



**HAL**  
open science

# Méthode d'analyse spectroscopique InfraRouge innovante pour le développement de nouveaux OLIGOsaccharides comme additifs alimentaires fonctionnels pour les animaux d'élevage

Baptiste Moge

► **To cite this version:**

Baptiste Moge. Méthode d'analyse spectroscopique InfraRouge innovante pour le développement de nouveaux OLIGOsaccharides comme additifs alimentaires fonctionnels pour les animaux d'élevage. Physique [physics]. Université Claude Bernard - Lyon I, 2023. Français. NNT : 2023LYO10328 . tel-04761928

**HAL Id: tel-04761928**

**<https://theses.hal.science/tel-04761928v1>**

Submitted on 31 Oct 2024

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

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

**THESE de DOCTORAT DE  
L'UNIVERSITE CLAUDE BERNARD LYON 1**

**Ecole Doctorale N° 52  
Physique et Astrophysique de Lyon**

**Discipline** : Physique

Soutenue publiquement le 11/12/2023, par :  
**Baptiste MOGE**

---

**Méthode d'analyse spectroscopique InfraRouge  
innovante pour le développement de nouveaux  
OLIGOsaccharides comme additifs alimentaires  
fonctionnels pour les animaux d'élevage**

---

Devant le jury composé de :

FOURNIER, Isabelle PU Université de Lille

Rapporteure

DANIEL, Régis DR CNRS, Evry

Rapporteur

AFONSO, Carlos PU Université de Rouen

Président

CHIROT, Fabien MCF-HDR Université de Lyon

Examineur

COMPAGNON, Isabelle MCF-HDR Université Lyon I

Directrice de thèse

DARBLADE, Benoit Responsable Partenariats et business  
développement Elicityl

Invité







## REMERCIEMENTS

---

Ces trois dernières années passées à l'iLM ont été pour moi très enrichissantes tant d'un point de vue scientifique que d'autres compétences « connexes » également importantes dans le cadre de la gestion d'un projet. J'ai énormément appris pendant cette période. Pour cela, je souhaite vivement remercier les personnes suivantes :

En tout premier lieu, les membres de mon jury. Isabelle Fournier et Régis Daniel de m'avoir fait l'honneur d'accepter d'être les rapporteurs de mon manuscrit pour votre temps, votre travail de relecture et vos remarques. Fabien Chirot d'avoir accepté de prendre part à cette soutenance de thèse en tant qu'examineur, également à Carlos Afonso d'avoir assuré la présidence durant cette soutenance de thèse. Merci à tous pour les discussions qui ont suivies ma présentation ainsi que les différentes remarques constructives.

Ensuite, je tiens bien naturellement à exprimer ma profonde gratitude à Isabelle Compagnon ma directrice de thèse, merci de m'avoir fait confiance pour ce projet et de m'avoir accompagné tout au long de ces trois années de thèse.

J'adresse également mes sincères remerciements à Franck Lépine, directeur de l'équipe Structure et Dynamique Multi-Echelle des Edifices Moléculaires, merci de m'avoir accueilli dans l'équipe Dynamo pendant la durée de ma thèse.

Merci aux membres permanents de l'équipe Dynamo : Bruno, Christian, Gulabi, Richard, Saikat, Vincent pour tous les échanges que nous avons pu avoir.

Merci également aux non-permanents : Alexie, Corentin, Sreelakshmi pour les doctorants et Emilien et Yaowei pour les post-doctorants pour les différentes discussions et les coups de mains et astuces sur les appareils et la paillasse.

Un merci tout particulier à Alicia et Ozgur les doctorantes « sucres » avec lesquelles j'ai eu la chance de travailler quotidiennement dans le bureau mais aussi au labo. Bon courage pour la suite !

Merci également aux stagiaires que j'ai pu rencontrer pendant ces trois années de thèse : Hugo, Mohammed, Sarah, Alizée bonne continuation dans vos projets respectifs.

Merci également aux collègues d'autres équipes de l'iLM avec qui j'ai pu interagir. Mention spéciale pour Clothilde de l'équipe Spectrobio.

Je profite de ces remerciements pour avoir une pensée pour tout le personnel de l'institut Lumière Matière avec qui j'ai pu avoir affaire en particulier le personnel technique dont Séverine, Marie-Jo, Daniel, Christian, Farys, Marc, Florian, Véronique...

Merci à nos partenaires d'Elicityl et en particulier à Benoit Darblade, Ludovic Bastide, Yoann Pascal et Sylvère Bonnet avec qui j'ai pu interagir tout au long de cette thèse, leurs réponses à mes questions. Merci d'avoir proposé un projet complexe qui nous a permis de travailler sur de nouveaux développements sur la spectroscopie IRMPD afin de repousser les limites de cette méthode.

J'ai également en tête les différents collaborateurs avec qui nous avons pu travailler et échanger sur des projets concernant la spectroscopie IRMPD de Marseille à Rennes en passant par Grenoble ou Evry.

Merci à ma famille : papa, maman, mon frère (me voilà docteur moi aussi !) et ma sœur pour votre soutien inconditionnel. Ainsi que mes amis Amanda, Benoit, Agathe, Théo, Guillaume, Florian, Camille, Arnaud, Dimitri et en particulier Romane et Romain (merci pour les nombreux covoiturages, le logement temporaire et tous ces verres de digestifs !) pour leur présence depuis toujours et en toutes circonstances.

# Table des matières

Liste des abréviations.....	10
Introduction.....	12
<b>Chapitre 1 Instrumentation principe et utilisation de l'IRMPD.....</b>	<b>21</b>
1 Couplages MS/IR.....	22
1.1 Principe de la spectroscopie d'action.....	22
1.2 Exemples de méthodes de spectroscopie IR en phase gazeuse.....	23
1.2.1 Messenger tagging.....	23
1.2.2 Helium droplet.....	23
1.2.3 IR-UV.....	23
1.2.4 L'IRMPD.....	24
2 La spectroscopie IRMPD.....	24
2.1 L'IRMPD comme moyen de fragmentation.....	24
2.2 La spectroscopie IRMPD.....	24
2.3 Acquisition d'un spectre IRMPD.....	25
2.4 Gamme de nombres d'onde d'intérêt .....	26
3 Instrumentation.....	27
3.1 Différentes sources lasers.....	27
3.2 Spectromètres de masse utilisés.....	28
3.3 Montage expérimental mis en place à l'iLM.....	28
4 L'IRMPD et les questions de physico-chimie : identification de fragments d'oligosaccharides fucosylés.....	30
4.1 La migration du fucose.....	30
4.2 Matériel et méthode.....	31
4.3 Migration interosidique.....	34
4.4 Migration intraosidique.....	35
4.5 Comparaison de ces deux types de migration.....	36
4.6 Conclusion et ouverture.....	36
5 Applications de la spectroscopie IRMPD à la chimie analytique.....	38
5.1 Applications analytiques.....	38
5.2 Identification de sucres à l'iLM.....	38
5.3 Méthode de séquençage.....	39
5.3.1 Principe du séquençage.....	39
5.3.2 Séquençage du Lacto-N-tetraose à des fins d'illustration.....	39
6 Conclusion.....	43
7 Bibliographie.....	44
<b>Chapitre 2 Développements expérimentaux de l'IRMPD pour les analyses haut débit.....</b>	<b>50</b>
1 Introduction.....	51
2 Accélération de l'acquisition des données.....	51
2.1 Acquisition rapide de spectres IRMPD.....	51
2.1.1 Nouveau laser et nouveau type de piège à ions : Nouveau montage.....	51
2.1.2 Matériel et méthode.....	53
2.1.3 Acquisition menée en 14 secondes.....	55
2.1.4 Limites.....	56
2.2 Ajout du laser CO <sub>2</sub> .....	57
2.2.1 Fonctionnement du laser CO <sub>2</sub> .....	57
2.2.2 Nouvelle fenêtre.....	57

2.2.3 Montage expérimental.....	58
2.2.4 Effet de l'irradiation avec deux lasers.....	59
2.2.5 Application à des espèces lithiées sodées et déprotonées.....	59
3 Reconnaissance de spectres.....	62
3.1 Approche par les scores.....	63
3.1.1 Calcul de la ressemblance.....	63
3.1.2 Calcul de l'erreur.....	64
3.2 Approche par ACP.....	65
3.2.1 La méthode.....	65
3.2.2 Application sur les hexosamines.....	65
3.2.3 L'analyse en ACP.....	66
3.2.4 Confirmation de l'identification.....	69
3.3 Approche par apprentissage machine.....	71
3.3.1 Production des données.....	71
3.3.2 Architecture du modèle.....	73
3.3.3 Exactitude de la classification du modèle.....	73
3.3.4 Incertitude et précision du modèle.....	74
3.3.5 Conclusion et ouverture.....	75
4 Conclusion.....	76
5 Bibliographie.....	77
<b>Chapitre 3 Chapitre III : Analyse d'oligosaccharides au moyen de la spectroscopie IRMPD après séparation par chromatographie liquide.....</b>	<b>79</b>
1 Introduction.....	80
2 La chromatographie liquide.....	80
2.1 L'appareil.....	80
2.2 Contraintes techniques pour le couplage entre chromatographie liquide et spectroscopie IRMPD.....	81
2.3 Approches déjà envisagées.....	82
2.3.1 Collecte de fractions.....	82
2.3.2 Stockage dans des boucles d'injection.....	83
2.3.3 Utilisation du stop flow.....	84
3 Analyse des produits de digestion d'un pentasaccharide synthétique.....	85
3.1 Structure du Fondaparinux.....	85
3.2 Acquisition des spectres en infusion directe.....	86
3.3 Les digestions.....	86
3.4 La méthode chromatographique.....	89
3.5 Résultats.....	90
4 Analyse en ligne des composés en sortie de colonne : exemple sur les HMOs.....	91
4.1 La méthode chromatographique.....	91
4.2 Premiers résultats du couplage LC MS/MS-IR.....	92
4.2.1 Intérêt de fragmenter le composé d'intérêt.....	92
4.2.2 Mesures des spectres IRMPD de référence.....	93
4.2.3 Séparation chromatographique des deux isomères.....	94
4.2.4 Mesure en ligne du spectre IRMPD du premier pic chromatographique.....	94
4.3 Réduction des composés.....	95
4.3.1 Les composés.....	95
4.3.2 La préparation des composés : réduction.....	95
4.3.3 Résultats.....	96

4.4 Nouveaux développements conduits sur le couplage LC-MS-IR.....	98
4.4.1 Spectres IRMPD de monosaccharides réduits : le cas des hexoses.....	98
4.4.2 Détection d'une forme rare de galactose composant des polysaccharides bactériens par screening à longueur d'onde fixe.....	100
5 Conclusion.....	102
6 Bibliographie.....	104
<b>Chapitre 4 Analyse de composés glycoconjugués l'exemple d'une glycoprotéine : la fétuine bovine.....</b>	<b>107</b>
1 Introduction.....	108
2 Les protéines.....	108
2.1 Les modifications post traductionnelles.....	108
2.1.1 Cas général.....	108
2.1.2 Cas particulier de la glycosylation.....	109
2.2 La fétuine bovine.....	109
2.3 Etudes des protéine bottom up.....	110
3 Recherche de glycosylation sur la protéine.....	111
3.1 Résultat sur une protéine non glycosylée.....	112
3.2 Résultats sur une protéine glycosylée.....	112
4 Etude des peptides : identification de glycopeptides par screening IRMPD.....	113
4.1 Proposition.....	113
4.1.1 Structure des acides aminés contenant un groupement hydroxyle.....	114
4.1.2 Protocoles similaires déjà proposés.....	114
4.1.3 Spectre IRMPD du glycopeptide de poulet.....	115
4.2 Protocole.....	116
4.2.1 Digestion de la protéine par des enzymes protéolytiques.....	116
4.2.2 Passage en LCMS.....	120
4.2.3 Irradiation par le laser.....	120
4.3 Résultats.....	120
4.3.1 Chromatogramme des produits de digestion de la fétuine bovine.....	120
4.3.2 Spectre IRMPD d'un peptide, pouvant porter un acide aminé avec un hydroxyle.....	122
4.3.3 Spectre IRMPD de candidats glycopeptides.....	123
4.3.4 Screening de glycopeptides.....	127
5 Conclusion.....	118
6 Bibliographie.....	130
<b>Conclusion et perspectives.....</b>	<b>133</b>
Annexe 1.....	136
Annexe 2.....	139
Liste des publications.....	142
Liste présentations orales.....	142
Résumé.....	
Abstract.....	

# Table des abréviations

Les abréviations utilisées dans ce manuscrit sont présentées dans ce glossaire.

AmBic	Bicarbonate d'Ammonium	ADN	Acide Désoxyribonucléique
ARN	Acide Ribonucléique	BGHI	Blood Group Antigen Type 1
BGHII	Blood Group Antigen Type 2	2'FL	2'-Fucosyllactose
3'SL	3-Sialyllactose	acac	Acide Acétique
ACN	Acétonitrile	ACP	Analyse en Composantes Principales
CaF <sub>2</sub>	Fluorure de Calcium	CID	Collision Induced Dissociation
CSC	Chondroïtine sulfate C	DDA	Data Dependent Acquisition
DTT	Dithiothréitol	ECD	Electron Capture Dissociation
ESI	Electrospray Source Ionization	ETD	Electron Transfer Dissociation
FELIX	Free Electron Laser for Infrared Experiments	FHI FEL	Free-Electron Laser at the Fritz Haber Institute
FTICR	Fourier Transform Ion Cyclotronic Resonance	FTIR	Fourier Transform Infrared
Fuc	Fucose	Gal	Galactose (sous sa forme de cycle à six membres, la forme la plus commune)
Gal <sub>f</sub>	Galactofuranose (le galactose est sous la forme d'un cycle à cinq membres, qui est une forme rare)	GalN	Galactosamine
GalNAc	N-Acétylgalactosamine	Glc	Glucose
GlcN	Glucosamine	GlcNAc	N-Acetylglucosamine
HILIC	Hydrophilic Interaction Chromatography	HMOs	Human Milk Oligosaccharide
IA	Intelligence Artificielle	IAA	Iodoacetamide
Ifrag	Intensité De L'ion Fragment	li	Intensité Incidente
ILM	Institut Lumière Matière	IMS	Ion Mobility Spectrometry
Iprec	Intensité de l'ion Précurseur	IRMPD	Infrared Multiphoton Dissociation

Table des abréviations

IT	Ion Trap	It	Intensité Transmise
IVR	Internal Vibrational Relaxation	KDP	Monopotassium Phosphate
KTA	Potassium Titanyl Arseniate	KTP	Potassium Titanyl Phosphoryl
LC	Liquid Chromatography	LC-MS	Chromatographie Liquide Couplée à la Spectrométrie de Masse
LC-MS-IR	Spectroscopie IR en Phase Gazeuse après Séparation en Chromatographie Liquide	LC-MS/MS-IR	Spectroscopie IR en Phase Gazeuse d'un Ion Produit par Fragmentation après Séparation en Chromatographie Liquide
LCQ	Spectromètre de Masse de la Marque ThermoScientific dans lequel les ions sont concentré en un point	Lea	Lewis A
LEL	Laser à Electrons Libres	Lex	Lewis X
Ley	Lewis Y	LiCl	Chlorure De Lithium
LID	Laser Induced Dissociation	LNT2	Lacto-N-Triose De Type 2
LTD	Spectromètre de Masse de la Marque ThermoScientific dans lequel les ions sont piégés radialement	MALDI	Matrix Assisted Laser Desorption Ionization
ManN	Mannosamine	MeOH	Méthanol
MgF <sub>2</sub>	Fluorure de Magnésium	MRM	Multiple Reaction Monitoring
MS	Spectrométrie De Masse	MS/IR	Spectroscopie Infrarouge en Phase Gazeuse
MS2	Spectre de fragmentation	MSn	Spectres de masse de fragments de fragment à l'étage n
Nd:YAG	Neodyme Doped Yttrium Aluminium Garnet	OPA	Amplificateur Paramétrique Optique
OPO	Oscillateur Paramétrique Optique	Neu5Ac	Acide N-Acétylneuraminique
PGC	Porous Graphitic Column	PTM	Post Translation Modification
QqQ	Triple Quadrupole	RELU	Rectifier Linear Unit
RPLC	Reverse Phase Liquid Chromatography	SRM	Single Reaction Monitoring
RVI	Redistribution Vibrationnelle Intramoléculaire	TIL	Tuneable Infrared Laser
SPE	Solid Phase Extraction	UV	UltraViolet
tr	Temps de Rétention	u. arb.	Unité Arbitraire

# Introduction

## 1.1 LES MOLECULES IMPLIQUEES DANS LA VIE

Le monde du vivant repose sur l'assemblage complexe et fin de biomolécules. Ces molécules sont classées en différentes familles. Il s'agit notamment des protéines, des acides nucléiques, des lipides, des sucres... La connaissance de ces molécules surtout d'un point de vue de leur structure donne des indications sur leurs fonctions. Les particularités de chacune de ces classes de biomolécules confèrent des contraintes différentes quant à leur caractérisation. Il y a trente ans, le projet de séquençage complet du génome humain (plus de 3 milliards de paires de bases azotées) a nécessité des milliards de dollars d'investissement, la formation d'un consortium international et plus de dix années de recherches pour être mené à bien<sup>1</sup>. De nos jours, le séquençage est bien moins onéreux (<1000 \$<sup>2</sup>) peut être effectué en quelques jours en routine<sup>3</sup> (récemment, le séquençage a même été réalisé en moins d'une journée<sup>4</sup>). Concernant les protéines, l'avènement de nouvelles sources d'ionisation<sup>5,6</sup> a permis d'analyser des molécules de poids moléculaires importants. Les protéines sont depuis couramment analysées par spectrométrie de masse<sup>7</sup>. En ce qui concerne les sucres, leur caractérisation structurale se heurte à la complexité intrinsèque de cette famille de biomolécules. En effet, les sucres présentent de très nombreuses isoméries, en outre, contrairement aux protéines, les sucres peuvent présenter des structures branchées. De plus il n'est pas toujours possible de disposer de ces molécules en grandes quantités ni même de manière pure. Toutes ces raisons expliquent que la caractérisation structurale de ce type de biomolécules souffre d'un important retard comparativement aux acides nucléiques ou aux protéines car aucune méthode analytique n'apporte de réponse universelle au défi qu'est la caractérisation structurale de ces molécules.

## 1.2 LES SUCRES

Cette classe de biomolécules remplit de multiples fonctions, par exemple : le glucose est impliqué dans le métabolisme énergétique, il constitue également la cellulose, ainsi il assure le maintien de la structure des cellules végétales. Des sucres sont impliqués dans le support de l'information génétique, le désoxyribose constitue les briques de l'ADN (acide désoxyribonucléique) tandis que le ribose constitue celles de l'ARN (acide ribonucléique). Les sucres médient la reconnaissance entre cellules et sont par exemple impliqués dans la reconnaissance de l'hôte par le pathogène<sup>8-10</sup>.

Comme dit plus haut, la présence de nombreux isomères parmi les monosaccharides complexifie l'analyse. Parmi ces isomères, on peut citer les épimères qui diffèrent uniquement par la configuration d'un de leur carbone asymétrique (Figure 1a). L'anomérisation caractérise la configuration du carbone réducteur du sucre et doit également être considérée (Figure 1b). Deux monosaccharides peuvent se lier l'un à l'autre via une liaison glycosidique (formation d'une liaison éther avec perte d'une molécule d'eau) cette liaison peut se faire de très nombreuses manières différentes. Les liaisons entre deux monosaccharides se font généralement entre l'extrémité réductrice d'un sucre et la position 2, 3, 4 ou 6 de l'autre monosaccharide, selon cette position, on parle de régiochimie (illustrée entre milieu et droite de la Figure 1c). On parle de stéréochimie quand le disaccharide diffère selon l'anomérisation du monosaccharide non réducteur (milieu et bas de la Figure 1c). Deux disaccharides de même masse peuvent présenter un contenu différent (milieu et haut de la Figure 1c). La variation peut avoir lieu dans la séquence des monosaccharides (milieu et gauche de la Figure 1c). Dans tous les exemples de la Figure 1c, la masse du disaccharide est la même. En considérant un disaccharide composé uniquement de deux hexoses, il existe plus de 1000 possibilités de liaisons entre ces hexoses. Bien que certaines de ces combinaisons n'existent pas dans la Nature, il est aisé de se rendre compte qu'au vu

de la quantité de structures différentes, la caractérisation structurale de sucre est ardue. En fait, la famille des sucres présente également des molécules de taille plus importante : la cellulose est un homopolymère linéaire, ou de structure encore plus complexes par exemple : le glycogène est un homopolymère présentant des ramifications, les glycosaminoglycanes sont des polymères constitués de deux monomères portant des groupements sulfates situés à différentes positions sur les sucres (à l'exception de l'acide hyaluronique qui ne comporte aucune sulfatation).

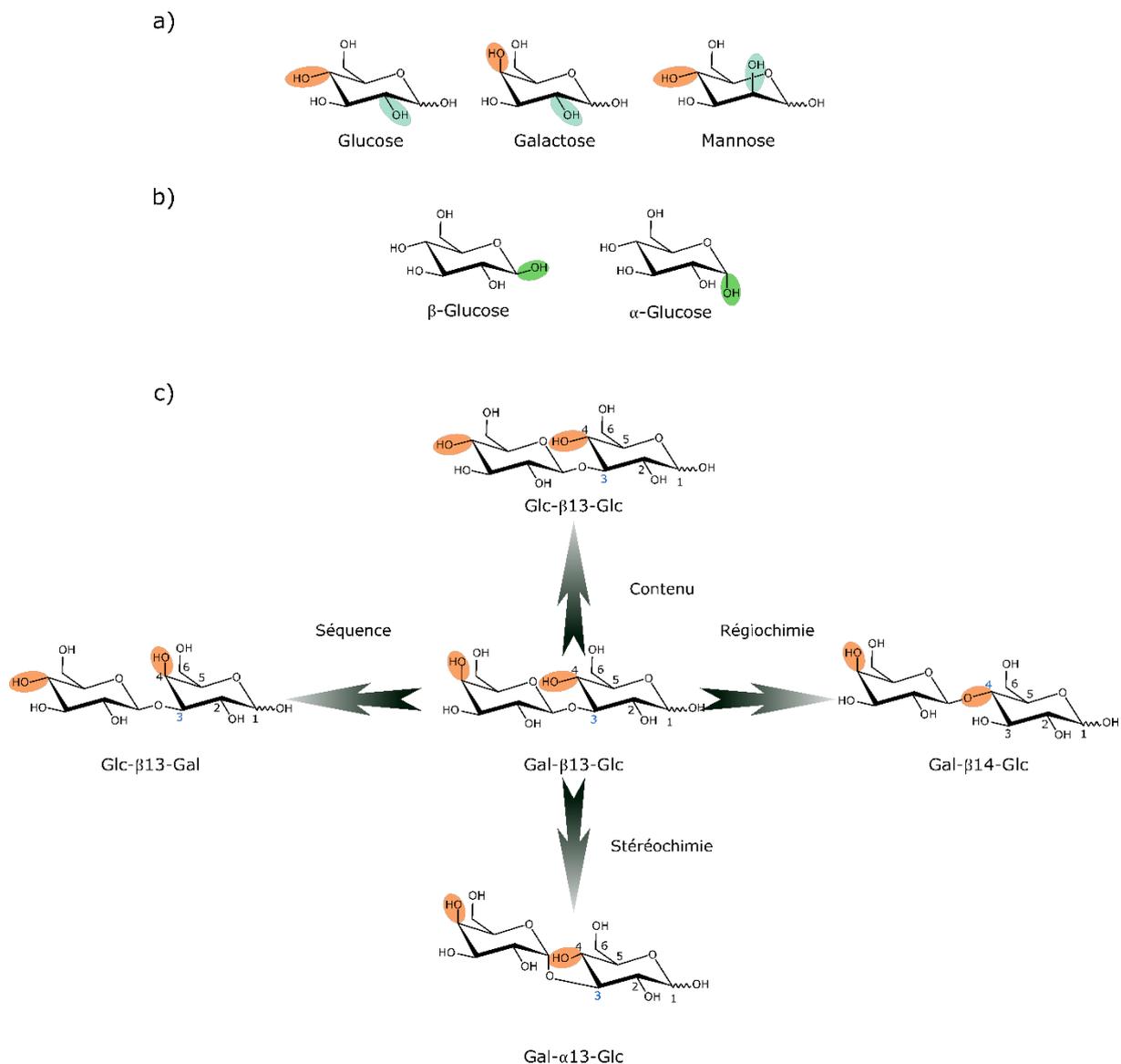


Figure 1 : Isoméris retrouvées dans la famille des sucres avec l'exemple d'épimères a), d'anomères b) et les différentes isoméris éventuellement retrouvées au niveau de la liaison entre deux sucres c).

### 1.3 LES METHODES ADAPTEES A L'ANALYSE DE SUCRES

Comme nous venons de le voir, la caractérisation structurale des sucres s'avère difficile à appréhender. Lorsque l'on parle de caractérisation structurale de molécules, la spectroscopie RMN vient naturellement à l'esprit. Cette méthode peut en effet être utilisée pour étudier la structure des sucres<sup>11,12</sup>. Les spectroscopies IR et Raman sont également des méthodes de choix pour caractériser la structure de tels composés<sup>13</sup>. Ces méthodes de spectroscopie sont très intéressantes dans l'analyse de

telles structures. Toutefois, ces techniques nécessitent de disposer de l'échantillon sous forme pure et en quantité suffisamment importante ce qui peut être rédhibitoire lors de l'analyse de sucres qui peuvent être présents en mélange et en quantité variables. Les étapes de préparation d'échantillons s'avèrent alors laborieuses et ne sont malheureusement pas toujours fructueuses.

Du fait de sa sensibilité, la spectrométrie de masse présente un intérêt tout particulier, en effet cette méthode permet de n'utiliser que de faibles quantités d'échantillons. La spectrométrie de masse mesure le rapport  $m/z$  du composé d'intérêt. Une première information est tirée de cette observation qui permet de remonter à la formule brute du composé. L'utilisation de la spectrométrie de masse en tandem apporte une information supplémentaire. En effet, l'étude des ions produits par fragmentation permet de connaître la nature des groupements composant l'échantillon d'intérêt. Toutefois, là aussi, la méthode n'est pas infaillible. Les sucres sont caractérisés par la présence de nombreuses isoméries qui ne sont pas toujours identifiables même après fragmentation.

La spectrométrie de masse est généralement couplée avec la chromatographie liquide, ainsi, en amont de la spectrométrie de masse les composés présents dans un mélange sont séparés ce qui facilite l'identification des composés contenus dans le mélange. La chromatographie est capable de séparer des isomères<sup>14,15</sup>.

Plus récemment, les auteurs ont couplé de nouvelles méthodes orthogonales à la spectrométrie de masse. C'est le cas de la spectrométrie de mobilité ionique, une méthode séparatives capable de séparer des ions (en fonction de leur charge, masse et structure tridimensionnelle) et de résoudre des isoméries<sup>16-19</sup>. Du côté des méthodes spectroscopiques, le couplage entre la spectrométrie de masse avec des lasers a été décrite dans le domaine de l'UV<sup>20</sup>. L'irradiation par des photons dans le domaine de l'UV est intéressant, en effet l'absorption de photons UV par un ion entraîne la photodissociation de ce dernier avec production de fragments différents de ceux observés en CID apportant une complémentarité d'informations. Dans ce projet de thèse, nous nous intéresserons à une toute autre gamme de longueurs d'onde : l'IR (InfraRouge). Cette gamme permet d'obtenir des informations sur la structure fine des oligosaccharides<sup>21-23</sup>.

### 1.4 LA SPECTROSCOPIE VIBRATIONNELLE EN PHASE GAZEUSE

Afin d'apporter une réponse adaptée à la caractérisation des sucres, l'équipe DYNAMO se sert d'une méthode de caractérisation structurale de composés qui a fait ses preuves dans l'identification de sucres<sup>23</sup>. La méthode développée au sein de notre équipe est un couplage entre la spectrométrie de masse et la spectroscopie infrarouge. Cette méthode n'est pas commercialement disponible, mais elle fait toutefois l'objet de nombreux développements afin de la rendre plus rapide, plus ergonomique, plus facile à utiliser tant d'un point de vue acquisition des spectres que lors du retraitement et de l'interprétation des données.

La méthode utilisée combine donc un spectromètre de masse de type piège à ions et un laser IR accordable dans la gamme des 3  $\mu\text{m}$ . Cette gamme contient les bandes d'absorption des modes de vibration des liaisons C-H, N-H et O-H toutes ces liaisons sont retrouvées dans les biomolécules et en particulier les sucres. L'intérêt de procéder à des analyses en phase gazeuse est que le composé est isolé de son environnement et est sous vide, les espèces potentiellement interférentes sont donc éliminées. De plus, cette analyse permet de s'affranchir d'interactions entre le composé et son environnement, seul le composé d'intérêt est étudié. Un autre avantage est que l'eau est éliminée, en effet, la gamme des 3  $\mu\text{m}$  est très largement dominée par l'absorption engendrée par la présence d'eau. En plus d'éliminer les espèces interférentes, la spectrométrie de masse permet de manipuler des ions et notamment de les contenir pendant suffisamment longtemps afin de permettre leur

irradiation et l'acquisition d'un spectre. Cette méthode en conjonction avec des calculs théoriques, est utilisée dans les laboratoires de physique moléculaire afin de répondre à des questions de physico-chimie et de déterminer la structure de composés en phase gazeuse.

Dans notre cas, en plus de répondre à des questions de physico-chimie, la spectroscopie IRMPD est utilisée pour distinguer des isomères en fonction de leur spectre IRMPD. Des spectres de références sont obtenus sur des échantillons connus, ces spectres sont entrés dans une base de données. Lorsque le spectre IRMPD d'un composé inconnu est acquis, il est comparé avec les spectres présents dans la base de données et une identification peut être obtenue. Toutefois, la constitution d'une base de données de spectres IRMPD de tous les sucres présents dans la Nature est un vœu pieux. Cette approche se heurte à plusieurs inconvénients. Tout d'abord la quantité de structures est bien trop importante pour espérer un jour générer une telle base de données. De plus en augmentant la taille des sucres étudiés, la quantité de groupements chimique augmente et avec elle le nombre de mode de vibration absorbant dans l'IR augmente aussi. Cette augmentation encombre les spectres IRMPD qui deviennent alors inexploitable. Enfin, la chimie de synthèse des sucres n'est pas toujours chose aisée, il n'est pas toujours possible de synthétiser le composé voulu.

C'est pour ces raisons qu'il est en fait plus facile de procéder au séquençage du composé d'intérêt. Pour cela il est nécessaire de diminuer la taille du système étudié, cela peut être fait en utilisant une digestion enzymatique, une hydrolyse chimique, ou en se servant de la fragmentation dans le spectromètre de masse. Ainsi nous nous replaçons dans le cas de plus petits composés, leur analyse permet de remonter à la structure du composé initial. Dans le cas du séquençage il est important de tenir compte des éventuelles modifications apportées par ces méthodes utilisées pour réduire la taille du sucre d'intérêt afin d'interpréter correctement les données<sup>24-26</sup>.

### 1.5 PROJET IROLIGO

La découverte des antibiotiques en 1928 a permis de sauver des millions de vies. Toutefois, à cause de leur utilisation à grande échelle, des souches bactériennes résistantes émergent. L'antibiorésistance s'impose comme un enjeu de santé publique mondial. En 2019 il a été estimé que 5 millions de morts sont imputables au phénomène d'antibiorésistance dont plus d'un million de manière directe<sup>27</sup> et ce chiffre risque d'augmenter dans les années à venir (10 millions de morts par ans prévues pour 2050)<sup>28</sup>. Les antibiotiques agissent sur les bactéries en les tuant ou en inhibant leur croissance. Toutefois, lors du mésusage de ces produits, certaines bactéries développent des résistances à ces antibiotiques qui deviennent alors inefficaces. Certaines bactéries deviennent même multirésistantes (voir *pan-résistantes*), aucun antibiotique n'a alors d'effets.

Les antibiotiques sont également administrés aux animaux d'élevage, en effet, peu de temps après la découverte des antibiotiques, il a été démontré que la supplémentation de la nourriture des animaux d'élevage par des antibiotiques en conjonction avec d'autres molécules accélère la croissance des animaux<sup>29,30</sup>. Les antibiotiques ont donc naturellement été utilisés en prophylaxie et comme promoteurs de croissance pour les animaux d'élevage. Depuis 2006, ces usages sont interdits en union européenne<sup>31</sup>, et de plus en plus réglementés aux Etats-Unis<sup>32,33</sup>. Toutefois, ces usages restent pratiqués dans d'autres régions du monde notamment dans les pays en voie de développement avec une tendance à l'augmentation de la consommation d'antibiotiques dans les années à venir<sup>34</sup>. Pour contrer ce phénomène, plusieurs solutions ont été proposées telles que l'utilisation de phages, le développement de vaccins, l'utilisation de peptides antimicrobiens ou bien encore l'utilisation d'additifs zootechniques (additifs alimentaires à destination des animaux) afin de diminuer l'adhésion des bactéries aux cellules de l'hôte<sup>35-37</sup>. Ce travail de thèse repose sur cette dernière option – utiliser un additif zootechnique – pour empêcher l'infection de l'animal par un agent pathogène.

Ce projet de thèse est une collaboration avec la société Elicityl basée à Crolles en Isère (38). Elicityl travaille au développement d'un additif zootechnique qui serait administré aux animaux d'élevage afin de prévenir l'infection par des agents pathogènes. Comme nous l'avons vu précédemment, les sucres remplissent de nombreux rôles vitaux pour les organismes vivants. L'un de ces rôles est la reconnaissance<sup>38,39</sup> et la communication<sup>40</sup> entre cellules. Des sucres sont exprimés à la surface des cellules et permettent à deux cellules de se reconnaître et d'interagir. Ce sont ces sucres exposés à l'environnement de la cellule qui sont ciblés par les agents pathogènes<sup>41,42</sup>. L'agent reconnaît le sucre exposé à la surface de la cellule à l'aide de protéine qu'il exprime lui-même à la surface de sa membrane et peut alors adhérer à la cellule pour initier l'infection de cette dernière. Une approche pour empêcher cette reconnaissance serait soit de tapisser les cellules de l'hôte pour dissimuler les sucres reconnus par le pathogène. L'autre approche serait de venir gêner les protéines du pathogène dans la reconnaissance de sa cible.

Le but de ce projet est donc de développer un mime des sucres exprimés à la surface de la cellule de l'hôte, afin d'empêcher la reconnaissance par le pathogène. L'additif zootechnique agit alors comme un leurre, les pathogènes reconnaissent ce composé et interagissent avec. Ils sont alors incapables de reconnaître leur cible initiale (les sucres présents à la surface des cellules de l'animal). L'adhésion du pathogène sur la cellule cible est évitée, l'infection n'a pas lieu et les pathogènes sont éliminés naturellement de l'hôte. En pratique ces composés seraient ajoutés à la nourriture ou l'eau de boissons des animaux. Ces composés sont particulièrement prometteurs pour l'industrie agroalimentaire car ils protégeraient les animaux d'élevage. Par exemple dans les élevages porcins, les porcelets sont spécifiquement vulnérables aux infections lors du sevrage<sup>43</sup>, c'est à ce moment que de nombreuses infections surviennent<sup>44</sup>. Les conséquences vont d'un retard de croissance, des diarrhées, et même à une déshydratation potentiellement mortelle pour l'individu.

L'additif zootechnique en cours de développement est un glycoconjugué, il s'agit d'un polymère sur lequel sont greffés des oligosaccharides. Son administration aux animaux d'élevage qui seront ensuite consommés par l'Homme implique que ce composé soit parfaitement caractérisé ainsi que les éventuelles impuretés qui pourraient être générées durant la synthèse du glycoconjugué. Cette caractérisation assurera l'innocuité du produit sur le consommateur. Sa caractérisation structurale est un véritable défi en raison de l'hétérogénéité de tels produits glycoconjugués, mais également par les différentes isoméries attendues dans les produits composant les impuretés. Au sein de l'institut Lumière Matière, nous avons pour objectif de développer les méthodes permettant de caractériser ce type de composé.

Pour des raisons de confidentialité, les molécules développées par Elicityl ne seront pas présentées dans ce manuscrit. Nous présenterons les développements méthodologiques effectués au cours de ce projet qui seront illustrés sur un système modèle pertinent.

Pour répondre à ces objectifs, le manuscrit se décompose de la manière suivante :

Dans une première partie, nous introduirons la spectroscopie IRMPD. Nous verrons son utilisation afin de répondre à des questions de physico-chimie avec l'identification de la structure du produit issu d'un mécanisme de réarrangement en phase gazeuse d'un oligosaccharide fucosylé : la migration du fucose. Le produit résultant de ce phénomène demeure inconnu bien que ce la migration du fucose en phase gazeuse soit maintenant connue depuis une trentaine d'années<sup>45</sup>. Ensuite nous verrons un exemple plus appliqué de la spectroscopie IRMPD afin de reconnaître deux isomères de sucres, ainsi que le séquençage d'un tétrasaccharide pour remonter à sa structure.

La seconde partie traitera de nouveaux développements instrumentaux avec notamment la présentation d'un nouveau montage développé dans le but de diminuer le temps d'acquisition des

spectres IRMPD<sup>46</sup>, et son amélioration avec l'ajout d'un second laser<sup>47</sup>. Ces développements n'impliquent pas que des changements instrumentaux, ils ont aussi été effectués d'un point de vue de l'interprétation des données afin de faciliter l'utilisation de la spectroscopie IRMPD par des utilisateurs non experts. L'identification de molécules à partir de leur spectre IRMPD peut alors se faire au moyen d'un calcul de score de ressemblance (ou d'erreur) entre un spectre inconnu et un spectre de référence contenu dans une base de données<sup>48</sup>. Cette identification peut également se faire avec des méthodes plus « élaborées » telles qu'une analyse de données non supervisée, ou alors par une approche d'intelligence artificielle<sup>49</sup>. Tous ces développements seront présentés sur des monosaccharides ou des petits oligosaccharide ( $\leq 3$  résidus). Ces oligosaccharides ont été sélectionnés car ils sont semblables aux composés pouvant être obtenus après synthèse du glycoconjugué et servent ainsi de modèle pour notre méthode. L'analyse se fait sur des composés purs de manière à évaluer uniquement l'identification finale des produits.

Nous verrons ensuite dans une troisième partie le couplage entre la chromatographie liquide et la spectroscopie IRMPD. Notamment, les contraintes techniques qu'il a fallu résoudre afin de mener à bien un tel couplage. Nous verrons par la suite différents modes permis par ce couplage avec l'application sur des échantillons complexes comme les produits de digestion d'un pentasaccharide, des mélanges dont l'un contenant des isomères avec l'étude d'un fragment d'un des deux isomères par spectroscopie IRMPD en ligne après la séparation chromatographique et un mode de screening permettant de détecter rapidement un motif d'intérêt au sein d'un sucre complexe. Ici nous utilisons la chromatographie liquide pour séparer des composés en amont de la spectroscopie IRMPD, en effet après une telle synthèse, l'échantillon produit ne sera pas un composé d'une seule espèce, mais il présentera diverses espèces provenant de la synthèse : impuretés, sous-produits, réactifs n'ayant pas réagi... cette séparation permet de simplifier l'analyse.

Finalement dans la dernière partie, nous étudierons des composés encore plus complexes. Les espèces considérées seront des composés glycoconjugués : à la diversité de structures des sucres, une difficulté supplémentaire est ajoutée avec la recherche de la localisation de la liaison composé-glycane. Pour cela nous étudierons une glycoprotéine modèle. Tout d'abord, la recherche d'une glycosylation sur cette glycoprotéine est faite au moyen de la spectroscopie IR en phase condensée. Les modes de vibration des sucres ont été localisés dans la région caractéristique des sucres du spectre IR, c'est alors la présence de cette bande qui indique la glycosylation de la protéine. Dans une seconde partie, nous cherchons à identifier des glycopeptides après digestion par des enzymes protéolytiques d'une protéine. Le mélange peptides et glycopeptides ainsi obtenu est analysé en chromatographie liquide avant d'être injecté dans le piège à ions. Les peptides et glycopeptides sont alors irradiés à une longueur d'onde fixe dans l'IR avant détection. Là aussi, la longueur d'onde est choisie spécialement pour exciter les modes de vibration des liaisons O-H présents dans les sucres. La glycoprotéine considérée permet de modéliser la complexité de l'analyse du composé d'intérêt.

## 2 BIBLIOGRAPHIE

- 1 International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, *Nature*, 2004, **431**, 931–945.
- 2 The Cost of Sequencing a Human Genome, <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>, (accessed September 15, 2023).
- 3 J.-Y. L. E. Gall and P. Debré, Séquençage des génomes et médecine personnalisée : perspectives et limites., *Bull. Académie Natl. Médecine*, 2014, **198**, 101–117.
- 4 J. E. Gorzynski, S. D. Goenka, K. Shafin, T. D. Jensen, D. G. Fisk, M. E. Grove, E. Spiteri, T. Pesout, J. Monlong, G. Baid, J. A. Bernstein, S. Ceresnak, P.-C. Chang, J. W. Christle, H. Chubb, K. P. Dalton, K. Dunn, D. R. Garalde, J. Guillory, J. W. Knowles, A. Kolesnikov, M. Ma, T. Moscarello, M. Nattestad, M. Perez, M. R. Z. Ruzhnikov, M. Samadi, A. Setia, C. Wright, C. J. Wusthoff, K. Xiong, T. Zhu, M. Jain, F. J. Sedlazeck, A. Carroll, B. Paten and E. A. Ashley, Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting, *N. Engl. J. Med.*, 2022, **386**, 700–702.
- 5 Michael. Karas and Franz. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal. Chem.*, 1988, **60**, 2299–2301.
- 6 J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, Electrospray Ionization for Mass Spectrometry of Large Biomolecules, *Science*, 1989, **246**, 64–71.
- 7 R. Aebersold and M. Mann, Mass spectrometry-based proteomics, *Nature*, 2003, **422**, 198–207.
- 8 W. Weis, J. H. Brown, S. Cusack, J. C. Paulson, J. J. Skehel and D. C. Wiley, Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid, *Nature*, 1988, **333**, 426–431.
- 9 N. K. Sauter, G. D. Glick, R. L. Crowther, S. J. Park, M. B. Eisen, J. J. Skehel, J. R. Knowles and D. C. Wiley, Crystallographic detection of a second ligand binding site in influenza virus hemagglutinin., *Proc. Natl. Acad. Sci.*, 1992, **89**, 324–328.
- 10 K. Moonens, J. Bouckaert, A. Coddens, T. Tran, S. Panjekar, M. De Kerpel, E. Cox, H. Remaut and H. De Greve, Structural insight in histo-blood group binding by the F18 fimbrial adhesin FedF, *Mol. Microbiol.*, 2012, **86**, 82–95.
- 11 B. Meyer and T. Peters, NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors, *Angew. Chem. Int. Ed.*, 2003, **42**, 864–890.
- 12 C. Fontana and G. Widmalm, Primary Structure of Glycans by NMR Spectroscopy, *Chem. Rev.*, 2023, **123**, 1040–1102.
- 13 E. Wiercigroch, E. Szafraniec, K. Czamara, M. Z. Pacia, K. Majzner, K. Kochan, A. Kaczor, M. Baranska and K. Malek, Raman and infrared spectroscopy of carbohydrates: A review, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, 2017, **185**, 317–335.
- 14 A. V. Everest-Dass, J. L. Abrahams, D. Kolarich, N. H. Packer and M. P. Campbell, Structural Feature Ions for Distinguishing N- and O-Linked Glycan Isomers by LC-ESI-IT MS/MS, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 895–906.
- 15 K. Wongtrakul-Kish, I. Walsh, L. C. Sim, A. Mak, B. Liau, V. Ding, N. Hayati, H. Wang, A. Choo, P. M. Rudd and T. Nguyen-Khuong, Combining Glucose Units, m/z, and Collision Cross Section Values: Multiattribute Data for Increased Accuracy in Automated Glycosphingolipid Glycan Identifications and Its Application in Triple Negative Breast Cancer, *Anal. Chem.*, 2019, **91**, 9078–9085.
- 16 F. Zhu, S. Lee, S. J. Valentine, J. P. Reilly and D. E. Clemmer, Mannose7 Glycan Isomer Characterization by IMS-MS/MS Analysis, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 2158–2166.
- 17 G. Nagy, I. K. Attah, S. V. B. Garimella, K. Tang, Y. M. Ibrahim, E. S. Baker and R. D. Smith, Unraveling the isomeric heterogeneity of glycans: ion mobility separations in structures for lossless ion manipulations, *Chem. Commun.*, 2018, **54**, 11701–11704.
- 18 C. Manz, M. Grabarics, F. Hoberg, M. Pugini, A. Stuckmann, W. B. Struwe and K. Pagel, Separation of isomeric glycans by ion mobility spectrometry – the impact of fluorescent labelling, *Analyst*, 2019, **144**, 5292–5298.
- 19 A. Mookherjee, S. S. Uppal, T. A. Murphree and M. Guttman, Linkage Memory in

- Underivatized Protonated Carbohydrates, *J. Am. Soc. Mass Spectrom.*, 2021, **32**, 581–589.
- 20 D. R. Klein, F. E. I. Leach, I. J. Amster and J. S. Brodbelt, Structural Characterization of Glycosaminoglycan Carbohydrates Using Ultraviolet Photodissociation, *Anal. Chem.*, 2019, **91**, 6019–6026.
- 21 E. Mucha, A. I. González Flórez, M. Marianski, D. A. Thomas, W. Hoffmann, W. B. Struwe, H. S. Hahm, S. Gewinner, W. Schöllkopf, P. H. Seeberger, G. von Helden and K. Pagel, Glycan Fingerprinting via Cold-Ion Infrared Spectroscopy, *Angew. Chem. Int. Ed.*, 2017, **56**, 11248–11251.
- 22 C. Masellis, N. Khanal, M. Z. Kamrath, D. E. Clemmer and T. R. Rizzo, Cryogenic Vibrational Spectroscopy Provides Unique Fingerprints for Glycan Identification, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 2217–2222.
- 23 B. Schindler, L. Barnes, G. Renois, C. Gray, S. Chambert, S. Fort, S. Flitsch, C. Loison, A.-R. Allouche and I. Compagnon, Anomeric memory of the glycosidic bond upon fragmentation and its consequences for carbohydrate sequencing, *Nat. Commun.*, 2017, **8**, 973.
- 24 A. J. Rhomberg, Z. Shriver, K. Biemann and R. Sasisekharan, Mass spectrometric evidence for the enzymatic mechanism of the depolymerization of heparin-like glycosaminoglycans by heparinase II, *Proc. Natl. Acad. Sci.*, 1998, **95**, 12232–12237.
- 25 D. J. Harvey, T. S. Mattu, M. R. Wormald, L. Royle, R. A. Dwek and P. M. Rudd, “Internal Residue Loss”: Rearrangements Occurring during the Fragmentation of Carbohydrates Derivatized at the Reducing Terminus, *Anal. Chem.*, 2002, **74**, 734–740.
- 26 M. J. Amicucci, E. Nandita, A. G. Galermo, J. J. Castillo, S. Chen, D. Park, J. T. Smilowitz, J. B. German, D. A. Mills and C. B. Lebrilla, A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis, *Nat. Commun.*, 2020, **11**, 3963.
- 27 C. J. L. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, G. Robles Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool, S. C. Johnson, A. J. Browne, M. G. Chipeta, F. Fell, S. Hackett, G. Haines-Woodhouse, B. H. Kashef Hamadani, E. A. P. Kumaran, B. McManigal, S. Achalapong, R. Agarwal, S. Akech, S. Albertson, J. Amuasi, J. Andrews, A. Aravkin, E. Ashley, F.-X. Babin, F. Bailey, S. Baker, B. Basnyat, A. Bekker, R. Bender, J. A. Berkley, A. Bethou, J. Bielicki, S. Boonkasidecha, J. Bukosia, C. Carvalheiro, C. Castañeda-Orjuela, V. Chansamouth, S. Chaurasia, S. Chiurchiù, F. Chowdhury, R. Clotaire Donatien, A. J. Cook, B. Cooper, T. R. Cressey, E. Criollo-Mora, M. Cunningham, S. Darboe, N. P. J. Day, M. De Luca, K. Dokova, A. Dramowski, S. J. Dunachie, T. Duong Bich, T. Eckmanns, D. Eibach, A. Emami, N. Feasey, N. Fisher-Pearson, K. Forrest, C. Garcia, D. Garrett, P. Gastmeier, A. Z. Giref, R. C. Greer, V. Gupta, S. Haller, A. Haselbeck, S. I. Hay, M. Holm, S. Hopkins, Y. Hsia, K. C. Iregbu, J. Jacobs, D. Jarovsky, F. Javanmardi, A. W. J. Jenney, M. Khorana, S. Khusuwan, N. Kissoon, E. Kobeissi, T. Kostyanov, F. Krapp, R. Krumkamp, A. Kumar, H. H. Kyu, C. Lim, K. Lim, D. Limmathurotsakul, M. J. Loftus, M. Lunn, J. Ma, A. Manoharan, F. Marks, J. May, M. Mayxay, N. Mturi, T. Munera-Huertas, P. Musicha, L. A. Musila, M. M. Mussi-Pinhata, R. N. Naidu, T. Nakamura, R. Nanavati, S. Nangia, P. Newton, C. Ngoun, A. Novotney, D. Nwakanma, C. W. Obiero, T. J. Ochoa, A. Olivas-Martinez, P. Olliaro, E. Ooko, E. Ortiz-Brizuela, P. Ounchanum, G. D. Pak, J. L. Paredes, A. Y. Peleg, C. Perrone, T. Phe, K. Phommason, N. Plakkal, A. Ponce-de-Leon, M. Raad, T. Ramdin, S. Rattavong, A. Riddell, T. Roberts, J. V. Robotham, A. Roca, V. D. Rosenthal, K. E. Rudd, N. Russell, H. S. Sader, W. Saengchan, J. Schnall, J. A. G. Scott, S. Seekaew, M. Sharland, M. Shivamallappa, J. Sifuentes-Osornio, A. J. Simpson, N. Steenkeste, A. J. Stewardson, T. Stoeva, N. Tasak, A. Thaiprakong, G. Thwaites, C. Tigoi, C. Turner, P. Turner, H. R. Van Doorn, S. Velaphi, A. Vongpradith, M. Vongsouvath, H. Vu, T. Walsh, J. L. Walson, S. Waner, T. Wangrangsamakul, P. Wannapinij, T. Wozniak, T. E. M. W. Young Sharma, K. C. Yu, P. Zheng, B. Sartorius, A. D. Lopez, A. Stergachis, C. Moore, C. Dolecek and M. Naghavi, Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis, *The Lancet*, 2022, **399**, 629–655.
- 28 J. O’Neill, *Tackling drug-resistant infections globally : final report and recommendations / the Review on Antimicrobial Resistance*, 2016.
- 29 E. L. L. Stokstad and T. H. Jukes, Growth-Promoting Effect of Aureomycin on Turkey Poults, *Poult. Sci.*, 1950, **29**, 611–612.
- 30 R. W. Luecke, F. Thorp, H. W. Newland and W. N. Mcmillen, The Growth Promoting Effects of Various Antibiotics on Pigs, *J. Anim. Sci.*, 1951, **10**, 538–542.

- 31 S. Van Puyvelde, S. Deborggraeve and J. Jacobs, Why the antibiotic resistance crisis requires a One Health approach, *Lancet Infect. Dis.*, 2018, **18**, 132–134.
- 32 Forty years later, FDA finally restricts use of antibiotics in livestock, [https://thefern.org/ag\\_insider/forty-years-later-fda-finally-restricts-use-antibiotics-livestock/](https://thefern.org/ag_insider/forty-years-later-fda-finally-restricts-use-antibiotics-livestock/), (accessed March 6, 2024).
- 33 K. Hill, FDA Changes to Livestock Antibiotic Regulations, <https://polk.extension.wisc.edu/2022/10/24/fda-changes-to-livestock-antibiotic-regulations/>, (accessed March 6, 2024).
- 34 T. P. Van Boeckel, C. Brower, M. Gilbert, B. T. Grenfell, S. A. Levin, T. P. Robinson, A. Teillant and R. Laxminarayan, Global trends in antimicrobial use in food animals, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 5649–5654.
- 35 H. K. Allen, J. Trachsel, T. Looft and T. A. Casey, Finding alternatives to antibiotics, *Ann. N. Y. Acad. Sci.*, 2014, **1323**, 91–100.
- 36 G. Cheng, H. Hao, S. Xie, X. Wang, M. Dai, L. Huang and Z. Yuan, Antibiotic alternatives: the substitution of antibiotics in animal husbandry?, *Front. Microbiol.*
- 37 A. Coddens, M. Loos, D. Vanrompay, J. P. Remon and E. Cox, Cranberry extract inhibits in vitro adhesion of F4 and F18+ *Escherichia coli* to pig intestinal epithelium and reduces in vivo excretion of pigs orally challenged with F18+ verotoxigenic *E. coli*, *Vet. Microbiol.*, 2017, **202**, 64–71.
- 38 P. R. Crocker and T. Feizi, Carbohydrate recognition systems: functional triads in cell–cell interactions, *Curr. Opin. Struct. Biol.*, 1996, **6**, 679–691.
- 39 R. L. Schnaar, Glycobiology simplified: diverse roles of glycan recognition in inflammation, *J. Leukoc. Biol.*, 2016, **99**, 825–838.
- 40 Y. Gao, X. Luan, J. Melamed and I. Brockhausen, Role of Glycans on Key Cell Surface Receptors That Regulate Cell Proliferation and Cell Death, *Cells*, 2021, **10**, 1252.
- 41 L. Byrd-Leotis, R. D. Cummings and D. A. Steinhauer, The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase, *Int. J. Mol. Sci.*, 2017, **18**, 1541.
- 42 J. Poole, C. J. Day, M. von Itzstein, J. C. Paton and M. P. Jennings, Glycointeractions in bacterial pathogenesis, *Nat. Rev. Microbiol.*, 2018, **16**, 440–452.
- 43 J.-P. Lallès, G. Boudry, C. Favier, N. Le Floc’h, I. Luron, L. Montagne, I. P. Oswald, S. Pié, C. Piel and B. Sève, Gut function and dysfunction in young pigs: physiology, *Anim. Res.*, 2004, **53**, 301–316.
- 44 J. M. Fairbrother, É. Nadeau and C. L. Gyles, *Escherichia coli* in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies, *Anim. Health Res. Rev.*, 2005, **6**, 17–39.
- 45 B. Moge, B. Schindler, O. Yeni and I. Compagnon, Fucose Migration Pathways Identified Using Infrared Spectroscopy, *Angew. Chem. Int. Ed.*, 2023, **62**, e202300538.
- 46 O. Yeni, B. Schindler, B. Moge and I. Compagnon, Rapid IRMPD (InfraRed multiple photon dissociation) analysis for glycomics, *Analyst*, 2022, **147**, 312–317.
- 47 B. Moge, O. Yeni, A. Infantino and I. Compagnon, CO<sub>2</sub> laser enhanced rapid IRMPD spectroscopy for glycan analysis, *Int. J. Mass Spectrom.*, 2023, **490**, 117071.
- 48 O. Yeni, S. Ollivier, B. Moge, D. Ropartz, H. Rogniaux, L. Legentil, V. Ferrières and I. Compagnon, Ring-Size Memory of Galactose-Containing MS/MS Fragments: Application to the Detection of Galactofuranose in Oligosaccharides and Their Sequencing, *J. Am. Chem. Soc.*, 2023, **145**, 15180–15187.
- 49 T. Barillot, B. Schindler, B. Moge, E. Fadda, F. Lépine and I. Compagnon, GIAIcomics: a deep neural network classifier for spectroscopy-augmented mass spectrometric glycans data, *Beilstein J. Org. Chem.*, 2023, **19**, 1825–1831.

# Chapitre I : Instrumentation, principe et utilisation de l'IRMPD

## TABLE DES MATIERES

---

<b>1</b>	<b>Couplages MS/IR .....</b>	<b>22</b>
1.1	Principe de la spectroscopie d'action .....	22
1.2	Exemples de méthodes de spectroscopie IR en phase gazeuse .....	23
1.2.1	Messenger tagging.....	23
1.2.2	Helium droplet .....	23
1.2.3	IR-UV .....	23
1.2.4	L'IRMPD.....	24
<b>2</b>	<b>La spectroscopie IRMPD.....</b>	<b>24</b>
2.1	L'IRMPD comme moyen de fragmentation .....	24
2.2	La spectroscopie IRMPD .....	24
2.3	Acquisition d'un spectre IRMPD.....	25
2.4	Gamme de nombres d'onde d'intérêt.....	26
<b>3</b>	<b>Instrumentation .....</b>	<b>27</b>
3.1	Différentes sources lasers .....	27
3.2	Spectromètres de masse utilisés .....	28
3.3	Montage expérimental mis en place à l'iLM .....	28
<b>4</b>	<b>L'IRMPD et les questions de physico-chimie : identification de fragments d'oligosaccharides fucosylés .....</b>	<b>30</b>
4.1	La migration du fucose .....	30
4.2	Matériel et méthode .....	31
4.3	Migration Interosidique.....	34
4.4	Migration intraosidique.....	35
4.5	Comparaison de ces deux types de migration.....	36
4.6	Conclusion et ouverture .....	36
<b>5</b>	<b>Applications de la spectroscopie IRMPD à la chimie analytique .....</b>	<b>38</b>
5.1	Applications analytiques.....	38
5.2	Identification de sucres à l'iLM.....	38
5.3	Méthode de séquençage .....	39
5.3.1	Principe du séquençage .....	39
5.3.2	Séquençage du Lacto-N-tetraose à des fins d'illustration .....	40
<b>6</b>	<b>Conclusion .....</b>	<b>43</b>
<b>7</b>	<b>Bibliographie .....</b>	<b>44</b>

# 1 COUPLAGES MS/IR

## 1.1 PRINCIPE DE LA SPECTROSCOPIE D'ACTION

La spectroscopie IR en phase gazeuse est une méthode de caractérisation structurale qui couple la spectrométrie de masse et la spectroscopie IR. Cette technique allie ainsi les avantages de ces deux méthodes. La spectrométrie de masse apporte la sensibilité et la sélectivité. Elle apporte aussi la possibilité de manipuler des ions en phase gazeuse et par fragmentation d'étudier la structure du composé. La spectroscopie IR sert à l'analyse structurale et à l'identification des isomères grâce à leur empreinte spectroscopique.

Un autre intérêt de la spectroscopie IR en phase gazeuse est que la molécule est isolée de son environnement. L'absence d'environnement permet d'éliminer l'absorbance du milieu. Par exemple, en phase condensée, la présence d'eau, rend la gamme des 3  $\mu\text{m}$  inexploitable. Cette gamme très informative devient exploitable en phase gazeuse. De plus, l'élimination du solvant permet également d'éliminer son influence sur le composé (liaisons intermoléculaires). Le spectre infrarouge obtenu en phase gazeuse présente des pics plus fins que si on l'observait en phase condensée.

La spectroscopie « classique » consiste à irradier un échantillon et à mesurer l'absorption en fonction de la longueur d'onde de la lumière incidente. Pour calculer cette absorption, il est nécessaire de connaître l'intensité de lumière incidente et de mesurer l'intensité de lumière transmise. La différence d'intensité entre la lumière incidente et la lumière transmise doit être significative (Figure 1a).

L'acquisition de spectres IR en phase gazeuse se heurte à une limite. La densité d'ions dans un piège à ions est très faible par rapport à la densité de photons incidents. La mesure de l'absorption directe de la lumière par l'échantillon est impossible, la différence d'intensité entre la lumière incidente et la lumière transmise est trop ténue pour être observée. C'est pourquoi, la spectroscopie utilisée ici est une spectroscopie d'action. Ce type de spectroscopie mesure de manière indirecte l'absorption de lumière par l'échantillon en mesurant l'effet de la lumière sur l'échantillon (Figure 1b).

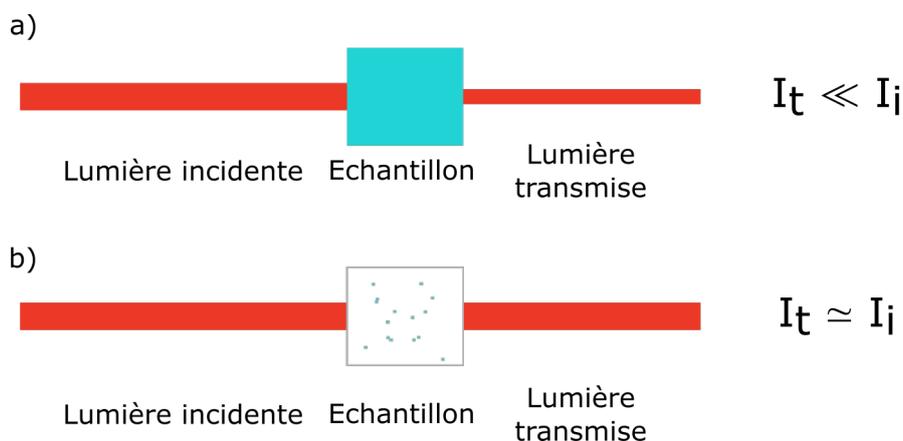


Figure 1 : Spectroscopie d'absorption en phase condensée, manière « conventionnelle » de faire de la spectroscopie, l'intensité de lumière transmise ( $I_t$ ) est bien inférieure à l'intensité de lumière incidente ( $I_i$ ) a), spectroscopie en phase diluée, la différence entre l'intensité de la lumière incidente et celle de la lumière transmise est indécélable l'approche par la spectroscopie d'action est alors nécessaire b).

Avec le couplage laser/MS, le nombre de photons incidents dans une impulsion laser est de l'ordre de  $10^{17(*)}$  tandis que le nombre d'ions dans le spectromètre de masse est plutôt de l'ordre de  $10^3$  à  $10^4$ . Cette différence est trop importante pour observer une quelconque différence d'intensité entre lumière incidente et lumière transmise.

La mesure est donc faite de manière indirecte. Le faisceau IR irradie les ions en phase gazeuse. L'effet de la lumière sur les ions est alors observé, le suivi des effets de l'irradiation est effectué par spectrométrie de masse. Différents couplages MS/IR existent, ils sont succinctement présentés ci-après.

## 1.2 EXEMPLES DE METHODES DE SPECTROSCOPIE IR EN PHASE GAZEUSE

### 1.2.1 Messenger tagging

Dans ce type de spectroscopie, un marqueur : typiquement un gaz ( $\text{He}$ ,  $\text{H}_2$  ou  $\text{N}_2$ ) se fixe sur les ions dans le spectromètre de masse, la température est basse (autours de 10 K). En cas d'irradiation résonante, la liaison entre l'ion et le marqueur se rompt. Le complexe s'étant dissocié, le rapport  $m/z$  a changé ce qui peut être mesuré par spectrométrie de masse<sup>1</sup>.

### 1.2.2 Helium droplet

Là aussi, les ions sont soumis à de basses températures (0.37 K), cette fois-ci, les ions sont emprisonnés dans une gouttelette d'hélium. La gouttelette d'hélium est transparente à l'IR, l'irradiation du complexe par un photon de nombre d'onde résonant avec un mode de vibration de l'ion va entraîner l'évaporation de la gouttelette d'hélium (partielle ou totale). L'ion est détecté au rapport  $m/z$  attendu et le spectre IR est mesuré en traçant l'efficacité d'éjection de l'ion en fonction du nombre d'onde<sup>2</sup>.

### 1.2.3 IR-UV

Il existe plusieurs variantes à cette méthode, mais elles ont toutes en commun l'usage de deux photons (deux lasers sont nécessaires). Dans cette technique, les composés sont présents sous forme neutre. Mis en phase gazeuse puis entraînés au moyen d'un jet moléculaire, l'échantillon subit deux irradiations successives<sup>3,4</sup>. Les deux irradiations sont faites à des longueurs d'onde différentes. Tout d'abord un photon  $h\nu_1$  provenant d'un premier laser porte le niveau électronique de l'état  $S_0$  vers l'état  $S_1$ , un photon  $h\nu_2$  provenant du second laser permet d'ioniser la molécule depuis son état excité. En faisant varier la longueur d'onde du photon  $h\nu_1$  il est possible d'en déduire les transitions électroniques.

Une variante de cette spectroscopie est de fixer la longueur d'onde du photon  $h\nu_2$  dans l'UV. L'irradiation d'un composé dans son état fondamental par ce photon  $h\nu_2$  ionise le composé. Le photon  $h\nu_1$  balaie les longueurs d'onde dans l'IR. Lorsque  $h\nu_1$  est résonant avec une transition vibrationnelle du composé, le photon est absorbé. L'état fondamental est dépeuplé, la population d'électrons dans l'état initial diminue, le photon  $h\nu_2$  ne sera pas absorbé et donc l'intensité des composés ionisés diminue, ce qui est visible par une diminution du signal sur le spectre. On obtient ainsi des informations sur les états des niveaux vibrationnels du composé étudié. Il existe d'autres variantes à cette méthode.

Dans ce type de spectroscopie, les composés passent en phase gazeuse au moyen d'un jet moléculaire ce qui permet de les observer à basse température. Cette méthode nécessite par ailleurs la présence

---

\* En prenant en compte un laser avec une cadence de 10 Hz, émettant une puissance de sortie de 100 mW et des photons de nombre d'onde =  $3500 \text{ cm}^{-1}$ .

d'un groupement chromophore (absorption de photons UV) dans la structure du composé analysé (le greffage d'un chromophore est possible).

#### 1.2.4 L'IRMPD

D'une manière générale, ces spectroscopies vibrationnelles ont démontré leur pertinence dans l'analyse structurale de sucres<sup>5-7</sup>. En effet, elles permettent d'obtenir une très bonne résolution spectrale notamment grâce au fait que les températures soient très basses. En revanche elles nécessitent une instrumentation conséquente (présence de deux lasers pour la spectroscopie IR-UV, nécessité d'introduire une gouttelette d'hélium, ou un marqueur/messenger dans les deux autres types de spectroscopie). C'est pour cette raison, que notre choix s'est porté sur un autre type de spectroscopie vibrationnelle : la spectroscopie IRMPD (InfraRed MultiPhoton Dissociation) qui nécessite un équipement moins important et plus simple à mettre en œuvre, elle est décrite dans la partie suivante.

## 2 LA SPECTROSCOPIE IRMPD

---

### 2.1 L'IRMPD COMME MOYEN DE FRAGMENTATION

Initialement, la spectroscopie IRMPD a consisté en un couplage entre un laser IR et un spectromètre de masse de type FT-ICR (Fourier Transform Ion Cyclotron Resonance)<sup>8</sup>. Ce couplage a été utilisé pour la photofragmentation d'ions confinés dans la cellule FT-ICR (initialement utilisé comme alternative à la CID). L'irradiation des ions piégés dans la cellule d'un FT-ICR par un laser IR permet d'apporter l'énergie nécessaire pour la fragmentation des ions. Un tel couplage a l'avantage d'éviter un cycle d'injection et de repompage du gaz de collision dans la cellule nécessaire à un événement de CID.

Une telle méthode a trouvé de nombreuses applications<sup>9</sup> d'un point de vue caractérisation structurale de molécules notamment dans le domaine des sciences de la vie avec l'étude des oligonucléotides<sup>10,11</sup>, des peptides<sup>12</sup>, de la protéomique<sup>10,13,14</sup> ainsi que de la glycomique<sup>15,16</sup>. Les premières mesures utilisaient des lasers CO<sub>2</sub> émettant à 10,6  $\mu\text{m}$  et capables de délivrer de grandes puissances.

