous-types de matrice extracellulaire : les membranes basales et les matrices interstitielles.

Les membranes basales sont organisées en couches de 50 à 100 nm d'épaisseur soutenant les monocouches cellulaires des épithéliums ou des endothéliums. Elles sont hétérogènes et multifonctionnelles, ainsi leur composition diffère en fonction des organes. On y retrouve majoritairement du collagène de type IV (jusqu'à 50% des protéines de la membrane basale) [20, 23]. Les matrices interstitielles enveloppent les cellules des tissus et leur fournissent un soutien structural et biochimique. La figure 2.1 résume l'organisation globale de la matrice extracellulaire dans les organes.

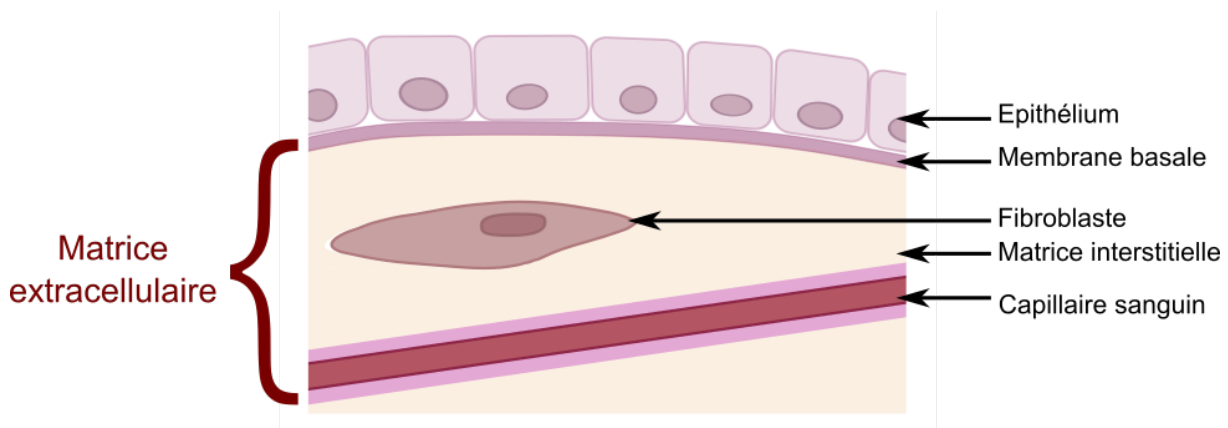


FIGURE 2.1 – **Organisation globale de la matrice extracellulaire.** La membrane basale soutient les cellules des épithéliums et leur sert de point d'ancrage. La matrice interstitielle plus épaisse, enveloppe les cellules et leur apporte un soutien structural mécanique et biochimique. Ces deux composants forment la matrice extracellulaire. Les fibroblastes contenus dans la matrice interstitielle synthétisent les molécules de la MEC.

2.1.2 Les protéoglycanes

À l'exception du hyaluronane qui existe sous forme libre, tous les protéoglycanes adoptent la forme d'un corps protéique auquel est greffé une ou plusieurs chaînes de glycosaminoglycane (GAG). Ces chaînes possèdent de nombreux groupements sulfates ou carboxylates, leur conférant une importante charge négative qui attire les molécules d'eau environnantes. Cette hydratation joue un rôle important dans les tissus subissant des variations de pression [24].

On retrouve trois grands groupes de protéoglycanes au sein de la MEC : les protéoglycanes des membranes basales, les hyalectanes et les SLRP (*Small Leucine Rich Protein*). Ces protéines sont notamment impliquées dans l'organisation des tissus, la croissance tissulaire et la maturation des tissus spécialisés. Elles sont également impliquées dans la fibrillogenèse et modulent l'action des facteurs de croissance [25]. Dans le cadre de cette thèse, nous nous sommes en particulier intéressés à la fibromoduline, un protéoglycane de la famille des SLRP. Cette famille comporte 18 membres répartis en 5 classes dont les membres comportent une forte homologie de structure. La fibromoduline fait partie de la classe II avec le lumican, deux protéines très étudiées au sein de l'unité MEDyC. La classe II des SLRPs se définit principalement par la présence de chaînes de keratan sulfate ou polylactosamine. Cette classe possède 3 sous-groupes définis en fonction de l'homologie de séquence des protéines les composant. La fibromoduline porte donc des chaînes de keratan sulfate et se lie aux collagènes pour réguler leur fibrillogenèse [25, 26]. Le lumican porte également des chaînes de keratan sulfate et est essentiellement présent dans les tissus mésenchymateux et dans les tumeurs (cancers du sein, mélanomes) [27, 28].

2.1.3 Les glycoprotéines

Les glycoprotéines sont des protéines auxquelles sont greffées des saccharides de plus petite taille que les glycosaminoglycanes. La synthèse et la composition de ces oligosaccharides sont détaillées dans la section 2.2. Ces macromolécules constituent une part importante de la MEC et regroupent notamment la fibronectine, les fibrillines, les laminines et trombospondines [29].

2.1.4 Les collagènes

Les collagènes sont les protéines fibreuses les plus abondantes dans la matrice extracellulaire, constituant jusqu'à 30% des protéines totales chez l'homme [30, 31]. La superfamille des collagènes comporte 28 types de collagènes différents composés par au moins 46 chaînes α différentes. Cette superfamille très complexe montre une grande diversité en terme d'organisation supramoléculaire, de fonction et de distribution au sein des tissus [32]. La caractéristique commune des collagènes est leur organisation en triple-hélice grâce au triplet G-X-Y, où X et Y sont respectivement une proline et une hydroxyproline [33]. La structure est maintenue par des liaisons hydrogènes interchaînes. Les extrémités N-terminales et C-terminales peuvent présenter des domaines globulaires non collagéniques appelés « domaine NC » [34].

2.1.5 L'élastine

Les fibres élastiques sont d'importantes structures de la matrice extracellulaire qui remplissent un rôle principalement mécanique : elles confèrent l'élasticité aux tissus soumis à des elongations ou des contractions comme les vaisseaux sanguins, les poumons ou le cœur. Ces fibres très stables

sont principalement composées d'un cœur d'élastine (polymères de tropoélastine) et de fibrillines associées en microfibrilles qui entourent ce cœur [35, 36]. Les microfibrilles sont composées de petites glycoprotéines comme les fibrillines, les MAGP (*Microfibril-associated glycoproteins*) ou certains glycosaminoglycanes comme la décorine ou le biglycane. La dégradation de l'élastine conduit à la production de nombreux fragments protéiques actifs appelés élastokines et impliqués dans certaines pathologies (par exemple l'athérosclérose ou des cancers) [37, 38].

Ainsi, ces grandes classes de molécules s'organisent et interagissent entre elles et avec leur environnement, formant ainsi un milieu complexe et dynamique, structuré par les réseaux formés par les collagènes, laminines et fibres élastiques. Les glycanes, présents sous forme libre (héparane sulfate) ou liés à la surface des protéines jouent un rôle important dans les fonctions des composants de la MEC, comme la migration ou l'adhésion cellulaire [39–42]. Présents sous la forme de longues chaînes de glycosaminoglycanes (comme dans la famille des SLRP, par exemple) ou de chaînes plus courtes et ramifiées, les oligo- et poly-saccharides sont ainsi des éléments clés de la matrice extracellulaire.

2.2 Glycosylations

La glycosylation est une modification post-traductionnelle consistant à ajouter à une protéine une ou plusieurs chaînes de saccharides. Ces chaînes sont très flexibles et réactives en raison de leur structure et leur composition.

2.2.1 Liaison glycosidique et diversité structurale des glycanes

Les glycanes sont des polymères de monosaccharides reliés par une liaison glycosidique et présents sous forme libre ou liés à des protéines. Sous forme libre, les monosaccharides existent sous forme linéaire ou cyclique. Lorsqu'ils sont liés à une protéine, les monosaccharides sont présents sous forme cyclique : cette forme crée un centre anomérique, utilisé pour former une liaison glycosidique et qui peut être situé au niveau du carbone C1 ou C2 en fonction de la nature du saccharide (aldose ou cétose) [43].

La liaison glycosidique permet ainsi de relier un monosaccharide à un autre composant hydroxylé comme un autre saccharide (permettant ainsi la formation de polymères complexes), un acide aminé (sérine, thréonine, asparagine) ou un lipide. Un seul monosaccharide peut porter plusieurs liaisons glycosidiques, créant un point de branchement dans la chaîne et augmentant la complexité structurale du glycanes. De plus, les deux stéréoisomères possibles de la liaison glycosidique, appelés α et β , contribuent à augmenter la diversité structurale des glycanes [43].

2.2.2 Représentation schématique et nomenclature SNFG

La complexité structurale des glycanes, leur longueur et la diversité de leur composition a conduit à l'élaboration d'une nomenclature, appelée nomenclature SNFG (pour *Symbol Nomenclature For Glycans*) [44], basée sur la représentation des monosaccharides à l'aide de formes géométriques colorées. Cette nomenclature permet de s'affranchir des représentations chimiquement exactes (représentation de Fischer ou Haworth, par exemple), qui tendent à devenir confuses pour des glycanes complexes. La première version standardisée de cette nomenclature a été introduite avec la première édition du livre *Essentials of Glycobiology*, en 1999 [45]. Actuellement, la nomenclature SNFG est largement utilisée par la communauté scientifique et les récents développements étendent l'utilisation de symboles blancs (dédiés aux saccharides dont la nature n'est encore pas identifiée) en permettant l'ajout de lettres dans le symbole afin d'apporter d'éventuelles précisions ou détails en annotation [46]. Dans ce manuscrit, nous avons adopté cette nomenclature pour les schémas simplifiés qui sont présentés.

2.2.3 Protéoglycanes et glycoprotéines

Comme évoqué en début de chapitre, les protéines glycosylées peuvent être classées en deux grandes catégories en fonction des chaînes qu'elles portent. Les protéoglycanes portent de longues chaînes (plusieurs dizaines de saccharides) très peu ramifiées et généralement composées de répétitions de disaccharides. Ces longues chaînes, appelées GAG, sont principalement présentes sur les protéines de la matrice extracellulaire (Section 2.1). Les GAG peuvent représenter jusqu'à 90% du poids moléculaire du protéoglycane. Les protéoglycanes peuvent également porter des N- ou des O-glycosylations, en plus des GAG [47, 48]. Les glycoprotéines représentent l'ensemble des protéines glycosylées qui ne comportent pas de chaînes de GAG. Les N- et O-glycosylations peuvent également représenter une importante part du poids moléculaire de ces protéines, mais les chaînes de glycanes sont plus courtes et ramifiées. Alors que les GAG sont essentiellement présents sur les protéines de la MEC, les N- et O-glycosylations sont présentes sur les protéines de tous les milieux (cellulaires et extracellulaires) [49].

2.2.4 N-glycosylations

Nos travaux se concentrent essentiellement sur l'impact des N-glycosylations chez l'homme et/ou les vertébrés. Cependant, cette modification post-traductionnelle a été retrouvée dans tous les domaines du vivant. Chez les vertébrés, les N-glycanes présentent un cœur ou noyau commun composé de 2 N-acetylglucosamines et 3 mannoses. A partir de cette caractéristique commune, une grande diversité de structures de N-glycanes existe.

La N-glycosylation consiste à créer une liaison glycosidique entre un glycane et une asparagine d'une protéine. Les sites de glycosylation partagent une séquence consensus commune : Asn – X – Thr/Ser, où Asn est l'asparagine portant la glycosylation, X peut être n'importe quel acide aminé sauf une proline [50]. Même si cette séquence est présente aux sites de N-glycosylations, toutes les séquences correspondantes d'une protéine ne sont pas forcément glycosylées. En effet, des contraintes telles que des gênes stériques peuvent empêcher l'ajout d'un glycane sur un site. Des outils permettant d'évaluer l'accessibilité d'un site de glycosylation potentiel existent, comme détaillé dans la section 2.3.

Synthèse des N-glycanes

La synthèse des N-glycanes se fait en deux étapes, dont la première est hautement conservée. Un oligosaccharide est assemblé sur un transporteur lipidique (dolichol phosphate) dans la membrane interne du réticulum endoplasmique [50]. Ce glycane, composé de 2 N-acétyl glucosamines, 9 mannoses et 3 glucoses, est ensuite transféré sur une protéine pendant sa translocation dans le réticulum endoplasmique. Cette première phase est ensuite suivie d'une phase de transformation ou maturation impliquant différentes glycosidases et glycosyltransférases dans la lumière du réticulum endoplasmique et également dans l'appareil de Golgi. Dans un premier temps, les trois glucoses à l'extrémité du bras α 1-3 sont hydrolysés. Les deux mannoses terminaux sur les deux branches α 1-6 peuvent également être hydrolysés. Les protéines sortent du réticulum endoplasmique avec 8 ou 9 mannoses et rejoignent l'appareil de Golgi où ces branches vont pouvoir être modifiées [51].

Le processus ainsi engagé va permettre d'ajouter des sucres au cœur du glycane, notamment l'ajout de α fucose sur la première N-acétyl glucosamine. Les branches vont également pouvoir être étendues par l'addition de disaccharides (par exemple, un dissaccharide composé d'un galactose et une N-acétyl glucosamine). Des « décorations » terminales comme des acides sialiques peuvent également être ajoutées à l'extrémité des branches. Ce processus de maturation est connu pour ses résultats hétérogènes, mais trois types de glycanes peuvent être distingués :

1. Les N-glycanes complexes présentent plusieurs branches ou antennes débutant par une N-acétylglucosamine.
2. Les N-glycanes de type oligomannose ne sont composés que de mannoses.
3. Les types hybrides enfin, portent une antenne « complète » sur le bras α 1-3 et des mannoses sur le bras α 1-6 [49].

La figure 2.2 résume le processus décrit ci-dessus et présente ces trois types de glycanes.

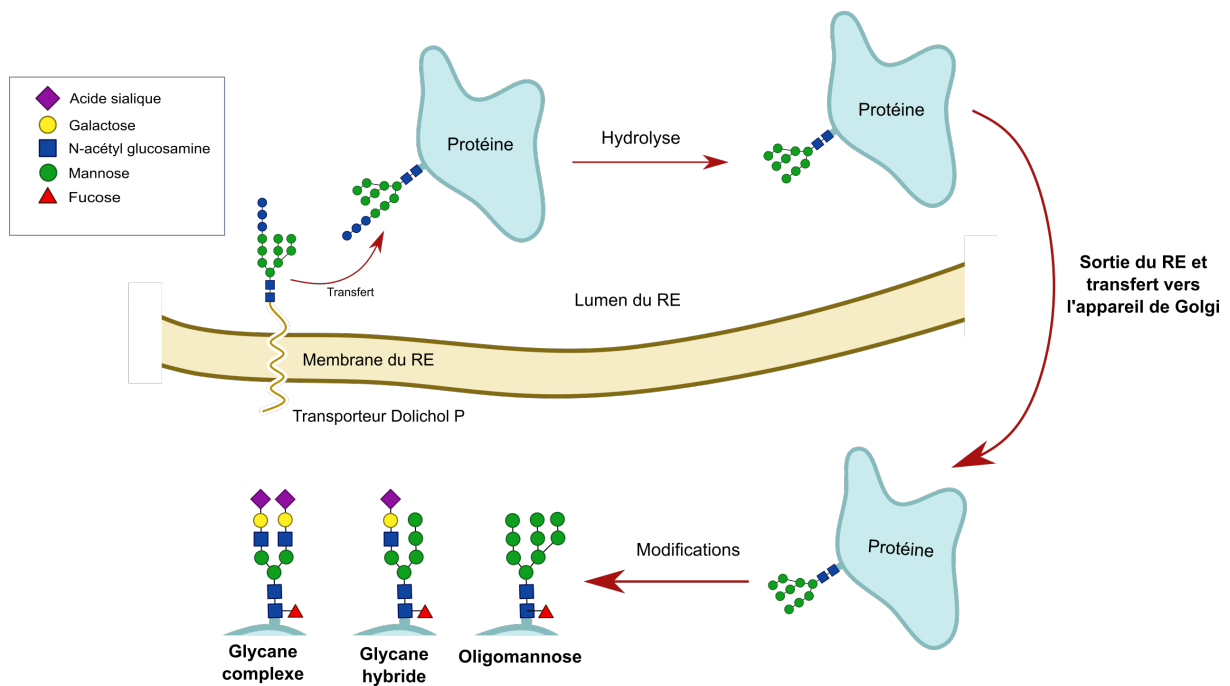


FIGURE 2.2 – **Synthèse des N-glycanes.** Dans un premier temps, un précurseur est synthétisé sur un transporteur lipidique Dolichol P. Ce précurseur est composé de 9 mannoses (en vert) et 3 glucoses (en bleu). Ce précurseur est ensuite transféré sur une protéine. Les glucoses sont ensuite hydrolysés. Quand la protéine quitte le réticulum endoplasmique (Abrégé en RE sur le schéma) et rejoint l'appareil de Golgi, le glycan est modifié pour devenir un glycane complexe, hybride, ou oligomannose.

Rôles biologiques

Les N-glycanes sont impliqués dans de nombreux processus biologiques et il a été montré que le glycome (ensemble des glycanes présents dans une cellule) varie en fonction du contexte cellulaire, et peut évoluer dans un contexte pathologique donné. Par exemple, les travaux de Lee et al. ont montré que la proportion d'acides sialiques liés en α 2-3 aux antennes des glycanes augmentait dans les cellules de lignées cancéreuses (cancers du sein) au détriment des acides sialiques en α 2-6, présents en très grande majorité dans les cellules saines [52]. Cette même étude a également mis en évidence que le nombre de branches des glycanes augmentait dans le contexte pathologique cancéreux, avec l'apparition notamment de glycanes complexes à 3 ou 4 antennes. Sate et al. ont également montré que, dans le contexte de la maladie d'Alzheimer, le taux de glycosylation de protéines impliquées dans la maladie pouvait augmenter ou diminuer de manière significative [53]. Chez les virus, la présence de nombreux glycanes à la surface des protéines de la capsid pourrait également jouer un rôle dans l'échappement au système immunitaire en masquant les épitopes reconnus par les anticorps [54, 55]. De nombreuses études s'intéressent également le rôle des glycosylations sur les immunoglobulines et le système immunitaire.

2.3 Glycosylations et modélisation moléculaire

Cette section a pour but d'introduire les outils déjà disponibles permettant d'étudier les glycosylations. Ces outils permettent d'identifier les sites de glycosylation, analyser les structures des saccharides, de les intégrer à une structure protéique et de les visualiser. Plusieurs de ces outils sont répertoriés sur le portail Glycosciences (www.glycosciences.de). Plus généralement, le portail Glycopédia (<https://www.glycopedia.eu>) répertorie de nombreuses ressources (actualités, librairies, ...) dédiées à l'étude des saccharide.

2.3.1 Identification des sites de glycosylation

Parfois, les premiers résidus saccharides d'un glycanes sont présents dans la structure pdb disponible sur la Protein Data Bank (PDB, www.rcsb.org, [56]). Ceci permet d'identifier certains sites de glycosylation mais la plupart du temps, les structures de glycanes ne sont pas visibles : en raison de leur grande flexibilité, les glycanes ne sont pas assez ordonnés pour être détectés par les méthodes classiques utilisées pour résoudre les structures des protéines, comme la cristallographie aux rayons X [57]. Il est donc nécessaire d'utiliser des outils de prédiction afin d'identifier les sites de glycosylation potentiels.

Le programme GPP (*Glycosylation Prediction Program*) [58] utilise des algorithmes couramment utilisés en bio-informatique (*pairwise pattern* et *random forest*) afin de déterminer si les sites d'intérêt comportent des N- ou O-glycosylations. Ces algorithmes se basent respectivement sur la détection de motifs similaires à l'aide d'une fenêtre glissant le long de la séquence et d'un ensemble d'arbres de décisions afin de détecter les sites potentiels de glycosylation. NetNGlyc [59] utilise des réseaux neuronaux artificiels afin d'étudier la séquence protéique entourant les sites potentiels de glycosylation. NetNGlyc est disponible en tant que programme autonome ou serveur web. NGlycPred [60] utilise également l'algorithme *random forest* et prend en compte les propriétés structurales autour des sites de glycosylation potentiels, comme la structure secondaire de la protéine ou l'accessibilité de la surface. Ce programme détermine ensuite si une chaîne de glycosylation peut y prendre place sans générer de gênes ou de clashes stériques.

2.3.2 Analyse des structures

Plusieurs outils permettent d'analyser et de comprendre la structure des protéines glycosylées. Notamment, pdb-care [61] recherche et trouve les résidus carbohydrates et identifie les erreurs conformationnelles. Les auteurs ont développé cet outil après avoir montré que plus de 30% des structures de carbohydrates dans les structures de la pdb comportaient des erreurs [62].

Une fois que les structures possèdent des conformations correctes, CSS (*Carbohydrate Structure Suite*) [63] met à disposition une série d'outils dédiés à l'analyse des carbohydrates : pdb2LINUCS lit les coordonnées des atomes et donne la représentation LINUCS (*Linear Notation for Unique description of Carbohydrate Sequences*) [64] des résidus saccharides. Cet outil compense donc le manque de consensus pour les noms des saccharides dans les fichiers de structures. Les résultats obtenus peuvent ensuite être utilisés dans l'outil de CSS Glytorsion. Glytorsion interroge la base de données GlytorsionDB afin d'identifier et analyser les angles de torsion des structures de carbohydrates. Il est possible d'obtenir facilement des valeurs telles que les valeurs préférentielles pour un angle donné. CARP (*Carbohydrates Ramachandran Plot*) génère des graphiques de Ramachandran pour chaque liaison dans la structure d'intérêt. Les résultats sont mis en parallèle avec des cartes provenant de la PDB ou de la base de données GlycoMapsDB [65] qui réunit des cartes conformationnelles obtenues par expériences *in silico*. Cet outil permet de détecter les valeurs d'angles incorrectes dans les liaisons glycosidiques.

Les carbohydrates possèdent de nombreux groupements hydroxyles. Ces groupes sont très réactifs et peuvent interagir avec leur environnement. C'est pourquoi il est également important de comprendre comment les glycanes peuvent interagir avec les résidus de la protéine. GlyVicinity [66] analyse statistiquement quels acides aminés sont les plus proches d'un type de

carbohydre choisi. Pour ce faire, il utilise des structures issues de la PDB, permettant ainsi de donner un aperçu des atomes impliqués dans les interactions glycane/protéine.

Bien que ces outils soient très utiles et aident à comprendre l'impact des glycanes sur la structure des protéines, ils sont principalement dédiés à l'analyse de structures statiques. Nous savons cependant que les glycanes et les protéines sont des objets dynamiques. Il est donc essentiel de compléter ces résultats à l'aide de données obtenues par des techniques de simulation de dynamique moléculaire et de les analyser avec des méthodes spécifiques. La première étape vers cet objectif est de construire la structure 3D de glycanes.

2.3.3 Construction des systèmes Protéine/Glycanes

Une fois les sites de glycosylation identifiés, l'étape suivante est de construire puis greffer les chaînes de glycosylation au niveau des sites d'intérêt. Même si certains fichiers pdb contiennent les premiers saccharides du glycane, le reste de la chaîne doit être reconstruite, tout en évitant les structures incorrectes ou le recouvrement avec des atomes de la protéine.

SHAPE [67] est un outil qui permet à un utilisateur de construire une structure de glycane et de prédire plusieurs conformères de basse énergie en utilisant un algorithme génétique. SWEET II, disponible sur le portail glycosciences.de, est un autre outil qui peut être utilisé pour construire une structure de glycane [68,69]. Les fichiers ainsi générés sont exploitables sur la page de Glyprot. Glyprot permet à un utilisateur de charger un fichier pdb de protéine et d'ajouter le glycane créé à un ou plusieurs sites de glycosylation [70]. Le fichier obtenu est une structure de protéine glycosylée, minimisée avec le champ de forces MM3. Récemment, l'ensemble d'outils DoGlycan [71] a été développé afin d'accomplir la même tâche mais avec le champ de forces Glycam [72], un champ de forces dédié à l'étude des structures de carbohydrates. Glycam-Web (<http://glycam.org/>) est un portail web qui propose des outils développés autour de ce même champ de forces et qui permet la construction et l'ajout d'un glycane sur une protéine.

Charmm-gui (www.charmm-gui.org/) est un autre portail web permettant les mêmes fonctions grâce à l'outil Glycan-Reader [73], qui lit les structures de carbohydrates depuis le fichier donné en entrée. Glycan-Reader ne prédit donc pas les sites de glycosylation et nécessite la présence d'au moins un carbohydrate pour construire le glycane sur un ou plusieurs sites choisis. Cependant, ce portail permet de solvater une molécule et de minimiser la structure pour préparer des travaux de simulation. Charmm-gui utilise le champ de forces Charmm36 [74].

2.3.4 Travaux connexes

La plupart des logiciels de visualisation dédiés aux protéines et édifices moléculaires biologiques permettent de représenter des saccharides en mode tout-atomes, comme les licorices ou les représentations en sphères de van der Waals (vdW)... Cependant, peu d'entre eux proposent des modes de visualisation qui mettent en avant les spécificités structurales des sucres et / ou des glycosylations.

Le plugin Azahar [75], disponible sous PyMol peut être utilisé pour créer un glycane à partir d'une liste de modèles de structures de saccharides. Ce plug-in intègre également trois modes de représentation spécifiques qui visent à simplifier la représentation des structures de glycanes : la représentation *cartoon* représente les cycles comme des polygones non plats et colorés, reliés par des tubes. La représentation *wire* est très similaire mais les tubes sont plus fins et les polygones des cycles sont vides. La représentation en perles (*beads*) représente chaque cycle comme des sphères. Azahar offre également la possibilité de calculer plusieurs paramètres comme le rayon de gyration, les diagrammes de Ramachandran ou la position de liaisons hydrogènes. Ces valeurs peuvent être évaluées pour des structures statiques ou des simulations de dynamiques moléculaires. Enfin, en utilisant une routine Monte-Carlo associée à une minimisation d'énergie, Azahar peut effectuer une recherche des conformations de plus basse énergie potentielle sur un glycane.

Avec le plugin 3D-SNFG [76], le logiciel VMD implémente sa propre représentation graphique basée sur la nomenclature développée dans la seconde édition du livre *Essentials of Glycobiology* [77]. Chaque monosaccharide est représenté par un objet 3D basé sur le symbole correspondant à la nature du sucre. Cette représentation simplifie beaucoup la visualisation des glycanes, donnant une vue claire de leur organisation tout en conservant les informations de structure. Cependant, le coût de calcul requis pour calculer les formes 3D pour chaque image extraite d'une trajectoire est élevé. Cette représentation est donc peu adaptée à des études sur des trajectoires de dynamique moléculaire.

UnityMol est une plateforme *open source* codée en C# et développée avec le moteur Unity (<http://unity3d.com/>) [78]. Cette plateforme est dédiée à la visualisation de molécules biologiques et utilise la visualisation *Hyperballs* [79]. Dans une version récente appelée SweetUnityMol [80], ce logiciel intègre des modes de représentation dédiés aux saccharides, notamment une représentation en rubans et la possibilité de colorer les saccharides en fonction leur nature. Pour chaque saccharide, la couleur associée est déterminée par le symbole correspondant dans la nomenclature SNFG.

Les outils répertoriés ci-dessus présentent différentes manières de construire, analyser et visualiser les structures de protéines glycosylées. Il est possible de mettre en place un processus aboutissant à la construction d'un système de protéine glycosylée prêt pour des travaux de simulation de dynamique moléculaire, comme celui présenté dans Mazola et al. [81]. Cependant, les champs de forces et les nomenclatures peuvent différer entre tous ces outils et portails, rendant difficile parfois la transition de l'un à l'autre. De plus, ces outils sont principalement dédiés à la construction de protéines glycosylées et à l'analyse de structures statiques, cristallographiques. Bien qu'il soit possible d'obtenir des données de dynamique moléculaire sur des glycanes et des protéines glycosylées, les outils spécifiques dédiés à l'analyse des trajectoires font défaut. Développer de nouveaux outils de visualisation scientifique dédiés aux glycanes est ainsi nécessaire.

2.4 Visualisation scientifique

Dans cette section, nous allons définir ce qu'est la visualisation scientifique et quels sont ses enjeux principaux au sein de la communauté scientifique. Nous introduirons ensuite les logiciels de visualisation utilisés en modélisation moléculaire et les principaux modes de représentation disponibles sur ces plateformes.

2.4.1 Visualisation scientifique : définition et enjeux principaux

La visualisation permet de présenter, sous forme de production visuelle (image, graphique, animation, ...), un ensemble d'informations afin de les rendre compréhensibles et claires auprès d'un public ciblé. Les expériences scientifiques génèrent une grande quantité de données qu'il faut traiter et rendre lisibles à l'aide de méthodes de visualisation.

Ainsi, nous pouvons définir la visualisation scientifique au sens large comme un domaine dédié à la mise en avant d'informations pertinentes qui vont permettre d'apporter un nouvel éclairage sur un sujet donné. Les représentations classiques et très répandues prennent la forme de graphiques, courbes ou images fixes. Ces représentations sont restreintes à un espace en deux dimensions.

Historiquement, l'une des premières représentation scientifique en 3D et préfigurant ce qui est fait de façon plus moderne de nos jours est la sculpture de plâtre créée par James Clerck Maxwell, en 1874 (figure 2.3 A). Cette sculpture fait suite aux travaux de thermodynamique de Josiah Willard Gibbs en 1871 sur la relation entre le volume (axe x), l'entropie (axe y) et l'énergie (axe z) d'une substance [82]. La publication de Gibbs ne comportait pas de schéma ou de représentation de la surface 3D considérée. Maxwell proposa alors son interprétation de cette surface en se

basant sur les propriétés d'une substance imaginaire aux propriétés physico-chimiques similaires à celle de l'eau. Maxwell traça sur cette surface des lignes de même température (isothermes), même pression (isobares) et même énergie.

Plus récemment, les développements de l'informatique et des logiciels de visualisation et de modélisation ont fait évoluer les enjeux de la visualisation scientifique. En effet, il est maintenant crucial d'apporter, en plus des méthodes de représentation, des méthodes aidant à l'interprétation et l'analyse des données numériques. Il s'agit donc de leur donner une forme appréciable par l'œil humain afin de permettre le raisonnement scientifique et la compréhension du processus étudié.

Ainsi, il est possible de définir deux enjeux principaux pour la visualisation scientifique :

1. Tout d'abord, les aspects liés à la communication scientifique impliquent la production d'éléments graphiques qui visent à mettre en avant une ou plusieurs informations pertinentes sous une forme accessible et compréhensible pour un public ciblé.
2. Ensuite, l'analyse et le traitement de données, souvent issues de simulations et donc sous forme numérique, cherchent à rendre observable quelque chose qui ne l'est pas. Ainsi, il est possible d'utiliser simultanément plusieurs modes de représentation (par exemple une surface et une coloration spécifique dépendante des données) pour permettre de clarifier et d'explicitier les résultats.

C'est notamment le cas des logiciels de visualisation moléculaire qui permettent de représenter les molécules biologiques étudiées sous différentes formes, soit comme finalité soit comme étape intermédiaire afin d'analyser les résultats.

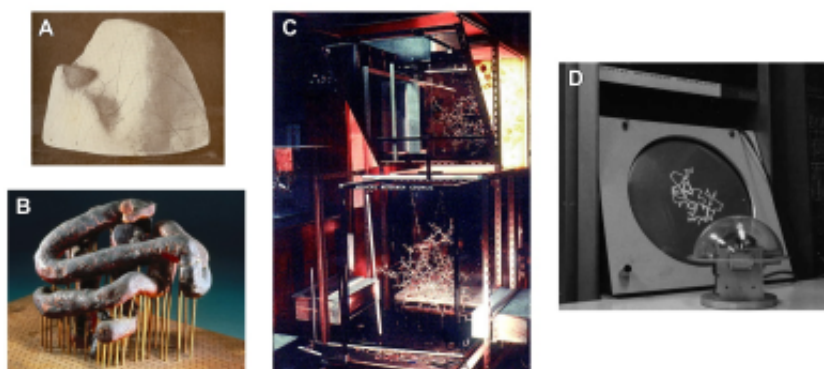


FIGURE 2.3 – **A** : Photographie historique de la surface sculptée par Maxwell, représentant le lien entre volume, entropie et énergie d'une substance. **B** : Sculpture en plastiline de la myoglobine, réalisée par John Kendrew en 1957. **C** : Photographie originale de la "Richard's Box" de F. Richards représentant la RNase S. **D** : Modèle en fil de fer développé par C. Levinthal et son équipe.

2.4.2 Logiciels de visualisation moléculaire : historique et évolutions

La compréhension de la structure des macromolécules biologiques est un élément essentiel dans le décryptage de leurs fonctions. Tout comme la surface sculptée par Maxwell, les premiers modèles de structures prenaient une forme physique. La première structure modélisée ainsi a été élaborée en 1957. Ce modèle de la myoglobine, réalisée en plasticine par John Kendrew (figure 2.3 B), fut créé un an avant la résolution de la structure par cristallographie aux rayons X. Avec l'augmentation de la précision des méthodes de détermination de structures, les sculptures en plasticines n'étaient plus suffisantes. C'est pourquoi d'autres modèles, comme ceux créés par Frederic Richards, furent utilisés, notamment dans le domaine de la cristallographie. Ces derniers, appelés *Richards Box* et (composés de plaques transparentes montées verticalement) permettaient de construire des modèles physiques des structures de protéines en visualisant la densité électronique à travers un système de miroirs semi-argentés. (figure 2.3 C).

L'arrivée des premiers modèles virtuels permit de s'affranchir du support physique (sculpture ou procédé optique), souvent impossible à transporter. Le premier recensé date de 1964 et a été développé au MIT par Cyrus Levinthal et son équipe. Un écran d'oscilloscope monochrome permettait d'afficher les structures protéiques avec un rendu en fil de fer (figure 2.3 D).

Pendant les décennies suivantes, des programmes comme CHAOS ou MANOSK [83] furent développés sur des ordinateurs devenant progressivement de plus en plus performant, permettant de s'éloigner des modèles physiques. Cependant, les ordinateurs capables d'exécuter ces programmes étaient très chers et donc peu accessibles.

L'arrivée de RasMol [84] va cependant permettre de palier ce problème. Développé par Roger Sayle pendant ses travaux de thèse à l'université d'Edimbourg et sous la direction d'Andrew Coulson, RasMol pouvait être exécuté par des ordinateurs de puissance moindre. Ce fait, en plus des développements de l'informatique qui participèrent à réduire le coût des ordinateurs, consolida la place de l'informatique dans la visualisation scientifique. Le programme fut ainsi développé et implémenté sur les plateformes Windows, Linux et Macintosh, évoluant jusqu'à devenir un logiciel de visualisation moléculaire complet.

En 1993, après la soutenance de thèse de Roger Sayle, RasMol devient libre de droit et le code source est ainsi publié [84]. RasMol sera donc intégré à de nombreux programmes dérivés sur toutes les plateformes, ouvrant la voie au développement d'autres logiciels de visualisation moléculaire. De nos jours, il en existe un nombre très important, permettant d'afficher les données issues d'expériences de cristallographie aux rayons X, résonance magnétique nucléaire ou encore dynamique moléculaire. On peut par exemple citer VMD [85], PyMol [86],

Jmol (<http://www.jmol.org/>), ChimeraX [87].

Ces logiciels sont souvent distribués en *open source*, ce qui permet aux utilisateurs de proposer des outils et fonctionnalités nouvelles pour ces derniers. Dans le cadre des développements effectués pendant cette thèse, nous avons utilisé le logiciel de visualisation UnityMol qui sera présenté dans la section 3.2.2.

2.4.3 Modes de représentations classiques des objets biologiques

Les macromolécules biologiques sont des objets complexes, composés d'un grand nombre d'atomes (de quelques milliers à plusieurs dizaines de milliers, voire plus). Ainsi, utiliser différents modes de visualisation permet de simplifier leur structure et de faciliter l'observation et l'analyse des résultats.

Fil de fer / bâtons :

La représentation en fil de fer (ou lignes) et en bâtons représente les liaisons entre les atomes par des lignes et des cylindres, respectivement (figure 2.4 A et B). Le rendu fil de fer étant fait de lignes, il n'est pas le plus adapté pour la représentation de la profondeur et du volume, au contraire du mode en bâtons. En effet, les cylindres utilisés pour représenter les liaisons permettent, grâce à leur volume en trois dimensions, d'y apposer des ombres et des reflets et de faire varier leur diamètre apparent pour rendre l'illusion de profondeur.

Boules et bâtons :

Dans les deux modes de représentation précédents, la position des atomes est suggérée par l'intersection des lignes ou cylindres représentant les liaisons. Le mode boules et bâtons est très similaire au mode bâtons et représente les liaisons de la même façon mais les atomes seront cette fois représentés par des sphères de diamètre équivalent pour tous les types d'atomes (figure 2.5 A).

VdW :

La représentation en sphères de van der Waals (VdW) représente les atomes sous forme de sphères et ne représente pas les liaisons entre atomes. Le diamètre des sphères est basé sur le rayon de van der Waals associé à l'atome considéré et représente donc son volume approximatif (figure 2.5 B).

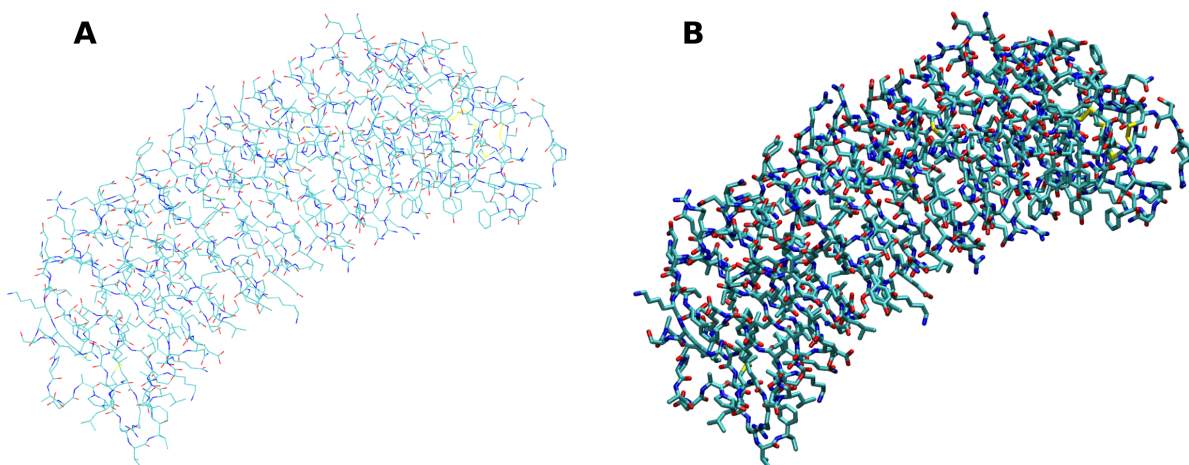


FIGURE 2.4 – **Représentations en fil de fer et en bâtons sur la fibromoduline.** L'identifiant de cette structure sur la PDB est 5MX0. **A** : Représentation en fil de fer. Les liaisons entre atomes sont représentées par des lignes. **B** : Représentation en bâtons. Les liaisons entre atomes sont cette fois représentées par des objets cylindriques en trois dimensions.

Cartoon :

La représentation en cartoon permet de simplifier la représentation des protéines en représentant la chaîne principale et les éléments de structure secondaire associés en utilisant des représentations spécifiques. Ainsi, les hélices α sont représentées par des serpentins et les brins β sont représentés par des flèches suivant le sens de la chaîne principale (figure 2.5 C).

Surface :

La représentation en surface permet d'évaluer l'enveloppe globale d'une protéine. Les surfaces sont souvent calculées en fonction du volume occupé par les sphères de van der Waals associés aux atomes de la protéine. Cette surface peut également être utilisée pour déterminer la surface accessible au solvant (*Solvent Accessible Surface* ou SAS) à l'aide d'une sonde représentant le solvant. Un rayon de sonde de 1,4 Å est couramment utilisé pour reproduire le rayon de la molécule d'eau. Cette sonde parcourt la surface de van der Waals et la surface accessible au solvant est définie par les positions accessibles par le centre de la sonde (figure 2.6).

Souvent, plusieurs de ces représentations sont utilisées sur une seule image afin de mettre en avant des événements ou des éléments de structure particuliers. On pourra par exemple représenter la structure globale d'une protéine en mode cartoon et représenter des acides aminés

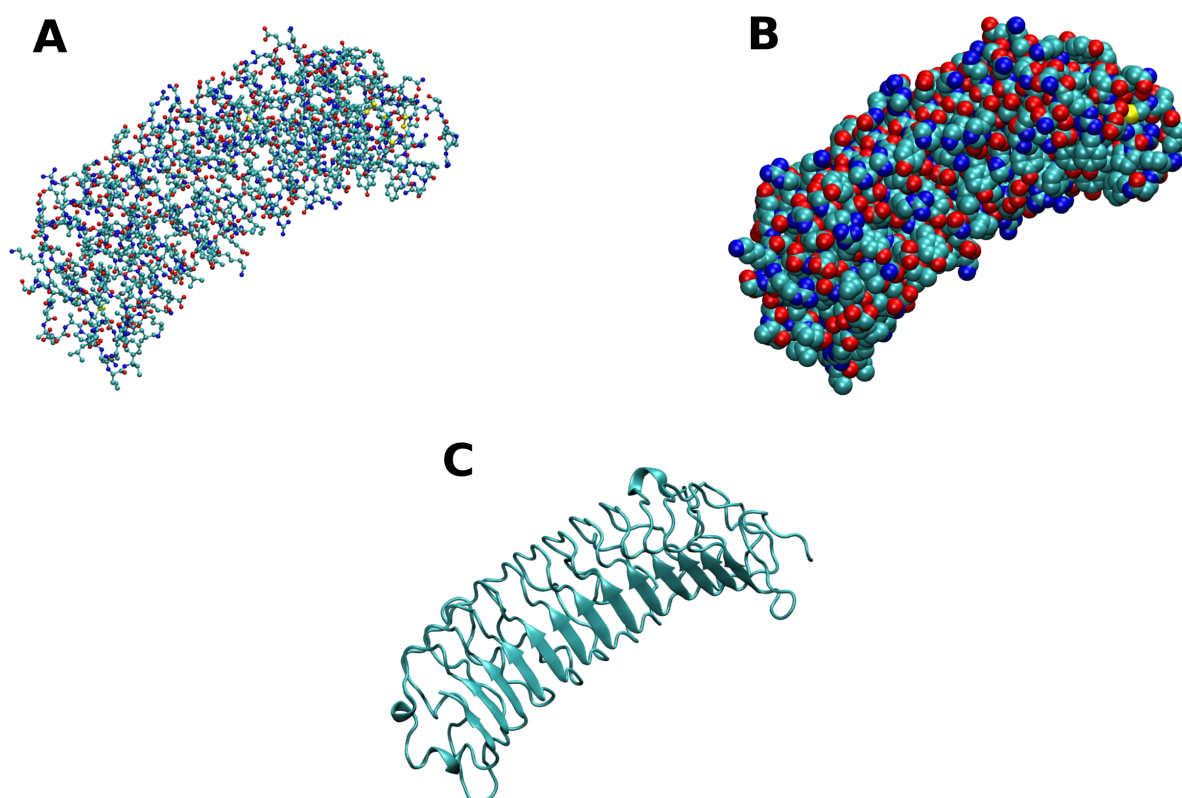


FIGURE 2.5 – **Représentations appliquées à la fibromoduline.** L'identifiant de cette structure sur la PDB est 5MX0. **A** : Représentation en boules et bâtons. **B** : Représentation en sphères de van der Waals. **C** : Représentation en cartoon.

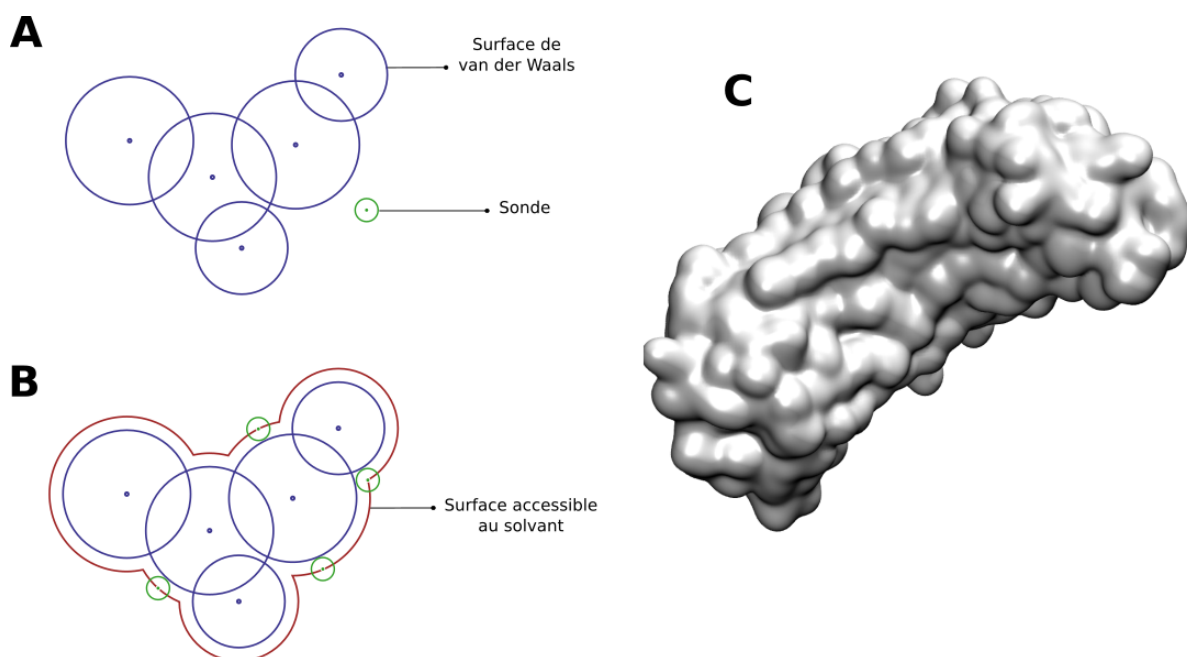


FIGURE 2.6 – **Principe de la représentation en surface.** **A** : La surface de van der Waals est représentée par l'espace occupé par les sphères de van der Waals associés aux atomes (en bleu). La sonde en vert, est utilisée pour déterminer la surface accessible au solvant. **B** : Calcul de la surface accessible au solvant (en rouge) à l'aide de la sonde. **C** : Représentation de la fibromoduline en utilisant le mode surface (identifiant PDB 5MX0).

d'intérêts (triade catalytique par exemple) en fil de fer. Comme mentionné précédemment, les modes de visualisation peuvent également être utilisés à des fins d'analyses. En effet, il est possible par exemple de colorer des éléments de structure en fonction d'information comme la nature des acides aminés, leur charge ou leur facteur thermique (aussi appelée *B factor*). En superposant ces modes de visualisation à la structure, il est possible d'en tirer des informations pertinentes. Par exemple, la figure 2.7 montre la structure cristallographique de la fibromoduline en représentation surface et colorée en fonction du facteur thermique associé aux atomes.

Ainsi, la visualisation est une partie intégrante du travail scientifique. Elle permet de mettre en valeur et analyser des résultats en jouant sur la forme des objets, leurs couleurs et / ou leur opacité. Grâce à la puissance des ordinateurs actuels, il est bien plus facile de développer et d'utiliser des modes représentation spécifiques à certains objets biologiques. La visualisation est donc devenue un aspect essentiel en modélisation moléculaire, au travers de logiciels complets. Au cours de ces travaux de thèse, nous avons ainsi utilisé les possibilités offertes par la visualisation pour tenter d'apporter un éclairage original sur la dynamique et la structure de glycanes. Ces objets complexes et flexibles sont des acteurs clés des fonctions cellulaires et, au travers de travaux de modélisation et de visualisation moléculaire, nous avons ainsi cherché à apporter un nouveau regard sur ces derniers.

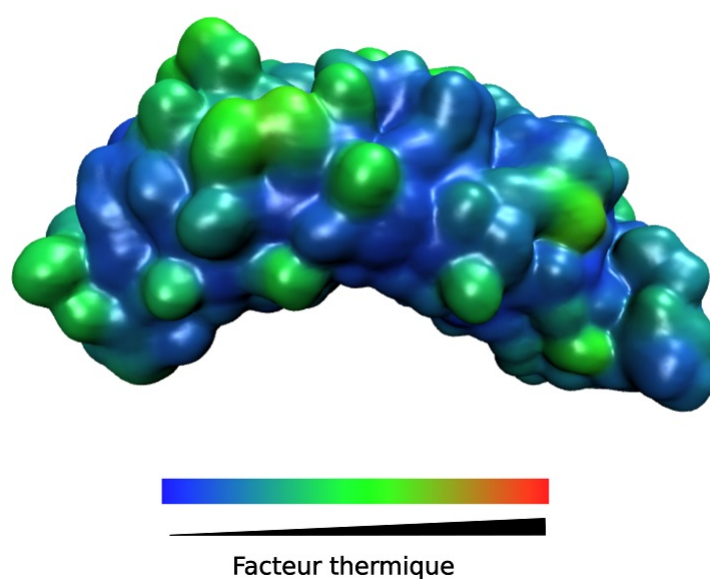


FIGURE 2.7 – **Exemple de coloration de la fibromoduline en fonction de son facteur thermique.** La fibromoduline est représentée en surface de van der Waals. La couleur superposée à cette surface correspond au facteur thermique de la structure cristallographique (identifiant PDB : 5mx0) et traduit l'agitation ou le désordre de ces régions. La couleur bleue indique des régions très stables et donc peu désordonnées.

Chapitre 3

Matériel et méthodes

3.1 Dynamique moléculaire : principes et usages

La modélisation moléculaire est un vaste domaine scientifique regroupant plusieurs méthodes et techniques basées sur les propriétés physiques et chimiques des atomes afin de réaliser des expériences de simulation permettant de comprendre et prédire les comportements de divers systèmes.

Deux grandes familles d’approches peuvent être définies : la mécanique moléculaire d’une part et la dynamique moléculaire d’autre part. Tandis que le but de la mécanique moléculaire est de déterminer pour une structure moléculaire donnée les positions des atomes qui permettent de minimiser l’énergie potentielle de celle-ci (parmi les techniques courantes on peut citer la minimisation en énergie ou encore l’amarrage moléculaire), la dynamique moléculaire a pour objectif l’étude de l’évolution de cette structure au cours du temps. Il est bien sûr possible de coupler différentes techniques provenant de ces deux approches afin d’accélérer la recherche de structures optimisées (*replica exchange*, *simulated annealing*, ...).

Le but de cette section est de détailler les principes et étapes d’une expérience de dynamique moléculaire.

3.1.1 Principes de la dynamique moléculaire

Dynamique *ab initio*

Un atome est composé d’un noyau entouré d’un nuage d’électrons. Un électron possédant une masse bien inférieure à celle d’un proton (environ 0,00055 unité atomique (ua) pour l’électron contre 1,00730 ua pour le proton, soit une masse 1836 fois inférieure), il est admis que les électrons vont s’adapter aux mouvements du noyau. Cette approximation, publiée en 1927 par

Born et Oppenheimer [88], permet ainsi de découpler les mouvements des électrons et du noyau.

En dynamique *ab initio* (aussi appelée AIMD pour *ab initio molecular dynamics*), les noyaux des atomes sont considérés comme étant fixes à un instant t et les calculs se concentrent essentiellement sur les mouvements des nuages électroniques en fonction de leur environnement. Ces méthodes prennent en compte la structure complète des orbitales électroniques (électrons de valence et électrons de cœur). Ainsi, ce sont des calculs qui ne sont pas adaptés à des gros systèmes puisque le nombre d'atomes multiplie le nombre d'électrons à considérer et donc les temps de calcul. Ces calculs sont cependant très précis et permettent de modéliser des réactions chimiques impliquant la formation ou la disparition de liaisons covalentes. Les méthodes *ab initio* permettent également l'accès à de nombreuses informations sur l'état des couches électroniques des atomes.

Dynamique moléculaire semi-empirique

Là où les méthodes *ab initio* prennent en compte tous les électrons des atomes, les méthodes semi-empiriques simplifient ces calculs en ne traitant explicitement que les électrons de valence (dernière couche électronique). Les électrons des couches internes sont inclus dans le cœur nucléaire. Ces approximations nécessitent l'introduction de paramètres issus de données expérimentales ou de calculs *ab initio* sur des molécules de référence. Ces méthodes sont ainsi jusqu'à trois fois plus rapides que les méthodes *ab initio*, applicables à des molécules de taille plus importante et permettent d'étudier l'apparition ou la disparition de liaisons covalentes [89]. Cependant, ces méthodes sont encore trop lentes pour les systèmes les plus imposants. Dans ce cas, les méthodes dites empiriques, jusqu'à trois fois plus rapides, peuvent être utilisées [89].

Dynamique moléculaire empirique

A la différence des méthodes *ab initio*, les méthodes dites empiriques simplifient la structure des atomes et les considèrent comme des points possédant des coordonnées cartésiennes dans l'espace, une masse et une charge. Ces méthodes se basent sur l'utilisation d'une fonction d'énergie potentielle (décrite ci-après) décrivant les interactions entre atomes liés et distants au sein du système considéré. Parce qu'elles ne prennent pas en compte explicitement le nuage électronique, les méthodes empiriques ne permettent pas d'étudier la création ou la disparition de liaisons covalentes au cours de la simulation. Pour ce faire, des méthodes hybrides QM/MM (pour *Quantum Mechanics / Molecular Mechanics*) permettent de repasser à l'échelle quantique au niveau de la liaison à modifier. Les méthodes empiriques permettent cependant un gain de temps significatif et la modélisation de systèmes bien plus importants (environ 10^6 particules pour des simulations pouvant atteindre la microseconde) que les méthodes *ab-initio* (limitées à environ 10^2 particules

pour des simulations de l'ordre de la picoseconde) et semi-empiriques [90].

Dans le cadre de nos travaux, nous avons ainsi utilisé les méthodes classiques de dynamique moléculaire utilisant une fonction d'énergie potentielle empirique.

Fonction d'énergie potentielle

La seconde loi de Newton permet de décrire le mouvement d'un objet en fonction des forces qui s'y appliquent. Plus précisément, dans le cas d'un système de masse constante, la somme des forces $\vec{F}$ exercée sur une particule de masse m est égale à la masse de la particule multipliée par l'accélération $\vec{a}$ de cette dernière :

$$\sum \vec{F} = m\vec{a}$$

Une force résulte de l'expression d'un gradient d'énergie $\vec{F} = -\vec{\text{grad}}E_{\text{totale}}$ (avec E_{totale} l'énergie potentielle). Ainsi, évaluer l'énergie potentielle des particules d'un système permet, pour chaque particule du système, d'évaluer la force à laquelle il est soumis et ainsi obtenir l'accélération puis la vitesse et les coordonnées des atomes au cours du temps (voir figure 3.4 illustrant le principe de la dynamique moléculaire).

Cette fonction d'énergie potentielle, telle que définie dans les champs de force moléculaire, est composée de cinq termes décrivant les énergies au sein du système entre atomes liés par une ou plusieurs liaisons covalentes (interactions liées) et atomes considérés comme non-liés par une liaison covalente (interactions non-liées).

