



**HAL**  
open science

# Caractérisation structurale de HSulf par microscopie électronique, cristallographie et RMN

Yoann Cretinon

► **To cite this version:**

Yoann Cretinon. Caractérisation structurale de HSulf par microscopie électronique, cristallographie et RMN. Biochimie, Biologie Moléculaire. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALV076 . tel-04149845

**HAL Id: tel-04149845**

**<https://theses.hal.science/tel-04149845v1>**

Submitted on 4 Jul 2023

**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.

THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

École doctorale : CSV- Chimie et Sciences du Vivant

Spécialité : Biologie Structurale et Nanobiologie

Unité de recherche : Institut de Biologie Structurale

**Caractérisation structurale de HSulf par microscopie électronique,  
cristallographie et RMN**

**Structural characterization of HSulf by electron microscopy,  
crystallography and NMR**

Présentée par :

**Yoann CRETINON**

Direction de thèse :

**Romain VIVES**

DIRECTEUR DE RECHERCHE, Université Grenoble Alpes

Directeur de thèse

**Elodie CRUBLET**

INGENIEUR DOCTEUR, NMR-Bio

Co-encadrante de  
thèse

Rapporteurs :

**Sylvie RICARD-BLUM**

PROFESSEUR DES UNIVERSITES, Université Lyon 1 - Claude Bernard

**Fabien LECAILLE**

PROFESSEUR DES UNIVERSITES, Université de Tours

Thèse soutenue publiquement le **23 novembre 2022**, devant le jury composé de :

**Romain VIVES**

DIRECTEUR DE RECHERCHE, CNRS délégation Alpes

Directeur de thèse

**Sylvie RICARD-BLUM**

PROFESSEUR DES UNIVERSITES, Université Lyon 1 - Claude  
Bernard

Présidente

**Véronique ROSSI**

MAITRE DE CONFERENCE HDR, Université Grenoble Alpes

Examinatrice

**Fabien LECAILLE**

PROFESSEUR DES UNIVERSITES, Université de Tours

Rapporteur

Invités :

**Cédric LAGURI**

CHARGE DE RECHERCHE, CNRS

**Elodie CRUBLET**

INGENIEUR DOCTEUR, NMR-Bio











## Résumé

Les Héparanes Sulfate se lient à un large éventail de protéines de signalisation, modulant ainsi leur disponibilité, leur stabilité et leur structure. Ces interactions se produisent par l'intermédiaire de motifs de sulfatation spécifiques, présents dans le polysaccharide. La structure de ces polysaccharides peut, toutefois, être régulée par différentes enzymes, dont les sulfatases extracellulaires de la famille Sulf. Les Sulfs catalysent l'élimination sélective des groupements 6-O-sulfatés, participant à la reconnaissance de nombreuses protéines, et sont, de ce fait, impliquées dans de nombreux processus physiopathologiques, comme le cancer. Ces enzymes se composent d'un domaine catalytique (CAT) contenant le site actif et d'un domaine basique hydrophile (HD) responsable de la liaison aux HS. Malgré un intérêt croissant pour cette famille d'enzymes, la structure des Sulfs reste inconnue. Dans ce contexte, ce projet a visé à réaliser une étude structurale de la protéine humaine HSulf-2 en utilisant une approche de biologie structurale intégrative combinant cristallographie, microscopie électronique et RMN.

HSulf-2 étant une protéine de haut poids moléculaire (~ 118 kDa), une stratégie de marquage isotopique spécifique des acides aminés portant un ou plusieurs groupements méthyles (marquage  $^{13}\text{CH}_3$ ) a été utilisée pour étudier HSulf-2 par RMN. A ce jour, il n'existe cependant pas de milieux de culture commercial pour réaliser ce type de marquage isotopique spécifique de protéines exprimées en cellules mammifères, système requis pour produire une protéine HSulf-2 active. Un des premiers objectifs de cette thèse a donc été de développer ce type de milieu et d'en optimiser la composition et les conditions d'utilisation afin de permettre un marquage  $^{13}\text{C}$  spécifique des acides aminés portant des méthyles. Ce milieu a été validé en réalisant le marquage et l'acquisition de spectres 2D d'une protéine modèle : Hsp90 N-terminal. Il a été constaté que le milieu développé entraînait une perte de rendement protéique d'environ 20 à 25 % (comparé au milieu de culture standard des cellules HEK 293), corrélé à une croissance cellulaire réduite.

Ce milieu a, par la suite, été utilisé pour exprimer le variant protéique HSulf-2  $\Delta\text{HD}$  (correspondant au domaine catalytique de l'enzyme), mais les spectres obtenus dans différentes conditions n'ont pas pu être exploités. L'hypothèse retenue est que l'enzyme

HSulf-2 présente une trop grande flexibilité et/ou une trop faible structuration pour générer des spectres RMN exploitables.

En parallèle de ces analyses, des expériences en cryo-microscopie électronique ont permis l'acquisition de deux jeux de données et la production de deux modèles structuraux résolus entre 4 Å et 7 Å. Bien que ces modèles n'autorisent pas un calcul de la structure, ils permettent tout de même d'identifier l'organisation tridimensionnelle d'un certain nombre de structures secondaires, dont certaines pourraient être calculées, et sont représentatifs de la protéine d'intérêt. De plus, grâce à l'ensemble des analyses structurales réalisées et aux nombreux points clés résolus, une nouvelle stratégie en deux étapes a pu être dessinée pour la caractérisation de HSulf-2 par cette méthode.

Enfin, des essais de cristallogénèse ont été réalisés sur trois constructions de la protéine cible, HSulf-2  $\Delta$ SG, HSulf-2  $\Delta$ HD et HSulf-2  $\Delta$ HDc, pour lesquelles la chaîne de GAG présente sur le domaine HD (HSulf-2  $\Delta$ SG), ou des régions non structurées (HSulf-2  $\Delta$ HD et HSulf-2  $\Delta$ HDc), ont été tronquées. Ces tests ont révélé des cristaux non-exploitable ou aucun cristal, mais la qualité prometteuse des échantillons, et les concentrations atteintes, laissent imaginer de nombreux tests supplémentaires et optimisations.

Ces travaux ont fourni des informations sur la structuration de HSulf-2, ont permis d'identifier quelles méthodes d'analyse structurale seront les plus appropriées pour l'étude de cette enzyme, et quels points d'optimisation devront être étudiés pour obtenir une structure haute résolution de l'enzyme. Cette thèse a, par ailleurs, permis de développer une nouvelle stratégie de marquage pour les protéines exprimées en cellules de mammifère HEK 293. Ce milieu nécessite encore quelques optimisations, notamment dans l'incorporation de résidus deutérés, afin d'accroître la résolution des spectres acquis. Il ouvre néanmoins des perspectives importantes pour l'étude par RMN de protéines complexes exprimées en système mammifère.

## Abstract

Heparan Sulfate binds to a wide variety of signaling proteins, modulating their availability, stability and structure. These interactions are carried out through saccharide motifs of specific sulfation patterns, present within the polysaccharide. The structure of the polysaccharide can be regulated by different enzymes, including extracellular sulfatases of the Sulf family. Sulfs catalyze the selective removal of 6-O-sulfat  groups, participating in the recognition of many proteins, and are thereby involved in many pathophysiological processes, such as cancer. These enzymes feature a catalytic domain (CAT) containing the active site, and a basic hydrophilic (HD) domain responsible for binding to HS. Despite growing interest for this family of enzymes, the structure of Sulfs remains unknown. In this context, this project aimed to determine the structure of the human protein HSulf-2, using an integrative structural biology approach combining crystallography, electron microscopy and NMR.

HSulf-2 is a high molecular weight protein (~ 118 kDa); thus, a  $^{13}\text{CH}_3$  methyl specific isotopic labeling strategy was used to study this enzyme by NMR. To date, there is no commercial media for specific isotopic labeling of proteins expressed in mammalian cells, a system required to produce an active HSulf-2 protein. So, one of the first objectives of this thesis was to develop this type of medium and to optimize its composition in order to allow a specific  $^{13}\text{C}$  labeling of methyl amino acids. This medium was validated by the labeling and the acquisition of 2D spectra of a model protein: Hsp90 N-terminal. The new medium was found to cause a loss of protein yield of approximately 20 to 25% (compared to standard culture medium for HEK 293 cells), correlated to a reduced cell growth.

This medium was then used to express the protein variant HSulf-2  $\Delta$ HD (corresponding to the CAT domain of the enzyme), but the resulting spectra obtained in different conditions could not be exploited. The hypothesis proposed is that the enzyme HSulf-2 exhibits too much flexibility and/or too little structuration to generate exploitable NMR spectra.

In parallel to these analyzes, cryo-electron microscopy experiments enabled the acquisition of two data sets and the production of two structural models resolved between 4   and 7  . Although these models do not allow a calculation of the structure,

they make possible to identify the three-dimensional organization of a certain number of secondary structures, some of which could be calculated, and are representative of the protein. In addition, thanks to all the structural analysis and the many key points resolved, a new two-step strategy could be designed for the characterization of HSulf-2 by this method.

Finally, crystallogenes assays were conducted on three constructs of the target protein, HSulf-2  $\Delta$ SG, HSulf-2  $\Delta$ HD and HSulf-2  $\Delta$ HDc, for which the GAG chain present on the HD domain (HSulf-2  $\Delta$ SG), or unstructured regions (HSulf-2  $\Delta$ HD and HSulf-2  $\Delta$ HDc), were truncated. These tests revealed non-exploitable crystals or no crystals, but the quality of the samples and the concentrations reached suggest many additional tests and optimizations.

This work provided informations on the structuration of HSulf-2, made it possible to identify which methods of structural biology will be the most appropriate for the study of this enzyme, and which points of optimization should be studied to obtain a high resolution structure of the enzyme. During this thesis, we also proposed a new labeling strategy for proteins expressed in mammalian cells. Even if this medium still requires some optimizations, in particular for the incorporation of deuterated residues in order to increase the resolution of the acquired spectra. It definitively opens new perspectives for NMR studies of proteins expressed in mammalian expression system.





## Table des matières

Résumé	5.
Abstract	7.
Table des figures	15.
Table des abréviations et acronymes	19.
<i>Introduction</i>	21.
<b>I. Les glycosaminoglycanes, leur structure et leur fonction</b>	23.
I.1. Les Protéoglycanes	23.
I.1.1. Classification et structure des PGs	24.
I.1.2. Classification et structure des GAGs	30.
I.2. Biosynthèse des HS/HP	38.
I.2.1. Initiation	38.
I.2.2. Elongation	40.
I.2.3. Maturation des HS/HP	41.
I.2.3.1. NDSTs	42.
I.2.3.2. Epimérase	43.
I.2.3.3. 2-O-sulfotransférase	45.
I.2.3.4. 3-O-sulfotransférases	45.
I.2.3.5. 6-O-sulfotransférases	47.
I.2.4. Le concept de gagosome	50.
I.3. Processus de régulation post-synthétique des HS	52.
I.3.1. « Shedding »	52.
I.3.2. Héparanase	53.
I.4. Les interactions protéines/HS	54.
I.4.1. Les facteurs de croissance et la 6-O-sulfatation	57.
I.4.2. Les chimiokines, les morphogènes et la 6-O-sulfatation	58.
I.4.3. Les autres ligands	59.



<b>II. Les sulfatases et les Sulfs</b>	61.
II.1. Les sulfatases	61.
II.1.1. Généralités sur les sulfatases	61.
II.1.2. Sulfs : la découverte d'une enzyme de régulation	64.
II.1.2.1. Sulfs : organisation des domaines	65.
II.1.2.2. Sulfs : modifications post-traductionnelles	67.
II.1.2.3. Sulfs : catalyseurs de la 6-O-désulfatation	69.
II.1.2.4. Sulfs : régulateurs de l'interaction HS/Ligand	71.
II.1.2.5. Régulation des Sulfs	73.
II.1.2.6. Sulfs et cancer	74.
II.1.2.7. Sulfs et autres pathologies	77.
II.1.2.8. HSulf-1 et HSulf-2 : récapitulatif des différences	79.
II.2. Caractérisation structurale des sulfatases	81.
II.3. Récents progrès dans la caractérisation structurale et fonctionnelle de HSulf-2	83.
II.3.1. Informations structurales sur HSulf-2	83.
II.3.2. Rôle biologique de la chaîne de CS	86.
<b>III. RMN et marquage isotopique des protéines de haute masse moléculaire en système eucaryote</b>	89.
III.1. Historique de la RMN et du marquage isotopique	90.
III.2. Acides aminés portant des groupements méthyles et leur méthode de marquage isotopique en bactérie	94.
III.2.1. Les voies métaboliques irréversibles	95.
III.2.2. Les voies métaboliques réversibles et intermédiaires	99.
III.2.3. Le marquage combinatoire	100.
III.3. Marquage isotopique en cellules eucaryotes : stratégies de marquage et formulation des milieux	101.
III.3.1. Les différents systèmes d'expression et leurs avantages	101.
III.3.2. Le métabolisme de biosynthèse des acides aminés	105.
III.3.3. Formulation des milieux de marquage	106.

III.3.3.1. Les sources d'acides aminés à considérer pour un marquage isotopique	106.
III.3.3.2. Les stratégies de marquage en cellules eucaryotes supérieures	108.
III.3.3.3. La deutération	113.
III.3.4. Marquage spécifique $^{13}\text{CH}_3$ des acides aminés portant un groupement méthyle en cellules d'insectes et de mammifères	114.
III.3.4.1. Marquage $^{13}\text{CH}_3$ par ajout en excès des acides aminés portant un groupement méthyle	114.
III.3.4.2. Marquage $^{13}\text{CH}_3$ par remplacement des acides aminés portant un groupement méthyle	115.
<b>IV. Objectifs de la thèse</b>	<b>119.</b>
<b>V. Matériels et méthodes</b>	<b>123.</b>
V.1. Expression et purification de la protéine HSulf-2	123.
V.1.1. Expression et purification de la protéine HSulf-2 $\Delta\text{SG}$	125.
V.1.2. Purification de la protéine HSulf-2 $\Delta\text{SG}$ clivée à la trypsine	127.
V.1.3. Test d'activité de la protéine HSulf-2 $\Delta\text{SG}$ et HSulf-2 $\Delta\text{SG}$ trypsinée	128.
V.1.4. Expression, purification de l'enzyme « Peptide N-glycosidase F » (PNGase F) et traitement de la protéine HSulf-2	128.
V.2. Développement du milieu de marquage spécifique des $^{13}\text{CH}_3$ -acides aminés	129.
V.2.1. Composition du mélange réactionnel d'acides aminés	129.
V.2.2. Protocole d'expression et de purification de la protéine Hsp90 N-terminal	136.
V.2.3. Mesure du taux d'incorporation des $^{13}\text{CH}_3$ -acides aminés	140.
V.2.4. Application du milieu de marquage spécifique à la protéine HSulf-2	145.
V.3. Etudes structurales de la protéine HSulf-2 par cryo-EM	146.
<b>VI. Résultats</b>	<b>151.</b>
VI.1. Essais cristallographiques : les stratégies d'optimisation	151.

VI.1.1. Essais cristallographiques : HSulf-2 $\Delta$ SG	151.
VI.1.2. Essais cristallographiques : HSulf-2 $\Delta$ HDc	157.
VI.1.3. Essais cristallographiques : HSulf-2 $\Delta$ HD	160.
VI.2. Développement de nouveaux milieux de culture pour l'étude par RMN de protéines exprimées en cellules HEK 293	164.
VI.2.1. Développement du milieu de marquage spécifique	164.
VI.2.2. Validation des rendements d'expression de la protéine Hsp90 N-terminal dans le milieu de culture développé	169.
VI.2.3. Mesure des taux d'incorporation des $^{13}\text{CH}_3$ -acides aminés	176.
VI.2.4. Développement du milieu de culture uniformément marqué $^{15}\text{N}$	187.
VI.2.5. Application du milieu de marquage spécifique à la protéine HSulf-2 : marquage de HSulf-2 $\Delta$ HD et HSulf-2 $\Delta$ HDc	189.
VI.3. Etudes structurales de la protéine HSulf-2 par cryo-EM	198.
<b>VII. Discussions et perspectives</b>	213.
VII.1. Développement d'un milieu de marquage spécifique et uniforme pour des protéines exprimées en cellules de mammifère	213.
VII.2. Etudes structurales de la protéine HSulf-2	219.
<b>VIII. Annexes</b>	233.
<b>IX. Remerciements</b>	245.
<b>X. Bibliographie</b>	249.

## Table des figures

Figure 1 : Classification des PGs au regard de leur localisation	25.
Figure 2 : Structure et composition des GAGs	31.
Figure 3 : Structure de l'HS	33.
Figure 4 : Biosynthèse de HP/HS	39.
Figure 5 : Gagosome	50.
Figure 6 : Structure 4-méthylumbelliferyl sulfate	62.
Figure 7 : Représentation schématique des Sulfs	66.
Figure 8 : Modèle de l'activité processive de HSulf et de la régulation des propriétés de liaison aux HS	71.
Figure 9 : Interprétation générique du site actif de la sulfatase représentant les résidus hautement conservés retrouvés dans les structures résolues	82.
Figure 10 : PTMs identifiées et modèle structural hypothétique de HSulf-2	85.
Figure 11 : Représentation schématique des régions de résonance des acides aminés portant un groupement méthyle sur un spectre RMN 2D	93.
Figure 12 : Représentation des principales réactions des voies métaboliques d' <i>E. coli</i>	96.
Figure 13 : Composition en acides aminés des extraits de levure et d'algue	110.
Figure 14 : Aperçu des voies métaboliques des acides aminés non essentiels issus de la glycolyse	111.
Figure 15 : Constructions de la protéine HSulf-2	124.
Figure 16 : Purification de HSulf-2 $\Delta$ SG	153.
Figure 17 : Purification de HSulf-2 $\Delta$ SG traitée par la trypsine	155.
Figure 18 : Tests de cristallogénèse de la protéine HSulf-2 $\Delta$ SG tryptsinée en présence de Sulfocalix-[4]-arène	156.
Figure 19 : Purification de HSulf-2 $\Delta$ HDC	158.
Figure 20 : Amas protéique obtenu pour la protéine HSulf-2 $\Delta$ HDC	159.
Figure 21 : Purification de la PNGase F, vérification de son activité et traitement de la protéine HSulf-2 $\Delta$ HD	161.
Figure 22 : Purification de la protéine HSulf-2 $\Delta$ HD	162.
Figure 23 : Deux amas protéiques obtenus pour la protéine HSulf-2 $\Delta$ HD	163.
Figure 24 : Western blot du test d'expression de la protéine Hsp90 N-terminal avec une quantité variable en acides aminés	166.
Figure 25 : Western blot du test d'optimisation de la croissance cellulaire et de	

l'expression de la protéine Hsp90 N-terminal	167.
Figure 26 : Western blot du test d'optimisation de la croissance cellulaire et de l'expression de la protéine Hsp90 N-terminal	169.
Figure 27 : Purification de la protéine Hsp90 N-terminal (non marquée) à partir du milieu de marquage développé	170.
Figure 28 : Purification de la protéine Hsp90 N-terminal (non marquée) à partir du milieu de culture standard (FreeStyle 293)	171.
Figure 29 : Spectre RMN 2D de la protéine Hsp90 N-terminal marquée $^{13}\text{CH}_3\text{-Met}/\text{-Ile}^{\delta 1}/\text{-Val}^{\text{proS}}/\text{-Thr}/\text{-Ala}$	173.
Figure 30 : Superposition des spectres 2D de la protéine Hsp90 N-terminal	175.
Figure 31 : Courbes d'incorporation des résidus $^{13}\text{CH}_3\text{-Ile}^{\delta 1}$ , $\text{Val}^{\text{proS}}$ et Thr	177.
Figure 32 : Comparaison des taux d'incorporation des $^{13}\text{CH}_3\text{-Ala}$ (relatifs au $^{13}\text{CH}_3\text{-Met}$ ) mesurés d'après différentes conditions	178.
Figure 33 : Spectre RMN 2D de l'optimisation de l'incorporation des $^{13}\text{CH}_3\text{-Ala}$	179.
Figure 34 : Spectres ESI-TOF de la protéine Hsp90 N-terminal	183.
Figure 35 : Spectre RMN 2D de la protéine Hsp90 N-terminal	186.
Figure 36 : Marquage uniforme $^{15}\text{N}$ de la protéine Hsp90 N-terminal	188.
Figure 37 : Spectre 2D de la protéine HSulf-2 $\Delta\text{HD}$ marquée	190.
Figure 38 : Spectre 2D de la protéine SNAP-HSulf-2 $\Delta\text{HD}$ trypsinée et marquée	192.
Figure 39 : Spectre 2D de la protéine SNAP-HSulf-2 $\Delta\text{HDc}$ marquée	195.
Figure 40 : Purification et spectre 2D de la protéine HSulf-1 $\Delta\text{HD}$ marquée	197.
Figure 41 : Classification 2D de la protéine HSulf-2 $\Delta\text{SG}$	200.
Figure 42 : Modèle structural de la protéine HSulf-2 $\Delta\text{SG}$	202.
Figure 43 : Densités correspondantes aux N-glycosylations de HSulf-2 $\Delta\text{SG}$	203.
Figure 44 : Classification 2D de la protéine HSulf-2 $\Delta\text{SG}$	204.
Figure 45 : Modèle structural de la protéine HSulf-2 $\Delta\text{SG}$	206.
Figure 46 : Deux approches de réticulation de la protéine HSulf-2 $\Delta\text{SG}$	208.
Figure 47 : Grilles de microscopie électronique en colorations négatives de HSulf-2 $\Delta\text{SG}$ avec différents ratios d'HP	210.
Figure 48 : Classes 2D de la protéine HSulf-2 $\Delta\text{SG}$	224.

—

Tableau 1 : Comparaison des caractéristiques de l'HP et de l'HS	32.
Tableau 2 : Positions et fonctions proposées des acides aminés et des cations métalliques dans le site actif	81.
Tableau 3 : Données de CD de HSulf-2	84.
Tableau 4 : Caractéristiques des plaques de cristallogenèse	127.
Tableau 5 : Composition du mélange réactionnel et comparaison avec des milieux dont la composition en acides aminés est connue	131.
Tableau 6 : Composition des mélanges réactionnels d'acide aminé	132.
Tableau 7 : Optimisation de la concentration en acides aminés	134.
Tableau 8 : Composition du milieu d'expression	136.
Tableau 9 : Composition du milieu de marquage spécifique	138.
Tableau 10 : Composition du milieu de marquage uniforme	139.
Tableau 11 : Composition des milieux de marquage spécifique pour l'établissement des courbes d'incorporation des $^{13}\text{CH}_3$ -Ile, Val et Thr	141.
Tableau 12 : Composition des milieux de marquage spécifique pour le calcul du taux d'incorporation de chaque acide aminé marqué	144.
—	
Annexe 1 : Séquence en acide aminé de la protéine HSulf-2	236.
Annexe 2 : Optimisation de la croissance cellulaire et des rendements protéiques par ajout de vitamines et de glucose	237.
Annexe 3 : Optimisation de la croissance cellulaire et des rendements protéiques par ajout d'extrait de levure	238.
Annexe 4 : Composition des milieux de marquage spécifique pour l'établissement de la courbe d'incorporation de la $^{13}\text{CH}_3$ -Ala	239.
Annexe 5 : Spectres ESI-TOF de la protéine Hsp90 N-terminal marquée sur les Ile et Thr	240.
Annexe 6 : Séquence en acide aminé de la protéine Hsp90 N-terminal et masses ajoutées par le marquage de la protéine Hsp90 N-terminal	241.



## Table des abréviations et acronymes

4MUS : 4-méthylumbelliferyl sulfate	ELISA : « Enzyme Linked ImmunoSorbent Assay »	HAT : Histone acétyl transférase
ADAMs : « A Desintegrin And Metalloprotease »	ERK : « Extracellular signal-regulated kinases »	HB-EGF : « Heparin binding Epidermal Growth Factor »
AHAS : Acéto-hydroxy-acide synthase	EXT : Extostosine	HCC : Hépatocellulaire
AKT : « Sérine/threonine-protein kinase »	Fab : « Fragment antigen-binding »	HDAC : Histone désacétylase
ARS : Arylsulfatase	FACS : « Fluorescence activated cell sorting »	HEK : « Human Embryonic Kidney »
ARSB : N-acétylgalactosamine-4-O-sulfatase	FGF(R) : (Récepteur)	HGF : « Hepatocyte Growth Factor »
AT : Anti-thrombine	Facteurs de croissance des fibroblastes	HH : « Hedgehog »
AvtA : Transaminase Alanine-valine	FGH : Hydrate de FGly	HMW-HA : HA de haut poids moléculaire
BACE : « Beta-siteAPP cleaving enzyme »	FGly : Formylglycine	HP : Héparine
BCAT : Aminotransférase	FIV : Fécondation <i>in vitro</i>	Hpa2 : Héparanase II
BMP4 : Protéine morphogénique osseuse 4	Fz : Frizzled	HRP : Horseradish Peroxidase
CD : Dichroïsme circulaire	G6S : Glucosamine-6-sulfatase	HS : Héparane sulfate
CHO : « Chinese Hamster Ovary »	GAG : Glycosaminoglycane	HSBP : « HS-binding proteins »
CN : Coloration Négative	Gal : Galactose	Hsp90 N-terminal : Heat shock protein 90 N-terminal
CS : Chondroïtine sulfate	GalNAc : N-acétyl galactosamine	HSPE : Héparanase
D <sub>2</sub> O : Deutérium	GDNF : « Glial cell-derived neurotrophic factor »	HSV : Virus de l'herpès simplex
DHAD : Dihydroxyacid déshydrogénase	GlcA : Acide glucuronique	IdoA : Acide iduronique
DS : Dermatane Sulfate	GlcNAc : N-acétyl glucosamine	KARI : Kétol-acid réductoisomérase
EDC : 1-Ethyl-3-[3-diméthylaminopropyl]-carbodiimide hydrochloride	GPCR : Récepteur couplé aux protéines G	KBL : 2-amino-3-kétobutyrate ligase
EGF(R) : (Récepteur)	GPI : Glycosyl-phosphatidyl-inositol	KS : Kératine Sulfate
Facteur de croissance épidermique	<sup>2</sup> H : Proton deutéré	LDS : Lithium dodecyl sulfate
	HA : Hyaluronate	
	HAS : Hyaluronate Synthase	



LMW-HA : HA de bas poids moléculaire  
 LRR : Régions répétées riches en leucine  
 ESI-TOF : Ionisation par électro-nebuliseur - temps de vol  
 MAPK : « Mitogen-activated protein kinases »  
 MBP : « Maltose binding protein »  
 MEC : Matrice Extra-Cellulaire  
 MEF : Fibroblastes embryonnaires  
 MLD : Leucodystrophie métochromatique  
 MMP : Métalloprotéinases matricielles  
 MS : Spectrométrie de masse  
 MSD : « Multiple sulfatase deficiency »  
 NDST : N-désacétyl sulfotransférase  
 NHS : N-hydroxysuccinimide  
 NOE : Effet Overhauser nucléaire  
 OST : O-sulfotransférase  
 PAPS : 3'phosphoadenosine 5'-phosphosulfate  
 PEG : Polyéthylène glycol  
 PEI : Polyéthylèneimine  
 PDGF(R) : « Platelet Derived Growth Factor (Receptor)»  
 PG : Protéoglycanes  
 PI3 : Phosphoinositide-3-kinase  
 PNGase F : « Peptide : N-glycosidase F »  
 PTM : Modification post-traductionnelle  
 RCC : Carcinome rénal  
 RE : Réticulum endoplasmique  
 RMN : Résonance magnétique nucléaire  
 RTK : Récepteurs tyrosine kinase  
 SAIL : « Stereo-Array Isotope Labelling »  
 SAM : Adénosyl-L-méthionine  
 SAXS : Diffusion des rayons X aux petits angles  
 SCCHN : Carcinome épidermoïde  
 SCID : Déficit immunitaire combiné sévère  
 SDS-PAGE : Electrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium  
 SEC-MALS : Chromatographie d'exclusion couplée avec la mesure de la diffusion de la lumière sous multiples angles  
 Sf : *Spodoptera frugiperda*  
 SHH : « Sonic Hedgehog »  
 SLRP : Petits protéoglycanes riches en leucine  
 SOFAST-HMQC : « Selective Optimized Flip Angle Short Transient Heteronuclear » Single Quantum Coherence »  
 SP : Sulfopropyl  
 SST : Sodium Silico Tungstate  
 SUMF1 : « Sulfatase modifying factor 1 »  
 $\tau_c$  : Temps de corrélation  
 TCEP : Tris(2-carboxyethyl)phosphine  
 TD : Thréonine désaminase  
 TDH : Thréonine déshydrogénase  
 TEV : Virus de la gravure du tabac  
 TGFB : « Transforming growth factor beta »  
 VEGF(R) : (Récepteur) Facteur de croissance endothélial vasculaire  
 Wnt : « Wingless type »  
 Xyl : Xylose  
 XylT : Xylosyltransférases  
 YE : Extrait de levure

## **INTRODUCTION**

***I. Les glycosaminoglycanes, leur structure et leur fonction (p. 23)***

***II. Les sulfatases et les Sulfs (p. 61)***

***III. RMN et marquage isotopique des protéines de haute masse moléculaire en système eucaryote (p. 89)***

***IV. Objectifs de la thèse (p. 119)***



## I. Les glycosaminoglycanes, leur structure et leur fonction

### I.1. Les Protéoglycanes

La communication cellulaire est la pierre angulaire de la croissance des organismes pluricellulaires. Cette communication est réalisée par de nombreuses protéines, tels que les facteurs de croissance, les cytokines, les chimiokines. Cependant, afin d'atteindre leurs récepteurs cellulaires, ces protéines doivent diffuser au travers de la matrice extra-cellulaire (MEC) et/ou le glycocalyx, présent à la surface des cellules. Or, ces deux structures extracellulaires sont composées en grande partie par des protéoglycanes (PG). Ces derniers jouent, entre autres, un rôle central dans l'activité et la régulation de la diffusion de la plupart des protéines de signalisation.

Les PGs sont des glycoprotéines ubiquitaires, composés d'un noyau protéique sur lequel est associé, de manière covalente, une ou plusieurs chaînes de glycosaminoglycanes (GAG). Chez les mammifères, les noyaux protéiques sont synthétisés à partir d'une cinquantaine de gènes, y compris des variants résultant d'un épissage alternatif. Notons également que ce noyau protéique est connu pour aider à la localisation du PG, mais aussi dans la transduction du signal au travers de la membrane cellulaire (Alexopoulou *et al.*, 2007). Les GAGs peuvent être de différents types (tels que la chondroïtine sulfate (CS), la dermatane sulfate (DS), la kératane sulfate (KS), l'héparine (HP), l'héparane sulfate (HS) et le hyaluronate (HA, GAG particulier car non associé à une protéine)), et permettent d'induire des fonctions biologiques grâce à leur capacité à interagir avec une large variété de protéines de signalisation. Cela contrôle leur biodisponibilité, l'accès à leurs récepteurs, le stockage et leur protection contre la protéolyse. En revanche, il est important de souligner que les fonctions biologiques peuvent être également déclenchées grâce à l'interaction d'un ligand avec le noyau protéique du PG, et ce dans le but de stabiliser la cohésion cellule-matrice (Whitelock *et al.*, 2008).

Ainsi, les PGs sont impliqués dans de nombreux processus biologiques importants, tels que la prolifération cellulaire, la différenciation, la migration, l'adhésion, la chimio-attraction, l'inflammation, les réponses immunitaires, le contrôle de l'angiogenèse et de la coagulation (Sarrazin *et al.*, 2011 ; Iozzo et Schaefer, 2015). Les PGs peuvent également fournir des sites de fixation pour une variété de micro-organismes

pathogènes, tels que les parasites (*Toxoplasma gondii*...), les bactéries (*Pseudomonas aeruginosa*, *Streptococcus pneumoniae* ...) et les virus (VIH, Zika, HSV, hépatite C ...) (Bartlett et Park, 2010 ; Shao *et al.*, 2012 ; Iozzo et Schaefer, 2015 ; Karamanos *et al.*, 2018).

### 1.1.1. Classification et structure des PGs

Les propriétés fonctionnelles des PGs et leur classification dépendent de leur noyau protéique, de la nature de leurs chaînes de GAGs (HSPGs, CSPGs, DSPGs...), ainsi que de leur localisation (extracellulaires, membranaires ou intracellulaires) (Figure 1).

Les PGs matriciels correspondent, généralement, à des protéines modulaires porteuses de chaînes d'HS ou de CS/DS. Cette famille est divisée en trois groupes.

Le premier groupe correspond aux Hyaléctanes qui contiennent une famille distincte de PGs présentant des similitudes structurelles aux niveaux génomique et protéique. Cette famille est composée de quatre gènes distincts, à savoir l'agrécane, le versicane, le neurocane et le brevicane (Iozzo et Murdoch, 1996). Ces PGs partagent une caractéristique commune qui est leur structure à trois domaines. En effet, ils possèdent des domaines globulaires aux extrémités N- et C-terminale par lesquels ils lient divers composants de la MEC, y compris les HA *via* le domaine N-terminal et les lectines *via* la région C-terminale, formant de grands complexes supramoléculaires (Iozzo, 1998 ; Karamanos *et al.*, 2018). Ces PGs peuvent également dépasser les 1 MDa, et dans le cas des agrécanes peuvent porter jusqu'à une centaine de GAGs (CS et KS) sur la région centrale du PG (Iozzo, 1998). Les hyaléctanes présentent des différences structurales significatives et une distribution spécifique aux tissus. En effet, l'agrécane se trouve principalement dans le cartilage, mais aussi dans les réseaux périneuronaux du cerveau. Le versicane est quant à lui présent principalement dans les MEC interstitielles, et les brevicane et neurocane sont exprimés dans le cerveau et contiennent un noyau protéique portant une à sept chaînes de CS (Iozzo et Schaefer, 2015 ; Theocharis *et al.*, 2017 ; Karamanos *et al.*, 2018).

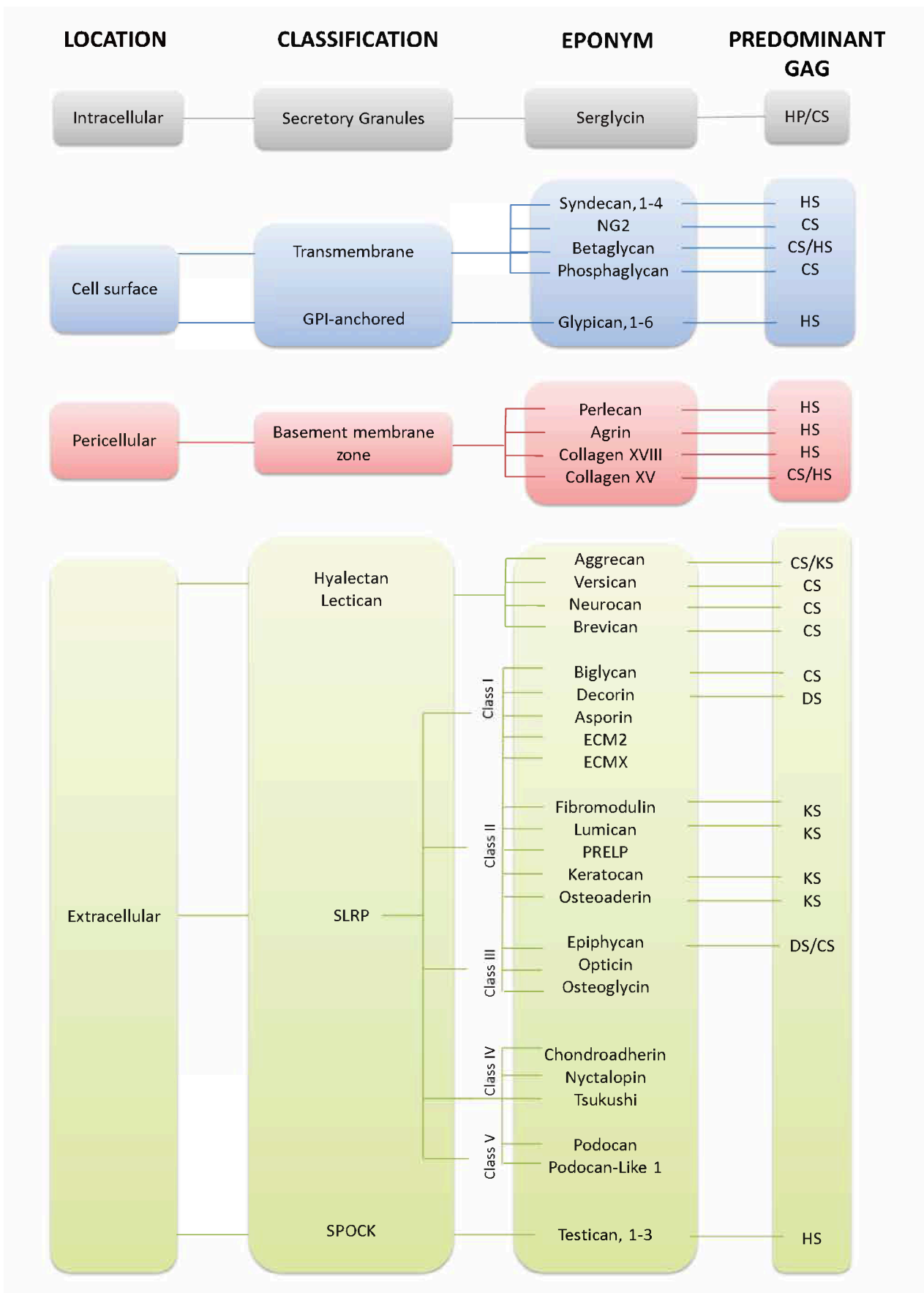


Figure 1 : Classification des PGs au regard de leur localisation (Figure du manuscrit de thèse de Rana El Masri, HAL Id : tel-03507533, version 1).

Le deuxième groupe rassemble les petits protéoglycanes riches en leucine (SLRP) comprenant 18 membres répartis en cinq classes (Neill *et al.*, 2015). Suivant l'appartenance à l'une de ces cinq classes, les SLRP peuvent porter à la fois des chaînes de CS et DS, ou bien seulement des chaînes de CS, DS et KS. Ces PGs sont impliqués dans de nombreuses fonctions biologiques, comme la stabilisation des fibres de collagène (Iozzo, 1998), la croissance tumorale ou le contrôle de l'angiogenèse (Grant *et al.*, 2002). La décorine, appartenant à la classe I des SLRP, est composée de deux régions : une région N-terminale portant une chaîne CS ou DS et une région centrale concave contenant 12 régions répétées riches en leucine (LRR). La partie C-terminale de la protéine semble avoir un rôle critique, étant donné qu'une décorine tronquée (dépourvue des 33 acides aminés C-terminaux) a été identifiée chez des patients atteints de dystrophie cornéenne stromale congénitale (Bredrup *et al.*, 2005). De surcroît, la décorine est impliquée dans l'intégrité structurelle de nombreux organes et régule les processus physiologiques et pathologiques, grâce à son interaction avec de nombreuses protéines de signalisation (Gubbiotti *et al.*, 2016). Notons également que la décorine peut interagir *via* certaines de ses régions centrales LRR spécifiques au collagène de type I et aux récepteurs tyrosine kinase (RTK) comme le récepteur 2 du facteur de croissance endothélial vasculaire (VEGFR2) et le récepteur du facteur de croissance épidermique (EGFR). Ces interactions entraînent l'activation de plusieurs voies de signalisation induisant une autophagie dans les cellules endothéliales, épithéliales et gliomales, et la mitophagie dans les cellules cancéreuses (Buraschi *et al.*, 2019).

Enfin, le troisième groupe de PGs matriciels correspond à la famille SPOCK. Elle comprend trois HSPGs qui se lient au calcium et sont appelés Testicans (Iozzo et Schaefer, 2015).

Les HSPGs (perlécane, agrine, collagènes XVIII et XV) peuvent également être associés à la matrice péricellulaire *via* les intégrines ou d'autres récepteurs, jouant ainsi un rôle important dans le maintien de la membrane basale en interagissant les uns avec les autres, ainsi qu'avec d'autres constituants de cette membrane basale, tels que les laminines, le collagène de type IV et le nidogène (Iozzo et Schaefer, 2015). Le perlécane et l'agrine contiennent un assortiment de modules protéiques et portent jusqu'à trois chaînes de HS/CS (perlécane) ou jusqu'à trois HS (agrine). Le collagène de type XVIII

possède jusqu'à trois chaînes de HS, tandis que le collagène de type XV porte des chaînes de CS (Karamanos *et al.*, 2018).

Le perlécane est localisé au niveau de la membrane vasculaire basale et des espaces péricellulaires (Arikawa-Hirasawa *et al.*, 2022). Cet HSPG a une architecture modulaire qui facilite une multitude de processus de régulation, incluant l'adhésion cellulaire, la formation de cartilage articulaire, l'ossification, l'arthrose, l'endocytose, l'assemblage ganglionnaire périphérique, le développement cardiovasculaire, entre autres processus physiologiques et pathologiques (Karamanos *et al.*, 2018). Deux maladies humaines liées au perlecan ont été signalées. L'une d'elles est la dysplasie dyssegmentaire de type Silverman-Handmaker, résultant de la perte complète de la fonction du gène *HSPG2* qui code pour la protéine (perlecan). L'autre est le syndrome de Schwartz-Jampel résultant de la perte partielle de la fonction du gène *HSPG2*. Le perlécane est également un intervenant fondamental de l'angiogenèse développementale et cancéreuse, où il peut exercer une double fonction en tant que facteur pro-angiogénique et anti-angiogénique (Gubbiotti *et al.*, 2017 ; Pozzi *et al.*, 2017 ; Theocharis *et al.*, 2017), ainsi que dans la régulation autophagique (Poluzzi *et al.*, 2014 ; Ning *et al.*, 2015). De ce fait, le perlécane s'avère être une cible thérapeutique d'intérêt (Arikawa-Hirasawa *et al.*, 2022). L'agrine, quant à elle, joue un rôle important au niveau des jonctions neuromusculaires en permettant le rassemblement des récepteurs à l'acétylcholine et la formation de l'appareil post-synaptique (Bezakova et Ruegg, 2003).

Les PGs membranaires sont des HSPGs regroupés dans deux grandes familles : les syndécans et les glypicans. Les syndécans sont des protéoglycanes transmembranaires *via* leur noyau protéique. Les glypicans sont associés à la membrane *via* une ancre glycosyl-phosphatidyl-inositol (GPI).

Les quatre syndécans de mammifères sont des PGs membranaires qui régulent de nombreux processus biologiques. Ils comprennent un ectodomaine allant de 20 à 40 kDa, un domaine transmembranaire et un domaine cytosolique composé d'une région variable et de deux régions constantes. Les ectodomains ne partagent entre eux que 10 à 20 % d'homologie de séquence. En revanche, les deux autres domaines sont bien conservés entre les syndécans (Iozzo et Schaefer, 2015). Toutefois, malgré cette faible homologie des ectodomains, certaines séquences restent très conservées. C'est le cas des



séquences impliquées dans l'attachement des chaînes de GAGs, qui sont liées de manière covalente à leur noyau protéique par des motifs conservés de sérine-glycine. Ainsi, les quatre syndécans portent des chaînes d'HS à leur extrémité N-terminale, et pour les syndécans-1 et 3, des chaînes de CS dans la région juxtamembranaire (Couchman *et al.*, 2015 ; Afratis *et al.*, 2017a).

Comme introduit précédemment, les syndécans interviennent dans de nombreux processus biologiques *via* un large réseau d'interaction qui comprend 347 partenaires (Gondelaud *et al.*, 2019). Ce large réseau d'interaction permet d'impliquer ces syndécans, entre autres, dans la signalisation cellulaire (Couchman *et al.*, 2015 ; Afratis *et al.*, 2017a), l'angiogenèse (De Rossi *et al.*, 2014), le recrutement et la présentation des facteurs de croissance à leurs récepteurs (Kwon *et al.*, 2012) et l'assemblage de la matrice extracellulaire (Yang *et al.*, 2016). Par ailleurs, leur rôle a été souligné dans plusieurs processus pathologiques comme la tumorigenèse (Afratis *et al.*, 2017a), les interactions hôte-pathogène (Bartlett et Park, 2010), les maladies neurodégénératives (Hudak *et al.*, 2019 ; Letoha *et al.*, 2019) et l'inflammation (Gopal, 2020).

Malgré la variété et l'importance de leurs rôles biologiques, très peu d'informations structurales sont disponibles pour les syndécans (Gondelaud *et al.*, 2019) ; ce qui est un frein dans l'étude de leurs mécanismes d'action moléculaire. En effet, bien que les structures du domaine transmembranaire des syndécans-2 et -4 et du domaine cytoplasmique des syndécans-1 et -4 (seuls ou en complexe avec d'autres protéines) aient été résolues (Gondelaud *et al.*, 2019), aucune donnée structurale n'est disponible pour les ectodomaines des syndécans. Il est vrai que ces domaines se comportent comme des protéines intrinsèquement désordonnées et adoptent plusieurs ensembles de conformation en solution (Leonova *et al.*, 2015 ; Gondelaud *et al.*, 2021). Or, sans ces informations, la compréhension des processus biologiques dont les ectodomaines sont à l'origine est limitée, notamment l'action des ectodomaines solubles libérés par des protéases : des métalloprotéinases matricielles (MMP) et des membres de la famille des « A Disintegrin And Metalloprotease » (ADAMs) dans le milieu extracellulaire, où ils peuvent agir comme effecteurs paracrines ou autocrines et inhibiteurs compétitifs des PGs de la surface cellulaire (Asundi *et al.*, 2003 ; Manon-Jensen *et al.*, 2010). En effet, les ectodomaines jouent un rôle dans la cicatrisation des plaies et modulent le comportement des cellules cancéreuses et leur micro-environnement (Couchman *et al.*, 2015). Par exemple, l'ectodomaine S2 soluble inhibe l'angiogenèse (De Rossi *et al.*, 2014)

et améliore les activités tumorigènes des cellules cancéreuses (Choi *et al.*, 2015), tandis que l'ectodomaine S4 soluble favorise le recrutement des cellules immunitaires, le remodelage de la matrice extracellulaire et atténue les dysfonctionnements cardiaques en réponse aux lipoglycans (Strand *et al.*, 2015).

Néanmoins, une étude récente a pu caractériser leurs paysages conformationnels et démontré que les ectodomaines des quatre syndécans humains acquièrent différentes conformations en fonction des contextes moléculaire et cellulaire (des conformations de type globulaire à des conformations étendues) qui seraient ou non biologiquement actives. En effet, les séquences des syndicats impliquées dans la formation de complexes ternaires avec les intégrines ne sont pas accessibles dans toutes les conformations. En revanche, en conformation étendue, les ectodomaines des syndicats ont une taille relativement similaire à celle des récepteurs de surface cellulaire auxquels ils se lient comme les intégrines (Gondelaud *et al.*, 2021).

La deuxième famille, les Glypicans, est composée de six membres très conservés (plus de 90 % d'homologie entre les espèces) et comportent tous un noyau protéique entièrement extracellulaire relié à la membrane plasmique par un ancrage de type GPI. Les glypicans sont des HSPGs stricts (Gallagher *et al.*, 1996 ; De Cat *et al.*, 2001), à l'exception des Glypican-5 également porteur de chaînes de CS/DS (Saunders *et al.*, 1997). L'action de protéases (« furin-like convertase ») ou bien de lipases (Notum, pouvant cliver l'ancre GPI) peuvent conduire, respectivement, à la libération de l'ectodomaine ou à la libération du noyau protéique (Filmus et Capurro, 2014). Ainsi, ces HSPG sont impliqués dans de nombreuses voies de signalisation, tels que « Hedgehog » (HH), « Wnt » (Wnt) et les facteurs de croissance des fibroblastes (FGF) (Capurro *et al.*, 2005, 2008).

La serglycine est un PG intracellulaire important, avec un rôle établi dans l'adressage des protéines dans les granules sécrétoires et/ou la direction de la sécrétion de ces molécules vers la matrice extracellulaire.

La serglycine est principalement exprimée par les cellules d'origine hématopoïétique (Kolset et Pejler, 2011), mais son expression a également été confirmée dans les cellules endothéliales (Schick *et al.*, 2001 ; Meen *et al.*, 2011), les cellules acineuses pancréatiques (Biederbick *et al.*, 2003) et les cellules cancéreuses nasopharyngées agressives (Li *et al.*, 2011). Fait intéressant, bien que la serglycine soit considérée

comme un PG intracellulaire, elle est également sécrétée dans la matrice extracellulaire par les cellules tumorales affectant les propriétés cellulaires, la signalisation oncogénique et la charge des exosomes (Manou *et al.*, 2020).

La serglycine a un noyau protéique court avec un domaine caractéristique riche en répétitions sérine/glycine sur lequel jusqu'à huit chaînes de GAGs peuvent être attachées (Kolset *et al.*, 2008). Ce PG peut, ainsi, porter soit des chaînes de CS, de DS, d'HS et d'HP selon le type de cellules (Kolset *et al.*, 2004 ; Kolset et Pejler, 2011). Fait intéressant, la serglycine peut également exister sous forme de PG hybride portant simultanément différents types de GAGs (Lidholt *et al.*, 1995).

Lorsqu'elle est exprimée dans les tissus conjonctifs, la serglycine est riche en HP hautement sulfatés (Yurt *et al.*, 1977). Les chaînes d'HP portées par la serglycine dans les granules de sécrétion des mastocytes sont en complexe avec des protéases qui peuvent être libérées lors de réactions inflammatoires (Kolset *et al.*, 2008 ; Pejler *et al.*, 2009). La serglycine est également présente dans d'autres granules de cellules inflammatoires, telles que les lymphocytes et les monocytes, entre autres, mais elles sont associées dans ce cas à des chaînes de CS (Kolset et Gallagher, 1990). Ce PG est également présent dans d'autres cellules, telles que les macrophages, les neutrophiles, les lymphocytes T cytotoxiques et les plaquettes (Kolset *et al.*, 2004). Enfin, comme souligné précédemment, la serglycine est également présente dans les cellules du myélome. Dans ces cellules, la serglycine favorise l'adhésion des cellules myélomateuses au collagène de type I et est impliquée dans la progression du myélome *via* l'inhibition de la minéralisation osseuse (Theocharis *et al.*, 2006 ; Skliris *et al.*, 2013).

### **1.1.2. Classification et structure des GAGs**

Les GAGs sont des polysaccharides complexes, linéaires et hétérogènes. Ils sont tous constitués d'une répétition d'unités disaccharides comprenant un acide uronique et une hexosamine ; à l'exception des KS où l'acide uronique est remplacé par du galactose (Gal). L'acide uronique peut se présenter sous deux formes : l'acide iduronique (IdoA) ou bien l'acide glucuronique (GlcA). Pour l'hexosamine, il peut s'agir soit de N-acétyl glucosamine (GlcNAc), soit de N-acétyl galactosamine (GalNAc) (Figure 2).

La diversité structurale des GAGs est renforcée par différents degrés de modification des sous-unités disaccharidiques. En effet, les groupes hydroxyles en position C2 des acides uroniques et en position C3, C4 et C6 de l'hexosamine peuvent être O-sulfatés, et les glucosamines peuvent être présentes sous forme N-acétylée ou N-sulfatée (ou plus rarement N-non substituée). Ainsi, la nature des sucres et les types de liaisons glycosidiques permettent de définir les différents membres de la famille des GAGs.

Glycosaminoglycanes	Unités disaccharidiques et modifications	Nombre de répétitions disaccharidiques et masse moléculaire
Acide Hyaluronique	<p>GlcA-β(1→3) – GlcNAc-β(1→4)</p>	<p>n = 8 - 16000 MW = 4 - 8000 kDa</p>
Kératane Sulfate	<p>Gal-β(1→4) – GalNAc-β(1→3)</p>	<p>n = 9 - 45 MW = 4 - 20 kDa</p>
Chondroïtine Sulfate Dermatane Sulfate	<p>GlcA-β(1→3) – GalNAc-β(1→4) IdoA-α(1→3) – GalNAc-β(1→4)</p>	<p>n = 10 - 130 MW = 5 - 60 kDa</p>
Héparine Héparane sulfate	<p>GlcA-β(1→4) – GlcNAc-α(1→4) IdoA-α(1→4) – GlcNAc-α(1→4)</p>	<p>n = 20 - 100 MW = 10 - 50 kDa</p>

Figure 2 : Structure et composition des GAGs (Figure adaptée du manuscrit de thèse de Rana El Masri, HAL Id : tel-03507533, version 1).

## Héparane sulfate et Héparine

L'HS est naturellement présent dans toutes les cellules et varie énormément en termes de degré de sulfatation et de longueur de chaîne selon son origine biologique. Les chaînes d'HS sont généralement constituées de 50 à 250 unités disaccharidiques (20 à 100 kDa). En comparaison, l'HP se compose généralement de chaînes plus courtes et contient des domaines sulfatés (domaine S) plus grands et portant davantage de modification, ce qui donne lieu à de grandes portions de chaînes contenant des disaccharides entièrement sulfatés (trisulfatés) et modifiés (IdoA) (Tableau 1) (Sarrazin *et al.*, 2011 ; Xu et Esko, 2014).

Characteristics	Heparan sulfate	Heparin
Site of synthesis	Virtually all cells	Connective tissue-type mast cells
Core protein	Many (~17)	Serglycin
Cell membrane attached?	Yes	No
Size	20–100 kDa	7–20 kDa
Alternate NS/NAc domains	Yes	Minimal
Average sulfate/disaccharide	0.6–1.8	1.8–2.6
IdoA	20–50%	≥80%
N-Sulfate	30–60%	≥80%
2-O-Sulfate	10–40%	≥80%
6-O-Sulfate	10–40%	≥80%
Binding to antithrombin	0–0.3%	~30%
Commercial availability	Milligrams	Kilograms

Tableau 1 : Comparaison des caractéristiques de l'HP et de l'HS (Figure : Xu et Esko, 2014).

L'unité disaccharidique de base des HS est constituée d'une répétition de GlcA (liaison  $\beta$  1->4) et de GlcNAc (liaison  $\alpha$  1->4). Au cours du processus de biosynthèse, les chaînes peuvent subir une série de modifications durant lesquelles les disaccharides subissent une dé-acétylation et une N-sulfatation des résidus glucosamines, une épimérisation du GlcA en IdoA, une O-sulfatation en position C6 (plus rarement en C3) de la glucosamine et en position C2 du résidu IdoA. Cependant, ces réactions de modification ne sont pas totales formant ainsi, des domaines S ( $\pm$  sulfatés et modifiés) entrecoupés de domaines NS dépourvus de tout ou partie de ces modifications, et des domaines de transitions

intermédiaires NS/S entre ces deux types de domaines (Figure 3) (Maccarana *et al.*, 1996 ; Murphy *et al.*, 2004). Fait intéressant, bien que la plupart des cellules animales produisent des HS, la taille, la composition et la distribution des domaines S varient considérablement car l'assemblage de HS n'est pas codé génétiquement (Xu et Esko, 2014). Ainsi, il est supposé que l'organisation et la spécificité des enzymes de biosynthèse, la disponibilité des précurseurs et le flux à travers l'appareil de Golgi déterminent la longueur de la chaîne, le degré de sulfatation et l'épimérisation ainsi que la taille et l'espacement des domaines sulfatés (Esko *et al.*, 2002 ; Carlsson *et al.*, 2012 ; Kreuger *et al.*, 2012).



Figure 3 : Structure de l'HS. Les rectangles rouges représentent les domaines S, les traits noirs les domaines NS et les rectangles roses les domaines de transition NS/S.

Des centaines de protéines ont développé la capacité d'interagir avec les HS, parfois avec une grande spécificité. Le réseau de « HS-binding proteins » (HSBP), appelé interactome héparane-sulfaté (Ori *et al.*, 2011), comprend des protéines impliquées dans l'attachement cellulaire, la migration, l'invasion et la différenciation, la morphogenèse, l'organogenèse, la coagulation sanguine, le métabolisme des lipides, l'inflammation et les infections par différents agents pathogènes (Bishop *et al.*, 2007 ; Ori *et al.*, 2011 ; Afratis *et al.*, 2017b ; Weiss *et al.*, 2017 ; Lanzi *et al.*, 2018 ; Chhabra *et al.*, 2021). Tous ces processus impliquent une interaction physique des HSBP aux séquences saccharidiques spécifiques présentes au sein des chaînes d'HS.

Plusieurs cartographies du réseau d'interaction des HP/HS (mais aussi des autres GAGs) avec des partenaires protéiques ont été produites au cours de ces dernières années. Ainsi, en 2011 et 2018, deux réseaux d'interaction HP/HS ont été construits, permettant de caractériser 435 et 262 protéines (de la MEC, des intégrines, des cytokines, des facteurs de croissance ...) se liant aux HP/HS (Ori *et al.*, 2011 ; Karamanos *et al.*, 2018). En 2021, le premier réseau de l'interactome des six GAGs de mammifères a été

construit. Il comprend 832 biomolécules (827 protéines et 5 GAGs) et 932 interactions protéine/GAG (Vallet *et al.*, 2021). Il a été déterminé 580 partenaires pour les HP/HS, 218 et 72 partenaires pour les KS et CS, respectivement, et 43 et 19 partenaires pour les HA et DS, respectivement. Ce réseau a été, toutefois, récemment mis à jour et élargi à 4290 interactions dont 3464 interactions uniques protéine/GAG (Vallet *et al.*, 2022), faisant apparaître l'importance des GAGs dans les voies métaboliques de nombreuses protéines et laissant imaginer de nouvelles voies de régulations.

### *Acide hyaluronique*

La structure des HA est la plus simple des GAGs : c'est une répétition d'unités GlcA-GlcNAc. Cependant, les HA se démarquent des autres GAGs car ils ne sont pas associés à un noyau protéique et ne sont pas sulfatés. De surcroît, la polymérisation des HA n'est pas réalisée, à l'instar d'autres GAGs, dans la lumière de l'appareil de Golgi, mais elle se produit sur la face interne de la membrane plasmique par trois différentes synthèses HA (HAS) (Weigel *et al.*, 1997).

Suite à une lésion tissulaire ou à une tumorigenèse, l'élévation de l'activité de l'hyaluronidase peut provoquer une accumulation excessive d'HA de bas poids moléculaire (LMW-HA, 10-500 kDa), dérivant de la dégradation d'HA de haut poids moléculaire (HMW-HA, >500 kDa) (Jiang *et al.*, 2007). Or, ces différents HA sont à l'origine de fonctions biologiques bien distinctes (Wolny *et al.*, 2010 ; Yang *et al.*, 2012). Par exemple, il a été démontré que les LMW-HA induisaient la prolifération des astrocytes, une étape majeure dans la réactivation gliale (Struve *et al.*, 2005). Tandis que les HMW-HA supprimaient la formation de cicatrices gliales (Khaing *et al.*, 2011), ainsi que diminuaient la production d'IL-1 $\beta$ , IL-6, TNF $\alpha$  et d'oxyde nitrique dans les cellules microgliales exposées au lipopolysaccharide (Austin *et al.*, 2012).

Egalement notons, des concentrations élevées de LMW-HA ont été détectées dans l'urine de patients atteints d'un cancer de la vessie (Lokeshwar *et al.*, 1997) et dans la salive de patients atteints d'un carcinome épidermoïde de la tête et du cou à un stade élevé (Franzmann *et al.*, 2003). Plus important encore, une étude a examiné les niveaux de fragment de hyaluronane, allant de 6 à 25 disaccharides, dans le liquide interstitiel des tumeurs colorectales, et a constaté que des niveaux élevés de LMW-HA étaient associés

à l'invasion des vaisseaux lymphatiques par ces cellules tumorales et aux métastases des ganglions lymphatiques (Schmaus *et al.*, 2014).

### *Kératane Sulfate*

Le KS est un type de GAG largement présent dans la MEC de certains tissus, tels que la cornée, les cartilages et les os. Le KS est composé d'un squelette poly-N-acétyl lactosamine sulfaté. Sa structure est formée par l'alternance d'unités  $\beta$ -galactose (liaisons :  $\beta 1 \rightarrow 3$ ) et de N-acétyl- $\gamma$ -glucosamine (liaison :  $\beta 1 \rightarrow 4$ ), et organisée en blocs disaccharidiques répétés au sein d'une chaîne polysaccharidique (Funderburgh, 2002). Bien que les deux unités puissent être 6-O-sulfatées, cette modification se produit plus souvent au niveau des unités GlcNAc (Pomin *et al.*, 2012). Le KS est le seul GAG qui ne porte pas de résidu acide (Funderburgh, 2002 ; Pomin *et al.*, 2012) tel que l'acide glucuronique couramment observé dans les CS, DS, HA et HP/HS (Pomin, 2014), ou l'acide iduronique couramment observé dans le DS et l'HP/HS (Pomin, 2014). Au lieu de ces unités acides, la chaîne de KS a un sucre neutre (Gal) (Funderburgh, 2002 ; Pomin *et al.*, 2012, 2014), et cette caractéristique donne au KS un potentiel moins acide en solution (du fait que la sulfatation soit le seul composant acide de sa structure).

Il existe trois types de KS caractérisés par des sites d'attachement différents. Pour le KS-1, un GlcNAc forme une liaison N avec une asparagine du noyau protéique, et pour le KS-2, l'initiation se produit avec la formation d'une liaison O entre un GlcNAc et un résidu sérine/thréonine (Varki *et al.*, 2008). Quant au KS-3, une liaison O est réalisée entre un mannose et une sérine (Caterson et Melrose, 2018).

La biosynthèse du KS implique l'action séquentielle de deux transférases : la  $\beta$ -1,3-N-acétylglucosaminyl transférase et la  $\beta$ -1,4-galactosyl transférase, et de deux sulfotransférases (maturation des chaînes) : la N-acétylglucosaminyl-6-sulfotransférase et la KS galactosyl sulfotransférase. Ainsi, KS-1 et -2 possèdent un mélange d'unités disaccharidiques non sulfatées (Gal-GlcNAc), monosulfatées (Gal-GlcNAc6S) et disulfatées (Gal6S-GlcNAc6S) dans leurs chaînes (Varki *et al.*, 2008). Le KS-1 est plus fréquemment retrouvé au niveau du tissu cornéen alors que le KS-2 est davantage présent au niveau des cartilages (Funderburgh, 2002).



Enfin, les chaînes de KS peuvent présenter une structure de type « capping » à leur extrémité non réductrice qui les protège d'une dégradation par la kératanase I, la kératanase II, et l'endo- $\beta$ -D-galactosidase (Caterson et Melrose, 2018).

Les KSPGs peuvent être exclusifs (c'est le cas du kératocane, mimécane, lumicane, fibromoduline et ostéomoduline) ou hybrides, comme l'agrécane, qui est un à la fois un CSPG (catégorie à laquelle il est le plus souvent rattaché) et un KSPG (Funderburgh, 2002 ; Varki *et al.*, 2008 ; Pomin, 2014).

Dans la cornée, la forte abondance de KSPGs est liée au maintien des niveaux d'hydratation de ce tissu pour le maintien de la transparence de ce dernier (Quantock *et al.*, 2010). Dans le cas de changements structurels indésirables sur les chaînes de KS des PGs cornéens, ceci entraîne des dysfonctionnements visuels. C'est le cas de certains troubles visuels, comme la dystrophie maculaire cornéenne (pertes ou changements dans les profils de sulfatation) et le kératocône (dysfonctionnement dans la formation de la chaîne de KS) (Varki *et al.*, 2008 ; Quantock *et al.*, 2010). Dans des tissus comme dans la cornée et les cartilages, les KS sont également bien connus pour jouer un rôle dans le maintien de l'équilibre des propriétés d'hydratation du tissu, en particulier en tant que composant de l'agrécane qui est grandement efficace pour conférer une résistance au stress physique et aux charges sur le tissu.

Outre la fonction principale des KS dans les tissus, ce type de GAG participe également à la biologie du développement et à la migration cellulaire (Funderburgh, 2000). Par exemple, les KSPGs contenant des KS constituent une barrière à la croissance des neurites *in vitro* (Olsson *et al.*, 1996 ; Funderburgh, 2000). De plus, il a été rapporté que le KS interagissait aussi avec des protéines de signalisation telles que les « sonic hedgehog » (SHH), les FGFs, un certain nombre de kinases, et des protéines solubles impliquées dans le développement des tissus (Conrad *et al.*, 2010 ; Weyers *et al.*, 2013).

### *Chondroïtine Sulfate et Dermatine Sulfate*

Les CSPGs et DSPGs sont distribués de manière ubiquitaire dans la matrice extracellulaire, ainsi qu'à la surface des cellules (Rodén, 1980 ; Iozzo, 1998). Plus précisément, les CSPGs sont abondamment distribués dans le cartilage (Rodén, 1980),

tandis que les DSPGs sont principalement distribués dans la peau, l'aorte, et les vaisseaux sanguins (Fransson *et al.*, 1993).

Le squelette des CS est composé d'unités disaccharidiques répétées d'acide D-glucuronique et de N-acétyl-D-galactosamine [-4GlcAB1 -> 3GalNAcB1-]. Quant au DS, ce dernier est un stéréoisomère de CS et se compose donc d'acide L-iduronique au lieu de GlcA et de GalNAc [-4IdoAα1 -> 3GalNAcβ1-]. Les chaînes de CS/DS présentent également des modifications, par sulfatation, au niveau de divers groupements hydroxyles (classification A, B, C, D et E suivant la combinaison des sulfatations présente sur les unités disaccharidiques, Djerbal *et al.*, 2017). Ceci donne lieu à une diversité structurale importante, jouant ainsi un rôle dans une variété de processus biologiques, comme les interactions avec divers facteurs de croissance, cytokines et morphogènes, la prolifération cellulaire, la morphogénèse tissulaire et les infections par des virus (Sugahara *et al.*, 2003 ; Sugahara et Mikami, 2007 ; Malavaki *et al.*, 2008 ; Yamada et Sugahara, 2008 ; Malmström *et al.*, 2012 ; Mizumoto et Sugahara, 2013 ; Thelin *et al.*, 2013 ; Mizumoto *et al.*, 2015, 2017 ; Schaefer *et al.*, 2017 ; Kosho *et al.*, 2019).

Diverses glycosyltransférases, épimérasas, sulfotransférases et enzymes apparentées dans la biosynthèse de CS et DS ont été identifiées et caractérisées (Kusche-Gullberg et Kjellén, 2003 ; Mikami et Kitagawa, 2013 ; Mizumoto, 2018). En effet, des analyses fonctionnelles de CS et DS à l'aide d'organismes modèles ont révélé que ces deux GAGs sont indispensables au développement normal (Sugahara et Schwartz, 1979 ; Bernhardt et Schachner, 2000 ; Hwang *et al.*, 2003 ; Mizuguchi *et al.*, 2003 ; Sugahara *et al.*, 2003 ; Takemae *et al.*, 2003 ; Olson *et al.*, 2006 ; Maccarana *et al.*, 2009 ; Mizumoto *et al.*, 2009 ; Li *et al.*, 2010 ; Tian *et al.*, 2010 ; Watanabe *et al.*, 2010 ; Wilson *et al.*, 2012 ; Takemura *et al.*, 2020). Ainsi, des souris déficientes en chondroïtine synthase 1 sont viables, mais présentent une chondrodysplasie, une progression de la bifurcation des doigts, une ossification endochondrale retardée et une densité osseuse réduite (Wilson *et al.*, 2012). Par ailleurs, des souris déficientes en DS-épimérase 1 présentent une masse plus petite, une réduction des structures contenant de l'IdoA dans la peau, des fibrilles de collagène plus épaisses dans le derme et l'hypoderme, des queues pliées, une altération de la migration directionnelle des cellules musculaires lisses aortiques, et des anomalies de la paroi abdominale fœtale, de l'exencéphalie et de la colonne vertébrale

(pathologie spina bifida : anomalie de la fermeture du tube neural) (Maccarana *et al.*, 2009 ; Bartolini *et al.*, 2013 ; Gustafsson *et al.*, 2014).

## 1.2. Biosynthèse des HS/HP

La biosynthèse des chaînes de HS/HP est un processus complexe impliquant de nombreux systèmes enzymatiques qui se déroulent dans le « late » réticulum endoplasmique (RE) et l'appareil de Golgi. Ce processus comporte trois grandes étapes (Figure 4). L'initiation consiste en l'assemblage d'un motif linker tétrasaccharidique sur le core protéique (core protéique préalablement synthétisé dans le RE puis transféré vers l'appareil de Golgi) permettant le futur attachement de la chaîne polysaccharidique. Puis, les étapes d'élongation et de maturation impliquent de nombreuses enzymes, telles que l'épimérase, les sulfotransférases et les glycosyltransférases.

### 1.2.1. Initiation

L'initiation de la synthèse des chaînes polysaccharidique est commune à tous les GAGs et commence par le transfert d'un xylose sur les résidus sérines du dipeptide spécifique sérine-glycine du core protéique (Figure 4A). Ce transfert a lieu dans le réticulum endoplasmique ou dans le compartiment *cis* de l'appareil de Golgi *via* deux xylosyltransférases (XylT-I et XylT-II) qui utilisent l'UDP-xylose comme donneur (Pönighaus *et al.*, 2007).

De plus, le processus de xylosylation n'implique pas tous les sites d'attachement potentiels de GAG. En effet, la nature des acides aminés situés autour du dipeptide sérine-glycine, ainsi que des modifications post-traductionnelles (PTMs) de la sérine comme la phosphorylation ou la glycosylation, peuvent bloquer le transfert du xylose et donc l'amorçage de la biosynthèse des chaînes (Silbert et Sugumaran, 1995).

Suite à la xylosylation de la protéine, il s'ensuit le transfert séquentiel de deux résidus galactose, successivement ajoutés par la Galactosyltransférase-I/b4-Galactosyltransférase 7 et la Galactosyltransférase-II/b3-Galactosyltransférase 6, et enfin l'ajout d'un résidu GlcA par la Glucuronyltransférase-I, pour former le linker GlcAb1-3Gal-b1-3Gal-b1-4Xyl-bl-O-Ser (Marques *et al.*, 2021). De plus, au cours de cet

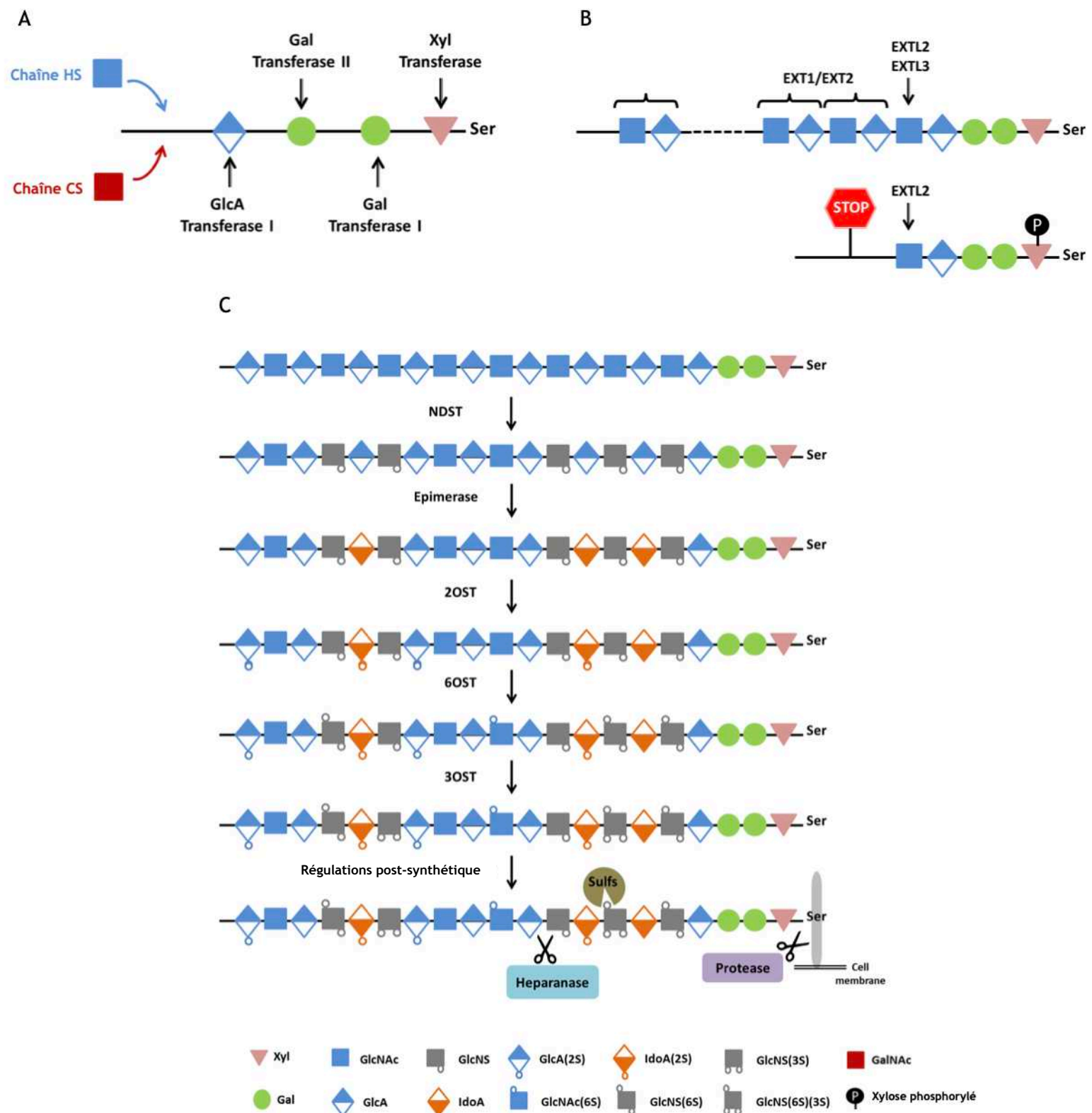


Figure 4 : Biosynthèse de HP/HS. (A) Initiation de la synthèse des HP/HS et CS/DS. La synthèse des GAGs est réalisée par l'ajout successif d'un xylose sur une sérine, de deux Gal, d'un GlcA et d'un GlcNAc ou GalNAc qui déterminera alors le type de GAG. (B) Allongement des HP/HS par l'action de EXTs. EXTL2 et EXTL3 peuvent ajouter le premier saccharide GlcNAc. L'ajout de GlcNAc et GlcA est catalysé par le complexe EXT1/EXT2. EXTL2 peut mettre fin à l'allongement des HS. (C) Maturation des HP/HS. Les chaînes HP/HS allongées subissent successivement la N-désacétylation/N-sulfatation de la glucosamine par les NDST, l'épimérisation de l'GlcA en IdoA par la C5 épimérase, la 2-O-sulfatation d'IdoA par la 2OST, la 6-O-sulfatation par la 6OSTs et enfin la 3-O-sulfatation par les 3OSTs. Les chaînes maturées peuvent ensuite être modifiées post-synthétiquement par l'action de protéases, d'héparanases et des Sulfs (Figure adaptée du manuscrit de thèse de Rana El Masri, HAL Id : tel-03507533, version 1).

assemblage, une phosphorylation transitoire du résidu Xyl par la kinase FAM20B et une déphosphorylation par la 2-Phosphoxylose phosphatase ont également lieu, ce qui améliore l'activité de b3Gal-T6, favorisant davantage la maturation du linker (Koike *et al.*, 2014 ; Wen *et al.*, 2014). La nature des chaînes de GAGs synthétisées est dépendante du type de saccharide ajouté à la suite du GlcA du linker tétrasaccharidique. Ainsi, si un GlcNAc est ajouté, alors une chaîne de HS/HP est assemblée. Cependant, si c'est un GalNAc, alors une chaîne CS/DS sera initiée. Toutefois, il n'est pas encore bien défini si le core protéique influence la nature de la chaîne de GAG synthétisée. En effet, suivant les tissus, un même core protéique peut porter différents GAGs. La serglycine porte des chaînes d'HP dans les mastocytes, mais des chaînes de CS dans d'autres types de cellules (Prydz, 2015). De plus, si le dipeptide sérine-glycine est entouré d'acides aminés acides et/ou hydrophobes, alors ceci semble favoriser la synthèse de chaînes d'HS (Kokenyesi *et al.*, 1994 ; Zhang *et al.*, 1994).

### 1.2.2. Elongation

L'élongation des chaînes de HS/HP se poursuit par addition séquentielle de GlcA et de GlcNAc par les enzymes EXT1 et EXT2 (Figure 4B). En effet, dans l'appareil de Golgi, ces enzymes forment un complexe hétéro-oligomérique capable de polymériser les HS car présentant la double activité glucuronyl- et glucosaminylnyl-transférase (Esko *et al.*, 2009). Toutefois, l'activité d'une seule enzyme n'est pas suffisante pour la bonne polymérisation de la chaîne, ces enzymes ne peuvent donc se substituer l'une à l'autre (McCormick *et al.*, 2000 ; Wei *et al.*, 2000). EXT2 semble également pouvoir interagir avec d'autres protéines impliquées dans la biosynthèse des GAGs comme l'enzyme de maturation N-désacétyl sulfotransférase 1 (NDST1). Cette interaction guide NDST1 vers l'extrémité non réductrice de la chaîne où elle exercera son activité (Busse *et al.*, 2007 ; Presto *et al.*, 2008).

De nombreuses pathologies ont été associées aux nombreuses mutations des gènes EXTs. En effet, une analyse de la répartition des mutations sur les gènes *EXT1* et *EXT2* révèle que toutes les mutations sont distribuées presque au hasard sur les 6 premiers exons, tandis que les 5 derniers exons, contenant la région carboxyterminale conservée, abritent beaucoup moins de mutations (Wuyts et Van Hul, 2000). Par exemple, une mutation des gènes EXTs (perte ou réarrangement de la région 8q24 et 11p du chromosome) peut être à l'origine de la maladie des exostoses multiples, qui est une

maladie génétique caractérisée par la croissance de nombreuses tuméfactions osseuses (exostoses ou ostéochondromes) au niveau des métaphyses des os longs.

Plus récemment, des homologues aux enzymes EXT ont été identifiés : il s'agit des EXT-like (EXTL1, EXTL2 et EXTL3). EXTL2 et EXTL3 semblent être impliquées dans l'initiation de HS étant donné qu'elles peuvent toutes les deux ajouter la première GlcNAc au linker tétrasaccharidique. De plus, EXTL2 pourrait jouer un rôle dans la terminaison de l'élongation des HS car l'ajout d'un GlcNAc dans une région contenant un xylose phosphorylé empêche le transfert de saccharides (Nadanaka *et al.*, 2013a). Pour finir, cette enzyme semblerait être, également, impliquée dans un autre système de régulation des HS. Ainsi, une expression nulle ou réduite d'EXTL2 entraîne respectivement davantage de chaînes d'HS produites (Nadanaka *et al.*, 2013b), ou des chaînes d'HS plus longues (Katta *et al.*, 2015).

Une étude très récente, étudiant les mécanismes de régulation sous-jacents des voies de biosynthèse des HSPGs dans le contexte du cancer gastrique, a montré que des cellules cancéreuses gastriques dépourvues d'EXTL2 et EXTL3 présentaient une altération de leurs niveaux en HS et CS ainsi que des caractéristiques structurales des chaînes de GAGs, suggérant que les deux enzymes exerceraient un large rôle de régulateur dans les voies de biosynthèse des GAGs. Plus précisément, il a été observé que si EXTL3 est la clé pour initier la synthèse des HS, au détriment de la biosynthèse des CS, EXTL2 fonctionne comme un régulateur négatif de la biosynthèse des HS. Ainsi, les rôles cruciaux d'EXTL2 et d'EXTL3 dans la modulation de l'expression des PGs et des profils de glycosylation aberrants affichés par les cellules cancéreuses, impacteraient le comportement malin des ces dernières (Marques *et al.*, submitted).

### **1.2.3. Maturation des HS/HP**

Lors de la dernière étape de la synthèse, les disaccharides HP/HS subissent des modifications réalisées par diverses enzymes (Figure 4C). La première de ces réactions est la N-désacétylation/N-sulfatation de la glucosamine, catalysée par les NDST (4 isoformes). Celle-ci est suivie de l'action de la C5 épimérase et de la 2-O-sulfotransférase (2OST) réalisant, respectivement, l'épimérisation de GlcA en un résidu IdoA et la 2-O-sulfatation de IdoA. Enfin, les groupements sulfates en position C6 et C3 de la glucosamine sont ajoutés, respectivement, par les familles des 6-O-

sulfotransférases (6OST, 3 isoformes) et des 3-O-sulfotransférases (3OST, 7 isoformes). Cette sulfatation des différentes positions est rendue possible par le transfert du groupement sulfate du substrat 3'-phosphoadenosine 5'-phosphosulfate (PAPS). Le PAPS est synthétisé dans le cytosol puis transféré vers l'appareil de Golgi par la PAPS translocase.