### 2.2 LA SPECTROSCOPIE IRMPD

En utilisant l'étroite accordabilité des lasers CO<sub>2</sub> (900-1100  $\text{cm}^{-1}$ ) il a été possible d'ajouter une dimension spectroscopique à la photofragmentation et donc d'acquérir les premiers spectres IRMPD permettant de discriminer des isomères<sup>17,18</sup>. Par cette méthode, dans les années 2010 Stefan et Eyler ont même pu différencier des isomères de sucres<sup>19,20</sup>. La spectroscopie IRMPD nécessite l'utilisation de sources lasers capables de délivrer de grandes puissances sur une large gamme de nombres d'onde. Ainsi des mesures de spectroscopie IRMPD se font au moyen de lasers à électrons libres (principalement dans la gamme comprise entre 900 et 1800  $\text{cm}^{-1}$ ). Mais c'est le développement de nouveaux lasers accordables bien plus petits qui a permis de l'avènement de la spectroscopie IRMPD

La fragmentation IRMPD a lieu lorsque la longueur d'onde  $\lambda$  des photons provenant du laser est en résonance avec un mode de vibration de l'ion piégé. L'ion va donc absorber un photon son énergie interne augmente (transitions  $\nu_0 \rightarrow \nu_1$ ). Le gain d'énergie est alors redistribué sur les autres modes de vibration de l'ion piégé par des mécanismes de RVI (redistribution vibrationnelle intramoléculaire) (Figure 2a). Le seuil de dissociation de l'ion ne peut pas être atteint après l'absorption d'un seul photon

(les photons IR sont trop peu énergétiques :  $0.43 \text{ eV}^\dagger$ , la liaison C-H dans le méthane nécessite elle environ  $4.7 \text{ eV}$  pour se dissocier<sup>21</sup>). La photofragmentation nécessite l'absorption de plusieurs photons, le processus IRMPD est un processus multiphotonique. Après la RVI, le mode de vibration qui a absorbé le premier photon est de nouveau disponible pour absorber un deuxième photon. Après absorption d'un photon, le processus de redistribution de l'énergie a de nouveau lieu. Ce phénomène fait progressivement gagner de l'énergie à l'ion. Ce gain d'énergie se poursuit jusqu'à ce que l'ion se fragmente lorsqu'il atteint son seuil de dissociation (Figure 2b). La fragmentation a lieu au niveau de la liaison la plus faible qui n'est pas forcément la liaison dont le mode de vibration a été excité. C'est cette fragmentation qui est mesurée par spectrométrie de masse, un fragment apparaît sur le spectre de masse (marqué en rouge sur la Figure 2c). Dans tout le processus, seule une transition vibrationnelle a été impliquée.

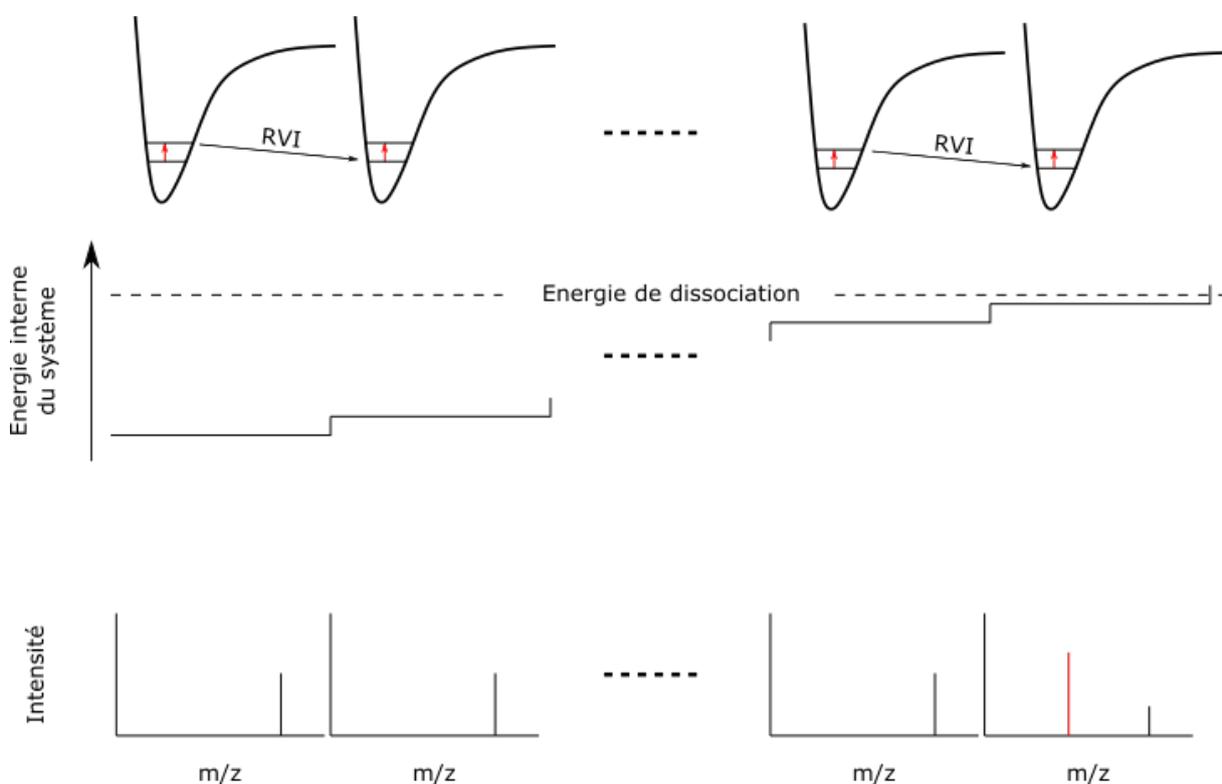


Figure 2 : Principe de l'IRMPD, l'énergie absorbée est redistribuée dans l'ensemble de l'ion, ce qui permet le gain d'énergie et la disponibilité du mode de vibration pour absorption d'un nouveau photon a), le processus se répète jusqu'à fragmentation de l'ion. Le niveau d'énergie de l'ion augmente photon par photon avant d'atteindre le seuil de fragmentation b), un fragment apparaît alors sur le spectre de masse. Le processus peut être suivi par spectrométrie de masse c).

### 2.3 ACQUISITION D'UN SPECTRE IRMPD

La photofragmentation par IRMPD vient de l'absorption de photons résonants avec un mode de vibration de l'ion au nombre d'onde porté par les photons. Cette absorption est suivie par la redistribution au sein de l'ion de l'énergie gagnée. Il est possible de faire varier le nombre d'onde d'un laser afin de mesurer le rendement de photofragmentation en fonction de la longueur d'onde du laser. En suivant l'évolution de ce rendement dont le calcul dérive de la loi de Beer-Lambert (voir équation 1) avec  $I_{prec}$  = intensité de l'ion précurseur et  $\Sigma I_{frag}$  = somme de l'intensité des ions fragments en fonction du nombre d'onde, il est possible d'obtenir le spectre IRMPD de l'espèce d'intérêt.

<sup>†</sup> Dans le cas d'un photon au nombre d'onde  $3500 \text{ cm}^{-1}$

$$\text{rendement de photofragmentation} = -\ln \frac{I_{prec}}{I_{prec} + \Sigma I_{frag}} \quad (1)$$

Comme nous venons de le voir, les spectres de fragmentation obtenus par irradiation IR peuvent porter le nom de spectre IRMPD. Dans la suite de ce manuscrit, l'utilisation du terme « spectre IRMPD » fera référence à la mesure d'un spectre IR en phase gazeuse représentant le rendement de photofragmentation en fonction du nombre d'onde comme nous allons le présenter maintenant.

La Figure 3d) présente un exemple de spectre IRMPD. Le rendement de photofragmentation varie selon la longueur d'onde. La longueur d'onde pour laquelle le rendement de photofragmentation est nul : par exemple la zone rouge ne présente aucune photofragmentation sur le spectre de masse correspondant (Figure 3b). A l'inverse, lorsque le rendement de photofragmentation est non nul, des pics se dessinent sur le spectre IRMPD, des fragments sont présents sur les spectres de masse correspondants (Figure 3a) pour le pic vert et Figure 3c) pour le pic bleu). Les spectres de photofragmentation montrent que les mêmes fragments sont observés quelques soit la longueur d'onde résonante (la dissociation se fait au niveau de la liaison la plus faible, qui ne change pas quelques soit le nombre d'onde des photons incidents).

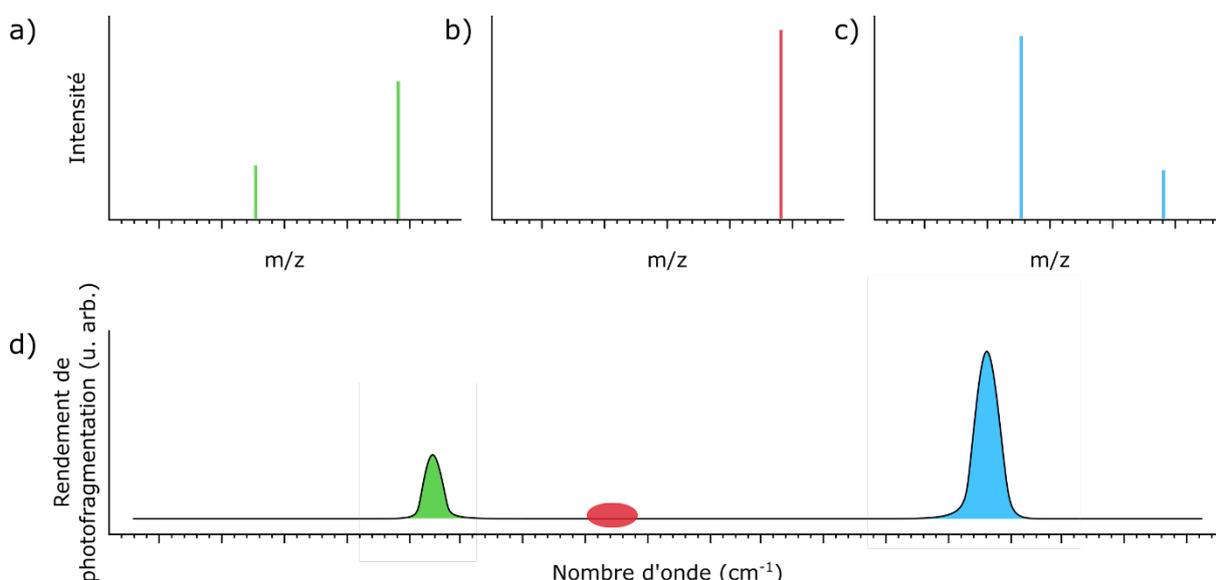


Figure 3 : Spectres de masse obtenus après irradiation d'un ion à des nombres d'onde différents bandes résonantes. Le rendement de photofragmentation n'est pas identique à ces deux nombres d'onde a) et c), bande non résonante b) La mesure du rendement de photofragmentation à différents nombres d'onde permet de tracer le spectre IRMPD correspondant d).

## 2.4 GAMME DE NOMBRES D'ONDE D'INTERET

En spectroscopie IR, les modes de vibration des groupements fonctionnels absorbent dans des gammes de nombres d'onde connues<sup>22</sup>. La gamme de nombres d'onde d'intérêt est donc choisie en fonction de la nature des groupements chimiques des composés. Les sucres sont des molécules organiques comportant de nombreux groupements hydroxyle, ils peuvent également porter des groupements amine ou diverses modifications telles que des sulfatations, phosphorylations ou carboxylations. Il existe donc deux gammes de nombres d'onde particulièrement intéressantes pour les étudier en spectroscopie IR en phase gazeuse : la gamme autour de 900-1800 cm<sup>-1</sup> qui est accessible par les lasers à électrons libres et étudiée par les équipes de FELIX (Free-Electron Lasers for Infrared eXperiments)<sup>23</sup>, celles de FHI FEL (Free-Electrons Laser at the Fritz Haber Institute)<sup>24</sup> ainsi qu'au Centre

Laser Infrarouge d'Orsay<sup>25</sup>. La gamme autour de  $3000\text{ cm}^{-1}$  est également intéressante<sup>26-28</sup>. Ces deux gammes de nombres d'onde présentent les bandes d'absorption de groupements fonctionnels présents dans les sucres. La gamme autour de  $900-1800\text{ cm}^{-1}$  contient des informations sur les modes d'élongation des liaisons C-O, C-C, N-H mais également des liaisons C=O, voire des groupements P-O/P=O et S-O/S=O (Figure 4). La gamme autour de  $3000\text{ cm}^{-1}$  contient les informations des modes d'élongation des liaisons C-H, N-H et O-H (Figure 4). La gamme autour de  $3\text{ }\mu\text{m}$  est inutilisable en phase condensée en présence d'eau ou d'un solvant polaire. Récemment la gamme de nombres d'onde dans le lointain IR ( $300-1000\text{ cm}^{-1}$ ) a été étudiée afin de distinguer des isomères de sucres de différentes tailles<sup>29</sup>. Dans cette gamme de nombres d'onde l'absorption se fait au niveau des modes de vibration des liaisons entre un carbone et un hétéroatome ou au niveau des modes de vibration des groupements sulfate ou phosphate.

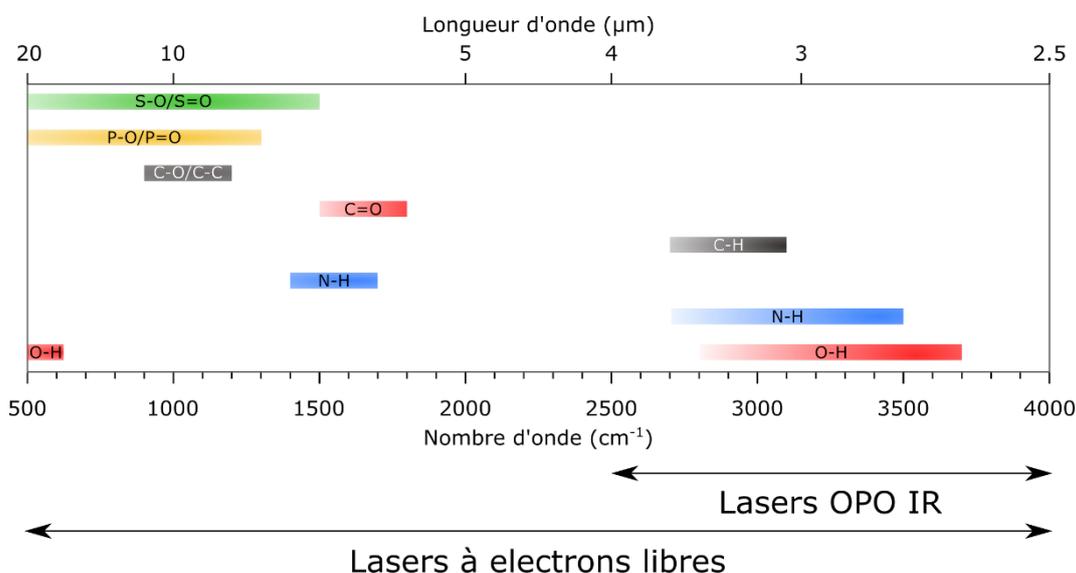


Figure 4 : Gammes des nombres d'onde intéressants, deux zones se distinguent l'une autour de  $1000\text{ cm}^{-1}$ , l'autre autour de  $3000\text{ cm}^{-1}$ .

### 3 INSTRUMENTATION

#### 3.1 DIFFÉRENTES SOURCES LASERS

Comme dit précédemment, la spectroscopie IRMPD a bénéficié des développements de sources lasers accordables capables de délivrer suffisamment d'énergie. Les premiers lasers disponibles : lasers  $\text{CO}_2$  avec une faible accordabilité ont été remplacés par des lasers couvrant des gammes de nombres d'onde bien plus larges avec suffisamment de puissance pour mener à bien les mesures. Les LEL (lasers à électrons libres) sont de grands instruments répondant à ces deux critères. Le laser à électrons libres de FELIX est accordable entre  $66$  et  $3600\text{ cm}^{-1}$ , celui de CLIO couvre la gamme de nombres d'onde allant de  $66$  à  $2000\text{ cm}^{-1}$  tandis que le laser du FHI est accordable de  $166$  à  $3300\text{ cm}^{-1}$ .

Des lasers plus compacts : les OPO/OPA (oscillateur paramétriques optiques/amplificateur paramétrique optique dont une représentation est visible représenté en Figure 5)) permettent l'accès à différentes gammes de longueurs d'onde (de l'UV jusqu'à l'IR). Dans notre cas ce type de laser nous permet de sonder la gamme  $2500-4000\text{ cm}^{-1}$ . L'appareil dont nous disposons dans l'équipe est pompé

par un laser pulsé Nd:YAG (Neodyme Doped Yttrium Aluminium Garnet) à 1064 nm, la durée de l'impulsion est d'environ 8 à 10 ns, la cadence du Nd:YAG est de 10 Hz. Le faisceau IR provenant du YAG est introduit dans l'OPO, dans lequel il est scindé en deux. La première partie du faisceau passe dans un doubleur de fréquence (un cristal de KDP : MonoPotassium Phosphoryl) : deux photons incidents (dont la longueur d'onde est de 1064 nm) forment un photon à la longueur d'onde 532 nm. Il passe ensuite dans une première cavité : l'OPO comportant deux cristaux de KTP (Potassium Titanyl Phosphoryl). Ces cristaux divisent ce faisceau en deux faisceaux de fréquences complémentaires. Le premier est accordable entre 710 et 880  $\text{cm}^{-1}$  selon l'orientation des cristaux de KDP (et donc l'angle d'incidence du faisceau). C'est le faisceau complémentaire qui est injecté dans la deuxième cavité : l'OPA avec la deuxième partie du faisceau provenant du Nd:YAG (1064 nm). L'OPA est constituée de 4 cristaux de KTA (Potassium Titanyl Arseniate) produisant un laser accordable dans la gamme des 3  $\mu\text{m}$ . Dans nos expériences, les lasers OPO nous permettent de couvrir la gamme de nombres d'onde comprise entre 2700 et 3700  $\text{cm}^{-1}$ .

### 3.2 SPECTROMETRES DE MASSE UTILISES

Les premières mesures d'IRMPD ont eu lieu sur des FT-ICR<sup>18</sup>. Dans un FT-ICR, les ions sont piégés et peuvent être manipulés à l'aide de champs électrique et magnétique. Ce piégeage permet donc une irradiation optimale des ions afin de leur apporter suffisamment d'énergie pour les fragmenter et mesurer les spectres de masse obtenus.

Depuis les premières mesures d'IRMPD, les pièges à ions ont également pu être utilisés. Bien qu'ils présentent un moins bon pouvoir de résolution que les FT-ICR, ils sont plus rapides pour acquérir un spectre de masse, nécessitent moins d'hélium et sont plus abordables. Les ions sont thermalisés par collision avec de l'hélium injecté dans le piège afin d'améliorer le piégeage. Dans le cas de l'utilisation de piège à ions, la présence d'hélium va jouer un rôle dans le processus de photofragmentation. En cas de collision entre l'ion et un atome de gaz, des échanges d'énergie ont lieu entre l'ion et le gaz, l'ion peut se relaxer et céder son énergie au gaz de collision ce qui est délétère pour le phénomène de photofragmentation.

### 3.3 MONTAGE EXPERIMENTAL MIS EN PLACE A L'ILM

Pour permettre des analyses en IRMPD au sein de notre équipe, un montage a été développé. Ce montage « historique » (présenté en Figure 5) est un couplage entre un OPO/OPA doté d'un laser de pompe YAG (LaserVision) et un piège à ions de type LCQ de chez ThermoFinnigan. Le piège à ions permet d'obtenir des spectres de masse, des spectres de fragmentation ( $MS^2$ ) et même des spectres de masse des fragments de fragments au niveau n ( $MS^n$ ). Nous verrons plus loin l'intérêt de disposer de pièges à ions avec de telles capacités pour nos expériences.

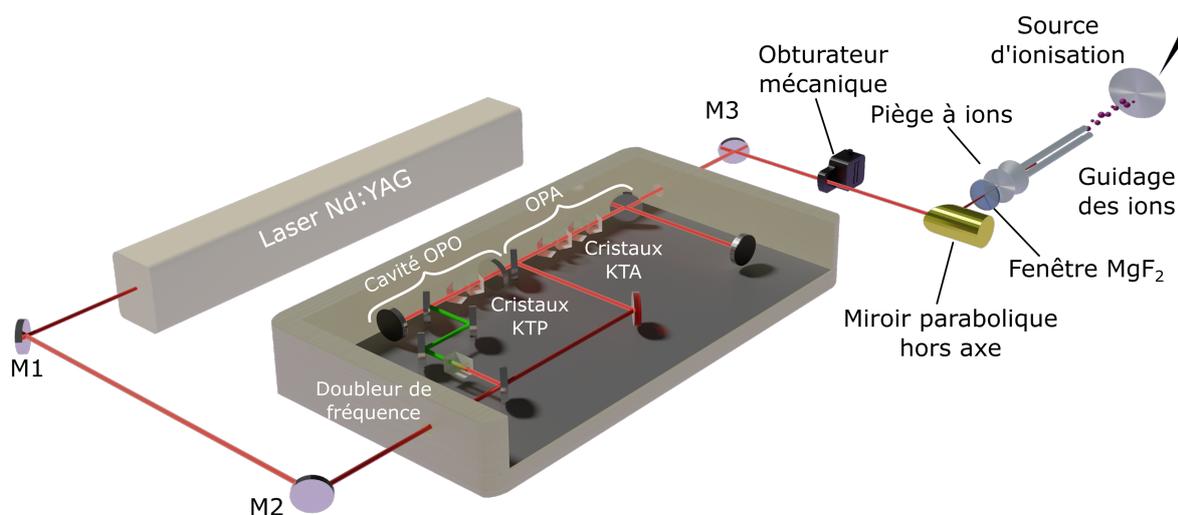


Figure 5 : Montage expérimental « historique » dans l'équipe. Un OPO/OPA est pompé par un laser Nd:YAG et délivre un laser accordable dans la gamme  $2700\text{-}3700\text{ cm}^{-1}$ , le faisceau IR est ensuite introduit dans le piège à ions au moyen d'une ouverture pratiquée dans le spectromètre de masse. Les ions peuvent alors être irradiés. M1, M2, M3 sont des miroirs argent, le miroir parabolique hors axe focalise le faisceau IR dans le piège à ions (dans un souci de simplicité tous les miroirs n'ont pas été numérotés).

Comme expliqué précédemment, l'OPO/OPA est pompé par un laser Nd:YAG à  $1064\text{ nm}$  avec une cadence de  $10\text{ Hz}$ . La puissance en sortie du YAG est d'environ  $7\text{ W}$ . L'OPO/OPA scanne la gamme de nombre d'onde d'intérêt ( $2700\text{-}3700\text{ cm}^{-1}$ ). Les lasers de type OPO/OPA permettent d'obtenir une bonne résolution (quelques  $\text{cm}^{-1}$ ). En sortie d'OPA, la puissance mesurée est de l'ordre de  $100\text{ mW}$ . Du fait de la cadence de ce type de laser il faut piéger les ions suffisamment longtemps typiquement  $800\text{ ms}$  pour que la durée d'irradiation assure une photofragmentation suffisante. Le spectromètre de masse est un piège à ions de type LCQ équipé d'une source d'ionisation par électrospray et d'un piège 3D. Une ouverture a été pratiquée dans le spectromètre de masse et une fenêtre en  $\text{MgF}_2$  transparente au faisceau IR est installée pour permettre l'injection du faisceau laser dans le piège. Les ions formés par électrospray sont ensuite guidés jusqu'au piège. Les ions d'intérêt sont alors isolés avant d'être irradiés. Un obturateur mécanique (SH05 Thorlabs) contrôlé par un signal provenant du spectromètre de masse assure la synchronisation de l'activation des ions au sein du spectromètre de masse et de leur irradiation. Le pointé du laser est stable selon les différents nombres d'onde émis par le laser.

A des fins d'illustration, un spectre expérimental est présenté en Figure 6. Dans la suite de ce manuscrit (sauf mention contraire), les spectres IRMPD seront tous représentés par des points. Chaque point correspond à la moyenne de 3 mesures et une courbe (moyenne glissante sur 5 points consécutifs) est ajoutée afin de faciliter la lecture du spectre. Les bandes d'absorption présentes sur ce spectre sont caractéristiques des modes de vibration des liaisons O-H ( $3666, 3628\text{ et }3545\text{ cm}^{-1}$ ) C-H ( $< 3000\text{ cm}^{-1}$ ) et la bande à  $3545\text{ cm}^{-1}$  est également représentative de la présence d'un groupement sulfate dans l'ion. Les bandes à  $3190, 3247\text{ et }3328\text{ cm}^{-1}$  sont caractéristiques du groupement amine<sup>30</sup>.

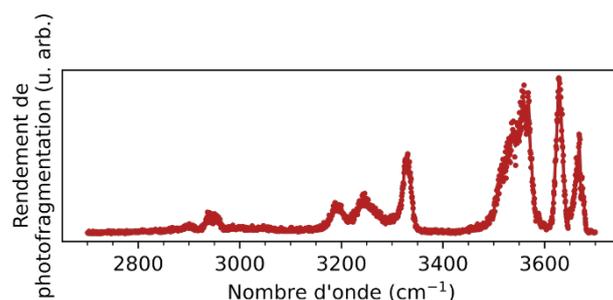


Figure 6 : Exemple de spectre IRMPD acquis dans l'équipe : spectre de la glucosamine-3 sulfate protonée dans la gamme des 3  $\mu\text{m}$ . Le spectre a été acquis en infusion directe dans le spectromètre de masse au moyen du montage avec l'OPO/OPA.

## 4 L'IRMPD ET LES QUESTIONS DE PHYSICO-CHIMIE : IDENTIFICATION DE FRAGMENTS D'OLIGOSACCHARIDES FUCOSYLES

Après avoir utilisé le processus IRMPD à des fins de fragmentation, la spectroscopie IRMPD a été développée et utilisée pour répondre à des questions physico-chimiques principalement pour la caractérisation de la structure de composés complexes<sup>31,32</sup>. Les mesures expérimentales sont comparées à des calculs théoriques afin de déterminer les conformations que le composé adopte. Pour illustrer cette approche, l'équipe a étudié deux composés isobariques<sup>33</sup> : la glucosamine-6 phosphate et la glucosamine-6 sulfate. Ces deux composés sont retrouvés dans les organismes vivants. Ils présentent deux empreintes spectroscopiques bien différentes ce qui permet de discriminer les composés sans avoir besoin de passer sur des appareils de spectrométrie de masse à haute résolution. Plus intéressant, la comparaison de l'empreinte MS/IR mesurée avec les calculs théoriques montrent la correspondance entre spectres IRMPD et conformation du composé. D'autres expérimentations permettent l'étude structurale complexes supramoléculaires en phase gazeuse, que ce soit des dimères<sup>34</sup> ou des composés bien plus gros tels que des peptides amyloides<sup>35</sup>. L'IRMPD permet aussi le suivi de réactivité comme dans l'identification d'intermédiaires de réactions<sup>36,37</sup>, ou encore l'identification de fragment issus de la CID<sup>38</sup> ou de l'ETD<sup>39</sup>.

Alternativement, les spectres IRMPD ne sont pas comparés à des calculs de théories, mais à d'autres spectres de composés connus. Ainsi, ce type de comparaison a été utilisé pour l'identification de produits de migration du fucose<sup>40</sup> comme nous allons le voir dans cette partie.

### 4.1 LA MIGRATION DU FUCOSE

Cette partie décrit l'article "**Fucose Migration Pathways Identified Using Infrared Spectroscopy**" paru dans la revue *Angewandte Chemie* par Baptiste Moge, Baptiste Schindler, Ozgur Yeni et Isabelle Compagnon en 2023.

La fucosylation désigne l'attachement d'un résidu fucose (voir structure en Figure 7) sur un oligosaccharide. La fucosylation est une des modifications communes sur les glycoprotéines ou les glycolipides. La fucosylation a différents rôles<sup>41</sup> notamment dans les interactions intramoléculaires, elle aide également au repliement de la protéine<sup>42,43</sup>. Elle permet également les interactions intermoléculaires ce qui a différentes conséquences : développement tissulaire<sup>42,44</sup>, régulation du système immunitaire<sup>45</sup>, rôle dans les cancers<sup>43,46-48</sup>. Le fucose est également retrouvé sur des oligosaccharides libres tels que les HMOs (human milk oligosaccharide). Il s'agit de sucres présents

dans le lait maternel qui participent au bon développement du nourrisson<sup>49</sup>. Au vu de l'importance de ces composés, la connaissance de leur structure est primordiale.

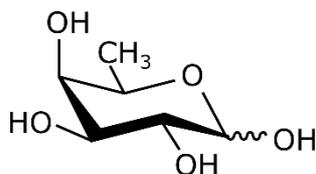


Figure 7 : structure du fucose

Toutefois, les analyses structurales par spectrométrie de masse souffrent d'un problème : la migration du fucose. Ce phénomène autrefois appelé « internal residue loss » est connu depuis maintenant une trentaine d'années<sup>50-53</sup>. Il s'agit d'un processus observé lors de l'analyse d'oligosaccharides contenant du fucose, mais qui a aussi été mis en évidence pour d'autres desoxyhexoses (comme le rhamnose)<sup>51</sup> ainsi que des pentoses comme le xylose<sup>51,54</sup>. Ce phénomène complique l'analyse de tels composés puisqu'il conduit à la formation de fragments résultant d'une modification de la structure du composé initial. La migration a lieu dans le cas d'espèces protonées<sup>55-58</sup> ou ammoniées<sup>53,58</sup> la présence d'un proton mobile semble donc être un prérequis. La migration est indépendante de l'anométrie de la liaison ou de la régiochimie de la liaison entre le fucose et le résidu qui le porte<sup>52</sup>. Le fucose migre vers des résidus spatialement proches : ils peuvent être à proximité sur la séquence de l'oligosaccharide, mais aussi être plus loin, puisque la fragmentation par CID peut rapprocher spatialement des résidus loin l'un de l'autre sur la chaîne d'oligosaccharide<sup>59</sup>.

Il est possible d'éviter la migration du fucose de plusieurs manières telles que : l'utilisation d'adduits métalliques en tant que porteurs de charge<sup>2,57,60</sup>, l'utilisation de groupes fonctionnels ou d'alkylammonium<sup>58</sup>, le marquage des sucres par des marqueurs fluorescents contenant des sites basiques<sup>61</sup> dont le but est de limiter la mobilité du proton ainsi que la production de radicaux libres<sup>62</sup>.

Malgré toutes les connaissances acquises sur ce phénomène, un mécanisme consensuel n'a pas encore été accepté, seules des propositions ont été faites<sup>55,56,63</sup>. De même l'identification du produit de migration n'a pas encore été faite, seul le résidu vers lequel le fucose migre est connu. L'émergence de nouvelles méthodes couplées à la spectrométrie de masse : mobilité ionique et spectroscopie IR offre un regard nouveau sur la migration du fucose. Ainsi la spectroscopie IR en phase gazeuse<sup>64</sup> et la mobilité ionique<sup>65</sup> ont mis en évidence la migration du fucose sur des composés considérés comme intacts. Les composés n'ont pas subi d'étape de fragmentation mais présentent une signature IR ou un temps d'arrivée similaire à des standards « mimant » la migration du fucose.

Dans cet article nous avons montré à l'aide de la spectroscopie IR en phase gazeuse comment le produit de réaction pouvait être identifié par comparaison avec des standards de référence.

## 4.2 MATERIEL ET METHODE

Le montage expérimental est décrit précédemment (voir Figure 5). Le laser OPO/OPA est couplé à un spectromètre de masse LCQ de chez ThermoFinnigan. Dans cette étude, quatre trisaccharides de même composition comportant un Gal (galactose), un Fuc (fucose) et un GlcNAc (N-Acétyleglucosamine) ne différant que selon la régiochimie de leurs liaisons sont considérés (voir Figure 8). Il s'agit des antigènes de groupe sanguin H de type 1 et 2 : Blood group antigen H type 1 (BGHI) et Blood group antigen type 2 (BGHII) ainsi que des espèces de Lewis : Lewis A et Lewis X. Les antigènes de groupes sanguins ont été choisis pour étudier la migration du fucose. Les espèces de

Lewis quant à elles doivent servir de témoins, *a priori*, on ne s'attend pas à observer de migration pour ces deux espèces. Les trois standards disaccharidiques Fuca1-xGlcNAc (avec x = 3, 4 ou 6) servent de références pour comparaison avec les produits de réaction. Les composés proviennent tous de chez Carbosynth, ils sont dissous à 1 mg.mL<sup>-1</sup> dans de l'eau sans purification supplémentaire avant dilution à 20 μM dans un mélange eau/MeOH (50/50 v/v) avec 0.1 % d'acide acétique. Les composés sont infusés directement à 3 μL.min<sup>-1</sup>. Tout d'abord, les spectres de fragmentation par CID des espèces protonées sont acquis. Ensuite, les spectres IRMPD des fragments sont obtenus au moyen du montage décrit en Figure 5. Brièvement, le spectre IRMPD est mesuré sur la gamme 2700-3700 cm<sup>-1</sup> le laser irradie les ions avec une cadence de 10 Hz, l'énergie des impulsions varie de 10 à 15 mJ le long de la gamme spectrale. La puissance des impulsions peut changer au jour le jour, les intensités sur les spectres IRMPD sont donc recalibrées en fonction de la puissance mesurée. Les ions sont piégés pendant 700 ms. La synchronisation avec le laser est faite au moyen d'un obturateur mécanique. Les spectres IRMPD des standards disaccharidiques protonés (m/z = 368) sont mesurés dans les mêmes conditions.

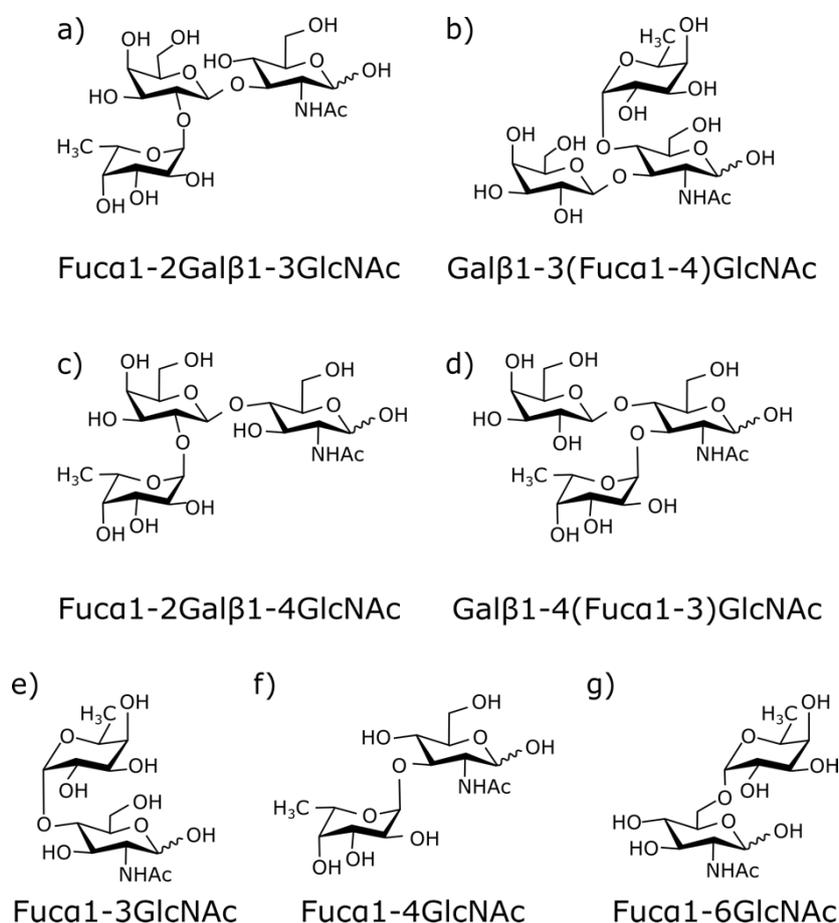


Figure 8 : Représentation des trisaccharides étudiés : BGHI a), Lewis A b) BGHII c), Lewis X d) ainsi que les disaccharides servant de références Fuca1-3GlcNAc e), Fuca1-4GlcNAc f) et Fuca1-6GlcNAc g). Les trisaccharides contiennent les trois mêmes résidus : galactose, fucose et N-Acetylglucosamine, seul l'arrangement des liaisons entre ces monosaccharides diffère.

Tout d'abord, il est important de vérifier que les trois standards disaccharidiques peuvent être différenciés par spectroscopie IR en phase gazeuse, leur empreinte est donc acquise. Pour cela, chacun d'eux est infusé séparément dans le piège à ions. Les précurseurs protonés (m/z = 368) sont alors isolés et irradiés pendant 800 ms avant détection. Le laser balaie alors les nombres d'onde de 3700 à 2700 cm<sup>-1</sup> afin d'obtenir le spectre complet. Les spectres résultant de cette manipulation sont

présentés en Figure 9. Ces trois spectres présentent des signaux dans la gamme des C-H : deux pics aux alentours de 2900-2950  $\text{cm}^{-1}$ , ainsi qu'une bande bien résolue à 3000  $\text{cm}^{-1}$  caractéristique des disaccharides contenant du fucose. Les spectres sont aussi caractéristiques des N-Acétylexosamines avec la bande intense autour de 3400  $\text{cm}^{-1}$  dans la gamme des modes de vibration des liaisons N-H pour finir, ils partagent également la signature des modes de vibration des liaisons O-H autour de 3600  $\text{cm}^{-1}$ . Le composé  $\text{Fu}\alpha 1\text{-6GlcNAc}$  (Figure 9c) se distingue des autres structures principalement par sa bande d'absorption à 3400  $\text{cm}^{-1}$  qui est très large (caractéristique des disaccharides de type X-6GlcNAc<sup>27,66-68</sup> où X est un monosaccharide ou un groupement chimique). De plus, le  $\text{Fu}\alpha 1\text{-6GlcNAc}$  présente une bande d'absorption des O-H à 3590  $\text{cm}^{-1}$  très intense, cette bande est moins intense dans les autres spectres. La bande d'absorption à 3400  $\text{cm}^{-1}$  est très fine pour les deux autres structures (Figure 9a et b) qui peuvent également être distinguées l'une de l'autre. Tout d'abord, l'absorption de cette bande est plus intense pour le  $\text{Fu}\alpha 1\text{-4GlcNAc}$  (Figure 9b) que pour le  $\text{Fu}\alpha 1\text{-3GlcNAc}$  (Figure 9a). Ensuite, la bande de d'absorption des O-H à 3540  $\text{cm}^{-1}$  présente sur le spectre du  $\text{Fu}\alpha 1\text{-3GlcNAc}$  (Figure 9a) est absente du spectre du  $\text{Fu}\alpha 1\text{-4GlcNAc}$  (Figure 9b). De plus, dans la gamme des modes de vibration des liaisons C-H le spectre de  $\text{Fu}\alpha 1\text{-3GlcNAc}$  présente des bandes d'absorption très intenses à 2900 et 3000  $\text{cm}^{-1}$  qui sont bien moins intenses dans le cas du  $\text{Fu}\alpha 1\text{-4GlcNAc}$ .

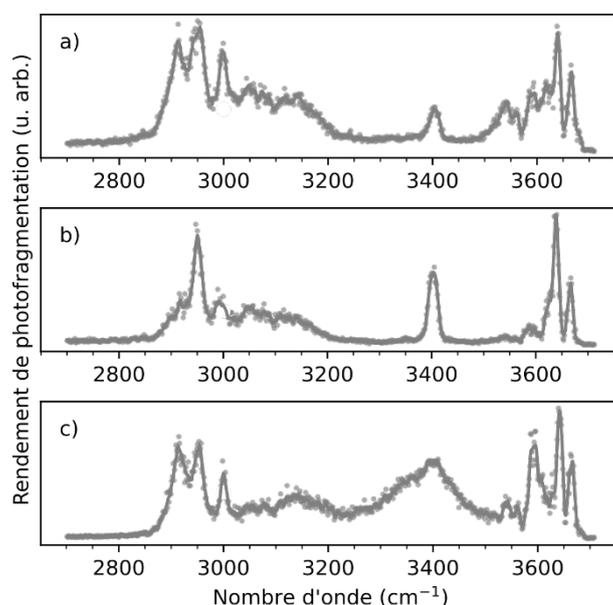


Figure 9 : Spectres IRMPD des différents standards protonés des disaccharides considérés :  $\text{Fu}\alpha 1\text{-3GlcNAc}$  a),  $\text{Fu}\alpha 1\text{-4GlcNAc}$  b) et  $\text{Fu}\alpha 1\text{-6GlcNAc}$  c). Les trois disaccharides sont étudiés sous leur forme protonée ( $m/z = 368$ ). Ces trois spectres présentent des signatures bien différentes, la spectroscopie IRMPD permet donc de discriminer ces standards qui ne varient que de la régiochimie de la liaison glycosidique.

La spectroscopie IRMPD permet donc de discriminer ces trois disaccharides ne différant que par la régiochimie de la liaison. L'utilisation de standards pour l'étude d'oligosaccharides fucosylés par spectroscopie IR en phase gazeuse est donc possible, ce qui n'était pas trivial quand on sait que le fucose est capable de migrer même sans que l'oligosaccharide ne subisse de fragmentation. Ainsi, Mucha *et al.* n'ont pas pu se servir de standards pour l'étude de la fragmentation de Ley car les standards subissaient une migration sous leur forme « intacte »<sup>64</sup>. Le phénomène a également été observé en mobilité ionique par Toraño *et al.*<sup>65</sup>.

### 4.3 MIGRATION INTEROSIDIQUE

Dans cette partie, nous avons étudié les antigènes de groupe sanguin H de type 1 et 2. Ces deux espèces sont de la forme Fuc-Gal-GlcNAc et ne varient que par la régiochimie de la liaison entre le galactose et le N-Acétyleglucosamine (voir Figure 8).

Le spectre de fragmentation de ces espèces (Figure 10a et c) présente les mêmes rapports  $m/z$  pour les ions fragments avec des intensités relatives qui varient. Tout d'abord, le fragment à  $m/z = 512$  provient de la perte de l'hydroxyle en position anomérique ( $-H_2O$ ), le fragment à  $m/z = 384$  vient de la perte du fucose, celui à  $m/z = 368$  de la perte du galactose (le fucose a donc migré du Gal vers le GlcNAc) enfin le fragment à  $m/z = 222$  correspond au N-Acétyleglucosamine, les deux autres monosaccharides ont donc été éliminés. Les fragments à  $m/z = 366$ , 350 et 204 correspondent respectivement à ces mêmes fragments mais avec une perte d'eau supplémentaire.

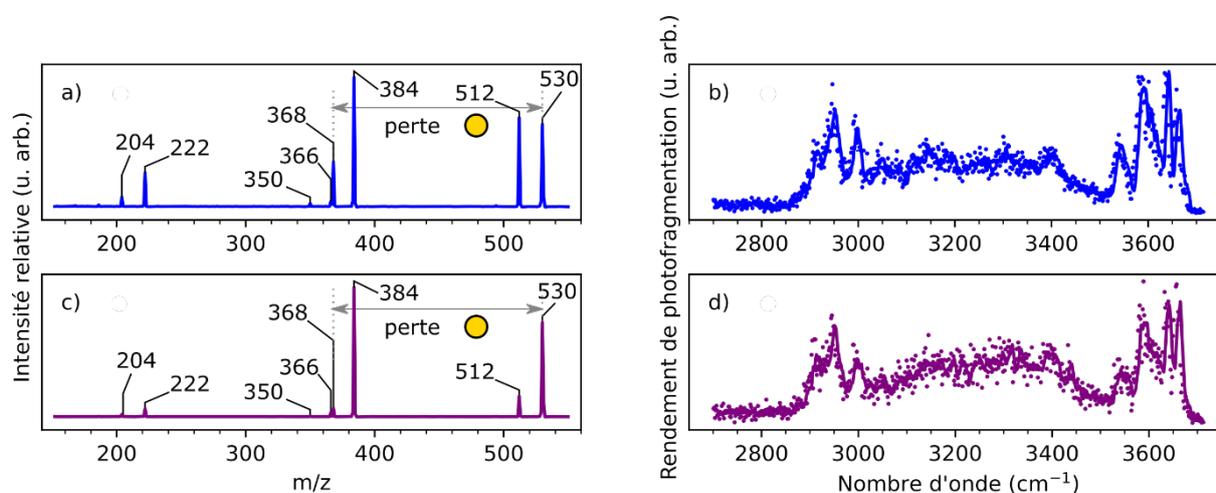


Figure 10 : Spectres de fragmentation des trisaccharides protonés des espèces de groupe sanguin BGHI a), BGHII c) et les spectres IRMPD du fragment  $m/z = 368$  résultant de la perte du galactose et donc migration du fucose BGHI c), BGHII d). Les spectres de masse présentent les mêmes fragments qui résultent de la perte d'eau, ou de la rupture de liaisons glycosidiques. Ces fragments présentent des intensité différentes. Les spectres IRMPD présentés dans cette figure présentent une signature similaire. Le cercle jaune représente un galactose selon la SNFG<sup>69</sup>.

Le spectre IRMPD de ces espèces (Figure 10b et d) sont similaires. Ils présentent tous deux trois pics dans la région des C-H (2900, 2950 et 3000  $cm^{-1}$ ), suivi d'une large bande d'absorption de 3100 à 3500  $cm^{-1}$  ce qui suggère une forte signature de Fuc $\alpha$ 1-6GlcNAc, cette identification est renforcée par la présence d'une bande d'absorption intense à 3590  $cm^{-1}$ . On note aussi la présence de la légère bande d'absorption à 3540  $cm^{-1}$  présente sur le Fuc $\alpha$ 1-3GlcNAc et le Fuc $\alpha$ 1-6GlcNAc ainsi que de deux bandes d'absorption à 3640 et 3670  $cm^{-1}$ . Seule la diminution d'absorption à 3250  $cm^{-1}$  change entre ces spectres et le spectre du Fuc $\alpha$ 1-6GlcNAc.

Les signatures IRMPD des deux fragments de trisaccharides ont été obtenues et comparées avec les signatures IRMPD des standards (Figure 12b et d). Une forte ressemblance est observée avec le disaccharides standard Fuc $\alpha$ 1-6GlcNAc. Le fucose migre donc vers l'hydroxyle n°6 du N-Acétyleglucosamine après fragmentation. Des différences entre les spectres du standard et les spectres des fragments (notamment la forme du spectre aux alentours de 3400  $cm^{-1}$ ) seront discutées dans la suite de ce chapitre.

#### 4.4 MIGRATION INTRAOSIDIQUE

Dans cet article, nous avons également examiné les espèces de Lewis. Dans le cas de ces deux espèces, le fucose et le galactose sont tous deux liés au N-Acetylglucosamine. Puisque le fucose est directement lié au GlcNAc, la perte du galactose ne requiert pas forcément migration du fucose et donc réarrangement. C'est pourquoi, il n'a jamais été considéré comme sujet à la migration du fucose.

L'examen des spectres de fragmentation (Figure 11a et c) montre des signaux à des  $m/z$  identiques. Le précurseur peut perdre une molécule d'eau ( $m/z = 512$ ), un résidu fucose ( $m/z = 384$ ) un galactose ( $m/z = 368$ ) ou encore ces deux monosaccharides ( $m/z = 222$ ). Là encore, les fragments correspondent à la perte d'eau et du résidu fucose ( $m/z = 366$ ), perte d'eau et du résidu galactose ( $m/z = 350$ ) ou des deux monosaccharides et d'une molécule d'eau ( $m/z = 204$ ). Seules les intensités varient entre ces deux spectres.

Comme dit plus haut, le fucose étant lié au GlcNAc, la perte du galactose n'implique pas forcément de migration du résidu fucose.

L'analyse des spectres IRMPD des ions  $m/z = 368$  provenant de Lea (Figure 11b) et Lex (Figure 11d) montre deux spectres similaires. Cette observation est étonnante. Tout d'abord, trois signaux sont présents dans la région des C-H ( $2900$ ,  $2950$  et  $3000\text{ cm}^{-1}$ ), dans la gamme des N-H, un gros massif s'étend de  $3100$  à  $3500\text{ cm}^{-1}$  (cette fois-ci avec diminution d'intensité aux alentours de  $3250\text{ cm}^{-1}$ ) avant d'observer une bande d'absorption à  $3540\text{ cm}^{-1}$ , une plus intense à  $3590\text{ cm}^{-1}$  et deux autres bandes d'absorption à  $3640$  et  $3670\text{ cm}^{-1}$ .

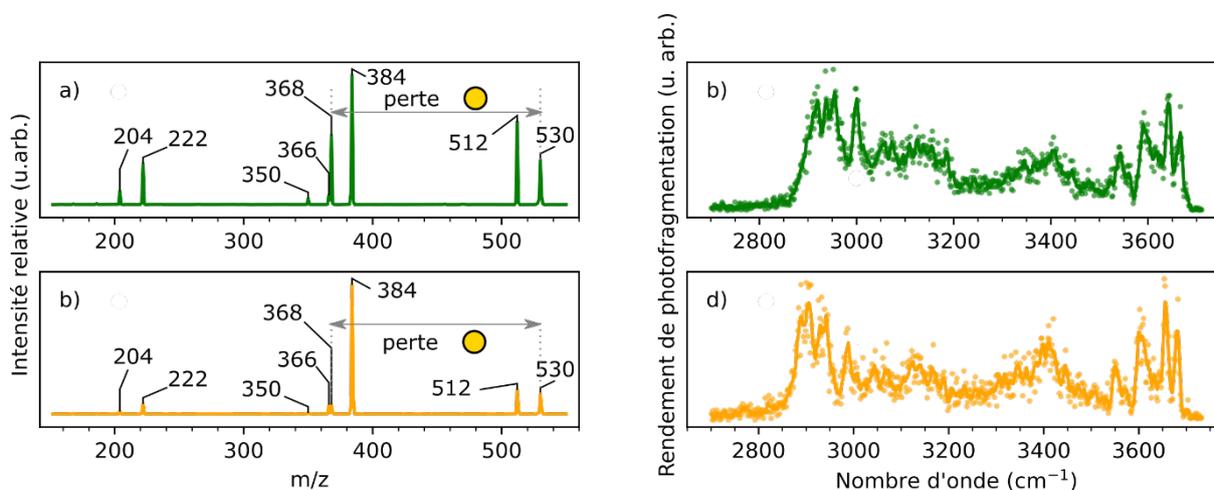


Figure 11 : Spectres de fragmentation des trisaccharides protonés d'espèces de Lewis Lea a), Lex c) et les spectres IRMPD du fragment  $m/z = 368$  (fragment protoné des trisaccharides) résultant de la perte du galactose Lea b), Lex d). Les spectres de masse présentent les mêmes fragments qui résultent de la perte d'eau, ou de la rupture de liaisons glycosidiques. Ces fragments présentent des intensité différentes. Les spectres IRMPD présentés dans cette figure présentent une signature similaire, ce qui est étonnant. Le cercle jaune représente un galactose selon la SNFG.

La fragmentation par CID ne produit pas de fragments de la forme  $\text{Fu}\alpha 1\text{-3GlcNAc}$  pour le Lex et  $\text{Fu}\alpha 1\text{-4GlcNAc}$  pour le Lea. Dans les deux cas, le spectre obtenu ressemble au spectre du  $\text{Fu}\alpha 1\text{-6GlcNAc}$  (Figure 12a et c). Il semble que le fucose soit capable de migrer sur une autre position du résidu que celle sur laquelle il est initialement.

#### 4.5 COMPARAISON DE CES DEUX TYPES DE MIGRATION

Afin d'étudier plus avant la migration du fucose, nous avons comparé les spectres IRMPD des quatre espèces étudiées (ligne et points en couleur), avec le Fuca1-6GlcNAc en fond (ligne grise). Une combinaison linéaire entre les Fuca1-6GlcNAc et le Fuca1-3GlcNAc ou le Fuca1-4GlcNAc ou encore du mélange de ces deux espèces ont également été tracées (Figure 12)

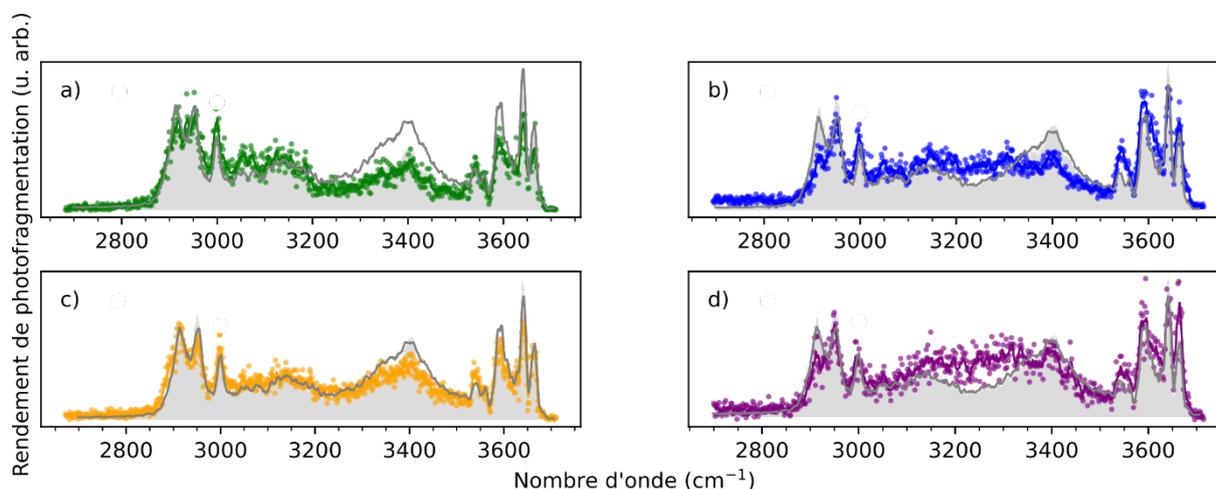


Figure 12 : Spectres IRMPD des disaccharides protonés (fragments obtenus par CID) de Lea (vert) a), BGHI (bleu) b), Lex (orange) c) et BGHII (violet) d). Spectre IRMPD du disaccharide standard Fuca1-6GlcNAc (gris). Les aires grisées correspondent à des combinaisons linéaires des standards de disaccharides : Fuca1-6GlcNAc 60 % et Fuca1-3GlcNAc 40 % a), Fuca1-6GlcNAc 80 % et Fuca1-4GlcNAc 20 % c) ou Fuca1-6GlcNAc 80 % Fuca1-3GlcNAc 10 % et Fuca1-4GlcNAc 10 % b) et d). Tous les spectres IRMPD montrés sont acquis sur les fragments protonés.

Visuellement, la meilleure correspondance entre le spectre expérimental et le spectre simulé est observée pour les espèces de Lewis (Figure 12a et c). En effet pour ces deux espèces, les combinaisons entre Fuca1-6GlcNAc (60 %) et Fuca1-3GlcNAc (40 %) (Figure 12a), ainsi que la combinaison Fuca1-6GlcNAc (80 %) et Fuca1-4GlcNAc (20 %) (Figure 12c) montrent un bon recouvrement avec les spectres expérimentaux obtenus sur les fragments de Lea et de Lex. Ce qui suggère que les deux espèces présentes dans les combinaisons linéaires sont très majoritaires dans le fragment observé expérimentalement. A l'inverse, aucune combinaison linéaire ne recouvre aussi bien le spectre expérimental des fragments d'antigènes de groupe sanguin (Figure 12b et d). Pour ces deux espèces nous avons tracé une combinaison Fuca1-6GlcNAc (80 %), Fuca1-3GlcNAc (10 %) et Fuca1-4GlcNAc (10 %). D'autres espèces doivent certainement coexister sous le pic à  $m/z = 368$ , ce qui explique que les deux spectres (expérimental et simulé) ne se recouvrent pas parfaitement.

#### 4.6 CONCLUSION ET OUVERTURE

En conclusion, dans le cas présent, les empreintes IR des disaccharides standards se sont révélées différentes. Il a été possible d'identifier les fragments majoritaires obtenus par spectroscopie IRMPD pour les quatre oligosaccharides étudiés. La spectrométrie de masse montre la migration du fucose pour les espèces d'antigènes de groupe sanguins H de type 1 et 2, car un fragment à  $m/z = 368$  et un autre à  $m/z = 350$  sont observés. Ces fragments sont clairement dus au phénomène de migration du fucose. En revanche, dans le cas des antigènes de Lewis, les fragments à  $m/z = 368$  et  $m/z = 350$  sont attendus, la migration du fucose n'est pas évidente pour ces deux espèces. La comparaison entre les spectres IRMPD des fragments à  $m/z = 368$  et les disaccharides standards permet d'identifier les fragments des antigènes de groupe sanguin H. Dans le cas des antigènes de Lewis, les empreintes IR sont déroutantes au premier abord. En effet, le fucose est initialement situé sur le N-

Acétylglucosamine. Il s'avère qu'il est quand même capable de migrer. Les empreintes IRMPD suggèrent une majorité de fragments correspondant à une migration du fucose vers l'hydroxyl n°6 du N-Acétylglucosamine. Ce type de migration « intrarésiduelle » était jusqu'alors inconnue. L'ajout d'une dimension spectroscopique à la MS a permis de répondre à la question « où le fucose migre-t-il ? » et a mis en évidence un autre mode de migration. De manière plus générale, nous nous attendons à ce que les méthodes orthogonales à la MS apportent de nouvelles réponses dans les mécanismes de fragmentation notamment dans le champ de la glycomique avec les phénomènes tels que la migration du fucose.

Plus récemment, nous avons voulu savoir quels étaient les fragments obtenus par CID du 2'fucosyllactose (Fuc $\alpha$ 1-2Gal $\beta$ 1-4Glc). Notamment quel était l'hexose observé à  $m/z = 181$ . Pour cela, l'extrémité réductrice est marquée isotopiquement ( $H_2^{18}O$ ). Le marquage s'est fait tel que décrit ci-après : Le 2'FL est dissout à  $1\text{mg}\cdot\text{mL}^{-1}$  dans de  $H_2^{18}O$ , la solution est laissée à incuber pendant deux semaines à température ambiante pour laisser le temps à l'échange isotopique de se produire. La solution est alors diluée à  $20\ \mu\text{M}$  dans une solution d'eau/MeOH 50:50 v/v avec de l'acide acétique 0.1 % avant infusion en spectrométrie de masse. Le précurseur marqué protoné visible à  $m/z = 491$  est fragmenté par CID (voir Figure 13). Les fragments observés proviennent de la rupture de liaisons glycosidiques ( $m/z = 345, 329$  et  $183$ ), ainsi que de la perte de molécule d'eau ( $m/z = 471$ ) et de ces deux phénomènes ( $m/z = 325, 309$  et  $163$ ). Dans le cas, où le fucose ne migre pas, on s'attend à observer un fragment à  $m/z = 327$ . Or, le fragment observé se trouve à  $m/z = 329$ . Le fucose a donc migré depuis le galactose jusque sur le glucose. C'est une observation étonnante, en effet, la migration du fucose n'a jamais été observée en l'absence de résidu basique.

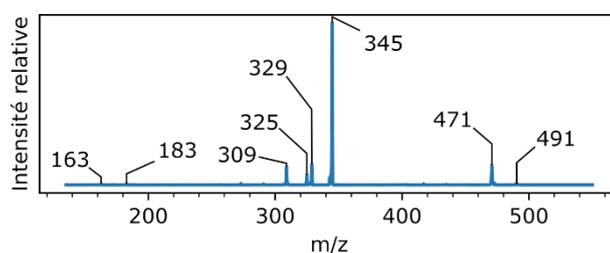


Figure 13 : Spectre de fragmentation du 2'fucosyllactose protoné marqué à  $^{18}O$  (au niveau de son extrémité réductrice). Les fragments observés proviennent de la rupture de liaisons glycosidiques ou de la perte de l'hydroxyl anomérique, à l'exception du fragment à  $m/z = 329$  qui provient du réarrangement du fucose vers le glucose.

Dans un article publié récemment<sup>70</sup>, les trisaccharides Lex et BGHII sont étudiés par différentes méthodes : spectroscopie IR en phase gazeuse, dissociation dirigée par des radicaux, mobilité ionique et approche théorique. Les auteurs démontrent que dans le cas du BGHII, le fucose est capable de migrer sur une autre position du Gal alors que pour le Lex, le fucose est capable de passer du GlcNAc jusque sur le résidu Gal.

Ces deux observations laissent la porte ouverte à de nouveaux questionnements sur la migration du fucose.

## 5 APPLICATIONS DE LA SPECTROSCOPIE IRMPD A LA CHIMIE ANALYTIQUE

Après avoir démontré l'utilité de l'IRMPD pour répondre à des questions physico-chimiques, cette partie montrera des exemples d'utilisation de la spectroscopie IRMPD appliquée à la chimie analytique.

### 5.1 APPLICATIONS ANALYTIQUES

Des équipes travaillent sur des applications de la spectroscopie IRMPD telles que l'identification de composés au moyen de la spectroscopie IRMPD. C'est le cas pour des métabolites associés à certaines pathologies<sup>71,72</sup>. Les composés chiraux peuvent également être distingués par IRMPD<sup>73-75</sup>. Polfer *et al.* ont pu mettre au point une méthode de screening sur un piège à ions pour distinguer des peptides sulfatés et phosphorylés<sup>76,77</sup> sans besoin de spectrométrie de masse à haute résolution pour être distingués.

### 5.2 IDENTIFICATION DE SUCRES A L'ILM

Comme l'a démontré l'analyse des disaccharides Fuc-GlcNAc, la spectroscopie IRMPD est très puissante dans la distinction de composés isomériques. Pour tirer parti de ces capacités, l'équipe DYNAMO à l'institut Lumière Matière utilise le montage décrit en Figure 5 dans le but d'identifier des oligosaccharides. La spectroscopie IRMPD s'est révélée particulièrement utile dans la discrimination d'isomères de sucres, que ce soient des monosaccharides tels que des épimères<sup>23,78-80</sup>, des anomères<sup>27,66</sup>, une différence dans la taille du cycle<sup>81,82</sup>, une différence de position d'un groupement fonctionnel<sup>30,68</sup>. Il est également possible de discriminer des composés de plus grande taille tels que des disaccharides<sup>40,66,83,84</sup> voire même des trisaccharides comme le 3'-sialyllactose et son isomère, le 6'-sialyllactose ou le couple 3'-sialyllactosamine et le 6'-sialyllactosamine<sup>85</sup>. Un autre exemple d'identification de trisaccharides est présenté ci-dessous avec l'exemple des trisaccharides d'antigènes de groupe sanguins.

Les oligosaccharides d'antigène de groupe sanguin portés à la surface des cellules participent aux phénomènes de reconnaissance entre cellules. Ils sont principalement composés de Gal, Fuc et GlcNAc (l'antigène de groupe sanguin A présente également du GalNAc (N-Acétalgalactosamine)).

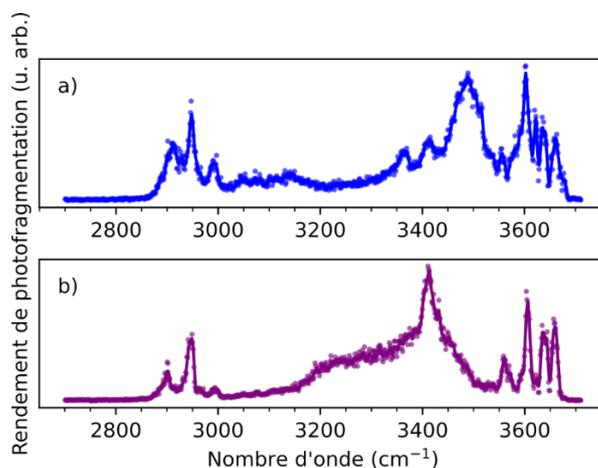


Figure 14 : Spectres IRMPD des espèces protonées  $m/z = 530$  des trisaccharides de BGHI a), BGHII b). La structure de ces espèces est présentée dans la Figure 8 a) et c). Ces spectres présentent des profils bien différents permettant de les discriminer.

La Figure 14 présente les spectres IRMPD des trisaccharides des antigènes de groupe sanguin H de type 1 et 2 (structures présentées Figure 8a et c). Seule la liaison entre Gal et GlcNAc varie entre ces deux espèces. Les spectres présentés en Figure 14 sont différents. Le BGHI présente trois pics dans la région des C-H 2910, 2950 et 2990  $\text{cm}^{-1}$ , un massif se dessine entre 3000 et 3200  $\text{cm}^{-1}$  avant d'observer une vallée puis une large bande apparaît culminant à 3486  $\text{cm}^{-1}$ , deux pics résolus sont présents sur le flanc de cette bande à 3366 et 3407  $\text{cm}^{-1}$ . Un pic est encore observé à 3555  $\text{cm}^{-1}$ . Le reste de la région des O-H est dominée par quatre pics bien résolus dont les intensités décroissent progressivement 3602, 3623, 3636 et 3661  $\text{cm}^{-1}$ . Dans la région des C-H, le BGHII présente des pics à 2906, 2950 et 2996  $\text{cm}^{-1}$  seul le pic à 2950  $\text{cm}^{-1}$  correspond, les deux autres sont légèrement décalés, et leur intensité est bien diminuée. Une large bande d'absorption est présente autour de 3200  $\text{cm}^{-1}$  et un grand pic se détache à 3412  $\text{cm}^{-1}$ . Un pic à 3555  $\text{cm}^{-1}$  est également observé dans le spectre du BGHII. Seuls trois pics sont encore observés dans la région des O-H : 3606, 3636 et 3659  $\text{cm}^{-1}$ .

L'une des limites de la spectroscopie IRMPD est que l'augmentation de la taille des composés va congestionner le spectre mesuré. De plus à des fins analytiques, la constitution de la base de données est importante, elle doit être la plus exhaustive que possible. Le problème est que lors de l'augmentation de la taille d'un oligosaccharide, la diversité des possibilités de structures devient exponentielle, on parle d'explosion combinatoire. La base de données sera nécessairement énorme pour être la plus exhaustive possible. De plus, la synthèse de tels composés est de plus en plus compliquée à mesure que la taille augmente. Pour pallier ces problèmes, une méthode de séquençage a été développée dans l'équipe.

## 5.3 METHODE DE SEQUENÇAGE

### 5.3.1 Principe du séquençage

Le séquençage permet de se passer de grandes bases de données les plus complètes possibles comprenant de nombreux standards d'oligosaccharides de grandes tailles. Il permet également de travailler sur les spectres IRMPD de petits oligosaccharides donc avec des spectres moins congestionnés. Après injection (Figure 15a), l'oligosaccharide d'intérêt est isolé (Figure 15b) avant d'être fragmenté par CID (Figure 15c). L'empreinte IRMPD de chacun de ces fragments est acquise (Figure 15d) afin de l'identifier (Figure 15e). La connaissance de chacun des fragments permet de remonter à la structure du composé d'intérêt (Figure 15f).

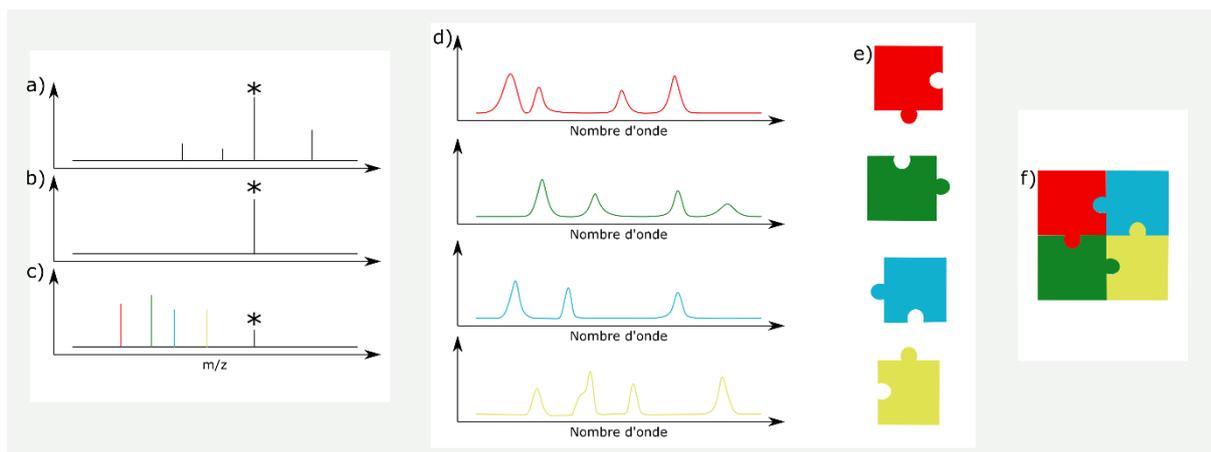


Figure 15 : Stratégie suivie pour identifier un sucre par séquençage : Le composé est infusé en spectrométrie de masse, son spectre de masse est acquis a), l'ion d'intérêt est isolé b) puis fragmenté en CID c). Chacun de ses fragments est ensuite isolé afin d'acquérir son empreinte IR d), la connaissance de la structure de ces fragments e) permet de reconstruire l'oligosaccharide de départ f).