L'équation finale de l'énergie potentielle correspondra ainsi à la somme de l'ensemble de ces termes d'énergie :

$$E_{\text{totale}} = E_{\text{liaisons}} + E_{\text{angles}} + E_{\text{torsions}} + E_{\text{Elec.}} + E_{\text{LJ}}$$

Terme d'interactions liées : Ce terme est composé de trois contributions principales (E_{liaisons} , E_{angles} , E_{torsions}) décrivant respectivement les vibrations des liaisons covalentes, les variations des valeurs des angles de valence et des angles de torsion (ou angle dièdre). Chacun de ces termes est calculé en fonction de la géométrie de la molécule et des paramètres définis par le champ de forces utilisé.

1. Énergie potentielle des liaisons covalentes. Ce terme décrit l'oscillation de la liaison l_i autour de la valeur idéale l_0 (longueur d'équilibre) et prend la forme d'une fonction har-

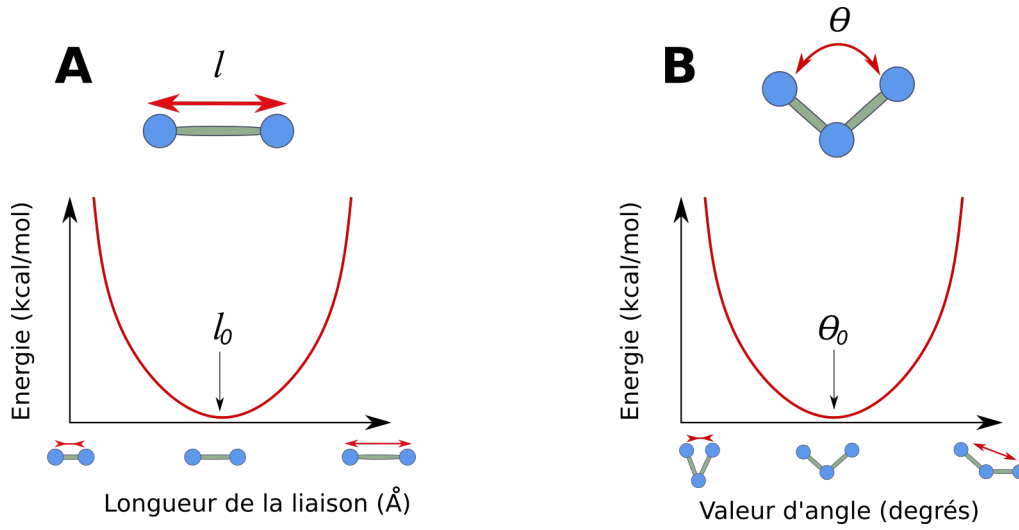


FIGURE 3.1 – **A** : Fonction harmonique décrivant l'énergie en fonction de la longueur de la liaison entre deux atomes. La longueur optimale (longueur d'équilibre) l_0 se situe au minimum d'énergie. **B** : Fonction harmonique décrivant l'énergie en fonction de la valeur de l'angle formé par trois atomes. La valeur d'angle optimale θ_0 se situe au minimum d'énergie.

monique. k^l est une constante définie dans le champ de forces représentant la raideur de la liaison (figure 3.1 A).

$$E_{liaisons} = \sum_{liaisons\ i} k_i^l (l_i - l_0)^2$$

2. Énergie potentielle des angles de valence. De la même façon que pour l'énergie de la longueur des liaisons covalentes, la fonction harmonique utilisée décrit l'énergie associée à l'angle θ_i comme une fonction de la variation de celui-ci autour de la valeur idéale θ_0 . k^a est une constante définie dans le champ de forces et représente la "raideur" de l'angle autour de sa valeur idéale (figure 3.1 B).

$$E_{angles} = \sum_{angles\ i} k_i^a (\theta_i - \theta_0)^2$$

3. Énergie potentielle des angles de torsion. Un angle dièdre implique 4 atomes et est défini par l'angle entre deux plans définis par trois atomes. Sur la figure 3.2, ces plans sont définis par les atomes A, B, C et B, C, D. La fonction cosinus permet de représenter par un phénomène périodique la rotation des atomes sur 360 degrés. Les constantes n et δ dépendent des atomes qui composent l'angle de torsion et ϕ_i est la valeur d'angle variable. V_m est une constante (figure 3.2).

$$E_{torsions} = \sum_{torsions\ i} \frac{V_m}{2} (1 + \cos(n\phi_i - \delta))$$

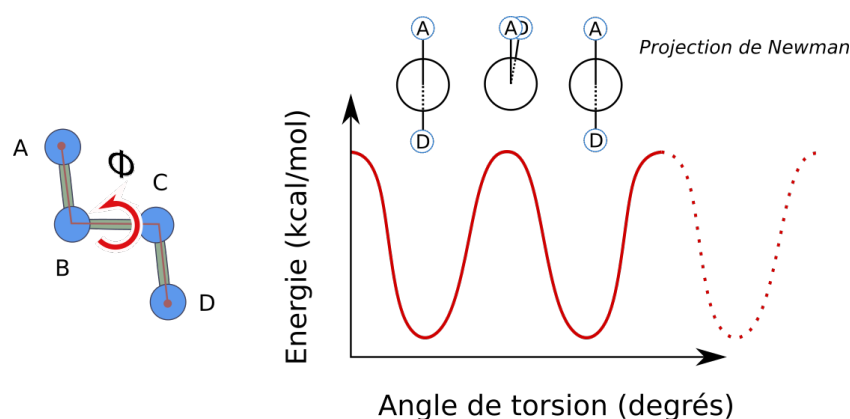


FIGURE 3.2 – Fonction périodique décrivant l'énergie en fonction de la position des atomes d'un angle dièdre considéré. A gauche, les positions de 4 atomes sont nécessaires pour définir l'angle dièdre considéré. A droite, au dessus du graphe de l'énergie, les projections de Newman selon la liaison BC sont présentées.

Terme d'interactions non-liées : Ce terme d'énergie est composé deux contributions décrivant les forces électrostatiques (définies par la loi universelle de Coulomb) et les forces de van der Waals (d'origine quantique et résultant de l'impossibilité de superposer des électrons dans un même état).

4. Interactions électrostatiques. La loi de Coulomb décrit les forces d'attraction d'atomes de charges opposées et les forces de répulsion d'atomes de charges identiques. La distance entre les atomes est appelée r , q_i et q_j sont les charges des atomes considérés; ϵ ($\epsilon = \epsilon_r \epsilon_0$) correspond à la constante diélectrique qui représente l'effet de l'environnement faisant écran aux charges des atomes. La constante ϵ_r peut être utilisée pour simuler l'environnement de manière implicite et peut prendre une valeur entre 1 (pour le vide) et 80 (pour l'eau). Si le solvant est modélisé explicitement, elle est fixée à 1. Le profil d'énergie de ce terme est décrit dans la figure 3.3 A.

$$E_{elec.} = \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi \epsilon r_{ij}}$$

5. Potentiel de Lennard-Jones. Ce potentiel est utilisé pour décrire l'énergie potentielle liée aux forces de van der Waals et est composé de deux termes. Le premier décrit la répulsion à courte distance (en r^{-12}) des nuages électroniques de deux atomes, empêchant ainsi les recouvrements stériques (répulsion de Pauli). Le second terme représente les forces de van der Waals qui agissent à plus longue distance (en r^{-6}). La valeur de σ_{ij} dépend de la nature des atomes i et j . ϵ_{ij} décrit la profondeur du puit de potentiel associé à ces

atomes, comme décrit sur la figure 3.3 B.

$$E_{LJ} = \sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

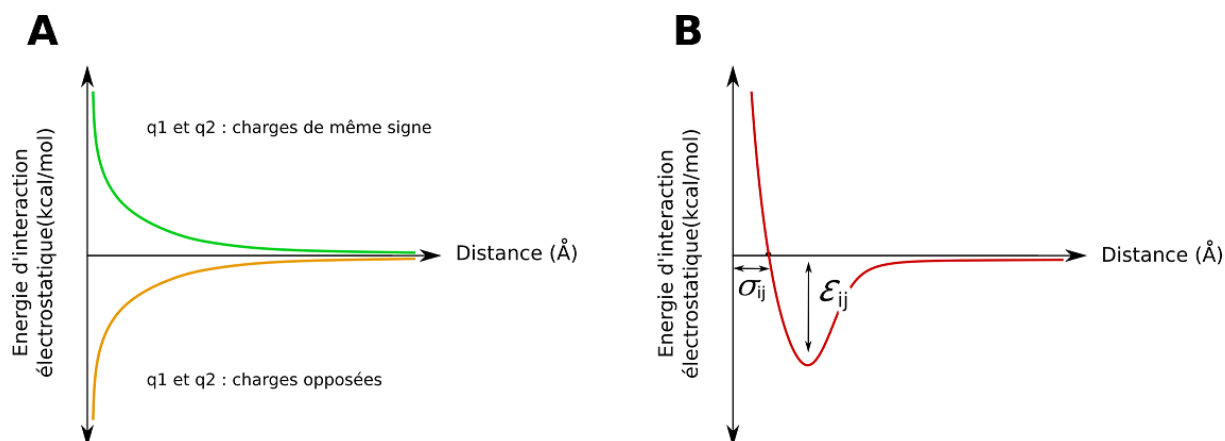


FIGURE 3.3 – **Fonction d'énergie des interactions non-liées.** **A** : Evolution de l'énergie d'interaction entre deux atomes en fonction de la distance inter-atomique. Des atomes de charges opposées ont une énergie d'interaction négative, et donc favorable (courbe orange) tandis que des atomes de même charge se repoussent en raison de leur énergie d'interaction positive (courbe verte). **B** : Variation du potentiel de Lennard-Jones entre deux atomes en fonction de la distance inter-atomique.

Pour chaque type d'atome, les paramètres de masse, charge, longueurs de liaisons, valeurs d'angles... sont décrits par le champ de forces. Les valeurs de ces paramètres peuvent être dérivées à partir de calculs *ab-initio* ou déterminées à l'aide de données expérimentales (par exemple des expériences de cristallographie aux rayons X).

C'est ensuite l'application de la seconde loi de Newton qui va nous permettre de déterminer les mouvements des atomes. Comme décrit précédemment, celle-ci relie la somme des forces qui s'appliquent à un objet à sa masse et son accélération. La masse des atomes étant connue et définie dans le champ de forces, il est donc possible de calculer l'accélération de chaque atome à un temps t et de l'utiliser pour calculer la position de l'atome au temps $t + \delta t$ (figure 3.4). L'intervalle de temps δt est appelé "pas d'intégration".

Le choix du pas d'intégration est une étape importante. En effet, les expériences de dynamique moléculaire visent à décrire correctement le mouvement des atomes les uns par rapport aux autres, y compris les mouvements moléculaires les plus rapides. Un pas d'intégration trop grand conduirait à manquer certains mouvements, créant des approximations et des instabilités dans le système. Un pas trop petit augmenterait drastiquement le temps de calcul en multipliant les étapes. Le mouvement intramoléculaire le plus rapide, situé autour des 1 fs, est la vibration

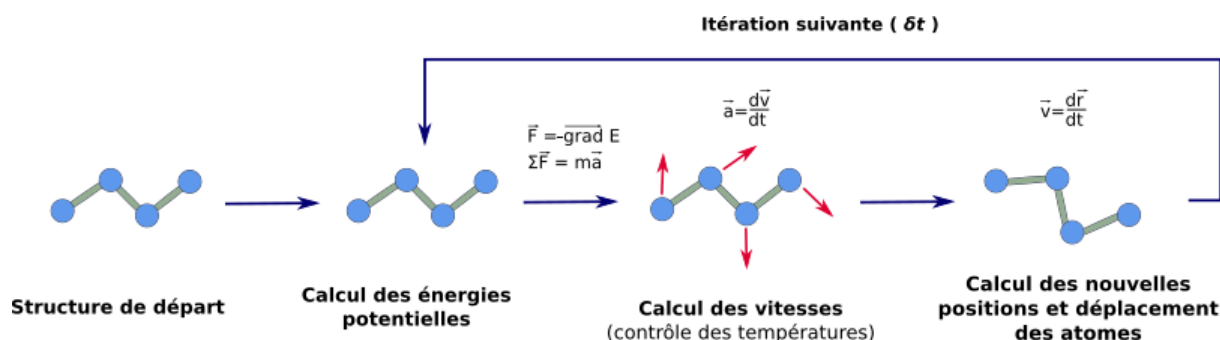


FIGURE 3.4 – **Processus d’une étape de dynamique moléculaire.** A chaque étape, les énergies potentielles de tous les atomes sont calculées et utilisées pour déterminer leur accélération puis leur vitesse et la direction de leur déplacement. Leurs nouvelles positions sont alors calculées et appliquées. Ces étapes sont répétées un grand nombre de fois N et permettent d’obtenir une trajectoire de dynamique moléculaire de durée $N\delta t$, avec δt le pas d’intégration.

entre les atomes d’hydrogène et les atomes lourds du système. Cependant, étant donné le petit volume et la faible masse des atomes d’hydrogène, on peut supposer que leurs mouvements n’influenceront que peu les mouvements au sein d’une protéine. Il est donc possible de fixer la longueur des liaisons entre atomes d’hydrogène et atomes lourds. La plus petite fréquence intramoléculaire est alors celle entre les atomes de carbone, située autour de 2 fs, et permettant de doubler le temps de la trajectoire pour le même temps de calcul.

Le nombre de pas d’intégration qu’il est possible de réaliser dans un temps donné dépendra de nombreux paramètres, notamment du nombre d’atomes du système et du nombre de processeurs utilisés pour les calculs. Grâce à la puissance des supercalculateurs, il est possible d’effectuer ces calculs sur des systèmes de plusieurs milliers d’atomes, voire plus. Dans nos travaux, nous avons par exemple utilisé un système composé d’une fibromoduline glycosylée de près de 260 000 atomes (figure 3.5).

3.1.2 Mise en place d’une expérience de dynamique moléculaire

Avant la production de données exploitables, le système doit cependant passer par plusieurs étapes de préparation visant à s’assurer que la structure de départ est stable et ne présente pas d’aberrations de structure. Elle est ensuite mise en présence de solvant s’approchant des conditions biologiques.

Structure de départ

La première étape est d’obtenir une structure de la / des molécules d’intérêt. Les structures protéiques résolues grâce à la cristallographie aux rayons X, la Résonance Magnétique

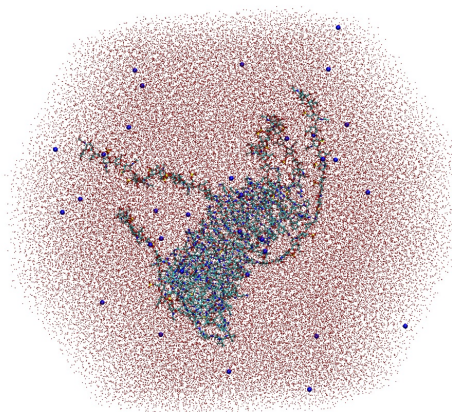


FIGURE 3.5 – **Exemple de système de grande taille étudié par des trajectoires de dynamique moléculaire.** La protéine d'intérêt (fibromoduline portant des chaînes de keratan sulfate) est représentée sous forme de bâtons. Les molécules d'eau sont représentées sous forme de points rouges et les ions sodium sous forme de sphères bleues. Ce système est composé de près de 260 000 atomes.

Nucléaire (RMN) ou la cryo microscopie sont déposées sur la Protein Data Bank (ou PDB, <http://www.rcsb.org/>). Les structures ainsi obtenues se présentent sous la forme de fichiers textuels au format pdb listant les atomes de la molécule et leurs positions respectives. La figure 3.6 illustre le format de ces fichiers sur les atomes d'une alanine. Chaque ligne correspond à un atome. Les premières colonnes répertorient les informations sur cet atome et le résidu (acide aminé) auquel il appartient. Les trois colonnes suivantes, les plus importantes, (labelisées x, y et z sur la figure 3.6) donnent les coordonnées cartésiennes dans l'espace (en Angstroem Å) de cet atome. Les deux colonnes suivantes sont un facteur d'occupation et un facteur d'agitation thermique reflétant l'oscillation moyenne de l'atome autour d'une position d'équilibre. Plus cette valeur est élevée, plus on peut supposer qu'une région protéique est flexible.

Ces fichiers peuvent être lus par les logiciels de visualisation moléculaire afin d'afficher la structure de la molécule associée.

Préparation du système

Avant de démarrer l'étape de production, il faut dans un premier temps s'assurer que la structure de la protéine est stable. Dans le cas inverse (conflits stériques par exemple), la fonction d'énergie potentielle telle que définie précédemment prendra une valeur positive très élevée, indiquant le caractère défavorable de cette conformation. Une étape de minimisation d'énergie permet d'évaluer et de trouver des valeurs d'énergie potentielle totale correspondant à des

Atome Résidu					x	y	z		
ATOM	1468	N	ALA	X 92	116.900	100.070	93.440	0.00	0.00
ATOM	1469	H	ALA	X 92	117.420	100.870	93.130	0.00	0.00
ATOM	1470	CA	ALA	X 92	117.640	98.810	93.320	0.00	0.00
ATOM	1471	HA	ALA	X 92	116.960	97.990	93.090	0.00	0.00
ATOM	1472	CB	ALA	X 92	118.570	98.870	92.050	0.00	0.00
ATOM	1473	HB1	ALA	X 92	119.330	99.640	92.190	0.00	0.00
ATOM	1474	HB2	ALA	X 92	119.010	97.880	91.930	0.00	0.00
ATOM	1475	HB3	ALA	X 92	117.990	99.130	91.160	0.00	0.00
ATOM	1476	C	ALA	X 92	118.450	98.490	94.630	0.00	0.00
ATOM	1477	O	ALA	X 92	118.460	97.360	95.190	0.00	0.00

FIGURE 3.6 – **Exemple de fichier au format pdb sur une Alanine.** Les trois premières colonnes permettent d'identifier les atomes. Les trois suivantes sont dédiées à l'acide aminé auquel cet atome appartient. Les trois colonnes x, y et z donnent les coordonnées de ces atomes dans l'espace en Angstroem.

structures stables, d'énergie inférieure à celle de la structure de départ et représentées par des minima locaux sur la fonction d'énergie potentielle (figure 3.7). Le minimum global correspond à la conformation la plus stable adoptée par le système. Étant donné le grand nombre de termes à minimiser (énergie potentielle de chacun des atomes du système), trouver le minimum global de cette fonction est très difficile mais différents algorithmes permettent de rejoindre l'un des minima locaux à partir de la dérivée de la fonction d'énergie potentielle. Par exemple, la méthode *steepest descent*, très utilisée, évalue la dérivée première de la fonction d'énergie afin d'évaluer localement la pente la plus forte et faire évoluer le système dans ce sens pour rejoindre le minimum local associé.

Cette première étape, réalisée en absence de solvant, ne prend en compte que les interactions intramoléculaires et permet de résoudre d'éventuels conflits stériques présents dans la structure.

La dynamique moléculaire implique de simuler les mouvements des molécules biologiques, ce qui implique la prise en compte d'un solvant recréant les conditions cellulaires. L'étape suivante est donc l'ajout de molécules d'eau et d'ions. La protéine est ainsi placée dans une boîte imaginaire et l'espace vide autour de la protéine est ensuite comblé par les molécules d'eau et les ions. Les limites de cette boîte sont cependant toujours au contact du vide, créant des artefacts appelés effets de bords. Des conditions périodiques aux limites sont donc appliquées à cette boîte : le système est représenté périodiquement et chaque "face" de la boîte est mise en contact avec son côté opposé. Ainsi, si une molécule d'eau sort par la face droite de la boîte, son image périodique y ré-entrera par la face gauche (figure 3.8). Dans ce cas de figure, seule la boîte centrale est modélisée explicitement.

Dans les travaux présentés ici, le solvant a été modélisé explicitement. Le modèle utilisé pour

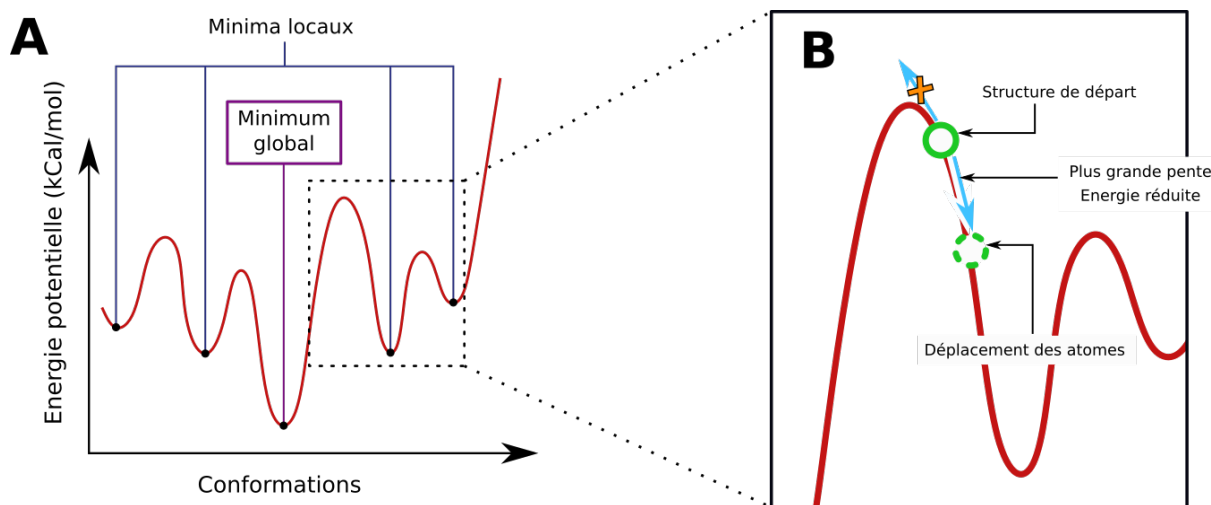


FIGURE 3.7 – **Principe de la minimisation d'énergie.** **A** : Représentation simplifiée d'une surface d'énergie potentielle. Les puits de potentiel correspondent aux minima locaux que l'on cherche à atteindre grâce aux algorithmes de minimisation. Ces puits correspondent à des structures stables. Les pics représentent des barrières d'énergie associées à des conformations peu stables. **B** : Exemple de la méthode de minimisation *steepest descent*. La pente (dérivée première de la fonction d'énergie) est évaluée localement autour de la structure de départ. Les atomes de la molécule sont ensuite déplacés pour rejoindre cette conformation d'énergie réduite. Ces étapes sont répétées jusqu'à ce qu'un minima local soit trouvé.

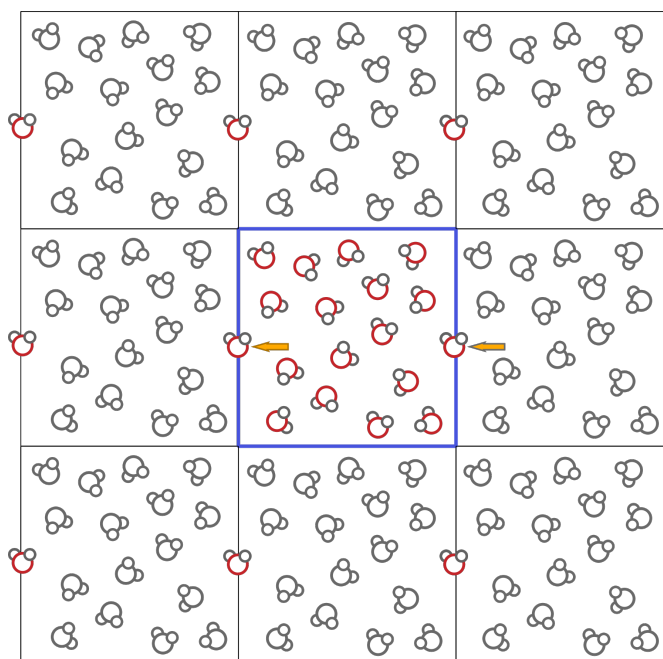


FIGURE 3.8 – **Illustration des conditions périodiques aux limites.** Seule la boîte centrale est modélisée explicitement. Pour chaque molécule (ici, de l'eau) qui sort de cette boîte par une face, son image identique rentre dans la boîte par l'autre face.

les molécules d'eau est appelé TIP3P : ce modèle rigide (signifiant que l'angle formé par les trois atomes ainsi que la longueur des liaisons ne varient pas pendant la trajectoire) comporte trois points possédant une charge et, pour l'oxygène, des paramètres de Lennard-Jones. La charge négative est portée par l'oxygène et la charge positive est portée par les deux atomes d'hydrogène.

Une fois le solvant ajouté, l'étape d'équilibration permet de s'assurer que le solvant est réparti uniformément au sein de la boîte afin d'éviter l'apparition de bulles de vide. Pour cela, dans un premier temps, l'ensemble NVT ou ensemble canonique est utilisé. Cela signifie que le nombre d'atomes N , le volume V et la température T sont constants. Cette étape permet aux molécules de solvant de se relaxer à la température T autour de la protéine puisque cette dernière est restreinte dans ses mouvements par des contraintes afin de laisser le solvant se répartir correctement autour de la protéine.

La deuxième étape utilise l'ensemble NPT et permet de s'assurer que la pression (et donc la densité) du système est stable. L'ensemble NPT ou ensemble isotherme-isobare signifie que le nombre d'atomes N , la pression P et la température T sont constants.

L'ensemble NPT est celui qui se rapproche le plus des conditions expérimentales et sera ainsi utilisé pour l'étape de production. Une fois le système équilibré, toutes les contraintes sont ainsi relâchées afin de laisser le système évoluer librement. Cette étape, permettant de générer jusqu'à plusieurs centaines de nanosecondes, est la plus longue et nécessite l'utilisation de supercalculateurs.

3.1.3 Préparation et paramètres utilisés lors des simulations

Glycanes isolés en solvant

Dans un premier temps, nous avons réalisé des trajectoires de dynamique moléculaire sur des glycanes tétra-antennés et bisectants en chaînes isolées. Les structures de départ ont été générées avec l'outil Glycan Reader [73] du serveur Charmm-Gui (<http://www.charmm-gui.org>). Une première étape de minimisation dans le vide avec la méthode *Steepest Descent* pendant 5000 pas a permis de résoudre les éventuels recouvrements stériques. Puis le système a été solvatoé dans une boîte cubique en solvant explicite et en présence de contre-ions. La table 3.1 résume les paramètres de remplissage des boîtes de simulation pour chacun des systèmes simulés. Deux étapes d'équilibration ont été réalisées : la première avec l'ensemble NVT et la seconde avec l'ensemble NPT pendant 100 ps par étape. La température du système a été portée à 310 K et la pression à 1 bar. Le cut-off utilisé pour les interactions électrostatiques et le potentiel de Lennard-Jones est de 1,2 nm. La préparation du système et la production de la trajectoire ont été réalisées avec Gro-

macs (package 5.4) [91,92] et le champ de forces Charmm36 [74]. La production des trajectoires a été réalisée grâce au centre de calcul Romeo en mobilisant 4 processeurs de 16 cœurs (64 cœurs).

Nous avons ainsi produit des trajectoires sur un total de 5 structures, toutes fucosylées : deux glycanes bi-antennés avec une N-acétylglucosamine bisectante (une structure avec acides sialiques, une structure sans acides sialiques) et trois glycanes tétra-antennés : une structure sans acides sialiques, une structure avec les acides sialiques liés en $\alpha 2-3$ et une structure avec les acides sialiques liés en $\alpha 2-6$. La figure 3.9 répertorie ces systèmes et leurs structures.

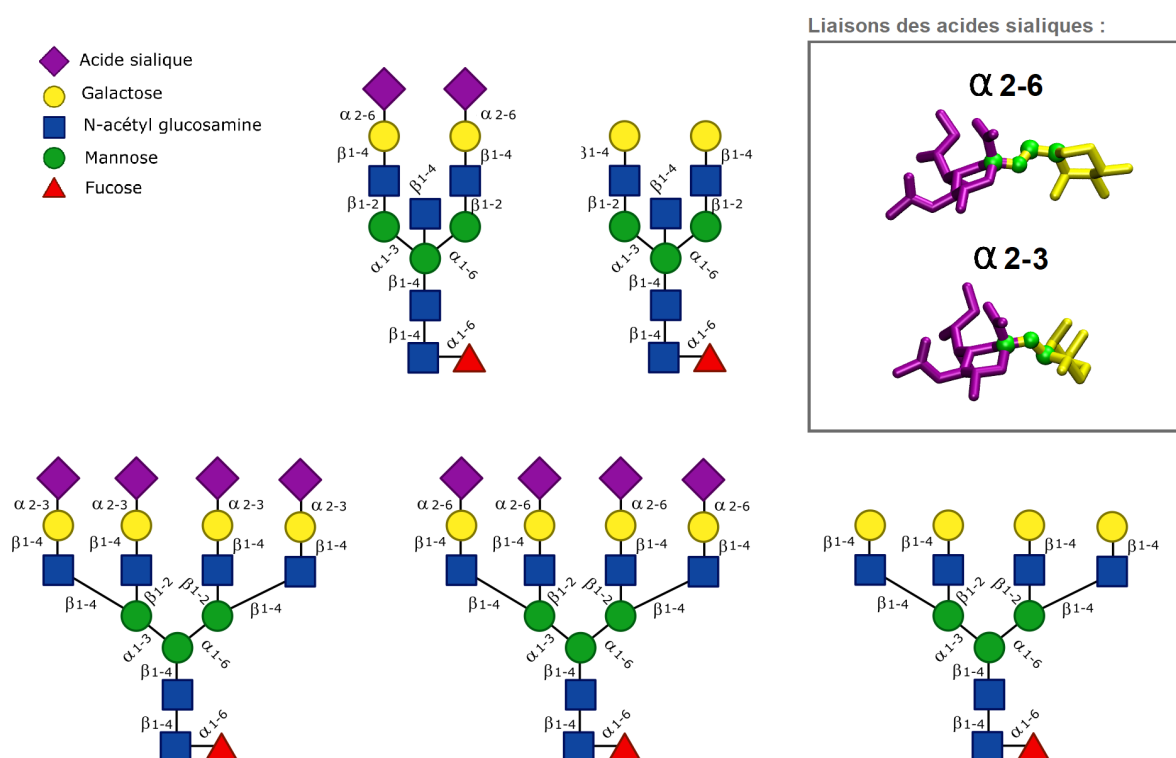


FIGURE 3.9 – **Structures des glycanes étudiés par les simulations de dynamique moléculaire.** Deux glycanes bisectants (sans acides sialiques et avec acides sialiques liés en $\alpha 2-6$) et trois glycanes tétra-antennés (sans acides sialiques, avec acides sialiques liés en $\alpha 2-6$ et avec acides sialiques liés en $\alpha 2-3$) ont été utilisés. La nomenclature SNFG est utilisée. Les types de chaque liaisons osidiques sont indiqués. Les deux types de liaisons des acides sialiques sont illustrés dans l’encadré. Les monosaccharides sont colorés selon la nomenclature SNFG et les atomes de la liaison sont représentés par les sphères vertes. La liaison en $\alpha 2-6$ comporte un atome de plus que la liaison en $\alpha 2-3$.

Nous avons généré une trajectoire de 500 ns pour les systèmes bi-antennés bisectants. Pour chacun des systèmes tétra-antennés, nous avons généré trois trajectoires de 500 ns chacune avec une structure de départ différente. Les tables 3.1 et 3.2 présentent la composition des systèmes

solvatés.

	Tétra-antenné	Tétra-antenné + AS α 2-3	Tétra-antenné + AS α 2-6
Nombre d'atomes	17 590	26 021	35 432
Nombre de molécules d'eau	5 748	8 505	11 646
Nombre et type d'ions	0	4 ions K ⁺	4 ions K ⁺

TABLE 3.1 – Détails des systèmes de glycanes étudiés par les trajectoires de dynamique moléculaire des glycanes tétra-antennés. Les mentions α 2-3 et α 2-6 décrivent la liaison utilisée pour les acides sialiques terminaux présents sur ces structures (respectivement, liaison en α 2-3 et α 2-6).

	Bi-antenné bisectant	Bi-antenné bisectant + AS α 2-6
Nombre d'atomes	18 550	19 353
Nombre de molécules d'eau	6 090	6 333
Nombre et type d'ions	0	2 ions K ⁺

TABLE 3.2 – Détails des systèmes de glycanes utilisés pour les trajectoires de dynamique moléculaire des glycanes bi-antennés bisectant. La mention AS α 2-6 décrit la liaison utilisée pour les acides sialiques terminaux présents.

Fibromoduline et fibromoduline glycosylée en solvant

Les trajectoires de dynamique moléculaire sur la fibromoduline glycosylée ont été réalisées à partir de la structure de la fibromoduline humaine disponible sur la PDB (identifiant 5MX0) [93]. Les sites de glycosylations sélectionnés ont été choisis en fonction des données présentées par Pietraszek-Gremplewicz *et al.* [94]. Sur les cinq sites identifiés sur la banque de données Uniprot (identifiant Q06828), nous avons choisis les 4 sites présentant au moins un saccharide sur la structure pdb, c'est à dire les sites N127, N166, N201 et N341.

Les glycanes ont été construits avec l'outil *Carbohydrate Builder* sur le portail web Glycam (<http://glycam.org/>). La longueur et la composition des chaînes ont été choisies en fonction des données présentées dans Lauder *et al.* [95]. Les N-Acétyle-glucosamines ont toutes été phosphorylées. La conformation des sucres, et plus particulièrement du fucose, présents dans le fichier pdb d'origine a été conservée (connectivité α 1-6 pour le fucose). La figure 3.10 A représente schématiquement la composition des chaînes de keratan sulfate greffées sur la protéine et la figure 3.10 B représente la structure de départ utilisée pour les trajectoires de dynamique moléculaire.

Afin de neutraliser le système, nous avons redistribué les charges sur l'asparagine glycosylée : en utilisant le résidu "NLN" du champ de forces Glycam (modèle de l'asparagine glycosylée), le système obtenu n'était pas neutre. En effet, afin de greffer le glycanes sur cet acide aminé, le groupement hydroxyle terminal du glycanes isolé est supprimé. La disparition de cette charge n'a pas été compensée sur les atomes du système. Nous avons donc recréé un fichier lib pour répartir cette charge sur l'asparagine glycosylée. Ces fichiers permettent de définir les paramètres (charges, géométrie...) d'un résidu ou d'une molécule qui n'existe pas dans le champ de forces. Ce type de fichier est souvent utilisé pour définir des petits ligands ou des résidus modifiés non-standard. Ici, nous avons utilisé ces fonctionnalités pour redéfinir les charges de l'asparagine portant la glycosylation. La charge de l'atome d'oxygène terminal OD1 et du carbone CG ont ainsi été modifiées. La figure 3.3 met en évidence l'emplacement de ces atomes sur la structure de l'asparagine. La table 3.11 présente la modification apportée aux charges de ces atomes afin d'obtenir un système neutre.

Les deux systèmes (fibromoduline glycosylée et fibromoduline seule) ont été placés dans une boîte de solvant en forme d'octaèdre tronqué et minimisés avec l'algorithme *Steepest Descent* pendant 500 000 pas. Le tableau 3.4 résume la composition des systèmes. Ensuite, deux étapes d'équilibration ont été effectuées : 250 ps avec l'ensemble NVT et 250 ps avec l'ensemble NPT. L'étape de production a été calculée avec l'ensemble NPT et un pas d'intégration de 2 fs à une température de 310 K et une pression de 1 bar. Ces calculs ont été effectués sur les supercal-



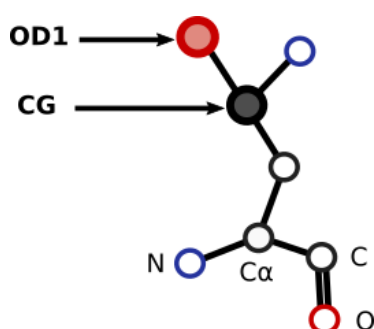


FIGURE 3.11 – Structure de l'asparagine. Les atomes portant les charges modifiées sont identifiés par les flèches.

Nom de l'atome	CG	OD1
Type de l'atome	C	O
Charge initiale (C)	0,7130	-0,5931
Charge modifiée (C)	0,6970	-0,6101

TABLE 3.3 – Modifications effectuées sur les charges de l'asparagine glycosylée afin de neutraliser le système.

culateurs du CINES avec le moteur Amber et les champs de forces ffb14 (pour la protéine) et GLYCAM06-j (pour les glycanes). Un cut-off de 1,3 nm a été utilisé pour les interactions non liées.

Nous avons généré 64 ns de trajectoire pour la fibromoduline glycosylée et avons mobilisé 12 nœuds de 24 cœurs chacun (pour un total de 288 cœurs) pendant environ 300 heures de calcul, soit 86 400 heures de calcul en temps processeur.

Pour la fibromoduline sans chaînes de keratan sulfate, nous avons généré 80 ns de trajectoire et mobilisé 8 nœuds de 24 cœurs chacun (pour un total de 192 cœurs) pendant environ 70 heures de calcul soit 13 440 heures de calcul en temps processeur.

	Fibromoduline glycosylée	Fibromoduline seule
Nombre d'atomes	258 934	34865
Nombre de molécules d'eau	84 065	9 999
Nombre et type d'ions	33 ions NA^+	0

TABLE 3.4 – Composition des systèmes de fibromoduline utilisés pour les trajectoires de dynamique moléculaire.

Récepteur à l'insuline glycosylé en solvant

Les trajectoires de dynamique moléculaire ont été générées sur l'ectodomaine (partie extracellulaire) du récepteur à l'insuline glycosylé. L'application de contraintes au niveau des trois derniers résidus C-terminaux de chaque monomère a permis de simuler l'ancrage à la membrane cellulaire. Huit types de glycanes différents ont été greffés au récepteur afin de former huit systèmes : quatre glycanes bi-antennés et quatre glycanes tri-antennés (glycane fucosylé, non

fucosylé, avec acides sialiques liés en α 2-6 non fucosylé et avec acides sialiques liés en α 2-6 et fucosylé). Chaque dimère du récepteur a été glycosylé sur deux sites, au niveau de l'asparagine 893. Ces trajectoires ont été générées pendant les travaux d'Alexandre Guillot [8]. La figure 3.12 A présente la structure du dimère et l'emplacement des points de glycosylation et la figure 3.12 B présente les huit glycanes considérés.

Les trajectoires de dynamique moléculaire du récepteur à l'insuline glycosylé ont été générées avec le package 5.0.2 du logiciel Gromacs et le champ de forces OPLS-AA [96,97]. L'ensemble NPT à 1 bar et 310 K a été utilisé et 250 ns de trajectoire ont été obtenues pour chaque système.

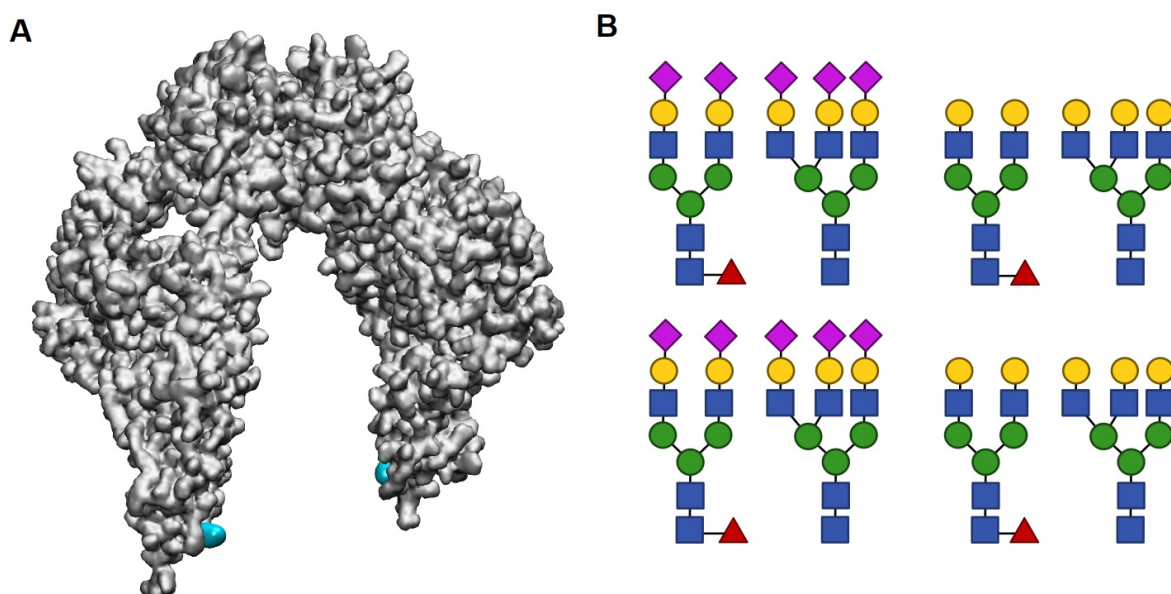


FIGURE 3.12 – **Structure du récepteur à l'insuline et glycanes utilisés pour les trajectoires.** **A** : Représentation en surface du dimère de l'ectodomaine du récepteur à l'insuline. Les asparagines glycosylées (une par monomère) sont représentées en cyan. **B** : Représentation schématique des huit glycanes étudiés.

3.1.4 Analyses des trajectoires

Clustering

Pour les glycanes en chaînes isolées et le récepteur à l'insuline glycosylé, la recherche des clusters a été réalisée à l'aide du module *cluster* de gromacs couplé à l'algorithme Gromos et en ne prenant en compte que les atomes lourds. Les *cut-offs* ont été choisis de sorte que la population cumulée des cinq premiers clusters obtenus représente 80 à 90% de la totalité des structures. Les représentants de chaque cluster ont été validés visuellement. Les structures utilisées ont été sélectionnées sur toute la longueur de la trajectoire à intervalles de 100 ps. La table 3.5 répertorie

le nombre de structures utilisées pour le clustering ainsi que les cut-offs utilisés pour analyser les conformations des chaînes glycosylées greffées sur le récepteur à l'insuline et la table 3.6 répertorie ces informations pour les trajectoires de glycanes isolés. L'alignement avant détermination des clusters a été fait sur le cœur du glycanes afin d'observer uniquement le mouvement des branches.

Système	Nombre de structures	Cut-off G1 (nm)	Cut-off G2 (nm)
RI + bi-antenné	2500	0,30	0,20
RI + bi-antenné + fucose	2500	0,30	0,20
RI + bi-antenné + AS	2500	0,25	0,30
RI + bi-antenné + fucose + AS	2500	0,35	0,25
RI + tri-antenné	2500	0,25	0,31
RI + tri-antenné + fucose	2500	0,35	0,34
RI + tri-antenné + AS	2500	0,33	0,33
RI + tri-antenné + fucose + AS	2500	0,31	0,31

TABLE 3.5 – Cut-off des clusters sur les trajectoires du récepteur à l'insuline (RI), en nm. G1 et G2 représentent le glycanes 1 et le glycanes 2 du dimère du récepteur à l'insuline.

Système	Nombre de structures	cut-off
Glycanes bisectant	5000	0,30
Glycanes bisectant + AS α 2-6	5000	0,40
Glycanes tétra-antenné	15000	0,35
Glycanes tétra-antenné + AS α 2-3	15000	0,47
Glycanes tétra-antenné AS α 2-3	15000	0,45

TABLE 3.6 – Cut-off des clusters sur les trajectoires des glycanes isolés, en nm.

Calcul des RMSD

Pour les trajectoires du récepteur à l'insuline glycosylé, les calculs de RMSD sur les atomes lourds des glycanes ont été effectués à l'aide du module *rms* de Gromacs. L'alignement a été effectué sur le noyau du glycanes en ignorant les atomes d'hydrogène et le RMSD a été calculé par rapport à la structure de départ.

Mise en œuvre et tests de performances de l'*Umbrella Visualization*

Pour ces travaux, nous avons utilisé les deux versions de l'*Umbrella Visualization*. Pour l'étude des glycanes en chaîne isolées, nous avons utilisé l'*Umbrella Visualization* en 2D (version développée par Alexandre Guillot, présentée dans la section 3.2.1) sur 30 000 clichés de trajectoires

pour les glycanes tétra-antennés et 10 000 clichés pour les glycanes bisectants (1 cliché toutes les 50 ps). Les positions des centres de masse des galactoses ont été utilisés pour la projection.

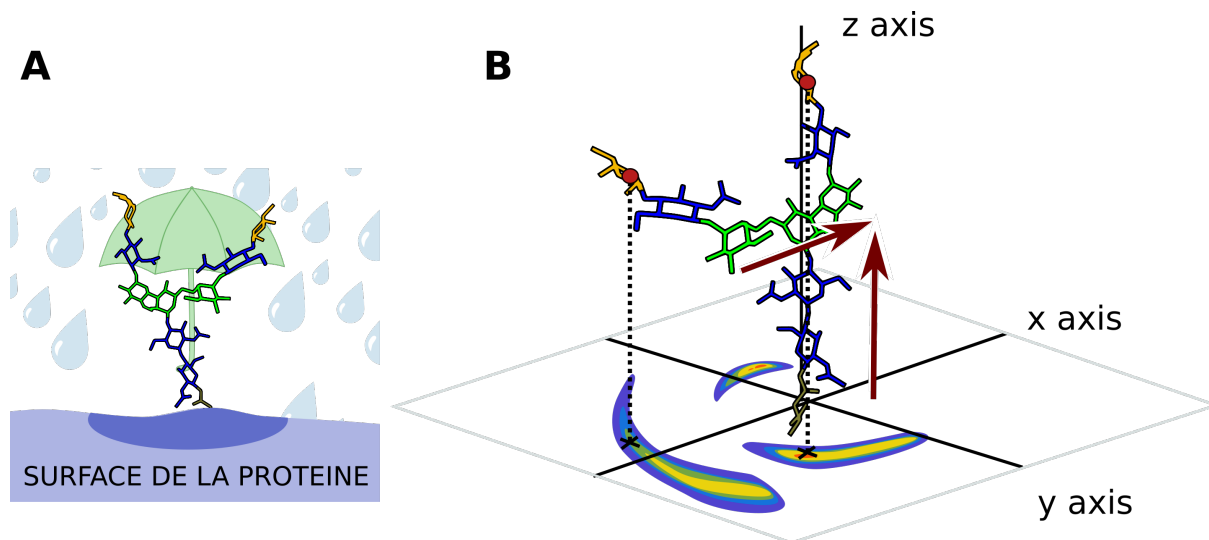
Pour le développement de l'*Umbrella Visualization* et sa mise en application, nous avons utilisé un ordinateur doté d'un processeur Intel Xeon W-2145 3.7 GHz, 64 Go de mémoire vive DDR4 et une carte graphique Nvidia GeForce GTX 1080Ti. Ces analyses ont été effectuées sur 12 500 clichés issues de trajectoires de dynamique moléculaire du récepteur à l'insuline glycosylé de 250 ns. Ces trajectoires ont été générées pendant les travaux de thèse d'Alexandre Guillot [8]. Le champ de forces utilisé est OPLS-AA et huit trajectoires ont été calculées avec des glycanes bi-antennés et tri antennés : avec et sans acides sialiques et / ou avec et sans fucose.

3.2 Méthodes de visualisation appliquées aux sucres

3.2.1 *Umbrella Visualization* : une approche 2D

L'*Umbrella Visualization* (ou visualisation en parapluie) est une méthode développée précédemment au sein du laboratoire et dédiée à l'étude des glycanes en chaîne isolée. Cette méthode a permis de mettre en avant les variations de flexibilité de chaînes de glycosylation en fonction de leur état de sialylation [17]. Le principe de cette méthode est simple : le cœur du glycanes est aligné le long de l'axe z d'un repère orthonormé. Puis le glycanes subit une rotation autour de l'axe z jusqu'à ce que l'axe défini par les deux beta mannose et l'alpha mannose soit colinéaire à l'axe x. Ensuite, les extrémités des positions des branches du glycanes sont projetées sur le plan (x,y) (figure 3.13). Cette étape est répétée pour chaque structure extraite d'une simulation de dynamique moléculaire, permettant ainsi d'obtenir un graphique en deux dimensions représentant la flexibilité relative des glycanes.

La première version de cette méthode se présentait sous un ensemble de programmes écrits en Fortran90. L'utilisation de ces programmes nécessitait au préalable la détermination de vecteurs à l'aide du module de distances de Gromacs. Ces vecteurs sont définis par rapport aux sucres composant le glycanes, tels que présentés dans la figure 3.14. Le vecteur définissant le noyau du glycanes est utilisé pour aligner le glycanes sur l'axe de référence z. Les vecteurs définissant les branches sont utilisés pour projeter les positions des extrémités de ces dernières sur le plan (x,y) pour chaque cliché extrait de la simulation de dynamique moléculaire. Finalement, un graphique de densité est calculé et tracé, nous permettant ainsi d'évaluer la flexibilité du glycanes au cours de la trajectoire et de corrélérer ces résultats aux familles conformationnelles majoritaires.


 FIGURE 3.13 – Principe de l'*Umbrella Visualization*.

A : On considère que le glycane protège ou masque une partie de la surface de la protéine, comme le fait un parapluie. C'est cette surface couverte que nous cherchons à évaluer avec l'*Umbrella Visualization*. **B** : Les deux vecteurs rouges correspondent aux vecteurs utilisés pour aligner le glycane le long de l'axe z, puis les positions des extrémités des branches (points rouges) sont projetées sur le plan (x, y) pour chaque structure extraite de la trajectoire de simulation de dynamique moléculaire. Le vecteur rouge horizontal, défini par l'axe traversant les mannoses en vert, est colinéaire à l'axe x. Le vecteur vertical, défini par l'axe traversant les N-acetylglucosamines et le mannose central, est aligné sur l'axe z.

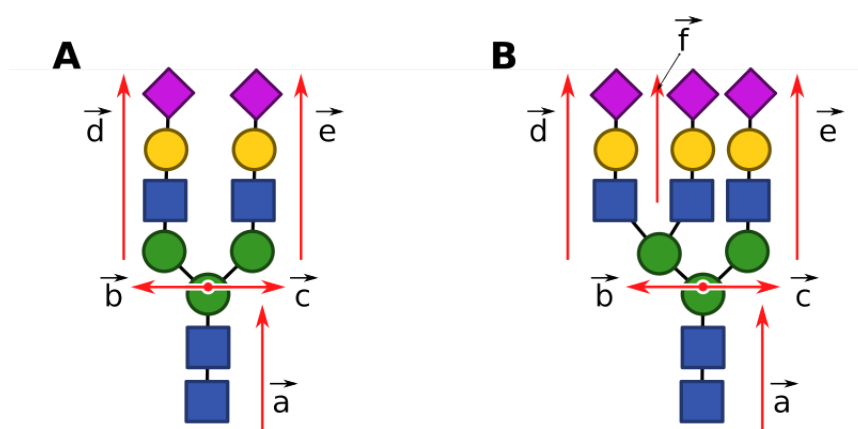


FIGURE 3.14 – Vecteurs entre les centres de masse des glycans. Les vecteurs $\vec{a}$, $\vec{b}$ et $\vec{c}$ définissent le cœur du glycane et sont utilisés pour aligner le glycane sur un axe z et orienter le plan (x, y). **A** : Vecteurs définis pour un glycane bi-antenné. Les vecteurs $\vec{d}$ et $\vec{e}$ permettent de projeter la position de l'extrémité des branches sur le plan (x,y). **B** : Vecteurs définis pour un glycane tri-antenné. Le vecteur $\vec{f}$ correspond à la branche supplémentaire. Les vecteurs $\vec{d}$, $\vec{e}$ et $\vec{f}$ permettent de projeter la position de l'extrémité des branches sur le plan (x,y).

Bien que cette méthode ait démontré son efficacité, un certain nombre de problèmes ont été soulevés à son utilisation. Tout d’abord, les programmes développés étaient adaptés à l’étude de glycanes bi- et tri-antennés et nécessitaient d’importantes modifications pour être adaptés à d’autres structures. Le nombre d’étapes et de scripts nécessaires à l’obtention des résultats réduisait l’accessibilité de la méthode.

L’inconvénient principal est le fait que cette méthode donne un résultat en deux dimensions, sous la forme d’un graphique de densité ne prenant en compte que les extrémités du glycanes. En réalité, les groupements hydroxyles présents sur toute la longueur de la chaîne de saccharides peuvent potentiellement influencer la dynamique du glycanes et / ou interagir avec le solvant et / ou la surface de la protéine. Cette surface est elle-même un objet en trois dimensions, peu comparable à une surface plane. De plus, aligner le glycanes le long d’un axe z avant de calculer les positions des branches permet d’évaluer la flexibilité relative de ces branches mais empêche également de prendre en compte des éventuelles inclinaisons du glycanes par rapport à la surface de la protéine. Ainsi, malgré les informations pertinentes apportées par cette méthode, cette représentation ne donne pas une idée suffisante de l’impact structural du glycanes sur la surface de la protéine.

Afin de prendre en compte tous les mouvements du glycanes ainsi que la topologie de la protéine, nous avons donc décidé d’intégrer et d’adapter l’*Umbrella Visualization* dans un logiciel de visualisation moléculaire. Pour ce faire, nous avons travaillé avec le logiciel UnityMol et la plateforme Unity. L’objectif est de proposer un outil accessible permettant de visualiser automatiquement la surface de la protéine couverte par un glycanes au cours d’une trajectoire de dynamique moléculaire. Le mode de visualisation choisi devra également permettre de garder une information statistique, en mettant en avant les zones les plus souvent couvertes par le glycanes au cours de la trajectoire.

3.2.2 UnityMol : un logiciel pour une implémentation en 3D

Plutôt que de programmer entièrement un nouveau logiciel de visualisation moléculaire, nous avons fait le choix d’implémenter l’*Umbrella Visualization* dans un logiciel déjà existant. La plupart de ces logiciels est *open source* et permet aux utilisateurs d’ajouter des extensions personnalisées, offrant ainsi plusieurs choix. Nous avons sélectionné des logiciels déjà utilisés dans la communauté scientifique, régulièrement mis à jour et bénéficiant du soutien des développeurs et de la communauté d’utilisateurs afin de s’assurer que notre méthode puisse évoluer dans le temps et être accessible sur différentes plateformes. La table 3.7 résume les principaux logiciels auxquels nous nous sommes intéressés.

Logiciel	Langage	Gestion des sucres native	Open Source	Site web (lien)
VMD	C / C++	Non (plug-in)	Oui	www.ks.uiuc.edu/Research/vmd
PyMol	C / C++ et Python	Non (plug-in)	Non	www.pymol.org/2
UCSF ChimeraX	C++ et Python	Non	Oui	www.rbvi.ucsf.edu/chimerax
UnityMol	C#	Sélection, affichage	Oui	www.baaden.ibpc.fr/umol

TABLE 3.7 – **Logiciels considérés pour l’implémentation de l’*Umbrella Visualization*.** Les caractéristiques de langages de programmation et les options de gestions des sucres qui ont guidé notre réflexion sont listés.