En termes de structure, l'ensemble des enzymes de biosynthèse des HS (à l'exception d'une isoforme 3OST) sont des protéines membranaires de type II, composées d'une courte queue cytoplasmique, d'un domaine transmembranaire hydrophobe et d'une région en tige portant le domaine catalytique. Notons toutefois que ces isoformes ne sont pas exprimées en même temps et présentent au contraire des profils d'expression différentiels (Li et Kusche-Gullberg, 2016).

#### I.2.3.1. NDSTs

Comme indiqué précédemment, les NDSTs interviennent dans les tous premiers processus de maturation des HS. En effet, sans leur activité les processus de maturation, comme l'épimérisation de GlcA en IdoA et les sulfatations 2-O et 3-O, sont diminués (Bame *et al.*, 1991a, 1991b, 1994). Cependant, certains groupements 6-O-sulfatés peuvent être ajoutés malgré l'absence préalable d'activité NDSTs (Holmborn *et al.*, 2004).

Les NDSTs représentent une famille de protéines membranaires de type II composée de 4 isoformes chez les mammifères, partageant toute une grande similarité de séquences (Aikawa et Esko, 1999 ; Aikawa *et al.*, 2001). Ces protéines catalysent l'élimination du groupement N-acétyl de la GlcNAc, suivie de l'ajout d'un groupement sulfate à la même place. Pour ce faire, deux sites catalytiques distincts fonctionnant en coopérativité sont nécessaires. Le premier site présent dans le domaine N-terminal permet l'activité N-désacétylase, alors qu'un site dans le domaine C-terminal permet l'activité N-sulfotransférase (Berninsone et Hirschberg, 1998). Notons que si la coopérativité des sites n'est pas optimale, comme avec l'inhibition de l'activité N-sulfotransférase, alors le taux de N-désacétylation diminue également (Dou *et al.*, 2015).

Toutefois, des différences existent entre les isoformes. NDST1 et NDST2 sont fortement exprimées au stade embryonnaire et adulte, mais NDST2 est exprimée dans les mastocytes alors que NDST1 est plus largement distribuée. Quant à NDST3 et NDST4, elles sont principalement produites au cours du développement embryonnaire, bien que NDST3 puisse être trouvée dans le cerveau adulte. Des études *in vivo* réalisées sur des souris KO *NDST1* et *NDST3* ont mis en lumière des altérations physiologiques ou comportementales. En effet, l'inactivation de *NDST1* génère une détresse respiratoire fatale (Fan *et al.*, 2000), alors qu'une inhibition de *NDST3* n'entraîne aucune létalité (Pallerla *et al.*, 2008). Notons que dans le cas d'une réponse inflammatoire, une augmentation importante de l'expression de NDST1 a été identifiée (Adhikari *et al.*, 2008). Quant à NDST2, son inactivation n'est pas létale mais conduit à une réduction des mastocytes dans le tissu conjonctif et à une longueur de chaîne d'HS réduite (Forsberg *et al.*, 1999 ; Deligny *et al.*, 2016).

Des études sur l'activité des différentes NDST ont également montré des différences significatives. En effet, NDST4 présente un très faible rapport d'activité désacétylase/sulfotransférase (0,04), au contraire de NDST3 pour laquelle ce rapport est très élevé (10,5). Pour NDST1 et NDST2, ce rapport est relativement équilibré (Aikawa *et al.*, 2001). Ceci suggère donc l'existence d'une coopérativité entre isoformes, probablement plus importante que la coopérativité inter-domaine. Ainsi, certaines isoformes réaliseraient la désacétylation du GlcNAc, mais d'autres réaliseraient sa sulfatation. Un exemple bien documenté reste la préférence de l'action de NDST1. Cette enzyme conduit préférentiellement à la synthèse de chaînes d'HS relativement peu sulfatés et modifiés, ainsi qu'une participation importante dans la distribution des domaines S et NS le long de la chaîne de HS (Sheng *et al.*, 2011). En revanche, la maturation des chaînes d'HP (très sulfatées) est préférentiellement réalisée par NDST2 (Forsberg *et al.*, 1999 ; Deligny *et al.*, 2016).

### 1.2.3.2. Epimérase

La conversion de GlcA en IdoA est réalisée par la C5-épimérase, et exercerait une influence importante sur la flexibilité des HS (le résidu IdoA pouvant adopter différentes conformations) ce qui faciliterait l'interaction des ligands aux HS. Cette réaction enzymatique, post activité NDST, consiste en l'élimination et la ré-addition du proton en



position C5 *via* un intermédiaire carbanion, avec une inversion du groupement carboxyle par rapport au plan formé par l'hétérocycle du pyranose (Lindahl *et al.*, 1976). Contrairement à des systèmes solubles, dans un système cellulaire, cette réaction est irréversible et n'atteint pas un équilibre (Hagner-Mcwhirter *et al.*, 2000, 2004). En effet, il est suggéré qu'une O-sulfatation supplémentaire favorise la génération d'IdoA, en bloquant la réversibilité de l'épimérisation. Ainsi, les résidus IdoA sont plus susceptibles d'être 2-O-sulfatés que les résidus GlcA. Il a également été montré que la C5 épimérase peut s'associer à la 2-O-sulfotransférase afin d'améliorer sa stabilité, son efficacité, et sa translocation vers l'appareil de Golgi (Pinhal *et al.*, 2001 ; Préchoux *et al.*, 2015).

La structure cristallographique de l'épimérase humaine et du poisson zèbre a été résolue avec ou sans la présence d'un oligosaccharide comme ligand. Ainsi, elles ont été identifiées comme des protéines transmembranaires de type II présentant une organisation dimérique. De ces structures, il a pu être déterminé que chaque sous-unité présente une hélice  $\alpha$  C-terminale comprenant trois tyrosines cruciales pour l'activité enzymatique. De plus, ces sous-unités sont chargées positivement, permettant ainsi l'interaction avec le ligand chargé négativement (Qin *et al.*, 2015 ; Debarnot *et al.*, 2019).

L'épimérase a été mise en évidence dans de nombreux processus physiologiques comme la lymphangiogenèse, le développement neuronal, la maturation des cellules B (Bülow et Hobert, 2004 ; Feyerabend *et al.*, 2006 ; Reijmers *et al.*, 2010, 2011). En effet, la mutation ciblée du gène murin de la C5-épimérase a pour résultat la mort néonatale de la souris, avec des anomalies rénales, pulmonaires et du développement squelettique. Les souris KO présentent des HS dépourvus de résidus IdoA, avec une diminution de la 2-O-sulfatation et une augmentation des N- et 6-O-sulfatation (Li *et al.*, 2003). Ce phénotype létal peut être dû à l'inhibition de l'interaction entre les HS et les protéines de signalisation étant donné que le résidu IdoA semble être un résidu important pour de nombreux sites d'interaction.

De surcroît, il a été observé que le gène de l'épimérase peut agir comme un suppresseur de tumeur, car une régulation négative de l'enzyme a été signalée dans les cancers du sein et du poumon (Grigorieva *et al.*, 2008, 2011).

### 1.2.3.3. 2-O-sulfotransférase

La 2OST catalyse l'ajout de groupements sulfates en position C2 de l'acide uronique. Des études structurales ont permis l'identification de résidus de la 2OST impliqués dans la spécificité du substrat (PDB ID code : 3F5F ; Bethea *et al.*, 2008). En effet, le groupement N-sulfaté du polysaccharide en synthèse est reconnu par trois acides aminés (deux arginines et une lysine) de la 2OST. A l'inverse, la 2OST n'établit pas de contact avec les groupements 6-O-sulfatés, ce qui suggère que l'activité 2OST se produit avant celle des 6OSTs (Liu *et al.*, 2014). Toujours d'après les données structurales, il a également été démontré que deux histidines sont essentielles pour l'activité catalytique. De surcroît, la structure de la 2OST du poulet met en avant une enzyme trimérique qui serait essentiel pour son activité enzymatique.

La sulfatation des résidus GlcA est très rare, sauf dans le cortex cérébral humain adulte et dans les fractions nucléaires des hépatocytes (Fedarko et Conrad, 1986 ; Lindahl *et al.*, 1995). Au contraire, l'IdoA (2S) se trouve dans la plupart des motifs protéiques de liaison aux HS. Il intervient notamment dans la liaison à l'ATIII, au FGF2, à la lipoprotéine lipase, au HGF et aux facteurs de croissance dérivés des plaquettes (Maccarana *et al.*, 1993 ; Lyon *et al.*, 1994 ; Kreuger *et al.*, 2001 ; Jemth *et al.*, 2002).

En termes fonctionnels, le KO de la 2OST entraîne une mort néonatale, une agénèse rénale, un retard de développement embryonnaire chez la souris (Bullock *et al.*, 1998), et une perturbation du développement du système nerveux chez *C. elegans* (Kinnunen *et al.*, 2005). Toutefois, chez la drosophile, le KO de la 2OST est moins impactant, probablement dû au fait d'un effet compensatoire de la 6-O-sulfatation (Kamimura *et al.*, 2006). Cet effet compensatoire a également été démontré dans les cellules endothéliales de souris (Axelsson *et al.*, 2012).

### 1.2.3.4. 3-O-sulfotransférases

Lors des dernières étapes de la maturation des HS, l'addition de groupements sulfate en position C3 de la GlcNAc est réalisée par les 3OSTs (Thacker *et al.*, 2013). Malgré le grand nombre d'isoenzymes 3OSTs, sept isoformes chez les mammifères (3OST1, 2, 3A, 3B, 4, 5 et 6) et huit chez le poisson zèbre (Cadwallader et Yost, 2006), la

sulfatation en position C3 est relativement rare par rapport à d'autres modifications (par exemple, la N-sulfatation des unités de glucosamine et la 2-O-sulfatation des acides iduroniques). En effet, bien que 10 % des glucosamines des HS sont 3-O-sulfatées dans les membranes de Reichert, cette modification touche ~ 1 % des glucosamines des HS des cellules endothéliales et n'est pas détectée dans les cellules ovariennes de hamster chinois (CHO) (Thacker *et al.*, 2016).

Les membres de la famille des 3OSTs sont des protéines transmembranaires, à l'exception de 3OST1 qui n'a pas de région transmembranaire. Pour cette dernière, sa localisation dans l'appareil de Golgi pourrait être due à des interactions avec d'autres protéines qui y résident (Liu *et al.*, 1999). Ces enzymes agissent tardivement dans la voie de biosynthèse et dépendent donc d'une modification préalable par les autres sulfotransférases et l'épimérase (Shworak *et al.*, 1996 ; Liu *et al.*, 1999). Elles ont également des spécificités de substrat différentes, ce qui entraîne le positionnement des 3-O-sulfates dans différents contextes structuraux (au sein du polysaccharide) (Esko *et al.*, 2002 ; Thacker *et al.*, 2013).

Malgré la description d'un très grand nombre de protéines interagissant avec les HS (Gómez Toledo *et al.*, 2021 ; Vallet *et al.*, 2021, 2022), seuls quelques ligands protéiques ont montré une prédilection pour les HS 3-O-sulfatés (Thacker *et al.*, 2013). Par exemple, les 3OST1 et 3OST5 sont impliquées dans la formation de sites de fortes affinités pour l'anti-thrombine (AT) (Shworak *et al.*, 1997 ; Xia *et al.*, 2002 ; Chen *et al.*, 2005). Ainsi, la liaison de l'AT avec le motif pentasaccharidique 3-O-sulfaté de type [Glc (NS ou NAc) (6S), GlcA, GlcNS (3S ± 6S), IdoA (2S), GlcNS (6S)] induit un changement conformationnel et potentialise son activité inhibitrice, ce qui en fait un anti-coagulant puissant.

Les 3OST2, -3A, -3B, -4, -5 et -6 peuvent créer le site de liaison de la glycoprotéine D et du virus de l'herpès Simplex 1, et il est supposé que leur action est également associée à d'autres fonctions (Thacker *et al.*, 2013). D'autres ligands se sont révélés préférer les HS 3-O-sulfatés, comme le FGF7, l'ectodomaine du récepteur FGF-(FGFR)1, la cyclophiline B et la stabiline (McKeehan *et al.*, 1999 ; Deligny *et al.*, 2010 ; Pempe *et al.*, 2012). Des études récentes sur les cellules « kit+ progenitor » dans la glande salivaire indiquent que les interactions avec les motif 3-O-sulfatés stabilisent les complexes FGF10/FGFR2b (Patel *et al.*, 2014), et des études génétiques chez le poisson zèbre suggèrent que la

protéine morphogénique osseuse 4 (BMP4) et le FGF8 préfèrent également les HS 3-O-sulfatés (Neugebauer *et al.*, 2013 ; Samson *et al.*, 2013). Pour finir, la 3-O-sulfatation module divers processus développementaux et pathologiques, y compris la différenciation des cellules souches, la ciliogenèse, le ciblage neuronal et la progression tumorale, mais les protéines qui se lient aux séquences 3-O-sulfatées et qui participent à ces processus restent largement inconnues (Miyamoto *et al.*, 2003 ; Hirano *et al.*, 2012 ; Neugebauer *et al.*, 2013 ; Samson *et al.*, 2013 ; Teclé *et al.*, 2013). Toutefois, une étude récente a développé une approche expérimentale pour étendre l'identification et la compréhension du « protéome 3-O-sulfaté » (Thacker *et al.*, 2016). Ainsi, elle a pu caractériser de nouvelles protéines préférant les HS 3-O-sulfatés, comme la neuropiline-1, une protéine qui participe à la vasculogenèse et au guidage axonal (Kolodkin *et al.*, 1997 ; Kawasaki *et al.*, 1999), mais aussi valider l'importance de la 3-O-sulfatation dans les interactions NRP1/HS dans divers tests de liaison, en montrant que l'effondrement du cône de croissance axonal induit par la sémaphorine-3a dépend de l'action de la 3OST2.

En termes de structure, les structures hautes résolutions des sulfotransférases (3OST1, 3OST3 et 3OST5) en présence de PAPS et d'oligosaccharides ont révélées la présence de deux domaines essentiels. Les domaines catalytiques de toutes les 3OST utilisent un résidu glutamate, conservé, comme base catalytique. Celui-ci est à l'origine de la liaison à l'accepteur et est essentiel à l'activité enzymatique (Edavettal *et al.*, 2004). Le second domaine, hors du domaine catalytique, est appelé «the gate» et contient des résidus impliqués dans la spécificité du substrat. Fait intéressant, inverser ces résidus entre 3OST1 et 3OST5 est suffisant pour convertir la spécificité de substrat d'une enzyme à l'autre (Moon *et al.*, 2004, 2012 ; Xu *et al.*, 2008).

#### **1.2.3.5. 6-O-sulfotransférases**

Les 6OSTs catalysent le transfert d'un groupement sulfate vers la position C6 des unités de glucosamine à partir d'un donneur PAPS. La 6OST1 fut initialement caractérisée à partir de cellules CHO (Habuchi *et al.*, 1995), puis, grâce aux études d'homologie de séquence, les isoformes 6OST2 et 6OST3 ont été identifiées par la suite (Habuchi *et al.*, 2000), ainsi qu'une forme tronquée de la 6OST2 contenant une délétion de 40 acides aminés, 6OST2-S (Habuchi *et al.*, 2003). Suivant les organismes, les 6OST

peuvent présenter différentes isoformes, avec deux isoformes décrites chez le poulet, quatre chez le poisson zèbre et trois chez l'Homme et la souris, ainsi que des orthologues chez la Drosophile, le Xénope et *C. elegans*. Chez l'Homme, les isoformes 6OST sont codées par trois gènes présents sur les chromosomes 2, X et 13 (Habuchi *et al.*, 2003). A l'inverse des autres glucosaminyl-sulfotransférases, la 6OST1 n'affiche que 51 % et 57 % d'identité de séquence avec 6OST2 et 6OST3 respectivement, tandis que la 6OST2 et 6OST3 ne partagent que 50 % d'homologie (Habuchi *et al.*, 2000). Cependant, la région centrale, contenant potentiellement la séquence de liaison au PAPS, est très conservée entre les trois isoformes (Habuchi, 2000).

A l'instar de la majorité des sulfotransférases, les 6OSTs sont des protéines transmembranaires de type II qui résident dans l'appareil de Golgi. De la même manière, elles présentent une région N-terminale en tige indispensable à la formation des complexes oligomériques et au maintien de l'enzyme dans un état actif (Nagai *et al.*, 2004). Toutefois, contrairement à la plupart des autres enzymes de biosynthèse des HS, la 6OST3 peut également être trouvée dans l'environnement extracellulaire (Habuchi *et al.*, 1995). En revanche, les mécanismes sous-jacents à cette sécrétion restent flous, mais impliquent la  $\beta$ -sécrétase comme enzyme de clivage du court domaine cytoplasmique hydrophobe de la 6OST3. En effet, l'inhibition de la  $\beta$ -sécrétase entraîne une accumulation de la 6OST3 dans l'appareil de Golgi et une augmentation de la 6-O-sulfatation des HS. Par conséquent, il peut être suggéré que l'équilibre intracellulaire/extracellulaire de la 6OST3 contribue à la régulation de la 6-O-sulfatation des HS.

De manière assez surprenante, les études des 6OST n'ont pas révélé de différences majeures dans l'activité et les spécificités de substrats des différentes isoformes. Ainsi, elles peuvent toutes ajouter des groupements sulfates sur les unités GlcNAc et GlcNS, quelle que soit la nature de l'acide uronique adjacent (Smeds *et al.*, 2003). Cependant, des études utilisant des oligosaccharides de structure bien définies ont révélé que la 6OST1 agit préférentiellement sur les motifs [IdoA, GlcNS]. Quant à la 6OST2, celle-ci est plus active envers les substrats 2-O-sulfatés (Jemth *et al.*, 2003). Enfin, la 6OST3 présente une spécificité de substrat intermédiaire, entre la 6OST1 et la 6OST2.

En revanche, les isoformes 6OST se distinguent par leur distribution spatio-temporelle (Habuchi *et al.*, 2000, 2003 ; Sedita *et al.*, 2004). Chez les souris adultes, la 6OST1 est principalement présente dans le foie mais modérément exprimée dans le cerveau, le

cœur, les reins et les poumons. Pour la 6OST2, son expression est spécifique au cerveau et à la rate. En revanche, son variant 6OST2-S, issu d'un épissage alternatif, est plus présent dans les ovaires, le placenta et le rein fœtal. Enfin, la 6OST3 est exprimée de manière ubiquitaire. Il convient également de noter que l'expression de ces isoformes varie considérablement au cours du développement. Ainsi, cette expression différentielle peut engendrer des HS différemment sulfatés, à l'origine de propriétés de liaison spécifiques pour certains ligands, et donc capables d'induire des processus de développement spécifiques. Dans l'ensemble, ces données indiquent clairement que la 6-O-sulfatation est catalysée par des enzymes dont les activités se chevauchent et se complètent.

La régulation de ces protéines n'est pas clairement définie, mais une étude a pu mettre en évidence une létalité élevée et des défauts de migration des cellules trachéales chez la drosophile où la dHS6ST est inhibée (Kamimura *et al.*, 2001). En effet, le KO de certaines 6OST conduit à des défauts anatomiques. Chez la souris, le KO de la 6OST1 conduit à un niveau élevé de létalité au cours des stades embryonnaires tardifs. Si toutefois les souris dépassent le stade embryonnaire, elles présentent des retards de croissance et des phénotypes morphologiques anormaux, tels qu'une altération de l'ossification, une réduction de la masse corporelle et une morphologie pulmonaire altérée entre autres malformations (Habuchi *et al.*, 2007 ; Pratt *et al.*, 2006). Malgré l'absence d'anomalie phénotypique pour le KO de la 6OST2, le double KO 6OST1/6OST2 entraîne une mortalité totale des souris, plus précoce que celle observée chez les souris simple KO 6OST1 (Sugaya *et al.*, 2008).

Pour finir, l'analyse de la composition des HS de divers organes de souris KO a montré une réduction de la teneur en 6-O-sulfatation. Pour les souris KO 6OST1, la diminution de la 6-O-sulfatation est modérée sur tous les disaccharides contenant la 6-O-sulfatation. Ceci étant, les unités disaccharidiques [IdoA (2S), GlcNS (6S)] se sont révélées être moins affectées. En revanche, l'analyse des HS de fibroblastes embryonnaires (MEF) de souris KO 6OST2 a montré une diminution marquée de la 6-O-sulfatation, l'unité [IdoA (2S), GlcNS (6S)] étant la plus affectée. Enfin, le double KO 6OST1/6OST2 a entraîné une perte quasi complète de 6-O-sulfatation des HS des MEF. Ces derniers présentent cependant une augmentation de 2-O-sulfatation, suggérant un possible effet compensatoire entre 6OST et 2OST (Do *et al.*, 2006). Cependant, une analyse transcriptionnelle des souris KO 6OST1 n'a montré aucun mécanisme de

compensation entre 6OST. Par conséquent, les effets de compensation ou non entre OST et isoformes ne sont pas encore clairs et peuvent varier considérablement entre les tissus et les espèces. Toutefois, cela souligne clairement les liens étroits entre les différentes étapes de modification enzymatique pendant la biosynthèse des HS.

#### 1.2.4. Le concept de gagosome

Au cours de la biosynthèse HS, la plupart des étapes d'assemblage et de maturation des polysaccharides sont contrôlées par des familles d'enzymes situées dans l'appareil de Golgi (Figure 5). Cependant, considérer toutes ces enzymes (avec un tel

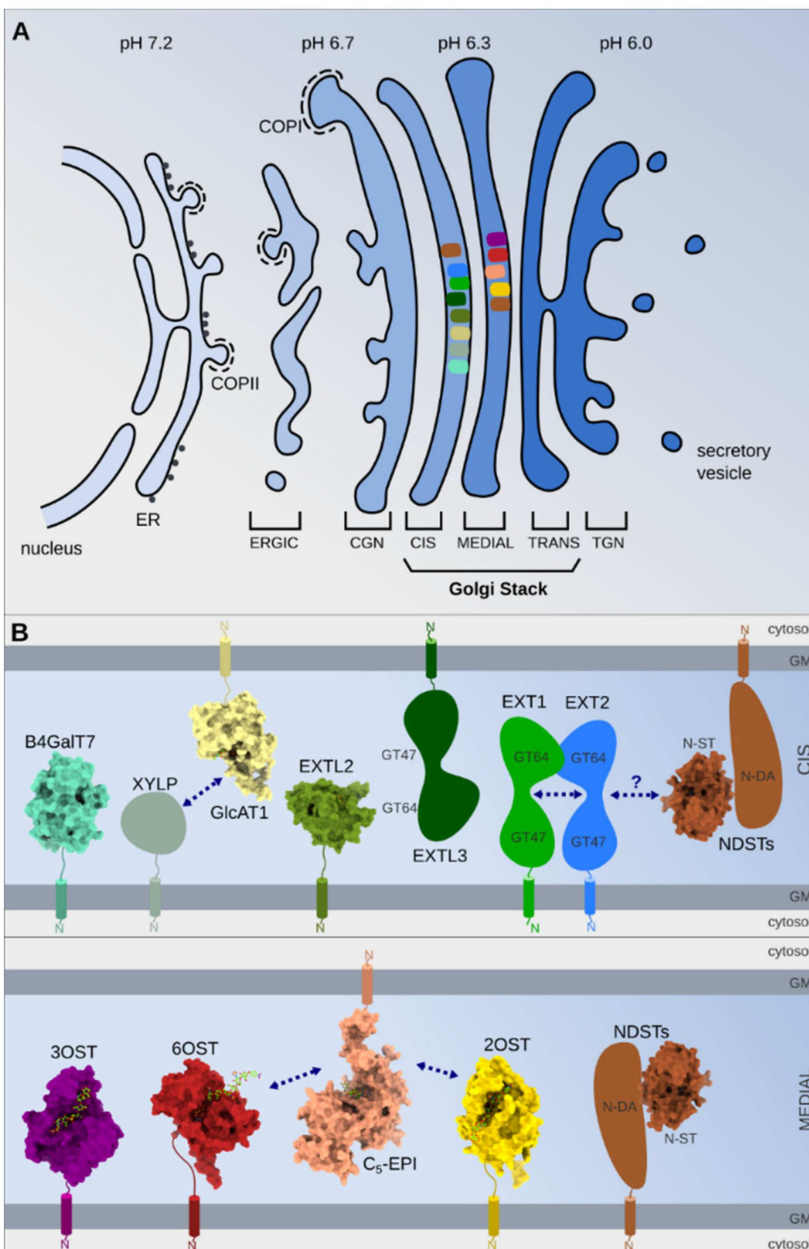


Figure 5 : Gagosome. (A) Schéma illustrant le complexe de Golgi. Les vésicules recouvertes de COPII assurent le transport des protéines du RE vers les amas vésiculaires (ERGIC), qui fusionnent pour former le réseau cis-Golgi (CGN). Le « Golgi Stack » est constitué de trois compartiments distincts : cis-, medial- et trans-cisternae. Le dernier élément de Golgi, faisant face à la membrane plasmique, est le réseau trans-Golgien (TGN). Certaines localisations enzymatiques sont représentées schématiquement par différentes couleurs et elles ont été suggérées en utilisant l'immunocytochimie confocale et différents marqueurs de Golgi ou traitements médicamenteux de Golgi (nocodazole, briefeldine A). (B) Schéma de l'organisation des enzymes impliquées dans la biosynthèse des HS. Les protéines, pour lesquelles une structure de cristallographie aux rayons X a été rapportée, sont présentées en représentation de surface et les ligands liés en bâton. ID PDB (B4GalT7 : 4IRQ, GlcAT1 : 1V84, EXTL2 : 10N8, NST : 1NST, 3OST : 3UAN, 6OST : 5T0A, 2OST : 4NDZ, C<sub>5</sub>-EPI : 6I01). Les flèches indiquent les interactions rapportées entre les protéines. GM représente la membrane de l'appareil de Golgi (Figure : Annaval *et al.*, 2020).

nombre d'isoformes, des activités spécifiques, des préférences de substrat et une grande variété dans leur distribution tissulaire) comme des acteurs indépendants dans la biosynthèse du polysaccharide n'est pas l'hypothèse la plus vraisemblable. Au contraire, il est de plus en plus suggéré la présence de grands complexes regroupant toutes ou parties de ces enzymes, définissant ainsi le concept de gagosome. Ces derniers contrôleraient ainsi de manière coopérative l'élongation et la maturation des HS, selon les besoins des cellules, et permettraient de générer une grande diversité des motifs structuraux d'HS avec des fonctions précises.

Ce concept n'est cependant, à ce jour, pas encore formellement démontré. En effet, les études se limitent à la caractérisation de paires d'enzymes de biosynthèse et non à des complexes plus larges. Néanmoins, l'hypothèse de l'existence du gagosome repose sur des bases solides. En effet, la surexpression de NDST1 et NDST2 dans les cellules Human Embryonic Kidney 293 (HEK 293) affecte différemment la sulfatation des HS, suggérant leur interaction avec d'autres enzymes de maturation (Pikas *et al.*, 2000 ; Ledin *et al.*, 2006). De plus, une interaction entre NDST1 et EXT2 a été mise en évidence. Celle-ci permettrait de protéger NDST1 de la dégradation et pourrait la guider vers son site d'action dans l'appareil de Golgi. En effet, la surexpression de EXT2 déclenche l'expression de NDST1. Cependant, si EXT1 est surexprimée, alors la quantité de NDST1 diminue. Ceci suggère donc une compétition dans l'interaction de ces deux protéines avec EXT2, entraînant de ce fait une dégradation des NDST1 non liées dans le RE (Presto *et al.*, 2008).

EXT1 et EXT2 forment un complexe hétéro-oligomère s'accumulant dans l'appareil de Golgi. Or, ce complexe présente une activité glycosyltransférase plus élevée que EXT1 seul (McCormick *et al.*, 2000).

L'existence du gagosome a été suggérée par la synthèse des résidus IdoA (2S) mettant en lumière un lien entre l'ajout de groupements 2-O-sulfatés et l'épimérisation. Dans ce contexte, il a été démontré que la 2OST et la C5 épimérase interagissent physiquement avec une forte affinité, permettant ainsi la formation d'un complexe à l'origine d'une stabilité accrue de l'épimérase. Cette interaction serait également à l'origine de l'activité des enzymes et permettrait la translocation vers l'appareil de Golgi, ceci générant des domaines de plus en plus étendus de [GlcNS, IdoA (2S)]. Par conséquent, cette association d'enzymes pourrait assurer une maturation des HS plus efficace (Pinhal *et al.*, 2001 ; Préchoux *et al.*, 2015).



### I.3. Processus de régulation post-synthétique des HS

Une fois la biosynthèse des HS réalisée, des protéines, comme les sheddases, l'héparanase et les 6-O-endosulfatases extracellulaires, viennent également réguler la structure des HS afin de générer davantage de diversités structurales et fonctionnelles (Rosen et Lemjabbar-Alaoui, 2010 ; Hammond *et al.*, 2014 ; Vivès *et al.*, 2014). Notons que, bien que les endosulfatases extracellulaires sont également à l'origine de régulation post-synthétique des HS, elles ne seront détaillées que dans la partie suivante (cf. II. p. 61).

#### I.3.1. « Shedding »

L'élimination protéolytique des éctodomains des protéines membranaires (excrétion des ectodomains) est une modification post-traductionnelle qui contrôle les niveaux et la fonction de centaines de protéines membranaires. Les protéases contributives, appelées sheddases, agissent comme des interrupteurs moléculaires importants dans des processus allant de la signalisation à l'adhésion cellulaire, et liées à des pathologies telles que l'inflammation et la maladie d'Alzheimer (Lichtenthaler *et al.*, 2018). Alors que les protéases des familles ADAM et "beta-siteAPP cleaving enzyme" (BACE) sont largement considérées comme des sheddases, ces dernières années, d'autres protéases, y compris des protéases intramembranaires (certaines MMPs, « Meprin B ») et solubles (Légumine (δ-secretase), Cathepsines S et L) semblent médier le processus de « Shedding » (Lichtenthaler *et al.*, 2018).

Des études portant sur syndécan-1, cible avérée des sheddases, ont montré une corrélation entre la présence de syndécan-1 solubles et la croissance tumorale (Yang *et al.*, 2002). De plus, dû au fait que syndécan-1 ait été trouvé dans le noyau des cellules des myélomes et des mésothéliomes, il a été suggéré que les « shed » syndécan-1 possèdent un récepteur non identifié permettant son transport vers le noyau où il peut jouer des fonctions spécifiques (Chen et Sanderson, 2009 ; Zong *et al.*, 2009). En effet, il a été démontré qu'avec l'aide de l'héparanase, les HS présents dans le noyau peuvent modifier l'activité de l'ADN topoisomérase I et de l'histone acétyl transférase (Buczek-Thomas *et al.*, 2008 ; Chen et Sanderson, 2009).

### 1.3.2. Héparanase

L'héparanase (HSPE) est une endoglycosidase qui clive les HS et participe donc à leur dégradation et au remodelage de la MEC. Elle clive les longues chaînes d'HS en fragments plus courts de 10 à 20 unités saccharidiques en ciblant les liaisons [GlcA-GlcNS]. La récente structure cristallographique de l'HSPE (une endo- $\beta$ -glucuronidase de la famille des glycoside hydrolase 79) en complexe avec un oligosaccharide d'HP a montré que l'héparanase reconnaît spécifiquement un trisaccharide portant des groupements N-sulfatés et 6-O-sulfatés (Wu *et al.*, 2015). Ainsi, l'héparanase cible principalement les régions de transition, libérant de larges fragments de type domaine S.

L'héparanase joue également un rôle important dans la régulation des PGs matriciels, car elle est capable de digérer à la fois les HSPGs extracellulaires et les HSPGs liés à la membrane. Ainsi, en déstabilisant la cohésion matricielle, elle facilite la migration cellulaire, l'invasion et la formation de nouveaux vaisseaux sanguins, et libère les cytokines et facteurs de croissance (VEGF, FGFs...) alors associés à la matrice. Cette dernière va ainsi pouvoir induire de nombreux processus biologiques, tels que la croissance tumorale et la métastatisation de ces dernières (Sanderson *et al.*, 2004 ; Vlodaysky *et al.*, 2006), de même que le remodelage de la matrice extracellulaire au niveau des sites inflammatoires (Ihrcke *et al.*, 1998). De plus, et comme indiqué précédemment, l'héparanase catalyse également la libération de fragments d'HS capables de promouvoir l'activité de facteurs de croissance impliqués, notamment, dans la communication cellulaire ou l'autophagie (Vlodaysky *et al.*, 2018).

Une surexpression de l'héparanase a, aussi, pu être corrélée aux processus métastatiques de cancers et à la chimiorésistance, ainsi qu'à diverses pathologies comme la thrombose, l'athérosclérose, la fibrose et le diabète (Vlodaysky *et al.*, 2018).

Par conséquent, l'héparanase représente une cible thérapeutique potentielle dans les maladies citées. En effet, il a été démontré qu'une inhibition de l'héparanase pouvait normaliser la structure vasculaire des tumeurs solides, empêchant la chimiorésistance et favorisant l'administration d'agents chimiothérapeutiques (Zhang *et al.*, 2018). De ce fait, des médicaments anticancéreux (PG545, Pixatimod) inhibant l'héparanase ont été développés. Ces derniers imitent les HS et induisent l'apoptose des cellules de lymphome (Weissmann *et al.*, 2018).

Une fois synthétisée, l'héparanase est d'abord adressée vers le RE grâce à son peptide signal, puis vers l'appareil de Golgi où elle est sécrétée sous une forme latente par des vésicules Golgiennes. Notons que chez l'Homme, l'héparanase présente six sites de glycosylation vraisemblablement impliqués dans la sécrétion de cette dernière, mais qui ne sont pas indispensables pour son activité (Simizu *et al.*, 2004).

Une fois sécrétée, l'héparanase interagit rapidement avec les HSPGs membranaires et induit leur internalisation par endocytose. Une fois internalisé, le core des PGs est digéré par protéolyse et les chaînes d'HS subissent l'action de glycosyl-hydrolases et de sulfatases. Cependant, l'héparanase n'est pas encore active à ce stade. En effet, ce n'est qu'après la conversion des vésicules endosomales en lysosomales que cette dernière devient active. Ainsi, les chaînes d'HS sont processées afin de générer des extrémités non réductrices supplémentaires pour l'activité des exoglycosidases lysosomales (Vlodavsky *et al.*, 2018).

Un homologue de l'héparanase, l'héparanase II (Hpa2), a été identifiée et pourrait jouer un rôle dans la régulation de l'enzyme. En effet, Hpa2 est capable de se lier aux HS, et ce avec une affinité plus élevée que l'héparanase, mais sans activer leur internalisation (McKenzie *et al.*, 2000). Ainsi, en se liant aux HS, Hpa2 peut entrer en compétition avec l'héparanase et inhiber son activité endoglycosidase. Hpa2 peut également inhiber l'activité de l'héparanase en s'y liant directement (Levy-Adam *et al.*, 2010).

#### 1.4. Les interactions protéines/HS

Grâce à cette fine régulation des HS par les enzymes de biosynthèse et de post-synthèse, ces derniers présentent des caractéristiques structurales et une organisation moléculaire en domaines permettant la reconnaissance d'une large variété de protéines (Gallagher, 2015).

Les interactions protéines/HS sont le plus souvent de nature électrostatique et réalisées entre un motif disaccharidique précis, situé dans les domaines S du polysaccharide, et des acides aminés basiques de la protéine (principalement les arginines et les lysines, et occasionnellement les histidines).

Un modèle, relativement ancien, présente deux séquences consensus pouvant être impliquées dans la liaison à l'héparine (Cardin et Weintraub, 1989). Ces séquences sont X-B-B-X-B-X et X-B-B-B-X-X-B-X, où B est un résidu basique (arginine ou lysine) et X est

hydrophobe. Cette étude montre qu'outre la charge globale, la disposition des résidus basiques est importante pour l'interaction électrostatique. Ces sites de liaison peuvent être localisés dans des régions non structurées (cas des interférons et du VEGF) ou bien dans des boucles, comme pour le FGF et l'ATIII (Gallagher, 2015). Cependant, certaines chimiokines acides, telles que CCL3 et CCL4, se lient aux HS. Par ailleurs, il a été montré que les interactions de type électrostatique ne contribuaient qu'à 30 % de l'énergie nécessaire à la formation du complexe FGF-2/HS, le reste étant apporté par d'autres types de liaisons. Cela signifie donc que les interactions protéines/HS ne sont pas exclusivement basées sur des interactions de charges globales, mais impliquent également des liaisons de Van der Waals et hydrogène (Thompson *et al.*, 1994 ; Proudfoot, 2006).

En plus des interactions électrostatiques, il a été établi que les interactions HS/protéine impliquent des motifs saccharidiques avec des motifs de sulfatation définis, principalement situés dans les domaines S et enrichis en disaccharide trisulfaté [IdoA(2S), GlcNS(6S)] (Xu et Esko, 2014). Ainsi, les domaines S se distinguent par la conformation qu'ils adoptent. En effet, ils peuvent former une symétrie hélicoïdale double où les trois groupements sulfates sont sur des faces opposées de l'axe hélicoïdal. Cette conformation permet donc aux protéines de se lier aux deux côtés de la chaîne saccharidique. De plus, la plasticité du cycle iduronate d'IdoA(2S) peut modifier la localisation spatiale du groupe carboxyle et 2-O-sulfaté, ce qui entraîne diverses caractéristiques structurales avec différentes spécificités et affinités pour les ligands des HS (Gallagher, 2015). Quant à certains ligands dimériques, comme l'IL-8, il a été montré que l'interaction ne peut se faire que si la chaîne saccharidique présente deux domaines S séparés par un domaine de transition (ou NS) de taille définie. Dans ce cas, grâce à la flexibilité du domaine de transition, les deux interactions IL-8 monomère/domaine S sont rendues possibles (Spillmann *et al.*, 1998).

Les sulfatations des HS ne semblent pas avoir le même degré d'importance dans les propriétés de spécificités de substrat. En effet, les N-sulfatation et le motif IdoA(2S) se trouvent dans la plupart des motifs des protéines de liaison aux HS (Maccarana *et al.*, 1993 ; Lyon *et al.*, 1994 ; Kreuger *et al.*, 2001 ; Jemth *et al.*, 2002). Il ne peut donc constituer un facteur discriminant pour une interaction, et nécessite la présence

d'autres types de sulfatation, comme les motifs 6-O-sulfatés et 3-O-sulfatés qui sont moins abondants (ceci est particulièrement vrai pour le motif 3-O-sulfatés).

L'identification de motifs saccharidiques spécifiques pour une interaction donnée demeure encore très difficile à réaliser. Pour l'interaction ATIII/HP aucune séquence consensus n'est définie, même si une très forte spécificité est observée pour les interactions impliquant des glucosamines 3-O-sulfatées. De la même manière, aucune séquence saccharidique consensus n'est définie pour la reconnaissance du FGF-2, bien que la présence de groupements 6-O-sulfatés soit indispensable pour l'activation de cette protéine (et une chaîne de HS formant un coude (Raman *et al.*, 2003)). Toutefois, les interphases de 20 protéines de liaison à l'héparine ont été analysées à la recherche d'un motif structural conservé, et un motif, le motif CPC clip, a été défini faisant intervenir un résidu polaire et deux résidus cationiques. Il a été proposé que ce motif CPC clip fonctionnait comme une agrafe, et était un contributeur principal dans la fixation de l'héparine et d'autres GAGs sulfatés aux protéines de liaison (Torrent *et al.*, 2012).

Le motif CPC clip est bien défini dans les structures cristallines de CCL5 et de CXCL12, et semble être préservé dans l'ensemble des familles de chimiokines CCL et CXCL. Ainsi, parce que le motif CPC clip s'est avéré décrire correctement les sites de liaison à l'héparine des chimiokines, alors ceci a permis l'obtention d'indices supplémentaires sur la caractérisation des motifs discontinus. Mais une limitation majeure dans l'identification des domaines de liaison aux GAGs réside dans le faible degré de conservation de la séquence ; et aux divers sites de reconnaissance qui peuvent exister au sein d'une protéine, par exemple les chimiokines sont connues pour oligomériser ces sites. Des limitations sont également observées du côté des GAG, où malgré la présence fréquente d'une préférence pour un GAG, d'autres GAGs sont également reconnus, engendrant un scénario de modèles à faible conservation et de spécificité limitée.

En résumé, un ensemble important de preuves montre que les GAG sulfatés exercent leur influence en interagissant avec d'autres composants de la MEC, principalement des protéines, et que le degré de sulfatation des GAG, en particulier leur charge négative, est le principal moteur de ces interactions (Soares da Costa *et al.*, 2017). Ainsi, la sulfatation peut influencer la communication entre GAGs et protéines : soit indirectement, par exemple en régulant le repliement des protéines *via* l'encombrement

stérique, l'exclusion ou le recrutement, soit directement, par le biais d'interactions électrostatiques souvent spécifiques à la séquence.

Pour la suite de cette thèse, l'angle d'étude se focalisera sur la 6-O-sulfatation des HS. Cette sulfatation est importante pour l'interaction de nombreuses protéines aux GAGs, et est la cible de régulation exercée par la famille d'endosulfatases Sulfs.

#### **1.4.1. Les facteurs de croissance et la 6-O-sulfatation**

Bien que les différentes sulfatations des HS interviennent dans divers processus biologiques, la contribution du groupement 6-O-sulfaté dans l'activité des HS a été clairement illustrée dans le modèle d'interaction FGF/HS (Gallagher, 2015 ; Li *et al.*, 2016).

Les FGFs, représentant une large famille (23 facteurs de croissance), se lient aux HS. Ce sont grâce aux études réalisées sur FGF1 et FGF2 qu'il a été établi que les HS jouaient un rôle de co-récepteur indispensable à l'activité de ces facteurs de croissance (Rapraeger *et al.*, 1991 ; Yayon *et al.*, 1991). FGF1 et FGF2 nécessitent tous les deux des résidus GlcNS et IdoA (2S) pour se fixer aux HS (Maccarana *et al.*, 1993 ; Faham et Hileman, 1996 ; DiGabriele *et al.*, 1998). En ce qui concerne la 6-O-sulfatation, elle est essentielle pour une interaction optimale entre HS et FGF1 (Ishihara, 1994 ; Kreuger *et al.*, 2001 ; Ashikari-Hada *et al.*, 2004), mais n'est pas indispensable pour l'interaction avec FGF2. Cependant, l'induction de l'activité du FGF2 nécessite une interaction avec des oligosaccharides plus longs (10 à 12 unités de sucre) et présentant des groupements 6-O-sulfatés (Pye *et al.*, 1998 ; Sugaya *et al.*, 2008). Ainsi, ces résultats ont mis en évidence la 6-O-sulfatation comme le déterminant essentiel pour discriminer les propriétés interactives des HS et leurs fonctions biologiques. Dans le cas du FGF2, en plus du motif d'interaction à la protéine, l'activité biologique nécessite une extension saccharidique et une 6-O-sulfatation définissant un site de liaison supplémentaire pour le récepteur FGF, et permettant ainsi la formation d'un complexe ternaire FGF/FGFR/HS capable de déclencher la signalisation cellulaire.

Récemment, il a pu être mis en évidence qu'un positionnement spécifique de groupements 6-O-sulfatés sur des résidus glucosamines de l'extrémité réductrice d'un oligosaccharide d'héparine était critique pour l'activation de FGF2 (Seffouh *et al.*,

2013). Ceci suggère donc que la présence, mais également la distribution de ces groupements, pouvaient contribuer aux activités biologiques des HS.

Outre cet exemple, la 6-O-sulfatation des HS a été impliquée dans l'interaction et/ou l'activation d'autres FGFs, tels que FGF4, 7, 9, 10 et 18 (Ishihara, 1994 ; Ashikari-Hada *et al.*, 2004 ; Patel *et al.*, 2008 ; Sugaya *et al.*, 2008 ; Xu *et al.*, 2012), ainsi que d'autres facteurs de croissance, dont le « Platelet Derived Growth Factor » (PDGF) (Feyzi *et al.*, 1997), HGF (Lyon *et al.*, 1994), GDN (Rickard *et al.*, 2003 ; Ai *et al.*, 2007), « Heparin binding Epidermal Growth Factor » (HB-EGF) (Cole *et al.*, 2014) et VEGF (Ono *et al.*, 1999 ; Robinson *et al.*, 2006 ; Ferreras *et al.*, 2012). Il a également été rapporté que la 6-O-sulfatation de HS module l'activité du « Transforming growth factor beta » (TGFβ). En effet, l'interaction avec les HS favorise l'activité de TGFβ1 et les groupements 6-O-sulfatés sont des déterminants structuraux importants pour cette liaison (Yue *et al.*, 2008). Ceci est en accord avec une étude récente montrant qu'une diminution de la 6-O-sulfatation des HS réduisait la réponse des fibroblastes primaires à TGFβ1 (Yue *et al.*, 2013 ; Lu *et al.*, 2014).

#### **1.4.2. Les chimiokines, les morphogènes et la 6-O-sulfatation**

La 6-O-sulfatation participe également à l'interaction avec d'autres protéines de signalisation, telles que les morphogènes et les chimiokines.

Les morphogènes sont des protéines de signalisation jouant un rôle essentiel au cours du développement embryonnaire. A l'instar des FGFs, la 6-O-sulfatation des HS module l'activité de Wnt. En effet, cette 6-O-sulfatation peut agir comme un régulateur négatif de la signalisation Wnt, qui se lie avec une affinité élevée aux HS 6-O-sulfatés, empêchant de ce fait l'interaction avec son récepteur de surface cellulaire Frizzled (Fz) (Ai *et al.*, 2003). Or, la 6-O-désulfatation catalysée par les Sulfs réduit cette affinité Wnt/HS, permettant ainsi la formation du complexe Wnt/Fz et l'induction de la réponse cellulaire (Ai *et al.*, 2003). A l'inverse, il a également été montré que la 6-O-désulfatation des HS par les Sulfs pourrait avoir un effet inhibiteur sur Wnt en facilitant sa libération et sa dégradation (Kleinschmit *et al.*, 2010).

Les chimiokines sont de petites protéines impliquées dans le développement, l'inflammation et l'immunosurveillance (Zlotnik et Yoshie, 2012). Pour ces protéines, la 6-O-sulfatation participe au mécanisme d'interaction aux HS. En effet, les HS peuvent séquestrer les chimiokines et les protéger du clivage enzymatique, ce qui augmente leur concentration locale dans la MEC pour former des gradients de chimiokine guidant les cellules migrantes, ceci permettant, par la suite, d'induire leur oligomérisation pour faciliter leur interaction avec leurs récepteurs (Sadir *et al.*, 2004 ; Monneau *et al.*, 2016). Ainsi, la contribution de la 6-O-sulfatation a été démontrée pour CXCL12 (Sadir *et al.*, 2001 ; Uchimura *et al.*, 2006 ; Roy *et al.*, 2014b), CXCL8 (Pichert *et al.*, 2012) et CXCL4 (Pempe *et al.*, 2012). Pour CXCL12 et CXCL8, une étude récente a montré que l'ajout d'un seul groupement 6-O-sulfaté à l'extrémité non réductrice d'un oligosaccharide d'héparine synthétique [IdoA (2S), GlcNS]<sub>6</sub> modifiait les propriétés d'interaction de ces chimiokines au substrat (Jayson *et al.*, 2015). De plus, il a été montré que les Sulfs diminuaient la capacité de l'héparine à se lier à la chimiokine CXCL12 *in vitro* (Uchimura *et al.*, 2006). Enfin, la structure cristallographique de CCL5 en complexe avec des disaccharides d'héparine a révélé des interactions électrostatiques entre un groupement 6-O-sulfaté et le résidu K45 de CCL5 (Shaw *et al.*, 2004).

### 1.4.3. Les autres ligands

Un dernier exemple grandement étudié est l'interaction HS/ATIII qui aura permis de caractériser les propriétés anticoagulantes des HP (Petitou *et al.*, 2003). L'interaction HP/ATIII nécessite un pentasaccharide présentant une 3-O-sulfatation de la glucosamine centrale, mais également une 6-O-sulfatation du résidu présent à l'extrémité non-réductrice du motif. La fixation à ce pentasaccharide provoque l'activation de l'ATIII en modifiant sa conformation, conduisant à l'inhibition de la thrombine et d'autres membres impliqués dans la coagulation (notamment le facteur Xa) (Zhang *et al.*, 2001).

Pour finir, il a également été constaté que la 6-O-sulfatation est impliquée dans l'interaction avec différentes protéines, comme la L-sélectine (Wang *et al.*, 2002 ; Zhang *et al.*, 2012b), l'endostatine (Blackhall *et al.*, 2003), mais aussi dans l'attachement du virus de l'hépatite E (Kalia *et al.*, 2009). Enfin, une étude a suggéré son implication dans la maladie d'Alzheimer, car la 6-O-sulfatation participe à la liaison aux peptides  $\beta$ -amyloïdes (Lindahl *et al.*, 1999).





## II. Les sulfatases et les Sulfs

### II.1. Les sulfatases

Les 6OST, au cours de la biosynthèse, permettent la 6-O-sulfatation des HS. Or, il a pu être mis en évidence que ce groupement sulfaté était important pour de nombreuses voies physiopathologiques. Cependant, ce groupement 6-O-sulfaté est également la cible d'une régulation négative réalisée par des sulfatases extracellulaires de la famille unique des Sulfs. La désulfatation des HS par les Sulfs n'implique pas d'importante modification structurale du polysaccharide, mais, en ciblant spécifiquement les groupements 6-O-sulfatés présents au cœur des domaines S, elle a de grandes conséquences fonctionnelles.

#### II.1.1. Généralités sur les sulfatases

La famille des sulfatases comprend des enzymes, eucaryotes et procaryotes, catalysant l'hydrolyse des liaisons esters sulfates à partir de substrats sulfatés tels que les stéroïdes, les glycolipides et les PGs.

Ces enzymes sont impliquées dans de nombreux processus physiologiques comme la régulation hormonale, la dégradation cellulaire et le contrôle des voies de signalisation. Les sulfatases ne partagent toutefois pas de fortes homologues de séquence (20 à 60 %). En revanche, elles présentent toutes, au sein de la région N-terminale, deux séquences hautement conservées importantes pour l'activité des sulfatases. La première séquence est un peptide de cinq acides aminés C/S-X-P-S/X-R qui commence par une cystéine. Ce résidu sera converti, au niveau post-traductionnel, en une formylglycine (FGly) essentielle pour l'activité enzymatique de par sa fonction aldéhyde (Dierks *et al.*, 1999). Chez certains procaryotes, ce résidu peut aussi résulter de l'oxydation d'une sérine (Dierks *et al.*, 1998a). La deuxième séquence, représentée par G-K-X-X-H, implique une lysine et une histidine importantes pour la catalyse des esters sulfatés (Waldow *et al.*, 1999).

Chez les eucaryotes, le gène de la « sulfatase modifying factor 1 » (SUMF1) code pour une enzyme (du même nom) indispensable à la conversion de la cystéine en FGly. SUMF1 a une importance toute particulière dans le développement de maladie telle que la

« multiple sulfatase deficiency » (MSD). En effet, une mutation du gène *SUMF1* empêche la formation du résidu FGly et entraîne l'accumulation incontrôlée de composés sulfatés du fait de l'altération de l'activité de toutes les sulfatases (Schmidt *et al.*, 1995). *SUMF2*, un paralogue de *SUMF1*, a été identifié (Cosma *et al.*, 2003). Ce dernier présente la même activité que *SUMF1*, mais avec une efficacité moindre, ce qui suggérerait qu'il est responsable de la faible activité sulfatase trouvée chez les patients atteints de MSD. Chez les procaryotes, des gènes homologues de *SUMF1* ont été trouvés. Chez ces organismes, les enzymes impliquées dans la conversion cystéine/FGly semblent être plus actives que chez les eucaryotes, car toutes les cystéines sont converties (Cosma *et al.*, 2003 ; Dierks *et al.*, 2003).

La structure cristallographique de certaines arylsulfatases (ARS) a été résolue pour certains organismes (cf. II.2. p. 81). Ces structures présentent une organisation similaire, avec une grande région N-terminale comprenant un feuillet  $\beta$  anti-parallèle (de 10 feuillets) entouré d'hélices  $\alpha$ . La région C-terminale est, quant à elle, plus petite et contient à la fois des feuillets  $\beta$  anti-parallèles et des hélices  $\alpha$ . Le site actif de l'enzyme, toujours très conservé, est présent au centre de la protéine dans une poche où se trouvent le résidu FGly et un ion métallique (Stressler *et al.*, 2016).

Notons que la nomination ARS est due à l'activité arylsulfatase des enzymes qui représente leur capacité à désulfater les composés aryles tels que le substrat 4-méthylumbelliferyl sulfate (4MUS) (Figure 6).

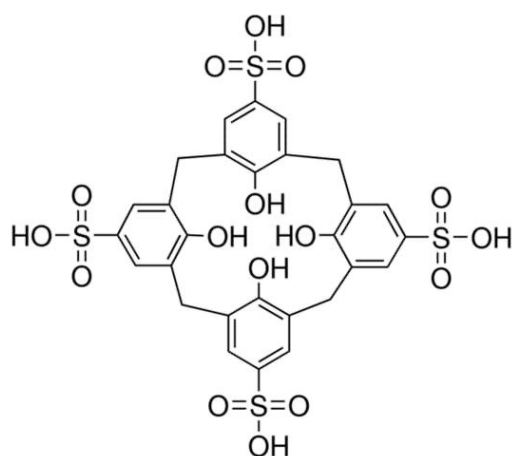


Figure 6 : Structure 4-méthylumbelliferyl sulfate (4MUS).

Contrairement aux sulfatases procaryotes, les sulfatases eucaryotes subissent des modifications post-traductionnelles. En effet, elles sont d'abord glycosylées, puis sécrétées dans les différents compartiments cellulaires ou dans la MEC après la coupure du peptide signal. L'oxydation de la cystéine arrive tardivement lors de la phase de traduction après la translocation de la sulfatase vers le RE, et avant le repliement des protéines (Dierks *et al.*, 1997, 1998b ; Fey *et al.*, 2001).

La localisation des sulfatases est corrélée à leur rôle biologique. En conséquence, elles peuvent être liées à la membrane dans l'appareil de Golgi et le RE ou bien être soluble dans le lysosome et la MEC. Les sulfatases procaryotes se retrouvent également sous une forme soluble dans le cytoplasme ou le périplasma et présentent une variété d'activités, comme des exosulfatases ciblant l'héparine (Raman *et al.*, 2003 ; Myette *et al.*, 2009a, 2009b), ou bien des exosulfatases agissant sur les résidus GlcNAc 6-O-sulfatés des HS ou sur les N-acétyl galactosamines 6-O-sulfatés des CS/DS (Ulmer *et al.*, 2014). En outre, il est notable de signaler que cette dernière étude a également identifié la première endosulfatase bactérienne. Cette enzyme élimine les groupements sulfates en position 4-O- des disaccharides CS/DS. Cette activité endosulfatase rend cette enzyme unique et différente de la CS 4-O-sulfatase caractérisée chez *P. vulgaris*, qui n'est elle active que sur l'extrémité réductrice des oligosaccharides jusqu'aux hexasaccharides, ainsi que de la N-acétylgalactosamine-4-O-sulfatase (ARSB) qui est une exosulfatase éliminant les groupements sulfates présents uniquement sur l'extrémité non réductrice des GAGs (Ulmer *et al.*, 2014). Il est également possible de noter que, par rapport aux endosulfatases eucaryotes, cette enzyme diffère et ce notamment en raison de l'absence du domaine HD présent chez les enzymes Sulfs.

Les sulfatases lysosomales sont importantes pour la dégradation des GAGs et des glycolipides. Les sulfatases lysosomales agissent à pH acide. En effet, l'arylsulfatase A (ARSA) présente la capacité de désulfater des substrats comme les sulfatides, en particulier les sphingolipides (Roy, 1975). Le défaut de l'activité des ARSA entraîne l'accumulation de sulfatide et est à l'origine de la maladie génétique de la leucodystrophie métochromatique (MLD). D'autres sulfatases lysosomales telles que l'ARSB et la glucosamine-6-sulfatase (G6S) hydrolysent les esters de sulfate en position 4 des résidus GalNAc trouvés dans les DS et CS (Matalon *et al.*, 1974), et catabolisent les esters de sulfate en position 6 de GalNAc des DS/CS, ainsi que les groupements 6-O-

sulfatés trouvés sur les résidus galactoses des KS (Bielicki et Hopwood, 1991), respectivement.

Contrairement aux sulfatases lysosomales, les sulfatases liées au RE et à l'appareil de Golgi fonctionnent à un pH proche de la neutralité. Cette famille de sulfatases est composée des ARSC, ARSD, ARSF et ARSG pour les sulfatases du RE, et de l'ARSE pour la sulfatase de l'appareil de Golgi. Fait intéressant, aucune d'entre elle n'agit sur les GAGs. L'ARSC est un ARS majeur (aussi connu sous le nom de stéroïde sulfatase) qui est présent dans de nombreux tissus (Ballabio *et al.*, 1987 ; Yen *et al.*, 1987 ; Stein *et al.*, 1989), y compris le placenta où sa fonction principale est de participer à la biosynthèse des œstrogènes (Kuss *et al.*, 1994). Son activité a également été démontrée sur des substrats comme le sulfate d'idothyronine, retrouvé principalement lors du développement fœtal (Richard *et al.*, 2001).

Les sulfatases extracellulaires correspondent à la famille des enzymes appelées Sulfs qui se sont révélées importantes de par leurs rôles dans la régulation des propriétés structure/fonction des HS impliqués dans de nombreux processus de signalisation cellulaire. Contrairement aux autres exosulfatases, les Sulfs sont les seules enzymes eucaryotes qui présentent une activité endosulfatase, mais aussi les seules à afficher un domaine basique hydrophile supplémentaire appelé HD. En outre, exception faite des Sulfs dont la séquence contient plus de 800 acides aminés, toutes les sulfatases sont constituées de 500 à 600 acides aminés.

### **II.1.2. Sulfs : la découverte d'une enzyme de régulation**

La première enzyme, QSulf-1, appartenant à la famille des Sulfs, a été découverte en 2001 chez la caille (Dhoot *et al.*, 2001). Cette dernière a été classifiée comme enzyme de la famille des sulfatases de par l'homologie de sa région N-terminale avec les sulfatases lysosomales. De surcroît, QSulf-1 s'est révélée avoir de nombreux orthologues retrouvés chez la souris, le rat, le poussin, *C. elegans*, le poisson zèbre et chez l'Homme. Chez l'Homme, deux gènes distincts ont été découverts. Bien que présentant une organisation structurale commune (Morimoto-Tomita *et al.*, 2002), ces derniers codent pour deux protéines distinctes : HSulf-1 et HSulf-2 (Morimoto-Tomita *et al.*, 2002).

Les Sulfs sont sécrétées dans le milieu extracellulaire et présentent une activité arylsulfatase sur un substrat sulfaté fluorescent (4MUS), ainsi qu'une activité endosulfatase sur les chaînes d'HS/HP (Morimoto-Tomita *et al.*, 2002). De plus, les Sulfs se sont révélées être impliquées dans la régulation de différentes voies de signalisation, la première découverte étant celle du morphogène Wnt. Pour finir, elles sont uniques de par leur organisation structurale, avec la présence d'un domaine HD dans la région C-terminale.

#### II.1.2.1. Sulfs : organisation des domaines

Les Sulfs sont initialement synthétisées sous une forme immature contenant une séquence signal clivable de 22 et 24 acides aminés pour HSulf-1 et HSulf-2, respectivement. Une fois cette séquence clivée, la protéine forme un corps protéique de 98 kDa qui va subir un clivage supplémentaire par une protéase, la furine (Tang et Rosen, 2009). La furine est une protéase à sérine qui est présente et active dans plusieurs compartiments cellulaires comme le réseau trans-Golgien, la surface cellulaire et l'endosome. Cette dernière clive au niveau d'une séquence basique : -Arg-X-(Arg/Lys)-Arg- où X représente un résidu quelconque. Or, deux sites de ce type existent chez les Sulfs (cf. p.68), conduisant à une protéine mature composée de deux sous-unités : une première de 59 kDa et une deuxième de 39 kDa, liées par un ou plusieurs ponts disulfures (cf. p. 84-85) (Frese *et al.*, 2009 ; Tang et Rosen, 2009) (Figure 7). Ces deux sous-unités se composent de trois domaines : le domaine catalytique (CAT), le domaine basique/hydrophile (HD) et le domaine C-terminal.



Figure 7 : Représentation schématique des Sulfs. Le domaine C-terminal sera nommé C-ter dans l'ensemble des figures du manuscrit. Le **trait rouge** représente le résidu FGly indispensable à l'activité arylsulfatase de la protéine (cf. p. 67). Le trait noir correspond aux ponts disulfures liant les deux sous unités de la protéine (cf. II.3.1. p. 81).

- Le domaine CAT est composé de 392 et 391 acides aminés pour HSulf-1 et HSulf-2, respectivement. Ce domaine partage une forte homologie de séquence avec toutes les Sulfs (~ 80 %), mais également avec d'autres sulfatases. Il comprend le site actif de l'enzyme, les deux séquences signatures très conservées des sulfatases ainsi que le résidu FGly strictement conservé et essentiel pour l'activité arylsulfatase. Notons que ce résidu est en position 65 et 64 pour HSulf-1 et HSulf-2, respectivement.
- Le domaine C-terminal composé de 130 à 148 résidus pour HSulf-1 et HSulf-2, respectivement, est défini comme étant une continuité du domaine CAT. Cette région présente une homologie de séquence avec la G6S et la GlcNAc transférase d'*Arabidopsis thaliana*, suggérant une implication de ce domaine dans la reconnaissance des motifs glucosamines (Morimoto-Tomita *et al.*, 2002).
- Le domaine HD, comme indiqué précédemment, est caractéristique des Sulfs en comparaison des autres sulfatases, mais aussi de toutes les autres protéines. Ce domaine est hydrophile et très basique : il comprend 327 acides aminés pour HSulf-1, dont 23 % de résidus basiques et 14 % de résidus acides, et 307 acides aminés pour HSulf-2, dont 21 % de résidus basiques et 13 % de résidus acides. En termes d'activité,

le domaine HD n'est pas impliqué dans l'activité arylsulfatase mais est responsable de la reconnaissance et de la liaison aux substrats HS. Ainsi, ce domaine est indispensable pour l'activité endosulfatase de l'enzyme, c'est-à-dire la capacité à catalyser la désulfatation des groupements 6-O-sulfatés des HS (Morimoto-Tomita *et al.*, 2002 ; Ai *et al.*, 2006 ; Frese *et al.*, 2009 ; Tang et Rosen, 2009). Toutefois, le domaine HD présente peu d'homologie de séquences, tant entre les enzymes (43 % entre les domaines HD de HSulf-1 et HSulf-2) qu'avec les orthologues (souris, caille ...), ce qui suggérerait l'existence d'une spécificité de substrat propre à chaque protéine Sulf.

Une fois Sulf sécrétée, elle exerce son activité sur les HS présents à la fois à la surface cellulaire et dans la MEC. Comme préalablement indiqué la fixation des Sulfs sur la surface cellulaire est due à l'interaction entre le domaine HD et les HS. En effet, il a été démontré par FACS que le traitement à l'héparinase de fibroblastes empêche la fixation du domaine HD de HSulf-1 à la surface cellulaire (Frese *et al.*, 2009). Par ailleurs, la suppression du domaine HD de HSulf-1 entraîne la libération de l'enzyme dans le milieu extracellulaire.

Toutefois, la reconnaissance et l'interaction HD/HS ne sont pas exclusives. En effet, un traitement à l'héparinase des cellules exprimant QSulf-1 s'est révélé insuffisant pour libérer l'enzyme dans le milieu extracellulaire (Dhoot *et al.*, 2001). Une autre étude a également démontré que les Sulfs se lient préférentiellement à leurs substrats (HP ou HS), mais que HSulf-1 était capable de se lier faiblement aux CS/DS, principalement *via* son domaine CAT (Milz *et al.*, 2013).

### II.1.2.2. Sulfs : modifications post-traductionnelles

Comme mentionné précédemment, pour que la protéine Sulf devienne mature, celle-ci subit plusieurs modifications post-traductionnelles.

Une première modification, commune à toutes les sulfatases et essentielle à l'activité arylsulfatase, est la conversion cystéine<sub>64</sub>/FGly. Le résidu FGly ainsi généré est indispensable à l'activité enzymatique. Le mécanisme détaillé de la désulfatation des HS par les Sulfs n'est pas clairement défini et se base à l'heure actuelle sur le mécanisme général de l'activité arylsulfatase des sulfatases. Dans ce cadre, deux mécanismes ont été suggérés : (1) le résidu FGly sert directement comme un électrophile formant ainsi un



diester de sulfate *via* un mécanisme d'addition-hydrolyse, (2) le résidu FGly agit comme un aldéhyde-hydrate par un mécanisme d'élimination de la trans estérification (Hanson *et al.*, 2004).

L'enzyme subit aussi un clivage à la furine, une endoprotéase cellulaire qui catalyse la maturation de nombreuses protéines sécrétées. Les Sulfs présentent deux sites consensus pour le clivage furine situés dans la région interne du domaine HD (Nagamine *et al.*, 2010). Pour HSulf-2, le clivage se réalise après la dernière arginine du site R<sub>511</sub>-S-I-R<sub>514</sub> et correspond à la séquence cible la plus courante de la furine. La seconde séquence, R<sub>536</sub>-N-L-T-K-R<sub>541</sub>, voit également une coupure après la dernière arginine mais cette séquence semble être plus rarement ciblée. Le rôle de la furine dans la maturation de HSulf-2 n'est pas clair. En effet, la mutation du premier site bloque partiellement le clivage enzymatique, tandis que la mutation du second ne semble pas affecter ce clivage. Néanmoins, avec une mutation des deux sites, le clivage est complètement bloqué (Nagamine *et al.*, 2010). De surcroît, l'analyse de la protéine en western blot a permis de mettre en évidence la présence de deux bandes à ~ 50 kDa, suggérant ainsi une action de la furine sur les deux sites du domaine HD (Tang et Rosen, 2009). Cependant, la protéine non clivée est également détectée dans le milieu extracellulaire et son activité, sa sécrétion, ou sa solubilité ne semblent pas affectées. En revanche, sa localisation à la surface cellulaire est modifiée, diminuant ainsi la signalisation Wnt (Tang et Rosen, 2009). Il a également été montré que la suppression du site de clivage de la furine n'affectait pas la transduction du signal FGF2 (Frese *et al.*, 2009). Pour finir, l'absence du clivage de la furine pourrait augmenter la possibilité de la protéine à dimériser. Cela peut être dû à la sous-unité C-terminale qui aurait tendance à se dimériser, contrairement à la sous-unité N-terminale. Toutefois, de plus amples études seront nécessaires pour clarifier le rôle du clivage furine dans la maturation des Sulfs.

Les Sulfs comportent enfin une dizaine de sites potentiels de N-glycosylation. Ces N-glycosylations sont importantes pour les protéines car elles contribuent à leur stabilité, et elles interviendraient aussi dans la localisation et l'activité enzymatique, comme il l'a été montré pour QSulf-1 (Ambasta *et al.*, 2007). Les sites potentiels de N-glycosylation sont situés principalement sur le domaine N-terminal, et la N-glycosylation représente environ 15 à 20 % de la masse moléculaire des protéines Sulfs. En effet, des analyses de chromatographie d'exclusion couplée avec la mesure de la diffusion de la lumière sous

multiangles (SEC-MALS) et de spectrométrie de masse (MS) ont pu mettre en évidence environ 14 kDa et 19,61 kDa de N-glycosylations pour HSulf-1 et HSulf-2, respectivement (données non publiées, Rana El Masri, HAL Id : tel-03507533, version 1). De surcroît, grâce à des analyses de MS réalisées sur HSulf-2, certaines asparagines porteuses de N-glycans ont pu être déterminées sur le domaine CAT : N<sub>88</sub>, N<sub>108</sub>, N<sub>147</sub>, N<sub>174</sub> et N<sub>217</sub> (Seffouh *et al.*, 2019). Pour finir, de très récents travaux permettent de porter un tout nouveau regard sur la protéine HSulf-2. En effet, une étude du laboratoire à laquelle j'ai participé, a montré que la protéine HSulf-2 est en réalité un CSPG, car cette enzyme porte une chaîne polysaccharidique liée à son domaine HD (El Masri *et al.*, 2022). La présence de cette chaîne de GAG a pu être démontrée par un traitement de l'enzyme à la chondroïtinase ABC qui modifie radicalement ses propriétés hydrodynamiques (cf. II.3.2. p. 86). Les mêmes effets ont également été observés suite à la mutation des deux motifs dipeptidiques SG (S<sub>584</sub>G et S<sub>559</sub>G) présents au niveau du domaine HD de HSulf-2 qui constituent les sites d'ancrage des chaînes de GAGs sur les PGs (Esko et Zhang, 1996). De plus, des mutations uniques du premier et deuxième site SG ont permis de localiser la chaîne de CS (~ 25 kDa) sur le résidu S<sub>559</sub>G de l'enzyme (El Masri *et al.*, 2022).

La présence de cette chaîne sur HSulf-2 redéfinit entièrement notre compréhension de son activité biologique. Ainsi, les travaux de El Masri *et al.* montrent que cette chaîne de GAG régule l'activité de l'enzyme *in vitro*, mais également dans un modèle *in vivo* de progression tumorale (El Masri *et al.*, 2022 et cf. II.3.2. p. 86).

### II.1.2.3. Sulfs : catalyseurs de la 6-O-désulfatation

Les Sulfs catalysent la 6-O-désulfatation des unités disaccharides trisulfatées [UA (2S), GlcNS (6S)] situées au sein des domaines S. Il a également été observé que ces enzymes pouvaient cibler dans une moindre mesure les disaccharides disulfatés [UA, GlcNS (6S)] (Ai *et al.*, 2003 ; Staples *et al.*, 2011 ; Pempe *et al.*, 2012 ; Seffouh *et al.*, 2013). En revanche, les Sulfs ne présentent aucune spécificité pour les disaccharides disulfatés contenant un GlcNAc ou des disaccharides monosulfatés, ce qui suggère un besoin des groupements N-sulfatés.

Une fois l'interaction entre l'enzyme et le polysaccharide réalisée, des zones d'ombres persistent sur le processus de 6-O-désulfatation. Cependant, ce dernier peut être décrit de la manière suivante : le domaine HD permet de lier l'enzyme à l'HS avec une forte affinité. En effet, il s'est avéré que le domaine HD se lie à l'héparine avec un  $K_d$  dans la gamme du nanomolaire, conduisant à la formation de complexes enzyme-substrat stables (Seffouh *et al.*, 2019). Notons que cette interaction s'est avérée plus complexe qu'un simple modèle de liaison 1:1. Ceci est en adéquation avec plusieurs études suggérant que la formation d'un tel complexe impliquerait plusieurs sites de liaison au sein du domaine HD (Frese *et al.*, 2009 ; Milz *et al.*, 2013).

Une étude impliquant une analyse de l'interaction HSulf-1 HD/HS par « Atomic Force Microscopy » a pu démontrer que cette liaison impliquait plusieurs sites d'interaction et était de type « catch-bond » (sous l'action de force allant de 10 à 18 pN), c'est à dire présentant une durée de vie de dissociation plus importante lorsqu'elle est soumise à des forces externes. Cette étude a également pu établir que le phénomène d'interaction était bien associé aux groupements 6-O-sulfatés, car d'autres substrats dépourvus de ces groupements présentaient un comportement de type « slip » (pas de « catch-bond ») sous toute la gamme de force (Harder *et al.*, 2015).

Une fois le complexe formé, la désulfatation débute de l'extrémité non-réductrice des domaines S et se dirige vers l'extrémité réductrice du domaine de manière processive (Seffouh *et al.*, 2013) (Figure 8). Cela impliquerait que, pour présenter la première glucosamine 6-O-sulfatée au site actif, le domaine HD se lierait à un motif saccharidique en aval de cette première glucosamine. Pour la processivité de la réaction, plusieurs modèles ont été proposés et définissent les groupements 6-O-sulfatés comme des effecteurs allostériques induisant des états conformationnels de transition, permettant de toujours présenter un groupement 6-O-sulfaté au site catalytique sans complètement dissocier l'interaction Sulf/HS (Seffouh *et al.*, 2013 ; Harder *et al.*, 2015). Une fois l'extrémité réductrice du domaine S atteinte, et en absence de groupements 6-O-sulfatés, l'affinité du domaine HD pour la chaîne de HS diminuerait drastiquement afin de libérer l'enzyme. Notons également que le domaine CAT présente des épitopes de liaison aux HS permettant des interactions supplémentaires de faibles affinités. Ceux-ci contribueraient à la bonne présentation du polysaccharide au site actif (Seffouh *et al.*, 2019).

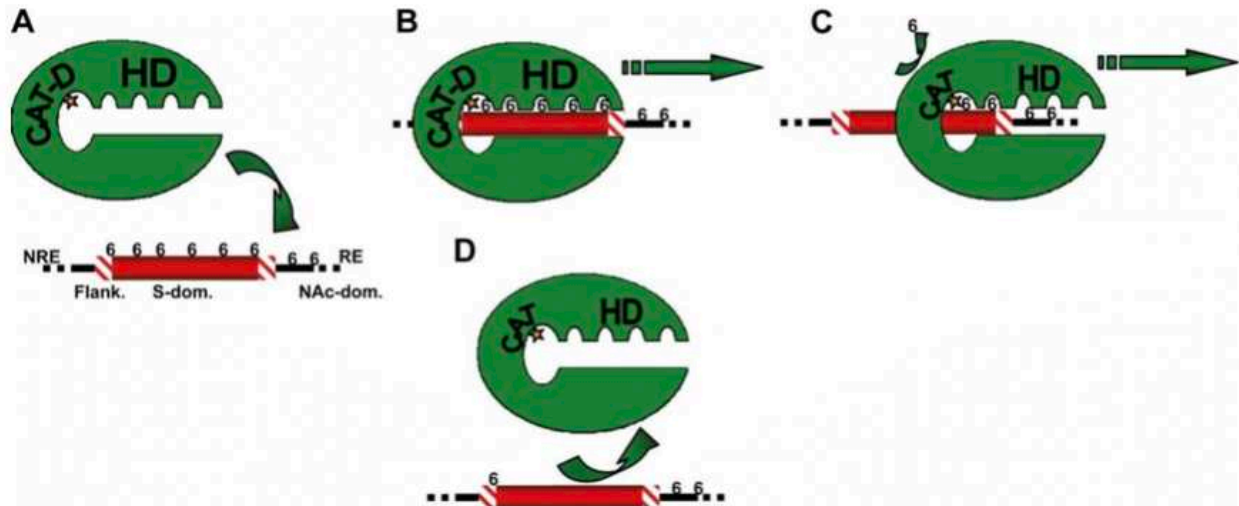


Figure 8 : Modèle de l'activité processive de HSulf et de la régulation des propriétés de liaison aux HS. HSulf (vert) reconnaît et se lie, par son domaine HD, préférentiellement aux 6-O-sulfates portés sur les disaccharides trisulfatés présents dans les domaines S (des HS) (A), ce qui permet au site catalytique (CAT-D) de l'enzyme d'être positionné en face du résidu 6-O-sulfaté à l'extrémité non-réductrice du motif (B). Après désulfatation, l'enzyme progresse le long de la chaîne polysaccharidique pour accueillir d'autres groupements 6-O-sulfatés (C). Une fois l'extrémité réductrice du domaine S atteinte, l'absence de 6-O-sulfatés appropriés (car ceux présents sur la « flanking » région (Flank) ou sur le domaine NS/S (NAc-dom.) sont de mauvais substrats pour l'enzyme), en aval de CAT-D, se traduit par une forte diminution de l'affinité du domaine HD pour le polysaccharide, ce qui engendre la dissociation du complexe. (Figure : Seffouh *et al.*, 2013).

#### II.1.2.4. Sulfs : régulateurs de l'interaction HS/Ligand

Les motifs trisulfatés substrats des Sulfs ne représentent que ~ 5 à 10 % des disaccharides des HS. Ainsi, d'un point de vue structural, l'action des Sulfs n'a que peu d'influence. En revanche, du fait de la présence de ces motifs au sein des domaines S fonctionnels du polysaccharide, les Sulfs modifient considérablement les propriétés d'interactions des HS et par conséquent leur fonction. Certains ligands voient leur capacité à lier les HS régulée favorablement ou défavorablement.

Les Sulfs inhibent une variété d'interaction comme les facteurs de croissance tels que FGF1 (Uchimura *et al.*, 2006 ; Seffouh *et al.*, 2013), FGF2 (Lai *et al.*, 2003 ; Wang *et al.*, 2004 ; Li *et al.*, 2005 ; Narita *et al.*, 2006), HGF (Lai *et al.*, 2004a ; Narita *et al.*, 2006), HB-EGF (Lai *et al.*, 2003 ; Dai *et al.*, 2005), amphiréguline (Narita *et al.*, 2007) ou TGFB

(Yue *et al.*, 2008). Ici, la régulation négative est la conséquence d'une altération structurale des HS de la surface cellulaire agissant comme récepteur/co-récepteur pour ces facteurs de croissance. En effet, un exemple pertinent est celui du FGF-2. Pour le FGF-2, la 6-O-sulfatation du HS n'est pas requise pour la liaison, mais elle est nécessaire pour la formation d'un complexe ternaire FGF/FGFR/HS capable de déclencher la signalisation cellulaire. Ainsi, l'activité des Sulfs pourrait réguler l'activité du FGF-2 en catalysant l'élimination rapide des HS 6-O-sulfatés, permettant, de ce fait, l'interaction et la séquestration du facteur de croissance tout en inhibant la réponse cellulaire.

L'action des Sulfs pourrait avoir des conséquences directes pour des protéines telles que les chimiokines dont l'activité dépend de la formation d'un gradient protéique stabilisé par des interactions avec les HS. En effet, l'inhibition de l'interaction des HS avec CXCL12 et CCL21 par HSulf-2 a été démontrée *in vitro* (Uchimura *et al.*, 2006). De plus, étant donné que de nombreuses chimiokines nécessitent des groupements 6-O-sulfatés pour se lier efficacement aux HS, il peut être attendu que les Sulfs modulent d'autres interactions chimiokines/HS, avec des conséquences fonctionnelles également importantes sur le stockage des chimiokines, la protection contre la dégradation protéolytique ou l'oligomérisation (El Masri *et al.*, 2020).

Les Sulfs peuvent aussi induire des voies de signalisation comme la voie Wnt (Ai *et al.*, 2003). Comme expliqué précédemment, Wnt se lie avec une grande affinité aux HS 6-O-sulfatés qui agissent comme un régulateur négatif du morphogène en empêchant l'accès à son récepteur de surface cellulaire Fz. Or, la modification des motifs de 6-O-sulfatation des HS par les Sulfs entraîne une réduction de l'affinité de l'interaction HS/Wnt, ce qui permet la formation d'un complexe fonctionnel HS/Wnt/Fz.

Depuis, des mécanismes d'activation similaires ont été rapportés pour d'autres protéines se liant aux HS, notamment le GDNF (Langsdorf *et al.*, 2011) et la BMP (Otsuki *et al.*, 2010). En revanche, une autre étude a également suggéré que la 6-O-désulfatation des HS, catalysée par Sulf, pourrait avoir un effet inhibiteur sur Wnt en facilitant sa libération et sa dégradation (Kleinschmit *et al.*, 2010). Ces activités contradictoires illustrent clairement la complexité de la régulation des HS par les Sulfs.

Grâce à leurs capacités à modifier les interactions HS/protéines, les Sulfs sont aussi fortement impliquées dans de nombreux processus physiopathologiques tels que le

développement, la régénération tissulaire/l'homéostasie et le cancer. En effet, l'importance des Sulfs au cours du développement a été mise en évidence dans des études, avec des rôles rapportés durant le développement neuronal (Danesin *et al.*, 2006 ; Ai *et al.*, 2007 ; Kalus *et al.*, 2009, 2015), le développement du squelette et du cartilage (Zhao *et al.*, 2006 ; Holst *et al.*, 2007), la formation de l'oreille interne (Freeman *et al.*, 2015) et la dentinogenèse (Hayano *et al.*, 2012). En effet, il a été montré que des souris double KO *Sulf-1* et *Sulf-2* présentent une mortalité néonatale élevée et de graves anomalies morphologiques, y compris des anomalies rénales, pulmonaires, squelettiques et/ou neuronales. Cependant, un KO unique *Sulf-1* ou *Sulf-2* n'entraîne que des phénotypes relativement peu altérés, suggérant des fonctions chevauchantes/redondantes des deux enzymes au cours du développement (Lamanna *et al.*, 2006).