La fragmentation par CID conserve l'information sur l'anométrie de la liaison entre les fragments<sup>27</sup> des études en mobilité ionique ont confirmé cet effet mémoire et en plus ont démontré que certaines informations concernant la régiochimie de la liaison étaient également conservées après fragmentation<sup>86</sup>. La fragmentation par CID ne modifie pas la taille du cycle<sup>87</sup>. Il est toutefois judicieux de faire attention lors de l'analyse d'oligosaccharides fucosylés en raison du phénomène de migration du fucose (comme indiqué précédemment page 30). Pour contrer ce problème il peut être intéressant de considérer des espèces sodées ou déprotonées par exemple.

### 5.3.2 Séquençage du Lacto-N-tetraose à des fins d'illustration

Le spectre de masse d'un oligosaccharide inconnu est obtenu, le signal est observé à  $m/z = 708$ . Ce spectre de masse est compatible avec un tetrasaccharide composé d'un N-Acétylexosamine et de trois hexoses ( $3 \times 162 + 203 + 18 + 1 = 708$ ). Son spectre de fragmentation par CID est acquis et est présenté en Figure 16. Le précurseur ( $m/z = 708$ ) perd une molécule d'eau : signal à  $m/z = 690$ . La perte consécutive de trois hexoses donne respectivement les signaux à  $m/z = 528$  ( $528 + 18 = 546$ ), 366 et 204. L'espèce à  $m/z = 204$  correspond à un N-Acétylexosamine, celle à  $m/z = 366$  correspond à un N-Acétylexosamine lié à un hexose.

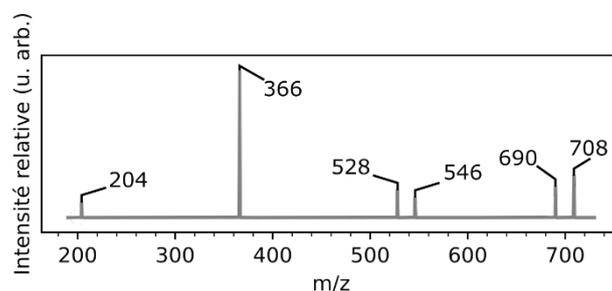


Figure 16 : Spectre de fragmentation par CID d'un oligosaccharide inconnu ( $m/z = 708$ ). Le précurseur est vraisemblablement un tetrasaccharide protoné constitué de trois hexoses et d'un N-acétylexosamine. Les fragments observés correspondent à la rupture de liaisons glycosidiques.

La fragmentation par CID permet de confirmer la nature des monosaccharides constituant le tetrasaccharide. Pour aller plus loin, l'empreinte IRMPD de chacun de ces fragments est obtenue avant d'être comparée avec l'empreinte IRMPD des standards présents dans la base de données (Figure 17). Les spectres sont normalisés par rapport à l'intensité la plus haute afin de pouvoir être comparés. Les

spectres montrent les points, puis une courbe est tracée, il s'agit d'un lissage par l'algorithme de Savitsky-Golay avec un polynôme de degré 3 et une largeur de 11 points afin de faciliter la lecture des spectres.

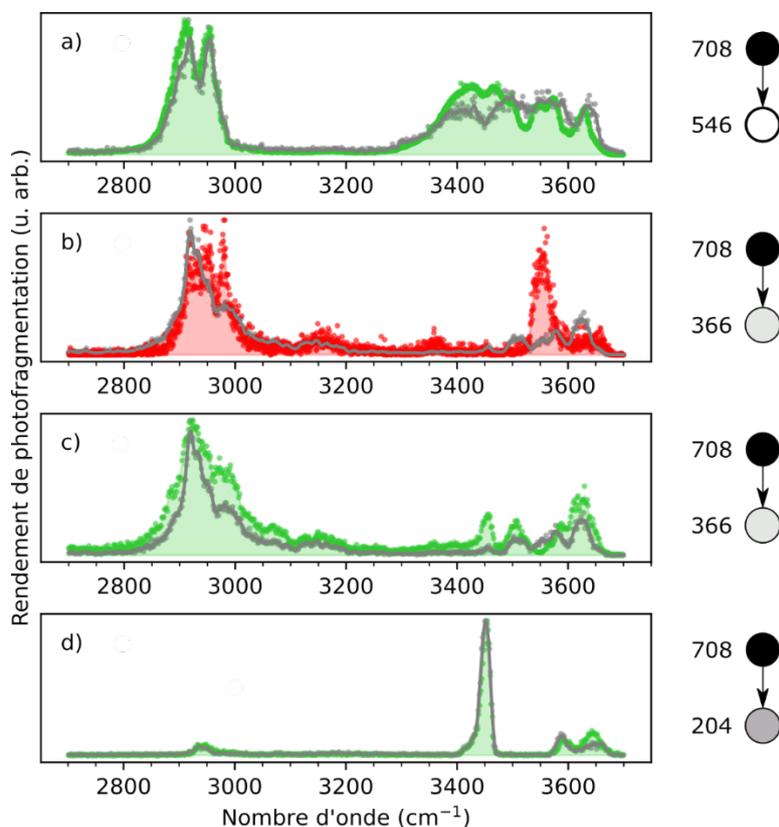


Figure 17 : L'oligosaccharide ( $m/z = 708$ ) est fragmenté et les fragments obtenus  $m/z = 546$  a),  $m/z = 366$  b) et c) et  $m/z = 204$  d) sont irradiés. Le fragment irradié est représenté sur la droite de la figure (l'ion précurseur est représenté par un rond noir quand le fragment est représenté par un rond gris avec différentes nuances). Le spectre du fragment est représenté en gris, tandis que les spectres de référence sont présentés en fond vert (si correspondance entre le spectre du fragment et le spectre de référence) et en rouge (correspondance non faite).

Le spectre IRMPD du fragment à  $m/z = 546$  (Figure 17a) présente deux pics dans la région des C-H à 2917 et 2954  $\text{cm}^{-1}$ , la deuxième partie du spectre présentant de l'absorption se situe au-delà de 3300  $\text{cm}^{-1}$  avec un massif peu résolu allant jusqu'à 3650  $\text{cm}^{-1}$ . Cette signature est reconnue comme étant caractéristique de l'oligosaccharide  $\text{GlcNAc}\beta 1\text{-}3\text{Gal}\beta 1\text{-}4\text{Glc}$ . L'empreinte IR du fragment à  $m/z = 366$  (Figure 17b) montre un maximum d'absorption dans la région des C-H à 2918  $\text{cm}^{-1}$ , suivie d'un épaulement à 2985  $\text{cm}^{-1}$ . Une faible bande d'absorption est présente à 3140  $\text{cm}^{-1}$ . Le spectre se finit dans la région des O-H avec quatre légers pics d'absorption à 3455, 3506, 3578 et 3624  $\text{cm}^{-1}$ . Pour illustrer la mauvaise identification d'un fragment, le spectre IRMPD de l'espèce  $\text{GlcNAc}\beta 1\text{-}3\text{Gal}$  est représenté en rouge. Ce spectre présente des non-concordances avec l'empreinte du composé inconnu tel qu'un maximum d'absorption dans la gamme des O-H à 3555  $\text{cm}^{-1}$  ainsi que la présence de trois maxima d'absorption dans la gamme des C-H : 2922, 2948 et 2979  $\text{cm}^{-1}$ . La Figure 17c montre la correspondance du fragment  $m/z = 366$  avec le  $\text{Gal}\beta 1\text{-}3\text{GlcNAc}$  espèce ayant perdu une molécule d'eau, ce dernier présente un pic dans la région des C-H plus large que le spectre expérimental, le signal à 3455  $\text{cm}^{-1}$  est également plus marqué avec une intensité d'absorption environ trois fois supérieure. Enfin le fragment à  $m/z = 204$  est comparé au spectre du  $\text{GlcNAc}$  (ayant perdu une molécule d'eau). Sur la Figure 17d, ces deux spectres présentent quatre pics 2942, 3453, 3587 et 3644

cm<sup>-1</sup>. La reconnaissance est ici la plus frappante de tous les fragments, seule une intensité d'absorption varie pour le pic à 3644 cm<sup>-1</sup>.

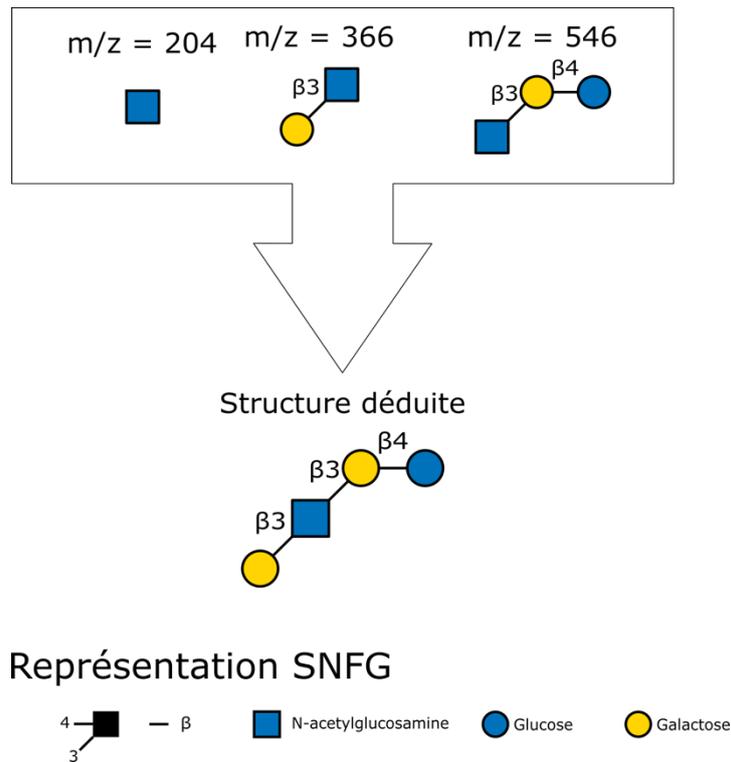


Figure 18 : Représentation selon la SNFG de l'identification des différents fragments par spectroscopie IRMPD et proposition de structure pour le tetrasaccharide inconnu.

En connaissant la nature des fragments du tetrasaccharide inconnu, il est possible de remonter à la structure vraisemblable du composé (Figure 18). Dans notre exemple, les trois fragments indiquent que le N-Acétylhexosamine est en fait du GlcNAc. Cette redondance est intéressante dans le cas où une attribution est incertaine. Ici nous identifions le composé inconnu comme étant du lacto-N-tetraose, un HMO de structure Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc. Identifier le plus de fragments possibles permet d'avoir de la redondance et donc d'apporter du poids à l'identification.

Le fait de travailler sur des fragments et non pas sur des grands oligosaccharides a deux principaux avantages. En augmentant la taille du composé, le nombre de groupements fonctionnels qui absorbe augmente également. Le nombre de bandes d'absorption va donc augmenter et, les spectres vont être davantage congestionnés, il sera difficile de les interpréter, ou d'identifier le composé inconnu. Dans le cas des sucres, le nombre d'hydroxyles va donc augmenter et avec lui, la capacité d'engager des liaisons hydrogène, ce qui a pour effet d'élargir les pics et faire perdre de la résolution sur les spectres IRMPD. Le fait de travailler sur des petits composés offre la possibilité de travailler sur des spectres peu encombrés. D'un autre côté, le fait d'avoir à identifier uniquement des fragments permet de se servir d'une base de données qui soit moins exhaustive que ce qui serait nécessaire dans le cas de l'étude des oligosaccharides entiers. En effet le nombre de combinaisons possibles pour la structure des oligosaccharides augmente de manière exponentielle avec la taille du composé. Il est donc extrêmement difficile pour ne pas dire impossible d'obtenir de telles bases de données. En fragmentant les oligosaccharides, on diminue cette complexité.

## 6 CONCLUSION

---

Initialement, la spectroscopie IRMPD était une méthode nécessitant une lourde instrumentation avec des LEL et des FT-ICR. Elle servait à répondre à des questions de physico-chimie en comparant les mesures expérimentales avec des calculs de théorie. Dans certaines applications bien particulières pour lesquelles il n'y a pas besoin d'une résolution spectrale extrême, il est possible de substituer les FT-ICR et des LEL à d'autres appareils. Ces développements autorisent aujourd'hui de faire des mesures en IRMPD avec un appareillage bien moins lourd. Un montage expérimental entre un OPO/OPA et un piège à ions a été développé dans l'équipe.

Ce montage permet toujours de répondre à des questions de physico-chimie. Dans notre cas, la spectroscopie IRMPD a d'une part identifié le produit de la migration du fucose sur quatre isomères de trisaccharides comme étant majoritairement un disaccharide de structure  $Fu\alpha 1-6GlcNAc$ . Et d'autre part il a mis en évidence un nouveau mode de migration du fucose dans lequel, le fucose ne migre pas vers un autre résidu, mais vers un autre groupement hydroxyle du même résidu. Une telle découverte pose de nouveaux questionnements et la spectroscopie IRMPD participera très certainement à leur réponse.

La spectroscopie IRMPD a été utilisée pour résoudre différentes isoméries. Elle permet aujourd'hui l'identification de composés au moyen d'une comparaison avec des bases de données ou des calculs théoriques. Elle est actuellement utilisée dans des approches de chimie analytique notamment dans l'identification de métabolites qui pourraient être des biomarqueurs de pathologies. Dans notre équipe nous l'utilisons pour répondre à des questions de physico-chimie sur les sucres, notamment dans l'identification de conformations ou dans le contexte plus appliqué de l'identification de sucres.

Cette méthode a du potentiel pour être utile à la communauté scientifique. Il s'avère que de toutes les spectroscopies d'action vibrationnelles, la spectroscopie IRMPD est la moins lourde à mettre en œuvre. Un tel couplage ne nécessite qu'un spectromètre de masse et un laser IR, nul besoin de descendre à des températures basses, de marquer ou d'emprisonner les ions avec des gaz, ni de faire des montages à plusieurs lasers. C'est la méthode qui sera la plus facile à disséminer dans les laboratoires d'analyses.

Toutefois, des progrès sont encore nécessaires, pour que la spectroscopie IRMPD puisse convaincre un public plus large et ainsi essaimer dans les laboratoires d'analyses. De nombreux développements instrumentaux ont eu lieu au cours de cette thèse pour faciliter son utilisation, l'automatiser, la rendre plus ergonomique, plus rapide, la coupler avec des workflows analytiques de référence notamment la chromatographie liquide. Ces efforts sont détaillés dans les chapitres suivants.

---

## 7 BIBLIOGRAPHIE

---

- 1 M. Z. Kamrath, E. Garand, P. A. Jordan, C. M. Leavitt, A. B. Wolk, M. J. Van Stipdonk, S. J. Miller and M. A. Johnson, Vibrational characterization of simple peptides using cryogenic infrared photodissociation of H<sub>2</sub>-tagged, mass-selected ions, *J. Am. Chem. Soc.*, 2011, **133**, 6440–6448.
- 2 E. Mucha, A. I. González Flórez, M. Marianski, D. A. Thomas, W. Hoffmann, W. B. Struwe, H. S. Hahm, S. Gewinner, W. Schöllkopf, P. H. Seeberger, G. von Helden and K. Pagel, Glycan Fingerprinting via Cold-Ion Infrared Spectroscopy, *Angew. Chem. Int. Ed.*, 2017, **56**, 11248–11251.
- 3 G. Meijer, M. S. de Vries, H. E. Hunziker and H. R. Wendt, Laser desorption jet-cooling of organic molecules, *Appl. Phys. B*, 1990, **51**, 395–403.
- 4 O. V. Boyarkin, Cold ion spectroscopy for structural identifications of biomolecules, *Int. Rev. Phys. Chem.*, 2018, **37**, 559–606.
- 5 J. P. Simons, R. A. Jockusch, P. ÇarÇabal, I. Hünig, R. T. Kroemer, N. A. Macleod and L. C. Snoek, Sugars in the gas phase. Spectroscopy, conformation, hydration, co-operativity and selectivity, *Int. Rev. Phys. Chem.*, 2005, **24**, 489–531.
- 6 C. J. Gray, I. Compagnon and S. L. Flitsch, Mass spectrometry hybridized with gas-phase InfraRed spectroscopy for glycan sequencing, *Curr. Opin. Struct. Biol.*, 2020, **62**, 121–131.
- 7 K. Greis, C. Kirschbaum, G. Von Helden and K. Pagel, Gas-phase infrared spectroscopy of glycans and glycoconjugates, *Curr. Opin. Struct. Biol.*, 2022, **72**, 194–202.
- 8 R. L. Woodin, D. S. Bomse and J. L. Beauchamp, Multiphoton dissociation of molecules with low power continuous wave infrared laser radiation, *J. Am. Chem. Soc.*, 1978, **100**, 3248–3250.
- 9 J. Laskin and J. H. Futrell, Activation of large ions in FT-ICR mass spectrometry, *Mass Spectrom. Rev.*, 2005, **24**, 135–167.
- 10 D. P. Little, J. Paul. Speir, M. W. Senko, P. B. O'Connor and F. W. McLafferty, Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing, *Anal. Chem.*, 1994, **66**, 2809–2815.
- 11 D. P. Little, D. J. Aaserud, G. A. Valaskovic and F. W. McLafferty, Sequence Information from 42–108-mer DNAs (Complete for a 50-mer) by Tandem Mass Spectrometry, *J. Am. Chem. Soc.*, 1996, **118**, 9352–9359.
- 12 Y. Bai, D. Galetskiy, E. Damoc, C. Paschen, Z. Liu, M. Griese, S. Liu and M. Przybylski, High resolution mass spectrometric alveolar proteomics: Identification of surfactant protein SP-A and SP-D modifications in proteinosis and cystic fibrosis patients, *PROTEOMICS*, 2004, **4**, 2300–2309.
- 13 J. R. Johnson, F. Meng, A. J. Forbes, B. J. Cargile and N. L. Kelleher, Fourier-transform mass spectrometry for automated fragmentation and identification of 5–20 kDa proteins in mixtures, *ELECTROPHORESIS*, 2002, **23**, 3217–3223.
- 14 C. A. Lutomski, T. J. El-Baba, J. D. Hinkle, I. Liko, J. L. Bennett, N. V. Kalmankar, A. Dolan, C. Kirschbaum, K. Greis, L. H. Urner, P. Kapoor, H.-Y. Yen, K. Pagel, C. Mullen, J. E. P. Syka and C. V. Robinson, Infrared Multiphoton Dissociation Enables Top-Down Characterization of Membrane Protein Complexes and G Protein-Coupled Receptors, *Angew. Chem. Int. Ed.*, 2023, **62**, e202305694.
- 15 D. Goldberg, M. Bern, B. Li and C. B. Lebrilla, Automatic Determination of O-Glycan Structure from Fragmentation Spectra, *J. Proteome Res.*, 2006, **5**, 1429–1434.
- 16 K. S. Lancaster, H. J. An, B. Li and C. B. Lebrilla, Interrogation of N-Linked Oligosaccharides Using Infrared Multiphoton Dissociation in FT-ICR Mass Spectrometry, *Anal. Chem.*, 2006, **78**, 4990–4997.
- 17 M. Bensimon, J. Rapin and T. Gäumann, Comparison of infrared photodissociation in a Fourier transform mass spectrometer with metastable ion decay in a double-focusing mass spectrometer, *Int. J. Mass Spectrom. Ion Process.*, 1986, **72**, 125–135.
- 18 S. K. Shin and J. L. Beauchamp, Identification of Mn(CO)<sub>n</sub>CF<sub>3</sub>- (n = 4, 5) structural isomers by IR multiphoton dissociation, collision-induced dissociation, and specific ligand displacement reactions: studies of the trifluoromethyl migratory decarbonylation reaction in the gas phase, *J. Am. Chem. Soc.*, 1990, **112**, 2057–2066.

- 19 J. R. Eyler, Infrared multiple photon dissociation spectroscopy of ions in Penning traps, *Mass Spectrom. Rev.*, 2009, **28**, 448–467.
- 20 S. E. Stefan and J. R. Eyler, Differentiation of glucose-containing disaccharides by infrared multiple photon dissociation with a tunable CO<sub>2</sub> laser and Fourier transform ion cyclotron resonance mass spectrometry, *Int. J. Mass Spectrom.*, 2010, **297**, 96–101.
- 21 R. T. Sanderson, Interrelation of bond dissociation energies and contributing bond energies, *J. Am. Chem. Soc.*, 1975, **97**, 1367–1372.
- 22 Table de correspondance des spectres IR, <https://www.sigmaaldrich.com/FR/fr/technical-documents/technical-article/analytical-chemistry/photometry-and-reflectometry/ir-spectrum-table>, (accessed June 21, 2023).
- 23 J. Martens, G. Berden, R. E. van Outersterp, L. A. J. Kluijtmans, U. F. Engelke, C. D. M. van Karnebeek, R. A. Wevers and J. Oomens, Molecular identification in metabolomics using infrared ion spectroscopy, *Sci. Rep.*, 2017, **7**, 3363.
- 24 M. Lettow, M. Grabarics, E. Mucha, D. A. Thomas, Ł. Polewski, J. Freyse, J. Rademann, G. Meijer, G. von Helden and K. Pagel, IR action spectroscopy of glycosaminoglycan oligosaccharides, *Anal. Bioanal. Chem.*, 2020, **412**, 533–537.
- 25 B. Chiavarino, M. E. Crestoni, S. Fornarini, F. Lanucara, J. Lemaire, P. Maitre and D. Scuderi, Infrared spectroscopy of isolated nucleotides. 1. The cyclic 3',5'-adenosine monophosphate anion, *Int. J. Mass Spectrom.*, 2008, **270**, 111–117.
- 26 C. Masellis, N. Khanal, M. Z. Kamrath, D. E. Clemmer and T. R. Rizzo, Cryogenic Vibrational Spectroscopy Provides Unique Fingerprints for Glycan Identification, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 2217–2222.
- 27 B. Schindler, L. Barnes, G. Renois, C. Gray, S. Chambert, S. Fort, S. Flitsch, C. Loison, A.-R. Allouche and I. Compagnon, Anomeric memory of the glycosidic bond upon fragmentation and its consequences for carbohydrate sequencing, *Nat. Commun.*, 2017, **8**, 973.
- 28 F. Lanucara, M. Elisa Crestoni, B. Chiavarino, S. Fornarini, O. Hernandez, D. Scuderi and P. Maitre, Infrared spectroscopy of nucleotides in the gas phase 2. The protonated cyclic 3',5'-adenosine monophosphate, *RSC Adv.*, 2013, **3**, 12711–12720.
- 29 R. E. van Outersterp, P. C. Kooijman, J. Merx, U. F. H. Engelke, N. Omidikia, M.-L. H. Tonneijck, K. J. Houthuijs, G. Berden, T. M. A. Peters, D. J. Lefeber, M. A. A. P. Willemsen, J. Mecinovic, J. J. Jansen, K. L. M. Coene, R. A. Wevers, T. J. Boltje, J. Oomens and J. Martens, Distinguishing Oligosaccharide Isomers Using Far-Infrared Ion Spectroscopy: Identification of Biomarkers for Inborn Errors of Metabolism, *Anal. Chem.*, 2023, **95**, 9787–9796.
- 30 B. Schindler, L. Barnes, C. J. Gray, S. Chambert, S. L. Flitsch, J. Oomens, R. Daniel, A. R. Allouche and I. Compagnon, IRMPD Spectroscopy Sheds New (Infrared) Light on the Sulfate Pattern of Carbohydrates, *J. Phys. Chem. A*, 2017, **121**, 2114–2120.
- 31 J. Oomens, D. T. Moore, G. Meijer and G. von Helden, Infrared multiple photon dynamics and spectroscopy of cationic PABA and its dehydroxylated fragment ion, *Phys. Chem. Chem. Phys.*, 2004, **6**, 710.
- 32 B. Lucas, G. Grégoire, J. Lemaire, P. Maître, J.-M. Ortega, A. Rupenyan, B. Reimann, J. Pierre Schermann and C. Desfrancois, Investigation of the protonation site in the dialanine peptide by infrared multiphoton dissociation spectroscopy, *Phys Chem Chem Phys*, 2004, **6**, 2659–2663.
- 33 B. Schindler, J. Joshi, A.-R. Allouche, D. Simon, S. Chambert, V. Brites, M.-P. Gageot and I. Compagnon, Distinguishing isobaric phosphated and sulfated carbohydrates by coupling of mass spectrometry with gas phase vibrational spectroscopy, *Phys Chem Chem Phys*, 2014, **16**, 22131–22138.
- 34 Å. Andersson, M. Poline, M. Kodambattil, O. Rebrov, E. Loire, P. Maître and V. Zhaunerchyk, Structure of Proton-Bound Methionine and Tryptophan Dimers in the Gas Phase Investigated with IRMPD Spectroscopy and Quantum Chemical Calculations, *J. Phys. Chem. A*, 2020, **124**, 2408–2415.
- 35 J. Seo, W. Hoffmann, S. Warnke, X. Huang, S. Gewinner, W. Schöllkopf, M. T. Bowers, G. Von Helden and K. Pagel, An infrared spectroscopy approach to follow  $\beta$ -sheet formation in peptide amyloid assemblies, *Nat. Chem.*, 2017, **9**, 39–44.

- 36 T. Hansen, H. Elferink, J. M. A. van Hengst, K. J. Houthuijs, W. A. Remmerswaal, A. Kromm, G. Berden, S. van der Vorm, A. M. Rijs, H. S. Overkleeft, D. V. Filippov, F. P. J. T. Rutjes, G. A. van der Marel, J. Martens, J. Oomens, J. D. C. Codée and T. J. Boltje, Characterization of glycosyl dioxolenium ions and their role in glycosylation reactions, *Nat. Commun.*, 2020, **11**, 2664.
- 37 F. T. Braak, H. Elferink, K. J. Houthuijs, J. Oomens, J. Martens and T. J. Boltje, Characterization of Elusive Reaction Intermediates Using Infrared Ion Spectroscopy: Application to the Experimental Characterization of Glycosyl Cations, *Acc. Chem. Res.*, 2022, **55**, 1669–1679.
- 38 A. L. Patrick, C. N. Stedwell, B. Schindler, I. Compagnon, G. Berden, J. Oomens and N. C. Polfer, Insights into the fragmentation pathways of gas-phase protonated sulfoserine, *Int. J. Mass Spectrom.*, 2015, **379**, 26–32.
- 39 J. Martens, J. Grzetic, G. Berden and J. Oomens, Structural identification of electron transfer dissociation products in mass spectrometry using infrared ion spectroscopy, *Nat. Commun.*, 2016, **7**, 11754.
- 40 B. Moge, B. Schindler, O. Yeni and I. Compagnon, Fucose Migration Pathways Identified Using Infrared Spectroscopy, *Angew. Chem. Int. Ed.*, 2023, **62**, e202300538.
- 41 Protein O-fucosylation: structure and function, *Curr. Opin. Struct. Biol.*, 2019, **56**, 78–86.
- 42 H. Takeuchi, H. Yu, H. Hao, M. Takeuchi, A. Ito, H. Li and R. S. Haltiwanger, O-Glycosylation modulates the stability of epidermal growth factor-like repeats and thereby regulates Notch trafficking, *J. Biol. Chem.*, 2017, **292**, 15964–15973.
- 43 K. T. Schjoldager, Y. Narimatsu, H. J. Joshi and H. Clausen, Global view of human protein glycosylation pathways and functions, *Nat. Rev. Mol. Cell Biol.*, 2020, **21**, 729–749.
- 44 M. Schneider, V. Kumar, L. U. Nordstrøm, L. Feng, H. Takeuchi, H. Hao, V. C. Luca, K. C. Garcia, P. Stanley, P. Wu and R. S. Haltiwanger, Inhibition of Delta-induced Notch signaling using fucose analogs, *Nat. Chem. Biol.*, 2018, **14**, 65–71.
- 45 J. Li, H.-C. Hsu, J. D. Mountz and J. G. Allen, Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases, *Cell Chem. Biol.*, 2018, **25**, 499–512.
- 46 Y.-C. Liu, H.-Y. Yen, C.-Y. Chen, C.-H. Chen, P.-F. Cheng, Y.-H. Juan, C.-H. Chen, K.-H. Khoo, C.-J. Yu, P.-C. Yang, T.-L. Hsu and C.-H. Wong, Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells, *Proc. Natl. Acad. Sci.*, 2011, **108**, 11332–11337.
- 47 E. Miyoshi, K. Moriwaki, N. Terao, C.-C. Tan, M. Terao, T. Nakagawa, H. Matsumoto, S. Shinzaki and Y. Kamada, Fucosylation Is a Promising Target for Cancer Diagnosis and Therapy, *Biomolecules*, 2012, **2**, 34–45.
- 48 K. Madunić, T. Zhang, O. A. Mayboroda, S. Holst, K. Stavenhagen, C. Jin, N. G. Karlsson, G. S. M. Lageveen-Kammeijer and M. Wuhrer, Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype, *Cell. Mol. Life Sci.*, 2021, **78**, 337–350.
- 49 M. Wiciński, E. Sawicka, J. Gębalski, K. Kubiak and B. Malinowski, Human Milk Oligosaccharides: Health Benefits, Potential Applications in Infant Formulas, and Pharmacology, *Nutrients*, 2020, **12**, 266.
- 50 M. McNeil, Elimination of internal glycosyl residues during chemical ionization-mass spectrometry of per-O-alkylated oligosaccharide-alditols, *Carbohydr. Res.*, 1983, **123**, 31–40.
- 51 V. Kováčik, J. Hirsch, P. Kováč, W. Heerma, J. Thomas-Oates and J. Haverkamp, Oligosaccharide characterization using collision-induced dissociation fast atom bombardment mass spectrometry: Evidence for internal monosaccharide residue loss, *J. Mass Spectrom.*, 1995, **30**, 949–958.
- 52 L. P. Brüll, W. Heerma, J. Thomas-Oates, J. Haverkamp, V. Kováčik and P. Kováč, Loss of internal 1 → 6 substituted monosaccharide residues from underivatized and per-O-methylated trisaccharides, *J. Am. Soc. Mass Spectrom.*, 1997, **8**, 43–49.
- 53 B. Ernst, D. R. Müller and W. J. Richter, False sugar sequence ions in electrospray tandem mass spectrometry of underivatized sialyl-Lewis-type oligosaccharides, *Int. J. Mass Spectrom. Ion Process.*, 1997, **160**, 283–290.
- 54 E. S. Hecht, P. L. Loziuk and D. C. Muddiman, Xylose Migration During Tandem Mass

- Spectrometry of N-Linked Glycans, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 729–732.
- 55 Y.-L. Ma, I. Vedernikova, H. Van den Heuvel and M. Claeys, Internal glucose residue loss in protonated O-diglycosyl flavonoids upon low-energy collision-induced dissociation, *J. Am. Soc. Mass Spectrom.*, 2000, **11**, 136–144.
- 56 M. Wuhler, A. M. Deelder and Y. E. M. van der Burgt, Mass spectrometric glycan rearrangements, *Mass Spectrom. Rev.*, 2011, **30**, 664–680.
- 57 L. P. Brüll, V. Kováčik, J. E. Thomas-Oates, W. Heerma and J. Haverkamp, Sodium-cationized oligosaccharides do not appear to undergo “internal residue loss” rearrangement processes on tandem mass spectrometry, *Rapid Commun. Mass Spectrom. RCM*, 1998, **12**, 1520–1532.
- 58 M. Lettow, E. Mucha, C. Manz, D. A. Thomas, M. Marianski, G. Meijer, G. von Helden and K. Pagel, The role of the mobile proton in fucose migration, *Anal. Bioanal. Chem.*, 2019, **411**, 4637–4645.
- 59 A. H. Franz and C. B. Lebrilla, Evidence for long-range glycosyl transfer reactions in the gas phase, *J. Am. Soc. Mass Spectrom.*, 2002, **13**, 325–337.
- 60 J. Hofmann, A. Stuckmann, M. Crispin, D. J. Harvey, K. Pagel and W. B. Struwe, Identification of Lewis and Blood Group Carbohydrate Epitopes by Ion Mobility-Tandem-Mass Spectrometry Fingerprinting, *Anal. Chem.*, 2017, **89**, 2318–2325.
- 61 C. Manz, M. Grabarics, F. Hoberg, M. Pugini, A. Stuckmann, W. B. Struwe and K. Pagel, Separation of isomeric glycans by ion mobility spectrometry – the impact of fluorescent labelling, *Analyst*, 2019, **144**, 5292–5298.
- 62 N. Desai, D. A. Thomas, J. Lee, J. Gao and J. L. Beauchamp, Eradicating mass spectrometric glycan rearrangement by utilizing free radicals, *Chem. Sci.*, 2016, **7**, 5390–5397.
- 63 D. J. Harvey, T. S. Mattu, M. R. Wormald, L. Royle, R. A. Dwek and P. M. Rudd, “Internal Residue Loss”: Rearrangements Occurring during the Fragmentation of Carbohydrates Derivatized at the Reducing Terminus, *Anal. Chem.*, 2002, **74**, 734–740.
- 64 E. Mucha, M. Lettow, M. Marianski, D. A. Thomas, W. B. Struwe, D. J. Harvey, G. Meijer, P. H. Seeberger, G. von Helden and K. Pagel, Fucose Migration in Intact Protonated Glycan Ions: A Universal Phenomenon in Mass Spectrometry, *Angew. Chem. Int. Ed Engl.*, 2018, **57**, 7440–7443.
- 65 J. S. Toraño, I. A. Gagarinov, G. M. Vos, F. Broszeit, A. D. Srivastava, M. Palmer, J. I. Langridge, O. Aizpurua-Olaizola, V. J. Somovilla and G.-J. Boons, Ion-Mobility Spectrometry Can Assign Exact Fucosyl Positions in Glycans and Prevent Misinterpretation of Mass-Spectrometry Data After Gas-Phase Rearrangement, *Angew. Chem. Int. Ed.*, 2019, **58**, 17616–17620.
- 66 B. Schindler, G. Laloy-Borgna, L. Barnes, A.-R. Allouche, E. Bouju, V. Dugas, C. Demesmay and I. Compagnon, Online Separation and Identification of Isomers Using Infrared Multiple Photon Dissociation Ion Spectroscopy Coupled to Liquid Chromatography: Application to the Analysis of Disaccharides Regio-Isomers and Monosaccharide Anomers, *Anal. Chem.*, 2018, **5**.
- 67 G. Renois-Predelus, B. Schindler and I. Compagnon, Analysis of Sulfate Patterns in Glycosaminoglycan Oligosaccharides by MS<sub>n</sub> Coupled to Infrared Ion Spectroscopy: the Case of GalNAc4S and GalNAc6S, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 1242–1249.
- 68 O. Yeni, A. Gharbi, S. Chambert, J. Rouillon, A.-R. Allouche, B. Schindler and I. Compagnon, O - Acetylated sugars in the gas phase: stability, migration, positional isomers and conformation, *Phys. Chem. Chem. Phys.*, 2022, **24**, 1016–1022.
- 69 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. Lütteke, 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*, 2015, **25**, 1323–1324.
- 70 M. Lettow, K. Greis, E. Mucha, T. R. Lambeth, M. Yaman, V. Kontodimas, C. Manz, W. Hoffmann, G. Meijer, R. R. Julian, G. von Helden, M. Marianski and K. Pagel, Decoding the Fucose Migration Product during Mass-Spectrometric analysis of Blood Group Epitopes, *Angew. Chem. Int. Ed.*, 2023, **62**, e202302883.
- 71 U. F. H. Engelke, R. E. van Outersterp, J. Merx, F. A. M. G. van Geenen, A. van Rooij, G.

- Berden, M. C. D. G. Huigen, L. A. J. Kluijtmans, T. M. A. Peters, H. H. Al-Shekaili, B. R. Leavitt, E. de Vrieze, S. Broekman, E. van Wijk, L. A. Tseng, P. Kulkarni, F. P. J. T. Rutjes, J. Mecinović, E. A. Struys, L. A. Jansen, S. M. G. Jr, S. Mercimek-Andrews, K. Hyland, M. A. A. P. Willemsen, L. A. Bok, C. D. M. van Karnebeek, R. A. Wevers, T. J. Boltje, J. Oomens, J. Martens and K. L. M. Coene, Untargeted metabolomics and infrared ion spectroscopy identify biomarkers for pyridoxine-dependent epilepsy, <https://www.jci.org/articles/view/148272/pdf>, (accessed March 16, 2023).
- 72 J. Merx, R. E. van Outersterp, U. F. H. Engelke, V. Hendriks, R. A. Wevers, M. C. D. G. Huigen, H. W. A. H. Waterval, I. M. L. W. Körver-Keularts, J. Mecinović, F. P. J. T. Rutjes, J. Oomens, K. L. M. Coene, J. Martens and T. J. Boltje, Identification of  $\Delta$ -1-pyrroline-5-carboxylate derived biomarkers for hyperprolinemia type II, *Commun. Biol.*, 2022, **5**, 1–8.
- 73 S.-S. Lee, S. Park, Y. Hong, J. Lee, J.-H. Kim, D. Yoon, X. Kong, S. Lee and H. Bin Oh, Chiral differentiation of d - and l -alanine by permethylated  $\beta$ -cyclodextrin: IRMPD spectroscopy and DFT methods, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14729–14737.
- 74 S.-S. Lee, J. Lee, J. Hyeon Oh, S. Park, Y. Hong, B. Ki Min, H. H. L. Lee, H. I. Kim, X. Kong, S. Lee and H. Bin Oh, Chiral differentiation of d - and l -isoleucine using permethylated  $\beta$ -cyclodextrin: infrared multiple photon dissociation spectroscopy, ion-mobility mass spectrometry, and DFT calculations, *Phys. Chem. Chem. Phys.*, 2018, **20**, 30428–30436.
- 75 L. Sun, F. Huang, W. Liu, L. Lin, Y. Hong and X. Kong, Chiral differentiation of - and -penicillamine by  $\beta$ -cyclodextrin: Investigated by IRMPD spectroscopy and theoretical simulations, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, 2020, **241**, 118653.
- 76 C. N. Stedwell, A. L. Patrick, K. Gulyuz and N. C. Polfer, Screening for Phosphorylated and Nonphosphorylated Peptides by Infrared Photodissociation Spectroscopy, *Anal. Chem.*, 2012, **84**, 9907–9912.
- 77 A. L. Patrick, C. N. Stedwell and N. C. Polfer, Differentiating Sulfopeptide and Phosphopeptide Ions via Resonant Infrared Photodissociation, *Anal. Chem.*, 2014, **86**, 5547–5552.
- 78 L. Barnes, A.-R. Allouche, S. Chambert, B. Schindler and I. Compagnon, Ion spectroscopy of heterogeneous mixtures: IRMPD and DFT analysis of anomers and conformers of monosaccharides, *Int. J. Mass Spectrom.*, 2020, **447**, 116235.
- 79 L. Barnes, B. Schindler, S. Chambert, A.-R. Allouche and I. Compagnon, Conformational preferences of protonated N-acetylated hexosamines probed by InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and ab initio calculations, *Int. J. Mass Spectrom.*, 2017, **421**, 116–123.
- 80 B. Schindler, N. Bagdadi, S. Melizi, L. Barnes, S. Chambert, A.-R. Allouche, I. Compagnon and G. Renois-Predelus, MS/IR, a new MS-based hyphenated method for analysis of hexuronic acid epimers in glycosaminoglycans, *Glycoconj. J.*, 2017, **34**, 421–425.
- 81 B. Schindler, L. Legentil, A.-R. Allouche, V. Ferrières and I. Compagnon, Spectroscopic diagnostic for the ring-size of carbohydrates in the gas phase: furanose and pyranose forms of GalNAc, *Phys. Chem. Chem. Phys.*, 2019, **21**, 12460–12467.
- 82 J. S. Ho, A. Gharbi, B. Schindler, O. Yeni, R. Brédy, L. Legentil, V. Ferrières, L. L. Kiessling and I. Compagnon, Distinguishing Galactoside Isomers with Mass Spectrometry and Gas-Phase Infrared Spectroscopy, *J. Am. Chem. Soc.*, 2021, **143**, 10509–10513.
- 83 L. Jaravel, B. Schindler, J. Randon, I. Compagnon, C. Demesmay and V. Dugas, Off-line coupling of capillary isotachopheresis separation to IRMPD spectroscopy for glycosaminoglycans analysis: Application to the chondroitin sulfate disaccharides model solutes, *J. Chromatogr. A*, 2020, **1617**, 460782.
- 84 B. Favreau, O. Yeni, S. Ollivier, J. Boustie, F. L. Dévéhat, J.-P. Guégan, M. Fanuel, H. Rogniaux, R. Brédy, I. Compagnon, D. Ropartz, L. Legentil and V. Ferrières, Synthesis of an Exhaustive Library of Naturally Occurring Gal f -Man p and Gal p -Man p Disaccharides. Toward Fingerprinting According to Ring Size by Advanced Mass Spectrometry-Based IM-MS and IRMPD, *J. Org. Chem.*, 2021, **86**, 6390–6405.
- 85 A. Depraz Depland, G. Renois-Predelus, B. Schindler and I. Compagnon, Identification of sialic acid linkage isomers in glycans using coupled InfraRed Multiple Photon Dissociation (IRMPD)

spectroscopy and mass spectrometry, *Int. J. Mass Spectrom.*, 2018, **434**, 65–69.

86 A. Mookherjee, S. S. Uppal, T. A. Murphree and M. Guttman, Linkage Memory in Underivatized Protonated Carbohydrates, *J. Am. Soc. Mass Spectrom.*, 2021, **32**, 581–589.

87 O. Yeni, S. Ollivier, B. Moge, D. Ropartz, H. Rogniaux, L. Legentil, V. Ferrières and I. Compagnon, Ring-Size Memory of Galactose-Containing MS/MS Fragments: Application to the Detection of Galactofuranose in Oligosaccharides and Their Sequencing, *J. Am. Chem. Soc.*, 2023, **145**, 15180–15187.

# Chapitre II : développements expérimentaux de l'IRMPD pour les analyses haut débit

## TABLE DES MATIERES

---

<b>1</b>	<b>Introduction .....</b>	<b>51</b>
<b>2</b>	<b>Accélération de l'acquisition des données .....</b>	<b>51</b>
2.1	Acquisition rapide de spectres IRMPD .....	51
2.1.1	Nouveau laser et nouveau type de piège à ions : Nouveau montage .....	51
2.1.2	Matériel et méthode.....	53
2.1.3	Acquisition menée en 14 secondes.....	55
2.1.4	Limites.....	56
2.2	Ajout du laser CO <sub>2</sub> .....	57
2.2.1	Fonctionnement du laser CO <sub>2</sub> .....	57
2.2.2	Nouvelle fenêtre .....	57
2.2.3	Montage expérimental .....	58
2.2.4	Effet de l'irradiation avec deux lasers.....	59
2.2.5	Application à des espèces lithiées, sodées et déprotonées.....	59
<b>3</b>	<b>Reconnaissance de spectres.....</b>	<b>62</b>
3.1	Approche par les scores .....	63
3.1.1	Calcul de la ressemblance .....	63
3.1.2	Calcul de l'erreur.....	64
3.2	Approche par ACP .....	65
3.2.1	La méthode .....	65
3.2.2	Application sur les hexosamines.....	65
3.2.3	L'analyse en ACP .....	66
3.2.4	Confirmation de l'identification.....	69
3.3	Approche par apprentissage machine.....	71
3.3.1	Production des données .....	71
3.3.2	Architecture du modèle .....	73
3.3.3	Exactitude de la classification du modèle .....	73
3.3.4	Incertitude et précision du modèle.....	74
3.3.5	Conclusion et ouverture.....	75
<b>4</b>	<b>Conclusion .....</b>	<b>76</b>
<b>5</b>	<b>Bibliographie .....</b>	<b>77</b>

## 1 INTRODUCTION

---

Comme vu dans le chapitre précédent, la spectroscopie IRMPD a un fort potentiel dans le domaine de la chimie analytique. Mais elle se heurte encore à des problèmes qui freinent son plein essor. Par exemple, actuellement, les utilisateurs doivent être experts en spectroscopie pour interpréter des données, ou encore aligner le laser dans le spectromètre de masse, ce qui n'est pas compatible avec les opérateurs procédant aux analyses en routine dans un laboratoire. De plus, l'acquisition d'un spectre prend du temps, au moyen de l'OPO/OPA, on parle de plusieurs dizaines de minutes, ce qui peut être rédhibitoire pour certaines applications. Dans le but de faciliter les analyses en IRMPD, nous disposons par exemple d'un logiciel dédié à l'acquisition. Ce programme rend le traitement et l'interprétation des données plus simples, ces étapes seraient fastidieuses à faire manuellement. Dans cette partie, nous montrerons les différents développements mis en œuvre pour rendre la spectroscopie IRMPD plus attractive. Les développements des lasers accordables dans la gamme des 3  $\mu\text{m}$  qui sont de plus en plus petits, ergonomiques avec une plus grande cadence, facilitent l'utilisation de l'IRMPD. De tels lasers diminuent grandement le temps d'acquisition des données. Nous décrirons également de nouvelles approches mises en place afin d'identifier un composé inconnu à partir de son spectre.

## 2 ACCELERATION DE L'ACQUISITION DES DONNEES

---

### 2.1 ACQUISITION RAPIDE DE SPECTRES IRMPD

Le but dans cette partie, est d'accélérer l'acquisition des spectres IRMPD dans le but de rendre cette spectroscopie compatible avec des analyses en routine dans des laboratoires avec un gros débit d'analyses. Cette partie couvre l'article : **"Rapid IRMPD (InfraRed multiple photon dissociation) analysis for glycomics"** publié dans la revue *Analyst* par Ozgur Yeni, Baptiste Schindler, Baptiste Moge et Isabelle Compagnon en 2022.

L'acquisition d'un spectre IRMPD au moyen du montage « historique » de l'équipe décrit dans la partie instrumentation du chapitre I prend environ 45 minutes. Des efforts sont faits pour diminuer ce temps d'acquisition. Dans cet article nous diminuons grandement ce temps d'acquisition d'un facteur 4 en changeant seulement de laser, mieux ce temps peut encore être amélioré jusqu'à gagner cette fois-ci un facteur 200 par rapport à l'ancien montage pour arriver à seulement 14 secondes si on choisit de diminuer l'échantillonnage.

#### 2.1.1 Nouveau laser et nouveau type de piège à ions : Nouveau montage

En se servant de pièges à ions, au lieu de FT-ICR, le temps d'acquisition d'un spectre de masse est grandement réduit. L'acquisition d'un spectre de masse requière différentes étapes, dont l'injection des ions, leur activation suivies de la détection. Les durées de ces différentes étapes ne sont pas les mêmes entre ces deux technologies de spectromètres de masse. Ce qui résulte du fait qu'un spectre de masse obtenu au moyen d'un FT-ICR nécessite des durées de l'ordre de la seconde (pour les résolutions les plus hautes), tandis que dans le cas d'un piège à ions, le temps typique est seulement de quelques millisecondes. L'un des points critiques dans l'acquisition d'un spectre IRMPD se trouve maintenant dans la durée nécessaire à l'activation des ions. En effet, pour obtenir une photofragmentation suffisante des ions, 700 ms d'irradiation étaient nécessaires avec un laser de type OPO/OPA. Ce temps est dû aux spécificités intrinsèques de l'OPO/OPA notamment à la cadence de

10 Hz du laser de pompe. Une des pistes alors envisagées a été de diminuer ce temps d'activation. Pour cela, un nouveau laser a été utilisé : le laser Firefly de chez MSquared, il s'agit d'un laser OPO/OPA fibré, beaucoup plus compact qu'un laser OPO/OPA classique. Ce laser a une cadence qui peut aller jusqu'à 300 kHz, cadence bien supérieure à celle rencontrée dans le cas d'OPO/OPA classiques fonctionnant à 10 Hz (Figure 1), avec une durée d'impulsion laser d'environ 10 ns, il est accordable sur la gamme 2500-3735  $\text{cm}^{-1}$ . La puissance laser en sortie de Firefly est d'environ 150 mW. Sa résolution spectrale est inférieure à 10  $\text{cm}^{-1}$ . Aucun alignement de la part de l'opérateur à l'intérieur du laser n'est requis. Des spectres IRMPD ont pu être acquis au moyen de lasers à haute cadence<sup>1</sup>. Ces lasers peuvent même autoriser une diminution du temps d'irradiation<sup>2</sup>, de même que les lasers continus<sup>3</sup>. Certains aspects du processus IRMPD ne sont pas encore totalement expliqués, le fait d'augmenter la cadence des lasers utilisés pour photofragmenter des ions dans un piège est nouveau, et nous place dans un tout nouveau paradigme. En effet, la spectroscopie IRMPD avec une cadence de l'ordre de la dizaine de Hz suppose qu'entre chaque impulsion laser (environ 100 ms) les ions ont le temps de refroidir par collision avec les atomes de gaz présents dans le piège à ions. Les ions passant le seuil de dissociation sont photofragmentés au moyen d'une seule impulsion laser. En revanche, dans le cadre de l'utilisation de lasers à haute cadence (6,7  $\mu\text{s}$  dans le cas d'une irradiation à 150 kHz, avec une énergie bien moindre), nous ne pouvons affirmer que les ions absorbent suffisamment de photons pour franchir le seuil de dissociation en une seule impulsion laser. De même, les ions ne sont pas forcément complètement refroidis entre deux impulsions laser à une telle cadence. Ainsi une impulsion laser peut potentiellement fournir l'énergie manquante à un ion pré-excité par une ou des impulsion(s) précédente(s) pour franchir le seuil de dissociation. Des études approfondies permettront de mieux appréhender les processus de spectroscopie IRMPD dans de tels régimes de cadence laser.

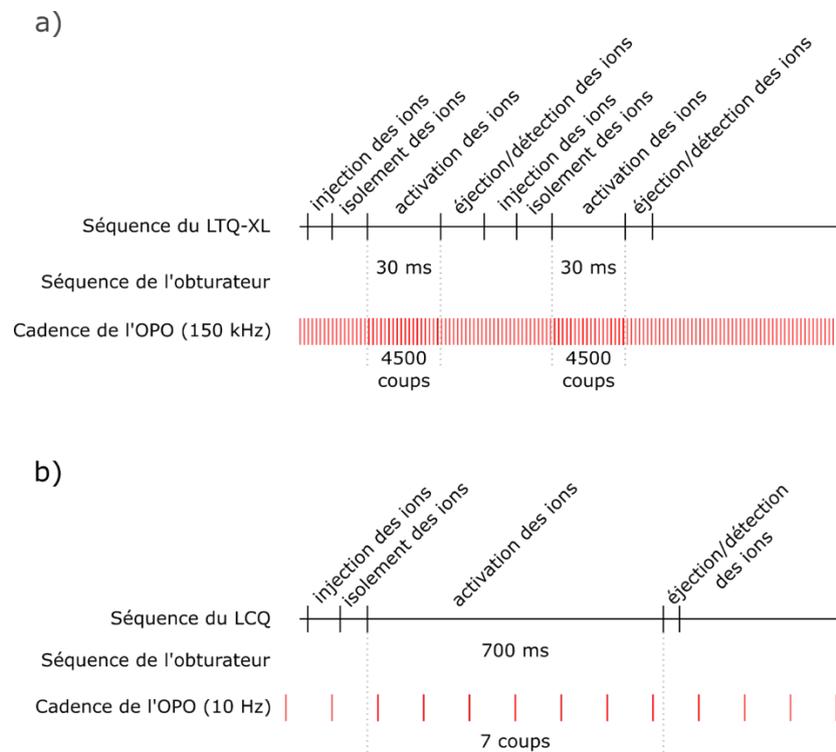


Figure 1 : Comparaison des séquences des deux couplages le montage avec le laser à haute cadence a) et le montage original (10 Hz de cadence) b). La cadence du laser présent sur le nouveau montage est de très loin supérieure à la cadence du montage original (environ 15 000 fois). La durée d'irradiation est toujours le facteur dictant le temps d'acquisition des spectres. Ce qui veut dire qu'avant de changer de technologie de spectromètres de masse pour accélérer l'acquisition des données, des efforts pour diminuer cette durée d'irradiation permettront de réduire le temps d'acquisition des spectres.

Le montage expérimental (Figure 2) consiste en un spectromètre de masse de type piège à ions équipé d'une source d'ionisation par electrospray LTQ-XL de chez ThermoFinnigan. Cette fois-ci, le piège à ion est de type linéaire, les ions ne sont pas concentrés en un point, mais sont piégés de manière linéaire. Le piège à ions prend la forme d'un quadrupôle, avec une électrode d'entrée et une électrode à l'opposé de cette entrée. La fenêtre transparente à l'infrarouge installée sur le LTQ-XL est une fenêtre en  $MgF_2$  dont la transmission est constante dans la gamme des  $3 \mu m$  (Figure 6 haut). Le laser Firefly opère à 150 kHz, avec une puissance en sortie de laser de 150 mW, chaque tir laser dure moins de 10 ns et porte une énergie de  $1 \mu J$ . C'est 10 000 fois moins que les 10 mJ par tir laser de l'OPO. En revanche en raison de la cadence du Firefly, 30 ms d'irradiation (4500 tirs laser) apporte 4,5 mJ d'énergie dans le piège à ions. Ce qui est du même ordre de grandeur que les 10 mJ apportés toutes les 100 ms par un tir laser de l'OPO à 10 Hz, l'avantage est donc qu'une irradiation de 30 ms avec le laser Firefly (300 kHz) photofragmente autant qu'une irradiation de 700 ms avec le montage OPO à 10 Hz, ce qui permet d'acquérir des données plus rapidement.

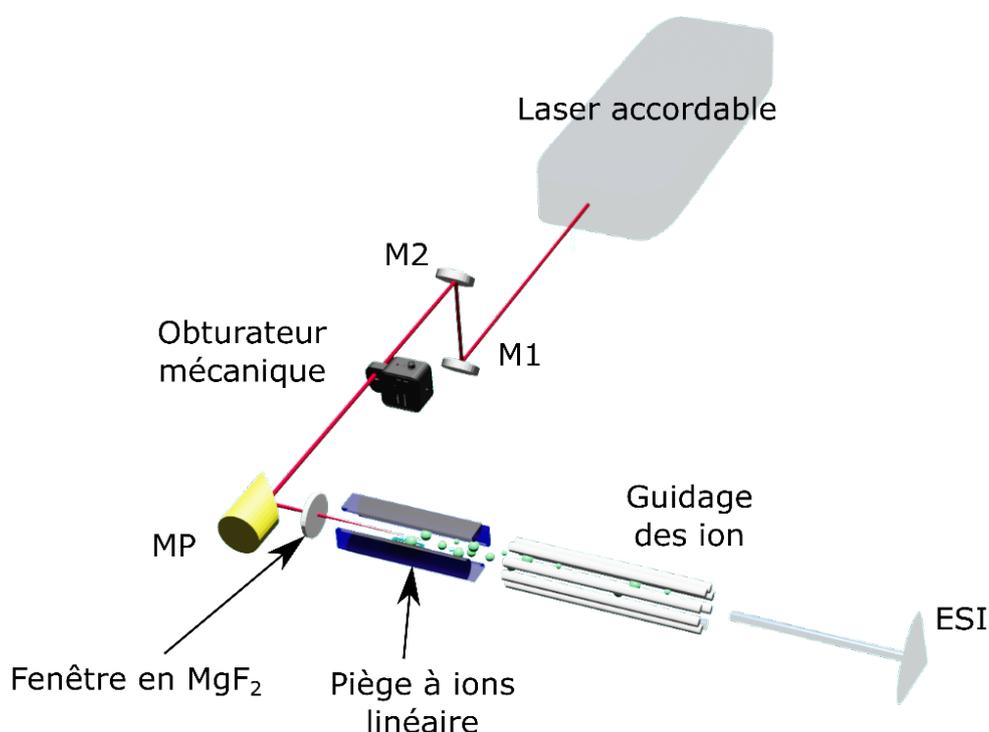


Figure 2 : Nouveau montage développé afin d'accélérer l'acquisition de spectres IRMPD. Le laser accordable est plus compact que le laser OPO/OPA pompé par YAG, M1 et M2 sont des miroirs recouvert d'or, MP est un miroir parabolique hors axe focalisant le faisceau au cœur du piège à ions. Le  $MgF_2$  est transparent dans la gamme des  $3 \mu m$ .

### 2.1.2 Matériel et méthode

L'échantillon étudié ici est un monosaccharide : la glucosamine. La glucosamine hydrochloride vient de chez Sigma. L'échantillon est dissous dans de l'eau pure avant d'être dilué à  $50 \mu M$  dans un mélange eau/MeOH 50:50 v/v contenant 0.1 % d'acide acétique. La glucosamine est le standard de référence dans l'équipe. Le signal de la glucosamine protonée en mode positif apparaît à  $m/z = 180$ , c'est donc cet ion qui est isolé et irradié avec le nouveau montage.

La Figure 3 montre la photofragmentation de la glucosamine avec une cadence de 10 Hz ou une cadence de 150 kHz. Le rendement de photofragmentation est calculé en considérant le fragment à  $m/z = 162$  qui correspond à la perte de la molécule d'eau à l'extrémité réductrice de la glucosamine. Étonnamment, pour atteindre un rendement de photofragmentation d'environ 50 % seulement 30 ms

sont nécessaire (Figure 3 haut) avec le nouveau montage (environ 140 mW et 150 kHz) contre 700 ms d'irradiation avec le montage OPO/LCQ (10 Hz). Le spectre de photofragmentation acquis au moyen de l'OPO (Figure 3 bas) présente également un pic à  $m/z = 144$  plus marqué que ce qui est observé sur le spectre obtenu au moyen du laser à haute cadence.

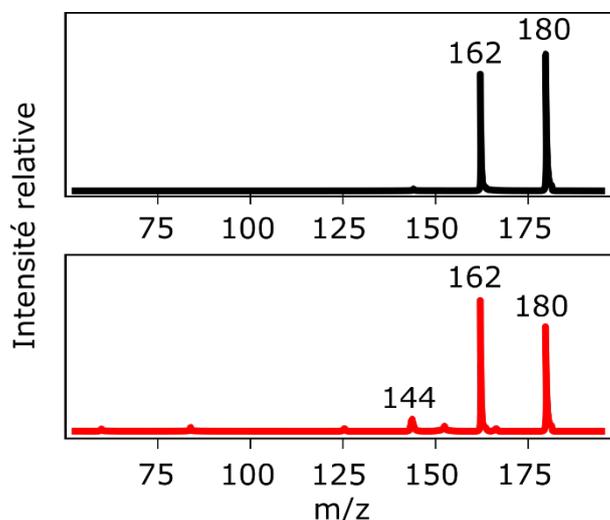


Figure 3 : Spectre de photofragmentation de la glucosamine protonée ( $m/z = 180$ ) sous irradiation du faisceau infrarouge à  $3345\text{ cm}^{-1}$  avec le laser haute cadence (noir) pendant 30 ms ou avec l'OPO/OPA (rouge) pendant 700 ms. Les intensités relatives sont comparables d'un montage à l'autre.

Le spectre IRMPD dans la gamme des  $3\text{ }\mu\text{m}$  de cette espèce est acquis au moyen du montage comprenant le laser Firefly ainsi qu'au moyen du laser OPO (Figure 4 haut et bas respectivement). Dans les deux cas, les bandes d'absorption sont observées dans les fréquences des groupements C-H, N-H et O-H. Les deux spectres sont similaires : ils montrent tous deux une intense bande d'absorption à  $3640\text{ cm}^{-1}$  et un autre maximum d'absorption à  $3345\text{ cm}^{-1}$ . Les intensités relatives varient quelques peu entre ces deux spectres. En considérant l'ensemble des bandes, il ressort que la bande à  $3345\text{ cm}^{-1}$  est relativement plus intense au moyen du nouveau montage. Le spectre obtenu au moyen du laser Firefly, présente une seule bande dans la gamme des O-H à  $3640\text{ cm}^{-1}$ . Ce qui cohérent avec les spécifications du laser Firefly qui offre une résolution de  $5\text{ cm}^{-1}$  tout le long du spectre à l'exception de la gamme au-dessus de  $3600\text{ cm}^{-1}$  qui est ici de  $10\text{ cm}^{-1}$ . Une autre différence est que la bande aux alentours de  $3050\text{ cm}^{-1}$  n'est pas présente dans le spectre IRMPD obtenu avec le nouveau montage. En IRMPD, qui est un processus multiphotonique, la photofragmentation est très dépendante de l'énergie fournie au système, la moindre variation dans l'intensité des tirs lasers peut engendrer une différence significative dans le rendement de photofragmentation. Ici, la faible absorption à  $3050\text{ cm}^{-1}$  disparaît complètement.

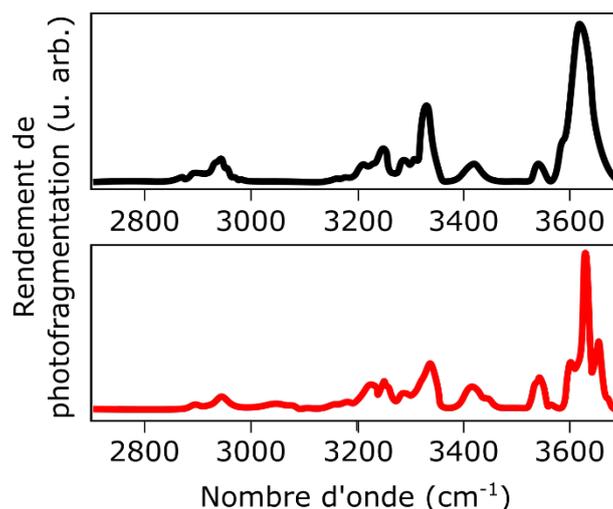


Figure 4 : Spectre IRMPD de la glucosamine protonée avec le Firefly (noir) ou l'OPO (rouge), le spectre acquis au moyen de l'OPO est mieux résolu que le spectre acquis au moyen du Firefly notamment dans la région des O-H (aux alentours de 3600  $\text{cm}^{-1}$ ). Cette résolution découle de la résolution spectrale du laser utilisé.

### 2.1.3 Acquisition menée en 14 secondes

Les spectres contiennent classiquement 1700 points. En diminuant le temps d'irradiation de 700 ms à seulement 30 ms, le temps d'acquisition passe à 18 min (Figure 5a). Une des solutions pour gagner encore du temps sur l'acquisition d'un spectre IRMPD est de diminuer le nombre de points acquis : le fait d'être à température ambiante, implique que les bandes d'absorption sont relativement larges ( $> 5 \text{ cm}^{-1}$ ), et donc une diminution de l'échantillonnage (augmentation de la vitesse de scan du laser) n'a pas d'influence sur la résolution spectroscopique du spectre IRMPD. Ainsi, le nombre de points peut être diminué à seulement 300 (Figure 5b). La qualité du spectre acquis en seulement une minute (Figure 5b) est très bonne, elle ne diffère que peu de ce qui est obtenu en 18 minutes (Figure 5a). La vitesse de scan est passée de 0,9  $\text{cm}^{-1}/\text{seconde}$  à 20  $\text{cm}^{-1}/\text{seconde}$ . Il est encore possible d'augmenter la vitesse de balayage du laser, et de descendre à seulement 30 secondes (Figure 5c), 20 secondes (Figure 5d) ou 14 secondes (Figure 5e) pour obtenir un spectre IRMPD. Avec seulement 14 secondes d'acquisition, le spectre ne comporte que 75 points, il est toujours reconnaissable avec sa signature à 2950  $\text{cm}^{-1}$  dans la gamme des C-H, les deux principales bandes à 3260 et 3345  $\text{cm}^{-1}$  dans la région des NH ainsi que les trois motifs présents dans la gamme des O-H : 3425, 3550 et 3640  $\text{cm}^{-1}$  (Figure 5e). On note, que seuls sur les spectres de la Figure 5a, b et c une ligne représentant une moyenne glissante est intégrée, sur les deux derniers spectres, les points sont directement reliés sans plus de traitement de données.

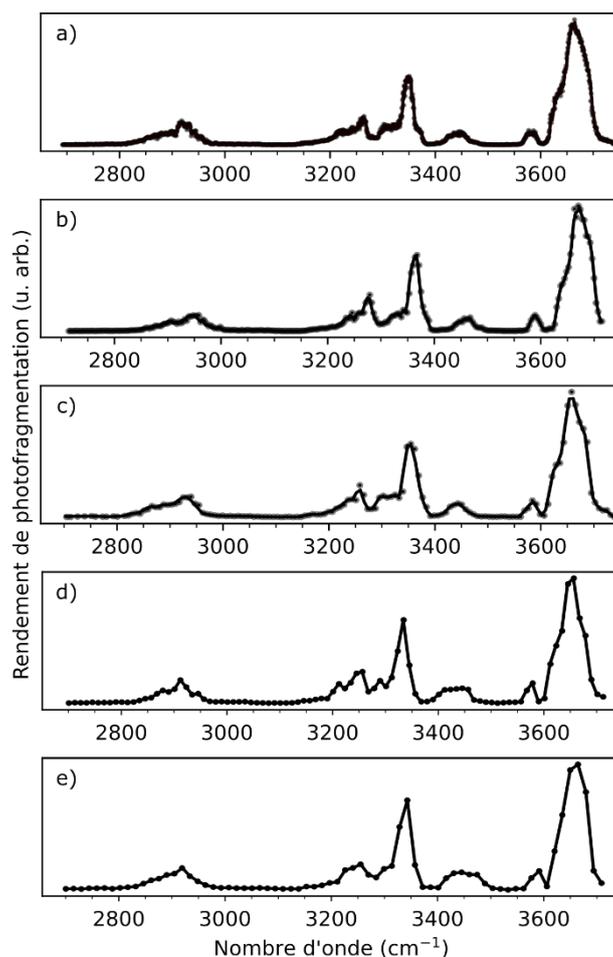


Figure 5 : Spectres IRMPD de la glucosamine avec différents temps d'acquisition 18 min a), une minute b), 30 secondes c), 20 secondes d) et 14 secondes e), ces spectres ont été obtenus au moyen du laser à haute cadence. Dans un premier temps entre 18 minutes et 30 secondes, l'échantillonnage ne semble pas avoir tellement diminué, les spectres présentent un aspect similaire. C'est lorsque l'on diminue encore le temps d'acquisition que la différence commence sérieusement à se voir avec des spectres bien moins échantillonnés.

Nous avons pu montrer que le temps d'acquisition d'un spectre IRMPD peut descendre à seulement 14 secondes. Selon l'application souhaitée, l'échantillonnage du spectre n'est pas le critère le plus important. En effet, un spectre acquis en 14 secondes n'est pas suffisamment échantillonné pour pouvoir être comparé avec précision à des calculs théoriques, mais il demeure néanmoins intéressant dans l'application analytique avec une empreinte reconnaissable. Parfois, une gamme du spectre n'est pas forcément intéressante, il n'est alors pas nécessaire de l'acquérir, cette approche peut aussi faire gagner du temps, car moins de points sont acquis. Ainsi la spectroscopie IRMPD peut arriver à des temps d'acquisition de quelques secondes, ces temps sont compatibles avec un couplage en ligne de la chromatographie liquide en amont de la spectroscopie.

#### 2.1.4 Limites

Les espèces sodées, lithiées ou déprotonées sont fréquemment caractérisées par une énergie de dissociation plus élevée que les espèces protonées. La puissance délivrée par le laser à haute cadence ne suffit pas à photofragmenter ces espèces. C'est l'un des inconvénients de ce montage qui se limite à l'étude d'espèces protonées et ammoniées. C'est pour cette raison qu'un nouveau montage présenté ci-après a été développé.