Notre sujet d’étude étant les sucres, un intérêt particulier a été porté sur les fonctionnalités spécifiques aux sucres mises à notre disposition par ces logiciels. ChimeraX [87] et PyMol [86] ne proposent pas de fonctionnalités dédiées à l’étude des sucres dans leurs fonctions par défaut. VMD permet la sélection de sucres grâce au mot clé "sugar" mais aucun mode de visualisation spécifique n’est proposé dans la version de base. Dans la section 2.3.4, nous avons cependant présenté des plug-ins adaptés à la visualisation des sucres pour PyMol et VMD. Cependant, nous souhaiterions intégrer l’*Umbrella Visualization* à un logiciel de façon native, sans avoir à télécharger de plug-in au préalable.

UnityMol est le logiciel qui répond à l’ensemble des critères que nous nous sommes fixés : il présente des fonctionnalités adaptées à l’étude des protéines glycosylées sans installation de plug-in supplémentaire, et permet l’intégration directe de l’*Umbrella Visualization*. De plus, utiliser UnityMol nous permet de profiter de la plateforme Unity ainsi que des outils et ressources associés.

UnityMol est un logiciel disponible pour les plateformes Windows, Mac et Linux [78]. Codé en C# à l’aide de la plateforme Unity (présentée ci-après), UnityMol intègre les modes de représentation et de visualisation classiques ainsi qu’une représentation en hyperballs [79] et peut lire des trajectoires issues de simulations de dynamique moléculaire. Ce logiciel a connu plusieurs versions, dont la version SweetUnityMol qui intègre également des modes de visualisation dédiés

aux chaînes de saccharides. Les spécificités structurales liées aux saccharides comme les possibilités de branchements et les atomes d'oxygène intracycliques peuvent être spécifiquement mis en évidence. Il est également possible de colorer les cycles des sucres en fonction de leur nature, en respectant les couleurs correspondant à la nomenclature SNFG.

Nous avons donc choisi d'intégrer l'*Umbrella Visualization* dans Unitymol afin de profiter des avantages présentés par ce logiciel et la plateforme Unity dont nous présentons les avantages ci-après.

3.2.3 Moteur de jeux vidéos Unity

UnityMol utilise le moteur Unity, qui est une plateforme de développement complète qui propose de nombreux outils permettant très aisément l'accès aux dernières technologies. Cette plateforme possède également une très vaste communauté d'utilisateurs, garantissant l'accès à de nombreuses ressources en ligne et de nombreux tutoriels. Unity permet aussi aux logiciels développés d'être accessibles à la fois sur Windows, Linux, Mac et Android en exportant le fichier exécutable approprié durant l'étape de compilation. Les logiciels ainsi créés sont donc accessibles à tous types d'utilisateurs sans étape de développement supplémentaire.

L'interface graphique Unity intègre également une fenêtre simulant l'exécution du logiciel, appelée fenêtre de jeu ou *Game view*, sans avoir à le compiler au préalable. Ceci permet de tester rapidement et efficacement les modifications et ajustements apportés au code. Unity est également reconnu pour sa facilité d'utilisation pour les nouveaux utilisateurs, partiellement grâce à son interface mais également de par sa versatilité. En effet, tous les paramètres et composants disponibles depuis l'interface graphique sont accessibles *via* des scripts C# ou Javascript.

Ainsi, le logiciel Unitymol développé sur cette plateforme est facilement accessible et modifiable grâce aux avantages de la plateforme Unity et au travail des créateurs de ce projet.

Chapitre 4

Caractérisation du comportement dynamique de chaînes sucrées : de la chaîne isolée aux glycosaminoglycanes

4.1 Constitution d'une librairie de glycanes

Le projet Dogme visait à l'origine à développer un module de greffage de glycanes ou de glycosaminoglycanes implémenté au logiciel de modélisation Gromacs, dans un esprit similaire aux fonctionnalités proposées par Glycam-web. Pour que l'utilisation puisse au préalable disposer d'un jeu de chaînes réalistes qui soit représentatif de divers contextes, nous nous sommes attelés à la construction d'une librairie de glycanes répertoriant des structures identifiées dans la bibliographie et caractéristiques de pathologies liées à l'âge (maladies cardio-vasculaires, diabète de type II, cancers ...). La librairie ainsi élaborée est de taille modeste. Cependant, elle s'inscrit dans le cadre des travaux menés au sein de notre unité de recherche. Quelques exemples de structures intégrées à cette librairie sont présentés dans la figure 4.1.

Le panel A est composé de structures de glycanes courantes : un bi-antenné bisectant et un glycanes portant le motif LacNac (Lactose et N-acétylglucosamine). Ce motif peut être prolongé plusieurs fois pour former de longues chaînes [49]. Les travaux de Han et. al [98] sur le glycome de la protéine membranaire CD44 et ceux de Kathri et. al [99] sur l'hémagglutinine du virus de la grippe ont montré que l'accessibilité réduite de certains sites de glycosylation entraîne une hausse de la proportion de glycanes immatures (composés essentiellement de mannoses) comme celle du panel B.

Otto et. Al [100] ont étudié l'isoforme soluble de la protéine ICAM-1 (*Intercellular Adhesion*

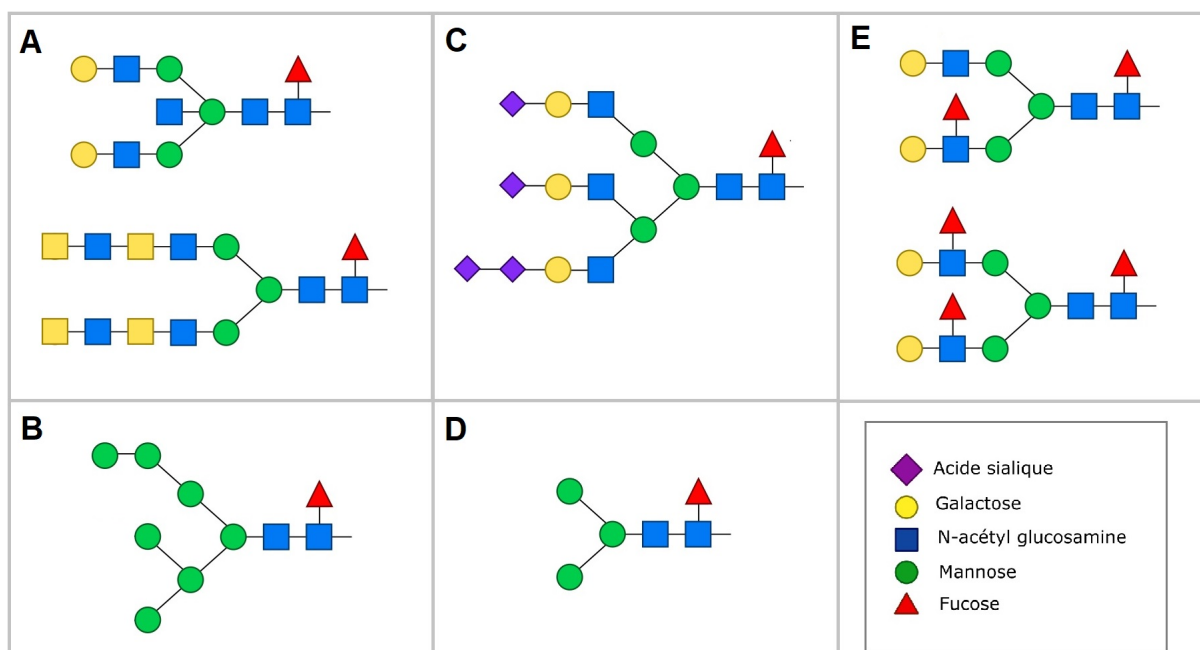


FIGURE 4.1 – **Exemples de glycanes intégrés à la librairie.** Les structures intégrées ont été identifiées dans la bibliographie comme structures communes ou comme marqueurs de pathologies liées à l'âge.

Molecule-1, une protéine transmembranaire impliquée dans l'adhésion cellulaire) et montré que les sites de glycosylation de cette protéine sont principalement occupés par des glycanes bi- et tri-antennés sialylés. Environ 4% des structures comportaient un nombre plus important d'acides sialiques que d'antennes, formant un motif composé de deux acides sialiques (figure 4.1 C).

Le glycanes de type paucimannose illustré dans le panel D est, en temps normal, très peu présent sur les protéines matures des cellules de vertébrés mais les travaux de Lee et al. [52] ont montré que certaines lignées cancéreuses présentaient une quantité accrue de ces glycanes. Enfin, les travaux de Sethi et al. [101] ont, dans le cadre de l'étude du glycôme de lignées cellulaires cancéreuses (cancer du côlon) permis de mettre en évidence la présence de glycanes fucosylés sur leurs extrémités non-réductrices rappelant les déterminants de Lewis des O-glycosylations. Leurs résultats montrent également une expression différentielle de ces glycanes en fonction de la souche cellulaire.

A travers la sélection de certains glycanes constituant cette librairie, nous avons pour objectif d'approfondir l'étude de l'impact des acides sialiques sur la dynamique et la flexibilité des glycanes. En parallèle des développements de l'*Umbrella Visualization* présentés dans le chapitre suivant, nous avons ainsi mis en place diverses simulations de dynamique moléculaire dont les résultats sont présentés dans les deux sections suivantes.

4.2 Analyses d'expériences de dynamique moléculaire portant sur des glycanes en chaînes isolées

Les travaux effectués par Alexandre Guillot relatifs à l'influence des glycanes bi-antennés et tri-antennés en présence ou non d'acides sialiques aux extrémités ont mis en évidence l'impact de ces derniers sur les conformations préférentielles et la flexibilité des glycanes [17]. Ces travaux ont également montré l'impact variable qu'ont ces résidus en fonction du nombre de branches des glycanes : là où les deux branches des glycanes bi-antennés sont affectées par la présence des acides sialiques terminaux, seule une branche des glycanes tri-antennés montre un comportement différent.

Dans l'optique d'approfondir ces travaux, nous nous sommes intéressés à deux autres types de structures de glycanes : des glycanes bi-antennés bisectants, comparables aux glycanes bi-antennés, et des glycanes tétra-antennés.

4.2.1 Caractérisation de glycanes bi-antennés bisectants

Dans un premier temps, nous nous sommes intéressés à des glycanes portant une N-acétyl glucosamine bisectante entre les deux branches du glycanes (figure 3.9). Ces glycanes sont très étudiés en raison de leur rôle dans différentes fonctions biologiques, notamment dans la progression tumorale [102–104]. Considérant le rôle important de ces glycanes et des acides sialiques dans la régulation de différentes fonctions cellulaires, nous avons ainsi décidé d'étudier les caractéristiques dynamiques et structurales de glycanes bisectants avec et sans acides sialiques. Élucider leurs principales caractéristiques en terme de structure et de dynamique pourrait apporter de nouvelles perspectives sur les fonctions associées à ces glycanes.

Ces premiers travaux avaient également pour objectif, dans un premier temps, de se familiariser avec les outils développés, en particulier l'*Umbrella Visualization*, et ont servi à la fois de point de comparaison avec les travaux sur les glycanes bi-antennés et de simulations tests pour les modifications apportées à l'*Umbrella Visualization*. Des trajectoires de 500 ns ont ainsi été générées avec Gromacs et le champ de forces Charmm36 sur les deux structures étudiées.

Dans un premier temps, ces trajectoires ont été analysées à l'aide de méthodes de clustering sur 5 000 structures extraites de la simulation. Alors que les résultats obtenus sur les glycanes bi-antennés montraient une préférence pour la conformation *Broken Wing* en présence d'acides sialiques et une transition vers des conformations étendues (*Bird*) en absence d'acides sialiques. Les résultats du clustering sur ces trajectoires montrent tous les deux une conformation principale

Bird. Ce premier cluster contient 58% des structures pour le glycané bisectant sans acides sialiques et 73% pour le glycané bi-antenné bisectant avec acides sialiques. Dans les deux cas, le bras α 1-3 semble plus contraint à proximité de la N-acétyl glucosamine bisectante sur les deux structures (figure 4.2).

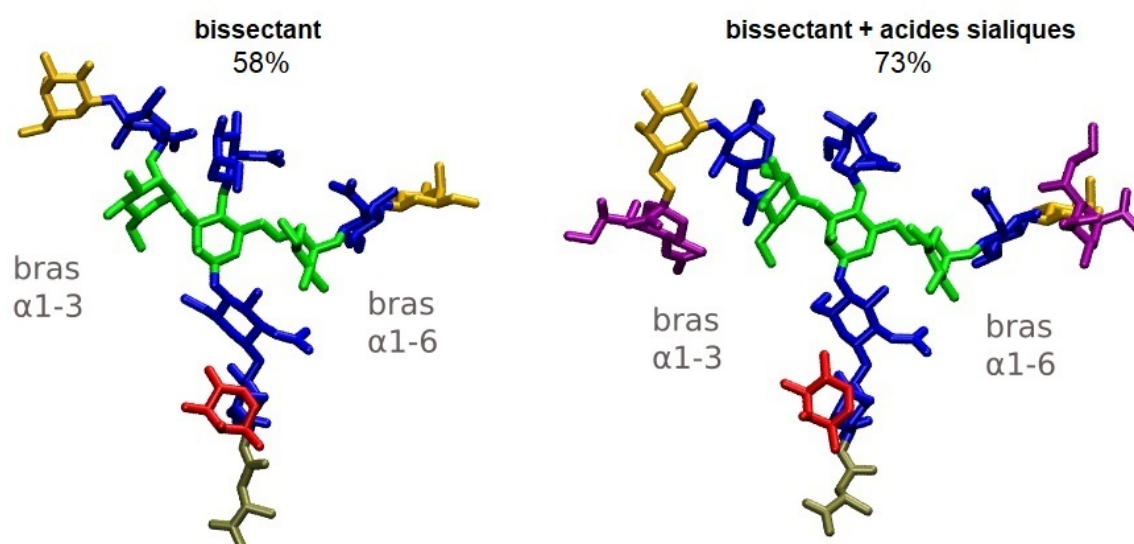


FIGURE 4.2 – **Résultat du clustering sur les glycanes bisectants.** La conformation principale adoptée pour les deux glycanes rappelle la conformation *Bird*, avec les deux branches étendues vers le solvant. Le bras α 1-3 semble contraint par la N-acétyl glucosamine bisectante dans les deux cas.

La projection des positions des centres de masse des galactoses avec la méthode 2D de l'*Umbrella Visualization* permet d'obtenir une appréciation de la zone explorée par ces glycanes durant la trajectoire. Les graphiques (figure 4.3) ainsi obtenus permettent de séparer la contribution des différentes branches. Dans le cas des glycanes bisectants, il semble que la zone explorée par la branche α 1-3 est plus compacte et restreinte autour d'une même zone (quart supérieur gauche à proximité des axes x et y) alors que la branche α 1-6 explore une surface présentant un profil plus étendu (quart supérieur droit du graphe). On remarque par ailleurs que la présence des acides sialiques réduit la zone explorée par le bras α 1-3 (ce phénomène est mis en évidence à l'aide des cadres rouges sur la figure 4.3) et modifie également le profil de la zone explorée par le bras α 1-6 : à distance du cœur du glycané, l'amplitude de la zone explorée est étendue aux valeurs de y négatives tandis que la zone explorée à proximité du cœur disparaît. Les deux observations sont mises en exergue à l'aide des cadres rouges sur la figure 4.3.

Ces glycanes ont été largement étudiés [103, 105] et une étude précédente a montré que la N-acétylglucosamine centrale restreint les conformations adoptées par le bras α 1-3 [105]. Ces premiers résultats obtenus avec l'*Umbrella Visualization* 2D vont dans ce sens puisque cette ap-

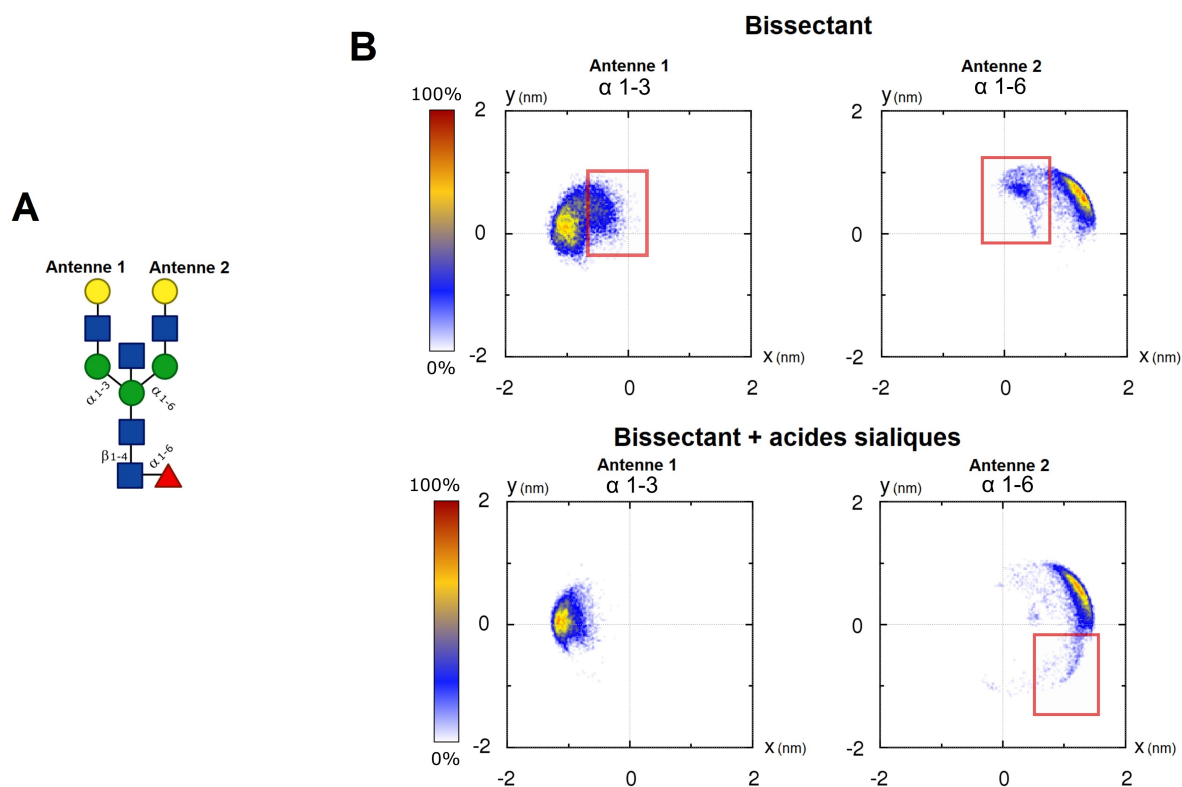


FIGURE 4.3 – **Résultat de l'Umbrella Visualization sur les glycanes bisectants.** **A** : Représentation schématique du glycan bisectant, l'antenne 1 correspond au bras $\alpha 1-3$ et l'antenne 2 au bras $\alpha 1-6$. **B** : Résultat de l'Umbrella Visualization pour les deux types de glycanes bisectants. La zone explorée par l'antenne 1 (bras $\alpha 1-3$) est réduite en présence d'acides sialiques tandis que l'antenne 2 (bras $\alpha 1-6$) explore une région plus large et plus éloignée du cœur du glycan. Les mannoses sont colorés en vert, les galactoses en jaune, les N-acétylglucosamines en bleu, le fucose en rouge et les acides sialiques en violet, selon la nomenclature SNFG. Les populations respectives ainsi que les bras $\alpha 1-3$ et $\alpha 1-6$ sont indiqués par les labels. Les cadres rouges indiquent les zones impactées. Le glycan est représenté selon la nomenclature SNFG. La couleur des points des graphiques indique le taux d'occupation de la position pendant la trajectoire.

proche montre que cette branche semble explorer un espace plus réduit. Bien que les résultats de clustering donnent des conformations principales similaires, l'*Umbrella Visualization* suggère cependant que la flexibilité de ces glycanes est affectée par la présence des acides sialiques, puisque le bras $\alpha 1-3$ perd de sa flexibilité tandis que le bras $\alpha 1-6$ explore différentes zones éloignées du glycanes.

Nous avons ensuite poursuivi ces travaux en nous intéressant à des chaînes de glycosylations plus atypiques, essentiellement retrouvées dans des contextes pathologiques tels que des cancers ou encore la maladie d'Alzheimer [102, 106] : des structures de glycanes tétra-antennés.

4.2.2 Caractérisation de glycanes tétra-antennés

Les glycanes complexes tétra-antennés sont très peu présents dans les cellules saines et sont considérés comme des marqueurs de pathologies. Approfondir la connaissance de leurs propriétés dynamiques et structurales pourrait être une première étape pour comprendre le rôle des glycanes dans un contexte pathologique et les modulations induites par ces modifications (perturbation des interactions protéine/protéine par exemple). Nous avons ainsi choisis trois structures de glycanes complexes tétra-antennés : la première structure sans acides sialiques et les deux suivantes avec deux types d'acides sialiques différant par la nature de leur liaison au galactose : $\alpha 2-3$ ou $\alpha 2-6$.

Ces deux types d'acides sialiques existent tous les deux naturellement au sein du domaine du vivant, les premiers étant présents en grande majorité dans les cellules. Cependant, il a été montré que le ratio entre les liaisons $\alpha 2-3$ et $\alpha 2-6$ peut varier dans certains contextes pathologiques comme les cancers [52]. Nous avons ainsi généré, pour chaque structure, trois trajectoires de 500 ns avec Gromacs et le champ de forces Charmm36.

Dans un premier temps, nous avons analysé l'ensemble de ces trajectoires avec l'*Umbrella Visualization*, sur un total de 30 000 structures issues des 1 500 ns de trajectoire. Les branches 1 et 2 correspondent au bras portant la liaison mannose $\alpha 1-3$ tandis que les branches 3 et 4 sont sur le bras portant la liaison mannose $\alpha 1-6$ (figure 4.4 A).

Sur les quatre branches, on observe des profils très similaires entre le glycanes tétra-antenné sans acides sialiques et le glycanes tétra-antenné avec des acides sialiques en $\alpha 2-6$ (figure 4.4 B et D). Seul le glycanes portant les acides sialiques en $\alpha 2-3$ présente un profil différent sur les branches 3 et 4 : les deux profils des branches pour cette structure présentent des zones explorées supplémentaires, absentes sur les deux autres trajectoires (figure 4.4 C, cadres rouges).

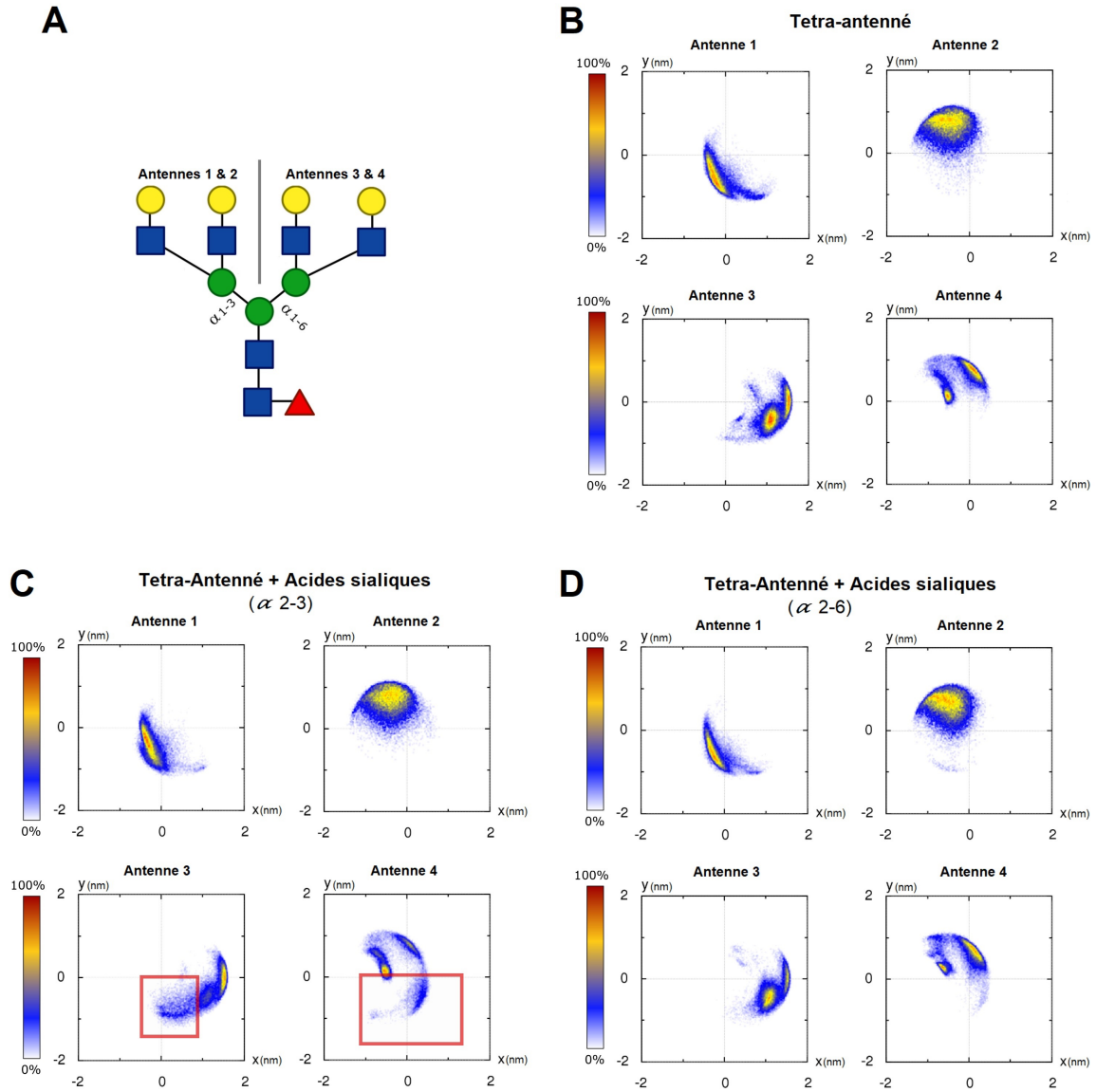


FIGURE 4.4 – Résultat de l'*Umbrella Visualization* sur les glycanes tétra-antennés, sur 30 000 structures. **A** : Représentation schématique d'un glycan tétra-antenné, les bras $\alpha 1-3$ et $\alpha 1-6$ sont identifiés. **B**, **C** et **D** : Résultats de l'*Umbrella Visualization* pour les 4 antennes des 3 structures de glycanes. Les nouvelles zones explorées par la trajectoire avec les acides sialiques liés en $\alpha 2-3$ sont encadrées en rouge. La couleur des points des graphiques indique le taux d'occupation de la position pendant la trajectoire.

Puisque cette différence n'apparaît que sur les branches 3 et 4 du glycanes, nous nous sommes intéressés à la liaison mannose α 1-6 : cette liaison comporte en effet un atome de plus que la liaison α 1-3 (figure 1.4), lui conférant un degré de liberté en plus. Nous avons donc mesuré la valeur des trois angles dièdres de la liaison α 1-6 au cours du temps, sur la totalité des trois trajectoires. Ces dernières ont été réalisées de manière indépendantes avec un point de départ différent. Sur la figure 4.5, les données de la trajectoire 1 occupent l'axe de 0 à 500 ns, la trajectoire 2 de 500 à 1000 ns et la trajectoire 3 de 1000 à 1500 ns.

Alors qu'aucune différence n'est observée pour les angles Phi et Psi pour les trois systèmes, l'angle Omega de la trajectoire 1 du glycanes tétra-antenné avec acides sialiques en α 2-3 montre un décalage soudain, à partir de environ 125 ns et jusqu'à la fin de la trajectoire. Cet angle passe d'une valeur centrée aux alentours des 50° vers une valeur centrée autour de -60° . Le glycanes semble ici adopter une autre conformation stable. Dans les trajectoires 2 et 3, ce phénomène ne se reproduit pas. On peut également remarquer que, sur les trajectoires des glycanes sans acides sialiques et avec acides sialiques liés en α 2-6, l'angle Omega occupe transitoirement cette position. La figure 4.6 montre la distribution des valeurs des angles dièdres de la liaison α 1-6 sur la totalité des trajectoires. La variation observée ici est bien visible sur l'angle Omega (cadres rouges).

Pour évaluer l'impact de la conformation de l'angle Omega observée majoritairement lors de la trajectoire 1 du glycanes portant les acides sialiques liés en α 2-3, la méthode de l'*Umbrella Visualization* 2D est appliquée sur un jeu de structures remanié. La trajectoire 1 est mise de côté et les 20 000 clichés des trajectoires 2 et 3 sont alors utilisés pour générer la figure 4.7. Afin d'avoir un résultat statistiquement comparable en terme de nombre de clichés pris en compte, nous avons refait ce calcul sur deux trajectoires du glycanes sans acides sialiques. Nous avons fait ce choix plutôt que de supprimer 10 000 clichés aléatoirement car le résultat obtenu sur les deux trajectoires sélectionnées est presque identique à celui obtenu sur les trois trajectoires. Ainsi, ceci nous permet d'avoir des résultats comparables entre les deux types de glycanes d'un point de vue statistique (même nombre de structures prises en compte) tout en conservant un résultat pour la trajectoire "témoin" (sans acides sialiques) similaire à celui obtenu sur 30 000 structures.

Cette seconde analyse, présentée sur la figure 4.7, montre un profil de surface couverte similaire pour les deux types de trajectoires. Ceci confirme donc que l'information contenue dans la trajectoire 1 est responsable des surfaces couvertes pour les antennes 3 et 4 (cadres rouges sur la figure 4.7), en lien avec la valeur canonique de l'angle dièdre Omega mise en évidence au travers de l'analyse dynamique.

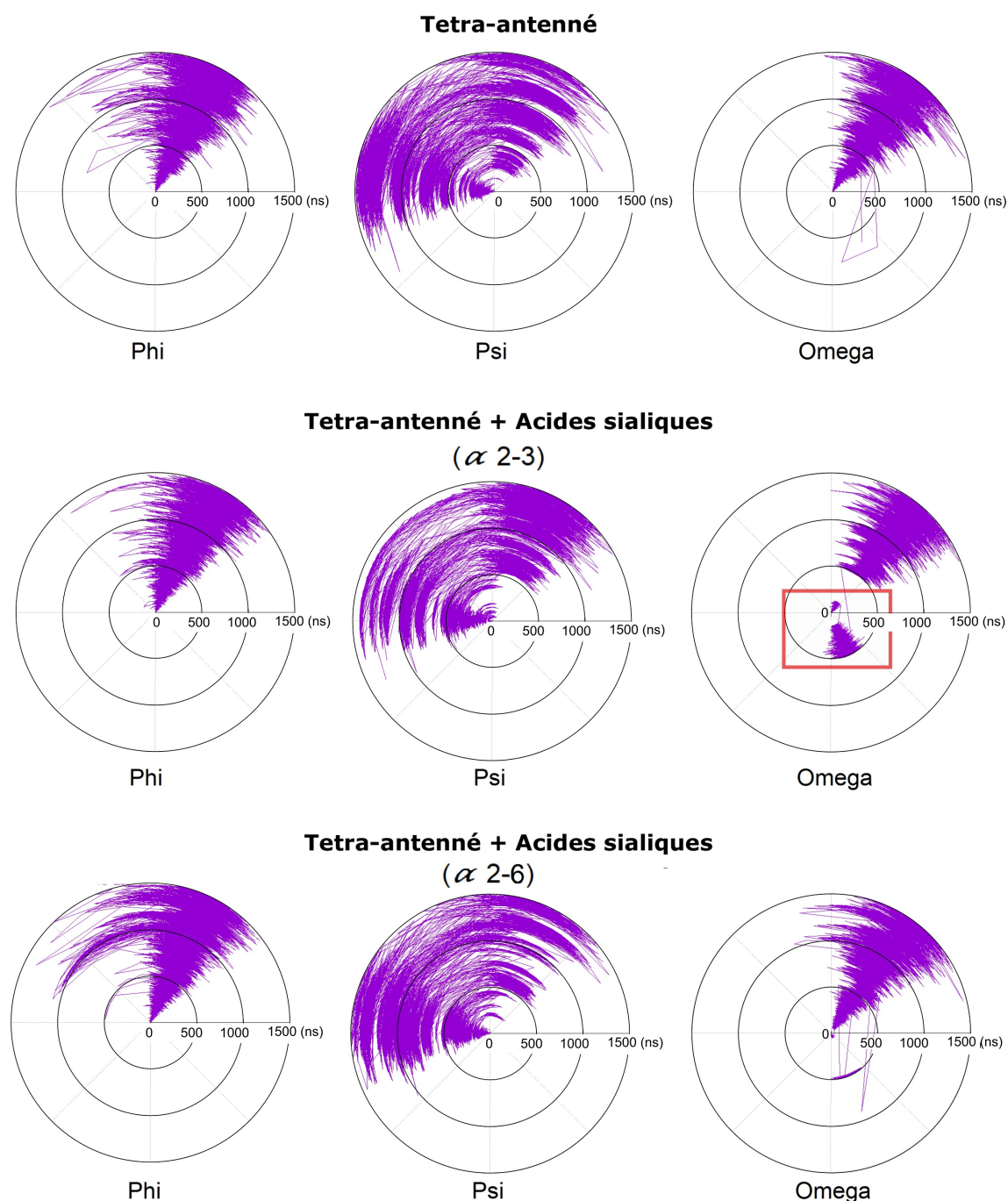


FIGURE 4.5 – Mesure des valeurs d’angles au cours du temps pour la liaison mannose α 1-6. Pour les trois types de glycanes tétra-antennés, on observe des valeurs d’angle similaires. Seul l’angle Omega pour la trajectoire 1 du glycanes tétra-antenné avec acides sialiques liés en α 2-3 montre une différence : un décalage après environ 125 ns de trajectoire a lieu au niveau de l’angle Omega (encadré en rouge). Le rayon du cercle correspond à la durée de la trajectoire et les valeurs d’angle correspondantes sont reportées sur le cercle. Les trajectoires indépendantes sont séparées par les cercles noirs. La trajectoire 1 correspond aux premières 500 ns, la trajectoire 2 est entre 500 et 1000 ns et la trajectoire 3 est entre 1000 et 1500 ns.

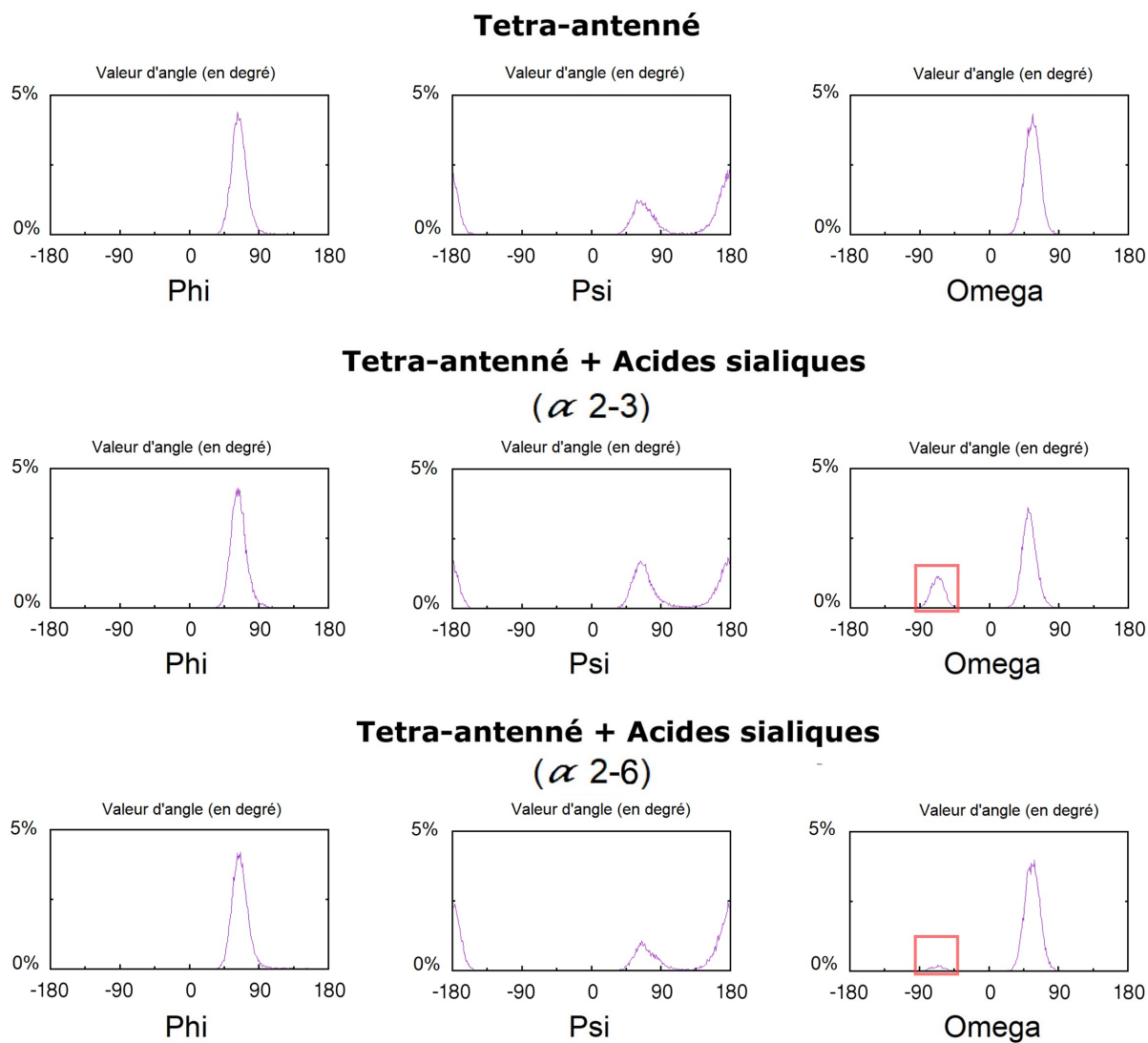


FIGURE 4.6 – Distribution des valeurs d'angles adoptées par la liaison mannose α 1-6. Le pic secondaire autour des -60° pour l'angle omega est bien visible pour les deux trajectoires en présence d'acides sialiques (cadres rouges).

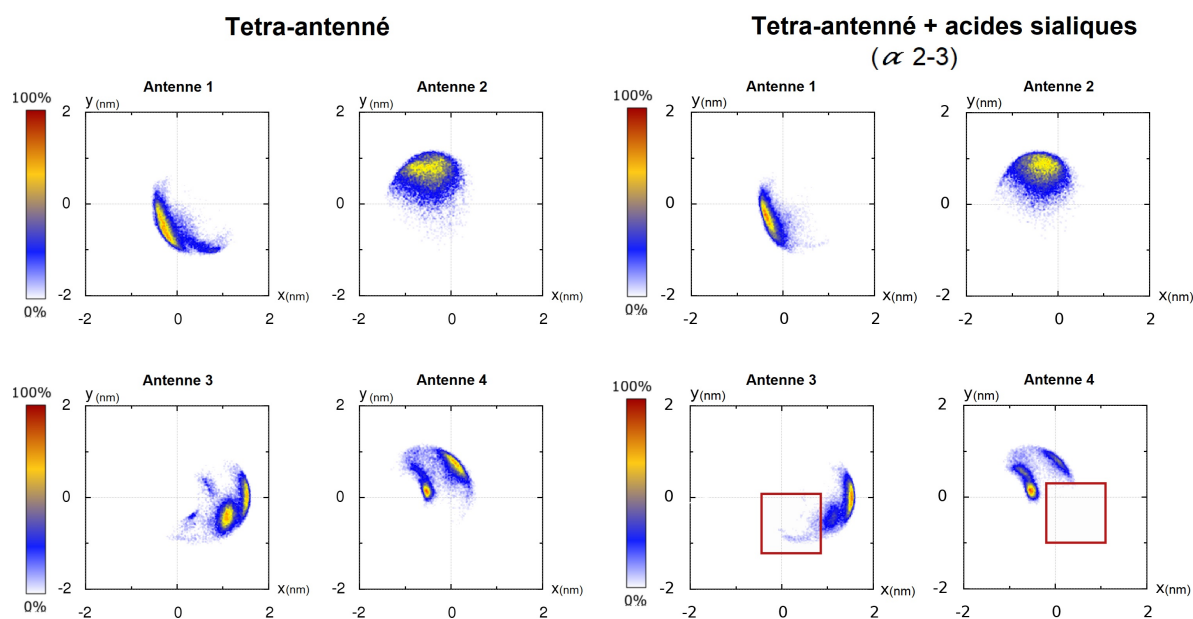


FIGURE 4.7 – Résultat de l'*Umbrella Visualization* sur les glycanes tétra-antennés sans acides sialiques et avec acides sialiques liés en $\alpha 2-3$, sur 20 000 structures. Les cadres rouges indiquent l'emplacement des taches disparues avec la suppression de la trajectoire 1 du glycanes tétra-antenné avec acides sialiques liés en $\alpha 2-3$. La couleur des points des graphiques indique le taux d'occupation de la position pendant la trajectoire.

Afin d'appuyer ces premiers résultats par des informations de structure, nous avons utilisé une approche de clustering menée conformément à la méthode employée avec les glycanes bi-sélectants. Les conformations majoritaires ont été déterminées sur un total de 15 000 structures issues des trajectoires (5 000 structures par trajectoire).

Pour les trois types de glycanes, les conformations majoritaires sont en effet très similaires et montrent une conformation où les branches sont éloignées les unes des autres, avec une branche repliée contre le cœur du glycanes (figure 4.8). Le bras $\alpha 1-3$ tend à rester étendu vers le solvant tandis que le bras $\alpha 1-6$ replie l'une de ses branches contre le cœur du glycanes. On observe également que les acides sialiques liés en $\alpha 2-6$ sont repliés contre les branches du glycanes, sous la N-acétylglucosamine, contrairement aux acides sialiques liés en $\alpha 2-3$. Ces derniers sont ainsi plus exposés au solvant, en particulier sur le bras $\alpha 1-3$.

Afin de visualiser la variation d'angle de 50° vers -60° observée sur la trajectoire du glycanes tétra-antenné avec acides sialiques en $\alpha 2-3$ (figure 4.4), le cluster majoritaire obtenu sur la trajectoire 1 de ce glycanes a été extrait à partir des 5000 structures correspondantes. La figure 4.9 permet d'observer la variation au niveau du bras $\alpha 1-6$: la figure 4.9 A présente le cluster obtenu sur cette trajectoire, la figure 4.9 B présente le cluster majoritaire sur la totalité des trois trajec-

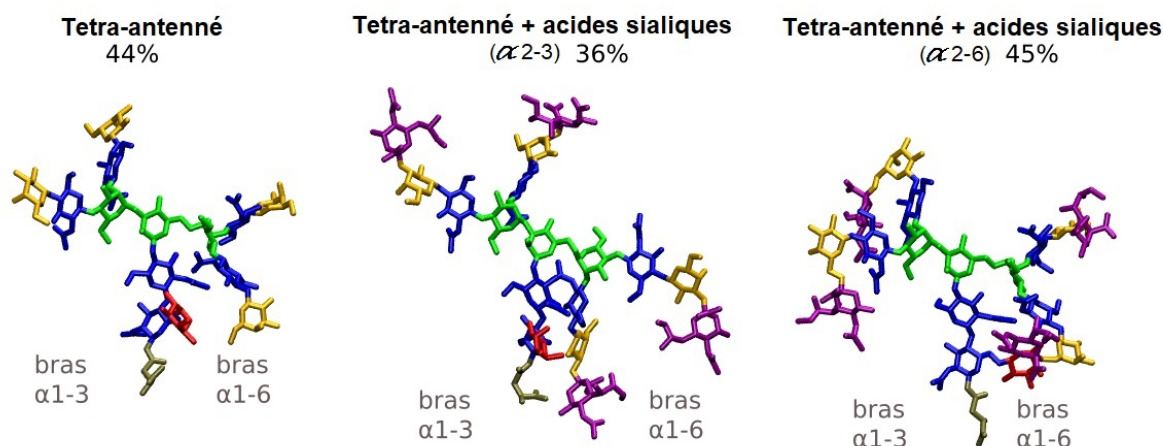


FIGURE 4.8 – **Conformation majoritaire issue du clustering pour les trois types de glycanes.** Les conformations majoritaires des trois glycanes sont très similaires avec les deux branches du bras $\alpha 1-3$ étendues vers le solvant. Le bras $\alpha 1-6$ tend à avoir une conformation qui rapproche une des branches du cœur du glycanes. Les conformations des deux types d'acides sialiques diffèrent : les acides sialiques liés en $\alpha 2-3$ sont étendus vers le solvant tandis que les acides sialiques liés en $\alpha 2-6$ sont repliés le long des branches. Les N-acétylglucosamines sont en bleu, les mannoses en vert, les galactoses en jaune et les acides sialiques en violet, selon la nomenclature SNFG.

toires. Sur cette dernière, l'une des branches du bras $\alpha 1-6$ est plus proche du cœur du glycanes et du fucose. La figure 4.9 C superpose les structures des deux panels précédents et montre les orientations totalement différentes des branches du bras $\alpha 1-6$. Il est possible que des interactions avec le fucose aient lieu, stabilisant cette conformation. En effet, au cours d'une étude sur la conformation de ces glycanes, Harbison et al. [107] ont observé que ce monosaccharide pouvait être impliqué dans des liaisons hydrogènes avec les branches de glycanes bi-antennés. On pourrait supposer ici que la présence du fucose stabilise sous certaines conditions, la position du bras $\alpha 1-6$ dans une conformation alternative.

La taille imposante des glycanes tétra-antennés implique une surface couverte plus importante que celle associée à des glycanes plus communs comme les structures bi-antennées. Ici, les résultats du clustering suggèrent que ces glycanes adoptent principalement des structures étendues et donc capables de couvrir une large surface. On peut supposer que le désordre de ces chaînes complexes et flexibles tend à réduire l'impact des acides sialiques sur les chaînes tétra-antennées. Le grand nombre de groupements polaires présents complexifie les interactions et déstabilise la structure globale. Il est ainsi possible que les trajectoires obtenues ici ne représentent pas de manière exhaustive l'espace exploré par ces glycanes.

De plus, la version 2D de l'*Umbrella Visualization* utilisée ici projette uniquement la position

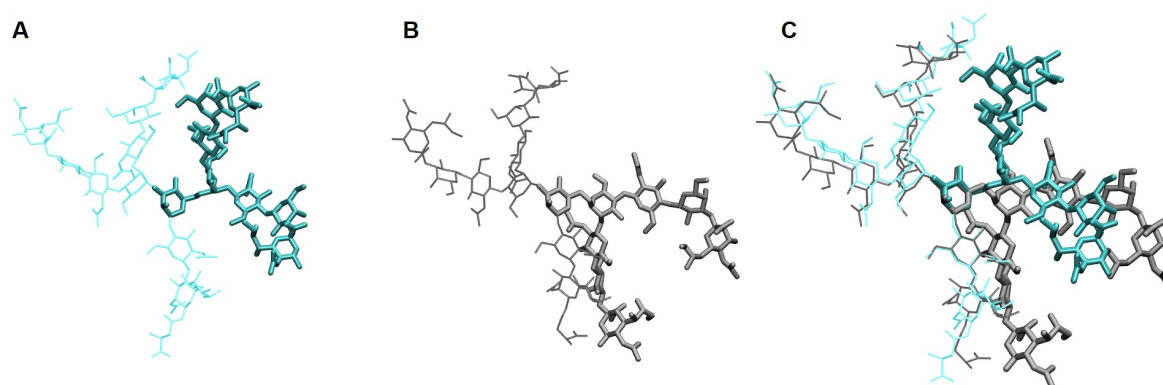


FIGURE 4.9 – Comparaison des conformations principales du glycane tétra-antenné avec acides sialiques liés en $\alpha 2-3$ obtenues sur la trajectoire 1 d’une part et sur l’ensemble des trois trajectoires d’autre part. Le mannose central, le bras $\alpha 1-6$ et les branches associées sont représentés sous forme de bâtons et le reste du glycane (cœur et bras $\alpha 1-3$) en lignes. **A** : Conformation majoritaire issue de la trajectoire 1. **B** : conformation majoritaire issue de l’ensemble des trajectoires pour ce glycane. **C** : Superposition des deux structures après alignement sur les atomes lourds du cœur du glycane. Alors que les deux branches du bras $\alpha 1-3$ se superposent bien, le bras $\alpha 1-6$ présente une conformation principale différente sur la trajectoire 1.

des centres de masse des galactoses des branches du glycane. Ainsi, bien que les informations apportées soient pertinentes en termes de surface explorée, les acides sialiques eux-mêmes ne sont pas pris en compte avec ce mode de projection. Or, les acides sialiques montrent des conformations par rapport au glycane très variables en fonction de leur liaison au glycane. La liaison en $\alpha 2-6$ plus longue permet à ces saccharides de se replier le long des branches du glycane, contrairement à la liaison en $\alpha 2-3$. On peut donc supposer que le type de la liaison affecte l’étendue de la surface couverte, même si les régions explorées par les branches restent similaires. De plus, plusieurs études ont rapporté l’augmentation de la proportion d’acides sialiques liés en $\alpha 2-3$ dans un contexte pathologique [41, 101, 108–110] et considérant l’importance des acides sialiques dans la régulation des fonctions protéiques, on peut supposer que la différence de conformation due à cette liaison affecte les interactions entre une protéine d’intérêt et ses partenaires. Il est également possible que cette différence de conformation soit à l’origine de la stabilisation de la conformation alternative du bras $\alpha 1-6$ en présence d’acides sialiques liés en $\alpha 2-3$.

Ces travaux focalisés sur les glycanes isolés nous ont permis de compléter les résultats précédemment obtenus avec des glycanes bi- et tri-antennés en présence ou non d’acides sialiques. Les glycanes bi-antennés bisectants ayant été déjà largement étudiés [101, 103, 105, 107, 111], nous nous sommes concentrés principalement sur trois types de glycanes tétra-antennés. Nous montrons que leur composition affecte peu la surface explorée. Cependant, l’existence d’une conformation stable alternative du bras $\alpha 1-6$ sur le glycane tétra-antenné avec acides sialiques liés en $\alpha 2-3$ démontre dans le cadre de notre échantillonnage l’influence de ces derniers. Considérant la nature très

flexible des chaînes glycosylées de ces molécules, il semble néanmoins raisonnable d'envisager un échantillonnage conformationnel plus exhaustif en générant d'autres trajectoires, éventuellement à l'aide d'échanges de répliques.

4.3 Fibromoduline et chaînes de keratan sulfate

Au sein de la MEC et à la surface des protéines constituant cette dernière, des glycanes complexes et également des glycosaminoglycanes comme ceux portés par les SLRP sont présents. Ces protéines font partie de la superfamille des protéines LRR (*Leucine Rich Repeat*, répétitions riches en leucine) qui est caractérisée par des répétitions de 20 à 30 acides aminés comportant de nombreuses leucines. Ces motifs répétés leur donnent une forme de solénoïde courbé. La fibromoduline, une SLRP comportant 11 LRR composés chacun d'une vingtaine d'acides aminés, possède deux sites de liaison au collagène (au niveau des LRR7 et 11) ainsi que 4 asparagines glycosylées portant des chaînes de keratan sulfate [112].

Afin de caractériser l'effet des chaînes de keratan sulfate nous avons ajouté 4 chaînes de glycosylation à la fibromoduline et généré 64 ns de trajectoire de dynamique moléculaire à l'aide des champs de forces Amber14 (description de la protéine) et GLYCAM06 (description des glycanes). Nous avons choisi ces champs de force en raison de leur fiabilité et leur compatibilité pour la simulation de protéines glycosylées. En parallèle, nous avons généré 80 ns de trajectoire avec la fibromoduline sans glycanes en utilisant le champ de forces Amber14. En raison de la grande taille des systèmes (en particulier la fibromoduline glycosylée), le temps de trajectoire que nous avons pu obtenir est plus court que pour les glycanes isolés. Le nombre d'atomes de chacun des systèmes est répertorié dans la section 3.1.3.

La visualisation de la trajectoire du système portant les chaînes de keratan sulfate permet d'observer que sur les quatre chaînes de glycosylation, deux se replient autour de la fibromoduline et forment de nombreux points de contact avec la protéine. Plutôt que de couvrir indifféremment la surface protéique, ces glycanes se replient essentiellement autour de la face convexe de la protéine et laissent la face concave accessible. Une partie de la chaîne reste ainsi étendue vers le solvant ou interagit avec une autre chaîne de glycosylation proche. La figure 4.10 présente les vues de chacune des faces du système protéique glycosylé en fin de la trajectoire.

Lorsque nous cherchons à caractériser plus précisément les contacts entre les chaînes de glycosaminoglycanes et la protéine, nous observons que les contacts au niveau de la chaîne de glycosylation n'impliquent pas tous les saccharides de la chaîne mais seulement certains points particuliers : les sulfates des N-acétyl glucosamine semblent être à l'origine de ces contacts et

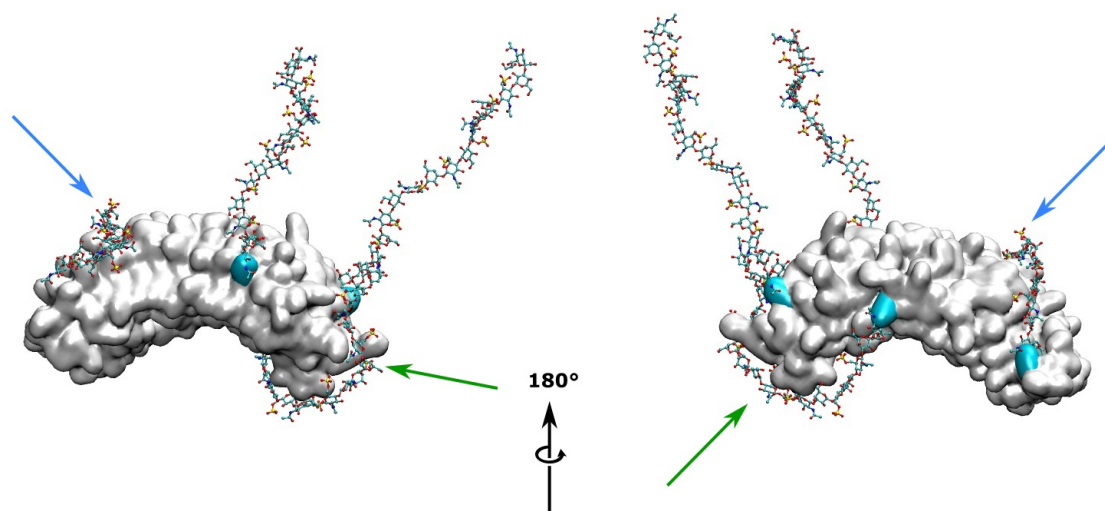


FIGURE 4.10 – **Repliement des chaînes de glycosaminoglycane autour de la fibromoduline.** La protéine est affichée sous forme de surface blanche. Les asparagines glycosylées sont colorées en cyan. Les glycosaminoglycane sont représentés en boule-bâtons. Durant la trajectoire, deux des quatre chaînes (identifiées par les flèches bleues et vertes) se replient autour de la protéine. Les clichés présentés (2 orientations différentes) correspondent à la fin de la trajectoire de la dynamique moléculaire ($t = 64$ ns)

se placent à proximité d'acides aminés chargés positivement ou capables d'être impliqués dans des liaisons hydrogène. La figure 4.11 illustre ces contacts ainsi que les acides aminés impliqués. Plusieurs histidines et arginines sont localisées à proximité des groupements sulfates. Afin de décrire ce repliement, nous avons identifié des résidus impliqués dans ces contacts et mesuré la distance au cours de la trajectoire entre les centres de masse du sulfate et d'un résidu du points de contact. Les résidus des points de contacts pour le glycan 1 (correspondant aux contacts de la figure 4.11 A) sont deux asparagines. Pour le glycan 2 (figure 4.11 B), le premier résidu du point de contact est une asparagine. Le second site est situé à la base d'une autre chaîne de keratan sulfate (figure 4.11 B2), nous avons donc mesuré la distance entre le centre de masse de l'acide sialique terminal du glycan 2 et une N-acétylglucosamine sulfatée de l'autre chaîne. Les courbes ainsi obtenues permettent de mettre en avant le rapprochement des saccharides de la protéine (figure 4.6 et, en particulier dans le cas du point de contact 2 du glycan 1, une stabilisation de la position du groupement sulfate et donc du saccharide associé.