En revanche, des différences majeures entre *Sulf-1* et *Sulf-2* ont été rapportées dans le contexte du cancer. *Sulf-1* a été principalement associée à des effets anti-oncogènes, tandis que *Sulf-2* présente des activités pro-oncogènes (Rosen *et al.*, 2010 ; Vivès *et al.*, 2014). Par conséquent, il n'est à ce jour pas encore clairement défini comment ces deux enzymes peuvent présenter des activités similaires *in vitro*, mais des propriétés antagonistes au cours du développement tumoral. Des différences dans les spécificités du substrat *Sulf-1/Sulf-2*, les niveaux d'expression et/ou la distribution tissulaire pourraient cependant expliquer ces propriétés antagonistes.

#### II.1.2.5. Régulation des Sulfs

Du fait de l'implication des Sulfs dans de nombreux processus physiopathologiques, différents mécanismes permettent une régulation de leur expression.

Des travaux ont tout d'abord mis en évidence un contrôle de l'expression des Sulfs par des cytokines. Ainsi, la cytokine TGF $\beta$ 1 peut induire l'expression de *Sulf-1* dans les fibroblastes pulmonaires, probablement car non identifiés, grâce à des suppresseurs de transcription nouvellement synthétisés. Ainsi, une fois *Sulf-1* exprimée, l'enzyme peut agir comme un régulateur négatif de l'activité TGF $\beta$ 1 (Yue *et al.*, 2008), l'activité du TGF $\beta$ 1 dépendant des groupements N-, 2-O- ou 6-O-sulfatés des HS (Lyon *et al.*, 1997).

Des mécanismes épigénétiques semblent également jouer un rôle important dans le contrôle de l'expression de Sulf. En effet, il a été montré qu'une inhibition de l'ADN méthylase, dans le contexte d'un carcinome hépatocellulaire primaire, induisait l'expression de HSulf-1 alors que l'ARNm de HSulf-1 était initialement régulé à la baisse (Lai *et al.*, 2004a ; Staub *et al.*, 2007). De plus, une étude comparative de patients atteints du cancer du sein ou de l'estomac avec des sujets contrôles a suggéré que le promoteur de HSulf-1 subissait une hyperméthylation comme mécanisme de régulation négative de l'expression de l'enzyme dans ces tumeurs (Chen *et al.*, 2009). Il est notable de signaler qu'un mécanisme de régulation similaire a été identifié dans l'adénocarcinome pulmonaire pour le promoteur de HSulf-2 (Wang *et al.*, 2013).

Enfin, les facteurs de transcription et les gènes suppresseurs de tumeurs sont également capables de réguler l'expression de Sulf. Grâce au profil d'expression génétique et à la prédiction du site de liaison du facteur de transcription, il a été démontré que le suppresseur de tumeur p53 active la transcription de Sulf-2 (Chau *et al.*, 2009). En cas de dommage à l'ADN, le suppresseur de tumeur p53 répond en régulant directement la transcription de HSulf-2. Ainsi, quand HSulf-2 est inhibée, la sénescence des cellules due au stress diminue (Chau *et al.*, 2009). De plus, il a été signalé, dans le cancer de l'ovaire, que la suppression de l'ARNm de HSulf-1 était médiée par le variant du facteur nucléaire hépatique 1 capable de se lier au promoteur de l'enzyme (Liu *et al.*, 2009).

#### **II.1.2.6. Sulfs et cancer**

Les Sulfs ont été associées à une grande variété de cancer et semblent impliquées dans presque toutes les étapes de la maladie : la transformation tumorale (Rosen et Lemjabbar-Alaoui, 2010), la croissance, l'invasion et le développement métastatique (Li *et al.*, 2005 ; Abiatari *et al.*, 2006 ; Nawroth *et al.*, 2007 ; Phillips *et al.*, 2012 ; Khurana *et al.*, 2012a, 2013a). Elles semblent aussi responsables de la sensibilité/résistance accrue des cellules tumorales aux médicaments (Lai *et al.*, 2003 ; Moussay *et al.*, 2010). Comme préalablement discuté, HSulf-1 et HSulf-2 présentent de manière générale des activités antagonistes, et HSulf-2 en particulier représente une cible thérapeutique d'intérêt pour certains cancers agressifs.

- Sulf-1

Dans le carcinome hépatocellulaire (HCC), Sulf-1 agit comme un suppresseur de tumeur. En effet, il a été démontré que la surexpression de Sulf-1 dans le cancer HCC murin entraînait une régulation négative de la mésothéline au niveau d'ARNm et de protéines dans les cellules HCC surexprimant HSulf-1. La mésothéline est une glycoprotéine de la surface cellulaire qui activerait la voie de signalisation de la prolifération des protéines « Mitogen-activated protein kinases » (MAPK), favorisant ainsi la croissance tumorale (Chang *et al.*, 2009). Il a été suggéré que Sulf-1, en supprimant les groupements 6-O-sulfatés des HSPGs tels que les syndécans, inhibait la formation du complexe HS/Wnt-1 en empêchant de ce fait la libération et l'induction de mésothéline et donc du signal de prolifération. Ainsi, il a été observé, *in vitro*, une diminution de la prolifération, de la division, de la migration et de l'invasion cellulaires et, *in vivo*, une réduction de la progression tumorale et des métastases ganglionnaires (Mahmoud *et al.*, 2016). De plus, la ré-expression de HSulf-1 dans les cellules HCC HSulf-1 négatives et HCC xéno greffes diminue la phosphorylation des « sérine/threonine-protein kinase » (AKT) et des « extracellular signal-regulated kinases » (ERK), mais aussi inverse l'action des FGFs exogènes, régulant à la baisse leurs gènes cibles, la *cycline* et la *survivin*. Ceci entraîne l'inhibition de la prolifération cellulaire et l'induction de l'apoptose (Liu *et al.*, 2013 ; Xu *et al.*, 2014).

HSulf-1 a également été liée à la régulation de la voie de signalisation MAPK par le biais de mécanismes épigénétiques. En effet, chez la souris, la transfection de cellules HCC avec HSulf-1 induit l'acétylation de l'histone nucléaire H4 en modifiant l'équilibre entre les activités de l'histone désacétylase (HDAC) et de l'histone acétyl transférase (HAT). Ceci entraîne donc une régulation négative de la phosphoinositide-3-kinase (PI3) et de la MAPK kinase qui conduit à la réduction de la croissance tumorale et à l'apoptose. De surcroît, HSulf-1 a favorisé le rôle des inhibiteurs de HDAC pendant la croissance tumorale et l'angiogenèse, tant *in vitro* que *in vivo* (Lai *et al.*, 2006).

HSulf-1 a aussi été identifiée comme un gène régulé à la baisse dans plusieurs types de tumeurs, notamment les carcinomes de l'ovaire, du sein, et hépatocellulaires. La perte de HSulf-1 augmente les interactions entre les facteurs de croissance (comme FGF-2 et VEGF165) et les HS, conférant une résistance à l'apoptose induite par la chimiothérapie.



Cependant, une surexpression de HSulf-1 dans les lignées cellulaires MDA-MB-468 (carcinome du sein) conduit à une prolifération réduite *in vitro* et à une charge tumorale réduite chez les souris *in vivo*. Une densité de vaisseaux réduite, une nécrose marquée et une apoptose, indiquant une inhibition de l'angiogenèse, ont été aussi constatées (Narita *et al.*, 2006).

De nombreux autres exemples existent. Le niveau d'ARNm de HSulf-1 est réduit dans le cancer du poumon (Zhang *et al.*, 2012a), le cancer gastrique (Ma *et al.*, 2011), le carcinome épidermoïde (SCCHN) (Lai *et al.*, 2004b) et le cancer ovarien primaire (Lai *et al.*, 2003). Cependant, grâce à une ré-expression de HSulf-1 au sein de ces cellules, des voies de régulation sont réduites ou supprimées. Pour exemple, la ré-expression de Sulf-1 dans les cellules SCCHN est à l'origine de la 6-O-désulfatation des HS permettant la diminution des voies de signalisation FGF2 et HFG (Lai *et al.*, 2004b), ou la suppression des voies de signalisation HH (Ma *et al.*, 2011) et Wnt/ $\beta$ -caténine (Li *et al.*, 2011) dans les cellules cancéreuses gastriques. Par conséquent, ces altérations permettent la diminution de la prolifération cellulaire et l'augmentation de l'apoptose.

- Sulf-2

A l'instar de Sulf-1, Sulf-2 intervient dans de nombreux cancers. En effet, l'expression de *HSulf-2* a été identifiée et étudiée chez des patients atteints de carcinome rénal (RCC). Il a été montré que le niveau d'ARNm de *HSulf-2* était significativement plus élevé dans les tissus cancéreux que dans les tissus normaux. Cependant, au cours de l'avancement du cancer, l'expression de *HSulf-2* devient moins importante dans les tissus cancéreux que dans les tissus normaux. L'augmentation de l'ARNm peut être due à une réaction excessive des tissus normaux contre le cancer afin de provoquer une inflammation et/ou le début du développement de métastases dans les tissus normaux. En effet, il a été démontré que de faibles niveaux d'expression de *HSulf-2* dans les cellules RCC améliorent l'activation des voies de signalisation VEGF et FGF, *in vitro*, et permettent la croissance tumorale et métastatique (Kumagai *et al.*, 2016).

Dans les tumeurs primaires du glioblastome humain (GBM), l'expression de *Sulf-2* est élevée tandis que l'expression de *Sulf-1* est inchangée. L'inhibition de l'expression de

*Sulf-2* par les cellules cancéreuses GBM conduit à une inhibition de la croissance tumorale qui peut être rétablie par la surexpression de *Sulf-2* au sein de ces cellules (Johansson *et al.*, 2005). De plus, une étude *in vivo* chez la souris a montré que la protéine Sulf-2 était fortement exprimée dans les gliomes invasifs, ce qui entraînait une croissance tumorale et une létalité plus importante. Ceci pourrait être dû à l'activation, par Sulf-2, de nombreux RTK dans les gliomes invasifs (Phillips *et al.*, 2012).

De plus, la protéine Sulf-2 active la signalisation Wnt dans les cellules tumorales de l'adénocarcinome pancréatique *in vitro*. Au contraire, une diminution de Sulf-2 inhibe la croissance tumorale *in vivo*. Ces données suggèrent donc que la protéine Sulf-2 favorise la croissance tumorale en induisant la signalisation de Wnt (Nawroth *et al.*, 2007 ; Kumagai *et al.*, 2016).

Cependant, HSulf-2 semble avoir un comportement plus complexe et plus interactif avec son micro-environnement, que d'être « simplement » un des acteurs intervenant dans la croissance tumorale et métastatique dans de nombreux cancers. En effet, une étude a montré que la surexpression de *HSulf-1* et *HSulf-2* au sein de cellules de myélome chez des souris présentant un déficit immunitaire combiné sévère (SCID) conduisait à la répression de la croissance et la progression tumorale, suggérant donc une influence du micro-environnement tumoral sur les Sulfs. Ceci est d'autant plus vrai qu'*in vitro*, cet effet suppresseur n'a pas été observé (Dai *et al.*, 2005).

Pour finir, une étude faisant intervenir des patients atteints de cancer de l'œsophage a montré, grâce à des analyses immunohistochimiques, que le taux de survie était corrélé avec une expression élevée de la protéine Sulf-2. Ainsi, Sulf-2 pourrait être considérée comme une enzyme biomarqueur intéressante pour le diagnostic et la prévention de ce type de cancer agressif (Lui *et al.*, 2012).

#### II.1.2.7. Sulfs et autres pathologies

Une augmentation de l'expression de *Sulf-1* et *Sulf-2* a également été signalée dans l'osteoarthrite. Dans l'arthrose, une activation anormale des chondrocytes et une dégradation du cartilage peuvent résulter de l'action des Sulfs sur les HS, modifiant ainsi

leur structure et propriété d'interaction et impactant de ce fait la voie de signalisation de nombreux facteurs de croissance.

Sulf-2 a aussi été associée au diabète de type 2. En effet, l'enzyme inhibe l'interaction entre des lipoprotéines de très basse densité avec les HS des hépatocytes (Hassing *et al.*, 2012). Plus récemment, le polymorphisme génétique de Sulf-1 a été associé à un échec de la fécondation *in vitro* (FIV) (Taghizadeh *et al.*, 2015). Pour finir, il a été suggéré que les Sulfs avaient des effets protecteurs dans des conditions pathologiques telles que la maladie d'Alzheimer et l'amylose rénale. Dans ces maladies, l'accumulation de domaines S des HS dans les plaques  $\beta$  rénales et amyloïdes cérébrales favorise la formation de fibrilles amyloïdes toxiques. Le remodelage enzymatique des HS par les Sulfs peut donc empêcher cette accumulation (Kameyama *et al.*, 2019).

De plus, il s'est avéré que Sulf-1 contribue au développement du disque intervertébral et maintient son homéostasie (Otsuki *et al.*, 2019). Toutefois, une autre étude a montré que Sulf-1 était également exprimée dans les cellules dégénératives du disque intervertébral (Tsai *et al.*, 2015). Par conséquent, le lien entre développement et maintien de l'homéostasie avec la dégénérescence du cartilage reste à clarifier. Ces résultats contradictoires soulignent cependant parfaitement le rôle complexe des Sulfs suivant les nombreuses voies de régulation où elles interviennent.

Pour finir, les Sulfs semblent être impliquées dans les mécanismes de réponse inflammatoire. En effet, il peut être rappelé que la 6-O-sulfatation est critique pour la liaison de nombreuses chimiokines et que les Sulfs peuvent donc réguler ces interactions comme cela a été démontré *in vitro* pour CXCL12 et CCL21 (Uchimura *et al.*, 2006). La 6-O-sulfatation des HS est également requise pour l'interaction avec la L-sélectine qui intervient dans les premiers événements de fixation à l'endothélium précédant la diapédèse des leucocytes (Wang *et al.*, 2002), ainsi que l'interaction avec CXCL12 qui régit le trafic des leucocytes (Zhang *et al.*, 2012b). Ceci suggère que les Sulfs pourraient agir comme des modulatrices de la migration et de l'adhésion des leucocytes pour activer l'endothélium. D'autres études font état de la régulation de l'expression et du rôle des Sulfs, comme la répression de Sulf-1 dans l'inflammation engendrée au cours d'une transplantation rénale (Celie *et al.*, 2007). Sulf-2 a également été impliquée dans la fibrose pulmonaire idiopathique *via* une régulation de la signalisation TGF $\beta$ 1 dans les

cellules épithéliales alvéolaires de type 2 (Yue *et al.*, 2013). En effet, l'expression de Sulf-2 dans ces cellules jouerait un rôle important dans la protection contre les lésions pulmonaires épithéliales, l'inflammation, et la mortalité (Yue, 2017).

#### II.1.2.8. HSulf-1 et HSulf-2 : récapitulatif des différences

Les Sulfs sont encore des enzymes très mal caractérisées, tant dans leurs fonctions de régulation des différentes voies de signalisation dans lesquelles elles sont impliquées et des maladies qui peuvent y être associées, que dans les différences fonctionnelles et structurales entre les deux protéines de HSulf. Ainsi, des différences existent dans leurs modes d'expressions, leurs activités *in vitro* et leurs implications *in vivo*.

En termes d'expression des enzymes, HSulf-1 est fortement exprimée dans les testicules, l'estomac, les muscles squelettiques, les poumons et les reins. Quant à HSulf-2, cette dernière est principalement exprimée dans les ovaires, les muscles squelettiques, l'estomac, le cerveau, l'utérus, le cœur, les reins et le placenta (Morimoto-Tomita *et al.*, 2002).

HSulf-1 et HSulf-2 partagent la même organisation moléculaire comprenant les deux domaines CAT et HD. Le domaine CAT présente des résidus strictement conservés dans le site actif, y compris la cystéine modifiée FGly. Néanmoins, une légère différence existe dans la conservation des épitopes de reconnaissance et d'interaction au substrat présents sur le domaine CAT. En effet, chacune des deux séquences voit un acide aminé modifié suivant l'enzyme considérée de HSulf, avec V<sub>155</sub>KEK et L<sub>377</sub>KKK pour HSulf-2 mais I<sub>155</sub>KEK et L<sub>377</sub>NKK pour HSulf-1. De plus, le domaine HD, responsable pour HSulf-1 et -2 de l'interaction au substrat, n'est pas aussi conservé que le domaine CAT entre les deux HSulf. Mais, la différence majeure entre les deux enzymes reste l'absence, pour HSulf-1, de la chaîne de GAG présente sur le domaine HD de HSulf-2 (El Masri *et al.*, 2022). Ces différences pourraient être à l'origine des activités antagonistes entre les deux formes car responsables de spécificités de substrats propres à chacune des enzymes.

Il est vrai que des études indiquent que HSulf-1 et HSulf-2 ne ciblent pas la même population d'HS (Morimoto-Tomita *et al.*, 2002 ; Lamanna *et al.*, 2006 ; Milz *et al.*, 2013). HSulf-1 a tendance à se lier et à rester sur les HS présents sur la surface

cellulaire, tandis que HSulf-2 est plus rapidement libérée dans le compartiment extracellulaire et pourrait ainsi agir préférentiellement sur les HS de la MEC.

*In vitro*, les deux enzymes ont toutes deux un mécanisme catalytique processif. Ainsi, l'analyse de l'activité *in vitro* de HSulf-1 et HSulf-2 montre que ces dernières peuvent catalyser la 6-O-désulfatation complète d'un oligosaccharide d'héparine, mais avec des efficacités différentes (Seffouh *et al.*, 2013). Ceci suggère donc que le domaine HD agit comme un régulateur clé de l'activité enzymatique et des préférences de substrats entre HSulf-1 et HSulf-2, et ce probablement dû au fait que ce domaine est le moins conservé entre les Sulfs et que ce dernier est responsable de la reconnaissance et de l'interaction aux substrats.

*In vivo*, des études sur des souris KO Sulf ont également fourni des indications sur ces différences entre les deux enzymes. En effet, des souris simple KO *Sulf-1* ou *Sulf-2* ne présentent aucune anomalie grave, alors que des souris double KO sont caractérisées par un phénotype marqué et une mortalité néonatale élevée (Holst *et al.*, 2007 ; Lum *et al.*, 2007). Ainsi, il a été déterminé une certaine redondance fonctionnelle des enzymes Sulf-1 et Sulf-2 au cours du développement. Toutefois, cela est nuancable car une expression accrue de Sulf-1 pourrait compenser la perte de Sulf-2, alors que Sulf-2 ne pourrait pas complètement remplacer l'absence de Sulf-1 (Lamanna *et al.*, 2006).

Cependant, il a également été mis en évidence une non-redondance des fonctions Sulf-1 et Sulf-2 au cours du développement du cerveau de la souris, car les deux enzymes contribuent différemment à la prolifération des neurites des neurones cérébelleux et hippocampiques, à la plasticité synaptique et à l'activité motrice (Kalus *et al.*, 2009).

Pour finir, et comme expliqué auparavant, des différences majeures entre les activités Sulf-1 et Sulf-2 ont également été signalées dans le cancer.

Ces différences et similitudes entre les deux enzymes pourraient provenir des spécificités subtiles de substrat et/ou de leurs actions sur différentes populations d'HS. Sachant que les HS agissent comme des co-récepteurs des facteurs de croissance, ou sont impliqués dans la séquestration et le stockage de ces facteurs, chaque spécificité de substrat peut générer des activités et des fonctions radicalement différentes.

## II.2. Caractérisation structurale des sulfatases

A ce jour, un certain nombre de sulfatases ont vu leur structure résolue, en complexe, ou non, avec leur substrat (Hwa *et al.*, 1992 ; Sugahara *et al.*, 1996 ; Bond *et al.*, 1997 ; Lukatela *et al.*, 1998 ; Boltes *et al.*, 2001 ; Hernandez-Guzman *et al.*, 2003 ; Mayette *et al.*, 2003, 2009a, 2009b ; Ulmer *et al.*, 2014 ; Wang *et al.*, 2015, 2019 ; Ndeh *et al.*, 2020 ; Luis *et al.*, 2021). Les structures des ARS sont très similaires et révèlent un monomère globulaire, comprenant deux domaines (N- et C-terminal) avec une topologie mixte d'hélices  $\alpha$  et de feuillets  $\beta$ . Toutefois, il est notable de signaler qu'ARSC présente un domaine transmembranaire unique auquel se rattache le domaine soluble (Hernandez-Guzman *et al.*, 2003).

Le domaine N-terminal est composé d'un grand feuillet  $\beta$  entouré de plusieurs hélices  $\alpha$ , dont une grande hélice  $\alpha$  exposée au solvant. Le domaine C-terminal, plus petit, contient plusieurs feuillets  $\beta$  antiparallèles, mais aussi différentes hélices  $\alpha$ . Le site actif est, quand à lui, présent au sein d'une étroite fente de l'extrémité N-terminale du grand feuillet  $\beta$ , avec le résidu FGly situé au bas de cette dernière et bordé d'acides aminés chargés.

Par ces études, des informations sur le site actif des sulfatases ont pu être mises en évidence : l'organisation spatiale des résidus catalytiques et la géométrie sont étroitement conservées, ce qui soutient la notion d'un mécanisme conservé pour l'hydrolyse des esters sulfates. De surcroît, le site actif est composé de 10 résidus polaires fortement interconnectés et d'un cation métallique divalent (Tableau 2).

Residue <sup>[a]</sup>	PARS	HARSA	HARSB	HARSC	Proposed function
FGly	51	69	91	75	catalytic nucleophile as FGH
M	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Ca <sup>2+</sup>	substrate binding and activation, stabilization of FGH
AsnA	Asn318	Asn282	Asn301	Gln343	metal coordination, activation of FGH
AspA	Asp317	Asp281	Asp300	Asp342	metal coordination
AspB	Asp14	Asp30	Asp54	Asp36	metal coordination
AspC	Asp13	Asp29	Asp63	Asp35	metal coordination
ArgA	Arg55	Arg73	Arg95	Arg79	stabilization of FGH
HisA	His115	His125	His147	His136	stabilization of FGH, elimination of FGS
HisB	His211	His229	His242	His290	substrate binding and activation, alcohol protonation
LysA	Lys113	Lys123	Lys145	Lys134	substrate binding and activation, stabilization of FGH
LysB	Lys375	Lys302	Lys318	Lys368	substrate binding and activation, alcohol protonation

Tableau 2 : Positions et fonctions proposées des acides aminés et des cations métalliques dans le site actif. PARS, HARSA, HARSB et HARSC sont des enzymes. Le tableau doit être mis en relation avec la figure 9 (cf. p. 82) afin de comprendre l'organisation du site actif et la fonction proposée pour chaque acide aminé (Figure : Hanson *et al.*, 2004).

Dans le cas de HARSA, une région de liaison à l'ion  $Mg^{2+}$  est coordonnée de manière octaédrique par les résidus AspA, AspB, AspC et AsnA (Waldow *et al.*, 1999). Cependant, si l'un des résidus liant l'ion  $Mg^{2+}$  est muté, alors de sévères diminutions de l'activité catalytique et d'interaction au substrat sont observées, ce qui souligne le rôle central que joue le métal dans le clivage des esters sulfate (vraisemblablement en liant le substrat et en l'activant pour une attaque nucléophile) (Waldow *et al.*, 1999). Pour HARSB, HARSC et PARS, la coordination heptavalente du cation  $Ca^{2+}$  est médiée par les résidus liant l'ion ainsi que le résidu FGly et un atome d'oxygène provenant du sulfate (ou de l'eau en son absence).

De surcroît, dans la poche catalytique (Figure 9), plusieurs résidus chargés positivement, dont LysA et LysB, sont impliqués dans la liaison au sulfate anionique. Avec l'ion métallique divalent et HisB, ces résidus neutralisent probablement l'anion sulfate et retirent la densité électronique pour créer un centre de soufre électrophile. Egalement, ils positionnent le sulfate tétraédrique pour une attaque nucléophile par FGly (von Bülow *et al.*, 2001). De plus, HisB et LysB sont impliqués dans la protonation de l'atome d'oxygène exocyclique.

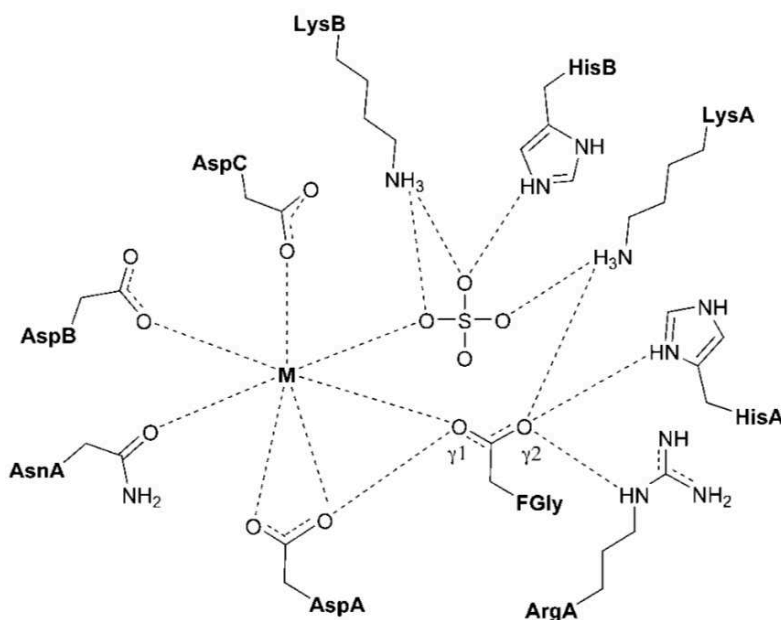


Figure 9 : Interprétation générique du site actif de la sulfatase représentant les résidus hautement conservés retrouvés dans les structures résolues. Les principales interactions (liaisons hydrogènes) sont représentées par des lignes pointillées. La lettre M fait référence au cation métallique divalent, cf. Tableau 2, ligne 2, p. 81. (Figure : Hanson *et al.*, 2004).

Quant au résidu catalytique FGly, celui-ci repose au fond de la poche sous forme d'hydrate de FGly (FGH). Le diol géminale remplit deux fonctions distinctes : FGH-O $\gamma$ 1, qui interagit avec le centre métallique et AspA, qui est orienté idéalement pour l'attaque nucléophile de l'atome de soufre du substrat. FGH-O $\gamma$ 2, qui est ancré par HisA, ArgA et LysA, pointe vers le centre de la poche et est impliqué dans l'élimination de l'intermédiaire enzyme-ester sulfate (Hanson *et al.*, 2004).

D'après différentes études publiées sur les sulfatases dont les Sulfs (Seffouh *et al.*, 2013 ; El Masri *et al.*, 2017), la reconnaissance des substrats naturels se produit en dehors de la fente contenant les résidus conservés du site actif, et plus précisément dans la région C-terminale de l'enzyme. En effet, cette région est structurellement très diverse au sein de la famille des sulfatases, ce qui génère une spécificité dans la reconnaissance et l'interaction enzyme/substrat. C'est également le cas de HSulf avec son domaine HD.

### **II.3. Récents progrès dans la caractérisation structurale et fonctionnelle de HSulf-2**

Malgré un intérêt croissant, de nombreux aspects des Sulfs restent peu voire pas connus comme leurs structures, leurs mécanismes catalytiques et leurs spécificités de substrat. En termes de structure, quelques informations ont, tout de même, pu être mises en évidence.

#### **II.3.1. Informations structurales sur HSulf-2**

Pour rappel, la protéine HSulf-2 porte environ 20 kDa de N-glycosylations et il s'avérera que ces dernières sont indispensables pour le maintien de la protéine HSulf-2 en solution (cf. p. 160-161).

De précédentes analyses de dichroïsme circulaire (CD) ont pu établir que la protéine HSulf-2  $\Delta$ SG (sans chaîne de CS, car le site de fixation de la chaîne de CS (S<sub>559</sub>G) a été muté) était partiellement structurée (Tableau 3). En effet, d'après ces analyses, les domaines CAT-C-terminal sont majoritairement structurés alors que le domaine HD est largement non structuré.

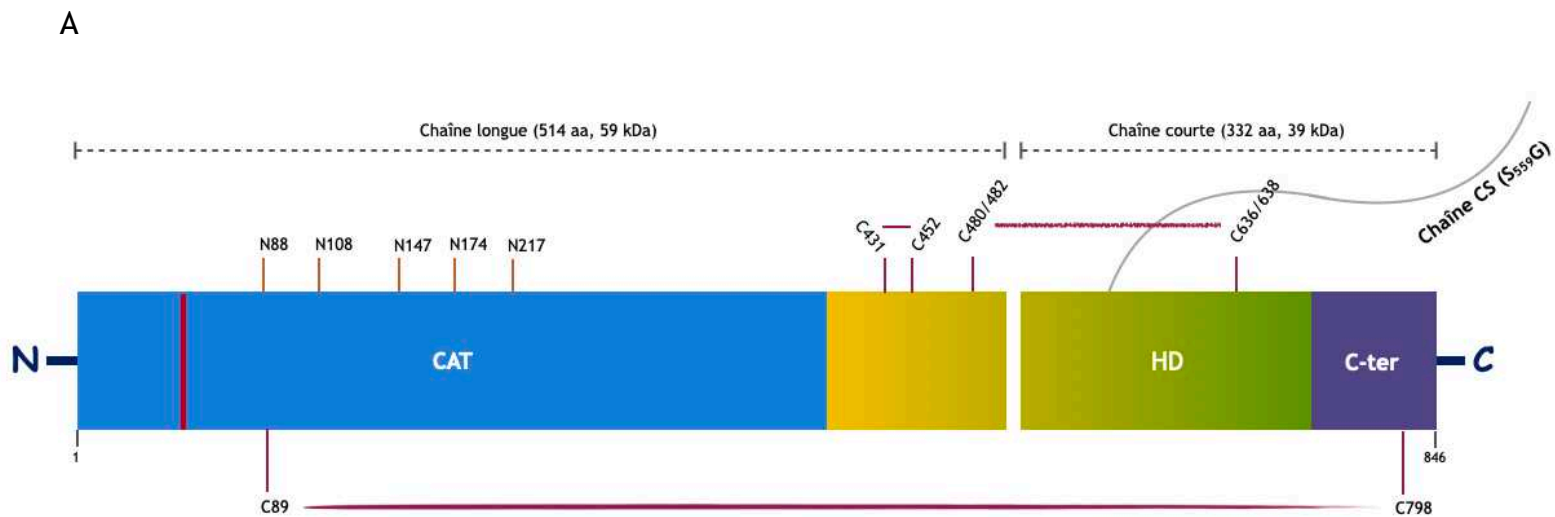


	<b>HSulf-2 ΔSG (sans chaîne de CS)</b>	<b>HSulf-2 ΔHD (Domaines CAT et C-terminal)</b>	<b>Domaine HD (sans chaîne de CS)</b>
<b>Helice α</b>	20 %	19,2 %	12,8 %
<b>Feuillet β</b>	35 %	39,3 %	31 %
<b>Boucle</b>	6 %	2,7 %	1,9 %
<b>Aléatoire (Non structurée)</b>	38,2 %	38,7 %	54 %

Tableau 3 : Données de CD montrant le taux de structuration de différentes formes de la protéine HSulf-2 (Figure adaptée du manuscrit de thèse de Rana El Masri, HAL Id : tel-03507533, version 1).

Ceci étant, des analyses par MS (Figure 10A et cf. Annexe 1A p. 233) ont permis de mettre en évidence quatre ponts disulfures intra-domaine HD. Les acides aminés impliqués sont : C<sub>422</sub> et C<sub>440</sub>, C<sub>431</sub> et C<sub>452</sub>, C<sub>480</sub> et C<sub>636</sub>, C<sub>482</sub> et C<sub>638</sub>. Par conséquent, les deux ponts disulfure C<sub>422</sub> et C<sub>440</sub>, C<sub>431</sub> et C<sub>452</sub> sont situés avant le clivage furine, ce qui n'entraîne, vraisemblablement, que peu d'influence sur la structuration du domaine HD et/ou de la protéine. En revanche, les deux autres ponts, reliant les deux parties du domaine HD clivé, devraient davantage influencer la structuration de ce domaine et de l'enzyme. Pour finir, ces analyses ont mis en évidence un dernier pont disulfure entre les domaines CAT et C-terminal (C<sub>89</sub> et C<sub>798</sub>), suggérant très fortement un repliement et une interaction entre ces deux domaines (Figure 10B).

Des études menées au laboratoire (antérieurement à la thèse) ont montré que le domaine CAT comprend, en plus de la présence du site actif de l'enzyme, deux sites de liaison aux HS : V<sub>155</sub>KEK et L<sub>377</sub>KKK (Figure 10B). En effet, ces deux épitopes s'alignent avec le site actif situé dans un sillon intermédiaire et peuvent participer à l'activité en guidant et en permettant une présentation correcte du polysaccharide au résidu FGly catalytique (Seffouh *et al.*, 2019). De surcroît, il a été montré que le domaine CAT seul présentait une activité de type exosulfatase sur des oligosaccharides, suggérant que le domaine HD n'est pas strictement requis pour la reconnaissance des HS, mais est essentiel pour une désulfatation efficace et processive du polysaccharide (Rana El Masri, HAL Id : tel-03507533, version 1).



B

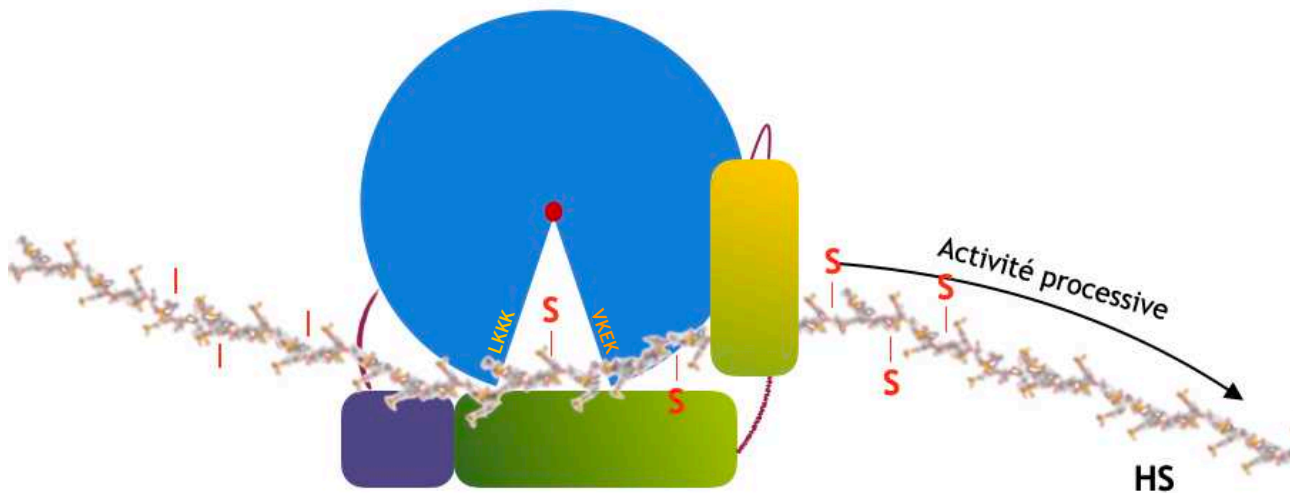


Figure 10 : PTMs identifiées et modèle structural hypothétique de HSulf-2. (A) Séquence HSulf-2 annotée, ici est représenté le domaine CAT (bleu) avec le site actif (résidu C<sub>64</sub>/FGly, trait rouge), le domaine hydrophobe HD (jaune/vert) et le domaine C-terminal (violet), ainsi que les ponts disulfures entre cystéines, la chaîne de CS (ancrée sur le domaine HD, S<sub>559</sub>G) et les différents sites de N-glycosylations identifiés. (B) Ici est schématisé le modèle structural hypothétique de HSulf-2 (forme holo et désulfatation de groupements 6-O-sulfatés (notés S)), avec une organisation globulaire de la protéine, notamment, grâce au repliement et à l'interaction du domaine C-terminal avec le domaine CAT, médié par le pont disulfure C<sub>89</sub>-C<sub>798</sub>. Notons également la présence des deux sites de basse affinité aux HS alignés au site actif (VKEK et LKKK).

*Remarque : dans l'ensemble du manuscrit, la numérotation de la séquence en acides aminés ne prend pas en compte le peptide signal.*

Par la suite, des analyses RMN sur le domaine HD ont été entreprises. Pour cela, le domaine HD a été segmenté en quatre peptides d'environ 10 kDa, et les protéines correspondantes exprimées en bactéries. Ainsi, dans un premier temps, ces données ont permis de confirmer la faible structuration du domaine HD. De plus, elles ont aussi pu identifier la présence de structures secondaires dans ses parties N- et C-terminales (régions portant les cystéines identifiées dans la formation des ponts disulfures). De surcroît, ces analyses RMN ont mis en évidence les résidus impliqués dans l'interaction enzyme/substrat, et ces derniers semblent être repartis en différents sites sur la quasi-totalité du domaine HD (analyses d'identification et d'attribution en cours).

### II.3.2. Rôle biologique de la chaîne de CS

Des analyses de la protéine HSulf-2 WT par diffusion des rayons X aux petits angles (SAXS) et par chromatographie d'exclusion couplée avec la mesure de la diffusion de la lumière sous multiples angles (SEC-MALS) ont montré que la protéine présentait une forme étendue, pour une masse moléculaire apparente d'environ 700 kDa. La détermination *ab initio* de la forme moléculaire des protéines a suggéré la présence de deux entités au sein de la protéine : un domaine globulaire, qui pourrait correspondre à la taille attendue de HSulf-2 WT, et une partie flexible allongée correspondant très probablement à la chaîne de CS fixée au domaine HD.

Afin d'étudier le rôle fonctionnel de cette chaîne de GAG, des tests enzymatiques ont été réalisés sur la protéine HSulf-2 WT (avec GAG) et HSulf-2  $\Delta$ SG (sites SG mutés donc perte de la chaîne de CS) (El Masri *et al.*, 2022).

*In vitro*, les deux enzymes présentent toutes deux une activité arylsulfatase sur un substrat 4-MUS, bien que cette activité soit supérieure pour la protéine HSulf-2 dépourvue de chaîne de GAG. Ces résultats démontrent un effet de régulation de cette chaîne. Cette différence d'activité a aussi été observée lors de tests endosulfatases réalisés sur des substrats naturels comme l'HP et l'HS. Ces tests montrent une 6-O-désulfatation accrue du polysaccharide avec la forme HSulf-2  $\Delta$ SG. De la même manière, lors de tests « Enzyme Linked ImmunoSorbent Assay » (ELISA), la protéine HSulf-2  $\Delta$ SG s'est significativement mieux liée à l'héparine, et des analyses en FACS ont également montré une fixation accrue aux HS de la surface cellulaire de cellules « Wish ». Ces différents résultats démontrent donc le rôle régulateur de cette chaîne de CS sur les

propriétés d'interaction de HSulf-2 vis-à-vis de ses substrats. En effet, la présence de la chaîne de GAG sur le domaine HD limiterait, par répulsion électrostatique ou encombrement stérique, l'accès des substrats sulfatés au site actif.

*In vivo*, des études ont été réalisées sur la progression tumorale dans un modèle de xénogreffe orthotopique chez la souris. Pour cela, des cellules MDA-MB-231 d'adénocarcinome mammaire humain, qui n'expriment pas de HSulfs de manière endogène (Peterson *et al.*, 2010), ont été modifiées afin de surexprimer HSulf-2 WT, HSulf-2  $\Delta$ SG ou une protéine contrôle, puis implantées dans la glande mammaire de souris Squid. L'analyse des tumeurs induites a pu montrer que l'expression de HSulf-2 WT n'a peu ou pas d'influence sur la taille de la tumeur, bien que des travaux antérieurs rapportaient des effets anti-oncogènes (Peterson *et al.*, 2010) ou pro-oncogènes (Zhu *et al.*, 2016) de l'expression de HSulf-2 WT dans ce même modèle. Ces données contradictoires illustrent clairement la complexité des fonctions régulatrices de HSulf-2 et le biais résultant de différences dans la conception expérimentale de ces trois études. En revanche, les tumeurs exprimant le mutant HSulf-2  $\Delta$ SG ont crû plus rapidement et ont atteint une taille plus grande par rapport aux tumeurs contrôles et exprimant HSulf-2 WT. De plus, la vascularisation était significativement augmentée dans les tumeurs HSulf-2  $\Delta$ SG. Enfin, l'analyse de l'invasion tumorale a montré une fréquence plus élevée de métastases pour les tumeurs exprimant les HSulfs (WT ou  $\Delta$ SG) comparées aux tumeurs contrôles. Cependant, les tumeurs secondaires induites par la zone de métastase étaient significativement plus importantes dans les tumeurs exprimant HSulf-2  $\Delta$ SG. Ces données indiquent donc que la chaîne de CS nouvellement identifiée influence le rôle biologique de HSulf-2 et module la capacité de l'enzyme à promouvoir la croissance tumorale mais aussi l'invasion métastatique *in vivo*.

Toutefois, il est notable de signaler que ces différences d'activités peuvent être imputées à d'autres facteurs, comme l'interaction de la chaîne CS à d'autres partenaires protéiques présents à la surface cellulaire et la MEC, ou bien la régulation par cette chaîne de la localisation, la diffusion et la disponibilité de l'enzyme dans les tissus. De plus, et comme indiqué dans cette étude, des formes dégradées de HSulf-2 apparaissent lors de la purification de l'enzyme : formes qui ont perdu la chaîne de CS. Il a enfin été montré que les hyaluronidases bovines pouvaient digérer efficacement la chaîne de CS de HSulf-2 et améliorer son activité endosulfatase (El Masri *et al.*, 2022). Les

hyaluronidases de mammifères, qui sont une famille de six enzymes catalysant la dégradation de l'acide hyaluronique (HA), présentent également la capacité de dépolymériser les CS (Csoka *et al.*, 2001 ; Jedrzejewski et Stern, 2005 ; Kaneiwa *et al.*, 2010 ; Bilong *et al.*, 2021). De plus, l'expression de la hyaluronidase est augmentée dans certains cancers (McAtee *et al.*, 2014), avec des rôles suggérés dans l'invasion tumorale et l'inflammation associée à la tumeur (Dominguez-Gutierrez *et al.*, 2021 ; McAtee *et al.*, 2014). Cependant, leur contribution précise reste mal connue et contradictoire. Par conséquent, d'après l'étude de El Masri *et al.*, 2022 une nouvelle fonction pourrait être attribuée à ces enzymes : l'activation de HSulf-2 grâce à leur capacité à « libérer » l'enzyme de sa chaîne de GAG.

Cependant, afin d'identifier le mécanisme réel et les influences de la présence de ce GAG, de nouvelles analyses sont indispensables telle que la résolution de la structure de la protéine HSulf-2 avec et sans la chaîne de CS. Pour résoudre la structure de HSulf-2, différentes approches biophysiques peuvent être envisagées, comme la microscopie électronique, la cristallographie et la RMN. Cette dernière technique, contrairement aux deux autres, nécessite un marquage isotopique de la protéine, ce qui représente un challenge puisque la protéine est exprimée dans un système d'expression mammifère.

### III. RMN et marquage isotopique des protéines de haute masse moléculaire en système eucaryote

La RMN est une méthode de choix pour l'étude à l'échelle atomique de macromolécules biologiques en solution. Contrairement à la cristallographie, elle ne nécessite pas l'obtention de cristaux et permet d'obtenir des informations sur la structure, la dynamique et les interactions des protéines. Cette technique utilise les propriétés de résonance des atomes placés dans un champ magnétique. L'interaction entre le noyau atomique et le champ magnétique est quantifiée, et seule une fréquence particulière permet cette interaction. Cette fréquence dépend du type de noyau observé (proton, carbone, azote, etc.), mais aussi de l'environnement chimique de ce noyau, c'est-à-dire des autres noyaux présents autour, et des électrons impliqués dans la liaison chimique, ce qui signifie par exemple qu'un proton porté par un groupement méthyle ne résonnera pas à la même fréquence qu'un proton d'aldéhyde. La fréquence de résonance permet donc de remonter à la structure de la molécule. Les techniques de RMN multidimensionnelles ont permis de corrélérer les fréquences de plusieurs spins et de résoudre les ambiguïtés liées aux superpositions spectrales. Depuis sa découverte en 1945, la RMN a connu d'incroyables développements technologiques, en particulier l'augmentation du champ magnétique des aimants RMN (qui est passé des électroaimants générant un champ de 0,9 T, correspondant à une fréquence de résonance de 40 MHz pour les protons, à des aimants supraconducteurs créant un champ de 1,2 GHz aujourd'hui), ainsi que le développement de séquences d'acquisition et de méthodes de marquage isotopique innovantes.

En effet, hormis les  $^1\text{H}$ , la plupart des spins observés en RMN ne sont pas abondants naturellement. Il faut donc que l'échantillon soit enrichi isotopiquement. Ainsi, des sources isotopiques  $^{13}\text{C}$  et  $^{15}\text{N}$  peuvent être introduites pour être détectées lors de l'acquisition de spectres. Mais aussi, des sources de protons deutérés ( $^2\text{H}$ ) qui vont permettre de masquer le signal des protons  $^1\text{H}$  et décomplexifier les spectres, en particulier pour les protéines de grande taille (cf. III.3.3.3. p. 113). Les atomes  $^{13}\text{C}$  et  $^{15}\text{N}$  peuvent remplacer tous les  $^{12}\text{C}$  et  $^{14}\text{N}$  (marquage uniforme) ou seulement certains (marquage spécifique). Le choix du marquage est lié au type de données structurales attendues et à la taille de la protéine étudiée. En effet, plus la protéine est grosse, plus la résolution et la sensibilité des spectres acquis diminuent. Ainsi, un marquage

uniforme peut être appliqué à une petite protéine (jusqu'à 40-50 kDa) afin d'acquérir des données d'interactions, de dynamique et des données permettant de résoudre la structure de la protéine. En revanche, pour des protéines plus importantes (supérieures à 40-50 kDa) le marquage uniforme, bien que techniquement applicable, ne permet pas l'acquisition de spectres exploitables et informatifs. En effet, si tous les acides aminés sont marqués ( $U$ -[ $^1\text{H}$ ,  $^{15}\text{N}$ ], par exemple), alors un pic apparaît sur le spectre pour chacun des résidus, entraînant de trop nombreux chevauchements de pics et donc une perte de l'information. Par conséquent, pour analyser de telles protéines, le marquage spécifique a été développé, où seuls quelques acides aminés d'une protéine sont marqués. Il permet l'obtention, pour des protéines jusqu'à plusieurs centaines de kDa, de spectres résolus présentant peu (voire pas) de chevauchement entre les pics (Weigelt *et al.*, 2002), et permet la caractérisation des interactions protéine/ligand, la détermination approximative du site de liaison, la détermination du  $K_d$  du ligand par titration, des informations structurales et/ou la détection des changements conformationnels de la protéine (Pellecchia *et al.*, 2002). Notons toutefois que ces deux marquages peuvent être complémentaires. En effet, le marquage spécifique permet d'acquérir des contraintes de distance de longue portée (non acquises par le marquage uniforme) qui sont indispensables pour un certain nombre de problématiques structurales.

### III.1. Historique de la RMN et du marquage isotopique

Jusqu'aux années 90, la RMN homonucléaire multidimensionnelle  $^1\text{H}$  était couramment utilisée, mais limitée aux protéines de taille allant jusqu'à 10 kDa (Qian *et al.*, 1989). Pour des molécules plus grandes, la relaxation transversale des noyaux, le chevauchement des résonances, la présence de trop nombreux protons et l'élargissement du signal RMN (à cause du temps de corrélation augmenté) engendraient des spectres RMN saturés et non exploitables. Ainsi, pour étudier de telles protéines, il devenait indispensable de faire évoluer les équipements et les méthodes expérimentales. C'est ainsi que plusieurs innovations ont été développées comme les sondes cryogéniques, des spectromètres plus puissants (jusqu'à 1,2 GHz), de nouvelles séquences d'impulsion, ou encore le marquage isotopique des protéines qui permet une RMN hétéronucléaire multidimensionnelle. Par conséquent, avec cet ensemble d'innovations, la taille des systèmes d'études et la qualité des données qui en résultent ont été considérablement accrues.

Au début des années 90, les premières études RMN faisant état de l'utilisation de la RMN hétéronucléaire multidimensionnelle (où les résonances des sondes  $^{15}\text{N}$ ,  $^{13}\text{C}$  et  $^1\text{H}$  sont corrélées) ont permis une meilleure dispersion du signal et, par conséquent, des études de protéines avec des tailles légèrement plus grandes (~ 20 kDa) (Torchia *et al.*, 1988 ; Kay *et al.*, 1989). Toutefois, pour atteindre des protéines plus grandes encore (~ 40 kDa), une deutération partielle de la protéine a été nécessaire (Yamazaki *et al.*, 1994 ; Muchmore *et al.*, 1996). En effet, lorsque le deutérium ( $\text{D}_2\text{O}$ ) remplace les protons, en raison de son faible rapport gyromagnétique et de l'élimination des voies de relaxation concurrentes dans le réseau de protons (diminution des interactions dipolaires  $^1\text{H}$ - $^1\text{H}$ ), les taux de relaxation des protons restants ( $[^1\text{H}, ^{13}\text{C}]$  ou  $[^1\text{H}, ^{15}\text{N}]$ , par exemple) sont considérablement réduits, ce qui conduit à des largeurs de raie plus petites et à un rapport signal sur bruit plus élevé (c'est-à-dire une intensité de signal RMN plus élevée, un spectre mieux résolu) (Sattler *et al.*, 1996).

Cependant, une limite à la deutération des protéines est apparue lors de l'étude d'objets encore plus grands. En effet, une deutération accrue de la protéine est requise mais celle-ci, bien qu'avantageuse pour la résolution des spectres, engendre une perte d'informations (car trop de sites  $^1\text{H}$ , et donc d'interactions dipolaires  $^1\text{H}$ - $^1\text{H}$ , ont été perdues). En conséquence, afin de maintenir un taux de deutération en adéquation avec l'acquisition de spectres résolus et informatifs, une méthode alternative a été développée. Cette dernière consiste en la réintroduction de protons échangeables dans une protéine perdeutérée pour permettre la protonation partielle de la protéine. Ceci offre la possibilité d'acquérir des spectres RMN restreints mais de haute qualité. Notons que la réintroduction de protons échangeables, tels que les protons amides de la chaîne principale dans la protéine perdeutérée, est l'approche la plus couramment utilisée (Venters *et al.*, 1995 ; Garrett *et al.*, 1997).

Pour repousser encore la limite de taille de la RMN des protéines, une stratégie de marquage sélectif  $^{13}\text{C}$  du carbone couplé à une protonation sélective des groupements méthyles dans une protéine entièrement deutérée a été proposée (Rosen *et al.*, 1996 ; Gardner et Kay, 1997). Cette approche, qui repose sur la multiplicité de protons des groupements méthyles et les propriétés de relaxation favorables qui en découlent, a permis d'obtenir des spectres RMN de haute qualité pour des protéines allant jusqu'à 1 MDa. Notons toutefois que les données obtenues ne permettent pas la détermination de la structure tridimensionnelle de la protéine (bien qu'une protonation sélective



combinée des groupes amides et méthyles ait contribué à repousser la limite de taille de la détermination de la structure par RMN (monomère de 83 kDa) (Tugarinov *et al.* 2005) mais permettent, tout de même, l'obtention d'informations structurales (Sprangers et Kay, 2007a ; Velyvis *et al.*, 2007 ; Macek *et al.*, 2017), de dynamiques (Sprangers et Kay, 2007b ; Tugarinov *et al.*, 2007) et d'interactions (Sprangers *et al.*, 2007c ; Mas *et al.*, 2018).

Les acides aminés contenant des groupements méthyles (**méthionine, isoleucine, leucine, valine, thréonine, alanine**) représentent jusqu'à 40 % des acides aminés totaux d'une protéine (par exemple, chez HSulf-2 ces acides aminés représentent 33 % de la séquence peptidique). Ces derniers représentent également jusqu'à 50 % du noyau hydrophobe des protéines (Miller *et al.*, 1987). De plus, ils peuvent occuper jusqu'à 24 % des interfaces protéine-protéine (Lo Conte *et al.*, 1999). Ainsi, bien que plus de la moitié de la protéine ne fournisse aucune donnée, il a été démontré que cette stratégie de marquage permettait de résoudre des questions structurales, car une empreinte générale de la protéine était acquise (au vu de la bonne répartition des acides aminés marqués au sein de cette dernière).

Les groupements méthyles sont de bons candidats pour une protonation sélective. En effet, de par leur localisation (cœur hydrophobe et interface d'interaction), leurs protons fournissent des contraintes de distance, NOEs (Effet Overhauser nucléaire), indispensables pour le calcul du repliement des protéines. Notons que ces contraintes de distances sont de longues portées. En effet, contrairement aux NOEs NH-NH et  $^{13}\text{CH}_3\text{-NH}$  qui indiquent des contraintes entre résidus séparés de quelques acides aminés, les NOEs  $^{13}\text{CH}_3\text{-}^{13}\text{CH}_3$  impliquent des résidus dont la séparation est d'environ 30 acides aminés (Gardner et Kay, 1997 ; Kerfah *et al.*, 2015).

Ces groupements permettent une acquisition sensible des spectres et des résonances résolues (Hajduk *et al.*, 2000 ; Tugarinov et Kay, 2004 ; Gelis *et al.*, 2007) grâce à la multiplicité de protons (des groupements méthyles) qui entraîne une résonance trois fois plus intense qu'un seul proton. De plus, en raison de leur rotation rapide autour de leur axe de symétrie triple, et en raison de leur flexibilité (lorsqu'elle est possible suivant sa localisation), les groupements méthyles montrent une relaxation transversale et longitudinale plus favorable qu'un seul proton de la chaîne principale ou d'une chaîne latérale. Pour finir, les acides aminés portant un groupement méthyle résonnent dans

une région précise du spectre 2D [ $^1\text{H}$ ,  $^{13}\text{C}$ ] (Figure 11). Par conséquent, une meilleure résolution des spectres est attendue.

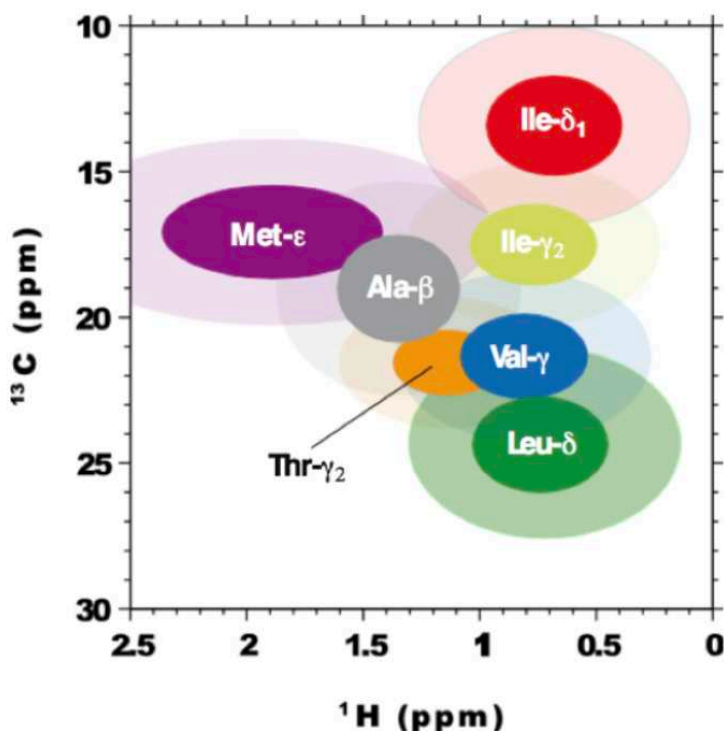


Figure 11 : Représentation schématique des régions de résonance des acides aminés portant un groupement méthyle sur un spectre RMN 2D (Figure adaptée du manuscrit de thèse de Rime Kerfah, HAL Id : tel-01343865, version 1).

Le marquage isotopique des acides aminés portant un groupement méthyle d'une protéine est réalisé par ajout de précurseurs d'acides aminés ou d'acides aminés dans le milieu de culture. Il est donc indispensable de maîtriser les voies métaboliques de biosynthèse des acides aminés car les chevauchements entre substrats et enzymes sont très fréquents, générant potentiellement des marquages indésirables d'autres acides aminés (fuite isotopique), ou bien, un faible enrichissement isotopique du résidu cible (dilution isotopique). Ceci est tout particulièrement vrai pour le système bactérien qui peut engendrer de nombreuses fuites isotopiques. En effet, les bactéries peuvent croître en milieu minimum, c'est-à-dire qu'elles peuvent synthétiser l'ensemble des biomolécules indispensables à leur croissance/fonction (dont les acides aminés à partir de glucose et d'ammonium). Les voies métaboliques sont bien régulées et largement décrites (cf. III.2. p. 94). Ceci permet donc l'utilisation de nombreux précurseurs et acides aminés libres pour le marquage isotopique (tant uniforme que spécifique) des protéines exprimées dans ce système.

Les cellules eucaryotes (insectes et mammifères), en revanche, ne peuvent croître dans un milieu minimum car elles ont perdu au cours de l'évolution les enzymes responsables

de la synthèse de nombreuses biomolécules, dont une large partie des acides aminés. En effet, ces derniers sont classifiés en deux groupes : les acides aminés dits essentiels (non synthétisés par la cellule) et non essentiels (synthétisés par la cellule) (cf. III.3.2. p. 105). Les acides aminés méthionine, isoleucine, valine, thréonine et leucine (portant un groupement méthyle) étant tous essentiels, leur ajout sous forme marquée au milieu ne générera aucune fuite isotopique. En revanche, pour l'alanine, qui est synthétisée au cours du cycle de l'acide citrique, une régulation de la voie métabolique sera indispensable afin d'éviter toute dilution de l'isotope (cf. III.3.3.2. p. 108).

### III.2. Acides aminés portant des groupements méthyles et leur méthode de marquage isotopique en bactérie

Parmi les systèmes d'expression disponibles, l'utilisation d'*Escherichia coli* (*E. coli*) est de loin le système le plus populaire et le mieux référencé. Les connaissances accumulées concernant sa génétique, son métabolisme, ainsi que son temps de génération court et sa capacité à croître en milieu minimum font d'*E. coli* un candidat robuste pour l'expression de protéines perdeutéérées marquées.

Lorsque la perdeutération de la protéine est requise (en plus du marquage des méthyles), la surexpression des protéines est induite dans un milieu de culture deutéré où la seule source de protons provient du précurseur ou de l'acide aminé spécifiquement protoné et marqué  $^{13}\text{C}$  sur le carbon du méthyle. Cela permet le marquage des résidus cibles. Le choix entre l'utilisation de précurseurs ou d'acides aminés marqués dépend de la disponibilité du précurseur/acide aminé marqué et de l'existence d'une voie métabolique permettant son incorporation efficace sans fuite isotopique (ce sujet sera largement traité dans les prochaines parties).

Le marquage des acides aminés portant un groupement méthyle, chez les bactéries, est plus ou moins compliqué suivant l'acide aminé souhaité. En effet, les acides aminés qui résultent de réactions irréversibles (Met, Ile, Leu, Val) sont facilement incorporés (si ajoutés à la culture) dans la protéine exprimée et sans aucune fuite isotopique. Cependant, les acides aminés peuvent être onéreux, voire non disponibles. Ainsi, il est plus intéressant, quand cela est possible, d'ajouter un précurseur d'acide aminé marqué (Gardner et Kay, 1997). Cette stratégie est possible si le précurseur porte un groupement  $\text{CH}_3$  et qu'il génère, après son incorporation dans la voie de biosynthèse,

des acides aminés cibles marqués et aucune fuite isotopique. Avec ces pré-requis, seuls les résidus Ile, Leu et Val sont marqués à l'aide de précurseurs  $^{13}\text{CH}_3$ , tandis que les autres sondes sont ajoutées sous forme libre.

### III.2.1. Les voies métaboliques irréversibles

L'**isoleucine** porte deux groupements  $\text{CH}_3$  aux positions delta 1 ( $\delta_1$ ) (Gardner et Kay, 1997) et gamma 2 ( $\gamma_2$ ) (Rosen *et al.*, 1996 ; Gans *et al.*, 2010b ; Ayala *et al.*, 2012), pouvant tous deux être marqués. Ce résidu est synthétisé à partir d'un précurseur, l' $\alpha$ -cétobutyrate, et ce dernier va subir cinq réactions enzymatiques (Figure 12). Dans *E. coli*, il existe trois acéto-hydroxy-acide synthases (AHAS) et ces synthases catalysent la première étape de biosynthèse commune des acides aminés Ile, Leu, et Val (Bar-Ilan *et al.*, 2001). Ces AHAS, dont l'activité dépend des substrats (Barak *et al.*, 1987 ; Steinmetz *et al.*, 2010), assurent la condensation du pyruvate avec l' $\alpha$ -cétobutyrate ou une autre molécule de pyruvate, produisant, respectivement, du 2-acéto-2-hydroxybutanoate (précurseur de l'Ile) ou de l'acétolactate (précurseur de la Leu et de la Val).

Le groupement méthyle Ile $\delta_1$ , synthétisé post activités AHAS, représente une sonde de choix pour l'analyse RMN, en raison de propriétés de relaxation longitudinales et transversales favorables, des signaux bien dispersés sur un spectre RMN, et est une sonde favorable pour les analyses NOEs (Gardner et Kay, 1997). Le marquage Ile $\gamma_2$  est moins évident. Celui-ci reste, certes, intéressant pour l'acquisition d'un signal RMN de qualité, mais les différents protocoles pour marquer ce groupement  $\gamma_2$  présentent tous des inconvénients (fuites isotopiques, synthèse de produits non métabolisés par la bactérie et faible dispersion des résonances lors d'un marquage combinatoire Ile $\gamma_2$ , Ala et Val) (Rosen *et al.*, 1996 ; Gans *et al.*, 2010b ; Ayala *et al.*, 2012).

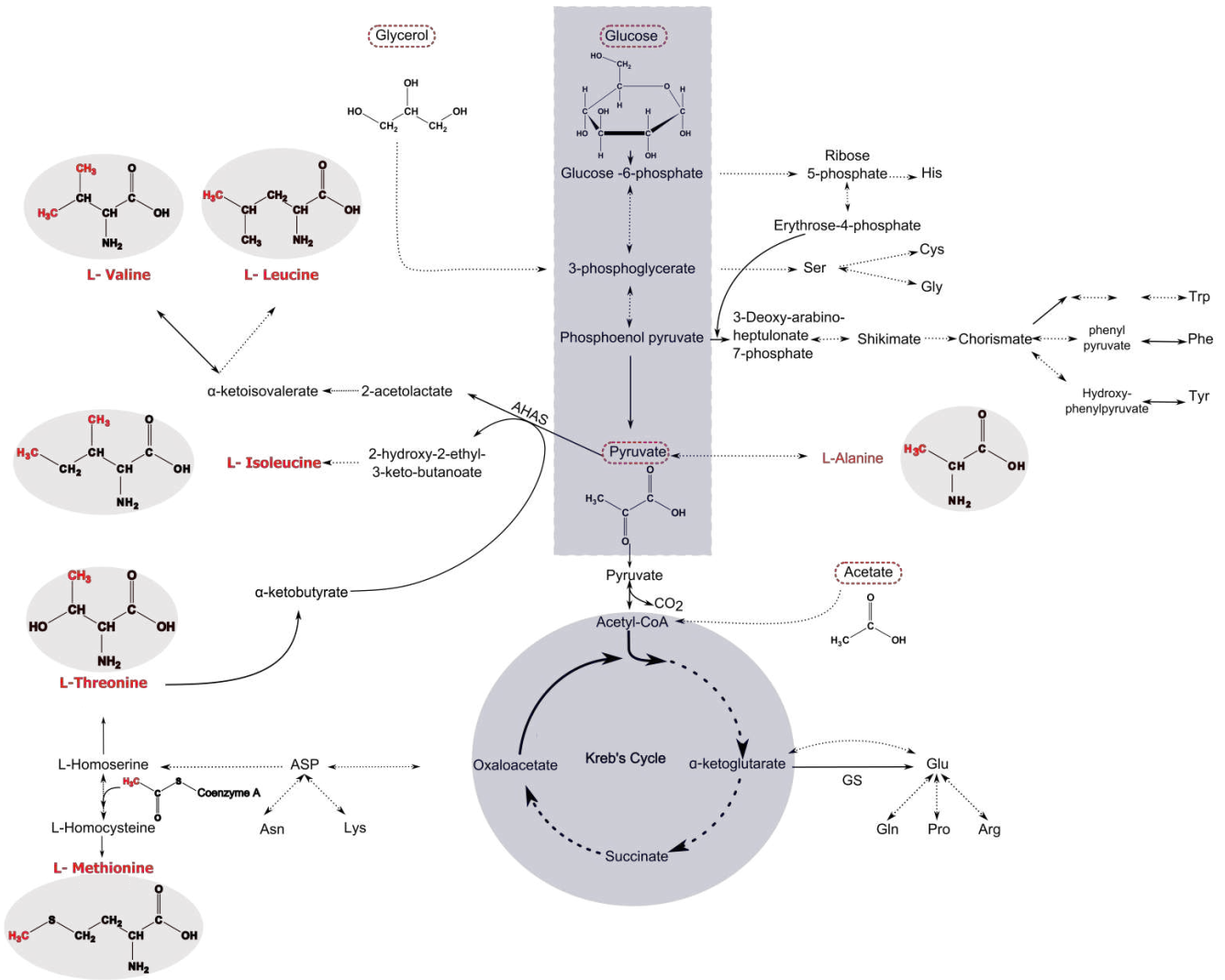


Figure 12 : Représentation des principales réactions des voies métaboliques d'*E. coli* qui conduisent à la synthèse des acides aminés portant un groupement méthyle. Abréviations des enzymes : AHAS : Acéto-hydroxy-acide synthase, GS : Glutamate Synthase. Les sources de carbone qui peuvent être utilisées pour un enrichissement isotopique extensif et sélectif sont entourées de rectangles en pointillés (Figure adaptée du manuscrit de thèse de Rime Kerfah, HAL Id : tel-01343865, version 1).

Les leucines et valines portent deux groupements  $\text{CH}_3$  et représentent plus de 50 % des sondes  $\text{CH}_3$  présentes au sein d'une protéine (McCaldon et Argos, 1988). Ces acides aminés possèdent une série de précurseurs communs qui permettent un marquage simultané simple des deux résidus (Figure 12). En effet, même si le précurseur  $\alpha$ -cétoisovalérate a été privilégié un temps (Goto *et al.*, 1999 ; Gross *et al.*, 2003 ; Tugarinov et Kay, 2004), il s'est avéré que le 2-( $\text{D}_3$ )-méthyl-2,4-( $^{13}\text{C}_2$ )-acetolactate et le 2-( $^{13}\text{C}$ )-méthyl-4-( $\text{D}_3$ )-acetolactate permettaient, respectivement, le marquage

stéréospécifique des résidus  $U$ -[ $^2\text{H}$ ] Leu/Val-[ $^{13}\text{C}^1\text{H}_3$ ] $^{\text{proR}}$  et  $U$ -[ $^2\text{H}$ ], Leu/Val-[ $^{13}\text{C}^1\text{H}_3$ ] $^{\text{proS}}$  (marquage d'un des deux groupements  $\text{CH}_3$  présent sur la chaîne latérale du résidu). Ainsi, il en résulte une diminution des chevauchements et une augmentation de la résolution des spectres obtenus, dû au fait de la stéréospécificité et d'un enrichissement isotopique élevé (> 95 %) (Gans *et al.*, 2010 ; Mas *et al.*, 2013).

Des protocoles de marquage de la valine mais pas de la leucine, et inversement, ont également été proposés. Le marquage stéréospécifique de la valine ou de la leucine seules peut être réalisé à partir de l'utilisation de précurseur 2-acétolactate et de leucines deutérées (marquage de la valine, Mas *et al.*, 2013) ou de 2-cétoisocaproate (marquage de la leucine, Lichtenecker *et al.*, 2013), ou bien d'acides aminés  $^{13}\text{CH}_3$ -valine (Miyanoiri *et al.*, 2013) ou  $^{13}\text{CH}_3$ -leucine (Miyanoiri *et al.*, 2013).

Contrairement au système bactérien, pour lequel on ajoute généralement les précurseurs des acides aminés Ile, Leu et Val, pour marquer ces résidus dans la protéine d'intérêt, l'utilisation des acides aminés  $^{13}\text{CH}_3$ -Leu, Val et Ile à la place de précurseurs est indispensable pour les systèmes eucaryotes pour lesquels ces acides aminés sont essentiels. Ces résidus peuvent être obtenus par synthèse chimique ou enzymatique et conçus à façon (marquage stéréospécifique des méthyles, deutération ...). Ces synthèses ont initialement été développées par le groupe du Dr. Kainosho et sont connues sous le nom de SAIL (Stereo-Array Isotope Labelling). Cette approche permet simultanément de multiplier par quatre à sept le rapport signal/bruit, d'obtenir des pics plus fins et de réduire de 40 à 60 % le nombre de signaux, sans sacrifier les informations essentielles sur le squelette et les chaînes latérales de tous les types d'acides aminés (Kainosho *et al.* 2006). Pour cela, la technique SAIL utilise des acides aminés marqués de façon stéréospécifique et région-spécifique. Les 20 acides aminés constitutifs des protéines sont préparés par des synthèses chimiques et enzymatiques de telle sorte que dans tous les groupes méthylènes, un  $^1\text{H}$  est remplacé stéréosélectivement par un  $^2\text{H}$ , dans tous les groupes méthyles simples, deux  $^1\text{H}$  sont remplacés par des  $^2\text{H}$ , et dans les groupes méthyles prochiraux de Leu et Val, un méthyle est stéréosélectivement  $^{12}\text{C}(^2\text{H})_3$  et l'autre est  $^{13}\text{C}^1\text{H}(^2\text{H})_2$ . Dans les cycles aromatiques à six chaînons, les fragments  $^{12}\text{C}-^2\text{H}$  et  $^{13}\text{C}-^1\text{H}$  alternent l'un avec l'autre (Torizawa *et al.* 2005 ; Kainosho *et al.* 2006). Malgré le caractère innovant et prometteur de ces acides aminés stéréospécifiquement marqués, leurs prix les rendent peu utilisables par les laboratoires (~ 10.000 €/échantillon).

Par conséquent, de nouveaux protocoles dérivés de cette méthode ont été développés et ont permis la synthèse chimique des quatre types de *U*-[<sup>2</sup>H, <sup>15</sup>N]-leucine et -valine marquées stéréospécifiquement <sup>13</sup>CH<sub>3</sub> (Miyanoiri *et al.*, 2013), ainsi que de leucines localement deutérées marquées stéréospécifiquement <sup>13</sup>CH<sub>3</sub> (Dubey *et al.*, 2021). De ces méthodes résultent des coûts de production significativement moins élevés et des résultats similaires à ceux observés lors du marquage SAIL, c'est-à-dire une augmentation de la résolution des spectres et des informations de contraintes et une diminution des nombreuses contributions <sup>1</sup>H.

Ces acides aminés peuvent également être obtenus par synthèse purement enzymatique à partir des précurseurs disponibles pour le système bactérien. La stratégie est de reproduire *in vitro* la voie de biosynthèse de la valine et de l'isoleucine. En effet, ces deux acides aminés partagent la même voie de biosynthèse où trois enzymes interviennent successivement : Kétol-acide réductoisomérase (KARI, EC 1.1.1.86), Dihydroxy-acide déshydratase (DHAD, EC 4.2.1.9), et une Aminotransférase (BCAT, EC 2.6.1.42). C'est l'approche qu'a choisi de développer l'entreprise NMR-Bio en produisant de façon recombinante ces 3 enzymes pour convertir les précurseurs commercialisés en valine et isoleucine marquées stéréospécifiquement et localement deutérées. Ces acides aminés ont été mis à disposition par NMR-Bio pour le développement du milieu de culture et l'étude RMN de HSulf-2 au cours de cette thèse.

La **méthionine** possède la chaîne latérale la plus dynamique des résidus portant un groupement méthyle. Or, cette propriété se traduit par l'acquisition de spectres RMN résolus (dû au fait d'un rapport signal/bruit élevé) (Religa *et al.*, 2010), mais aussi par une résonance dans une région peu peuplée du spectre de corrélation [<sup>1</sup>H, <sup>13</sup>C] (Stoffregen *et al.*, 2012) (cf. Figure 11 p. 93).

La protonation sélective et le marquage isotopique du groupement méthyle (-ε) de la méthionine peuvent être obtenus en ajoutant un précurseur marqué (Fischer *et al.*, 2007). Néanmoins, supplémenter le milieu de culture avec l'acide aminé marqué est, aujourd'hui, une stratégie privilégiée car elle est simple, peu onéreuse, et des incorporations supérieures à 90 % sont observées (Gifford *et al.*, 2011 ; Weininger *et al.*, 2012).

En outre, notons que la méthionine est impliquée dans la voie métabolique de la S-Adénosyl-L-méthionine (SAM). Toutefois, cette fuite isotopique impliquant des

composants cellulaires non protéiques, il n'y aura pas de répercussion sur les spectres RMN (Gifford *et al.* 2011 ; Weininger *et al.* 2012).

### III.2.2. Les voies métaboliques réversibles et intermédiaires

L'alanine est un résidu abondant au sein des protéines. Il représente environ 8 % du nombre total d'acides aminés répartis dans le cœur hydrophobe et à la surface de la protéine (Miller *et al.*, 1987 ; McCaldon et Argos, 1988).

Ce résidu peut également être marqué spécifiquement *via* son groupement méthyle  $\beta$ . Cependant, contrairement aux précédents acides aminés, l'alanine résulte d'une seule réaction réversible du pyruvate propice aux fuites isotopiques (Figure 12). En effet, il existe trois différentes transaminases : la transaminase Alanine-valine (AvtA), la YfbQ et la YfdZ, qui permettent toutes la synthèse et la réversion de l'alanine *via* une activité, vraisemblablement, coordonnée (Kim *et al.*, 2010 ; Yoneyama *et al.*, 2011).

Par conséquent, le marquage spécifique du groupement méthyle de l'alanine nécessite d'ajouter l'alanine en excès et 2 précurseurs (l' $\alpha$ -cétoisovalérate et le « 2-hydroxy-2-[1',2'-(D<sub>5</sub>)] ethyl-3-oxo-4-(D<sub>3</sub>) butanoic acid ») afin de saturer les voies induisant le phénomène de fuite isotopique vers les résidus Val, Leu et Ile, permettant une incorporation efficace de la <sup>13</sup>CH<sub>3</sub>-Ala (Isaacson *et al.*, 2007 ; Ayala *et al.*, 2009). De cette approche résulte une incorporation des alanines de plus de 95 % et un taux de fuite réduit à moins de 0,1 %.

Au regard de ses groupements hydroxyles et méthyles, la thréonine participe aux liaisons hydrogènes et aux interactions non polaires. Ceci implique donc que ce résidu se retrouve davantage à la surface que dans le cœur hydrophobe de la protéine (Miller *et al.*, 1987).

La thréonine étant un intermédiaire de réaction de l'Ile (Figure 12), l'ajout de <sup>13</sup>CH<sub>3</sub>-Thr entraîne un marquage des groupements méthyles Ile <sup>$\delta$ 1</sup> dû à l'action de la thréonine désaminase (TD). Pour pallier cette fuite, il est nécessaire d'ajouter de l' $\alpha$ -cétobutyrate non marqué en excès qui va entrer en compétition avec la thréonine et être converti en isoleucine. La thréonine peut, par ailleurs, être dégradée par différentes enzymes, dont la voie de dégradation de la thréonine déshydrogénase (TDH) et 2-amino-3-cétobutyrate ligase (KBL). Il en découle différents composés marqués (Glycine et Acétyl-CoA marqués) qui vont intégrer différents cycles (comme le cycle TCA) et être à l'origine d'une fuite



isotopique  $^{13}\text{C}$  car ils interviennent dans la biosynthèse de l'Ala et de la Met par la voie Ser et C1. De plus, il est important de souligner que ces réactions consomment de la thréonine marquée et réduisent donc le marquage des thréonines *in fine*. Pour éviter cela, il faut ajouter de la glycine deutérée. Ainsi, en ajoutant 50 mg/L de [ $\alpha,\beta\text{-}^2\text{H}, \gamma\text{-}^{13}\text{C}^1\text{H}_3$ ]-L-Thr avec 50 mg/L de [ $4\text{-}^{13}\text{C}, 3\text{-}^2\text{H}_2$ ]- $\alpha$ -cétobutyrate et 100 mg/L de Glycine deutérée, il a été possible d'obtenir un double marquage, Thr $\gamma^2$  et Ile $\delta^1$ , du protéasome 20S (670 kDa) avec une incorporation supérieure à 90 % (Velyvis *et al.*, 2012).

### III.2.3. Le marquage combinatoire

Comme introduit précédemment, de manière isolée, ces marquages  $^{13}\text{CH}_3$  (tant dans le système bactérien que dans le système eucaryote présenté ci-après) ne permettent pas l'obtention de données structurales. Mais, de manière combinatoire, ils peuvent apporter des informations de cet ordre. Néanmoins, diverses difficultés peuvent apparaître lors du marquage de la totalité des sondes méthyles. Au delà des fuites et dilutions isotopiques propres à chaque acide aminé, le marquage combinatoire peut avoir des conséquences délétères lors de l'analyse et du traitement des spectres.

Premièrement, les acides aminés portant un groupement méthyle résonnent dans une région étroite du spectre RMN 2D (cf. Figure 11 p. 93). Or, plus le système étudié est de haute masse moléculaire, plus il y aura de signaux détectés et enregistrés et donc, possiblement, un large chevauchement des régions où résonnent les  $^{13}\text{CH}_3\text{-aa}$ . Dans ce cas, un compromis dans le marquage combinatoire doit être réalisé afin de limiter ces chevauchements. Pour ce faire, les propriétés de dispersion de résonance des acides aminés portant un groupement méthyle doivent être prises en compte. Par exemple, les groupes  $^{13}\text{CH}_3\text{-Met}$  et  $^{13}\text{CH}_3\text{-Ile}^{\delta^1}$  apparaissent dans une fenêtre spectrale bien séparée, ce qui les rend faciles à combiner avec d'autres résidus. Mais, les signaux de l'Ala, la Val<sup>proS</sup>, et la Thr apparaissent dans une région très réduite (tout particulièrement pour la Val<sup>proS</sup> et la Thr où les régions sont quasiment superposées). Ainsi, faire un marquage combinatoire avec ces trois derniers acides aminés complexifiera significativement l'analyse des spectres RMN.

Deuxièmement, afin de simplifier les spectres, il est possible d'avoir recours à un marquage stéréospécifique des groupements méthyles des valines, leucines et isoleucines (comme présenté précédemment). En effet, si ces acides aminés (qui

possèdent chacun deux groupements méthyles) ne sont pas marqués de manière stéréospécifique, alors les fortes interactions dipolaires  $^1\text{H}$ - $^1\text{H}$  présentes entre deux groupements méthyles très proches vont détériorer la qualité du spectre en raison d'un taux de relaxation rapide et d'un chevauchement des signaux. En pratique, les groupes méthyles pro-S de la leucine et de la valine présentent une dispersion de déplacement chimique plus large que les groupes pro-R (London *et al.*, 2008). Pour les groupes méthyles Ile $\alpha^2$ , le point le plus critique est le fait que leurs résonances chevauchent celles des Ala, Val, Met, et Thr, rendant l'Ile $\delta^1$  plus pertinent pour un marquage combinatoire. De plus, la position atomique de l'Ile $\alpha^2$  le rend moins flexible et, par conséquent, moins sensible que l'Ile $\delta^1$ .

Pour les cellules eucaryotes, les mêmes problématiques s'appliqueront lors de la réalisation de marquages combinatoires. Toutefois, grâce au fait que la plupart des acides aminés sont devenus essentiels et donc non synthétisables par la cellule, les fuites et les dilutions isotopiques seront quasiment inexistantes (cf. III.3.3.2. p. 108).