## 2.2 AJOUT DU LASER CO<sub>2</sub>

Dans cette partie, l'ajout d'un deuxième laser au montage développé est décrit afin de franchir le seuil de dissociation de l'ion d'intérêt. Le montage est décrit dans l'article "**CO<sub>2</sub> laser enhanced rapid IRMPD spectroscopy for glycan analysis**" par Baptiste Moge, Ozgur Yeni, Alicia Infantino et Isabelle Compagnon publié dans *International Journal of Mass Spectrometry* en 2023.

L'idée pour franchir le seuil de fragmentation plus élevé est d'apporter plus d'énergie au système au moyen d'un laser CO<sub>2</sub>. Ce laser va permettre de préchauffer le système afin de porter l'énergie interne du système juste sous le seuil de dissociation de l'ion. L'absorption d'un photon provenant du laser accordable apportera l'énergie nécessaire pour franchir le seuil de dissociation pour permettre de fragmenter le composé. Le laser CO<sub>2</sub> utilisé émet à 10,6  $\mu\text{m}$ , ce qui est une toute autre région du spectre IR que la gamme des 3  $\mu\text{m}$  étudiée dans notre équipe. Cette stratégie a largement été employée pour augmenter le taux de photofragmentation dans des montages entre un FT-ICR et un OPO lors des débuts de la spectroscopie IRMPD.

### 2.2.1 Fonctionnement du laser CO<sub>2</sub>

Le laser CO<sub>2</sub> est un laser à gaz contenant un mélange entre du CO<sub>2</sub> (10-20 %), de l'azote (10-20 %) et de l'hélium. Ce type de laser fonctionne par décharges électriques dans ce mélange de gaz. Un tel laser permet d'obtenir des puissances de sortie élevées. Sa puissance peut être facilement modulée. De tels lasers irradient de manière continue autour de 10,6  $\mu\text{m}$ .

### 2.2.2 Nouvelle fenêtre

Du fait de l'utilisation d'un second laser émettant dans un autre domaine de nombres d'onde, la fenêtre en fluorure de magnésium n'est pas adaptée. Le MgF<sub>2</sub> est parfaitement transparent dans la gamme des 3  $\mu\text{m}$ , la transmission est constante dans cette gamme. En revanche sa transmission diminue lorsque la longueur d'onde augmente, jusqu'à être nulle au-delà de 9  $\mu\text{m}$  (Figure 6a). Les fenêtres en germanium traitées avec un traitement antireflet sont plus adaptées à de tels montages. Dans la gamme des 3  $\mu\text{m}$ , la transmission n'est pas aussi optimale qu'avec le MgF<sub>2</sub>, la transmission varie de quelques % en restant supérieure à 95 % (Figure 6b). On s'attend à ce que la faible perte dans la gamme des 3  $\mu\text{m}$  soit compensée par la puissance apportée par le laser CO<sub>2</sub>. Cette fenêtre est donc choisie pour effectuer ces mesures.

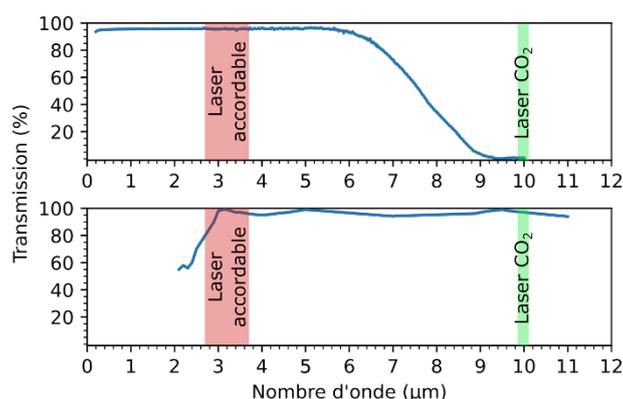


Figure 6 : Transmission en fonction de la longueur d'onde d'une fenêtre en MgF<sub>2</sub> a) ou en germanium b). La fenêtre en MgF<sub>2</sub> présente une très bonne transmission dans la gamme des 3  $\mu\text{m}$ , mais cette transmission chute entre 6 et 9  $\mu\text{m}$  pour être nulle à 10  $\mu\text{m}$ . La fenêtre en germanium est, elle moins performante dans la gamme des 3  $\mu\text{m}$  mais la transmission est suffisamment haute dans la gamme des 10  $\mu\text{m}$ .

Les miroirs utilisés pour apporter les faisceaux IR ne souffrent pas de ce problème, en effet ils présentent une réflexion similaire dans les deux gammes de longueurs d'onde.

### 2.2.3 Montage expérimental

Le montage est présenté en Figure 7, le faisceau du laser CO<sub>2</sub> est réfléchi sur M'1 et M'2 avant d'être rendu colinéaire avec le faisceau du laser accordable M'3. Ensuite, les deux lasers sont réfléchis sur M1 et M2 avant de passer à travers un obturateur mécanique. Cet obturateur assure la synchronisation entre l'étape d'activation des ions et l'irradiation réelle des ions. Les faisceaux arrivent finalement sur le miroir parabolique hors axe qui les fait entrer dans le piège à ions en les focalisant au même endroit dans le piège à ions. Le laser CO<sub>2</sub> irradie en continu à 10,6 μm, tandis que le laser Firefly est accordable dans la gamme 2700-3700 cm<sup>-1</sup> et irradie avec une cadence de 150 kHz en faisant varier le nombre d'onde.

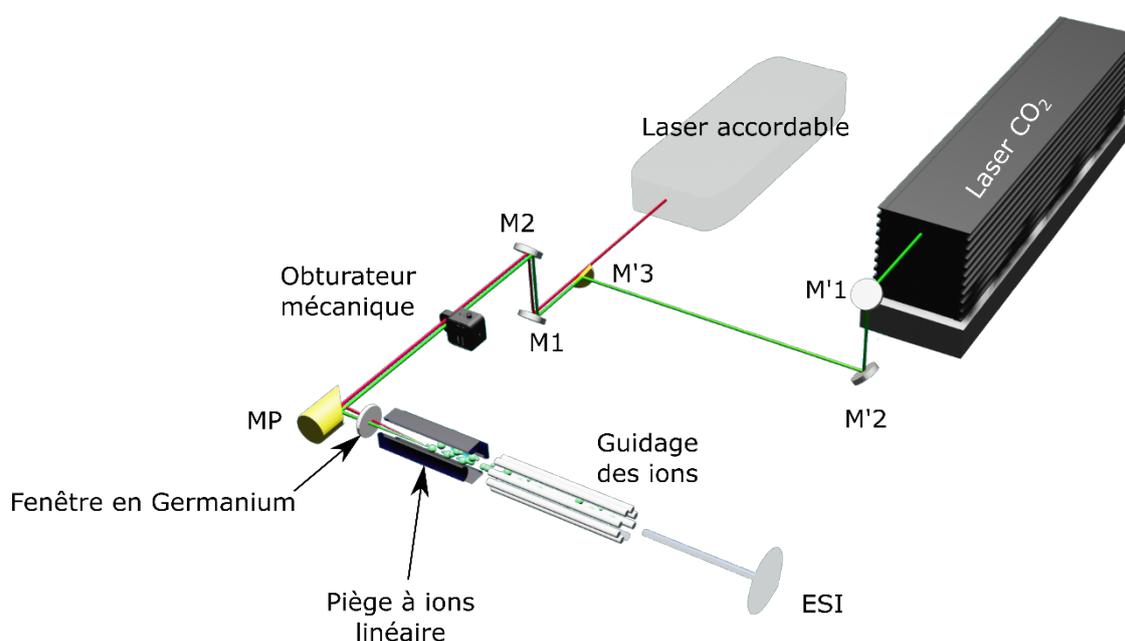


Figure 7 : Le montage représenté sur cette figure représente le couplage entre les 2 lasers et le spectromètre de masse développé au cours de cette thèse. Les faisceaux lasers sont rendus colinéaires par l'utilisation d'un miroir demi lune. La fenêtre installée au niveau du spectromètre de masse est cette fois ci une fenêtre en germanium.

D'autres couplages entre un laser accordable et un laser CO<sub>2</sub> continu ont déjà été proposés. Dans ces montages, deux obturateurs sont disposés de manière à avoir deux irradiations consécutives<sup>4-8</sup>. Une première irradiation des ions avec le laser accordable (OPO ou FEL) pour apporter de l'énergie aux ions au cas où ceux-ci absorbent à ce nombre d'onde, suivie d'une irradiation du laser CO<sub>2</sub> pour franchir le seuil de dissociation et donc déclencher la photofragmentation des ions préexcités. Les irradiations peuvent aussi avoir lieu simultanément<sup>9</sup>. Dans notre cas, le fait de se faire succéder les deux irradiations n'aidait en rien la photofragmentation. Aucun changement significatif n'est observé entre le fait d'irradier uniquement avec le laser Firefly ou d'ajouter une deuxième étape d'irradiation avec le laser CO<sub>2</sub>. Si le laser Firefly seul photofragmente, le rendement de photofragmentation n'est pas augmenté par l'ajout d'une étape d'irradiation avec le laser CO<sub>2</sub>. Si le laser Firefly ne photofragmente pas les ions seul, l'ajout d'une deuxième étape d'irradiation ne permet pas d'obtenir de la photofragmentation. Vraisemblablement les composés préchauffés par le laser CO<sub>2</sub> se thermalisent beaucoup plus rapidement par collision avec le gaz dans le piège à ions que dans une configuration FT-ICR entre les deux irradiations, ce qui annule le gain d'énergie apporté par la préchauffe du système.

En effet, les articles décrivant des irradiations successives<sup>4-8</sup> se servaient de FT-ICR dans lesquels le vide est bien plus poussé que dans le cas de pièges à ions dont la référence<sup>9</sup> et nous nous servons.

#### 2.2.4 Effet de l'irradiation avec deux lasers

La puissance en sortie de laser CO<sub>2</sub> peut être choisie manuellement. Cette puissance est sélectionnée de manière à apporter le plus d'énergie interne possible aux ions sans qu'ils ne franchissent le seuil de dissociation. En pratique, l'énergie est choisie en vérifiant le spectre de masse en direct, dès qu'un fragment apparaît, la puissance est alors légèrement diminuée. L'irradiation de la glucosamine protonée par le laser CO<sub>2</sub> n'engendre pas de photofragmentation, aucun fragment n'apparaît sur le spectre de la Figure 8a. Lors de l'irradiation de la glucosamine protonée à 3622 cm<sup>-1</sup> (mode de vibration d'un hydroxyl), le fragment à m/z = 162 (perte de l'hydroxyl en position anomérique) apparaît (Figure 8b). Enfin, la glucosamine est irradiée avec les deux lasers de manière simultanée (Figure 8c), la perte de l'extrémité anomérique (fragment à m/z = 162) est plus intense que lors de l'irradiation avec uniquement le laser Firefly. Le pourcentage de photofragmentation passe de 19 % pour une irradiation au laser Firefly seul à 38 % pour une irradiation avec les deux lasers ce qui représente le double.

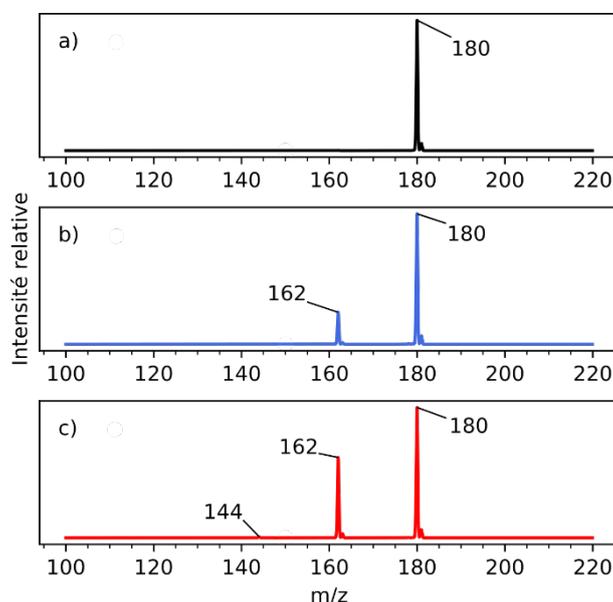


Figure 8 : Spectre de masse de la glucosamine protonée ( $m/z = 180$ ) avec irradiation (30 ms) par le laser CO<sub>2</sub> seul a), le laser Firefly à 3622 cm<sup>-1</sup> seul b) ou de la combinaison de ces deux lasers c). La puissance du laser CO<sub>2</sub> est choisie de manière à ce que le rendement de photofragmentation soit nul sous irradiation avec ce laser. Le pourcentage de photofragmentation sous irradiation avec le laser à haute cadence est de 19 %. Ce pourcentage double (38 %) en cas d'irradiation simultanée des deux lasers.

#### 2.2.5 Application à des espèces lithiées, sodées et déprotonées

L'ajout dans l'échantillon d'un cation métallique pour favoriser son ionisation est courant lors d'analyses en spectrométrie de masse. En plus de favoriser l'ionisation, l'ajout de lithium (Li<sup>+</sup>) augmente le seuil de dissociation des espèces ce qui offre l'accès à des signatures de fragmentation différentes que celles retrouvées dans le cas d'espèces protonées. Dans le cas des sucres, les espèces lithiées vont se fragmenter en rompant des liaisons intracycles, ce qui offre des informations structurales très intéressantes sur la position des groupements fonctionnels ainsi que des liaisons entre les sucres<sup>10,11</sup>. Dans cette partie, nous avons étudié trois hexoses : galactose, glucose et mannose (structures sur la gauche de la Figure 9) qui présentent deux épiméries (en C4 entre Gal et Glc et en C2 pour le Glc et le Man). Les spectres de fragmentation présentent les mêmes fragments à m/z = 169, 127 et 109 et dans une moindre mesure m/z = 157, 139 et 151 (Figure 9a, c et e), la fragmentation en

CID ne permet pas de les distinguer l'un de l'autre. La spectroscopie IRMPD a montré son intérêt dans l'analyse de sucres présentant de tels adduits<sup>12,13</sup>. Ici nous avons étudié chacun de ces hexoses avec notre montage expérimental. Les hexoses sont préparés à 50  $\mu\text{M}$  dans une solution d'eau méthanol (50:50 v/v) avec 50  $\mu\text{M}$  de chlorure de lithium. Les hexoses lithiés ( $m/z = 187$ ) sont isolés puis irradiés par les deux lasers simultanément. Les spectres IRMPD sont présentés en Figure 9b, d et f. Dans les trois cas, les spectres présentent une absorption maximale dans la région des O-H. Le reste du spectre est peu fourni. L'absorption maximale est centrée autour de 3630  $\text{cm}^{-1}$  dans les trois cas : 3626  $\text{cm}^{-1}$  pour le glucose, 3638  $\text{cm}^{-1}$  pour le galactose et 3632  $\text{cm}^{-1}$  pour le mannose. Le glucose présente une faible bande d'absorption (3564  $\text{cm}^{-1}$ ) peu résolue de la bande d'absorption maximale (3626  $\text{cm}^{-1}$ ) tandis que le galactose présente une faible bande d'absorption relativement étendue dans la gamme des C-H (maximum à 2965  $\text{cm}^{-1}$ ). Le mannose lui ne présente ni l'une ni l'autre de ces bandes secondaires caractéristiques. Ce sont ces caractéristiques (présence ou non d'un épaulement sur le pic principal et présence ou non de la faible bande d'absorption dans la gamme des C-H qui permet de discriminer ces espèces.

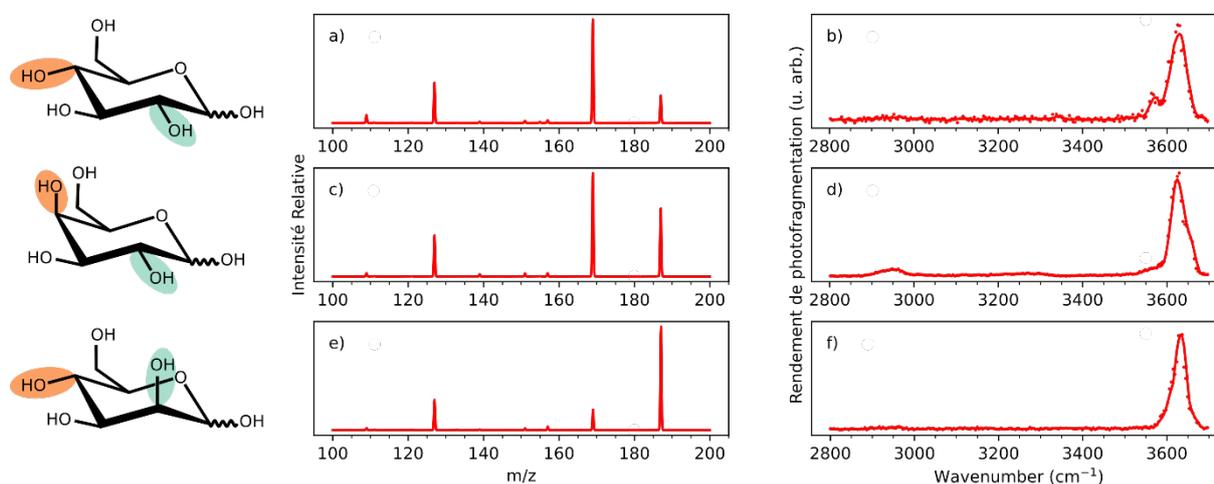


Figure 9 : Spectres de fragmentation par CID de trois hexoses lithiés ( $[M+Li]^+$   $m/z = 187$  : glucose a), galactose c) et mannose f) et spectres IRMPD de ces mêmes espèces : glucose b), galactose d) et mannose f). Les structures correspondantes sont présentées sur la gauche de la figure. Les spectres de fragmentation par CID de ces isomères sont similaires, les hexoses ne peuvent pas être discriminés par ce moyen. En revanche, des bandes d'absorption ou des épaulements apparaissent sur les spectres IRMPD de ces hexoses ce qui permet de les discriminer.

Comme nous venons de le voir, l'ajout d'un cation métallique est intéressant notamment dans le cas où les composés ne se protonent pas facilement et pour discriminer des isomères par spectroscopie IRMPD.

Il arrive cependant que l'adduit ne soit pas maîtrisé, en particulier dans le cas d'échantillons biologiques. C'est le cas par exemple avec le sodium qui est très souvent rencontré en spectrométrie de masse. Dans notre exemple, le 3'SL (3'sialyllactose) est étudié. Il s'agit d'un trisaccharide contenant un acide N-acétylneuraminique, un glucose et un galactose. L'acide N-acétylneuraminique possède un groupement carboxyle. Ce groupement peut se protoner pour donner un groupement carboxyl mais il peut également contenir un sodium. Ainsi sur un spectre de masse, nous observons trois signaux pour ce composé  $m/z = 634$  le groupement carboxyl est protoné et la charge est portée par un proton, dans le cas du signal à  $m/z = 656$  un proton et un sodium sont lié au composé, finalement, le signal du 3'SL à  $m/z = 678$  porte un sodium sur son carboxylate et est chargé au moyen d'un autre sodium. La spectrométrie de masse indique que dans le cas de l'espèce à  $m/z = 656$ , le groupement carboxyle est

protoné et un sodium apporte sa charge à l'espèce. En effet sur le spectre de fragmentation (non montré), nous observons un fragment à  $m/z = 365$  ce qui correspond à deux hexoses avec un adduit sodium, le sodium n'est donc pas sur le carboxyl, nous verrons que la spectroscopie IRMPD confirme cette affirmation.

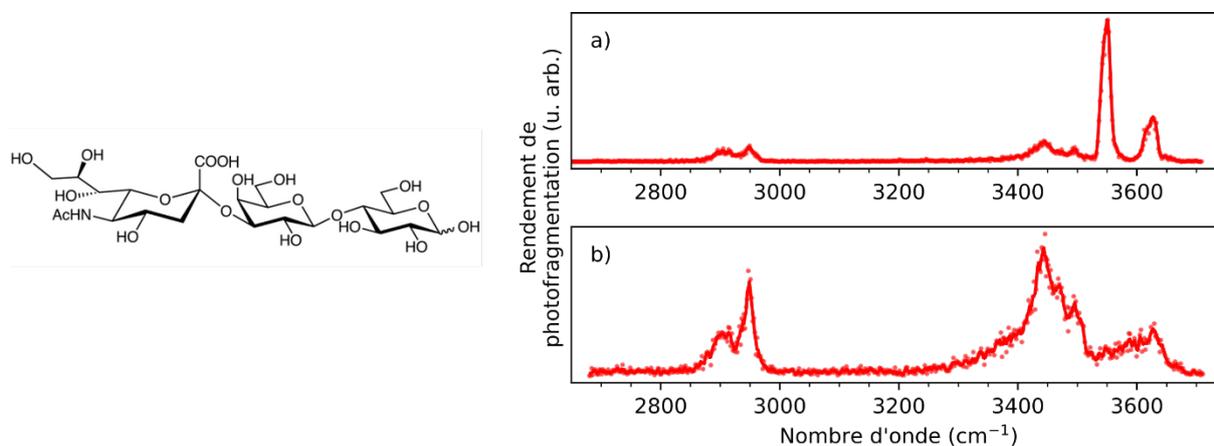


Figure 10 : Spectres IRMPD du 3'SL sur les espèces à  $m/z = 656$  (le 3'SL porte à la fois un proton et un ion sodium, la présence d'une bande à  $3550\text{ cm}^{-1}$  montre que le proton est situé sur l'acide carboxylique ce qui implique que le porteur de charge est un sodium) a) et  $m/z = 678$  (le 3'SL porte deux ions sodium) b), la structure (non chargée portant un proton sur son groupement carboxyle) du composé est présenté sur la gauche.

Le spectre IRMPD de l'espèce à  $m/z = 656$  (Figure 10 a) présente un maximum d'absorption à  $3550\text{ cm}^{-1}$ , dans la gamme des OH, une autre bande est observée à  $3626\text{ cm}^{-1}$  deux bandes sont présentes dans la gamme des CH : entre  $2820$  et  $2970\text{ cm}^{-1}$  ainsi que deux autres bandes vers  $3444$  et  $3493\text{ cm}^{-1}$ . La bande à  $3550\text{ cm}^{-1}$  est caractéristique du proton porté sur un acide carboxylique, ce qui confirme le fait que ce groupement est occupé par un proton. Ces résultats sont cohérents avec les mesures effectuées par Depraz Depland *et al.*<sup>14</sup> : dans leur analyse du 3'SL, les auteurs relevaient des bandes situées dans la bande des C-H entre  $2820$  et  $3000\text{ cm}^{-1}$ , une large bande présentant l'absorption maximale centrée autour de  $3439\text{ cm}^{-1}$  une bande à  $3550\text{ cm}^{-1}$  ainsi que deux autres bandes à  $3598$  et  $3645\text{ cm}^{-1}$ . Dans notre cas, la différence de résolution peut expliquer que les deux bandes à  $3598$  et  $3645\text{ cm}^{-1}$  soient confondues en une seule. Les processus multiphotoniques sont très sensibles aux différences de puissance, ainsi une variation même légère de puissance peut engendrer une différence significative sur l'intensité d'absorption d'une bande, c'est ce qui a dû se passer quand on observe l'inversion de maxima entre la mesure au moyen du couplage (Figure 10 a) et celle obtenue à l'OPO<sup>14</sup>

L'espèce à  $m/z = 678$  comporte deux adduits sodium : un sodium qui agit en contre ion sur le carboxyle et un sodium porteur de charge. Ici on n'observe pas de bande à  $3550\text{ cm}^{-1}$ , le proton est remplacé par un sodium. Le maximum d'absorption est présent à  $3430\text{ cm}^{-1}$  sur une large bande présentant un épaulement à  $3495\text{ cm}^{-1}$ . Deux pics sont présents dans la gamme des C-H :  $2910$  et  $2949\text{ cm}^{-1}$ . Pour finir une large bande entre  $3549$  et  $3653\text{ cm}^{-1}$  est aussi présente sur ce spectre.

Les sucres acides perdent facilement un proton et sont donc facilement observables en mode négatif. Dans ce dernier exemple, les spectres IRMPD de l'acide glucuronique déprotoné (glucuronate) et celui de l'acide iduronique déprotoné (iduronate) ont été acquis au moyen du nouveau montage. Ces spectres sont comparés avec les spectres obtenus au moyen du montage original avec l'OPO (Figure 11).

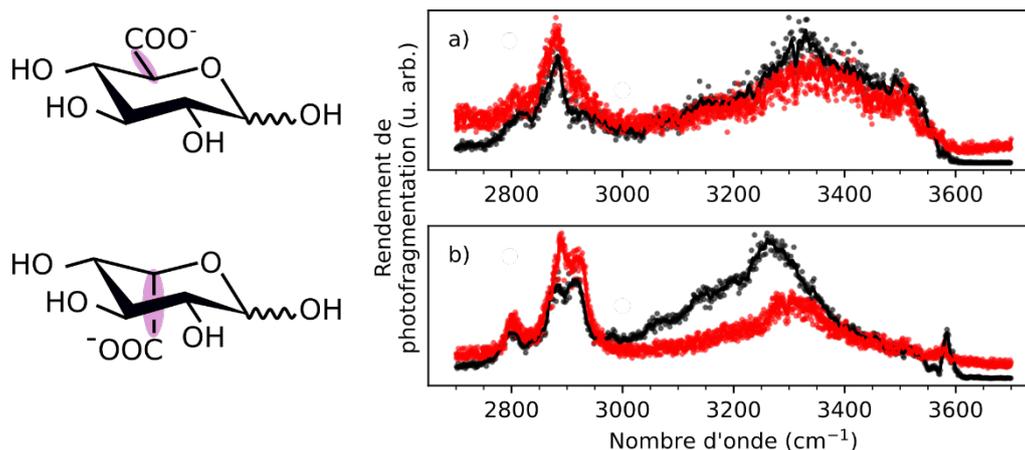


Figure 11 : Spectres IRMPD (espèces  $[M-H]^-$ ) du glucuronate a) et de l'iduronate b). Les spectres ont été acquis au moyen du montage avec l'OPO (noir) et avec le couplage laser  $CO_2$  et laser haute cadence (rouge). Les structures de ces deux composés sont visibles sur la gauche de la figure. Dans les deux cas, l'identification des composés à partir de leur spectre IRMPD est non ambiguë.

L'acquisition de ces spectres n'est pas possible au moyen du seul laser accordable, aucune photofragmentation n'est mesurée. L'ajout du laser  $CO_2$  a permis l'acquisition des spectres IRMPD de ces deux hexuronates. Le glucuronate (Figure 11 a) présente deux bandes d'absorption dans la région des C-H 2808 et 2884  $cm^{-1}$ . De plus une large bande d'absorption est présente à 3333  $cm^{-1}$  ainsi qu'un épaulement à 3510  $cm^{-1}$ . Ce spectre présente un bon accord avec le spectre acquis au moyen de l'OPO, les deux courbes rouge et noir se recouvre bien. Concernant l'iduronate Figure 11 b) le spectre IRMPD acquis au moyen de l'OPO et celui acquis avec le nouveau montage présente bien la même empreinte dans la région des C-H. On observe une bande à 2798  $cm^{-1}$  ainsi qu'un doublet à 2887 et 2920  $cm^{-1}$ . Toutefois, une large bande est observée avec un maximum à 3300  $cm^{-1}$  contre 3265  $cm^{-1}$  au moyen de l'OPO. De plus la bande observée à 3580  $cm^{-1}$  avec l'OPO se détache à peine du bruit de fond lors de l'acquisition au moyen du Firefly et du laser  $CO_2$ . Ces différences seraient très intéressantes à approfondir pour une meilleure connaissance des processus ayant lieu lors d'analyses IRMPD. Mais dans une démarche analytique, le fait d'obtenir des signatures répétables et diagnostiques est suffisant.

### 3 RECONNAISSANCE DE SPECTRES

L'acquisition de spectres IRMPD s'est grandement accélérée. Les premiers spectres nécessitaient environ 1 h d'acquisition. Actuellement, les récents développements ont fait tomber cette durée à seulement quelques dizaines de secondes<sup>15</sup>. Les quantités de données obtenues nécessitent d'être interprétées dans le but d'obtenir une identification pour les composés analysés. La première approche, la plus simple est de les inspecter visuellement. Cette méthode nécessite du temps, un personnel qualifié, et les analyses sont dépendantes de l'expérimentateur. Plus récemment, différentes méthodes ont été développées pour interpréter de telles données. Les critères importants sont la fiabilité, les composés doivent être correctement identifiés ; la robustesse : même en changeant d'appareil l'identification doit demeurer la même ; et la rapidité : l'identification doit être la plus rapide possible. Dans cette partie nous verrons trois approches différentes que nous pouvons effectuer afin d'identifier un composé inconnu. Il s'agit du calcul du score de ressemblance entre un spectre de référence présent dans une base de données et le spectre de l'espèce inconnue, d'une analyse en composantes principales (ACP) et enfin de l'utilisation de l'intelligence artificielle basée sur un réseau de neurones.

Il est important de noter que les spectres que nous traçons et qui sont présentés dans ce manuscrit comportent une courbe qui représente un lissage des points afin de faciliter la lecture du spectre. En revanche, à aucun moment, les identifications de spectres par les méthodes présentées ci-dessous ne font intervenir de lissage. Seuls les points bruts sont interprétés dans ces identifications.

### 3.1 APPROCHE PAR LES SCORES

#### 3.1.1 Calcul de la ressemblance

Une façon relativement simple d'identifier un composé à partir d'un spectre est de constituer une base de données. Cette base de données permettra le calcul d'un score de ressemblance entre le spectre inconnu et les spectres présents dans la base de données. Pour mener à bien un tel calcul, il est nécessaire de normaliser les données (la puissance laser peut varier d'un jour à l'autre et donc le rendement de photofragmentation peut être affecté). Les données sont aussi interpolées afin de faire correspondre les nombres d'onde entre spectre de référence et spectre inconnu. Le calcul des scores se fait selon l'équation n°1 décrite dans l'article<sup>16</sup>.

$$score = 999 \times \left( 1 - \sqrt{\frac{\sum_i^n (I_{\bar{\nu}_i}^{ref} - I_{\bar{\nu}_i}^{exp})^2}{n}} \right) \quad (1)$$

Ce calcul de score se base sur la différence de rendement de photofragmentation entre une référence (présente dans la base de données)  $I_{\bar{\nu}_i}^{ref}$  et une espèce inconnue  $I_{\bar{\nu}_i}^{exp}$  à un nombre d'onde donné  $\bar{\nu}_i$  avec  $n$  est le nombre de points comparés.

Dans ce calcul, chaque rendement de photofragmentation de l'échantillon inconnu est comparé avec le rendement de photofragmentation de la référence. Plus le score est élevé, plus les spectres se ressemblent. Un regard d'expert pour confirmer l'identification proposée par le programme est toujours indispensable. Cette approche a le mérite d'être automatisée, mais elle demeure biaisée, en effet un seul spectre est choisi arbitrairement et est utilisé pour l'identification de spectres inconnus futurs.

Le programme de calcul des scores est également équipé d'un filtre. Ainsi le calcul des scores ne se fait que sur les standards de même  $m/z$  ce qui diminue grandement le nombre de résultats (simplification de l'affichage) et surtout diminue le temps de calcul, il ne se fait que sur quelques entrées de la base de données et non sur son intégralité.

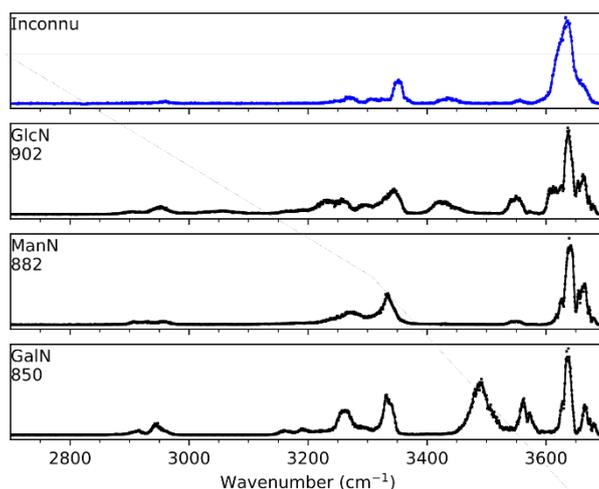


Figure 12 : Identification d'un composé inconnu ( $m/z = 180$ ) à partir de son spectre IRMPD (spectre de l'échantillon inconnu bleu) spectres IRMPD des références (noir). Le nom des références de même  $m/z$  et le score correspondant sont indiqués en haut à gauche des spectres.

Le spectre IRMPD de l'échantillon inconnu a été acquis au moyen du montage « rapide », les références ont, elles été acquises au moyen du montage original. Le spectre IRMPD du composé inconnu présente un large pic aux alentours de  $3622 \text{ cm}^{-1}$  avec un épaulement à  $3656 \text{ cm}^{-1}$ . Une faible bande d'absorption est présente à  $3545 \text{ cm}^{-1}$  ainsi qu'une à  $3425 \text{ cm}^{-1}$ . Un massif se détache entre  $3200 \text{ cm}^{-1}$  et  $3400 \text{ cm}^{-1}$  avec un minimum d'absorption à  $3280 \text{ cm}^{-1}$  et deux maxima l'un à  $3344 \text{ cm}^{-1}$  et l'autre à  $3260 \text{ cm}^{-1}$ . Un dernier pic se détache dans la zone des C-H avec une faible absorption autour de  $2920 \text{ cm}^{-1}$ . Les spectres de référence ont été obtenus au moyen du montage LCQ-OPO/OPA développé dans l'équipe. Les pics des spectres IRMPD de référence sont plus fins que le spectre du composé inconnu. Le programme indique que le composé inconnu correspond à de la glucosamine avec un score de 902, contre 882 et 850 pour les deux autres propositions. L'écart de score entre la première proposition (glucosamine) et la deuxième (mannosamine) n'est pas très important. Il peut être intéressant de regarder plus en détail les spectres pour confirmer cette identification. Ainsi, la glucosamine est le seul spectre à présenter une bande d'absorption à  $3425 \text{ cm}^{-1}$  bande retrouvée sur le spectre du composé inconnu. De plus, le pic d'absorption à  $3260 \text{ cm}^{-1}$  est décalé à  $3280 \text{ cm}^{-1}$  dans le cas du mannosamine. Finalement, le spectre de la galactosamine présente un maximum d'absorption aux alentours de  $3490 \text{ cm}^{-1}$ , bande qui est absente du spectre du composé inconnu.

### 3.1.2 Calcul de l'erreur

Plus récemment, dans l'équipe une autre manière d'identifier un composé inconnu a été envisagé, il s'agit d'un calcul d'erreur (voir équation n°2). Ce calcul a été mis en œuvre pour identifier un hexose après fragmentation d'un disaccharide<sup>17,18</sup>.

$$erreur = \frac{100}{3} \sum (I_{\bar{\nu}i}^{ref} - I_{\bar{\nu}i}^{exp})^2 \quad (2)$$

Dans cette équation,  $I_{\bar{\nu}i}^{ref}$  représente la valeur du rendement de photofragmentation au nombre d'onde  $\bar{\nu}i$  du spectre de référence (celui entré dans la base de données) et  $I_{\bar{\nu}i}^{exp}$  représente celle du rendement de photofragmentation au même nombre d'onde du spectre du composé inconnu. Ici, le calcul ne tient pas compte du nombre de points acquis ( $n$  dans l'équation n°1), lors de la comparaison entre deux calculs d'erreur, il est alors important de s'assurer que le nombre de points acquis soit le même (nombres d'onde pour lesquels un rendement de photofragmentation a été mesuré).

Ces identifications par des mesures de ressemblance ou d'erreur sont très intéressantes et sont utilisées en routine au sein de l'équipe car elles sont simples à mettre en place, elles souffrent toutefois d'un biais, le spectre utilisé en tant que référence est unique, il a été choisi car il représentait le mieux l'espèce analysée. Ce choix est le choix de l'expérimentateur, il n'est pas forcément le plus représentatif du composé, notamment dans la mesure où le spectre de celui-ci peut être influencé par des paramètres qui varient au jour le jour. C'est pourquoi d'autres techniques d'identification de spectres sont testées, ces techniques se servent de plusieurs spectres de référence pour procéder à l'identification du composé inconnu.

## 3.2 APPROCHE PAR ACP

### 3.2.1 La méthode

L'ACP (analyse en composantes principales) est une méthode d'analyse multivariée utilisée pour résumer l'information présente dans un grand nombre de variables<sup>19</sup>. L'ACP est une méthode non supervisée, les données ne sont pas étiquetées. Cette méthode est couramment utilisée en chimométrie par exemple en conjonction avec la spectroscopie IR en phase gazeuse pour l'identification de métabolites<sup>20</sup> mais aussi de sucres<sup>3,21,22</sup>.

Dans le cas de spectres IRMPD,  $n$  individus (échantillons) sont décrits par  $p$  variables (rendement de photofragmentation en fonction du nombre d'onde). La matrice décrivant ces individus est donc composée de  $n$  lignes et  $p$  colonnes. Dans le cas où  $p$  est petit, il est possible de représenter un nuage de points. Par exemple si  $p = 2$ , les individus seront représentés sur deux axes : un plan, dans le cas où  $p = 3$  les individus seront représentés dans un repère en trois dimensions. La représentation se complique dans le cas où  $p$  dépasse 3, il est alors impossible de visualiser une telle représentation. Dans notre cas, les spectres IRMPD,  $p$  est de l'ordre de plusieurs centaines de variables (environ un millier de nombres d'onde). Les individus devraient être représentés sur environ 1000 axes.

L'analyse en composantes principales vise alors à compresser l'information obtenue dans l'ensemble des variables  $p$  dans un sous espace de dimensions  $k$ . La représentation doit être la meilleure possible tout en perdant le moins d'information possible. Ainsi  $k$  représentent les nouvelles variables caractérisant les individus aussi appelées « composantes principales ». Il est intéressant de centrer réduire les variables (en particulier si les unités ne sont pas les mêmes entre variables, cela permet de ne pas donner plus de poids à une variable en particulier), pour cela, on soustrait la valeur moyenne à chacune des variables et on divise par l'écart type. L'inertie totale est égale à la somme des variances des variables étudiées. Afin de représenter au mieux les données sur un petit nombre de composantes principales, l'inertie doit être la plus grande possible sur les premiers axes. L'ACP construit donc de nouvelles variables en s'appuyant sur la matrice de variance-covariance.

### 3.2.2 Application sur les hexosamines

Dans cet exemple, les mêmes trois hexosamines : GlcN, GalN et ManN (Glucosamine, Galactosamine et Mannosamine respectivement) sont considérées. Cette fois-ci, leur spectre IRMPD est obtenu au moyen du nouveau montage obtenu avec le laser à haute cadence. Le spectre de chacune de ces espèces est acquis trois fois, afin d'augmenter le nombre d'individus.

Les hexosamines sont préparées à 50  $\mu$ M dans un mélange eau/MeOH 50:50 v/v avec 0.1 % d'acide acétique pour favoriser la protonation des espèces. Les spectres sont acquis en environ 10 minutes. Les spectres des trois hexosamines sont acquis trois fois afin de procéder à l'ACP ensuite, le spectre d'une hexosamine inconnue est ensuite intégré au jeu de données afin d'identifier l'inconnu. Les spectres obtenus sont présentés ci-dessous (Figure 13). Visuellement ces spectres sont bien différents

selon les espèces considérées. Brièvement, les trois spectres présentent des pics d'absorption dans les trois régions caractéristiques de la gamme des 3  $\mu\text{m}$  (C-H, N-H et O-H). Ces trois spectres ont un maximum d'absorption dans la région des 3600  $\text{cm}^{-1}$ , un massif autour de 3300  $\text{cm}^{-1}$  et une signature non discriminante en dessous de 3000  $\text{cm}^{-1}$  dans la région caractéristique des C-H. Les différences entre ces espèces sont, la GalN (Figure 13b) présente une large bande d'absorption à 3488  $\text{cm}^{-1}$  (attribuée à l'hydroxyle  $n^{\circ}4^{23}$ ) absente sur les spectres des deux autres hexosamines. La GlcN (Figure 13a) elle présente une légère bande d'absorption à 3430  $\text{cm}^{-1}$  (attribuée à l'hydroxyle  $n^{\circ}3^{23}$ ) qui est elle aussi absente des deux autres spectres. Il est intéressant de voir sur ces spectres que les différentes acquisitions ne se superposent pas parfaitement. En effet, il est facile de voir sur le spectre de la GlcN des différences d'absorption (au niveau des régions autour de 3300  $\text{cm}^{-1}$  et 3600  $\text{cm}^{-1}$ ), de même que sur celui de la ManN, la puissance du laser peut avoir changée, ou le laser être mieux aligné dans le piège à ions dans un cas que dans l'autre. Dans le cas de la GalN, un spectre présente une ligne de base qui ne revient pas à zéro, nous remédions à cela en appliquant une correction de la ligne de base. Tous ces paramètres doivent être considérés si on veut identifier un composé à partir de la comparaison avec un spectre présent dans une base de données. Si maintenant, un seul spectre est entré dans une base de données pour servir de référence, dans le but d'identifier un composé inconnu il peut y avoir un biais car on ne tient pas compte de la variabilité de ces paramètres.

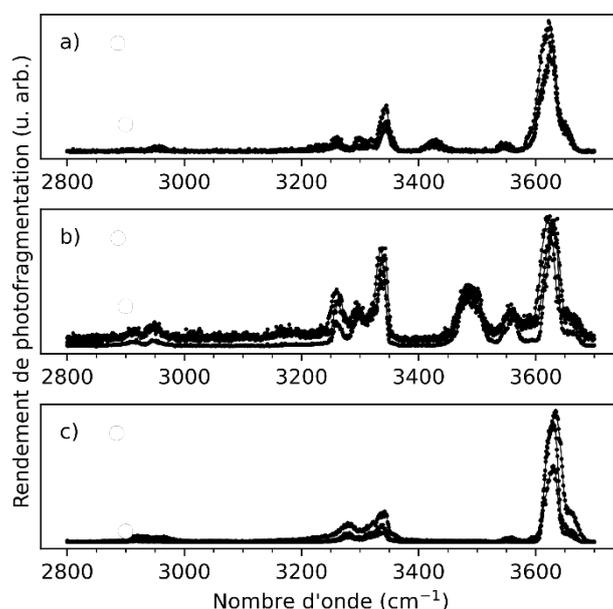


Figure 13 : Spectres IRMPD de trois hexosamines protonées ( $m/z = 180$ ) GlcN a), GalN b) et ManN c) obtenus à des jours différents, différents facteurs peuvent jouer sur le reproductibilité. Visuellement, les spectres IRMPD de ces trois hexosamines sont différenciables, notamment dans la gamme comprise entre 3400 et 3600  $\text{cm}^{-1}$ .

### 3.2.3 L'analyse en ACP

Tout d'abord, Les spectres sont interpolés de manière à observer un point pour chaque 1  $\text{cm}^{-1}$  ainsi, ils sont tous décrits par les mêmes variables. Les intensités sont normalisées et standardisées avant de procéder à l'analyse. Ainsi en cas de changement dans les conditions lasers avec un rendement de photofragmentation plus ou moins important, un individu n'aura pas plus de poids qu'un autre lors de l'analyse en ACP. Comme observé sur la Figure 14, les deux premières composantes expliquent presque 68 % de la variance contenue dans les données. Ici on ne considère ces deux composantes.

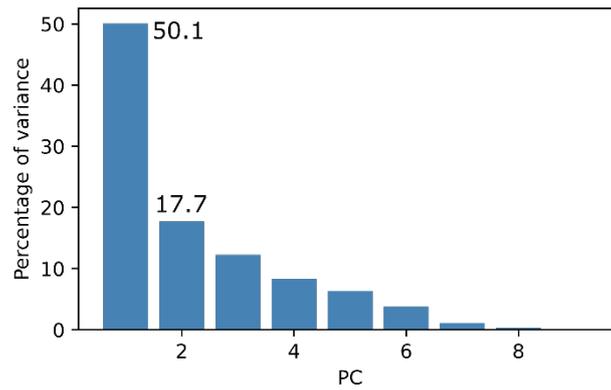


Figure 14 : Contribution des composantes principales pour expliquer la variance dans le jeu de données. Les deux premières composantes expliquent près de 68 % de la variance contenues dans les données.

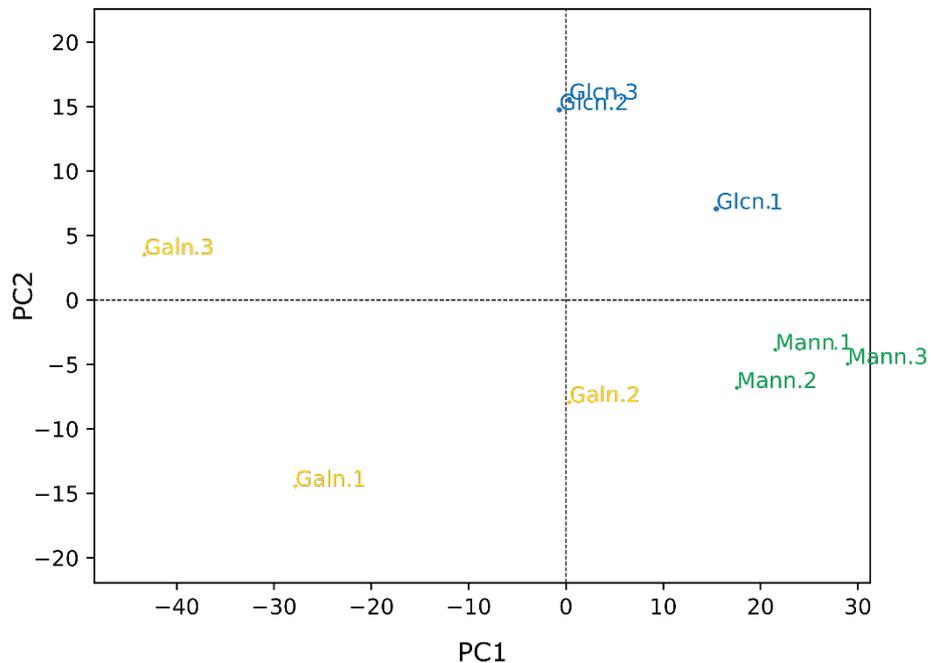


Figure 15 : Représentation des scores selon les composantes principales 1 et 2. Les trois hexosamines différentes sont bien séparées selon cette représentation. Les mannosamines sont les monosaccharides les mieux regroupés, les glucosamines sont plus espacées et les galactosamines sont les espèces les plus dispersées dans le plan PC1/PC2.

La Figure 15 représente le tracé des scores selon les axes 1 et 2 (près de 68 % de la variance contenue dans les données). Dans ce plan, les trois composés (GalN en jaune, GlcN en bleu et ManN en vert acquis en triplicats) forment des groupes bien distincts. On note que les individus sont plus ou moins dispersés en fonction de la classe de composés. Afin de pouvoir comparer la projection d'un individu inconnu avec les groupes formés par les différentes classes d'hexosamine, il est nécessaire de trouver un point de comparaison, pour cela, nous utiliserons le barycentre de chacune des classes de ces composés. Le barycentre des différents groupes est calculé selon la formule suivante :

$$x_{barycentre} = \frac{\sum x_i}{n}, y_{barycentre} = \frac{\sum y_i}{n} \quad (3)$$

Dans le cas où on considère le plan Oxy (dans notre cas le plan formé par les deux premières composantes). Ces barycentres sont projetés sur la Figure 16.

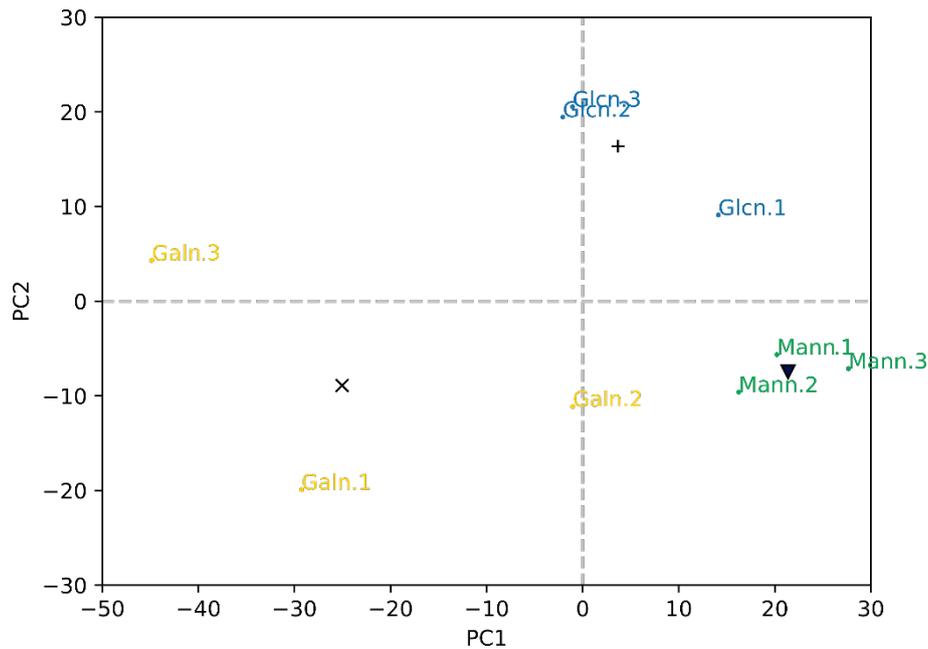


Figure 16 : Carte des individus présentant les barycentres des différents groupes le symbole « x » représente le barycentre des galactosamines, « + » représente celui des glucosamines tandis que le triangle plein représente celui des mannosamines.

Les barycentres sont représentés par des symboles géométriques.

Le composé inconnu subit la même transformation que les données des individus lors de l'analyse en ACP afin de pouvoir le projeter sur le même graphique (Figure 17). Le composé inconnu semble être plus proche spatialement des glucosamines.

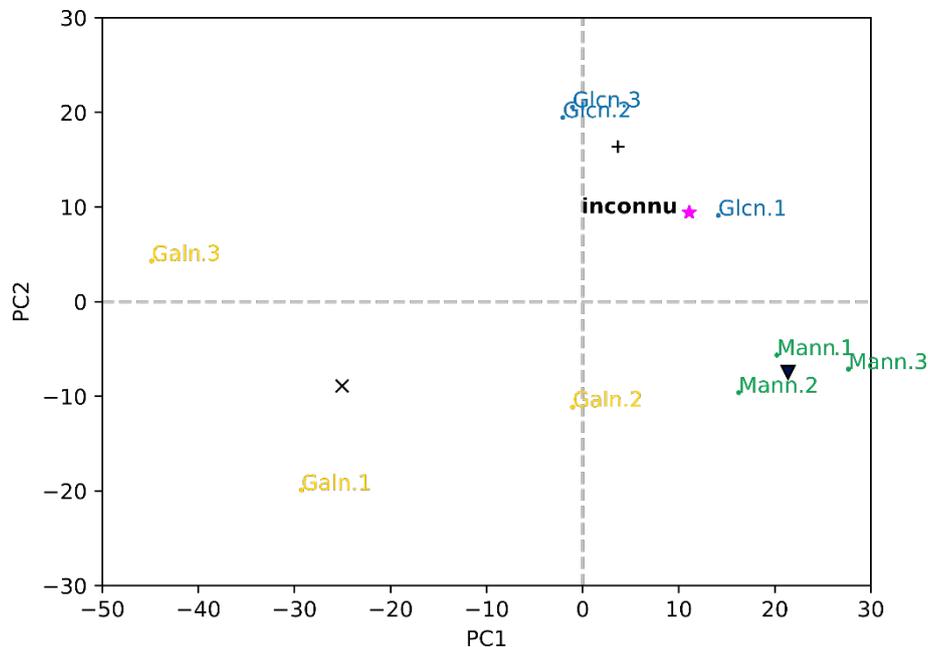


Figure 17 : Carte des individus présentant les différents groupes, les différents barycentres et la projection d'un échantillon inconnu. « x » représente le barycentre des galactosamines, Le symbole « + » représente celui des glucosamines tandis que le triangle plein représente celui des mannosamines. Le composé inconnu est représenté par une étoile rose.

Pour se faire une idée de la similarité entre le composé inconnu et chacun des composés de référence, la distance euclidienne entre la projection du composé inconnu et le barycentre des différentes classes d'hexosamines est calculée de la manière suivante :

$$Distance\ euclidienne = \sqrt{(x_i - x_{barycentre})^2 + (y_i - y_{barycentre})^2} \quad (4)$$

Avec  $x_i$  et  $y_i$  les coordonnées du point  $i$  et  $x_{barycentre}$  et  $y_{barycentre}$  les coordonnées du barycentre (toujours dans le plan Oxy).

Dans le cas présenté, l'identification n'est pas très ambiguë, les composés sont bien regroupés par famille, et le composé inconnu est plus près du groupe glucosamine que des autres. Les distances euclidiennes sont représentées en Figure 18. La distance euclidienne entre le composé inconnu et le barycentre de GlcN est la plus faible des distances euclidiennes (10,16 contre 40,52 et 19,77 pour respectivement les distances entre le composé inconnu et le barycentre de GlcN, entre le composé inconnu et le barycentre de GalN et entre le composé inconnu et le barycentre de ManN).

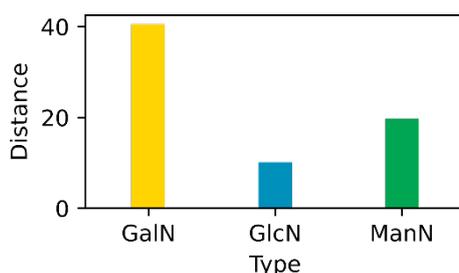


Figure 18 : Distance euclidienne du composé inconnu par rapport au barycentre des différentes hexosamines. La distance est minimale dans le cas de la glucosamine.

### 3.2.4 Confirmation de l'identification

Afin de confirmer cette identification, les spectres IRMPD des différentes espèces sont tracés et inspectés par l'expérimentateur un exemple est présenté en Figure 19. Visuellement, le composé ne correspond pas à de la galactosamine les différences principales sont le léger décalage entre le pic à  $3622\text{ cm}^{-1}$  de l'échantillon inconnu le pic de la galactosamine à  $3630\text{ cm}^{-1}$ , mais surtout la présence du pic d'absorption à  $3485\text{ cm}^{-1}$  pour l'échantillon inconnu. La galactosamine ne semble pas absorber à  $3430\text{ cm}^{-1}$  contrairement à l'échantillon inconnu. Le spectre de l'échantillon inconnu recouvre presque parfaitement le spectre de la glucosamine, seule une légère différence d'intensité sur la bande à  $3344\text{ cm}^{-1}$  est présente. Le spectre de la mannosamine présente un pic dans la région des OH décalé par rapport au pic à  $3622\text{ cm}^{-1}$  du composé inconnu, mais surtout, la mannosamine n'absorbe pas à  $3430\text{ cm}^{-1}$ . Au premier coup d'œil, il est possible de confirmer l'identification par l'ACP à savoir que le composé est de la glucosamine.

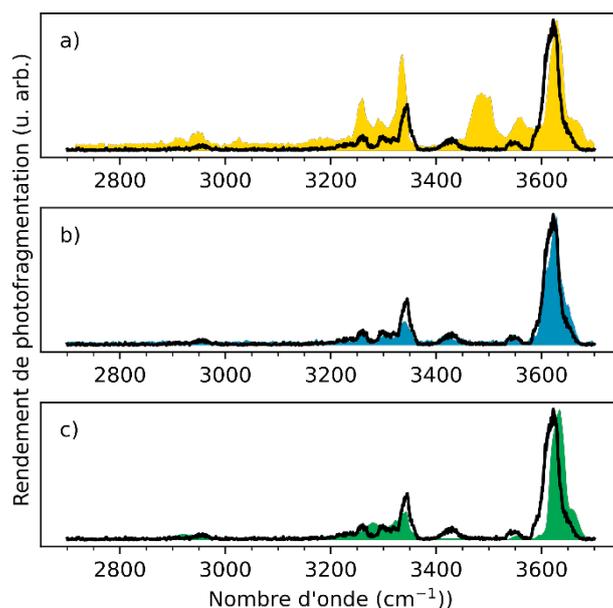


Figure 19 : tracé du spectre IRMPD du composé inconnu :  $m/z = 180$  (ligne noire) et des spectres des différentes hexosamines  $m/z = 180$  (fond coloré) galactosamine a), glucosamine b) et mannosamine c).

Pour aller un peu plus loin, le score de similarité est calculé (Equation 1). La Figure 20 représente les scores de similarité retrouvés en comparant le composé inconnu avec les trois hexosamines de référence. Il ressort qu'en moyenne, le composé inconnu ressemble plus à la glucosamine. Comme vu lors de l'examen manuel des spectres, visuellement et par le calcul, le spectre de la mannosamine ressemble plus au spectre du composé inconnu que le spectre de la galactosamine.

Comme nous avons pu le voir dans cet exemple, il est intéressant de disposer de plusieurs spectres de référence afin de s'assurer qu'un spectre aberrant n'est pas présent lors de l'identification d'un composé inconnu.

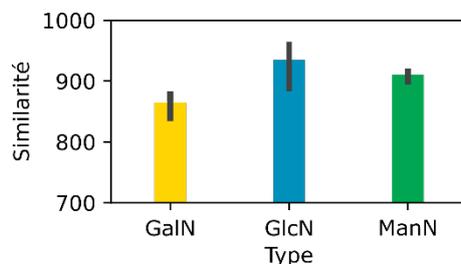


Figure 20 : Score de similarité entre le composé inconnu et les trois hexosamines investiguées.

En observant les différences entre les spectres, l'expérimentateur confirme que le spectre inconnu correspond à de la glucosamine, identification confirmée par le score de similarité.

Ici de premières identifications ont été menées au moyen de l'ACP, il serait intéressant d'aller plus loin, en introduisant plus de spectres pour une même classe de sucres afin d'effectuer l'identification. Il serait également intéressant de vérifier comment les identifications se font lorsque le spectre du composé inconnu est acquis au moyen d'un autre montage expérimental que le montage avec lequel les standards ont été acquis.

Dans le cas de spectres d'hexosamines, l'analyse en composantes principales identifie le bon composé. Une prochaine étape serait également de tenter l'identification de nouvelles espèces pour confirmer l'utilité d'une telle méthode pour l'identification des espèces inconnues d'intérêt.

Une autre approche actuelle utilisée dans l'identification de composés est l'intelligence artificielle. Nous avons exploré cette approche en collaboration avec des partenaires.

### 3.3 APPROCHE PAR APPRENTISSAGE MACHINE

Dans cette partie, nous détaillerons l'interprétation de spectres IRMPD au moyen d'une approche par l'apprentissage machine dans le but de faciliter l'identification de monosaccharides. Un article « **GIAlcomics: a deep neural network classifier for Spectroscopy-Augmented Mass Spectrometric glycans data** » par Thomas Barillot, Baptiste Schindler, Baptiste Moge, Elisa Fadda, Franck Lépine et Isabelle Compagnon publié dans le journal *Beilstein Journal of Organic Chemistry* en 2023.

L'utilisation de l'IA (intelligence artificielle) a été imaginée pour interpréter des données de spectroscopie vibrationnelle dès le début des années 90<sup>24</sup>. L'utilisation de ces méthodes dans le cas de spectres de masse remonte elle au milieu des années 2000<sup>25</sup>. La classification de tissus cancéreux peut être effectuée en se servant de l'intelligence artificielle sur des données de spectroscopie vibrationnelle<sup>26</sup>. Depuis des efforts sont faits pour simuler des structures moléculaires ou des spectres vibrationnels avec de telles approches<sup>27,28</sup>. A ce jour, une seule étude montre l'utilisation de l'intelligence artificielle dans le cadre de l'analyse de sucres par spectroscopie vibrationnelle en phase gazeuse<sup>29</sup>, par un modèle d'IA appelé Random Forest.

#### 3.3.1 Production des données

Les données ont été obtenues au moyen du couplage OPO/A-LCQ<sup>12</sup>. Trois jeux de données différents ont été acquis. Le premier jeu de données contient les trois épimères d'hexosamines précédemment étudiés (GlcN, GalN et ManN) ainsi que le GlcNAc. En tout ce jeu de données contient trente-trois spectres de ces composés. La Figure 21 présente un exemple de spectres IRMPD de chacun des composés contenus dans le jeu de données. Ce premier jeu de données sert pour l'entraînement et la validation du modèle.

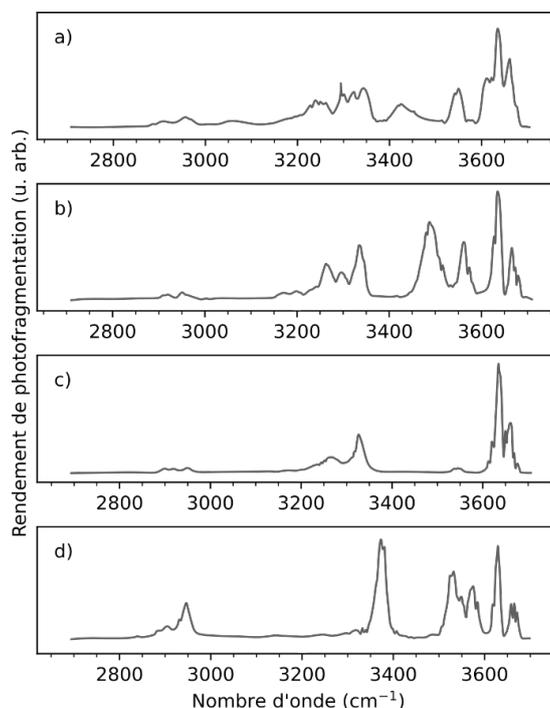


Figure 21 : Spectres IRMPD de différentes espèces protonées ( $m/z = 180$ ) GlcN a), GalN b), ManN c) et GlcNAc d). Ces spectres sont dans le premier jeu de données acquis au moyen du montage laser OPO/OPA-LCQ. Ce jeu de données est utilisé afin d'entraîner et de valider le modèle.

Le second jeu de données apporte de l'hétérogénéité. L'appareil utilisé pour acquérir les données est celui couplant le Firefly au LTQ-XL<sup>15</sup> et différentes conditions expérimentales sont utilisées. Le taux de répétition du laser est plus élevé et les bandes sont plus larges. Ce jeu de données est dit « exogène » car il n'a pas été obtenu sur le même appareil que le « jeu d'entraînement ».

Enfin le troisième jeu de données a été acquis avec le montage OPO/OPA-LCQ. Il contient lui des spectres de GlcN, GalN, ManN et GlcNAc ainsi que sept autres composés : des disaccharides, un monosaccharide sulfaté et du paracétamol. Ces composés sont vus comme d'éventuelles pollutions dans l'analyse de polysaccharides. Ce jeu de données est vu comme endogène étant donné qu'il a été acquis avec le même appareillage que le « jeu d'entraînement ».

Pour améliorer l'entraînement de l'algorithme, ces trois jeux de données expérimentaux ont été enrichis en produisant des variants de spectres synthétiques (*in silico*). Lors de l'analyse d'un jour à l'autre, des paramètres peuvent changer, ce sont ces paramètres que nous avons modifiés pour enrichir les jeux de données.

- Le rapport signal sur bruit peut changer, pour mimer ce défaut, un bruit gaussien blanc a été ajouté.
- L'intensité du laser peut aussi fluctuer d'un jour à l'autre, voire même le long de la gamme des nombres d'onde. Cela a pour effet de changer l'amplitude des bandes d'absorption. Une variation linéaire du signal le long du spectre est ajoutée (contenue dans +/- 10 %).
- Le fait d'augmenter la vitesse d'acquisition du spectre change l'échantillonnage. Visuellement la résolution semble moins bonne. Pour simuler la différence d'échantillonnage, des points sont retirés régulièrement le long des spectres avant de refaire une interpolation entre les points restants pour retrouver le nombre de points initiaux. Le facteur de diminution de l'échantillonnage est choisi aléatoirement entre 1 et 5.
- Enfin de petites variations peuvent avoir lieu dans la calibration en longueur d'onde du laser. Les changements de calibration ont pour effet de décaler les pics d'absorption sur le spectre IRMPD. Cela a été simulé en décalant linéairement les nombres d'onde de +/-10 cm<sup>-1</sup>.

Finalement, les spectres synthétiques sont normalisés et interpolés pour obtenir 1200 points sur la gamme de nombres d'onde.

En tout 8000 spectres sont générés aléatoirement pour le jeu de données n°1 (2000 par monomère). Des lots sont produits en piochant de manière aléatoire les différents spectres afin de ne pas se retrouver avec des lots composés d'une seule molécule. 70 % des spectres sont utilisés pour effectuer l'entraînement du modèle et 30 % pour valider le modèle. Les jeux de données 2 et 3 sont également enrichis. La composition des différents jeux de données est donnée dans le Tableau 1.

Tableau 1 : Tableau présentant les caractéristiques des différents jeux de données utilisés dans cette étude.

Jeu de données	1	2	3
Objectif du jeu de données	Entraînement : 70 % Validation : 30 %	Test de classification	Test de discrimination
Catégories	4	1	10
Acquisition	Standard	Plus faible échantillonnage	Standard
Spectres IRMPD expérimentaux	33	4	12
Jeu de données augmenté	8000	8000	1300

### 3.3.2 Architecture du modèle

Le modèle choisi est un réseau de neurones entièrement connecté à propagation avant basé sur une architecture de type perceptron multicouche<sup>30</sup> avec une approche probabilistique (réseau bayésien de neurones profonds). Ce modèle permet de quantifier son incertitude intrinsèque. Il s'agit d'un réseau de neurones composé de trois couches cachées avec respectivement 300, 225, 100 neurones ainsi qu'une fonction d'activation de type redresseur linéaire (ReLU pour Rectified Linear Unit) pour chaque couche. C'est cette fonction qui déterminera si un neurone est activé.

Deux couches de « dropout » sont intercalées, l'une entre la 1<sup>ère</sup> et la 2<sup>nde</sup> et l'autre entre la 2<sup>nde</sup> et la 3<sup>ème</sup> couche cachée. Le dropout est utilisé lors de la phase d'entraînement du modèle, pendant cette phase, des neurones sont inactivés aléatoirement. Cette inactivation proposée dans les années 2010<sup>31</sup> permet de contrecarrer les problèmes de surapprentissage. Dans notre modèle, la valeur du dropout est de 25 %, ce qui revient à dire que 1 neurone sur 4 est inactivé au niveau de la 1<sup>ère</sup> couche, ainsi que sur la 2<sup>nde</sup> couche. Autrement, le modèle perdrait ses capacités de prédiction sur de nouveaux échantillons, il serait trop spécifique aux échantillons d'apprentissage. Le dropout n'est pas actif lors de la phase de test, phase pendant laquelle, tous les neurones sont pris en compte. L'objectif de l'entraînement est la construction du modèle afin d'identifier un composé au moyen de son spectre IRMPD.

Pour prendre en compte la nature probabilistique des réseaux de neurones profonds, nous avons utilisé une technique d'inférence variationnelle. Le poids de chaque nœud n'est pas une valeur exacte mais est remplacé par une distribution centrée en  $\mu$  avec un écart type  $\sigma$  optimisés par la méthode Bayes-by-Backprop<sup>32</sup>. Cette méthode permet de quantifier l'incertitude de la prédiction des modèles. Pour cela, un même spectre est identifié plusieurs fois avec le modèle, à chaque identification les poids associés à chaque nœud ne seront pas strictement identiques il est alors possible de déduire une incertitude associée à l'identification du composé inconnu.

### 3.3.3 Exactitude de la classification du modèle

Le modèle GAIcomics présente une exactitude de prédiction de 100 % sur le lot de validation (30 % des spectres et de 99.98 % sur le jeu de données n°2 (jeu de données exogène, qui permet le test sur la classification dans le Tableau 1). Les 8000 spectres synthétiques du jeu de données n°2 sont triés en fonction de leur niveau de bruit, modulation de l'amplitude (fluctuation de la puissance laser), variation de l'énergie et baisse de l'échantillonnage. L'exactitude moyenne du modèle est tracée en fonction de ces 4 paramètres et présentée Figure 22.