La fibromoduline est une protéine impliquée dans la fibrillogenèse du collagène et capable de s'y lier grâce à deux sites de liaisons au niveau des LRR 7 et 11. Le LRR11 est défini comme un site de haute affinité pour le collagène et le LRR7 comme un site de basse affinité pour le collagène. Sur les 4 sites de glycosylation de la fibromoduline, 3 sont éloignés de ces régions. Le glycan porté par le quatrième s'enroule autour de la surface de la fibromoduline à proximité du

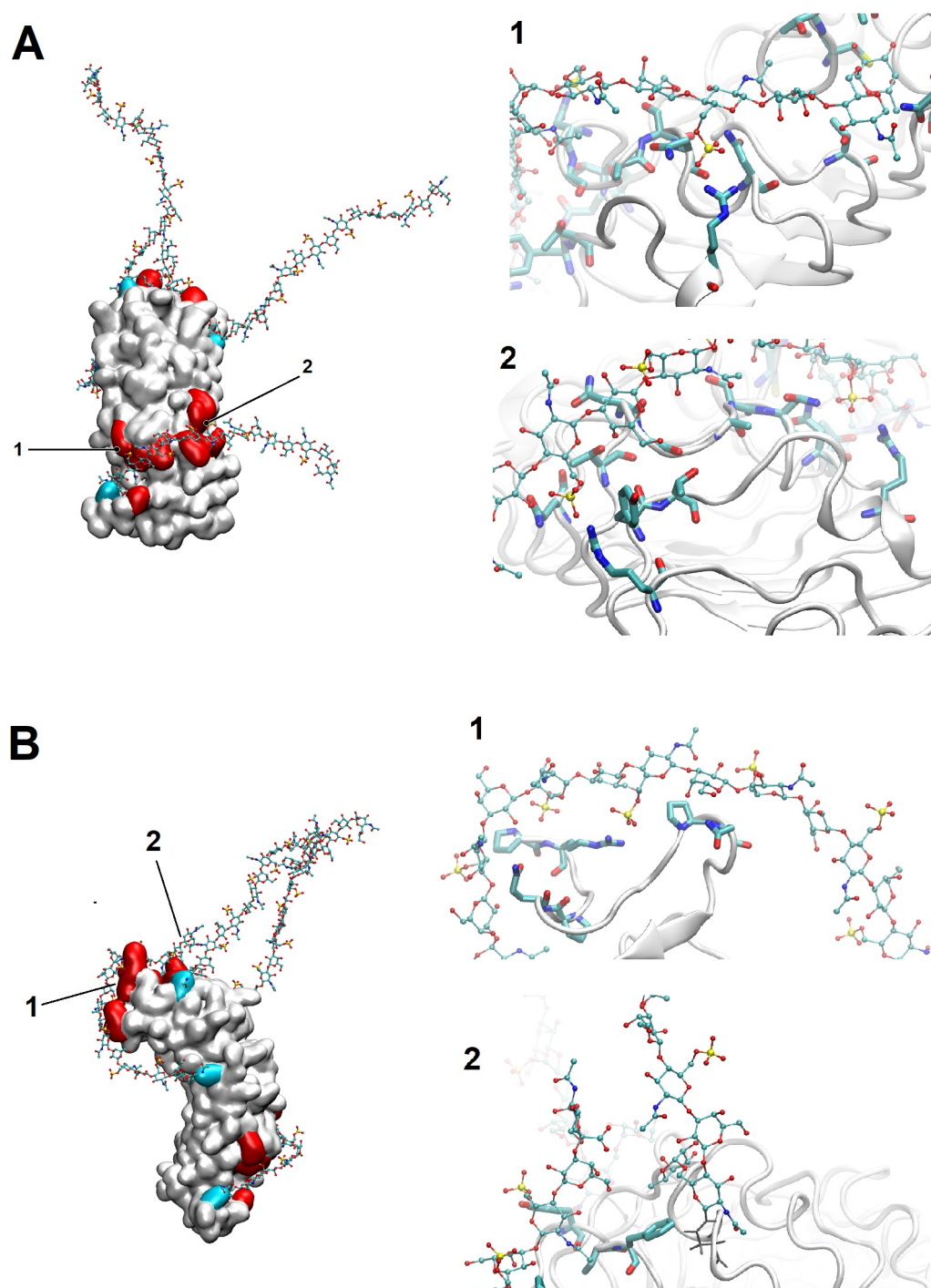


FIGURE 4.11 – **Détail des points de contacts des glycosaminoglycanes repliés sur la fibromoduline.** Les groupements sulfates du glycosaminoglycane sont impliqués dans des contacts avec des résidus polaires. Les acides aminés proches des points de contact sont colorés en rouge sur la surface. **A** : Premier glycan replié, l'extrémité de la chaîne reste dirigée vers le solvant. Pour ce premier glycan, deux points de contacts (labels **1** et **2**) ont été identifiés et s'organisent autour de 2 groupements sulfate. **B** : Deuxième glycan replié, les groupements sulfates du glycosaminoglycane sont impliqués dans des contacts avec des résidus polaires au niveau d'une boucle (label **1**) et l'extrémité de la chaîne de glycosaminoglycane rejoint une autre chaîne au niveau du point de contact **2**.

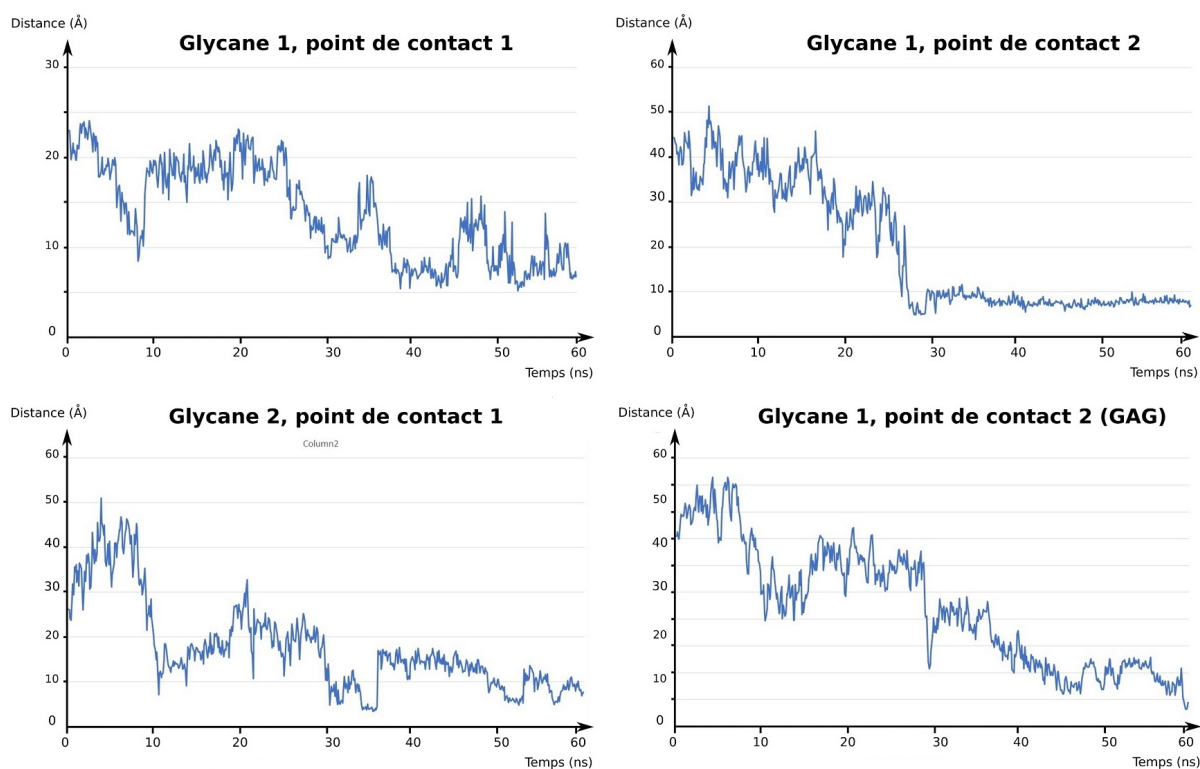


FIGURE 4.12 – Distance au cours du temps entre les centres de masse des groupements sulfate et des résidus des points de contact. La distance pour les deux sites du glycan 1 et le point de contact 1 du glycan 2 a été mesurée entre un sulfate du GAG et une asparagine de la protéine. Le dernier point de contact (glycan 2, site 2) correspond à un contact avec une autre chaîne de glycosylation. La distance a été mesurée entre l'acide sialique terminal du glycan 2 et une N-acétylglucosamine sulfatée de l'autre chaîne. Dans tous les cas, la diminution de la distance traduit le rapprochement entre la chaîne de keratan sulfate et la protéine.

LRR7 (figure 4.13). La proximité de la chaîne de keratan sulfate avec ce LRR pourrait jouer un rôle dans les interactions avec ce site de liaison au collagène.

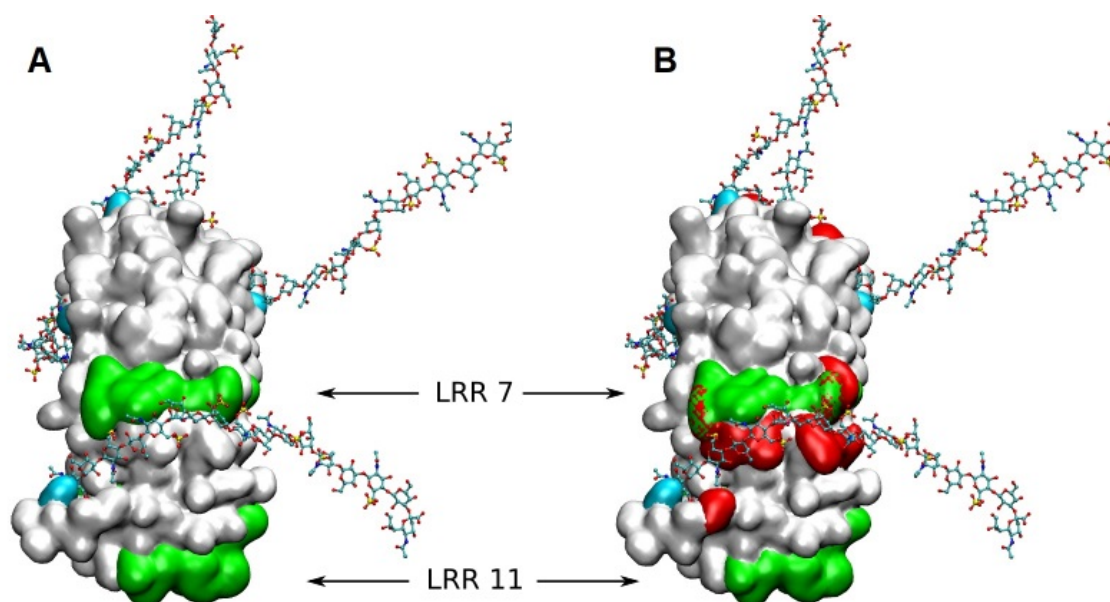


FIGURE 4.13 – **Comparaison entre la position du premier glycanes replié et le LRR7 de la fibromoduline.** Les LRR7 et 11 sont colorés en vert, les asparagines glycosylées en cyan.

Le repliement des chaînes de keratan sulfate observé ici montre que celles-ci sont capable à la fois de couvrir une large surface de la protéine et également d’interagir entre elles. L’importance de la surface ainsi couverte et la taille de ces glycanes par rapport à la protéine illustre l’importance des chaînes de glycosaminoglycanes. D’après nos premiers résultats de dynamique moléculaire, dans le cas de la fibromoduline, il semble que le keratan sulfate s’enroule principalement autour de la face convexe de la protéine, laissant la face concave accessible au solvant ou à tout autre type d’interaction.

Ces premiers résultats montrent que, malgré la courte durée de la trajectoire, les chaînes de glycosaminoglycanes se replient très vite sur la protéine et ce malgré leur grand taux de sulfatation. En effet, les groupements sulfates sont un élément clé de l’hydratation des matrices extracellulaires et tendent à attirer les molécules d’eau. Ceci, ainsi que les contre-ions attirés par les charges des glycosaminoglycanes, pourrait limiter ou influencer la conformation des chaînes à la surface de la protéine. Cependant, la présence d’interactions avec la protéine, en particulier à proximité des sites de liaison au collagène montre que les glycosaminoglycanes peuvent potentiellement masquer une région clé de la surface protéique. Ces premiers résultats nous encouragent

à poursuivre ces travaux avec des expériences supplémentaires afin d'étudier l'impact des glycosaminoglycanes sur les protéines de la matrice extracellulaire et d'étendre ces recherches, d'une part en augmentant l'échantillonnage de conformations sur la fibromoduline et d'autre part en étendant ces recherches au lumican, une autre SLRP étudiée au sein de l'unité.

Les méthodes de dynamique moléculaire nous ont ainsi permis de mettre en avant certains phénomènes dans la dynamique et la structure des glycanes isolés ou de la fibromoduline glycosylée. Afin de caractériser les interactions entre les deux parties des systèmes glycosylés, nous avons également développé un nouvel outil dédié à la visualisation des interactions entre la surface protéique et les glycosylations.

Chapitre 5

Développement d'un nouvel outil de visualisation intégré à UnityMol

5.1 Problématique liée à la représentation des chaînes de glycosylation

Les expériences de dynamique moléculaire menées durant la thèse ont généré une grande quantité de données qui doivent ensuite être traitées et analysées. Les logiciels de visualisation moléculaire sont des outils qui permettent de suivre et de visualiser les mouvements des molécules d'intérêt au cours des trajectoires de dynamique moléculaire. Les différents modes de représentation des surfaces moléculaires, atomes et acides aminés permettent de mettre en évidence des éléments de structure ou des interactions clés qui régissent le comportement de systèmes moléculaires. Développer des outils de visualisation spécifiques apparaît ainsi comme un aspect essentiel du processus de modélisation moléculaire. Un axe prépondérant de ce travail de thèse a consisté à développer une méthode de visualisation permettant d'évaluer au cours d'une trajectoire de dynamique moléculaire la totalité de la surface couverte par un glycané lié à une protéine.

L'outil développé et présenté dans ce chapitre se base sur une méthode mise en œuvre précédemment au sein de l'unité MEDyC et qui permet d'obtenir une représentation 2D sous forme de graphique de la flexibilité d'un glycané au cours d'une trajectoire. Cette méthode a permis de mettre en avant l'influence des acides sialiques sur la flexibilité et la dynamique de glycanes bi- et tri-antennés [17]. Cependant, cette méthode ne permettait pas de prendre en compte la topologie des protéines ni de voir comment les glycanes l'impactent. C'est pourquoi nous nous sommes inspirés de ce premier outil pour mettre en place une méthode de visualisation et un mode de représentation avec le logiciel UnityMol qui permet de prendre en compte la surface protéique. Grâce à cet outil, nous voulons mettre en évidence l'influence des glycosylations sur

la structure et la dynamique des protéines et de leurs interactions.

5.2 Adaptation de l'*Umbrella Visualization* à UnityMol

5.2.1 Présentation des éléments clés utilisés lors de l'implémentation

Pendant le développement de l'*Umbrella Visualization*, nous avons utilisé de nombreux outils proposés par le moteur Unity (présenté dans la section 3.2.3). Cette section détaille ceux que nous avons utilisés afin d'implémenter l'*Umbrella Visualization* dans UnityMol (logiciel présenté dans la section 3.2.2). Nous avons ainsi pu recréer un système en 3 dimensions comportant un plan et un axe de référence z permettant de collecter l'information désirée d'une manière similaire à celle de la version 2D (cf. section 3.2.1). Comme indiqué dans la section 3.2.3, chacun de ces éléments ainsi que leurs propriétés sont accessibles et modifiables au travers de scripts en C# ou Javascript.

GameObject

Les *GameObjects* sont les objets fondamentaux de Unity, auxquels différents composants seront ajoutés. Un *GameObject* quel qu'il soit possède toujours un composant de type *Transform*, qui définit sa position et son orientation dans l'espace de la scène. C'est le seul composant d'un *GameObject* qui ne peut pas être retiré. En lui-même, un *GameObject* n'a donc pas de fonctions définies et c'est l'ajout d'autres composants tels que les lumières, caméras, volumes géométriques ou textures qui lui donnent sa fonction et son apparence. Enfin, différents *GameObjects* peuvent être combinés et organisés pour former un modèle qui pourra être instancié au besoin : un *Prefab*.

Prefab

Les *Prefabs* sont un outil dans Unity permettant le regroupement d'objets (*GameObjects*) et la définition de leurs propriétés individuelles. Ceci permet donc d'organiser plusieurs composants les uns par rapport aux autres et de les instancier plusieurs fois en conservant cette organisation. Il est possible de manipuler les instances du *Prefab* modèle de manière indépendante et d'opérer des modifications sur l'ensemble ou sous-partie des composants d'une instance individuelle, sans modifier le modèle d'origine. Dans le cas de l'*Umbrella Visualization*, sauvegarder l'ombre du glycanes se fait à l'aide d'un *Prefab* composé d'une lumière directionnelle, d'une caméra et d'un projecteur placés au dessus d'un plan, émulant ainsi un repère orthonormé. Ceci nous permet de créer un ou plusieurs plans et de les placer puis les orienter indépendamment par rapport aux différents glycanes étudiés.

L'utilisation des *Prefabs* nous permet donc de définir un modèle capturant l'ombre d'un glycanes puis d'effectuer cette opération sur plusieurs glycanes de façon identique sans répéter toutes les opérations mais en orientant le modèle selon le glycanes. L'information recherchée ici, l'ombre du glycanes sur le plan, est sauvegardée grâce aux propriétés de la caméra et des textures.

Quaternion

Afin de s'adapter aux données chargées dans le logiciel, il est nécessaire de déplacer et pivoter le plan du *Prefab* dans l'espace pour le positionner correctement sous le glycanes. Les rotations associées à ce positionnement sont stockées dans des objets en quatre dimensions (coordonnées cartésiennes dans l'espace x, y z et un paramètre w qui caractérise la rotation autour du vecteur défini par ces coordonnées) appelés *Quaternions*. Ces objets sont dédiés au calcul et stockage de rotations qu'il est ensuite possible d'appliquer au *Transform* attaché au *GameObject* d'intérêt. Dans notre cas, ce *GameObject* est un *Prefab*.

Textures

Les textures sont des fichiers d'images qui sont appliqués à la surface des objets pour donner à ceux-ci une surface colorée et / ou texturée. Les formats d'images supportés par Unity comprennent les formats classiques (BMP, TIF, TGA, JPG, PNG,...). Certains formats supportant les calques (comme les PSD générés par le logiciel Photoshop) peuvent aussi être utilisés comme textures. Dans ce cas, les calques sont accessibles depuis Unity. Cependant, ces textures sont aplaties au lancement du logiciel, donc l'accès aux calques de l'image n'est pas possible pendant l'exécution du logiciel. La plupart des formats d'images supportés permettent de gérer la transparence des pixels, sauf le format JPG.

Lorsque la texture est appliquée sur un objet, le mode d'enveloppement de cette texture définit la façon dont elle est appliquée sur l'objet : le mode *Repeat* reproduit la texture à l'identique plusieurs fois, donnant un effet de "carrelage"; le mode *Clamp* répète les pixels des bords de la texture tout autour de l'objet. Ce dernier mode permet de n'afficher qu'une seule fois le motif dans sa taille d'origine au centre d'une texture.

Différents types de textures existent, chacun avec leurs propres spécificités. Dans le cadre de l'*Umbrella Visualization*, nous utilisons une *Render Texture* et deux *Texture2D*. La principale différence entre ces deux types de texture est la capacité ou non à modifier l'apparence des pixels depuis des scripts pendant l'exécution. Dans le cas d'une *Render Texture*, son apparence est déterminée uniquement par la vue d'une caméra. Ses pixels ne peuvent donc pas être modifiés directement. Les *Texture2D* sont des textures classiques, modifiables à l'exécution.

Pixels et canaux rgba

Les textures étant des images, elles sont donc composées de pixels. La couleur et l'opacité de chacun de ces pixels sont définies par quatre canaux : r, g, b et a, pour *red*, *green*, *blue* et *alpha*. Les trois premiers canaux permettent de définir la couleur du pixel et le dernier, *alpha*, détermine l'opacité du pixel. Chacun de ces canaux est encodé dans un byte et peut donc prendre une valeur entre 0 et 255, 0 signifiant que la couleur correspondante n'est pas représentée dans le pixel. Pour le canal *alpha*, 0 signifie que le pixel est complètement transparent et 255 signifie que le pixel est complètement opaque. Ceci permet donc de moduler à la fois la couleur et la transparence des pixels de la texture pour lui donner l'aspect désiré. Ces canaux sont accessibles et modifiables indépendamment les uns des autres à partir des scripts C#.

Il est ainsi possible de modifier l'apparence d'une texture de type *Texture2D* pendant l'exécution du programme et d'appliquer ces modifications en conséquence. Nous sommes ici principalement intéressés par l'affichage de pixels en niveau de gris. Ceci est possible en appliquant aux composants r, g et b une valeur identique. Ainsi, lorsque ces trois composants sont définis à une valeur de 255, le pixel est blanc. Pour une valeur de 0, il est noir. Les valeurs intermédiaires correspondent à 254 nuances de gris plus ou moins foncés.

Shaders

Les *Shaders* sont de courts programmes exécutés par la carte graphique d'un ordinateur. Ils sont utilisés afin de pouvoir afficher les textures des objets, les reflets ou la réfraction de la lumière à leur surface, les ombres...

Projecteurs et caméras

Les caméras et projecteurs sont des objets permettant respectivement de capturer un point de vue et de projeter une image pouvant prendre la forme d'une *RenderTexture* ou d'une *Texture2D*. L'utilisation des *Culling Masks* et des *Layers* (définis ci-après) permettent de contrôler quels objets sont capturés par la vue de la caméra et quels objets reçoivent l'image du projecteur.

Calques et *Culling masks*

Les calques (*Layers*) permettent de regrouper les objets d'une scène. Ils sont ensuite utilisés par les lumières et les caméras pour illuminer ou afficher à l'écran des éléments en fonction de leur présence ou non dans un calque. Sélectionner les calques qui seront illuminés ou affichés à l'écran se fait en configurant le *Culling mask* des lumières et des caméras. Ce dernier paramètre étant accessible depuis les scripts, il est possible de changer dynamiquement, à l'exécution du

programme, les objets affectés par les lumières et les caméras. Il est ainsi possible d'utiliser ces fonctions pour cacher ou montrer des éléments à l'utilisateur sans avoir à supprimer un objet.

5.2.2 Implémentation de la méthodologie *Umbrella Visualization* au logiciel UnityMol

Afin d'adapter l'*Umbrella Visualization* à UnityMol, nous avons traduit le principe de cette méthode avec les outils proposés par Unity. Nous avons ensuite développé ce travail pour la lecture de la trajectoire et la compilation des ombres sur toutes les structures issues de la trajectoire.

Premières tentatives

Dans un premier temps, nous avons essayé de récupérer l'ombre du glycanes sur une seule structure directement sur la surface de la protéine. Unity propose en effet l'utilisation de *Lightmaps*, un outil permettant de précalculer une ombre sur un objet puis de sauvegarder l'apparence de la texture. Nous espérons ainsi pouvoir sauvegarder l'information voulue directement, sans utiliser d'outil ou objet intermédiaire.

De plus, nous avons besoin des données de dynamique moléculaire chargées par l'utilisateur pendant l'exécution, pour permettre la collecte et compilation des ombres pendant la lecture d'une trajectoire. Or, l'utilisation des *Lightmaps* n'est pas possible pendant l'exécution du logiciel et nous n'avons donc pas pu utiliser cette méthode.

Traduction du principe

Bien que le résultat de la première version de l'*Umbrella Visualization* [17] donne un graphique en deux dimensions, l'alignement du glycanes s'opérait dans un espace en trois dimensions. Nous avons donc conservé le repère orthonormé (O,x,y,z). Afin d'émuler ce repère, nous avons construit un *Prefab* composé d'un plan, d'une caméra, d'une lumière directionnelle et d'un projecteur. Le plan contient les axes (Ox,Oz) de la version 2D de l'*Umbrella Visualisation*. L'axe y est représenté par la lumière directionnelle, placée au dessus du plan, perpendiculairement à celui-ci. La caméra, placée au même endroit que la lumière directionnelle, permet de capter l'ombre projetée sur le plan, la sauvegardant dans une *RenderTexture*. Finalement, un script en C# intégré au *Prefab* contrôle les fonctions associées à chacun de ces éléments et permet de lancer le processus d'analyse de façon automatisée.

Les propriétés des *Culling masks* sont utilisées pour projeter l'ombre du glycanes sur le plan sans la lumière principale de UnityMol, permettant à l'utilisateur de voir et de manipuler sa molécule, sans que cela n'affecte cette ombre. De la même manière, cela permet de configurer

la caméra du *Prefab* du plan orienté pour qu'elle sauvegarde uniquement l'image de l'ombre sur le plan, sans influence parasite d'autres éléments du système étudié (surface de la protéine traversant le plan, molécules d'eau,...). Afin de ne pas encombrer la visualisation, le plan est masqué à l'utilisateur grâce à un dernier *Culling mask* appliqué sur la caméra principale de UnityMol qui définit la vue de l'utilisateur du logiciel.

Mise en place du *Prefab*

Le *Prefab*, une fois instancié, doit être placé et orienté sous le glycanes. Pour ce faire, c'est l'asparagine portant une glycosylation qui est utilisée comme point de repère. Dans la première version de l'*Umbrella Visualization*, le plan était orienté par rapport au cœur du glycanes. Cette partie du glycanes est rigide [18] et l'utiliser comme axe Oy permet de ne considérer et projeter sur le plan (Ox,Oy) que le mouvement des branches du glycanes. Cependant, dans le contexte protéique, il est nécessaire de prendre en compte l'orientation du cœur du glycanes par rapport à la protéine. En effet, même si le cœur du glycanes est rigide, les structures pdb montrent qu'il peut se plaquer le long de la surface protéique [62]. C'est pourquoi le point de repère choisi pour aligner et orienter le plan est l'asparagine portant la glycosylation. En prenant cet acide aminé comme point de référence, il est ainsi possible de prendre en compte à la fois les mouvements des branches et le degré d'inclinaison de la chaîne de glycosylation par rapport à la surface de la protéine. A l'heure actuelle, l'*Umbrella Visualization* sélectionne uniquement la première asparagine glycosylée rencontrée à la lecture de la structure.

Les données de la structure chargée dans UnityMol permettent de détecter les résidus asparagine dont l'amine terminale forme une liaison avec un sucre. Pour cet acide aminé, les coordonnées du carbone alpha (noté C) et de l'atome d'azote (noté N) permettant la liaison glycosidique seront sauvegardées et utilisées pour calculer les coordonnées d'un vecteur $\vec{NC}$. Ces coordonnées et ce vecteur nous permettent de déterminer la position et l'orientation du plan sous le glycanes. Le vecteur $\vec{NC}$ est en effet utilisé pour définir l'axe Oy du repère orthonormé (figure 5.1 A). Choisir l'asparagine glycosylée comme point de repère permet de placer le plan de façon relativement indépendante du glycanes, permettant de prendre en compte les variations d'inclinaison du glycanes sur la surface de la protéine. De plus, la présence du glycanes sur l'asparagine nous assure de son orientation globale vers le solvant et donc de la bonne orientation du plan par rapport au glycanes et à la surface de la protéine.

A l'initialisation du *Prefab*, aucune rotation n'est appliquée à l'ensemble des éléments, et la normale au plan est donc un vecteur de coordonnées (0,1,0) (figure 5.1 B). La rotation entre la normale du plan et le vecteur $\vec{NC}$ est calculée et stockée dans un *Quaternion* dédié (figure

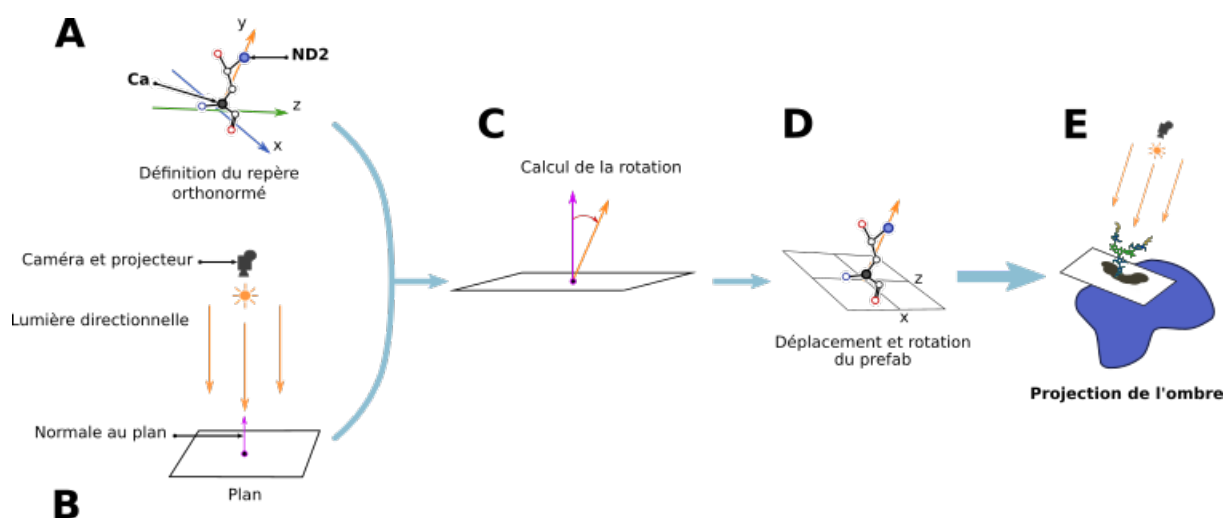


FIGURE 5.1 – **Rotation et déplacement du plan par rapport à l'asparagine glycosylée.** **A** : Les atomes du carbone alpha de l'asparagine et de l'azote portant la liaison glycosydique permettent de définir un axe Oz, normale d'un plan (Ox,Oy). **B** : Le *Prefab* est composé d'un plan, d'une lumière directionnelle, d'une caméra et d'un projecteur. A l'initialisation du plan, la normale au plan est un vecteur vertical de coordonnées (0,1,0). **C** : La normale au centre du plan est utilisée afin de calculer l'angle de rotation à appliquer au *Prefab*. **D** : La rotation est appliquée au *Prefab* et le centre du plan est déplacé au niveau du carbone alpha de l'asparagine. **E** : L'ombre du glycan est projetée sur le plan.

5.1 C). En utilisant la normale partant du plan du *Prefab* pour calculer la rotation, nous nous assurons que le plan soit toujours correctement orienté dans la même direction que l'asparagine, de sorte que la caméra et le projecteur soient bien au dessus du glycan. L'étape suivante est de déplacer et d'orienter le plan : le carbone alpha de l'asparagine sélectionnée est utilisé comme point de repère. Ses coordonnées sont utilisées pour placer le centre du plan sur cet atome et la rotation calculée à l'étape précédente est appliquée au *Prefab* (figure 5.1 D). Ces éléments une fois positionnés et orientés, l'ombre du glycan peut être projetée sur le plan, sauvegardée et visualisée sur la surface de la protéine (figure 5.1 E).

La dernière étape est ensuite de configurer l'affichage des sucres : afin de simuler leur volume, le mode de représentation en sphères de van der Waals (figure 5.2) est utilisé pour le calcul de l'ombre puis les atomes sont masqués pour n'afficher que leur ombre dans le champ de vision de la caméra.

Structure fixe d'un fichier pdb simple

Dans un premier temps, nous avons cherché à collecter l'ombre du glycan sur une seule structure fixe (pdb, sans trajectoire). Le *Prefab*, une fois instancié et placé sous le glycan comme expliqué auparavant, permet de sauvegarder l'ombre du glycan grâce aux éléments qui le com-

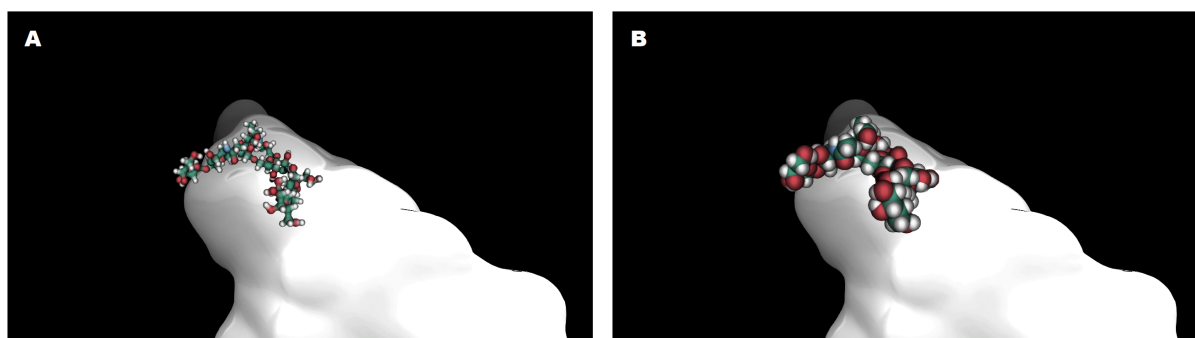


FIGURE 5.2 – **Représentation des atomes pour l'*Umbrella Visualization*.** La surface de la protéine apparaît en blanc. Les atomes du glycan sont en couleurs : les atomes de carbone en vert, les atomes d'oxygène en rouge et les atomes d'hydrogène en blanc. **A** : Représentation *Hyperballs* classique dans UnityMol. Les liens entre les atomes sont bien visibles mais le volume des atomes n'est pas reflété. **B** : Représentation en sphères de van der Waals : les liens entre les atomes ne sont plus visibles, mais le volume occupé par les atomes est mieux représenté. Cette représentation est utilisée afin d'évaluer la couverture du glycan pendant la trajectoire de dynamique moléculaire.

posent : la lumière directionnelle projette cette ombre sur le plan et la caméra capte cette image et la sauvegarde dans la *RenderTexture*. La *RenderTexture* est ensuite utilisée dans le projecteur afin d'afficher l'ombre du glycan sur la surface de la protéine. L'utilisation des calques tels que définis précédemment permet de conserver uniquement l'information de l'ombre, sans influence parasite des atomes ou de la surface de la protéine. Ces étapes sont résumées par le procédé représenté par les flèches bleues sur la figure 5.4.

La figure 5.3 montre le résultat obtenu sur un glycan d'une structure unique du récepteur à l'insuline porteur de deux glycosylations, extraite d'une trajectoire de dynamique moléculaire. Ceci peut permettre de mettre en avant une structure particulière ou une conformation représentative d'une trajectoire.

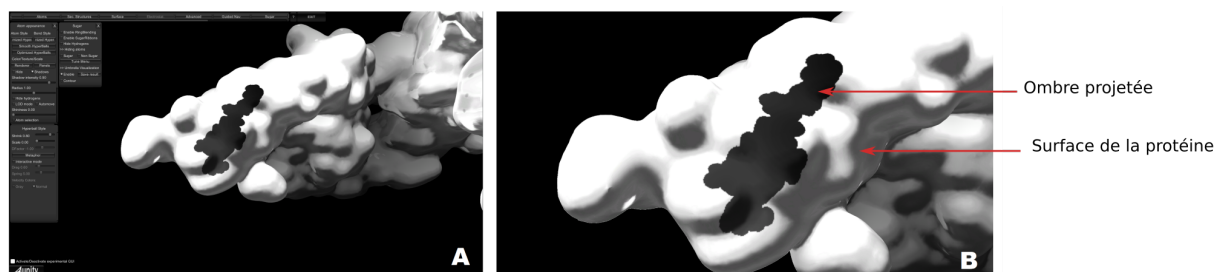


FIGURE 5.3 – ***Umbrella Visualization* pour une seule structure.** **A** : Ombre projetée sur la surface de la protéine dans UnityMol. Le menu *Sugar* est visible sur la gauche. **B** : Détail de l'ombre projetée sur la surface.

A partir de ce simple flux utilisant une seule texture, nous avons ensuite travaillé à développer une méthode plus complète permettant de lire et d'analyser les données d'une trajectoire de dynamique moléculaire.

Lecture d'une trajectoire

Dans le cas de la lecture d'une trajectoire, nous devons conserver l'information de plusieurs images afin de pouvoir les compiler et obtenir un cliché final statistiquement pertinent. La méthode pour instancier, placer et orienter le *Prefab* reste la même que celle décrite précédemment.

Dans ce cas, à chaque nouvelle image, la *RenderTexture* est transcrite en *Texture2D* afin de pouvoir lire les pixels et détecter ceux faisant partie de l'ombre et une texture finale est ensuite utilisée pour compiler les informations sur l'ensemble de la trajectoire. Ces trois textures (*RenderTexture*, *Texture2D* et texture finale) doivent faire la même taille en pixels afin de ne pas perdre d'information au moment de la conversion. Sur la figure 5.4, les flèches violettes résument ce procédé.

Le script présent dans le *Prefab* permet de créer un tableau d'entier à l'activation de l'*Umbrella Visualization*. La taille de ce tableau est définie par le nombre de pixels dans les textures utilisées par le *Prefab*. De cette manière, chaque case du tableau est associée à un pixel de la texture. Pendant la lecture de la trajectoire, à chaque nouvelle image (cliché) affichée à l'écran, l'ombre du glycane est capturée. Ainsi, pour chaque cliché, cette fonction va transférer les pixels de la *RenderTexture* sur une *Texture2D* temporaire. Les pixels de cette dernière texture sont ensuite lus un par un afin de détecter s'ils font partie de l'ombre du glycane pour l'image considérée : lorsque c'est le cas, (c'est-à-dire si ce pixel n'est pas blanc), la valeur de la case correspondante du tableau est augmentée de 1, comptant ainsi le nombre de fois qu'un pixel est masqué par le glycane au cours de la trajectoire. Comme expliqué dans le paragraphe suivant, c'est ce tableau une fois complété qui est utilisé pour déterminer l'apparence de la texture finale.

Au cours de la trajectoire de dynamique moléculaire, la protéine et le glycane se déplacent dans l'espace. Afin de compenser ces mouvements, pour chaque cliché lu pendant la trajectoire de dynamique moléculaire, le vecteur de l'asparagine est redéfini en fonction des coordonnées du carbone alpha et de l'atome d'azote N pour ce cliché. Ce vecteur, appelé $N\vec{C}_n$, est utilisé afin de calculer la nouvelle rotation entre la normale du plan (Ox,Oz) (définie par le $N\vec{C}$ du cliché précédent), ou $N\vec{C}_{n-1}$. La rotation depuis $N\vec{C}_{n-1}$ vers $N\vec{C}_n$ est appliquée au *Prefab* permettant ainsi de réorienter le *Prefab* pour chaque nouveau cliché de la trajectoire. La position du *Prefab* est également modifiée en fonction de la nouvelle position du carbone alpha de l'asparagine étu-

diée.

Bien que la protéine se déplace, la surface sur laquelle l'ombre est projetée se base uniquement sur la structure pdb initiale et reste fixe durant la trajectoire. C'est pourquoi la position du projecteur reste fixe tout au long du processus. Ainsi, le plan et la caméra traquent les mouvements de la protéine et du glycanes dans l'espace pendant que le projecteur reste en position au dessus de la surface et y projette les ombres captées.

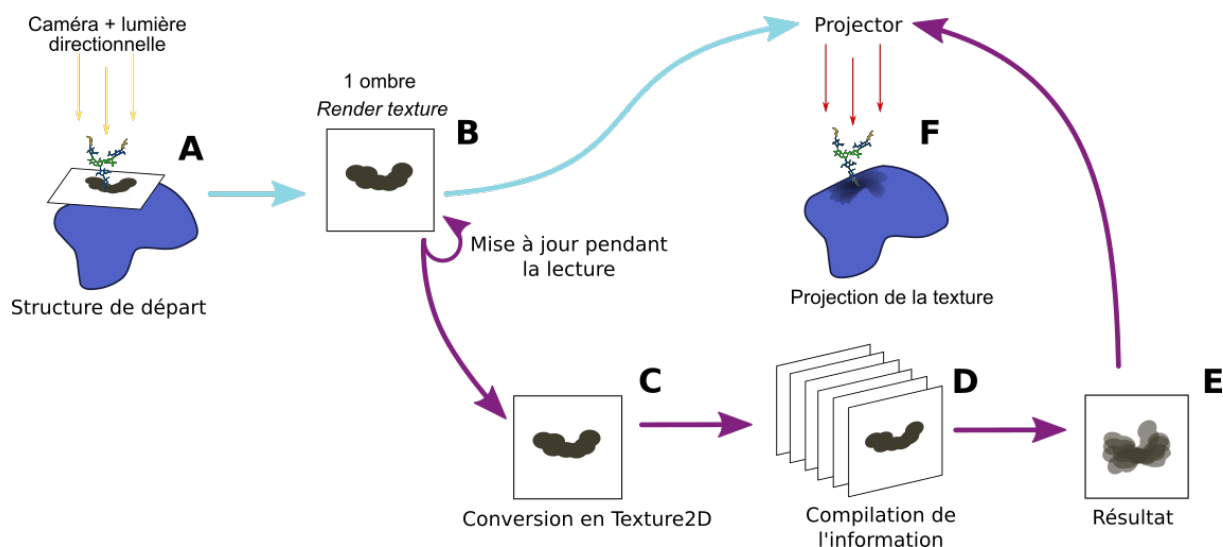


FIGURE 5.4 – Méthodes de projection en fonction des données à analyser. Les flèches bleues correspondent à la méthode pour afficher l'analyse de structures seules (fichiers pdb). Les flèches violettes correspondent à la méthode utilisée pour l'analyse de trajectoires de dynamique moléculaire. **A** : Structure de départ, le *Prefab* est instancié et placé sous le plan. **B** : L'ombre projetée est sauvegardée par la caméra dans la *RenderTexture*. **C** : Cette texture est ensuite convertie en *Texture2D* afin de détecter les pixels de l'ombre du glycanes. **D** et **E** : Les informations de l'ensemble des clichés sont compilées afin de déterminer l'apparence de la texture finale. Les zones les plus souvent masquées pendant la trajectoire apparaissent en noir. **F** : Projection du résultat sur la surface de la protéine. Pour une seule structure, la *RenderTexture* de l'étape B est projetée, pour une trajectoire, c'est le résultat final obtenu après compilation des ombres de la trajectoire.

Compilation de l'information

A la fin de la lecture de la trajectoire, le tableau d'entiers nous permet donc de représenter le nombre de fois qu'un pixel a été couvert par l'ombre du glycanes. Plus la valeur contenue par la case sera élevée, plus le pixel a été couvert pendant la trajectoire. L'étape suivante est donc de représenter les pixels les plus souvent masqués par le glycanes avec une teinte plus sombre. Cela permet de conserver la représentation sous forme d'ombre, tout en mettant en valeur les régions les plus impactées par le glycanes.

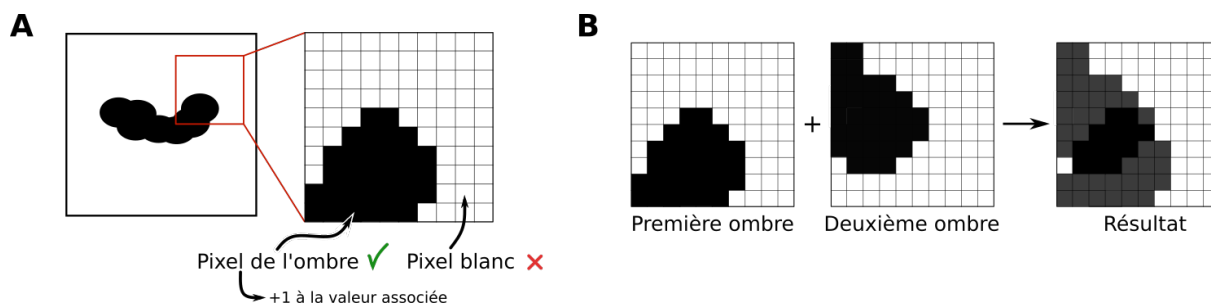


FIGURE 5.5 – **Méthode de compilation des ombres.** **A** : Détection des pixels éclairés ou assombris par le glycane. Au cours d'une trajectoire, le nombre de fois où un pixel est obscurci par le glycane est compté. La valeur finale est utilisée à la fin de la lecture de la trajectoire pour déterminer la couleur du pixel associé. **B** : Exemple avec deux clichés compilés à l'aide de méthode. Les pixels obscurcis une seule fois apparaissent en gris. Ceux obscurcis sur les deux clichés apparaissent en noir.

Comme expliqué précédemment, les couleurs sont encodées sous forme de bits prenant une valeur entre 0 et 255, 0 correspondant à l'absence totale d'une couleur. Afin d'obtenir un résultat en niveau de gris, les composants r, g et b d'un pixel prennent la même valeur, comme expliqué dans le paragraphe 5.2.1. Le grand nombre de clichés pouvant dépasser largement la valeur de 255, les valeurs du tableau sont donc normalisées pour être comprises entre 0 et 255. Cette valeur est soustraite à la valeur maximale de 255 pour que les pixels ayant été obscurcis le plus souvent apparaissent en noir (valeur proche de 0). Le calcul est donc le suivant :

$$Valeur\ du\ pixel = 255 - \left(\frac{Valeur\ de\ la\ case}{Nombre\ de\ clichés} \times 255 \right)$$

La valeur du pixel issue de ce calcul est un entier, obtenu après troncature du résultat. Ce calcul est appliqué à chaque valeur du tableau et permet ainsi de définir l'apparence de tous les pixels de la texture finale, comme résumé sur la figure 5.5. Cette texture est finalement utilisée dans le projecteur du *Prefab*, permettant ainsi de la projeter sur la surface de la protéine.

Représentation et affichage du résultat

Une fois le *Prefab* placé et l'ombre capturée, cette dernière est projetée à la surface de la protéine. Dans le cas d'un fichier pdb, une seule ombre est capturée et projetée. Lorsqu'une trajectoire de dynamique moléculaire est chargée par l'utilisateur, l'ombre de chacun des clichés individuels est projetée successivement durant la lecture de la trajectoire. Une fois la lecture terminée, c'est finalement le résultat compilé qui est affiché.

La surface des protéines est souvent irrégulière et peut présenter un aspect rugueux. La topologie locale de la surface est ainsi un facteur qu'il peut être important de prendre en compte

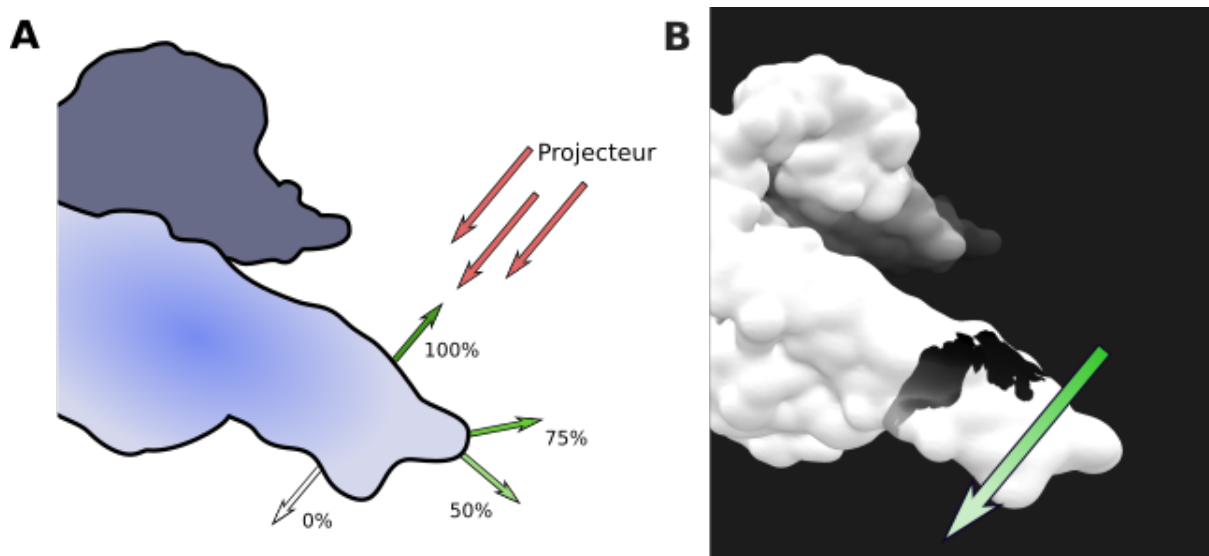


FIGURE 5.6 – Atténuation de l'ombre en fonction de l'orientation de la surface et de la distance. **A** : les normales à la surface représentées par des flèches en nuances de vert permettent de définir avec quelle intensité la texture est projetée sur la surface. **B** : Visualisation de l'atténuation en fonction de la distance au projecteur. En réduisant l'intensité de l'ombre sur les parties éloignées, nous évitons de surestimer l'impact du glycane sur ces régions. La flèche verte donne la direction dans laquelle l'ombre est projetée (et donc atténuée).

pour ne pas surévaluer la couverture de zones orientées vers le solvant plutôt que vers le glycane. Une option a donc été mise à disposition, permettant de moduler l'intensité de l'ombre projetée en fonction de la distance et l'orientation de la surface par rapport au projecteur. Le *Shader* du projecteur a été adapté. Ainsi, quand cette option est activée (voir section 5.2.3) pour chaque point de la surface affecté par le projecteur, la normale à ce point est calculée et son orientation définit l'intensité de l'image projetée. Si cette normale pointe directement vers le projecteur, l'intensité de l'ombre est maximale. Si cette normale a une direction différente, l'image projetée est atténuée (figure 5.6). L'intensité de l'ombre projetée est également modifiée selon la distance de la surface au projecteur. En combinant cette propriété avec la prise en compte de l'orientation de la surface, nous pouvons ainsi moduler l'intensité de l'ombre en fonction de la distance entre la surface et le glycane, mais aussi en fonction de l'orientation de la surface. En plus de la projection de l'information statistique projetée, nous pouvons ainsi prendre en compte les informations en 3D liées à la topologie de la surface protéique.

5.2.3 Intégration et rendu au sein de l'interface de UnityMol

Afin de rendre l'*Umbrella Visualization* accessible à tout types d'utilisateurs, nous l'avons ensuite intégrée à l'interface du logiciel UnityMol.

Affichage pour l'utilisateur

Afin de permettre l'instanciation du *Prefab* et la lecture de la trajectoire, un bouton a été ajouté dans le menu dédié aux sucres (*Sugar*, figure 5.7 A). En fonction des données chargées dans UnityMol par l'utilisateur, deux types de réponse sont possibles : si un pdb seul est chargé, le nombre de *clichés* de la trajectoire de dynamique moléculaire est défini à 0, le *Prefab* est instancié et le résultat (une seule ombre) affiché directement. Si une trajectoire est chargée, la réponse au clic permettra d'accéder aux fonctions qui contrôlent la lecture et la manipulation de la trajectoire. Ainsi, le *Prefab* du plan est instancié et le nombre de s de la trajectoire est utilisé pour la lecture de la trajectoire. Ces actions sont déclenchées automatiquement et ne nécessitent pas d'autre information de la part de l'utilisateur. En effet, le script contenu dans le *Prefab* dispose de toutes les fonctions permettant de gérer le calcul du résultat ainsi que les modes de représentation des atomes, de la surface et du résultat (figure 5.7 B).

Ainsi, pour utiliser l'*Umbrella Visualization*, un utilisateur quelconque peut simplement charger ses données de dynamique moléculaire et accéder à l'*Umbrella Visualization* dans le menu *Sugar* de UnityMol. Comme décrit précédemment, l'utilisation de calques et de *Culling mask* permet de masquer le *Prefab* à l'utilisateur. Afin de permettre l'observation des mouvements du glycane pendant la trajectoire, la *Texture2D* temporaire, utilisée pour détecter les pixels à chaque cliché, est également utilisée dans le projecteur et affichée sur la surface de la protéine. A la fin de la lecture de la trajectoire, la texture projetée est automatiquement modifiée pour faire apparaître le résultat final.

Une fois le calcul statistique terminé, un autre mode de visualisation est proposé : le mode en courbes de niveau (*Contour*). Ce mode de représentation permet de simplifier l'ombre calculée au cours de la trajectoire en définissant des zones bien délimitées. La figure 5.8 permet de comparer les deux représentations.

Les différentes courbes de densité sont calculées en fonction de la valeur associée à chaque pixel enregistré dans le tableau calculé dans la section 5.2.2. Ces paramètres pouvant prendre une valeur entre 0 et 255, nous avons défini 8 intervalles, détaillés dans la table 5.1, ou seuils de valeur qui permettent de créer ces délimitations. Conformément au calcul développé ici, les valeurs les plus basses correspondent aux zones les plus souvent couvertes par le glycane pendant la trajectoire.