### **III.3. Marquage isotopique en cellules eucaryotes : stratégies de marquage et formulation des milieux**

La majorité des protéines étudiées par RMN ont été classiquement produites en utilisant des systèmes d'expression bactériens ou des systèmes d'expression « Cell Free » (Carlson *et al.*, 2012). Or, pour la biologie structurale, la production de protéines recombinantes dans des conformations et des fonctions natives est devenue un enjeu majeur. En effet, la demande d'informations sur les récepteurs de surface cellulaire et/ou la structure de glycoprotéines de mammifères a significativement augmenté. C'est pourquoi, le passage à des systèmes d'expression eucaryotes a été nécessaire pour atteindre les informations portant sur les modifications post-traductionnelles (telles que les glycosylations, les phosphorylations ou la formation de ponts disulfures ...).

#### **III.3.1. Les différents systèmes d'expression et leurs avantages**

i. Bien que les levures soient des organismes eucaryotes, les méthodes utilisées pour le marquage isotopique des protéines sont largement dérivées du système

d'expression en bactérie, et s'éloignent, de ce fait, des autres systèmes d'expression eucaryotes « classiques ».

Les levures peuvent être adaptées à une croissance en présence de deutérium, ce qui permet la deutération de la protéine et induit une diminution significative des interactions dipolaires  $^1\text{H}$ - $^1\text{H}$ , ce qui est indispensable pour certaines analyses RMN. Les levures peuvent également croître dans des milieux de culture minimum car elles ont la capacité de synthétiser tous les acides aminés à partir de précurseurs que sont le glucose et l'ammonium. Ceci facilite donc leurs substitutions par des précurseurs différemment marqués pour des études RMN. De plus, comme les bactéries, les levures présentent également des temps de génération rapides et atteignent des densités cellulaires élevées (Ahmad *et al.*, 2014 ; Byrne *et al.*, 2015 ; Puxbaum *et al.*, 2015).

Outre ces caractéristiques se rapprochant de celles des bactéries, les levures disposent de la machinerie des eucaryotes, c'est-à-dire qu'elles permettent le bon repliement des protéines, la formation de ponts disulfures, et les principales modifications post-traductionnelles (Cregg *et al.*, 2009 ; Felber *et al.*, 2014).

La souche la plus largement utilisée est *Pichia pastoris* (*P. pastoris*). Cette dernière, classée en trois phénotypes (Mut+, MutS et Mut-), est capable d'utiliser du glucose (Emami *et al.*, 2013 ; Fan *et al.*, 2011, 2015) ou du glycérol (Wood *et al.*, 1999) comme seule source de carbone dans un milieu minimal pendant la croissance cellulaire (ces deux sources sont très généralement utilisées pour l'ensemble des protocoles de marquage isotopique ( $^{13}\text{C}$ ) établis chez la bactérie). L'expression de la protéine recombinante est induite par le méthanol, car ce dernier induit l'activité du promoteur AOX1 (et dans une moindre mesure AOX2). Ainsi, ce promoteur est utilisé pour piloter l'expression de la protéine souhaitée en remplaçant le gène AOX1 par l'ADN cible (Cregg *et al.*, 1985).

En conséquence, *P. pastoris* permet l'expression et le marquage de protéines recombinantes. Par exemple, le marquage uniforme  $^{13}\text{C}/^{15}\text{N}$  chez cette souche est très simple et suit des protocoles bien établis pour l'utilisation de milieux minimums (Wood *et al.*, 1999 ; Rodriguez *et al.*, 2001 ; Pickford *et al.*, 2004 ; Sugiki *et al.*, 2012). En revanche, en raison de la compartimentation du métabolisme de biosynthèse des acides aminés, le marquage spécifique des acides aminés portant un groupement méthyle n'est pas aussi simple. En effet, le métabolisme central du carbone de *P. pastoris* est divisé en deux sous-réseaux : un cytosolique et un mitochondrial, où différentes enzymes,

essentielles à la biosynthèse des acides aminés y sont exclusivement localisées (Solà *et al.*, 2004). De ce fait, les précurseurs classiquement utilisés chez la bactérie et permettant d'enrichir spécifiquement les résidus Ile, Leu et Val avec des groupes  $^{13}\text{CH}_3$  dans la protéine perdeutérée que sont l' $\alpha$ -cétobutyrate marqué et l'acétolactate marqué, ne sont que faiblement incorporés. Un taux d'incorporation de 45 % pour la  $^{13}\text{CH}_3\text{-Ile}^{\delta 1}$  est observé chez la souche *P. pastoris* (Clark *et al.*, 2015, 2017, 2018 ; Ali *et al.*, 2019 ; Dikiy *et al.*, 2019), tandis qu'un taux de 67 % avec la souche *K. lactis* (Suzuki *et al.*, 2018) est retrouvé. De la même manière, l'incorporation des valines et des leucines est faible avec des taux d'incorporation de 57 % et 14 %, respectivement (Ali *et al.*, 2019).

Outre l'utilisation de précurseurs (qui ne permettent pas des taux d'incorporation importants), l'ajout d'acides aminés marqués peut se révéler être une meilleure stratégie pour un taux d'incorporation plus important des groupements  $^{13}\text{CH}_3$ . Pour la valine, par exemple, le résidu marqué n'aura pas à entrer dans les mitochondries pour être métabolisé (à l'inverse de son précurseur) (Miyazawa-Onami *et al.*, 2013 ; Clark *et al.*, 2015 ; Zhang *et al.*, 2017 ; Suzuki *et al.*, 2018). Selon cette même stratégie, le marquage des méthionines, des alanines ou des thréonines pourrait être réalisé.

En conclusion, les levures sont des hôtes d'expression prometteurs, en particulier pour les protéines membranaires, comme l'illustrent les travaux réalisés sur les GPCR. Néanmoins, de nouvelles études seront nécessaires afin d'identifier de nouveaux protocoles permettant le marquage de l'ensemble des acides aminés portant un groupement méthyle, mais aussi dans l'accroissement de leur incorporation.

Contrairement aux levures, les systèmes d'expression eucaryotes supérieurs (les cellules d'insectes et les cellules de mammifères) sont capables d'effectuer des glycosylations plus évoluées voire identiques à l'Homme. Par conséquent, dans certains cas d'étude, ces systèmes d'expression sont indispensables.

Ainsi, de nouvelles méthodes de marquage isotopique dédiées à ces systèmes d'expression au métabolisme cellulaire particulier doivent être développées. Quelques études se sont, d'ores et déjà, concentrées sur le développement d'un protocole de marquage uniforme ou spécifique des acides aminés portant un groupement méthyle, et plus récemment, couplé à une deutération (cf. III.3.3. p. 106).

ii. A l'inverse des bactéries, **les cellules d'insectes** présentent une génétique plus compliquée et un métabolisme ne permettant pas un développement en milieu minimum. Toutefois, les cellules d'insectes permettent la production de protéines cytosoliques. De grands complexes protéiques correctement repliés peuvent, également, être produits par ces cellules car leurs machineries disposent de chaperonnes et de cofacteurs similaires à ceux des cellules humaines (O'Reilly *et al.*, 1994 ; Jarvis *et al.*, 1997). Quant aux protéines sécrétées, les cellules d'insectes engendrent des profils de glycosylation homogènes et limités, facilitant les études structurales. Par conséquent, une large fraction des protéines humaines est accessible avec l'utilisation des cellules d'insectes comme système d'expression (Assenberg *et al.*, 2013). Toutefois, ces glycosylations peuvent être dites « limitées » car elles ne sont pas équivalentes à celles produites par les cellules de mammifères (Shi *et al.*, 2007 ; Yanaka *et al.*, 2018). En effet, alors que les glycoprotéines des mammifères portent souvent des N-glycans présentant des acides sialiques, les glycoprotéines recombinantes dérivées de cellules d'insectes ont généralement des chaînes de sucres beaucoup plus simples connues sous le nom de N-glycans paucimannosidiques. Ainsi, ces différences structurales entre les N-glycans synthétisés par les cellules d'insectes et de mammifères doivent être prises en compte dans le choix du système d'expression afin d'obtenir des glycoprotéines recombinantes où les fonctions, les interactions avec les partenaires physiologiques, et autres activités, sont maintenues.

Il existe deux principaux types de cellules d'insectes : les cellules de *Spodoptera frugiperda* (Sf9 ou Sf21) qui sont infectées par un baculovirus recombinant (BV) pour initier l'expression de protéines cibles, et les cellules de *Drosophila Schneider S2* qui peuvent être transfectées avec des vecteurs d'expression simples pour permettre une expression stable ou transitoire de protéines recombinantes. Pour des études RMN, l'infection par un baculovirus est utilisé dans la majorité des cas (une seule étude fait état des cellules S2 comme outil pour un marquage uniforme  $^{15}\text{N}$  ; Meola *et al.*, 2014).

Le marquage d'acides aminés est relativement bien établi pour ce système d'expression et fournira, notamment pour le marquage spécifique, de meilleurs résultats que dans *E. coli* en termes d'incorporation et de fuite isotopique. Cependant, pour le marquage uniforme d'une protéine, ce système peut se révéler plutôt onéreux et présenter une

incorporation en isotopes plus faible suivant les acides aminés, bien que pouvant aussi présenter des taux d'incorporation jusqu'à 90 % (cf. III.3.3.2. p. 108).

iii. **Les cellules de mammifères** les plus couramment utilisées sont les cellules CHO et les cellules HEK 293. Les CHO ont été largement modifiées pour améliorer les rendements et la qualité des protéines et sont désormais capables d'exprimer de grandes quantités de protéines. Cela en fait un système de choix pour la production industrielle d'anticorps. La lignée cellulaire des HEK 293 est, quant à elle, utilisée pour les protéines qui nécessitent des modifications post-traductionnelles fonctionnellement sensibles.

Ces cellules peuvent être cultivées en suspension (augmentant significativement la densité cellulaire au regard des cultures adhérentes) et ne font pas intervenir des virus pour la transfection mais de simples réactifs. Le réactif chimique le plus couramment utilisé pour la transfection est le polyéthylèneimine (PEI), qui est utilisé à la fois pour la transfection transitoire et pour l'établissement de lignées cellulaires stables. Le PEI condense l'ADN en particules chargées positivement qui se lient aux résidus anioniques de la surface cellulaire et sont introduites, par la suite, dans la cellule par endocytose. Pendant l'expression transitoire, la séquence d'acide nucléique n'est pas intégrée dans le génome de la cellule hôte. Par conséquent, l'effet sur l'expression du gène cible est temporaire (48-96 heures), mais l'établissement de lignées stables permet une expression en continu de la protéine cible.

### **III.3.2. Le métabolisme de biosynthèse des acides aminés**

Les cellules d'insectes et de mammifères présentent un métabolisme de biosynthèse de leurs acides aminés différent de celui décrit pour les levures (Solà *et al.*, 2004). Pour rappel, les bactéries et les levures peuvent synthétiser toutes les molécules du vivant à partir de simples précurseurs. En revanche, les cellules d'insectes et de mammifères ont perdu cette capacité de synthèse de bon nombre de ces molécules. Ainsi, les acides aminés sont répartis en deux groupes. Les acides aminés essentiels (Arg, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Val, Tyr et Cys) sont tous les acides aminés que la cellule ne peut synthétiser (Mitsuhashi, 1982), ce qui implique que ces derniers doivent être apportés à la cellule par ajout d'extraits ou sous forme libre. Les acides aminés non essentiels (Ala, Asp, Asn, Glu, Gln, Gly, Pro, Ser) sont synthétisés par la cellule, dans des

concentrations variables, *via* différentes voies de biosynthèse découlant de la glycolyse (Ser, Gly) et du cycle de l'acide citrique (Ala, Glu, Gln, Pro, Asp, Asn) (Lehninger, 1975 ; Öhman *et al.*, 1996 ; Drews *et al.*, 2000). Enfin, les résidus Cys et Tyr sont dits semi-essentiels car ils peuvent être synthétisés par les cellules d'insectes (mais pas par les cellules de mammifères) si des résidus précurseurs Met/Ser (synthèse Cys) et Phe (synthèse Tyr) sont présents (Mitsuhashi, 1982 ; Öhman *et al.*, 1996 ; Doverskog *et al.*, 1998).

Ainsi, le marquage des groupements méthyles des Ile, Val, et Leu (pour lesquels un précurseur marqué peut être utilisé chez la bactérie) sera réalisé par apport direct de l'acide aminé marqué pour les cellules d'insectes et de mammifères.

En conséquence, l'utilisation de cellules de mammifères pour la production de protéines marquées nécessite le développement de nouvelles stratégies de marquage, et de nouveaux milieux de culture (les cellules d'insectes ont déjà de nombreux milieux et stratégies validés).

### **III.3.3. Formulation des milieux de marquage**

#### **III.3.3.1. Les sources d'acides aminés à considérer pour un marquage isotopique**

Les cellules eucaryotes (insectes et mammifères) ne pouvant croître en milieu minimum, leurs milieux de culture doivent apporter l'ensemble des constituants métabolites indispensables à leur vie. Ainsi, les milieux de culture pour les cellules d'insectes et de mammifères sont relativement similaires et contiennent généralement plus de 50 constituants, dont des sels et différents glucides (avec 5-10 g/L de glucides pour les milieux de culture des cellules d'insectes et 2-3 g/L pour les milieux de culture des cellules de mammifères). A ces formulations s'ajoutent des acides aminés sous forme pure avec, encore une fois, davantage d'acides aminés pour les milieux insectes (10-15 g/L) que mammifères (1-2 g/L). Ces milieux sont également riches en hydrolysats, comme l'extrait de levure (2-8 g/L) pour les milieux de culture des cellules d'insectes. Cet extrait de levure est riche en acides aminés mais aussi en peptides, glucides, vitamines et lipides, et permet de stimuler la croissance et la production des protéines (O'Reilly *et al.*, 1994). Pour les cellules de mammifères, l'extrait de levure est parfois

remplacé par de la peptone (hydrolysats de protéines dérivés de plantes). La peptone apporte les mêmes constituants que l'extrait de levure mais son action sur les cellules n'est pas clairement définie (Mosser *et al.*, 2013 ; Kim *et al.*, 2013 ; Shirsat *et al.*, 2013). Enfin, du sérum peut être présent dans les milieux de culture destinés aux deux systèmes d'expression. Ceci étant, l'utilisation de sérum, de protéines et de peptides dans les milieux de culture de cellules pose de nombreux problèmes, en particulier pour les applications industrielles. En effet, l'utilisation de sérum dans les milieux de culture induit un risque en termes d'innocuité puisque sa composition est indéfinie et qu'il peut contenir des agents pathogènes indétectables par les méthodes conventionnelles. Le sérum est également une source importante de contamination (protéique notamment) qui rend parfois difficile la purification de molécules biopharmaceutiques. Ces dernières années, des milieux ne contenant donc aucune protéine ni peptide, et dont la composition est parfaitement maîtrisée, ont été mis au point. Ces milieux sont connus sous le nom de « CD media » pour « chemically defined ». De ce fait, une large majorité de cellules sont, aujourd'hui, adaptées à des milieux de culture sans sérum.

A l'exception des sels et des vitamines, tous les composants entrant dans la formulation des milieux représentent des sources de carbone et/ou d'azote, possibles. Or, pour la réalisation d'un marquage, tant spécifique qu'uniforme, l'ensemble des composants doit être pris en compte et, si nécessaire, substitué ou éliminé afin de limiter les dilutions isotopiques. Un certain nombre de ces molécules sont néanmoins présentes dans des quantités infimes et peuvent donc être négligées (Gossert et Jahnke, 2012). En revanche, d'autres éléments sont non négligeables, comme les acides aminés, les sucres et les hydrolysats éventuels.

Le glucose, par exemple, qui est le principal sucre présent, est à l'origine de la voie de glycolyse et du cycle de l'acide citrique qui entraînent la biosynthèse des acides aminés non essentiels (Strauss *et al.*, 2005). Par conséquent, réaliser un marquage  $^{13}\text{C}$  des acides aminés issus de ces voies nécessite l'utilisation d'une source de carbone marquée, ou bien de réguler à la baisse ces voies métaboliques conjointement à l'ajout d'un mélange d'acides aminés uniformément  $^{13}\text{C}$  marqués.

Les acides aminés présents dans le milieu de culture découlent de deux sources : une forme libre et, parfois, des extraits de levure ou de peptone. Pour les cellules d'insectes, bien que cet extrait de levure ne puisse être éliminé, il a été montré qu'il pouvait être diminué (quant à la peptone, son addition au milieu n'est pas indispensable



pour les cellules de mammifères). En effet, en réduisant sa quantité par 10, la dilution isotopique est grandement limitée (Gossert *et al.*, 2011). De plus, la dialyse ou la filtration de ces extraits permet également d'éliminer la plupart des acides aminés le composant, améliorant ainsi l'incorporation des acides aminés marqués (Zhang *et al.*, 2017). En effet, il a pu être démontré que cette double action sur l'extrait de levure (diminution de la quantité et filtration) permettait le maintien des rendements d'expression des protéines (comparée à l'expression obtenue dans le milieu commercial standard), ainsi qu'une augmentation significative du taux d'incorporation des isotopes. Par exemple, pour la  $^{13}\text{C}$ -Thr, le taux d'incorporation est passé de 62,4 % (extrait de levure non dialysé mais quantité réduite ; Gossert *et al.*, 2011) à 89,6 % dans la condition où l'extrait de levure a été dialysé.

Pour les acides aminés libres dans le milieu, de nombreux milieux sont aujourd'hui commercialisés sous une forme « free amino acids ». De ce fait, un simple ajout d'acides aminés sous une forme marquée permettra une incorporation maximale au sein des protéines.

En résumé, pour un marquage isotopique dans les cellules d'insectes et de mammifères, les acides aminés libres peuvent être éliminés dès la formulation du milieu, puis ajoutés sous une version marquée ou non. Les hydrolysats peuvent, eux, être réduits voire éliminés, et les voies de biosynthèse des acides aminés non essentiels peuvent être régulées à la baisse, comme avec la synthèse endogène des alanines qui peut être inhibée par l'ajout d'inhibiteur ( $\beta$ -chloro-L-alanine/L-cyclo-sérine, cf. Figure 14 p.111).

### III.3.3.2. Les stratégies de marquage en cellules eucaryotes supérieures

Afin de marquer uniformément ( $U$ - $^{15}\text{N}$  et/ou  $U$ - $^{13}\text{C}$ ), ou spécifiquement ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^{13}\text{CH}_3$ , entre autres), une protéine exprimée en cellules d'insectes ou de mammifères, une stratégie générale a été établie. Elle repose sur le contrôle de la formulation des milieux de culture afin d'obtenir un marquage optimal des protéines d'intérêt avec des taux d'incorporation élevés et un rendement protéique maintenu. Ce dernier point est bien souvent un frein à l'innovation et au développement de milieux de marquage car, en voulant réduire au maximum les sources de fuites et de dilutions isotopiques, les milieux deviennent moins nutritifs, moins optimaux pour la croissance cellulaire, et les rendements protéiques peuvent être drastiquement diminués. Initialement, les milieux pour la RMN étaient formulés à la main à partir de composants individuels (Bruggert *et*

*al.*, 2003 ; Walton *et al.*, 2006), avec peu de reproductibilité du fait de toutes les variations induites par les manipulateurs. Depuis, ces milieux ont été commercialisés et sont disponibles à façon, déplétés d'un, plusieurs, ou de la totalité des acides aminés (par exemple le milieu ESF-SFM Mammalian Cell Culture Medium All Amino Acid Deficient, réf. 98-003, commercialisé par Expression System). Dans ce cas, l'utilisateur complète le milieu de culture avec un mélange d'acides aminés marqués ou non (Strauss *et al.*, 2003, 2005 ; Werner *et al.*, 2007, 2008 ; Kato *et al.*, 2010 ; Yamaguchi et Kato, 2010 ; Gossert *et al.*, 2011 ; Kofuku *et al.*, 2012, 2014 ; Yagi *et al.*, 2015 ; Zhang *et al.*, 2017). Ces acides aminés peuvent être ajoutés, soit un à un sous une forme pure, engendrant un coût très élevé pour un marquage uniforme où tous les acides aminés doivent être marqués, soit par l'ajout d'extraits de levures ou d'algues (sources moins coûteuses d'acides aminés marqués). La composition en acides aminés de ces extraits dépend de l'organisme source et de la méthode d'hydrolyse (autolyse, hydrolyse). En effet, l'autolyse (incubation des levures à 50°C et dégradation des protéines internes par des protéases) est l'approche la plus douce car elle conserve la plupart des acides aminés, mais une majorité de ces derniers sont retrouvés dans de faibles quantités (Egorova-Zachernyuk *et al.*, 2009, 2011). En revanche, l'hydrolyse acide (extraits de protéines purifiées soumis à des températures élevées (120°C) à faible pH) entraîne de plus grandes quantités d'acides aminés (65 % contre 35 % pour l'autolyse) mais est aussi à l'origine de la dégradation de nombres d'entre eux tels que Gln, Asn, Cys, et Trp (LeMaster et Richards, 1982).

Pour marquer uniformément une protéine, l'utilisation d'extraits de levure marqués est possible, mais les extraits d'algues représentent une alternative intéressante car les extraits d'algues marqués sont moins chers à produire que l'extrait de levure (Kofuku *et al.*, 2014 ; Sitarska *et al.*, 2015). En effet, les algues peuvent utiliser les molécules les plus simples comme précurseurs des acides aminés ( $^{15}\text{NH}_3$ ,  $^{13}\text{CO}_2$  et  $\text{D}_2\text{O}$ ). En revanche, les levures nécessitent des glucides plus complexes comme le glucose, ce qui rend les marquages  $^{13}\text{C}$  et  $^2\text{H}$  plus coûteux. Il est vrai que seulement ~ 10 % du glucose utilisé est converti en acides aminés (Egorova-Zachernyuk *et al.*, 2009). Ceci étant, des protocoles de production et d'utilisation d'extraits de levure marqués existent pour la réalisation de nombreux marquages uniformes.

De la même manière, il existe des autolysats de levures marquées  $^{15}\text{N}$  ou  $^{15}\text{N}$ ,  $^{13}\text{C}$  (Protein labeling innovation and CortecNet, Meola *et al.*, 2014), mais aussi des

hydrolysats d'algues, moins chers, disponibles et ceci dans une large gamme de marquages ( $^{15}\text{N}$  ;  $^{13}\text{C}$ ,  $^{15}\text{N}$  ;  $^2\text{H}$ ,  $^{15}\text{N}$  ;  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , Sigma-Aldrich and Cambridge Isotope Laboratories). En pratique, il peut être ajouté jusqu'à 10-12 g/L d'extrait d'algues au milieu de culture. Au-delà de cette valeur, les cellules d'insectes ne tolèrent plus l'osmolarité du milieu (Sitarska *et al.*, 2015).

Toutefois, un déficit, voire une absence, d'acides aminés existe au sein de ces extraits (Figure 13), ce qui peut influencer défavorablement les marquages isotopiques à partir de ces derniers. Néanmoins, d'après plusieurs études documentant particulièrement bien le métabolisme des cellules d'insectes (constatation théoriquement transposable pour les cellules de mammifères car leurs métabolismes sont proches), ce déficit en acides aminés n'est pas problématique. En effet, la Cys marquée peut être synthétisée par la cellule d'insecte à partir de Ser marquée grâce à un intermédiaire de réaction, la cystathionine, et deux enzymes de biosynthèse (cystathionine  $\beta$ -synthase/ $\alpha$ -lyase)

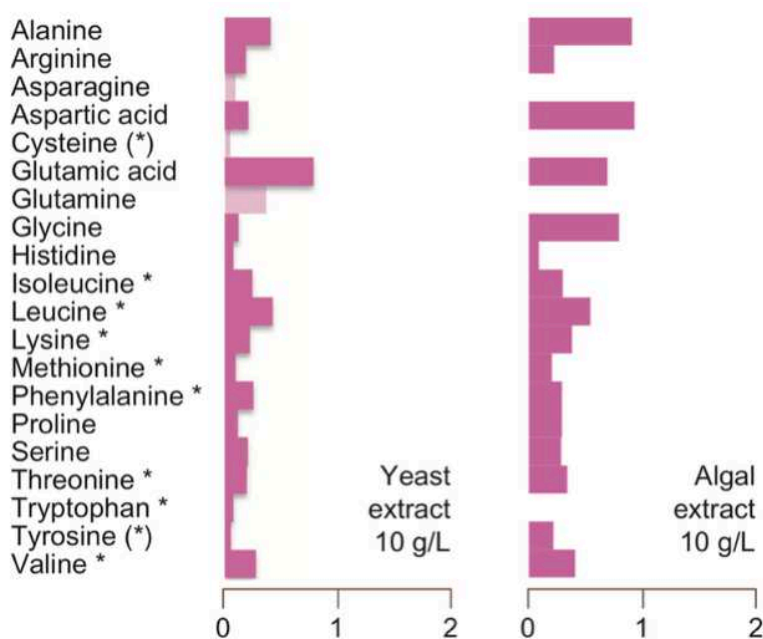


Figure 13 : Composition en acides aminés des extraits de levure et d'algue. Les astérisques sur le nom de l'acide aminé marquent les acides aminés essentiels. Notons que dans les hydrolysats d'algues, la teneur globale en acides aminés est plus élevée, mais Asn, Gln, Cys et Trp sont manquants en raison du traitement par hydrolyse acide. En revanche, dans les extraits de levure préparés par autolyse, ces acides aminés sont partiellement conservés (violet clair) mais la teneur globale en acides aminés est moindre (Figure : Skora *et al.*, 2015).

(Doverskog *et al.*, 1998). Grâce à cette biosynthèse, il n'est donc pas nécessaire d'ajouter de la Cys marquée car la Ser et la Met sont présentes en grandes quantités dans ces extraits. La Gln marquée est synthétisée à partir de L-Glu et d'ammonium marqués (Öhman *et al.*, 1996 ; Doverskog *et al.*, 2000 ; Drews *et al.*, 2000). Puis, cette Gln va intervenir comme précurseur, avec l'Asp, dans la synthèse de l'Asn marquée, grâce à l'enzyme asparagine synthétase (Hansen *et al.*, 1992). En résumé, pour pallier la faible présence des différents résidus (donc la faible incorporation finale), il

suffit de deux précurseurs (présents en grande quantité dans l'extrait) et de  $\text{NH}_4\text{Cl}$  marqué  $^{15}\text{N}$  (Hansen *et al.*, 1992 ; Meola *et al.*, 2014 ; Sitarska *et al.*, 2015).

Le Trp doit, lui, être supplémenté marqué étant donné qu'il est essentiel, mais sa quantité peut être réduite à 10-20 mg/L sans affecter l'incorporation isotopique ou la croissance cellulaire (Hansen *et al.*, 1992 ; Sitarska *et al.*, 2015).

Pour finir, une dernière problématique apparaît, et ceci pour l'ensemble des marquages (uniforme et spécifique) au sein des deux systèmes d'expression : l'impact des acides aminés non essentiels synthétisés à partir du catabolisme du glucose, *via* des réactions de transaminations ou d'amidations réductrices (Figure 14) (Öhman *et al.*, 1995 ; Drews *et al.*, 2000). En effet, dans le cas d'un marquage  $^{13}\text{C}$  et/ou  $^{15}\text{N}$ , si la source de glucose et de chlorure d'ammonium n'est pas marquée (seuls les acides aminés ajoutés au milieu sont marqués), les acides aminés synthétisés par la cellule à

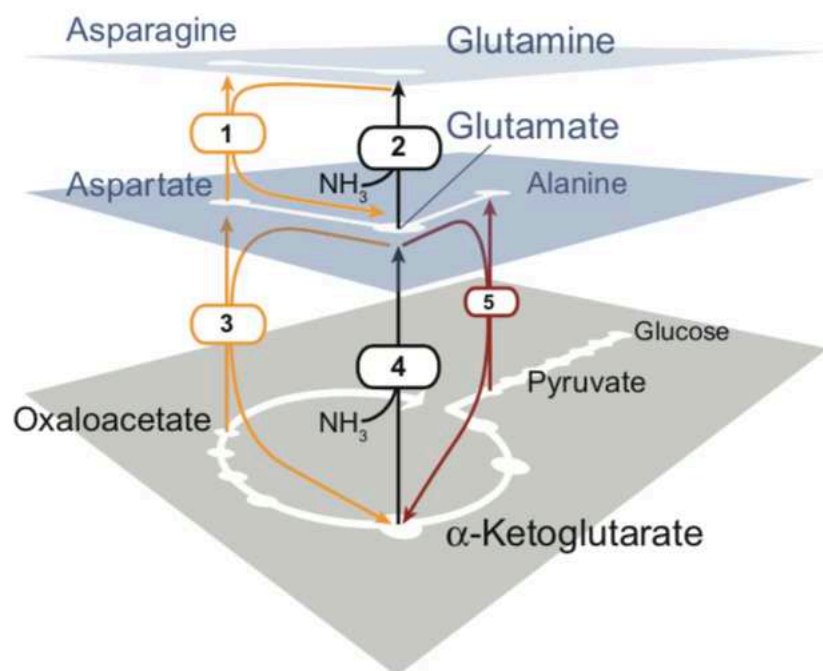


Figure 14 : Aperçu des voies métaboliques des acides aminés non essentiels issus de la glycolyse. Le plan inférieur (gris) représente la glycolyse et le cycle de l'acide citrique. Le niveau intermédiaire (bleu) montre les acides aminés qui sont synthétisés dans une réaction en une étape à partir des précurseurs respectifs par addition d'un azote. Le niveau supérieur (bleu clair) montre les acides aminés avec encore un azote supplémentaire. Les réactions enzymatiques sont marquées par des flèches et les enzymes sont énumérées comme suit : 1 : Asparagine synthétase, 2 : Glutamine synthétase, 3 : Aspartate transaminase, 4 : Glutamate déshydrogénase, 5 : Alanine transaminase (Figure : Skora *et al.*, 2015).

partir du glucose ne seront pas marqués et engendreront une dilution isotopique d'environ 10 à 20 %. L'ajout de glucose  $^{13}\text{C}$  permet donc de remédier à cette dilution. Cependant, pour atteindre 80 % de marquage  $^{13}\text{C}$ , 5 à 10 g/L de glucose seront nécessaires, augmentant significativement le coût du milieu (Kato *et al.*, 2010 ; Skora *et al.*, 2015). Pour le marquage  $^{15}\text{N}$ , le précurseur marqué  $^{15}\text{NH}_4\text{Cl}$  (intermédiaire de réaction du cycle de l'acide citrique et de l'alanine) permettra un marquage  $^{15}\text{N}$  simultané de l'acide glutamique, la glutamine, l'acide aspartique, l'asparagine et l'alanine (Bruggert *et al.*, 2003).

En pratique, il a pu être observé que les acides aminés Asp, Asn, Glu et Gln ne souffrent que d'une dilution relativement faible. De plus, les acides aminés non essentiels dérivés du glucose n'affectent pas l'incorporation du  $^{13}\text{C}$  du fait de leurs voies de biosynthèse qui sont peu/pas actives dans certaines conditions (Skora *et al.*, 2015). En revanche, l'alanine est massivement synthétisée à partir du glucose grâce à l'alanine transaminase. Cette synthèse peut être sélectivement inhibée si de la L-cyclo-sérine, un inhibiteur de la pyruvate-alanine transaminase (Wong *et al.*, 1973), est ajoutée. De ce fait, l'addition de glucose marqué au milieu n'est plus indispensable, ce qui permet de réduire les coûts et d'atteindre des niveaux d'incorporation comparables (73 %) (Sitarska *et al.*, 2015).

Ainsi, ces protocoles s'avèrent être relativement faciles à mettre en place et robustes pour le marquage sélectif et uniforme des acides aminés essentiels. En effet, des taux d'incorporation proches de 80 % à 90 % pour le marquage uniforme  $^{15}\text{N}$  et  $^{13}\text{C}$  et des rendements d'expression comparables à ceux des milieux non marqués ont par exemple pu être obtenus (Strauss *et al.*, 2003, 2005 ; Werner *et al.*, 2008 ; Kato *et al.*, 2010, 2018 ; Gossert *et al.*, 2011 ; Hamatsu *et al.*, 2013 ; Sitarska *et al.*, 2015 ; Zhang *et al.*, 2017).

De plus, ces approches reposent sur des composés peu coûteux, disponibles et largement distribués dans les laboratoires et commerces. Egalement, si le milieu de culture est disponible sous une version sans acides aminés, ces protocoles sont applicables à de nombreux milieux de cultures et systèmes d'expression.

Pour l'étude des grandes protéines, il est en outre nécessaire d'apporter de la deutération afin de réduire les interactions dipolaires  $^1\text{H}$ - $^1\text{H}$  et de pouvoir acquérir des spectres de haute résolution.

### III.3.3.3. La deutération

La deutération dans les cellules eucaryotes reste l'un des objectifs majeurs de la RMN pour répondre aux problématiques actuelles et à venir. Mais, contrairement aux bactéries ou aux levures, la deutération ( $^2\text{H}$ ) uniforme des protéines est loin d'être aisée puisque les cellules eucaryotes ne peuvent croître dans un milieu de culture deutéré (Murphy *et al.*, 1977). En effet, le glucose deutéré est toxique pour les cellules. Afin d'obtenir une deutération partielle des protéines, seul l'ajout d'acides aminés deutérés (sous forme pure ou d'un mélange réactionnel d'algues ou de levures  $^2\text{H}$ ) peut être envisagé. Toutefois, ceci reste une tâche difficile car, à cause du métabolisme des cellules eucaryotes, la deutération des protéines obtenues n'est pas uniforme (Sitarska *et al.*, 2015). En effet, sans glucose deutéré, les acides aminés Asp, Asn, Glu et Gln, et Ala seront synthétisés protonés. De plus, dès lors que les cellules sont cultivées dans un milieu protoné, les acides aminés impliqués dans les réactions des transaminases verront leurs C $\alpha$  protonés (Kofuku *et al.*, 2014). Cela se traduira par un modèle de deutération hétérogène diminuant la résolution et/ou la sensibilité des acquisitions RMN. L'utilisation de L-cyclo-sérine peut néanmoins restaurer l'incorporation des  $^2\text{H}$ -Ala en inhibant la synthèse d'alanine protonée endogène. Aucun autre inhibiteur n'a en revanche été défini comme efficace pour empêcher la protonation des autres acides aminés synthétisés *in vivo*.

Par ailleurs, du fait des réactions réversibles des aspartates transaminases, des alanines transaminases et des glutamates déshydrogenases, les  $^2\text{H}$ -acides aminés présents dans le milieu réduisent les rendements protéiques de 30 à 60 %. En effet, ces enzymes conduisent à la synthèse d' $\alpha$ -cétoglutarate deutéré ré-utilisable dans le cycle de l'acide citrique. Or, ces intermédiaires deutérés agissent comme de faibles inhibiteurs en réalisant la rupture des liaisons hydrogènes, ce qui engendre une diminution de la synthèse protéique (Sitarska *et al.*, 2015).

Au regard de ces différentes limites, deux premières approches de marquages ont tout de même été publiées en utilisant des extraits d'algues deutérés. Une première approche repose sur l'ajout de la quantité maximale d'extrait d'algue deutéré possible, et d'inhibiteur des alanines transaminases, tout en réalisant un marquage [ $^2\text{H}$ ,  $^{15}\text{N}$ ] et [ $^2\text{H}$ ,  $^{13}\text{C}$ ]. Il en résulte des taux d'incorporation de 76 % et 73 % pour les marquages [ $^2\text{H}$ ,  $^{15}\text{N}$ ] et [ $^2\text{H}$ ,  $^{13}\text{C}$ ], respectivement (Sitarska *et al.*, 2015). Cela se traduit par des spectres

mieux résolus et permettant la réalisation d'expériences de triple résonance. Cependant, il convient de noter que la deutération n'est pas uniforme et les taux d'incorporation sont différents suivant les acides aminés. Ainsi, bien que cette stratégie de marquage uniforme soit efficace, ce premier protocole ne permet pas d'étudier avec succès des protéines de masse moléculaire supérieure à 100 kDa. En conséquence, une deuxième approche a été développée pour l'étude de tels objets (Kofuku *et al.*, 2014). Dans cette étude, une protéine membranaire de haute masse moléculaire a été analysée dans le but d'enregistrer les résonances des groupements méthyles, des méthionines, de manière résolue. Alors que le protocole ne visait pas une deutération uniforme, il s'est avéré qu'une majorité des acides aminés hydrophobes étaient deutérés. Cette approche sera plus amplement introduite dans la partie suivante, mais elle représente une alternative prometteuse pour l'étude des macromolécules.

### **III.3.4. Marquage spécifique $^{13}\text{CH}_3$ des acides aminés portant un groupement méthyle en cellules d'insectes et de mammifères**

Deux méthodes de marquages isotopiques spécifiques d'acides aminés ont été décrites pour ces deux systèmes. La première consiste en un marquage des acides aminés portant un groupement méthyle par ajout, dans un milieu de culture commercial standard, d'un excès d'acides aminés marqués afin d'obtenir des taux d'incorporation satisfaisants. Toutefois, peu d'acides aminés marqués sont suffisamment bon marché pour que cette méthode soit économiquement viable. La deuxième méthode consiste en un remplacement du résidu cible par sa version marquée, dans un milieu déplété en acides aminés.

#### **III.3.4.1. Marquage $^{13}\text{CH}_3$ par ajout en excès des acides aminés portant un groupement méthyle**

Cette méthode est très facilement réalisable : il suffit d'ajouter l'acide aminé à marquer, en excès, dans le milieu de culture commercial complet. Ceci étant, cette approche nécessite une connaissance, au moins approximative, de la quantité de l'acide aminé ciblé non marqué présent dans le milieu de culture. Pour atteindre 90 % d'incorporation, il est nécessaire d'ajouter en excès jusqu'à 9 fois l'acide aminé ciblé

sous sa forme marquée. Cependant, seuls quelques acides aminés marqués peuvent convenir à une telle approche, du fait des coûts importants engendrés.

La méthionine est l'un des premiers résidus qui a été proposé pour ce type de marquage dans les cellules eucaryotes. En effet, outre le coût abordable dans sa version non deutérée, la  $^{13}\text{CH}_3\text{-Met}$  est une sonde RMN intéressante, car la flexibilité de sa chaîne latérale compense la lente rotation des grosses protéines, améliorant, de ce fait, la résolution des spectres (Tugarinov *et al.*, 2003). Ainsi, pour le système d'expression en cellules de mammifères avec l'ajout, en excès, de 1 g/L de  $^{13}\text{CH}_3\text{-Met}$  dans le milieu de culture, les protéines se sont révélées être largement marquées, et présentant un taux d'incorporation supérieur à 80 % (Kofuku *et al.*, 2012 ; Nygaard *et al.*, 2013 ; Solt *et al.*, 2017 ; Kleist *et al.*, 2019).

Cependant, cette stratégie coûteuse n'étant pas applicable pour l'ensemble des résidus, les chercheurs ont développé et optimisé la stratégie basée sur le remplacement des acides aminés.

#### **III.3.4.2. Marquage $^{13}\text{CH}_3$ par remplacement des acides aminés portant un groupement méthyle**

L'approche publiée par Kofuku *et al.* en 2014 s'inscrit dans l'amélioration du marquage d'acides aminés essentiels et non essentiels dans des protéines partiellement deutérées. Dans cette étude les états conformationnels d'un récepteur couplé aux protéines G (GPCR), le récepteur  $\beta_2$ -adrénergique, sont étudiés par RMN. Pour ce faire, en plus du marquage spécifique  $^{13}\text{CH}_3\text{-Met}$ , la protéine a été deutérée en vue d'une forte augmentation de la résolution des signaux RMN. Or, afin d'accroître la qualité et la sensibilité du signal tout en considérant les limites de la deutération en système eucaryote, les auteurs ont préparé (en ajoutant séparément les composants dans une base dépourvue en acides aminés) un milieu de culture comportant des acides aminés deutérés ou non, de l'extrait d'algue deutéré, ainsi que de la  $^{13}\text{CH}_3\text{-Met}$ .

Le choix des acides aminés deutérés est basé sur une réflexion et une connaissance approfondie de la protéine cible. En effet, il s'est avéré qu'une deutération de quelques acides aminés définis sur la base des distances  $^1\text{H}\text{-}^1\text{H}$  entre les groupements  $\text{CH}_3$  des méthionines, mais aussi d'après leur capacité à être incorporés par les protéines,



permettait la diminution de 80 à 90 % des interactions  $^1\text{H}$ - $^1\text{H}$ , et donc une augmentation (d'environ 3 à 7 fois) de la sensibilité du signal des interactions dipôle-dipôle ( $^1\text{H}$ - $^1\text{H}$ ) des groupes  $^{13}\text{CH}_3$ -Met ( $M_{82}$ ,  $M_{215}$  et  $M_{279}$ ). Il faut signaler que limiter la quantité d'acides aminés deutérés est avantageux pour les coûts expérimentaux, mais aussi pour ne pas réduire drastiquement les rendements protéiques.

Sachant que l'activation des GPCR s'accompagne de changements conformationnels dans diverses régions des hélices transmembranaires, la méthionine ne s'est pas avérée être la meilleure sonde pour suivre cette activation. En revanche, les alanines sont une cible de choix puisqu'elles sont fréquemment distribuées dans le cœur hydrophobe et la surface hydrophile, et représentent 9 % des résidus transmembranaires (contre 3 % pour la Met). Par conséquent, un nouveau protocole de marquage combinatoire  $^{13}\text{CH}_3$ -Ala et  $^{13}\text{CH}_3$ -Met au sein de protéines hautement deutérées a été mis au point (Kofuku *et al.*, 2018). Ce protocole a permis un taux d'incorporation de 85 % des  $^{13}\text{CH}_3$ -Met, mais celui des alanines s'est révélé nettement inférieur, car aucun inhibiteur de l'alanine transaminase n'a été ajouté au milieu. Ainsi, le marquage  $^{13}\text{C}$  à la position  $\beta$  de l'alanine était inférieur à 15 %, même en ajoutant 0,5 g/L d'alanine marquée dans le milieu. De plus, les efficacités de marquage  $^2\text{H}$  à la position  $\alpha$  de l'alanine marquée étaient inférieures à 30 % car dans le milieu protoné, les enzymes catalysent l'échange des  $^2\text{H}$  en  $^1\text{H}$  à cette position  $\alpha$ . Afin de limiter ces effets, de la  $\beta$ -chloro-L-alanine a été ajoutée au milieu. Cet inhibiteur, comme la L-cyclo-sérine, est indispensable pour limiter la dilution de l'alanine marquée ajoutée dans le milieu de culture, car il inhibe les alanines transaminases responsables de la synthèse endogène des alanines non marquées (Kato *et al.*, 1991). Ainsi, ce nouveau protocole de marquage a conduit à plus de 45 % d'incorporation des  $\beta$ - $^{13}\text{C}$  Ala, à 80 % d'incorporation des  $^2\text{H}$  à la position  $\alpha$  et à aucune fuite isotopique (Kofuku *et al.*, 2018).

Sur cette même base, d'autres études font état de marquages spécifiques  $^{13}\text{CH}_3$  (Velyvis *et al.*, 2012 ; Miyanoiri *et al.*, 2013). Récemment, une étude a décrit le marquage des groupements méthyles des leucines dans des protéines exprimées en systèmes d'expression bactériens, eucaryotes (cellules d'insectes) ou acellulaires (Dubey *et al.*, 2021). La méthode proposée repose sur un marquage  $^{13}\text{C}$  stéréosélectif et une deutération de la leucine, ce qui atténue le besoin d'une deutération supplémentaire de la protéine, et réduit donc significativement les coûts. Pour valider leur approche, les avantages spectroscopiques de la deutération locale ont été examinés au travers

d'expériences et des simulations de « Forbidden Coherence Transfer ». Il en résulte une réduction de la largeur de raie (50 %) due à la réduction de l'interaction dipôle-dipôle des protons méthyles. Cette méthode de marquage représente donc une nouvelle voie pour l'étude de protéines nécessitant d'être deutérées.

Le marquage spécifique  $^{13}\text{CH}_3$  d'acides aminés en cellules de mammifères n'a été que peu documenté. Une étude récente a, toutefois, décrit la relaxation transversale d'IgG (masse moléculaire de 150 kDa) exprimée en cellules de mammifères (cellules d'hybridome de souris) (Yanaka *et al.*, 2018). Les IgG ont été exprimées dans un milieu de culture « fait maison » et marquées en utilisant la [ $\delta_2$ - $^{13}\text{C}$ ;  $\text{H}_\alpha$ ,  $\text{H}_\beta$ ,  $\text{H}_\gamma$ ,  $\text{H}_{\delta_1-2}\text{H}_7$ ]-Leu ou son homologue non deutéré (acide aminé SAIL - Kainosho et Güntert 2009 ; Miyanoiri *et al.*, 2016). Il en résulte des améliorations significatives de la qualité des spectres pour l'échantillon deutéré sélectivement au niveau des résidus de leucine. En effet, ceci a éliminé des interactions dipolaires  $^1\text{H}$ - $^1\text{H}$  intenses entre les groupements méthyles  $^{13}\text{CH}_3$ . Enfin, une deutération supplémentaire à l'aide de 2 g/L de glucose [ $^2\text{H}_7$ ] a été réalisée mais n'a entraîné que peu ou pas d'améliorations supplémentaires des spectres (notons toutefois que le protocole utilisé n'a pas généré de toxicité du glucose [ $^2\text{H}_7$ ], comme initialement décrit par Murphy *et al.*, 1977).

Actuellement, aucune étude ne fait mention d'un marquage de l'ensemble des acides aminés portant un groupement méthyle, de l'utilisation de cellules humaines comme système d'expression, ni d'une formulation de milieu précise indiquant les quantités d'acides aminés à compléter afin d'obtenir une incorporation maximale, tant pour les acides aminés essentiels que non essentiels. Par conséquent, c'est dans ce contexte que se placent ces travaux de thèse.



## IV. Objectifs de la thèse

Malgré leurs importances croissantes et leurs implications dans de nombreux processus physiologiques et pathologiques, les Sulfs sont des enzymes encore mal caractérisées en raison de leurs propriétés structurales et fonctionnelles uniques, et de la complexité de leurs substrats. Un des problèmes majeurs entravant les progrès dans l'étude des Sulfs est l'absence de toute structure haute résolution. Ainsi, il est difficile de comprendre l'importance de leurs PTMs, leurs spécificités de substrat et leurs implications dans diverses pathologies, de même que d'expliquer les nombreux antagonismes qui sont observés entre les deux enzymes de HSulf. C'est dans ce contexte que des travaux visant à caractériser la structure de HSulf-2 ont été réalisés.

Antérieurement à ce projet de thèse, des essais de cristallogénèse avaient été réalisés, mais n'avaient pas permis d'obtenir un cristal diffractant. De ce fait, un projet de biologie structurale intégrative, réunissant la cristallographie, la RMN et la microscopie électronique, a été proposé afin de caractériser la structure de l'enzyme HSulf-2, ainsi que ses propriétés d'interaction, notamment par RMN.

Dans cet objectif, une collaboration a été initiée entre le groupe SAGAG et NMR-Bio (co-responsable et financeur du projet de thèse à hauteur de 50 %, le reste étant pris en charge par le programme glycoAlps de l'UGA (IDEX Recherche, projet-CDP)), une entreprise hébergée dans le groupe de RMN à l'IBS et spécialisée dans le marquage isotopique de protéines de haute masse moléculaire et leurs études par RMN.

La caractérisation structurale de HSulf-2 par RMN nécessitera au préalable le développement et la caractérisation d'un milieu de culture permettant le marquage isotopique spécifique des méthyles (marquage  $^{13}\text{CH}_3$ ) de protéines exprimées en cellules mammifères (HEK 293). Pour ce faire, un milieu de culture devra être développé et optimisé, afin de permettre une croissance cellulaire et des rendements protéiques optimaux, ainsi que des taux d'incorporation isotopique en adéquation avec les besoins des études RMN (attribution, études d'interaction et de dynamique). Pour ce faire, nous utiliserons comme protéine modèle le domaine N-terminal de la protéine heat shock protein 90 (Hsp90 N-terminal). Cette protéine a été choisie car elle a été largement étudiée et caractérisée par le groupe de RMN de l'IBS (Grenoble) (expression,

purification, activité et structures), et pour laquelle nous disposons déjà de spectres RMN obtenus sur la protéine produite en bactérie.

Une fois la mise au point d'un tel milieu réalisée, celui-ci sera appliqué à l'étude de la protéine HSulf-2 dans le but d'acquérir des données structurales et d'interactions.

En parallèle, une optimisation du protocole de purification de la protéine HSulf-2 sera réalisée afin d'obtenir une protéine pure de tous contaminants (protéique et ADN/ARN) pour son étude par cristallographie et microscopie électronique. Un protocole optimisé de préparation de grilles de (cryo-)microscopie électronique sera ensuite établi pour l'acquisition de jeux de données en vue de la résolution de la structure de la protéine. De front, de nouvelles constructions protéiques de HSulf-2, ainsi que des protocoles de traitement enzymatique de la protéine d'intérêt (déglycosylation, digestion ménagée ...) seront étudiés, afin de limiter la flexibilité et la proportion de régions non structurées de la protéine HSulf-2, et d'identifier des conditions de cristallisation favorables à la formation de cristaux délivrant un diagramme de diffraction exploitable.

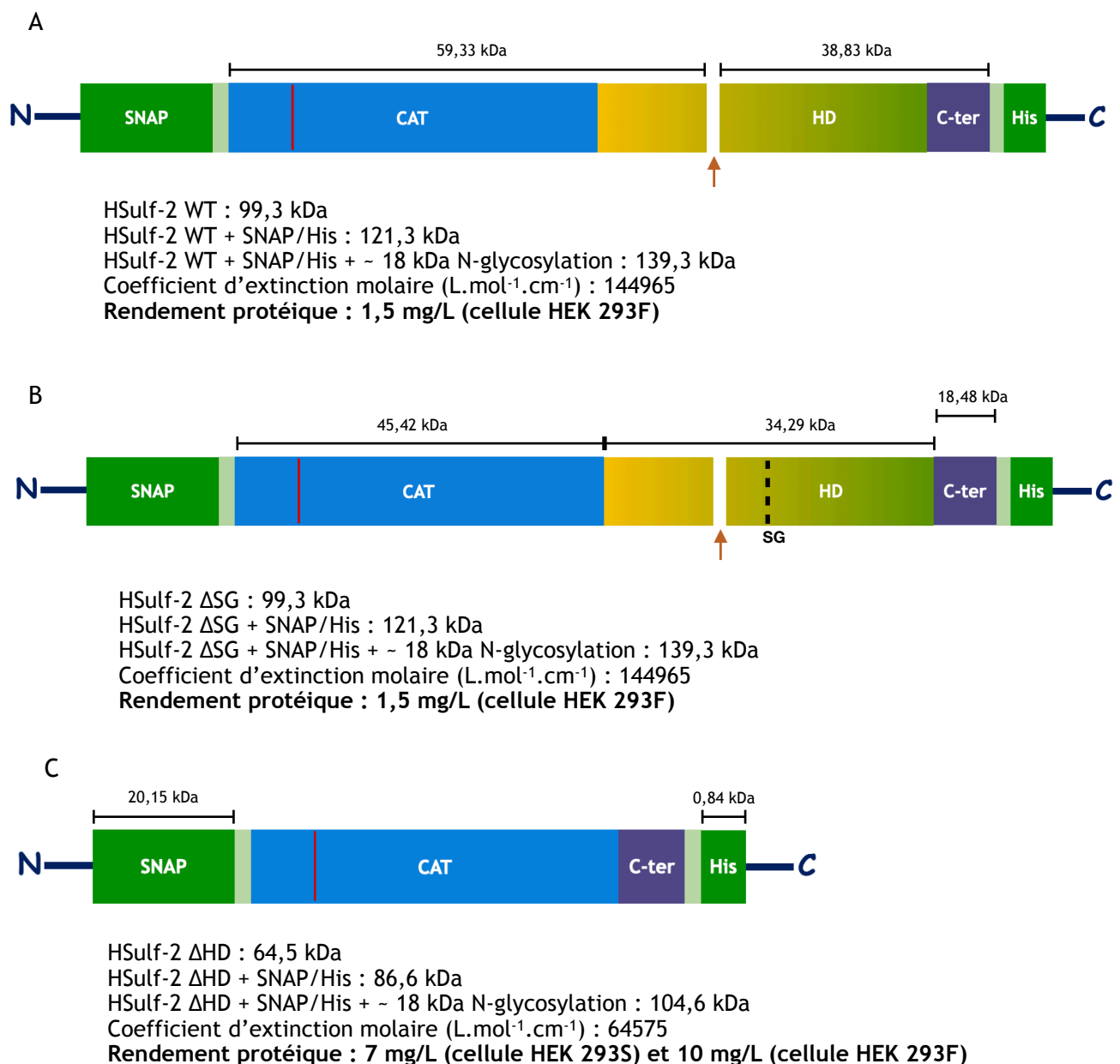
## ***Matériels et méthodes***



## V. Matériels et méthodes

### V.1. Expression et purification de la protéine HSulf-2

Afin de répondre à la problématique de cette thèse, différentes constructions de la protéine HSulf-2 (cf. séquences en Annexe 1B à F p. 233-236) ont été conçues (Figure 15), exprimées et purifiées. L'ensemble de ces protéines s'exprime et se purifie de la même manière, tout comme la caractérisation de leur activité ou les différents traitements qui peuvent leur être appliqués. Ainsi, seule l'expression, la purification et la caractérisation de la protéine HSulf-2  $\Delta$ SG sera détaillée dans la suite de cette section.



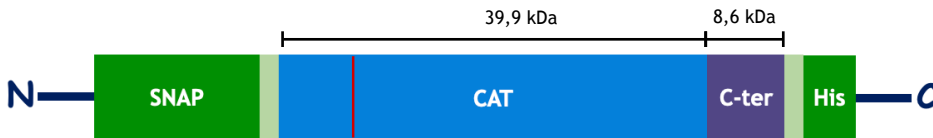


D



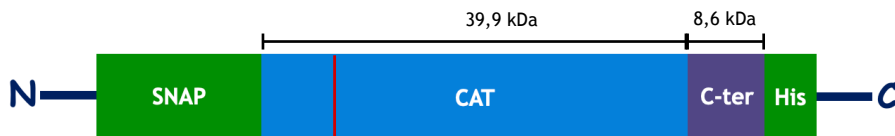
HSulf-2  $\Delta$ HDc : 82,5 kDa  
 HSulf-2  $\Delta$ HDc + SNAP/His : 107,2 kDa  
 HSulf-2  $\Delta$ HDc + SNAP/His + ~ 18 kDa N-glycosylation : 125,2 kDa  
 Coefficient d'extinction molaire ( $L \cdot mol^{-1} \cdot cm^{-1}$ ) : 82470  
 Rendement protéique : 1 mg/L (cellule HEK 293F)

E



HSulf-2  $\Delta$ HD V1 : 49,6 kDa  
 HSulf-2  $\Delta$ HD V1 + SNAP/His : 77 kDa  
 HSulf-2  $\Delta$ HD V1 + SNAP/His + ~ 18 kDa N-glycosylation : 95 kDa  
 Coefficient d'extinction molaire ( $L \cdot mol^{-1} \cdot cm^{-1}$ ) : 68230  
 Rendement protéique : 2 mg/L (cellule HEK 293F)

F



HSulf-2  $\Delta$ HD V2 + SNAP/His : 74,7 kDa  
 HSulf-2  $\Delta$ HD V2 + SNAP/His + ~ 18 kDa N-glycosylation : 92,7 kDa  
 Coefficient d'extinction molaire ( $L \cdot mol^{-1} \cdot cm^{-1}$ ) : 87960  
 Rendement protéique : 2 mg/L (cellule HEK 293F)

Figure 15 : Constructions de la protéine HSulf-2 utilisées pour les études structurales. (A) HSulf-2 WT : protéine totale portant une chaîne de GAG sur le domaine HD et pouvant subir un clivage furine (flèche orange). Le site actif est présent sur le domaine CAT (trait rouge). Pour finir, les rectangles verts et verts clairs sont, respectivement, l'étiquette d'expression (SNAP) et de purification (His), et les sites de coupure au virus de la gravure du tabac (TEV). (B) HSulf-2  $\Delta$ SG : le site  $S_{559}G$  portant la chaîne GAG a été muté (perte de la chaîne de CS, pointillés noirs). (C) HSulf-2  $\Delta$ HD : le domaine HD a été éliminé. (D) HSulf-2  $\Delta$ HDc : le domaine HD a été réduit afin d'éliminer la région centrale non structurée (séquence tronquée entre  $D_{483}$  et  $C_{636}$ ), remplacée par un linker SGSGGS permettant de lier les deux parties du domaine HD. Les ponts disulfures intra-HD récemment identifiés ont été conservés (cf. II.3.1. p. 83). (E) HSulf-2  $\Delta$ HD V1 : cette construction est dérivée de la construction C., où quelques dizaines d'acides aminés ont été éliminés sur les extrémités N et C-terminales des domaines CAT et C-ter. (F) HSulf-2  $\Delta$ HD V2 : identique à HSulf-2  $\Delta$ HD V1, mais ici les sites de clivées à la TEV ont été éliminés.

### V.1.1. Expression et purification de la protéine HSulf-2 $\Delta$ SG

La protéine HSulf-2  $\Delta$ SG (cf. Figure 15B p. 123) est exprimée de manière constitutive par des cellules HEK 293F transfectées avec un vecteur PcDNA3 codant pour l'enzyme, puis sélectionnées à la généticine pour obtenir une lignée cellulaire transfectée stable. Les cellules HEK 293 sont d'origine humaine (cellules épithéliales de rein embryonnaire humain), et permettent donc la modification post-traductionnelle de la protéine HSulf-2.

Ainsi, une fois les cellules (HEK 293F – HSulf-2  $\Delta$ SG) amplifiées jusqu'à un volume de 700 mL à une densité de  $1,5 \cdot 10^6$  cellules/mL, les cellules sont cultivées durant 4 jours pour permettre l'expression de la protéine (incubateur paramétré à 37°C et 8 % de CO<sub>2</sub>). Ce temps d'expression a été défini après la réalisation de tests de viabilité cellulaire et permet une viabilité cellulaire >99 % (données non présentées). Au bout de ce temps d'expression, le milieu conditionné est récupéré (en réalisant une première centrifugation du milieu de culture à 310 g puis une deuxième à 2000 g, afin d'éliminer les cellules et les débris cellulaires) et purifié.

La purification de HSulf-2  $\Delta$ SG nécessite trois étapes successives : une chromatographie échangeuse d'ions, une chromatographie d'affinité et une chromatographie d'exclusion. Le milieu conditionné est chargé (à 4°C, à un débit d'environ 1 mL/min), grâce à une pompe péristaltique sur une colonne HiTrap sulfopropyl (SP) High Performance (5 mL) échangeuse de cations. L'utilisation de cette colonne est compatible avec le point isoélectrique de la protéine HSulf-2  $\Delta$ SG (9,31) et le pH du milieu conditionné (pH 8) : la charge globale de la protéine d'intérêt étant positive, ce qui permet sa fixation sur la résine Sepharose SP. La colonne HiTrap SP High Performance est ensuite connectée à un appareil de chromatographie AKTA Pure (GE Healthcare Life Sciences) afin de laver la colonne (Tampon A) et d'éluer la protéine par l'application d'un gradient de sel (NaCl 0,1 M à 1 M sur un volume de 50 mL, à 4°C). Pour ce faire, deux tampons sont utilisés : le tampon A (Tris 50 mM pH 8, NaCl 100 mM) et le tampon B (Tris 50 mM pH 8, NaCl 1 M). Des fractions de 1 mL sont ensuite collectées et analysées par électrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium (SDS-PAGE). Pour l'analyse des fractions par SDS-PAGE en condition dénaturantes, les réactifs Invitrogen™ Agent réducteur d'échantillons NuPAGE™ (10X) et Invitrogen™ Tampon d'échantillon Lithium dodecyl sulfate (LDS) NuPAGE™ (4X) sont rajoutés aux échantillons (de 15  $\mu$ l). Ainsi, une

fois les échantillons préparés (dénaturés, à 100°C pendant 10 minutes, et en présence du bleu de charge), ces derniers sont chargés sur un gel d'acrylamide 10 % et la migration des échantillons est réalisée grâce au système Bio-Rad Mini-Protein® Tetra System, dans un tampon : Tris 50 mM, Glycine 380 mM et SDS 20 % (v/v). Le gel est coloré au bleu de Coomassie (InstantBlue® Coomassie Protein Stain) afin de révéler la ou les espèces protéiques composant les échantillons. Pour finir, les fractions contenant la protéine d'intérêt sont rassemblées et concentrées par centrifugation jusqu'à 1 mL à 16100 g, dans un Amicon® Ultra (ici, limite de séparation : 30 kDa). Notons que l'ensemble des analyses par SDS-PAGE présentées par la suite (de même pour la concentration de toutes les protéines) suivront le même protocole.

Le produit d'élution de la chromatographie échangeuse d'ions (concentré à 1 mL) est ensuite incubé avec 1 mL de résine Ni-NTA équilibrée en tampon A. Après 30 minutes à 4°C, la fraction non retenue est collectée puis différentes étapes de lavages (à 4°C) sont réalisées. Les deux premiers lavages de la résine (où est fixée HSulf-2 ΔSG) sont réalisés avec 5 mL de phosphate de potassium (200 mM) et 5 mL de chlorure de potassium (Tris 50 mM pH 8, KCl 1 M) afin d'éliminer toutes contaminations à l'ADN de la protéine cible. Puis, un dernier lavage de la résine est réalisé avec 10 mL de tampon C (Tris 50 mM pH 8, NaCl 200 mM) + 5 mM d'imidazole, dans le but d'éliminer toutes protéines contaminantes interagissant de manière non spécifique avec les ions Ni<sup>2+</sup> de la résine. Pour terminer, la protéine est éluée (4°C) avec 15 mL de tampon C supplémenté avec 250 mM d'imidazole. Les fractions collectées (1,5 mL) sont analysées par SDS-PAGE, puis les fractions d'intérêt sont rassemblées et concentrées jusqu'à 350 µL.

La dernière étape consiste à cliver les étiquettes (SNAP et His). Pour ce faire, 10 % de TEV (m/m) sont ajoutées à l'échantillon protéique concentré. Après, une digestion sur la nuit à 4°C et une centrifugation de l'échantillon à 16100 g pendant 10 minutes, la solution protéique est injectée sur colonne d'exclusion stérique Superdex 200 10/300 GL (dans un environnement à 4°C), préalablement équilibrée en tampon C. De la même manière, les fractions collectées sont analysées par SDS-PAGE, et les fractions correspondant à la protéine HSulf-2 ΔSG pure sont rassemblées et concentrées. La protéine est alors quantifiée par mesure de l'absorbance à 280 nm à l'aide d'un appareil Thermo Scientific™ NanoDrop™ 2000 Spectrophotometers, en saisissant le coefficient d'extinction molaire (L.mol<sup>-1</sup>.cm<sup>-1</sup>) et la masse moléculaire (kDa), soit 144,965 et 99,292 respectivement. Notons que l'ensemble des quantifications protéiques sont réalisées de la même manière, avec toutefois le coefficient d'extinction molaire et la masse

moléculaire qui varient car propres à chaque protéine. Enfin, l'échantillon est congelé par la technique de « Flash-freezing » dans de l'azote liquide avant d'être stockée à -20°C.

### V.1.2. Purification de la protéine HSulf-2 ΔSG clivée à la trypsine

Une stratégie de réduction des régions flexibles/désordonnées de HSulf-2 ΔSG a été mise au point et repose sur une digestion à la trypsine de la protéine. La protéine purifiée a été incubée avec un rapport molaire 1:50 de trypsine (1 molécule de trypsine pour 50 molécules de HSulf-2), et après une incubation de 40 minutes à 37°C, le produit de la réaction est centrifugé et injecté sur colonne Superdex 200 10/300 GL. Après l'analyse du SDS-PAGE, les fractions correspondant à une forme protéique stable et pure ont été concentrées jusqu'à ~ 3,5 mg/mL dans un volume final de ~ 200 µL.

Afin d'essayer de structurer davantage la protéine, la solution protéique a été divisée en deux et 1 mM de sulfocalix-[4]-arène (v/v) a été rajouté à un de ces deux échantillons. Le sulfocalixarène, un macrocycle synthétique connu pour améliorer l'assemblage et la cristallisation des protéines (Alex *et al.*, 2019) grâce à ses groupes phénols, pourrait aider à stabiliser la protéine. Ces deux échantillons protéiques (avec ou sans ligand), ont donc été utilisés pour réaliser un large criblage de conditions favorables à la formation de cristaux diffractants (criblage réalisé par la plateforme de cristallographie HTXlab (EMBL, Grenoble) dans six plaques commerciales de cristallogénèse (Tableau 4).

Plaques	Descriptions
The Classics Suite (Qiagen)	Produits chimiques les plus populaires en cristallographie (plaque anciennement connu sous le nom de plate_1_qiagen)
The JCSG-plus™	Combinaisons de polyéthylènes glycols (PEG), de sels, d'acides organiques neutralisés et de précipitants organiques sur une plage de pH de 4,0 à 10,5, et comprenant une gamme d'additifs salins.
The PACT (MD)	Combinaisons de PEG/ions et PEG/pHs
The PEGs I (Qiagen)	Sels faiblement concentrés contre PEG 3350 (PEG-ion) + une large gamme de PEG (200 à 20 000) à 4 pH
Salt-Grid_hampton	Plaque composée de sels, polymères organiques et une variation de pH (ex : Ammonium Sulfate, Sodium Malonate de « Hampton Research » et un screen propre à HTXlab de Sodium Formate) (plaque anciennement connue sous le nom de plate_4_hampton)
Wizard I & II (Rigaku)	Matrice aléatoire pour la cristallisation de macromolécules biologiques

Tableau 4 : Caractéristiques des plaques de cristallogénèse.

### V.1.3. Test d'activité de la protéine HSulf-2 $\Delta$ SG et HSulf-2 $\Delta$ SG trypsinée

L'activité de la protéine peut être caractérisée par un test spectrofluorimétrique (4MUS) permettant de mesurer l'activité arylsulfatase de la protéine. En effet, la désulfatation du substrat 4MUS par l'enzyme conduit à la formation d'un produit (4MU) fluorescent. Par conséquent, l'activité arylsulfatase de la protéine HSulf-2 peut être suivie par mesure de cette fluorescence (excitation à 340-360 nm ; émission à 460 nm). Pour étudier l'activité arylsulfatase de la protéine purifiée, les protéines pures HSulf-2  $\Delta$ SG, ainsi que HSulf-2  $\Delta$ SG traitée à la trypsine, sont incubées avec 50  $\mu$ L de solution 4MUS (4MUS 20mM resuspendus dans un tampon Tris 50 mM pH 7,5, MgCl<sub>2</sub> 20 mM) et 2  $\mu$ g de protéine pure q.s.p. (quantité suffisante pour) 100  $\mu$ L de tampon Tris 50 mM pH 7,5 (à 37°C pendant 7 heures). Une condition témoin (où le substrat est incubé sans enzyme) est réalisée en parallèle. Une fois la réaction terminée, une mesure de fluorescence pour chaque condition est réalisée et permet de valider (après soustraction du contrôle négatif) l'activité de la protéine.

### V.1.4. Expression, purification de l'enzyme « Peptide N-glycosidase F » (PNGase F) et traitement de la protéine HSulf-2

Afin de réduire la flexibilité et l'hétérogénéité engendrée par la présence d'environ 18 kDa de N-glycosylations, un protocole d'expression et de purification de la PNGase F a été mis au point.

Des souches bactériennes *E. coli* BL21 (DE3)-RIL ont été transformées avec le vecteur codant pour la PNGase F (pBluescript KS(-), Plasmid #114274, Addgene). Puis, l'expression de l'enzyme a été réalisée à partir de cette souche transformée sur une culture de 1 litre de milieu LB, avec une induction par ajout d'IPTG (0,5 mM) lors de la phase de croissance exponentielle de la culture (DO = 0,7). Une fois l'expression terminée (18 heures à 22°C), la culture est centrifugée et le culot resuspendu dans 50 mL de tampon PNGase F (Tris 50 mM pH 7,8, NaCl 350 mM), puis lysée par sonication (1 seconde pulse, 4 secondes d'attente, ceci pendant 3 minutes à une amplitude de 40 %). Après centrifugation (20 minutes à 41400 g) et filtration (0,45  $\mu$ m) du surnageant, ce dernier est chargé (5 mL/min) sur une colonne Ni-NTA (Ni-NTA Superflow, QIAGEN) et un premier lavage est réalisé avec le tampon PNGase F + 10 mM d'imidazole (20 mL de

lavage). Enfin, la protéine cible est éluée avec le tampon PNGase F + 250 mM d'imidazole (20 mL) et concentrée jusqu'à 500 µL, avant d'être injectée sur une colonne Superdex 200 10/300 GL. Après analyse par SDS-PAGE, les fractions correspondant à la protéine purifiée sont concentrées. La protéine est quantifiée par mesure d'absorbance, puis congelée par « Flash-freezing » dans de l'azote liquide, avant d'être stockée à -20°C.

La validation de l'expression et de la purification de la PNGase F a été réalisée avec un test de déglycosylation en conditions non dénaturantes sur la protéine HSulf-2 ΔHD. Pour cela, 10 % de PNGase F (m/m) ont été ajoutées à l'échantillon protéique et incubées sur la nuit à 4°C. Cette condition a été privilégiée car les protéines étudiées au cours de cette thèse sont déjà traitées avec 10 % de protéase TEV (m/m) sur la nuit à 4°C afin de cliver les étiquettes SNAP et His. L'utilisation de la PNGase F dans ces mêmes conditions permettrait donc de réaliser les deux digestions simultanément.

## **V.2. Développement du milieu de marquage spécifique des <sup>13</sup>CH<sub>3</sub>-acides aminés**

### **V.2.1. Composition du mélange réactionnel d'acides aminés**

Un milieu de culture sans acides aminés doit être disponible et adapté à la lignée cellulaire HEK 293. Cependant, pour ces cellules, l'industriel (Thermofisher, Gibco®) ne commercialise pas de milieu FreeStyle 293 déplété. Par conséquent, un milieu de culture alternatif, le ESF-SFM (référence : 98-003, Expression System) qui existe sous deux versions (avec et sans acides aminés), a été utilisé.

Afin d'identifier une formulation du milieu de marquage la plus favorable possible (en termes de rendements protéiques et d'incorporations isotopiques), différents tests d'optimisation ont été réalisés. Tous ces tests ont été réalisés en comparant les rendements protéiques obtenus pour la protéine modèle Hsp90 N-terminal exprimée dans le milieu de marquage développé et dans le milieu de culture commercial (FreeStyle 293).

Hsp90 N-terminal est un domaine de la protéine Hsp90 (90 kDa). Cette dernière, présente chez de nombreux organismes (dont l'Homme), est un élément crucial de la carcinogénèse et sert de chaperonne moléculaire pour faciliter la maturation de protéines tout en protégeant les cellules contre le stress induit par la température

(Magwenyane *et al.*, 2020). Hsp90 est dimérique et se compose de trois domaines : (1) le domaine N-terminal, (2) le domaine intermédiaire et (3) le domaine C-terminal, où chaque domaine a un rôle fonctionnel dans l'activité de Hsp90. Le domaine N-terminal contient la poche de liaison à l'ATP, qui sera indispensable à la réalisation du cycle de l'ATPase et donc à l'activité d'Hsp90 (Magwenyane *et al.*, 2020).

Sachant que Hsp90 est une protéine de haute masse moléculaire (d'autant plus que la protéine est dimérique), les possibilités d'études par RMN sont limitées car, au vu des limites de la deutération en système eucaryote (cf. III.3.3.3. p. 113), seuls des marquages spécifiques peuvent être réalisés (Karagöz *et al.*, 2010). En revanche, le domaine N-terminal de la protéine (26,5 kDa) est d'importance majeure pour l'activité de la protéine et est en adéquation avec la réalisation de marquages, tant spécifiques qu'uniformes, permettant ainsi d'acquérir de nombreuses informations, notamment sur l'inhibition par déplacement de l'ATP au moyen de liaisons compétitives (Delmotte *et al.*, 1953 ; Schulte *et al.*, 1998 ; Karagöz *et al.*, 2010 ; Magwenyane *et al.*, 2020). De surcroît, Hsp90 N-terminal (exprimée en bactérie) est une protéine très largement étudiée et caractérisée par le groupe de RMN et l'entreprise NMR-Bio présents au sein du laboratoire, ce qui implique la mise à disposition de protocoles d'expression et de purification validés, ainsi que de nombreuses données RMN (différents marquages et expériences RMN).

La composition des milieux de culture (tels que le milieu FreeStyle 293 ou le milieu ESF-SFM) n'étant pas connue, le point de départ pour développer le mélange réactionnel d'acides aminés a été le milieu MEM Gibco®. Puis différents milieux, dont la formulation est connue (comme le milieu DMEM, Jong-Seon *et al.*, 2013 ; Advanced D-MEM/F-12, Gibco®), ont servi de points de comparaison (Tableau 5).

Composition	Milieu développé (en mg/L)	MEM Gibso® (en mg/L)	Advanced D-MEM/ F-12 Gibso® (en mg/L)	DMEM (Jong- Seon <i>et al.</i> , 2013) (en mg/L)
L - Arg	126	126	147	84
L - Cys	24	31	48	63
L - His	42	42	31,4	42
L - Leu	52	52	59	105
L - Lys	72	72,5	91	146
L - Phe	33	32	35	66
L - Trp	10	10	9	16
L - Thr	47	48	53	95
L - Tyr	130	52	55	104
L - Met	15	15	17	30
L - Val	46	46	52	94
L - Ile	52	52	54	105
L - Gly	7,5	7,5	18,7	30
L - Asn	13	15	7,5	-
L - Asp	13	13	6,5	-
L - Pro	11	11,5	17	-
L - Ser	10	10,5	26	42
L - Gln	300	292	868	584
L - Glu	12	14	7,3	-
L - Ala	8,9	8,9	4	-

Tableau 5 : Composition du mélange réactionnel et comparaison avec des milieux dont la composition en acides aminés est connue.



Avec cette formulation théorique déterminée, un protocole d'expression faisant intervenir différentes concentrations d'acides aminés (Tableau 6) a été produit dans le but d'optimiser cette composition en acides aminés et d'obtenir les rendements protéiques les plus importants possible.

Acides aminés essentiels (EAA) Mix	Quantité en acide aminé (g/L)
L - Arg	12,6
L - Cys	2,4
L - His	4,2
L - Lys	7,2
L - Phe	3,3
L - Trp	1

Acides aminés non essentiels (NEAA) Mix	Quantité en acide aminé (g/L)
L - Gly	0,75
L - Asn	1,3
L - Asp	1,3
L - Pro	1,15
L - Ser	1,05
L - Glu	1,24
L - Gln	30

Acides aminés essentiels (Met) Mix	Quantité en acide aminé (g/L)
L - Met	1,5

Acides aminés essentiels (Thr) Mix	Quantité en acide aminé (g/L)
L - Thr	4,7

Acides aminés essentiels (Ile) Mix	Quantité en acide aminé (g/L)
L - Ile	5,2

Acides aminés non essentiels (Ala) Mix	Quantité en acide aminé (g/L)
L - Ala	0,89

Acides aminés essentiels (Val) Mix	Quantité en acide aminé (g/L)
L - Val	4,6

Acides aminés essentiels (Leu) Mix	Quantité en acide aminé (g/L)
L - Leu	5,2

Acides aminés essentiels (Tyr)	Quantité en acide aminé (g/L)
L - Tyr*	0,13

Tableau 6 : Composition des mélanges réactionnels d'acide aminé. Chacun des huit mélanges réactionnels (jaune) présente une concentration de 100x (100 fois la concentration en acides aminés détaillée dans le tableau 5, cf. p. 131). \* La L-tyrosine (gris) est ajoutée sous forme de poudre, directement dans le milieu final, car elle est très peu soluble.

Huit solutions d'acides aminés ont été préparées : un mélange réactionnel d'acides aminés essentiels, un mélange réactionnel avec les acides aminés non essentiels et six solutions concentrées (également appelées, par la suite, mélange réactionnel par souci de compréhension et de simplification) de chacun des acides aminés portant un groupement méthyle, de façon à pouvoir marquer indépendamment ou bien de façon combinatoire ces résidus.

Ces mélanges réactionnels sont préparés dans de l'eau distillée (pH ajusté à 8), à une concentration 100 fois supérieure à la concentration finale utilisée dans le milieu de culture, et stérilisés par filtration sur filtre 0,22 µm. Notons que la L-Tyr est le seul acide aminé non marqué à être supplémenté au milieu sous forme libre (poudre), car sa solubilisation étant difficile aucune solution concentrée ne peut être préparée. En résumé, le milieu de marquage final préparé est composé du milieu ESF-SFM déplété (sans acides aminés) et recomplémenté avec l'ensemble des mélanges réactionnels, sauf celui de la L-Met qui est substitué par de la  $^{13}\text{CH}_3$ -Met libre (dans le cas d'un marquage des méthionines, par exemple), ainsi que de la L-Tyr en poudre. Puis, une fois que le milieu est homogène (résidus libres dissous), son pH est ajusté à pH 8 et est stérilisé par filtration (filtration 0,22 µm, sous hotte en environnement stérile).

Afin d'identifier si la quantité en acides aminés définie dans le mélange réactionnel permet de maintenir la croissance cellulaire et les rendements protéiques, un test d'optimisation a été réalisé avec différentes quantités de ce mélange. Pour ce test, 28 mL de culture cellulaires (cellules HEK 293F à une densité de  $1,5 \cdot 10^6$  cellules/mL) ont été transfectées avec le vecteur codant pour la protéine Hsp90 N-terminal. Cette culture a ensuite été divisée en sept (7 fois 4 mL de cellules transfectées), puis chaque culture a été centrifugée (5 minutes à 410 g) et lavée (culot cellulaire repris dans 10 mL de PBS puis re-centrifugé) afin d'éliminer toute trace du milieu FreeStyle 293. Pour finir, les sept culots de cellules transfectées et lavées ont été resuspendus dans les différents milieux d'expression (Tableau 7).

Conditions	Temoin +	1	2	3	4	5	6	Temoin -
Volume culture	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL Cellules non transfectées
Milieu de culture	FreeStyle 293	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	FreeStyle 293
Concentration des mélanges réactionnels pour reconstituer le milieu	-	1/2x*	1x*	2x*	4x*	6x*	8x*	-
L-Tyrosine**	-	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	-

Tableau 7 : Optimisation de la concentration en acides aminés. \* Les huit mélanges réactionnels d'acides aminés (cf. Tableau 6 p. 132) dilués au 1/200<sup>ème</sup> = 1/2x ; au 1/100<sup>ème</sup> = 1x ; au 1/50<sup>ème</sup> = 2x ; au 1/25<sup>ème</sup> = 4x ; au 1/16,7<sup>ème</sup> = 6x ; au 1/12,5<sup>ème</sup> = 8x. \*\* 3 mg de L-Tyr sont dilués dans 3 mL de milieu ESF-SFM déplété, avant d'ajuster le pH (du milieu + tyrosine) à 8, et de stériliser cette solution par filtration (0,22 µm, sous hotte en environnement stérile).

Après quatre jours d'expression, un comptage des cellules est réalisé pour chaque condition, et une même quantité de cellules est prélevée dans chaque milieu (1 000 000 de cellules) et centrifugée (5 minutes à 310 g), ceci afin de quantifier et comparer l'expression de la protéine Hsp90 N-terminal. Une fois la lyse des cellules (1 mL de tampon de lyse : Tris HCl 25 mM pH 7,4, NaCl 150 mM, SDS 0,1 %, IGEPa 1 %, DNase/RNase 1 %, pendant 20 minutes à 4°C) et la centrifugation du produit de la lyse réalisées (20 minutes à 16100 g), une analyse par western blot du surnageant est effectuée. Pour cela, les différents échantillons (15 µl) sont supplémentés avec les réactifs Invitrogen™ NuPAGE™ Sample Reducing Agent (10x) et LDS Sample Buffer (4x), dénaturés (à 100°C pendant 10 minutes) et chargés sur un gel d'acrylamide 15 %. La migration des échantillons est réalisée sur un système Bio-Rad Mini-Protein® Tetra System, dans un tampon Tris 50 mM, Glycine 380 mM et SDS 20 % (v/v). Les protéines sont par la suite transférées sur une membrane Amersham™ Hybond™ P 0,45 PVDF, grâce au système Bio-Rad Mini Trans-Blot® (100 V, 200 mA, pendant 1,25 heures), dans un tampon Tris 50 mM, Glycine 380 mM et Ethanol 15 %. La membrane est ensuite saturée en protéines (PBS 1x,

lait écrémé en poudre 5 % et Tween20 0,05 %) et incubée avec un anticorps (ici, seul un anticorps est utilisé :  $\alpha$ -His-HRP (« Horseradish Peroxidase »), A7058, Sigma (dil. 1/1000), PBS 1x et lait écrémé en poudre 5 %). Après un lavage extensif de la membrane (PBS 1x, Tween20 0,05 %), un agent révélat (Pierce™ Western Blot Signal Enhancer, Thermo Scientific™) est appliqué pour une visualisation par chimioluminescence (Chemi Hi Sensitivity ChemiDoc™ MP Imaging System Bio-Rad). Notons que l'ensemble des analyses par western blot présentées par la suite suivront le même protocole (à l'exception des anticorps qui peuvent changer).