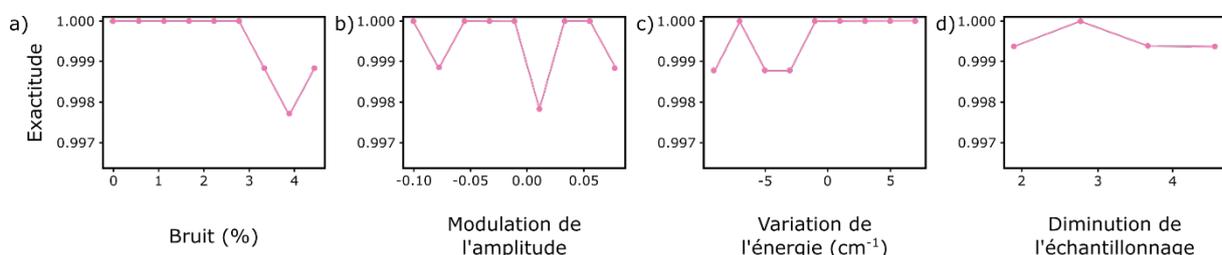


Figure 22 : Exactitude du modèle avec différentes conditions expérimentales simulés par les jeux de données enrichis. L'exactitude est toujours supérieure à 0.997.

Les données en Figure 22 montrent l'exactitude de la prédiction du modèle en fonction des différents paramètres affectant l'acquisition des spectres IRMPD. Sans surprise, l'exactitude diminue quand le bruit augmente (Figure 22a). Dans le cas de la modification de l'amplitude des pics, l'exactitude ne varie pas de manière monotone, elle présente des minima aux alentours de -7 %, 1 % et 8 % de

variation d'amplitude, toutes les autres valeurs donnent une exactitude de 100 % (Figure 22b). Le changement de l'énergie le long de la gamme de nombre d'onde a un effet sur l'exactitude des prédictions quand cette variation est négative, l'exactitude a tendance à être moins bonne que dans le cas d'une variation d'énergie positive le long de la gamme des nombres d'onde (Figure 22c). Enfin la diminution de la fréquence d'échantillonnage diminue légèrement l'exactitude, elle demeure supérieure à 0.999 dans tous les cas (Figure 22d). Les spectres synthétiques sont obtenus avec différentes combinaisons de paramètres. C'est peut-être ces combinaisons qui sont responsables des variations non monotones de l'exactitude (par exemple l'exactitude est meilleure avec un bruit supérieur à 4 % que dans le cas où le bruit est compris entre 3 et 4 %).

Ce modèle est donc bien capable de prédire la structure d'un monosaccharide à partir de son spectre IRMPD, le tout en se servant d'une résolution variable, en ajoutant du bruit ou une variation de l'énergie qui entraîne une fluctuation du signal. Dans le cadre de la glycomique, il n'est pas possible d'être exhaustif, il arrivera forcément un moment où le modèle sera utilisé sur une molécule pour laquelle il n'a pas été entraîné. Pour savoir comment le modèle réagira dans un tel cas, le modèle a été évalué de manière plus approfondie.

### 3.3.4 Incertitude et précision du modèle

D'un point de vue analytique, il est important de déterminer la proportion de faux positifs et celle de faux négatifs. On s'attend à n'observer que peu de faux négatifs, leur analyse manuelle est envisageable afin d'annoter correctement le spectre et le réinjecter dans le modèle pour l'améliorer. En revanche dans le cas de faux positifs, il est nécessaire d'évaluer le pouvoir de discrimination du modèle. C'est dans ce but que nous nous servons du troisième jeu de données. Pour rappel, il consiste en 1300 spectres simulés à partir des spectres IRMPD obtenus expérimentalement de 12 molécules acquis au moyen du montage historique. Le modèle n'a jamais rencontré ces données que ce soit pour s'entraîner, ou au moment de la phase de validation du modèle. Ce jeu de données contient les spectres de GlcN, GalN et ManN ainsi que sept autres molécules qui peuvent être vues comme des impuretés.

La variabilité de la probabilité de prédiction est effectuée en interrogeant le modèle de nombreuses fois sur le même spectre. Si au cours des interrogations, le modèle prédit la même, alors l'échantillon appartient vraisemblablement à cette catégorie et l'incertitude est basse. En revanche, si au cours des interrogations le modèle ne prédit pas toujours la même catégorie, alors l'incertitude est haute et l'échantillon n'appartient vraisemblablement pas à cette catégorie. Le modèle est interrogé 200 fois pour chaque échantillon. La moyenne des probabilités de prédiction ainsi que sa variabilité représentée par l'interpercentile (5-95 %) sont représentées dans la Figure 23.

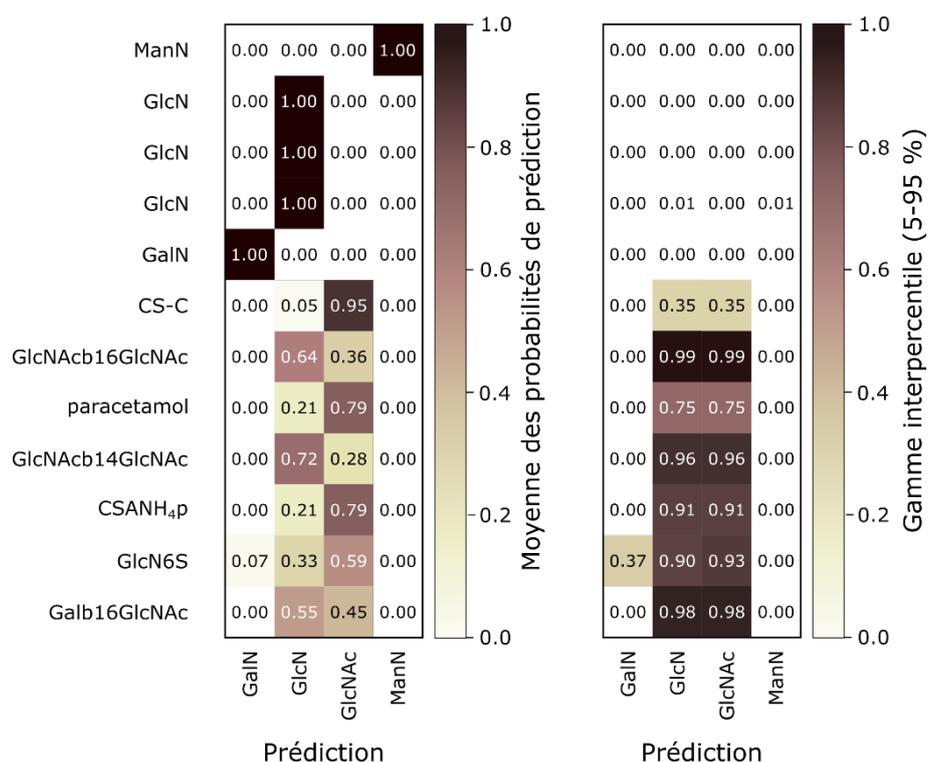


Figure 23 : Prédiction par le modèle sur le troisième jeu de données (contient 5 hexosamines et 7 autres molécules). La carte de gauche présente la moyenne des probabilités de prédiction pour chaque catégorie, celle de droite présente l'interpercentile de 5 à 95 % des probabilités de prédiction de chaque catégorie. Toutes les hexosamines sont correctement prédites avec une moyenne de 1 et une gamme interpercentile très proche de 0, ce qui donne une bonne confiance dans la prédiction.

La première carte montre la probabilité de prédiction des différentes espèces inconnues. Les molécules pour lesquelles le modèle a été entraîné (mannosamine, glucosamine et galactosamine) sont attribuées aux bonnes catégories avec une probabilité de 1. De plus l'incertitude de prédiction est nulle. Le modèle reconnaît donc correctement ces espèces.

Dans le cas des molécules pour lesquelles le modèle n'a pas été entraîné, l'interrogation du modèle ne présente pas de probabilité de prédiction haute, le modèle ne reconnaît donc pas de faux positifs. Une seule espèce fait exception : la CS-C qui est prédite comme étant du GlcNAC avec une probabilité de 0,95 ce qui est une valeur proche de 1. Nous pouvons donc nous poser la question est-ce que cette identification est bonne ? Dans ce cas, les multiples interrogations du modèle montrent que la variabilité des prédictions est significative : 0,35 ce qui est bien plus élevé que pour les assignations correctes. Dans cet exemple, si on choisit une variabilité maximale de 0,35, aucune mauvaise identification ne sera effectuée. Le cas de la CS-C est ici la limite du modèle, c'est la seule prédiction qui peut poser question, cette interrogation est éliminée avec l'examen de l'incertitude. En calibrant correctement le modèle, on s'assure qu'aucune erreur ne soit faite.

### 3.3.5 Conclusion et ouverture

Un modèle de réseau de neurones profonds bayésien a été évalué pour l'analyse automatique et l'identification de spectres IRMPD de monosaccharides. Trois jeux de données ont été utilisés comme suit : un jeu de données divisé en deux utilisé à des fins d'entraînement et de validation. Un second jeu de données issu d'un autre appareil permet de déterminer l'exactitude du modèle. Enfin un troisième jeu de données provenant du même appareil que le premier jeu de données mais contenant des spectres d'autres molécules évalue l'incertitude du modèle.

Le modèle décrit dans cet article permet d'identifier des composés présents dans la base de données de manière robuste et exacte. L'identification se fait sur un jeu de données obtenu avec un autre appareillage. Le modèle peut identifier des spectres bruités, présentant une moins bonne résolution ou bien encore sur des spectres distordus. Après avoir identifié des composés connus, la spécificité du modèle est évaluée à l'aide d'un troisième jeu de données contenant des molécules connues par le modèle ainsi que d'autres inconnues. Nous concluons qu'un réseau de neurones relativement simple est une solution intéressante dans l'analyse et l'identification de spectres IRMPD de sucres. Ce modèle peut être facilement intégré à un workflow analytique entre l'acquisition des données et l'algorithme de séquençage. Les spectres non identifiés peuvent être interprétés manuellement afin de nourrir le modèle en tant que nouvelles données pour l'entraînement du modèle.

L'addition du rapport  $m/z$  à ce modèle pourrait être très judicieuse, ainsi un premier filtre éliminerait toutes les molécules d'autres tailles. On pourrait faire de même avec d'autres données telles que les temps de rétention en chromatographie liquide, les temps d'arrivée en mobilité ionique ou toute autre donnée pertinente issue de l'analyse des échantillons.

## 4 CONCLUSION

---

Les analyses menées dans le but d'identifier et de caractériser des sucres ne nécessitent pas forcément d'avoir la meilleure résolution spectrale possible. Dans ce cas les mesures peuvent être grandement accélérées et rendues compatibles avec les contraintes des laboratoires d'analyses (d'un point de vue de la vitesse d'acquisition des données). Les spectres ainsi obtenus ne permettront pas de répondre à des questions de physico-chimie sur les composés analysés notamment à cause du manque d'échantillonnage du spectre. Toutefois, ils sont tout à fait convenables pour faire la différence entre deux isomères. Afin de pallier le manque de puissance en sortie du nouveau laser, un montage à deux lasers a été mis au point dans l'équipe. Ces améliorations simplifient et accélèrent l'acquisition des données. La vitesse d'acquisition des données de spectroscopie IRMPD est maintenant de l'ordre de quelques secondes, ce qui est compatible avec le couplage entre chromatographie liquide et spectroscopie IRMPD. Ce couplage est l'un des prochains développements envisagés dans le cadre de ce projet de thèse.

D'un point de vue de l'interprétation des données, l'expérimentateur n'est plus obligé d'examiner les données manuellement, le système de calcul de scores est dorénavant utilisé en routine dans l'équipe. L'ACP et l'IA sont des approches encore exploratoires mais prometteuses. Il faut encore explorer d'autres classes de sucres afin de bien évaluer ces méthodes. De plus, l'incorporation d'autres données telles que le rapport  $m/z$ , le temps de rétention (lorsqu'il sera possible de coupler la chromatographie liquide à la spectroscopie IRMPD) est de premier intérêt pour avoir des sources d'identification orthogonales et donc maximiser la confiance dans l'identification du produit inconnu.

## 5 BIBLIOGRAPHIE

- 1 Y. Yang, G. Liao and X. Kong, Charge-state Resolved Infrared Multiple Photon Dissociation (IRMPD) Spectroscopy of Ubiquitin Ions in the Gas Phase, *Sci. Rep.*, 2017, **7**, 16592.
- 2 R. E. van Outersterp, J. Martens, A. Peremans, L. Lamard, F. Cuyckens, J. Oomens and G. Berden, Evaluation of table-top lasers for routine infrared ion spectroscopy in the analytical laboratory, *The Analyst*, 2021, **146**, 7218–7229.
- 3 S. Warnke, A. Ben Faleh and T. R. Rizzo, Toward High-Throughput Cryogenic IR Fingerprinting of Mobility-Separated Glycan Isomers, *ACS Meas. Sci. Au*, 2021, **1**, 157–164.
- 4 A. Lagutschenkov, R. K. Sinha, P. Maitre and O. Dopfer, Structure and Infrared Spectrum of the Ag<sup>+</sup>-Phenol Ionic Complex, *J. Phys. Chem. A*, 2010, **114**, 11053–11059.
- 5 R. K. Sinha, E. Nicol, V. Steinmetz and P. Maître, Gas phase structure of micro-hydrated [Mn(ClO<sub>4</sub>)<sup>+</sup>] and [Mn<sub>2</sub>(ClO<sub>4</sub>)<sub>3</sub>]<sup>+</sup> ions probed by infrared spectroscopy, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 758–772.
- 6 R. K. Sinha, U. Erlekam, B. J. Bythell, B. Paizs and P. Maître, Diagnosing the Protonation Site of b2 Peptide Fragment Ions using IRMPD in the X–H (X = O, N, and C) Stretching Region, *J. Am. Soc. Mass Spectrom.*, DOI:10.1007/s13361-011-0173-1.
- 7 D. Scuderi, J. M. Bakker, S. Durand, P. Maitre, A. Sharma, J. K. Martens, E. Nicol, C. Clavaguéra and G. Ohanessian, Structure of singly hydrated, protonated phospho-tyrosine, *Int. J. Mass Spectrom.*, 2011, **308**, 338–347.
- 8 R. K. Sinha, D. Scuderi, P. Maitre, B. Chiavarino, M. E. Crestoni and S. Fornarini, Elusive Sulfurous Acid: Gas-Phase Basicity and IR Signature of the Protonated Species, *J. Phys. Chem. Lett.*, 2015, **6**, 1605–1610.
- 9 T. C. Penna, G. Cervi, A. F. Rodrigues-Oliveira, B. D. Yamada, R. Z. C. Lima, J. J. Menegon, E. L. Bastos and T. C. Correra, Development of a photoinduced fragmentation ion trap for infrared multiple photon dissociation spectroscopy, *Rapid Commun. Mass Spectrom.*, 2020, **34**, e8635.
- 10 G. E. Hofmeister, Z. Zhou and J. A. Leary, Linkage position determination in lithium-cationized disaccharides: tandem mass spectrometry and semiempirical calculations, *J. Am. Chem. Soc.*, 1991, **113**, 5964–5970.
- 11 S.-T. Tsai, H.-S. Nguan and C.-K. Ni, Identification of Anomerism and Linkage of Arabinose and Ribose through Collision-Induced Dissociation, *J. Phys. Chem. A*, 2021, **125**, 6109–6121.
- 12 B. Schindler, L. Barnes, G. Renois, C. Gray, S. Chambert, S. Fort, S. Flitsch, C. Loison, A.-R. Allouche and I. Compagnon, Anomeric memory of the glycosidic bond upon fragmentation and its consequences for carbohydrate sequencing, *Nat. Commun.*, 2017, **8**, 973.
- 13 J. S. Ho, A. Gharbi, B. Schindler, O. Yeni, R. Brédy, L. Legentil, V. Ferrières, L. L. Kiessling and I. Compagnon, Distinguishing Galactoside Isomers with Mass Spectrometry and Gas-Phase Infrared Spectroscopy, *J. Am. Chem. Soc.*, 2021, **143**, 10509–10513.
- 14 A. Depraz Depland, G. Renois-Predelus, B. Schindler and I. Compagnon, Identification of sialic acid linkage isomers in glycans using coupled InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and mass spectrometry, *Int. J. Mass Spectrom.*, 2018, **434**, 65–69.
- 15 O. Yeni, B. Schindler, B. Moge and I. Compagnon, Rapid IRMPD (InfraRed multiple photon dissociation) analysis for glycomics, *Analyst*, 2022, **147**, 312–317.
- 16 K. Varmuza, P. N. Penchev and H. Scsibrany, Maximum Common Substructures of Organic Compounds Exhibiting Similar Infrared Spectra, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 420–427.
- 17 O. Yeni, S. Ollivier, B. Moge, D. Ropartz, H. Rogniaux, L. Legentil, V. Ferrières and I. Compagnon, Ring-Size Memory of Galactose-Containing MS/MS Fragments: Application to the Detection of Galactofuranose in Oligosaccharides and Their Sequencing, *J. Am. Chem. Soc.*, 2023, **145**, 15180–15187.
- 18 O. Yeni, Analyse par IRMPD de polysaccharides de lichen contenant des unités galactofuranose, Thèse soutenue, Lyon 1, 2023.
- 19 S. Wold, K. Esbensen and P. Geladi, Principal Component Analysis.

- 20 A. Ben Faleh, S. Warnke, T. Van Wieringen, A. H. Abikhodr and T. R. Rizzo, New Approach for the Identification of Isobaric and Isomeric Metabolites, *Anal. Chem.*, 2023, **95**, 7118–7126.
- 21 R. E. van Outersterp, P. C. Kooijman, J. Merx, U. F. H. Engelke, N. Omidika, M.-L. Tonnejck, K. J. Houthuijs, G. Berden, T. M. A. Peters, D. J. Lefeber, M. A. A. P. Willemsen, J. Mecinović, J. J. Janssen, K. L. M. Coene, R. A. Wevers, T. J. Boltje, J. Oomens and J. Martens, *En Révis.*, 2022.
- 22 R. E. van Outersterp, P. C. Kooijman, J. Merx, U. F. H. Engelke, N. Omidika, M.-L. H. Tonnejck, K. J. Houthuijs, G. Berden, T. M. A. Peters, D. J. Lefeber, M. A. A. P. Willemsen, J. Mecinovic, J. J. Jansen, K. L. M. Coene, R. A. Wevers, T. J. Boltje, J. Oomens and J. Martens, Distinguishing Oligosaccharide Isomers Using Far-Infrared Ion Spectroscopy: Identification of Biomarkers for Inborn Errors of Metabolism, *Anal. Chem.*, , DOI:10.1021/acs.analchem.3c00363.
- 23 L. Barnes, A.-R. Allouche, S. Chambert, B. Schindler and I. Compagnon, Ion spectroscopy of heterogeneous mixtures: IRMPD and DFT analysis of anomers and conformers of monosaccharides, *Int. J. Mass Spectrom.*, 2020, **447**, 116235.
- 24 H. J. Luinge, Automated interpretation of vibrational spectra.
- 25 M. Hilario, A. Kalousis, C. Pellegrini and M. Müller, Processing and classification of protein mass spectra, *Mass Spectrom. Rev.*, 2006, **25**, 409–449.
- 26 M. Sattlecker, Support vector machine ensembles for breast cancer type prediction from mid-FTIR micro-calcification spectra, *Chemom. Intell. Lab. Syst.*
- 27 W. Fu and W. S. Hopkins, Applying Machine Learning to Vibrational Spectroscopy, *J. Phys. Chem. A*, 2018, **122**, 167–171.
- 28 K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari and P. Rinke, Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra, *Adv. Sci.*, 2019, **6**, 1801367.
- 29 J. Riedel, M. Lettow, M. Grabarics, M. Götze, R. L. Miller, G.-J. Boons, G. Meijer, G. Von Helden, G. P. Szekeres and K. Pagel, Predicting Structural Motifs of Glycosaminoglycans using Cryogenic Infrared Spectroscopy and Random Forest, *J. Am. Chem. Soc.*, 2023, **145**, 7859–7868.
- 30 I. Goodfellow, Y. Bengio and A. Courville, *MIT Press*, 2016.
- 31 N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting.
- 32 C. Blundell, J. Cornebise, K. Kavukcuoglu and D. Wierstra, 2015.

# Chapitre III : Analyse d'oligosaccharides au moyen de la spectroscopie IRMPD après séparation par chromatographie liquide

## TABLE DES MATIERES

<b>1</b>	<b>Introduction .....</b>	<b>80</b>
<b>2</b>	<b>La chromatographie liquide .....</b>	<b>80</b>
2.1	L'appareil.....	80
2.2	Contraintes techniques pour le couplage entre chromatographie liquide et spectroscopie IRMPD.....	81
2.3	Approches déjà envisagées .....	82
2.3.1	Collecte de fractions .....	82
2.3.2	Stockage dans des boucles d'injection.....	83
2.3.3	Utilisation du stop flow .....	84
<b>3</b>	<b>Analyse des produits de digestion d'un pentasaccharide synthétique .....</b>	<b>85</b>
3.1	Structure du Fondaparinux.....	85
3.2	Acquisition des spectres en infusion directe .....	86
3.3	Les digestions .....	86
3.4	La méthode chromatographique.....	89
3.5	Résultats .....	90
<b>4</b>	<b>Analyse en ligne des composés en sortie de colonne : exemple sur les HMOs .....</b>	<b>91</b>
4.1	La méthode chromatographique.....	91
4.2	Premiers résultats du couplage LC-MS/MS-IR.....	92
4.2.1	Intérêt de fragmenter le composé d'intérêt.....	92
4.2.2	Mesures des spectres IRMPD de référence .....	93
4.2.3	Séparation chromatographique des deux isomères .....	94
4.2.4	Mesure en ligne du spectre IRMPD du premier pic chromatographique .....	94
4.3	Réduction des composés.....	95
4.3.1	Les composés .....	95
4.3.2	La préparation des composés : réduction.....	95
4.3.3	Résultats.....	96
4.4	Nouveaux développements conduits sur le couplage LC-MS-IR .....	98
4.4.1	Spectres IRMPD de monosaccharides réduits : le cas des hexoses .....	98
4.4.2	Détection d'une forme rare de galactose composant des polysaccharides bactériens par screening à longueur d'onde fixe .....	100
<b>5</b>	<b>Conclusion .....</b>	<b>102</b>
<b>6</b>	<b>Bibliographie .....</b>	<b>104</b>

# 1 INTRODUCTION

---

Au cours des parties précédentes, la spectroscopie IRMPD a montré son potentiel en matière de discrimination d'isomères dans le domaine de la glycomique. Toutefois, une limite à cette discrimination demeure. En effet, la manipulation des ions en phase gazeuse se fait au moyen de la spectrométrie de masse. Il est donc par définition impossible d'analyser individuellement deux composés isomériques s'ils sont présents simultanément dans le spectromètre de masse. Le spectre IRMPD résultant de cette analyse sera le spectre pondéré de ces deux espèces.

C'est pour passer outre cette barrière que l'utilisation de méthodes séparatives en amont de la spectroscopie IRMPD est intéressante. Les méthodes séparatives sont couramment utilisées pour analyser des échantillons complexes. Cela donne également la possibilité d'étudier des isomères séparément et donc d'obtenir le spectre propre de chacune de ces espèces. Il existe une grande variété de méthodes séparatives dans le domaine de la chimie analytique. Certaines méthodes séparatives ont été testées dans l'équipe en amont de la spectroscopie IRMPD. C'est le cas de la mobilité ionique<sup>1</sup>, de l'électrophorèse capillaire<sup>2</sup> et de la chromatographie liquide<sup>3-5</sup>.

Dans ce chapitre, nous nous sommes intéressés au couplage en ligne de la chromatographie liquide au montage expérimental sans étape intermédiaire. Nous verrons également comment le développement a pu être mené à bien, quelles ont été les contraintes et quels sont les avantages de l'ajout de la chromatographie liquide au montage existant.

Dans le cas des sucres (composés très polaires) deux types de chromatographie sont intéressants : la chromatographie HILIC (hydrophilic interaction chromatography) et la chromatographie PGC (porous graphitic carbon). Ces deux types de colonnes ont montré leur intérêt dans la séparation de composés polaires<sup>6,7</sup>, notamment les sucres<sup>8,9</sup>. Nous verrons des applications sur les sucres dont la caractérisation des produits de digestion d'un pentasaccharide synthétique : le Fondaparinux après collecte de fractions, ou en ligne en sortie de colonne chromatographique. Ensuite, nous verrons comment la régiochimie d'une liaison responsable de l'isomérisation entre deux composés a pu être analysée au moyen de ce couplage. Enfin un mélange de deux composés a été modifié chimiquement afin de procéder à l'analyse en LC-MS-IR (spectroscopie IR en phase gazeuse en ligne avec la chromatographie liquide).

Pour finir, nous verrons deux développements récents. Certains modes chromatographiques requièrent une modification chimique des sucres. Dans notre étude, nous avons réduits des hexoses et les avons analysés par spectroscopie IRMPD. Il ressort que la spectroscopie IRMPD est capable de discriminer les produits de ces modifications. Enfin, l'analyse de petits oligosaccharides issus de l'hydrolyse chimique de polysaccharides a été menée au moyen de ce montage dans lequel, une bande d'absorption d'un composé d'intérêt est identifiée et utilisée pour irradier les ions arrivant dans le spectromètre de masse. Ce screening a permis de détecter ce composé bien particulier dans des di- et des trisaccharides.

## 2 LA CHROMATOGRAPHIE LIQUIDE

---

### 2.1 L'APPAREIL

L'appareil de chromatographie utilisé est l'Ultimate 3000 commercialisé par ThermoFisher Scientific. Cet appareil contient un compartiment pompe (Figure 1b) qui permet de délivrer le débit de phase mobile (solvants présents Figure 1a) au travers de la colonne, la phase mobile entraîne les composés au cours de leur migration. Les pompes sont programmées par l'utilisateur avant l'analyse et assurent

le débit et la composition de phase mobile voulus (jusqu'à un mélange de trois solvants pour cet appareil) afin d'avoir la composition optimale pour séparer des composés. Ces concentrations peuvent varier au cours du temps, on parle alors de gradient. Un passeur (système de vannes et de boucles d'injection) permet l'introduction de l'échantillon dans la colonne ainsi que la collecte des fractions d'intérêt (la phase mobile est collectée alors en sortie de colonne après séparation chromatographique) (Figure 1c). Après injection, les échantillons passent dans le compartiment colonne qui consiste en un four thermostaté ce four régule la température sélectionnée (Figure 1d). C'est dans ce compartiment que la séparation des composés a lieu. En fonction de leur affinité avec la phase stationnaire, les composés sont élués plus ou moins rapidement de la colonne. Après séparation, les composés traversent alors un détecteur UV (Figure 1e) avant de rejoindre la source d'ionisation par ESI pour détection en spectrométrie de masse et dans notre cas analyse en MS-IR.

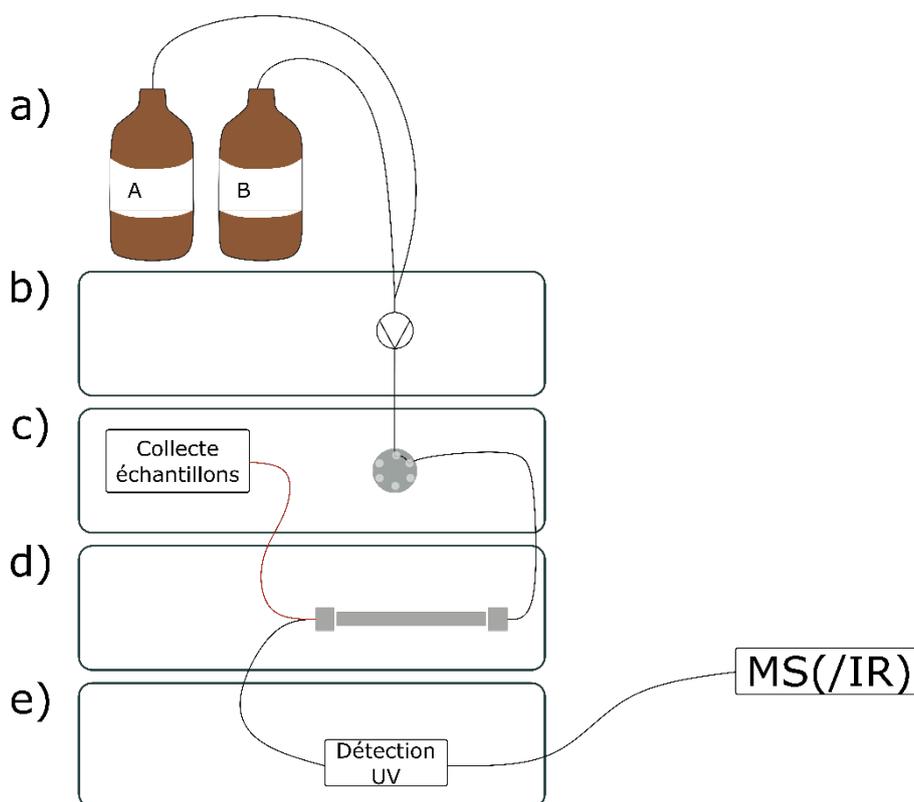


Figure 1 : Schéma de la chaîne de chromatographie liquide utilisée dans notre équipe, l'appareil est constitué de différents modules ayant chacun sa fonction. Les pompes b) entraînent le solvant à travers la boucle d'injection c), jusque dans la colonne d), en sortie de colonne les composés passent au niveau du détecteur (à travers la détection UV puis au niveau de la source d'ionisation e) alternativement, la solution peut être collectée au niveau du module du passeur c).

Le couplage entre la chromatographie liquide et la spectroscopie IRMPD se heurte à des contraintes qu'il faut absolument résoudre avant de pouvoir mener à bien des analyses par ce moyen.

## 2.2 CONTRAINTES TECHNIQUES POUR LE COUPLAGE ENTRE CHROMATOGRAPHIE LIQUIDE ET SPECTROSCOPIE IRMPD

Les pics chromatographiques durent quelques dizaines de secondes (Figure 2a). En chromatographie liquide, la finesse des pics est obtenue en travaillant avec une vitesse de phase mobile optimale (les analyses doivent être menées le plus rapidement possible). Les colonnes chromatographiques sont conçues pour obtenir les pics les plus fins possibles. Toutefois, au début de l'activité *caractérisation de sucres par spectroscopie IRMPD* dans l'équipe, l'acquisition d'un spectre nécessitait plus de 45 min (Figure 2b).

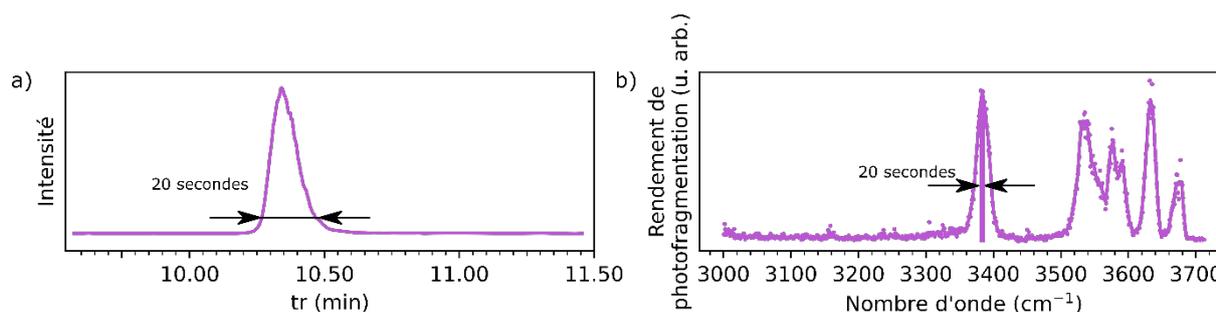


Figure 2 : Chromatogramme a) et spectre IRMPD d'un sucre b), les 20 secondes sont représentées par les deux flèches et le rectangle violet.

Les échelles de temps entre les deux méthodes étaient alors incompatibles, en effet en vingt secondes, uniquement  $7 \text{ cm}^{-1}$  sont couverts par le laser, des développements sont alors nécessaires. Actuellement, des méthodes ont été développées afin d'autoriser le couplage entre la spectroscopie IRMPD et une séparation chromatographique.

## 2.3 APPROCHES DEJA ENVISAGÉES

### 2.3.1 Collecte de fractions

La première approche est la plus simple, après séparation, le composé est récupéré dans un vial. Le contenu de ce vial est alors introduit en infusion dans le spectromètre de masse afin d'acquérir son spectre IRMPD<sup>3,4,10</sup> (voir Figure 3). La concentration de la fraction récupérée dépend de la concentration initiale, mais si l'on considère une injection de  $10 \mu\text{L}$  d'une solution à  $50 \mu\text{M}$  avec un débit de phase mobile réglé à  $300 \mu\text{L}\cdot\text{min}^{-1}$  et que l'on souhaite recueillir l'ensemble de l'échantillon (élué dans un créneau d'une minute), on recueille alors  $300 \mu\text{L}$  d'une solution à  $1,7 \mu\text{M}$ .

Si un gradient a été programmé, la composition de la solution dépendra du temps de rétention du produit et ne sera peut-être pas la plus adaptée pour une ionisation pour l'analyse par spectrométrie de masse. Il peut alors être nécessaire d'évaporer la solution pour reprendre le composé dans une solution idoine afin de faciliter l'ionisation des composés dans la source d'ionisation du spectromètre de masse. Cette étape d'évaporation/reprise dans un volume de solution peut également permettre d'augmenter la concentration en composé. Cette étape de concentration prend du temps car il est nécessaire d'évaporer la solution. De même, plusieurs séparations peuvent être programmées pour récupérer une plus grande quantité de composés, cette option accroît le temps passé entre la collecte de fractions et l'infusion directe de ces fractions.

L'avantage de la collecte de fractions est qu'elle est très simple, et peut se faire de manière automatique. La méthode chromatographique se déroule sans modification et on peut s'attendre à une séparation optimale des composés. En revanche une des limitations est que le temps écoulé entre la récupération de l'échantillon et son passage en spectroscopie peut être suffisant pour empêcher l'étude des effets d'anomérisation. En effet, initialement en solution, les anomères  $\alpha$  et  $\beta$  sont tous deux présents. Il est possible de séparer les deux anomères  $\alpha$  et  $\beta$  par chromatographie liquide<sup>5</sup>. Si maintenant les deux anomères sont collectés dans des vials différents, alors ces vials contiendront uniquement un anomère ou l'autre. Mais au bout d'un certain temps, les deux vials présenteront tous deux les mêmes proportions d'anomères  $\alpha$  et  $\beta$  car l'équilibre aura de nouveau été atteint<sup>11</sup>. L'équilibre de mutarotation est atteint entre 30 min et quelques heures en fonction de la température et du solvant.

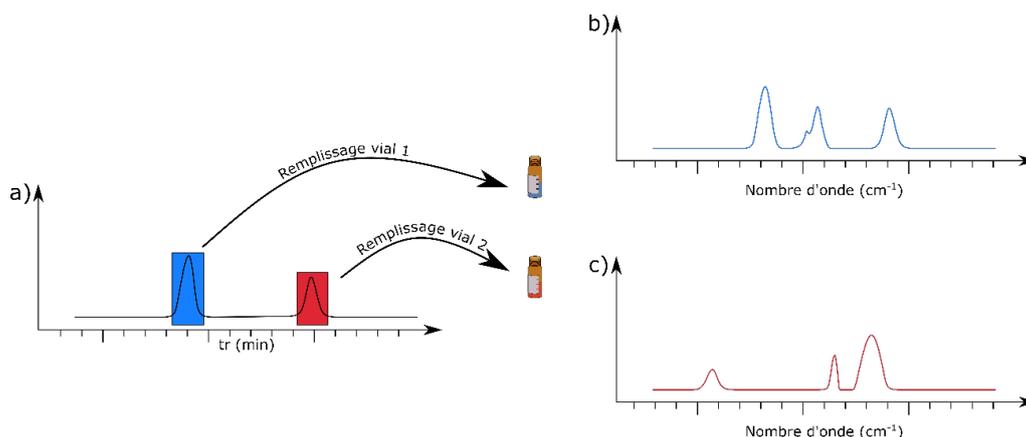


Figure 3 : Chromatogramme de deux composés a), lors de l'élution de chacun des pics, la phase mobile est collectée, infusion du vial contenant le premier composé à avoir été élué b), infusion du vial contenant le deuxième composé à avoir été élué c).

### 2.3.2 Stockage dans des boucles d'injection

Ici, la chaîne chromatographique est à la manière d'appareils dédiés à l'analyse en LC 2D équipée de plusieurs boucles d'injection (Figure 4). Le débit de phase mobile passe à travers la colonne et rejoint la spectrométrie de masse. Le programme de contrôle de la chromatographie liquide suit les consignes de l'opérateur, ainsi à certains temps de rétention la phase mobile est collectée ou alors elle est dirigée vers le spectromètre de masse. La collecte de fractions peut être répétée autant de fois que l'appareil contient de boucles d'injection. A la fin de la séparation, le contenu des boucles d'injection est successivement injecté dans le spectromètre de masse en infusion directe<sup>12</sup>.

Cette configuration a l'avantage d'être automatisée (recueil des fractions et infusions). Toutefois, cette méthode requiert l'utilisation d'un appareil spécialement conçu pour recueillir les fractions dans des boucles d'injection. Ce type d'appareil est moins répandu que les appareils de chromatographie liquide « classiques » qui ne sont pas équipés d'un tel système. Le nombre de composés pouvant être recueillis est physiquement limité par le nombre de boucles d'injection présentes dans le système.

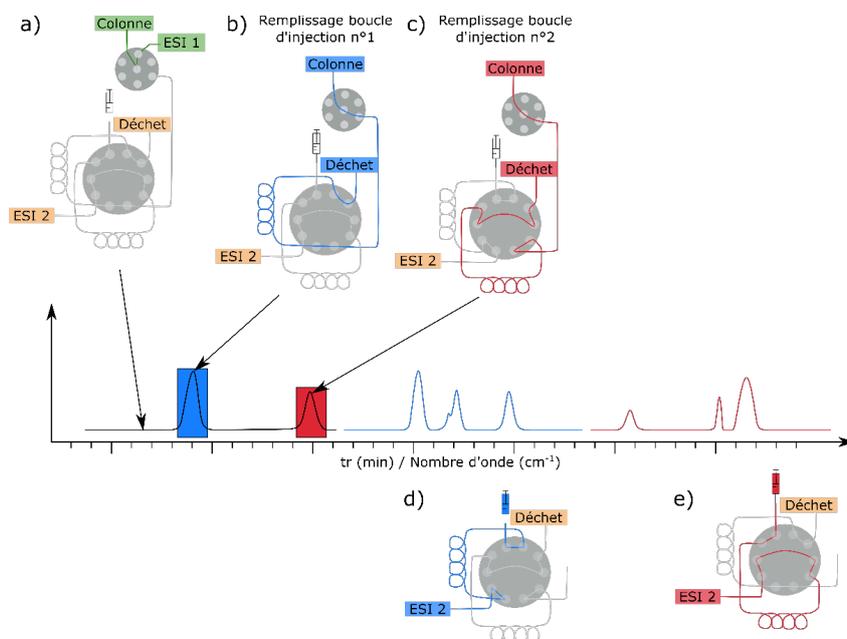


Figure 4 : Chromatogramme et spectre IRMPD de deux composés. La phase mobile est directement injectée dans le spectromètre de masse a), un composé a été détecté par MS, la phase mobile est alors collectée dans des boucles d'injection lors de l'élution de chacun des pics b) et c), enfin le contenu des boucles d'injection est infusé successivement dans le spectromètre de masse pour acquisition des spectres IR en phase gazeuse d) et e).

Avec ces deux méthodes, les analytes sont prélevés et l'analyse est faite ultérieurement, le temps d'acquisition du spectre IRMPD n'est donc pas contraignant car les spectres sont ensuite acquis en infusion directe.

### 2.3.3 Utilisation du stop flow

Cette méthode nécessite de modifier le débit de phase mobile en cours de séparation. Lorsqu'un composé sort de la colonne le débit de phase mobile diminue afin d'élargir le pic chromatographique (Figure 5). Le composé sera ainsi disponible plus longtemps devant le détecteur pour permettre d'acquérir son spectre IRMPD. Dans l'article de Schindler *et al.*<sup>5</sup> les auteurs ont réussi à séparer et acquérir le spectre IRMPD de deux disaccharides isomères l'un de l'autre présentant uniquement une variation de la position de la liaison entre les deux monosaccharides. Ils ont également réussi à séparer les deux anomères de la glucosamine et acquérir leur empreinte IR. Dans cette méthode, le spectre est acquis en environ 6,7 minutes, ce qui est presque 7 fois plus rapide que les spectres jusqu'alors acquis dans l'équipe. Des efforts ont été faits tant d'un point de vue de la LC afin d'élargir les pics que d'un point de vue de la spectroscopie IRMPD afin d'accélérer l'acquisition du spectre.

Toutefois, cette méthode fait perdre de la résolution chromatographique. En effet, la résolution chromatographique est directement dépendante de la vitesse de la phase mobile (l'efficacité de la colonne change). Ainsi, le fait de modifier le débit de phase mobile va inévitablement jouer sur la séparation des composés. C'est une des limites de l'analyse en « stop flow ». Dans notre cas, il est possible de séparer convenablement quelques composés, mais dans le cas où l'on souhaiterait séparer un grand nombre de composés, cette limite pourrait devenir gênante.

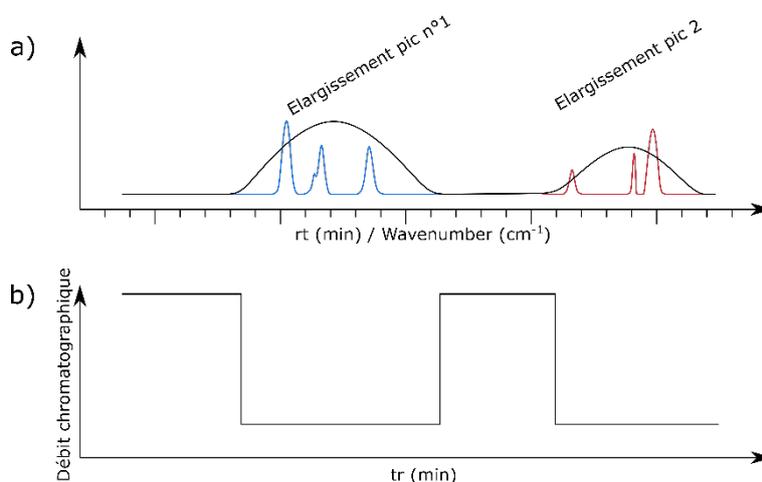


Figure 5 : Chromatogramme de deux composés, lors de l'élution d'un composé, le débit est diminué afin d'élargir le pic chromatographique pour avoir suffisamment de temps pour acquérir le spectre IRMPD a), variation du débit en fonction du temps b).

Comme nous avons pu le voir, différentes méthodes ont vu le jour afin de coupler chromatographie liquide et spectroscopie IR en phase gazeuse, aucune d'entre elle n'est parfaite, à chaque fois un compromis est recherché. Selon les applications voulue il faudra choisir la plus adaptée.

Dans la suite de cette partie, nous verrons l'application de la collecte de fractions sur l'analyse de la digestion d'un pentasaccharide. Nous avons également travaillé à la mise en place d'une nouvelle manière d'acquérir des spectres IRMPD en sortie de séparation chromatographique sans étapes intermédiaires entre la séparation et l'analyse ni la modification des paramètres de chromatographie, ce développement sera présenté sur les produits de digestion de ce même pentasaccharide, mais également dans un second temps sur des petits oligosaccharides.

### 3 ANALYSE DES PRODUITS DE DIGESTION D'UN PENTASACCHARIDE SYNTHETIQUE

Les glycosaminoglycanes (GAGs) sont des polysaccharides constitués d'une alternance de deux sucres, un sucre aminé et un sucre acide. Ces polysaccharides sont lourdement sulfatés (à l'exception de l'acide hyaluronique qui ne possède aucun sulfate dans sa structure). L'analyse de la structure des GAGs fait l'objet de nombreuses études on peut citer les approches par LC-MS<sup>13-16</sup> (chromatographie liquide couplée à la spectrométrie de masse). D'autres approches ont aussi été proposées pour étudier de telles structures notamment en MS/IR<sup>2,17-21</sup>.

Ces études sont motivées notamment par le fait que les GAGs représentent un intérêt thérapeutique. En effet, ils sont impliqués dans de nombreux processus biologiques par exemple les chondroïtines sulfatées jouent un rôle dans le développement neuronal<sup>22,23</sup>, les héparines sont, elles connues et utilisées depuis près de 100 ans pour leurs propriétés anticoagulantes<sup>24,25</sup>. Les héparines se lient à l'antithrombine III ce qui l'active et permet son activité anticoagulante. En fait, il existe des chaînes d'héparines de tailles différentes qui ont plus ou moins d'affinités pour l'antithrombine ainsi, lors d'expériences de fractionnement de l'héparine toutes les fractions ne montrent pas la même activité anticoagulante<sup>26</sup>. Seules quelques fractions d'héparine agissent sur l'antithrombine et ont un effet anticoagulant. Tout l'intérêt est alors de trouver la séquence d'intérêt sur une longue chaîne polysaccharidique afin de s'en servir comme médicament. Après des études intensives, un composé actif : le Fondaparinux a pu être synthétisé, purifié et commercialisé (sous le nom d'Arixtra) il est couramment utilisé afin de limiter les risques de thromboses notamment en cas d'intervention chirurgicale<sup>27-29</sup>.

#### 3.1 STRUCTURE DU FONDAPARINUX

Le Fondaparinux est un pentasaccharide synthétisé chimiquement. Sa structure consiste en une alternance de glucosamines et d'acides hexuroniques (Figure 6). Il est hautement sulfaté avec pas moins de 8 groupements sulfate, de plus il contient deux groupements carboxyls. Ces sulfates lui confèrent la possibilité de se lier à la protéine d'intérêt notamment par des liaisons hydrogènes avec les chaînes latérales des acides aminés basique (arginines et lysines)<sup>30</sup>. Le Fondaparinux présente un groupement méthyl au niveau de son extrémité anomérique. Ce groupement n'est pas présent naturellement mais facilite la synthèse et la purification du composé<sup>28,29</sup>. Sous forme de poudre, le Fondaparinux est stabilisé par 10 ions sodium qui agissent comme contre ions sur les groupements acides (sulfates et carboxylates), sa masse moléculaire est alors de 1728 Da. En solution, les ions sodium quittent leur site de liaison avec le Fondaparinux, qui est alors principalement observé en spectrométrie de masse sous ses formes  $[M-4H]^{4-}$  et  $[M-3H]^{3-}$  ( $m/z = 376$  et  $m/z = 501,7$ ).

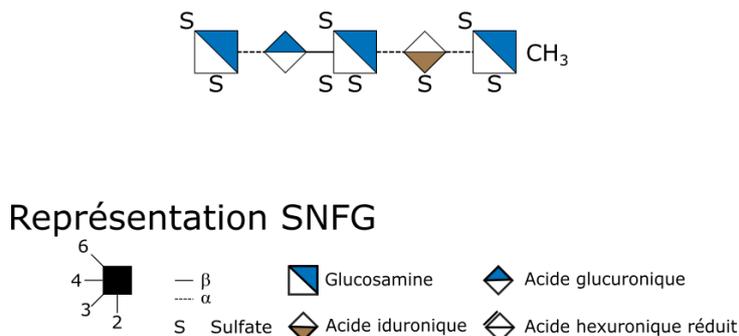


Figure 6 : structure selon la SNFG du Fondaparinux. Un groupement méthyl est présent en lieu et place de l'extrémité anomérique.

### 3.2 ACQUISITION DES SPECTRES EN INFUSION DIRECTE

La présence de nombreux sites acides sur la structure du fondaparinux (2 groupements carboxyle des acides hexuroniques et 8 groupements sulfate) fait qu'en solution, l'espèce peut être présente sous plusieurs degrés de déprotonation. Typiquement, le Fondaparinux est observé en MS à  $m/z = 376$  et  $m/z = 501,7$  ce qui correspond respectivement à  $[M-4H]^{4-}$  et  $[M-3H]^{3-}$ . Ainsi nous avons acquis les spectres IRMPD du Fondaparinux sous ces deux formes (Figure 7).

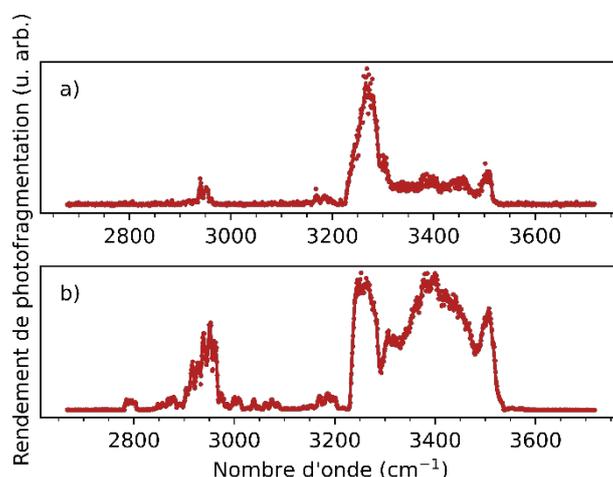
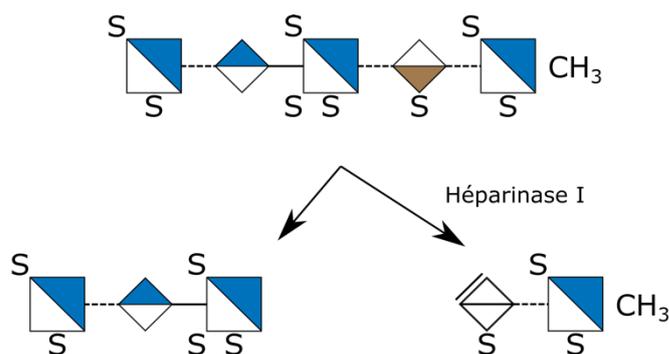


Figure 7 : Spectres IRMPD du Fondaparinux en mode négatif des espèces quatre fois déprotonée  $[M-4H]^{4-}$  a) et trois fois déprotonée  $[M-3H]^{3-}$  b).

L'espèce quatre fois déprotonée (Figure 7a) présente un pic d'absorption majoritaire à  $3270\text{ cm}^{-1}$ , deux pics mineurs sont également présents à  $3177$  et  $2946\text{ cm}^{-1}$  de l'autre côté, une large bande d'absorption faible est délimitée entre le pic majoritaire et un pic pas entièrement résolu à  $3496\text{ cm}^{-1}$ . L'espèce 3 fois déprotonée présente une bande très intense et large (autour de  $3387\text{ cm}^{-1}$ ), trois autres pics intenses sont également présents à  $2946$ ,  $3253$  et  $3504\text{ cm}^{-1}$  d'autres pics mineurs sont aussi présents entre  $3000$  et  $3200\text{ cm}^{-1}$  (Figure 7b).

### 3.3 LES DIGESTIONS

Certaines enzymes peuvent cliver le Fondaparinux, c'est le cas de certaines héparinases dont l'héparinase I et l'héparinase IV (cette dernière est aussi appelée héparanase). L'héparinase I clive le Fondaparinux au niveau de la liaison glycosidique ( $\alpha$ ) entre la glucosamine trois fois sulfatée et l'acide iduronique sulfaté (Figure 8). L'acide iduronique est alors modifié avec apparition d'une double liaison, le trisaccharide lui n'est pas modifié. Après digestion, les deux composés sont séparés par chromatographie liquide. Les fractions contenant ces composés sont collectées, afin d'acquérir leur spectre IRMPD en infusion directe.



## Représentation SNFG

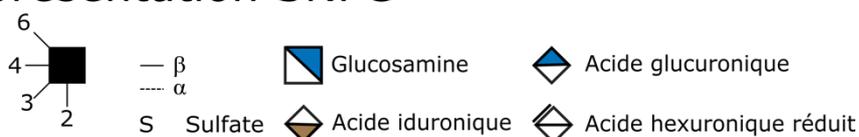


Figure 8 : Représentation selon la SNFG du Fondaparinux et des produits de digestion du Fondaparinux par l'héparinase I. Deux produits de digestion sont obtenus un trisaccharide côté extrémité « non réductrice » et un disaccharide côté extrémité « réductrice », ce disaccharide est modifié lors de la digestion enzymatique, une double liaison est alors formée.

Le trisaccharide obtenu est observé à  $m/z = 457$  ( $[M-2H]^{-2}$ ). L'autre produit de digestion est, lui détecté à  $m/z = 504$ .

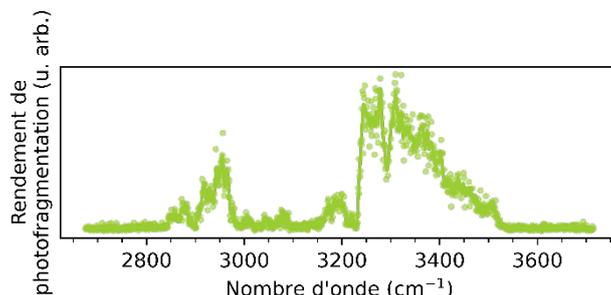
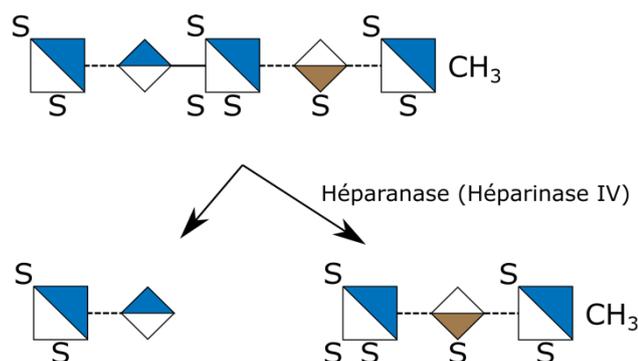


Figure 9 : Spectre IRMPD en mode négatif du produit de digestion du Fondaparinux extrémité « réductrice »

Le spectre IRMPD du produit de digestion du Fondaparinux détecté à  $m/z = 504$  est présenté en Figure 9. Ce spectre présente deux massifs dans la région des -CH : deux doublets, l'un autour de  $2854 \text{ cm}^{-1}$  ( $2849$  et  $2859 \text{ cm}^{-1}$ ) et l'autre autour de  $2876 \text{ cm}^{-1}$  ( $2872$  et  $2880 \text{ cm}^{-1}$ ) un doublet à  $2922 \text{ cm}^{-1}$  ( $2916$ ,  $2928 \text{ cm}^{-1}$ ) suivi d'un triplet à  $2956 \text{ cm}^{-1}$  ( $2943$ ,  $2953$  et  $2965 \text{ cm}^{-1}$ ). La vallée entre  $3000$  et  $3122 \text{ cm}^{-1}$  est encombrée de pics d'intensité modeste ( $3002$ ,  $3040$  et  $3076 \text{ cm}^{-1}$ ) avant de rencontrer un pic plus intense à  $3187 \text{ cm}^{-1}$ . Finalement un massif important est présent au-dessus de  $3228 \text{ cm}^{-1}$  avec un pic plus élevé à  $3258 \text{ cm}^{-1}$  suivi d'un creux ( $3291 \text{ cm}^{-1}$ ). Un second pic imposant se trouve à  $3311 \text{ cm}^{-1}$  avant que le rendement de photofragmentation ne diminue progressivement jusqu'à  $3534 \text{ cm}^{-1}$  (avec quelques ressauts à  $3362$  et  $3502 \text{ cm}^{-1}$ ).

Une autre enzyme de la famille des héparinases est aussi capable de digérer le Fondaparinux : l'héparinase IV. Les produits obtenus sont là aussi un disaccharide et un trisaccharide, cette fois-ci, le disaccharide est présent à l'extrémité « non réductrice » et le trisaccharide est du côté de l'extrémité « réductrice » (Figure 10). Lors de cette digestion, la structure de ces deux produits n'est pas modifiée.



## Représentation SNFG

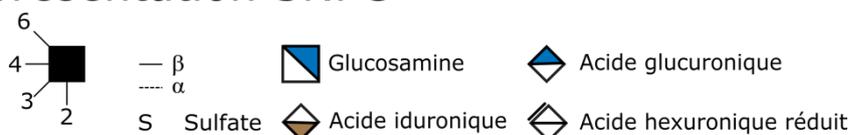


Figure 10 : Représentation selon la SNFG du Fondaparinux et des produits de digestion du Fondaparinux par l'héparinase IV. Digestion du Fondaparinux par l'héparinase IV. Deux produits de digestion sont obtenus un trisaccharide côté extrémité « réductrice » et un disaccharide côté extrémité « non réductrice ».

L'héparinase IV hydrolyse le Fondaparinux en deux produits d'une part l'extrémité « non réductrice » est un disaccharide contenant une glucosamine deux fois sulfatée et un acide glucuronique (Figure 10) avec  $[M-H]^-$   $m/z = 514$ . L'extrémité « réductrice » est un trisaccharide contenant 2 glucosamines sulfatées et l'acide iduronique lui aussi sulfaté avec  $[M-3H]^{3-} = 335$   $m/z$ . Les deux produits ont été séparés l'un de l'autre par chromatographie liquide avec collecte de fractions afin de purifier ces composés. Les spectres IRMPD de ces espèces sont présentés dans la Figure 11a pour le disaccharide et b pour le trisaccharide).

Le disaccharide issu de la digestion par l'héparinase IV (GlcN sulfatée en 2 et 6 liée à un acide glucuronique) est détecté à  $m/z = 514$  (espèce une fois déprotonée). Son spectre IRMPD (Figure 11a) présente deux pics principaux  $3293$  et  $2947$   $cm^{-1}$ . Le pic à  $2947$   $cm^{-1}$  présente un épaulement autour de  $2909$   $cm^{-1}$ . Une bande assez large mais faible en intensité se dessine entre  $3121$  et le pic à  $3293$   $cm^{-1}$ . Enfin un pic faible en intensité est présent à  $3508$   $cm^{-1}$ . Le signal du trisaccharide issu de la digestion du Fondaparinux est observé à  $m/z = 335$ , il s'agit du trisaccharide  $[M-3H]^{3-}$  (GlcN sulfatée en 2, 3 et 6 liée à un acide iduronique sulfaté en 2 lui-même lié à une GlcN sulfatée en 2 et 6 et méthylée en 1). Cet ion est isolé afin d'acquérir son spectre IRMPD (Figure 11b). Le spectre ainsi obtenu présente des pics à  $2943$  et  $3259$   $cm^{-1}$ , un large massif entre  $3312$  et  $3476$   $cm^{-1}$  enfin un pic est également présent à  $3510$   $cm^{-1}$ .

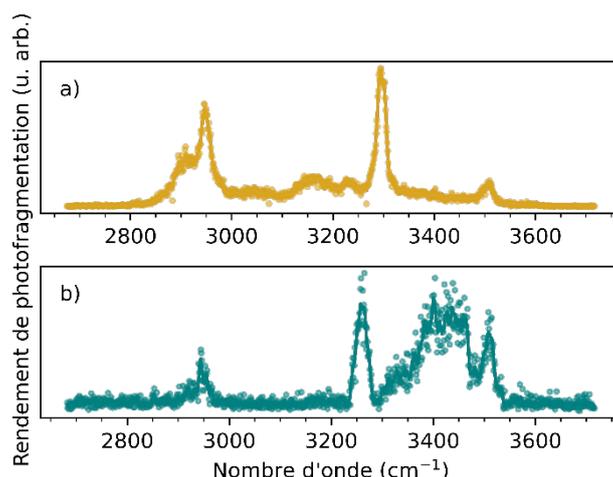


Figure 11 : Spectre IRMPD des produits de digestion du Fondaparinux par l'héparinase IV disaccharide à l'extrémité « non réductrice » a) et trisaccharide présent au niveau de « l'extrémité réductrice » b).

### 3.4 LA METHODE CHROMATOGRAPHIQUE

Comme nous l'avons vu dans la partie précédente, des développements instrumentaux ont été réalisés dans l'équipe d'un point de vue de la rapidité d'acquisition des données<sup>31,32</sup>. Les spectres IRMPD sont maintenant acquis en seulement quelques secondes. Les échelles de temps entre LC et spectroscopie IRMPD sont compatibles. Après avoir acquis les spectres IRMPD des différents produits de digestion du Fondaparinux, nous avons voulu voir jusqu'où pouvait aller notre méthode d'IRMPD afin de savoir si nous étions capables d'acquérir un spectre IRMPD en ligne après séparation chromatographique. Pour cela, nous avons utilisé une méthode chromatographique basée sur de la HILIC. Brièvement, cette méthode consiste en un gradient durant 13 minutes (Figure 12) pendant lesquelles, la proportion de solvant B passe de 65 % à 45 %, avant passage à 100 % de A puis rééquilibrage de la colonne de 15 à 20 min. Le débit est de 200  $\mu\text{L}\cdot\text{min}^{-1}$ . Les solvants sont A : eau contenant de l'ammonium formate (7,5 mM) ajusté à pH 4,0 avec de l'acide formique et B : acétonitrile pendant la séparation, la colonne est maintenue à 20 °C, la spectrométrie de masse est utilisée en mode négatif. La colonne est une ZIC-cHILIC de chez SeQuant dont les dimensions sont L = 150 mm, d = 2,1 mm avec des particules de 3  $\mu\text{m}$  de diamètre.

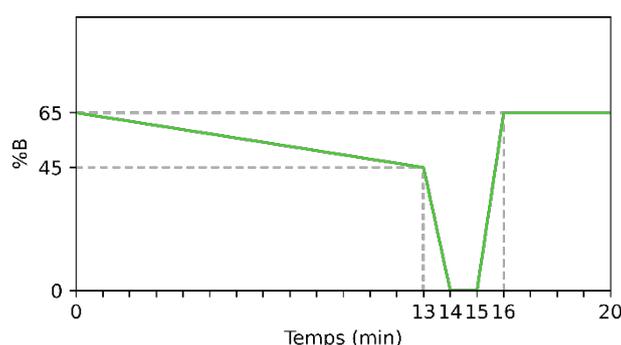


Figure 12 : Représentation du gradient utilisé pendant la séparation chromatographique des produits de digestion du fondaparinux par les deux héparinases utilisées. La proportion de phase mobile B passe de 65 à 45 % en treize minutes. La proportion de B diminue devient nulle en une minute, elle reste nulle pendant une autre minute avant de remonter à la proportion initiale pendant quatre minutes. La phase mobile B est composée d'acétonitrile pendant que le reste de la phase mobile (A) est une solution d'ammonium formate (7,5 mM) ajustée à pH 4,0.

Le chromatogramme de la digestion du Fondaparinux est présenté en Figure 13. On distingue 3 pics tout d'abord, le produit de digestion par l'héparinase IV sur le Fondaparinux ( $m/z = 514$ ) est observé à 4,07 min, ensuite un composé à  $m/z = 464$  est détecté à 7,30 min enfin, le produit de la digestion du

Fondaparinux par l'héparinase 1 sort de la colonne à 8,91 min. Il s'agit d'un chromatogramme dans lequel le mode single ion monitoring est utilisé, les ions sont suivis pendant un certain temps avant de suivre un autre ion, c'est pourquoi on observe des brisures abruptes du signal après les pics correspondants aux ions à  $m/z = 464$  et  $m/z = 504$ .

Le but est alors d'isoler chacun des ions lors de leur arrivée dans le spectromètre de masse afin de pouvoir l'irradier et d'acquérir son spectre de masse. Le chromatogramme est ici découpé en trois segments de 0 à 5,5 min on isole le pic à 514, de 5,5 à 8 min on isole le composé à  $m/z = 464$  et enfin le dernier segment de 8 min à la 10 min suit le composé à  $m/z = 504$ .

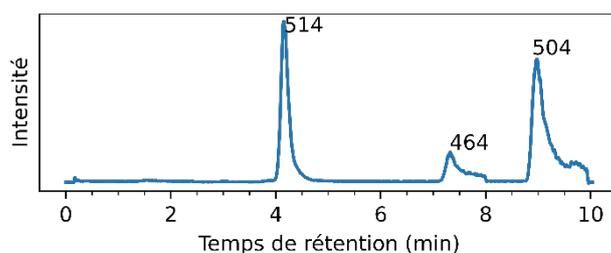


Figure 13 : Chromatogramme des composés issus de la digestion du Fondaparinux par l'héparinase I et l'héparinase IV. Ce chromatogramme présente 3 pics séparés.

### 3.5 RESULTATS

Les pics chromatographiques durent de plusieurs dizaines de secondes à une minute, ce qui donne suffisamment de temps pour acquérir une portion du spectre IRMPD. Nous avons choisi d'acquérir une gamme de nombres d'onde comprise entre  $3224$  et  $3382$   $\text{cm}^{-1}$  sur le composé à  $m/z = 504$ . De cette manière, le scan des nombres d'onde n'est pas trop rapide, et l'échantillonnage demeure assez important. En agissant de la sorte, le nombre de points présents sur le spectre IRMPD acquis en couplage avec la chromatographie liquide est de 171 pour  $158$   $\text{cm}^{-1}$  (1,08 point par  $\text{cm}^{-1}$ ) ce qui est similaire à l'échantillonnage obtenu en infusion directe (1112 points pour le spectre acquis en infusion ce qui représente un peu plus de 1,11 point pour  $1$   $\text{cm}^{-1}$ ). A l'instar de Rizzo *et al.*<sup>33</sup> et de Oomens *et al.*<sup>34</sup> qui sont parvenus à acquérir des spectres IR en phase gazeuse en ligne après séparation des composés en chromatographie liquide, nous avons acquis les spectre IRMPD de composés présents en mélange dans une solution après séparation par LC en ligne. La Figure 14 représente le spectre obtenu en infusion directe (courbe verte) et le spectre obtenu lors du couplage entre la méthode séparative avec la spectroscopie IRMPD (courbe et points noirs).

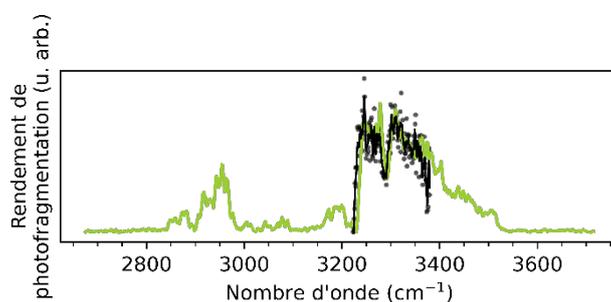


Figure 14 : Spectre IRMPD du produit de digestion à  $m/z = 504$  en infusion directe (vert) et portion de spectre IRMPD du produit de digestion à  $m/z = 504$  en couplage chromatographie liquide-spectroscopie IRMPD (noir, ligne et pointillés).

L'acquisition des données se fait manuellement, ce qui explique un léger décalage de quelques  $\text{cm}^{-1}$  entre le spectre obtenu en couplage et le spectre obtenu en infusion directe. Excepté ce décalage, ces deux spectres présentent un bon accord. En effet, les principales caractéristiques du spectre IRMPD du spectre de référence sont observées avec notamment le premier pic principal observé à  $3256$   $\text{cm}^{-1}$

(observé en infusion directe à  $3258\text{ cm}^{-1}$ ) suivi du creux à  $3289\text{ cm}^{-1}$  (contre  $3292\text{ cm}^{-1}$  en infusion directe) avant le pic à  $3301\text{ cm}^{-1}$  (observé en infusion directe à  $3309\text{ cm}^{-1}$ ) suivie de la diminution du rendement de photofragmentation après ce pic.

Ici nous avons vu que nous sommes capables de procéder à l'acquisition d'une portion de spectre IRMPD après séparation en LC, avec l'application sur des produits de digestion du Fondaparinux au sein de notre équipe. Nous passons maintenant sur une autre classe de sucres analysés au moyen d'un autre mode de chromatographie liquide.