Grâce au mode de représentation en contours, la délimitation entre la zone impactée par le glycane et le reste de la protéine apparaît nettement, définissant clairement la zone d'influence

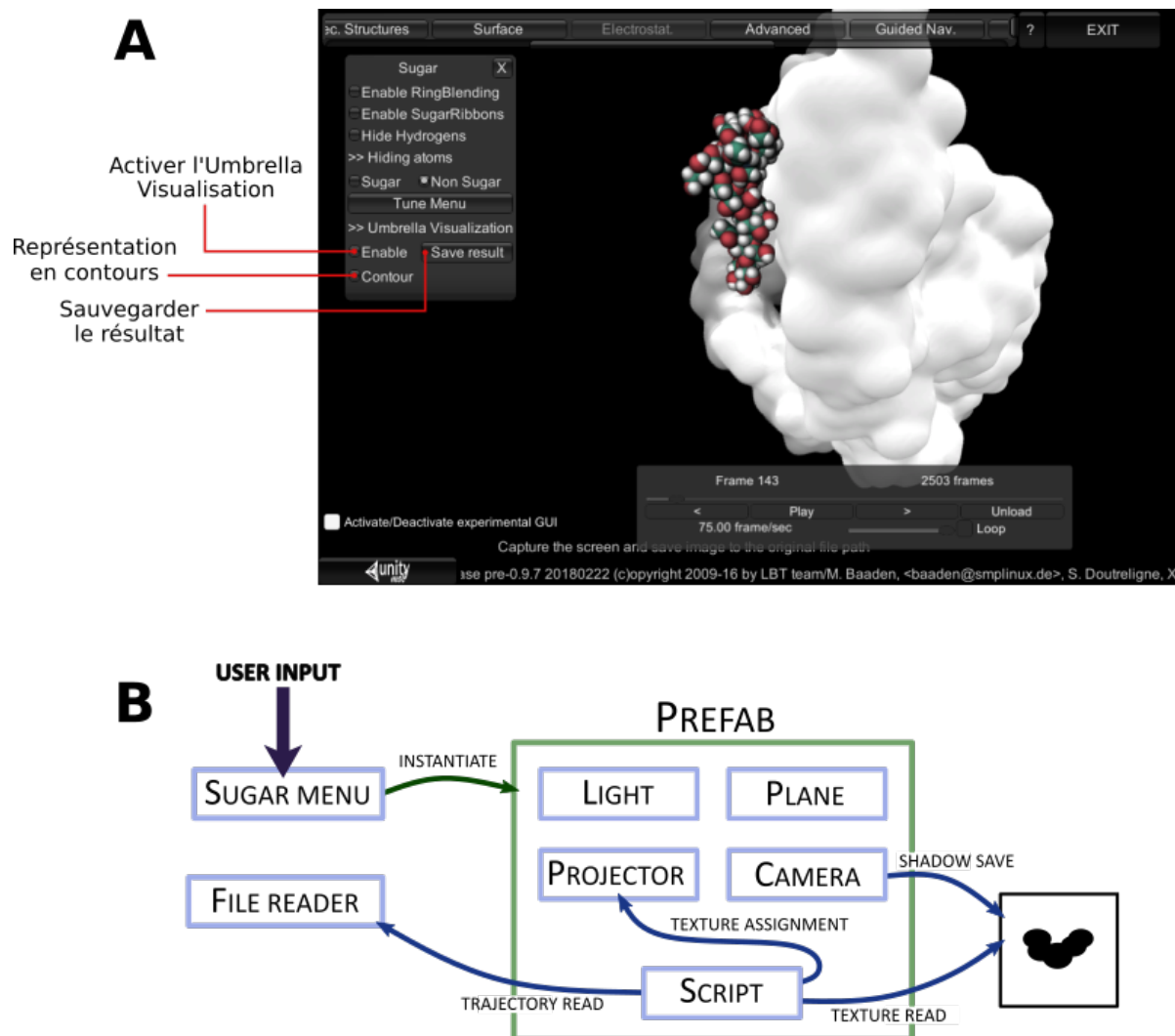


FIGURE 5.7 – **Interface utilisateur de UnityMol.** **A** : Dans le menu *Sugar*, les options de l'*Umbrella Visualization* sont disponibles directement. Le bouton *Enable* permet d'activer l'*Umbrella Visualization* pour des structures simples ou des trajectoires automatiquement. Le bouton *Contour* permet d'activer la représentation en contours pour les résultats de trajectoire de dynamique moléculaire et enfin, l'option *Save Result* permet de sauvegarder la texture finale dans un fichier jpeg. **B** : Schéma représentant les actions déclenchées par l'utilisateur au moment du démarrage de l'*Umbrella Visualization*. Le *Prefab* est instancié depuis le menu *Sugar* puis le script contenu dans le *Prefab* contrôle les différents éléments impliqués dans le processus de compilation ainsi que l'affichage.

Intervalle	0-31	32-63	64-95	96-127	128-159	160-191	192-223	224-255
Valeur r,g,b	0	32	64	96	128	160	192	255
Couleur associée								

TABLE 5.1 – Tableau récapitulant les intervalles de valeurs r,g,b des pixels pour la représentation en courbes de niveaux. Plus la couleur est sombre plus le pixel est masqué par le glycane pendant la trajectoire.

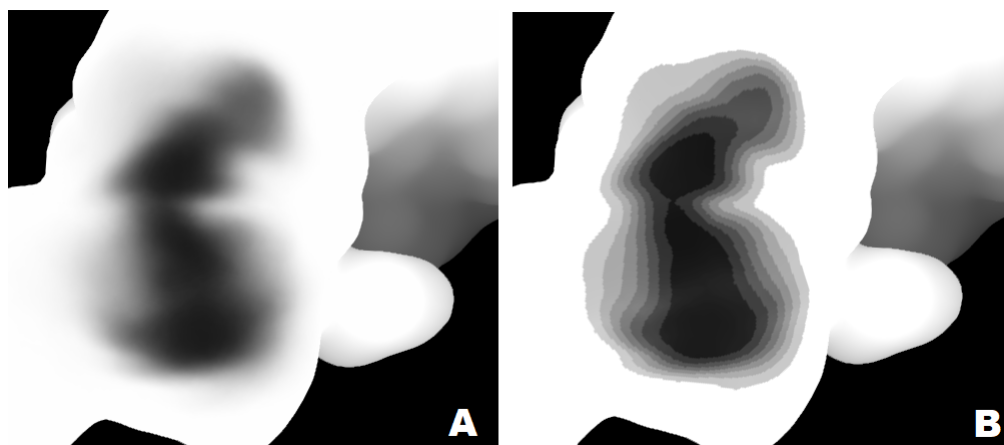


FIGURE 5.8 – **Méthodes de représentation du résultat final.** **A** : Représentation classique, les régions les moins impactées par le glycane sont assez peu visibles, contrairement à la partie centrale de l'ombre, qui représente les régions les plus souvent couvertes. **B** : Représentation en contours ou courbes de niveau : la délimitation de la zone d'influence du glycane apparaît bien plus nettement, même dans les parties les moins souvent masquées.

du glycane. En présentant les résultats sous forme de courbes de niveau en fonction de l'intensité de la couverture au cours de la trajectoire de dynamique moléculaire, nous espérons augmenter la lisibilité des résultats obtenus à des fins d'illustration ou de communication et améliorer l'identification des résidus impactés par la présence de glycanes à la surface de la protéine.

Enfin, il est également possible pour l'utilisateur de sauvegarder la texture finale dans un fichier jpeg séparé : les données de la texture finale peuvent ainsi être sauvegardées très facilement. En complément de l'information structurale, l'information en deux dimensions reste donc facilement disponible et peut être utilisée, à des fins de comparaisons entre différentes trajectoires par exemple.

Performances de l'outil d'analyse implémenté

Afin d'évaluer le temps de calcul nécessaire à l'obtention des résultats, nous avons testé l'*Umbrella Visualization* sur trois trajectoires de dynamique moléculaire du récepteur à l'insuline glycosylé. Nous avons utilisé une texture de 512 x 512 pixels et une structure de départ (protéine + glycanes) comportant 26 030 atomes. Nous avons évalué le temps de calcul pour un glycane sur une trajectoire comportant 250, 2 500 ou 10 000 clichés.

La figure 5.9 présente les résultats obtenus sur un glycane pour ces trois versions de la trajectoire. La lecture et l'affichage du résultat a respectivement pris 30 secondes, 5 minutes et 25 minutes. Bien qu'utiliser une texture plus petite (256 x 256 pixels) nous ait permis de gagner un peu de temps (18 minutes pour 10 000 clichés, soit un gain de temps de 28%), la principale limite

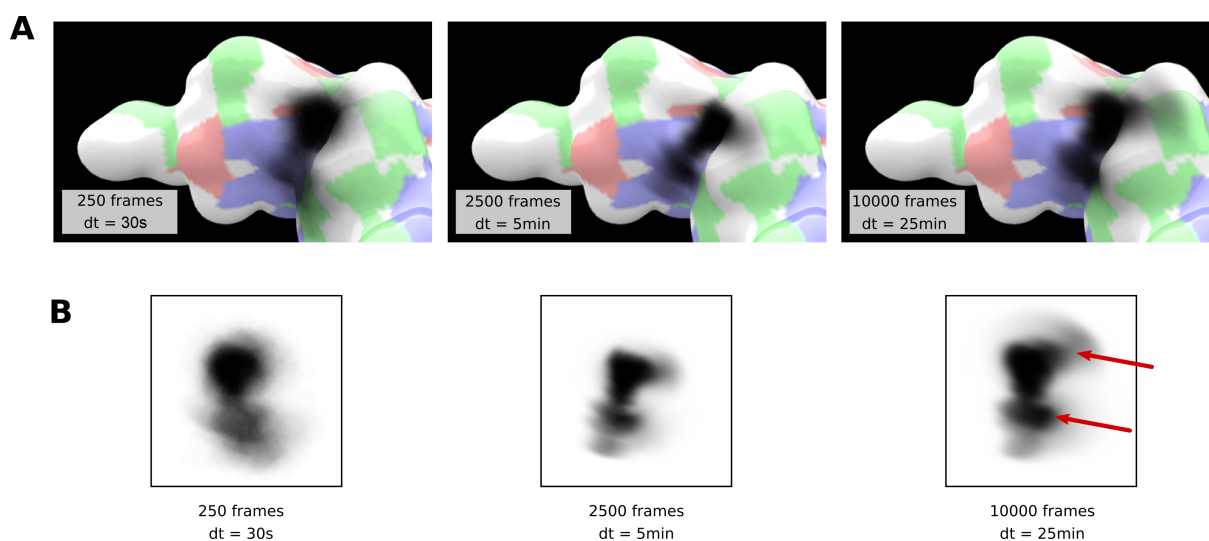


FIGURE 5.9 – Résultats de l'*Umbrella Visualization* en fonction du nombre d'images combinées. **A** : Résultats obtenus pour les différentes trajectoires. La texture projetée apparaît en gris et noir sur la surface de la protéine. Les zones bleues sont constituées de résidus basiques. Les régions rouges sont constituées de résidus acides. Les régions vertes sont constituées de résidus polaires et les régions blanches, de résidus apolaires. **B** : Textures projetées sur la surface plane correspondante. Les flèches rouges mettent en évidence des régions masquées par le glycane qui ne sont pas visibles avec 250 images. Les temps de calculs dt nécessaires pour la compilation et le rendu statistique sont également indiqués.

est le nombre d'atomes contenus par la structure. Ici, la protéine que nous avons utilisée comporte un grand nombre d'atomes, la lecture de la structure et de la trajectoire nécessite donc plus de ressources que pour un système plus petit. Cette méthode nous permet d'obtenir des résultats significatifs dans une échelle de temps raisonnable, en particulier si l'on considère les semaines de calculs qui peuvent être nécessaires à l'obtention des données de dynamique moléculaire. De plus, l'utilisation de l'*Umbrella Visualization* ne nécessite pas le recours à un supercalculateur et peut être réalisée sur un ordinateur classique.

Ces trois résultats nous permettent également de mettre en exergue l'importance du choix du nombre de clichés sur la trajectoire à analyser : bien que le résultat obtenu sur 250 clichés (figure 5.9 A) permette déjà de mettre en avant les plus grandes tendances conformationnelles du glycane au cours de la trajectoire, le résultat obtenu sur 2 500 clichés nous permet de voir des détails plus fins (figure 5.9 B). Finalement, le résultat obtenu sur 10 000 clichés (figure 5.9 C) affine encore le rendu visuel, et certaines zones couvertes qui n'apparaissent pas lors de l'analyse des 250 clichés sont alors mises en évidence.

5.3 Application et utilisation de l'*Umbrella Visualization* pour décrypter l'impact de la désialylation

Dans le cadre de l'étude de l'impact de la désialylation du récepteur à l'insuline, une première étude de dynamique moléculaire a été réalisée, permettant de mettre en avant la variation de flexibilité de glycanes en chaînes isolées. Ces travaux ont permis de démontrer l'impact de l'hydrolyse des acides sialiques sur les conformations préférentielles de glycanes bi- et tri-antennés. Les méthodes de clustering ont mis en évidence qu'en présence d'acides sialiques, les glycanes bi-antennés [17] (avec et sans fucose sur le cœur du glycan) adoptent préférentiellement une conformation « *Broken Wing* », avec le bras $\alpha 1-6$ du glycan replié contre le cœur. Sans acides sialiques, ces glycanes adoptent majoritairement une conformation « *Bird* », avec les branches étendues vers le solvant. L'*Umbrella Visualization* a également permis de montrer que l'hydrolyse des acides sialiques augmentent la surface explorée par les branches du glycan, traduisant ainsi une hausse de la flexibilité.

Suite à ces travaux, nous avons effectué des expériences de dynamique moléculaire sur le récepteur à l'insuline glycosylé et y avons appliqué la méthode de l'*Umbrella Visualization* présentée dans la section précédente. Nous avons également extrait des trajectoires les conformations majoritaires à l'aide de méthodes de clustering et comparé ces résultats à ceux fournis par l'*Umbrella Visualization* appliquée à 12 500 clichés issus de la trajectoire. Dans chacune des simulations, le récepteur à l'insuline porte deux glycanes ; les glycanes bi-antennés sont appelés Gb1 et Gb2 tandis que les glycanes tri-antennés sont appelés Gt1 et Gt2.

5.3.1 Comparaison de l'*Umbrella Visualization* avec d'autres méthodes

L'*Umbrella Visualization* étant une méthode statistique qui permet de mettre en avant les zones les plus couvertes dans la trajectoire, nous avons comparé les clusters majoritaires des trajectoires à l'ombre obtenue sur la durée de toute la trajectoire. Dans tous les cas, le cluster représentant la conformation la plus représentée correspond à la forme de l'ombre projetée sur la surface de la protéine. L'information obtenue par l'*Umbrella Visualization* étant une information statistique qui met en avant les conformations les plus représentées, ce résultat est attendu. Cependant, cette méthode permet également de prendre en compte les conformations moins représentées et les transitions entre les conformations, comme mis en évidence sur la figure 5.10.

Sur deux des trajectoires (récepteur + glycanes bi-antennés fucosylés et récepteur + glycanes tri-antennés fucosylés), l'ombre projetée du glycan Gx1 reste assez ramassée et presque circulaire (figure 5.11 A et B). Ce résultat peut s'expliquer par le fait que, au cours de la trajectoire, le

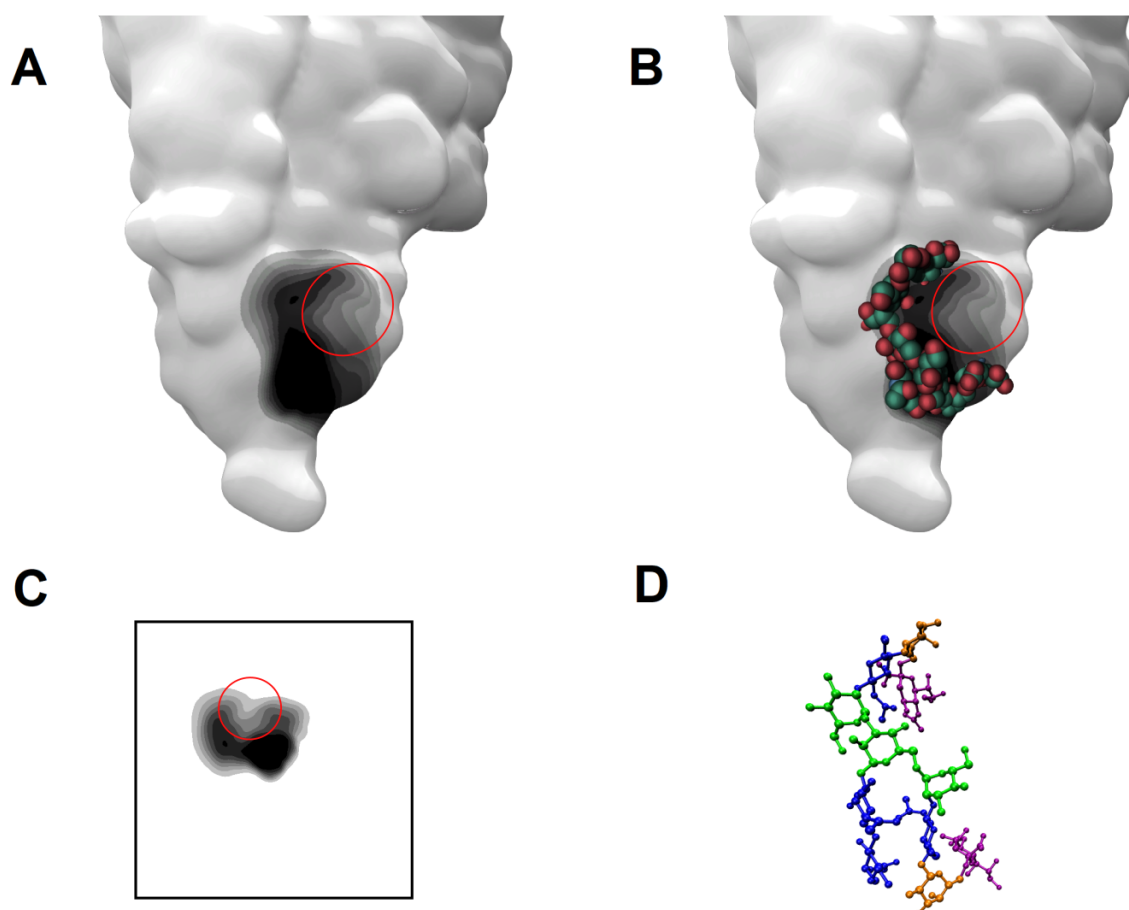


FIGURE 5.10 – **Comparaison entre les résultats de clustering et l'*Umbrella Visualization*.** Sur les panels **A** et **B**, le récepteur à l'insuline est représenté par la surface blanche sur laquelle est projeté le résultat de la compilation des ombres extraites de la trajectoire de dynamique moléculaire d'un glycane bi-antenné en présence d'acides sialiques. Le cercle rouge entoure la région de l'ombre projetée générée par les conformations intermédiaires ou minoritaires. **A** : Projection sur la surface de la protéine, le glycane n'est pas affiché. **B** : Projection sur la surface de la protéine, le glycane est ici représenté en sphères de van der Waals (atomes d'hydrogène masqués). Sa conformation correspond à celle de la conformation majoritaire issue des résultats du clustering. **C** : Texture projetée sur la surface de la protéine. **D** : Structure du glycane issue du cluster majoritaire en représentation boules-bâtons. Les N-acétylglucosamines sont en bleu, les mannoses en vert, les galactoses en jaune et les acides sialiques en violet, selon la nomenclature SNFG.

glycane reste étendu vers le solvant et ne se couche pas sur la surface de la protéine. Cette hypothèse est d'ailleurs confirmée par les résultats du clustering, qui montrent que le noyau du glycane est bien orienté vers le solvant (figure 5.11). La forme de l'ombre projetée et son étendue réduite sont, dans ce cas, un indicateur de l'orientation du glycane par rapport à la surface de la protéine et au solvant.

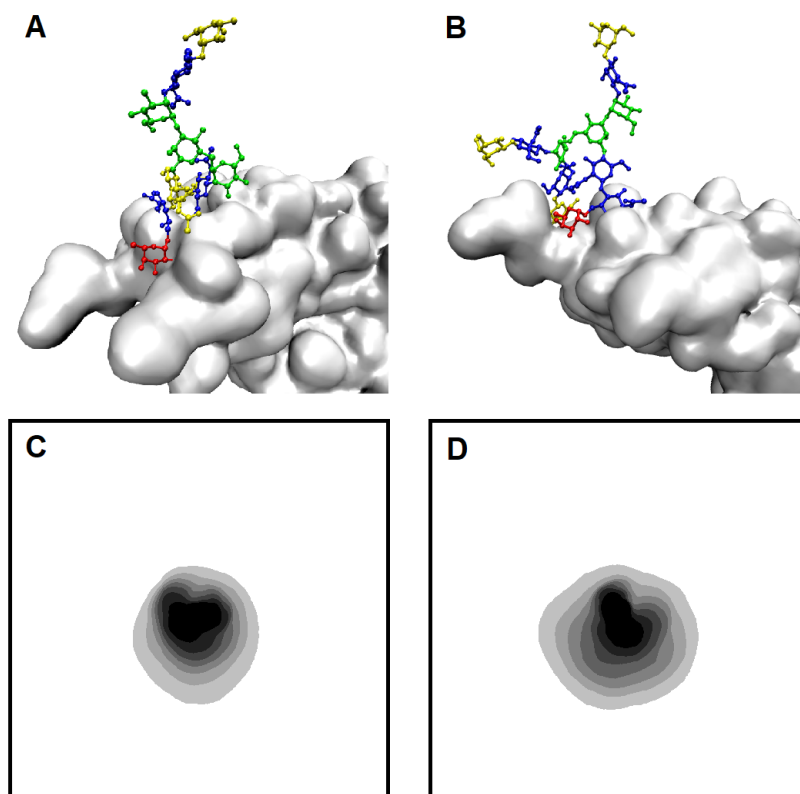


FIGURE 5.11 – Résultats du clustering des glycanes Gb1 et Gt1. **A** et **B** : Conformation principale des glycanes Gx1 bi- et tri-antenné avec fucose. Dans les deux cas, la branche α 1-3 est étendue vers le solvant. Les N-acétylglucosamines sont en bleu, les mannoses en vert et les galactoses en jaune, selon la nomenclature SNFG. **C** et **D** : Ombre compilée pendant la trajectoire de dynamique moléculaire pour le glycane Gx1 bi- et tri-antenné avec fucose. La forme globale presque circulaire de l'ombre observée sur les deux types de glycanes traduit l'orientation des chaînes de glycosylation vers le solvant.

5.3.2 Comportement des glycanes bi-antennés

Nos résultats de clustering montrent que les glycanes bi-antennés adoptent principalement une conformation en *Broken Wing*, à l'exception du glycane Gb1 sur la structure sans acides sialiques ni fucose qui adopte une conformation plus étendue (conformation *Bird*, figure 5.12)

Cette différence entre les conformations préférentielles pourrait être liée à la présence de la

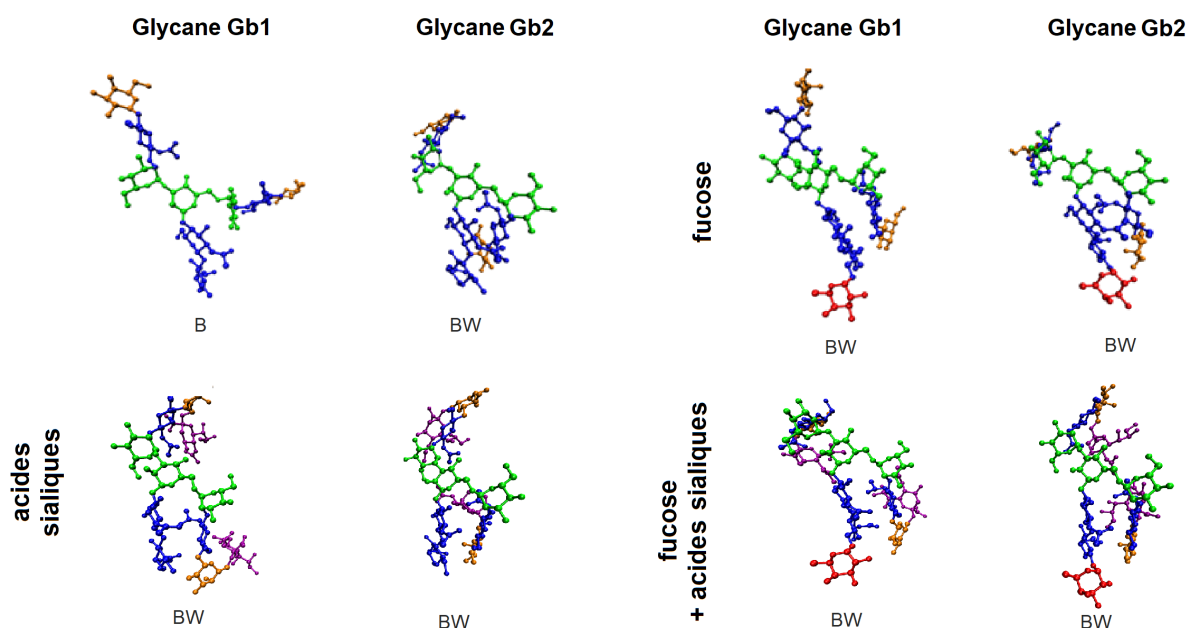


FIGURE 5.12 – **Conformations majoritaires des glycanes bi-antennés.** Tous les glycanes semblent adopter majoritairement une conformation *Broken Wing* (BW), sauf le glycane Gb1 de la structure sans fucose ni acides sialiques qui adopte une conformation avec les deux branches étendues, ou conformation *Bird* (B). Les N-acétylglucosamines sont en bleu, les mannoses en vert, les galactoses en jaune et les acides sialiques en violet, selon la nomenclature SNFG.

protéine qui favorise certaines conformations au dépend des autres. En effet, sur la plupart des trajectoires, le glycane est couché le long de la surface de la protéine, ce qui peut contraindre fortement les mouvements de sa chaîne. La position du glycane peut être liée à la formation de liaisons hydrogène avec un glutamate située près du cœur du glycane. De la même façon, plusieurs aspartates plus éloignés du noyau semblent pouvoir établir des liaisons hydrogènes avec les branches du glycane. Ces acides aminés, situés autour du site de glycosylation, pourraient ainsi favoriser certaines conformations du glycane, similaires à celles observées sur la figure 5.13 A et B. Ces exemples de structure correspondent au premier cluster obtenu pour les trajectoires de dynamique moléculaire.

En terme de flexibilité, la plupart des glycanes semblent adopter plusieurs conformations stables dans le cas des structures de bi-antennés avec fucose (figure 5.14). En l'absence d'acides sialiques, la fluctuation du RMSD pour le glycane Gb1 semble avoir une amplitude plus faible qu'en présence d'acides sialiques (figure 5.14 1A et 1B). Ainsi l'ombre projetée obtenue par l'*Umbrella Visualization* semble plus nette et la zone d'intensité maximale (parties les plus sombres) de plus grande surface (figure 5.14 2A et 2B). En effet, pour un glycane très flexible explorant un vaste espace, la zone de forte couverture est restreinte à proximité du site de gly-

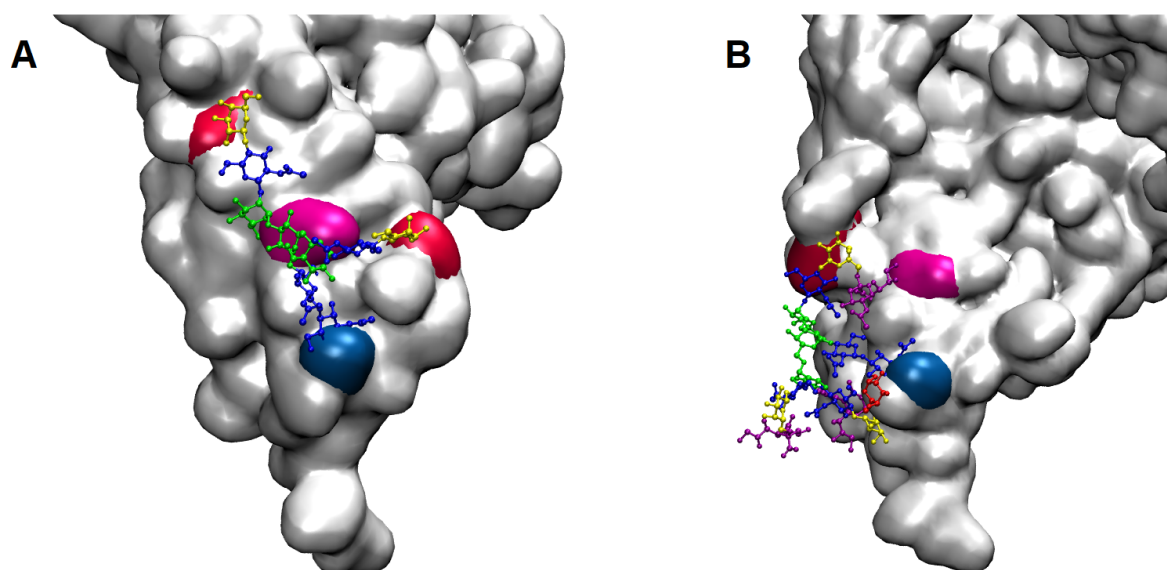


FIGURE 5.13 – **Présence de résidus acides accepteurs de liaisons hydrogène à proximité du site de glycosylation et pouvant influencer la conformation des glycanes.** Le glutamate situé à proximité du site de glycosylation est coloré en magenta. Les aspartates interagissant avec les branches des glycanes sont colorés en rouge. L'asparagine glycosylée est colorée en bleu. Le panel **A** présente un exemple de conformation d'un glycane bi-antenné et le panel **B** un exemple de conformation d'un glycane tri-antenné, sialylé et fucosylé. Ces exemples sont issus des résultats de clustering. Les N-acétylglucosamines sont en bleu, les mannoses en vert, les galactoses en jaune et les acides sialiques en violet, selon la nomenclature SNFG.

cosylation tandis que les zones de faibles couvertures seront plus étendues. Dans le cas d'un glycan stable, l'espace exploré étant faible, on s'attendra alors à ce que la zone de forte couverture soit bien plus étendue et reflète cette conformation stable. De manière similaire, le glycan Gb2 semble osciller entre deux conformations lorsque les acides sialiques ne sont pas présents (figure 5.14 C) mais reste stable tout au long de la trajectoire lorsqu'ils sont présents (figure 5.14 D).

Sur les trajectoires de glycan bi-antenné sans fucose ni acides sialiques, l'ombre projetée ainsi que les fluctuations du RMSD montre qu'une conformation particulière prédomine (figure 5.15 1A et 2A pour le glycan Gb1 et figure 5.15 1C et 2C pour le glycan Gb2). Pour le glycan Gb1 sans acides sialiques, le glycan adopte aux environs des 140 ns la conformation la plus représentée pendant la trajectoire. Pour le glycan Gb2 (figure 5.15 A), la conformation la plus représentée est présente entre 75 et 215 ns (figure 5.15 1C).

En présence d'acides sialiques, les glycanes semblent gagner en flexibilité : le glycan Gb1 en particulier, oscille entre plusieurs conformations stables, séparées par des états de transitions plus flexibles (figure 5.15 1B). Le glycan Gb2 semble stable jusqu'à environ 125 ns puis montre une plus grande flexibilité sur toute la fin de la trajectoire (figure 5.15 1D), ce qui peut traduire une plus grande flexibilité. L'ombre projetée est ainsi plus étendue, avec des zones de forte couvertures moins étalées (5.15 2D).

5.3.3 Comportement des glycanes tri-antennés

Les résultats du clustering montrent que la conformation principale des glycanes tri-antennés présente, dans la plupart des cas, une structure qui pourrait se rapprocher de la conformation *Bird* des glycanes bi-antennés, avec notamment deux branches étendues loin du noyau du glycan. Les deux glycanes portant à la fois des acides sialiques et un fucose diffèrent et ont une structure plus ramassée, avec les branches regroupées autour du noyau (figure 5.16).

Dans le cas des glycanes tri-antennés fucosylés, on observe sur les courbes de RMSD que le glycan Gt1 semble plus flexible en l'absence d'acides sialiques, (passage par plusieurs conformations différentes), qu'en leur présence, où les fluctuations sont moins importantes et une conformation majoritaire semble prédominer au cours de la trajectoire (figure 5.17 1A et 1B). Cela se traduit par une ombre projetée sur la protéine plus étendue (figure 5.17 2B). Le glycan Gt2 montre un comportement similaire, avec des fluctuations de plus forte intensité en l'absence d'acides sialiques et une stabilisation du glycan autour des 100 ns en présence d'acides sialiques (respectivement, figure 5.17 1C et 1D). Les ombres projetées sont ainsi similaires au cas du glycan Gt1, avec une ombre présentant des zones de forte intensité réduite sans acides sialiques contre une

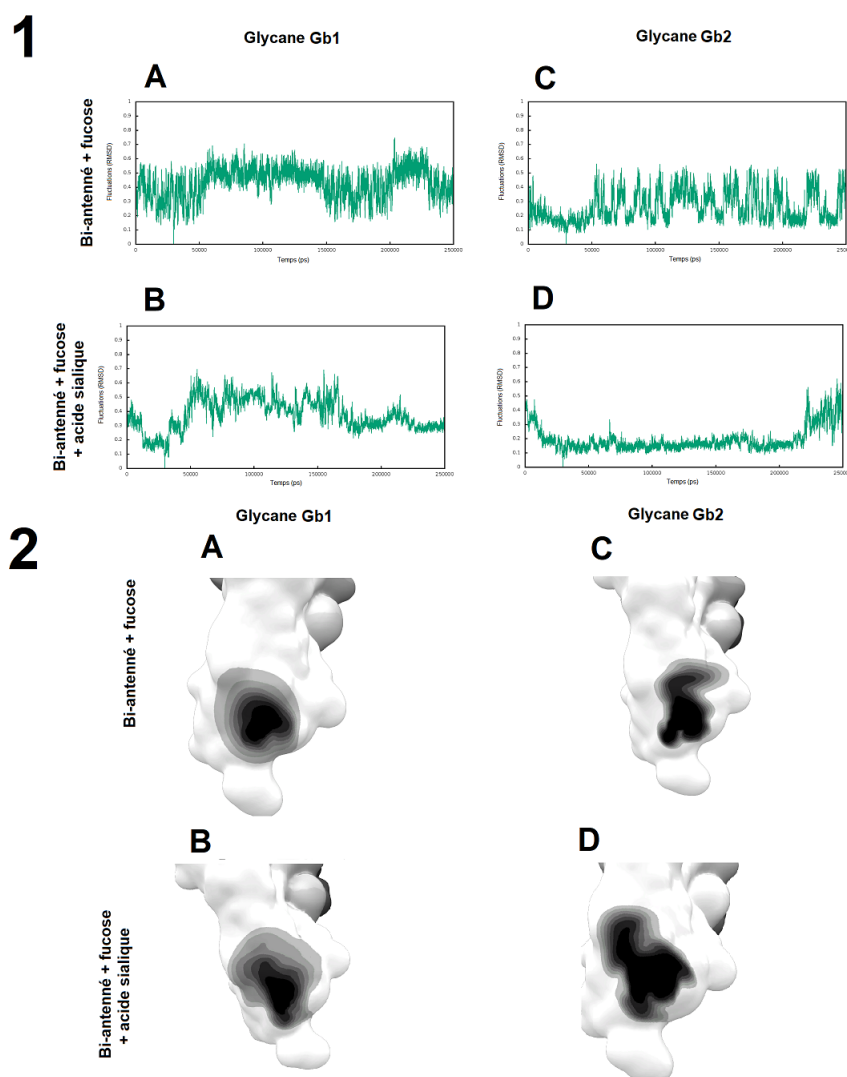


FIGURE 5.14 – **Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l'Umbrella Visualization pour les deux glycanes bi-antennés fucosylés sans et avec acides sialiques.** **1A et 1B** : Fluctuations du glycanes Gb1 fucosylé sans et avec acides sialiques, respectivement. Les fluctuations du glycanes Gb1 sont réduites en présence d'acides sialiques. **1C et 1D** : Fluctuations du glycanes Gb2 fucosylé sans et avec acides sialiques, respectivement. Sans acides sialiques, le glycanes oscille entre deux conformations. En présence d'acides sialiques, le glycanes reste stable tout au long de la trajectoire. **2A et 2B** : Résultats obtenus avec l'Umbrella Visualization pour le glycanes Gb1 fucosylé sans et avec acides sialiques, respectivement. La diminution de l'amplitude des fluctuations se traduit par une zone de forte intensité plus étendue. **2C et 2D** : Résultats obtenus avec l'Umbrella Visualization pour le glycanes Gb2 fucosylé sans et avec acides sialiques, respectivement. De manière similaire, la stabilisation de la conformation du glycanes en présence d'acides sialiques entraîne une augmentation de la surface de la zone de forte intensité.

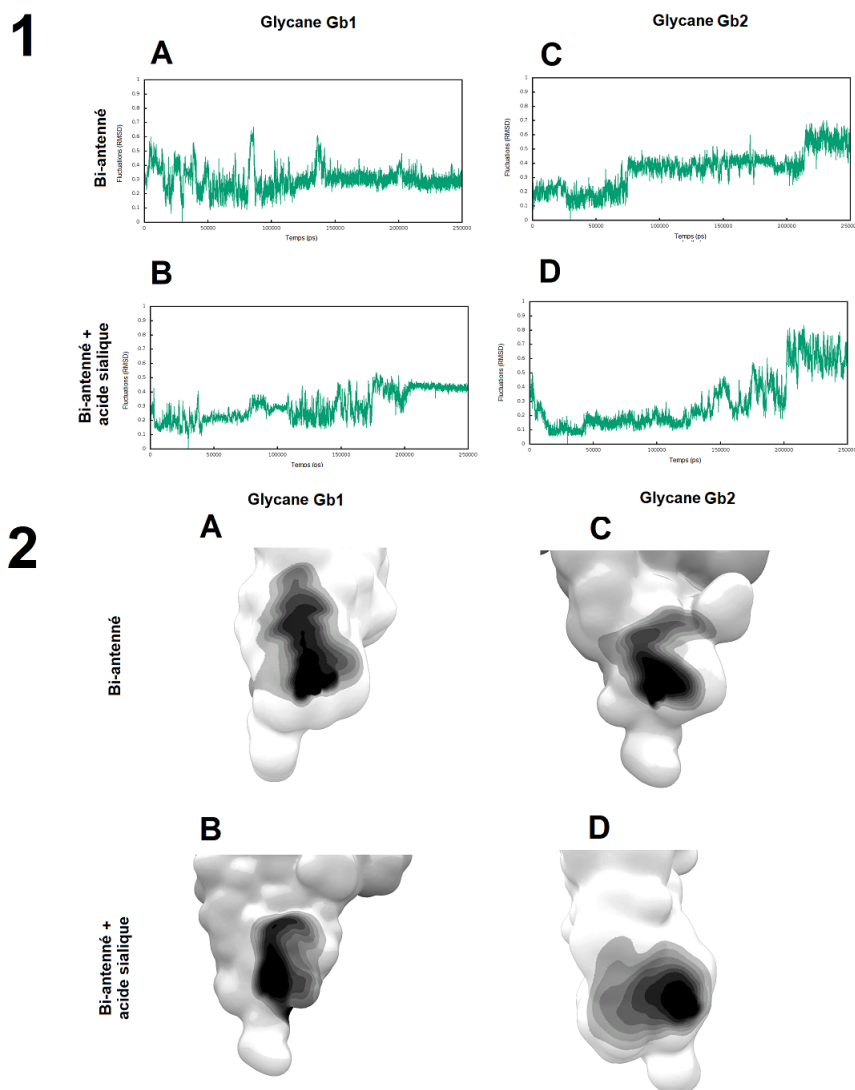


FIGURE 5.15 – Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l'*Umbrella Visualization* pour les deux glycanes bi-antennés non-fucosylés sans et avec acides sialiques. **1A et 1B** : Fluctuations du RMSD du glycanes Gb1 sans et avec acides sialiques, respectivement. Sans acides sialiques, le glycanes Gb1 présente une conformation principale stabilisée à environ 140 ns. Avec acides sialiques, le glycanes oscille entre plusieurs conformations. **1C et 1D** : Fluctuations du RMSD du glycanes Gb2 sans et avec acides sialiques, respectivement. Sans acides sialiques, le glycanes Gb2 oscille entre plusieurs états stables, la conformation majoritaire étant située en 75 et 215 ns. Avec acides sialiques, le glycanes est stable au long de la trajectoire mais se déstabilise à partir d'environ 240 ns. **2A et 2B** : Résultats obtenus avec l'*Umbrella Visualization* pour le glycanes Gb1 sans et avec acides sialiques, respectivement. Dans les deux cas, les zones de forte intensité reflètent les conformations majoritaires adoptées par le glycanes. **2C et 2D** : Résultats obtenus avec l'*Umbrella Visualization* pour le glycanes Gb2 sans et avec acides sialiques, respectivement. Sans acides sialiques, la conformation majoritaire est clairement délimitée par la zone de forte intensité. En présence d'acides sialiques, la zone couverte est plus dispersée, reflétant la perte de flexibilité.

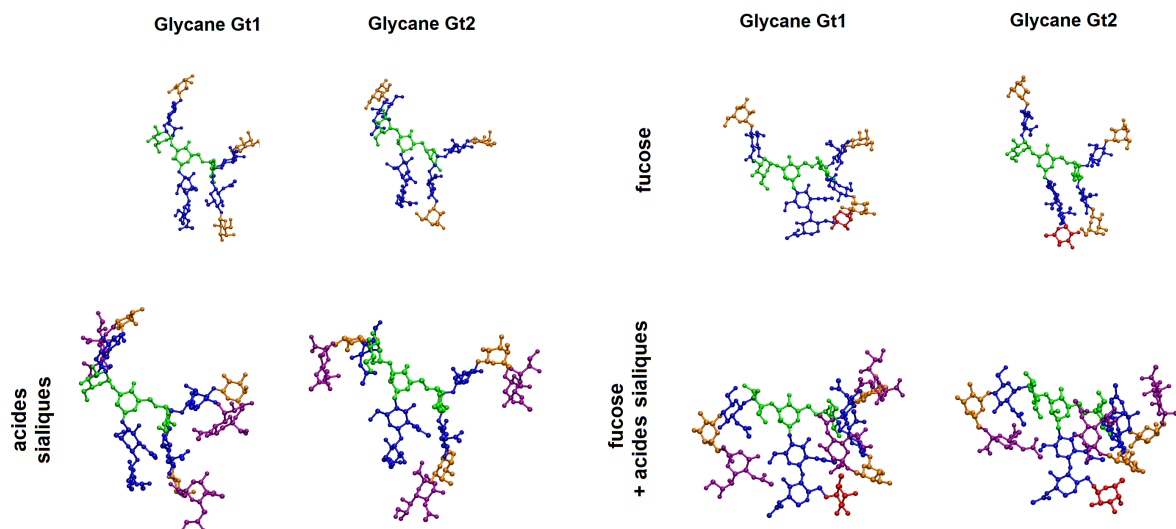


FIGURE 5.16 – **Conformations majoritaires des glycanes tri-antennés.** Les N-acétylglucosamines sont en bleu, les mannoses en vert, les galactoses en jaune et les acides sialiques en violet, selon la nomenclature SNFG.

ombre plus intense traduisant le mouvement réduit des glycanes en présence d'acides sialiques (respectivement, figure 5.17 2C et 2D).

En l'absence de fucose, les ombres projetées obtenues par application de l'*Umbrella Visualization* semblent plus étendue et moins intenses en l'absence d'acides sialiques. Pour le glycan Gt1, on observe plusieurs transitions rapides sur la courbe de RMSD (figure 5.18 1A), suggérant des oscillations entre deux conformations en l'absence d'acides sialiques. En présence d'acides sialiques, ces transitions sont moins rapides et l'amplitude des variations est plus faible (figure 5.18 1B). Pour le glycan Gt2, il semble toujours osciller entre plusieurs états en l'absence d'acides sialiques (figure 5.18 1C). En présence d'acides sialiques, la chaîne de glycosylation très flexible se stabilise dès 150 ns au cours de la trajectoire (figure 5.18 1D). Ceci se traduit donc par des ombres aux zones très sombres plus étendues en présence d'acides sialiques (figure 5.18 2B et 2D) qu'en leur absence (figure 5.18 2A et 2C).

Ces résultats montrent un impact des acides sialiques sur la flexibilité des glycanes bi- et tri-antennés. Leur présence semble stabiliser la conformation des glycanes sur la protéine, conduisant à l'apparition de zones de couvertures plus denses grâce à l'*Umbrella Visualization*. Cependant, sur les glycanes bi-antennés en absence de fucose, il semblerait que la présence des acides sialiques tende à augmenter la flexibilité des chaînes de glycanes. Pour les glycanes tri-antennés, les résultats de clustering suggèrent que la présence simultanée d'acides sialiques et de fucose entraîne la présence de conformations plus compactes, avec des branches plus regroupées autour

du cœur du glycanes. D'autres études ont montré que le fucose avait une influence sur la structure des glycanes [107,113]. Il est donc possible que ce saccharide puisse jouer un rôle supplémentaire dans la flexibilité des glycanes.

Il est également intéressant de noter que, sur deux trajectoires (bi-antenné et tri-antenné sans fucose ni acides sialiques), le glycanes Gx1 (x étant b ou t) est étendu vers le solvant. Que cette conformation n'apparaisse pas sur les autres trajectoires suggère que la présence d'acides sialiques et / ou du fucose puisse contraindre le glycanes à rester penché sur la protéine.

De manière générale, les résultats obtenus par application de l'*Umbrella Visualization* et de méthodes plus classiques comme les méthodes de clustering ou les mesures de RMSD se complètent efficacement et ont permis de mettre en évidence les variations de flexibilité en fonction de la composition des glycanes. Nous avons également pu voir que, dans le cas de chaînes de glycosylations tournées vers le solvant, la forme de l'ombre projetée est à la fois plus réduite et circulaire que dans les autres cas, et la zone de forte couverture est de surface plus réduite. Cette indication peut être importante, étant donné le rôle des glycanes comme sites de liaisons pour d'autres protéines telles que les lectines. L'information statistique apportée par l'application de l'*Umbrella Visualization* montre une bonne corrélation avec les méthodes de clustering et les mesures de RMSD en montrant l'impact des conformations majoritaires (zones les plus couvertes donc plus sombres) et les états transitoires ou conformations mineures (zones plus claires car moins couvertes). De plus, par son intégration à un logiciel de visualisation moléculaire, l'*Umbrella Visualization* procure également une information visuelle claire et rapidement compréhensible, aidant à l'interprétation des données.

Nous avons également pu voir que des acides aminés accepteurs de liaisons hydrogènes pourraient avoir un rôle stabilisateur des conformations préférentielles et influencer ainsi les zones couvertes par le glycanes pendant la trajectoire de dynamique moléculaire.

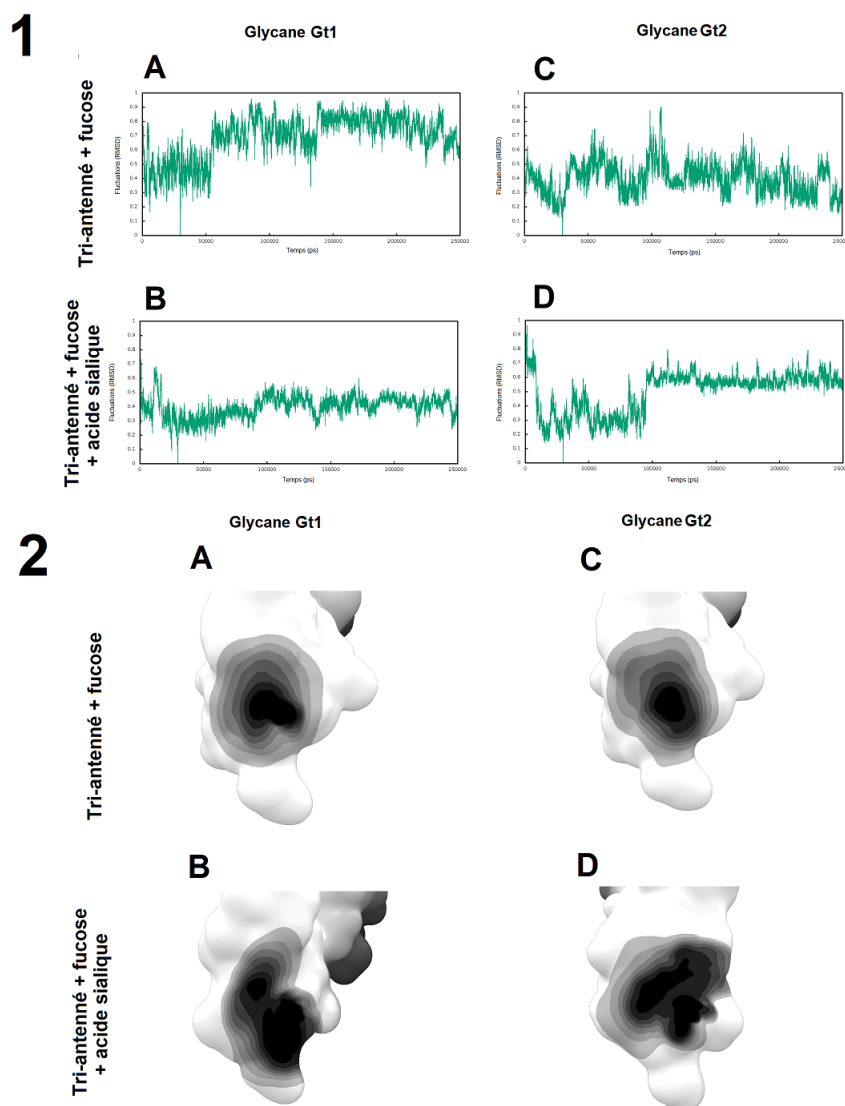


FIGURE 5.17 – **Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l'Umbrella Visualization pour les deux glycanes tri-antennés fucosylés sans et avec acides sialiques.** **1A et 1B** : Fluctuations du RMSD du glycanes Gt1 fucosylé sans et avec acides sialiques, respectivement. Sans acides sialiques, le glycanes passe par plusieurs conformations. Avec acides sialiques, l'amplitude des fluctuations diminue et le glycanes reste stable au cours de la trajectoire. **1C et 1D** : Fluctuations du RMSD du glycanes Gt2 fucosylé sans et avec acides sialiques, respectivement. Le comportement du glycanes Gt2 est similaire à celui du glycanes Gt1, avec des fluctuations entre plusieurs états sans acides sialiques et une stabilisation en présence d'acides sialiques. **2A et 2B** : Résultats obtenus avec l'Umbrella Visualization pour le glycanes Gt1 fucosylé sans et avec acides sialiques, respectivement. La stabilisation du glycanes en présence d'acides sialiques se traduit par une zone de forte intensité plus étendue que sans acides sialiques. **2C et 2D** : Résultats obtenus avec l'Umbrella Visualization pour le glycanes Gt2 fucosylé sans et avec acides sialiques, respectivement. De manière similaire, la stabilisation du glycanes est visible sur le résultat de l'Umbrella Visualization.

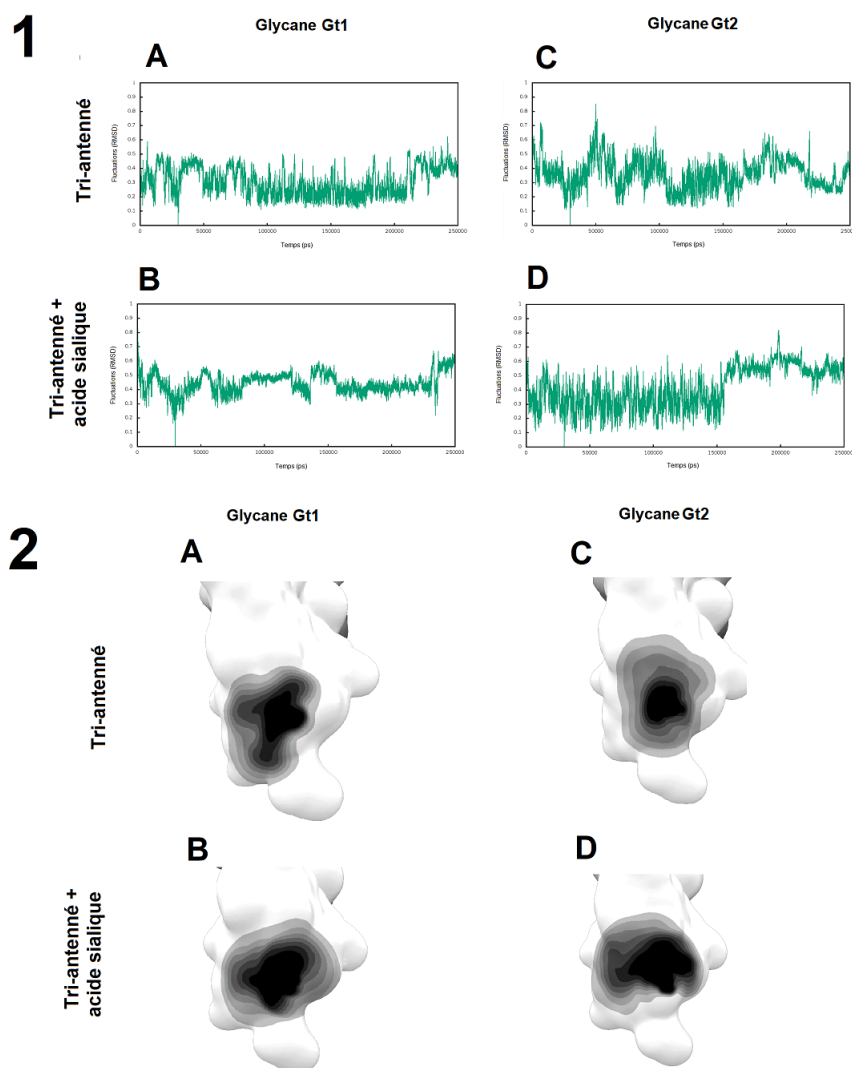


FIGURE 5.18 – Evaluation temporelle du RMSD des glycanes et résultats obtenus par application de l'*Umbrella Visualization* pour les deux glycanes tri-antennés non-fucosylés sans et avec acides sialiques. **1A et 1B** : Fluctuations du RMSD du glycanes Gt1 sans et avec acides sialiques, respectivement. Sans acides sialiques, le glycanes Gt1 effectue beaucoup de transitions entre deux états. Avec les acides sialiques, le glycanes est plus stable. **1C et 1D** : Fluctuations du RMSD du glycanes Gt2 sans et avec acides sialiques, respectivement. Sans acides sialiques, le glycanes oscille entre plusieurs états. Avec acides sialiques, il se stabilise fortement à environ 150 ns. **2A et 2B** : Résultats obtenus avec l'*Umbrella Visualization* pour le glycanes Gt1 sans et avec acides sialiques, respectivement. La zone de forte couverture est plus étendue en présence d'acides sialiques, traduisant la relative stabilisation du glycanes. **2C et 2D** : Résultats obtenus avec l'*Umbrella Visualization* pour le glycanes Gt2 sans et avec acides sialiques, respectivement. L'ombre projetée sans acides sialiques a une forme relativement circulaire qui traduit la flexibilité du glycanes. Avec acides sialiques, la zone de forte intensité est plus étendue, correspondant à un glycanes plus stable.

Chapitre 6

Discussion et conclusion

6.1 Caractérisation de la flexibilité et de la dynamique de chaînes sucrées

6.1.1 Propriétés intrinsèques de glycanes isolés

Dans la continuité des travaux menés précédemment au sein de l'unité MEDyC, nous avons poursuivi la caractérisation de l'influence des acides sialiques sur la dynamique et la flexibilité des glycanes à travers la mise en place et l'analyse de trajectoires de dynamique moléculaire appliquées à différentes chaînes de glycanes isolées, à savoir deux glycanes bisectants et trois glycanes tétra-antennés.