L'influence du glucose et des vitamines a également été testée (cf. Annexe 2 p. 237). Après quatre jours d'expression, un comptage des cellules est réalisé pour chaque condition et un même volume de milieu de culture est prélevé (volume de 1,5 mL) afin de quantifier et comparer l'expression de la protéine Hsp90 N-terminal. Ainsi, une fois la lyse des cellules (20 minutes à 4°C) et la centrifugation du produit de la lyse réalisées (20 minutes à 16100 g), une analyse des taux d'expression a été réalisée par western blot du surnageant (anticorps :  $\alpha$ -His-HRP, A7058, Sigma).

D'après la littérature (Gossert *et al.*, 2011 ; Zhang *et al.*, 2017), l'ajout d'extrait de levure (dialysé ou non) génère une croissance cellulaire et une expression protéique identiques une fois les cellules d'insectes transférées dans le milieu de marquage spécifique (cf. III.3.3. p. 113). Un dernier test d'optimisation a donc été mené sur cette base, en ajoutant les mélanges réactionnels d'acides aminés et de l'extrait de levure dialysé ou non (cf. Annexe 3 p. 238). Après quatre jours d'expression, un comptage des cellules est réalisé pour chaque condition et une même quantité de cellules est prélevée dans chaque milieu (1 000 000 de cellules) pour quantifier et comparer l'expression de la protéine Hsp90 N-terminal par western blot (anticorps :  $\alpha$ -His-HRP, A7058, Sigma), comme précédemment décrit.

## V.2.2. Protocole d'expression et de purification de la protéine Hsp90 N-terminal

Dans le but de valider la composition du milieu développé et de comparer les rendements protéiques, la purification de la protéine Hsp90 N-terminal, exprimée dans le milieu de culture standard (FreeStyle 293) et dans le milieu de marquage spécifique (ESF-SFM déplété et reconstitué en acides aminés non marqués), a été réalisée.

Dans un premier temps, 100 mL de cellules HEK 293F, à une densité de  $1,5 \cdot 10^6$  cellules/mL, ont été transfectées avec l'ADNc de la protéine cible dans le milieu FreeStyle 293F (200 µg PEI pour 100 µg d'ADNc Hsp90 N-terminal). Après quatre heures sous agitation en incubateur (temps nécessaire pour permettre la transfection des cellules), la culture a été centrifugée (5 minutes à 310 g) et les cellules lavées au PBS (10 mL puis re-centrifugées) afin d'éliminer toute trace de milieu FreeStyle 293. En parallèle, le milieu de culture destiné au marquage spécifique des protéines a été composé. Pour cette expérience visant simplement à étudier les niveaux d'expression protéiques, seuls des acides aminés non marqués ont été ajoutés au milieu de culture. Ainsi, les cellules (transfectées et lavées) ont été re-suspendues dans le milieu ci-après (Tableau 8), pour une expression sur quatre jours.

Condition	Hsp90 N-terminal non marquée
Volume culture	100 mL
Milieu de culture	ESF-SFM déplété
L-Tyrosine	13 mg
Concentration des mélanges réactionnels pour reconstituer le milieu	1x

Tableau 8 : Composition du milieu d'expression. Les acides aminés supplémentés au milieu sont non marqués.

Au bout de quatre jours d'expression, la culture est centrifugée (5 minutes à 310 g) et les cellules lysées (20 mL de tampon de lyse, cf. p. 134). Le lysat est de nouveau centrifugé (20 minutes à 41400 g) et le surnageant filtré (0,22 µm) avant d'être dilué avec 80 mL de tampon HSP A (Sodium phosphate 20 mM pH 7,4, NaCl 500 mM, MgCl<sub>2</sub> 5 mM, 2-mercaptoéthanol 0,05 %). Cette nouvelle solution (de 100 mL) est chargée sur une colonne Ni-NTA (1 mL de résine), puis deux lavages sont réalisés, le premier avec le tampon HSP A (10 mL) et le second avec le tampon HSP A + 17 mM d'imidazole (10 mL). Pour finir, 15 mL de tampon HSP A + 250 mM d'imidazole sont ajoutés à la colonne afin d'éluer les protéines.

Les fractions sont analysées par SDS-PAGE. Celles contenant la protéine d'intérêt sont rassemblées et concentrées jusqu'à 500 µL, puis centrifugées et injectées sur une colonne Superdex 200 10/300 GL (équilibrée en tampon GF : HEPES 20 mM pH 7,5, NaCl 150 mM, Tris(2-carboxyethyl)phosphine (TCEP) 1 mM). Les fractions correspondant à la protéine purifiée ont été concentrées. La protéine est alors quantifiée par mesure de l'absorbance à 280 nm, puis congelée et stockée à -20°C.

Afin de comparer les rendements protéiques engendrés par les différents milieux de culture (milieu commercial ou milieux développés pour le marquage isotopique), la protéine Hsp90 N-terminal a aussi été exprimée et purifiée dans un milieu de culture marqué en ajoutant 100 mg/L de *U*-[<sup>1</sup>H], [<sup>13</sup>C<sup>1</sup>H<sub>3</sub>]<sup>ε</sup>-Met (299146, Sigma Aldrich) ; 100 mg/L de *U*-[<sup>1</sup>H], [<sup>13</sup>C<sup>1</sup>H<sub>3</sub>]<sup>β</sup>-Ala (CLM-117-pk) ; 100 mg/L de <sup>15</sup>N-2,3,3,4,4,4-(D5)-proS-<sup>13</sup>CH<sub>3</sub>-L-Val (NMR-Bio) ; 100 mg/L de <sup>15</sup>N-2,3,4,4-(D4)-3-méthyl-(D3) 5-<sup>13</sup>CH<sub>3</sub>-L-Ile<sup>δ1</sup> (NMR-Bio) et 100 mg/L de 2,3-(D<sub>2</sub>) 4-(<sup>13</sup>CH<sub>3</sub>)-L-Thr (NMR-Bio). La nomenclature de ces acides aminés est simplifiée pour la suite en : <sup>13</sup>CH<sub>3</sub>-Met, <sup>13</sup>CH<sub>3</sub>-Ala, <sup>13</sup>CH<sub>3</sub>-Val<sup>proS</sup>, <sup>13</sup>CH<sub>3</sub>-Ile<sup>δ1</sup>, <sup>13</sup>CH<sub>3</sub>-Thr. Pour cela, les cellules HEK 293F (transfectées et lavées) sont re-suspendues dans le milieu de marquage suivant (Tableau 9).

Condition	Hsp90 N-terminal marquée $^{13}\text{CH}_3\text{-Met}/\text{-Ile}^{\delta 1}/\text{-Val}^{\text{proS}}/\text{-Thr}/\text{-Ala}$
Volume culture	100 mL
Milieu de culture	ESF-SFM déplété
Concentration des mélanges réactionnels EAA, NEAA et Leu pour reconstituer le milieu	1x
L-Tyrosine*	130 mg/L
$^{13}\text{CH}_3\text{-Met}^*$	100 mg/L
$^{13}\text{CH}_3\text{-Ile}^{\delta 1*}$	100 mg/L
$^{13}\text{CH}_3\text{-Val}^{\text{proS}*}$	100 mg/L
$^{13}\text{CH}_3\text{-Thr}^*$	100 mg/L
$^{13}\text{CH}_3\text{-Ala}^*$	100 mg/L
$\beta$ -chloro-alanine* (inhibiteur des alanines transaminases)	5 mg/L

Tableau 9 : Composition du milieu de marquage spécifique. Les acides aminés supplémentés sont marqués. \* Chaque constituant a été prélevé (solution) ou pesé (poudre) puis dissous dans 30 mL de milieu de culture (ESF-SFM déplété), avant que le pH ne soit ajusté à 8 et la solution stérilisée.

Une fois la protéine Hsp90 N-terminal exprimée, purifiée et quantifiée, celle-ci est analysée par RMN. Pour ce faire, elle est concentrée jusqu'à 90  $\mu\text{L}$  (~ 100  $\mu\text{M}$ ) et 10 % de  $\text{D}_2\text{O}$  sont ajoutés à l'échantillon avant que ce dernier ne soit placé dans un microtube Shigemi 3 mm. Cet ajout de  $\text{D}_2\text{O}$  est indispensable pour bloquer le signal électromagnétique du spectromètre sur l'échantillon protéique lors du paramétrage et de l'acquisition du spectre RMN.

Pour finir, un spectre 2D de l'échantillon est acquis sur un spectromètre 700 MHz équipé d'une cryosonde, grâce à une expérience « Selective Optimized Flip Angle Short

Transient Heteronuclear Single Quantum Coherence » (SOFAS-HMQC). Cette expérience permet l'enregistrement de spectres de corrélation hétéronucléaire bidimensionnels de protéines en quelques secondes, avec un gain significatif de sensibilité du signal  $^{15}\text{N}$  et  $^{13}\text{C}$  en comparaison avec des expériences de détection  $^1\text{H} - ^{15}\text{N}$  et  $^1\text{H} - ^{13}\text{C}$  standards (Schanda *et al.*, 2005). Par conséquent, bien que la concentration de l'échantillon soit suffisante pour des analyses RMN standards, cette expérience permet de réduire significativement la durée d'acquisition (1 à 2 heures suffisent), tout en maintenant une importante résolution du signal des groupements  $^{13}\text{CH}_3$  présents sur Hsp90 N-terminal.

En parallèle du développement du milieu de marquage spécifique, un milieu de marquage uniforme a également été mis au point. Pour cela, l'expression de la protéine a été réalisée dans un milieu reconstitué avec de l'extrait de levure marqué. Ainsi, pour un marquage uniforme  $^{15}\text{N}$  de la protéine Hsp90 N-terminal, les cellules (transfectées et lavées) ont été re-suspendues dans le milieu suivant (Tableau 10).

Condition	Hsp90 N-terminal marquée U-[ $^1\text{H}$ , $^{15}\text{N}$ ]
Volume culture	100 mL
Milieu de culture	ESF-SFM déplété
Extrait de levure U-[ $^1\text{H}$ , $^{15}\text{N}$ ]* (Collaboration Dr. Nadine ASSRIR, ICSN Gif/Yvette)	4 g/L
$^{15}\text{NH}_4\text{Cl}$ *	5 mM

Tableau 10 : Composition du milieu de marquage uniforme. \* Chaque constituant a été pesé puis dissous dans 30 mL de milieu de culture (ESF-SFM déplété) avant que le pH ne soit ajusté à 8 et la solution stérilisée.

Une fois la protéine exprimée et purifiée (en utilisant le même protocole que précédemment détaillé), 10 % de  $\text{D}_2\text{O}$  sont ajoutés à la protéine purifiée et concentrée (à  $\sim 113 \mu\text{M}$ ) afin de permettre l'acquisition d'un spectre RMN 2D de la protéine U-[ $^1\text{H}$ ,  $^{15}\text{N}$ ]-Hsp90 N-terminal (sur un spectromètre 700 MHz et à partir d'une expérience SOFAST-HMQC).



### V.2.3. Mesure du taux d'incorporation des $^{13}\text{CH}_3$ -acides aminés

Dans le processus de développement d'un premier milieu de marquage spécifique de protéines exprimées en cellules HEK 293, il est nécessaire de caractériser et quantifier le taux d'incorporation de chaque  $^{13}\text{CH}_3$ -acide aminé supplémenté au milieu de marquage. Pour une telle quantification, deux étapes sont nécessaires. Dans un premier temps, une courbe d'incorporation doit être établie pour chaque  $^{13}\text{CH}_3$ -acide aminé. Cette courbe, dont l'allure attendue présente une phase croissante puis un plateau, permet d'identifier la quantité en acides aminés marqués pour laquelle tout ajout supplémentaire de résidus marqués au milieu n'augmentera pas le taux d'incorporation de ce dernier. Dans un deuxième temps, un nouvel échantillon protéique sera produit avec cette quantité d'acide aminé définie comme minimale (pour une incorporation maximale) avant d'être analysé par spectrométrie de masse. De cette analyse pourra être déduit (en faisant la différence de masse entre la protéine non marquée et marquée) le nombre d'acides aminés marqués incorporés à la protéine, et donc le taux d'incorporation de chaque sonde.

Pour produire les courbes d'incorporation de chaque  $^{13}\text{CH}_3$ -acide aminé, différents marquages de la protéine Hsp90 N-terminal ont été réalisés.

i. Afin d'identifier le maximum d'incorporation des résidus  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ , -Val $^{\text{proS}}$  et -Thr, quatre marquages combinatoires  $^{13}\text{CH}_3$ -Met/-Ile $^{\delta 1}$ /-Val $^{\text{proS}}$ /-Thr de la protéine Hsp90 N-terminal ont été réalisés, avec quatre concentrations différentes en  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ , Val $^{\text{proS}}$  et Thr : 40 mg/L, 60 mg/L, 80 mg/L, 100 mg/L. La concentration en  $^{13}\text{CH}_3$ -Met a, quant à elle, été fixée pour chaque échantillon (200 mg/L). En effet, les courbes d'incorporation sont produites en réalisant la différence de la moyenne des intégrations des pics correspondant aux résidus  $^{13}\text{CH}_3$ -Met et celle des pics correspondant aux résidus  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$  ou  $^{13}\text{CH}_3$ -Val $^{\text{proS}}$  ou  $^{13}\text{CH}_3$ -Thr (ceci à l'aide du logiciel TopSpin<sup>®</sup>, version 4.1.3, Bruker). Ainsi, en fixant (arbitrairement) qu'une concentration de 200 mg/L de  $^{13}\text{CH}_3$ -Met permet une incorporation de 100 % de ce résidu, celui-ci peut servir de référence pour l'établissement des courbes d'incorporation de chacun des autres résidus. La protéine Hsp90 N-terminal a donc été exprimée (en utilisant le même protocole que précédemment détaillé) dans les quatre conditions suivantes (Tableau 11).

Conditions	Hsp90 N-terminal marquée : 100 mg/L <sup>13</sup> CH <sub>3</sub> -IVT	Hsp90 N-terminal marquée : 80 mg/L <sup>13</sup> CH <sub>3</sub> -IVT	Hsp90 N-terminal marquée : 60 mg/L <sup>13</sup> CH <sub>3</sub> -IVT	Hsp90 N-terminal marquée : 40 mg/L <sup>13</sup> CH <sub>3</sub> -IVT
Volume culture	100 mL	100 mL	100 mL	100 mL
Milieu de culture	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété
L-Tyrosine*	130 mg/L	130 mg/L	130 mg/L	130 mg/L
Concentration des mélanges réactionnels EAA, NEAA, Ala et Leu	1x	1x	1x	1x
Concentration en acides aminés marqués <sup>13</sup> CH <sub>3</sub> -Met*	200 mg/L	200 mg/L	200 mg/L	200 mg/L
Concentration en acides aminés marqués <sup>13</sup> CH <sub>3</sub> -IVT*	100 mg/L	80 mg/L	60 mg/L	40 mg/L

Tableau 11 : Composition des milieux de marquage spécifique pour l'établissement des courbes d'incorporation des <sup>13</sup>CH<sub>3</sub>-Ile, Val et Thr. \* Chaque constituant a été prélevé (solution) ou pesé (poudre) puis dissous dans 30 mL de milieu de culture (ESF-SFM déplété), avant que le pH ne soit ajusté à 8 et la solution stérilisée.

ii. Avant toute expérience visant à produire la courbe d'incorporation des <sup>13</sup>CH<sub>3</sub>-Ala, une optimisation du protocole de marquage a été nécessaire afin d'accroître le taux d'incorporation de cet isotope. Ainsi, différentes conditions ont été testées dans le but d'en identifier une générant une intensité du signal RMN comparable au signal des <sup>13</sup>CH<sub>3</sub>-Met (ce qui n'a pas été le cas lors du marquage des <sup>13</sup>CH<sub>3</sub>-Ala dans l'échantillon marqué <sup>13</sup>CH<sub>3</sub>-Met/-Ile<sup>δ1</sup>/-Val<sup>proS</sup>/-Thr/-Ala, cf. Tableau 9 p. 138).

- La première condition d'optimisation testée représentait un ajout de huit fois plus d'inhibiteur des alanines transaminases que dans la condition initiale <sup>13</sup>CH<sub>3</sub>-Met/-Ile/-Val<sup>proS</sup>/-Thr/-Ala (cf. Tableau 9 p. 138). Ainsi, les cellules (transfectées dans le milieu FreeStyle 293 et lavées) ont été re-suspendues dans le milieu de marquage formulé avec

le milieu ESF-SFM déplété (100 mL final), supplémenté avec 40 mg/L de  $\beta$ -chloro-alanine et tous les mélanges réactionnels d'acides aminés non marqués (à la concentration de 1x), exception faite des mélanges réactionnels Met et Ala qui sont substitués par 80 mg/L de  $^{13}\text{CH}_3$ -Met et 80 mg/L de  $^{13}\text{CH}_3$ -Ala.

- Les trois conditions suivantes reposent sur l'élimination d'une réserve (supposée) d'alanine non marquée au sein des cellules. Ainsi, les cellules HEK 293F, ont été transférées, 48 heures, 24 heures ou juste avant la transfection, dans du milieu ESF-SFM déplété, reconstitué avec 40 mg/L de  $\beta$ -chloro-alanine et tous les mélanges réactionnels d'acides aminés non marqués (concentration 1x), sauf le mélange réactionnel Ala. Ainsi, une fois cette réserve (supposée) d'alanine non marquée épuisée, les cellules sont transfectées (avec l'ADNc de Hsp90 N-terminal) dans ce même milieu (soit le milieu ESF-SFM avec la  $\beta$ -chloro-alanine et tous les mélanges réactionnels, sauf celui de l'alanine), avant d'être transférées dans le milieu de marquage (identique pour les trois conditions), composé du milieu ESF-SFM déplété (100 mL final), plus 40 mg/L de  $\beta$ -chloro-alanine et tous les mélanges réactionnels d'acides aminés non marqués (à la concentration de 1x), exception faite des mélanges réactionnels Met et Ala qui sont substitués par 80 mg/L de  $^{13}\text{CH}_3$ -Met et 80 mg/L de  $^{13}\text{CH}_3$ -Ala.

Notons toutefois que cette stratégie n'a pu être réalisée qu'une seule fois au cours de cette thèse, car un seul lot de milieu ESF-SFM déplété a généré une expression de la protéine Hsp90 N-terminal (bien que faible), une fois les cellules transfectées dans ce même milieu, c'est pourquoi toutes les transfusions des cellules ont été réalisées dans le milieu de culture FreeStyle 293 avant leur transfert dans les différents milieux de marquage.

- Une dernière condition a été testée avec l'ajout de L-cyclo-sérine comme inhibiteur des alanines transaminases. Pour rappel, cet inhibiteur est largement utilisé lors du marquage des alanines en cellules d'insectes. Ainsi, dans le cas présent, les cellules (transfectées, dans le milieu FreeStyle 293, et lavées) sont re-suspendues dans le milieu de marquage formulé avec le milieu ESF-SFM déplété (100 mL final), plus 5 mg/L de L-cyclo-sérine et tous les mélanges réactionnels d'acides aminés non marqués (à la concentration de 1x), exception faite des mélanges réactionnels Met et Ala qui sont substitués par 80 mg/L de  $^{13}\text{CH}_3$ -Met et 80 mg/L de  $^{13}\text{CH}_3$ -Ala.

Puis, pour réaliser les courbes d'incorporation, la même stratégie a été appliquée, soit la production de quatre échantillons contenant quatre concentrations différentes en  $^{13}\text{CH}_3\text{-Ala}$  et une même concentration en  $^{13}\text{CH}_3\text{-Met}$  (cf. Annexe 4 p. 239). Ainsi, une fois la protéine Hsp90 N-terminal exprimée dans ces différentes conditions, purifiée et analysée par RMN (d'après la même expérience SOFAST-HMQC et sur un spectromètre 700 MHz), il peut être calculé et produit la courbe d'incorporation des  $^{13}\text{CH}_3\text{-Ala}$  (d'après la même méthode que celle appliquée aux résidus Ile, Val et Thr, et avec le même logiciel de traitement, TopSpin® version 4.1.3, Bruker).

Dès lors qu'il a été déterminé pour chaque  $^{13}\text{CH}_3\text{-acide aminé}$  une valeur à laquelle le taux d'incorporation n'augmente pas, et ce malgré un accroissement des quantités d'acides aminés marqués supplémentés au milieu de marquage, sept nouvelles productions (de Hsp90 N-terminal marquée ou non) ont été entreprises dans le but de calculer le taux d'incorporation pour chacun de ces résidus.

Ces productions reposent sur les mêmes protocoles d'expression et de purification, avec, toutefois, un milieu de marquage composé différemment suivant les échantillons protéiques à produire (Tableau 12). L'utilisation des acides aminés marqués  $U\text{-}[^1\text{H}, ^{13}\text{C}, ^{15}\text{N}]$  s'explique par la nécessité d'accroître le plus possible la masse de chaque résidu incorporé. En effet, plus la différence de masse sera importante entre un résidu incorporé ou non, meilleure sera la confiance dans les taux d'incorporation calculés.

Il convient de noter l'absence de courbes d'incorporation réalisées pour la méthionine et la leucine. Cependant, ces deux acides aminés étant essentiels et, au vu des courbes d'incorporation identiques pour cette famille d'acides aminés, il a donc été supposé que 80 mg/L de  $^{13}\text{CH}_3\text{-Met}$  et  $^{13}\text{CH}_3\text{-Leu}$  (supplémentés au milieu) permettraient une incorporation maximale.

Conditions	Hsp90 N-terminal marquée : 80 mg/L <sup>13</sup> CH <sub>3</sub> -Met	Hsp90 N-terminal marquée : 80 mg/L [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Ile	Hsp90 N-terminal marquée : 80 mg/L [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Val	Hsp90 N-terminal marquée : 80 mg/L [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Thr	Hsp90 N-terminal marquée : 80 mg/L [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Leu	Hsp90 N-terminal marquée : 80 mg/L <sup>13</sup> CH <sub>3</sub> -Ala	Hsp90 N-terminal non marquée
Volume culture	100 mL	100 mL	100 mL	100 mL	100 mL	100 mL	100 mL
Milieu de culture	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	FreeStyle 293
L-Tyrosine	130 mg/L	130 mg/L	130 mg/L	130 mg/L	130 mg/L	130 mg/L	-
Concentration des mélanges réactionnels EAA et NEAA	1x	1x	1x	1x	1x	1x	-
Concentration en <sup>13</sup> CH <sub>3</sub> -Met OU Mélange réactionnel	<b>80 mg/L</b>	Met 1x	Met 1x	Met 1x	Met 1x	Met 1x	-
Concentration en [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Ile OU Mélange réactionnel	Ile 1x	<b>80 mg/L</b>	Ile 1x	Ile 1x	Ile 1x	Ile 1x	-
Concentration en [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Val OU Mélange réactionnel	Val 1x	Val 1x	<b>80 mg/L</b>	Val 1x	Val 1x	Val 1x	-
Concentration en [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Thr OU Mélange réactionnel	Thr 1x	Thr 1x	Thr 1x	<b>80 mg/L</b>	Thr 1x	Thr 1x	-
Concentration en [ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Leu OU Mélange réactionnel	Leu 1x	Leu 1x	Leu 1x	Leu 1x	<b>80 mg/L</b>	Leu 1x	-
Concentration en <sup>13</sup> CH <sub>3</sub> -Ala OU Mélange réactionnel	Ala 1x	Ala 1x	Ala 1x	Ala 1x	Ala 1x	<b>80 mg/L</b>	-
L-cyclo-sérine*	-	-	-	-	-	<b>5 mg/L</b>	-

Tableau 12 : Composition des milieux de marquage spécifique pour le calcul du taux d'incorporation de chaque acide aminé marqué.

Une fois l'expression de Hsp90 N-terminal au sein de ces milieux et les purifications réalisées, ces sept échantillons ont été remis à la plateforme de spectrométrie de masse de l'IBS afin de mesurer la masse de chaque protéine par ionisation par électro-ébuliseur - temps de vol (ESI-TOF). Cette mesure est indispensable pour entreprendre le calcul du taux d'incorporation. Pour rappel, ce taux est calculé à partir de la différence de masse entre la masse de la protéine marquée attendue si le taux d'incorporation est de 100 %, et la masse réelle calculée par MS.

#### V.2.4. Application du milieu de marquage spécifique à la protéine HSulf-2

Le milieu de marquage spécifique mis au point a permis d'initier l'étude de la protéine HSulf-2 par RMN.

La protéine HSulf-2  $\Delta$ HD marquée sur les  $^{13}\text{CH}_3\text{-Met}$  a été produite par ajout d'un excès de  $^{13}\text{CH}_3\text{-Met}$  (0,5 g/L) au milieu de culture FreeStyle 293 (marquage en excès). Après quatre jours d'expression, la protéine a été purifiée comme décrit précédemment (cf. V.1.1. p. 125), à l'exception de la chromatographie échangeuse d'ions qui n'a pas été réalisée (le milieu conditionné étant directement chargé sur colonne Ni-NTA). Une fois le résultat de la purification analysé par SDS-PAGE, les fractions correspondant à la protéine d'intérêt ont été concentrées (jusqu'à 90  $\mu\text{L}$ ) et 10 % de  $\text{D}_2\text{O}$  (v/v) ont été rajoutés à l'échantillon, avant qu'un spectre RMN 2D ne soit acquis (sur un spectromètre 700 MHz équipé d'une cryosonde, d'après la même méthode SOFAST-HMQC). Une fois pure, la protéine HSulf-2  $\Delta$ HD marquée  $^{13}\text{CH}_3\text{-Met}$  a été digérée à la trypsine (d'après le même protocole qu'appliqué pour HSulf-2  $\Delta$ SG) et purifiée par chromatographie d'exclusion (Superdex 75 10/300 GL), puis ré-analysée par RMN.

La protéine HSulf-2  $\Delta$ HD marquée sur les  $^{13}\text{CH}_3\text{-Met/-Ile}^{\delta^1}\text{-Thr}$  a été produite dans le milieu de marquage composé du milieu ESF-SFM déplété, plus tous les mélanges réactionnels d'acides aminés non marqués (à la concentration de 1x), exception faite des mélanges réactionnels Met, Ile et Thr qui ont été substitués par 80 mg/L de  $^{13}\text{CH}_3\text{-Met/-Ile}^{\delta^1}\text{-Thr}$ . Après quatre jours d'expression, la protéine a été purifiée et analysée par RMN.

Les protéines HSulf-2  $\Delta$ HD V1 et V2 (cf. Figure 15E et 15F p. 124) marquées  $^{13}\text{CH}_3$ -Met ont été produites et analysées selon les mêmes protocoles. Ces protéines sont des dérivées de la protéine HSulf-2  $\Delta$ HD où quelques dizaines d'acides aminés ont été éliminés en position N et C-terminale.

Une autre construction de la protéine HSulf-2, HSulf-2  $\Delta$ HDc (cf. Figure 15D p. 124) a été analysée par RMN. Pour cela, 300 mL de cellules HEK 293F, à une densité de  $1,5 \cdot 10^6$  cellules/mL, ont été transfectées (600  $\mu\text{g}$  PEI pour 300  $\mu\text{g}$  de vecteur HSulf-2  $\Delta$ HDc) dans du milieu FreeStyle 293 en présence de 0,5 g/L de  $^{13}\text{CH}_3$ -Met. Après quatre jours d'expression et un protocole de purification rigoureusement identique à celui développé pour HSulf-2  $\Delta$ SG, un spectre RMN 2D a été acquis. Puis, l'échantillon a été récupéré, et 1 mM de sulfocalix-[4]-arène a été rajouté avant l'acquisition d'un nouveau spectre RMN 2D (expérience SOFAST-HMQC sur un spectromètre 700 MHz équipé d'une cryosonde).

Pour finir, le domaine catalytique de HSulf-1, un homologue de HSulf-2, a été analysé par RMN. Ainsi, d'après la même stratégie que détaillée précédemment, un premier marquage en excès a été réalisé à partir d'une expression en milieu FreeStyle 293 supplémenté avec 0,5 g/L de  $^{13}\text{CH}_3$ -Met et d'une lignée cellulaire stabilisée exprimant HSulf-1  $\Delta$ HD de manière continue. Une fois la purification réalisée, d'après le même protocole qu'établi pour HSulf-2  $\Delta$ HD, un spectre RMN 2D de la protéine a été acquis (après ajout de 10 % de  $\text{D}_2\text{O}$  (v/v) et une expérience SOFAST-HMQC sur un spectromètre 700 MHz équipé d'une cryosonde).

### V.3. Etudes structurales de la protéine HSulf-2 par cryo-EM

Cette étude a été réalisée en collaboration avec le Dr. Rebekka Wild (chargée de recherche CNRS, IBS-SAGAG). Grâce à cette collaboration, deux jeux de données ont pu être acquis pour la protéine HSulf-2  $\Delta$ SG (exprimée et purifiée comme détaillé dans la première partie, cf. V.1.1. p. 125). Le premier jeu de données a été enregistré sur un microscope ThermoFisher Scientific GLACIOS 200kV FEG équipé d'une caméra de détection directe GATAN K2, et le deuxième sur un microscope Titan Krios 300 kV FEG équipé d'une caméra de détection directe GATAN K3.

Afin d'effectuer ces analyses, des conditions de préparation des grilles de microscopie ont tout d'abord été établies. Pour cela, plusieurs grilles ont été produites à l'aide d'un appareil Vitrobot System (ThermoFisher Scientific) et d'après le protocole suivant : 7  $\mu$ L d'échantillon sont déposés sur une grille Quantifoil 1.2/1.3 Cu mesh 300 carbon coated, dans une atmosphère où l'humidité et la température sont régulées (100 % et 4 °C). Notons que ces grilles ont préalablement été soumises à une décharge lumineuse (« glow discharge ») générant une ionisation des atomes de cuivre composant la grille. Après 5 secondes d'attente (une fois l'échantillon déposé), de manière automatique, la solution protéique présente sur la grille est enlevée par absorption sur papier filtre (pendant 6 secondes avec une force de 0). Une fois l'absorption de l'excédent de liquide réalisée, sans temps d'attente, la grille est plongée (rapidement et toujours de manière automatique) dans de l'éthane liquide. Pour finir, la grille est transférée dans de l'azote liquide et maintenue dans cet environnement avant d'être visualisée.

L'analyse des grilles puis la collecte d'images (réalisées par le Dr. Rebekka Wild sur un microscope Glacios (IBS, Grenoble) et au réseau et infrastructure nationale FRISBI (IGBMC, Strasbourg) sur un microscope Titan Krios) a permis l'acquisition de deux jeux de données qui ont été analysés par le logiciel de traitement cryoSPARC v3.2 (traitement réalisé par le Dr. Rebekka Wild).

Au vu des résultats, il a été tenté de produire une forme stable de complexe HSulf-2  $\Delta$ SG/Substrat afin d'accroître la structuration de la protéine et la qualité du jeu de données acquis. Ainsi, des tests de réticulations ont été entrepris.

Le premier test repose sur la stabilisation d'un complexe HSulf-2  $\Delta$ SG avec des chaînes d'HP ou un octasaccharide purifié d'héparine, DP8 (dont la composition est une répétition de quatre unités [IdoA(2S), GlcNS(6S)]). Pour ce faire, un traitement du substrat à l'EDC/NHS (EDC 12 mM et NHS 30 mM) a été réalisé pendant 10 minutes. L'utilisation de l'EDC (1-Ethyl-3-[3-diméthylaminopropyl]-carbodiimide hydrochloride) s'explique par sa propriété à réagir avec les groupements carboxyles pour former un intermédiaire de réaction pouvant interagir avec des groupements amines de la protéine. Cet intermédiaire étant toutefois instable, l'ajout de NHS (N-hydroxysuccinimide) permet de le stabiliser. La réaction est ensuite stoppée et les réactifs libres sont inactivés par ajout de 1 % de 2-mercaptoéthanol (20 mM final). La protéine d'intérêt (purifiée dans un tampon HEPES 20 mM pH 8, NaCl 200 mM et



concentrée à 700 nM finale) est ensuite ajoutée au substrat « activé » (concentration finale : 3  $\mu$ M) et incubée pendant 2 heures à 20°C, dans le but de former des complexes HSulf-2  $\Delta$ SG/Substrat covalents. Pour finir, la réaction est terminée par blocage des esters activés avec une solution de Tris 1 M pH 8 (concentration finale : 100 mM).

La deuxième méthode repose sur la formation de liaisons covalentes inter-HSulf-2  $\Delta$ SG. Pour ce faire, 300  $\mu$ g de protéines préalablement purifiées dans un tampon HEPES 20 mM pH 8, NaCl 200 mM, sont incubées avec de 0,5 % à 2,3 % de glutaraldéhyde fraîchement préparé pendant 1 à 7 minutes (suivant les tests) à 20°C. La réaction est terminée par l'addition d'une solution concentrée de Tris 1 M pH 8.

Une fois ces réactions analysées par western blot, d'après le même protocole que précédemment détaillé (exception faite d'une migration des échantillons sur un gel d'acrylamide 10 % et des anticorps utilisés (anticorps primaire : R-19, anticorps polyclonal de lapin ciblant spécifiquement la région interne du domaine HD, dil. 1/1000, et anticorps secondaire :  $\alpha$ -rabbit-HRP, JIR 111-035-045, dil. 1/10000, cf. protocole p. 134-135), des grilles en colorations négatives ont été réalisées d'après le protocole suivant : 10  $\mu$ L d'échantillon (HSulf-2  $\Delta$ SG/Substrat) concentrés à 0,013 mg/mL sont déposés sur une grille Quantifoil Cu mesh 400. Après 1 minute d'attente la solution protéique est retirée de la grille à l'aide d'un papier filtre (par absorption). Puis, deux lavages de la grille sont réalisés avec de l'eau distillée et cette eau est retirée par absorption (deux fois 20  $\mu$ L de lavage, avec absorption sans temps d'attente entre et pendant les lavages). Pour finir, 10  $\mu$ L de colorant Sodium Silico Tungstate (SST) 1 % sont déposés sur la grille pendant 30 secondes avant d'être retirés par absorption. Une fois la grille sèche, cette dernière est visualisée sur un microscope FEI Tecnai F20 200 kV FEG équipé d'une caméra Ceta FEI.

En parallèle des essais de formation de complexes HSulf-2  $\Delta$ SG/Substrat ont été entrepris avec un ratio croissant en HP (1:1 à 1:16) (m/m). Une fois ces réactions terminées (15 minutes à 4°C), des grilles de colorations négatives ont été réalisées suivant le même protocole que précédemment détaillé (grilles Cu mesh 400, 130  $\mu$ g HSulf-2  $\Delta$ SG/HP, coloration SST 1 %) et visualisées sur un microscope FEI Tecnai F20 200 kV FEG équipé d'une caméra Ceta FEI.

## *Résultats*



## VI. Résultats

### VI.1. Essais cristallographiques : les stratégies d'optimisation

Lors de précédents travaux sur le projet Sulf, des tests de cristallogenèse ont été menés sur différentes protéines, comme HSulf-2 WT (protéine totale avec sa chaîne de GAG) (cf. Figure 15A p. 123), HSulf-2  $\Delta$ SG (protéine totale sans chaîne de GAG) (cf. Figure 15B p. 123) et HSulf-2  $\Delta$ HD (maintien des domaines CAT et C-terminal, et délétion du domaine HD) (cf. Figure 15C p. 123). Cependant, aucune condition n'a permis l'obtention de cristaux diffractants, ce qui peut vraisemblablement s'expliquer par les propriétés de flexibilité et de faible structuration de la protéine HSulf-2 (cf. II.3.1. p. 83). De plus, ces protéines ont été exprimées dans des cellules HEK 293F. Or, l'utilisation de ces cellules génère des PTMs qui, très probablement, augmentent la flexibilité et l'hétérogénéité de la protéine (facteurs limitants pour la cristallogenèse et pouvant également diminuer la qualité des données RMN et EM acquises). Par conséquent, dans une perspective de diminuer ces facteurs défavorables, des optimisations ont été réalisées : (i) en réduisant les N-glycosylations de la protéine, (ii) en conservant seulement les régions structurées de HSulf-2 et (iii), et en stabilisant la protéine grâce à la présence de substrats.

#### VI.1.1. Essais cristallographiques : HSulf-2 $\Delta$ SG

Un premier test de cristallogenèse a été réalisé sur la protéine HSulf-2  $\Delta$ SG (Rappel Figure 15B) où seules les régions protégées de la protéolyse ont été conservées. Dans cet objectif, la protéine a été exprimée en cellules HEK 293F, et après quatre jours d'expression, le milieu conditionné a été filtré et chargé sur une colonne HiTRAP Sepharose SP High Performance avant qu'une élution par gradient de sel ne soit réalisée.



HSulf-2  $\Delta$ SG : 99,3 kDa  
HSulf-2  $\Delta$ SG + SNAP/His : 121,3 kDa  
HSulf-2  $\Delta$ SG + SNAP/His + ~ 18 kDa N-glycosylation : 139,3 kDa

Rappel Figure 15B : Représentation schématique de la protéine HSulf-2  $\Delta$ SG.

Le chromatogramme obtenu par mesure de l'absorbance à 280 nm présente deux pics élués après 380 mM et 530 mM de NaCl, respectivement (Figure 16A). L'analyse des fractions d'éluion par SDS-PAGE (Figure 16A) montre que le pic a correspond à des contaminants, car de nombreuses bandes apparaissent mais aucune ne présente la masse moléculaire attendue de la protéine d'intérêt. En revanche, les fractions correspondant au pic b présentent 4 bandes. La bande majoritaire observée entre 95 et 130 kDa du marqueur de taille peut correspondre aux domaines SNAP-CAT(-HD) (+ ~ 18 kDa de N-glycosylation), et le doublet de bande observé entre 43 et 55 kDa peut correspondre au double domaine (-HD)-C-terminal. En effet, à cause des différents sites de clivage furine que porte le domaine HD, en conditions dénaturantes, la protéine HSulf-2 peut présenter une bande à ~ 99 kDa, et une ou plusieurs bandes (suivant le clivage) entre ~ 40 et 50 kDa, ce qui est en adéquation avec les observations faites à partir du SDS-PAGE. Pour finir, une dernière bande de haute masse moléculaire, autour de 130 kDa, est présente et pourrait correspondre à la protéine n'ayant subi aucun clivage par la furine. Une chromatographie d'affinité a ensuite été réalisée sur une colonne Ni-NTA. Ici, les protéines sont chargées sur la colonne afin d'accroître la pureté de la protéine (avec la perte d'un contaminant de basse masse moléculaire ; Figure 16B), mais surtout afin de réaliser un traitement au phosphate de potassium et chlorure de potassium des protéines liées aux ions Ni<sup>2+</sup> ; ceci dans le but d'éliminer toutes contaminations à l'ADN. Dans l'étape suivante, les étiquettes His et SNAP sont clivés à l'aide de la protéine TEV. Bien que l'étiquette SNAP soit clivée, il apparaît que le His ne l'est pas (résultats non présentés), suggérant un manque d'accessibilité de la TEV pour le site de coupure C-terminal.

Une fois l'étape de clivage réalisée, une chromatographie d'exclusion (Superdex 200 10/300 GL) a été entreprise afin d'éliminer les derniers contaminants, dont les protéines SNAP et TEV. Il en résulte un pic majoritaire présent à un volume d'éluion de 10,3 mL (Figure 16C). La pureté de la protéine, ainsi que la bonne coupure de l'étiquette SNAP, ont été confirmées par SDS-PAGE coloré au bleu de Coomassie (Figure 16C). Sur le gel, trois bandes principales sont observées et correspondent toutes à la protéine cible (Figure 16D). En effet, après l'analyse de la protéine d'intérêt (avant la réalisation du clivage des étiquettes) par western blot, l'ensemble des bandes a été révélé par l'anticorps spécifique à la protéine (Anticorps primaire R-19 ciblant spécifiquement la région interne du domaine HD). De ce fait, il peut être avancé avec confiance que la bande majoritaire observée à ~ 70 kDa correspond aux domaines CAT - HD N-terminal

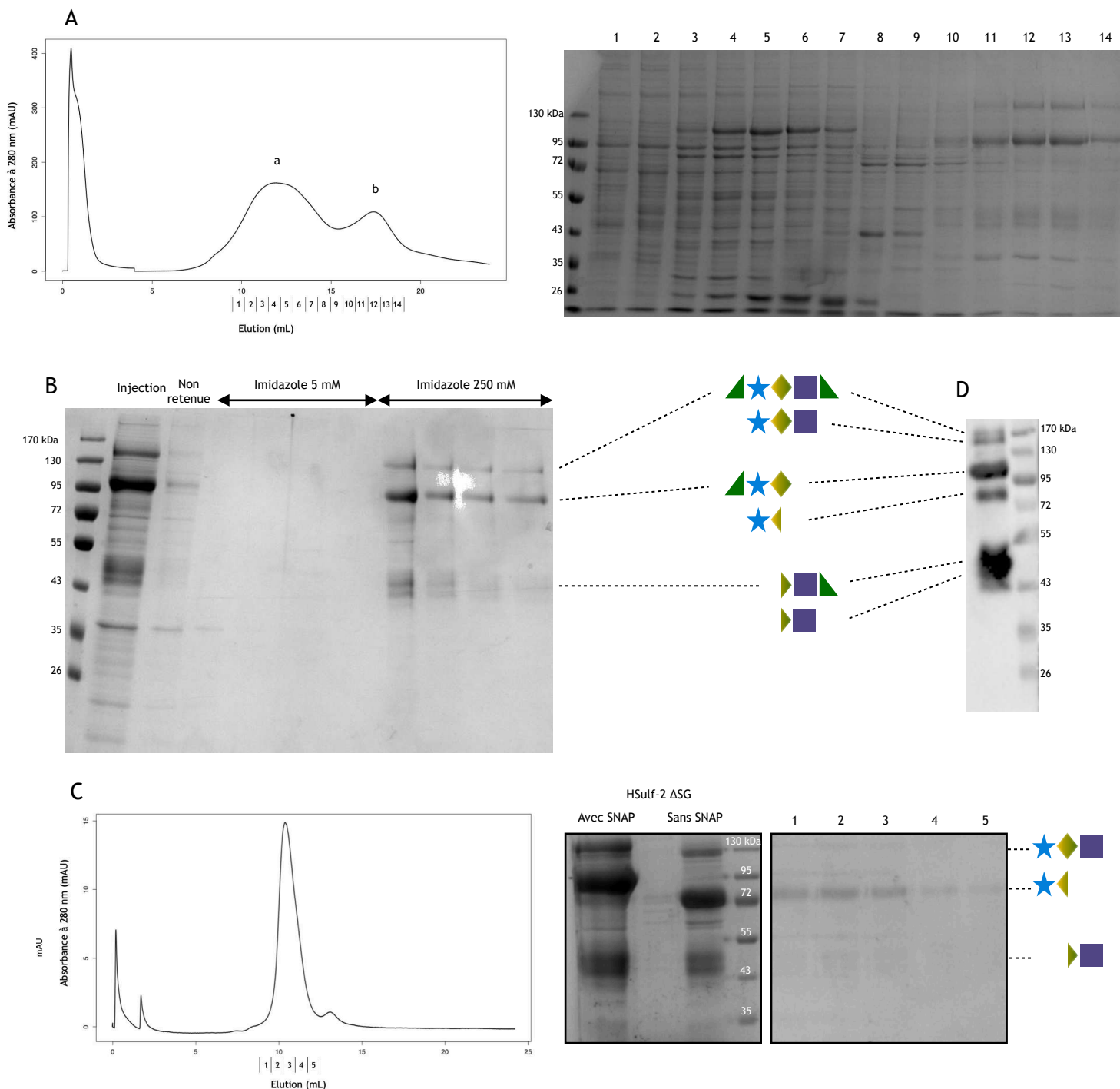
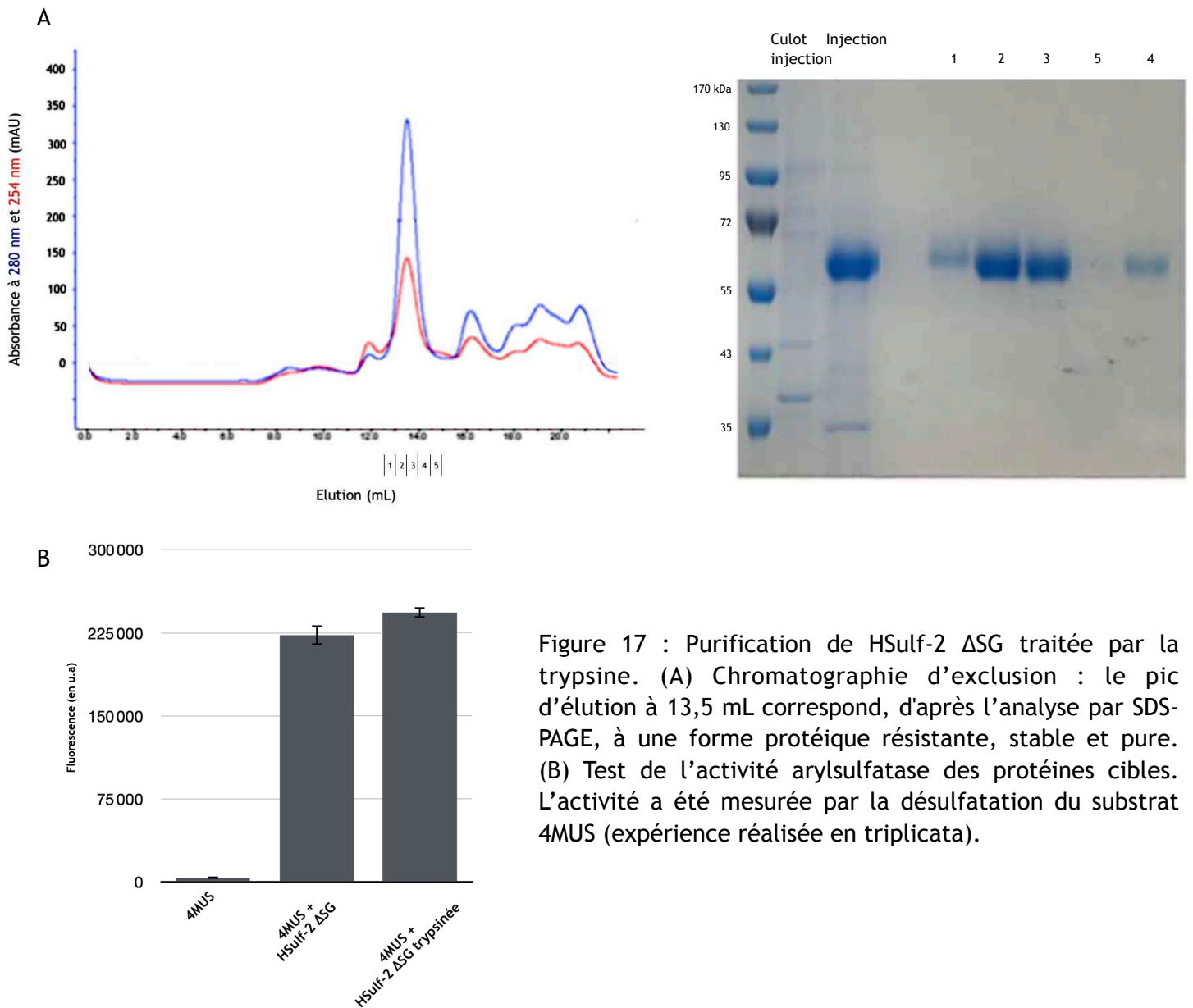


Figure 16 : Purification de HSulf-2  $\Delta$ SG. (A) Chromatographie échangeuse d'ions et SDS-PAGE associé. Les fractions 11 à 14 semblent correspondre à la protéine d'intérêt, à l'exception de la bande au-dessus du marqueur de taille 35 kDa qui pourrait être une protéine contaminante. (B) Chromatographie d'affinité (Ni-NTA) : la totalité de la protéine ne s'est pas fixée à la colonne (puits « non retenue »). En revanche, le lavage avec 5 mM d'imidazole permet d'éliminer le contaminant de basse masse moléculaire (puits Imidazole 5 mM). Pour terminer, la protéine a été éluée (puits Imidazole 250 mM) et un clivage à la TEV a été réalisé. (C) Chromatographie d'exclusion : le pic d'éluion à 10,3 mL correspond, d'après l'analyse par SDS-PAGE, à la protéine pure. (D) Western blot de l'échantillon HSulf-2  $\Delta$ SG avant coupure TEV, d'après le même protocole que celui présenté p. 144 (Anticorps primaire : R-19, anticorps polyclonal de lapin ciblant spécifiquement la région interne du domaine HD, dil. 1/1000, et anticorps secondaire :  $\alpha$ -rabbit-HRP, JIR 111-035-045, dil. 1/10000). L'utilisation de cet échantillon explique pourquoi deux bandes sont révélées entre les marqueurs de taille 130 et 170 kDa et autour du marqueur de taille 95 kDa, car un clivage naturel de l'étiquette SNAP existe.

Nomenclature : *Etoile* = domaine CAT ; *Losange* = domaine HD ; *Carré* = domaine C-terminal ; *Triangle* : étiquettes SNAP et His.

(HD en amont du clivage furine), le doublet de bande entre 43 et 55 kDa correspond aux domaines C-terminal - HD C-terminal (HD en aval du clivage furine), et pour finir, la bande à ~ 130 kDa correspond à la protéine non clivée par la furine. Ainsi, puisque l'ensemble des bandes observées sur SDS-PAGE a été identifié comme étant la protéine HSulf-2  $\Delta$ SG, les fractions ont été rassemblées et concentrées. Au final, un rendement d'environ 1,5 mg/L a été obtenu.

Afin d'éliminer les régions flexibles exposées au solvant, qui pourraient empêcher la cristallisation de la protéine, cette dernière a été incubée avec de la trypsine (dans un rapport molaire trypsine : HSulf-2 de 1:50) pendant 40 minutes à 37°C. Le produit de la digestion (HSulf-2  $\Delta$ SG - Tryp) a, par la suite, été purifié par chromatographie d'exclusion de taille (Superdex200 10/300 GL) (Figure 17A). Le profil d'élution montre un pic majoritaire élué à 13,5 mL. Une analyse par SDS-PAGE de ce pic révèle une bande pure à ~ 65 kDa (Figure 17A). Au vu de nos connaissances sur cette protéine et des données antérieures acquises par des collaborateurs du projet (Dr. Rana El Masri), la protéine purifiée correspondrait aux domaines CAT et C-terminal, avec une perte du domaine HD (sensible à la digestion tryptique du fait de sa faible structuration). Notons pour finir qu'un test d'activité spectrofluorimétrique (4MUS) permettant de mesurer l'activité arylsulfatase de la protéine a été réalisé sur les protéines HSulf-2  $\Delta$ SG et HSulf-2  $\Delta$ SG trypsinée. Les résultats de mesure de la fluorescence indiquent que deux enzymes sont bien actives (et d'activité équivalente), signifiant que le site actif n'est pas affecté par cette digestion (Figure 17B).

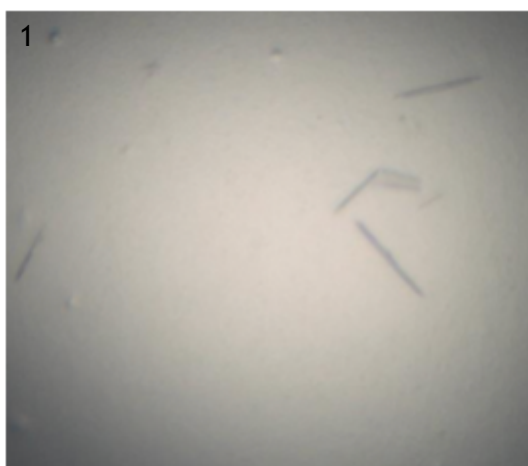


A l'issue de cette digestion partielle de H2Sulf-2 ΔSG par la trypsine, les fractions contenant le cœur hydrophobe de la protéique (~ 65 kDa) ont été concentrées jusqu'à 3,5 mg/mL. Cet échantillon protéique a ensuite été divisé en deux : un échantillon contenant la protéine seule, et un échantillon où 1 mM de sulfocalix-[4]-arène a été rajouté à la protéine.

A partir de ces deux échantillons, six plaques de criblage de conditions de cristallisation ont été réalisées grâce à la plateforme HTXlab (EMBL, Grenoble) (cf. Tableau 4 p. 127). Il est apparu que les plaques contenant H2Sulf-2 ΔSG trypsinée + Sulfocalix-[4]-arène présentaient trois conditions avec, au moins, un cristal (aiguilles,



oursins) dans chacune d'elles (Figure 18). Ces résultats sont d'autant plus intéressants qu'aucun cristal n'a été observé dans les conditions où la protéine sans Sulfocalix-[4]-arène est présente (y compris dans les trois conditions présentées dans la figure 18). Par conséquent, il est possible que les cristaux observés ne soient pas issus de sels présents dans les tampons testés. Ceci étant, l'observation des cristaux obtenus en lumière UV n'a pas montré de fluorescence caractéristique d'un cristal de protéines, ce qui pourrait présager que ces cristaux ne sont pas des cristaux de protéines mais peut-être de sulfocalix-[4]-arène. Toutefois, la fiabilité de ce test est très protéine dépendante et les cristaux étant très petits, il est possible que l'information délivrée par l'imagerie UV ne soit pas significative. Par conséquent, le cristal B (seul cristal pêchable et analysable) a été envoyé au Synchrotron SOLEIL afin de tester sa diffraction. Cependant, aucun diagramme de diffraction n'a pu être enregistré.



100  $\mu$ m



100  $\mu$ m



100  $\mu$ m

Figure 18 : Tests de cristallogénèse de la protéine HSulf-2  $\Delta$ SG trypsinée en présence de Sulfocalix-[4]-arène (à température ambiante). Pour les conditions 1 et 2, les micro-cristaux sont apparus après 48 heures, quant aux micro-cristaux de la condition 3, ils sont apparus après 24 heures. Condition 1 : Sodium acétate 0,1 M pH 4,5, lithium sulfate 0,2 M et PEG 400 50 % (v/v). Condition 2 : Tris 0,1 M pH 8,5, ammonium phosphate monobasic 0,2 M et (+/-)-2-Methyl-2,4-pentanediol 50 % (v/v). Condition 3 : HEPES 0,1 M pH 7,5, ammonium sulfate 0,5 M et MPD 30 % (v/v).

## VI.1.2. Essais cristallographiques : HSulf-2 $\Delta$ HDc

En parallèle, une nouvelle construction a été conçue (cf. Rappel Figure 15D p. 124) en tenant compte de l'ensemble des informations structurales obtenues pour la protéine HSulf-2 (cf. II.3.1. p. 83) et en tentant de se rapprocher de la forme protéique stable obtenue après digestion ménagée (seule condition ayant délivrée des cristaux prometteurs). Cette construction, nommée HSulf-2  $\Delta$ HDc, comprend les domaines CAT, C-terminal et une partie du domaine HD (séquence du domaine HD tronquée entre D<sub>483</sub> et C<sub>636</sub>). Pour rappel, de récentes analyses RMN et de MS ont pu mettre en évidence, à la fois la présence de ponts disulfures entre les domaines CAT-C-terminal et au sein du domaine HD, ainsi qu'une structuration des extrémités de ce dernier domaine. Notons que ces informations ont par la suite pu être corroborées avec la prédiction de la structure de la protéine HSulf-2 WT par le nouveau logiciel de prédiction AlphaFold (Jumper *et al.*, 2021 ; Varadi *et al.*, 2022).

En conséquence, avec une troncature des zones flexibles de la protéine et le maintien de régions structurellement importantes, une meilleure structuration de la protéine est attendue, et possiblement favorable à l'obtention de cristaux diffractants.



HSulf-2  $\Delta$ HDc : 82,5 kDa

HSulf-2  $\Delta$ HDc + SNAP/His : 107,2 kDa

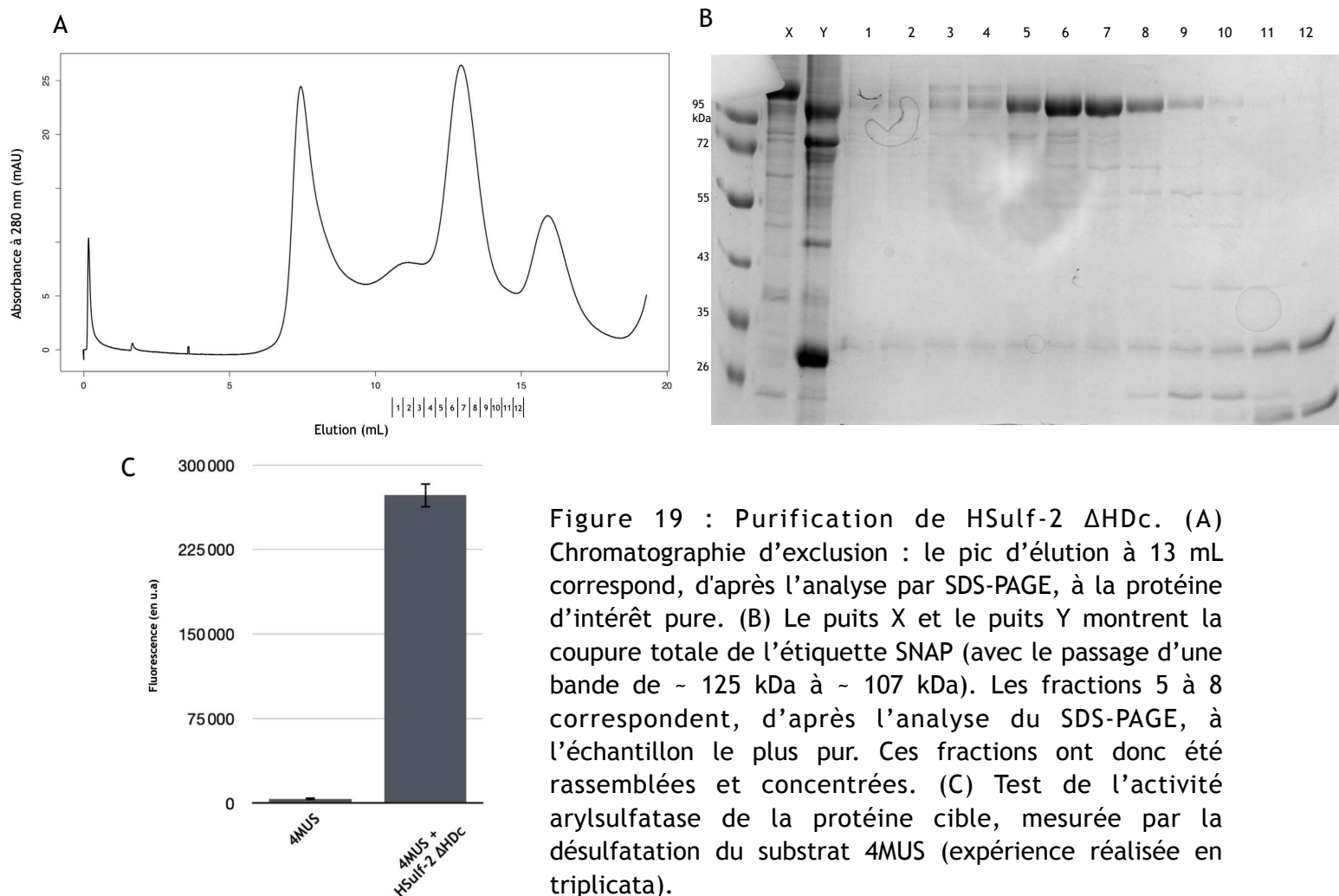
HSulf-2  $\Delta$ HDc + SNAP/His + ~ 18 kDa N-glycosylation : 125,2 kDa

Rappel Figure 15D : Représentation schématique de la protéine HSulf-2  $\Delta$ HDc.

La protéine HSulf-2  $\Delta$ HDc s'exprime et se purifie d'après le même protocole que décrit pour HSulf-2  $\Delta$ SG. Ainsi, après la réalisation d'une chromatographie d'exclusion sur colonne Superdex 200 10/300 GL, un pic majoritaire s'éluant à 13 mL est obtenu (Figure 19A). Une fois ce pic analysé par SDS-PAGE et le gel coloré au bleu de Coomassie (Figure 19B), une bande majoritaire est observée juste au-dessus de la bande 95 kDa du marqueur de taille, ce qui est en adéquation avec la masse attendue de la protéine cible. En effet, pour la protéine HSulf-2  $\Delta$ HDc sans l'étiquette SNAP et avec ~ 18 kDa de N-glycosylations (et sachant que les sites de clivage furine portés par le domaine HD ne

sont pas présents dans cette construction), une seule bande à une masse de 107 kDa est attendue. Ceci étant, malgré la présence de trois étapes de purification, des contaminants persistent, avec notamment la présence d'une bande entre 26 et 35 kDa qui correspond vraisemblablement à la protéine TEV. Plusieurs bandes sont également présentes entre les bandes 45 et 72 kDa du marqueur de taille. Cependant, la séquence reconnue par l'anticorps R19 ayant été éliminée de la séquence de HSulf-2  $\Delta$ HDC, conclure quant à la nature (dégradations et/ou contaminants) de celles-ci reste difficile. Néanmoins, en ne rassemblant que les fractions correspondant à la protéine d'intérêt les plus pures, un rendement d'environ 1 mg/L a pu être obtenu.

Pour finir, un test d'activité arylsulfatase sur un substrat sulfaté fluorescent (4MUS) a été réalisé et confirme que la protéine HSulf-2  $\Delta$ HDC est bien active (Figure 19C).



Des tests de cristallogenèse ont été réalisés à partir de cette protéine purifiée. Pour ce faire, 5 mM de Sulfocalix-[4]-arène ont été rajoutés à la protéine HSulf-2  $\Delta$ HDC concentrée à 23 mg/mL. Cependant, une précipitation de l'échantillon s'est produite, probablement dû à la forte concentration de la protéine (première fois qu'une telle concentration a été obtenue). L'échantillon a, toutefois, été centrifugé et le surnageant récupéré. Or, après une mesure de la concentration (par absorbance à 280 nm), la concentration est restée la même (23 mg/mL). Par conséquent, l'échantillon protéique a été délivré à la plateforme de cristallogenèse HTXlab (EMBL, Grenoble) afin d'identifier, par un large criblage (cf. Tableau 4 p. 127), de possibles conditions favorables à la formation de cristaux diffractants.

Après l'analyse des plaques, 50 % de gouttes sont claires, 30 % présentent une séparation de phase et 20 % une précipitation, et aucun cristal, ce qui suggère une concentration trop faible, et très probablement un échantillon non-optimal. Il est à noter qu'une condition a présenté un agrégat de protéines pêchable (un amas de micro-cristaux ou une agrégation autour d'un contaminant), car cet objet a émis de la fluorescence une fois soumis aux UV (Figure 20). Cependant, après l'analyse de sa diffraction au Synchrotron ESRF, Grenoble (Ligne MASSIF), aucun diagramme de diffraction n'a pu être enregistré.

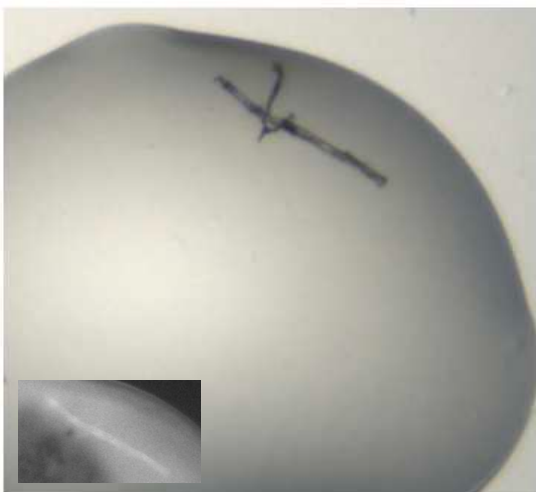


Figure 20 : Amas protéique obtenu pour la protéine HSulf-2  $\Delta$ HDC. Condition : SPG 0,1 M pH 4, PEG 1500 25% (m/v). L'encart en bas à gauche correspond à l'objet sous UVs.

100  $\mu$ m

### VI.1.3. Essais cristallographiques : HSulf-2 $\Delta$ HD

Bien que ces deux derniers tests tentent de limiter la flexibilité de la protéine HSulf-2, d'autres facteurs limitants persistent. En effet, ces deux protéines (HSulf-2  $\Delta$ SG et HSulf-2  $\Delta$ HDC) présentent un profil N-glycosidique complet (car exprimées en cellules HEK 293F). C'est pourquoi, une troisième construction, HSulf-2  $\Delta$ HD (Rappel Figure 15C), a été étudiée car cette dernière permet l'expression d'une protéine pure dans des rendements importants et avec un motif de N-glycosylation pouvant être réduit.



HSulf-2  $\Delta$ HD : 64,5 kDa  
HSulf-2  $\Delta$ HD + SNAP/His : 86,6 kDa  
HSulf-2  $\Delta$ HD + SNAP/His + ~ 18 kDa N-glycosylation : 104,6 kDa

Rappel Figure 15C : Représentation schématique de la protéine HSulf-2  $\Delta$ HD.

Afin de réduire la glycosylation de la protéine, deux stratégies ont été étudiées. La première méthode consiste en un traitement de la protéine cible avec de la PNGase F. Cette protéine est une amidase qui clive les chaînes N-glycanes entre l'asparagine et le premier GlcNAc du sucre. Toutefois, après avoir mis au point un protocole d'expression et de purification, ainsi qu'identifié des conditions non dénaturantes permettant la déglycosylation complète de la protéine HSulf-2 (Figure 21), il s'est avéré que la protéine (toutes constructions confondues) n'était pas stable dans le temps (précipitation de la protéine en quelques heures). Par conséquent, une deuxième approche a été étudiée, basée sur l'expression de la protéine dans des cellules HEK 293S. Ces cellules dérivées des cellules HEK 293 présentent une mutation de l'enzyme N-acétyl-glucosaminyltransférase I. Ainsi, ces cellules ne peuvent plus produire des structures saccharidiques complexes au-delà du noyau penta-saccharidique des N-glycans.

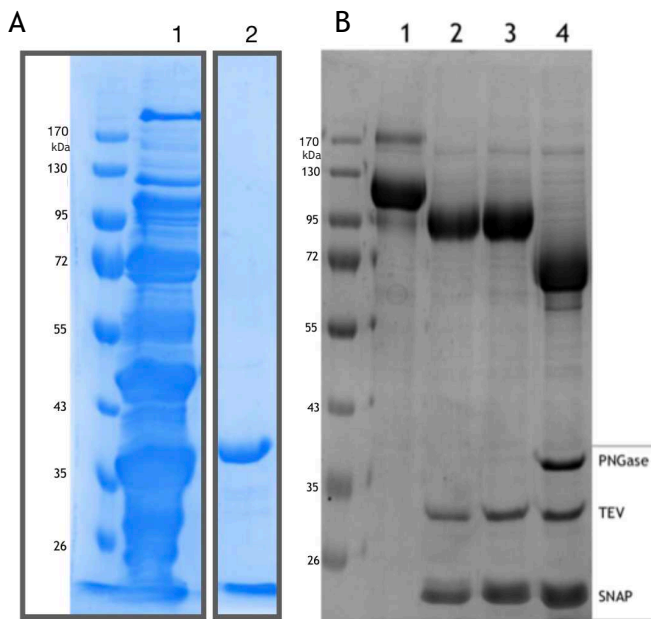


Figure 21 : Purification de la PNGase F, vérification de son activité et traitement de la protéine HSulf-2  $\Delta$ HD.

(A) SDS-PAGE de la chromatographie d'exclusion de la protéine PNGase F. Le puits 1 représente l'échantillon injecté sur colonne de filtration sur gel et le puits 2 la protéine d'intérêt. Il peut être conclu que le protocole d'expression et de purification permet d'obtenir une protéine PNGase F pure.

(B) Test d'activité de la PNGase F et observation sur SDS-PAGE. Puits 1 : HSulf-2  $\Delta$ HD avec l'étiquette SNAP et ses N-glycosylations. Puits 2 et 3 (deux fois le même échantillon) : HSulf-2  $\Delta$ HD sans l'étiquette SNAP, après un traitement TEV (TEV 10 % sur la nuit à 4°C) mais avec ses N-glycosylations. Puits 4 : HSulf-2  $\Delta$ HD double traitement TEV et PNGase F (TEV 10 % et PNGase F 10 % sur la nuit à 4°C). La taille observée sur SDS-PAGE est bien inférieure à 72 kDa, ce qui est en adéquation avec la masse théorique attendue (64,5 kDa).

Pour valider l'utilisation de ces cellules comme système d'expression permettant la production de protéines pures et actives tout en ayant une glycosylation réduite, la protéine HSulf-2  $\Delta$ HD a été exprimée (en cellules HEK 293S) et purifiée en suivant le protocole décrit pour HSulf-2  $\Delta$ SG, à l'exception de la première étape de chromatographie qui ne peut être réalisée avec la protéine HSulf-2  $\Delta$ HD (n'ayant pas le domaine HD basique, la protéine ne peut se fixer à la résine Sepharose SP). Ainsi, à l'issue de l'étape de chromatographie d'exclusion sur colonne Superdex 200 10/300 GL, un pic majoritaire à un volume d'élution de 13,7 mL est obtenu (Figure 22A). La pureté de la protéine, ainsi que la bonne coupure de l'étiquette SNAP ont été confirmées par SDS-PAGE (révélé au bleu de Coomassie) (Figure 22B). D'après l'analyse du gel, les fractions correspondant au pic d'élution à 13,7 mL présentent une bande majoritaire entre 72 et 95 kDa du marqueur de taille, ce qui est en adéquation avec la masse attendue de la protéine cible. En effet, la masse attendue pour la protéine HSulf-2  $\Delta$ HD sans l'étiquette SNAP et avec des N-glycosylations réduites est d'environ 77 kDa, estimation réalisée sur la base d'une analyse par spectrométrie de masse MALDI montrant une réduction d'environ 10 kDa des N-glycosylations présentes sur HSulf-2  $\Delta$ HD exprimée en HEK 293S (résultats obtenus par le Dr. Rana El Masri). Les fractions correspondant à la protéine d'intérêt ont été rassemblées et concentrées pour produire deux échantillons concentrés à 55 mg/mL.

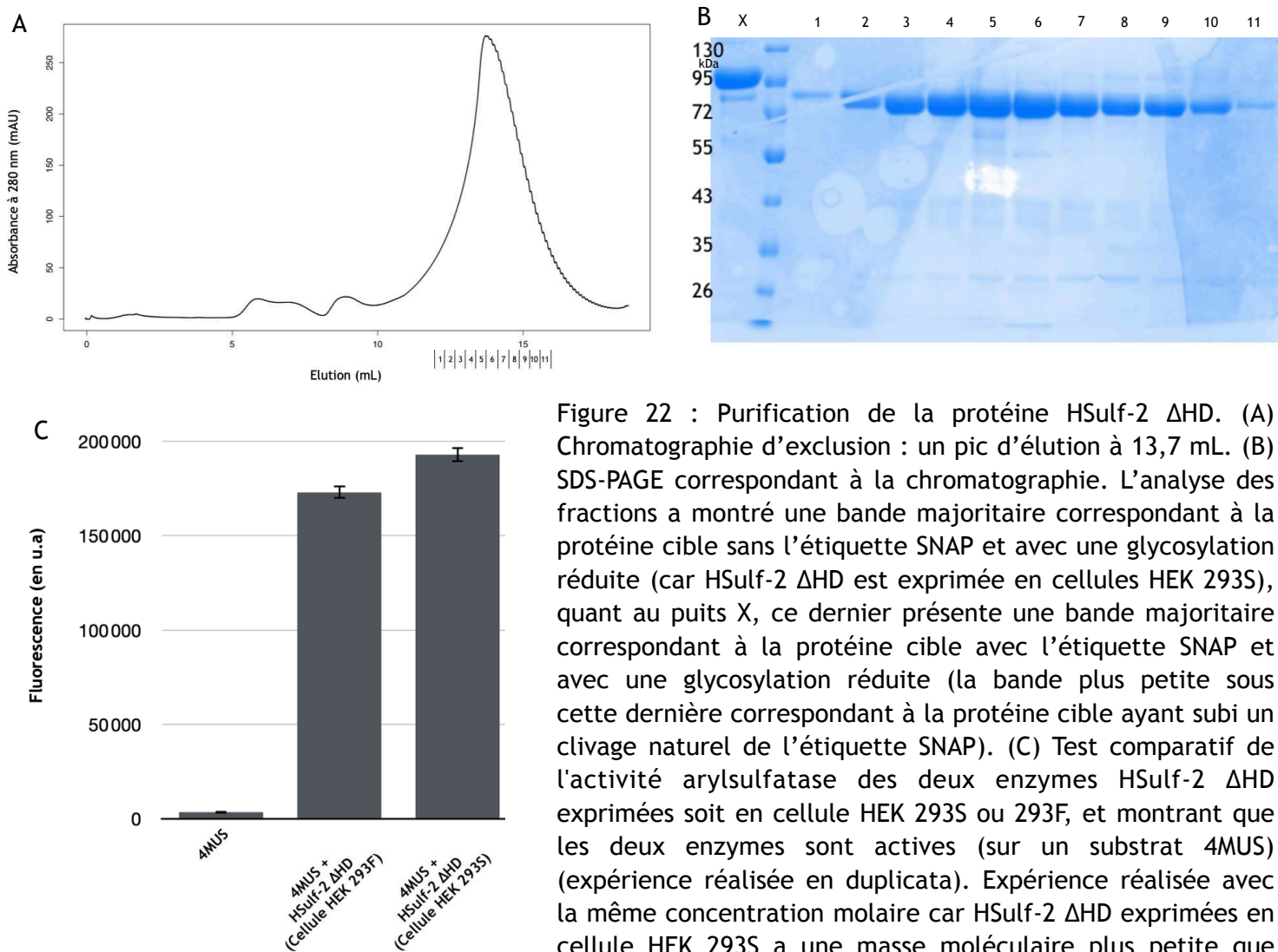


Figure 22 : Purification de la protéine H2Sulf-2 ΔHD. (A) Chromatographie d'exclusion : un pic d'élution à 13,7 mL. (B) SDS-PAGE correspondant à la chromatographie. L'analyse des fractions a montré une bande majoritaire correspondant à la protéine cible sans l'étiquette SNAP et avec une glycosylation réduite (car H2Sulf-2 ΔHD est exprimée en cellules HEK 293S), quant au puits X, ce dernier présente une bande majoritaire correspondant à la protéine cible avec l'étiquette SNAP et avec une glycosylation réduite (la bande plus petite sous cette dernière correspondant à la protéine cible ayant subi un clivage naturel de l'étiquette SNAP). (C) Test comparatif de l'activité arylsulfatase des deux enzymes H2Sulf-2 ΔHD exprimées soit en cellule HEK 293S ou 293F, et montrant que les deux enzymes sont actives (sur un substrat 4MUS) (expérience réalisée en duplicata). Expérience réalisée avec la même concentration molaire car H2Sulf-2 ΔHD exprimées en cellule HEK 293S a une masse moléculaire plus petite que H2Sulf-2 ΔHD exprimées en cellule HEK 293F (cf. Figure 22B).

Pour un des deux échantillons, 5 mM de Sulfocalix-[4]-arène ont été rajoutés à la solution protéique. Cependant, à l'instar de la protéine H2Sulf-2 ΔHDc, une précipitation massive a été obtenue. Toutefois, en centrifugeant cet échantillon et en utilisant l'excédant de solutions protéiques du premier échantillon, deux échantillons ont pu être envoyés à la plateforme HTXlab (EMBL, Grenoble) (le premier concentré à 55 mg/mL et le deuxième à 15 mg/mL) pour un criblage (cf. Tableau 4 p. 127) de possibles conditions favorables à la formation de cristaux diffractants. Il apparaît une fois les plaques de cristallogénèse analysées (pour l'échantillon sans ligand et à 55 mg/ml), que 27 % des gouttes sont claires, 50 % sont précipitées et 23 % présentent une séparation de phases, et aucun cristal de protéines. Ceci suggère que la concentration de la protéine est satisfaisante et qu'une optimisation de cette dernière (légère diminution) pourrait



permettre la formation de premiers cristaux de la protéine HSulf-2  $\Delta$ HD. Pour le deuxième échantillon, les mêmes pourcentages sont observés ce qui laisserait imaginer une influence du ligand sur les propriétés de cristallisation. Cependant, au vu de la qualité de l'échantillon et de la perte massive de protéines, aucune conclusion précise ne peut être avancée.

Il est à noter que six conditions ont également présenté un agrégat de protéines pêchable (un amas de micro-cristaux ou une agrégation autour d'un contaminant), car ces objets émettaient de la fluorescence une fois soumis aux UVs (Figure 23). Cependant, après l'analyse de ses différents objets par diffraction au Synchrotron ESRF, Grenoble (Ligne MASSIF), aucun diagramme de diffraction n'a pu être enregistré.

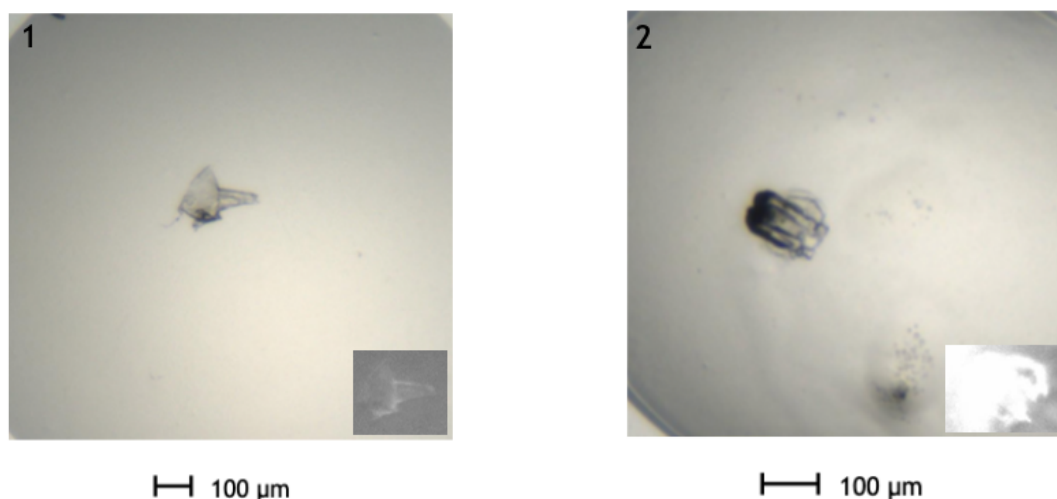


Figure 23 : Deux amas protéiques obtenus pour la protéine HSulf-2  $\Delta$ HD. 1. HSulf-2  $\Delta$ HD + Sulfocalix-[4]-arène, condition : Ammonium acétate 0,2 M, Trisodium citrate 0,1 M pH 5,6, MPD 30 % (v/v). 2. HSulf-2  $\Delta$ HD, condition : Calcium acétate 0,2 M, Tris base/HCl 0,1 M pH 7, PEG 3000 20 % (m/v). L'encart en bas à droite correspond à l'objet sous UVs.

Un test d'activité sur le substrat 4MUS a été réalisé en parallèle afin de valider l'activité arylsulfatase de la protéine. Les résultats montrent une protéine présentant un niveau d'activité équivalent à celui de HSulf-2  $\Delta$ HD exprimée en cellules HEK 293F (Figure 22C). Egalement, la stabilité de la protéine avec une glycosylation réduite a été confirmée par l'analyse de la protéine en SDS-PAGE et la réalisation d'un deuxième test d'activité 4MUS après un stockage de plusieurs mois à  $-20^{\circ}\text{C}$ .