## 4 ANALYSE EN LIGNE DES COMPOSES EN SORTIE DE COLONNE : EXEMPLE SUR LES HMOs

---

Dans cette partie, les composés analysés ne proviennent pas de la digestion d'un unique oligosaccharide, mais il s'agit d'un mélange d'oligosaccharides. Ici la nature de la colonne de chromatographie est différente de la chromatographie HILIC. Les conditions de spectroscopie ont permis d'acquérir une gamme de  $500\text{ cm}^{-1}$  sur un pic chromatographique durant moins de 20 secondes (plus fin que les pics présentés dans la partie 3). Le premier mélange est constitué de deux isomères de groupe sanguin différant l'un de l'autre par une liaison. Le deuxième mélange est composé de deux HMOs (Human Milk Oligosaccharide) des sucres sont présents dans le lait maternel. Ces sucres contribuent au bon développement du nouveau-né<sup>35-37</sup>.

Tout d'abord, un mélange de deux tétrasaccharides isomères l'un de l'autre au niveau d'une liaison glycosidique est analysé. Ces tétrasaccharides sont constitués de deux galactoses, d'un N-acétylglucosamine et d'un fucose. La chromatographie liquide sépare les deux isomères, la fragmentation ( $MS^3$ ) de ces deux tétrasaccharides permet d'obtenir le disaccharide contenant la liaison responsable de l'isomérisation. Enfin, la spectroscopie IRMPD est appliquée sur ce disaccharide.

Ensuite, deux composés sont choisis le Lacto-N-triaose de type 2 (LNT2) et le 2'Fucosyllactose (2'FL). Ces deux composés sont en solution, ils présentent donc un mélange de deux anomères. Cette anomérie se traduit par la présence de deux pics dans le mode de chromatographie utilisée. Pour lever cette anomérie, une réduction chimique est requise. Ainsi un seul pic sera observé pour chaque oligosaccharide, ce qui simplifie l'interprétation des données et augmente le signal. Après séparation chromatographique de ce mélange, une analyse en spectroscopie IRMPD est appliquée aux deux ions.

Enfin, des mesures ont été faites sur les composés réduits : trois hexoses afin de vérifier que les spectres IRMPD de ces composés soient bien différents. Un nouveau mode d'irradiation est décrit pour détecter un monosaccharide bien particulier dans un di ou trisaccharide.

### 4.1 LA METHODE CHROMATOGRAPHIQUE

La méthode utilisée ici repose sur une colonne dont la phase stationnaire est du carbone graphite : PGC, une technique de choix pour séparer des composés polaires<sup>38</sup> dont les sucres<sup>9,39</sup>.

Les colonnes de PGC sont basées sur une phase stationnaire en carbone, contrairement aux colonnes plus « conventionnelles » qui contiennent elles de la silice. Ce type de colonne a l'avantage de présenter une très bonne stabilité chimique, elles peuvent donc être utilisées avec de plus grandes gammes de pH et à plus haute température<sup>40</sup> comparées à des colonnes reposant sur des particules de silice. Les mesures effectuées avec des colonnes de PGC ne nécessitent pas de marquer le sucre, aucune modification du composé pour améliorer la rétention des sucres ou leur ionisation après la

séparation n'est nécessaire comme ça pourrait être le cas lors de l'utilisation des colonnes de type RPLC (Reverse Phase Liquid Chromatography)<sup>41</sup> ou HILIC.

La colonne utilisée est une colonne Hypercarb de chez Thermo Fisher Scientific dont les dimensions sont les suivantes : L = 10 cm, diamètre interne 1 mm avec des particules de 3  $\mu\text{m}$  de diamètre. Le débit est fixé à 200  $\mu\text{L}\cdot\text{min}^{-1}$  le four est maintenu à 30 °C. Le gradient se fait entre deux solvants, A : eau et B : acétonitrile, dans chacun des deux solvants 0.1 % d'acide formique est ajouté. Au début de la séparation, la phase mobile est constituée de 100 % de A, le gradient monte à 12 % de B en 10 min, puis 40 % de B en 21 min avant de rejoindre 100 % de B entre 26 min et 31 min. Le rééquilibrage se fait alors à 0 % de B (Figure 15). Le gradient choisi n'est pas critique dans notre cas, nous ne cherchons à séparer que quelques espèces présentes dans un mélange. La détection par spectrométrie de masse se fait en mode positif en ne cherchant que les rapports m/z des ions d'intérêt.

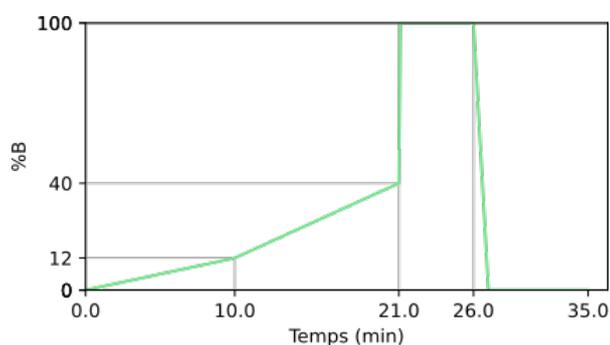


Figure 15 : Gradient utilisé pour la séparation à l'aide de la colonne PGC. La solution B est constituée de 0,1 % d'acide formique dans de l'acétonitrile (volume/volume), tandis que la solution A est constituée de 0,1 % d'acide formique dans de l'eau (volume/volume). La proportion de B passe de 0 à 12 % en 10 minutes, avant d'augmenter à 40 % en 11 minutes. Le lavage de la colonne dure 5 minutes avant de revenir aux conditions initiales pendant neuf minutes.

## 4.2 PREMIERS RESULTATS DU COUPLAGE LC-MS/MS-IR

### 4.2.1 Intérêt de fragmenter le composé d'intérêt

Comme nous le voyons depuis le début de ce manuscrit, de nombreuses isoméries sont présentes dans la structure des sucres. C'est notamment le cas pour les antigènes de groupe sanguins de type 1 et 2. Ces deux isomères contiennent un fucose, deux galactoses et un N-Acétylglucosamine et ne varient que de la régiochimie d'une liaison : la liaison entre le GlcNAc et le Gal comme c'est visible (en rouge) dans la Figure 16.

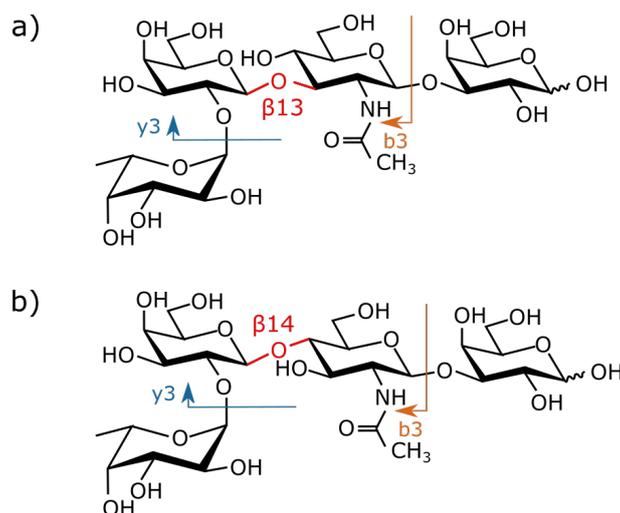


Figure 16 : Structure des antigènes de groupe sanguin H de type 1 a) et 2 b) sous forme tétrasaccharide. Ces deux isomères ne diffèrent que de la régiochimie de la liaison glycosidique entre le galactose et le N-Acétylelucosamine (représentée en rouge). Deux étapes de CID éliminent le galactose à l'extrémité réductrice, puis le fucose (fragment b3/y3 est symbolisé par les flèches bleues et oranges).

#### 4.2.2 Mesures des spectres IRMPD de référence

Les spectres IRMPD des références sur les disaccharides Gal $\beta$ 1-3GlcNAc et Gal $\beta$ 1-4GlcNAc sont acquis après élimination de l'hydroxyle anomérique (fragmentation du disaccharide protoné) pour obtenir l'espèce  $[M - 18 + H]^+$ , les spectres IRMPD de ces deux espèces sont présentés en Figure 17. Les deux signatures infrarouges sont similaires, avec trois maxima d'absorption dans la gamme des C-H (2921, 2956 et 2991  $\text{cm}^{-1}$  pour l'espèce Gal $\beta$ 1-3GlcNAc et 2913, 2953 et 2991  $\text{cm}^{-1}$  pour l'espèce Gal $\beta$ 1-4GlcNAc). Une vallée est ensuite visible entre ce massif et les prochaines bandes d'absorption au-dessus de 3350  $\text{cm}^{-1}$ . Le pic est large dans le cas de l'espèce Gal $\beta$ 1-3GlcNAc (bande de 3332 à 3519  $\text{cm}^{-1}$ ) tandis que deux pics pas entièrement résolus se distinguent pour l'espèce Gal $\beta$ 1-4GlcNAc (3379 et 3452  $\text{cm}^{-1}$ ) enfin deux derniers pics non résolus sont visibles dans les deux cas (3582 et 3634  $\text{cm}^{-1}$  pour Gal $\beta$ 1-3GlcNAc et 3590 et 3632  $\text{cm}^{-1}$  pour Gal $\beta$ 1-4GlcNAc. Il n'est donc pas pertinent d'acquérir l'intégralité de la gamme couverte par le laser, là encore, nous ne nous intéresserons qu'à la gamme comprise entre 3200 et 3700  $\text{cm}^{-1}$ .

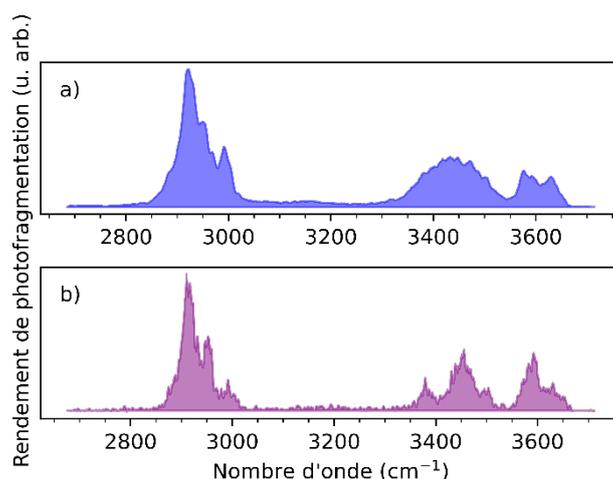


Figure 17 : Spectre de référence des deux disaccharides Gal $\beta$ 1-3GlcNAc a) et Gal $\beta$ 1-4GlcNAc b). Les deux spectres sont similaires, avec un massif dans la région des C-H (inférieure à 3000  $\text{cm}^{-1}$ ) et une signature qui diffère légèrement dans la gamme des O-H (supérieure à 3400  $\text{cm}^{-1}$ ).

#### 4.2.3 Séparation chromatographique des deux isomères

Afin d'étudier ces deux espèces en ligne, elles sont préparées dans une même solution à 50  $\mu\text{M}$ . La séparation chromatographique se fait selon les mêmes conditions que celles présentées en Figure 15. Après séparation, chacun des composés est fragmenté au moyen de deux étapes de CID successives. La première fragmentation élimine le galactose situé à l'extrémité réductrice (fragment b3), ce fragment est ensuite de nouveau fragmenté pour éliminer cette fois-ci le fucose. Après ces deux étapes de fragmentation seul le cœur du tétrasaccharide qui contient l'isomérisation (fragment  $\gamma_3/\text{b}_3$ ) est étudié (espèce à  $m/z = 366$  voir Figure 16). Les deux étapes de fragmentation successives allongent la durée d'acquisition des données. Toutefois, nous avons eu l'idée de compenser cette longue durée d'acquisition en augmentant artificiellement le temps de signal de chacune des espèces. Étant donné que les deux étapes de fragmentation éliminent l'extrémité anomérique des tétrasaccharides il n'y a pas besoin de réduire les composés. Le chromatogramme obtenu est présenté en Figure 18, chacune des deux espèces est présente sous deux pics, les deux anomères sont observés, ces deux pics ne sont pas complètement résolus, le fait d'observer deux pics permet d'observer du signal d'ions plus longtemps (Figure 18).

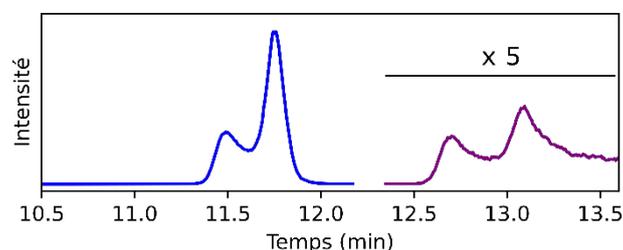


Figure 18 : Représentation du chromatogramme obtenu après séparation du mélange des deux isomères de tétrasaccharides BGHI (bleu) et BGHII (violet) dans les conditions présentées précédemment.

#### 4.2.4 Mesure en ligne du spectre IRMPD du premier pic chromatographique

Le spectre IRMPD du fragment  $\gamma_3/\text{b}_3$  issu de la fragmentation par CID de BGHI est présenté en Figure 19. La courbe noire (spectre IRMPD moyenné sur 3 points consécutifs) présente 2 bandes d'absorption dans la région  $3200\text{--}3700\text{ cm}^{-1}$ , l'une entre  $3335$  et  $3528\text{ cm}^{-1}$  et l'autre entre  $3553$  et  $3663\text{ cm}^{-1}$ . Cette signature correspond à ce que nous avons observé lors de la prise des références (Figure 17a). Il existe tout de même une différence observée au niveau de l'intensité des pics à  $3582$  et  $3634\text{ cm}^{-1}$ , les intensités sont inversées, de plus les pics obtenus après séparation chromatographique semblent être élargis d'un côté du pic : vers les nombres d'onde plus faibles (élargissement de chacun des deux pics à  $3555$  et  $3368\text{ cm}^{-1}$ ). Cet écart entre les deux spectres peut-être dû au plus faible nombre de points acquis, donc plus faible échantillonnage lors de l'analyse en ligne.

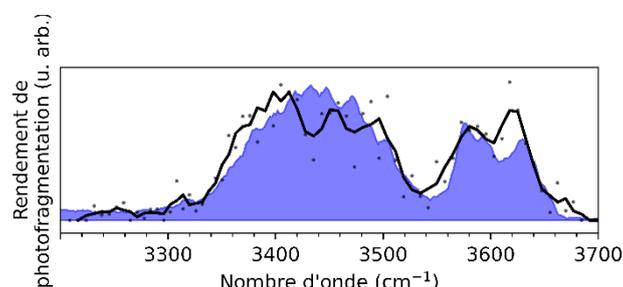


Figure 19 : Spectre IRMPD de la référence Gal $\beta$ 1-3GlcNAc obtenu en infusion directe après CID pour éliminer l'hydroxyle anomérique (fragment b2) fond bleu, spectre IRMPD du fragment  $\gamma_3/\text{b}_3$  du tétrasaccharide de BGHI obtenu après séparation chromatographique en quelques dizaines de secondes (points noirs, moyenne glissante sur 3 points : courbe).

### 4.3 REDUCTION DES COMPOSES

Comme vu sur la Figure 18, la chromatographie au moyen de la colonne de PGC sépare les anomères, ce qui induit un dédoublement des pics et une difficulté d'interprétation des données. Pour passer outre cette difficulté, il est nécessaire de réduire les oligosaccharides avant analyse en LC-MS.

Leur structure est présentée en Figure 20. Les HMOs partagent une structure commune : ils possèdent un résidu commun : le lactose (Gal $\beta$ 1-4Glc) au niveau de leur extrémité réductrice. Le LNT2 ([M+H]<sup>+</sup> m/z = 546) présente un GlcNAc lié au résidu lactose tandis que le 2'FL ([M+H]<sup>+</sup> = 489 m/z) contient un résidu fucose lié au lactose.

#### 4.3.1 Les composés

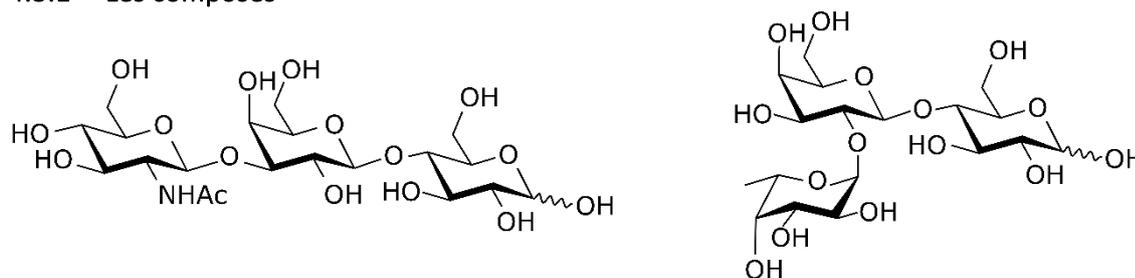


Figure 20 : Structures des deux HMOs : le LNT2 (gauche) et le 2'FL (droite), structures développées.

#### 4.3.2 La préparation des composés : réduction

La chromatographie liquide avec des colonnes de PGC sépare facilement les anomères. Pour un même sucre, les pics sont alors dédoublés. C'est pourquoi lors d'analyses LC-MS en glycomique avec de telles colonnes, une étape de réduction est nécessaire afin d'éliminer l'anomérisation et de diminuer le nombre de pics pour augmenter la sensibilité et simplifier l'interprétation des données<sup>9,42</sup>.

Afin de réduire les trisaccharides, nous nous sommes inspirés du protocole présenté par Jensen *et al.*<sup>42</sup>. La réduction se fait en milieu basique à l'aide de borohydrure de sodium : NaBH<sub>4</sub>. La solution de réduction est préparée : 1 M de NaBH<sub>4</sub> et 0.5 M de NaOH. Dans notre cas, nous avons préparé 50  $\mu$ g de chaque composé auquel nous avons ajouté 20  $\mu$ L de la solution de réduction. La réduction est faite en laissant la solution à incuber pendant 3 h à 50 °C. 2  $\mu$ L d'acide formique glacial sont ajoutés afin d'arrêter la réaction (un précipité blanc se forme) avant d'ajouter 1 mL d'eau. Il est alors nécessaire de dessaler la solution pour éliminer toute impureté. Pour cela nous nous servons d'une colonne SPE de type PGC (Strata GCB de Phenomenex). Les colonnes sont conditionnées au moyen d'une solution d'acétonitrile (80 %) contenant 1 % d'acide formique. La solution de 1 mL d'eau contenant la solution de réduction est chargée sur la colonne. La colonne est ensuite rincée à l'eau enfin l'élution est effectuée avec de l'acétonitrile (40 %) contenant 0,5 % d'acide formique l'éluat est alors séché au moyen d'un SpeedVac après cette étape d'évaporation, deux options sont possibles, soit l'échantillon est repris dans les conditions voulues pour analyse (volume et solvant), soit il est conservé au congélateur à -20 °C.

Après purification et reprise dans le volume voulu, les composés sont infusés en spectrométrie de masse pour vérification de la réduction (Figure 21a et c) les ions précurseurs sont bien observés à M + 2 par rapport à ce que nous observions pour les espèces non réduites. Le spectre de fragmentation du LNT2 présente la perte du glucose réduit : fragment b<sub>2</sub> à 366 m/z tandis que le pic à m/z = 204 présente le GlcNAc ayant perdu une molécule d'eau (fragment b<sub>1</sub>). Le cas du 2'FL est encore plus intéressant. Le fragment à m/z = 345 résulte de la perte du fucose (fragment  $\gamma$ <sub>2</sub>) et le fragment à m/z = 183

représente le glucose réduit : fragment  $\gamma_1$ . Maintenant le fragment à  $m/z = 329$  provient de la migration du fucose depuis le galactose jusque sur le glucose réduit, phénomène qui peut se produire comme nous l'avons montré lors du marquage au  $^{18}\text{O}$  du 2'FL (dans le premier chapitre de ce manuscrit) mais qui est invisible dans le cas du 2'FL non modifié.

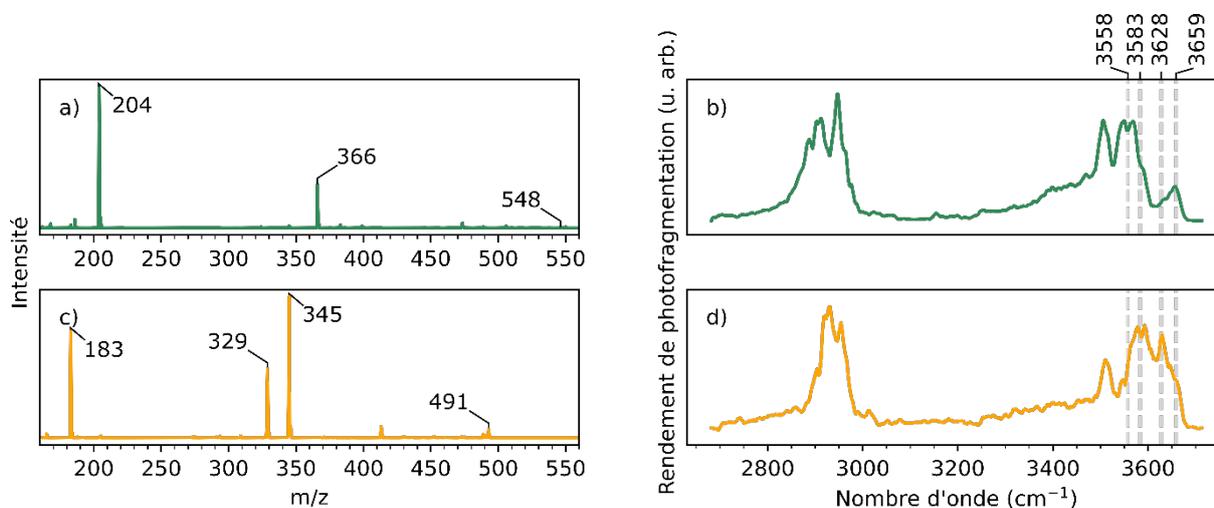


Figure 21 : Spectres de fragmentation (CID) du LNT2 a), du 2'FL c) et spectres IRMPD de ces deux mêmes espèces b) et d) respectivement. Les spectres de fragmentation montrent des fragments auxquels nous pouvons nous attendre (produits de la rupture de liaisons glycosidiques) dans les deux cas. Il est intéressant de noter dans le cas du 2'FL, la présence du pic à  $m/z = 329$  démontrant la migration du fucose de sa position initiale jusque sur le glucose réduit.

Les spectres IRMPD présentent une signature commune en dessous de  $3000\text{ cm}^{-1}$  avec trois pics à  $2945$ ,  $2910$  et  $2885\text{ cm}^{-1}$ , ensuite une large vallée est observée (avec un pic discret à  $3000\text{ cm}^{-1}$  caractéristique des composés fucosylés dans le cas du 2'FL) jusqu'à un nouveau maximum d'absorption à  $3506\text{ cm}^{-1}$  dans les deux cas. Ensuite les choses changent, un double pic centré autour de  $3558\text{ cm}^{-1}$  ( $3548$  et  $3566\text{ cm}^{-1}$ ) est observé pour le LNT2 réduit tandis qu'il est centré autour de  $3583\text{ cm}^{-1}$  pour le 2'FL ( $3576$  et  $3590\text{ cm}^{-1}$ ), le 2'FL présente un dernier pic d'absorption à  $3628\text{ cm}^{-1}$  tandis que le LNT2 présente un pic à  $3659\text{ cm}^{-1}$ . Comme nous venons de le voir, en dessous de  $3400\text{ cm}^{-1}$ , les deux spectres sont relativement similaires, il n'est donc guère intéressant d'enregistrer la signature IRMPD de ces composés dans cette région, en revanche la zone au-dessus de  $3400\text{ cm}^{-1}$  est bien plus intéressante à étudier.

Lors du couplage chromatographie liquide-spectroscopie IRMPD pour ces composés, nous n'acquerrons que la zone comprise entre  $3200$  et  $3700\text{ cm}^{-1}$ . Dans ce cas, le fait d'ignorer toute une gamme du spectre est intéressant car cela autorise un meilleur échantillonnage de la gamme d'intérêt, en effet sur un même intervalle de temps, autant de points sont acquis sur une gamme plus restreinte.

### 4.3.3 Résultats

#### 4.3.3.1 Séparation des deux composés

Le chromatogramme d'ions extraits des produits non réduits est présenté ci-dessous (Figure 22). L'ion suivi pour le 2'FL est l'ion à  $m/z = 489$  qui correspond à l'espèce protonée (orange sur la Figure), de même que pour le LNT2 (en vert sur la figure) à  $m/z = 546$ . Comme attendu, les espèces présentent chacune deux pics correspondant aux deux anomères.

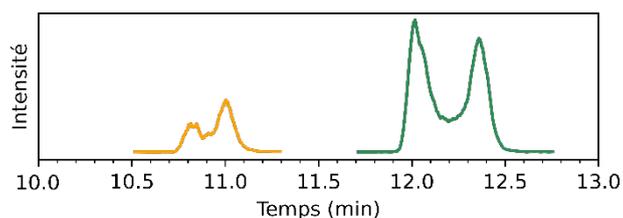


Figure 22 : Chromatogramme d'ions extraits de la séparation du LNT2 (vert) et du 2'FL (orange). Les pics apparaissent dédoublés, ces deux pics correspondent aux anomères.

La séparation des deux composés réduits est effectuée dans les mêmes conditions de chromatographie que celles présentées pour les oligosaccharides non réduits. Après réduction, nous voyons que l'ordre d'élution des composés a changé (Figure 23), le LNT2 réduit  $m/z = 548$  (vert) sort de la colonne avant le 2'FL réduit  $m/z = 491$  (orange). Après réduction, on n'observe maintenant plus qu'un pic par espèce, la réduction a bien levé l'isomérisation.

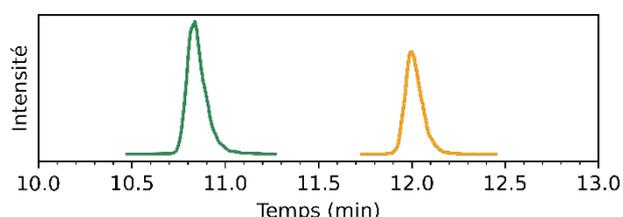


Figure 23 : Chromatogramme d'ions extraits des deux HMOs étudiés LNT2 ( $m/z = 548$ ) et 2'FL ( $m/z = 491$ ) respectivement en vert et orange. Il est intéressant de noter que l'ordre d'élution a changé après réduction des composés.

#### 4.3.3.2 Acquisition des spectres MS-IR

Après nous être assurés que les deux composés sont réduits, et qu'ils peuvent être séparés par LC, nous sommes passés à l'acquisition du spectre IRMPD des composés après séparation en LC. Sur la Figure 24, les ions parents ( $548$  et  $491$   $m/z$  pour le LNT2 et le 2'FL respectivement) sont présentés en noir tandis que les ions fragments sont présentés en gris, la couleur verte ou orange correspond à la somme des intensités entre l'ion parent et l'ion fragment (a et c). Le début de l'irradiation est déclenché manuellement, ce qui explique que le début de l'irradiation ne commence pas au début du pic chromatographique, mais après le maximum d'intensité du pic chromatographique. Les irradiations sont déclenchées manuellement car les temps de rétention n'étaient pas suffisamment répétables pour automatiser le début de l'acquisition du spectre IRMPD. Dans la deuxième partie de cette figure, les spectres IRMPD acquis en infusion sont présentés (fond vert ou orange) tandis que les spectres IRMPD acquis en ligne en sortie de chromatographie sont présentés par les points et les courbes. Le temps de début d'irradiation et la durée d'acquisition du spectre IRMPD sont relevés pour pouvoir effectuer la correspondance entre temps et nombres d'onde. Les spectres en ligne peuvent donc être tracés, ils ne sont pas moyennés car trop peu de points sont acquis.

Malgré un décalage entre le spectre acquis en infusion et le spectre acquis en ligne sur le pic d'absorption à  $3506$   $\text{cm}^{-1}$  qui se retrouve à  $3522$   $\text{cm}^{-1}$  le reste du spectre IRMPD du LNT2 présente un spectre IRMPD acquis en ligne similaire au spectre IRMPD acquis en infusion directe, le double pic centré à  $3558$   $\text{cm}^{-1}$  n'est pas aussi bien résolu visuellement, du fait d'un échantillonnage moindre, seule une cinquantaine de points est acquise en ligne contre environ 400 en infusion (Figure 24 b) le pic à  $3659$   $\text{cm}^{-1}$  se recouvre bien dans les deux cas. La Figure 24d présente le même décalage du pic à  $3506$   $\text{cm}^{-1}$  qui se retrouve cette fois-ci à  $3535$   $\text{cm}^{-1}$  dans le cas du 2'FL. Dans ce deuxième cas également, le doublet de pics à  $3583$   $\text{cm}^{-1}$  souffre d'un sous échantillonnage et ne présente qu'un pic. En revanche ces deux pics observés en infusion directe sont bien recouverts par l'enveloppe formée par le pic

d'absorption observé en ligne. Finalement, le pic observé à  $3628\text{ cm}^{-1}$  correspond bien entre les deux mesures, toutefois en ligne, le pic est bien élargi vers les fréquences plus élevées.

L'acquisition de données en ligne présente des spectres IRMPD similaires au spectres IRMPD acquis en infusion directe malgré quelques décalages dus au fait que l'acquisition ne soit pas encore été automatisée et se fait encore de manière manuelle.

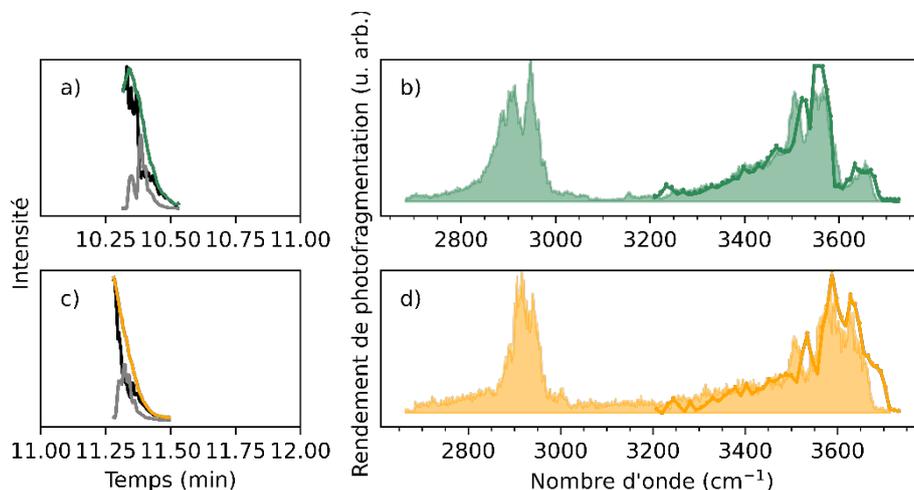


Figure 24 : Chromatogramme de l'ion parent (en noir) et de l'ion fragment (gris) et somme des intensités des ions parents et fragments (vert) du LNT2 a) et (orange) du 2'FL c) et spectre IRMPD en infusion (fond coloré) et en ligne (ligne et points) du LNT2 b) et du 2'FL d). L'irradiation a démarré au niveau du pic d'intensité des analytes.

Pour conclure sur ces deux espèces, le couplage entre chromatographie liquide et spectroscopie IRMPD permet d'acquérir des spectres en ligne qui présentent un bon recouvrement à l'exception d'un décalage observé pour le pic à  $3506\text{ cm}^{-1}$  qui est décalé dans le même sens dans les deux cas.

La séparation et l'analyse des ions est possible comme nous l'avons vu dans ces deux exemples. Maintenant nous allons voir que d'autres modes en MS sont également prometteurs.

## 4.4 NOUVEAUX DEVELOPPEMENTS CONDUITS SUR LE COUPLAGE LC-MS-IR

### 4.4.1 Spectres IRMPD de monosaccharides réduits : le cas des hexoses

Les spectres IRMPD des hexoses présentent des signatures différentes. Toutefois comme nous l'avons vu ci-dessus (Figure 18) les chromatogrammes sont plus facilement interprétables quand l'anomérisation est levée. Il est alors nécessaire de procéder à la réduction des sucres analysés. La réduction des sucres modifie sa structure avec notamment l'ouverture du cycle. La spectroscopie IRMPD est-elle capable de distinguer deux hexoses réduits ? Pour répondre à cette question, nous avons procédé à la réduction de ces espèces et avons acquis leur empreinte IR. Le mécanisme de la réduction est présenté en Figure 25 avec la réduction du glucose a), l'espèce réduite du glucose (sorbitol/glucitol) b), du galactose (galactitol) c) et du mannose (mannitol) d).

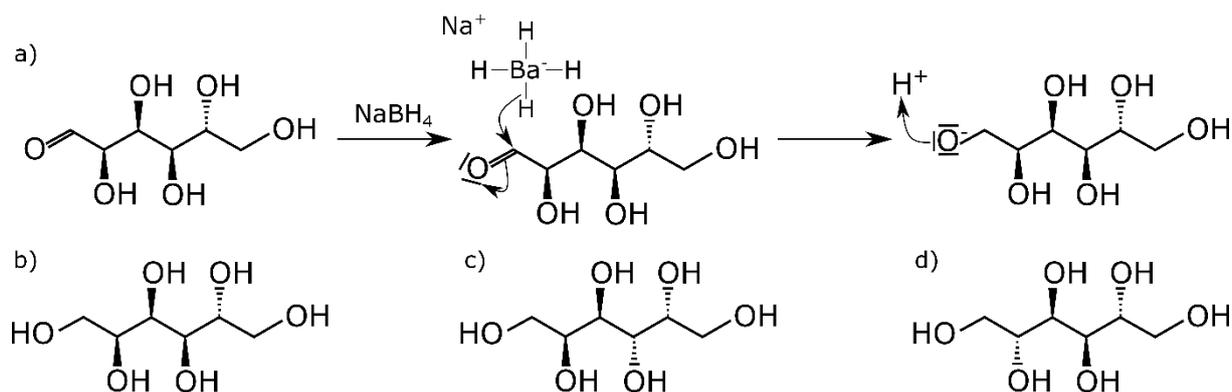


Figure 25 : Mécanisme de réduction au moyen de  $\text{NaBH}_4$  sur le glucose a), produit de la réduction sur le glucose b), sur le galactose c) et sur le mannose d).

Cette réduction est une étape supplémentaire dans la préparation d'échantillons et pose de nouvelles questions. La réaction de réduction au  $\text{NaBH}_4$  réduit l'aldéhyde porté par le monosaccharide. Nous voulons savoir s'il est possible de discriminer les produits de réduction au moyen de la spectroscopie IRMPD. Pour cela, les sucres sont réduits et les empreintes IRMPD des sucres réduits sont acquises. Les spectres IRMPD des espèces réduites protonées ( $m/z = 183$ ) sont présentés en Figure 26.

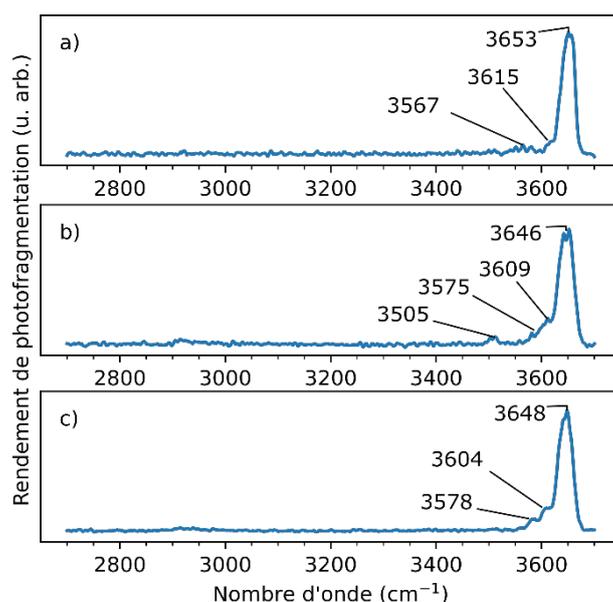


Figure 26 : Spectre IRMPD dans la gamme des 3  $\mu\text{m}$  obtenus en infusion directe des hexoses réduits protonés : sorbitol a), galactitol b) et mannitol c).

Comme nous l'avons vu précédemment dans la partie 2, la spectroscopie IRMPD fait la différence entre les hexoses lithiés. Dans le cas des hexoses réduits protonés, il existe également des différences entre ces trois espèces bien que ténues. Les spectres présentent tous une signature similaire, un maximum d'absorption aux alentours de 3650  $\text{cm}^{-1}$  (3653  $\text{cm}^{-1}$  pour le sorbitol Figure 26a), 3646  $\text{cm}^{-1}$  pour le galactitol Figure 26b) et 3648  $\text{cm}^{-1}$  pour le mannitol Figure 26c). Le sorbitol présente un pic relativement fin avec un épaulement discret (3615  $\text{cm}^{-1}$ ) ainsi qu'une bande d'absorption dont l'intensité est très faible à 3567  $\text{cm}^{-1}$ . Les spectres du galactitol et du mannitol eux présentent deux épaulements observés à 3575  $\text{cm}^{-1}$  et 3609  $\text{cm}^{-1}$  pour le galactitol et 3578  $\text{cm}^{-1}$  et 3604  $\text{cm}^{-1}$  pour le mannitol. Ces épaulements permettent de différencier le sorbitol des deux autres composés. Ces deux

derniers sont différenciables grâce à la présence d'un pic d'absorption à  $3505\text{ cm}^{-1}$  dans le spectre IRMPD du galactitol, bande qui est absente du spectre du mannitol.

Afin de confirmer ces observations (présence d'un épaulement sur le pic principal ou d'une bande d'absorption à  $3505\text{ cm}^{-1}$  dans le cas du galactitol), les spectres IRMPD du galactitol et du mannitol ont été mesurés de nouveau (Figure 27). Excepté un léger décalage en nombre d'onde, dans les deux cas, et un offset observé dans le cas du galactitol, les observations faites précédemment sont confirmées.

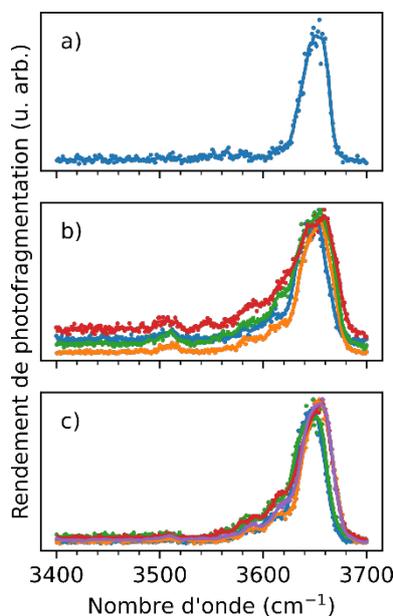


Figure 27 : Spectres IRMPD (entre  $3400$  et  $3700\text{ cm}^{-1}$ ) du sorbitol a), du galactitol b) et du mannitol c) les spectres ne sont acquis que dans la région des O-H.

La spectroscopie IRMPD est donc capable de distinguer les trois hexoses ayant subi une réduction. La distinction n'est faite qu'en se basant sur la région des modes de vibration des liaisons O-H (Figure 26) (les hexoses ne contiennent pas d'azote dans leur structure et les bandes d'absorption des modes de vibration des liaisons C-H ne sont pas visibles dans nos conditions). Ces informations sont contenues dans moins de  $250\text{ cm}^{-1}$  d'intervalle cela permet de ne se concentrer que sur 1/4 de la gamme habituellement considérée ce qui est intéressant dans le cas d'un couplage avec la LC pour lequel, la rapidité de l'acquisition est un des critères les plus importants.

#### 4.4.2 Détection d'une forme rare de galactose composant des polysaccharides bactériens par screening à longueur d'onde fixe

Un nouveau développement effectué au moyen du couplage entre la chromatographie liquide et la spectroscopie IRMPD est présenté dans cette partie : il s'agit d'un mode de screening. Ce développement a été effectué dans le cadre de la thèse de Ozgur Yeni<sup>43</sup> afin de détecter la présence d'un sucre rare dans de longues chaînes polysaccharidiques. Les composés sont séparés dans la colonne chromatographique (Figure 28a) avant d'être injectés dans le spectromètre de masse. Au sein du piège à ions, les composés d'intérêt sont irradiés par le laser IR à une longueur d'onde fixe (Figure 28b la longueur d'onde est non résonante en haut et la longueur d'onde est résonante en bas). En regardant le spectre de masse des composés irradiés, on observe ou non de la photofragmentation. Cette longueur d'onde est spécialement choisie pour entraîner la photofragmentation de l'espèce d'intérêt (spectres obtenus après irradiation Figure 28c).

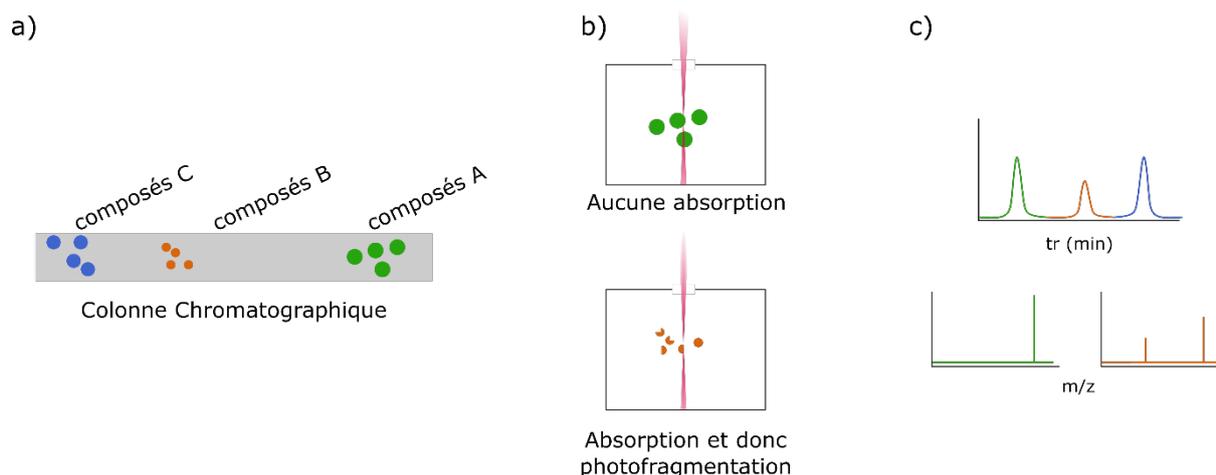


Figure 28 : Séparation de trois composés A, B et C au moyen de la chromatographie liquide a), irradiation de deux composés l'un n'absorbant pas à la longueur d'onde d'irradiation (haut) et l'autre absorbant à la longueur d'onde d'irradiation (bas) b) chromatogramme des trois composés (haut) et spectre obtenu après irradiation de deux de ces composés (bas) c).

Nous présentons ici une partie des développements instrumentaux effectués afin de détecter la présence d'une forme rare de galactose : le galactofuranose (Gal<sub>f</sub>) dont la structure est présentée ci-dessous.

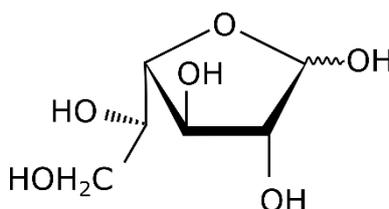


Figure 29 : Représentation de Haworth du galactofuranose.

Le galactose n'est pas présent sous cette forme chez les mammifères. Toutefois de nombreux organismes (virus, champignons, bactéries) synthétisent cette forme de galactose<sup>44</sup> (Figure 29). Le fait qu'une telle structure soit présente chez des organismes potentiellement pathogènes, mais absente chez les animaux, peut fournir une cible thérapeutique très intéressante contre ces organismes<sup>45,46</sup>, voire même de potentiels vaccins<sup>47</sup>. La masse d'un galactofuranose est celle d'un hexose, il est donc impossible de détecter la présence de ce composé parmi les autres hexoses au moyen de la spectrométrie de masse seule. Des mesures en spectroscopie IRMPD<sup>48-50</sup> ont été menées afin de discriminer la forme furane de la forme pyrane du galactose dans un échantillon.

Dans cette partie, des composés (di ou trisaccharides) contenant le mannose à l'extrémité réductrice et un galactofuranose ou un autre hexose (glucose, galactose ou mannose, tous sous forme pyranose) à l'extrémité non réductrice sont injectés en spectrométrie de masse. Le composé est en solution dans un mélange eau/méthanol (50/50 v/v) avec du LiCl afin de promouvoir la formation d'espèces lithiées  $[M + Li]^+$ . Ces espèces se chargent avec un adduit lithium, ce qui facilite leur ionisation et donc leur observation par spectrométrie de masse. Ces composés sont alors fragmentés par CID pour obtenir le fragment b1. Le fragment obtenu est l'espèce  $[Hex - 18 + Li]^+$   $m/z = 169$ . Ce produit de fragmentation est ensuite de nouveau isolé afin d'être irradié dans l'IR au moyen du laser accordable (gamme des 3  $\mu\text{m}$ ), ou du laser CO<sub>2</sub> (10,6  $\mu\text{m}$ ).

Nous présentons ici les résultats avec les conditions suivantes : nous avons irradié les ions avec le laser CO<sub>2</sub> (irradiation à 10,6  $\mu\text{m}$ ) pendant 130 ms, ici le laser délivre une puissance de 3.8 W, environ 40 spectres sont acquis. La Figure 30, compile les résultats obtenus sur les différents hexoses : glucose,

mannose, galactose sous sa forme pyranose (Galp) et galactose sous sa forme furanose (Galf) des ratios d'intensités entre ions fragments (If) et l'intensité de l'ion parent (Ip). Les ions à  $m/z = 169$  ne se fragmentent pas lorsqu'aucune étape d'irradiation n'est programmée.

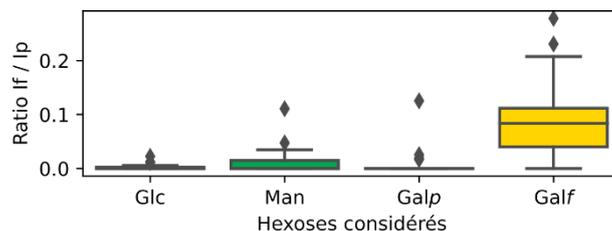


Figure 30 : Représentation des ratios entre l'intensité des ions fragments par rapport à l'ion précurseur pour les quatre hexoses considérés : glucose, mannose, galactopyranose et galactofuranose.

Dans ces conditions (irradiation avec une puissance de 3,8 W à 10,6  $\mu\text{m}$  pendant 130 ms) il apparaît clairement sur la Figure 30, que seul le Galf présente significativement de la photofragmentation. Des mesures effectuées avec d'autres puissances montrent des résultats semblables. En augmentant encore la puissance laser, les autres hexoses commencent à se photofragmenter.

Après intégration de ce type de screening dans un workflow de LC-MS, il a été possible de détecter des disaccharides contenant du galactofuranose après digestion de polysaccharides extraits de lichens récoltés dans la nature<sup>51</sup>.

## 5 CONCLUSION

Dans cette partie, nous avons vu le couplage entre la spectroscopie IRMPD avec la chromatographie liquide moyennant des développements instrumentaux et des compromis que ce soit d'un point de vue de la séparation des composés ou de l'acquisition du spectre IRMPD. Peu de groupes ont déjà mené à bien le couplage en ligne entre la chromatographie liquide et la spectroscopie vibrationnelle en phase gazeuse mais de tels développements sont prometteurs pour les domaines d'applications présentant de nombreuses isoméries telles que la glycomique, mais également comme on peut le retrouver dans d'autres classes de composés comme la métabolomique, et suscitera l'intérêt d'un nombre croissant de scientifiques.

Nous avons séparé des produits de la digestion enzymatique d'un pentasaccharide synthétique par HILIC avant d'acquérir le spectre IR en phase gazeuse de ces deux composés. De plus, une portion du spectre IR a été acquis au moyen du couplage LC-MS-IR en ligne. Plus récemment, l'acquisition du spectre IRMPD après séparation en PGC de HMOs réduits a aussi été réalisée.

Finalement de premiers résultats sur l'acquisition du spectre IRMPD d'un fragment issu de la CID en ligne après séparation chromatographique d'un mélange de tétrasaccharides ont été montrés.

Après avoir montré ces résultats, de nouveaux développements sont attendus, notamment l'automatisation de l'acquisition de spectres IRMPD. En effet, actuellement, l'opérateur déclenche l'acquisition du spectre IRMPD, en notant le début de l'acquisition pour tracer le spectre après acquisition. On pourrait imaginer faire un programme qui automatiserait l'acquisition du spectre IRMPD d'un composé inconnu, par exemple sous certaines conditions (niveau d'intensité du signal, stabilité du signal...) dès que la spectrométrie de masse détecte un rapport  $m/z$ , le programme déclenche l'acquisition du spectre IRMPD du composé inconnu. A l'aide d'un tel programme l'expérimentateur n'aurait pas besoin de procéder à une séparation initiale pour connaître les temps

de rétention du composé d'intérêt et tous les paramètres en rapport avec l'acquisition serait enregistrés afin de pouvoir tracer le spectre obtenu.

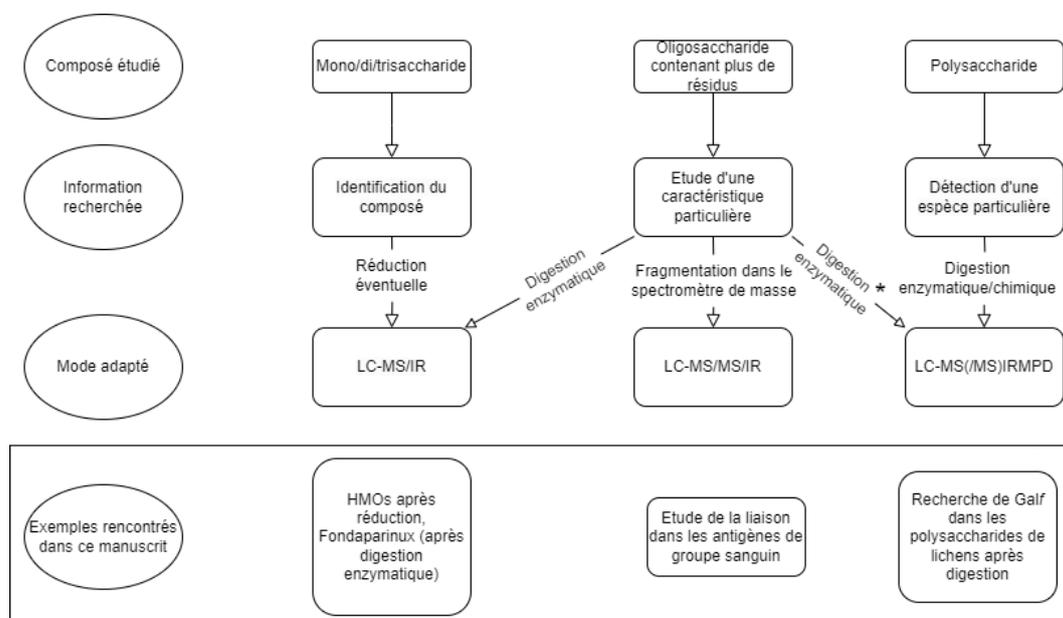
L'acquisition de spectres IRMPD de fragments après CID autorise l'acquisition des spectres IRMPD du monosaccharide situé à l'extrémité réduite (ou non en fonction du mode chromatographique choisi) de l'oligosaccharide afin de déterminer sa nature, avec une base de données de la plus petite taille possible. Cette mesure serait particulièrement intéressante dans le cas du séquençage d'un plus gros saccharide après hydrolyse chimique ou enzymatique.

Enfin concernant l'acquisition de spectres IRMPD de polysaccharides, un autre type de couplage a été implémenté dans lequel, le laser irradie les ions arrivant dans le spectromètre de masse à une longueur d'onde fixe. Il a été utilisé afin de détecter la présence de galactofuranose dans un di-/trisaccharide après fragmentation et sélection du résidu situé côté extrémité non réductrice. Le galactofuranose se photofragmente dans des conditions dans lesquelles les autres monosaccharides ne se photofragmentent pas<sup>51</sup>.

Ces deux nouveaux développements permettront de détecter le galactofuranose dans des produits de dégradation d'un polysaccharide et de caractériser l'extrémité « réductrice ». Dans le cas des polysaccharides contenant du galactofuranose, l'intégralité du produit de réaction pourra donc être identifié, ce qui nous l'espérons ouvrira la voie au séquençage de polysaccharides entiers. En effet, il en existe de nombreux dans la nature (chitosane, glycosaminoglycane...) et il demeure nécessaire de développer de nouveaux outils afin de les caractériser.

La détection d'une espèce particulière au moyen du screening IRMPD à longueur d'onde fixe est intéressante. Dans la partie suivante nous nous intéresserons de manière plus approfondie à une telle méthode. Elle fera l'objet d'une autre application, dans un tout autre domaine.

La Figure 31 représente un récapitulatif des modes présentés dans ce manuscrit en fonction des échantillons étudiés et des informations attendues.



\*Non présenté dans ce manuscrit

Figure 31 : Diagramme récapitulatif des différents modes utilisés au cours de ces développements.

## 6 BIBLIOGRAPHIE

- 1 B. Schindler, A. D. Depland, G. Renois-Predelus, G. Karras, B. Concina, G. Celep, J. Maurelli, V. Lorient, E. Constant, R. Bredy, C. Bordas, F. Lépine and I. Compagnon, FAIMS-MS-IR spectroscopy workflow: a multidimensional platform for the analysis of molecular isoforms, *Int. J. Ion Mobil. Spectrom.*, 2017, **20**, 119–124.
- 2 L. Jaravel, B. Schindler, J. Randon, I. Compagnon, C. Demesmay and V. Dugas, Off-line coupling of capillary isotachopheresis separation to IRMPD spectroscopy for glycosaminoglycans analysis: Application to the chondroitin sulfate disaccharides model solutes, *J. Chromatogr. A*, 2020, **1617**, 460782.
- 3 J. Wattjes, B. Schindler, S. Trombotto, L. David, B. M. Moerschbacher and I. Compagnon, Discrimination of patterns of N-acetylation in chitoooligosaccharides by gas phase IR spectroscopy integrated to mass spectrometry, *Pure Appl. Chem.*, 2017, **89**, 1349–1357.
- 4 B. Schindler, L. Barnes, G. Renois, C. Gray, S. Chambert, S. Fort, S. Flitsch, C. Loison, A.-R. Allouche and I. Compagnon, Anomeric memory of the glycosidic bond upon fragmentation and its consequences for carbohydrate sequencing, *Nat. Commun.*, 2017, **8**, 973.
- 5 B. Schindler, G. Laloy-Borgna, L. Barnes, A.-R. Allouche, E. Bouju, V. Dugas, C. Demesmay and I. Compagnon, Online Separation and Identification of Isomers Using Infrared Multiple Photon Dissociation Ion Spectroscopy Coupled to Liquid Chromatography: Application to the Analysis of Disaccharides Regio-Isomers and Monosaccharide Anomers, *Anal. Chem.*, 2018, **5**.
- 6 G. Gu and C. K. Lim, Separation of anionic and cationic compounds of biomedical interest by high-performance liquid chromatography on porous graphitic carbon, *J. Chromatogr. A*, 1990, **515**, 183–192.
- 7 B. Buszewski and S. Noga, Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique, *Anal. Bioanal. Chem.*, 2012, **402**, 231–247.
- 8 G. Qing, J. Yan, X. He, X. Li and X. Liang, Recent advances in hydrophilic interaction liquid chromatography materials for glycopeptide enrichment and glycan separation, *TrAC Trends Anal. Chem.*, 2020, **124**, 115570.
- 9 K. Madunić, T. Zhang, O. A. Mayboroda, S. Holst, K. Stavenhagen, C. Jin, N. G. Karlsson, G. S. M. Lageveen-Kammeijer and M. Wührer, Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype, *Cell. Mol. Life Sci.*, 2021, **78**, 337–350.
- 10 R. E. van Outersterp, K. J. Houthuijs, G. Berden, U. F. Engelke, L. A. J. Kluijtmans, R. A. Wevers, K. L. M. Coene, J. Oomens and J. Martens, Reference-standard free metabolite identification using infrared ion spectroscopy, *Int. J. Mass Spectrom.*, 2019, **443**, 77–85.
- 11 L. Barnes, A.-R. Allouche, S. Chambert, B. Schindler and I. Compagnon, Ion spectroscopy of heterogeneous mixtures: IRMPD and DFT analysis of anomers and conformers of monosaccharides, *Int. J. Mass Spectrom.*, 2020, **447**, 116235.
- 12 R. E. van Outersterp, J. Oosterhout, C. R. Gebhardt, G. Berden, U. F. H. Engelke, R. A. Wevers, F. Cuyckens, J. Oomens and J. Martens, Targeted Small-Molecule Identification Using Heartcutting Liquid Chromatography–Infrared Ion Spectroscopy, *Anal. Chem.*, 2023, **95**, 3406–3413.
- 13 N. Volpi, High-performance liquid chromatography and on-line mass spectrometry detection for the analysis of chondroitin sulfates/hyaluronan disaccharides derivatized with 2-aminoacridone, *Anal. Biochem.*, 2010, **397**, 12–23.
- 14 N. Volpi and R. J. Linhardt, High-performance liquid chromatography-mass spectrometry for mapping and sequencing glycosaminoglycan-derived oligosaccharides, *Nat. Protoc.*, 2010, **5**, 993–1004.
- 15 J. Wang, A. Bhalla, J. C. Ullman, M. Fang, R. Ravi, A. Arguello, E. Thomsen, B. Tsogtbaatar, J. L. Guo, L. L. Skuja, J. C. Dugas, S. S. Davis, S. B. Poda, K. Gunasekaran, S. Costanzo, Z. K. Sweeney, A. G. Henry, J. M. Harris, K. R. Henne and G. Astarita, High-Throughput Liquid Chromatography–Tandem Mass Spectrometry Quantification of Glycosaminoglycans as Biomarkers of Mucopolysaccharidosis II, *Int. J. Mol. Sci.*, 2020, **21**, 5449.

- 16 S. Poyer, I. Seffouh, C. Lopin-Bon, J.-C. Jacquinet, J. L. Neira, J.-Y. Salpin and R. Daniel, Discrimination of sulfated isomers of chondroitin sulfate disaccharides by HILIC-MS, *Anal. Bioanal. Chem.*, 2021, **413**, 7107–7117.
- 17 N. Khanal, C. Masellis, M. Z. Kamrath, D. E. Clemmer and T. R. Rizzo, Glycosaminoglycan Analysis by Cryogenic Messenger-Tagging IR Spectroscopy Combined with IMS-MS, *Anal. Chem.*, 2017, **89**, 7601–7606.
- 18 I. Compagnon, Lasers and ion mobility: new additions to the glycosaminoglycanomics toolkit, *Curr. Opin. Struct. Biol.*, 2018, **10**.
- 19 M. Lettow, M. Grabarics, E. Mucha, D. A. Thomas, Ł. Polewski, J. Freyse, J. Rademann, G. Meijer, G. von Helden and K. Pagel, IR action spectroscopy of glycosaminoglycan oligosaccharides, *Anal. Bioanal. Chem.*, 2020, **412**, 533–537.
- 20 M. Lettow, K. Greis, M. Grabarics, J. Horlebein, R. L. Miller, G. Meijer, G. von Helden and K. Pagel, Chondroitin Sulfate Disaccharides in the Gas Phase: Differentiation and Conformational Constraints, *J. Phys. Chem. A*, 2021, **125**, 4373–4379.
- 21 J. Riedel, M. Lettow, M. Grabarics, M. Götze, R. L. Miller, G.-J. Boons, G. Meijer, G. Von Helden, G. P. Szekeres and K. Pagel, Predicting Structural Motifs of Glycosaminoglycans using Cryogenic Infrared Spectroscopy and Random Forest, *J. Am. Chem. Soc.*, 2023, **145**, 7859–7868.
- 22 G. M. Miller and L. C. Hsieh-Wilson, Sugar-dependent modulation of neuronal development, regeneration, and plasticity by chondroitin sulfate proteoglycans, *Exp. Neurol.*, 2015, **274**, 115–125.
- 23 S. Miyata and H. Kitagawa, Formation and remodeling of the brain extracellular matrix in neural plasticity: Roles of chondroitin sulfate and hyaluronan, *Biochim. Biophys. Acta BBA - Gen. Subj.*, 2017, **1861**, 2420–2434.
- 24 W. H. Howell and E. Holt, TWO NEW FACTORS IN BLOOD COAGULATION—HEPARIN AND PRO-ANTITHROMBIN, *Am. J. Physiol.-Leg. Content*, 1918, **47**, 328–341.
- 25 G. B. Lim, Discovery and purification of heparin, *Nat. Rev. Cardiol.*, , DOI:10.1038/nrcardio.2017.171.
- 26 L.-O. Andersson, T. W. Barrowcliffe, E. Holmer, E. A. Johnson and G. E. C. Sims, Anticoagulant properties of heparin fractionated by affinity chromatography on matrix-bound antithrombin III and by gel filtration, *Thromb. Res.*, 1976, **9**, 575–583.
- 27 C. A. A. van Boeckel and M. Petitou, The Unique Antithrombin III Binding Domain of Heparin: A Lead to New Synthetic Antithrombotics, *Angew. Chem. Int. Ed. Engl.*, 1993, **32**, 1671–1690.
- 28 K. A. Bauer, D. W. Hawkins, P. C. Peters, M. Petitou, J.-M. Herbert, C. A. A. van Boeckel and D. G. Meuleman, Fondaparinux, a Synthetic Pentasaccharide: The First in a New Class of Antithrombotic Agents — The Selective Factor Xa Inhibitors, *Cardiovasc. Drug Rev.*, 2002, **20**, 37–52.
- 29 M. Petitou and C. A. A. van Boeckel, A Synthetic Antithrombin III Binding Pentasaccharide Is Now a Drug! What Comes Next?, *Angew. Chem. Int. Ed.*, 2004, **43**, 3118–3133.
- 30 L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike and R. W. Carrell, The anticoagulant activation of antithrombin by heparin, *Proc. Natl. Acad. Sci.*, 1997, **94**, 14683–14688.
- 31 C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, Rapid screening of classic galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma, *The Analyst*, 2015, **140**, 2280–2286.
- 32 B. Moge, O. Yeni, A. Infantino and I. Compagnon, CO<sub>2</sub> laser enhanced rapid IRMPD spectroscopy for glycan analysis, *Int. J. Mass Spectrom.*, 2023, **490**, 117071.
- 33 A. Ben Faleh, S. Warnke, T. Van Wieringen, A. H. Abikhodr and T. R. Rizzo, New Approach for the Identification of Isobaric and Isomeric Metabolites, *Anal. Chem.*, 2023, **95**, 7118–7126.
- 34 J. Martens, R. E. Van Outersterp, R. J. Vreeken, F. Cuyckens, K. L. M. Coene, U. F. Engelke, L. A. J. Kluijtmans, R. A. Wevers, L. M. C. Buydens, B. Redlich, G. Berden and J. Oomens, Infrared ion spectroscopy: New opportunities for small-molecule identification in mass spectrometry - A tutorial perspective, *Anal. Chim. Acta*, 2020, **1093**, 1–15.

- 35 L. Cheng, M. B. G. Kiewiet, A. Groeneveld, A. Nauta and P. De Vos, Human milk oligosaccharides and its acid hydrolysate LNT2 show immunomodulatory effects via TLRs in a dose and structure-dependent way, *J. Funct. Foods*, 2019, **59**, 174–184.
- 36 L. Cheng, R. Akkerman, C. Kong, M. T. C. Walvoort and P. de Vos, More than sugar in the milk: human milk oligosaccharides as essential bioactive molecules in breast milk and current insight in beneficial effects, *Crit. Rev. Food Sci. Nutr.*, 2021, **61**, 1184–1200.
- 37 Y. Zhu, L. Wan, W. Li, D. Ni, W. Zhang, X. Yan and W. Mu, Recent advances on 2'-fucosyllactose: physiological properties, applications, and production approaches, *Crit. Rev. Food Sci. Nutr.*, 2022, **62**, 2083–2092.
- 38 L. Pereira, Porous Graphitic Carbon as a Stationary Phase in HPLC: Theory and Applications, *J. Liq. Chromatogr. Relat. Technol.*, 2008, **31**, 1687–1731.
- 39 H. Hinneburg, S. Chatterjee, F. Schirmeister, T. Nguyen-Khuong, N. H. Packer, E. Rapp and M. Thaysen-Andersen, Post-Column Make-Up Flow (PCMF) Enhances the Performance of Capillary-Flow PGC-LC-MS/MS-Based Glycomics, *Anal. Chem.*, 2019, **91**, 4559–4567.
- 40 C. West, C. Elfakir and M. Lafosse, Porous graphitic carbon: A versatile stationary phase for liquid chromatography, *J. Chromatogr. A*, 2010, **1217**, 3201–3216.
- 41 L. Veillon, Y. Huang, W. Peng, X. Dong, B. G. Cho and Y. Mechref, Characterization of isomeric glycan structures by LC-MS/MS, *ELECTROPHORESIS*, 2017, **38**, 2100–2114.
- 42 P. H. Jensen, N. G. Karlsson, D. Kolarich and N. H. Packer, Structural analysis of N- and O-glycans released from glycoproteins, *Nat. Protoc.*, 2012, **7**, 1299–1310.
- 43 O. Yeni, Analyse par IRMPD de polysaccharides de lichen contenant des unités galactofuranose, Thèse soutenue à l'Université Claude Bernard Lyon 1, 2023.
- 44 P. Peltier, R. Euzen, R. Daniellou, C. Nugier-Chauvin and V. Ferrières, Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications, *Carbohydr. Res.*, 2008, **343**, 1897–1923.
- 45 L. L. Pedersen and S. J. Turco, Galactofuranose metabolism: a potential target for antimicrobial chemotherapy, *Cell. Mol. Life Sci. CMLS*, 2003, **60**, 259–266.
- 46 B. Tefsen, A. F. Ram, I. Van Die and F. H. Routier, Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact, *Glycobiology*, 2012, **22**, 456–469.
- 47 D. Laverde, F. Romero-Saavedra, D. A. Argunov, J. Enotarpi, V. B. Krylov, E. Kalfopoulou, C. Martini, R. Torelli, G. A. van der Marel, M. Sanguinetti, J. D. C. Codée, N. E. Nifantiev and J. Huebner, Synthetic Oligomers Mimicking Capsular Polysaccharide Diheteroglycan are Potential Vaccine Candidates against Encapsulated Enterococcal Infections, *ACS Infect. Dis.*, 2020, **6**, 1816–1826.
- 48 J. S. Ho, A. Gharbi, B. Schindler, O. Yeni, R. Brédy, L. Legentil, V. Ferrières, L. L. Kiessling and I. Compagnon, Distinguishing Galactoside Isomers with Mass Spectrometry and Gas-Phase Infrared Spectroscopy, *J. Am. Chem. Soc.*, 2021, **143**, 10509–10513.
- 49 B. Favreau, O. Yeni, S. Ollivier, J. Boustie, F. L. Dévéhat, J.-P. Guégan, M. Fanuel, H. Rogniaux, R. Brédy, I. Compagnon, D. Ropartz, L. Legentil and V. Ferrières, Synthesis of an Exhaustive Library of Naturally Occurring Gal f -Man p and Gal p -Man p Disaccharides. Toward Fingerprinting According to Ring Size by Advanced Mass Spectrometry-Based IM-MS and IRMPD, *J. Org. Chem.*, 2021, **86**, 6390–6405.
- 50 O. Yeni, S. Ollivier, B. Moge, D. Ropartz, H. Rogniaux, L. Legentil, V. Ferrières and I. Compagnon, Ring-Size Memory of Galactose-Containing MS/MS Fragments: Application to the Detection of Galactofuranose in Oligosaccharides and Their Sequencing, *J. Am. Chem. Soc.*, 2023, **145**, 15180–15187.
- 51 O. Yeni, B. Moge, A. Infantino, C. Guitteny, J. Boustie, F. Le Dévéhat, L. Legentil, V. Ferrières and I. Compagnon, LC-MS/MS-IR screening method in glycomics: application to the discrimination of galactofuranose-containing disaccharides, *En Préparation*.