Les résultats obtenus sur les glycanes bisectants sont en accord avec les études précédentes ayant montré que le bras $\alpha 1-3$ est contraint par la N-acétyl glucosamine bisectante. Ces glycanes ayant largement été étudiés, nous n'avons généré que 500 ns de trajectoire. En effet, Harbison et al [107] ont généré plus de 60 μs de trajectoire sur ces glycanes et ont montré que les acides sialiques n'affectent pas la conformation de ces derniers. Les résultats que nous avons obtenus ont suggéré au contraire un impact des acides sialiques sur la flexibilité du bras $\alpha 1-6$. Cependant, considérant la faible durée de nos trajectoires, il est probable que nous n'ayons accès qu'à un échantillonnage incomplet, expliquant ainsi ces différences avec les travaux cités.

Outre l'effet de taille associé aux glycanes tétra-antennés, nous avons également testé deux types de liaisons différentes des acides sialiques : des liaisons en $\alpha 2-3$ et des liaisons en $\alpha 2-6$. L'analyse des trajectoires associées aux glycanes tétra-antennés en présence d'acides sialiques a permis de mettre en évidence deux types de conformations et dynamiques différentes selon la nature de la liaison de ces derniers. Alors que les acides sialiques liés en $\alpha 2-3$ restent dans

le prolongement de la branche du glycane, les acides sialiques liés en α 2-6 se replient le long de la branche qui les porte. La liaison α 2-6 est plus longue, ce qui permet un tel repliement. Or, plusieurs études ont montré le changement de ratio entre ces deux types de liaison dans le cadre de pathologies telles que les cancers ou le diabète de type 2 [108, 110, 114]. La différence de conformation au niveau des acides sialiques implique très certainement un déplacement de la charge portée par ces monosaccharides et un encombrement stérique différent ; ces éléments pourraient perturber les fonctions de la protéine et/ou son interaction avec ses partenaires. La méthode en 2D de l'*Umbrella Visualization* ne prend en compte que le centre de masse du dernier résidu commun entre les différentes chaînes avec ou sans acides sialiques (ici un galactose), c'est pourquoi nos résultats ne permettent pas de voir si la surface couverte par les acides sialiques varie en fonction de la nature de la liaison des types d'acides sialiques.

Comme présenté dans l'introduction de ce manuscrit, les travaux qui ont précédés et motivés cette étude se sont principalement intéressés au récepteur à l'insuline et au diabète. C'est aussi dans le cadre du diabète de type 2 qu'une étude de Dotz et al. [110] a montré que le ratio entre les deux types de liaisons d'acides sialiques dans le glycome du plasma est modifié, avec une augmentation de glycanes présentant la liaison α 2-6 et une diminution de glycanes présentant la liaison α 2-3. En considérant à la fois le volume des acides sialiques et leur charge, il semble raisonnable d'imaginer que la nature de la liaison influence les conformations préférentielles. L'évaluation de l'énergie libre de la structure pourrait nous apporter de nouvelles informations sur l'influence structurale de ces deux types de liaison des acides sialiques. La mise en œuvre de la nouvelle méthode de l'*Umbrella Visualization* en 3D sur de tels glycanes placés à la surface du récepteur à l'insuline ou d'autres protéines d'intérêt pourrait également nous permettre de visualiser et décrypter l'impact de ces saccharides sur la surface protéique couverte en fonction de leur liaison au glycane. Ainsi, étudier l'impact de ces liaisons pourrait aider à comprendre leur rôle dans la régulation des fonctions protéiques dans le cadre de différentes pathologies.

Les résultats issus du clustering et des analyses par *Umbrella Visualization* 2D ont montré un impact réduit des deux types d'acides sialiques sur la dynamique et la conformation des glycanes tétra-antennés. Cependant, nous avons tout de même observé une variation de la valeur de l'angle omega de la liaison α 1-6 sur une des trajectoires des glycanes avec acides sialiques liés en α 2-3. Ceci soulève la question du rôle de ces derniers dans le processus de stabilisation de nouvelles conformations. Par ailleurs, ayant déjà observé sur les glycanes bi-antennés et tri-antennés un impact variable des acides sialiques [17], il est possible que ces derniers aient peu ou pas d'effet sur les structures tétra-antennées. Il est également possible qu'en raison de leur grande flexibilité, ces structures ne puissent être correctement décrites par 1 500 ns de trajectoire qui sont alors

insuffisantes pour accéder à l'ensemble de l'espace exploré par ces glycanes. Ainsi, des études supplémentaires devront être menées pour améliorer l'échantillonnage conformationnel et ainsi affiner les résultats relatifs à l'impact des acides sialiques dans la flexibilité et la dynamique des glycanes.

6.1.2 Étude de l'interaction réciproque protéine / glycane

La trajectoire de dynamique moléculaire qui décrit la fibromoduline glycosylée nous a permis d'observer le repliement de deux des quatre chaînes de keratan sulfate portées par la protéine. Bien que ces chaînes couvrent une large surface sur la partie convexe de la protéine, les points de contact direct avec la protéine sont localisés principalement autour des groupements sulfates et l'une des chaînes se replie sur le motif LRR7, qui est l'un des sites de liaison au collagène.

On sait que les N-glycanes ramifiés interagissent avec des acides aminés proches du point de glycosylation [113,115] : dans le cadre de cette étude, nous avons pu voir que les GAG sont capables d'interagir avec des acides aminés à bien plus grande distance. Ainsi, la disposition des points de contact loin du point de glycosylation ainsi que la longueur de la chaîne de glycosylation ne nous ont pas permis d'utiliser l'*Umbrella Visualization* telle que nous l'avons développée au cours de ce travail de thèse. D'autres méthodes doivent donc être développées pour s'adapter aux spécificités des structures de grande taille telles que les GAG. Une piste intéressante à suivre pourra consister en la mesure du volume exclu à la surface de la protéine lorsque la chaîne de glycosaminoglycane se replie.

Bien que la trajectoire soit relativement courte en regard de la flexibilité des chaînes de keratan sulfate, ces premiers travaux exploratoires nous permettent d'envisager des pistes d'études intéressantes et relatives aux protéoglycanes de la MEC. Outre l'étude de la sulfatation ou la modification des conditions ioniques pour refléter les conditions cellulaires, l'utilisation de méthodes comme l'échange de répliques permettant d'obtenir un échantillonnage optimisé de structures s'avère nécessaire et incontournable. Cela constitue malgré tout un matériel de départ nous permettant des développements dans le domaine du graphisme moléculaire.

Plus généralement, les résultats de cette étude pourront servir de base afin d'étendre nos travaux au lumican. Ce dernier, un autre SLRP présent dans la MEC, appartient à la même classe de SLRP que la fibromoduline et possède lui aussi 4 sites de glycosylation. Le lumican peut exister sous 4 états de glycosylation différents : un cœur protéique non glycosylé, un cœur protéique portant des oligosaccharides, un cœur protéique portant des polylactosamine (keratan

sulfate sans groupements sulfates), ou un cœur protéique portant des keratanes sulfates [94, 116]. Dans le cadre des travaux de l'unité, le lumicane est étudié pour son rôle dans la progression tumorale [28]. En raison des nombreuses similarités entre la fibromoduline et cette protéine, les résultats obtenus sur la première pourront également être étendus au second lors de futures études.

6.2 Apports et perspectives offerts grâce à l'implémentation de l'*Umbrella Visualization* sous UnityMol

L'*Umbrella Visualization* a, dans un premier temps, été développée comme une méthode 2D adaptée et dédiée à l'étude de la flexibilité de glycanes bi- et tri-antennés [17]. Bien que cette méthode ait apporté des résultats originaux et pertinents, la nécessité de faire évoluer celle-ci vers une approche tridimensionnelle qui permette d'intégrer la topologie de surface des protéines s'est imposée. Nous avons ainsi traduit et transféré le principe de l'*Umbrella Visualization* sur le logiciel UnityMol à l'aide d'un plan orienté, d'une lumière, d'une caméra et d'un projecteur. Cette implémentation dans la plateforme UnityMol, nous a permis de mettre en place un outil capable d'afficher sur la surface d'une protéine l'aire couverte par toute la chaîne du glycanes au cours d'une trajectoire de dynamique moléculaire. L'information extraite conserve l'aspect statistique de la version originelle 2D de l'*Umbrella Visualization* puisqu'il est possible de compiler les ombres individuelles de sorte que les régions les plus souvent couvertes par le glycanes apparaissent plus sombres que les zones moins couvertes.

L'utilisation de Unity nous a également permis d'implémenter et utiliser un mode de représentation en courbes de niveau qui délimite clairement des zones d'intensité sur l'ombre compilée. En comparant l'aspect de l'ombre avec les structures de glycanes portés par le récepteur à l'insuline et issues d'analyses de clustering, nous avons démontré que les résultats issus des deux méthodes sont corrélés et complémentaires. Cette adéquation semble logique puisque les méthodes de clustering identifient les conformations les plus récurrentes, et qui par conséquent ont un poids statistique plus élevé, au cours de la trajectoire de dynamique moléculaire. Il convient de noter que l'*Umbrella Visualization* permet également de mettre en évidence les conformations minoritaires ou intermédiaires et d'évaluer leur impact au même titre que les conformations majoritaires, reflétant plus précisément l'aspect dynamique des glycanes. La forme de l'ombre compilée est liée au comportement du glycanes par rapport à la protéine : lorsque ce dernier est étendu vers le solvant, comme observé sur deux de nos trajectoires du récepteur à l'insuline glycosylé (figure 5.11), l'ombre finale compilée est moins étendue et de forme circulaire, traduisant l'extension du glycanes dans une direction presque perpendiculaire et loin de la surface protéique.

Afin de placer le *prefab* par rapport à la glycosylation, nous avons choisi l'asparagine portant la glycosylation comme point de repère pour placer et orienter le plan sur lequel l'ombre du glycane est projetée. Ce choix nous permet de prendre en compte l'inclinaison du glycane sur la surface de la protéine. De plus, même si l'asparagine glycosylée peut elle aussi se réorienter pendant la trajectoire, la présence du glycane nous assure que cet acide aminé sera toujours orienté vers le solvant. Cependant, ce choix pourrait mener à des approximations, liées à l'orientation variable de cet acide aminé. Il serait ainsi judicieux de proposer d'autres solutions pour placer et orienter le *prefab*, en l'alignant par exemple sur le plan tangent à la surface au point de glycosylation. Cependant, calculer l'orientation de ce plan pourrait dépendre de la topologie de la surface locale de la protéine et / ou de la taille de la surface considérée.

Les résultats obtenus au cours de nos travaux sur le récepteur à l'insuline suggèrent cependant que le point de repère choisi (asparagine glycosylée) reste relativement fiable, comme le montrent la bonne concordance entre les résultats de l'*Umbrella Visualization* et les structures obtenues à l'aide des méthodes de clustering.

Ici, la surface sur laquelle l'ombre du glycane est projetée correspond à la surface de la première structure de la trajectoire de dynamique moléculaire et celle-ci ne varie pas pendant la lecture de la trajectoire. Or, les protéines sont des objets dynamiques et la topologie de la surface ainsi que les résidus exposés au solvant peuvent varier au cours de la trajectoire de dynamique moléculaire. Ainsi, l'utilisation d'une surface statique comporte un biais qu'il faut prendre en compte au moment d'interpréter les données. Une autre possibilité serait de calculer la surface moyenne de la protéine au cours de la trajectoire, une méthode qui serait particulièrement adaptée aux protéines comme le récepteur à l'Insuline qui présentent peu de mouvements de grande ampleur entre les domaines.

L'*Umbrella Visualization*, couplée aux méthodes de clustering et aux mesures de RMSD, nous a ainsi permis de caractériser le comportement dynamique de chaînes de glycosylations isolées ou portées par des protéines et d'illustrer leur impact dans le cas particulier du récepteur à l'insuline. Le principal avantage de l'*Umbrella Visualization* est la simplicité de lecture de la méthode : l'information compilée est directement visible sur la surface de la protéine et les fonctionnalités ajoutées permettent de sauvegarder la texture projetée dans un fichier séparé et donc de comparer les distributions obtenues sur plusieurs trajectoires. De plus, l'intégration de cet outil original à l'interface de UnityMol démocratise son utilisation puisque seul le chargement des données par l'utilisateur est nécessaire. Ces développements ont fait l'objet d'un article associé à une

présentation de congrès international (BIBM 2018) [117] et d’une publication dans le journal *Methods* [118] (cf. annexes).

L’outil que nous proposons à l’issue des trois années de thèse permet d’afficher sous une forme clairement lisible la zone couverte par un glycanes au cours d’une trajectoire de dynamique moléculaire et de relier de façon triviale l’information statistique (densité des zones couvertes) à l’information de structure contenue dans la topologie de la protéine. Cette nouvelle version de l’*Umbrella Visualization* comble ainsi certains des défauts soulevés par l’utilisation de la version précédente qui générerait uniquement des graphiques de densité en deux dimensions. De plus, nous prenons en compte avec cette nouvelle version la totalité de la structure du glycanes et non plus les centres de masse des extrémités des chaînes de glycosylation.

Parmi les futurs développements possibles, nous listons ci-après ceux qui nous semblent les plus pertinents.

Développement n°1 : extension de l’*Umbrella Visualization* à plus d’un glycanes

La version actuelle de l’*Umbrella Visualization* sélectionne automatiquement uniquement la première asparagine glycosylée présente sur la protéine à la lecture du fichier de structure afin d’appliquer les calculs et d’afficher le résultat de la compilation statistique. L’utilisation d’un *Prefab* dédié permettra d’appliquer indépendamment la méthode développée sur plusieurs glycanes. Cependant, l’impact en terme de performances de calcul devra être pris en compte.

Développement n°2 : ajout d’une fonction permettant de charger des résultats obtenus précédemment

Les différents modes de représentation proposés avec l’*Umbrella Visualization* permettent de générer des images claires décrivant le comportement du glycanes au cours d’une trajectoire de dynamique moléculaire. Bien qu’une option de sauvegarde de la texture du résultat final sous forme de fichier d’image soit proposée, l’*Umbrella Visualization* ne permet pas de charger directement un résultat préalablement sauvegardé et de l’afficher sur la surface protéique. La trajectoire doit être à nouveau lue pour que le résultat soit recalculé. Il nous semble pertinent d’ajouter la possibilité de charger une texture sur un point de glycosylation choisi ; ceci simplifiera l’utilisation de la méthode et offrira un gain de temps supplémentaire à l’utilisateur.

Développement n°3 : extension de l’*Umbrella Visualization* à d’autres types de glycosylations

La liaison glycosidique entre l’asparagine et le glycane est identifiée et permet de placer le plan orienté sur lequel l’ensemble des projections est réalisé. Une extension du travail actuel pourrait consister à étendre cette méthode à d’autres types de glycosylations, comme les O-glycosylations, en ajoutant l’identification d’autres types de liaisons et résidus qui pourront être utilisés comme repères afin de placer le *Prefab*. L’*Umbrella Visualization* pourra ainsi être étendue à l’étude et la visualisation de modifications post-traductionnelles avec pour objectif l’évaluation de leur impact sur la protéine.

Dans la section suivante, nous commentons les résultats associés à l’utilisation de l’*Umbrella Visualization* dans le cadre de l’étude et la caractérisation du récepteur à l’insuline.

6.3 Désialylation du Récepteur à l’insuline : pertinence de la visualisation scientifique dans la compréhension des conséquences moléculaires associées

Les résultats obtenus sur le récepteur à l’insuline à l’aide de l’*Umbrella Visualization* ont ainsi permis de montrer la stabilisation des conformations de glycanes bi- et tri-antennés à la surface de la protéine. Un glutamate localisé à proximité du site de glycosylation semble stabiliser la position du cœur du glycane tandis que des aspartates pourraient attirer les groupes des branches. Les résultats de clustering et l’analyse des courbes de RMSD, couplés à l’*Umbrella Visualization* ont permis d’observer les différences de flexibilité entre les glycanes en fonction de leur composition et état de sialylation. Ces résultats semblent en effet confirmer l’impact différentiel des acides sialiques en fonction du nombre de branches et de l’état de sialylation du glycane.

Cependant, bien que le dimère du récepteur à l’insuline soit un homodimère, la surface couverte par les glycanes sur les deux sites homologues diffère parfois. Par exemple, les glycanes Gb1 et Gb2 sur la trajectoire des bi-antennés avec acides sialiques n’ont pas le même comportement : là où le glycane Gb1 reste stable pendant la trajectoire, le glycane Gb2 gagne en stabilité à environ 125 ns. On peut également citer les glycanes Gx1 (x étant b ou t) qui maintiennent leur orientation en direction du solvant durant la trajectoire tandis que les glycanes correspondants Gx2 se stabilisent sur la surface protéique. Au total, 250 ns de trajectoire ont été générées pour chaque type de glycanes. Considérant la flexibilité des glycanes, il est fort probable que cette durée de simulation ne permette pas un échantillonnage exhaustif et donc d’observer la totalité des

conformations adoptées par les glycanes. Les variations observées entre les différentes trajectoires pourraient dépendre de la conformation de départ et étendre la durée de simulation ou encore considérer d'autres répliques permettraient de s'affranchir de ce biais. De manière similaire, il est possible que les deux glycanes Gx1 sur le site 893 sans acides sialiques ni fucose et restant étendus vers le solvant puissent adopter une conformation similaire à ce qui est observé sur les autres trajectoires et se replier le long de la surface protéique. La possibilité que les glycanes adoptent une conformation stable vers le solvant pourrait également jouer un rôle dans les interactions avec les partenaires de la protéine tels que Neu-1. Il est important de noter également que le site de glycosylation 893 est proche du point d'ancrage du récepteur à l'insuline à la membrane plasmique. Le CRE dont fait partie Neu-1 étant situé au niveau de cette membrane, il est possible que l'orientation des glycanes vers le solvant, comme observé sur nos trajectoires, puisse jouer un rôle dans la reconnaissance entre Neu-1 et son substrat.

Il est important également de noter que la présence du fucose et des acides sialiques augmente la surface couverte par le glycanes au cours des trajectoires de dynamique moléculaire. Ces deux saccharides sont impliqués dans la régulation des fonctions protéiques [108, 119–121] et leurs éventuelles interactions avec la surface de la protéine, stabilisant certaines conformations, pourraient être un des mécanismes impliqués dans les processus de régulation. Les résultats que nous avons obtenus ont permis de montrer qu'un résidu glutamate et deux résidus aspartate à proximité du site de glycosylation 893 jouent à priori un rôle dans la conformation et la position du glycanes à la surface de la protéine. La charge et l'encombrement stérique lié aux acides sialiques et / ou au fucose pourraient perturber voir modifier les interactions avec ces acides aminés et ainsi déplacer l'équilibre conformationnel du glycanes qui interviendrait alors dans la mise en place des interactions avec Neu-1.

Dans le cadre de l'étude que nous présentons, nous avons travaillé sur un dimère de l'ectodomaine du récepteur à l'insuline glycosylé sur un seul site. Or, le récepteur à l'insuline possède au total 18 points de glycosylation potentiels principalement présents sur sa région extracellulaire [12]. Considérant le volume important que peuvent représenter les glycanes sur une protéine, visualiser l'impact de l'ensemble des points de glycosylation pourrait nous éclairer sur la surface de la protéine qui reste accessible aux différents partenaires de la MEC. Le rôle des glycanes dans la protection contre la dégradation des protéines est connu, ainsi que leur rôle dans l'échappement au système immunitaire dans le cas des virus [122–124]. Les glycanes étant impliqués dans des interactions avec les acides aminés proches du point de glycosylation [125], visualiser leur impact sur la totalité d'une protéine pourrait aider à identifier les résidus impliqués dans ces mécanismes de protection.

6.4 Conclusion générale

Les travaux réalisés au cours de cette thèse ont permis d'intégrer l'*Umbrella Visualization* au logiciel UnityMol. Grâce aux fonctionnalités du moteur Unity et celles développées au sein du logiciel, nous avons pu créer un nouvel outil d'analyse et de visualisation qui permet d'évaluer visuellement et statistiquement l'impact d'un glycanes sur la surface de la protéine. L'*Umbrella Visualization* permet ainsi de caractériser les interactions entre glycanes et protéine et contribuera certainement à l'élucidation des mécanismes de régulation liés aux glycanes. Les travaux de dynamique moléculaire réalisés sur les glycanes en chaîne isolée soulignent le rôle de la liaison mannose $\alpha 1-6$ comme point de flexibilité clé du glycanes et permettent d'illustrer la différence de conformation des deux types d'acides sialiques étudiés. Dans tous les cas, la nécessité d'accéder à un échantillonnage exhaustif des structures s'impose afin d'élucider le rôle des glycosylations dans les fonctions protéiques.

Les travaux réalisés sur deux protéines glycosylées (récepteur à l'insuline et fibromoduline) ont permis d'illustrer les interactions entre protéine et glycanes. Dans le cas du récepteur à l'insuline, nous identifions notamment quelques acides aminés de la protéine capables de contraindre la conformation des glycanes. La courte trajectoire obtenue sur la fibromoduline portant des chaînes de keratan sulfate nous donne un premier aperçu du comportement de ces chaînes vis-à-vis de la protéine et ces premières observations servent de base à une étude future plus poussée et détaillée de l'impact des glycosaminoglycanes sur la structure et la dynamique des protéines de la matrice extracellulaire.

Annexes

Communications scientifiques

Présentations orales

- BIBM 2018, Madrid, Espagne, Décembre 2018 : *New visualization of dynamical flexibility of N-glycans : Umbrella Visualization in UnityMol*
- JJCC 2018, Reims, France, Octobre 2018 : *Umbrella Visualization : Visualization and Analysis of Glycan chains in Molecular Dynamics Trajectories*
- Journée Doctorale Transfrontalière, Reims, France, Mars 2018 : *DOGME : in silico DevelOpments for the study of Glycosylation and post-translational modifications applied to Extracellular Matrix proteins*
- Electronic Imaging 2018, Burlingame CA, USA, Janvier 2018 : *ViDy, ViGly : Visualization of dynamical flexibility of virtual N-Glycans on proteins.*
- JOBIM 2017, Lille, France, Juillet 2017 : *DOGME : in silico DevelOpments for the study of Glycosylation and post-translational modifications applied to Extracellular Matrix proteins*

Présentations poster

- MBE Conference, Manchester, United Kingdom, Juillet 2018 : *DOGME : towards the improvement of the atomic and molecular modelling of ECM proteins and components*
- Journée VISION, Reims, France, Novembre 2017 : *ViGly : Visualisation de la flexibilité dynamique de N-glycanes liés aux protéines*
- RCTGE 2017, Besançon, France, Juin 2017 : *DOGME - DévelOppements in silico dédiés à l'étude des Glycosylations et modifications post-traductionnelles des protéines de la Matrice Extracellulaire*
- GGMM 2017, Reims, France, Mai 2017 : *DOGME - DévelOppements in silico dédiés à l'étude des Glycosylations et modifications post-traductionnelles des protéines de la Matrice Extracellulaire*

Publications

New visualization of dynamical flexibility of N-Glycans: Umbrella Visualization in UnityMol.

Camille Besançon, Alexandre Guillot, Sébastien Blaise, Manuel Dauchez, Nicolas Belloy, Jessica Prévotéau-Jonquet, Stéphanie Baud
Univ. of Reims Champagne-Ardenne, UMR CNRS 7369 MEDyC.
 Reims, France
 {firstname.last}@univ-reims.fr

Abstract—N-glycosylations play an important role in protein functions and some alterations of glycosylation such as sialic-acid hydrolysis are related to protein dysfunction. *In vitro* study of N-glycans can be a challenging task because of the structural diversity and the many reactive groups of the glycan chains. Molecular dynamics is a useful tool and probably the only one in biology able to overcome this problem and give access to conformational informations through exhaustive sampling. To better decipher the impact of N-glycans, we have to visualize their influence over time on the protein structure. This is why we recently developed a new 2D graphical method called the Umbrella Visualization to assess, as a density graph, the protein surface covered by glycans during a molecular dynamics trajectory. Whereas this methodology brought relevant informations related to the glycans intrinsic flexibility and dynamics, we needed further developments in order to integrate an accurate description of the protein topology and its interactions. We propose here to transform this analysis method into a visualization mode in UnityMol. UnityMol is a molecular editor, viewer and prototyping platform, coded in C# with the Unity3D game engine. The new representation of glycan chains presented in this study takes into account both the main positions adopted by each antenna of the glycan and the intensity of their motions. The adapted Umbrella Visualization provides invaluable informations about the protein surface shadowed by the glycan antennas. Ultimately, the analysis of the collected data allows us to discuss both the flexibility of the glycan (=ability to explore distinct areas), and its stability (=ability to avoid spreading from its main conformational state) and offers the possibility of identifying the key elements in the protein/glycan interactions.

Index Terms—Molecular modeling, visualization, UnityMol, Unity3D

I. INTRODUCTION

In superior organisms, some cells like fibroblasts are able to produce and secrete fibrous proteins and polysaccharides that organize and form an environment called the extracellular matrix (ECM) [1]. The animal ECM is divided into two main elements. The basement membrane is a collection of proteins organized in sheets to which cells are anchored, forming epithelium. The interstitial matrix fills the space between cells [2]. This environment provides a structural support to cells because of its mechanical properties, contributing to mechanisms such as cell migration [3].

The ECM is also a very dynamical environment that contains many biologically active molecules such as enzymes or degradation products of its components. These molecules

are able to interact with the nearby cells through receptors, mostly transmembrane proteins which are anchored to the cellular membrane [4]. These proteins trigger molecular signals through the membrane, allowing cells to react to their environment. Understanding how these signaling processes take place and how cells are affected is essential in order to deepen our understanding of the biochemical role of the ECM.

Transmembrane proteins, as well as proteins composing the ECM, are modified after their synthesis by post-translational processes. Glycosylation is an enzymatic maturation process essential to the protein function, including ligand-receptor affinity and dimerisation. This process involves the linking of branched chains of saccharide residues on the protein surface. More precisely, N-glycosylations are defined by the covalent linking of a glycan chain on an asparagine amino-acid. These chains can be composed of several types of saccharide residues linked by a linear glycosidic linkage and arranged around a common, rigid core with highly variable and flexible branches. Because these chains are bulky, they have an impact on the protein folding, helping it to adopt its fully functional 3D structure. Glycosylations are also described as recognition patterns for glycosylation binding proteins such as lectins and are thus involved in signaling pathways. In pathological contexts such as age-related diseases like cancers, diabetes or cardiovascular diseases, alteration of the glycosylation patterns affects the protein functions [5].

Glycans are very flexible because of the linear glycosidic linkage between each saccharide residues [6]. They are also very reactive structures because of the many hydroxyl groups that compose saccharide residues. They are subject to a quick degradation under experimental conditions. Cells usually used in protein production and purification, such as *E. coli*, are not able to synthesize glycan structures found in animal organisms. For these reasons, studying glycosylations with classical experimental methods is very difficult.

Molecular dynamics (MD) is a useful tool to overcome this problem and give access to conformational information through exhaustive sampling. MD methods use empirical forcefields to describe molecules at the atomic level. Calculating motions with a high precision is possible thanks to the availability of heavy computational resources. Considering the biological roles of glycosylations and their big impact

on protein function, understanding and deciphering protein functions and regulations strongly motivates the integration of glycosylations in *in silico* studies and the development of tools dedicated to the study of MD simulations of glycosylated proteins.

Our previous *in silico* studies conducted on the impact of sialic acids on the structure of glycans has risen the lack of tools dedicated to the field [7]. This study also initiated the development of a method called the Umbrella Visualization. The aim of this method is to evaluate the zone covered by a glycan during a MD simulation. This representation is a way to appreciate the relative flexibility of the glycan during a simulation. Thanks to this method, we were able to demonstrate, for the first time, the impact of sialic acids on the flexibility of glycans. Since this study focused on isolated glycans in water, the results of this first application of the Umbrella Visualization were restricted to 2 dimensions graphical representation.

We introduce the improvement of this method that now takes into account the protein topology and the glycan motions above the protein surface. The main goal is to fully understand how a glycan will structurally impact the protein surface. The dynamic information collected during a MD simulation should help us identify the most impacted areas and understand which parts of the protein will be accessible to other molecules such as solvent molecules, hormones or even other proteins.

First, we will introduce the main tools already available in order to study glycosylated proteins. Many of these tools focus on the study of static, crystallographic structures. Visualization tools such as PyMol [8] and VMD [9] integrate visualization tools dedicated to the representation of glycan with the aim of better understanding their conformations. We then present UnityMol, developed with the video game engine Unity3D. With this software, we were able to upgrade the Umbrella Visualization from a 2D method generating density plot to a tool allowing the visualization of the impact of glycans onto a protein structure during a MD trajectory. This tool, in conjunction with insights of experimental studies, could help understand mechanisms involving N-glycans and design *in vitro* experiments aimed at understanding their impact on protein functions.

II. RELATED WORK

A. Glycosylation and Modelling

The first step needed to study glycosylated proteins is to identify the glycosylation sites. Sometimes, the first saccharides from a glycan are present on the .pdb structure files available on the Protein Data Bank at www.rcsb.org [10]. This can help identify some glycosylation sites but most of the time, glycan structures are not available: because of their great flexibility, glycans are not sufficiently ordered to be detected by classical methods used to determine protein structures such as X-ray crystallography [11]. It is thus necessary to use prediction tools in order to identify the glycosylation sites. Several tools are available as web-portals or standalone programs [6], [12], [13].

Before running MD simulation, it is useful to analyze and understand the structure of glycosylated proteins. Some tools can help identify conformational errors in structures [14], [15] while others are dedicated to the study of the torsion angles and saccharidic linkage [16]–[18]. Carbohydrates have many hydroxyl groups. These groups are very reactive and can interact with their surroundings. This is why it is also important to understand how glycans will interact with the protein residues. GlyVicinity [19] is a tool that statistically analyzes the aminoacid closest to a chosen type of carbohydrates.

Once the glycosylation sites are identified, it is necessary to build the glycan structure on the site of interest. Even if the first saccharide residues are present in the .pdb file, the remaining part of the glycan has to be reconstructed, avoiding any aberrant structures or steric clashes with the protein. SHAPE is a tool enabling a user to build any glycan structure [20]. SHAPE also predicts several energetically-favorable conformers using a genetic algorithm.

SWEET II on the glycosciences.de portal is another tool that can be used to build a glycan structure [21]. This glycan can then be used on the GlyProt page. GlyProt allows a user to upload a .pdb file and add the glycan created with SWEET II on one or more glycosylation sites [22]. Glycam-Web (<http://glycam.org/>) and Charmm-gui with the Glycan-Reader tool [23] are also web-servers allowing a user to build a glycosylated-protein structure.

Very recently, the DoGlycan set of tools [24] was developed in order to perform the same function with the Glycam forcefield, which is dedicated to the study of carbohydrate structures.

B. Visualization

Many of the main visualization softwares allow the visualization of saccharides in all-atom representations such as licorice, balls and sticks or van der Waals (VdW) sphere representations. But, few of them propose ways of visualization that highlight the positions and structures of glycosylation on the proteins' surface.

PyMOL which includes a plugin called Azahar [25] can be used to build a glycan from a template list of saccharide structures. The plugin adds three specific display modes aimed at simplifying the representation of glycan structures: the cartoon and wire representations show the cycles as non-flat polygons linked by rods and the bead representation shows each cycle as a sphere. This plugin also gives the possibility to compute relevant structural values in order to analyze both static structures or MD simulations.

With the 3D-SNFG plugin [26], the VMD software implements its own graphical representation based on the nomenclature developed in the 2nd edition of the Essentials of Glycobiology textbook [27]. Each monosaccharide is represented by a 3D object based on the corresponding symbol.

UnityMol, an open source platform developed with the Unity3D engine (<http://unity3d.com/>) in C# language [28] is dedicated to the visualization of biological molecules and uses HyperBalls [29] visualization. In

the latest versions called SweetUnitymol, this software integrates display modes and representations dedicated to saccharides such as ribbons representation or the possibility to color saccharide depending on the saccharide nature. For each saccharide, the associated color is determined by the symbol found in the SNFG nomenclature.

The tools described above offer different ways of building, analyzing and visualizing a glycosylated-protein structure. It is possible to have a workflow leading to the building of a glycosylated system ready for MD simulations, such as the one presented in [30]. However, the forcefields and nomenclatures can differ between all these tools and portals, making it hard to switch between them. Moreover, these tools are mostly dedicated to the building of glycosylated protein systems and the analysis of static, crystallographic structures. Even if there is a possibility to obtain MD simulations data on glycans and glycosylated proteins, there is a lack of tools dedicated to the analysis of such trajectories.

III. BACKGROUND

A. 2D Umbrella Visualization

It is well known that sialic acids, charged saccharides often found at the extremities of glycan chains, are involved in protein function's regulation [31]. Such mechanisms are involved in diseases such as type 2 diabetes. To better understand the influence of sialic acid on protein's function and structure, a MD study on sialylated and non-sialylated glycan chains: Trajectories for four glycan chains with two and three branches, with and without sialic acids were computed [7]. The starting carbohydrates structures were built using the Avogadro software. 500ns of trajectories were generated with the OPLS-AA forcefield and the Gromacs package 4.6.3. NPT ensemble was used with a 310K temperature and 1 bar pressure.

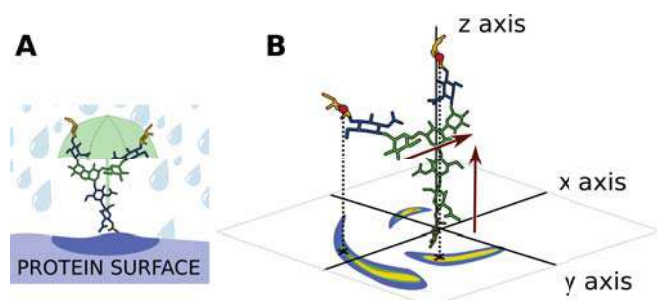


Fig. 1. A) Schematic principle of the Umbrella Visualization: The glycan here is shown in sticks representation and colored by residue type. The Umbrella Representation considers the glycan acting as an umbrella over the protein surface. B) Implementation of the Umbrella Visualization process: The glycan is represented in sticks. The core of the Glycan, shown in green, is aligned along the z axis thanks to vectors describing the core (red arrows). The center of mass (red dots) of the last saccharide residue on the branches is projected on the xy plane. The final result is a density plot reflecting the relative flexibility of the glycan during a MD trajectory.

In order to plot and evaluate the relative flexibility of glycan chains with and without sialic acids, the Umbrella Visualization was developed. The Umbrella Representation

considers the glycan acting as an umbrella protecting the protein surface so that an hypothetical molecule would not be able to interact directly with this part of the protein (Fig.1A). The glycan is divided into vectors representing its different components: Two vectors describe the glycan core. The other vectors describe the branches of the glycan. The first step is to measure the distance between the centers of mass of the saccharide residues defining these vectors for every trajectory step. This is done thanks to the Gromacs [32], [33] software and the distance module. Files listing the coordinates of these vectors during the simulation are generated and analyzed.

An in-house program reads those coordinates in order to place the chain on an orthogonal coordinate system, with the asparagine amino-acid at the glycan's base set on the origin. The two core vectors as described on Fig.1B, are used in order to orient the glycan within this coordinate system. First, the angle between the inner-core vector and the z-axis is calculated so this core can be oriented along this axis. Then the angle between the second vector and the x axis is calculated and the glycan chain is rotated to make this vector coplanar with the xz plan. Finally, the xy coordinates of the last common saccharide residue for both type of glycans (sialylated and non-sialylated) are written and displayed on a density graph.

The non-sialylated bi-antennary glycan showed increased flexibility and explored a much larger region of the plot compared to the sialylated bi-antennary glycan. Desialylation impacted the tri-antennary glycan differently: only one of the three branches showed a significant change in flexibility. Though the influence of sialic acids varies depending on the branching of the glycan, it is clear that sialic acids have an impact on the flexibility and stability of glycosylation chains. The Umbrella Visualization, combined with clustering methods and measurement of angles values during the trajectory, was thus able to demonstrate this influence [7].

B. UnityMol

In order to better understand how glycans and sialylation impact proteins, it was then decided to transform this 2D-analysis method into a visualization method aimed at identifying which regions of the protein surface are impacted by a glycan during a MD trajectory.

For this purpose, we have decided to work with an existing open source software, UnityMol, to implement our new functionality. Moreover, this would allow us to make our visualization easier to use by integrating it into a software already used by the community.

UnityMol is a molecular editor, viewer and prototyping platform, coded in C# with the Unity3D game engine. It already has all the classic visualization modes used in molecular modeling and is even able to automatically detect saccharide residues in order to provide user-friendly visualization modes dedicated to the representation and study of saccharide chains. It also provides a plug-in allowing a user to read MD trajectories. The main advantage of this software is the Unity3D engine that makes it easy to work either with built-in tools or by implementing new features. The wide use of

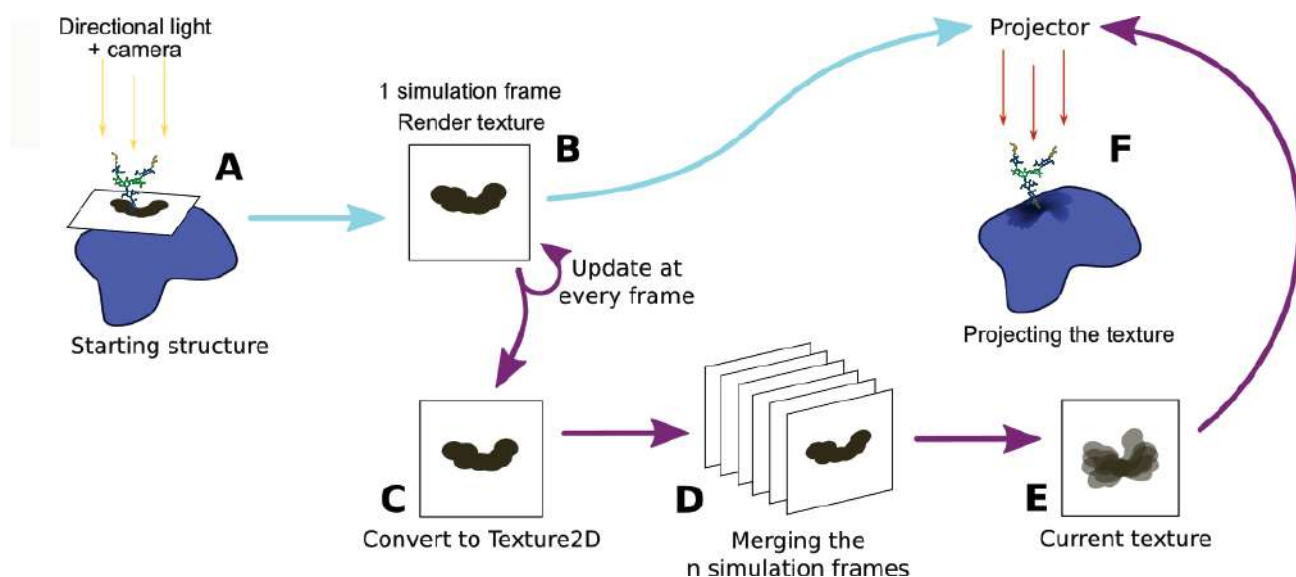


Fig. 2. Unity workflow: The blue arrows describe the workflow to project only one frame. The purple arrows describe the workflow to project the shadow from the whole simulation. A directional light project the glycan's shadow over an oriented plane (A). This shadow is rendered into a render texture by a camera (B) and either displayed on the surface for a static structure or, with a MD trajectory, converted into a Texture2D object (C). This object is read and combined with the images from previous frames (D). All the results acquired along the MD simulation are merged together into a final Texture2D (E) object which is then used in the projector, thus displaying it onto the protein surface (F).

this platform also warrants support from an active community and provides many online resources and tutorials. This engine supports many different platforms such as Android, Windows, Linux or Mac platforms. Exporting the executable file for these platform is done very simply at compilation, giving access to these platforms while saving a great amount of time by avoiding further developments. The Unity3D graphical interface also displays a Game Mode that allows a user to run its program without compiling it. This feature is especially useful in order to rapidly and easily test adjustments brought to the code. Unity3D is also known for being relatively user friendly, partially thanks to its interface, but it is also important to note that every parameters and components are accessible via scripting in C# or javascript languages, making it a very versatile tool. Thus, the UnityMol prototype developed with this engine is easily modifiable and extensible and provides every tool needed to implement the Umbrella Visualization.

IV. SYSTEM DESIGN

The aim of this work is to evaluate the covered protein surface during a MD simulation by taking into account the glycan's motions and the protein topology. This enables the display of the 3D structural information. First, the general principle of the shadow projection is detailed. Then, a second part will detail how the frames extracted from an entire MD trajectory are combined together in order to obtain a statistical information. Finally, we present a test-case from a MD trajectory of a glycosylated protein.

A. General principle

One of the many advantages of the Unity3D engine is the possibility to build prefabs that can be used for development

purposes. Prefabs are assets storing GameObjects components and setting their properties to a desired values. This way, when the prefab is instantiated into a scene, the GameObjects components are already positioned and oriented as desired. Among these components, C# or Javascript scripts can be assigned to a GameObject component and used in order to dictate a behavior to this component. Moreover, prefabs are very versatile: even if all the instances of a prefab will share its properties and ensure a certain homogeneity between the instances, individual instances are independently editable and it is thus possible to change some individual properties as needed. Building a prefab and assigning scripts to its component is a good way to build a template for an asset that would be instantiated several times into a Unity scene.

By using these built-in tools and components from the Unity3D engine, we were able to design a UnityMol prefab able to capture the glycan shadow and project it onto the protein surface. This prefab is composed of a camera, a projector and a directional light oriented perpendicular to a flat mesh defining a plane (Fig.2A). These components are organized together to project the shadow of a glycan on the protein.

Cameras are also defined as GameObjects components and can be used into prefabs. Cameras are usually used to capture and display a view to the user but it can be useful to render this view into a texture. To this end, a Render Texture can be assigned to a camera. Unity3D also provides layers that allow for lights and cameras to not illuminate or not render elements from the layers defined in their respective "culling masks".

The Projector prefab from the Unity3D "standard asset" package allows a user to project a desired material or tex-

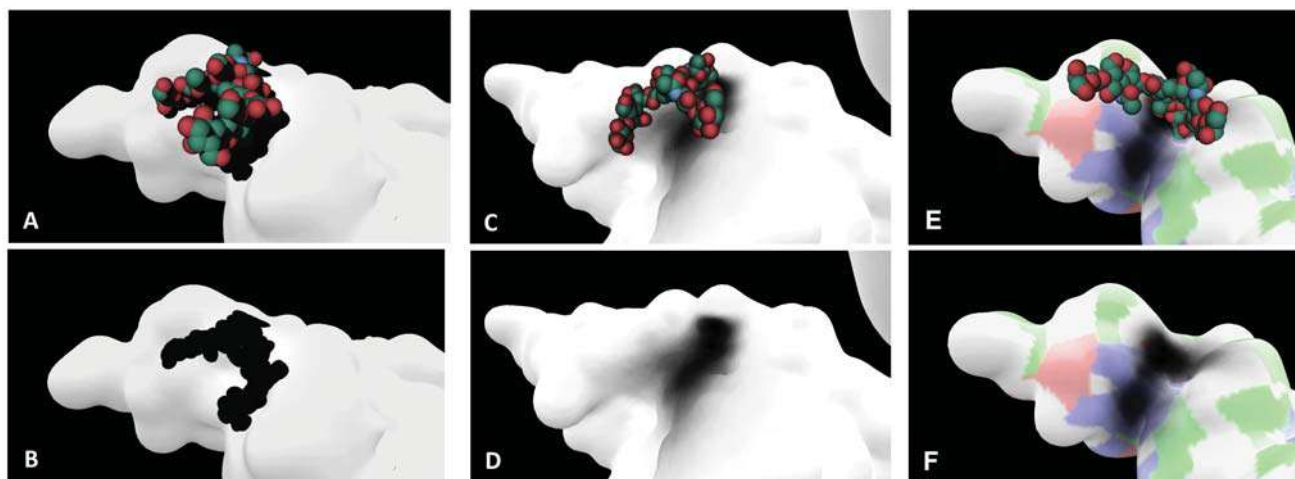


Fig. 3. Results: The white surface is the protein surface. The glycan displayed on VdW mode. The dark-gray areas corresponds to the projected texture. A): Starting structure used for the MD trajectory. The glycan is located near a shallow groove on the protein surface. B): Shadow of the glycan projected on the protein surface on the starting structure. C): This conformation is extracted from the MD simulation and shows how the glycan is positioned on the groove, in a conformation that looks like the starting structure. D): On this particular example, it is possible to see that, during this MD trajectory, the glycan's motions are mostly restricted to an area along a small groove on the protein surface. E) and F) show the glycan and the projected shadow over the groove with the surface colored depending on the residues underneath. Blue patches are for basic residues, red patches are for acid residues, green patches are for polar residues, white patches are for non-polar residues.

ture onto all objects intersecting its *frustum*. Projectors have dedicated shaders in which the projected texture is assigned.

One of the many features of the Unity3D engine is the possibility to quickly instantiate GameObjects of basic shapes such as cubes, spheres or flat planes. As we need a flat surface to project the glycan shadow, the prefab integrates a plane positioned under the glycan. Different types of lights, traditionally used to illuminate a scene, are also available on the Unity3D engine. We want the glycan's shadow size to be independent from the light to plane distance, we thus use a directional light. This type of light illuminates scenes with parallel rays, similar to sun rays. The light's orientation is perpendicular to the plane's surface, reproducing an orthogonal space used to orient the glycan.

Once the shadow is displayed, we need to have a representation mode that reflect the atomic volume of the glycan. To this end, we use the VdW representation mode which is based on the VdW radius, a theoretical value used to represent the atom's volume. In order to only display the shadow of the glycan on the plane, the glycan atom's mesh is set in "shadow only" mode. Doing so makes this mesh also invisible to UnityMol's main camera but it is possible to copy the glycan atoms mesh and set it on another layer in order to see the shadow and the atoms at the same time with the prefab's camera ignoring the glycan mesh. Likewise, to prevent the camera to capture parts of the surface passing through the plane, the protein surface is set on a dedicated layer and a culling mask on this layer is applied to the prefab's camera. This can happen often due to the irregularity of the protein surface and the plane being close to the surface. This set-up is able to save the glycan's shadow into a render texture (Fig.2B).

Finally, the shadow-projector uses the texture generated by

the prefab displaying the glycan shadow. This texture is then projected onto the protein surface, highlighting parts of the protein surface impacted by the glycan (Fig.2F). In order to avoid displaying unnecessary elements to the UnityMol's user, the lit-plane prefab is set on another layer and a culling mask is applied on this layer on UnityMol's main camera and light. This way, the plane is not visible to the user and only the molecules are displayed on the screen. Doing so also prevent the main light to interact with the lit plane prefab, which would alter the glycan shadow. With a similar goal, the camera's and projector's views are set on orthographic mode. With this mode, 3D objects are represented in two dimensions by the means of a parallel projection of the render texture's pixels. The picture projected on the protein surface will not vary with the relative distance to the protein surface.

This workflow projects the glycan's shadow onto the protein surface for a static structure. However, this tool is dedicated to the study of data extracted from MD. We have thus to take into account the whole length of a MD trajectory.

B. Frame merging

MD trajectories can have several thousands of frames to visualize and analyze. Our goal here is to provide a tool allowing the user to understand the behavior of the glycan during the whole trajectory to see which parts of the proteins are more affected by the glycan. We have to combine or merge the information from all the simulation frames and display the result.

To this end, a script is included on the lit-plane prefab. When the UnityMol program is started, an integer array is created. The array size is defined by the number of pixels of the final texture. When the simulation frames are loaded,

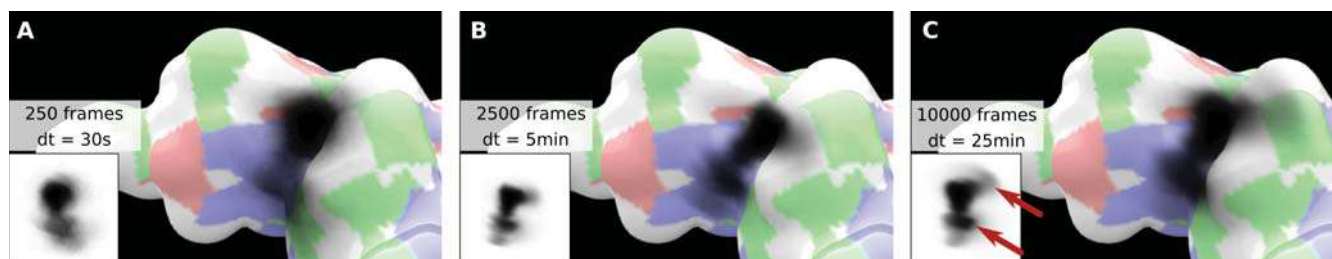


Fig. 4. Results of the Umbrella Visualization on three MD samples. The flat textures projected onto the surfaces are shown as insets. A): Results extracted from 250 frames. The overall tendency is already visible, highlighting the importance of the basic residues. B): Results extracted from 2500 frames. C): Results extracted from 10000 frames. By increasing the number of frames used for the computation, the lower area (bottom red arrow) becomes more compact and darker while a light spread zone (top red arrow) appears in the upper area. This clearly indicates that with a low amount of frames, the role of zones may be overestimated while others are missed. The dt label is the elapsed time during the calculation of these results.

for each frame, the prefab collects the glycan shadows and saves it in a render texture (Fig.2C). the scripts converts this render texture into a temporary Texture2D object in order to read the pixels from this texture. Then, for each pixel, the script will detect if the pixel is black (a shadow pixel) or not. The corresponding element on the array will be incremented by 1, thus counting the number of times where a pixel is shadowed by the glycan during the trajectory (Fig.2D). After having read all the frames, the array's values for each pixel are normalized. The higher the count, the darker the pixel will be, allowing us to see what parts of the protein are most often obscured by the glycan (Fig.2E). Pixels color are defined by three components: r, g, b and a. These components determine the final color of the pixel by adjusting the red, green and blue hues (r, g and b). Using the values from the array, we are able to define the r, g and b elements of the accumulation texture's pixels. Thus, throughout the trajectory, the accumulation of the different positions of the glycan is visualized and at the end of the trajectory, the distribution of the conformations of the glycan is obtained.

C. Test case

The Fig.3 shows the result of the Umbrella Visualization with only one frame displayed (A and B panels). The protein surface is displayed in white while the glycan is in all-atom representation. Each sphere corresponds to an atom from the glycan (For clarity, the hydrogen atoms are not displayed). The green spheres are carbon atoms, the red spheres are oxygen atoms. Thanks to the VdW representation, the shadow size is matching the volume of the glycan and helps approximate the surface covered for this specific conformation. On the A and B panels, the glycan is oriented above a shallow groove of the protein. Such a conformation on a .pdb file could indicate a preferential conformation due to favorable interactions with the protein surface. To confirm or reject this hypothesis, we ran a MD simulation of the insulin receptor ectodomain dimer with one bi-antennary glycan per subunit. We used the OPLS-AA forcefield and Gromacs in order to run a 500 nanoseconds simulation using the NPT ensemble.

The C and D panels show the projected shadow on a 250 frames trajectory. Because we normalize the r, g, b values

with the number of frames, the result is statistically relevant. On this example, we can conclude that during most of the trajectory, the glycan is extended above a shallow groove but doesn't explore the remaining part of the protein surface. This could indicate that particular protein residues in this region have favorable interactions with the saccharide residues of the glycan, hence favoring this conformation. These results thus corroborate the hypothesis made with the static structure. The Fig.3E and F shows that the groove where the glycan stays during most of the 250 frames is made by basic residues. In order to investigate the nature of residues interacting with the glycan, it is possible to apply a color scale corresponding to the physico-chemical nature of residues located under the surface. Thus, red, blue and green patches will show the presence of acid, basic and charged residues, respectively. The superimposition of this information and the shadow of the glycan emphasizes what are the properties of the residues below that interacts with the glycan. In the context of protein / protein interaction, this is a valuable information as charged amino-acids can be involved. This could highlight how a glycan will influence the protein / protein interaction. These residue would be able to interact with the many hydroxyl groups of the sugar residues, hence stabilizing the position of the glycan in the groove. Considering the importance of basic residues in protein function and interaction, such insight could be valuable in order to understand the impact of glycosylation on protein structure and function.

D. Performances

We tested this method on a protein system composed of a dimer of the insulin receptor ectodomain with one glycan per monomer. This corresponds to a number of 26.030 atoms, without water molecules used to calculate the motions during the trajectory. We worked with a 3.7GHz Intel Xeon W-2145 CPU machine with 64Gb of DDR4 RAM and a 1 Nvidia GeForce GTX 1080Ti graphic card. This analysis of 10000 frames with the Umbrella Visualization method took approximately 25 minutes, 5 minutes for 2500 frames and 30 seconds for 250 frames. We used 512 pixel wide texture. Using a smaller texture allowed some time gain (18 minutes for a 256 pixels wide texture). The main limitation is the

number of atoms contained in the structure. The dimer we used here is a relatively large system, thus processing the data for every frame of the simulation requires more computational resources than a smaller system. In spite of that size, this method still allowed us to process a significant number of frames in a reasonable time. We then tested this method on a higher, more significant number of frames. The Fig.4 displays the results we obtained for 250, 2500 frames and 10000 frames. As expected, more information is available with a higher number of frames. Though the more prominent conformations are already displayed on the 250 frames sample (Fig.4A), the two other samples bring to light the impact of less represented conformations (Fig.4B and Fig.4C). In every case, the combined frames show that the glycan will have a tendency to contact the patch of basic residues on the groove, highlighting the possible role of this part of the protein in stabilizing the glycan's conformations.

V. DISCUSSION AND CONCLUDING REMARKS

The Umbrella Visualization is a new method of analysis using a camera and a projector to project the shadow of a glycan on the protein surface. Shadows rendering of the glycan is done by counting the number of times a pixel is obscured by the glycan and determining the color of the final texture depending on this number.

We also wish to project the shadow corresponding to one single frame over the protein surface in order to both have the information given by the merged shadow and the information given by individual conformations. Highlighting one frame could help put into evidence rare events occurring during a trajectory or a representative conformation, for demonstration and illustration purposes.

The Umbrella Visualization brings structural information about the impact of glycans, but doesn't offer any quantifiable value that would help evaluate more concretely the space occupied by glycans. However, the 2D graphical method previously developed took the form of a density plot, giving statistical informations but no structural information. Correlating the results of the 2D graphical method and the results of the 3D visual method is a good way to provide quantifiable data to support the visualization results.

As glycans are bulky, the space they explore and occupy during a MD trajectory could also inform us on their influence on molecular complex formation. We plan to add methods to calculate values such as the occupied volume during the trajectory.

Local topology such as groove, bulges or the presence of other protein domains can also influence how this interaction would take place. On this first version of the Umbrella Visualization, the directional light is oriented perpendicularly to the plane, reproducing an orthogonal space. This introduces a bias into the results as it doesn't reflect how another molecule would approach a protein in order to interact or form a complex. To correct this, we wish to allow the prefab to move around the glycan, following a given position defined by the user, and then visualize the results with this new orientation.

Molecular modeling is a tool able to help designing *in vitro* experiments. Molecular docking, for example, is used to design new drugs thanks to its ability to test several molecules without having to synthesize them. Identification of the key interactions can further the understanding on how a drug will interact with a protein and influence its function. In a similar way, with the Umbrella Visualization, we wish to provide a tool able to help the design of studies focused on glycosylations. By highlighting which parts of the protein are affected by glycans, it is possible to identify important protein residues that could be mutated in order to understand their impact on glycans and protein structure and dynamics.