Simultanément à ces tests de cristallogénèse, le projet de thèse visait également à étudier la structure de la protéine HSulf-2 par RMN, avec le support de l'entreprise NMR-Bio.

## **VI.2. Développement de nouveaux milieux de culture pour l'étude par RMN de protéines exprimées en cellules HEK 293**

L'étude de protéines exprimées en cellules HEK 293 par RMN nécessite de développer un milieu de culture permettant un marquage efficace de ces dernières car, à ce jour, aucun milieu de ce type n'est commercialement disponible. HSulf-2 étant une protéine complexe, le développement de ce milieu a été effectué sur une protéine modèle dans un premier temps. Il s'agit de la protéine Hsp90 N-terminal, dont la masse moléculaire est compatible avec la RMN (26,5 kDa) et dont la version produite en bactéries a préalablement été caractérisée par RMN (collaboration NMR-Bio et le Dr. Jérôme Boisbouvier, IBS). Cette approche a permis de mettre au point un milieu de culture adapté au marquage spécifique, mais également un milieu de culture permettant un marquage uniforme  $^{15}\text{N}$ . Ce n'est que dans un second temps que le milieu de marquage spécifique nouvellement développé a été appliqué pour l'étude de la protéine HSulf-2.

### **VI.2.1. Développement du milieu de marquage spécifique ( $^{13}\text{CH}_3$ -acides aminés)**

La stratégie de marquage développée ici repose sur la même stratégie qu'introduite précédemment (cf. III.3.4. p. 114), soit l'utilisation d'un milieu de culture déplété de tous les acides aminés auquel est rajouté un mélange réactionnel composé d'acides aminés libres où certains peuvent être marqués, comme les  $^{13}\text{CH}_3$ -acides aminés.

Le milieu de culture standard commercialisé pour les cellules HEK 293 est le FreeStyle 293 (Gibco®). Cependant, ce milieu de culture n'existe pas dans une version sans acides aminés. C'est pourquoi, un milieu commercial alternatif a été choisi (ESF-SFM : référence 98-003, Expression System). Ce milieu existe sous deux formes : un milieu complet et un milieu sans acides aminés. La première étape a consisté à adapter les

cellules dans ce milieu ESF-SFM complet. Néanmoins, bien que les cellules HEK 293 aient pu être adaptées et maintenues dans ce milieu de culture, aucune expression, après transfection dans ce milieu, n'a été observée, tant pour HSulf-2 (expression continue, lignées cellulaires stables) que pour la protéine Hsp90 N-terminal (expression transitoire). De la même manière, si les cellules adaptées (au milieu ESF-SFM complet) sont transférées au milieu ESF-SFM déplété, ou bien les cellules directement transférées du milieu FreeStyle au milieu ESF-SFM déplété, et transfectées dans ce dernier milieu, aucune expression protéique n'a été observée (à l'exception d'un lot de milieu ESF-SFM déplété qui a généré des rendements très faibles, cf. p. 142).

En revanche, il a pu être montré que si les cellules HEK 293F transfectées dans le milieu FreeStyle 293, étaient directement transférées de ce milieu vers le milieu ESF-SFM déplété et reconstitué en acides aminés, alors l'expression protéique était maintenue. Ainsi, à la lumière de l'ensemble de ces informations, il a été choisi de maintenir la croissance des cellules HEK 293F ou S dans le milieu FreeStyle 293, de transfecter les cellules dans ce même milieu, puis de les transférer (au bout de 4 heures sous agitation en incubateur et après lavage des cellules pour éliminer toute trace du milieu FreeStyle 293) dans le milieu ESF-SFM déplété et reconstitué avec un mélange réactionnel d'acides aminés composé d'un ou plusieurs  $^{13}\text{CH}_3$ -acides aminés.

La première étape a été de mettre au point un mélange réactionnel d'acides aminés permettant un rendement d'expression satisfaisant tout en minimisant les quantités en acides aminés utilisées, ceci dans le but de maîtriser le coût final du milieu. La composition des milieux de culture (tels que le milieu FreeStyle 293 ou le milieu ESF-SFM) n'étant pas connue, le point de départ pour développer ce mélange réactionnel d'acides aminés a été le milieu MEM Gibco® (cf. Tableau 5 p. 131). Puis, différents milieux, dont la formulation est connue (comme le milieu DMEM, Jong-Seon et al., 2013 ; Advanced D-MEM/F-12, Gibco®), ont servi de point de comparaison.

Afin d'identifier une composition optimale du mélange réactionnel (concentration en composés, ajout de glucose, de vitamines et/ou d'extraits de levure), différents tests d'optimisation ont été réalisés sur la protéine Hsp90 N-terminal.

i. Le mélange réactionnel ainsi mis au point (défini à la concentration 1x) a été ajouté au milieu de culture ESF-SFM déplété dans différentes concentrations variant de 0,5x à 8x (Figure 24), afin de corrélérer la quantité d'acides aminés à l'expression protéique.

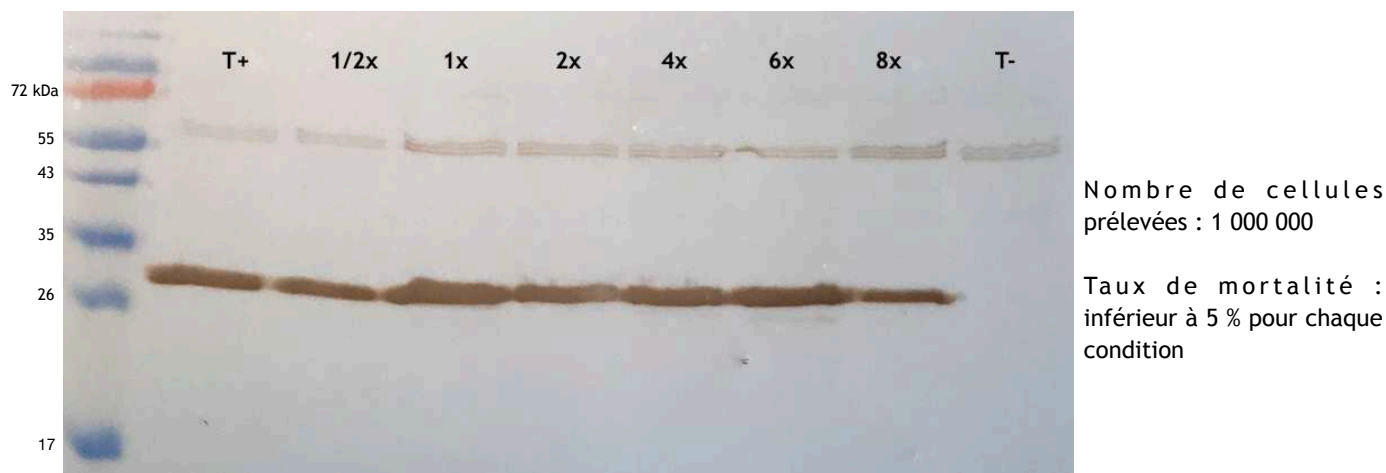


Figure 24 : Western blot du test d'expression de la protéine Hsp90 N-terminal avec une quantité variable en acides aminés supplémentés au milieu de culture (Anticorps :  $\alpha$ -His-HRP, A7058, Sigma (dil. 1/1000) et coloration de la membrane avec une pastille de 3,3'-Diaminobenzidine, SIGMAFAST™ (dilué dans 10 ml de PBS 1x, Tween20 0,05 %) car la détection par chimioluminescence n'a pas généré une image exploitable). Le puits T+ représente la protéine exprimée en FreeStyle 293 et le puits T- correspond à la condition sans transfection des cellules HEK 293F. Ceci permet d'identifier la bande présente entre les marqueurs de taille 26 et 35 kDa comme la protéine Hsp90 N-terminal. En revanche, la bande de plus haute masse moléculaire est, probablement, une reconnaissance non spécifique de l'anticorps. Le puits 3 (1x) correspond à la condition où la quantité définie dans le tableau 5 (cf. p. 131) a été rajoutée au milieu de culture. Les puits restants correspondent donc à une valeur inférieure ou supérieure de cette quantité d'acides aminés. Par exemple, 2x correspond à une concentration deux fois plus importante en acides aminés que dans la condition 1x.

*Pour rappel, cette expérience est réalisée avec un même nombre de cellules.*

Au vu de l'intensité des bandes révélées par western blot, il peut être conclu qu'une quantité en acides aminés inférieure à celle définie dans le mélange réactionnel (puits 2 - 1/2x) induit une expression légèrement plus faible que la protéine exprimée avec le mélange réactionnel 1x (puits 3 - 1x). En revanche, supplémenter le milieu avec plus d'une fois la quantité en acides aminés présente dans le mélange réactionnel ne semble pas améliorer le niveau d'expression de la protéine Hsp90 N-terminal. Par conséquent, le milieu de marquage spécifique sera supplémenté avec une concentration 1x en acides aminés correspondant à celle détaillée dans le tableau 5 (cf. p. 131). Néanmoins, il est à noter que le milieu de culture ainsi reconstitué présente, à la fin des quatre jours

d'expression, un taux de croissance cellulaire qui est inférieur de 20 à 25 % pour l'ensemble des conditions ESF-SFM reconstituées par rapport à la condition FreeStyle 293 (T+).

Un second test, où un même volume de culture a été prélevé (et non une même quantité de cellules), montre cette même différence d'expression (Figure 25, puits 1x et T+). Deux hypothèses peuvent être avancées pour expliquer cette différence : (i) il peut manquer un composant essentiel tel que du glucose ou des vitamines, ou (ii) le transfert brutal des cellules du milieu FreeStyle 293 au milieu "maison" n'est pas bien toléré par les cellules qui peinent à retrouver une croissance cellulaire optimale. Pour vérifier la première hypothèse, du glucose et/ou des vitamines ont été ajoutés à différentes concentrations au milieu de culture (Figure 25). Cependant, l'ajout de ces composés n'améliore pas la croissance cellulaire et, au contraire, semble être défavorable à l'expression de la protéine. Ceci étant, il est difficile d'expliquer ce résultat car l'ensemble des conditions faisant intervenir le milieu ESF-SFM dépleté et reconstitué montrait une densité de cellules d'environ  $3 \cdot 10^6$  cellules/mL (contre  $\sim 4 \cdot 10^6$  cellules/mL pour la condition T+). Par conséquent, il peut être imaginé que le glucose et les vitamines aient limité la transfection des cellules.

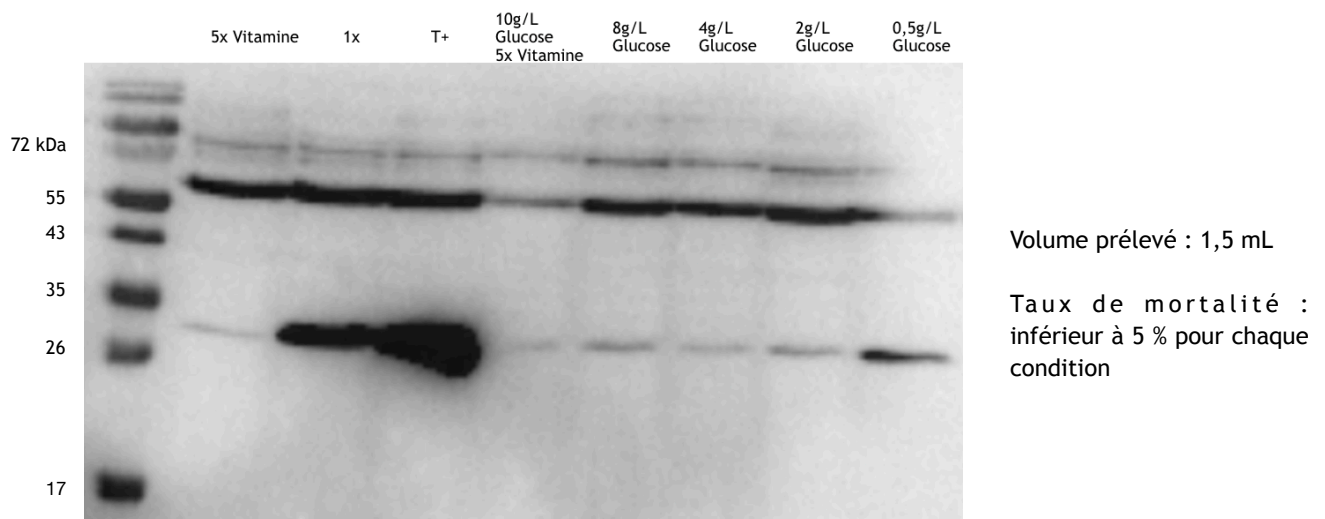


Figure 25 : Western blot du test d'optimisation de la croissance cellulaire et de l'expression de la protéine Hsp90 N-terminal (Anticorps :  $\alpha$ -His-HRP, A7058, Sigma (dil. 1/1000)). Le puits T+ représente la protéine exprimée en milieu FreeStyle 293. Le puits 1x représente la condition où seuls des acides aminés sont ajoutés au milieu de culture. Les puits restants comportent tous la même quantité d'acides aminés que dans le puits 1x mais des concentrations variables en vitamines et glucose ont été rajoutées dans le milieu de culture.

*Pour rappel, la lyse des cellules est réalisée à partir d'un même volume de culture cellulaire (1,5 mL) et non pas à partir d'une même quantité de cellules.*

ii. L'influence de l'extrait de levure a également été étudiée d'après les travaux de Gossert *et al.* (2011) et Zhang *et al.* (2017) (cf. III.3.3. p. 113). Ces études portent sur l'ajout d'extrait de levure dialysé ou non dans le milieu de culture de cellules d'insectes. Pour rappel, l'extrait de levure est un composant essentiel des milieux de culture de ce système d'expression et son élimination totale entraîne une mort cellulaire. Or, les extraits de levure contiennent beaucoup d'acides aminés (incompatibles avec la mise au point d'un milieu de marquage isotopique pour la RMN). Toutefois, il a été démontré qu'en diminuant la quantité d'extrait de levure ajoutée et en dialysant ce dernier (afin d'éliminer les acides aminés libres), le taux d'incorporation restait important et les rendements protéiques étaient équivalents au milieu commercial standard. Le milieu Freestyle 293 ne contenant vraisemblablement pas d'extrait de levure (sans information de l'industriel sur la formulation du milieu, il est nécessaire de nuancer cette information), un test a donc été réalisé afin d'identifier si un tel ajout permettait de maintenir la croissance cellulaire (Figure 26). Ainsi, après avoir supplémenté le milieu de culture avec le mélange réactionnel d'acides aminés ou le mélange réactionnel d'acides aminés plus de l'extrait de levure dialysé ou non à des concentrations de 1 g/L et 4 g/L, il apparaît des niveaux d'expression similaires entre les différentes conditions et une croissance cellulaire toujours réduite ( $\sim 2,9 \cdot 10^6$  cellules/mL contre  $\sim 3,7 \cdot 10^6$  cellules/mL).

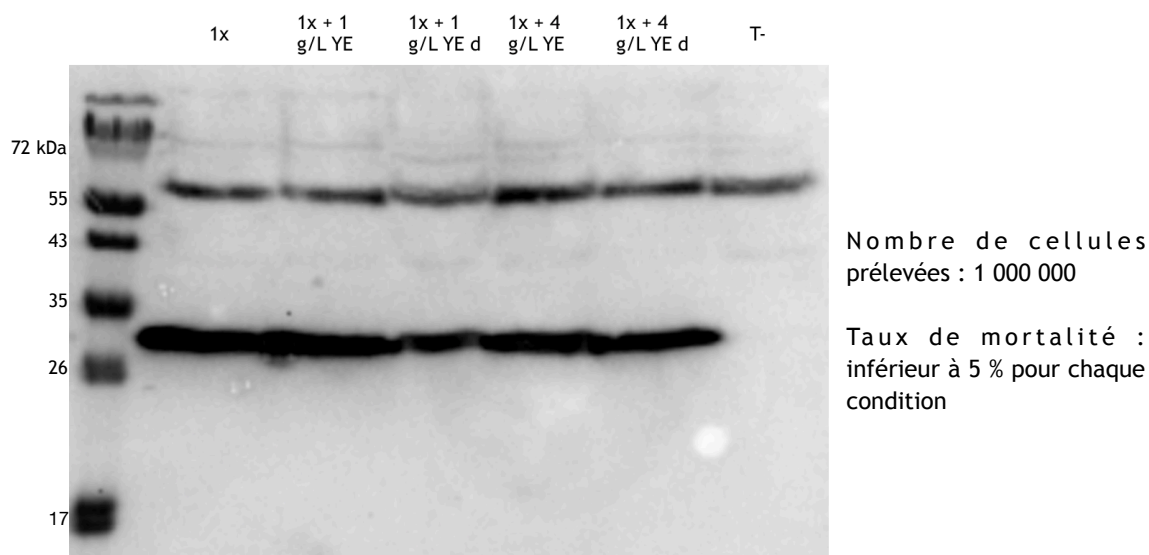


Figure 26 : Western blot du test d'optimisation de la croissance cellulaire et de l'expression de la protéine Hsp90 N-terminal (Anticorps :  $\alpha$ -His-HRP, A7058, Sigma (dil. 1/1000)). Le puits T- représente la condition sans transfection de la protéine cible. Le puits 1x représente la condition où seuls des acides aminés sont ajoutés au milieu de culture. Les puits restants comportent tous la même quantité d'acides aminés que dans le puits 1x, mais aux milieux de culture ont été supplémentés des concentrations variables en extrait de levure dialysé (YE d) ou non (YE).

*Pour rappel, cette expérience est réalisée avec un même nombre de cellules.*

En conséquence, le mélange réactionnel d'acides aminés 1x (cf. Tableau 5 p. 131) a été validé pour la suite des expériences ; tout en sachant que l'utilisation de ce milieu entraîne une chute de la croissance et donc du rendement protéique final d'environ 20 à 25 %. Notons toutefois qu'une adaptation du protocole d'expression permettrait d'atteindre les mêmes rendements protéiques que ceux observés dans le milieu de culture standard (cf. p. 171).

En résumé, dès lors qu'une expression de protéines marquées est réalisée, les cellules sont transfectées dans le milieu de culture FreeStyle 293, puis transférées dans le milieu marqué constitué du milieu de culture ESF-SFM déplété et reconstitué en acides aminés non marqués (à la concentration détaillée dans le tableau 5) et/ou marqués.

### VI.2.2. Validation des rendements d'expression de la protéine Hsp90 N-terminal dans le milieu de culture développé

Afin de valider le milieu de culture développé, la protéine Hsp90 N-terminal a été exprimée dans ce dernier et purifiée.

Après quatre jours d'expression, une lyse chimique et deux étapes de purification (chromatographie d'affinité sur NI-NTA et d'exclusion sur Superdex200 10/300 GL) ont été réalisées. Il en résulte un chromatogramme présentant un pic majoritaire (Figure 27A). Un échantillonnage complet de ce chromatogramme a été déposé et analysé par SDS-PAGE coloré au bleu de Coomassie (Figure 27B). Le pic dont le volume d'élution est de 15,5 mL montre une seule bande juste au-dessus du marqueur de taille 26 kDa, ce qui correspond à la protéine Hsp90 N-terminal (26,5 kDa). Le rendement final est estimé à 2,7 mg/L.

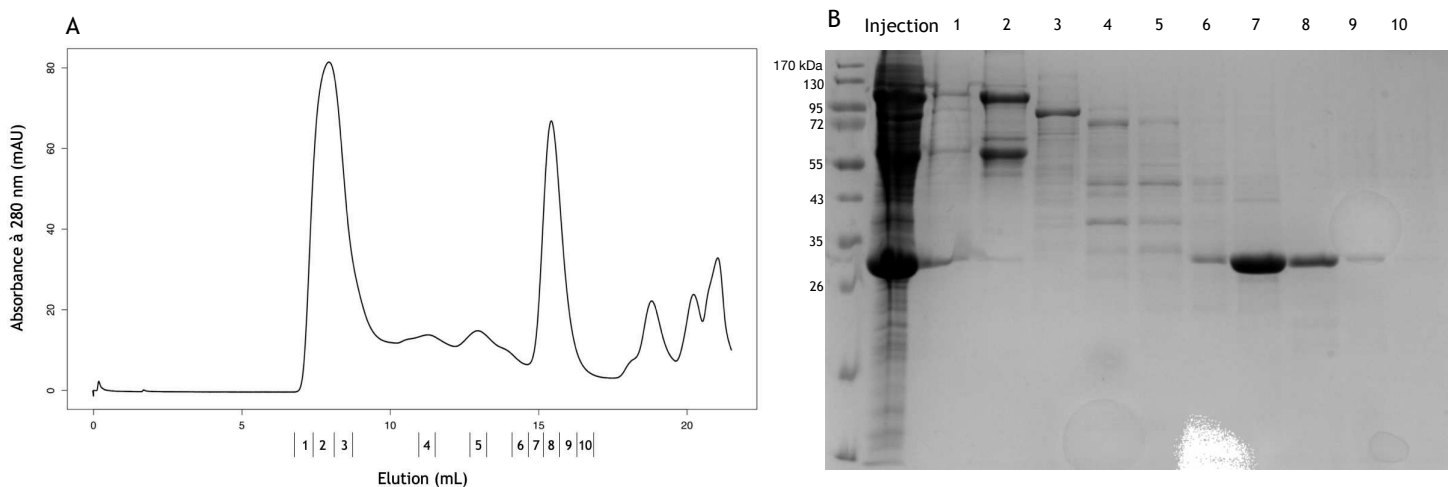


Figure 27 : Purification de la protéine Hsp90 N-terminal (non marquée) à partir du milieu de marquage développé. (A) La chromatographie d'exclusion présente deux pics : le premier à 7,5 mL (correspondant au volume mort) et le deuxième à 15,5 mL. (B) SDS-PAGE correspondant au chromatogramme (gel d'acrylamide 15 %). Les fractions 6 à 10 présentent une bande majoritaire dont la masse moléculaire correspond à la protéine Hsp90 N-terminal.

La comparaison des rendements protéiques a été réalisée en exprimant la même protéine en milieu FreeStyle 293 (et en cellules HEK 293F). En utilisant un protocole d'expression et de purification rigoureusement identiques, un chromatogramme présentant un pic majoritaire est également obtenu (Figure 28A). Un échantillonnage complet de ce chromatogramme a aussi été déposé et analysé par SDS-PAGE coloré au bleu de Coomassie (Figure 28B). L'analyse du gel reste identique à celle de la figure 27B. Après concentration des fractions contenant la protéine cible, un rendement de 3,5 mg/L est obtenu.

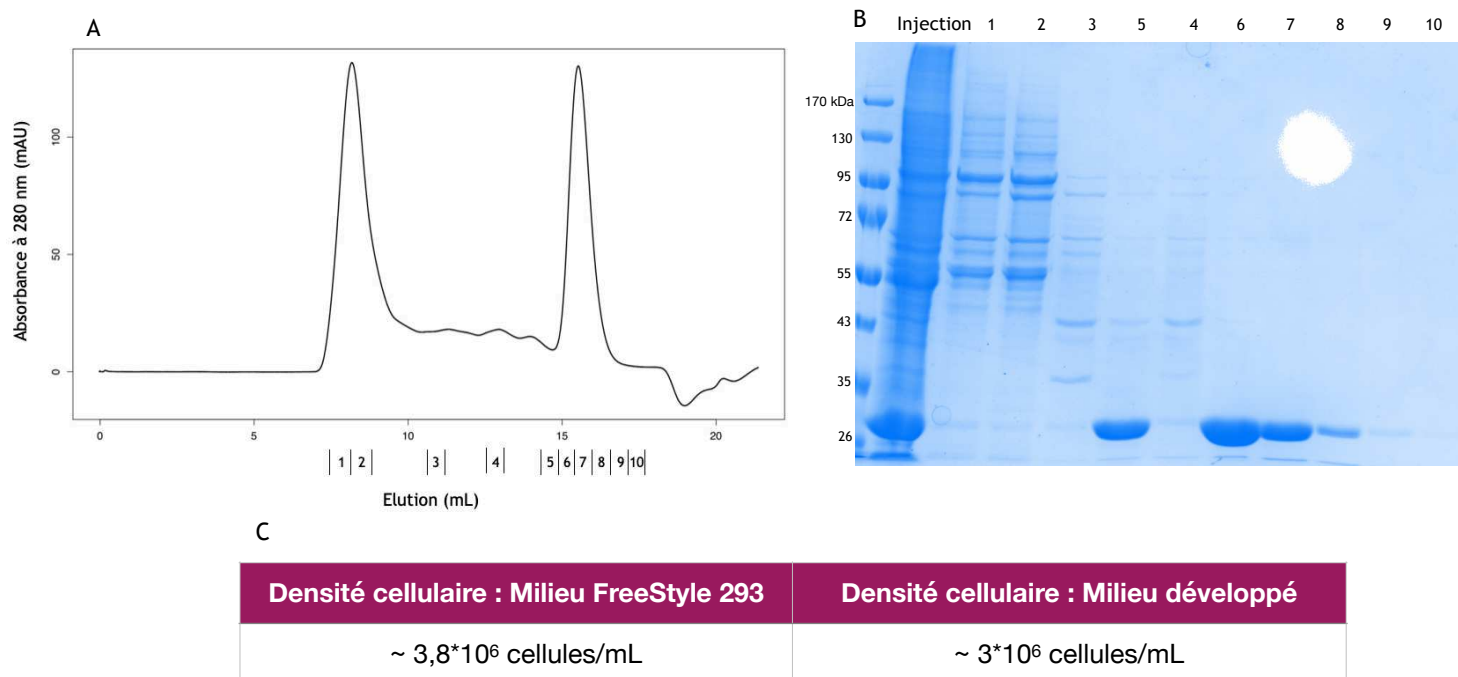


Figure 28 : Purification de la protéine Hsp90 N-terminal (non marquée) à partir du milieu de culture standard (FreeStyle 293). (A) La chromatographie d'exclusion présente deux pics : le premier à 7,5 mL (correspondant au volume mort) et le deuxième à 15,5 mL. (B) SDS-PAGE correspondant au chromatogramme (gel d'acrylamide 10 %), les fractions 5 à 10 présentent une bande majoritaire dont la masse moléculaire correspond à la protéine Hsp90 N-terminal. (C) Comparaison des densités cellulaires après quatre jours d'expression.

Ainsi, en comparant les deux tests d'expression de Hsp90 N-terminal, réalisés dans deux milieux de culture différents (milieu standard et milieu fait maison), une perte de rendement d'environ 20 à 25 % est observée avec le milieu développé. Cette perte peut être corrélée avec la diminution de la croissance cellulaire inhérente à ce milieu (Figure 28C). Cependant, il a pu être déterminé qu'une adaptation du protocole d'expression pouvait compenser cette diminution de l'expression et des rendements protéiques. En effet, il a été observé un temps d'adaptation (sans croissance cellulaire) de 24 heures des cellules HEK 293 dans le milieu développé avant qu'elles initient leur croissance et division. Or, après ce temps d'adaptation, la viabilité et la croissance cellulaire se révèlent identiques à celles des cellules HEK 293 dans le milieu standard (FreeStyle 293). De même, les rendements protéiques se révèlent identiques dès lors qu'un jour d'expression supplémentaire est appliqué aux cellules exprimées dans le milieu développé. Ce délai permet, en toute vraisemblance, de compenser les 24 heures d'adaptation des cellules au milieu.



En résumé, dès lors qu'une expression de protéines marquées est réalisée, les cellules sont transfectées dans le milieu de culture FreeStyle 293, puis transférées dans le milieu développé pour un temps d'expression supérieur de 24 heures, dans le but d'atteindre un rendement protéique identique à celui observé initialement en milieu FreeStyle 293.

Une fois démontré que le milieu de marquage spécifique développé permettait l'expression de la protéine Hsp90 N-terminal dans des rendements compatibles avec la RMN et comparable avec le milieu standard, il a été nécessaire de valider les rendements et les taux d'incorporation des acides aminés dans le milieu marqué.

Pour ce faire, le milieu ESF-SFM déplété a été supplémenté avec les mélanges réactionnels EAA, NEAA, et Leu, 0,13 g/L de L-Tyr, 5 mg/L de  $\beta$ -chloro-alanine (inhibiteur des alanines transaminases, afin d'inhiber la synthèse endogène des alanines) et 100 mg/L de *U*-[<sup>1</sup>H], [<sup>13</sup>C<sup>1</sup>H<sub>3</sub>] <sup>$\epsilon$</sup> -Met (299146, Sigma Aldrich) ; *U*-[<sup>1</sup>H], [<sup>13</sup>C<sup>1</sup>H<sub>3</sub>] <sup>$\beta$</sup> -Ala (CLM-117-pk) ; <sup>15</sup>N-2,3,3,4,4,4-(D5)-proS-<sup>13</sup>CH<sub>3</sub>-L-Val (NMR-Bio) ; <sup>15</sup>N-2,3,4,4-(D4)-3-méthyl-(D3) 5-<sup>13</sup>CH<sub>3</sub>-L-Ile <sup>$\delta$ 1</sup> (NMR-Bio) ; 2,3-(D<sub>2</sub>) 4-(<sup>13</sup>CH<sub>3</sub>)-L-Thr (NMR-Bio) (nomenclature simplifiée pour la suite en : <sup>13</sup>CH<sub>3</sub>-Met, <sup>13</sup>CH<sub>3</sub>-Ala, <sup>13</sup>CH<sub>3</sub>-Val<sup>proS</sup>, <sup>13</sup>CH<sub>3</sub>-Ile <sup>$\delta$ 1</sup>, <sup>13</sup>CH<sub>3</sub>-Thr). Une fois les cellules transfectées, elles sont transférées dans le milieu de marquage et laissées en culture pendant quatre jours. Après purification, un rendement d'expression protéique de 2,7 mg/L est obtenu.

Par la suite, l'échantillon a été concentré à 2,7 mg/mL (~ 100  $\mu$ M) et analysé par RMN (Figure 29).

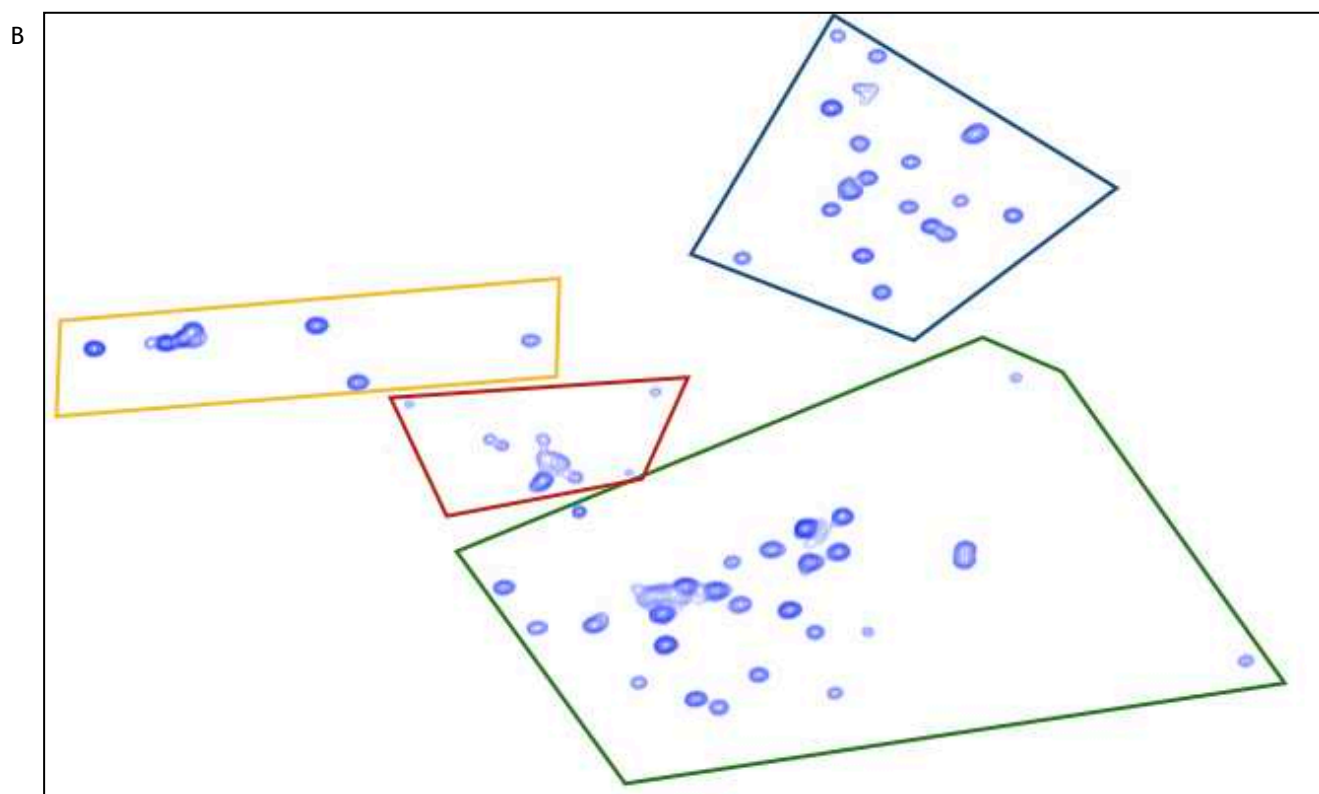
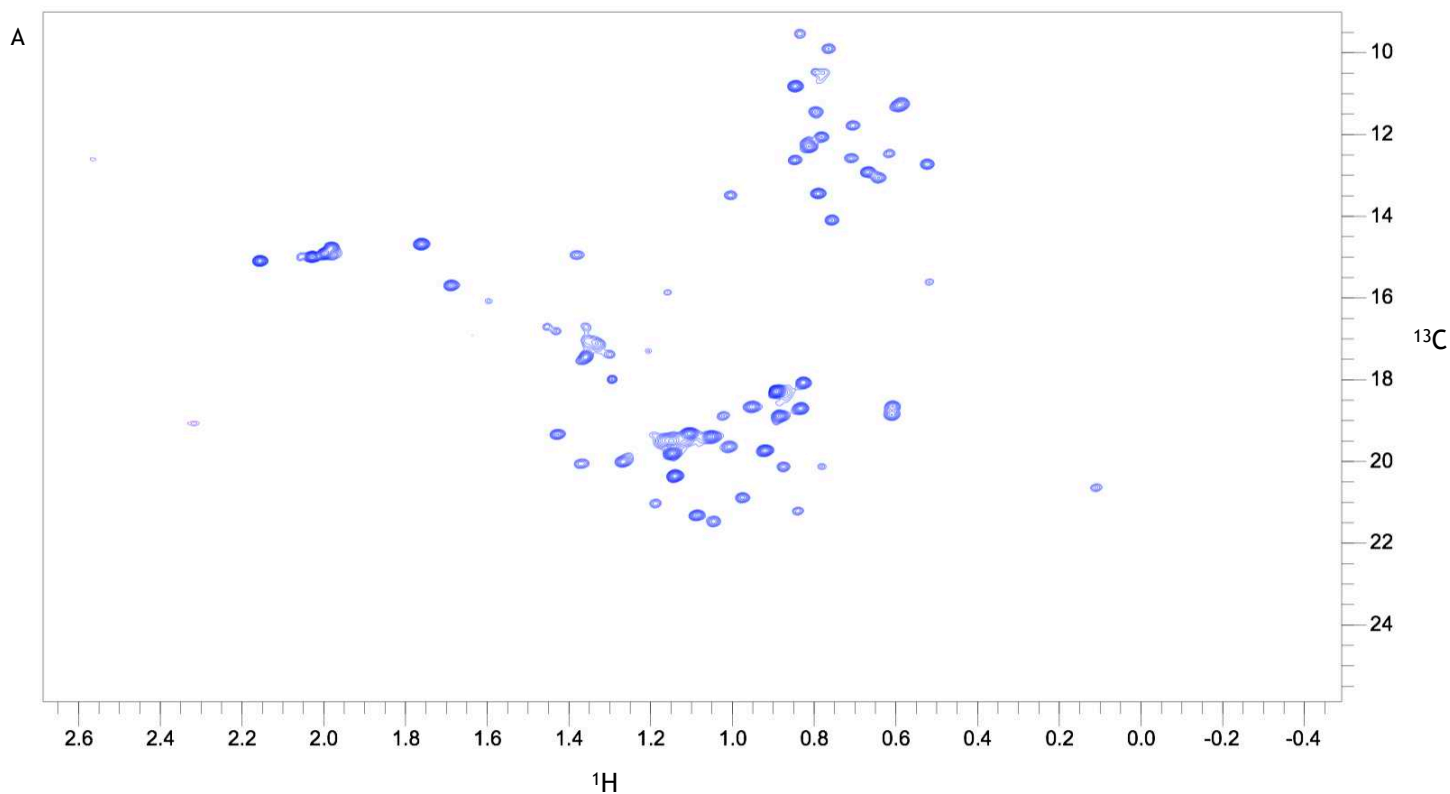


Figure 29 : (A) Spectre RMN 2D de la protéine Hsp90 N-terminal marquée  $^{13}\text{CH}_3\text{-Met/-Ile}^{\delta 1}\text{-Val}^{\text{proS}}\text{-Thr/-Ala}$ . (B) Agrandissement et classification des pics du spectre RMN 2D : les cases de différentes couleurs délimitent les régions spécifiques de résonance des 5 acides aminés marqués  $^{13}\text{CH}_3$ . **Jaune** :  $^{13}\text{CH}_3\text{-Met}$  ; **Rouge** :  $^{13}\text{CH}_3\text{-Ala}$  ; **Vert** :  $^{13}\text{CH}_3\text{-Thr}$  et  $^{13}\text{CH}_3\text{-Val}^{\text{proS}}$  ; **Bleu** :  $^{13}\text{CH}_3\text{-Ile}^{\delta 1}$ . Echantillon concentré à 100  $\mu\text{M}$ , acquisition à 25°C, tampon : HEPES 20 mM pH 7,5, NaCl 150 mM, TCEP 1 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une cryosonde.

Sur ce spectre, quatre régions de pics sont parfaitement distinguables et chacune correspond à un type de résidu portant un groupement méthyle, à l'exception de la région verte qui comprend à la fois les  $^{13}\text{CH}_3\text{-Val}^{\text{proS}}$  et les  $^{13}\text{CH}_3\text{-Thr}$  qu'on ne peut pas différencier sur la base de leurs déplacements chimiques. Cependant, seulement trois régions (sur les quatre définies) présentent des intensités de pics correctes. En effet, les régions des Met, Ile, Thr et Val sont très bien résolues et présentent des intensités de pics visuellement comparables entre elles. Ceci est cohérent (et était attendu) car ces quatre acides aminés étant essentiels, leurs incorporations devraient être similaires pour une même concentration. Notons toutefois que pour l'ensemble de ces régions, certains pics des Ile, Thr et Val sont plus intenses que d'autres. Ceci n'indique pas une différence d'incorporation entre mêmes résidus mais seulement une différence de sensibilité lors de l'acquisition. Il est vrai que suivant la position des acides aminés dans la structure 3D de la protéine (régions plus ou moins flexibles, par exemple), certains résidus peuvent être significativement mieux détectés du fait de propriétés de relaxation transversale favorables. De plus, certains pics se superposent. Dans le cas de la Met, par exemple, sept pics sont attendus mais seulement quatre sont parfaitement isolés et trois se superposent en partie. Ce type de superposition est aussi retrouvée pour les autres sondes. Ceci n'est toutefois pas un problème inhérent à la stratégie de marquage mais aux propriétés de résonance des sondes.

En revanche, pour la région des alanines (rouge) le marquage n'est pas satisfaisant. En effet, de nombreux pics manquent (15 pics sont attendus) et les quelques pics résolus présentent de très faibles intensités. Ceci pourrait s'expliquer par une dilution des  $^{13}\text{CH}_3\text{-Ala}$  par de l'alanine non marquée produite par les cellules à partir du glucose du milieu de culture. En effet, il apparaît, très probablement, que la concentration en inhibiteur ( $\beta$ -chloro-alanine à 5 mg/L) est trop faible. Or, si les alanines transaminases ne sont pas efficacement inhibées, alors la production d'alanine endogène non marquée diluera d'autant la  $^{13}\text{CH}_3\text{-Ala}$ . Par conséquent, une optimisation des conditions d'inhibition des alanines transaminases est indispensable afin de générer un marquage efficace de cet acide aminé (cf. Figure 32 et 33 p. 178-179).

Afin de valider la qualité de ce milieu de marquage et la stratégie développée, le spectre obtenu a été comparé à celui acquis sur la protéine Hsp90 N-terminal exprimée en bactéries et marquée  $^{13}\text{CH}_3\text{-Met}$ ,  $^{-}\text{Ile}^{\delta 1}$ ,  $^{-}\text{Val}^{\text{proS}}$ ,  $^{-}\text{Thr}$ ,  $^{-}\text{Ala}$  et  $^{-}\text{Leu}^{\text{proS}}$  grâce à 1 kit de marquage commercial (NMR-Bio) (Figure 30). Une très large majorité des pics sont

superposés, montrant ainsi que la protéine exprimée dans le milieu de culture développé est correctement structurée et marquée. Les pics seulement visibles sur le spectre RMN 2D rouge (0,2-1,1 ppm pour le  $^1\text{H}$  et 20-24 ppm pour le  $^{13}\text{C}$ ) appartiennent aux  $^{13}\text{CH}_3\text{-Leu}^{\text{proS}}$  qui ont été marquées seulement pour Hsp90 N-terminal exprimée en bactéries. L'absence de superpositions pour certains pics peut s'expliquer par la présence de PTMs pour la protéine exprimée en cellules eucaryotes (cf. Figure 34A p. 180) et des variations dans la préparation des échantillons et l'acquisition des spectres.

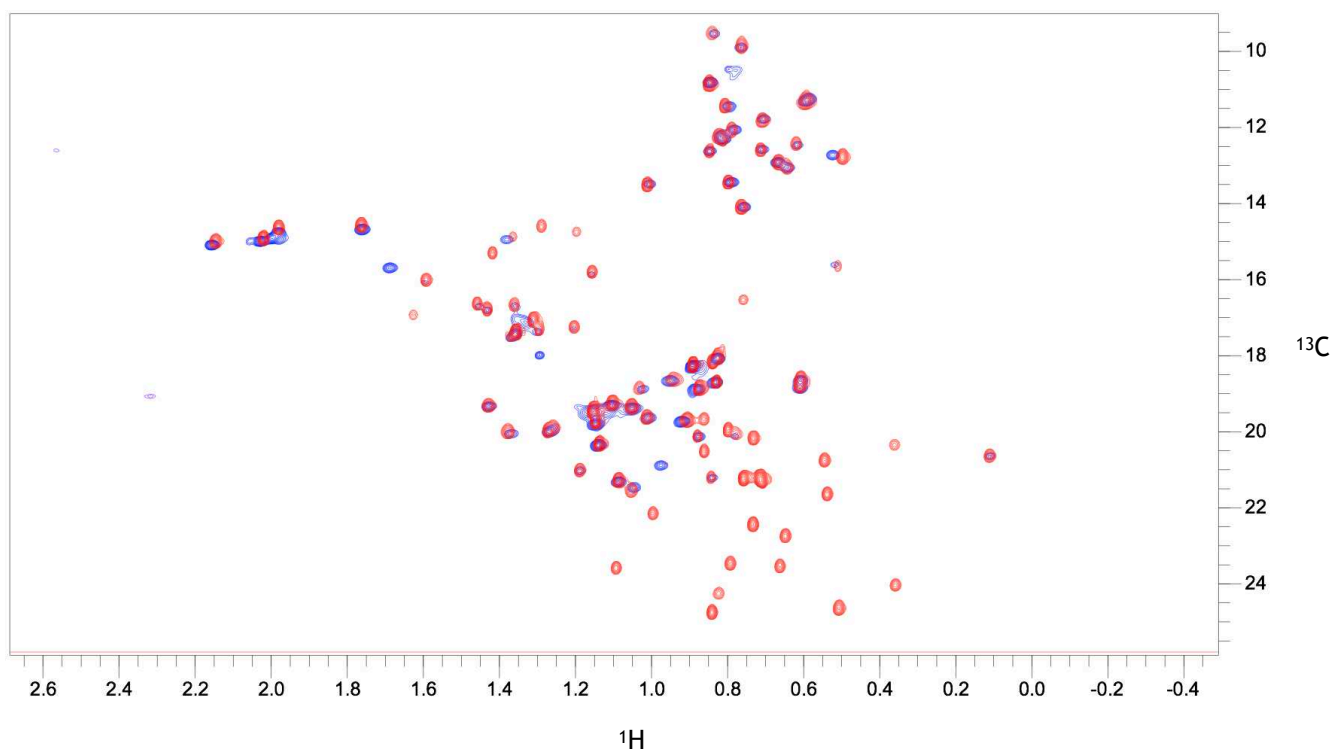


Figure 30 : Superposition des spectres 2D de la protéine Hsp90 N-terminal  $^{13}\text{CH}_3\text{-Met}$ , Ile, Val $^{\text{proS}}$ , Thr, Ala exprimée en cellules HEK 293F (spectre **bleu**) et Hsp90 N-terminal  $^{13}\text{CH}_3\text{-Met}$ , Ile, Val $^{\text{proS}}$ , Thr, Ala, Leu $^{\text{proS}}$  exprimée en bactérie (spectre **rouge**). Notons de nombreux pics (spectre **rouge**) non superposés en bas du spectre. Ces pics correspondent à des  $^{13}\text{CH}_3\text{-Leu}^{\text{proS}}$  qui ne sont pas marqués dans l'échantillon produit en cellules de mammifère.  
*L'échantillon Hsp90 N-terminal exprimé en bactéries a été fourni par le Dr Rime KERFAH, NMR-Bio.*

Développer et commercialiser un milieu de marquage performant et économiquement compétitif nécessite de définir la concentration minimale en acides aminés marqués ( $^{13}\text{CH}_3$ ) à ajouter au milieu et permettant un taux d'incorporation maximal.

Pour cela, l'approche utilisée s'est faite en deux étapes. La première a consisté à établir des courbes d'incorporation pour chaque  $^{13}\text{CH}_3$ -acide aminé pour obtenir un plateau d'incorporation (après analyse des échantillons marqués par RMN). Ce plateau d'incorporation signifie que, même si plus d'acides aminés marqués  $^{13}\text{CH}_3$  sont ajoutés au milieu, le taux d'incorporation n'augmentera pas. Une fois cette quantité minimale d'acides aminés marqués  $^{13}\text{CH}_3$  identifiée, une validation par spectrométrie de masse est indispensable pour analyser et mesurer quantitativement la masse des échantillons marqués afin d'en déduire un pourcentage d'incorporation pour chaque sonde.

### VI.2.3. Mesure des taux d'incorporation des $^{13}\text{CH}_3$ -acides aminés

Afin d'identifier un plateau d'incorporation pour les  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ , Val $^{\text{proS}}$  et Thr, quatre marquages combinatoires (Met, Ile $^{\delta 1}$ , Val $^{\text{proS}}$  et Thr) de la protéine Hsp90 N-terminal ont été réalisés avec quatre concentrations différentes en  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ , Val $^{\text{proS}}$  et Thr (40 mg/L, 60 mg/L, 80 mg/L, 100 mg/L). La  $^{13}\text{CH}_3$ -Met a été supplémentée en excès (200 mg/L) dans le milieu (en raison de son faible coût) pour servir de référentiel interne. Une fois les protéines exprimées et purifiées, les spectres RMN 2D (expérience SOFAST-HMQC) ont été acquis, et les aires de chaque pic ont été calculées. Ainsi, en fixant arbitrairement que la  $^{13}\text{CH}_3$ -Met, ajoutée en excès, est incorporée à 100 % et en calculant le *ratio* de la moyenne des aires des pics correspondant aux  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ , Val $^{\text{proS}}$  et Thr sur celle des  $^{13}\text{CH}_3$ -Met, des courbes d'incorporation ont pu être tracées pour chaque  $^{13}\text{CH}_3$ -acide aminé (Figure 31).

Pour l'ensemble des trois résidus Ile, Val et Thr, l'allure des courbes est celle attendue. Les valeurs d'incorporation augmentent bien avec la quantité croissante en  $^{13}\text{CH}_3$ -acides aminés, puis un plateau apparaît pour une concentration d'acides aminés marqués ajoutés de 80 mg/L. Ceci signifie qu'au-delà de 80 mg/L d'acides aminés essentiels ajoutés, il n'y aura pas davantage d'incorporation des sondes isotopiques. Notons l'absence d'une courbe pour la  $^{13}\text{CH}_3$ -Leu car cet acide aminé n'est pas commercialement disponible (synthèse à façon avec de longs délais de production).

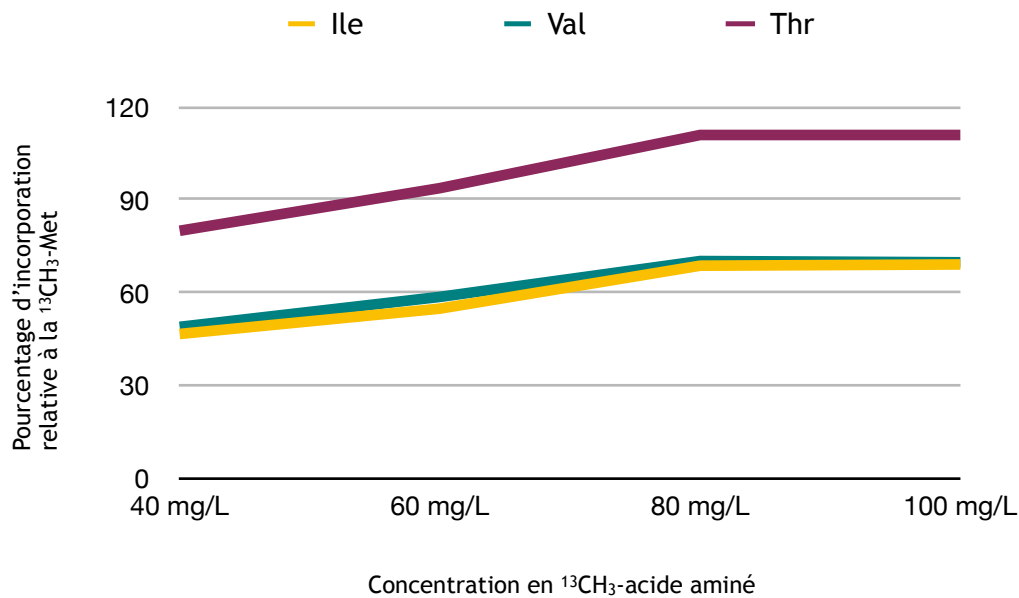


Figure 31 : Courbes d'incorporation des résidus  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ , Val $^{\text{proS}}$  et Thr. La concentration 80 mg/L représente la valeur minimale en acides aminés marqués à ajouter au milieu de marquage spécifique pour une incorporation maximale en isotopes (expérience réalisée une fois, n=1).

Pour définir la courbe d'incorporation de la  $^{13}\text{CH}_3$ -Ala, il faut à la fois tester différentes concentrations en alanine marquée et bloquer efficacement la synthèse de l'alanine par la cellule en ajoutant un inhibiteur de l'alanine transaminase dans le milieu de culture. Initialement, en ajoutant 5 mg/L de  $\beta$ -chloro-alanine dans le milieu de culture après la transfection, 22 % d'incorporation relative à l'incorporation des  $^{13}\text{CH}_3$ -Met ont été mesurés à partir du spectre RMN 2D de la figure 29. Afin d'accroître l'incorporation des alanines, trois stratégies différentes ont été mises en place (cf. p. 141-143) : (i.) une augmentation de la concentration en inhibiteur ( $\beta$ -chloro-alanine multipliée par huit), (ii.) une augmentation de la concentration en inhibiteur couplée à une incubation des cellules avec cet inhibiteur plusieurs heures avant la transfection (afin d'éliminer de possibles réserves d'alanine au sein des cellules), et (iii.) un changement d'inhibiteur des transaminases, soit l'utilisation de la L-cyclo-sérine (utilisée dans les cellules d'insectes, Wong *et al.*, 1973 ; Kato *et al.*, 1991) (Figure 32).

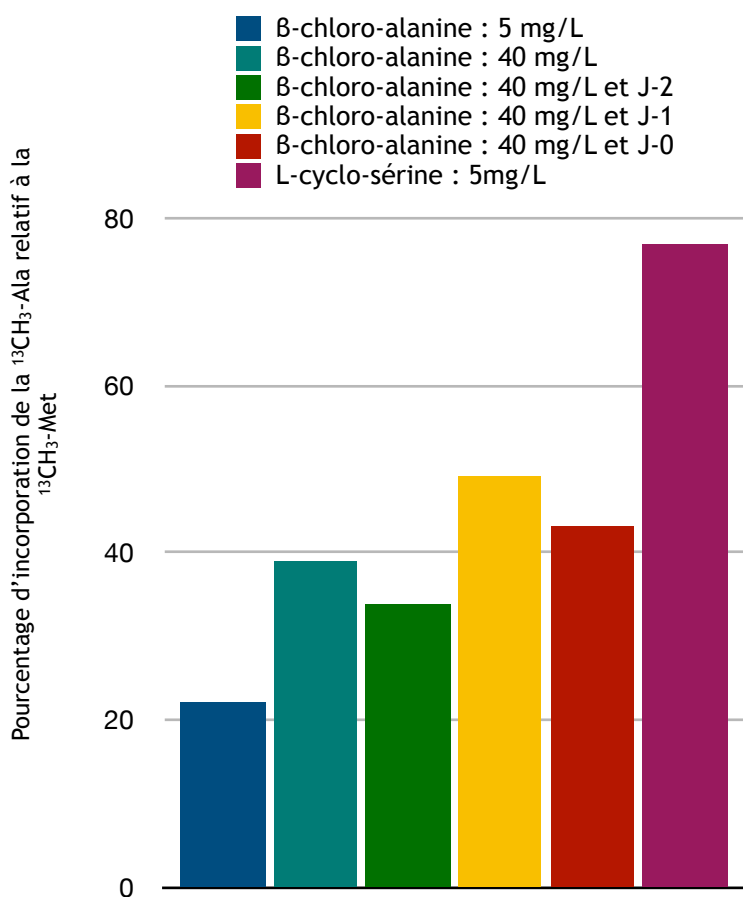


Figure 32 : Comparaison des taux d'incorporation des  $^{13}\text{CH}_3\text{-Ala}$  (relatifs au  $^{13}\text{CH}_3\text{-Met}$ ) mesurés d'après différentes conditions.

- **Bleu foncé** : 5 mg/L de  $\beta$ -chloro-alanine sont ajoutées (après transfection dans le milieu FreeStyle 293) dans le milieu de marquage.

- **Bleu** : 40 mg/L de  $\beta$ -chloro-alanine sont ajoutées (après transfection dans le milieu FreeStyle 293) dans le milieu de marquage.

- **Vert/Jaune/Rouge** : les cellules sont maintenues dans le milieu ESF-SFM déplété et reconstitué avec 19 acides aminés (sans l'alanine) et 40 mg/L de  $\beta$ -chloro-alanine pendant 48h (**vert**) et 24h (**jaune**) avant la transfection dans le milieu ESF-SFM déplété ; ou au moment de la transfection dans le milieu ESF-SFM déplété (**rouge** - même condition que l'histogramme bleu afin de comparer les résultats entre deux expériences).

- **Violet** : 5 mg/L de L-cyclo-sérine sont ajoutées au milieu de marquage, après transfection dans le milieu FreeStyle 293. (Résultats obtenus à partir d'un seul échantillon et d'un seul spectre RMN 2D).

L'augmentation de la concentration en  $\beta$ -chloro-alanine a permis d'augmenter le taux d'incorporation de 22 % à 40 % environ. Le fait d'incuber les cellules avec la  $\beta$ -chloro-alanine en amont de la transfection, dans le but d'éliminer la production endogène d'alanine, n'a en revanche pas considérablement amélioré l'incorporation de l'alanine marquée. De plus, le succès de la transfection en milieu ESF-SFM déplété n'ayant été observée qu'une seule fois et la  $\beta$ -chloro-alanine, à 40 mg/L, étant toxique pour les cellules HEK 293 (ce qui s'est traduit par une diminution de 50 % des rendements protéiques quand la  $\beta$ -chloro-alanine est ajoutée au moment de la transfection et de 70 % lorsque les cellules sont incubées avec cet inhibiteur 24 ou 48 heures avant), cette stratégie ne peut être viable.

En revanche, la troisième condition (ajout de 5 mg/L de L-cyclo-sérine au milieu de marquage sans phase d'incubation des cellules avec cet inhibiteur) engendre une croissance cellulaire, une viabilité cellulaire et un rendement protéique identiques à ceux obtenus sans inhibiteur, ainsi qu'un taux d'incorporation des  $^{13}\text{CH}_3\text{-Ala}$  (relatif à l'incorporation des  $^{13}\text{CH}_3\text{-Met}$ ) de 77 % (Figure 33). L'ajout de 5 mg/L de L-cyclo-sérine a donc été validé pour, dans un second temps, préparer des échantillons avec différentes concentrations d'alanine afin d'établir la courbe d'incorporation de ce résidu.

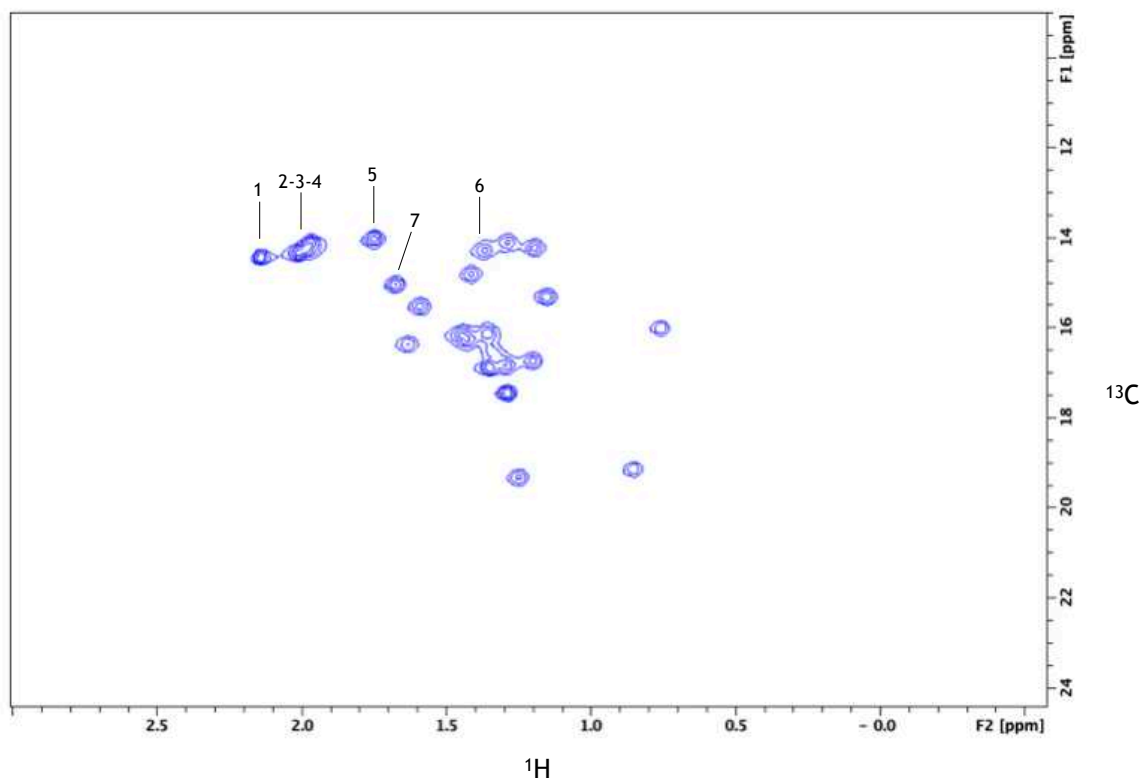


Figure 33 : Spectre RMN 2D de l'optimisation de l'incorporation des  $^{13}\text{CH}_3\text{-Ala}$ . Les pics 1 à 7 sont des  $^{13}\text{CH}_3\text{-Met}$  et les pics restants sont des  $^{13}\text{CH}_3\text{-Ala}$ . Présence de 5mg/L de L-cyclo-sérine dans un échantillon marqué  $^{13}\text{CH}_3\text{-Met/Ala}$  concentré à 90  $\mu\text{M}$ , acquisition à 25°C, tampon : HEPES 20 mM pH 7,5, NaCl 150 mM, TCEP 1 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une crysonde.

Ainsi, quatre conditions présentant une même concentration en  $^{13}\text{CH}_3\text{-Met}$  (supplémentée en excès à 200 mg/L) et en L-cyclo-sérine (5 mg/L), et des concentrations croissantes en  $^{13}\text{CH}_3\text{-Ala}$  (60 mg/L, 80 mg/L, 100 mg/L et 140 mg/L) ont été réalisées. Un cinquième échantillon, avec une concentration en alanine de 80 mg/L et deux fois plus d'inhibiteur (soit 10 mg/L) a également été produit. En rapportant la moyenne de l'intégration des pics de la  $^{13}\text{CH}_3\text{-Ala}$  à la moyenne de l'intégration des pics de la  $^{13}\text{CH}_3\text{-Met}$ , les conditions présentent toutes entre 70 et 73 % d'incorporation

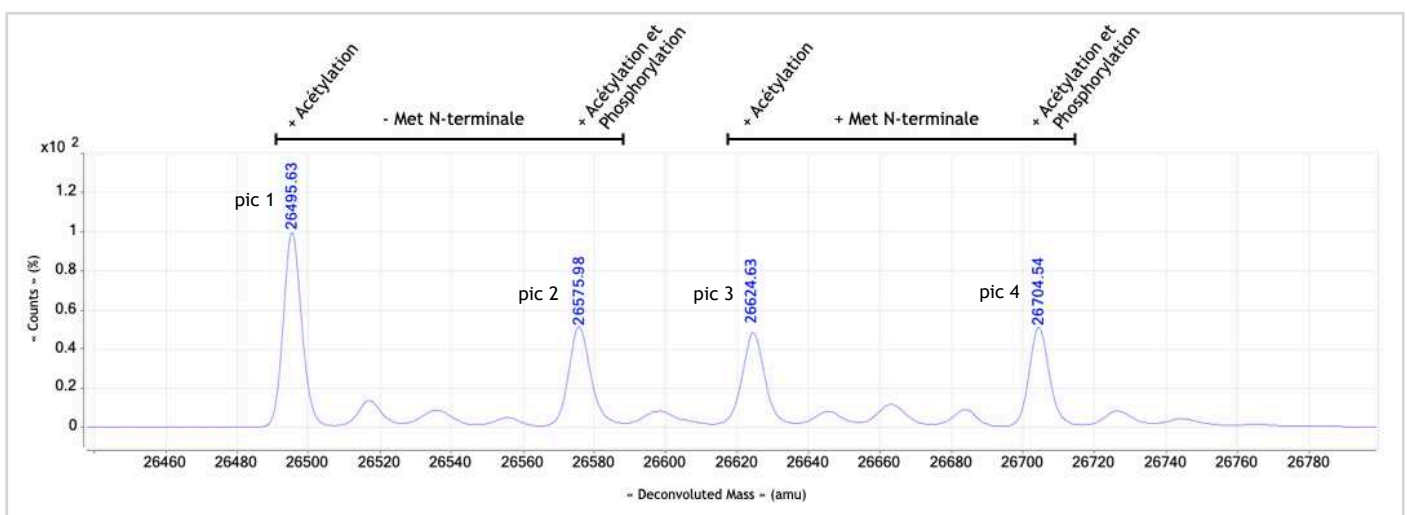


relative. Cela signifie qu'au-delà de 60 mg/L d'alanine marquées ajoutées, le taux d'incorporation n'augmente pas. Néanmoins, sans phase croissante sur la courbe, il est difficile de conclure que le plateau est atteint.

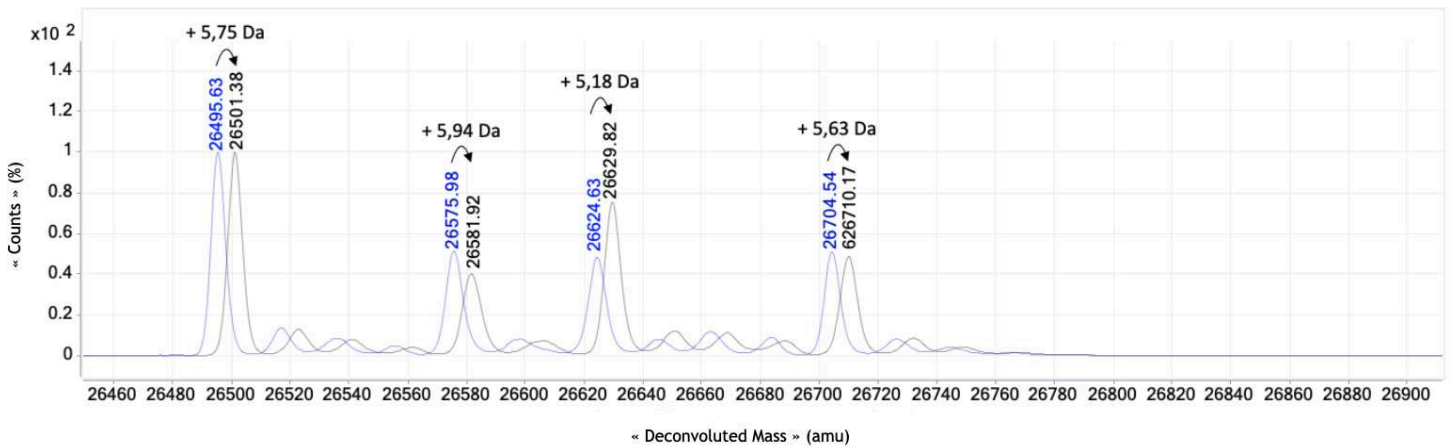
Pour tous les acides aminés, il s'est donc avéré nécessaire de valider les valeurs de plateau définies par RMN par des analyses plus quantitatives de spectrométrie de masse. Pour cela, le taux d'incorporation de chaque acide aminé est mesuré par ESI-TOF en comparant la masse de la protéine Hsp90 N-terminal non marquée avec la même protéine, mais marquée sur un seul type d'acide aminé (soit Met, soit Ala, etc). Ainsi, la différence de masse observée, entre la masse attendue pour 100 % d'incorporation et la valeur réelle mesurée par MS permet de calculer le taux d'incorporation pour chaque sonde. Pour cela, un échantillon non marqué et six échantillons marqués ont été produits avec 80 mg/L d'un des acides aminés suivants :  $^{13}\text{C}_3$ -Met,  $^{13}\text{C}_3$ -Ala + 5 mg/L de L-cyclo-sérine,  $U$ -[ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Leu,  $U$ -[ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Ile,  $U$ -[ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Val et  $U$ -[ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Thr et analysés par MS (plateforme MS de l'IBS, Dr Luca Signor) (Figure 34 et cf. Annexe 5 p. 240).

Les échantillons préparés sont différents de ceux produits pour les analyses RMN. En effet, les échantillons analysés par RMN, jusqu'ici, étaient enrichis avec des acides aminés marqués sur les méthyles et deutérés sur le reste du résidu (NMR-Bio). En revanche, ces acides aminés, du fait du métabolisme des cellules eucaryotes et du milieu de culture protoné, sont partiellement reprotés *in vivo*, entraînant une hétérogénéité et une incertitude sur la masse de la protéine. Il était donc nécessaire d'utiliser des acides aminés sans deutération.

A : Hsp90 N-terminal non marquée (bleu)



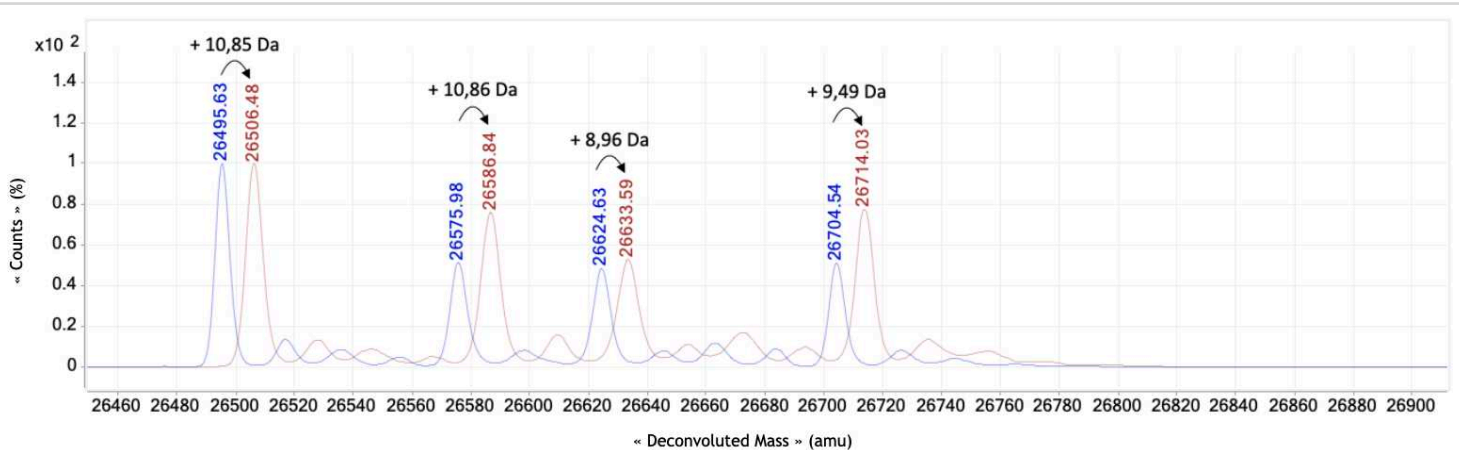
B. Hsp90 N-terminal marquée (noir) –  $^{13}\text{CH}_3\text{-Met}$  : 6 (pic 1 et 2) ou 7 (pic 3 et 4) méthionines dans la séquence, + 1 Da par  $^{13}\text{CH}_3\text{-Met}$  incorporée (+ Superposition de Hsp90 N-terminal non marquée - bleu).



Traitement du spectre :

	Différence de masse attendue pour 100 % d'incorporation	Pic 1	Pic 2	Pic 3	Pic 4
$^{13}\text{CH}_3\text{-Met}$ :					
Différence de masse observée (Da)	+ 7 Da	+ 5,75 Da	+ 5,94 Da	+ 5,18 Da	+ 5,63 Da
– Taux d'incorporation calculé (%)		95 %	99 %	74 %	80 %

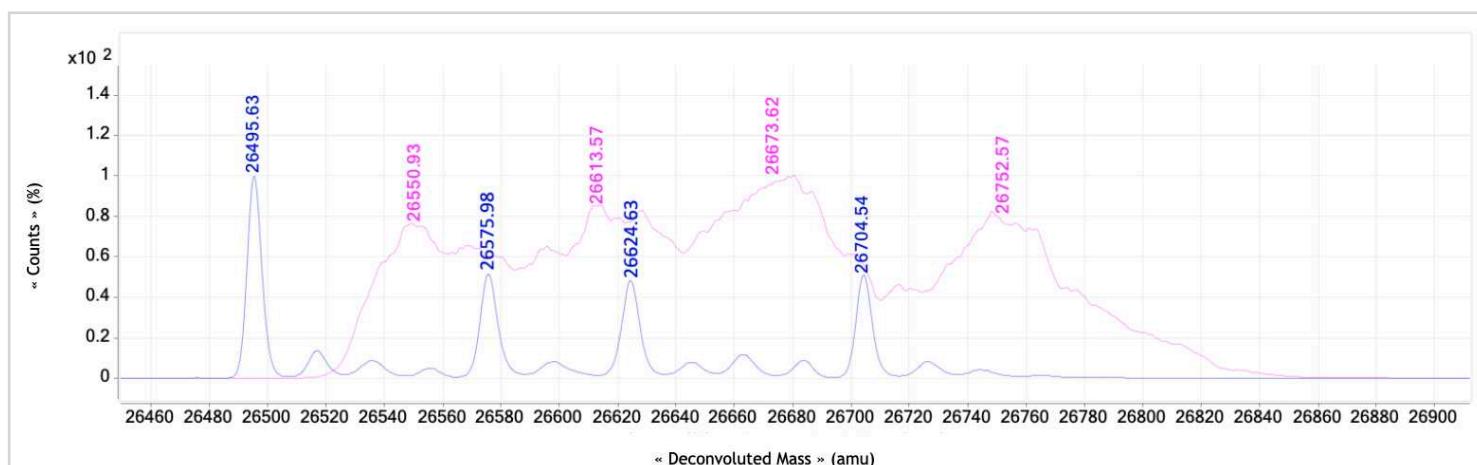
C. Hsp90 N-terminal marquée (marron) –  $^{13}\text{CH}_3\text{-Ala}$  : 15 alanines dans la séquence, + 1 Da par  $^{13}\text{CH}_3\text{-Ala}$  incorporée (+ Superposition de Hsp90 N-terminal non marquée - bleu).



Traitement du spectre :

	Différence de masse attendue pour 100 % d'incorporation	Pic 1	Pic 2	Pic 3	Pic 4
$^{13}\text{CH}_3\text{-Ala}$ :					
Différence de masse observée (Da)	+ 15 Da	+ 10,85 Da	+ 10,86 Da	+ 8,96 Da	+ 9,49 Da
– Taux d'incorporation calculé (%)		72 %	72 %	59 %	63 %

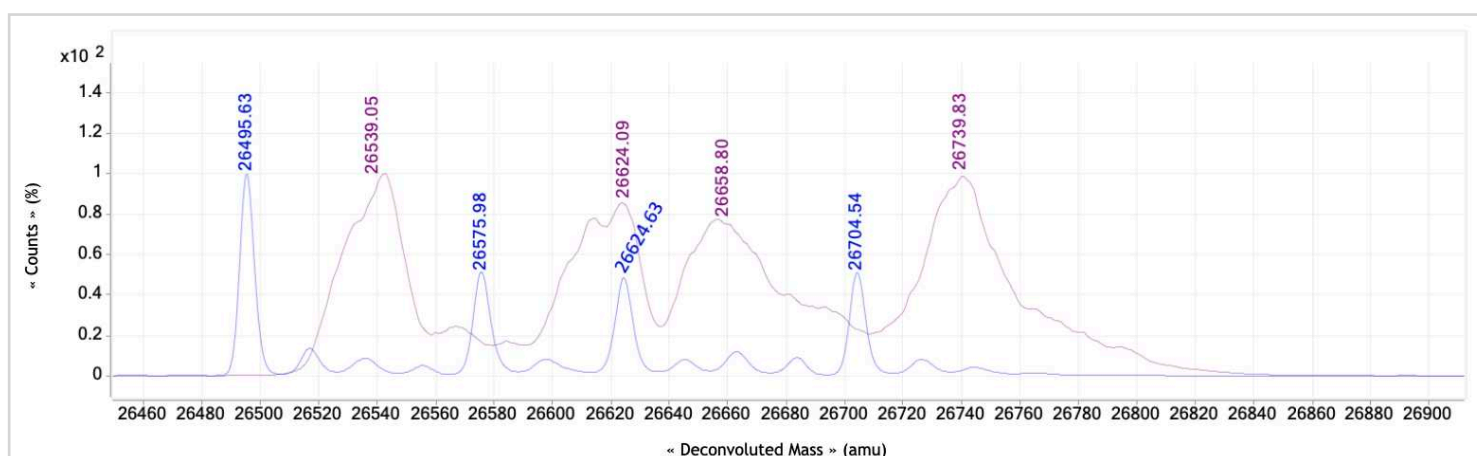
D. Hsp90 N-terminal marquée (rose) – [ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Leu : 18 leucines dans la séquence, + 7 Da par [ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Leu incorporée (+ Superposition de Hsp90 N-terminal non marquée - bleu).



Traitement du spectre :

	Différence de masse attendue pour 100 % d'incorporation	Pic 1	Pic 2	Pic 3	Pic 4
[ $^1\text{H}$ , $^{13}\text{C}$ , $^{15}\text{N}$ ]-Leu : Différence de masse observée (Da) – Taux d'incorporation calculé (%)	+ 126 Da	+ 55,3 Da – 43 %	+ 37,59 Da – 30 %	+ 48,99 Da – 38 %	+ 48,03 Da – 38 %

E. Hsp90 N-terminal marquée (violet) – [ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Val : 11 valines dans la séquence, + 6 Da par [ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Val incorporée (+ Superposition de Hsp90 N-terminal non marquée - bleu).



Traitement du spectre :

	Différence de masse attendue pour 100 % d'incorporation	Pic 1	Pic 2	Pic 3	Pic 4
[ $^1\text{H}$ , $^{13}\text{C}$ , $^{15}\text{N}$ ]-Val : Différence de masse observée (Da) – Taux d'incorporation calculé (%)	+ 66 Da	+ 43,42 Da – 65 %	+ 48,11 Da – 72 %	+ 34,17 Da – 51 %	+ 35,29 Da – 53 %

F.

Conditions	Taux d'incorporation (%)	Ecart-type	Justesse (spectromètre)
<sup>13</sup> CH <sub>3</sub> -Met	87	± 12	1 Da = ± 15%
<sup>13</sup> CH <sub>3</sub> -Ala	67	± 7	1 Da = ± 7 %
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Ile	50	± 12	1 Da = ± 1 %
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Val	60	± 10	1 Da = ± 1,5 %
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Thr	77	± 6	1 Da = ± 1,2 %
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Leu	37	± 5	1 Da = ± 1,8 %

Figure 34 : (A) à (E) Spectres ESI-TOF de la protéine Hsp90 N-terminal non marquée (bleu) et marquée (autres couleurs) sur un acide aminé portant un groupement méthyle et tableaux du traitement des spectres avec le résultat du taux d'incorporation pour chaque pic. (F) Taux d'incorporation moyen pour chaque acide aminé portant un groupement méthyle.

La méthionine et l'alanine ont été utilisées marquées sur le carbone des méthyles uniquement (entraînant un delta de masse de + 1 Da pour chaque résidu incorporé). Les autres résidus étaient, quant à eux, marqués sur tous les carbones et les azotes (avec un delta de masse compris entre + 5 à 7 Da par résidu, en fonction du nombre de carbones composant le résidu). De plus, étant donné que la valeur obtenue pour atteindre le plateau pour les acides aminés essentiels a toujours été de 80 mg/L, la méthionine et la leucine ont aussi été supplémentées à 80 mg/L dans le milieu de culture.

L'analyse de la protéine Hsp90 N-terminal non marquée par MS révèle quatre espèces protéiques (Figure 34A) : la protéine sans méthionine N-terminale et portant une PTM, une acétylation (+ 42 Da) (pic 1), la protéine sans méthionine N-terminale et portant deux PTMs, une acétylation et une phosphorylation (+ 80 Da) (pic 2), la protéine avec la méthionine N-terminale et une acétylation (pic 3) puis la protéine avec la méthionine N-terminale, une acétylation et une phosphorylation (pic 4).

Pour le calcul du taux d'incorporation de chaque acide aminé, il est nécessaire d'identifier le nombre de résidus et la masse ajoutée par chaque résidu marqué incorporé à la protéine (cf. Annexe 6A et 6B p. 241). Par exemple, chaque  $^{13}\text{CH}_3$ -alanine incorporée apporte un accroissement de la masse de 1 Da par rapport à la masse de la protéine non marquée, dans un maximum théorique de 15 Da si les alanines sont marquées à 100 % (15 alanines dans la séquence).

i. Les protéines Hsp90 N-terminal marquées  $^{13}\text{CH}_3$ -Met/-Ala présentent également quatre espèces protéiques (Figure 34B et 34C). Par exemple, la différence de masse observée entre le pic 1 de la figure 34A (26495,46 Da, qui correspond à la protéine non marquée avec une acétylation et sans la méthionine N-terminale) et le pic 1 de la figure 34B (26501,38 Da, même espèce protéique mais marquée sur les méthionines) est de + 5,75 Da. Or, la différence de masse attendue pour 100 % d'incorporation est de 6 Da (7 méthionines moins la méthionine N-terminale), ce qui permet de calculer un taux d'incorporation de 95 %. Le même calcul a été réalisé pour les différents échantillons et a permis d'obtenir le tableau de la figure 34F.

Mais ce tableau met en évidence la limite de la technique. En effet, pour un même marquage, les quatre espèces de Hsp90 N-terminal ne présentent pas des taux d'incorporation identiques, ce qui n'est expérimentalement pas concevable. Ces différences viennent du fait que Hsp90 N-terminal possède peu de méthionines et que le delta de masse observé est très faible, générant donc un écart-type important auquel s'ajoute la justesse de l'appareil qui augmente encore l'incertitude des taux d'incorporation calculés. Notons que ces facteurs défavorables s'appliquent à l'ensemble des sondes analysées.