# Chapitre IV : Analyse de composés glycoconjugués, l'exemple d'une glycoprotéine : la fétuine bovine

## TABLE DES MATIERES

---

<b>1</b>	<b>Introduction .....</b>	<b>108</b>
<b>2</b>	<b>Les protéines .....</b>	<b>108</b>
2.1	Les modifications post traductionnelles.....	108
2.1.1	Cas général.....	108
2.1.2	Cas particulier de la glycosylation .....	109
2.2	La fétuine bovine .....	109
2.3	Etudes des protéines bottom-up.....	110
<b>3</b>	<b>Recherche de glycosylations sur la protéine.....</b>	<b>111</b>
3.1	Résultat sur une protéine non glycosylée .....	112
3.2	Résultats sur une protéine glycosylée .....	112
<b>4</b>	<b>Etude des peptides : identification de glycopeptides par screening IRMPD .....</b>	<b>113</b>
4.1	Proposition .....	113
4.1.1	Structure des acides aminés contenant un groupement hydroxyle .....	114
4.1.2	Protocoles similaires déjà proposés.....	114
4.1.3	Spectre IRMPD du glycopeptide de poulet .....	115
4.2	Protocole .....	116
4.2.1	Digestion de la protéine par des enzymes protéolytiques.....	116
4.2.2	Passage en LC-MS .....	120
4.2.3	Irradiation par le laser.....	120
4.3	Résultats .....	120
4.3.1	Chromatogramme des produits de digestion de la fétuine bovine .....	120
4.3.2	Spectre IRMPD d'un peptide, pouvant porter un acide aminé avec un hydroxyle .....	122
4.3.3	Spectre IRMPD de candidats glycopeptides.....	123
4.3.4	Screening de glycopeptides .....	127
<b>5</b>	<b>Conclusion .....</b>	<b>128</b>
<b>6</b>	<b>Bibliographie .....</b>	<b>130</b>

# 1 INTRODUCTION

---

Les précédentes parties ont traité de l'analyse de sucres au moyen de la spectroscopie IRMPD, seule ou couplée avec la chromatographie liquide. Ces développements permettront de caractériser les impuretés. Le projet IROLIGO a également pour but de caractériser le composé synthétisé, il s'agit d'un glycoconjugué. Ces composés sont des hétéropolymères dans lesquelles une partie glucidique est liée de manière covalente à une autre classe de molécules, qui peut être une protéine ou un lipide par exemple. Ces composés forment une classe bien plus hétérogène que les homopolymères étudiés précédemment. Ici, nous prenons comme modèle de composés glycoconjugués une glycoprotéine. Ainsi la caractérisation doit être menée sur la protéine, mais également sur la nature et la position des sucres greffés sur cette protéine, d'où la difficulté de la caractérisation, il s'agit donc d'un modèle complexe dont l'étude s'avère intéressante.

Généralement, au moins trois analyses différentes sont nécessaires pour permettre la caractérisation structurale des glycoconjugués. L'analyse de la glycoprotéine « native » donne une vue d'ensemble de la protéine, est-elle glycosylée<sup>1,2</sup> ? quelle est la nature de ces glycosylations ?<sup>3</sup> La deuxième étape est de procéder à la digestion de la glycoprotéine par des protéases, afin d'identifier quels sont les sites de glycosylation occupés par ces sucres ? Ces deux étapes permettent d'avoir une bonne idée de la macrohétérogénéité présente sur la protéine avec des informations sur l'occupation d'un site de glycosylation<sup>4-6</sup>. L'utilisation de glycosidases sur la protéine entière donne des indications sur les différentes glycoformes de la protéine<sup>7</sup>. Si maintenant la déglycosylation a lieu sur les peptides, après extraction purification et analyse des sucres, la connaissance de la structure de ceux-ci est affinée<sup>5,6</sup>. Cette dernière étape de caractérisation permet de connaître la microhétérogénéité : quels sont les sucres présents sur tel site de glycosylation. Il est également possible de connaître la position de ces glycosylations, si la déglycosylation porte sur les peptides déglycosylés enzymatiquement.

Dans cette partie, nous nous intéressons aux deux premiers niveaux de caractérisation. Le but ici est de démontrer qu'il est possible de détecter la présence de sucres sur une protéine entière par spectroscopie IR en phase condensée. Dans une seconde partie, nous nous intéresserons à l'identification de glycopeptides présents dans un mélange de peptides produits par la digestion enzymatique d'une protéine. Pour cette identification, nous nous sommes servis du récent développement présenté dans la partie précédente traitant du screening à longueur d'onde fixe. Après séparation chromatographique, les ions sont irradiés par le laser IR dans la gamme des 3  $\mu\text{m}$ . Leur photofragmentation apporte une information sur la glycosylation de ces peptides

## 2 LES PROTEINES

---

### 2.1 LES MODIFICATIONS POST TRADUCTIONNELLES

#### 2.1.1 Cas général

Les PTMs (Post Translational Modification) sont des modifications qui ont lieu après traduction de la protéine au sein du ribosome. Ces modifications ne sont pas prévisibles en étudiant uniquement le génome de l'individu, mais elles répondent à des besoins bien précis, notamment en réponse à l'environnement de l'individu. Par exemple, la phosphorylation de certaines protéines, permet le transfert de l'information par une cascade d'activations de protéines. L'ubiquitination se fait par fixation d'une ou plusieurs ubiquitines (une petite protéine) sur une protéine afin d'adresser cette dernière au protéasome pour dégradation<sup>8</sup>. Ces PTMs changent la nature chimique d'un acide aminé,

voir même la structure de la molécule (formation de ponts disulfures). Il existe donc une grande variété de PTMs par exemple la sulfatation, l'acétylation, la succinylation...

### 2.1.2 Cas particulier de la glycosylation

La glycosylation est l'une des trois modifications post traductionnelles les plus fréquentes<sup>9</sup>. Certaines glycoprotéines nécessitent d'être glycosylées afin de se replier correctement<sup>10</sup>. Un défaut de glycosylation est observé dans certains cancers<sup>11-13</sup>. Le principal problème est que ces modifications ne sont pas prévisibles en étudiant uniquement le code génétique de l'individu, la glycosylation change en fonction de l'environnement de l'individu, de son état de santé de son âge ou de son sexe<sup>14,15</sup>.

Différents types de glycosylation ont lieu sur les protéines : la N-, la O- ou plus rare la C-glycosylation. Les deux types N et O-glycosylation présentent une localisation et une structure différente. La N-glycosylation a lieu sur une asparagine si celle-ci est située deux résidus en amont d'une sérine ou d'une thréonine (telle que la séquence est : asparagine-X-sérine/thréonine, avec X n'importe quel acide aminé excepté la proline). La séquence bien spécifique d'acides aminés sur lesquels peut venir se greffer un N-glycane permet en regardant la séquence protéique de se faire une idée d'où pourra avoir lieu la N-glycosylation. En revanche, dans le cas de la O-glycosylation, un N-Acétalgalactosamine est greffé sur une sérine ou une thréonine, le O-glycane peut alors être allongé par ajout de monosaccharides. Les O-glycanes présentent une plus grande variété de structures que les N-glycanes, ce qui rend leur prédiction et détection plus délicate.

La plupart des modifications post-traductionnelles consiste en l'ajout d'un groupement chimique (phosphate, sulfate...). La glycosylation au contraire ne se limite pas forcément à l'ajout d'un glycane unique. En effet, sur une même position, plusieurs glycanes différents peuvent être présents. Cela complexifie la tâche lorsqu'on doit mener à bien la caractérisation structurale de ces modifications.

Il faut souligner que la glycosylation est un processus biochimique nécessitant des enzymes. Au contraire de la glycation qui est un simple processus chimique dans lequel un sucre se greffe sur une protéine, ce greffage n'a rien de physiologique (la glycation est observée sur l'hémoglobine des individus diabétiques par exemple<sup>16</sup>).

## 2.2 LA FETUINE BOVINE

Chez les mammifères, les fétuines sont des protéines présentes dans le sérum des fœtus. La fétuine bovine est excrétée par le foie. Elle possède différentes fonctions. La fétuine bovine joue un rôle dans le métabolisme du calcium et donc dans la formation des os<sup>17</sup>. La fétuine bovine inhibe également la phosphorylation de récepteurs à insuline<sup>18</sup>.

Selon UniprotKB<sup>19</sup>, la fétuine bovine (entrée P12763) comporte 359 acides aminés, dont la séquence est donnée ci-dessous (Tableau 1). La masse moléculaire de la protéine est alors de 38.4 kDa.

Tableau 1 : séquence de la fétuine bovine

10	20	30	40	50	60
MKSFVLLFCL	AQLWGCHSIP	LDPVAGYKEP	ACDDPDTEQA	ALAAVDYINK	HLPRGYKHTL
70	80	90	100	110	120
NQIDSVKWP	RRPTGEVYDI	EIDTLETTCH	VLDPPLANC	SVRQQTQHAV	EGDCDI HVLK
130	140	150	160	170	180
QDGQFSVLFT	KCDSSPDSAE	DVRKLCPCP	LLAPLNSRV	VHAVEVALAT	FNAESNGSYL
190	200	210	220	230	240
QLVEISRAQF	VPLPVSVSVE	FAVAATDCIA	KEVDPTKCN	LLAEKQYGFC	KGSVIQKALG
250	260	270	280	290	300
GEDVRTCTL	FQTQVIPQP	QPDGAEAEAP	SAVPDAAGPT	PSAAGPPVAS	VVVGPSVVAV
310	320	330	340	350	359
PLPLHRAHYD	LRHTFSGVAS	VESSSGEAFH	VGKTPIVGQP	SIPGGPVRLC	PGRIRYFKI

La fétuine bovine présente des PTMs, après traduction, tout d'abord, elle se voit raccourcir de ses 18 premiers acides aminés<sup>20</sup> (la masse moléculaire passe alors à 36.3 kDa), les cystéines forment des ponts disulfures. Il s'agit de deux cystéines (dont la chaîne latérale contient un thiol) formant une liaison covalente entre les atomes de soufre présents sur leur chaîne latérale (6 ponts disulfures sont observés dans la structure de la fétuine bovine). Certaines sérines sont phosphorylées<sup>21,22</sup>. Et enfin certaines asparagines peuvent être glycosylées (N-glycane)<sup>21,23,24</sup> de même que certaines thréonines ou sérines<sup>21,22</sup> (O-glycane). Ces différentes PTMs apportent beaucoup de complexité à cette protéine. Ce qui est observé sur les spectres de masse de la fétuine bovine en électrospray<sup>1,25</sup> ou en MALDI<sup>26,27</sup> (différentes glycoformes de la fétuine bovine coexistent).

### 2.3 ETUDES DES PROTEINES BOTTOM-UP

Un des moyens d'étudier les protéines est de les digérer par des protéases : des enzymes capables de cliver des protéines en chaînes peptidiques. Le mélange de peptides résultant est alors analysé au moyen de la chromatographie liquide couplée à la spectrométrie de masse, afin de connaître la séquence de la protéine et de localiser les modifications post traductionnelles. Lors de mesures en spectrométrie de masse et lors de la fragmentation des peptides en spectrométrie de masse, des fragments bien spécifiques sont produits (Figure 1). Dans notre cas, la fragmentation obtenue par CID conduit principalement à la formation des séries d'ions y et b. Ces séries d'ions clivent la liaison peptidique côté N-ter de l'azote engagé dans la liaison, la série y correspond aux fragments situés en C-ter du clivage tandis que la série b correspond aux fragments situés côté N-ter du clivage. Les modes de fragmentation mettant en jeu des électrons conduisent plutôt aux séries d'ions z et c.

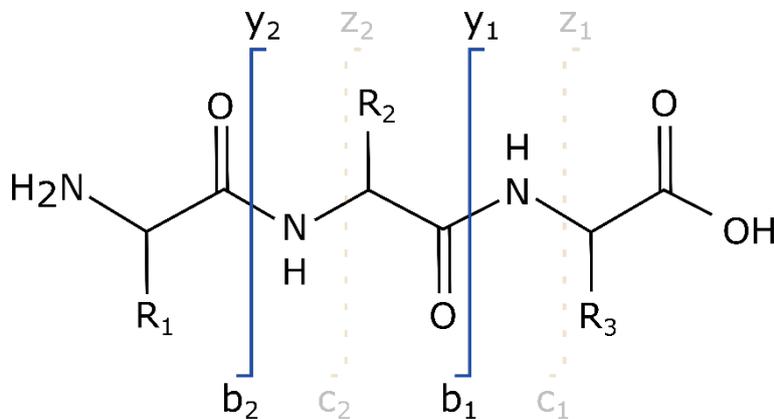


Figure 1 : Fragments obtenus lors de la fragmentation de peptides en spectrométrie de masse

### 3 RECHERCHE DE GLYCOSYLATIONS SUR LA PROTEINE

La spectroscopie IR en phase condensée de protéines a déjà été utilisée afin de connaître la structure secondaire de protéines<sup>28-30</sup>, principalement en étudiant les bandes amides I et II présentes à respectivement 1650 et 1550  $\text{cm}^{-1}$ . La spectroscopie IR a également mis en évidence la présence de sucres sur les protéines<sup>31-34</sup> avec la présence de pics caractéristiques des sucres entre 1000 et 1100  $\text{cm}^{-1}$ . Ici, nous voulons vérifier que nous arrivons à détecter la glycosylation sur une protéine au moyen de la spectroscopie IR en phase condensée. Nous nous attendons à ce que cette différence soit visible dans la région caractéristique des sucres, c'est dans cette gamme que de nombreuses bandes d'absorption caractéristiques des sucres notamment celles concernant les modes de vibration de la liaison C-O sont présentes. La spectroscopie IR a l'avantage de ne pas dégrader l'échantillon, après une mesure, il est possible de procéder à de nouvelles analyses sur le même échantillon.

Pour caractériser la structure de protéine, elle est généralement solubilisée dans un tampon afin de l'observer sous sa forme native dans laquelle sa structure est conservée. La protéine est dissoute dans le tampon à haute concentration ainsi elle absorbe suffisamment même avec seulement une faible épaisseur d'échantillon, en limitant donc la contribution venant du solvant. Pour effectuer ces mesures, une cellule est formée par deux fenêtres d'un matériau transparent aux IR, une fine cale de polyéthylène ( $< 5 \mu\text{m}$ ) est intercalée entre ces deux fenêtres. La cellule est remplie de l'échantillon dissout dans le tampon d'intérêt. Dans notre cas, nous n'avons pas travaillé en phase liquide. Le spectre est mesuré après séchage d'une goutte de solution de protéine. En fait, étant donné que notre objectif n'est pas l'étude fine de la structure des protéines, mais seulement de déterminer la présence de glycosylations sur la protéine. Il n'est pas nécessaire de disposer de la protéine parfaitement repliée dans un tampon. Aussi, nous nous sommes affranchis de cette difficulté et avons travaillé avec des échantillons séchés de protéines (ainsi on élimine l'absorption due à l'eau). Pour obtenir un dépôt homogène de l'échantillon, la protéine est dissoute dans l'eau à 1  $\text{mg}\cdot\text{mL}^{-1}$ . 80  $\mu\text{L}$  de cette solution sont déposés sur une fenêtre de  $\text{CaF}_2$  (fluorure de calcium). La goutte est ensuite séchée au moyen du Speedvac pendant 15 minutes avant mesure dans le FTIR (Fourier Transform InfraRed Spectrometer).

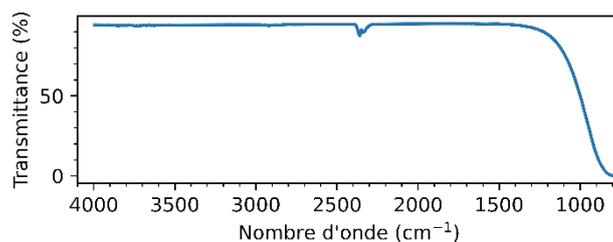


Figure 2 : Spectre de transmittance du  $\text{CaF}_2$  dans la gamme d'intérêt. Cette transmittance est optimale entre 4000 et 1100  $\text{cm}^{-1}$ , à l'exception de la bande à 2350  $\text{cm}^{-1}$ .

L'absorption du  $\text{CaF}_2$  entre 750 et 4000  $\text{cm}^{-1}$  est présenté en Figure 2. La transmittance est  $> 90\%$  sur la quasi-totalité de la gamme (1100-4000  $\text{cm}^{-1}$ ), à l'exception de la gamme autour de 2350  $\text{cm}^{-1}$  due à l'absorption du  $\text{CO}_2$  (la transmittance descend alors à 87 %). En revanche, en dessous de 1100  $\text{cm}^{-1}$ , la lame de  $\text{CaF}_2$  absorbe le rayonnement IR.

Dans cette partie, nous comparons une protéine et une glycoprotéine. La protéine choisie est la myoglobine de cheval (entrée P68082 sur UniprotKB<sup>35</sup>), cette protéine longue de 154 acides aminés (17 kDa) n'est pas glycosylée. La glycoprotéine utilisée est la fétuine bovine (décrite en 2.2). Le spectre IR de ces deux protéines est acquis comme décrit précédemment.

### 3.1 RESULTAT SUR UNE PROTEINE NON GLYCOSYLEE

Le spectre IR de la myoglobine est présenté en Figure 3. Ce spectre présente tout d'abord un important massif d'absorption dans la gamme des 3  $\mu\text{m}$  avec les bandes caractéristiques des modes de vibration des espèces X-H (avec X est C, N ou O). On retrouve dans cette gamme la bande provenant du mode de vibration stretching de la liaison N-H de l'amide A ( $3296\text{ cm}^{-1}$ )<sup>36</sup>. Il peut également s'agir des amines présentes sur les chaînes latérales des acides aminés basiques (lysines et arginines). L'épaule à  $3068\text{ cm}^{-1}$  peut correspondre au mode de vibration stretching =C-H. De tels modes de vibration sont retrouvés dans les acides aminés possédant un cycle aromatique sur leur chaîne latérale (tryptophane, phénylalanine et tyrosine). Dans cette gamme, nous retrouvons également des groupements C-H (stretching à  $2961\text{ cm}^{-1}$  : mode antisymétrique,  $2933\text{ cm}^{-1}$  : mode antisymétrique et  $2974\text{ cm}^{-1}$  : mode symétrique) des acides aminés portant des chaînes latérales aliphatiques. Les protéines possèdent également des acides aminés portant un hydroxyle sur leur chaîne latérale, c'est certainement au mode de vibration stretching de la liaison O-H de ces acides aminés que l'épaule à  $3450\text{ cm}^{-1}$  peut être attribué. Il ne faut pas non plus négliger la présence d'eau dans l'atmosphère. Dans la gamme 5 à 10  $\mu\text{m}$ , la myoglobine présente les bandes caractéristiques de la liaison de la bande amide I  $1653\text{ cm}^{-1}$ , de la bande amide II (composée des vibrations déphasées des modes de bending de la liaison N-H et stretching de la liaison C-N) à  $1546\text{ cm}^{-1}$ . De nombreux autres groupes sont présents, toutefois, on ne peut pas tous les attribuer avec certitude : les bandes amide III (composée des vibration en phase de bending de la liaison N-H et stretching de la liaison C-N) sont attendues entre  $1200$  et  $1400\text{ cm}^{-1}$ )<sup>28,37</sup>. Cette gamme contient également les bandes d'absorption du mode de vibration bending des liaisons C-H attendues autour de  $1400\text{ cm}^{-1}$ , stretching des liaisons C=C (dans un cycle aromatique) prévues à  $1500\text{ cm}^{-1}$ , bending des liaisons O-H d'un alcool primaire (chaîne latérale de la sérine) ou d'un alcool secondaire (chaîne latérale de la thréonine) supposées être entre  $1260$  et  $1350\text{ cm}^{-1}$ , le mode stretching de la liaison C=O d'un acide carboxylique ionisé aux alentours de  $1400\text{ cm}^{-1}$  ou encore le mode bending de la liaison O-H d'un alcool aromatique (chaîne latérale de la tyrosine).

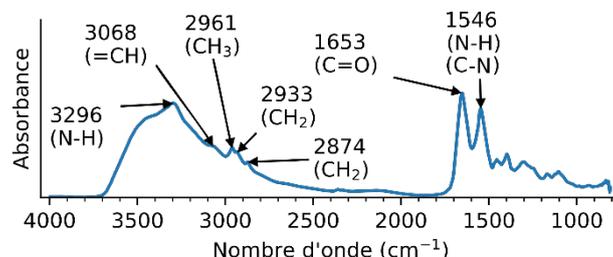


Figure 3 : Spectre IR du dépôt de la myoglobine, de nombreuses bandes d'absorption caractéristiques des protéines sont visibles dans la gamme des 3  $\mu\text{m}$  et dans celle comprise entre 5 et 10  $\mu\text{m}$ .

### 3.2 RESULTATS SUR UNE PROTEINE GLYCOSYLEE

Le spectre IR de La fétuine bovine (Figure 4) est similaire à ce que nous avons observé sur la myoglobine (Figure 3). Les bandes d'absorption des modes de vibration des liaisons X-H sont retrouvés aux mêmes nombres d'onde. La fétuine bovine présente également les bandes caractéristiques de la liaison amide, avec des bandes d'absorption légèrement décalées. La première bande est observée à  $1639\text{ cm}^{-1}$  (bande amide I) et la seconde à  $1544\text{ cm}^{-1}$  (bande amide II). Ensuite, comme dans le cas de la myoglobine, des pics non attribués sont présents sur le spectre. La différence d'avec le spectre IR de la myoglobine réside dans la présence de deux pics non résolus à  $1067$  et  $1034\text{ cm}^{-1}$ . Ces pics sont cohérents avec les bandes d'absorption des modes de vibration de C-O au sein d'un sucre<sup>32,33,36</sup>.

Les sucres présents sur cette glycoprotéine présentent un très grand nombre de groupements hydroxyles ce qui explique le fait que nous ne voyions pas de tels pics sur la myoglobine. En effet, la myoglobine ne contient qu'un nombre restreint de groupements hydroxyle sur sa séquence (les alcools portés par les acides aminés sérine, thréonine et tyrosine).

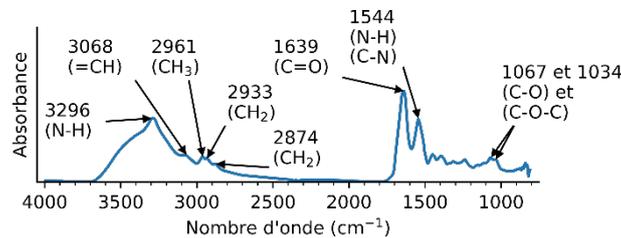


Figure 4 : Spectre IR en phase condensée de la fétuine bovine, dans ce cas aussi, des bandes d'absorption dans les gammes des 3  $\mu\text{m}$  et dans celle comprise entre 5 et 10  $\mu\text{m}$  sont visibles, ici un signal intéressant est observé entre 1000 et 1100  $\text{cm}^{-1}$  (bandes caractéristiques des sucres).

La myoglobine et la fétuine bovine présentent donc toutes deux un spectre IR avec beaucoup de ressemblances deux zones sont bien distinctes d'un côté la gamme des 3  $\mu\text{m}$  avec un recouvrement quasi parfait des deux spectres. L'autre côté du spectre : entre 1700 et 800  $\text{cm}^{-1}$ , présente des signatures similaires avec de légers décalages en nombre d'onde, mais la différence repose sur la présence de deux légères bandes d'absorption : les bandes 1067 et 1034  $\text{cm}^{-1}$ . Ces bandes sont caractéristiques des sucres et permettent de discriminer la myoglobine de la fétuine bovine et donc d'identifier une glycoprotéine.

## 4 ETUDE DES PEPTIDES : IDENTIFICATION DE GLYCOPEPTIDES PAR SCREENING IRMPD

Comme nous l'avons vu en introduction, l'analyse des peptides/glycopeptides permet de connaître la position des sucres sur la chaîne protéique. Pour cela une étape préalable de digestion enzymatique de la glycoprotéine est requise. Le produit de cette digestion est donc un mélange de peptides, de glycopeptides et de l'enzyme restant dans le mélange. Il est alors nécessaire de simplifier au maximum l'échantillon (éliminer les impuretés et procéder à une méthode séparative en amont de la spectrométrie de masse). Il reste néanmoins encore des difficultés pour connaître la position de la glycosylation. Tout d'abord, il est difficile de distinguer un peptide d'un glycopeptide. De plus, les glycopeptides peuvent porter différents sucres, la masse change selon qu'un même peptide porte un glycane ou un autre. En spectrométrie de masse, le signal est alors dilué sur plusieurs  $m/z$  et l'intensité de chacun de ces glycopeptides est alors amoindrie. Les bibliothèques de glycopeptides en MS sont mal documentées. Enfin, les spectres obtenus par fragmentation CID des glycopeptides sont riches en fragments de la partie glucide du glycoconjugué, mais demeurent finalement pauvres en détails sur la structure du peptide.

### 4.1 PROPOSITION

Ici, nous proposons de nous servir d'une caractéristique structurale des peptides et des sucres afin de faciliter la détection des glycopeptides. Les peptides sont composés d'acides aminés, chacun d'eux porte une chaîne latérale qui lui est propre (polaire chargée, polaire non chargée ou encore non polaire hydrophobe). Ce qui est intéressant dans notre cas, est que seuls trois acides aminés contiennent un groupement hydroxyle sur leur chaîne latérale. Ces acides aminés sont la sérine, la thréonine et la

tyrosine (la tyrosine contient plus exactement un groupement phénol sur sa chaîne latérale). Les structures de ces trois acides aminés sont présentées en Figure 5.

D'un autre côté, les sucres contiennent un grand nombre de groupements hydroxyles. En spectroscopie IR, les groupements hydroxyles absorbent dans la gamme des 3  $\mu\text{m}$  : élongation du mode de vibration de la liaison O-H (voir Chapitre I).

Nous avons imaginé un protocole similaire à ce que nous avons proposé pour la détection du galactofuranose (voir Chapitre III). En sortie de chromatographie, les ions sont détectés par spectrométrie de masse et éventuellement irradiés au sein du piège à ions. On s'attend à observer de la photofragmentation lors de l'irradiation d'un glycopeptide à une longueur d'onde correspondant à l'absorption du mode de vibration de la liaison O-H, mais pas ou alors très faiblement s'il s'agit d'un peptide.

#### 4.1.1 Structure des acides aminés contenant un groupement hydroxyle

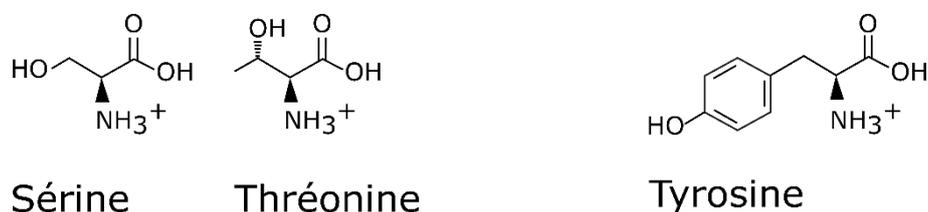


Figure 5 : Structure des trois acides aminés présentant un groupement hydroxyle sur leur chaîne latérale.

La sérine, la thréonine et la tyrosine sont les trois acides aminés contenant un hydroxyle sur leur chaîne latérale (Figure 5). La sérine contient un alcool primaire sur sa chaîne latérale ( $\text{R-CH}_2\text{-OH}$ ) et la thréonine un alcool secondaire ( $\text{R-C(-R')H-OH}$ ) tandis que la tyrosine contient un groupement phénol (l'hydroxyle est porté par un phényle). Les modes de vibration des liaisons O-H et C-O sont similaires entre un alcool et un phénol.

#### 4.1.2 Protocoles similaires déjà proposés

En fait, un tel montage (irradiation des ions par un laser à une longueur d'onde fixe au sein du spectromètre de masse pour la détection d'une caractéristique du composé) a déjà été proposé ailleurs. C'est le cas en métabolomique avec l'analyse de deux isomères de position que sont l'acide 3,4-dihydroxyphenylacétique et l'acide 2,5-dihydroxyphenylacétique et qui présentent chacun un maximum d'absorption dans l'infrarouge quand l'autre composé n'absorbe pas<sup>38</sup>. Dans le domaine de la protéomique, on peut citer la recherche des PTMs que sont les sulfatations et les phosphorylations. Par exemple, Polfer *et al.* ont discriminé des peptides phosphorylés et sulfatés sans avoir recours à la spectrométrie de masse à haute résolution (bien que ces composés soient isobariques) en se servant d'un laser IR accordable dans la gamme des 3  $\mu\text{m}$ <sup>39</sup>, chaque groupement absorbe dans une gamme bien particulière. Une conclusion similaire a été faite dans notre équipe à propos de sucres phosphorylés ou sulfatés<sup>40</sup>. Le screening par LC-IRMPD avec un laser  $\text{CO}_2$  d'un mélange de peptides phosphorylés et non phosphorylés a permis de détecter les peptides phosphorylés<sup>41</sup>. Récemment, nous avons montré que le screening à longueur d'onde fixe est capable de détecter le galactofuranose (une forme rare de galactose) dans des di- et trisaccharides après séparation chromatographique<sup>42,43</sup>. L'un des principaux intérêts de la spectroscopie IRMPD est qu'elle ne nécessite généralement pas le marquage par un chromophore.

Les gammes du visible et de l'UV ont aussi été utilisées pour détecter des peptides contenant une cystéine. Pour cela, les cystéines sont marquées au moyen d'un chromophore par un greffage

chimique. Le mélange peptides/peptides marqués est alors séparé en chromatographie liquide avant d'être analysé par spectrométrie de masse dans laquelle les ions sont irradiés<sup>44-46</sup>.

Enfin dans la gamme des UV, il a été possible d'identifier des peptides interagissant ensemble. En solution, un pontage est établi avant séparation chromatographique et injection dans le spectromètre de masse. La détection de deux peptides liés par un agent pontant se fait en phase gazeuse en irradiant l'édifice moléculaire avec un laser UV (l'agent pontant contient un groupement absorbant dans l'UV)<sup>47</sup>. L'identification de deux peptides liés l'un à l'autre a également été fait en incorporant un agent pontant contenant un phosphate puis en irradiant les ions au sein du piège à ions au moyen d'un laser IR (10,6  $\mu\text{m}$ ) pour effectuer un screening au moyen de l'IRMPD<sup>48</sup>.

La détection des ions d'intérêt se fait lors de l'exploitation des données, en comparant deux chromatogrammes, ou alors deux événements (un chromatogramme sans irradiation avec un chromatogramme avec irradiation, ou alors un événement sans irradiation avec un événement avec irradiation). Il est alors possible de voir quels ions ont été affectés par l'irradiation.

Dans notre cas, nul besoin de modifier les peptides. Nous nous attendons à observer la photofragmentation dans le cas d'un peptide glycosylé sous irradiation IR dans la gamme des 3  $\mu\text{m}$  et aucune photofragmentation dans le cas de peptides non glycosylés.

#### 4.1.3 Spectre IRMPD du glycopeptide de poulet

Ici nous prenons pour exemple le glycopeptide de poulet, un glycopeptide facilement isolé à partir d'œufs lyophilisés. Ce glycopeptide est donc accessible en grandes quantités. Sa structure est connue (Figure 6). Sa séquence peptidique est KVANKT, ce peptide est N-glycosylé au niveau de l'asparagine. Le glycane présent sur l'oligosaccharide contient 3 mannoses, 2 galactoses, 4 N-Acétyleglucosamines et 2 acides sialiques.

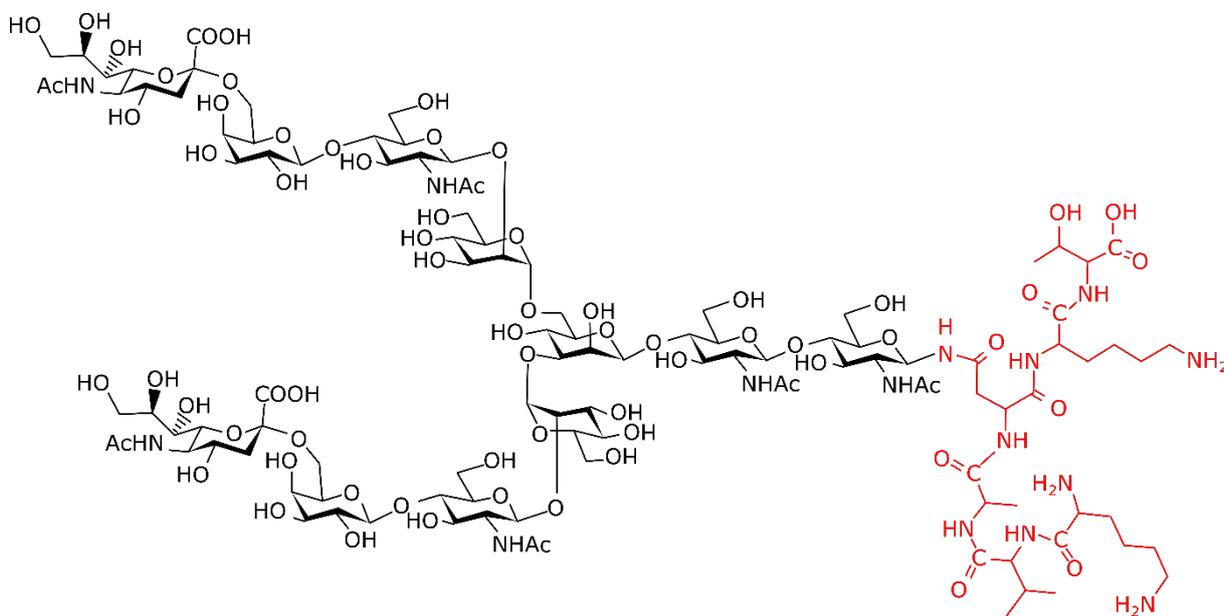


Figure 6 : Structure du glycopeptide de poulet, la séquence KVANKT (rouge) porte un N-glycane (noir) au niveau de la chaîne latérale de l'asparagine.

Après infusion directe en spectrométrie de masse, le glycopeptide de poulet présente deux états de charge :  $m/z = 1433$  ( $[M + 2H]^{2+}$ ) et  $m/z = 956$  ( $[M + 3H]^{3+}$ ) (Figure 7a). Dans un second temps, l'ion deux fois chargé ( $m/z = 1433$ ) est isolé puis irradié (30 ms) dans le but d'acquérir son spectre IRMPD (Figure 7b). Le spectre IRMPD de cette espèce montre une signature relativement simple. Deux

principaux pics sont présents sur ce spectre, ils se trouvent dans la zone des modes de vibration des liaisons O-H. Tout d'abord, le pic le plus intense est présent à  $3503\text{ cm}^{-1}$ , ensuite un second pic se trouve à  $3441\text{ cm}^{-1}$ . Enfin un massif de pics non résolus est observé ( $3314$ ,  $3335$ ,  $3354$ ,  $3373$  et  $3389\text{ cm}^{-1}$ ) est visible à des nombres d'onde plus faibles. On note que le spectre IRMPD présente des valeurs de rendement de photofragmentation élevées (jusqu'à 1). La somme des intensités des ions fragments est donc supérieure à l'intensité du précurseur (ce qu'on voit sur la Figure 7 c). Sur ce spectre, nous voyons différents fragments, mais le précurseur a totalement disparu, il est complètement photofragmenté. Le spectre de photofragmentation présente un pic à  $m/z = 657$  ce qui traduit la présence du trisaccharide GlcNAc-Gal-Neu5Ac (avec Neu5Ac : acide N-Acétylneuraminique qui est un acide sialique) contenu dans le glycane lié au peptide. Pour comparaison, la Figure 13 montre les spectres IRMPD de deux peptides et du glycopeptide de poulet sur la même échelle.

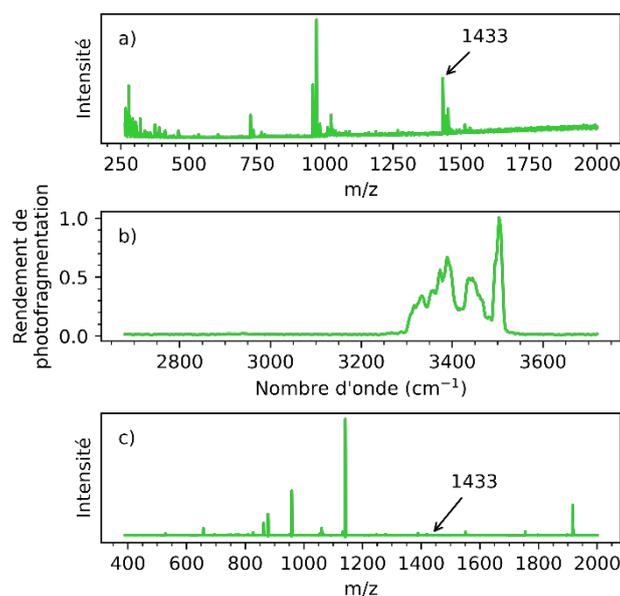


Figure 7 : Spectre de masse du glycopeptide de poulet avec les espèces deux fois protonée  $m/z = 1433$  et trois fois protonée  $m/z = 956$  a), spectre IRMPD de l'espèce deux fois protonée à  $m/z = 1433$  du glycopeptide de poulet (espèce  $[M+2H]^{2+}$ ) b) et spectre de photofragmentation (irradiation à  $3360\text{ cm}^{-1}$ ) de l'espèce deux fois protonée  $m/z = 1433$  du glycopeptide de poulet c).

## 4.2 PROTOCOLE

### 4.2.1 Digestion de la protéine par des enzymes protéolytiques

Le protocole est présenté en Figure 8. La fêtuine bovine (entre 200 et 500  $\mu\text{g}$ ) est dissoute dans 50  $\mu\text{L}$  de tampon de bicarbonate d'ammonium (50 mM à  $\text{pH} = 8,0$ ). 5,5  $\mu\text{L}$  de dithiothréitol (DTT) à 150 mM et 40  $\mu\text{L}$  d'urée à 8 M sont ajoutés à la solution, la solution est incubée pendant 40 min à  $60\text{ }^\circ\text{C}$ . 17  $\mu\text{L}$  d'iodoacétamide (IAA) 150 mM est ajouté à la solution. La solution est de nouveau laissée à incuber 40 min à température ambiante à l'obscurité. Enfin 3 mL de solution de bicarbonate d'ammonium sont ajoutés afin de faire chuter la concentration des agents dénaturants afin que les enzymes protéolytiques ne soient pas inhibées. 10  $\mu\text{L}$  de trypsine à  $2\text{ mg}\cdot\text{mL}^{-1}$  et entre 1 et 3  $\mu\text{L}$  de chymotrypsine à  $1\mu\text{g}\cdot\mu\text{L}^{-1}$  sont ajoutés à la solution. Ces deux enzymes digèrent les protéines en peptides. L'incubation se fait à  $37\text{ }^\circ\text{C}$  pendant toute une nuit.

La trypsine clive les protéines en C-ter d'acides aminés basiques à savoir la lysine ou l'arginine (excepté si l'acide aminé en C-ter de l'acide aminé basique est une proline). La chymotrypsine elle clive les protéines en C-Ter des acides aminés qui sont la tyrosine, le tryptophane, la phénylalanine, la

méthionine et dans une moindre mesure la leucine, là aussi, la chymotrypsine n'agit pas en cas de présence d'une proline en C-ter de ces acides aminés.

La protéine est mise en solution dans un tampon permettant son bon repliement 3D (Figure 8a), l'ajout d'urée à haute concentration dénature la protéine, de plus, l'utilisation de DTT qui rompt les ponts disulfures (en les réduisant) parachève cette dénaturer (Figure 8b). L'IAA permet alors d'alkyler les thiols des cystéines maintenant libres afin de prévenir la reformation de ponts disulfures (Figure 8c). Ainsi lors de la prochaine étape, la baisse de molarité de l'urée et du DTT ne s'accompagnera pas du repliement de la protéine. L'ajout d'une grande quantité de tampon diminue la concentration de l'urée et du DTT et permet aux protéases utilisées de cliver la protéine d'intérêt (Figure 8d).

A la suite de la digestion, il est encore nécessaire de purifier les produits de digestion. En effet, de nombreuses impuretés sont présentes dans le milieu de réaction. Pour cela, nous procédons à une étape de SPE (Solid Phase Extraction) au moyen de cartouches HLB de chez Supelco (3 mL). Grâce à cette étape de purification, les différents réactifs utilisés lors de notre manipulation sont éliminés (urée, DTT, IAA, trypsine et chymotrypsine...). Seuls les peptides et les glycopeptides sont conservés.

Après élimination des impuretés, les produits de digestion sont congelés, ou repris dans le volume voulu afin de procéder à une analyse en LC-MS.

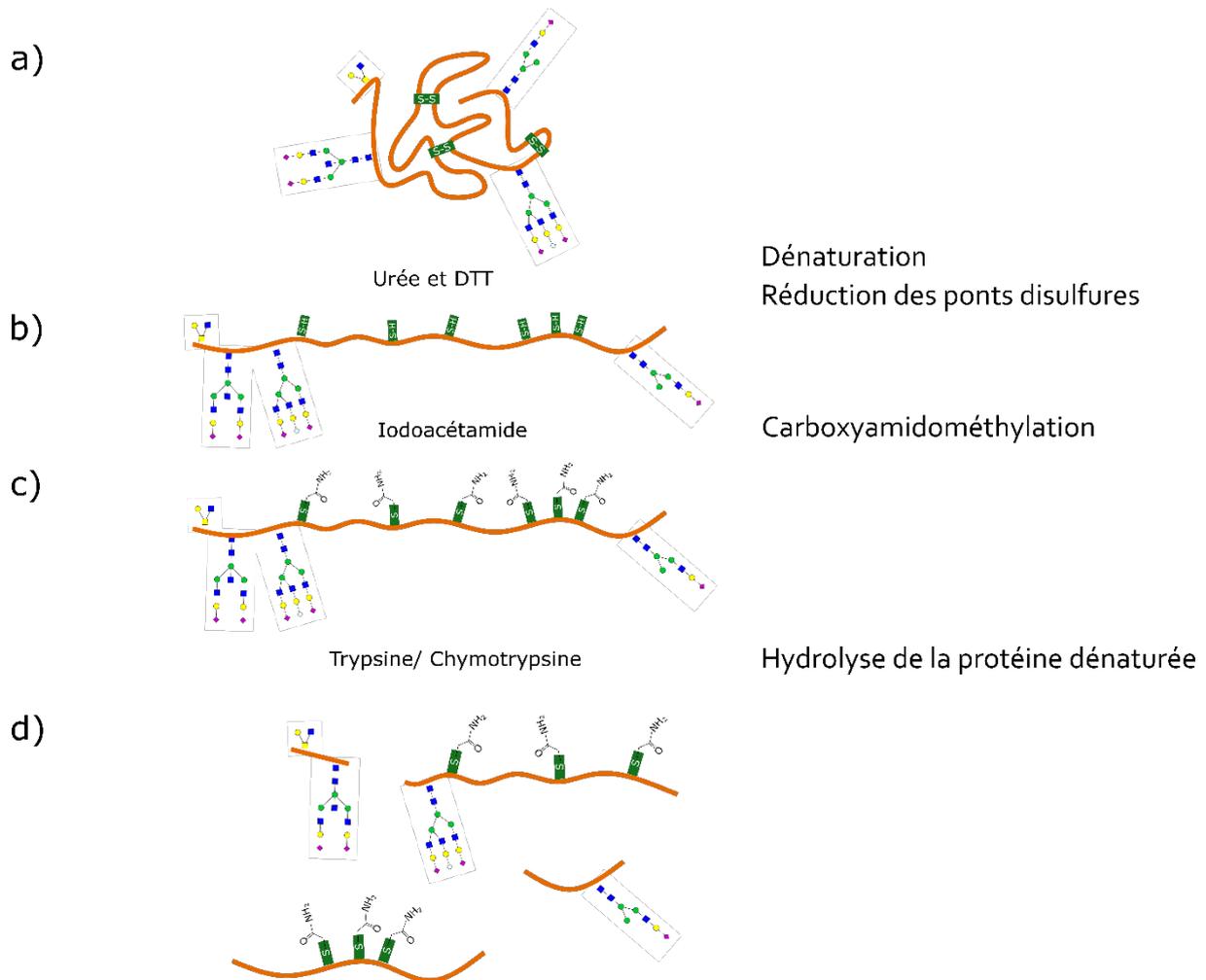


Figure 8 : Protocole de digestion de protéines au moyen de la trypsine ou de la chymotrypsine. La protéine est en solution dans un tampon permettant son bon repliement a), la protéine est dénaturée au moyen d'un agent chaotrope : l'urée, les ponts disulfures sont réduits par l'usage de DTT b), les thiols réduits sont alkylés afin d'empêcher la formation de nouveaux ponts disulfure c) enfin, la protéine subit une digestion par une protéase, les peptides sont alors formés d).

Après digestion et purification, les peptides attendus sont présentés dans le Tableau 2. Ce tableau est produit par l'outil Expasy de l'institut Suisse de bioinformatique<sup>49</sup> sur l'entrée p12763 de la base de données Uniprot KB<sup>19</sup>. Ce tableau génère les masses des peptides auxquels nous pouvons nous attendre, mais il compile également les modifications que les peptides portent, que ces modifications soient naturelles telles que la phosphorylation d'un peptide ou alors artificielles avec l'alkylation des cystéines. Ce tableau ne présente toutefois pas les masses des peptides portant un glycanes, en effet, pour un même peptide, la diversité des glycanes que ce sucre peut porter impliquerait un grand nombre de masses possibles et le tableau serait alors complexe à étudier. C'est pourquoi, les acides aminés susceptibles de porter un glycanes ont été colorés en rouge.

Compte tenu de la gamme de m/z observable, il est aisé de voir que certains ions vont être visibles en étant une seule fois chargé, tandis que d'autres ne seront observables que si l'ion se charge davantage (c'est notamment le cas pour les glycopeptides dont la chaîne peptidique excède déjà 2000 Da)

Tableau 2: Différents fragments attendus après digestion par les protéases utilisées.

Position	Masse	Modification dues au protocole	Masse après modification	PTM	Masse après PTM	Séquence du peptide
252-306	5293,7854					QTQPIVQRPDQGAENAFPS AVPDAAQF TPSAAGFPVVASV WGP <sup>S</sup> VWAVPLPLHR
79-103	2755,3277	Cys_CAM: 89, 100	2869,3706			DIETDLETTCHEMLDPTPLA <sup>N</sup> CSVR
29-47	1993,8331	Cys_CAM: 32	2050,8545			EPACDQPDTEQAALAAVDY
104-120	1920,9232	Cys_CAM: 114	1977,9447			QQTQHAVEGDCDIHWLK
145-159	1626,7978	Cys_CAM: 146, 149	1740,8407			LQPDGFLAFL <sup>N</sup> DSR
334-348	1474,8376					TPVGGP <sup>S</sup> IPGGPVR
316-329	1313,5855			PHOS: 316, 320,323,325	1633,4508	SGAVASVESSSGEAF
132-143	1280,5059	Cys_CAM: 132	1337,5274	PHOS: 134, 135,138	1520,4049	CDSSPDSAEDVR
160-171	1255,7044					VHAAVEVALATF
191-201	1172,6561					VPLPVSVSEF
58-67	1154,6164					HTLNQIDSVK
202-211	962,4975	Cys_CAM: 208	1019,5190			AVAATDQIAK
180-187	957,5727					LQVVEISR
19-27	944,5087					IPLDPVAGY
172-179	841,3322					NAES <sup>N</sup> KSSY
72-78	821,4152					RPTGEVY
238-245	816,4210					ALGGEDVR
219-225	790,4127	Cys_CAM: 219	847,4342			CNLLAEK
212-218	787,4196					EVVDPTK
246-251	683,3433	Cys_CAM: 248	740,3647			VICTLF
232-237	631,3773					GSVIQK
121-125	594,2518					QDQGF
68-71	557,3194					VWPR
349-353	545,2864	Cys_CAM: 350	602,3079			LQPR
51-54	522,3147					HLPR

Devant cette diversité, une première étape de séparation chromatographique avant analyse par spectrométrie de masse est requise. La méthode utilisée est présentée dans la partie suivante

#### 4.2.2 Passage en LC-MS

La séparation en chromatographie liquide se fait au moyen du même appareil que présenté dans le Chapitre III à savoir une chaîne chromatographique Ultimate 3000 de chez ThermoFisher Scientific. Cette fois-ci, le mode chromatographique utilisé est le mode inverse, la colonne est une XSelect CSH C18 de chez Waters dont les dimensions sont les suivantes diamètre de particules 3,5  $\mu\text{m}$  diamètre interne de la colonne 2,1 x longueur de la colonne 150 mm. Durant toute la séparation, la colonne est maintenue à 22°C. La phase mobile est un gradient (Figure 9) programmé entre A de l'eau et B de l'acétonitrile, tous deux contenant 0,5 % d'acide formique.

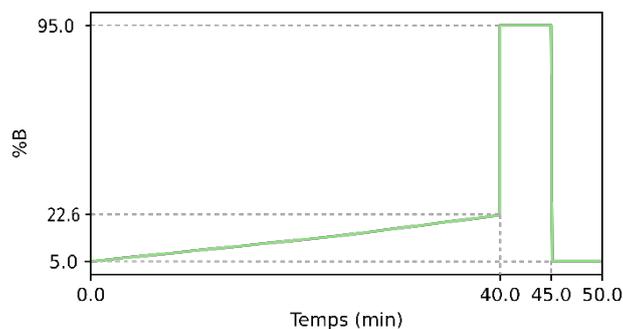


Figure 9 : Gradient utilisé pour la séparation des peptides/glycopeptides. La proportion d'acétonitrile acidifié passe de 5 à 22,6 % en 40 min. La proportion monte ensuite à 95 % pendant 5 min avant une étape de rééquilibration pendant 5 min pour procéder à de nouvelles injections.

Dans ces mesures, la spectrométrie de masse est opérée au moyen du piège à ions LTQ-XL de ThermoFinnigan. La spectrométrie de masse est utilisée en mode positif.

#### 4.2.3 Irradiation par le laser

Le montage est similaire à ce qui a été décrit précédemment (Chapitre 2) le laser utilisé est le laser haute cadence Firefly de chez MSquared. Le temps d'irradiation est contrôlé avec un shutter optique (SH05-ThorLabs).

### 4.3 RESULTATS

#### 4.3.1 Chromatogramme des produits de digestion de la fétuine bovine

Différentes mesures en MS ont eu lieu après la séparation des peptides, tout d'abord, un simple full scan est acquis afin de rechercher le plus de masses attendues (Figure 10). Le chromatogramme de la digestion de la fétuine bovine présente de nombreux pics (Figure 10a), parmi eux on retrouve les masses pouvant correspondre à des peptides prédits précédemment, trois sont choisis à des fins d'illustration (Figure 10b).

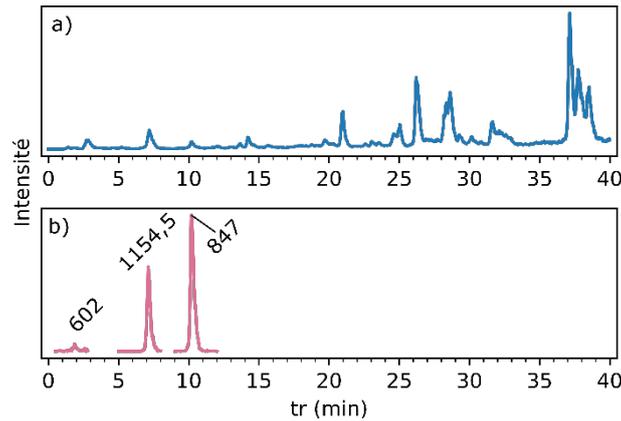


Figure 10 : Chromatogramme de la digestion de la fêtuine bovine a), suivi de trois peptides obtenus après digestion de la fêtuine bovine : LCPGR  $[M+H]^+$   $m/z = 602$ , CNLLAEK  $[M+H]^+$   $m/z = 847$  et HTLNQIDSVK  $[M+H]^+$   $m/z = 1154,5$  b).

L'appareil dont nous disposons, est équipé pour faire de la fragmentation par CID. Dans le but d'identifier formellement les signaux précédemment détectés, une nouvelle analyse de la solution de digestion de la fêtuine bovine est relancée. Cette fois-ci, les spectres de fragmentation des supposés peptides sont acquis afin de formellement les identifier. Dans ces conditions, les fragments générés sont principalement les séries d'ions b et y, ce sont ces séries d'ions qui nous permettront d'identifier formellement les peptides. Les conditions de chromatographie liquide sont les mêmes, seuls les paramètres de la spectrométrie de masse changent, deux évènements sont alors programmés. Tout d'abord un full scan permet de suivre l'ensemble des ions présents dans la solution. Ensuite, les masses intéressantes sont isolées et fragmentées par CID. Les spectres de fragmentation des supposés peptides montrent la présence de fragments b et y correspondant aux séquences attendues (Figure 11).

Ainsi, les signaux détectés correspondent bien aux peptides LCPGR (Figure 11a), CNLLAEK (Figure 11b) et HTLNQIDSVK (Figure 11c). Mis à part l'alkylation sur les cystéines des peptides LCPGR et CNLLAEK, aucune modification n'est présente sur ces peptides.

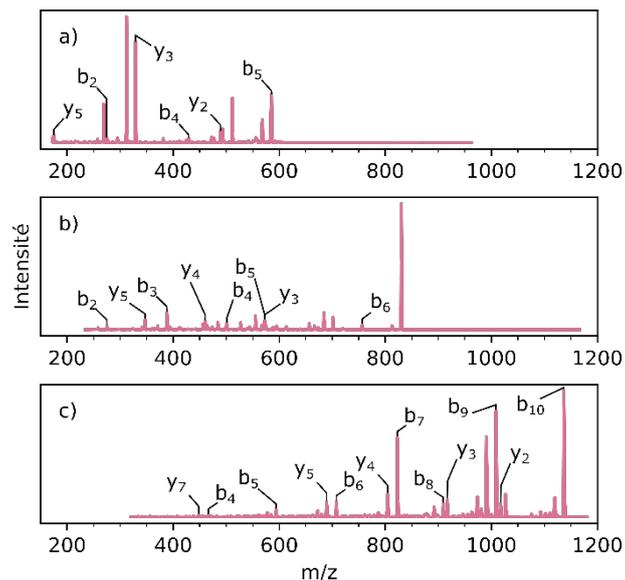


Figure 11 : Spectres de fragmentation par CID de différents peptides recherchés précédemment, LCPGR  $[M+H]^+$   $m/z = 602$  a), CNLLAEK  $[M+H]^+$   $m/z = 847$  b) et HTLNQIDSVK  $[M+H]^+$   $m/z = 1154,5$  c). Certains fragments sont annotés pour confirmer l'identification des peptides.

Après avoir identifié les peptides, une mesure supplémentaire a eu lieu. Cette nouvelle mesure a pour but de regarder l'effet de l'irradiation sur ces peptides. Les conditions de chromatographie liquide sont donc les mêmes, l'énergie de fragmentation en CID est nulle. Les ions ne sont soumis à aucune fragmentation par collision avec les atomes de gaz au sein du piège, ils sont seulement irradiés à  $3425\text{ cm}^{-1}$  afin d'étudier leur comportement sous irradiation IR (Figure 12). Le laser irradie les ions pendant 30 ms à une puissance d'environ 130 mW.

Avec ces conditions, le peptide LCPGR (Figure 12a) ne montre pas la présence de fragments. De même pour le peptide CNLLAEK (Figure 12b), bien que l'on observe un faible signal à 830 m/z, ce signal est aussi présent en absence d'irradiation, il peut s'agir de fragmentation résiduelle de ce fragment, ce qui n'a rien à voir avec l'irradiation des ions. Enfin, le peptide HTLNQIDSVK (Figure 12c) qui contient un hydroxyle sur la chaîne latérale de sa thréonine ne semble pas non plus présenter de photofragmentation dans ces conditions.

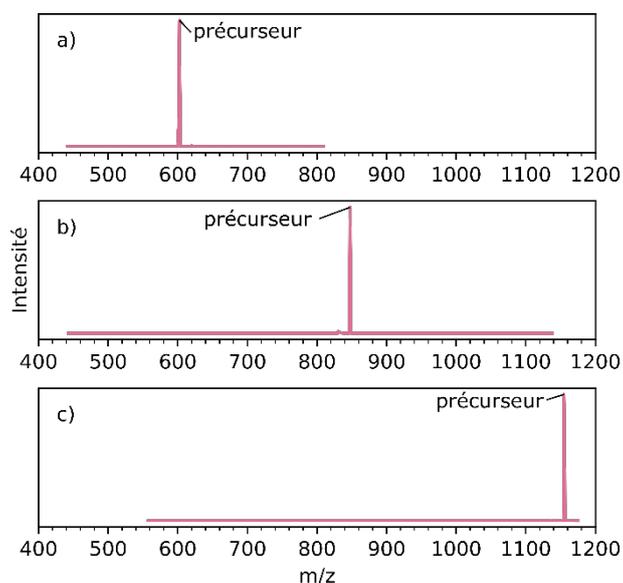


Figure 12 : Spectres de photofragmentation des peptides LCPGR  $[M+H]^+$   $m/z = 602$  a), CNLLAEK  $[M+H]^+$   $m/z = 847$  b) et HTLNQIDSVK  $[M+H]^+$   $m/z = 1154,5$  c). Les peptides sont isolés pour être irradiés dans le piège à ions (30 ms à 130 mW), aucun photofragment n'apparaît.

#### 4.3.2 Spectre IRMPD d'un peptide, pouvant porter un acide aminé avec un hydroxyle

Nous avons ensuite acquis en ligne après séparation chromatographique les spectres IRMPD de deux de ces peptides : CNLLAEK et HTLNQIDSVK entre  $3300$  et  $3600\text{ cm}^{-1}$ . Cette fois-ci, le temps d'irradiation est allongé à 80 ms (Figure 13). La Figure 13 présente ces spectres IRMPD, le spectre IRMPD du glycopeptide de poulet est ajouté pour avoir une référence à l'échelle (il n'est irradié que 30 ms) Les deux peptides ne présentent pas de photofragmentation qui soit comparable à ce qui est observé dans le cas du glycopeptide de poulet (Figure 13c). Le spectre IRMPD du peptide CNLLAEK (Figure 13a) semble présenter un rendement de photofragmentation relativement plus important aux alentours de  $3340\text{ cm}^{-1}$ . Ce rendement de photofragmentation demeure toutefois dérisoire par rapport à ce qui est observé pour le glycopeptide de poulet. De plus, bien qu'il contienne un groupement hydroxyle (sur sa thréonine), le peptide HTLNQIDSVK (Figure 13b) ne montre pas non plus de photofragmentation de son précurseur.

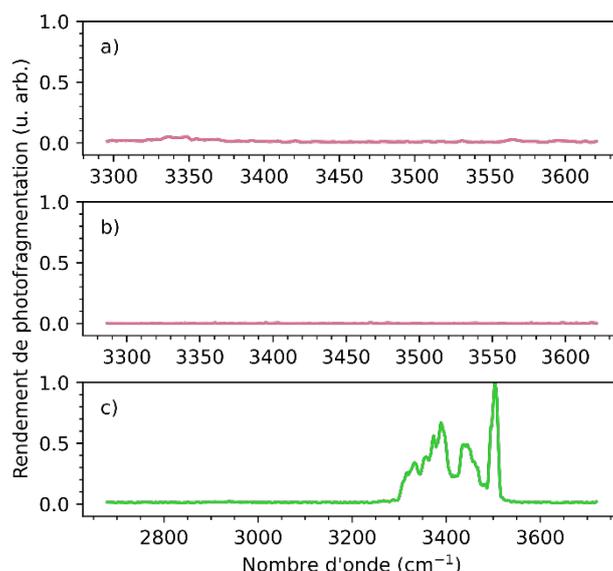


Figure 13 : Spectres IRMPD des peptides CNLLAEK  $[M+H]^+$   $m/z = 847$  a) et HTLNQIDSVK  $[M+H]^+$   $m/z = 1154,5$  b) et de l'espèce deux fois protonée  $m/z = 1433$  du glycopeptide de poulet c). Les peptides sont irradiés pendant 80 ms à 130 mW, le glycopeptide de poulet lui n'est irradié que pendant 30 ms. Contrairement au glycopeptide de poulet qui se photofragmente dans ces conditions, aucune photofragmentation n'advient dans le cas des peptides.

#### 4.3.3 Spectre IRMPD de candidats glycopeptides

Après avoir observé comment se comportent les peptides, nous nous sommes intéressés à des espèces dont le rapport  $m/z$  n'est pas référencé en tant que peptides. Ces espèces apparaissent toutefois comme intéressantes. En effet leur fragmentation par CID laisse apparaître l'ion à  $m/z = 657$ , ce qui suggère la présence d'un N-glycane contenant le trisaccharide Neu5Ac $\beta$ 1-3-Gal $\beta$ 1-4GlcNAc. Lors de la recherche de glycopeptides par LC-MS/MS, il est d'usage de rechercher des ions reporters<sup>24</sup>. Ces ions sont issus de la fragmentation du glycopeptide. L'ion à  $m/z = 657$  est un de ces ions reporter. Nous l'illustrons ici avec les espèces à  $m/z = 1875$  et  $m/z = 1436$  dont l'un de leurs produits de fragmentation est l'espèce à  $m/z = 657$ . Ces supposés glycopeptides sont détectés à des temps de rétention plus importants que les peptides précédemment identifiés (Figure 14).

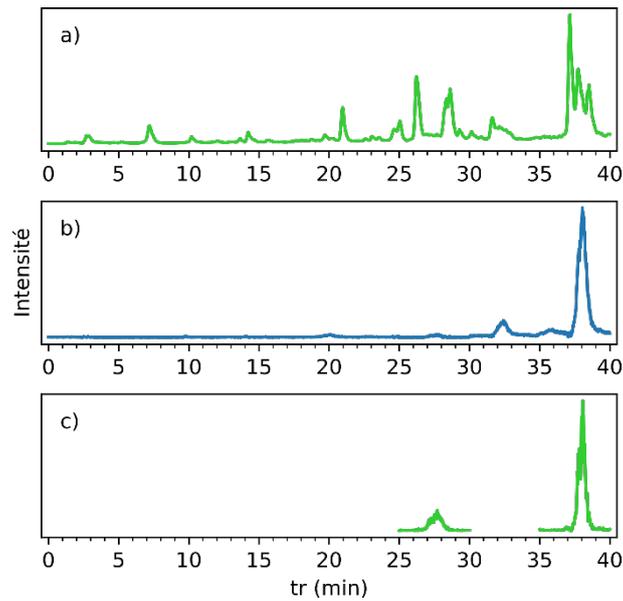


Figure 14 : Chromatogramme de la digestion de la fétuine bovine a), chromatogramme reconstruit suivant le fragment à  $m/z = 657$  b), les espèces à  $m/z = 1875$  et  $m/z = 1436$  sont ainsi observés après avoir recherché le fragment  $m/z = 657$  dans le chromatogramme des produits de digestion de la fétuine bovine c).

Ces espèces ainsi détectées sont isolées puis fragmentées par collision. Les fragments ainsi obtenus sont présentés dans la Figure 15. Ces deux espèces présentent dans leur spectre de fragmentation l'ion à  $m/z = 657$ .

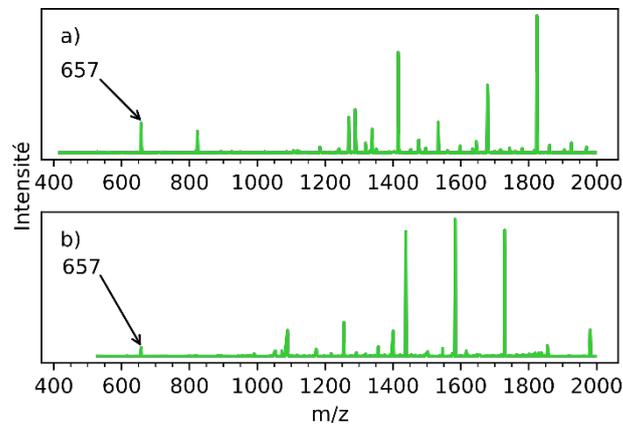


Figure 15 : Spectres CID des espèces à  $m/z = 1436$  a) et  $m/z = 1875$  b). La fragmentation par collision produit un ion à  $m/z = 657$  caractéristique d'un glycanes : le trisaccharide Neu5Ac-Gal $\beta$ -GlcNAc protoné.

Le trisaccharide Neu5Ac-Gal $\beta$ -GlcNAc est un trisaccharide souvent retrouvé au niveau de la terminaison du N-glycane, il peut être présent sous deux formes telles que présentées dans la Figure 16. Ces deux trisaccharides servent d'ions « reporter » des glycopeptides.

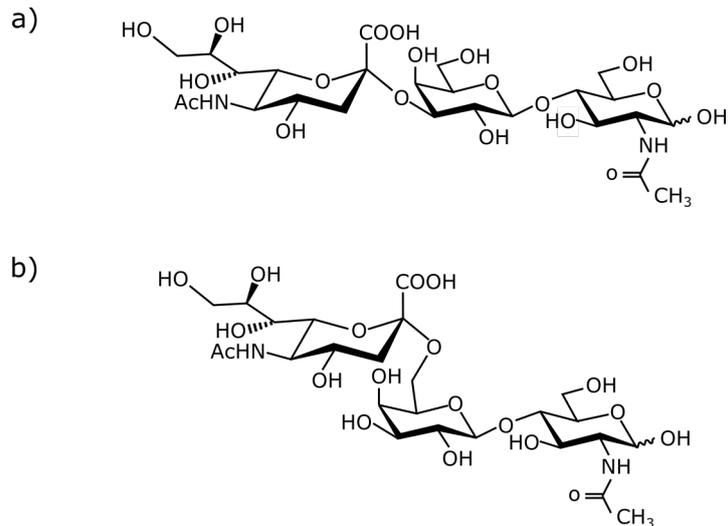


Figure 16 : Structure de Neu5Acβ1-3-Galβ1-4GlcNAc a) et de Neu5Acβ1-6-Galβ1-4GlcNAc b) deux formes retrouvées dans la terminaison des N-glycanes de la fêtuine bovine.

Comme dit plus haut, il est pertinent de rechercher des « ions reporter ». Le trisaccharide à  $m/z = 657$  est comme ses homologues les ions à  $m/z = 292$ , 366 et 454 un de « ion reporter » qui confirme la présence de sucres dans l'ion parent. Ils sont observés lorsque les ions parents subissent une fragmentation. Dans notre cas, nous ne pouvons pas voir les ions  $m/z = 292$ , 366 et 454 la gamme de  $m/z$  de notre appareil ne le permet pas toujours. Toutefois, la fragmentation de cet ion (présentée en Figure 17) montre des fragments intéressants. Le spectre de cet ion présente différents pics qui peuvent être assignés aux fragments attendus. Les ions à  $m/z = 274$ , 366 et 639 correspondent à la série des fragments b (b1, b2 et b3 respectivement). Les ions à  $m/z = 204$  et 366 correspondent à la série z (z2 et z1 respectivement). Enfin l'ion à  $m/z = 292$  correspond au fragment c1. D'autres ions sont aussi observables sur ce spectre :  $m/z = 621$  qui correspond à la perte d'un deuxième hydroxyle sur l'ion précurseur (ou b3 - 18) ainsi que deux ions à  $m/z = 256$  et 238 qui peuvent également provenir de la perte d'un ou deux hydroxyles depuis l'ion b1. La Figure 17b présente les fragments correspondants.