At this time, the Umbrella Visualization is usable for the study of one glycan at a time. Because some proteins are heavily glycosylated, we will optimize this method to enable the visualization and analysis of multi-glycosylated proteins. This process heavily relies on the use of a prefab, so it is possible to instantiate this object several times in the same UnityMol session while applying transformations to components of single instances of the prefab. Each instance of the plane assigned to a glycan can thus be oriented accordingly. The Umbrella Visualization method could help highlight how glycans impact protein-protein interactions in biological contexts. For example, it was shown that the human coronavirus NL63, a virus causing severe lower-respiratory-tract infections, uses a glycan shield to protect surface epitopes from the host immune system [34].

Among the methods used in molecular modeling, free-energy calculations are used to characterized interactions between partners [35]: it is possible to evaluate the contribution from each residue from both partners. It would be interesting to combine the Umbrella Visualization with free-energy calculation to highlight how the protein's amino acids would influence the glycan's behaviour by favoring interactions with the glycan's hydroxyl groups. By displaying both the results of such calculation and the results of the Umbrella Visualization on the protein surface, it would be possible to correlate the binding free energy with glycan's motions.

At the moment, the surface receiving the shadow is based on the initial .pdb file loaded in UnityMol and doesn't change during the course of the trajectory (reduced computational work). Because proteins are also dynamical objects, it would be interesting to implement a quicker surface calculation and alter the surface shape for each frame. This would indicate how the protein evolves and adapts to the glycan motions. This could also impact the regions on which the shadow is projected and thus better highlight the influence of the glycan on the protein.

We tested this method on three MD trajectory sample of 250, 2500 and 10000 frames. We were able to see significant results with smaller numbers of frames that allowed us to highlight the main conformational tendencies of the glycan of interest. The biggest samples showed more details and hinted at less represented conformations. This shows that this method has the potential to give informative results that could also either help design *in vitro* experiments and / or corroborate

data from other calculation. As the ECM is a very dynamic environment featuring heavily-glycosylated protein, we wish to use this method on related studies. We have shown that the insulin receptor's function was altered by the hydrolysis of sialic acids and that these saccharide [36] had an impact on the flexibility of glycans [7], the next step would be to investigate how this change in flexibility impacts the insulin receptor function. By applying the Umbrella Visualization on a glycosylated insulin receptor, we hope to highlight important interactions involved in this mechanism.

To conclude, the work presented here is the development of the Umbrella Visualization method in the UnityMol molecular viewer. This method is dedicated to the study of glycosylated proteins and emphasize hidden regions of the protein surface inaccessible to other molecules, thus highlighting the impact of glycans on protein surfaces during the length of a MD simulation trajectory. We wish to release this tool as a plugin for the UnityMol software.

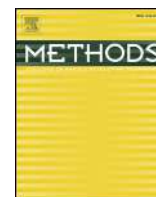
REFERENCES

- [1] J. Halper and M. Kjaer. Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. *Advances in Experimental Medicine and Biology*, 802:31-47, 2014.
- [2] A. D. Theocharis, S. S. Skandalis, C. Gialeli, and N. K. Karamanos. Extracellular matrix structure. *Advanced Drug Delivery Reviews*, 97:4-27, Feb. 2016.
- [3] S. V. Plotnikov, A. M. Pasapera, B. Sabass, and C. M. Waterman. Force fluctuations within focal adhesions mediate ECM-rigidity sensing to guide directed cell migration. *Cell*, 151(7):1513-1527, Dec. 2012.
- [4] C. Bonnans, J. Chou, and Z. Werb. Remodelling the extracellular matrix in development and disease. *Nature Reviews. Molecular Cell Biology*, 15(12):786-801, Dec. 2014.
- [5] P. Stanley, H. Schachter, and N. Taniguchi. N-Glycans. In A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, editors, *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2nd edition, 2009.
- [6] G.-Y. Chuang, J. C. Boyington, M. G. Joyce, J. Zhu, G. J. Nabel, P. D. Kwong, and I. Georgiev. Computational prediction of N-linked glycosylation incorporating structural properties and patterns. *Bioinformatics*, 28(17):2249-2255, Sept. 2012.
- [7] A. Guillot, M. Dauchez, N. Belloy, J. Jonquet, L. Duca, B. Romier, P. Maurice, L. Debelle, L. Martiny, V. Durlach, S. Baud, and S. Blaise. Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans. *Scientific Reports*, 6:35666, Oct. 2016.
- [8] Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. Nov. 2015.
- [9] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33-38, 27-28, Feb. 1996.
- [10] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235-242, Jan. 2000.
- [11] M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott, and R. A. Dwek. Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling. *Chemical Reviews*, 102(2):371-386, Feb. 2002.
- [12] S. E. Hamby and J. D. Hirst. Prediction of glycosylation sites using random forests. *BMC bioinformatics*, 9:500, Nov. 2008.
- [13] R. Gupta, E. Jung, and S. Brunak. Prediction of N-glycosylation sites in human proteins. 46:203-206, Jan. 2004.
- [14] T. Lutteke and C.-W. von der Lieth. pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC bioinformatics*, 5:69, June 2004.
- [15] T. Lutteke, M. Frank, and C.-W. von der Lieth. Data mining the protein data bank: automatic detection and assignment of carbohydrate structures. *Carbohydrate Research*, 339(5):1015-1020, Apr. 2004.
- [16] A. Böhne-Lang, E. Lang, T. Frster, and C. W. von der Lieth. LINUCS: linear notation for unique description of carbohydrate sequences. *Carbohydrate Research*, 336(1):1-11, Nov. 2001.
- [17] M. Frank, T. Lutteke, and C.-W. von der Lieth. GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Research*, 35(Database issue):287-290, Jan. 2007.
- [18] T. Lutteke, M. Frank, and C.-W. von der Lieth. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3d structures derived from the PDB. *Nucleic Acids Research*, 33(suppl 1):D242-D246, Jan. 2005.
- [19] M. A. Rojas-Macias and T. Lutteke. Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by GlyVicinity. *Methods in Molecular Biology (Clifton, N.J.)*, 1273:215-226, 2015.
- [20] J. Rosen, L. Miguet, and S. Prez. Shape: automatic conformation prediction of carbohydrates using a genetic algorithm. *Journal of Cheminformatics*, 1(1):16, Sept. 2009.
- [21] A. Böhne, E. Lang, and C.-W. v. d. Lieth. W3-SWEET: Carbohydrate Modeling By Internet. *Molecular modeling annual*, 4(1):33-43, Jan. 1998.
- [22] A. Böhne-Lang and C.-W. von der Lieth. GlyProt: in silico glycosylation of proteins. *Nucleic Acids Research*, 33(Web Server issue):W214-219, July 2005.
- [23] S. Jo, K. C. Song, H. Desaire, A. D. MacKerell, and W. Im. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. *Journal of computational chemistry*, 32(14):3135-3141, Nov. 2011.
- [24] R. Danne, C. Poojari, H. Martinez-Seara, S. Rissanen, F. Lolicato, T. Rg, and I. Vattulainen. doGlycans-Tools for Preparing Carbohydrate Structures for Atomistic Simulations of Glycoproteins, Glycolipids, and Carbohydrate Polymers for GROMACS. *Journal of Chemical Information and Modeling*, 57(10):2401-2406, Oct. 2017.
- [25] A. Arroyuelo, J. A. Vila, and O. A. Martin. Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules. *Journal of Computer-Aided Molecular Design*, 30(8):619-624, 2016.
- [26] D. F. Thieker, J. A. Hadden, K. Schulten, and R. J. Woods. 3d implementation of the symbol nomenclature for graphical representation of glycans. *Glycobiology*, 26(8):786-787, Aug. 2016.
- [27] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, editors. *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2nd edition, 2009.
- [28] Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent, and M. Baaden. Game on, science - how video game technology may help biologists tackle visualization challenges. *PloS One*, 8(3):e57990, 2013.
- [29] M. Chavent, A. Vanel, A. Tek, B. Levy, S. Robert, B. Raffin, and M. Baaden. GPU-accelerated atom and dynamic bond visualization using hyperballs: A unified algorithm for balls, sticks, and hyperboloids. *Journal of Computational Chemistry*, 32(13):2924-2935, Oct. 2011.
- [30] Y. Mazola, G. Chinea, and A. Musacchio. Integrating Bioinformatics Tools to Handle Glycosylation. *PLoS Computational Biology*, 7(12), Dec. 2011.
- [31] R. Schauer. Sialic acids as regulators of molecular and cellular interactions. *Current Opinion in Structural Biology*, 19(5):507514, Oct. 2009.
- [32] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1):43-56, Sept. 1995.
- [33] S. Pronk, S. Pli, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845-854, Apr. 2013.
- [34] A. C. Walls, M. A. Tortorici, B. Frenz, J. Snijder, W. Li, F. A. Rey, F. DiMaio, B.-J. Bosch, and D. Veisler. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nature Structural & Molecular Biology*, 23(10):899-905, Oct. 2016.
- [35] C. D. Christ, A. E. Mark, and W. F. van Gunsteren. Basic ingredients of free energy calculations: a review. *Journal of Computational Chemistry*, 31(8):15691582, June 2010.
- [36] S. Blaise, B. Romier, C. Kaweck, M. Ghirardi, F. Rabenoelina, S. Baud, L. Duca, P. Maurice, A. Heinz, C. E. Schmelzer, M. Tarpin, L. Martiny, C. Garbar, M. Dauchez, L. Debelle, and V. Durlach. Elastin-Derived Peptides Are New Regulators of Insulin Resistance Development in Mice. *Diabetes*, 62(11):3807-3816, Nov. 2013.



Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Umbrella Visualization: A method of analysis dedicated to glycan flexibility with UnityMol

Camille Besançon^{a,*}, Alexandre Guillot^a, Sébastien Blaise^a, Manuel Dauchez^{a,b}, Nicolas Belloy^{a,b}, Jessica PrévotEAU-Jonquet^a, Stéphanie Baud^{a,b}

^a Université de Reims Champagne Ardenne, CNRS, MEDyC UMR 7369, 51097 Reims, France

^b Université de Reims Champagne Ardenne, Plateau de Modélisation Moléculaire Multi-Echelle (P3M), Maison de la Simulation de Champagne Ardenne (MaSCA), 51097 Reims, France

ARTICLE INFO

Keywords:

UnityMol

Unity

Glycan flexibility

Covered surface

Molecular dynamics

ABSTRACT

N-glycosylation is a post-translational modification heavily impacting protein functions. Some alterations of glycosylation, such as sialic acid hydrolysis, are related to protein dysfunction. Because of their high flexibility and the many reactive groups of the glycan chains, studying glycans with *in vitro* methods is a challenging task. Molecular dynamics is a useful tool and probably the only one in biology able to overcome this problem and gives access to conformational information through exhaustive sampling. To better decipher the impact of N-glycans, the analysis and visualization of their influence over time on protein structure is a prerequisite. We developed the Umbrella Visualization, a graphical method that assigns the glycan intrinsic flexibility during a molecular dynamics trajectory. The density plot generated by this method brought relevant informations regarding glycans dynamics and flexibility, but needs further development in order to integrate an accurate description of the protein topology and its interactions. We propose here to transform this analysis method into a visualization mode in UnityMol. UnityMol is a molecular editor, viewer and prototyping platform, coded in C#. The new representation of glycan chains presented in this study takes into account both the main positions adopted by each antenna of a glycan and their statistical relevance. By displaying the collected data on the protein surface, one is then able to investigate the protein/glycan interactions.

1. Introduction

In superior organisms, the extracellular matrix (ECM) is an environment embedding fibrous proteins, glycoproteins and polysaccharides synthesized and secreted by cells like fibroblasts [1]. In the animal reign, the ECM is divided into two main components: the basement membrane and the interstitial matrix. The basement membrane is a combination of proteins forming sheets used to anchor epithelial cells [2]. The interstitial matrix fills the space between cells [3]. Because of its mechanical properties, this environment provides structural support to cells and contributes to mechanisms such as cell migration or cancer metastasis.

The ECM is a very dynamical environment with ongoing remodeling or degradation, making it a source of many active molecules. Enzymes or matrix degradation products, i.e. matrikines [4], are able to interact with the nearby cells through receptors, mostly transmembrane proteins which are located at the cell surface [5–8]. These proteins trigger molecular signals through the membrane, allowing cells to react

to their environment. Understanding how these signaling processes take place and how cells are affected is essential in order to deepen our understanding of the biochemical role of the ECM. As an example, the insulin receptor (IR) is a protein belonging to the class II tyrosine kinase receptors family. IR is synthesized as a precursor of 210 kDa. The pro-protein is cleaved into an α and a β subunits linked by a disulfide bridge. The α subunit is extracellular while the β subunit contains an extracellular part about 190 amino acids, a transmembrane helix and a cytoplasmic domain. The mature receptor is present at the membrane as a dimer with 2 α and β subunits. Transmembrane proteins, as well as proteins composing the ECM, are modified after their synthesis by post-translational processes. Glycosylation is an enzymatic maturation process essential to the protein function, including ligand-receptor affinity and dimerisation [9]. This process involves the attachment of branched chains of saccharide residues on the protein surface. More precisely, N-glycosylations are defined by the covalent linking of a glycan chain to the amide nitrogen of asparagine predominantly at the sequons Asn-X-Thr and Asn-X-Ser, where X is any amino acid but not proline [10]. In

* Corresponding author.

E-mail address: camille.besancon@univ-reims.fr (C. Besançon).

<https://doi.org/10.1016/j.ymeth.2019.07.010>

Received 1 March 2019; Received in revised form 9 May 2019; Accepted 9 July 2019

1046-2023/ © 2019 Elsevier Inc. All rights reserved.

the IR, 18 potential N-glycosylation sites are identified (14 in α subunit and 4 in β subunit). These chains are constituted of several types of saccharide residues connected by a linear glycosidic linkage and arranged around a common, rigid core with highly variable and flexible branches. N-glycans are classified into 3 types: high-mannose, complex and hybrid types. High-mannose types are glycans only elongated with mannose residues beyond the core structure. Complex types are extended with other monosaccharides. Hybrid glycans are bi/tri-antennary glycans with one high-mannose branch and one or two complex branches. Because these chains are bulky, they play an important role in protein folding, helping it to adopt its fully functional 3D structure. Glycosylations are sometimes bound by proteins, such as lectins [11], and act as recognition patterns in signaling pathways. Alterations of the glycosylation pattern are observed in pathological contexts and affect the protein functions in diseases such as cancers, diabetes or cardiovascular diseases [12].

Glycans are highly flexible because of the linear glycosidic linkage between each saccharide residues [13], granting them a high number of degrees of freedom. They are also very reactive structures because of the many hydroxyl groups composing these residues. Under experimental conditions, they are subject to a quick degradation because of this reactivity. Moreover, cells used in protein production and purification, such as *E. coli*, are not able to synthesize glycan structures found in animal organisms. It is therefore very difficult to study glycosylations with experimental methods.

Molecular dynamics (MD) is a useful tool to overcome this problem and give access to conformational information. By using empirical forcefields fitted on experimental observables, MD methods enable the description of molecules at the atomic level. Heavy computational resources and exhaustive sampling are used to calculate motions with a high precision. Considering the crucial impact of glycans on protein function, integrating glycosylations in *in silico* studies as well as developing tools dedicated to the study of MD simulations of glycosylated proteins is a key point towards the understanding and deciphering of protein function and regulation.

Our previous *in silico* studies investigating the impact of sialic acids on glycan's structure and dynamics brought to light the lack of tools dedicated to glycan in the field [14]. This study was the beginning of the development of the Umbrella Visualization method. This method aims at evaluating the area covered by a glycan during a MD trajectory.

This representation helps appreciate the relative flexibility of the glycan during the trajectory. This method demonstrated, for the first time, the impact of sialic acids on the flexibility of glycan chains. Because this study was focused on isolated, solvated glycan chains, the results of this first application of the Umbrella Visualization were restricted to a 2D graphical representation.

We present here the latest developments of this method. The Umbrella Visualization now takes into account both the protein topology and the glycan motions above the protein surface.

The main goal of this original and unique approach is to fully understand how a glycan moiety will structurally impact the protein surface. The dynamic information collected during a MD simulation should help identify the most impacted areas and understand which parts of the protein would be accessible to other molecules such as solvent molecules or even other proteins.

First, we report the main tools already available in the field of glycosylated proteins. Many of these tools focus on the study of static, crystallographic structures. Visualization softwares such as PyMol [15] and VMD [16] integrate visualization tools dedicated to the representation of glycan with the aim of better understanding their conformations. We then present UnityMol, developed with the video game engine Unity. Using this software, we upgraded the Umbrella Visualization from a 2D method giving results as density plots to an integrated tool allowing the visualization of the impact of glycans onto a protein structure during a MD trajectory. The Umbrella Visualization is fully implemented in the UnityMol interface in order to provide a user-

friendly, fully automated tool able to accommodate a wide variety of N-glycan structures. We tested this method on a glycosylated insulin receptor and highlighted the differential impact of bi-antennary and tri-antennary glycans on the protein.

This tool, in conjunction with insights of experimental studies, could help understand mechanisms involving N-glycans and design *in vitro* experiments aimed at understanding their impact on protein functions.

2. Related work

2.1. Glycosylation and modelling

The first step needed to study glycosylated proteins is to identify the glycosylation sites. Most of the time, only the first saccharides from a glycan are present on the .pdb structure files available on the Protein Data Bank at www.rcsb.org [17]. This can help identify some glycosylation sites but most of the time, glycan structures are not available: because of their great flexibility, glycans are not sufficiently ordered to be detected by classical methods used to determine protein structures such as X-ray crystallography [18]. It is thus necessary to use prediction tools in order to identify the glycosylation sites. Several tools are available as web-portals or standalone programs [13,19,20].

Before running MD simulation, it is useful to analyze and understand the structure of glycosylated proteins. Some tools can help identify conformational errors in structures [21,22] while others are dedicated to the study of the torsion angles and saccharidic linkage [23–25]. Carbohydrates have many hydroxyl groups that are very reactive and can interact with their surroundings. This is why it is also important to understand how glycans will interact with the protein residues. GlyVicinity [26] is a tool that statistically analyzes the amino acids closest to a chosen type of carbohydrates.

Once the glycosylation sites are identified, it is necessary to build the glycan structure on the site of interest. Even if the first saccharide residues are present in the .pdb file, the remaining part of the glycan has to be reconstructed, avoiding any aberrant structures or steric clashes with the protein. SHAPE is a tool enabling a user to build any glycan structure [27]. SHAPE also predicts several energetically-favorable conformers using a genetic algorithm.

SWEET II on the glycosciences.de portal is another tool that can be used to build a glycan structure [28]. This glycan can then be used on the GlyProt page. GlyProt allows a user to upload the .pdb file and add the glycan created with SWEET II on one or more glycosylation sites [29]. Glycam-Web (<http://glycam.org/>) and Charmm-gui with the Glycan-Reader tool [30] are also web-servers including a builder dedicated to glycosylated-protein structure.

Very recently, the DoGlycan set of tools [31] was developed in order to perform the same function with the Glycam forcefield, which is dedicated to the study of carbohydrate structures.

2.2. Visualization

Many of the main visualization software integrate the visualization of saccharides in all-atom representations but are often limited to licorice, balls and sticks or van der Waals (VdW) sphere representations. But, few of them propose ways of visualization that highlight the positions and structures of glycosylations on the proteins' surface.

PyMOL includes a plugin called Azahar [32] which is able to build a glycan from a template list of saccharide structures. The plugin adds three specific display modes aimed at simplifying the representation of glycan structures: the cartoon and wire representations show the cycles as non-flat polygons linked by rods and the bead representation shows each cycle as a sphere. This plugin also gives the possibility to compute relevant structural values in order to analyze both static structures or MD simulations.

With the 3D-SNFG plugin [33], the VMD software implements its own graphical representation based on the nomenclature developed

before the second edition of the Essentials of Glycobiology textbook [34,35]. Each monosaccharide is represented by a 3D object based on the corresponding symbol.

UnityMol, an open source platform developed with the Unity engine (<http://unity3d.com/>) in C# language [36] is dedicated to the visualization of biological molecules and uses HyperBalls [37] representation. In the latest versions called SweetUnitymol [38], this software integrates display modes and representations dedicated to saccharides such as ribbons representation or the possibility to color saccharide depending on its nature: for each saccharide, the associated color is determined by the symbol found in the SNFG nomenclature.

The tools described above offer different ways of building, analyzing and visualizing a glycosylated-protein structure. Mazola et al. [39] introduced a workflow leading to the building of a glycosylated system ready for MD simulations. However, the forcefields and nomenclatures can differ from one to another, making it hard to switch between them. Moreover, these tools are mostly dedicated to the building of glycosylated protein systems and the analysis of static, crystallographic structures. Despite the possibility to obtain MD simulations data on glycans and glycosylated proteins, there is a lack of tools dedicated to the analysis of such trajectories.

3. Background

3.1. 2D Umbrella Visualization

It is well known that sialic acids, charged saccharides often found at the extremities of glycan chains, are involved in protein function's regulation [40] or dysfunctions such as in type 2 diabetes [41]. To better understand the influence of sialic acid on protein's function and structure, a molecular dynamics study on sialylated and non-sialylated glycan chains has been initiated. Trajectories for four glycan chains were computed, with glycans bearing two and three branches, with and without sialic acids [14]. For each system, the starting carbohydrates structures were built using the Avogadro software. 500 ns of trajectories were generated with the OPLS-AA forcefield and the Gromacs package 4.6.3. NPT ensemble was used with a 310 K temperature and 1 bar pressure.

In order to plot and evaluate the relative flexibility of glycan chains with and without sialic acids, the Umbrella Visualization was developed. The Umbrella Representation considers the glycan acting as an umbrella protecting the protein surface so that an hypothetical molecule would not be able to interact directly with this part of the protein (Fig. 1A). The glycan is divided into vectors representing its different components: two vectors describe the glycan core. The other vectors describe the branches of the glycan. The first step is to measure the distance between the centers of mass of the saccharide residues defining these vectors for every trajectory step. This is done thanks to the Gromacs [42,43] software and the distance module. Files listing the

coordinates of these vectors during the simulation are generated and analyzed.

An in-house program reads those coordinates in order to place the chain on an orthogonal coordinate system, with the asparagine amino acid at the glycan's base set on the origin. The two core vectors as described on Fig. 1B, are used in order to orient the glycan within this coordinate system. First, the angle between the inner-core vector and the z-axis is calculated so this core can be oriented along this axis. Then the angle between the second vector and the x axis is calculated and the glycan chain is rotated to make this vector coplanar with the xz plan. Finally, the xy coordinates of the last common saccharide residue for both types of glycans (sialylated and non-sialylated) are written and displayed on a density graph.

The non-sialylated bi-antennary glycan showed increased flexibility and explored a much larger region of the plot compared to the sialylated bi-antennary glycan. Desialylation impacted the tri-antennary glycan differently: only one of the three branches showed a significant change in flexibility. Though the influence of sialic acids varies depending on the branching of the glycan, it is clear that sialic acids have an impact on the flexibility and stability of glycosylation chains. The Umbrella Visualization, combined with clustering methods and measurement of angles values during the trajectory, was thus able to demonstrate this influence [14].

3.2. UnityMol

In order to better understand how glycans and sialylation impact proteins, it was then decided to transform this 2D-analysis method into a visualization method aimed at identifying which regions of the protein surface are impacted by a glycan during a MD trajectory.

For this purpose, we have decided to work with an existing open source software, UnityMol, to implement this new functionality. By integrating the Umbrella Visualization to a software already used by the scientific community, we hope to make it user-friendly and broadly available.

As stated previously, UnityMol is a molecular editor, viewer and prototyping platform, coded in C# with the Unity game engine. It already has all the classic visualization modes used in molecular modeling and is even able to automatically detect saccharide residues in order to provide user-friendly visualization modes dedicated to the representation and study of saccharide chains. It also provides a plug-in reading MD trajectories. The Unity engine makes it easy to work either with built-in tools or to implement new features. The wide use of this platform also warrants support from an active community and provides online resources and tutorials. This engine supports different platforms such as Android, Windows, Linux or Mac platforms. Exporting the executable file for these platforms is done at compilation, and saves a great amount of time by avoiding further developments. The Unity graphical interface also displays a Game Mode that allows a user to run

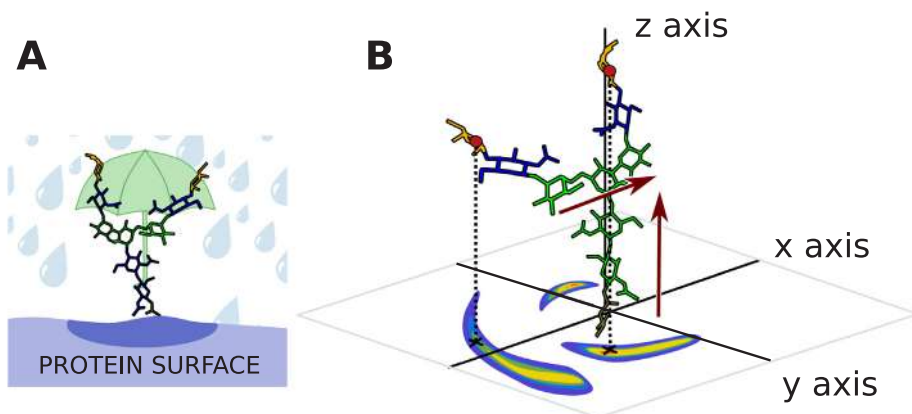


Fig. 1. A) Schematic principle of the Umbrella Visualization: The glycan here is shown in sticks representation and colored by residue type. The Umbrella Representation considers the glycan acting as an umbrella over the protein surface. B) Implementation of the Umbrella Visualization process: The glycan is represented in sticks. The core of the Glycan, shown in green, is aligned along the z axis thanks to vectors describing the core (red arrows). The center of mass (red dots) of the last saccharide residue on the branches is projected on the xy plane. The final result is a density plot reflecting the relative flexibility of the glycan during a MD trajectory. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

its program without compiling it. This feature is especially useful in order to rapidly and easily test adjustments brought to the code. Unity is also known for being relatively user-friendly, partially thanks to its interface, but it is also important to note that every parameters and components are accessible via scripting in C# or Javascript languages, making it a very versatile tool. Thus, the UnityMol prototype developed with this engine is easily modifiable and extensible and provides the means necessary for the implementation of the Umbrella Visualization.

4. System design

The aim of this implementation is to evaluate the covered protein surface during a MD simulation by taking into account the glycan's motions and the protein topology. This enables the display of the 3D structural information. First, the general principle of the shadow projection is detailed. The second part describes the method used to align the plane under the glycan. Then the merging of the information obtained from all the trajectory frames is explained. Finally, the modifications brought to the UnityMol interface in order to implement the Umbrella Visualization are detailed.

4.1. General principle

One of the many advantages of the Unity engine is the possibility to build prefabs that can be used for development purposes. Prefabs are assets storing GameObjects components and setting their properties to desired values. This way, when the prefab is instantiated into a scene, the GameObjects components are already positioned and oriented as desired. Among these components, C# or Javascript scripts can be assigned to a GameObject component and used in order to dictate a behavior to this component. Prefabs are very versatile: even if all the instances of a prefab will share its properties and ensure a certain homogeneity between the instances, individual instances are independently editable and it is thus possible to change some individual properties as needed. Building a prefab and assigning scripts to its component is a good way to build a template for an asset that would be instantiated several times into a Unity scene.

By using these built-in tools and components from the Unity engine, we designed a Unity prefab able to capture the glycan shadow and project it onto the protein surface. This prefab is composed of a camera, a projector and a directional light oriented perpendicular to a flat mesh defining a plane (Fig. 2A). These components are organized together to project the shadow of a glycan on the protein.

Cameras are also defined as GameObjects components and can be used into prefabs. Cameras usually capture and display a view to the user but can also render this view into a texture. To this end, a Render Texture is assigned to a camera. Unity also provides layers that allow for lights and cameras to not illuminate or not render elements from the layers defined in their respective "culling masks".

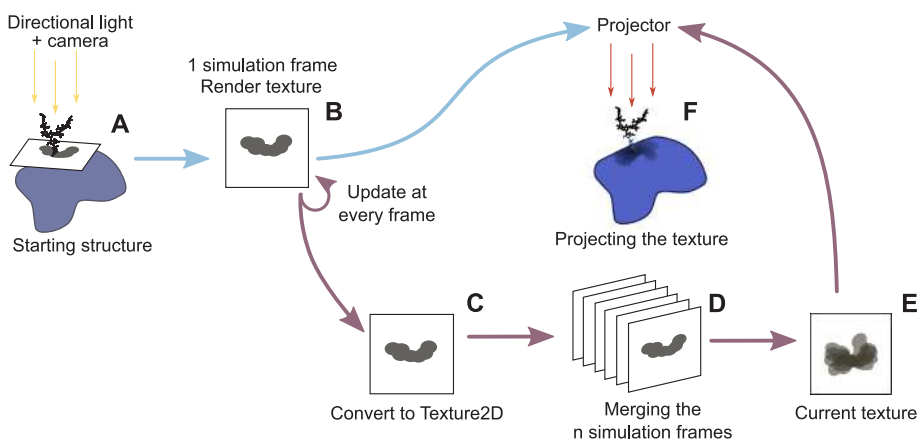


Fig. 2. Unity workflow: The blue arrows describe the workflow to project only one frame. The purple arrows describe the workflow to project the shadow processed from the whole simulation. A directional light projects the glycan's shadow over an oriented plane (A). This shadow is rendered into a render texture by a camera (B) and either displayed on the surface for a static structure or, with a MD trajectory, converted into a Texture2D object (C). This object is read and combined with the images from previous frames (D). All the results acquired along the MD simulation are merged together into a final Texture2D (E) object which is then used in the projector, thus displaying it onto the protein surface (F). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The Projector prefab from the Unity *Standard Assets* package projects a desired material or texture onto all objects intersecting its *frustum*. Projectors have dedicated shaders in which the projected texture is assigned.

One of the many features of the Unity engine is the possibility to quickly instantiate GameObjects of basic shapes such as cubes, spheres or flat planes. As we need a flat surface to project the glycan shadow, the prefab integrates a plane positioned under the glycan. Different types of lights, traditionally used to illuminate a scene, are also available on the Unity engine. As glycan's shadow size must be independent from the light to plane distance, we thus use a directional light. This type of light illuminates scenes with parallel rays, similar to sun rays. The light's orientation is perpendicular to the plane's surface, reproducing an orthogonal space used to orient the glycan.

Once the shadow is displayed, the VdW representation mode has been used in order to reflect the atomic volume of the glycan. The VdW representation mode is based on the VdW radius, a theoretical value representing the atom's volume. In order to only display the shadow of the glycan on the plane, the glycan atom's mesh is set to *Shadow Only* mode. Doing so makes this mesh also invisible to UnityMol's main camera but it is possible to copy the glycan atoms mesh and set it on another layer in order to see the shadow and the atoms at the same time with the prefab's camera ignoring the glycan mesh. Likewise, to prevent the camera to capture parts of the surface going through the plane, the protein surface is set on a dedicated layer and a "culling mask" on this layer is applied to the prefab's camera. This is often due to the irregularity of the protein surface and the plane being close to the surface. This set-up is able to save the glycan's shadow into a render texture (Fig. 2B).

Finally, the shadow-projector uses the texture generated by the prefab displaying the glycan shadow. This texture is then projected onto the protein surface, highlighting parts of the protein surface impacted by the glycan (Fig. 2F). In order to avoid displaying unnecessary elements to the UnityMol's user, the lit-plane prefab is set on another layer and a culling mask is applied on this layer on UnityMol's main camera and light. The plane is thus not visible to the user and only the molecules are displayed on the screen. Using a layer dedicated to the prefab's elements also prevents the main light to interact with the lit plane prefab, which would alter the glycan shadow.

With a similar goal, the camera's and projector's views are set on orthographic mode. With this mode, 3D objects are represented in two dimensions by the means of a parallel projection of the render texture's pixels. The picture projected on the protein surface will not vary with the relative distance to the protein surface. This workflow projects the glycan's shadow onto the protein surface for a static structure. However, this tool is dedicated to the study of data extracted from MD. We have thus to take into account the whole length of a MD trajectory, as described in Section 4.3.

4.2. Orienting the prefab

Whereas the 2D Umbrella Visualization aligned the glycan core along the z axis in order to evaluate the branches' motions, in UnityMol we use the glycosylated asparagine residue as a reference point to place and orient the lit plane prefab. By doing so, we take into account every motion of the glycan on the protein surface, including motions and tilts coming from the core of the glycan.

With the Unity engine, every GameObject has a Transform component. This component has three properties: the Position property determines the position of the Transform in x, y, and z coordinates. The Rotation components determine the rotation, in degrees, on the x, y and z axis. Finally, the Scale component determines the relative size of the object along the x, y and z axis. For the Umbrella Visualization, we use the Position and Rotation properties in order to move and orient the prefab around the glycosylated asparagine.

A normal to a surface is determined by the vector perpendicular to a surface, pointing outward. They are used on meshes to determine how an object will be illuminated by comparing the orientation of the normal to the light's vector. Here, normals define the lit plane's rotation in order to orient it under the glycan. We also use normals pointing outward from the protein surface in order to refine the intensity of the picture projected depending on the orientation of said surface.

Quaternions are based on complex numbers and are used to define and store rotations. Most of the time, their components are not directly modified via scripting, but are re-defined thanks to already existing vectors. Quaternions are easily built from vectors, allowing to compute rotations in a three dimensional space with ease.

As described in Fig. 3A, we define a vector going from the alpha carbon of the asparagine residue to the nitrogen atom involved in the glycosidic linkage (called the Asn vector). This vector is used as the y axis of our coordinate system. When the plane is instantiated in the Unity scene, its Transform component doesn't contain any rotation. The plane normal is thus defined by a vertical vector (0, 1, 0) (Fig. 3B). The rotation going from this normal to the Asn vector is computed for the starting structure and stored into a quaternion (Fig. 3C). The coordinates of the alpha carbon are used in the Transform component of the prefab, hence positioning the center of the plane at this atom's position. The previously built quaternion is then applied on the prefab's Transform component, thus orienting the plane under the glycan with respect to the asparagine's residue (Fig. 3D). As the Asn vector is computed only on the starting structure, it is necessary to align the MD trajectory on the glycan in order to ensure that the glycan and Asparagine residue both stay in the same area during the trajectory read.

Once the plane is instantiated and correctly positioned, the next step is to collect the information for every frame of a molecular dynamics trajectory and compute the statistical, final result (Fig. 3D).

4.3. Frame merging

MD trajectories can be constituted of several thousands of frames to visualize and analyze. Our goal is to provide a tool allowing the user to understand the behavior of the glycan and to see which parts of the proteins are more affected by the glycan's positions and motions. We have to combine or merge the information from all the simulation frames and display the result.

To this end, a script is included on the lit-plane prefab. When the Umbrella Visualization is activated, an integer array is created. The array size is defined by the number of pixels of the final texture. When the simulation frames are loaded, for each frame, the prefab collects the glycan shadow and saves it in a render texture (Fig. 2C). The script converts this render texture into a temporary Texture2D object in order to read the pixels from this texture. Then, for each pixel, the script will detect if the pixel is black (a shadow pixel) or not. The corresponding element in the array will be incremented by 1, thus counting the number of times a pixel is shadowed by the glycan during the trajectory (Fig. 2D). After having read all the frames, the array's values for each pixel are normalized. The higher the count, the darker the pixel will be, highlighting the parts of the protein that are most often obscured by the glycan (Fig. 2E). Pixels color are defined by three components: r, g and b. These components determine the final color of the pixel by adjusting the red, green and blue hues (r, g and b). Using the values from the array, we are able to define the r, g and b elements of the accumulation texture's pixels. Thus, throughout the trajectory, the accumulation of the different positions of the glycan is visualized and at the end of the trajectory, the distribution of the conformations of the glycan is obtained.

During the MD reading, the current shadow is projected onto the surface of the protein so that the user can visualize the motions of the glycan above the protein surface during the trajectory. At the end of the read, the projected texture is switched in order to display the final result compiling the information from all the individual frames.

In order to prevent the overestimation of the impact of the glycan on parts of the protein that are too far from the glycosylation site, the distance from the glycan is evaluated and the rendering of the projected texture is modified accordingly. The projector's shader modulates the projected shadow's intensity by comparing the direction of the projector and the direction of the normals to the protein surface. The intensity of the shadow is altered depending on the angle between those two vectors, as described in Fig. 4A.

The projector also allows the use of a "falloff" texture, meaning that it can modulate the projected shadow's intensity depending on the distance to the projector (Fig. 4B). As described on the next section, the UnityMol interface integrates a box in order to deactivate the falloff texture.

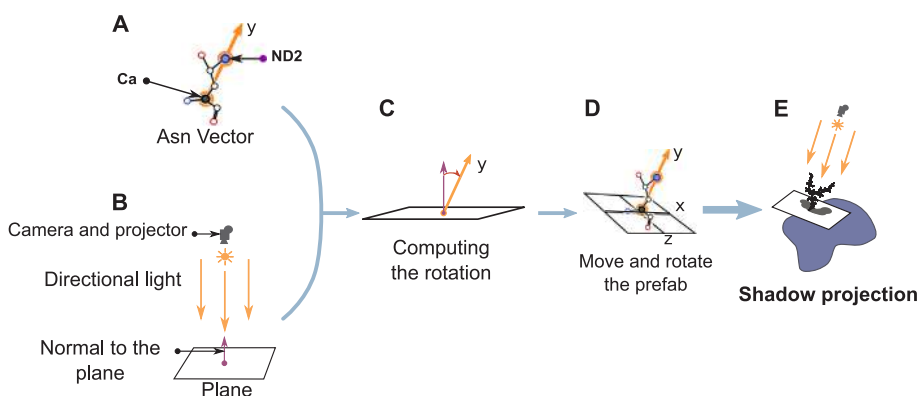


Fig. 3. Positioning and orienting the lit plane prefab with the glycosylated asparagine. A) The coordinates of the alpha carbon and terminal nitrogen atoms define a vector (Asn vector, orange arrow). This vector defines a (x, y, z) coordinate system. B) Instantiated prefab, the normal to the plane is the purple arrow and is defined by a (0,1,0) vector. C) We compute the rotation needed in order to go to the Asn vector from the normal to the plane. D) The center of the plane is moved to the α coordinates and the rotation defined at the previous step is applied to the prefab. E) Once the prefab is correctly positioned, we project the shadow on the plane and save it thanks to the camera. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

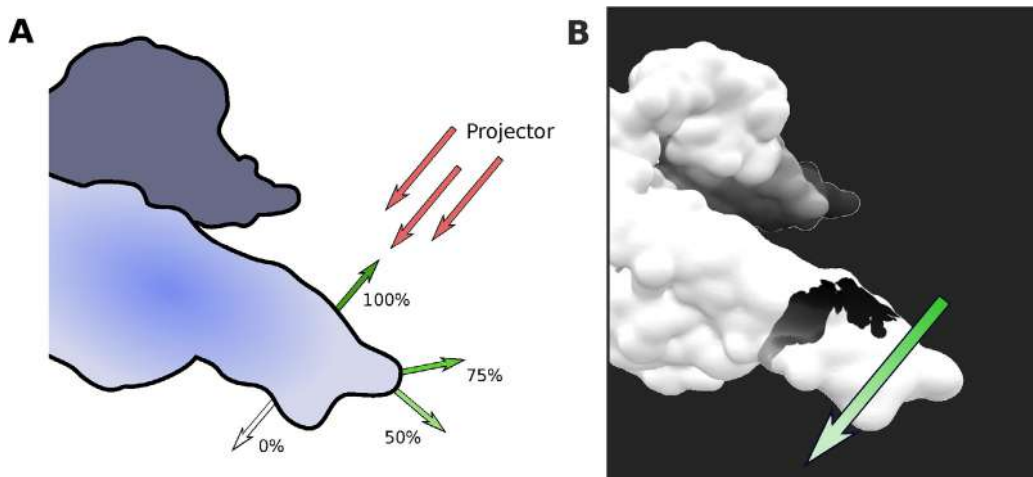


Fig. 4. Fading of the projected shadow on the protein surface. A) Variation of the intensity of the projecting shadow depending on the direction of the normal to the surface. When this normal is facing the projector, the shadow is not faded out. When the normals are facing the same direction as the projector, the shadow is not visible. For other normals, the intensity of the shadow is adjusted proportionally. B) Insulin receptor protein surface with a single-structure shadow projected on the surface. The shadow intensity is also modulated by the distance to the projector. The green and white arrow highlight this directional fading. This snapshot is extracted from the 250 ns MD simulations (for details, refer to Section

5.1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4. User interface

One of the aim of the Umbrella Visualization is to make this method accessible to every UnityMol users and enable the production of pictures for analysis and illustration purposes. This implies to link the described process to the UnityMol interface and works on display modes highlighting clearly the covered parts of the protein surface.

We expanded the UnityMol sugar menu, adding four new boxes. The first one (*Enable* box) starts the Umbrella Visualization once the data are loaded. Detection of the data type (single structure file or molecular dynamics trajectory) is automated so that no other input from the user is needed.

Depending on the input data, the limits of the covered surface may be blurry. In order to have a clear view of this surface, we added a display mode available through the *Contour* button. This display mode modifies the result texture in order to define clear intensity zones, as seen on the Fig. 5. Eight zones are delimited, ranging from black zones (most covered parts) to white/transparent zones (least covered parts). These zones are defined thanks to the number of times a pixel is obscured during the trajectory, as described previously. This gives a clear view of the protein surface impacted by the glycan during the MD trajectory.

Once the calculation is finished and the shadow displayed, the *Save Result* button saves the projected texture in a separate file, giving access to the 2D information besides the 3D information. The raw results can be saved as well as the *Contour* result.

Finally the *Falloff* box enables the fading of the projected shadow with the distance to the projector. Disabling this option gives access to the full, unaltered projected result while activating it could help highlight the influence of the glycan on parts of the protein near the glycosylation site.

5. Preliminary results and application on a biologically relevant system

In order to further our studies on the insulin receptor, we used the Umbrella Visualization on the insulin receptor ectodomain (extracellular part of the protein) dimer. After detailing our systems and the forcefields used in the MD trajectories, we tested the Umbrella Visualization on a small samples of frames of a mono-glycosylated IR in order to evaluate the relevance of this method. Then, thanks to the identification of two glycosylation sites on previous studies, we ran a comparative preliminary study on the impact of bi-antennary and tri-antennary glycans on the insulin receptor. Finally, we focus on the performances of this method by comparing the quality and time needed to compute the results depending on the number of frames to analyze.

5.1. Molecular dynamics set-up

To start our studies on the insulin receptor and the impact of glycosylation, we first ran docking experiments with the insulin receptor and the neuraminidase I subunit of the elastin receptor complex [8]. Correlating a map of the consensus sequence for N-glycosylation with the docking results led to the identification of two potential glycosylation sites on the insulin receptor dimer accessible to the neuraminidase I protein. Trajectories generated with the GROMACS software (5.0.2 package) and the OPLS-AA forcefields, using the NPT ensemble ($P = 1$ bar and $T = 310$ K) were used for this study. Each dimer was glycosylated on two sites, thus creating two systems: the IR with two non-sialylated bi-antennary glycans and the IR with two non-sialylated tri-antennary glycans. 250 ns of trajectories were generated for each system.

The MD trajectory used in the 5.2 is a sample of 250 frames

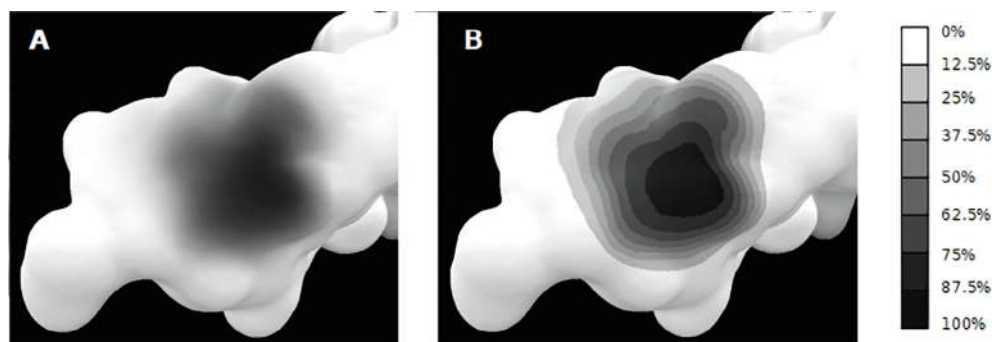


Fig. 5. Display modes available for the Umbrella Visualization. A) classical display, the glycan's shadow is very detailed but blurry. B) The *Contour* display mode gives a better idea of the covered surface by highlighting different zones. The darker parts correspond to the most covered area during the trajectory. The scale to the right gives the thresholds for each zones. These pictures are obtained averaging the data extracted from 12,500 frames of MD simulations of a glycosylated insulin receptor (for details, refer to Section 5.1).

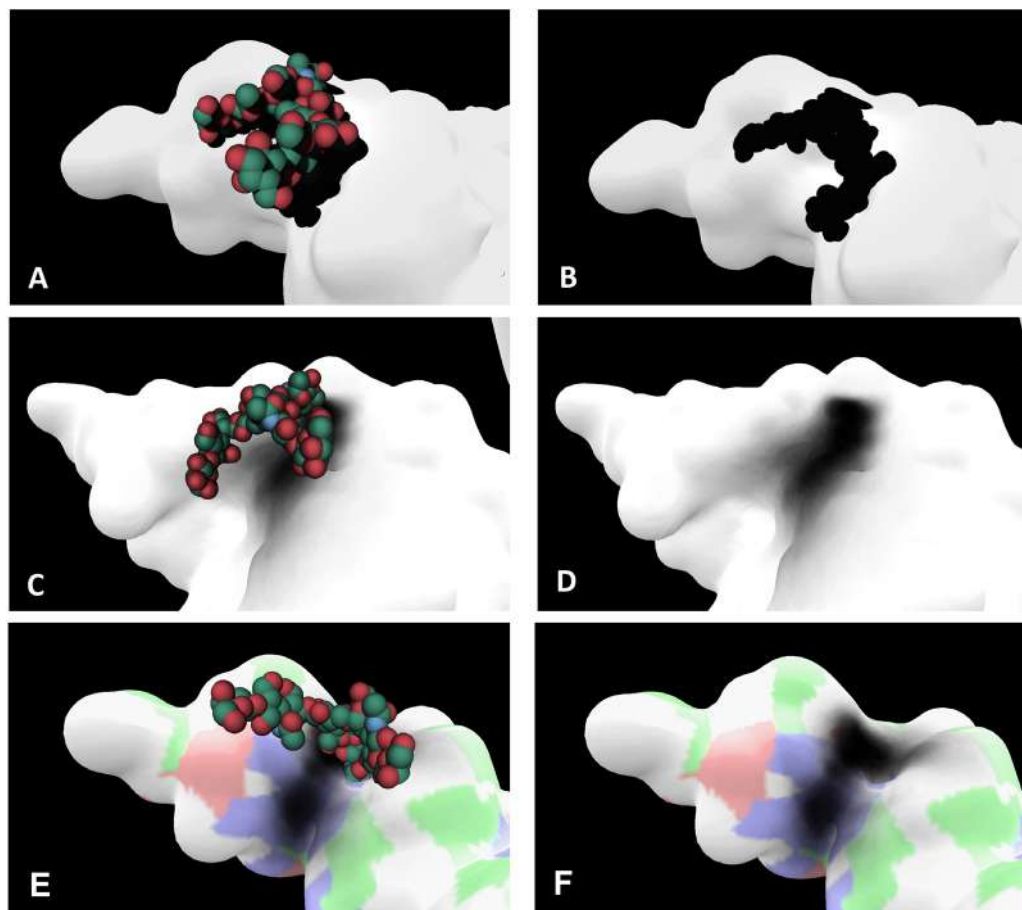


Fig. 6. Relevance of the Umbrella Visualization: On the A, B, C and D panels, the protein surface is shown in white. On the E and F panel, this surface is white with colored patches. The shadow projected on the C, D, E and F panels is the result obtained over the 250 frames (dark grey/black areas). The glycan is displayed on VdW mode. A) Starting structure used for the MD trajectory. The glycan is located near a shallow groove on the protein surface. B) Shadow of the glycan projected on the protein surface on the starting structure. C) This conformation is extracted from the MD simulation and shows how the glycan is positioned on the groove, in a conformation similar to the starting structure. D) On this particular example, it is possible to see that, during this MD trajectory, the glycan's motions are mostly restricted to an area along a small groove on the protein surface. E) and F) show the glycan and the projected shadow over the groove with the surface colored depending on the residues underneath. Blue patches are for basic residues, red patches are for acid residues, green patches are for polar residues, white patches are for non-polar residues. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extracted from the IR-dimer with two non-sialylated bi-antennary glycans.

The results presented in the Section 5.3 were computed on 12500 frames from a 250 ns trajectory (1 frame every 20 ps). These data were also used in order to evaluate the performances of the Umbrella Visualization (Section 5.4).

5.2. Relevance of the Umbrella Visualization

In order to evaluate the relevance of Umbrella Visualization, we used this method on the insulin receptor ectodomain dimer with one bi-antennary glycan per subunit. The Fig. 6 shows the result of the Umbrella Visualization with one frame displayed (A and B panels). The protein surface is displayed in white while the glycan's atoms are represented in VdW spheres. Each sphere corresponds to an atom from the glycan (For clarity, the hydrogen atoms are not displayed). The green spheres are carbon atoms, the red spheres are oxygen atoms. Thanks to the VdW representation, the shadow size matches the volume of the glycan and helps approximate the surface covered for this specific conformation. On the A and B panels, the glycan is oriented above a shallow groove of the protein. Such a conformation on the .pdb file could indicate a preferential conformation due to favorable interactions with the protein surface. To confirm or reject this hypothesis, we ran a MD simulation of the glycosylated insulin receptor ectodomain dimer.

The C and D panels show the projected shadow on a 250 frames trajectory. Because we normalize the r, g, b values with the number of frames, the result is statistically relevant. On this example, we can conclude that during most of the trajectory, as suggested by the .pdb structure, the glycan is extended above the groove but doesn't explore the surrounding part of the protein surface. This could indicate that particular protein residues in this region have favorable interactions

with the saccharide residues of the glycan, hence favoring this conformation. These results thus corroborate the hypothesis made with the static structure. The Fig. 6E and F show that the groove covered by the glycan during most of the 250 frames is composed of positively charged. In order to investigate the nature of residues interacting with the glycan, a color scale corresponding to the physico-chemical nature of residues located under the surface is applied. Thus, red, blue and green patches show the presence of acid, basic and polar residues, respectively. The superimposition of this information and the shadow of the glycan emphasize what are the properties of the residues below that interact with the glycan.

In the context of protein/protein interactions, this is a valuable information as basic, positively charged residues can be involved in these interactions. On this case, these residues would be able to interact with the many hydroxyl groups of the sugar residues, hence stabilizing the position of the glycan in the groove, hindering the interaction of other molecules with this patch. This could highlight how a glycan could influence and modulate protein/protein interactions.

5.3. Comparative study on the impact of bi-antennary and tri-antennary glycans

In order to evaluate the impact of different glycans on the insulin receptors, we ran molecular dynamics simulation on two systems: a glycosylated insulin receptor with bi-antennary glycans and a glycosylated receptor with tri-antennary glycans. We used the Umbrella Visualization on the computed trajectories and compared the results to highlight the differences in the covered protein surface.

On both insulin receptor structures, the global shape of the projected shadow is very similar between the two glycosylation sites, showing a similar behavior for both glycans. However, the number of

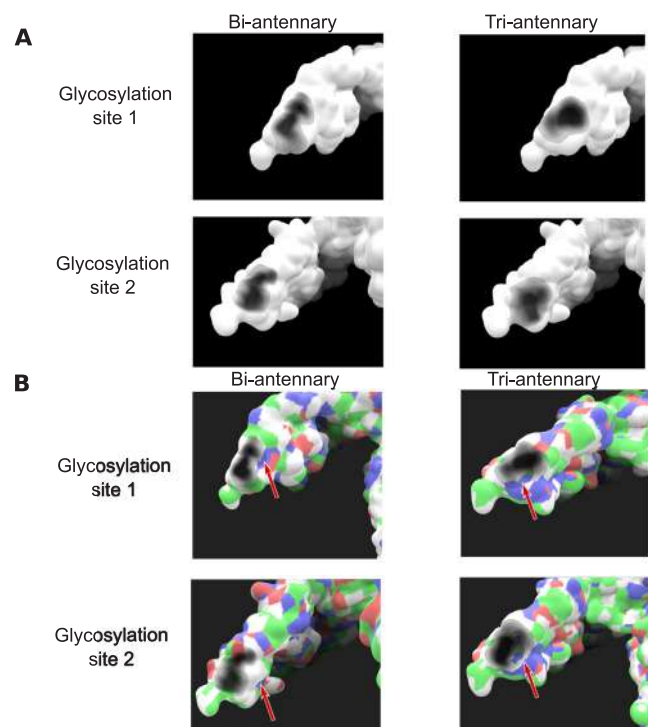


Fig. 7. Results obtained with the Umbrella Visualization on the glycosylated Insulin Receptor. A) Comparison of the different shadows computed. The global shape for the bi-antennary glycans is more stretched along the protein. Tri-antennary glycans display a wider shadow. This is consistent with the fact that this glycan has one extra branch able to explore a wider space. B) Patch of basic residues located near the glycan (red arrows). The glycan's shadow tends to hover around this part of the surface. This could indicate the role of these amino acids in stabilizing the glycan. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

branches seems to have a higher impact on the covered surface. For the bi-antennary glycans, the shadow is elongated lengthwise on the protein (Fig. 7A). The tri-antennary glycans covered surface takes a more rounded or oval shape, suggesting a higher coverage over the length of the trajectory. This is consistent with the higher number of branches on the tri-antennary glycan.

Coloring the protein surface with the amino acid's properties shows that the glycan is located on a part of the protein surface composed of apolar residues (colored in white). Though most of the covered surface is located on this apolar zone, a small patch made by basic residues is partially covered by the glycan (red arrows in the Fig. 7B). Both bi-antennary and tri-antennary glycans seem to cover part of this patch during the trajectory. However, the tri-antennary glycans cover up a larger portion of this patch. This might indicate that the glycan interacts with these amino acid in order to stabilize its conformation over the protein and limits the explored space during the trajectory. Indeed, along the trajectories, the glycans don't explore much of the space on the opposite side of this patch. For the bi-antennary glycan, this could also explain why the elongated shape of the covered surface is oriented towards it.

5.4. Performances

The investigated system on this study is a dimer of the insulin receptor ectodomain with one glycan per monomer. This corresponds to a number of 26030 atoms, not counting the water molecules used to calculate the motions along the trajectory. The Umbrella Visualization analysis was performed with a 3.7 GHz Intel Xeon W-2145 CPU machine with 64 Gb of DDR4 RAM and a Nvidia GeForce GTX 1080Ti graphic card.