Toutefois, en analysant le spectre RMN 2D de la protéine marquée avec 200 mg/L  $^{13}\text{CH}_3$ -Met et 80 mg/L de  $^{13}\text{CH}_3$ -Ala (cf. Figure 33 p. 179), la différence d'intensité des pics (entre méthionines et alanines) semblerait corroborer la tendance des résultats obtenus par MS, c'est-à-dire un écart d'environ 20 % dans le taux d'incorporation des  $^{13}\text{CH}_3$ -Met et  $^{13}\text{CH}_3$ -Ala.

ii. Pour la valine et la leucine (Figure 34D et 34E) (mais aussi la thréonine et l'isoleucine, spectres ESI-TOF présentés en annexes 5 p. 240), les spectres ESI-TOF ne présentent pas de pics résolus mais des massifs isotopiques. Ces massifs sont la conséquence du marquage  $^{13}\text{C}$  et  $^{15}\text{N}$  des résidus. En effet, le marquage générant un accroissement de la masse de la protéine de 6 Da (cas de la [ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Val) pour chaque acide aminé marqué incorporé, une largeur de pic plus importante est nécessairement attendue, en comparaison de l'augmentation de la largeur de pic lors de l'incorporation des méthionines et alanines (seulement +1 Da). Cependant, ce phénomène est aggravé par une hétérogénéité dans le nombre de résidus incorporés. Ainsi, dans le cas de l'isoleucine (et de façon moindre pour la valine), de nombreuses espèces protéiques de masses différentes sont présentes et détectées (car incorporant plus ou moins de résidus) en plus des espèces protéiques ayant ou non la méthionine N-terminale et des PTMs. La largeur du massif induit donc une incertitude sur le sommet du pic, empêchant ainsi une quantification précise du taux d'incorporation pour les résidus [ $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-Ile/ -Val/ -Thr/ -Leu.

Bien que la présence des massifs isotopiques limite la confiance du traitement des spectres, des taux d'incorporation peuvent tout de même être estimés par ces acides aminés (Figure 34F). Néanmoins, en analysant le spectre RMN 2D de la protéine marquée avec 200 mg/L  $^{13}\text{CH}_3$ -Met et 80 mg/L de  $^{13}\text{CH}_3$ -Ile $^{\delta 1}$ /-Val $^{\text{proS}}$ /-Thr (Figure 35), l'intensité des pics de chaque acide aminé semble être proche et ne corroborerait donc pas les taux d'incorporation très variables observés par MS.

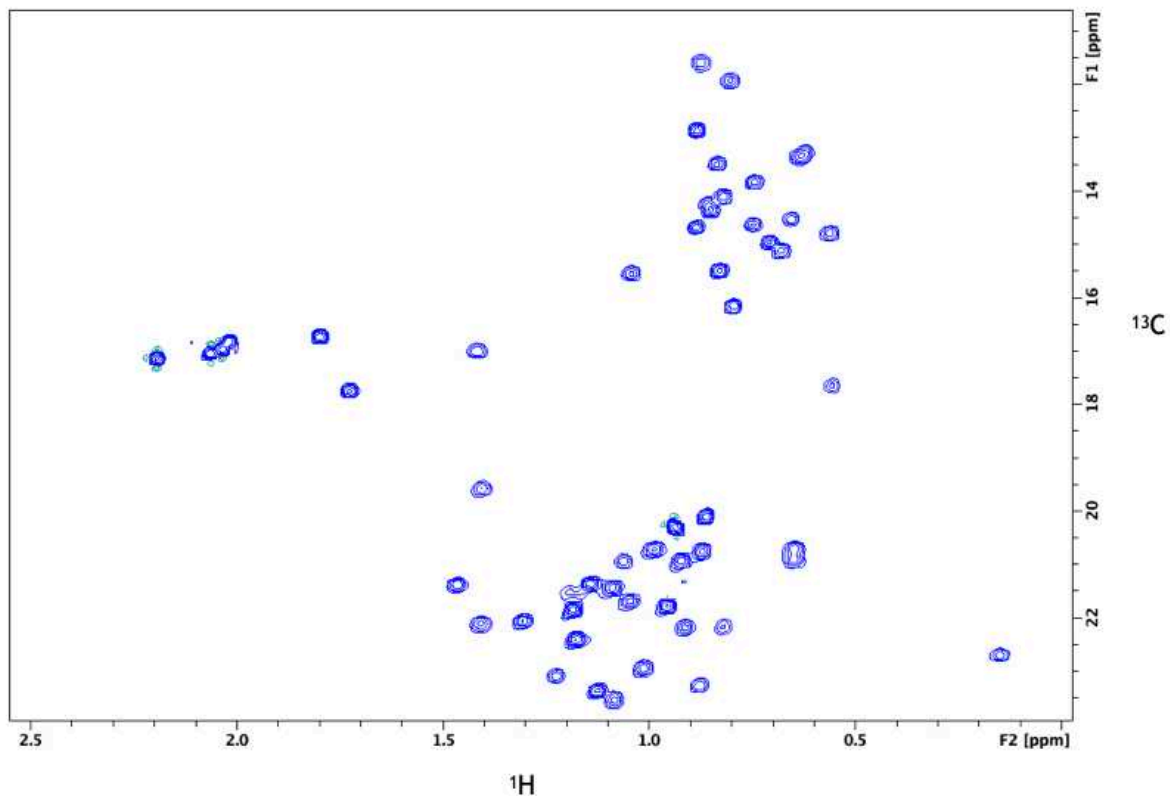


Figure 35 : Spectre RMN 2D de la protéine Hsp90 N-terminal. Protéine marquée avec 200 mg/L de  $^{13}\text{CH}_3\text{-Met}$  et 80 mg/L de  $^{13}\text{CH}_3\text{-Thr}$ ,  $^{13}\text{CH}_3\text{-Val}^{\text{proS}}$  et  $^{13}\text{CH}_3\text{-Ile}^{\delta 1}$ . Echantillon concentré à 100  $\mu\text{M}$ , acquisition à 25°C, tampon : HEPES 20 mM pH 7,5, NaCl 150 mM, TCEP 1 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une cryosonde.

Par conséquent, pour préciser ces résultats, nous avons entrepris de compléter les analyses par des expériences de FT-ICR sur la protéine intacte et digérée à la trypsine (TGE FT-ICR, collaboration Héloïse Dossmann, Sorbonne Université). Les 14 échantillons (7 échantillons de Hsp90 N-terminal marquées, ou non, et/ou non digérées) ont été analysés et les données sont actuellement en cours de traitement. Cette méthode, de par sa grande sensibilité, pourrait permettre de déconvoluer les spectres (résolution des massifs isotopiques) et d'obtenir des données plus précises sur les taux d'incorporation.

Néanmoins, malgré l'incertitude sur le taux exact d'incorporation des  $^{13}\text{CH}_3\text{-acides aminés}$ , un milieu de marquage spécifique a pu être développé. Sa composition n'engendre pas de fuites isotopiques et une dilution isotopique au niveau des alanines très limitée. Ce milieu permet également une expression protéique relativement proche du milieu de culture standard (dans les mêmes conditions d'expression), voire des rendements protéiques identiques en adaptant le protocole d'expression (24 heures d'expression supplémentaire avant l'arrêt de la culture). Il permet également

l'acquisition de spectres RMN résolus et exploitables dont la qualité a été validée en comparant ces derniers aux spectres acquis sur des protéines exprimées en bactérie. En conséquence, ce milieu de marquage peut être appliqué aux protéines exprimées en cellules HEK 293, et plus particulièrement à la protéine HSulf-2.

#### VI.2.4. Développement du milieu de culture uniformément marqué $^{15}\text{N}$

Parallèlement au développement d'un milieu de marquage spécifique, une stratégie de marquage uniforme a aussi été mise au point. Ces milieux de culture sont disponibles commercialement (BioExpress 6000, CIL) à des prix de 7000 à 15 000 euros le litre. Théoriquement, le mélange réactionnel d'acides aminés peut être préparé à partir de chacun des 20 acides aminés marqués  $^{15}\text{N}$ , mais cela entraîne un coût très conséquent. Ici, l'alternative choisie a été d'ajouter, au milieu ESF-SFM déplété, de l'extrait de levure  $^{15}\text{N}$  et du  $^{15}\text{NH}_4\text{Cl}$  afin de marquer uniformément la protéine Hsp90 N-terminal (Opitz *et al.*, 2015). Avec cette méthode, la protéine Hsp90 N-terminal obtenue est pure et présente des rendements équivalents à ceux observés précédemment, soit environ 2,7 mg/L.

Afin de valider la qualité du milieu de marquage et la stratégie développée, le spectre 2D de la protéine marquée  $^{15}\text{N}$  a été acquis (Figure 36A). Ce spectre est bien résolu et l'intensité des pics démontre l'efficacité de l'extrait de levure comme source d'acides aminés marqués. Un spectre similaire, acquis sur la protéine Hsp90 N-terminal uniformément marquée  $^{15}\text{N}$  et exprimée en milieu deutéré (bactérie) (collaboration Dr Boisbouvier et F. Hénot, groupe RMN à l'IBS), a été acquis dans les mêmes conditions (Figure 36B).

Une très large superposition des pics (quelques déplacements chimiques dus à la deutération partielle de la protéine exprimée en bactéries, et aux PTMs) est observée, ce qui montre que la protéine exprimée dans le milieu de culture développé est correctement structurée et marquée, et laisse la possibilité d'appliquer ce milieu aux petites protéines exprimées en cellules de mammifères.



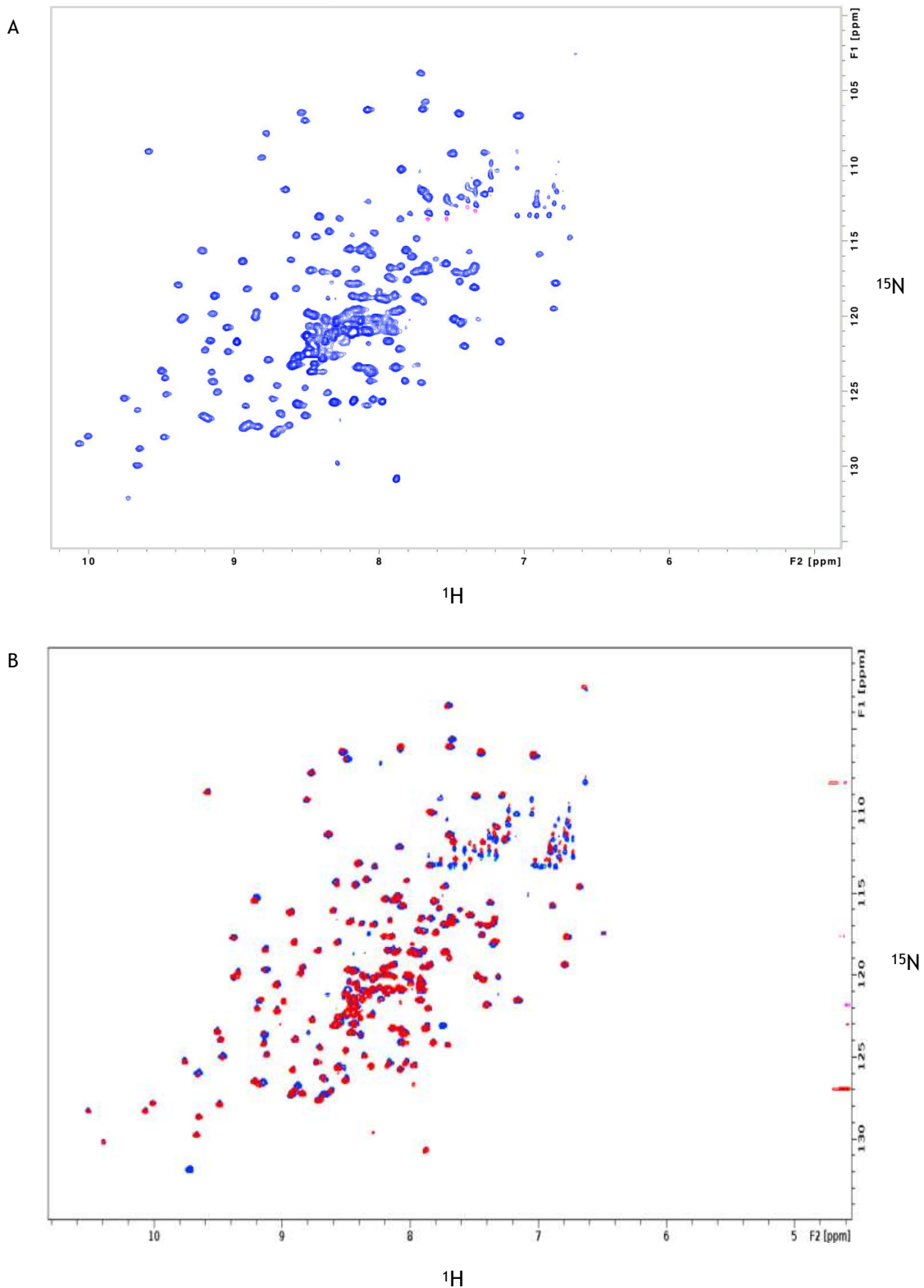


Figure 36 : Marquage uniforme  $^{15}\text{N}$  de la protéine Hsp90 N-terminal. (A) Spectre de la protéine Hsp90 N-terminal marquée  $U\text{-}[^1\text{H}, ^{15}\text{N}]$  et produite en cellules HEK 293F (échantillon concentré à  $113\ \mu\text{M}$ , acquisition à  $25^\circ\text{C}$ , tampon : HEPES 20 mM pH 7,5, NaCl 150 mM, TCEP 1 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une cryosonde). (B) Superposition de deux spectres de la protéine Hsp90 N-terminal marquée  $U\text{-}[^1\text{H}, ^{15}\text{N}]$  dans les deux cas. Le spectre rouge correspond à la protéine exprimée en eucaryote et le spectre bleu à la protéine exprimée en bactérie.

*Production de la protéine Hsp90 N-terminal exprimée en bactérie et acquisition des spectres réalisées par F. Hénot (IBS, groupe RMN).*

Cependant, et pour rappel, la protéine HSulf-2 étant une protéine de haute masse moléculaire, utiliser un marquage uniforme générerait donc des spectres non exploitables (nombreux chevauchements de signaux, faible résolution du spectre, entre autres), contrairement au marquage spécifique des  $^{13}\text{CH}_3$ -acides aminés qui permet l'acquisition de spectres RMN résolus. Toutefois, dans un objectif général de développer des stratégies de marquage isotopique appliquées aux protéines exprimées en cellules de mammifères, la mise au point d'un milieu de marquage uniforme reste indispensable.

#### **VI.2.5. Application du milieu de marquage spécifique à la protéine HSulf-2 : marquage de HSulf-2 $\Delta\text{HD}$ et HSulf-2 $\Delta\text{HDc}$**

Parallèlement à la mise au point des milieux de marquage pour la RMN, il a été réalisé dès les premiers mois de cette thèse un marquage de la protéine HSulf-2  $\Delta\text{HD}$  par ajout, dans un milieu commercial non dépleted en acides aminés (FreeStyle 293), de  $^{13}\text{CH}_3$ -Met en excès. La protéine a été produite en cellules HEK 293F et marquée en ajoutant 0,5 g/L de  $^{13}\text{CH}_3$ -Met au milieu de culture FreeStyle 293. Après purification (cf. Figure 22 p. 162), la protéine a été concentrée à 6,5 mg/mL (soit ~ 100  $\mu\text{M}$ ) et un spectre RMN 2D a été acquis (SOFAS-HMQC) (Figure 37). Le spectre obtenu présente un seul « pic » non résolu dans la région des  $^{13}\text{CH}_3$ -Met, alors que 18 pics résolus et dispersés (à l'instar des spectres de la protéine Hsp90 N-terminal) étaient attendus. Ainsi, ce spectre est typique d'une protéine non structurée, car ce « pic » non résolu signifie que le spectromètre enregistre le même environnement chimique autour de chaque résidu  $^{13}\text{CH}_3$ -Met.

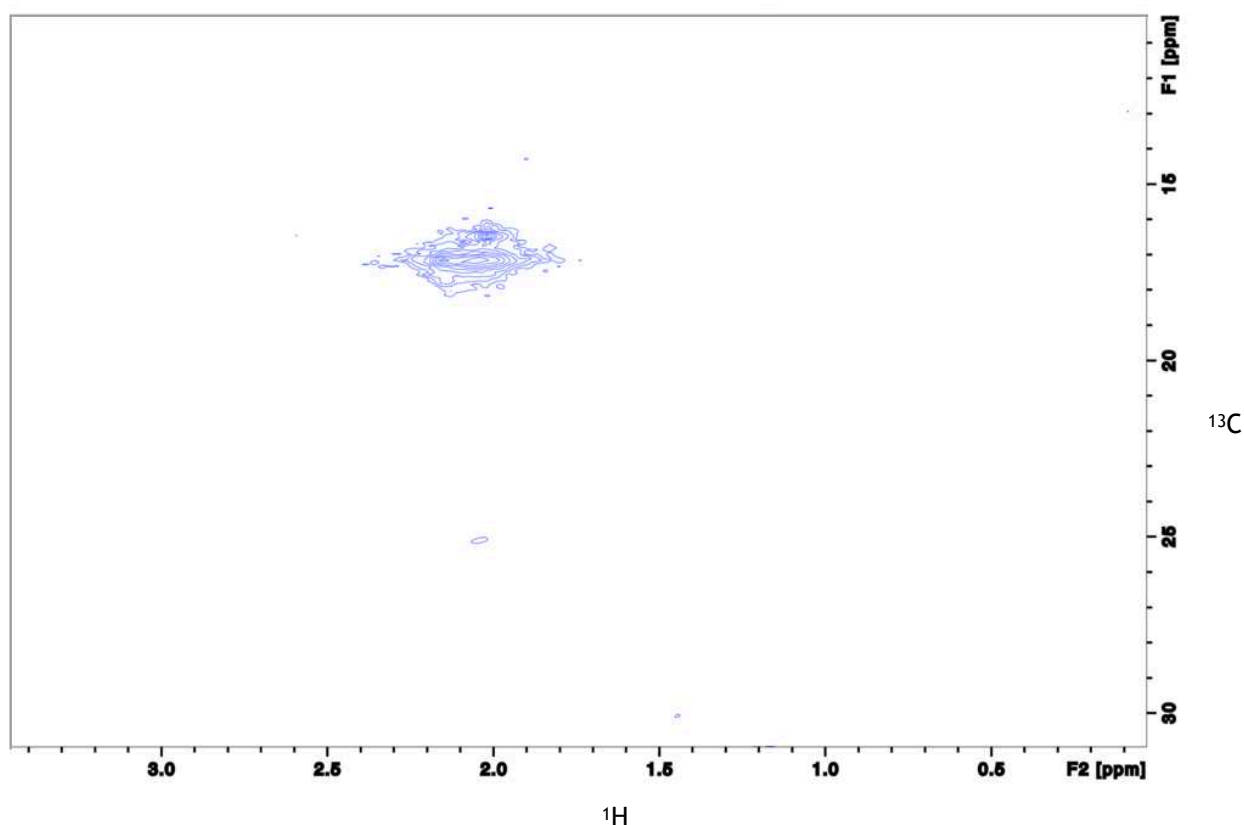


Figure 37 : Spectre 2D de la protéine HSulf-2  $\Delta$ HD marquée  $^{13}\text{CH}_3$ -Met et exprimée en cellules HEK 293F (échantillon concentré à 90  $\mu\text{M}$ , acquisition à 25°C, tampon : Tris 50 mM pH 8, NaCl 200 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une cryosonde).

Cependant, il a été démontré que la protéine HSulf-2  $\Delta$ HD était structurée et fonctionnelle. En effet, l'activité arylsulfatase de cet échantillon a été confirmée (de même que pour l'ensemble des échantillons produits et marqués durant ces travaux de thèse) (données non présentées). En revanche, après dénaturation de la protéine (5 minutes à 100°C) l'activité est perdue, confirmant de ce fait la structuration de la protéine HSulf-2  $\Delta$ HD marquée et analysée par RMN. Ainsi, pour expliquer ce résultat, une première hypothèse a été émise : la coupure de l'étiquette SNAP pourrait déstabiliser la protéine car cette étiquette représente près de 30 % de la masse de la protéine. De plus, l'étape de coupure à 4°C sur la nuit, ainsi que les douze heures d'attente entre la fin de la purification et l'acquisition du spectre, pourraient affecter défavorablement la qualité de l'échantillon. Pour le vérifier, la protéine a de nouveau été produite (en cellules HEK 293F) et marquée, puis purifiée sans le clivage des étiquettes et analysée au spectromètre le même jour. Cependant, le même type de spectre a été obtenu (résultats non présentés).

Une deuxième hypothèse pourrait être un impact majeur des N-glycosylations. Il est vrai que la protéine HSulf-2  $\Delta$ HD exprimée en cellules HEK 293F porte ~ 18 kDa de ces PTMs. Or, cela engendre une augmentation du rayon hydrodynamique de la protéine qui peut être délétère pour le temps de corrélation ( $\tau_c$ , temps moyen nécessaire pour qu'un carbone effectue une rotation de 1 radian). En effet, le  $\tau_c$  dépend de la masse des molécules, mais aussi de leur sphéricité (Canet *et al.*, 2010). Ainsi, avec un  $\tau_c$  plus grand, les variations d'orientations se font à des fréquences moyennes trop éloignées de la fréquence de résonance typique/d'observation du  $^{13}\text{C}$ . Ceci signifie qu'un temps de corrélation augmenté entraîne un élargissement du signal RMN (élargissement des raies/pics) et peut également être à l'origine de sa disparition complète. Pour vérifier cela, la protéine HSulf-2  $\Delta$ HD a été exprimée en cellules HEK 293S (cellules engendrant un motif glycosidique réduit) et marquée  $^{13}\text{CH}_3$ -Met, puis purifiée avec le clivage des étiquettes avant son analyse par RMN (SOFAS-HMQC). Toutefois, dans ce cas aussi, un spectre identique à celui présenté dans la figure 37 a été obtenu (résultats non présentés).

Une troisième hypothèse, pourrait être que la protéine HSulf-2  $\Delta$ HD est de trop grande masse moléculaire, d'autant que la protéine n'est pas deutérée. En effet, en l'absence d'une deutération de la protéine, il n'y a pas de réduction des interactions dipolaires, de la relaxation transversale et des chevauchements de signaux. Or, ceci peut conduire, à l'instar de l'hypothèse précédente, à l'élargissement des pics et/ou à seulement détecter les petites structures flexibles (comme les boucles). Pour vérifier cette hypothèse, la protéine SNAP-HSulf-2  $\Delta$ HD marquée  $^{13}\text{CH}_3$ -Met (et exprimée en cellules HEK 293S) a été soumise à une protéolyse limitée par digestion à la trypsine, afin d'essayer d'éliminer les parties flexibles de la protéine et d'isoler une forme protéique structurée (par exemple le cœur hydrophobe).

Une fois la digestion et la chromatographie d'exclusion réalisées (données non présentées), le chromatogramme montre un pic d'élution majoritaire (à 14,7 mL) et une fois les fractions correspondant à ce pic analysées par SDS-PAGE, quatre bandes sont visibles, indiquant la co-élution de quatre fragments de dégradation de la protéine. Afin de déterminer la masse exacte de ces espèces, l'échantillon a été analysé par MS (plateforme MS de l'IBS, Dr. Luca Signor). Cette analyse a confirmé la présence de quatre peptides de masse moléculaire : 50 kDa, 30 kDa, 18 kDa et 7 kDa. Ceci étant, à cause du marquage de la protéine, ainsi que de la présence de PTMs sur la protéine, ces peptides n'ont pas pu être attribués à un domaine/une séquence précise de la protéine

HSulf-2  $\Delta$ HD. Néanmoins, le peptide de 18 kDa pourrait être le domaine C-terminal. En effet, si une boucle entre le domaine CAT et C-terminal est clivée alors, en conditions dénaturantes, une bande à cette masse moléculaire devrait apparaître. En revanche, les autres fragments sont difficilement attribuables.

Cet échantillon, digéré à la trypsine, a en parallèle été analysé par RMN (Figure 38), et un spectre difficilement analysable a été obtenu. En effet, un pic peu résolu est une nouvelle fois observé (2,1 ppm  $^1\text{H}$ , 16 ppm  $^{13}\text{C}$ ), mais de nombreux autres pics apparaissent. Etant donné que seules les méthionines sont marquées et qu'il ne peut pas y avoir de fuites isotopiques, ces pics correspondent à des méthionines mais résonant dans une région étendue du spectre de résonance (0,8 à 1,5 ppm,  $^1\text{H}$ ) (cf. Figure 11 p. 93).

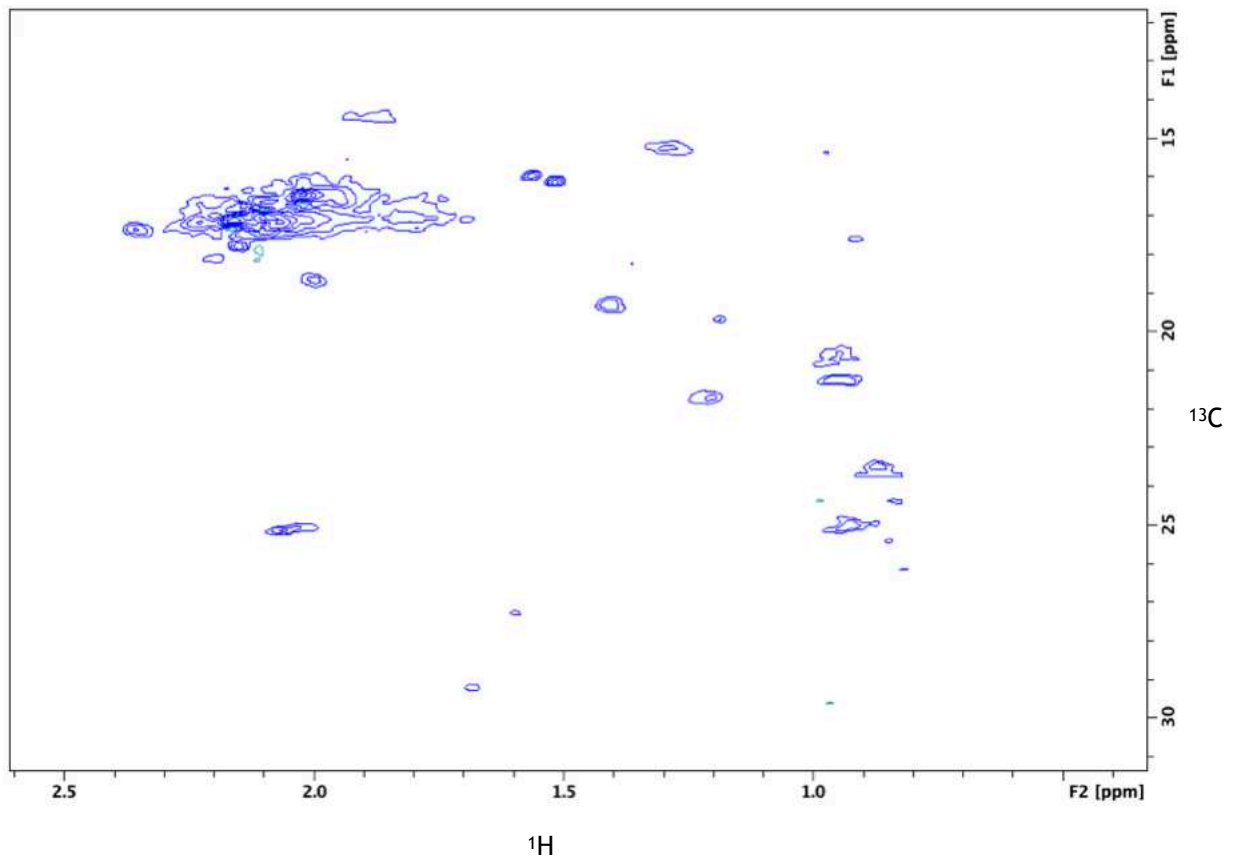


Figure 38 : Spectre 2D de la protéine SNAP-HSulf-2  $\Delta$ HD trypsinée et marquée  $^{13}\text{CH}_3\text{-Met}$  en excès (échantillon concentré à 30  $\mu\text{M}$ , acquisition à 25°C, tampon : Tris 50 mM pH 8, NaCl 200 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une cryosonde).

Ces pics pourraient correspondre à une dégradation de la protéine et/ou à des protéines contaminantes. En effet, des pics similaires (mais moins intenses) sont visibles sur le spectre RMN 2D de la figure 40B (cf. p. 197), confortant donc l'hypothèse d'une détection de petits peptides appartenant (ou non) à la protéine HSulf-2  $\Delta$ HD. Néanmoins, il peut être conclu avec confiance que la digestion de la protéine à la trypsine ne permet pas d'accroître la résolution du spectre acquis, bien que selon toute vraisemblance, des régions flexibles exposées au solvant aient été clivées.

Une quatrième hypothèse pourrait être un choix de sonde isotopique non approprié. En effet, une analyse de la structure de HSulf-2 WT prédite par Phyre2 a été réalisée. Ce logiciel prédit la structure tridimensionnelle d'une séquence protéique en utilisant les principes et les techniques de modélisation par homologie, c'est-à-dire, que la structure de la protéine cible est obtenue à partir de sa séquence en acides aminés et d'une (ou plusieurs) structure tridimensionnelle haute résolution d'une protéine homologue (<http://www.sbg.bio.ic.ac.uk>, Kelley *et al.*, 2015). Or, d'après cette prédiction, une majorité des résidus méthionines sont présents dans des régions non structurées ou flexibles (boucles). Ceci pourrait donc expliquer le même environnement chimique pour l'ensemble des isotopes  $^{13}\text{CH}_3\text{-Met}$  et les spectres obtenus par RMN, sans pour autant que la protéine ne soit pas structurée. Or, d'après ce même modèle, les résidus Thr et Ile sont davantage présents dans des régions structurées (60 et 72 %, respectivement). Par conséquent, un marquage combinatoire  $^{13}\text{CH}_3\text{-Ile}^{\delta 1}\text{-Met/-Thr}$  de HSulf-2  $\Delta$ HD (exprimée en cellules HEK 293S) a été réalisé. Ce marquage a été obtenu avec le milieu de marquage spécifique développé sur Hsp90 N-terminal.

Une fois que les fractions contenant HSulf-2  $\Delta$ HD pure et marquée ont été rassemblées et concentrées, cet échantillon a été analysé par RMN. Toutefois, un résultat similaire a été obtenu avec trois « pics » non résolus apparaissant dans les régions de résonance respective de chaque sonde (résultats non présentés). Ainsi, bien que les résidus Ile et Thr étaient, d'après les données de prédiction, plutôt situés dans des régions structurées, ceci n'aura pas permis d'accroître la résolution des spectres.

Ces résultats, difficilement explicables, pourraient indiquer que la construction CAT n'est pas suffisamment structurée ou induirait un mauvais repliement de la protéine HSulf-2  $\Delta$ HD (bien que des données antérieures à cette thèse, obtenues par dichroïsme circulaire, montrent une structuration supérieure à 60 % pour cette protéine). Ainsi,

avec l'expertise du Dr. Christine Gaboriaud, deux nouvelles constructions légèrement modifiées ont été produites. Ces constructions, nommées HSulf-2  $\Delta$ HD V1 et HSulf-2  $\Delta$ HD V2, sont des dérivées de la construction HSulf-2  $\Delta$ HD, où quelques dizaines d'acides aminés ont été éliminés sur les extrémités N et C-terminales des domaines CAT et C-terminal (HSulf-2  $\Delta$ HD V1), et sans les sites de coupure à la TEV (HSulf-2  $\Delta$ HD V2) (cf. Figure 15E et 15F p. 124 et Annexe 1D et 1E p. 233 et 235). Cependant, en plus de rendements protéiques nettement réduits ( $\sim 2$  mg/L) pour ces deux protéines, une fois ces dernières marquées  $^{13}\text{CH}_3$ -Met (en excès), les spectres RMN 2D obtenus ont encore une fois présenté les mêmes caractéristiques (résultats non présentés).

Suite à ces résultats, une deuxième construction, HSulf-2  $\Delta$ HDc (cf. Rappel Figure 15D, p. 157 et Annexe 1F p. 235), a été étudiée. Comparée à HSulf-2  $\Delta$ HD, cette protéine de 82,5 kDa comprend l'ensemble des ponts disulfures et régions structurées connues, mais ne comprend pas la région centrale du domaine HD qui est non structurée (17 kDa). Un premier marquage  $^{13}\text{CH}_3$ -Met (en excès en milieu FreeStyle 293) a été réalisé afin de vérifier la qualité du spectre RMN (Figure 39A). Notons que l'expression de la protéine a été réalisée en cellules HEK 293F, car les cellules HEK 293S ne permettent pas un rendement protéique suffisant (inférieur à  $\sim 700$   $\mu\text{g/L}$ ). Ainsi, une fois la protéine exprimée et purifiée, l'échantillon a été analysé par RMN (expérience SOFAST-HMQC), mais une fois encore le spectre obtenu est non exploitable. Toutefois, en comparant ce spectre avec les précédents, il semblerait que certains pics plus ou moins résolus apparaissent (il pourrait être identifié  $\sim 7$  pics sur 26 attendus), mais une fois de plus cette protéine n'est pas adaptée pour des analyses RMN.

Dans une tentative de diminuer la flexibilité de la protéine (diminution du rayon hydrodynamique, du  $\tau_c$ , et possiblement d'une dynamique non favorable), pour ne pas focaliser l'acquisition sur des régions types boucles ou avoir une perte/dissimulation du signal RMN des régions structurées, 1 mM de Sulfocalix-[4]-arène (v/v) a été ajouté à l'échantillon protéique. En observant les deux spectres obtenus en l'absence et en présence de Sulfocalix-[4]-arène (Figure 39B), un déplacement chimique de certains pics est visible, ce qui signifie que la protéine interagit bien avec le sulfocalix-[4]-arène. En revanche, aucun pic supplémentaire correspondant à des méthionines n'est apparu, ce qui suggère que la protéine n'a pas moins de propriétés défavorables pour l'acquisition de spectres résolus.

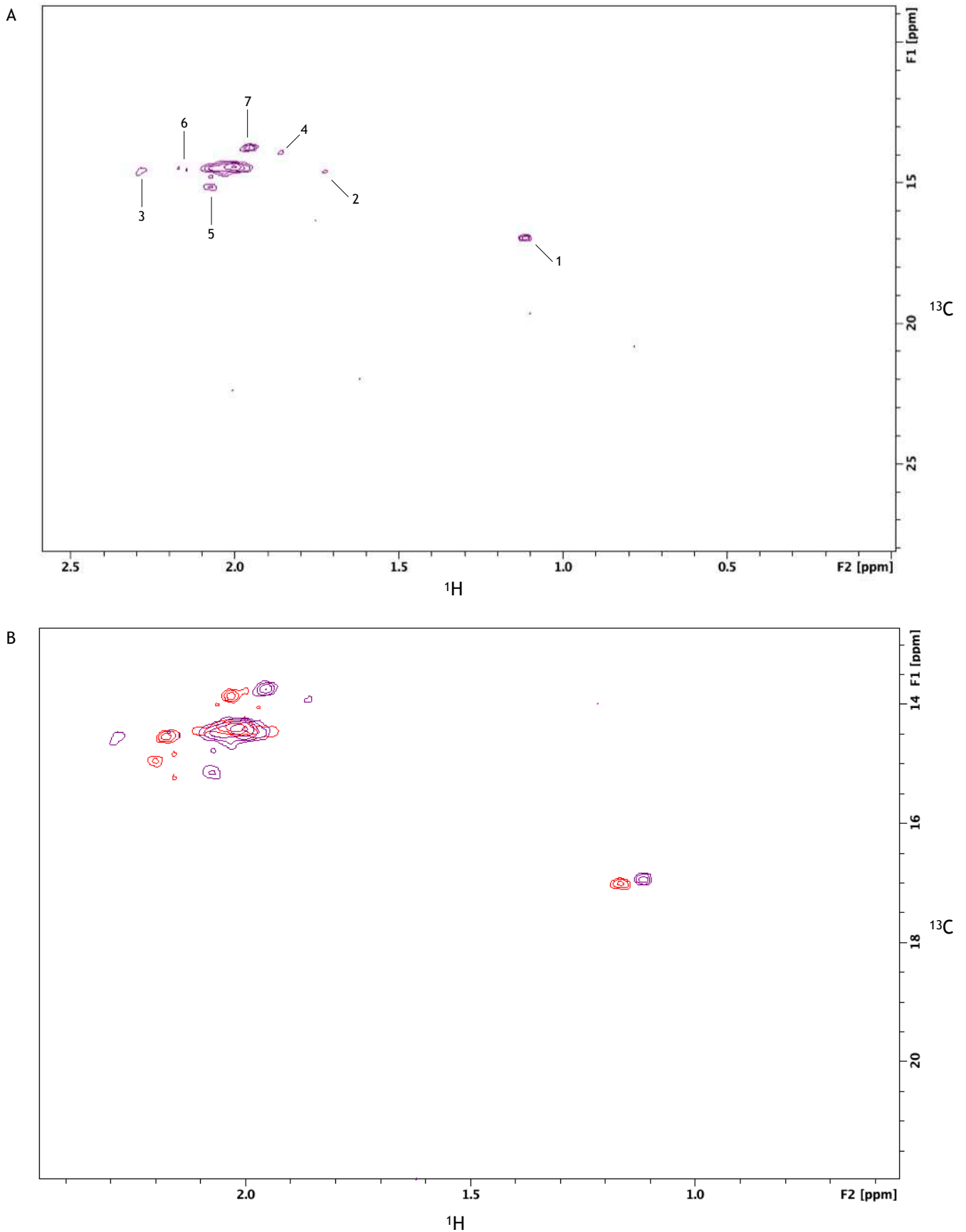
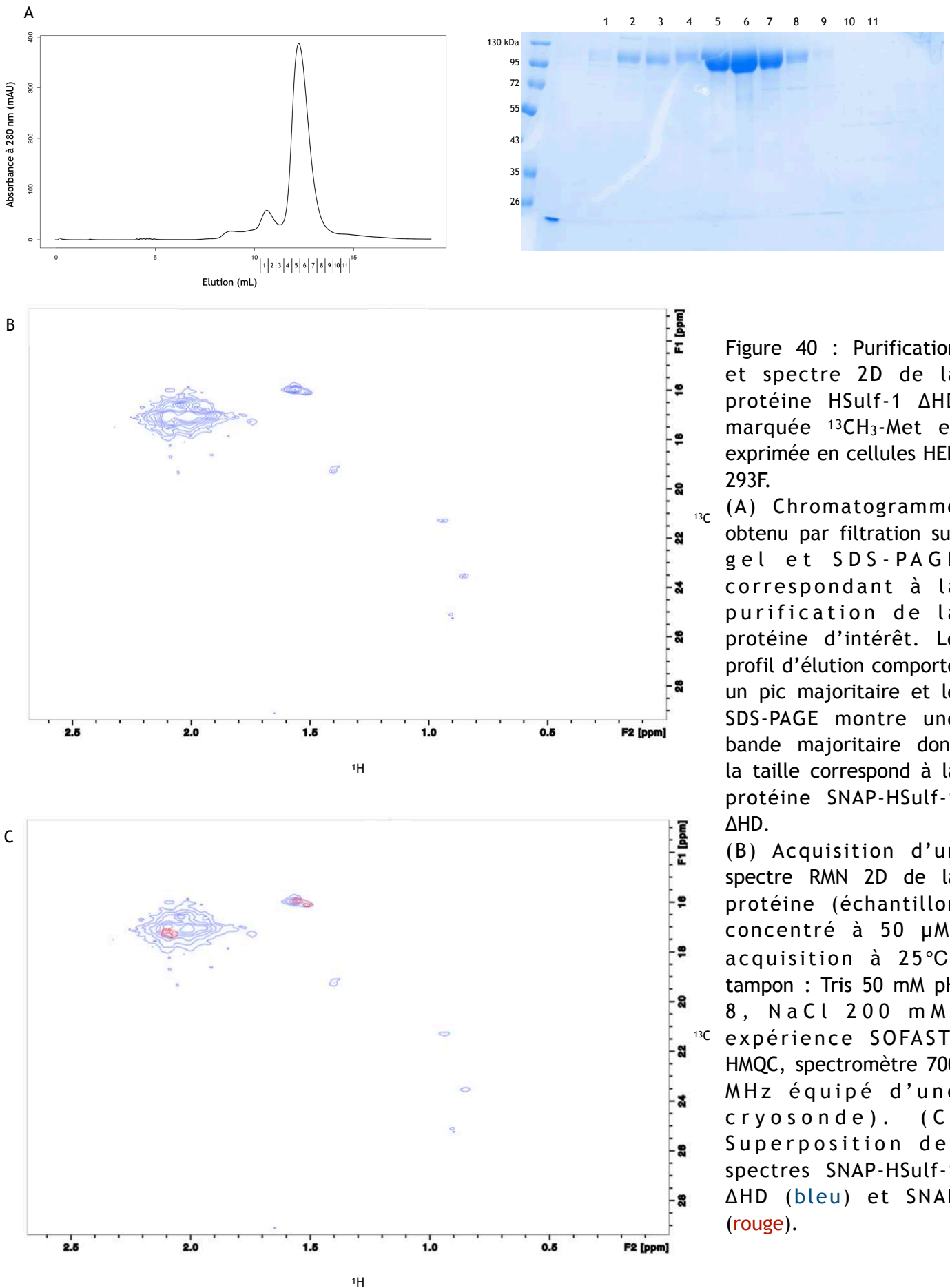


Figure 39 : (A) Spectre 2D de la protéine SNAP-HSulf-2  $\Delta\text{Hdc}$  marquée  $^{13}\text{CH}_3\text{-Met}$  en excès. Les pics 1 à 7 correspondraient aux méthionines de la protéine cible. (B) Superposition des spectres 2D de la protéine SNAP-HSulf-2  $\Delta\text{Hdc}$  marquée  $^{13}\text{CH}_3\text{-Met}$  sans (violet) et avec (rouge) sulfocalix-[4]-arène. Echantillon concentré à  $20\ \mu\text{M}$ , acquisition à  $25^\circ\text{C}$ , tampon : Tris 50 mM pH 8, NaCl 200 mM, expérience SOFAST-HMQC, spectromètre 700 MHz équipé d'une crysonde.



La deuxième enzyme de HSulf (HSulf-1) a également été étudiée, en émettant l'hypothèse que HSulf-1 pourrait présenter un spectre RMN de meilleure qualité que HSulf-2. En effet, bien que des différences existent entre ces deux protéines, les domaines CAT sont bien conservés et présentent la même activité arylsulfatase. Pour cela, HSulf-1  $\Delta$ HD (déjà étudiée au sein du groupe) a été produite en cellules HEK 293F, avec un marquage sur les méthionines (en excès), puis purifiée (sans clivage TEV) (Figure 40A). Le chromatogramme obtenu après filtration sur gel montre un seul pic dont l'analyse par SDS-PAGE montre une bande majoritaire entre les marqueurs de masse moléculaire 95 kDa et 130 kDa, ce qui correspond à la protéine SNAP-HSulf-1  $\Delta$ HD. Une fois la protéine concentrée, celle-ci est analysée par RMN (Figure 40B). Cependant, le spectre résultant est également caractéristique d'une protéine non structurée. Les deux pics résolus (autour de 1,6 ppm) appartiennent à la protéine SNAP. En effet, lorsque la protéine SNAP (comptant trois méthionines) est purifiée et analysée par RMN, son spectre RMN 2D (Figure 40C) montre des pics résolus qui se superposent avec les pics résolus observés sur le spectre de SNAP-HSulf-1  $\Delta$ HD.



A ce jour, aucune explication satisfaisante n'a pu être apportée pour justifier l'obtention de tels spectres, puisque les données sont contradictoires entre méthodes biophysiques-biochimiques et la RMN. Des tests supplémentaires d'optimisation pourraient être réalisés sur la protéine HSulf-2  $\Delta$ HDC, comme un échantillon exprimé en cellules HEK 293S (pour une glycosylation réduite), avec un marquage combinatoire, avec ou sans une digestion ménagée et la présence de ligands/substrats. En effet, cette protéine a été la seule à présenter un spectre partiellement résolu. Néanmoins, les faibles rendements protéiques obtenus limitent grandement les possibilités d'optimisation.

Ainsi, il peut être conclu que l'étude de la protéine HSulf-2 par RMN est compromise, probablement à cause de propriétés intrinsèques de la protéine (large rayon hydrodynamique, augmentation du temps de corrélation, possible dynamique du corps structuré de la protéine) qui sont à l'origine des spectres non exploitables (élargissement des signaux, chevauchement des pics, pertes de signaux, détection de signaux correspondant seulement aux régions flexibles). Néanmoins, deux milieux de marquage (spécifique et uniforme) ont pu être développés et validés sur une protéine modèle (Hsp90 N-terminal).

### **VI.3. Etudes structurales de la protéine HSulf-2 par cryo-EM**

En parallèle des études initiées par RMN et cristallographie, des analyses de cryo-EM ont été entreprises sur la protéine HSulf-2  $\Delta$ SG en collaboration avec le Dr Rebekka Wild (membre du groupe SAGAG).

La cryo-EM est une technique puissante dans la résolution de structures de grandes protéines (plusieurs centaines de kDa). Bien que des études fassent état de la détermination de structure haute résolution de protéines nettement inférieures à 100 kDa, ces dernières sont généralement présentes sous forme multimérique (ce qui, de fait, rend la taille des objets étudiés largement supérieure à 100 kDa) et/ou présentent des axes de symétrie (facilitant le traitement des données). Par conséquent, l'étude d'une protéine d'environ 100 kDa, partiellement non structurée et portant ~ 20 kDa de N-glycosylations constitue un réel challenge. Néanmoins, l'utilisation de cette méthode s'explique par une plus grande tolérance de protéines flexibles et peu structurées. De ce

fait, les N-glycosylations et la faible structuration globale de HSulf-2 ne devraient pas être limitantes, au moins dans les premières phases d'analyse.

La première étape dans l'analyse de HSulf-2  $\Delta$ SG par cryo-EM repose sur la préparation de grilles de haute qualité. Pour ce faire, la protéine pure HSulf-2  $\Delta$ SG (cf. Figure 16 p. 153) concentrée à 0,13 mg/mL a été utilisée avec un protocole standard de glaciation de grilles cryo (protocole obtenu auprès de la plateforme de cryo-EM de l'IBS). L'analyse des grilles cryo Quantifoil 1.2/1.3 Cu mesh, carbon coated 300 + HSulf-2  $\Delta$ SG a été réalisée à l'aide d'un microscope GLACIOS 200kV FEG et a permis de confirmer que l'épaisseur de glace vitreuse et la concentration de protéines utilisées, étaient optimales. En effet, cette concentration permet la présence de plusieurs centaines de particules par image, sans pour autant générer une trop grande proximité empêchant toutes analyses des micrographies. De plus, du fait de la faible épaisseur de glace, un contraste optimal des particules est observé, sans aucun chevauchement de particules dans la profondeur. Notons, pour finir, que le protocole de préparation des grilles est parfaitement reproductible.

Une fois les grilles préparées, un jeu de données a été acquis sur un microscope ThermoFisher Scientific GLACIOS 200kV FEG équipé d'une caméra de détection directe GATAN K2.

Ce premier jeu de données a permis d'acquérir 1500 micrographies. Une fois ces micrographies analysées par le logiciel de traitement Cryosparc, 2,8 millions de particules ont été extraites et une classification 2D de ces dernières a été entreprise. Après plusieurs cycles d'affinement, 10 classes 2D regroupant ~ 970 000 particules ont été sélectionnées, et elles représentent 10 orientations différentes de la protéine (Figure 41). Cependant, le rapport signal sur bruit obtenu n'est pas parfait, car sur la totalité des classes 2D, de fines densités diffuses sortent de la particule et altèrent significativement la résolution. Ces densités sont très probablement des boucles flexibles et/ou des N-glycosylations (également flexibles). Malheureusement, il est apparu au cours du traitement des données, que le logiciel n'arrivait pas à aligner les particules extraites exclusivement sur le cœur hydrophobe de la protéine. Si un tel alignement avait été réalisé, alors ces régions flexibles ne seraient que peu apparues et le rapport signal sur bruit (la netteté des classes 2D) aurait été significativement amélioré. Dans le cas présent, le logiciel a au contraire tendance à aligner les particules

sur ces régions, ce qui entraîne une perte massive de particules à classifier et une résolution finale des classes 2D ne permettant pas d'obtenir, avec les prochaines étapes, une structure 3D haute résolution. Pour tenter de pallier ce problème d'alignement, des masques ont été appliqués aux particules classifiées (c'est-à-dire que l'utilisateur filtre toutes les densités électroniques au-delà d'un certain diamètre), mais sans réelle amélioration sur l'alignement des particules et la classification. En effet, ces régions flexibles étant très petites/très proches du cœur hydrophobe de la protéine, aucun masque ne peut être ajouté sans tronquer cette région structurée des particules (et un masque ne générant pas ces troncatures ne filtre pas les régions flexibles).

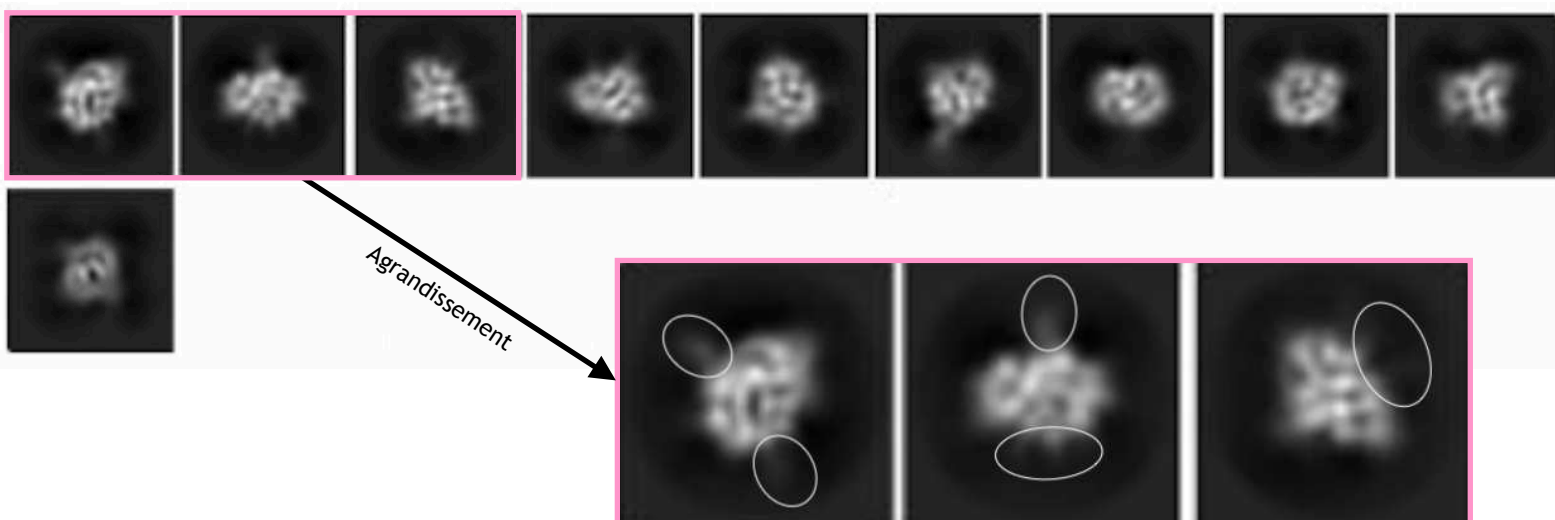


Figure 41 : Classification 2D de la protéine HSulf-2  $\Delta$ SG. L'agrandissement (en rose) des trois premières classes, par exemple, met en évidence des fines densités représentant des régions flexibles et/ou des boucles et/ou des N-glycosylations (ellipses grises). Ces structures sont délétères car elles empêchent un parfait alignement des régions structurées et donc une meilleure résolution finale des classes.

Les particules ont tout de même été extraites de ces 10 classes 2D, et une classification 3D appliquée à un modèle *ab initio* (sphère produite par le logiciel qui est par la suite affinée lors de nombreuses itérations) a été réalisée et a permis la génération, après affinement, d'un modèle structural de la protéine HSulf-2  $\Delta$ SG (Figure 42A). Ce modèle, résolu à  $\sim 7$  Å, ne permet pas la reconstruction de la structure mais autorise la validation de cette méthode et la validité des données expérimentales obtenues. En effet, un des problèmes rencontrés en cryo-EM est la production de modèles structuraux

faux car le logiciel ne raisonne pas sur la validité de la carte de densité électronique délivrée. Toutefois, dans le cas présent, en appliquant la structure prédite de HSulf-2 WT (Figure 42B) par le logiciel de prédiction AlphaFold à la carte de densité électronique de HSulf-2  $\Delta$ SG, une correspondance entre les structures secondaires de la structure prédite et du modèle est observée. Notamment, une très large correspondance est observée pour la longue hélice alpha N-terminale exposée au solvant et retrouvée chez les sulfatases (Hanson *et al.*, 2004). En revanche, le modèle produit n'est pas représentatif de la protéine HSulf-2  $\Delta$ SG dans sa totalité. En visualisant le taux de recouvrement entre densité électronique et structure prédite, seule la moitié de la protéine a été détectée. Ainsi, certaines régions du domaine CAT et du domaine C-terminal n'apparaissent pas, et aucune densité n'est observée pour le domaine HD.

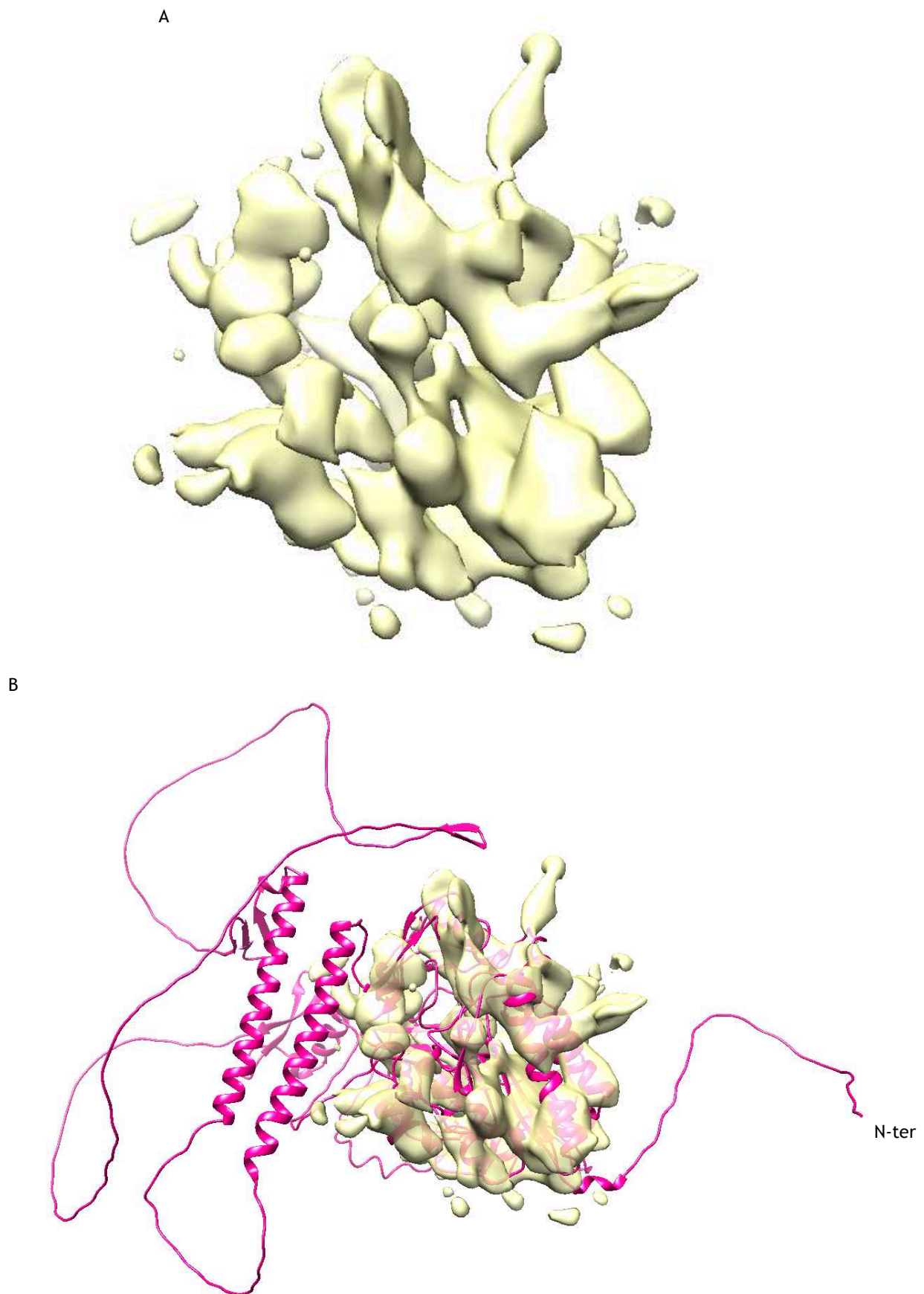


Figure 42 : Modèle structural de la protéine HSulf-2  $\Delta$ SG. (A) La densité électronique (jaune) correspond au modèle structural de HSulf-2  $\Delta$ SG. (B) Ajustement entre le modèle produit (jaune) et la structure de HSulf-2 WT prédite par le logiciel de prédiction structurale AlphaFold (rose).

Néanmoins, quelques informations intéressantes ressortent de ce modèle structural. En effet, avec une prise en compte plus importante du bruit de fond, certaines densités sortantes du modèle ressemblent à des N-glycosylations. Or, avec l'ajustement du modèle et de la structure prédite, l'Asn 241, par exemple, définie par MS comme porteuse d'une N-glycosylation, se voit correspondre parfaitement à une densité électronique attribuable à ce genre de PTMs (Figure 43A). En revanche, d'autres densités électroniques équivalentes (Figure 43B) sont aussi portées par des asparagines (cas de l'Asn 41) mais ne correspondent pas aux asparagines modifiées identifiées par MS. Cela pourrait s'expliquer par une mauvaise correspondance densité/structure prédite ou bien par la non détection de certaines N-glycosylations par MS. Par conséquent, bien que le modèle obtenu ait été validé, celui-ci ne permet pas d'obtenir d'information structurale précise sur HSulf-2 et notamment sur son domaine HD, ainsi que sur l'organisation des domaines entre eux.

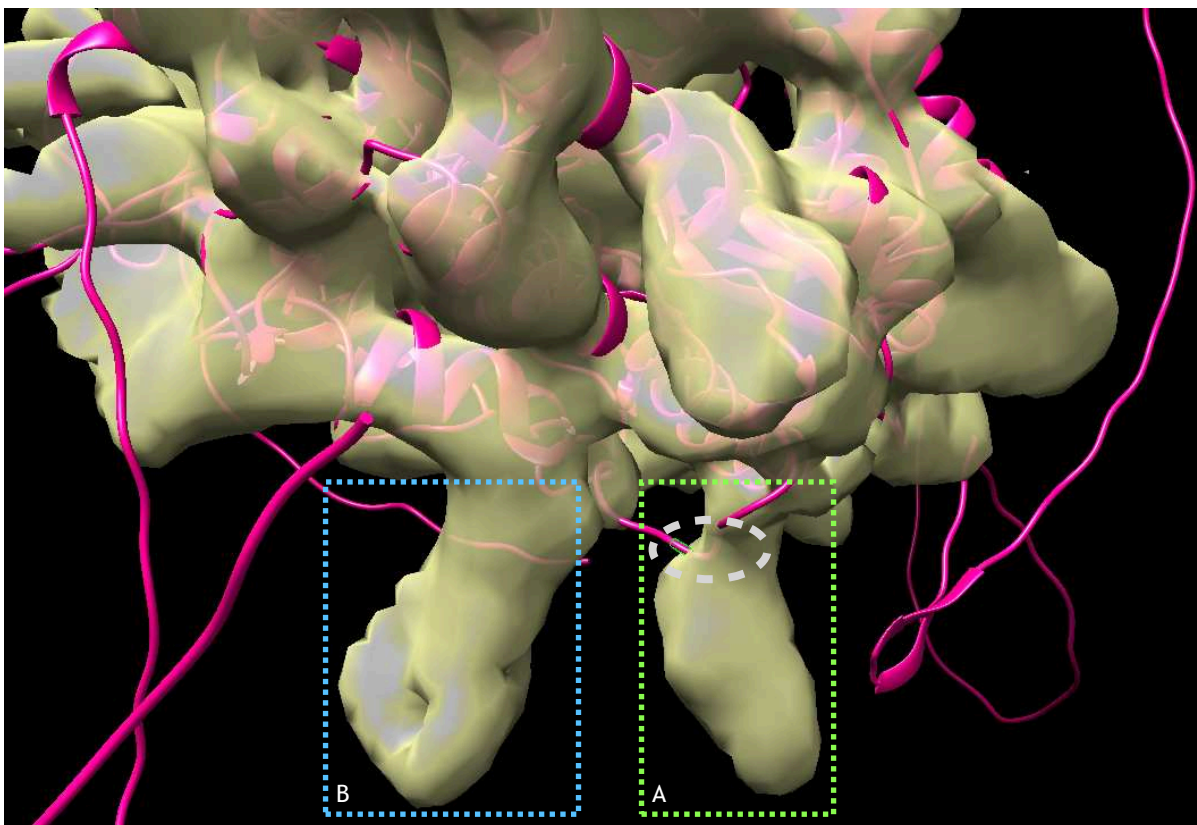


Figure 43 : Densités correspondantes aux N-glycosylations de HSulf-2  $\Delta$ SG. (A) L'ellipse grise représente le résidu N<sub>217</sub> et le rectangle vert correspond à la chaîne glycanée portée par ce résidu. (B) Le rectangle bleu correspondrait aussi à une N-glycosylation portée par le résidu (N<sub>41</sub>), mais ce résidu n'a pas été identifié comme N-glycosylé par MS. En effet, seuls les résidus N<sub>88</sub>, N<sub>108</sub>, N<sub>147</sub>, N<sub>174</sub> et N<sub>217</sub> ont été caractérisés (par MS) comme porteurs de N-glycosylations. **Rappel : la numérotation des acides aminés ne prend pas en compte le peptide signal.**



Au vu de la qualité des grilles et de la validité du modèle structural produit, un nouveau jeu de données a été acquis sur un microscope plus puissant (microscope Titan Krios 300 kV FEG équipé d'une caméra de détection directe GATAN K3) afin d'essayer d'accroître la résolution et la qualité du modèle de HSulf-2  $\Delta$ SG.

Cette acquisition a été réalisée par la plateforme de microscopie électronique de Strasbourg. Plusieurs étapes de traitement des données ont ensuite été réalisées dès les images collectées, dont l'extraction de 7 000 000 de particules. Ces particules ont été classifiées et 17 classes 2D ont été obtenues (Figure 44) regroupant ~ 1 300 000 de particules. Ces classes sont mieux résolues que les précédentes, mais les mêmes constatations peuvent être faites. En effet, des régions flexibles sont toujours visibles sur ces classes 2D, ce qui diminue encore une fois le rapport signal sur bruit et les possibilités d'atteindre une carte de densités électroniques hautement résolue.

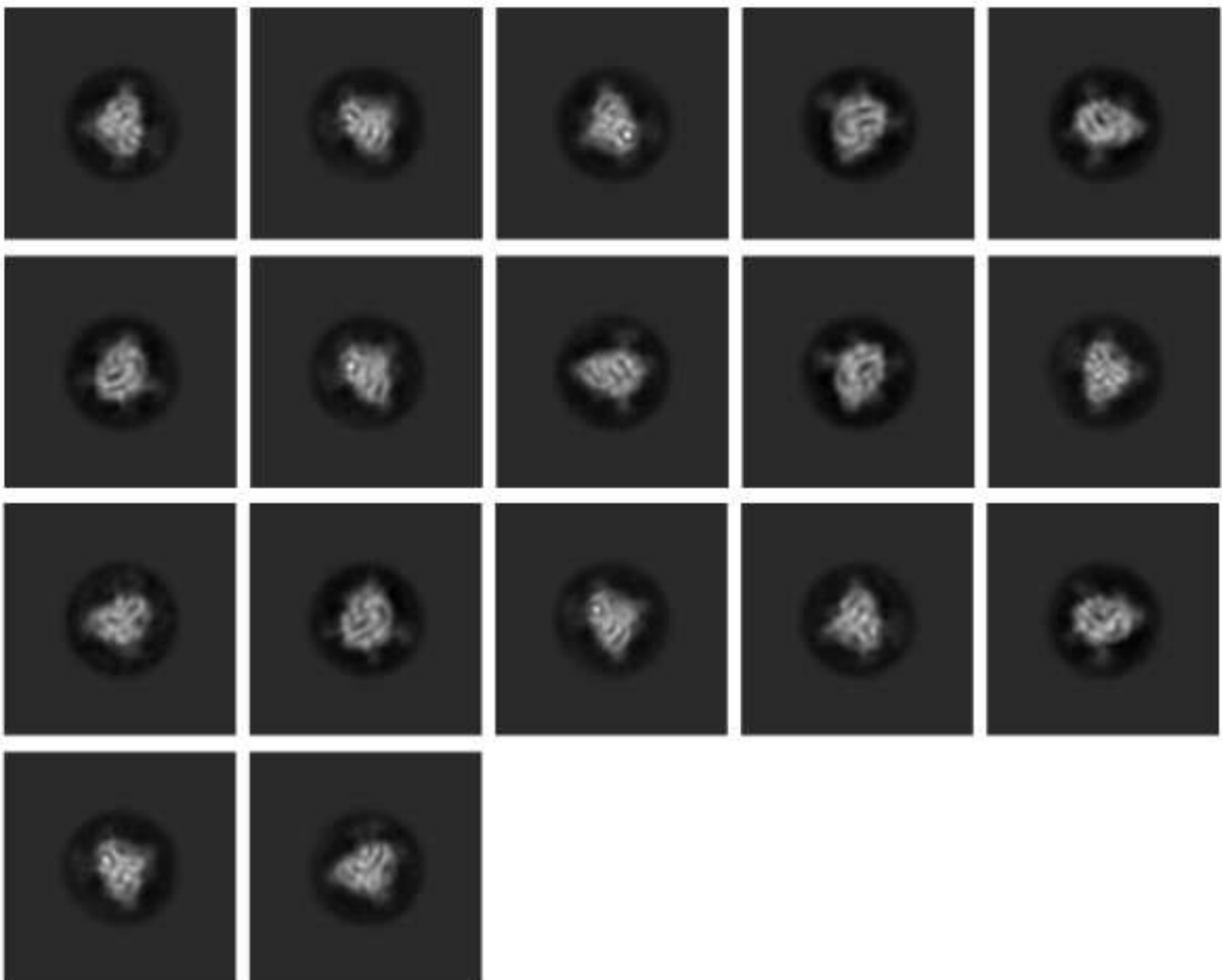


Figure 44 : Classification 2D de la protéine HSulf-2  $\Delta$ SG.

Cependant, au vu de la quantité accrue des classes 2D et des particules extraites à partir de ces classes, un nouveau modèle structural a été produit (Figure 45A<sub>i, ii</sub>). Ce modèle mieux résolu, ~ 4 Å, permet la visualisation très nette du grand feuillet  $\beta$  (10 feuillets en position N-terminale), seulement supposé dans le précédent modèle. Cependant, dans ce cas, les densités électroniques correspondant aux hélices alpha semblent être moins bien résolues et plus courtes. De surcroît, bien que le modèle soit localement mieux résolu, la carte de densités électroniques est encore plus petite que la précédente, ce qui signifie que davantage de régions de la protéine sont encore non définies et/ou non accessibles sur ce jeu de données. Ces constatations ont été corroborées par l'ajustement de la structure prédite au sein de cette nouvelle carte de densités électroniques de HSulf-2  $\Delta$ SG (Figure 45B<sub>i, ii</sub>).

Par conséquent, ce modèle est encore une fois validé par une bonne correspondance de structures secondaires (comme le large feuillet  $\beta$  et certaines hélices alpha), mais de nombreuses régions de la protéine manquent (domaine HD) ou sont réduites (domaines CAT/C-terminal). De surcroît, les informations sur l'organisation globale des domaines ne sont pas disponibles, tant entre le domaine CAT et HD qu'avec le domaine C-terminal. De ce fait, il semblerait que l'utilisation de cette méthode pour la résolution de la structure de HSulf-2 en forme apo ne soit, pour l'heure, pas possible. En effet, un calcul de la structure de HSulf-2 ne peut être effectué du fait de la présence de régions flexibles (boucles, régions non structurées, N-glycosylations) limitant, une nouvelle fois, la qualité des données acquises.

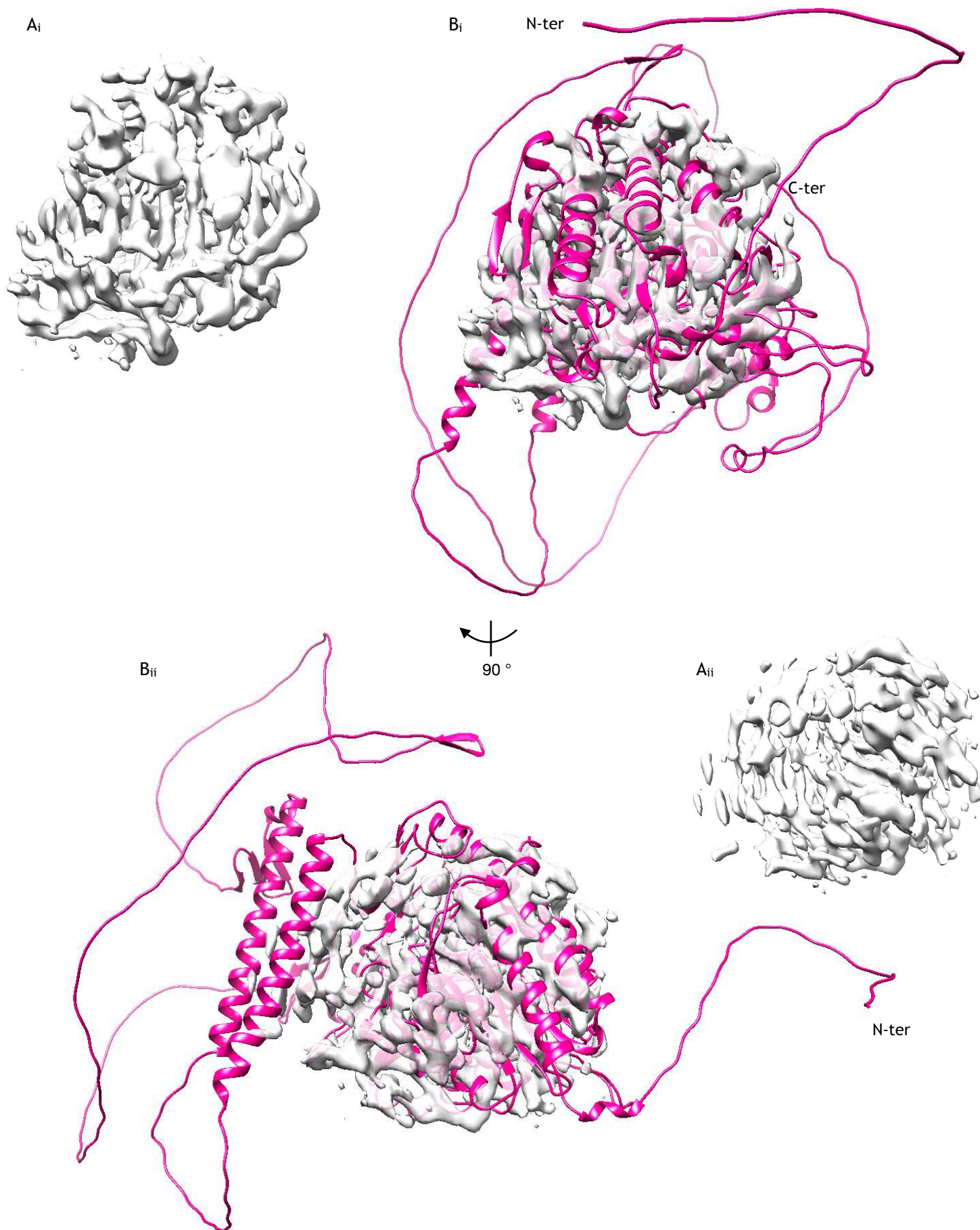


Figure 45 : Modèle structural de la protéine HSulf-2  $\Delta$ SG. (A<sub>i</sub>, ii) Deux angles de vue de la densité électronique (gris) qui correspondent au modèle structural de HSulf-2  $\Delta$ SG. (B<sub>i</sub>, ii) Ajustement entre le modèle produit (gris) et la structure de HSulf-2 WT prédite par le logiciel de prédiction structurale AlphaFold (rose), sous les deux mêmes angles de vue.

Afin d'essayer d'accroître la structuration de la protéine, et donc d'augmenter à la fois la résolution (car les régions flexibles n'apparaîtraient plus, ou moins) des classes 2D et l'étendue des régions caractérisées par des densités électroniques résolues, des tests d'optimisation de grilles ont été réalisés sur la forme holo de la protéine HSulf-2  $\Delta$ SG.

Pour cela, la protéine HSulf-2  $\Delta$ SG a été exprimée et purifiée, puis incubée avec différents substrats. Une première possibilité considérée a été l'ajout de substrat en excès dans l'échantillon protéique. Cependant, rien n'atteste que l'ensemble des protéines soit présente sous forme de complexe enzyme/substrat. Or, si ce n'est pas le cas, l'hétérogénéité de l'échantillon sera délétère lors de la classification des particules. De plus, il est probable que les substrats libres augmentent le bruit de fond et, à l'instar des régions flexibles, affectent négativement la résolution du modèle produit. Ainsi, l'application de cette méthode sans connaissances de la forme apo de la protéine reste difficile.

C'est pourquoi, afin de résoudre ce problème d'hétérogénéité, deux protocoles de réticulation ont été appliqués à la protéine dans le but de former des complexes covalents et homogènes. Ces approches reposent sur la réticulation de la protéine HSulf-2  $\Delta$ SG avec un substrat (un oligosaccharide ou une chaîne d'HP) *via* un couplage avec les réactifs l'EDC/NHS, ou bien entre les résidus glycines des protéines HSulf-2  $\Delta$ SG par un traitement au glutaraldéhyde. Ces deux approches conduisent à la formation de complexes impliquant la (quasi) totalité des protéines, car aucune bande correspondant à la protéine seule n'apparaît sur le western blot après ces traitements (à l'exception de très faibles bandes pour la condition 1) (Figure 46A). Toutefois, l'analyse des grilles en coloration négative (CN) montre une forte agrégation des différents complexes (Figure 46B). Notons que, malgré de nombreux tests d'optimisation (variation du temps de réaction, du rapport protéine-substrat et de la quantité d'agents réticulants), afin de limiter la quantité d'agrégats, aucune grille exploitable n'a pu être obtenue.

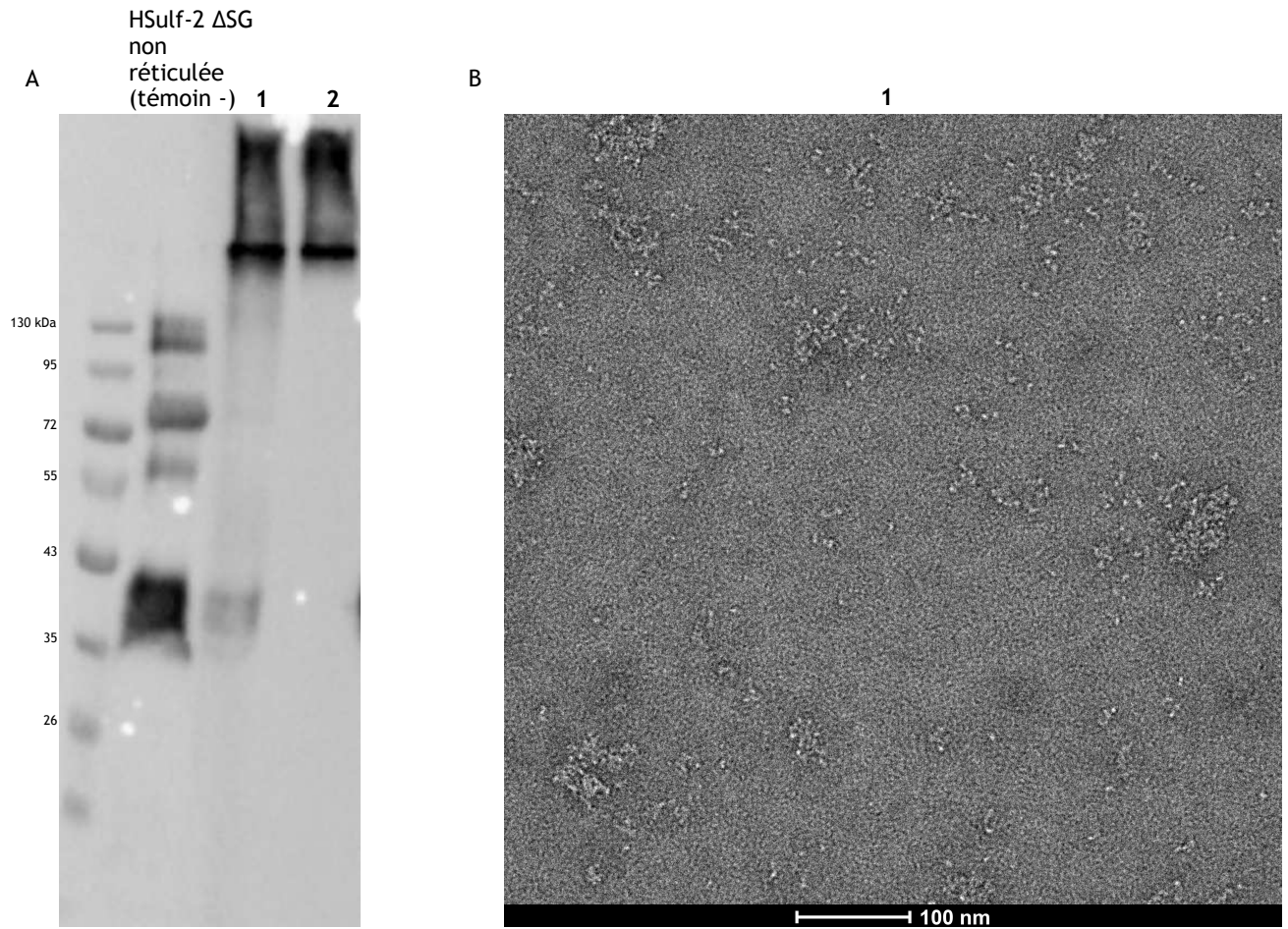
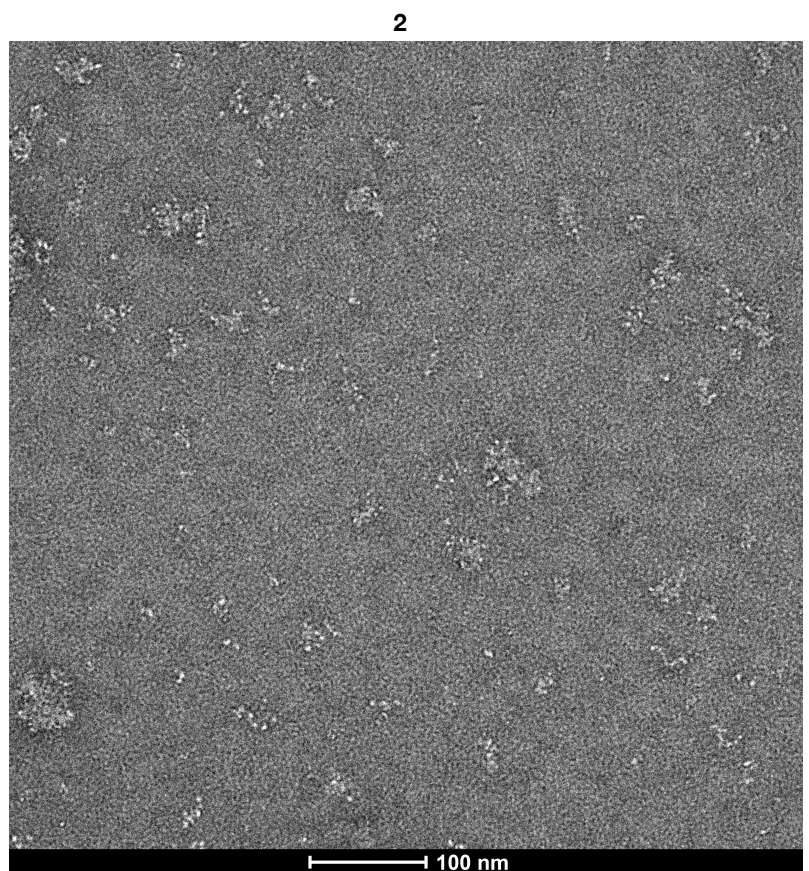


Figure 46 : Deux approches de réticulation de la protéine HSulf-2  $\Delta$ SG.