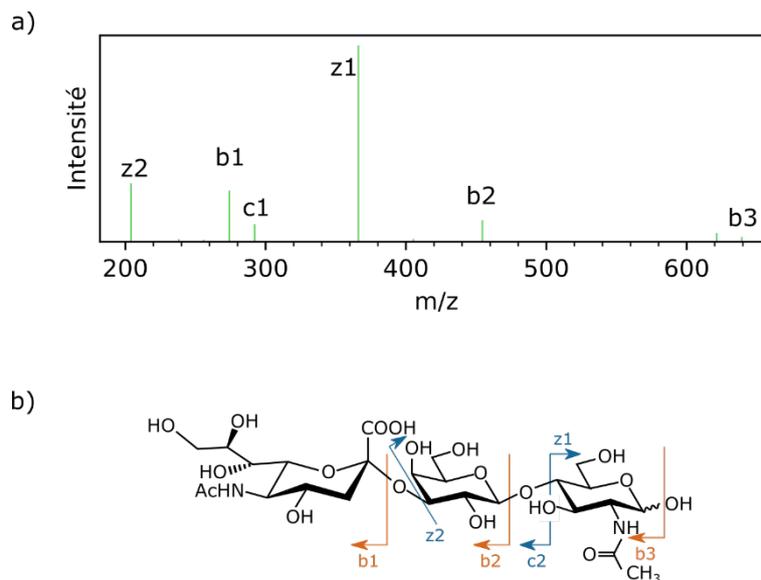


Figure 17 : Spectre CID de l'espèce à  $m/z = 657$  avec annotations des pics selon la nomenclature de Domon et Costello<sup>50</sup> a) et fragmentation du sucre correspondant (sur l'isomère Neu5Acβ1-3-Galβ1-4GlcNAc) b).

La recherche d'ions reporter n'est pas universelle, elle ne permet pas de détecter tous les types de glycanes : tout d'abord, selon, le type de N-glycanes, il ne portera pas forcément d'acide sialique sur sa structure. Les O-glycanes offrent une variété de structures bien plus importante que ce qui est observé dans le cas des N-glycanes, et ces structures ne portent pas forcément un acide sialique dans leur structure.

Après avoir repéré ces ions, nous avons acquis leur spectre IRMPD sur la gamme 3290-3620  $\text{cm}^{-1}$  (Figure 18). Au-dessus de 3575  $\text{cm}^{-1}$ , aucune photofragmentation n'est observée (a et b). Ensuite un faible pic s'élève (3530-3575  $\text{cm}^{-1}$ ) avant de voir un creux dans le spectre (3530  $\text{cm}^{-1}$ ) dans le cas des deux ions. Enfin un massif très important se présente : c'est le cas de l'espèce à  $m/z = 1436$ , dans le cas de l'espèce à  $m/z = 1875$ , un sillon se creuse à 3400  $\text{cm}^{-1}$ . Dans le cas de l'espèce à  $m/z = 1436$ , le massif est toujours très important en dessous de 3300  $\text{cm}^{-1}$ , tandis qu'il a déjà commencé à refluer à ces nombres d'onde dans le cas de l'espèce à  $m/z = 1875$ .

Ces irradiations ont été effectuées en irradiant uniquement pendant 30 ms les ions (ce qui est moins de la moitié de la durée d'irradiation des peptides en Figure 13) et le rendement de photofragmentation dépasse l'unité.

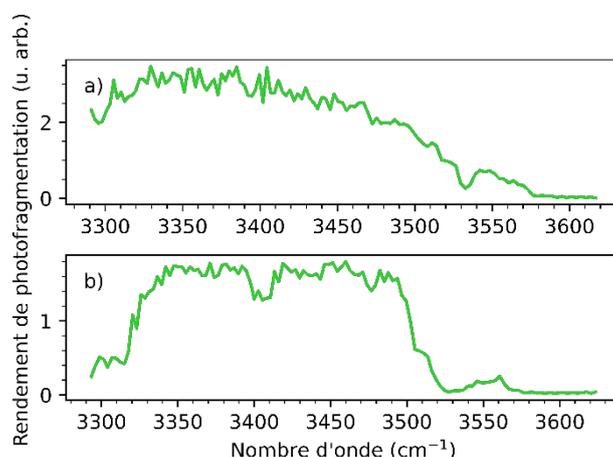


Figure 18 : Spectres IRMPD (entre 3290 et 3620  $\text{cm}^{-1}$ ) des deux composés d'intérêt l'espèce à  $m/z = 1436$  a) et l'espèce  $m/z = 1875$  b). Les spectres sont obtenus en irradiant les ions pendant 30 ms avec une puissance d'environ 110 mW.

Après avoir acquis les spectres IRMPD de ces deux espèces, nous nous sommes intéressés aux fragments provenant de cette irradiation (Figure 19). Comme c'est le cas pour la fragmentation par CID, ces deux espèces présentent un fragment à  $m/z = 657$ .

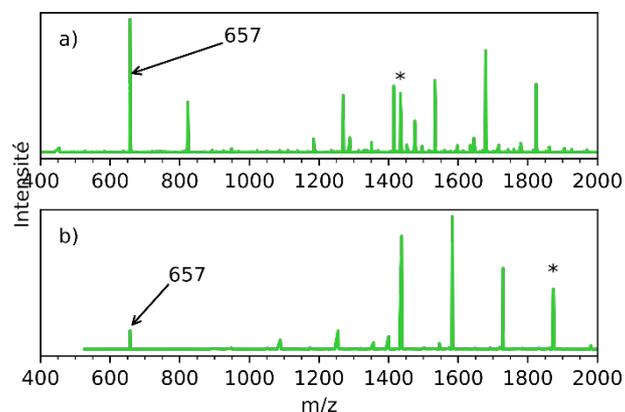


Figure 19 : Spectres de photofragmentation des deux composés d'intérêt l'espèce à  $m/z = 1436$  a) et l'espèce à  $m/z = 1875$  b). Le signal du précurseur est marqué d'un astérisque, ces deux spectres comportent du signal à  $m/z = 657$  (correspondant à GlcNAc-Gal-Neu5Ac qui est souvent la terminaison des N-glycanes).

Au vu des spectres de ces deux composés, on s'aperçoit qu'ils ont les mêmes caractéristiques que le glycopeptide de poulet. En effet, on note une forte absorption aux environs de  $3500\text{ cm}^{-1}$ . De plus ces spectres sont également caractérisés par un fort taux de photofragmentation tout le long du spectre.

#### 4.3.4 Screening de glycopeptides

Afin de retrouver d'autres espèces se photofragmentant également, une analyse en DDA (Data Dependent Acquisition) a été menée.

Le mode DDA est une analyse non ciblée qui repose sur l'alternance de deux types d'acquisition. Le spectromètre de masse réalise une alternance entre un full scan et des spectres MS/MS. Le spectre acquis en mode full scan indique au spectromètre de masse quel est l'ion à fragmenter lors du spectre suivant. Généralement le mode DDA est réglé tel que lorsque le full scan est acquis, les trois ions les plus intenses observés sur ce spectre sont successivement isolés et fragmentés avant que le spectromètre de masse ne recommence un nouveau cycle. Il est judicieux qu'après que les spectres MS/MS de ces trois ions soient acquis, ces ions soient inclus dans une liste d'exclusion pour qu'ils ne monopolisent pas les acquisitions des spectres MS/MS et qu'il soit possible d'acquérir les spectres MS/MS d'ions avec des intensités plus faibles.

Dans notre cas, nous avons procédé comme suit : deux protocoles ont été réalisés. Tout d'abord, le mélange de peptides et glycopeptides est introduit en chromatographie liquide dans les mêmes conditions que présentées précédemment (partie 4.2.2). Le spectromètre de masse est réglé en mode DDA, après le full scan, les trois ions les plus intenses sont isolés et fragmentés (par CID ou par LID). La fragmentation par CID sert de référence et confirme l'identification faite par la photofragmentation induite par le laser. En pratique, le spectre de fragmentation obtenu en LID est obtenu en réglant la CID à 0 % et le laser IR irradie à  $3425\text{ cm}^{-1}$  avec une puissance de 114 mW pendant 30 ms.

Les chromatogrammes obtenus en mode DDA contiennent de très nombreux pics. Les spectres LID des peptides sont présentés en Annexe 1. Les peptides ne se photofragmentent pas, l'Annexe 1 contient également le spectre MS/MS obtenu par CID sur ces précurseurs pour confirmer l'identification de ces composés.

Maintenant, les glycopeptides sont présentés en Annexe 2, de la même manière, les spectres LID sont tracés. La fragmentation par CID montre des spectres MS/MS semblables aux spectres obtenus par photofragmentation.

La Figure 20 présente les rendements de photofragmentation en fonction de la nature du composé : peptide ou glycopeptide. Tout d'abord, les peptides sont pour la plupart plus légers que les glycopeptides (ce à quoi on pourrait s'attendre car les glycopeptides résultent de l'addition d'un glycane sur un peptide). Il apparaît ensuite clairement que l'ensemble des peptides ne se photofragmente pas, tandis que les glycopeptides montrent un rendement de photofragmentation élevé dans ces conditions. Les rendements de photofragmentation des glycopeptides montrent une grande dispersité (les rendements sont étagés entre 0,5 et 3,7).

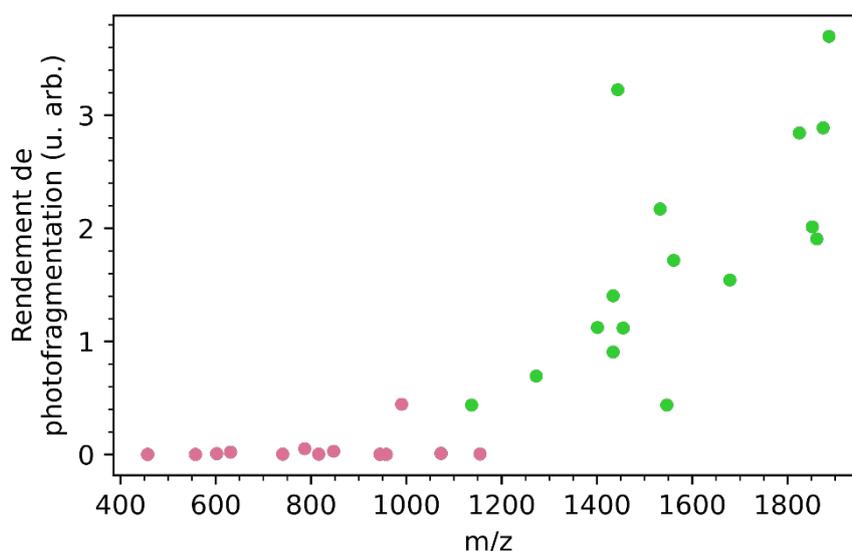


Figure 20 : Rendement de photofragmentation des peptides et des supposés glycopeptides sous une irradiation de 114 mW pendant 30 ms. Les peptides sont présentés en rose et les glycopeptides en vert.

Une ambiguïté est observée pour l'ion à  $m/z = 989,5$ , il s'agit d'un composé présentant de la photofragmentation, mais pas de fragments à  $m/z = 657$ . Son rapport  $m/z$  peut correspondre à un peptide obtenu après digestion enzymatique. En regardant le spectre CID de cet ion (Figure 21), on se rend compte qu'il s'agit bel et bien du spectre d'un peptide : le peptide QQTQHAVEGDCDIHVLIK.

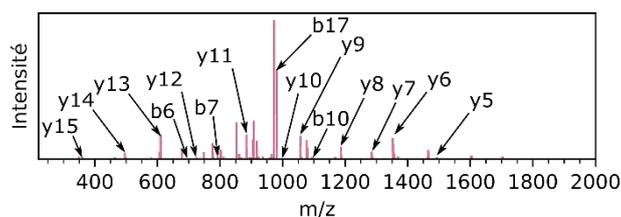


Figure 21 : Spectre de fragmentation du peptide QQTQHAVEGDCDIHVLIK  $[M+2H]^{2+}$   $m/z = 989,5$

La Figure 20 présente le rendement de photofragmentation de peptides (identité confirmée par CID) et des supposés glycopeptides. Nous voyons que tous les candidats glycopeptides se photofragmentent dans les conditions utilisées (irradiation 114 mW pendant 30 ms). Du côté des peptides, un seul se comporte de la même façon (rendement de photofragmentation d'environ 0,5). Les glycopeptides sont donc facilement discriminés par ce screening dans un mélange de peptides/glycopeptides provenant de la digestion de protéines.

Après cette première étape de screening, il est possible de se concentrer uniquement sur ces ions afin de les caractériser plus finement par des mesures en spectrométrie de masse à plus haute résolution. Dans notre cas, un faux positif (peptide présentant de la photofragmentation) est observé, nous nous attendons donc à en voir lors de l'analyse d'autres types de protéines mais ce n'est pas le cas de la majorité des peptides. Ici, seuls 13 peptides et 16 candidats glycopeptides sont présentés, ce qui ne représente pas une grande quantité de données, l'utilisation de cette méthode sur un plus grand nombre de protéines optimisera les conditions expérimentales, et affinera ces observations.

## 5 CONCLUSION

Les composés glycoconjugués sont des macromolécules complexes à caractériser. Dans notre cas, nous avons montré qu'il est toutefois possible de déterminer si une protéine contient ou non des

glycosylations simplement par spectroscopie IR en phase condensée. Ce type d'analyse permet en outre de ne pas détruire l'échantillon, qui reste disponible pour d'autres analyses ultérieures.

Pour caractériser plus finement un tel composé, la complexité augmente. Dans le développement présenté ici, nous proposons d'irradier les produits de digestion d'une protéine supposée glycosylée. Les peptides ne se photofragmentent pas sous irradiation laser à  $3425\text{ cm}^{-1}$ , tandis que les supposés glycopeptides se photofragmentent. Dans les cas de la fragmentation par CID, ou de la photofragmentation par IRMPD, des « glycopeptides », les fragments observés proviennent de la fragmentation des sucres, ils ne sont donc pas informatifs sur la structure du peptide portant le sucre. L'utilisation d'un autre mode de fragmentation (comme l'ETD par exemple) donnera ces informations manquantes. De plus, l'utilisation de spectromètres de masse à haute résolution offrira une plus grande facilité dans l'attribution des spectres de masse à un glycopeptide. La résolution du piège à ions n'offre pas une grande confiance sur la masse déconvoluée pour attribuer une structure à cette masse.

Les développements effectués dans cette partie, n'ont pas vocation à se limiter à l'analyse des sucres, mais pourront également être utilisés à d'autres domaines des -omiques. En effet, nous nous attendons à voir d'autres applications se développer au moyen d'un tel screening notamment la détection d'autres PTMs avec par exemple la sulfatation ou la phosphorylation. Ces deux modifications présentent des structures isobares (ions précurseurs et fragments) d'où la difficulté de les distinguer. Bien qu'ils présentent des masses très proches, la spectroscopie IR, a déjà permis de distinguer des peptides phosphorylés ou sulfatés<sup>39</sup>.

Le screening par IRMPD est un bon préfiltre aidant à l'élimination de nombreux signaux correspondants aux peptides. Bien qu'il arrive que des peptides puissent également se photofragmenter, nous n'en avons observés qu'un seul. Il est donc possible de ne s'attarder que sur les glycopeptides.

## 6 BIBLIOGRAPHIE

- 1 Y.-H. Lin, V. Franc and A. J. R. Heck, Similar Albeit Not the Same: In-Depth Analysis of Proteoforms of Human Serum, Bovine Serum, and Recombinant Human Fetuin, *J. Proteome Res.*, 2018, **17**, 2861–2869.
- 2 W. B. Struwe and C. V. Robinson, Relating glycoprotein structural heterogeneity to function – insights from native mass spectrometry, *Curr. Opin. Struct. Biol.*, 2019, **58**, 241–248.
- 3 D. Wu and C. V. Robinson, Understanding glycoprotein structural heterogeneity and interactions: Insights from native mass spectrometry, *Curr. Opin. Struct. Biol.*, 2022, **74**, 102351.
- 4 K. Hirayama, R. Yuji, N. Yamada, K. Kato, Y. Arata and I. Shimada, Complete and Rapid Peptide and Glycopeptide Mapping of Mouse Monoclonal Antibody by LC/MS/MS Using Ion Trap Mass Spectrometry, *Anal. Chem.*, 1998, **70**, 2718–2725.
- 5 H. J. An, T. R. Peavy, J. L. Hedrick and C. B. Lebrilla, Determination of N-Glycosylation Sites and Site Heterogeneity in Glycoproteins, *Anal. Chem.*, 2003, **75**, 5628–5637.
- 6 I. Loke, N. H. Packer and M. Thaysen-Andersen, Complementary LC-MS/MS-Based N-Glycan, N-Glycopeptide, and Intact N-Glycoprotein Profiling Reveals Unconventional Asn71-Glycosylation of Human Neutrophil Cathepsin G, *Biomolecules*, 2015, **5**, 1832–1854.
- 7 T. Wohlschlagler, K. Scheffler, I. C. Forstenlehner, W. Skala, S. Senn, E. Damoc, J. Holzmann and C. G. Huber, Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals, *Nat. Commun.*, 2018, **9**, 1713.
- 8 J. M. Lee, H. M. Hammarén, M. M. Savitski and S. H. Baek, Control of protein stability by post-translational modifications, *Nat. Commun.*, 2023, **14**, 201.
- 9 G. A. Khoury, R. C. Baliban and C. A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database, *Sci. Rep.*, 2011, **1**, 90.
- 10 A. Helenius and M. Aebi, Intracellular Functions of N-Linked Glycans, *Science*, 2001, **291**, 2364–2369.
- 11 B. Adamczyk, T. Tharmalingam and P. M. Rudd, Glycans as cancer biomarkers, *Biochim. Biophys. Acta BBA - Gen. Subj.*, 2012, **1820**, 1347–1353.
- 12 Y. Mechref, Y. Hu, A. Garcia and A. Hussein, Identifying cancer biomarkers by mass spectrometry-based glycomics, *ELECTROPHORESIS*, 2012, **33**, 1755–1767.
- 13 D. N. Tabang, M. Ford and L. Li, Recent Advances in Mass Spectrometry-Based Glycomic and Glycoproteomic Studies of Pancreatic Diseases, *Front. Chem.*
- 14 J. Krištić, F. Vučković, C. Menni, L. Klarić, T. Keser, I. Beceheli, M. Pučić-Baković, M. Novokmet, M. Mangino, K. Thaqi, P. Rudan, N. Novokmet, J. Šarac, S. Missoni, I. Kolčić, O. Polašek, I. Rudan, H. Campbell, C. Hayward, Y. Aulchenko, A. Valdes, J. F. Wilson, O. Gornik, D. Primorac, V. Zoldoš, T. Spector and G. Lauc, Glycans Are a Novel Biomarker of Chronological and Biological Ages, *J. Gerontol. Ser. A*, 2014, **69**, 779–789.
- 15 A. A. Merleev, D. Park, Y. Xie, M. J. Kailemia, G. Xu, L. R. Ruhaak, K. Kim, Q. Hong, Q. Li, F. Patel, Y.-J. Y. Wan, A. I. Marusina, I. E. Adamopoulos, N. N. Lal, A. Mitra, S. T. Le, M. Shimoda, G. Luxardi, C. B. Lebrilla and E. Maverakis, A site-specific map of the human plasma glycome and its age and gender-associated alterations, *Sci. Rep.*, 2020, **10**, 17505.
- 16 A. Hay-Lombardie and E. Bigot-Corbel, Biomarqueurs permettant le suivi de l'équilibre glycémique du patient diabétique, *Rev. Francoph. Lab.*, 2018, **2018**, 33–43.
- 17 K. Mori, M. Emoto and M. Inaba, Fetuin-A: A Multifunctional Protein, *Recent Pat. Endocr. Metab. Immune Drug Discov.*, 2011, **5**, 124–146.
- 18 S. T. Mathews, P. R. Srinivas, M. A. Leon and G. Grunberger, Bovine fetuin is an inhibitor of insulin receptor tyrosine kinase, *Life Sci.*, 1997, **61**, 1583–1592.
- 19 AHSG - Alpha-2-HS-glycoprotein - Bos taurus (Bovine) | UniProtKB | UniProt, <https://www.uniprot.org/uniprotkb/P12763/entry#sequences>, (accessed August 24, 2023).
- 20 D. I. Christie, K. m. Dziegielewska, R. m. Hill and N. r. Saunders, Fetuin: The bovine homologue of human  $\alpha$ 2HS glycoprotein, *FEBS Lett.*, 1987, **214**, 45–49.

- 21 M. Windwarder and F. Altmann, Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry, *J. Proteomics*, 2014, **108**, 258–268.
- 22 L.-J. Huang, J.-H. Lin, J.-H. Tsai, Y.-Y. Chu, Y.-W. Chen, S.-L. Chen and S.-H. Chen, Identification of protein O-glycosylation site and corresponding glycans using liquid chromatography–tandem mass spectrometry via mapping accurate mass and retention time shift, *J. Chromatogr. A*, 2014, **1371**, 136–145.
- 23 P. Hägglund, J. Bunkenborg, F. Elortza, O. N. Jensen and P. Roepstorff, A New Strategy for Identification of N-Glycosylated Proteins and Unambiguous Assignment of Their Glycosylation Sites Using HILIC Enrichment and Partial Deglycosylation, *J. Proteome Res.*, 2004, **3**, 556–566.
- 24 M. Thaysen-Andersen, S. Mysling and P. Højrup, Site-Specific Glycoprofiling of N-Linked Glycopeptides Using MALDI-TOF MS: Strong Correlation between Signal Strength and Glycoform Quantities, *Anal. Chem.*, 2009, **81**, 3933–3943.
- 25 E. Balaguer and C. Neusüss, Glycoprotein Characterization Combining Intact Protein and Glycan Analysis by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry, *Anal. Chem.*, 2006, **78**, 5384–5393.
- 26 E. Giménez, F. Benavente, J. Barbosa and V. Sanz-Nebot, Towards a reliable molecular mass determination of intact glycoproteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 2555–2563.
- 27 E. Giménez, F. Benavente, J. Barbosa and V. Sanz-Nebot, Ionic liquid matrices for MALDI-TOF-MS analysis of intact glycoproteins, *Anal. Bioanal. Chem.*, 2010, **398**, 357–365.
- 28 B. R. Singh, D. B. DeOliveira, F.-N. Fu and M. P. Fuller, Fourier transform infrared analysis of amide III bands of proteins for the secondary structure estimation, *Biomol. Spectrosc. III*, 1993, **1890**, 47–55.
- 29 M. Baldassarre, R. Galeazzi, B. Maggiore, F. Tanfani and A. Scirè, Bovine  $\alpha$ 1-acid glycoprotein, a thermostable version of its human counterpart: Insights from Fourier transform infrared spectroscopy and in silico modelling, *Biochimie*, 2014, **102**, 19–28.
- 30 H. Yang, S. Yang, J. Kong, A. Dong and S. Yu, Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy, *Nat. Protoc.*, 2015, **10**, 382–396.
- 31 A. Natalello, D. Ami, S. Brocca, M. Lotti and S. M. Doglia, Secondary structure, conformational stability and glycosylation of a recombinant *Candida rugosa* lipase studied by Fourier-transform infrared spectroscopy, *Biochem. J.*, 2005, **385**, 511–517.
- 32 A. Derenne, K.-M. Derfoufi, B. Cowper, C. Delporte and E. Goormaghtigh, FTIR spectroscopy as an analytical tool to compare glycosylation in therapeutic monoclonal antibodies, *Anal. Chim. Acta*, 2020, **1112**, 62–71.
- 33 B. S. McAvan, A. P. France, B. Bellina, P. E. Barran, R. Goodacre and A. J. Doig, Quantification of protein glycation using vibrational spectroscopy, *The Analyst*, 2020, **145**, 3686–3696.
- 34 L. Voronina, C. Leonardo, J. B. Mueller-Reif, P. E. Geyer, M. Huber, M. Trubetskov, K. V. Kepesidis, J. Behr, M. Mann, F. Krausz and M. Žigman, Molecular Origin of Blood-Based Infrared Spectroscopic Fingerprints\*\*, *Angew. Chem.*, 2021, **133**, 17197–17206.
- 35 MB - Myoglobin - Equus caballus (Horse) | UniProtKB | UniProt, <https://www.uniprot.org/uniprotkb/P68082/entry>, (accessed October 21, 2022).
- 36 C. Lacombe, V. Untereiner, C. Gobinet, M. Zater, G. D. Sockalingum and R. Garnotel, Rapid screening of classic galactosemia patients: a proof-of-concept study using high-throughput FTIR analysis of plasma, *The Analyst*, 2015, **140**, 2280–2286.
- 37 A. Barth, Infrared spectroscopy of proteins, *Biochim. Biophys. Acta BBA - Bioenerg.*, 2007, **1767**, 1073–1101.
- 38 J. Martens, R. E. Van Outersterp, R. J. Vreeken, F. Cuyckens, K. L. M. Coene, U. F. Engelke, L. A. J. Kluijtmans, R. A. Wevers, L. M. C. Buydens, B. Redlich, G. Berden and J. Oomens, Infrared ion spectroscopy: New opportunities for small-molecule identification in mass spectrometry - A tutorial perspective, *Anal. Chim. Acta*, 2020, **1093**, 1–15.
- 39 A. L. Patrick, C. N. Stedwell and N. C. Polfer, Differentiating Sulfopeptide and Phosphopeptide

- Ions via Resonant Infrared Photodissociation, *Anal. Chem.*, 2014, **86**, 5547–5552.
- 40 B. Schindler, J. Joshi, A.-R. Allouche, D. Simon, S. Chambert, V. Brites, M.-P. Gaigeot and I. Compagnon, Distinguishing isobaric phosphated and sulfated carbohydrates by coupling of mass spectrometry with gas phase vibrational spectroscopy, *Phys Chem Chem Phys*, 2014, **16**, 22131–22138.
- 41 M. C. Crowe and J. S. Brodbelt, Differentiation of Phosphorylated and Unphosphorylated Peptides by High-Performance Liquid Chromatography-Electrospray Ionization-Infrared Multiphoton Dissociation in a Quadrupole Ion Trap, *Anal. Chem.*, 2005, **77**, 5726–5734.
- 42 O. Yeni, S. Ollivier, B. Moge, D. Ropartz, H. Rogniaux, L. Legentil, V. Ferrières and I. Compagnon, Ring-Size Memory of Galactose-Containing MS/MS Fragments: Application to the Detection of Galactofuranose in Oligosaccharides and Their Sequencing, *J. Am. Chem. Soc.*, 2023, **145**, 15180–15187.
- 43 O. Yeni, B. Moge, A. Infantino, C. Guitteny, J. Boustie, F. Le Dévéhat, L. Legentil, V. Ferrières and I. Compagnon, LC-MS/MS-IR screening method in glycomics: application to the discrimination of galactofuranose-containing disaccharides, *En Préparation*.
- 44 Q. Enjalbert, R. Simon, A. Salvador, R. Antoine, S. Redon, M. M. Ayhan, F. Darbour, S. Chambert, Y. Bretonnière, P. Dugourd and J. Lemoine, Photo-SRM: laser-induced dissociation improves detection selectivity of selected reaction monitoring mode, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 3375–3381.
- 45 Q. Enjalbert, M. Girod, R. Simon, J. Jeudy, F. Chirot, A. Salvador, R. Antoine, P. Dugourd and J. Lemoine, Improved detection specificity for plasma proteins by targeting cysteine-containing peptides with photo-SRM, *Anal. Bioanal. Chem.*, 2013, **405**, 2321–2331.
- 46 V. C. Cotham, Y. Wine and J. S. Brodbelt, Selective 351 nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows, *Anal. Chem.*, 2013, **85**, 5577–5585.
- 47 M. W. Gardner and J. S. Brodbelt, Ultraviolet Photodissociation Mass Spectrometry of Bis-aryl Hydrazone Conjugated Peptides, *Anal. Chem.*, 2009, **81**, 4864–4872.
- 48 M. W. Gardner, L. A. Vasicek, S. Shabbir, E. V. Anslyn and J. S. Brodbelt, Chromogenic Cross-Linker for the Characterization of Protein Structure by Infrared Multiphoton Dissociation Mass Spectrometry, *Anal. Chem.*, 2008, **80**, 4807–4819.
- 49 Expasy - PeptideMass, [https://web.expasy.org/cgi-bin/peptide\\_mass/peptide-mass.pl](https://web.expasy.org/cgi-bin/peptide_mass/peptide-mass.pl), (accessed September 4, 2023).
- 50 B. Domon and C. E. Costello, A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates, *Glycoconj. J.*, 1988, **5**, 397–409.

## CONCLUSION ET PERSPECTIVES

---

Le projet IROLIGO a pour but de développer des outils capables de caractériser un composé glycoconjugué ainsi que les impuretés éventuellement générées au cours de sa synthèse. Afin de mener à bien ce projet, différentes méthodes ont été utilisées : spectrométrie de masse, chromatographie liquide, spectroscopie IR en phase condensée et en phase gazeuse (spectroscopie IRMPD). Au cours de ce projet, de nombreuses approches originales quant à la spectroscopie IRMPD ont été développées.

Beaucoup de développements ont été effectués depuis les débuts de la spectroscopie IRMPD dans l'équipe DYNAMO. Tout d'abord, les mesures de spectres IRMPD ont permis par comparaison avec des calculs de théorie de connaître la conformation de monosaccharides connus en phase gazeuse. Rapidement, il a été possible de développer de nouvelles approches importantes dans les analyses de glycomique telle que la constitution d'une base de données pour l'identification d'espèces inconnues. Dorénavant, avec les différents développements dont elle a bénéficié, la spectroscopie IRMPD peut s'intégrer dans des workflows analytiques plus complexes. Ces workflows peuvent être suffisamment poussés avec notamment la préparation d'échantillons ainsi que la séparation chromatographique avant analyse par spectroscopie IRMPD, de plus, le retraitement des données a aussi bénéficié de nouveaux développements. La méthode est en constant progrès avec des composés présentant des natures différentes ainsi que des tailles et complexités croissantes. Ces progrès ont été obtenus tant d'un point de vue instrumental (couplage de la spectroscopie IRMPD avec des méthodes séparatives) que d'un point de vue logiciel (automatisation de l'acquisition, du retraitement et de l'interprétation des données...).

Les progrès obtenus au cours de cette thèse (acquisition de spectres IRMPD en ligne avec la chromatographie liquide, identification d'espèces par screening LC-MS(-MS)-IR, identification d'inconnus avec une méthode supervisée : intelligence artificielle ou non supervisée : ACP) sont dans la lignée des efforts déjà effectués dans l'équipe au bénéfice de la spectroscopie IRMPD.

Après avoir présenté la spectroscopie IRMPD, nous avons démontré toute son utilité dans la réponse à des questions de physico-chimie avec l'identification du produit de migration du fucose en phase gazeuse. Nous avons également vu comment par comparaison avec des spectres présents dans une base de données, la spectroscopie IRMPD permet d'identifier un composé inconnu (dans certains cas, la comparaison avec des calculs de chimie théorique n'est pas absolument nécessaire). Au cas où le composé est de trop grande taille, le précurseur peut être fragmenté au sein du spectromètre de masse, les fragments alors obtenus sont à leur tour isolés et irradiés afin d'acquérir leur spectre IRMPD et effectuer l'identification sur ces fragments, ce qui permet de remonter à la structure du composé avant fragmentation.

Ensuite, nous nous sommes penchés sur des développements instrumentaux. Le but est d'améliorer les performances de notre méthode pour l'identification de composés inconnus. Un nouveau montage expérimental est développé contenant un nouveau laser. Le laser décrit dans cette partie permet d'augmenter la vitesse d'acquisition des spectres et s'inscrit donc dans un contexte analytique permettant un plus haut débit d'analyse. Les développements instrumentaux dans cette partie ne s'arrêtent pas là, il a également été possible d'augmenter le rendement de photofragmentation des ions au moyen de l'ajout d'un second laser émettant dans la gamme des 10  $\mu\text{m}$ . Après avoir développé un nouveau montage expérimental, les performances de l'interprétation des données ont été

discutées, tout d'abord, une méthode de calcul de score de ressemblance (ou d'erreur) est décrite, et deux autres méthodes ont été présentées une méthode non supervisée (ACP) et une méthode supervisée d'intelligence artificielle. Bien que ces deux dernières méthodes ont montré leur intérêt dans l'identification de spectres, elles sont relativement nouvelles dans l'équipe et de nouvelles mesures seront nécessaires pour s'assurer de leur robustesse avant de pouvoir s'en servir en routine lors de l'analyse d'autres classes de composés.

L'acquisition de nouveaux spectres IRMPD est toujours intéressante dans notre cas afin de constituer une base de données qui soit la plus exhaustive que possible. Des perspectives existent, par exemple, il serait intéressant de constituer d'autres bases de données propres pour d'autres classes de composés comme les métabolites, les lipides...

Ces démonstrations ont eu lieu sur des petits oligosaccharides (mono, di ou trisaccharides), mais nous sommes également allés plus loin dans la caractérisation de sucres, sur de plus gros oligosaccharides, des polysaccharides et même des glycoconjugués.

Dans une troisième partie, la chromatographie liquide a été couplée à la spectroscopie IRMPD. Les efforts faits dans la seconde partie ont permis de lever les verrous technologiques qui demeuraient et empêchaient de procéder à un tel couplage. L'analyse des produits de digestion du Fondaparinux a pu être faite par ce couplage. L'analyse a été faite de deux manières différentes : tout d'abord en collectant les fractions de manière hors ligne, comme cela a déjà été effectué, mais de manière prometteuse un spectre IRMPD avec une gamme de largeur  $200\text{ cm}^{-1}$  a pu être acquis au moyen de l'acquisition en ligne, où les spectres des composés sont obtenus directement en sortie de colonne chromatographique. Il a également été possible de séquencer la liaison entre deux résidus au sein d'un tétrasaccharide qui diffère de son isomère uniquement par cette liaison glycosidique. Les spectres IRMPD de deux trisaccharides modifiés par la réduction de l'extrémité anomérique du sucre ont été acquis. Un autre mode de LC-MS/MS/IR dans lequel l'irradiation se fait à longueur d'onde fixe a montré tout son potentiel dans la détection de la forme furanose du galactose sur des polysaccharides. Le couplage de la chromatographie liquide avec la spectroscopie IRMPD de manière automatisée et robuste constitue un très grand pas en avant dans les développements de la spectroscopie IRMPD. Cela autorise l'identification d'analytes dans une matrice complexe. Ces développements ne se limitent pas au domaine de la glycomique, mais également à d'autres domaines comme par exemple la métabolomique qui présente également de nombreuses isoméries dans les composés étudiés.

La dernière partie de ce manuscrit traite d'un composé de grande taille et hétérogène : une glycoprotéine. Nous proposons une méthode par spectroscopie IR permettant de détecter la présence de sucres sur une protéine. La position de la glycosylation est plus précisément déterminée par screening LC-MS/IR. Après digestion enzymatique, les peptides sont séparés par chromatographie liquide. Les ions sont alors successivement isolés dans le piège à ions avant d'être irradiés. La longueur d'onde ( $3425\text{ cm}^{-1}$ ) est choisie spécialement pour être absorbée par les modes de vibration des liaisons O-H très présents dans la structure des sucres. En cas de photofragmentation, il s'agit vraisemblablement d'un glycopeptide. Cette méthode permet de procéder à un premier filtrage des composés présents après digestion enzymatique de la protéine. Dans le domaine de la protéomique, la reconnaissance des glycopeptides pourrait être grandement facilitée.

Maintenant d'autres développements sur la spectroscopie IRMPD semblent atteignable. Par exemple d'autres applications pourraient être envisagées dans le domaine de la protéomique, on pourrait chercher à distinguer les PTMs que sont la phosphorylation et la sulfatation sur des peptides issus de la digestion enzymatique de protéines d'intérêt sans avoir recours à la spectrométrie de masse à haute résolution.

Concernant le retraitement et l'interprétation des données nos outils de reconnaissance de spectres sont pour le moment plutôt « bruts ». L'identification d'un composé inconnu se fait en entrant un spectre IRMPD, le programme indique alors un score en se basant sur la ressemblance entre ce spectre IRMPD et les références présentes dans notre base de données (ainsi que sur le rapport m/z du composé). Toutefois, plus d'informations sont disponibles, et ne se bornent pas à un simple spectre IRMPD du composé inconnu. Une nouvelle piste pourrait être d'intégrer ces paramètres dans l'identification du composé afin de conforter cette identification. Il faudrait que le programme en tienne compte afin d'affiner l'identification (spectre MS/MS, temps de rétention...).

# Annexe 1 : Identification et irradiation des peptides issus de la digestion de la feteine bovine.

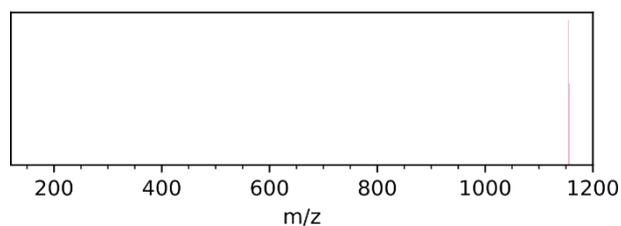
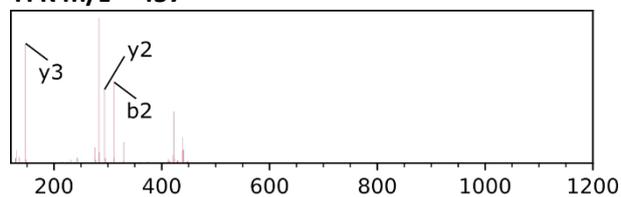
Séquences et rapport m/z des peptides identifiés.

Spectres de fragmentation des peptides pour confirmer leur identification (haut). Tous les fragments ne sont pas annotés dans un souci de clarté. Spectres des peptides sous irradiation infrarouge (bas).

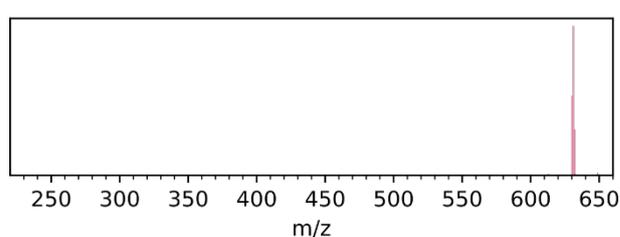
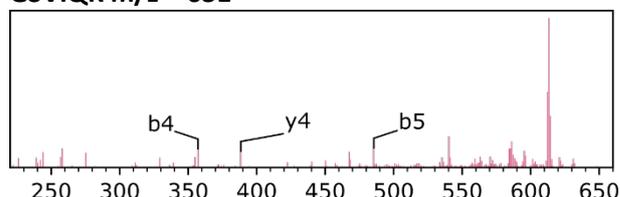
Les mêmes conditions de fragmentation et d'irradiation (durées et puissances) sont utilisées sur les peptides et les glycopeptides (Annexe 1 et Annexe 2).

CID : 30 ms, 35 % et LID : 30 ms, 114 mW à 3425 cm<sup>-1</sup>

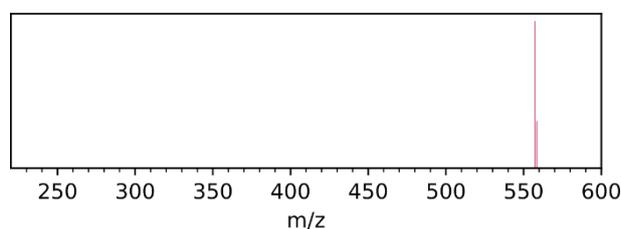
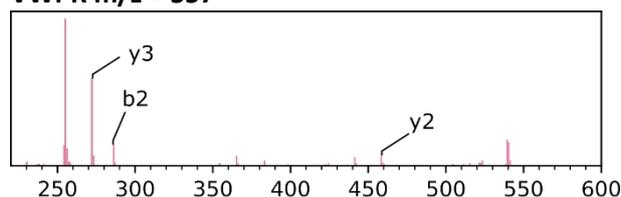
**YFK m/z = 457**



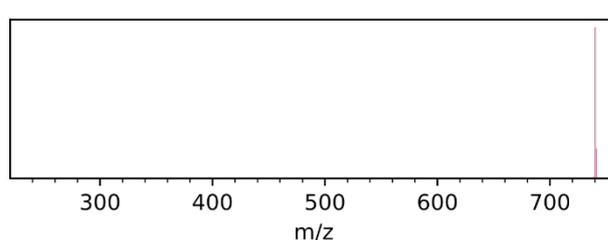
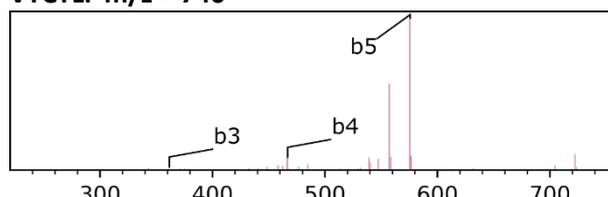
**GSVIQK m/z = 631**



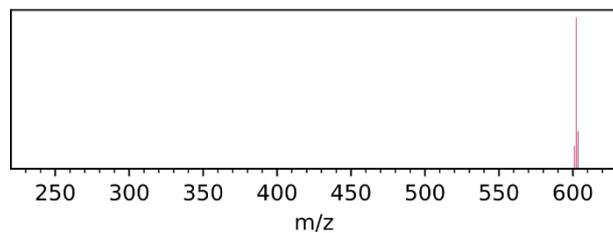
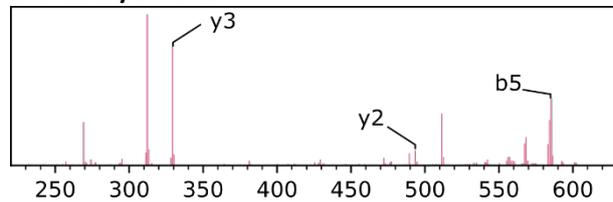
**VWPR m/z = 557**



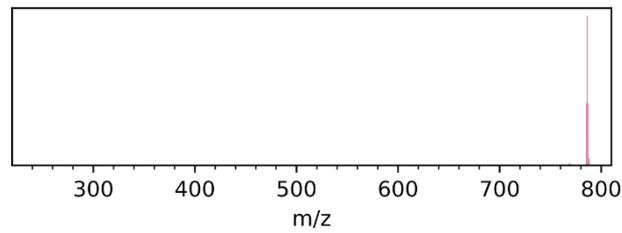
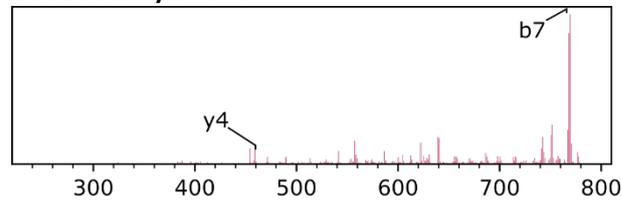
**VTCTLF m/z = 740**



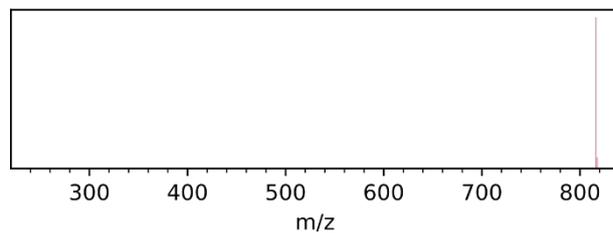
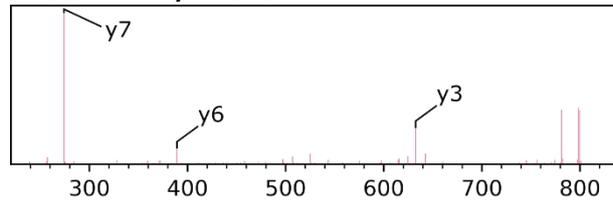
**LCPGR m/z = 602**



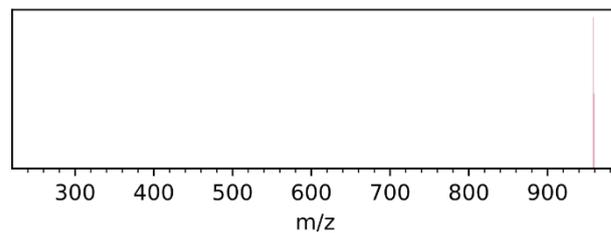
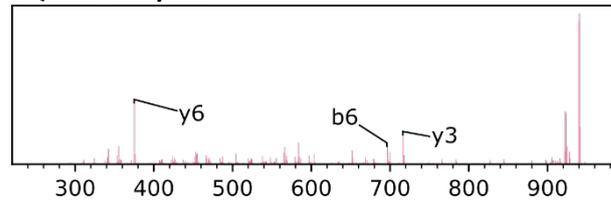
**EVVDPTK m/z = 787**



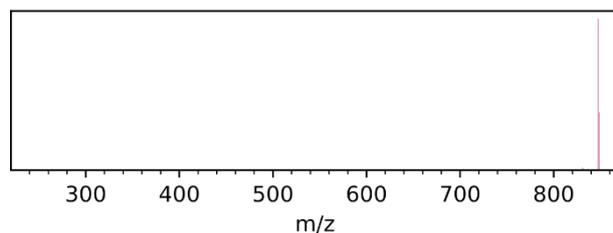
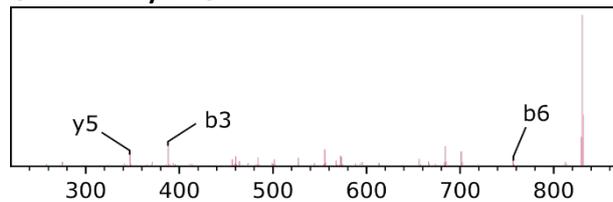
**ALGGEDVR m/z = 816**



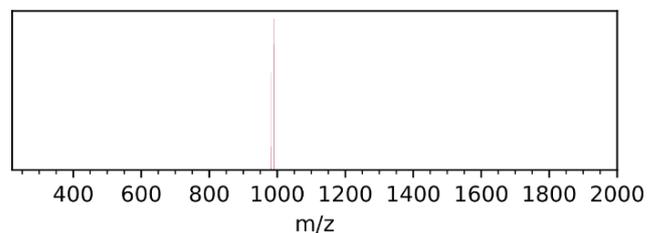
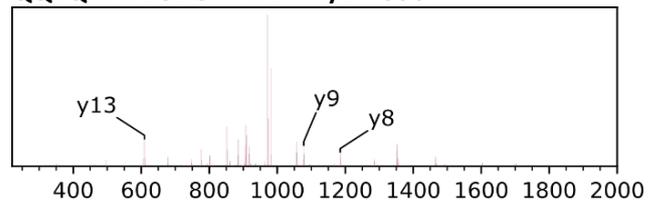
**LQLVEISR m/z = 957**



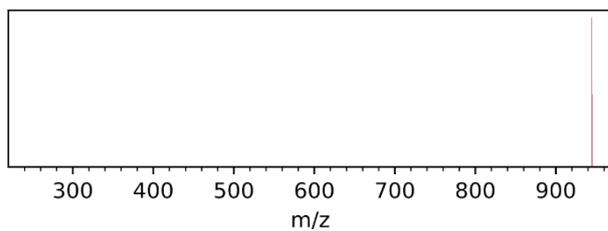
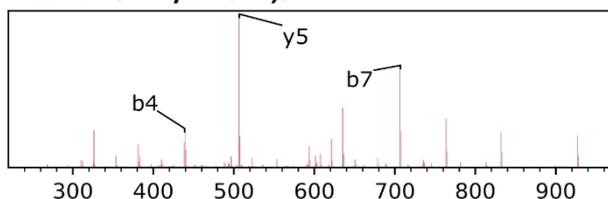
**CNLLAEK m/z = 847**



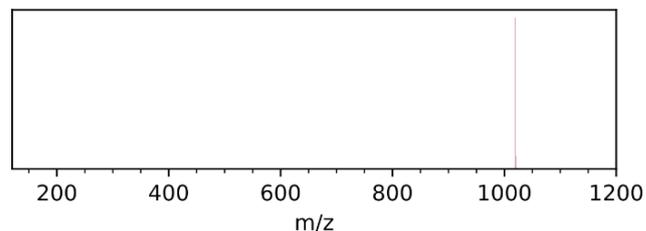
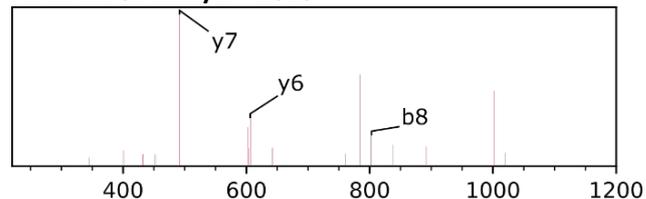
**QQTQHAVEGDCDIHVLK m/z = 990**



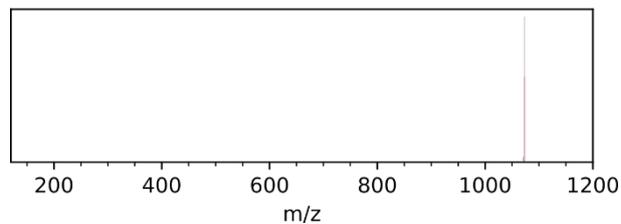
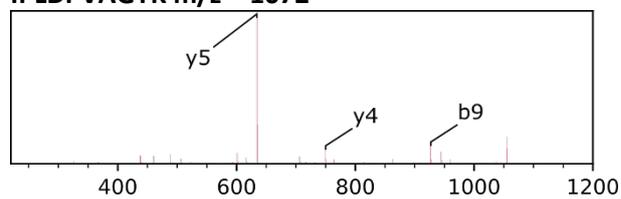
**IPLDPVAGY m/z = 944,5**



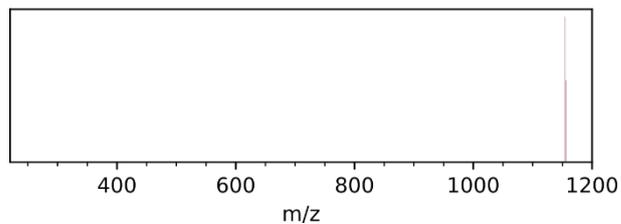
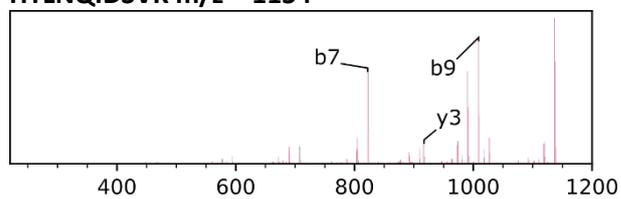
**AVAATDCIAK m/z = 1019**



**IPLDPVAGYK m/z = 1072**



**HTLNQIDSVK m/z = 1154**



## Annexe 2 : Identification et irradiation des supposés glycopeptides issus de la digestion de la fetuine bovine.

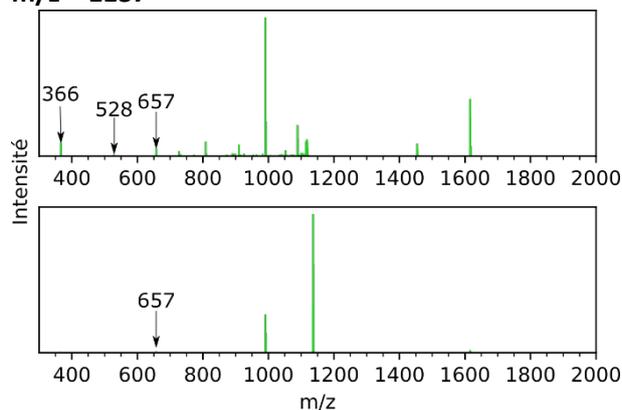
Les composés sont identifiés par leur rapports  $m/z$ .

Les spectres de fragmentation obtenus par CID sont présentés afin de vérifier que les ions « reporter » sont présents (haut). L'ion à  $m/z = 657$  est systématiquement observé. Spectres des composés sous irradiation IR (bas).

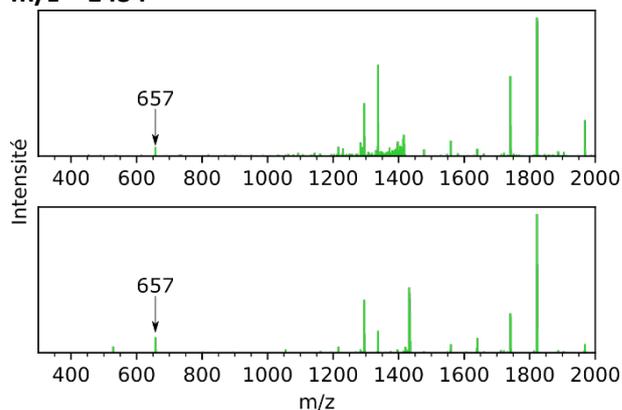
Les mêmes conditions de fragmentation et d'irradiation (durées et puissances) sont utilisées sur les peptides et les glycopeptides (Annexe 1 et Annexe 2).

CID : 30 ms, 35 % et LID : 30 ms, 114 mW à  $3425\text{ cm}^{-1}$

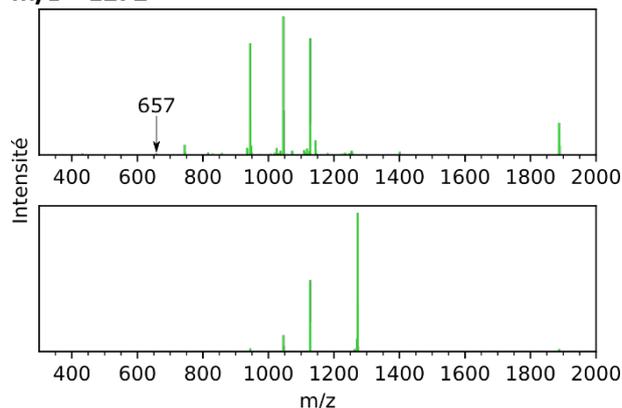
**$m/z = 1137$**



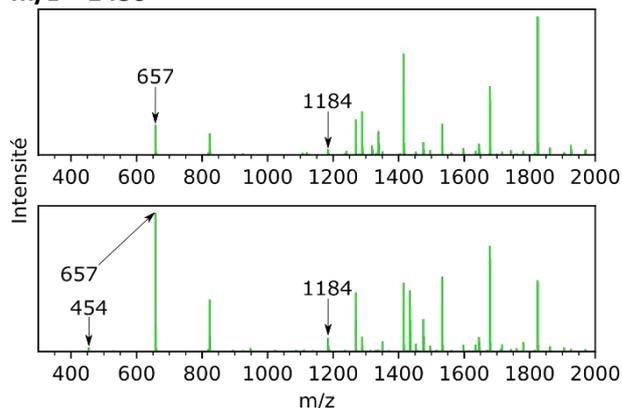
**$m/z = 1434$**



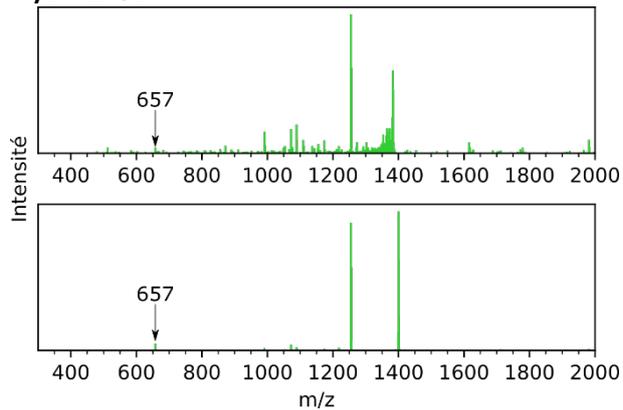
**$m/z = 1272$**



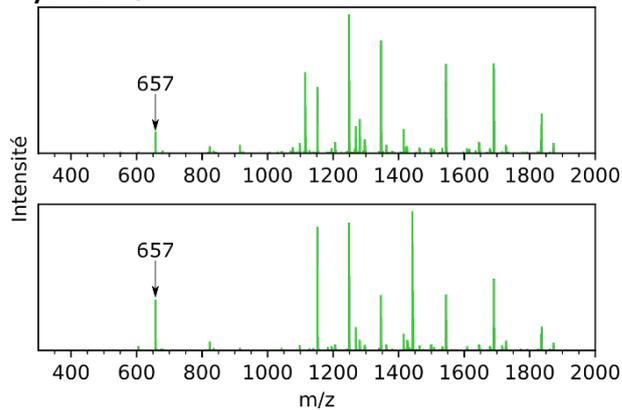
**$m/z = 1436$**



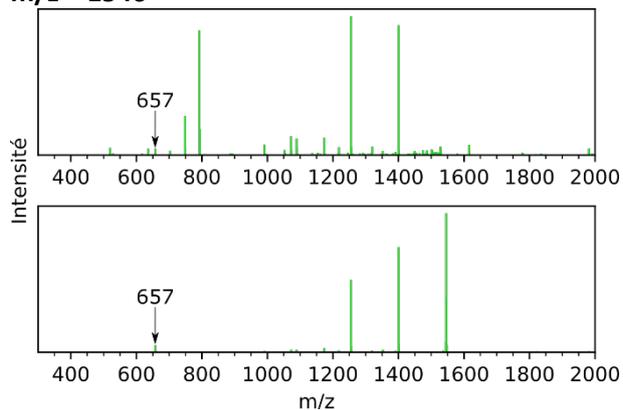
**m/z = 1401**



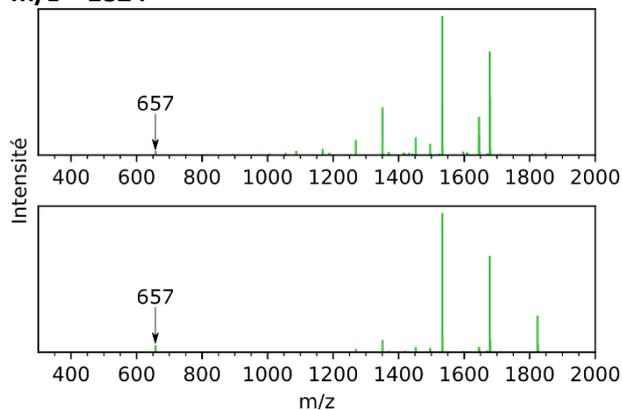
**m/z = 1443**



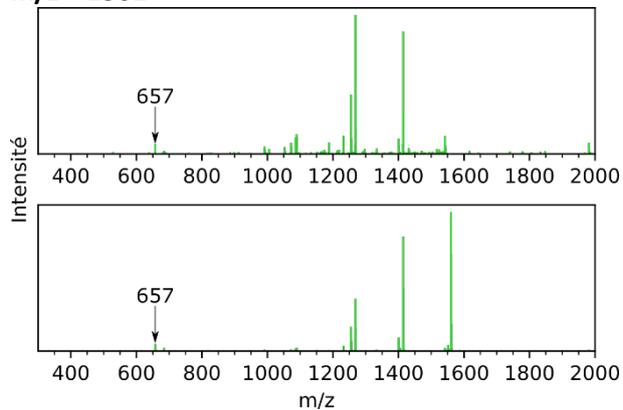
**m/z = 1546**



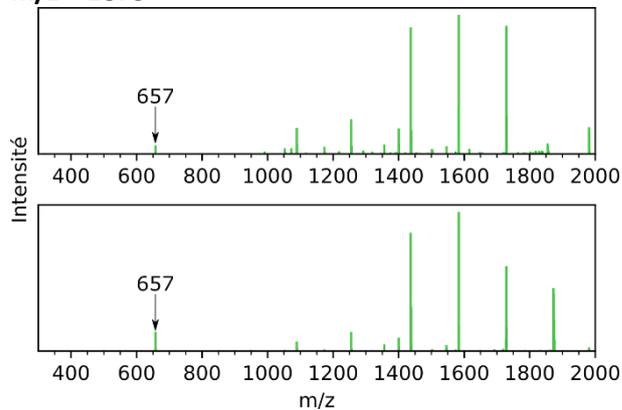
**m/z = 1824**



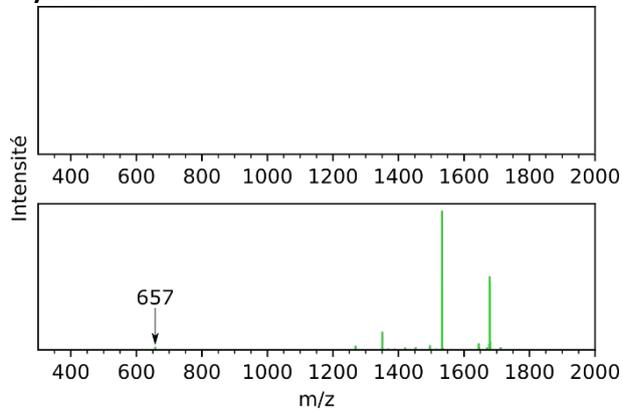
**m/z = 1561**



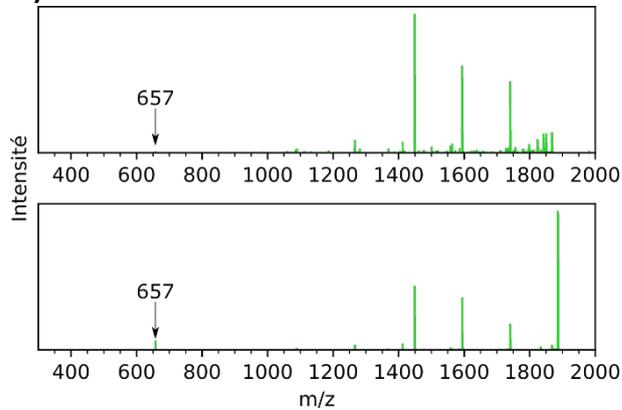
**m/z = 1875**



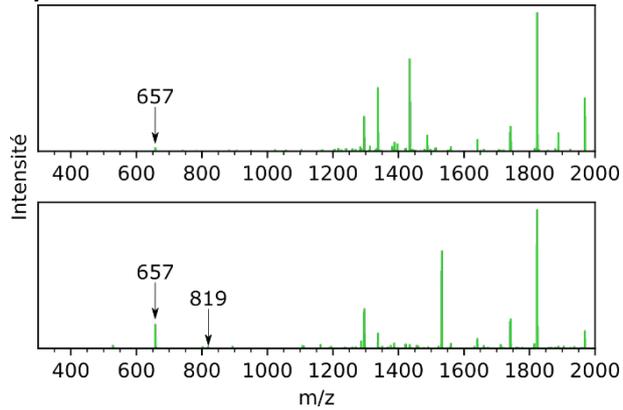
**m/z = 1679**



**m/z = 1887**



**m/z = 1532**



### Notes

Les ions annotés peuvent correspondre aux oligosaccharides suivants :

m/z = 366 disaccharide contenant un hexose et un N-acétylhexosamine (avec perte d'eau)

m/z = 454 disaccharide contenant un hexose et un acide N-Aétylneuraminique

m/z = 657 trisaccharide contenant un hexose un N-acétylhexosamine et un acide N-Acétylneuraminique

m/z = 1184 hexasaccharide contenant trois hexoses deux N-acétylhexosamines et un acide N-Acétylneuraminique

## Liste des publications :

### En préparation :

LCMSMSIR for the screening of galactofuranose in lichen polysaccharides

O. Yeni\*, **B. Moge\***, A. Infantino, C. Guitteny, J. Boustie, F. Le Dévéhat, L. Legentil, V. Ferrières, I. Compagnon

### En revision:

Vibrational and electronic spectra of protonated vanillin: Exploring protonation sites and isomerization

A. Gutiérrez-Quintanilla, **B. Moge**, A. Infantino, I. Compagnon, J. A. Noble

### Publiés:

GLIComics : a deep neural network classifier for Spectroscopy-Augmented Mass Spectrometric glycans data  
T. Barillot, B. Schindler, **B. Moge**, F. Lépine, I. Compagnon, Beilstein J Org Chem. 2023; 19: 1825–1831.

Ring-Size Memory of Galactose-Containing MS/MS Fragments: Application to the Detection of Galactofuranose in Oligosaccharides and Their Sequencing

O. Yeni, S. Ollivier, **B. Moge**, D. Ropartz, H. Rogniaux, L. Legentil, V. Ferrieres, I. Compagnon, Journal of American Chemical Sciences, 145, 2023, 28, 15180–15187

CO<sub>2</sub> laser enhanced rapid IRMPD spectroscopy for glycan analysis

**B. Moge**, O. Yeni, A. Infantino, I. Compagnon, International Journal of Mass Spectrometry, 490, 2023, 117071

Fucose Migration Pathways Identified Using Infrared Spectroscopy

**B. Moge**, B. Schindler, O. Yeni, I. Compagnon, Angewandte Chemie International Edition, 62, 17

Rapid IRMPD (InfraRed multiple photon dissociation) analysis for glycomics

O. Yeni, B. Schindler, **B. Moge**, I. Compagnon, Analyst 147, 312-317

## Présentation orales lors de conférences internationales :

Eurocarb 21, 2023 Paris

Recent developments in gas phase ion spectroscopy for structural analysis of sugars

**B. Moge**, A. Infantino, O. Yeni, I. Compagnon

International Mass Spectrometry Conference, 2022 Maastricht

Structural characterization of antigen-like oligosaccharide structures by gas phase infrared spectroscopy

**B. Moge**, O. Yeni, B. Schindler, I. Compagnon

# Résumé

A l'instar des protéines ou des acides nucléiques, les sucres représentent une des grandes familles de biomolécules indispensables à la vie. A ce titre leur caractérisation structurale est cruciale si l'on veut totalement comprendre les différentes fonctions que ces sucres remplissent. Les sucres sont notamment impliqués dans la reconnaissance cellulaire, c'est par les sucres exprimés à la surface des cellules de l'hôte que les pathogènes reconnaissent la cellule à infecter. Notre partenaire Elicityl souhaite exploiter cette propriété pour développer un nouveau composé : un glycoconjugué qui sera utilisé en tant qu'additif zootechnique afin d'éviter l'infection des animaux d'élevage et ainsi jouer un rôle dans la lutte contre l'antibiorésistance. La connaissance de la structure de ces composés est indispensable, de même que celle des éventuelles impuretés produites lors de la synthèse du glycoconjugué. Afin de mener à bien ce projet, de nombreux verrous devront être levés. Contrairement aux protéines et aux acides nucléiques, nous ne disposons actuellement pas d'une méthode universelle qui permettrait le séquençage des sucres. C'est notamment dû à la complexité intrinsèque de ces structures. En effet de nombreuses isoméries sont présentes dans leur structure. Ce travail de thèse a été l'occasion de développer plus en avant la méthode utilisée dans notre laboratoire pour permettre la caractérisation de tels composés. Cette méthode repose sur la combinaison entre la spectroscopie IR et la spectrométrie de masse. Des développements ont été effectués de manière à répondre aux exigences du projet. Au cours de ce projet elle s'est révélée utile pour la caractérisation de différentes classes de sucres différentes tailles et complexités de structures par exemple sur des monosaccharides ou des petits oligosaccharides ( $\leq 3$  monosaccharides), mais également sur des composés plus complexes tels que des oligosaccharides plus imposants, voire des polysaccharides et apportant même des informations sur des glycoconjugués. Ces développements s'inscrivent dans la lignée des efforts mis en place dans l'équipe afin de faciliter et accélérer l'acquisition de données par cette méthode.

# Abstract

Like proteins or nucleic acids, sugars represent one of the large families of biomolecules essential to life. As such, their structural characterization is crucial if we want to fully understand the different functions that these sugars perform. For example, sugars are notably involved in cellular recognition; it is through the sugars expressed on the surface of the host cells that pathogens recognize the cell to infect. Our partner Elicityl wants to exploit this property to develop a new compound: a glycoconjugate which will be used as a zootechnical additive to prevent infection of farm animals and thus play a role in the fight against antibiotic resistance. Knowledge of the structure of these compounds is essential, as well as any impurities produced during the synthesis of the glycoconjugate. To successfully complete this project, many obstacles need to be solved. Unlike proteins and nucleic acids, we currently do not have a universal method that would allow the sequencing of sugars. This is notably due to the intrinsic complexity of these structures. Indeed, numerous isomers exist in their structure. This thesis work was an opportunity to further develop the method used in our laboratory to allow the characterization of such compounds. This method is based on the combination of IR spectroscopy and mass spectrometry. Developments were made to meet the project requirements. During this project this method proved to be particularly useful for the characterization of different classes of sugars of different sizes and complexities of structures, for example on monosaccharides or small oligosaccharides ( $\leq 3$  monosaccharides), but also on more complex compounds such as larger oligosaccharides, or polysaccharides, and even providing information on glycoconjugates. These developments are in line with the efforts put in place by the team to facilitate and accelerate the acquisition of data by this method.