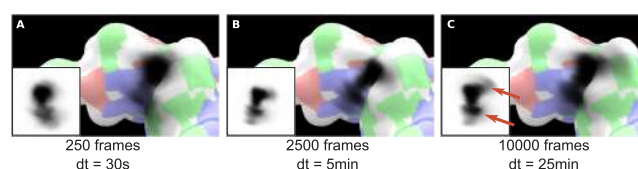


Fig. 8. Results of the Umbrella Visualization on three MD samples. The flat textures projected onto the surfaces are shown as insets. A) Results extracted from 250 frames. The overall tendency is already visible, highlighting the importance of the basic residues. B) Results extracted from 2500 frames. C) Results extracted from 10000 frames. By increasing the number of frames used for the computation, the lower area (bottom red arrow) becomes more compact and darker while a light spread zone (top red arrow) appears in the upper area. This clearly indicates that with a low amount of frames, the role of zones may be overestimated while others are missed. The dt label is the elapsed time for the calculation of these results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The performances of our methodology were calculated through the analysis of three different sets of frames: 250, 2500 and 10,000 frames. We used 512 pixels wide texture. Using a smaller texture saved time (18 min for a 256 pixels wide texture). The main limitation is the number of atoms contained in the structure. The dimer simulated here is a relatively large system, thus processing the data for every frame of the simulation requires more computational resources than a smaller system. In spite of that size, this method still process a significant number of frames in a reasonable time (roughly 30 min), especially when considering the amount of time necessary to collect MD data (almost 8 days accumulating 32 ns/day using 480 CPUs).

As indicated on Fig. 8, the rendering of the Umbrella Visualization through 250, 2500 and 10,000 frames took approximately 30 s, 5 min and 25 min respectively. As expected, more information is available with a higher number of frames. Though the more prominent conformations are already displayed on the 250 frames sample (Fig. 8A), the two other samples bring to light the impact of less represented conformations (Fig. 8B and Fig. 8C). In every case, the combined frames show that the glycan have a tendency to contact the patch of basic residues on the groove.

6. Discussion and concluding remarks

The Umbrella Visualization is a new method of analysis using a camera and a projector to project the shadow of a glycan on the protein surface. Shadows rendering of the glycan is performed and processed by counting the number of times a pixel is obscured by the glycan and determine the color of the final texture depending on this number. During the trajectory read, the individual frame's shadows are projected on the protein surface, highlighting all the single events of the simulation. As the Umbrella Visualization is also available for single-structure .pdb files, this helps to put into evidence rare events occurring during a trajectory or a representative conformation, for demonstration and illustration purposes.

The integration of the Umbrella Visualization on the UnityMol interface overcomes the lack of flexibility and the cumbersome process of the former 2D version [14] since it offers a fully automated and user-friendly method able to accommodate all N-glycan structures. Saving the computed results is possible thanks to the *Save result* button. Moreover, the UnityMol interface provides tools to save pictures of the projected results on the protein surface.

The design of clear pictures for illustration purposes is offered through two display modes related to the projected shadow: the native computed results and a simplified representation delimiting clear zones depending on the glycan's impact on the protein surface. Because comparing results between trajectories can be harder on a 3D protein surface, the projected texture can be saved in a picture file. As illustrated in this study, this helps emphasize interesting results and

comparisons between the results since both the 2D information and the 3D information are available.

The relevance and performance of the method was evaluated considering different number of frames within a MD trajectory: namely 250, 2500 and 10,000 frames. We were able to see significant results even with a small numbers of frames that allowed us to highlight the main conformational tendencies of the glycan of interest. The largest samples showed more details and hinted at less represented conformations.

This emphasizes the high importance of the number of frames. Because of the statistical nature of the Umbrella Visualization, working with a low number of frames could lead to the overestimation of some conformations and the underestimation of others, meaning that we could miss relevant information. The choice of trajectory length is thus a key point regarding the significance of the Umbrella Visualization results. Nonetheless, this first experiment showed that this method has the potential to give informative results that could also either help design *in vitro* experiments and/or corroborate data from other calculations.

The Umbrella Visualization was applied in order to investigate the impact of glycans on the insulin receptor. We ran molecular dynamics simulation on the glycosylated protein and analyzed the trajectories. The Umbrella Visualization showed the different behaviour between the bi-antennary and tri-antennary glycans on the insulin receptor surface. We were able to see the impact of the third branch, extending the surface covered during the trajectory. Common behaviours between these MD trajectories also help to highlight a patch of basic residues that could be involved in protein/glycans interactions.

Although the glycan structures we used here are very common, we also know that, in some pathological contexts such as cancers, alteration of the glycosylation process leads to the synthesis of abnormal structures. Adding more branches could heavily influence the way proteins (With the insulin receptor, a known interaction partner is the neuraminidase I, for instance) interact with each other. For example, the profiling of the N-glycome of breast cancer cells showed that tri-antennary and tetra-antennary glycans were detected only in cancerous cells [44]. Investigating how multi-antennary glycans impact protein surfaces could bring new light on the influence of disruption of the glycosylation process.

We also showed that the glycans had a tendency to shift toward a patch of basic residues above the protein surface. These residues would be able to interact with the many hydroxyl groups of the sugar residues, hence stabilizing the position of the glycan above this region. Knowing how basic residues can be key amino acids in protein/protein interactions [45], it would be interesting to investigate the role of the residues composing this patch. We can imagine that these basic residues play a role in stabilizing the glycan in order to allow other proteins such as the neuraminidase I to interact with it. By correlating these results with the docking experiment previously ran, we bring new informations regarding the interaction with the neuraminidase I and the glycosylated insulin receptor.

We have described in previous studies how the insulin receptor's function was altered by the hydrolysis of sialic acids [8] and that these saccharide had an impact on the flexibility of glycans [14], the next step will be to investigate how this change in flexibility impacts the insulin receptor function. By applying the Umbrella Visualization on an insulin receptor bearing fully sialylated glycans, we hope to point out important interactions involved in this mechanism. Because sialic acids are negatively charged, they have the potential to interact strongly with positively charged residues. It would be interesting to see if the basic patch we highlighted in our results has any influence on sialylated glycans.

The results presented in this paper showed the relevance and potential of the Umbrella Visualization in molecular modeling studies. In the near future, we plan to improve this method with further developments.

Currently, this methodology is usable for the study of one glycan at a time. Because some proteins are heavily glycosylated, we will optimize this to enable the visualization and analysis of multi-glycosylated proteins, meaning that their molecular weight is increased. This process heavily relies on the use of a prefab, so it is possible to instantiate one object several times in the same UnityMol session while applying transformations to components of single instances of the prefab. Each instance of the plane assigned to a glycan can thus be oriented accordingly. The Umbrella Visualization method could help investigate how glycans modulate protein-protein interfaces in biological contexts. For example, it has been shown that the human coronavirus NL63, a virus causing severe lower-respiratory-tract infections, uses a glycan shield to protect surface epitopes from the host immune system [46]. The role and impact of this shield could be unraveled with the Umbrella Visualization.

It is also important to note that the surface receiving the shadow is based on the initial .pdb file loaded in UnityMol and doesn't change during the course of the trajectory (reduced computational work). Because proteins are also dynamical objects, it would be interesting to implement a quicker surface calculation and alter the surface shape for each frame. This would indicate how the protein evolves and adapts to the glycan motions. This could also impact the regions on which the shadow is projected and thus better highlight the influence of the glycan on the protein.

In a similar fashion, the lit plane prefab's position and orientation is defined by the atom's position on the starting structure and is not moving during the MD trajectory. This means that, in order to have significant results, the MD trajectory must be centered on the glycan so that it stays on the same area during the trajectory. While this is easily done with most MD software, this could still introduce bias on the results. To avoid those bias and generate results as precise as possible, we wish to recompute the plane's position and orientation for every trajectory frame.

Local topology such as groove, bulges or the presence of other protein domains can also influence how this interaction would take place. On this first version of the Umbrella Visualization, the directional light is oriented perpendicularly to the plane, reproducing an orthogonal space. This introduces a bias into the results as it doesn't reflect how another molecule would approach a protein in order to interact or form a complex. To correct this, we wish to allow the prefab to move around the glycan, following a given position defined by the user, and then visualize the results with this new orientation.

The Umbrella Visualization brings valuable structural information about the impact of glycans, but doesn't offer any quantifiable value that would help evaluate more concretely the space occupied by glycans. However, the 2D graphical method previously developed took the form of a density plot [14], giving statistical informations but no structural information. Correlating the results of the 2D graphical method and the results of the 3D visual method is a good way to provide quantifiable data to support the visualization results.

To support this, we also plan to add methods to bring more informations to the Umbrella Visualization results, such as the occupied volume during the trajectory: as glycans are bulky, the space they explore and occupy during a MD trajectory could also inform us on their influence on molecular complex formation.

More generally, correlating the Umbrella Visualization results with other calculations could bring new insight into the impact of glycans on proteins. Among the methods used in molecular modeling, free-energy calculations are used to characterized interactions between partners [47]: it is possible to evaluate the contribution from each residue from both partners. It would be interesting to combine the Umbrella Visualization with free-energy calculation to emphasize how the protein's amino acids influence the glycan's behaviour and favor interactions with the glycan's hydroxyl groups. By displaying both the results of such calculation and the results of the Umbrella Visualization on the protein surface, it would be possible to correlate the binding free energy

with the glycan's motions.

Molecular modeling is also able to help designing *in vitro* experiments. Molecular docking, for example, is used to design new drugs thanks to its ability to test several molecules without having to synthesize them [48]. Identification of the key interactions can further the understanding on how a drug will interact with a protein and influence its function. In a similar way, with the Umbrella Visualization, we wish to provide a tool able to help the design of studies focused on glycosylations. By highlighting which parts of the protein are affected by glycans, it is possible to identify important protein residues that are good candidates for point mutations in order to understand their impact on glycans and protein structure and dynamics.

In conclusion, we propose a new method available on the UnityMol viewer called the Umbrella Visualization. This method is dedicated to the study of glycosylated protein and highlights regions of the protein surface hidden by the glycan during a molecular dynamics trajectory. These regions are thus inaccessible to other molecules, emphasizing the impact of glycans on protein surfaces. By implementing this method into the UnityMol interface, the Umbrella Visualization becomes user-friendly and efficient, bringing us first results about the impact of glycans on the insulin receptor.

Acknowledgements

This work was supported by grants from the Region Grand Est and the ERDF (European Regional Development Fund). The authors thank the HPC-Regional Center ROME, the Multiscale Molecular Modeling Platform (P3M) and the national HPC-CINES Center for providing CPU-time and support. The chair MaGICS is acknowledged for financial and technical support.

References

- [1] J. Halper, M. Kjaer, Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins, *Adv. Exp. Med. Biol.* 802 (2014) 31–47.
- [2] A. Pozzi, P.D. Yurchenco, R.V. Iozzo, The nature and biology of basement membranes, *Matrix Biol.* 57–58 (2017) 1–11.
- [3] A.D. Theodoris, S.S. Skandalis, C. Gialeli, N.K. Karamanos, Extracellular matrix structure, *Adv. Drug. Deliv. Rev.* 97 (2016) 4–27.
- [4] F.X. Maquart, A. Siméon, S. Pasco, J.C. Monboisse, Regulation of cell activity by the extracellular matrix: the concept of matrikines, *J. Soc. Biol.* 193 (1999) 423–428.
- [5] C. Bonnans, J. Chou, Z. Werb, Remodelling the extracellular matrix in development and disease, *Nat. Rev. Mol. Cell. Biol.* 15 (2014) 786–801.
- [6] C. Kaweck, O. Bocquet, C.E.H. Schmelzer, A. Heinz, C. Ihling, A. Wahart, B. Romier, A. Bennisroune, S. Blaise, C. Terryn, K.J. Linton, L. Martiny, L. Duca, P. Maurice, Identification of CD36 as a new interaction partner of membrane NEU1: potential implication in the pro-atherogenic effects of the elastin receptor complex, *Cell. Mol. Life Sci.* 76 (2019) 791–807.
- [7] B. Romier, C. Ivaldi, H. Sartelet, A. Heinz, C.E.H. Schmelzer, R. Garnotel, A. Guillot, J. Jonquet, E. Bertin, J.-L. Guéant, J.-M. Alberto, J.-P. Bronowicki, J. Amoyel, T. Hocine, L. Duca, P. Maurice, A. Bennisroune, L. Martiny, L. Debelles, V. Durlach, S. Blaise, Production of elastin-derived peptides contributes to the development of nonalcoholic steatohepatitis, *Diabetes* 67 (2018) 1604–1615.
- [8] S. Blaise, B. Romier, C. Kaweck, M. Ghirardi, F. Rabenoelina, S. Baud, L. Duca, P. Maurice, A. Heinz, C.E. Schmelzer, M. Tarpin, L. Martiny, C. Garbar, M. Dauchez, L. Debelles, V. Durlach, Elastin-derived peptides are new regulators of insulin resistance development in mice, *Diabetes* 62 (2013) 3807–3816.
- [9] J.B. Hwang, J. Hernandez, R. Leduc, S.C. Frost, Alternative glycosylation of the insulin receptor prevents oligomerization and acquisition of insulin-dependent tyrosine kinase activity, *Biochim. Biophys. Acta* 1499 (2000) 74–84.
- [10] Y. Gavel, G. von Heijne, Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering, *Protein Eng.* 3 (1990) 433–442.
- [11] M. Nagae, Y. Yamaguchi, Sugar recognition and protein-protein interaction of mammalian lectins conferring diverse functions, *Curr. Opin. Struct. Biol.* 34 (2015) 108–115.
- [12] P. Stanley, N. Taniguchi, M. Aebi, N-glycans, in: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger (Eds.), *Essentials of Glycobiology*, third ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015, pp. 99–111.
- [13] G.-Y. Chuang, J.C. Boyington, M.G. Joyce, J. Zhu, G.J. Nabel, P.D. Kwong, I. Georgiev, Computational prediction of N-linked glycosylation incorporating structural properties and patterns, *Bioinformatics* 28 (2012) 2249–2255.
- [14] A. Guillot, M. Dauchez, N. Belloy, J. Jonquet, L. Duca, B. Romier, P. Maurice, L. Debelles, L. Martiny, V. Durlach, S. Baud, S. Blaise, Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans, *Sci. Rep.* 6 (2016) 35666.
- [15] L.L.C. Schrodinger, The PyMOL Molecular Graphics System, Version 1.8, 2015.
- [16] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (33–38) (1996) 27–28.
- [17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [18] M.R. Wormald, A.J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott, R.A. Dwek, Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling, *Chem. Rev.* 102 (2002) 371–386.
- [19] S.E. Hamby, J.D. Hirst, Prediction of glycosylation sites using random forests, *BMC Bioinf.* 9 (2008) 500.
- [20] R. Gupta, E. Jung, S. Brunak, Prediction of N-glycosylation sites in human proteins, in preparation (2004).
- [21] T. Lutteke, M. Frank, C.-W. von der Lieth, Data mining the protein data bank: automatic detection and assignment of carbohydrate structures, *Carbohydr. Res.* 339 (2004) 1015–1020.
- [22] T. Lutteke, C.-W. von der Lieth, pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files, *BMC Bioinf.* 5 (2004) 69.
- [23] A. Böhne-Lang, E. Lang, T. Frster, C.W. von der Lieth, LINUCS: linear notation for unique description of carbohydrate sequences, *Carbohydr. Res.* 336 (2001) 1–11.
- [24] M. Frank, T. Lutteke, C.-W. von der Lieth, GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages, *Nucleic Acids Res.* 35 (2007) 287–290.
- [25] T. Lutteke, M. Frank, C.-W. von der Lieth, Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB, *Nucleic Acids Res.* 33 (2005) D242–D246.
- [26] M.A. Rojas-Macias, T. Lutteke, Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by GlyVicinity, *Methods Mol. Biol.* 1273 (2015) 215–226.
- [27] J. Rosen, L. Miguet, S. Perez, Shape: automatic conformation prediction of carbohydrates using a genetic algorithm, *J. Cheminform.* 1 (2009) 16.
- [28] A. Böhne, E. Lang, C.-W. von der Lieth, W3-SWEET: carbohydrate modeling by internet, *J. Mol. Model.* 4 (1998) 33–43.
- [29] A. Böhne-Lang, C.-W. von der Lieth, GlyProt: in silico glycosylation of proteins, *Nucleic Acids Res.* 33 (2005) W214–219.
- [30] S. Jo, K.C. Song, H. Desaire, A.D. MacKerell, W. Im, Glycan reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins, *J. Comput. Chem.* 32 (2011) 3135–3141.
- [31] R. Danne, C. Poojari, H. Martinez-Seara, S. Rissanen, F. Lolicato, T. Rág, I. Vattulainen, doGlycans-tools for preparing carbohydrate structures for atomistic simulations of glycoproteins, glycolipids, and carbohydrate polymers for GROMACS, *J. Chem. Inf. Model.* 57 (2017) 2401–2406.
- [32] A. Arroyuelo, J.A. Vila, O.A. Martin, Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules, *J. Comput. Aid Mol. Des.* 30 (2016) 619–624.
- [33] D.F. Thieker, J.A. Hadden, K. Schulten, R.J. Woods, 3D implementation of the symbol nomenclature for graphical representation of glycans, *Glycobiology* 26 (2016) 786–787.
- [34] A. Varki, R.D. Cummings, M. Aebi, N.H. Packer, P.H. Seeberger, J.D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J.J. Prestegard, R.L. Schnaar, H.H. Freeze, J.D. Marth, C.R. Bertozzi, M.E. Etzler, M. Frank, J.F. Vliegthart, T. Látteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Torkach, K.F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, S. Kornfeld, Symbol nomenclature for graphical representations of glycans, *Glycobiology* 25 (2015) 1323–1324.
- [35] A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, M.E. Etzler (Eds.), *Essentials of Glycobiology*, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.
- [36] Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent, M. Baaden, Game on, science – how video game technology may help biologists tackle visualization challenges, *PLoS One* 8 (2013) e57990.
- [37] M. Chavent, A. Vanel, A. Tek, B. Levy, S. Robert, B. Raffin, M. Baaden, GPU-accelerated atom and dynamic bond visualization using hyperballs: a unified algorithm for balls, sticks, and hyperboloids, *J. Comput. Chem.* 32 (2011) 2924–2935.
- [38] S. Perez, T. Tubiana, A. Imbert, M. Baaden, Three-dimensional representations of complex carbohydrates and polysaccharides–SweetUnityMol: a video game-based computer graphic software, *Glycobiology* 25 (2015) 483–491.
- [39] Y. Mazola, G. China, A. Musacchio, Integrating bioinformatics tools to handle glycosylation, *PLoS Comput. Biol.* 7 (2011) e1002285.
- [40] R. Schauer, Sialic acids as regulators of molecular and cellular interactions, *Curr. Opin. Struct. Biol.* 19 (2009) 507–514.
- [41] V. Dotz, R.F.H. Lemmers, K.R. Reidling, A.L. Hipgrave Ederveen, A.G. Lieverse, M.T. Mulder, E.J.G. Sijbrands, M. Wuhrer, M. van Hoek, Plasma protein N-glycan signatures of type 2 diabetes, *Biochim. Biophys. Acta* 1862 (2018) 2613–2622.
- [42] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56.
- [43] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit,

- Bioinformatics 29 (2013) 845–854.
- [44] L.Y. Lee, M. Thaysen-Andersen, M.S. Baker, N.H. Packer, W.S. Hancock, S. Fanayan, Comprehensive N-glycome profiling of cultured human epithelial breast cells identifies unique secretome N-glycosylation signatures enabling tumorigenic subtype classification, *J. Proteome Res.* 13 (2014) 4783–4795.
- [45] N. Sinha, S.J. Smith-Gill, Electrostatics in protein binding and function, *Curr. Protein Pept. Sci.* 3 (2002) 601–614.
- [46] A.C. Walls, M.A. Tortorici, B. Frenz, J. Snijder, W. Li, F.A. Rey, F. DiMaio, B.-J. Bosch, D. Veesler, Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy, *Nat. Struct. Mol. Biol.* 23 (2016) 899–905.
- [47] C.D. Christ, A.E. Mark, W.F. van Gunsteren, Basic ingredients of free energy calculations: a review, *J. Comput. Chem.* 31 (2010) 1569–1582.
- [48] P. Sledz, A. Caflisch, Protein structure-based drug design: from docking to molecular dynamics, *Curr. Opin. Struct. Biol.* 48 (2018) 93–102.

Bibliographie

- [1] M. P. Czech and J. Massague, "Subunit structure and dynamics of the insulin receptor," *Federation Proceedings*, vol. 41, pp. 2719–2723, Sept. 1982.
- [2] L. G. Sparrow, N. M. McKern, J. J. Gorman, P. M. Strike, C. P. Robinson, J. D. Bentley, and C. W. Ward, "The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain," *The Journal of Biological Chemistry*, vol. 272, pp. 29460–29467, Nov. 1997.
- [3] L. Schäffer and L. Ljungqvist, "Identification of a disulfide bridge connecting the alpha-subunits of the extracellular domain of the insulin receptor," *Biochemical and Biophysical Research Communications*, vol. 189, pp. 650–653, Dec. 1992.
- [4] C. W. Ward, "Members of the insulin receptor family contain three fibronectin type III domains.," *Growth factors (Chur, Switzerland)*, vol. 16, no. 4, pp. 315–322, 1999.
- [5] B. Zhang, J. M. Tavaré, L. Ellis, and R. A. Roth, "The regulatory role of known tyrosine autophosphorylation sites of the insulin receptor kinase domain. An assessment by replacement with neutral and negatively charged amino acids," *The Journal of Biological Chemistry*, vol. 266, pp. 990–996, Jan. 1991.
- [6] H. Benecke, J. S. Flier, and D. E. Moller, "Alternatively spliced variants of the insulin receptor protein. Expression in normal and diabetic human tissues.."
- [7] B. Vogt, J. M. Carrascosa, B. Ermel, A. Ullrich, and H. U. Häring, "The two isotypes of the human insulin receptor (HIR-A and HIR-B) follow different internalization kinetics," *Biochemical and Biophysical Research Communications*, vol. 177, pp. 1013–1018, June 1991.
- [8] A. Guillot, *Couplage "complexe récepteur de l'élastine / récepteur de l'insuline" : la désialylation des glycanes comme facteur d'insulino résistance*. thesis, Reims, Jan. 2017.
- [9] J. M. Tavaré, R. M. O'Brien, K. Siddle, and R. M. Denton, "Analysis of insulin-receptor phosphorylation sites in intact cells by two-dimensional phosphopeptide mapping," *The Biochemical Journal*, vol. 253, pp. 783–788, Aug. 1988.
- [10] T. Issad, J. M. Tavaré, and R. M. Denton, "Analysis of insulin receptor phosphorylation sites in intact rat liver cells by two-dimensional phosphopeptide mapping. Predominance

- of the tris-phosphorylated form of the kinase domain after stimulation by insulin,” *The Biochemical Journal*, vol. 275 (Pt 1), pp. 15–21, Apr. 1991.
- [11] J. Lee and P. F. Pilch, “The insulin receptor : structure, function, and signaling,” *Am. J. Physiol.*, vol. 266, pp. C319–334, Feb. 1994.
- [12] L. G. Sparrow, M. C. Lawrence, J. J. Gorman, P. M. Strike, C. P. Robinson, N. M. McKern, and C. W. Ward, “N-linked glycans of the human insulin receptor and their distribution over the crystal structure,” *Proteins*, vol. 71, pp. 426–439, Apr. 2008.
- [13] T. C. Elleman, M. J. Frenkel, P. A. Hoyne, N. M. McKern, L. Cosgrove, D. R. Hewish, K. M. Jachno, J. D. Bentley, S. E. Sankovich, and C. W. Ward, “Mutational analysis of the N-linked glycosylation sites of the human insulin receptor.,” *Biochemical Journal*, vol. 347, pp. 771–779, May 2000.
- [14] S. Blaise, B. Romier, C. Kawecki, M. Ghirardi, F. Rabenoelina, S. Baud, L. Duca, P. Maurice, A. Heinz, C. E. Schmelzer, M. Tarpin, L. Martiny, C. Garbar, M. Dauchez, L. Debelle, and V. Durlach, “Elastin-Derived Peptides Are New Regulators of Insulin Resistance Development in Mice,” *Diabetes*, vol. 62, pp. 3807–3816, Nov. 2013.
- [15] A. Scandolera, L. Odoul, S. Salesse, A. Guillot, S. Blaise, C. Kawecki, P. Maurice, H. El Btaouri, B. Romier-Crouzet, L. Martiny, L. Debelle, and L. Duca, “The Elastin Receptor Complex : A Unique Matricellular Receptor with High Anti-tumoral Potential,” *Frontiers in Pharmacology*, vol. 7, p. 32, 2016.
- [16] A. Bennasroune, B. Romier-Crouzet, S. Blaise, M. Laffargue, R. G. Efremov, L. Martiny, P. Maurice, and L. Duca, “Elastic fibers and elastin receptor complex : Neuraminidase-1 takes the center stage,” *Matrix Biology : Journal of the International Society for Matrix Biology*, June 2019.
- [17] A. Guillot, M. Dauchez, N. Belloy, J. Jonquet, L. Duca, B. Romier, P. Maurice, L. Debelle, L. Martiny, V. Durlach, S. Baud, and S. Blaise, “Impact of sialic acids on the molecular dynamic of bi-antennary and tri-antennary glycans,” *Scientific Reports*, vol. 6, p. 35666, Oct. 2016.
- [18] J. Mazurier, M. Dauchez, G. Vergoten, J. Montreuil, and G. Spik, “Molecular modeling of a disialylated monofucosylated biantennary glycan of the N-acetyllactosamine type,” *Glycoconj. J.*, vol. 8, pp. 390–399, Oct. 1991.
- [19] M. L. Tanzer, “Current concepts of extracellular matrix,” *J Orthop Sci*, vol. 11, pp. 326–331, May 2006.
- [20] A. D. Theocharis, S. S. Skandalis, C. Gialeli, and N. K. Karamanos, “Extracellular matrix structure,” *Advanced Drug Delivery Reviews*, vol. 97, pp. 4–27, Feb. 2016.

- [21] A. Page-McCaw, A. J. Ewald, and Z. Werb, "Matrix metalloproteinases and the regulation of tissue remodelling," *Nat. Rev. Mol. Cell Biol.*, vol. 8, pp. 221–233, Mar. 2007.
- [22] E. M. Culav, C. H. Clark, and M. J. Merrilees, "Connective tissues : matrix composition and its relevance to physical therapy," *Phys Ther*, vol. 79, pp. 308–319, Mar. 1999.
- [23] K. M. Mak and R. Mei, "Basement Membrane Type IV Collagen and Laminin : An Overview of Their Biology and Value as Fibrosis Biomarkers of Liver Disease," *Anat Rec (Hoboken)*, vol. 300, no. 8, pp. 1371–1390, 2017.
- [24] M.-P. Jacob, "Matrice extracellulaire et vieillissement vasculaire," *Med Sci (Paris)*, vol. 22, pp. 273–278, Mar. 2006.
- [25] S. Kalamajski and A. Oldberg, "The role of small leucine-rich proteoglycans in collagen fibrillogenesis," *Matrix Biol.*, vol. 29, pp. 248–253, May 2010.
- [26] A. Oldberg, S. Kalamajski, A. V. Salnikov, L. Stuhr, M. Morgelin, R. K. Reed, N.-E. Heldin, and K. Rubin, "Collagen-binding proteoglycan fibromodulin can determine stroma matrix structure and fluid balance in experimental carcinoma," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, pp. 13966–13971, Aug. 2007.
- [27] R. V. Iozzo and L. Schaefer, "Proteoglycan form and function : A comprehensive nomenclature of proteoglycans," *Matrix Biol.*, vol. 42, pp. 11–55, Mar. 2015.
- [28] S. Brezillon, L. Venteo, L. Ramont, M.-F. D’Onofrio, C. Perreau, M. Pluot, F.-X. Maquart, and Y. Wegrowski, "Expression of lumican, a small leucine-rich proteoglycan with antitumour activity, in human malignant melanoma," *Clin. Exp. Dermatol.*, vol. 32, pp. 405–416, July 2007.
- [29] J. Halper and M. Kjaer, "Basic components of connective tissues and extracellular matrix : elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins," *Adv. Exp. Med. Biol.*, vol. 802, pp. 31–47, 2014.
- [30] C. Frantz, K. M. Stewart, and V. M. Weaver, "The extracellular matrix at a glance," *J Cell Sci*, vol. 123, pp. 4195–4200, Dec. 2010.
- [31] K. E. Kadler, C. Baldock, J. Bella, and R. P. Boot-Handford, "Collagens at a glance," *J. Cell. Sci.*, vol. 120, pp. 1955–1958, June 2007.
- [32] S. Ricard-Blum, "The collagen family," *Cold Spring Harb Perspect Biol*, vol. 3, p. a004978, Jan. 2011.
- [33] S. Ricard-Blum and F. Ruggiero, "The collagen superfamily : from the extracellular matrix to the cell membrane," *Pathol. Biol.*, vol. 53, pp. 430–442, Sept. 2005.
- [34] M. D. Shoulders and R. T. Raines, "Collagen structure and stability," *Annu. Rev. Biochem.*, vol. 78, pp. 929–958, 2009.

- [35] J. E. Wagenseil and R. P. Mecham, “New insights into elastic fiber assembly,” *Birth Defects Res. C Embryo Today*, vol. 81, pp. 229–240, Dec. 2007.
- [36] L. D. Muiznieks, A. S. Weiss, and F. W. Keeley, “Structural disorder and dynamics of elastin,” *Biochem. Cell Biol.*, vol. 88, pp. 239–250, Apr. 2010.
- [37] F. X. Maquart, A. Siméon, S. Pasco, and J. C. Monboisse, “[Regulation of cell activity by the extracellular matrix : the concept of matrikines],” *J. Soc. Biol.*, vol. 193, no. 4-5, pp. 423–428, 1999.
- [38] J. M. Wells, A. Gaggar, and J. E. Blalock, “MMP generated matrikines,” *Matrix Biol.*, vol. 44-46, pp. 122–129, July 2015.
- [39] C. Singh, R. K. Shyanti, V. Singh, R. K. Kale, J. P. N. Mishra, and R. P. Singh, “Integrin expression and glycosylation patterns regulate cell-matrix adhesion and alter with breast cancer progression,” *Biochem. Biophys. Res. Commun.*, vol. 499, no. 2, pp. 374–380, 2018.
- [40] C.-T. Hsiao, H.-W. Cheng, C.-M. Huang, H.-R. Li, M.-H. Ou, J.-R. Huang, K.-H. Khoo, H. W. Yu, Y.-Q. Chen, Y.-K. Wang, A. Chiou, and J.-C. Kuo, “Fibronectin in cell adhesion and migration via N-glycosylation,” *Oncotarget*, vol. 8, pp. 70653–70668, Sept. 2017.
- [41] O. Suzuki, M. Abe, and Y. Hashimoto, “Sialylation and glycosylation modulate cell adhesion and invasion to extracellular matrix in human malignant lymphoma : Dependency on integrin and the Rho GTPase family,” *Int. J. Oncol.*, vol. 47, pp. 2091–2099, Dec. 2015.
- [42] Q. Hang, T. Isaji, S. Hou, Y. Wang, T. Fukuda, and J. Gu, “A Key Regulator of Cell Adhesion : Identification and Characterization of Important N-Glycosylation Sites on Integrin $\alpha 5$ for Cell Migration,” *Mol. Cell. Biol.*, vol. 37, no. 9, 2017.
- [43] P. H. Seeberger, “Monosaccharide Diversity,” in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, and P. H. Seeberger, eds.), Cold Spring Harbor (NY) : Cold Spring Harbor Laboratory Press, 3rd ed., 2015.
- [44] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar, H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. Vliegenthart, T. Lütke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, and S. Kornfeld, “Symbol Nomenclature for Graphical Representations of Glycans,” *Glycobiology*, vol. 25, pp. 1323–1324, Dec. 2015.
- [45] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, and P. H. Seeberger, eds., *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, 1999.

- [46] S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, T. Lütteke, N. O’Boyle, N. Packer, P. Stanley, P. Toukach, A. Varki, R. J. Woods, and SNFG Discussion group, “Updates to the Symbol Nomenclature For Glycans (SNFG) Guidelines,” *Glycobiology*, June 2019.
- [47] L. Zhang, “Glycosaminoglycan (GAG) biosynthesis and GAG-binding proteins,” *Prog Mol Biol Transl Sci*, vol. 93, pp. 1–17, 2010.
- [48] J. Casale and J. S. Crane, “Biochemistry, Glycosaminoglycans,” in *StatPearls*, Treasure Island (FL) : StatPearls Publishing, 2019.
- [49] P. Stanley, H. Schachter, and N. Taniguchi, “N-Glycans,” in *Essentials of Glycobiology* (A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds.), Cold Spring Harbor (NY) : Cold Spring Harbor Laboratory Press, 2nd ed., 2009.
- [50] M. Aebi, “N-linked protein glycosylation in the ER,” *Biochim. Biophys. Acta*, vol. 1833, pp. 2430–2437, Nov. 2013.
- [51] H. Schachter, “The joys of HexNAc. The synthesis and function of N- and O-glycan branches,” *Glycoconj. J.*, vol. 17, pp. 465–483, Sept. 2000.
- [52] L. Y. Lee, M. Thaysen-Andersen, M. S. Baker, N. H. Packer, W. S. Hancock, and S. Fannan, “Comprehensive N-glycome profiling of cultured human epithelial breast cells identifies unique secretome N-glycosylation signatures enabling tumorigenic subtype classification,” *Journal of Proteome Research*, vol. 13, pp. 4783–4795, Nov. 2014.
- [53] Y. Sato and T. Endo, “Alteration of brain glycoproteins during aging,” *Geriatrics & Gerontology International*, vol. 10 Suppl 1, pp. S32–40, July 2010.
- [54] J. D. Allen and T. M. Ross, “H3n2 influenza viruses in humans : Viral mechanisms, evolution, and evaluation,” *Hum Vaccin Immunother*, vol. 14, no. 8, pp. 1840–1847, 2018.
- [55] S. J. Zost, K. Parkhouse, M. E. Gumina, K. Kim, S. Diaz Perez, P. C. Wilson, J. J. Treanor, A. J. Sant, S. Cobey, and S. E. Hensley, “Contemporary H3n2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 47, pp. 12578–12583, 2017.
- [56] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “The Protein Data Bank,” *Nucleic Acids Research*, vol. 28, pp. 235–242, Jan. 2000.
- [57] M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott, and R. A. Dwek, “Conformational studies of oligosaccharides and glycopeptides : complementarity of NMR, X-ray crystallography, and molecular modelling,” *Chemical Reviews*, vol. 102, pp. 371–386, Feb. 2002.

- [58] S. E. Hamby and J. D. Hirst, "Prediction of glycosylation sites using random forests," *BMC bioinformatics*, vol. 9, p. 500, Nov. 2008.
- [59] R. Gupta, E. Jung, and S. Brunak, "Prediction of N-glycosylation sites in human proteins," *In preparation*, vol. 46, pp. 203–206, Jan. 2004.
- [60] G.-Y. Chuang, J. C. Boyington, M. G. Joyce, J. Zhu, G. J. Nabel, P. D. Kwong, and I. Georgiev, "Computational prediction of N-linked glycosylation incorporating structural properties and patterns," *Bioinformatics*, vol. 28, pp. 2249–2255, Sept. 2012.
- [61] T. Lütteke and C.-W. von der Lieth, "pdb-care (PDB carbohydrate residue check) : a program to support annotation of complex carbohydrate structures in PDB files," *BMC bioinformatics*, vol. 5, p. 69, June 2004.
- [62] T. Lütteke, M. Frank, and C.-W. von der Lieth, "Data mining the protein data bank : automatic detection and assignment of carbohydrate structures," *Carbohydrate Research*, vol. 339, pp. 1015–1020, Apr. 2004.
- [63] T. Lütteke, M. Frank, and C.-W. von der Lieth, "Carbohydrate Structure Suite (CSS) : analysis of carbohydrate 3d structures derived from the PDB," *Nucleic Acids Research*, vol. 33, pp. D242–D246, Jan. 2005.
- [64] A. Bohne-Lang, E. Lang, T. Förster, and C. W. von der Lieth, "LINUCS : linear notation for unique description of carbohydrate sequences," *Carbohydrate Research*, vol. 336, pp. 1–11, Nov. 2001.
- [65] M. Frank, T. Lütteke, and C.-W. von der Lieth, "GlycoMapsDB : a database of the accessible conformational space of glycosidic linkages," *Nucleic Acids Research*, vol. 35, pp. 287–290, Jan. 2007.
- [66] M. A. Rojas-Macias and T. Lütteke, "Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by GlyVicinity," *Methods in Molecular Biology (Clifton, N.J.)*, vol. 1273, pp. 215–226, 2015.
- [67] J. Rosen, L. Miguet, and S. Pérez, "Shape : automatic conformation prediction of carbohydrates using a genetic algorithm," *Journal of Cheminformatics*, vol. 1, p. 16, Sept. 2009.
- [68] A. Bohne, E. Lang, and C.-W. v. d. Lieth, "W3-SWEET : Carbohydrate Modeling By Internet," *Molecular modeling annual*, vol. 4, pp. 33–43, Jan. 1998.
- [69] A. Bohne, E. Lang, and C. W. von der Lieth, "SWEET - WWW-based rapid 3d construction of oligo- and polysaccharides," *Bioinformatics (Oxford, England)*, vol. 15, pp. 767–768, Sept. 1999.
- [70] A. Bohne-Lang and C.-W. von der Lieth, "GlyProt : in silico glycosylation of proteins," *Nucleic Acids Research*, vol. 33, pp. W214–219, July 2005.

- [71] R. Danne, C. Poojari, H. Martinez-Seara, S. Rissanen, F. Lolicato, T. Róg, and I. Vattulainen, “doGlycans-Tools for Preparing Carbohydrate Structures for Atomistic Simulations of Glycoproteins, Glycolipids, and Carbohydrate Polymers for GROMACS,” *Journal of Chemical Information and Modeling*, vol. 57, pp. 2401–2406, Oct. 2017.
- [72] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, and R. J. Woods, “GLYCAM06 : a generalizable biomolecular force field. Carbohydrates,” *J Comput Chem*, vol. 29, pp. 622–655, Mar. 2008.
- [73] S. Jo, K. C. Song, H. Desaire, A. D. MacKerell, and W. Im, “Glycan Reader : Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins,” *Journal of computational chemistry*, vol. 32, pp. 3135–3141, Nov. 2011.
- [74] J. Huang and A. D. MacKerell, “CHARMM36 all-atom additive protein force field : validation based on comparison to NMR data,” *J Comput Chem*, vol. 34, pp. 2135–2145, Sept. 2013.
- [75] A. Arroyuelo, J. A. Vila, and O. A. Martin, “Azahar : a PyMOL plugin for construction, visualization and analysis of glycan molecules,” *Journal of Computer-Aided Molecular Design*, vol. 30, no. 8, pp. 619–624, 2016.
- [76] D. F. Thieker, J. A. Hadden, K. Schulten, and R. J. Woods, “3d implementation of the symbol nomenclature for graphical representation of glycans,” *Glycobiology*, vol. 26, pp. 786–787, Aug. 2016.
- [77] A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, and M. E. Etzler, eds., *Essentials of Glycobiology*. Cold Spring Harbor (NY) : Cold Spring Harbor Laboratory Press, 2nd ed., 2009.
- [78] Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent, and M. Baaden, “Game on, science - how video game technology may help biologists tackle visualization challenges,” *PloS One*, vol. 8, no. 3, p. e57990, 2013.
- [79] M. Chavent, A. Vanel, A. Tek, B. Levy, S. Robert, B. Raffin, and M. Baaden, “GPU-accelerated atom and dynamic bond visualization using hyperballs : A unified algorithm for balls, sticks, and hyperboloids,” *Journal of Computational Chemistry*, vol. 32, pp. 2924–2935, Oct. 2011.
- [80] S. Pérez, T. Tubiana, A. Imbert, and M. Baaden, “Three-dimensional representations of complex carbohydrates and polysaccharides—SweetUnityMol : A video game-based computer graphic software,” *Glycobiology*, vol. 25, pp. 483–491, May 2015.
- [81] Y. Mazola, G. China, and A. Musacchio, “Integrating Bioinformatics Tools to Handle Glycosylation,” *PLoS Computational Biology*, vol. 7, Dec. 2011.

- [82] J. W. Gibbs, *A method of geometrical representation of the thermodynamic properties of substances by means of surfaces*. 1873. OCLC : 12301679.
- [83] J. Cherfils, M. C. Vaney, I. Morize, E. Surcouf, N. Colloc'h, and J. P. Mornon, "MANOSK : A graphics program for analyzing and modeling molecular structure and functions," *Journal of Molecular Graphics*, vol. 6, pp. 155–160, Sept. 1988.
- [84] R. A. Sayle and E. J. Milner-White, "RASMOL : biomolecular graphics for all," *Trends Biochem. Sci.*, vol. 20, p. 374, Sept. 1995.
- [85] W. Humphrey, A. Dalke, and K. Schulten, "VMD : visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 27–28, Feb. 1996.
- [86] Schrödinger, LLC, "The PyMOL Molecular Graphics System, Version 1.8." Nov. 2015.
- [87] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, and T. E. Ferrin, "UCSF ChimeraX : Meeting modern challenges in visualization and analysis," *Protein Science : A Publication of the Protein Society*, vol. 27, no. 1, pp. 14–25, 2018.
- [88] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927.
- [89] Q. Cui and M. Elstner, "Density Functional Tight Binding : values of semi-empirical methods in an ab initio era," *Physical chemistry chemical physics : PCCP*, vol. 16, pp. 14368–14377, July 2014.
- [90] C. Lorenz and N. L. Doltsinis, "Molecular Dynamics Simulation : From "Ab Initio" to "Coarse Grained"," in *Handbook of Computational Chemistry* (J. Leszczynski, ed.), pp. 195–238, Dordrecht : Springer Netherlands, 2012.
- [91] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5 : a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, pp. 845–854, Apr. 2013.
- [92] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS : A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, pp. 43–56, Sept. 1995.
- [93] P. Paracuellos, S. Kalamajski, A. Bonna, D. Bihan, R. W. Farndale, and E. Hohenester, "Structural and functional analysis of two small leucine-rich repeat proteoglycans, fibromodulin and chondroadherin," *Matrix Biol.*, vol. 63, pp. 106–116, 2017.
- [94] K. Pietraszek-Gremplewicz, K. Karamanou, A. Niang, M. Dauchez, N. Belloy, F.-X. Maquart, S. Baud, and S. Brézillon, "Small leucine-rich proteoglycans and matrix metalloproteinase-14 : Key partners?," *Matrix Biol.*, vol. 75-76, pp. 271–285, 2019.

- [95] R. M. Lauder, T. N. Huckerby, and I. A. Nieduszynski, "The structure of the keratan sulphate chains attached to fibromodulin from human articular cartilage," *Glycoconj. J.*, vol. 14, pp. 651–660, Aug. 1997.
- [96] D. Kony, W. Damm, S. Stoll, and W. F. Van Gunsteren, "An improved OPLS-AA force field for carbohydrates," *J Comput Chem*, vol. 23, pp. 1416–1429, Nov. 2002.
- [97] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *J. Am. Chem. Soc.*, vol. 110, pp. 1657–1666, Mar. 1988.
- [98] H. Han, M. Stapels, W. Ying, Y. Yu, L. Tang, W. Jia, W. Chen, Y. Zhang, and X. Qian, "Comprehensive characterization of the N-glycosylation status of CD44s by use of multiple mass spectrometry-based techniques," *Analytical and Bioanalytical Chemistry*, vol. 404, pp. 373–388, Aug. 2012.
- [99] K. Khatri, J. A. Klein, M. R. White, O. C. Grant, N. Leymarie, R. J. Woods, K. L. Hartshorn, and J. Zaia, "Integrated Omics and Computational Glycobiology Reveal Structural Basis for Influenza A Virus Glycan Microheterogeneity and Host Interactions," *Molecular & Cellular Proteomics : MCP*, vol. 15, pp. 1895–1912, June 2016.
- [100] V. I. Otto, E. Damoc, L. N. Cueni, T. Schürpf, R. Frei, S. Ali, N. Callewaert, A. Moise, J. A. Leary, G. Folkers, and M. Przybylski, "N-Glycan structures and N-glycosylation sites of mouse soluble intercellular adhesion molecule-1 revealed by MALDI-TOF and FTICR mass spectrometry," *Glycobiology*, vol. 16, pp. 1033–1044, Nov. 2006.
- [101] M. K. Sethi, M. Thaysen-Andersen, J. T. Smith, M. S. Baker, N. H. Packer, W. S. Hancock, and S. Fanayan, "Comparative N-Glycan Profiling of Colorectal Cancer Cell Lines Reveals Unique Bisecting GlcNAc and α -2,3-Linked Sialic Acid Determinants Are Associated with Membrane Proteins of the More Metastatic/Aggressive Cell Lines," *Journal of Proteome Research*, vol. 13, pp. 277–288, Jan. 2014.
- [102] D. Zhang, Q. Xie, Q. Wang, Y. Wang, J. Miao, L. Li, T. Zhang, X. Cao, and Y. Li, "Mass spectrometry analysis reveals aberrant N-glycans in colorectal cancer tissues," *Glycobiology*, vol. 29, no. 5, pp. 372–384, 2019.
- [103] H. E. Miwa, Y. Song, R. Alvarez, R. D. Cummings, and P. Stanley, "The bisecting GlcNAc in cell growth control and tumor progression," *Glycoconj J*, vol. 29, pp. 609–618, Dec. 2012.
- [104] N. Taniguchi and Y. Kizuka, "Glycans and cancer : role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics," *Adv. Cancer Res.*, vol. 126, pp. 11–51, 2015.
- [105] S. Hanashima, A. Suga, and Y. Yamaguchi, "Bisecting GlcNAc restricts conformations of branches in model N-glycans with GlcNAc termini," *Carbohydr. Res.*, vol. 456, pp. 53–60, Feb. 2018.

- [106] K. Kanninen, G. Goldsteins, S. Auriola, I. Alafuzoff, and J. Koistinaho, "Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach," *Neuroscience Letters*, vol. 367, pp. 235–240, Sept. 2004.
- [107] A. M. Harbison, L. P. Brosnan, K. Fenlon, and E. Fadda, "Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans : a molecular dynamics study," *Glycobiology*, vol. 29, no. 1, pp. 94–103, 2019.
- [108] T. Dědová, E. I. Braicu, J. Sehouli, and V. Blanchard, "Sialic Acid Linkage Analysis Refines the Diagnosis of Ovarian Cancer," *Front Oncol*, vol. 9, p. 261, 2019.
- [109] R. Zhao, X. Liu, Y. Wang, X. Jie, R. Qin, W. Qin, M. Zhang, H. Tai, C. Yang, L. Li, P. Peng, M. Shao, X. Zhang, H. Wu, Y. Ruan, C. Xu, S. Ren, and J. Gu, "Integrated glycomic analysis of ovarian cancer side population cells," *Clinical Proteomics*, vol. 13, p. 32, 2016.
- [110] V. Dotz, R. F. H. Lemmers, K. R. Reiding, A. L. Hipgrave Ederveen, A. G. Lieveise, M. T. Mulder, E. J. G. Sijbrands, M. Wuhler, and M. van Hoek, "Plasma protein N-glycan signatures of type 2 diabetes," *Biochim Biophys Acta Gen Subj*, vol. 1862, no. 12, pp. 2613–2622, 2018.
- [111] M. K. Sethi, H. Kim, C. K. Park, M. S. Baker, Y.-K. Paik, N. H. Packer, W. S. Hancock, S. Fanayan, and M. Thaysen-Andersen, "In-depth N-glycome profiling of paired colorectal cancer and non-tumorigenic tissues reveals cancer-, stage- and EGFR-specific protein N-glycosylation," *Glycobiology*, vol. 25, pp. 1064–1078, Oct. 2015.
- [112] P. A. McEwan, P. G. Scott, P. N. Bishop, and J. Bella, "Structural correlations in the family of small leucine-rich repeat proteins and proteoglycans," *Journal of Structural Biology*, vol. 155, pp. 294–305, Aug. 2006.
- [113] Y. Sakae, T. Satoh, H. Yagi, S. Yanaka, T. Yamaguchi, Y. Isoda, S. Iida, Y. Okamoto, and K. Kato, "Conformational effects of N-glycan core fucosylation of immunoglobulin G Fc region on its interaction with Fc γ receptor IIIa," *Sci Rep*, vol. 7, no. 1, p. 13780, 2017.
- [114] N. Rudman, O. Gornik, and G. Lauc, "Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes," *FEBS Lett.*, vol. 593, pp. 1598–1615, July 2019.
- [115] S. Jo, Y. Qi, and W. Im, "Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins," *Glycobiology*, vol. 26, pp. 19–29, Jan. 2016.
- [116] E. Leygue, L. Snell, H. Dotzlaw, S. Troup, T. Hiller-Hitchcock, L. C. Murphy, P. J. Roughley, and P. H. Watson, "Lumican and decorin are differentially expressed in human breast carcinoma," *J. Pathol.*, vol. 192, pp. 313–320, Nov. 2000.
- [117] C. Besançon, A. Guillot, S. Blaise, M. Dauchez, N. Belloy, J. Prévotau-Jonquet, and S. Baud, "New visualization of dynamical flexibility of n-glycans : Umbrella visualization

- in unitymol,” in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018* (H. J. Zheng, Z. Callejas, D. Griol, H. Wang, X. Hu, H. H. H. W. Schmidt, J. Baumbach, J. Dickerson, and L. Zhang, eds.), pp. 291–298, IEEE Computer Society, 2018.
- [118] C. Besançon, A. Guillot, S. Blaise, M. Dauchez, N. Belloy, J. Prévotau-Jonquet, and S. Baud, “Umbrella Visualization : A method of analysis dedicated to glycan flexibility with UnityMol,” *Methods*, July 2019.
- [119] A. Kurimoto, S. Kitazume, Y. Kizuka, K. Nakajima, R. Oka, R. Fujinawa, H. Korekane, Y. Yamaguchi, Y. Wada, and N. Taniguchi, “The absence of core fucose up-regulates GnT-III and Wnt target genes : a possible mechanism for an adaptive response in terms of glycan function,” *J. Biol. Chem.*, vol. 289, pp. 11704–11714, Apr. 2014.
- [120] M. Takahashi, Y. Kuroki, K. Ohtsubo, and N. Taniguchi, “Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core : their functions and target proteins,” *Carbohydr. Res.*, vol. 344, pp. 1387–1390, Aug. 2009.
- [121] R. Schauer, “Sialic acids as regulators of molecular and cellular interactions,” *Current Opinion in Structural Biology*, vol. 19, pp. 507–514, Oct. 2009.
- [122] A. Larkin and B. Imperiali, “The expanding horizons of asparagine-linked glycosylation,” *Biochemistry*, vol. 50, pp. 4411–4426, May 2011.
- [123] A. C. Walls, M. A. Tortorici, B. Frenz, J. Snijder, W. Li, F. A. Rey, F. DiMaio, B.-J. Bosch, and D. Veasler, “Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy,” *Nature Structural & Molecular Biology*, vol. 23, pp. 899–905, Oct. 2016.
- [124] Q. Zhou and H. Qiu, “The Mechanistic Impact of N-Glycosylation on Stability, Pharmacokinetics, and Immunogenicity of Therapeutic Proteins,” *J Pharm Sci*, vol. 108, pp. 1366–1377, Apr. 2019.
- [125] M.-E. Losfeld, E. Scibona, C.-W. Lin, T. K. Villiger, R. Gauss, M. Morbidelli, and M. Aebi, “Influence of protein/glycan interaction on site-specific glycan heterogeneity,” *The FASEB Journal*, p. fj.201700403R, July 2017.

DOGME - Développement d'un Outil de visualisation moléculaire et application à l'étude *in silico* des Glycosylations de protéines de la Matrice Extracellulaire.

Les N-glycosylations sont des modifications post-traductionnelles volumineuses liées aux résidus asparagine et pointant à la surface des protéines. Elles ont un rôle essentiel dans la fonction des protéines et certaines modifications des N-glycosylations, comme l'hydrolyse des acides sialiques, peuvent altérer ces fonctions. Cependant, l'étude *in vitro* des N-glycans est compliquée par la diversité structurale et les nombreux groupes réactionnels des chaînes de glycosylation. La dynamique moléculaire est un outil de simulation permettant de surmonter ces problèmes tout en fournissant de nombreuses informations conformationnelles grâce à un échantillonnage exhaustif.

Ce travail de thèse a consisté en l'amélioration de la caractérisation de l'impact des glycosylations sur la structure des protéines à travers le couplage de simulations de dynamique moléculaire et le développement d'une méthode originale de visualisation moléculaire : l'Umbrella Visualization. Celle-ci a été implémentée dans le logiciel de visualisation UnityMol et permet de visualiser la surface protéique couverte par les glycans. Cette méthode adopte un point de vue statistique qui met en valeur les régions les plus souvent masquées par le glycan. Nous avons ensuite utilisé cette méthode sur des objets biologiques fortement glycosylés et liés à la matrice extracellulaire (récepteur à l'insuline et fibromoduline). Nos premiers résultats et les travaux réalisés sur des chaînes de glycans isolées, couplés à la méthode de l'Umbrella Visualization, joueront un rôle important dans la suite de nos travaux visant à élucider les relations structure/fonction/dynamique des acteurs majeurs (pour certains porteurs de nombreuses glycosylations) de la matrice extracellulaire.

N-glycosylations, modélisation moléculaire, visualisation moléculaire, matrice extracellulaire, dynamique moléculaire

DOGME - Development of a molecular visualisation tool and application to the *in silico* study of glycosylations on extracellular matrix proteins.

N-glycosylations are voluminous post-translational modifications linked to asparagine residue on the protein surface. They have an essential impact in the function of proteins and some N-glycosylation modifications such as sialic acid hydrolysis, can alter these functions. However *in vitro* studies of N-glycans is challenging because of the structural diversity of glycosylation chains and their numerous reactive groups. Molecular dynamics is a simulation tool capable of overcoming these difficulties by giving access to a great amount of conformational data thanks to an exhaustive sampling.

This thesis work consists in the improvement of the characterization of glycosylations' impact on protein structures through the association of molecular dynamics simulations and the development of a new, original visualization method: The Umbrella Visualization. This method has been implemented into the molecular viewer UnityMol and has thus enabled the visualization of the protein surface covered by glycans. This method adopts a statistical approach and highlights the protein surface parts most often covered by the glycan. We then used this method on heavily glycosylated objects linked to the extracellular matrix (insulin receptor and fibromodulin). Our first results and the work done on isolated glycan chains paired with the use of the Umbrella Visualization will play an important role in the pursuit of our work in elucidating the structure/function/dynamics relations of major actors (some of them being heavily glycosylated) from the extracellular matrix.

N-glycosylations, molecular modeling, molecular visualization, extracellular matrix, molecular dynamics

Discipline : SCIENCES DE LA VIE ET DE LA SANTE

Spécialité : Bio-informatique

Moulin de la Housse, BP 1039 51687 Reims cedex 2, France.

Université de Reims Champagne-Ardenne

MEDYC - UMR CNRS 7369

UFR Sciences Exactes et Naturelles, Campus