(A) Western blot de l'échantillon HSulf-2  $\Delta$ SG (Anticorps primaire : R-19, anticorps polyclonal de lapin ciblant spécifiquement la région interne du domaine HD, dil. 1/1000, et anticorps secondaire :  $\alpha$ -rabbit-HRP, JIR 111-035-045, dil. 1/10000) avant et après deux protocoles de réticulation : 1. HSulf-2  $\Delta$ SG + Substrat (DP8 : répétition de quatre unités [IdoA(2S), GlcNS(6S)]) avec un traitement EDC/NHS. 2. Inter-HSulf-2  $\Delta$ SG avec un traitement au glutaraldéhyde. Condition : 0,7 % de glutaraldéhyde, réaction pendant 3 minutes à 20°C.

(B) Grilles de microscopie électronique en colorations négatives de HSulf-2  $\Delta$ SG différemment réticulée.

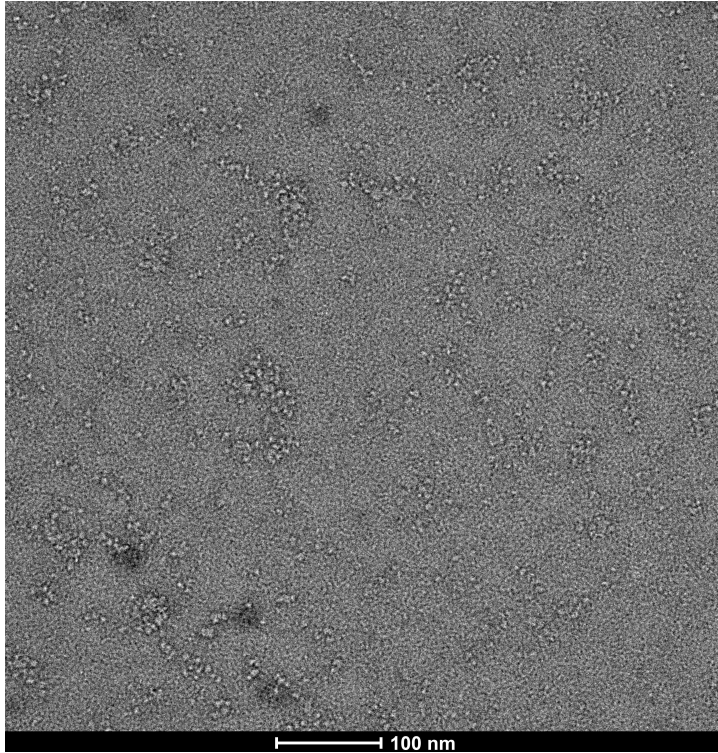


Les approches de réticulation ne permettant pas l'obtention de grilles exploitables, des expériences d'optimisation de formations de complexes protéine/HP ont été réalisées (sans agents chimiques). Pour cela, des grilles en CN ont été réalisées en utilisant des ratios variables d'HSulf-2  $\Delta$ SG/HP (Figure 47). Toutefois, l'analyse de ces grilles révèle également une forte agrégation de la protéine corrélée à une augmentation de la quantité d'HP. Ainsi, bien que le plus petit ratio protéine/HP utilisé permettrait l'obtention de grilles exploitables, rien n'affirme encore que l'ensemble des protéines interagissent avec un substrat. A l'inverse, l'utilisation d'un ratio plus fort pour lequel l'ensemble, ou une large majorité des protéines, sembleraient bien interagir avec l'HP conduit à un niveau d'agrégation qui ne permet pas une dispersion homogène des particules (indispensables pour toute analyse).

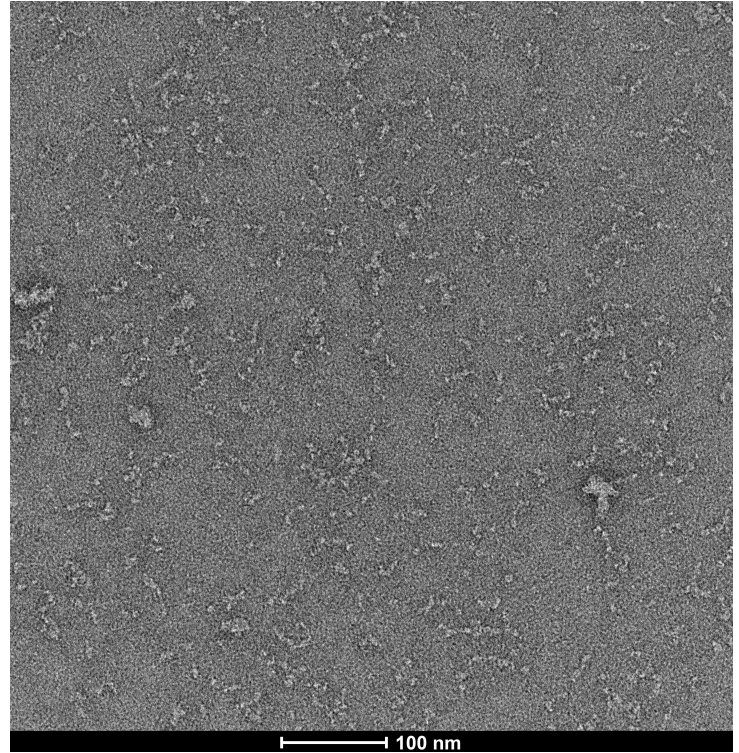
En conclusion, malgré des premiers modèles structuraux très intéressants et représentatifs de la protéine HSulf-2, la flexibilité de la protéine n'aura pas permis d'obtenir une carte de densités électroniques représentative de la totalité de la protéine, et à une résolution permettant le calcul de la structure. Néanmoins, ces premiers résultats ont permis de mieux comprendre les facteurs limitant la résolution finale des modèles. Ainsi, en corrélant les résultats obtenus après les différentes expérimentations, de nombreuses optimisations peuvent être envisagées, tout comme une nouvelle stratégie de caractérisation de HSulf-2 par cryo-EM (cf. p. 228-230).



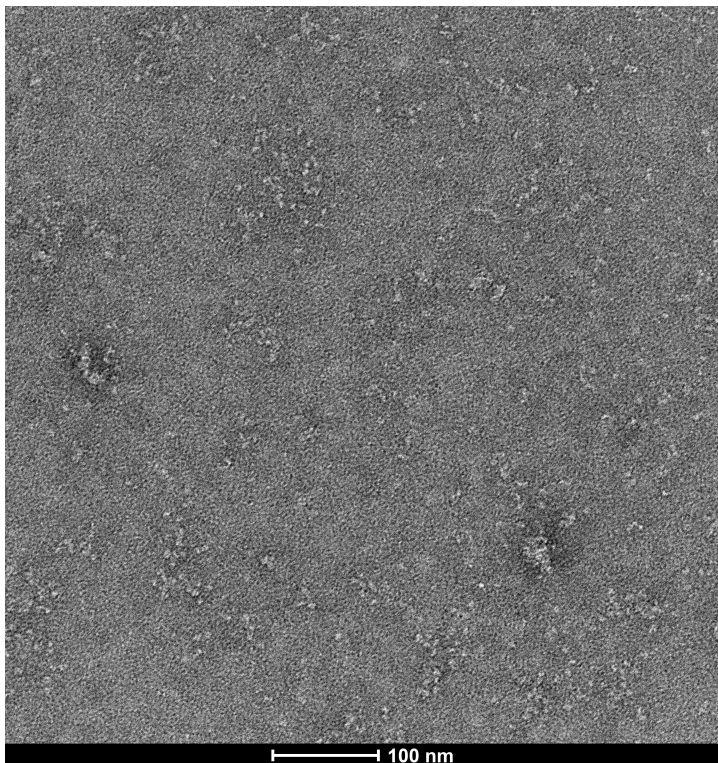
1



2



3



4

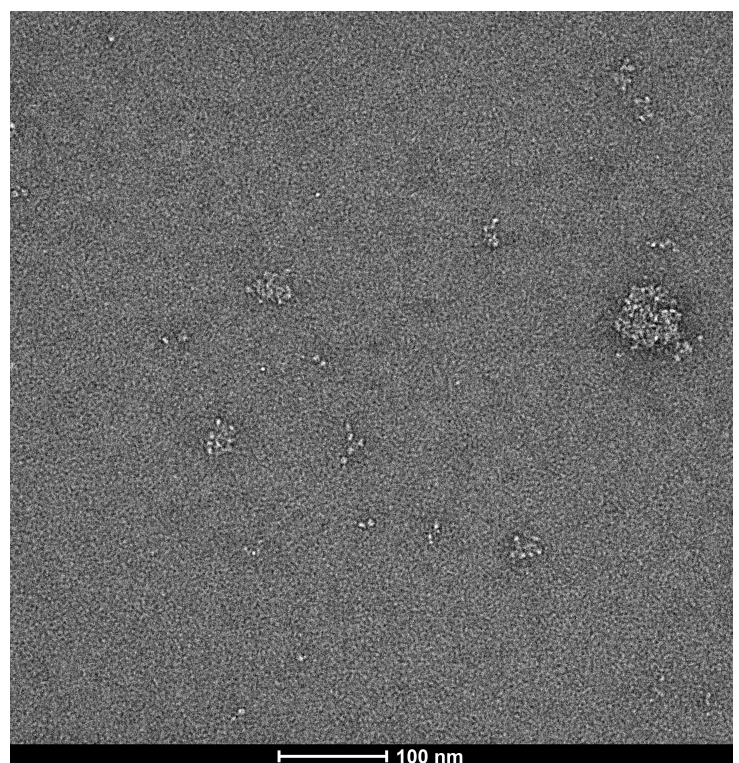


Figure 47 : Grilles de microscopie électronique en colorations négatives de HSulf-2  $\Delta$ SG avec différents ratios d'HP (taille du substrat : 1400 Da, réf. H3393, Sigma). La formation des complexes HSulf-2  $\Delta$ SG - HP a été réalisée (sans réticulation) pendant 15 minutes à 4°C.

1. HSulf-2  $\Delta$ SG - HP : ratio 1:1
2. HSulf-2  $\Delta$ SG - HP : ratio 1:4
3. HSulf-2  $\Delta$ SG - HP : ratio 1:8
4. HSulf-2  $\Delta$ SG - HP : ratio 1:16

## ***Discussions et perspectives***





## VII. Discussions et perspectives

Au cours de cette thèse, un large spectre de méthodes de caractérisation structurale a été balayé en vue de résoudre la structure de la protéine HSulf-2. Malgré de nombreuses optimisations portant sur l'expression, la purification, le traitement des PTMs, la production de nouvelles constructions et le développement d'un milieu de marquage pour analyser spécifiquement cette enzyme par RMN, aucune réponse n'a pu être apportée à notre problématique. Néanmoins, ces travaux ont permis d'identifier les méthodes d'analyse représentant des pistes prometteuses à investiguer en priorité.

### VII.1. Développement d'un milieu de marquage spécifique et uniforme pour des protéines exprimées en cellules de mammifère

Les travaux réalisés au cours de la thèse ont permis de mettre au point un milieu de culture adapté au marquage spécifique des acides aminés portant un groupement méthyle, ainsi qu'un milieu pour le marquage uniforme  $^{15}\text{N}$ . Le milieu de marquage spécifique est une innovation puisque, jusqu'à présent, aucun milieu n'était disponible commercialement, ce qui limitait donc l'utilisation du marquage spécifique des acides aminés portant un groupement méthyle (dans le système d'expression eucaryote) à une communauté très restreinte d'experts RMN. Or, les industriels pharmaceutiques manifestent un intérêt croissant pour les protéines humaines qui nécessitent la plupart du temps d'être exprimées dans des systèmes eucaryotes pour être fonctionnelles. Elles peuvent également être de haute masse moléculaire, comme les anticorps. Ainsi, l'absence de disponibilité d'un milieu de culture adapté limitait, jusqu'alors, l'étude par RMN de ces protéines. Il sera désormais possible d'étudier, à l'échelle atomique, des interactions de protéines d'intérêt thérapeutique avec des ligands, par exemple.

Un milieu permettant un marquage uniforme des  $^{15}\text{N}$  existe déjà commercialement (BioExpress 6000, Eurisotop). Celui-ci est probablement réalisé par ajout de chacun des 20 acides aminés  $^{15}\text{N}$  marqués purs, d'où son coût d'environ 7000 €/L. Par comparaison, même s'il est difficile de donner un prix fixe au marquage uniforme  $^{15}\text{N}$  réalisé (cf. Tableau 10 p. 139) (l'extrait de levure utilisé étant obtenu par collaboration car commercialement non disponible - Dr. Nadine ASSRIR, ICSN Gif/Yvette), en se basant sur notre protocole et les expériences chiffrées décrites dans la publication

de Opitz *et al.*, 2015, un coût d'environ 300 €/L peut être estimé (ce coût ne comprend toutefois pas la main d'oeuvre) ; le rendant accessible à une majorité de laboratoires.

Quant au milieu de marquage spécifique des acides aminés portant un groupement méthyle (sans deutération des autres résidus) développé, en considérant que 80 mg/L de chacun des acides aminés portant un groupement méthyle sont ajoutés, ce dernier pourrait être commercialisé autour de 5000 €/L.

Dans ce milieu, les acides aminés portant un groupement méthyle sont deutérés sur le reste de leurs carbones (autre carbone que le  $^{13}\text{CH}_3$ ). Cette deutération permet une augmentation du signal et de la résolution des spectres (Yanaka *et al.*, 2018 ; Dubey *et al.*, 2021) nécessaire à l'étude de protéines de haute masse moléculaire. Avec un tel marquage, il est par exemple possible d'étudier une interaction de la protéine d'intérêt avec un ligand. En comparant les deux spectres obtenus (avec et sans ligand), il sera possible de conclure sur une éventuelle interaction puisque les pics correspondant aux résidus impliqués dans l'interaction subiront une perturbation de leurs déplacements chimiques, sur le spectre RMN, en présence du ligand. C'est ce qui a été montré sur la figure 39B (cf. p. 195) avec le sulfocalix-[4]-arène et la protéine HSulf-2  $\Delta\text{HDc}$ .

En revanche, ces signaux ne seront pas attribués (il ne peut donc être identifié quels sont les résidus impliqués dans cette interaction). L'attribution nécessite, en effet, l'acquisition d'expériences 3D de RMN sur des échantillons les plus deutérés possibles afin d'augmenter la sensibilité et la résolution des signaux. Pour cela, il faut envisager d'apporter une source de deutération supplémentaire en remplaçant les 14 acides aminés protonés par leur version deutérée (car ces cellules ne peuvent croître en présence de deutérium ; Murphy *et al.*, 1977). Cela a pour conséquence d'augmenter le coût du milieu de culture mais également de diminuer les rendements (Sitaraka *et al.*, 2015). Toutefois, certains travaux décrivent des protocoles intermédiaires où seuls certains acides aminés sont ajoutés sous forme deutérée, notamment les acides aminés hydrophobes (Kofuku *et al.*, 2018).

Pour les protéines de plus faible masse moléculaire (<30 kDa), un marquage uniforme peut suffire. Pour attribuer le squelette (étape préalable au calcul de structure, par exemple), il est nécessaire que la protéine soit marquée *U*-[ $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ]. Pour cela, il faut ajouter des extraits de levure triplement marqués, du glucose  $^{13}\text{C}$  et des inhibiteurs de transaminase pour bloquer la synthèse des acides aminés non

essentiels non marqués. Or, les extraits de levure avec un tel marquage ne sont commercialement pas disponibles. Une étude a décrit leur préparation (Opitz *et al.*, 2015), mais elle implique un protocole coûteux et fastidieux. Il faut également, dans ce cas, ajouter une source d'ammonium marquée ( $^{15}\text{NHCl}_4$ ) car la préparation des extraits détruit certains acides aminés (cf. III.3.3.2. p. 108). Néanmoins, en utilisant des extraits de levure  $U$ - $^{2}\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ] et du glucose non deutéré, ces travaux ont décrit une deutération >60 % des protéines recombinantes et des taux d'incorporations importants des différents acides aminés. Une étude récente a permis de développer un protocole permettant l'utilisation de  $^{2}\text{H}_7$ -glucose (Yanaka *et al.*, 2018). En revanche, un certain nombre d'autres publications décrivent une toxicité du glucose deutéré. Il sera donc nécessaire de vérifier la possible utilisation du glucose marqué  $^2\text{H}$  avec le milieu de culture développé afin d'apporter une deutération la plus complète possible.

Dans tous les cas, la deutération apportée par les acides aminés diminue les rendements. Or, comme décrit dans cette thèse, à temps d'expression identique le milieu actuellement développé voit déjà son efficacité diminuer d'environ 25 %. En théorie, pour changer de milieu de culture, les cellules doivent généralement être adaptées progressivement. Cependant, dans le cas de notre travail, nous n'avons pas pu réaliser une adaptation progressive des cellules. Cela aurait probablement permis le maintien de la croissance cellulaire (dès les premières heures) et donc les rendements protéiques. Au contraire, les cellules ont été transfectées dans un milieu commercial (FreeStyle 293, Gibco®) et directement centrifugées puis re-suspendues dans le milieu de marquage (ESF-SFM déplété et reconstitué). Cela a donc très probablement engendré un stress sur des cellules déjà sensibilisées par les agents de transfection, et conduit à la diminution de la croissance cellulaire et donc des rendements protéiques. Notons toutefois qu'une stratégie simple et efficace a pu être proposée pour palier le problème de rendement protéique diminué avec une adaptation du protocole d'expression (cf. p. 171). Il serait également intéressant d'identifier une stratégie de protection des cellules (face au stress) afin de diminuer et/ou de gagner 24 heures d'expression supplémentaires (et possiblement gagner en rendement protéique).

Dans les protocoles de transfection « classiques » pour les cellules HEK 293, en milieu FreeStyle 293, de la peptone (composé riche en peptides à chaînes courtes et en acides aminés) est supplémentée au milieu de culture, au moment de la transfection, et entre

12 et 24 heures après (Pham *et al.*, 2005 ; Davami *et al.*, 2011, 2014). Cet ajout de peptone conduit, en plus d'une augmentation de 15 % de l'efficacité de transfection (si ce composé est supplémenté lors de la réaction PEI-ADNc), à une augmentation de 37 % de la densité cellulaire (Davami *et al.*, 2014). Toutefois, dans le cas d'une production de protéines marquées pour la RMN, cet ajout n'est pas envisageable puisque la peptone contient une grande quantité d'acides aminés non marqués qui viendraient « diluer » les acides aminés marqués. En revanche, la peptone pourrait être ajoutée au milieu FreeStyle 293 lors de la transfection pour accroître l'efficacité de cette dernière et, ainsi, augmenter les rendements protéiques sans pour autant accroître la densité cellulaire finale. Par ailleurs, comme décrit dans l'étude publiée par Zhang *et al.* (2017), il pourrait être imaginé de réduire la quantité ajoutée et/ou dialyser la peptone afin d'éliminer les acides aminés, mais garder d'autres facteurs de croissance métaboliques essentiels à la croissance cellulaire (et ainsi, augmenter les rendements protéiques).

De la même façon, il est nécessaire de tester différents agents de transfection, non pas pour influencer la croissance cellulaire, mais pour augmenter les rendements protéiques. En effet, au-delà du PEI (23966, CliniSciences) qui est l'agent de transfection utilisé pour l'ensemble des expériences décrites, un agent dérivé, le PEI MAX, a été développé et commercialisé et permet d'accroître les rendements protéiques (Baranyi *et al.*, 2013 ; Delafosse *et al.*, 2016). Ceci a également été confirmé au sein du laboratoire, sur la protéine humaine 6OST1, où le PEI MAX a permis un accroissement du rendement protéique de x1,8 (comparé au PEI). La Lipofectamine 2000, et récemment la Lipofectamine 3000 (Invitrogen™), permettent également une augmentation significative des expressions protéiques. Ainsi, bien que le prix de ces agents transfectants soit élevé (~ 150 euros pour 100 mL de culture), ils pourraient constituer une alternative intéressante si les rendements protéiques sont significativement augmentés, et sachant qu'un litre du milieu de marquage développé coûte environ 5000 euros.

Outre ces agents transfectants, des variations dans le protocole de transfection peuvent générer une augmentation des taux d'expression des protéines cibles, comme la variation du rapport PEI-ADN ou la concentration des cellules présentes lors de la transfection. L'utilisation d'une haute densité cellulaire au moment de la transfection (entre 5 et  $10 \times 10^6$  cellules/mL au lieu d'une densité de  $1,5 \times 10^6$  cellules/mL utilisée classiquement) a déjà été décrite pour accroître les rendements protéiques (Backliwal

*et al.*, 2008 ; Elshereef *et al.*, 2019). Les cellules sont ensuite re-diluées à  $1,5 \cdot 10^6$  cellules/mL, quatre à six heures après la transfection.

En revanche, bien que l'utilisation de nouveaux agents transfectants et/ou l'accroissement de la densité cellulaire lors de la transfection (avant de diluer la culture au bout de 4-6 heures) n'engendreront pas de dilution des isotopes supplémentés au milieu de marquage, l'ajout de peptone pour accroître la densité cellulaire aura un impact certain sur la dilution des isotopes (la peptone étant une source d'acides aminés non marqués). Par conséquent, de nouvelles courbes d'incorporation devront être établies.

Toutes ces variations dans le protocole sont susceptibles de modifier le taux d'incorporation des acides aminés. Il sera donc nécessaire de valider ces valeurs d'incorporation (courbes et taux d'incorporation) au travers du nouveau protocole établi.

Dans le milieu actuel, il a été défini que 80 mg/L de  $^{13}\text{CH}_3$ -acides aminés, supplémentés au milieu, représentait la concentration minimale en acides aminés générant un plateau maximal d'incorporation. Pour rappel, cette valeur a été calculée en établissant des courbes d'incorporation par rapport à une référence interne ajoutée en excès (la  $^{13}\text{CH}_3$ -Met). Il faut néanmoins prendre en considération qu'à cause de la différence de flexibilité de leurs chaînes latérales, les acides aminés portant un groupement méthyle ne possèdent pas les mêmes propriétés spectroscopiques, ce qui se traduit par des différences d'intensités de leurs signaux RMN. Cette différence peut parfois être exacerbée par l'existence de phénomènes dynamiques intrinsèques à la protéine d'intérêt, biaisant davantage les quantifications par RMN. Par ailleurs, il est important de préciser que l'expérience RMN type SOFAST (choisie pour sa sensibilité et facilitant l'obtention de spectres de qualité acquis sur des échantillons peu concentrés) n'est pas optimale pour des analyses de quantification.

Ainsi, pour une quantification précise de l'incorporation des acides aminés marqués, il a été nécessaire de produire des échantillons marqués (mais sans deutération, puisque les deutériums peuvent être ré-échangés par des protons par l'action de transaminases présentes dans le milieu de culture, et donc induire de l'hétérogénéité dans la masse de la protéine) afin de calculer leur masse et la comparer à celle de la protéine non marquée. Les échantillons ont été produits avec des acides aminés marqués seulement

sur le  $^{13}\text{CH}_3$  (alanine et méthionine), ou bien uniformément marqués sur tous les  $^{13}\text{C}$  et  $^{15}\text{N}$  (thréonine, valine, isoleucine, leucine). Théoriquement, en calculant la différence de masse entre la masse attendue pour 100 % de marquage et la valeur donnée par les analyses de spectrométrie de masse, il était possible de définir un pourcentage d'incorporation. Les résultats obtenus ont cependant été plus complexes à analyser. En effet, dans le cas de la méthionine et de l'alanine, bien que des taux d'incorporation aient pu être calculés pour ces résidus, la variabilité de ces taux, entre les quatre espèces protéiques présentes dans l'échantillon, ne permet pas de conclure quant à une valeur précise. Néanmoins, en se basant sur une analyse visuelle et d'intégration de la figure 33 (cf. p. 179), il est tout de même apparu que les valeurs calculées étaient cohérentes.

Ainsi, le taux d'incorporation calculé pour l'alanine (72 %) est très satisfaisant par rapport à ceux obtenus dans la littérature. En effet, en comparaison, en cellules d'insectes, l'ajout de  $^{13}\text{CH}_3$ -(-C<sup>2</sup>H)-Ala<sub>B</sub> n'aura conduit qu'à 45 % d'incorporation des B- $^{13}\text{C}$  (Kofuku *et al.*, 2018). Une autre étude fait état de 73 % d'incorporation des alanines dans le cadre d'un marquage uniforme  $U$ -[ $^1\text{H}$ ,  $^{13}\text{C}$ ] en présence de 50  $\mu\text{M}$  de L-cyclo-sérine (concentration identique à celle utilisée lors des expériences menées dans cette thèse ; Sitarska *et al.*, 2015), soit une valeur très proche de celle que nous avons obtenue.

Pour les autres acides aminés (Met, Ile, Val, Thr et Leu), les valeurs obtenues sont très variables. Néanmoins, il s'agit d'acides aminés essentiels, donc non synthétisés par la cellule. En théorie, le pourcentage d'incorporation devrait être de 100 % puisque seuls des acides aminés marqués sont apportés au milieu. Ainsi, la diminution et la variabilité du pourcentage d'incorporation calculé proviendrait (malgré la présence d'une étape de lavage des cellules au PBS, qui permet d'éliminer les acides aminés non marqués du milieu FreeStyle), d'une réserve intracellulaire d'acides aminés persistante et entrant en compétition avec les acides aminés marqués ajoutés dans le milieu après la transfection.

Par ailleurs, le choix de la méthode de quantification de l'incorporation peut être discuté. En effet, le fait que la protéine soit petite avec peu d'acides aminés de chaque type fait que la variabilité du marquage (massif isotopique) entraîne une erreur dans l'estimation de l'incorporation. Pour affiner ces résultats, un projet TGE-FT-ICR, en collaboration avec le Dr. Héloïse Dossmann (Sorbonne Université), est en cours afin

d'acquérir des spectres résolus (déconvolution des massifs isotopiques) et des taux d'incorporation plus précis.

Bien que des optimisations puissent encore être apportées au milieu de marquage spécifique, il a déjà permis à NMR-Bio d'exprimer d'autres protéines en cellules HEK 293 dans le cadre des services que l'entreprise propose. En revanche, un travail plus conséquent reste à effectuer pour finaliser le milieu de marquage uniforme, notamment, au regard de la deutération.

## VII.2. Etudes structurales de la protéine HSulf-2

Les récentes tentatives de cristalliser la protéine HSulf-2 représentent, à court terme, l'une des perspectives les plus prometteuses dans la résolution de sa structure (la deuxième étant présentée p. 220-221). En effet, les derniers tests de cristallogénèse ont permis de mettre en évidence, vraisemblablement, plusieurs cristaux pour HSulf-2  $\Delta$ SG, et des résultats encourageants pour HSulf-2  $\Delta$ HD et  $\Delta$ HDc. Pour la première fois, de la précipitation protéique a pu être observée (validée par l'analyse UVs), ainsi que des proportions gouttes claires/précipitées intéressantes et permettant d'imaginer de nouvelles expériences optimisées.

Ces nouvelles expériences seront d'autant plus intéressantes et cruciales, car un nouveau protocole d'expression a été mis au point, permettant d'augmenter par 2,5 (voire 3,5) les rendements protéiques. Dans ce protocole, les cellules suivent le même protocole d'initiation de l'expression (avec, entre autres, le même protocole de transfection (si expression transitoire) et la même densité cellulaire initiale de  $1,5 \cdot 10^6$  cellules/mL), mais après quatre jours d'expression, les cellules sont re-diluées à  $1,5 \cdot 10^6$  cellules/mL pour, de nouveau, trois à quatre jours d'expression supplémentaires. Ainsi, ceci autorise donc pour la première fois l'expression des protéines HSulf-2  $\Delta$ SG (ayant fournies les meilleurs résultats en cristallogénèse) et HSulf-2  $\Delta$ HDc (qui semblait être la plus favorable par RMN) en cellules HEK 293S. Pour rappel, l'utilisation de ces cellules engendre un motif glycosidique réduit des protéines exprimées, mais également une diminution d'environ 30 % des rendements protéiques (déjà faible  $\sim 1$  mg/L), limitant jusqu'ici leur utilisation à la seule protéine fortement exprimée HSulf-2  $\Delta$ HD. Dorénavant, l'ensemble des protéines pourra être exprimé sous forme partiellement



glycosylées, avec des rendements permettant de les concentrer davantage. Or, ces deux paramètres sont cruciaux pour la réussite de la cristallogenèse de HSulf-2.

i. Pour la protéine HSulf-2  $\Delta$ HD (exprimée en HEK 293S), trois concentrations différentes pourront être testées : 30 mg/L, 40 mg/L et 60 mg/L. Pour rappel, l'analyse des plaques de cristallogenèse a montré 27 % de gouttes claires, 50 % de précipitées et 23 % présentant une séparation de phases. Ainsi, il peut être supposé qu'en modulant la concentration en protéines, davantage de cristaux pourraient se former.

De surcroît, ces mêmes conditions seront testées avec la protéine incubée avec le ligand sulfocalix-[4]-arène. Ceci étant, il sera nécessaire d'identifier un nouveau protocole d'addition du ligand afin d'éviter une précipitation massive de la protéine. Pour ce faire, le ligand pourrait être ajouté sur la protéine diluée, puis le complexe formé serait concentré (au lieu d'ajouter le ligand directement sur la protéine concentrée), ce qui pourrait permettre de maintenir la solubilité de la protéine.

ii. Pour la protéine HSulf-2  $\Delta$ HDC, l'analyse des plaques révélait 50 % de gouttes claires, 30 % de séparations de phases et 20 % de précipitations. Cependant, au vu de l'échantillon utilisé pour ces tests (précipité apparu une fois le sulfocalix-[4]-arène ajouté), il est difficile de conclure quant à la fiabilité de ces statistiques. Par conséquent, de nouveaux tests devront être menés avec des échantillons protéiques, si possible davantage concentrés (avec et sans sulfocalix-[4]-arène, si un protocole est identifié), et exprimés en cellules HEK 293S.

iii. Pour la protéine HSulf-2  $\Delta$ SG, un affinement des trois conditions ayant présenté des cristaux devra être réalisé, et une nouvelle campagne de cristallogenèse sera réalisée (avec les mêmes plaques de criblage, cf. Tableau 4 p. 127) avec une protéine plus concentrée et présentant un motif glycosidique réduit (expression en cellules HEK 293S).

En résumé, les résultats encourageants observés, et le nouveau protocole d'expression établi, permettront pour la première fois de combiner l'ensemble des paramètres importants à la cristallogenèse de HSulf-2, soit : une protéine, la plus structurée possible (prise en compte des régions structurées seulement), délaissant les régions et

les structures flexibles (domaine HD et N-glycosylations) et pouvant être largement exprimée afin de ne pas être limité lors de la concentration de la protéine.

Un autre aspect intéressant de cette thèse réside dans le fait que l'ensemble des résultats et expérimentations (l'optimisation et le développement de nouveaux protocoles d'expression et de purification, le criblage de nouvelles conditions de cristallisation et le développement de milieu de marquage) ont contribué à faire avancer les connaissances requises pour résoudre, dans le futur, la structure de cette enzyme. En effet, ces stratégies ont mis en lumière que les propriétés intrinsèques de la protéine (faible structuration et grande flexibilité) empêchent toute acquisition de jeux de données (RMN et cristallographie) ou génèrent un jeu de données dont la qualité ne permet pas d'effectuer le calcul de la structure (cryo-EM). Par conséquent, de nouveaux efforts doivent être mobilisés afin de limiter au maximum la flexibilité de cette enzyme. Pour ce faire, différentes approches peuvent être envisagées.

Au vu de l'ensemble des analyses réalisées, il est difficile d'identifier quelles sont les caractéristiques de HSulf-2 (parmi les N-glycosylations, la présence de régions non structurées, ou une grande flexibilité inter-domaine de la protéine) qui constituent les principaux obstacles à la résolution de sa structure. Néanmoins, en se basant sur les résultats de RMN, il semblerait que la présence de N-glycosylations ne soit pas le principal facteur responsable de l'obtention de spectres non résolus. En effet, l'expression de la protéine en cellules HEK 293F ou S n'a montré aucune différence dans la résolution des spectres obtenus. Aussi, il n'est pas à exclure que le niveau réduit de N-glycosylations suffise à augmenter le temps de corrélation, empêchant donc l'obtention de spectres RMN résolus. Néanmoins, l'absence de toute amélioration du spectre suggère que les propriétés de structuration et de flexibilité intrinsèques de la protéine sont le principal facteur de tels résultats. De surcroît, la littérature semblerait aller dans ce sens. En effet, il a été démontré l'importance des N- et O-glycosylations dans le maintien de la stabilité et la diminution de la dynamique des protéines, mais pas dans la structuration de ces dernières (Lee *et al.*, 2015).

Il est tout de même nécessaire de souligner que, d'après les données de cryo-EM, certes les glycosylations ne semblent pas avoir une importance majeure dans la structuration de la protéine, mais elles altèrent considérablement le traitement de données, et donc notre capacité à résoudre la structure de HSulf-2.

Par conséquent, identifier une forme déglycosylée et stable serait une grande avancée pour l'obtention de données en cristallographie et de classes 2D mieux résolues en cryo-EM. Cela permettrait également de vérifier l'importance des N-glycosylations et leurs répercussions sur l'objectif du projet (et de savoir si les futurs efforts doivent se concentrer sur l'augmentation de la structuration de l'enzyme). Pour ce faire, différentes expériences peuvent être envisagées.

La protéine HSulf-2  $\Delta$ HD pourrait être traitée enzymatiquement afin d'éliminer les N-glycosylations. Pour ce faire, et sachant qu'un traitement à la PNGase F ne permet pas de maintenir la stabilité de la protéine, un traitement de HSulf-2 avec l'enzyme Endoglycosidase H (endoH) pourrait être entrepris. En effet, quand la PNGase clive la totalité de la chaîne glycanique, l'endoH laisse le premier résidu GlcNAc sur la protéine. Cependant, il est difficile de prédire si la présence d'un seul résidu GlcNAc suffira au maintien de la stabilité de l'enzyme. Toutefois, une étude a démontré qu'un traitement d'IgG de lapin avec cette enzyme permettait d'obtenir une protéine stable, mais également d'identifier, une fois l'acquisition de spectres RMN réalisée, que la digestion par endoH affectait le micro-environnement de la Leu<sub>235</sub> (qui est impliquée dans le site de liaison Fc $\gamma$ RIII) en induisant une altération conformationnelle (à cause de l'absence de la chaîne glucidique entière) et en rendant le site de liaison Fc $\gamma$ RIII inactif (Yamaguchi *et al.*, 2006).

Une alternative au traitement endoH, pourrait être un traitement des cellules avec de la tunicamycine (Yamaguchi *et al.*, 2007). Ici, au lieu d'un traitement par glycosidase de la protéine purifiée, l'antibiotique tunicamycine inhibe l'enzyme GlcNAc phosphotransférase, ce qui a pour conséquence d'inhiber la formation des N-glycosylations.

Une dernière approche, plus fastidieuse, serait de supprimer par mutagène dirigée tout ou partie des sites de N-glycosylations identifiés par MS (Seffouh *et al.*, 2019). Cette approche a été utilisée pour la déglycosylation d'IgG pour laquelle il a été démontré que la mutation de certains sites n'affectait en rien la stabilité et l'activité de la protéine, quand d'autres se sont révélés être critiques (Yamaguchi *et al.*, 2007).

En outre, au delà de la compréhension de l'impact des ces PTMs, ainsi que d'une possible amélioration du traitement de données, a minima pour la cryo-EM, l'élimination des glycosylations ouvrirait de nouvelles perspectives d'étude sur la protéine HSulf,

notamment sur l'impact du nombre élevé de N-glycosylations sur l'accessibilité de la chaîne de GAG, par exemple.

En parallèle, des tests sur l'accroissement de la structuration de la protéine doivent être entrepris. Dans un premier temps, une stratégie équivalente à celle de l'étude du domaine HD par RMN (cf. p. 86) pourrait être appliquée aux domaines CAT et C-terminal. Celle-ci consisterait à concevoir de petits fragments des domaines, en s'aidant de la structure prédite par AlphaFold, afin de couper la séquence de manière précise sans affecter de structures secondaires, puis d'analyser ces fragments par RMN (marquages uniformes  $^{13}\text{C}$ ,  $^{15}\text{N}$ , et deutération de la protéine). Ainsi, suivant les résultats obtenus, la structure tridimensionnelle de ces fragments pourrait être obtenue (par RMN et/ou cristallographie) et comparée à la structure de la protéine entière prédite par AlphaFold, dans le but de valider ou modifier cette dernière, ou de concevoir de nouvelles constructions plus favorables à une caractérisation structurale.

Des données, récemment obtenues (au sein du laboratoire) par photométrie de masse, suggèrent que la protéine HSulf-2 WT (avec chaîne de CS) et  $\Delta\text{SG}$  sont dimériques (données non présentées). En revanche, l'analyse des protéines HSulf-2  $\Delta\text{HD}$  et  $\Delta\text{HDc}$ , par cette même méthode, montre que ces protéines sont monomériques (données non présentées). Cela indique que le motif responsable de la dimérisation de la protéine se situerait dans la région interne du domaine HD (région éliminée pour les protéines HSulf-2  $\Delta\text{HD}$  et  $\Delta\text{HDc}$ ). Ceci est en accord avec l'étude de Morimoto *et al.* (2002) qui a reporté des séquences putatives de type « coiled-coils » qui pourraient servir d'éléments de multimérisation pour les enzymes HSulf-1 et HSulf-2. De plus, il a été suggéré qu'une multimérisation des HSulfs serait bien médiée par la région C-terminale (domaines HD et C-terminal) et facilitée par l'absence de clivage furine (Tang et Rosen, 2009). Cependant, la cryo-EM n'a pas fait état d'une telle organisation. D'après les modèles produits, la protéine serait monomérique. Toutefois, en inspectant les résultats de la classification 2D, de nombreuses classes montrent deux protéines très proches (Figure 48). Ainsi, ce qui avait été défini comme une proximité induite par la forte concentration de particules dans chaque trou de la grille pourrait, finalement, correspondre à une dimérisation de la protéine. Cependant, d'après ces classes 2D, ces dimères semblent être « lâches », et probablement même dynamiques, car une majorité des particules ne présente aucune proximité entre elles. En revanche, la localisation de

cet élément de dimérisation, au sein de la région hautement flexible du domaine HD, serait en accord avec un concept de dimère « lâche ». Toujours d'après ces classes, la position de cette hélice propice à la dimérisation est, vraisemblablement, bien du côté du domaine HD car un espace significatif est observé entre deux particules dimériques. Or, cet espace peut correspondre au domaine HD qui n'est pas détecté à cause de sa flexibilité et de sa faible structuration (aucun alignement de densité électronique n'est donc possible).

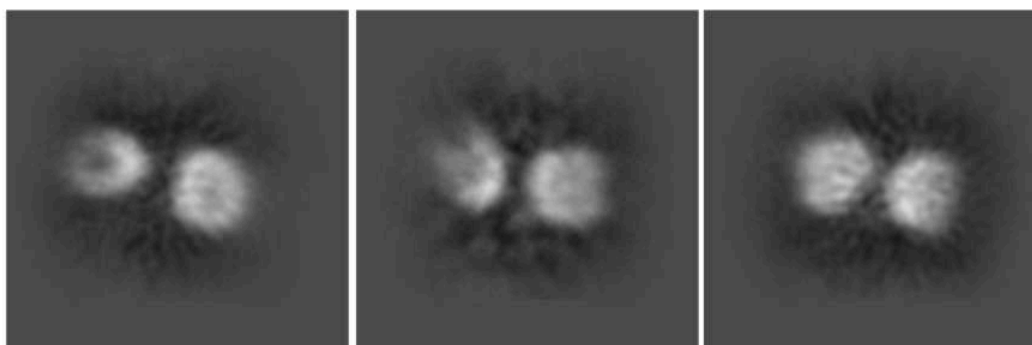


Figure 48 : Classes 2D de la protéine HSulf-2  $\Delta$ SG. Ces classes sont représentatives de nombreuses autres classes 2D montrant une proximité entre deux protéines. Il est supposé que ces classes représentent la forme dimérique de la protéine. Toutefois, ces classes ne peuvent être sélectionnées et utilisées pour la production d'une carte de densité électronique car la flexibilité, la dynamique supposée, de ce dimère génère une résolution trop faible.

De ce fait, identifier une méthode et/ou une construction permettant de favoriser et stabiliser cette dimérisation est une piste majeure à investiguer car, dans ce cas, la structuration serait, vraisemblablement, augmentée, tout comme la taille de l'objet. Par conséquent, cet échantillon protéique deviendrait un objet tout à fait propice à des études cryo-EM.

Pour ce faire, outre la possibilité d'investiguer davantage d'agents réticulants, comme le DSP (dithiobis succinimidyl propionate), le Diméthyl adipimidate (DMA) ou bien le Diméthyl suberimidate (DMS), une stratégie basée sur la substitution de quelques acides aminés pourrait être envisagée. En effet, à la fin des années 1990, un groupe de recherche a réussi, par « computational design », à réaliser la stabilisation du domaine B1 de la protéine streptococcique G grâce à sept substitutions : trois en position centrale (Dahiyat et Mayo, 1997) et quatre partiellement exposées (Malakauskas et Mayo, 1998). En 2009, une étude s'est focalisée sur ces quatre dernières substitutions E15V, T16L, T18I et N37L, et a élucidé la base moléculaire de cette forte stabilisation.

Ainsi, ces substitutions stabilisent la protéine par des contributions comprises entre 1,6 et 6,0 kJ.mol<sup>-1</sup> (à 70°C). Mais, ce sont principalement les résidus hydrophobes aux positions 16 et 37 qui fournissent les contributions majeures à cette stabilisation. De plus, ces deux résidus augmentent significativement la formation de dimères (Thoms *et al.*, 2009). Ainsi, l'élimination de résidus hydrophiles et la malléabilité de la surface hydrophobe créée constituent un socle favorable à la formation d'une structuration accrue de la protéine, et à la formation de dimères stables. Par conséquent, grâce à la structure prédite par AlphaFold, et à l'identification de l'hélice alpha impliquée dans la dimérisation (grâce aux recoupements des données), une substitution de quelques acides aminés hydrophiles par des résidus hydrophobes pourrait significativement améliorer la stabilité du dimère et permettre d'envisager de nouvelles analyses structurales.

D'autres constructions protéiques peuvent aussi être envisagées et analysées afin de répondre à la problématique de cette thèse. En effet, la dernière construction produite (HSulf-2 ΔHDc), où les régions flexibles et non structurées ont été réduites, a été à l'origine d'un accroissement (faible) de la résolution des spectres RMN.

Ainsi, la première construction possible serait le domaine CAT seul (sans le domaine C-terminal). En effet, il se pourrait que sans le domaine HD flexible, ce domaine C-terminal ne se replie pas correctement (pour rappel, il est imaginé que le domaine C-terminal interagit avec le domaine CAT pour former un seul domaine) et génère une déstabilisation globale de la protéine. Dans ce sens, une construction SNAP-CAT (sans C-terminal) a été réalisée (antérieurement à cette thèse). Néanmoins, l'expression quasi-nulle de la protéine correspondante n'a jamais permis d'envisager d'études structurales sur cette construction. Toutefois, il serait intéressant de tenter d'optimiser la construction et de réussir à exprimer le domaine CAT seul, puis de réaliser un marquage (<sup>13</sup>CH<sub>3</sub>-Met en excès) afin de déterminer si les problèmes observés en RMN ne découlent pas, en grande majorité, de ce domaine C-terminal qui ne se replie/s'organise pas correctement avec le domaine CAT.

En parallèle, une deuxième construction, le domaine C-terminal seul, pourrait être conçue et analysée. Ce domaine de petite taille (18 kDa) serait, si l'expression est suffisante, un très bon candidat pour des analyses RMN afin d'identifier si un spectre résolu peut être obtenu. Il serait également une cible protéique de choix pour le développement du milieu de marquage uniforme.

En conclusion, il serait très intéressant de produire et d'analyser par RMN ces deux constructions, car elles permettraient d'identifier avec certitude si c'est la structuration intrinsèque de chaque domaine qui ne permet pas l'obtention de spectres résolus, ou si c'est la conception de la construction HSulf-2  $\Delta$ HD qui pose problème. Notons de même que, si un seul des deux domaines s'avère être résolu par RMN, alors cette information pourrait significativement modifier la compréhension de la structuration générale de HSulf-2 et permettre d'optimiser de futures constructions et protocoles.

Enfin, une dernière alternative serait d'étudier l'orthologue de Sulf exprimé par *C. elegans* : la protéine Sul-1. Sul-1 présente un domaine CAT homologue à HSulf, mais un domaine HD de taille réduite. En effet, le domaine HD de Sul-1 conserve seulement les régions structurées du domaine HD des Sulf humaines (perte de la région centrale non structurée). Sul-1 pourrait donc représenter un candidat intéressant pour des essais de cristallisation et de RMN. De surcroît, des premiers tests réalisés, antérieurement à la thèse, ont montré de bons rendements protéiques (relativement similaires à HSulf-2  $\Delta$ HD). Cependant, suite à ces premiers tests, l'expression stable de cette protéine n'a pas pu être reproduite et les efforts de recherche ont été portés sur l'enzyme HSulf-2. Au vu des difficultés rencontrées avec cette dernière, il pourrait être intéressant de re-transfecter des cellules pour exprimer Sul-1, et réaliser un marquage RMN qui permettrait d'identifier si une protéine naturelle ressemblant fortement à la protéine recombinante HSulf-2  $\Delta$ HDc est, oui ou non, vue comme structurée par RMN. De surcroît, si cet orthologue permettait d'acquérir des données RMN exploitables, il serait intéressant de modifier la construction HSulf-2  $\Delta$ HD et/ou HSulf-2  $\Delta$ HDc en se basant sur les homologies de séquences avec Sul-1.

Pour terminer, une dernière approche prometteuse pour la détermination de la structure de HSulf serait l'utilisation de nanobodies.

Les nanobodies sont des anticorps miniatures (~ 15 kDa) dérivés d'anticorps à chaîne lourde présents chez les camélidés. Au cours de ces dernières années, les nanobodies se sont considérablement développés et avérés être des outils précieux dans le processus de détermination de la structure de protéines difficiles et de grands complexes protéiques, où initialement aucune méthode d'analyse structurale n'était applicable. Il est vrai que les nanobodies donnent accès aux épitopes conformationnels dans de multiples régions (notamment charnières), permettant alors une stabilisation de la

protéine dans une conformation unique. Cela permet d'atteindre une maille parfaitement structurée qui diffracte sous rayons X. Ces nanobodies agissent donc comme de réelles chaperonnes de cristallisation et ont permis, jusqu'à présent, la détermination de nombreuses structures à haute résolution (Traenkle *et al.*, 2016 ; Duhoo *et al.*, 2017 ; Hansen *et al.*, 2017 ; Uchański *et al.*, 2020).

Par conséquent, concevoir un ou plusieurs nanobodies spécifiques à des régions clés de la protéine HSulf-2, comme la région d'interaction CAT-HD, ou bien C-terminal-HD, dans le but de stabiliser les régions flexibles défavorables au processus de cristallisation, fournirait un avantage certain pour la résolution de la structure de notre protéine.

Les nanobodies peuvent également se prêter à des analyses cryo-EM. Cependant, il est apparu que l'utilisation d'un échantillon protéique stabilisé par un nanobody ne fournissait pas de résultats aussi avantageux que ceux observés en cristallographie. En effet, les nanobodies sont trop petits pour influencer favorablement la taille de l'objet étudié et pour permettre aux logiciels de traitement de se focaliser et de réaliser l'alignement des particules sur ces nanobodies pour augmenter significativement la résolution des classifications, et de la carte de densité électronique finale.

Ainsi, pour utiliser les nanobodies en cryo-EM, une approche alternative a été développée. Cette approche fait appel à de nouvelles molécules chimériques appelées mégabodies qui augmentent la masse moléculaire des protéines cibles. Elles peuvent aussi surmonter l'orientation préférentielle des particules dans la glace, améliorant la résolution et la reconstruction finale (Uchański *et al.*, 2021). Ces mégabodies sont en réalité des nanobodies dont la masse a été augmentée artificiellement car greffés de manière rigide sur des « scaffolds » protéiques sélectionnés (HopQ ou YgjK) pour augmenter leur masse moléculaire tout en conservant la spécificité et l'affinité de liaison à l'antigène (Uchański *et al.*, 2021).

Une étude récente décrit une méthode de synthèse de mégabodies basée sur la disponibilité d'un nanobody interagissant avec la protéine cible, qui est ensuite fixé de manière rigide à deux « scaffolds » : (1) un fragment FAB (« Fragment antigen-binding ») d'un anticorps dirigé contre le nanobody, et (2) un nanobody liant la « protein A fragment » fusionné à la maltose binding protein (MBP) et au « Fab-binding domains ». Cet assemblage d'environ 120 kDa ne perturbe pas l'interaction du nanobody avec la protéine cible et facilite l'alignement des particules dans le traitement d'images cryo-EM. Le succès de cette méthode a été démontré avec le récepteur KDE1, une



protéine membranaire de 23 kDa, où une carte résolue à 3,2 Å a été obtenue. Une deuxième protéine de 22 kDa, le domaine de liaison au récepteur de la protéine Spike de SARS-Cov2, a aussi permis l'obtention d'une carte de densité électronique résolue à 3,6 Å et des informations sur les interfaces d'interactions ont pu être obtenues (Wu *et al.*, 2021).

Les voies d'exploration décrites ci-dessus (telles que les méga/nanobodies, l'élimination partielle des N-glycosylations, la stabilisation du dimère et de nouvelles campagnes de cristallogénèse) nécessitent de grands investissements en temps, et cela sans garantie de réussite. A plus court terme, les résultats obtenus au cours de l'étude de la protéine HSulf-2 par cryo-EM méritent d'être approfondis car ils représentent une première étape de vraie valeur vers le calcul de sa structure. Il est vrai qu'un objet aussi petit et intrinsèquement désordonné que HSulf-2 était un défi majeur pour le laboratoire. Ceci étant, les différents modèles produits ont tout de même présenté une résolution très satisfaisante, voire inattendue, au vu de nos connaissances structurales de HSulf-2. Ces analyses ont, également, généré des modèles permettant d'identifier l'organisation tridimensionnelle d'un certain nombre de structures secondaires (ce qui n'avait jamais été observé), dont certaines pourraient être résolues/calculées en modifiant le traitement des données comme le long feuillet  $\beta$  anti-parallèle de dix feuillets. De plus, ce modèle s'est avéré être représentatif de la protéine d'intérêt au vu de la qualité de l'ajustement réalisé entre la structure prédite par AlphaFold (entre autres) et les données expérimentales. Ceci pourrait suggérer que le nouvel outil de prédiction de structures qu'est AlphaFold délivrerait le premier modèle de structure réaliste de la protéine HSulf-2. Par conséquent, grâce à l'ensemble des analyses structurales réalisées, et aux nombreux points clés résolus (notamment dans l'optimisation des protocoles d'expression et de purification), une nouvelle stratégie en deux étapes a pu être dessinée pour la caractérisation de HSulf-2 par cryo-EM.

Tout d'abord, la protéine HSulf-2  $\Delta$ HDC pourrait être exprimée avec un motif glycosidique réduit (expression en cellules HEK 293S) grâce au nouveau protocole d'expression permettant d'obtenir des rendements protéiques significativement améliorés. Il est ainsi imaginé que lors du traitement de données (après préparation et analyse de grilles), les structures limitant l'alignement, et donc la résolution des classes 2D, ne seront plus (aussi) présentes ; d'autant plus que cette construction ne contient

pas la partie interne du domaine HD qui apportait une flexibilité (du fait de ses caractéristiques) et des structures délétères à l'alignement des particules. Par conséquent, bien que cette protéine soit encore plus petite que la protéine HSulf-2  $\Delta$ SG, elle est très probablement davantage globulaire et moins flexible. Cette hypothèse peut être corrélée aux données de RMN montrant ~ 25 % des pics des méthionines résolues pour cette protéine contre aucune pour les autres constructions.

De surcroît, généralement en cryo-EM, les fractions issues de la chromatographie d'exclusion contenant la protéine ne sont pas rassemblées et concentrées, mais un criblage des fractions est fait et celle présentant une qualité protéique optimale est juste diluée avant la préparation des grilles. Or, avec les rendements protéiques initiaux cette méthode (destinée à l'augmentation de la qualité de l'échantillon) n'était pas réalisable, mais le sera dorénavant. Notons également que l'ajout de ligands en excès ou réticulés à la protéine pourrait être réalisé afin d'identifier de possibles changements conformationnels et/ou d'accroître la structuration de la protéine.

Par conséquent, si les résultats obtenus sont ceux escomptés, un premier modèle structural hautement résolu de la protéine pourrait être atteint. Quand au domaine HD, ce dernier fait l'objet actuellement d'une caractérisation structurale par RMN ; où les données préliminaires corroborent la structure du domaine HD prédite par le logiciel AlphaFold.

Pour terminer, la caractérisation totale de l'enzyme pourrait être réalisée à partir des données préexistantes de HSulf-2  $\Delta$ SG ou à partir de nouvelles données acquises sur HSulf-2  $\Delta$ SG modifiée (d'après les différentes hypothèses précédemment décrites). En effet, il a pu être établi que les cartes de densités électroniques n'étaient pas suffisamment résolues pour un calcul de structure, mais l'étaient suffisamment pour un ajustement de la structure ; notamment un ajustement des domaines CAT-C-terminal afin de comprendre l'organisation des domaines, mais aussi l'organisation de l'enzyme dans sa forme dimérique. Sur ce dernier point, il est vrai qu'avec les jeux de données actuels, et en se focalisant sur les particules dimériques, de nouveaux modèles pourraient être produits. Cela apporterait ainsi une compréhension globale de l'enzyme et une possible première structure haute résolution de HSulf-2.

En conclusion, bien que de nombreuses approches puissent être envisagées pour résoudre la structure de la protéine HSulf-2, les récents résultats obtenus en cryo-EM, corrélés aux données de cristallographie et RMN, permettraient d'envisager des résultats

à court terme. En cristallogénèse, les nombreuses voies d'exploration citées précédemment laissent entrevoir un avenir prometteur pour cette technique. En revanche, à ce jour, la RMN n'est plus envisagée comme une méthode de caractérisation structurale de l'enzyme HSulf-2.

## ***Annexes***



## VIII. Annexes

A.

- 24 **MGPPSLVLCLLSATVFSLLGGSSA**FLSHHRLKGRFQRDRRNIRPNIILVLTDD 29  
30 QDVELGSMQVMNKTRRIMEQGGAHFINAFVTT**PMCC**PSRSSILTGKYVHNHNT 82  
83 YTNNE**NC**SSPSWQAQHESTRFAVYLN**ST**GYRTAFFGKYLNEYNGSYVPPGWKE 135  
136 WVGLLKNSRFY**NY**TLCRNG**VKEK**HGSDYSKDYLTDLIT**ND**SVSFFRTSKKMY P 188  
189 HRPVLMVISHAAPHGPEDSAPQYSRLFP**NAS**QHITPSYNYAPNPDKHWIMRYT 241  
242 GPMKPIHMEFTNMLQQRKRLQTLMSVDDSMETIYNMLVETGELDNTYIVYTADH 294  
295 GYHIGQFGLVKGKSMPYEFDIRVPFYVRGPNVEAGCLNPHIVLNIDLAPTILD 347  
348 IAGLDIPADMDGKSIKLLDTERPVNRFH**LKKK**MRVWRDSFLVERGKLLHKRD 400  
401 NDKVDAQEENFLPKYQ**RV**KDLCQRAEYQ**TACE**QLGQKWQ**CV**EDATGK**LKLHKC** 453  
454 KGPMRLGGSRALSNLVPKYYGQGSE**ACTC**DSGDYKLSLAGRRK**KL**FKKKYKAS 506  
507 YVRSRSI**RS**VAIEVDGRVYHVGLGDA**AQ**PRNLTKRHWP**GAP**EDQDDKDG**DFS** 559  
560 GTGGLPDYSAANPIKVTHRCYILE**ND**TVQCDLDLYKSLQAWKD**HL**HDHEIE 612  
613 TLQNKIKNLREVRGHLKKKR**PEE**CD**CHK**ISYHTQHKGRLK**HR**GSSLHPFRKGL 665  
666 QEKDKVWLLREQKRKK**KL**RKLLKRLQ**ND**TCSMPGLTCFTHDNQHWQ**TAP**FWT 718  
719 LGPFCACTSANNNTYWCMRTINETHN**FL**CFEFATGFLEYFDLNTDPYQLMNAV 771  
772 NTLDRDVLNQLHVQLMELRSCKGYKQ**C**NPRTRNMDLGLKDGGSYEQYRQ**FQ**RR 824  
825 KWPEMKRPSSKSLGQLWEGWEG 846

Peptide signal

Domaine HD (avant : Domaine CAT ; après : Domaine C-terminal)

Cysteine convertie en Fgly

N-glycosyations

Sites d'interaction aux HP

Ponts disulfures

Clivage furine

O-glycosylation

## B. HSulf-2 WT et HSulf-2 ΔSG :

RSDKDCEMKRRTTLDSP LGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLMQATAW  
LNAYFHQPEIEEFPVPALHHPVFQQESFTRQVLWKLKLVVKFGEVISYQQLAALAGNPAATAAVKT  
ALSGNPVPILIPCHRVS SSGAVGGYEGGLAVKEWLLAHEGHRLGKPG LGPAGIGAPGSLGGGSEN  
LYFQSFLSHHRLKGRFQRDRRNIRPNIILVLTDDQDVELGSMQVMNKTRRIMEQGGAHFINAFVTT  
PMCCPSRSSILTGKYVHNHNTYTNNENCSSPSWQAQHE SRTFAVYLNSTGYRTAFFGKYLNEYNGS  
YVPPGWKEWVGLLKN SRFYNYTLCRNGVKEKHGSDYSKDYLTDLITNDSVSFFRTSKKMYPHRPVL  
MVISHAAPHGPEDSAPQYSRLFPNASQHITPSYNYAPNPKHWIMRYTGPMKPIHMEFTNMLQRKR  
LQTLMSVDDSMETIYNMLVETGELDNTYIVYTADHGYHIGQFGLVKGKSMPYEFDIRVPPFYVRGPN  
VEAGCLNPHIVLNIDLAPTILDIAGLDIPADMDGKSILKLLDTERPVNRFHLK KMRVWRDSFLVE  
RGKLLHKRDNDKVD AQEENFLPKYQRVKDLCQRAEYQTACEQLGQKWQCVEDATGKLLKHKCKGPM  
RLGGSRALSNLVPKYYGQSEACTCDSGDYKLSLAGRRK KLFK KKYKASYVRSRSIRSV AIEVDGR  
VYHVGLGDA AQPRNLTKRHWP GAPEDQDDKGGDFSGTGGLPDYSAANPIKVTHRCYILENDTVQC  
DLDLYKSLQAWKDHKLHIDHEIETLQNKIKNLREVRGHLK KKRPEECDCHKISYHTQHKGRLKHRG  
SSLHPFRKGLQEKDKVWLLREQKRKKLRKLLKRLQNN DTC SMPGLTCFTHDNQHWQTAPFWTLGP  
FCACTSANNTY WCMRTINETHNFLFC EFATGFLEYFDLNTDPYQLMNAVNTLDRDVLNQLHVQLM  
ELRSCKGYKQCNPRTRNMDLGLKDGGSYEQYRQFQR RKWPEMKRPSSKSLGQLWEGWEGGGGSEN  
LYFQSHHHHHH

## C. HSulf-2 ΔHD :

RSDKDCEMKRRTTLDSP LGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLMQATAW  
LNAYFHQPEIEEFPVPALHHPVFQQESFTRQVLWKLKLVVKFGEVISYQQLAALAGNPAATAAVKT  
ALSGNPVPILIPCHRVS SSGAVGGYEGGLAVKEWLLAHEGHRLGKPG LGPAGIGAPGSLGGGSEN  
LYFQSFLSHHRLKGRFQRDRRNIRPNIILVLTDDQDVELGSMQVMNKTRRIMEQGGAHFINAFVTT  
PMCCPSRSSILTGKYVHNHNTYTNNENCSSPSWQAQHE SRTFAVYLNSTGYRTAFFGKYLNEYNGS  
YVPPGWKEWVGLLKN SRFYNYTLCRNGVKEKHGSDYSKDYLTDLITNDSVSFFRTSKKMYPHRPVL  
MVISHAAPHGPEDSAPQYSRLFPNASQHITPSYNYAPNPKHWIMRYTGPMKPIHMEFTNMLQRKR  
LQTLMSVDDSMETIYNMLVETGELDNTYIVYTADHGYHIGQFGLVKGKSMPYEFDIRVPPFYVRGPN  
VEAGCLNPHIVLNIDLAPTILDIAGLDIPADMDGKSILKLLDTERPVNRFHLK KMRVWRDSFLVE  
NNDTCSMPGLTCFTHDNQHWQTAPFWTLGPFCACTSANNTY WCMRTINETHNFLFC EFATGFLEY  
FDLNTDPYQLMNAVNTLDRDVLNQLHVQLMELRSCKGYKQCNPRTRNMDLGLKDGGSYEQYRQFQR  
RKWPEMKRPSSKSLGQLWEGWEGGGGSENLYFQSHHHHHH

## D. HSulf-2 ΔHD V1 :

LENPLLTGLSKLIRLTIGRPKLASTMKLCILLAVVAFVGLSLPTALARSKD KDCEMKRRTTLDSP LGK  
LELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPAL  
HHPVFQQESFTRQVLWKLKLVVKFGEVISYQQLAALAGNPAATAAVKTALSGNPVPILIPCHRVS  
SSGAVGGYEGGLAVKEWLLAHEGHRLGKPG LGPAGIGAPGSLGGGSENLYFQSSGFLSHHRLKGRF

~~QRDRRNIRPNI~~ILVLTDDQDVELGSMQVMNKTRRIMEQGGAHFINAFVTTMCCPSRSSILTGKYV  
HNHNTYTNNENCSSPSWQAQHESTRFAVYLNSTGYRTAFFGKYLNEYNGSYVPPGWKEWVGLLKNS  
RFYNYTLCRNGVKEKHGSDYSKDYLTDLITNDSVSFFRTSKKMYPHRPVLMVISHAAPHGPEDSAP  
QYSRLFPNASQHITPSYNYAPNPKHWIMRYTGPMKPIHMEFTNMLQRKRLQTLMSVDDSMETIYN  
MLVETGELDNTYIVYTADHGYHIGQFGLVKGKSMPEYFDIRVPFYVRGPNVEAGCLNPHIVLNIDL  
APTILDIAGLDIPADMDGKSILKLLDTERPVNRFHLKKKMRVWRDSFLVENNDTCSMPGLTCFTHD  
NQHWTAPFWTLGPFCACTSANNNTYWCMRTINETHNFLFCEFATGFLEYFDLNTDPYQLMNAVNT  
LDRDVLNQLHVQLMELRSCKGYKQCNPRTNRNMDLGLKDCGSYEQYRQFQRRKWPEMKRPSSKSLGQ  
LWECWEGGENLYFQGPGGGSHHHHHH

**E. HSulf-2 ΔHD V2 :**

LENPLLTGLSKLIRLTIGRPKLASTMKLCILLAVVAFVGLSLPTALARSDKDCEMKRRTTLDSP LGK  
LELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPAL  
HHPVFQQESFTRQVLWKLKLVVKFGEVVISYQQLAALAGNPAATAAVKTALSGNPVPIILIPCHR  
VSSGAVGGYEGGLAVKEWLLAHEGHRLGKPGLPAGIGAPGSLGFLSHHRLKGRFQDRRNIRPNI  
ILVLTDDQDVELGSMQVMNKTRRIMEQGGAHFINAFVTTMCCPSRSSILTGKYVHNHNTYTNNENC  
SSPSWQAQHESTRFAVYLNSTGYRTAFFGKYLNEYNGSYVPPGWKEWVGLLKNSRFYNYTLCRNGV  
KEKHGSDYSKDYLTDLITNDSVSFFRTSKKMYPHRPVLMVISHAAPHGPEDSAPQYSRLFPNASQH  
ITPSYNYAPNPKHWIMRYTGPMKPIHMEFTNMLQRKRLQTLMSVDDSMETIYNMLVETGELDNTY  
IVYTADHGYHIGQFGLVKGKSMPEYFDIRVPFYVRGPNVEAGCLNPHIVLNIDLAPTILDIAGLDI  
PADMDGKSILKLLDTERPVNRFHLKKKMRVWRDSFLVENNDTCSMPGLTCFTHDNQHWTAPFWTL  
GPFCACTSANNNTYWCMRTINETHNFLFCEFATGFLEYFDLNTDPYQLMNAVNTLDRDVLNQLHVQ  
LMELRSCKGYKQCNPRTNRNMDLGLKDCGSYEQYRQFQRRKWPEMKRPSSKSLGQLWECWEGGGGSH  
HHHHH

**F. HSulf-2 ΔHDc :**

MKLCILLAVVAFVGLSLPTALARSDKDCEMKRRTTLDSP LGKLELSGCEQGLHEIKLLGKGTSAADA  
VEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLKLVVKF  
GEVVISYQQLAALAGNPAATAAVKTALSGNPVPIILIPCHR  
VSSGAVGGYEGGLAVKEWLLAHEGHRLGKPGLPAGIGAPGSLGGGSENLYFQSDIFLSHHRLKGRFQDRRNIRPNIILVLTDDQDVELGS  
MQVMNKTRRIMEQGGAHFINAFVTTMCCPSRSSILTGKYVHNHNTYTNNENCSSPSWQAQHESTR  
FAVYLNSTGYRTAFFGKYLNEYNGSYVPPGWKEWVGLLKNSRFYNYTLCRNGVKEKHGSDYSKDYLT  
DLITNDSVSFFRTSKKMYPHRPVLMVISHAAPHGPEDSAPQYSRLFPNASQHITPSYNYAPNPK  
HWIMRYTGPMKPIHMEFTNMLQRKRLQTLMSVDDSMETIYNMLVETGELDNTYIVYTADHGYHIGQ  
FGLVKGKSMPEYFDIRVPFYVRGPNVEAGCLNPHIVLNIDLAPTILDIAGLDIPADMDGKSILKLL  
DTERPVNRFHLKKKMRVWRDSFLVERGKLLHKRDNDKVD AQEENFLPKYQRVKDLCQRAEYQTACE  
QLGQKWQCVEDATGKLLHKCKGPMRLGGSRALS NLVPKYYGQGSEACTCDSSGGGSCDCHKISYH  
TQHKGR LKHGSSLHPFRKGLQEKDKVWLLREQKRKKKLRKLLKRLQNNDTCSMPGLTCFTHDNQH  
WTAPFWTLGPFCACTSANNNTYWCMRTINETHNFLFCEFATGFLEYFDLNTDPYQLMNAVNTLDR



DVLNQLHVQLMELRSCKGYKQCNPRTRNMDLGLKDGGSYEQYRQFQRRKWPEMKRPSSKSLGQLWE  
GWEGPGENLYFQGPGGGSHHHHHH

Annexe 1 : Séquence en acide aminé de la protéine HSulf-2. (A) Séquence annotée de HSulf-2. (B) Séquence de la construction HSulf-2 WT et HSulf-2  $\Delta$ SG avec les deux étiquettes de purification (vert), les sites de coupure à la protéine TEV (orange) et la protéine HSulf-2 (les domaines CAT et C-terminal en noir, le domaine HD en noir souligné). Remarque, la mutation **S559A** est présente pour HSulf-2  $\Delta$ SG afin d'éliminer la chaîne de CS. (C) Même nomenclature. Pour cette construction, la séquence correspondante au domaine HD a été tronquée. (D) et (E) Ces deux constructions (HSulf-2  $\Delta$ HD V1 et HSulf-2  $\Delta$ HD V2) découlent de la construction HSulf-2  $\Delta$ HD. Ainsi, dans un souci de compréhension des modifications réalisées, la séquence de HSulf-2  $\Delta$ HD a été réécrite mais avec trois portions de séquences tronquées correspondant aux modifications (séquences barrées). Notons l'absence de site de coupure à la TEV pour HSulf-2  $\Delta$ HD V2. (F) Séquence de la construction HSulf-2  $\Delta$ HDc, comprenant une partie du domaine HD (séquences soulignées) et un linker **SGSGGS** remplaçant la séquence tronquée du domaine HD comprise entre D<sub>483</sub> et C<sub>636</sub> :

SGDYKLSLAGRRKFLFKKYYKASYVRSRSIRSVAI EVDGRVYHVGLGDAAQPRNLTKRHWPGAPEDQDDKD  
GGDFSGTGGLPDYSAANPIKVTHRCYILENDTVQCDLDLYKSLQAWKDHKLHI DHEIETLQNKIKNLREVR  
GHLKKRPEE.

Conditions	1	2	Temoin +	3	4	5	6	7
Volume culture	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL
Milieu de culture	ESF-SFM déplété	ESF-SFM déplété	FreeStyle 293	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété
Concentration des mélanges réactionnels pour reconstituer le milieu	1x	1x	-	1x	1x	1x	1x	1x
Autres composés supplémentés	Vitamine 5x*	-	-	Vitamine 5x* et 10 g/L glucose**	8 g/L glucose**	4 g/L glucose**	2 g/L glucose**	0,5 g/L glucose**
L-Tyrosine	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	-	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)

Annexe 2 : Optimisation de la croissance cellulaire et des rendements protéiques par ajout de vitamines et de glucose. \* Vitamines obtenues après dilution de la solution de vitamine MEM (100X), Gibco®. \*\* Solution stock de glucose à 100 g/L dissous dans de l'eau distillée, pH 8 et stérilisée par filtration (0,22 µm, sous hotte en environnement stérile).

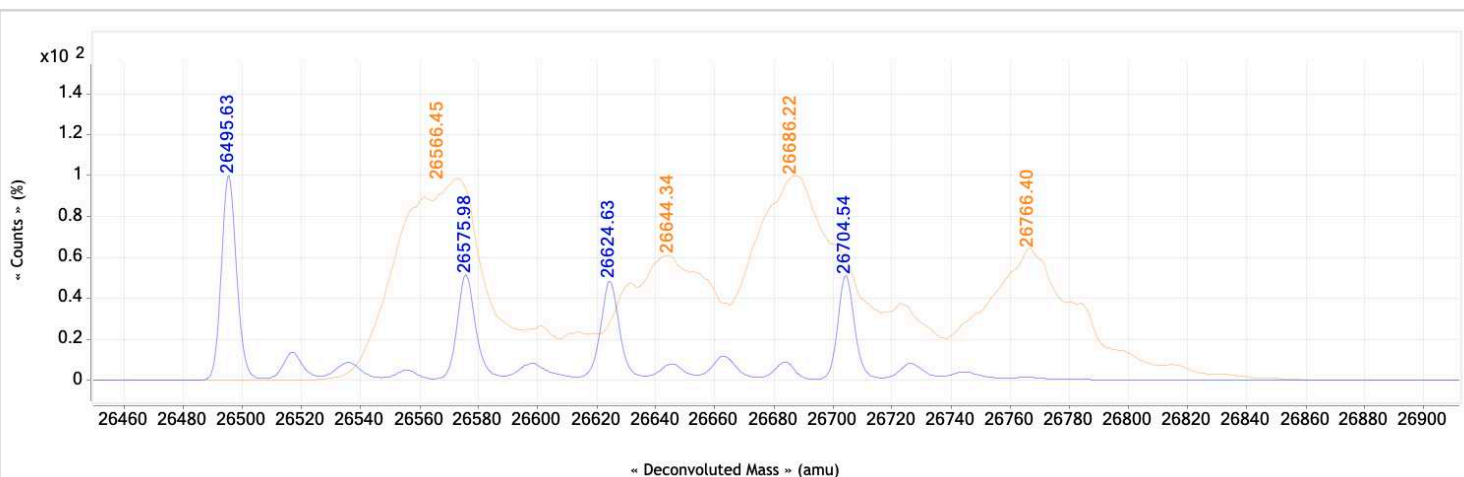
Conditions	1	2	3	4	5	Témoin -
Volume culture	4 mL	4 mL	4 mL	4 mL	4 mL	4 mL Cellules non transfectées
Milieu de culture	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	FreeStyle 293
Concentration des mélanges réactionnels pour reconstituer le milieu	1x	1x	1x	1x	1x	-
Autres composés supplémentés	-	1 g/L extrait de levure*	1 g/L extrait de levure dialysé**	4 g/L extrait de levure*	4 g/L extrait de levure dialysé**	-
L-Tyrosine	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	500 µL (soit 0,5 mg)	-

Annexe 3 : Optimisation de la croissance cellulaire et des rendements protéiques par ajout d'extrait de levure. \* Extrait de levure (Yeast Extracted, référence 70161, Sigma Aldrich) \*\* Solution stock à 100 g/L d'extrait de levure dialysé pendant 2 fois 5 heures avec un rapport 1/1000 (5 mL extrait de levure dans 5 L d'eau distillée). Après dialyse, le pH de la solution est ajusté à 8 et la solution stérilisée par filtration (0,22 µm, sous hotte en environnement stérile).

Conditions	Hsp90 N-terminal marquée : 140 mg/L <sup>13</sup> CH <sub>3</sub> -Ala	Hsp90 N-terminal marquée : 100 mg/L <sup>13</sup> CH <sub>3</sub> -Ala	Hsp90 N-terminal marquée : 80 mg/L <sup>13</sup> CH <sub>3</sub> -Ala	Hsp90 N-terminal marquée : 60 mg/L <sup>13</sup> CH <sub>3</sub> -Ala
Volume culture	100 mL	100 mL	100 mL	100 mL
Milieu de culture	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété	ESF-SFM déplété
L-Tyrosine*	130 mg/L	130 mg/L	130 mg/L	130 mg/L
Concentration des mélanges réactionnels EAA, NEAA, Ile, Val, Leu et Thr	1x	1x	1x	1x
Concentration en acides aminés marqués <sup>13</sup> CH <sub>3</sub> -Met*	200 mg/L	200 mg/L	200 mg/L	200 mg/L
Concentration en acides aminés marqués <sup>13</sup> CH <sub>3</sub> -Ala*	140 mg/L	100 mg/L	80 mg/L	60 mg/L
L-cyclo-sérine* (inhibiteur des alanines transaminases)	5 mg/L	5 mg/L	5 mg/L	5 mg/L

Annexe 4 : Composition des milieux de marquage spécifique pour l'établissement de la courbe d'incorporation de la <sup>13</sup>CH<sub>3</sub>-Ala. \* Chaque constituant a été prélevé (solution) ou pesé (poudre) puis dissous dans 30 mL de milieu de culture (ESF-SFM déplété), avant que le pH ne soit ajusté à 8 et la solution stérilisée par filtration (0,22 µm, sous hotte en environnement stérile).

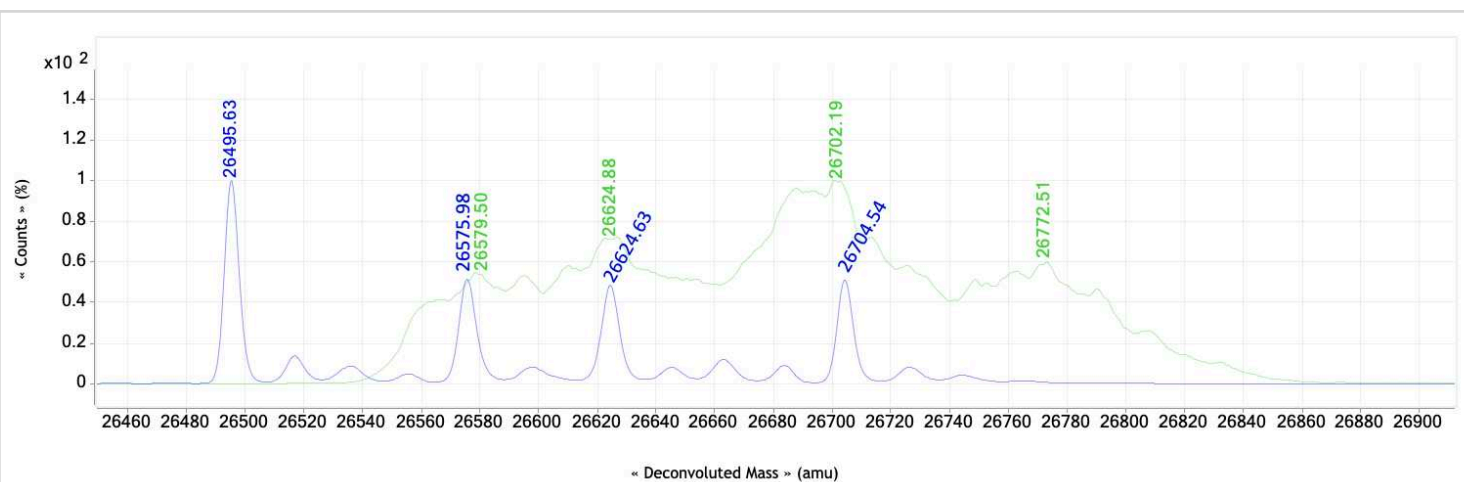
G. Hsp90 N-terminal marquée (orange) – [<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N]-Thr : 17 Thréonines dans la séquence, + 5 Da par [<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N]-Thr incorporée (+ Superposition de Hsp90 N-terminal non marquée - bleu).



Traitement du spectre :

	Différence de masse attendue pour 100 % d'incorporation	Pic 1	Pic 2	Pic 3	Pic 4
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Thr : Différence de masse observée (Da) – Taux d'incorporation calculé (%)	+ 85 Da	+ 70,82 Da – 83 %	+ 68,36 Da – 80 %	+ 61,59 Da – 72 %	+ 61,86 Da – 72 %

H. Hsp90 N-terminal marquée (vert) – [<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N]-Ile : 20 Isoleucines dans la séquence, + 7 Da par [<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N]-Ile incorporée (+ Superposition de Hsp90 N-terminal non marquée - bleu).



Traitement du spectre :

	Différence de masse attendue pour 100 % d'incorporation	Pic 1	Pic 2	Pic 3	Pic 4
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Ile : Différence de masse observée (Da) – Taux d'incorporation calculé (%)	+ 140 Da	+ 83,87 Da – 60 %	+ 48,9 Da – 35 %	+ 77,56 Da – 55 %	+ 67,97 Da – 48 %

Annexe 5 : (G) et (H) Spectres ESI-TOF de la protéine Hsp90 N-terminal marquée sur un acide aminé portant un groupement méthyle et tableaux du traitement des spectres avec le résultat du taux d'incorporation pour chaque pic.

A. Hsp90 N-terminal :

1 MHHHHHHDQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLRELISNSSD 53  
 54 ALDKIRYESLTDPSKLDGKELHINLIPNKQDRTLTIIVDTGIGMTKADLINNL 106  
 107 GTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEKVTVITKHNDDEQY 159  
 160 AWESSAGGSFTVRTDTGEPMGRGTKVILHLKEDQTEYLEERRIKEIVKKHSQF 212  
 213 IGYPIITLFVEKERDKEVSDDEAE 235

B.

Marquage réalisé de la protéine Hsp90 N-terminal	Nombre de résidus au sein de la séquence protéique	Différence de masse pour un résidu incorporé	Différence de masse attendue pour 100 % d'incorporation
<sup>13</sup> CH <sub>3</sub> -Met	7	+ 1 Da	+ 7 Da
<sup>13</sup> CH <sub>3</sub> -Ala	15	+ 1 Da	+ 15 Da
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Ile	20	+ 7 Da	+ 140 Da
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Val	11	+ 6 Da	+ 66 Da
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Thr	17	+ 5 Da	+ 85 Da
[ <sup>1</sup> H, <sup>13</sup> C, <sup>15</sup> N]-Leu	18	+ 7 Da	+ 126 Da

Annexe 6 : (A) Séquence en acide aminé de la protéine Hsp90 N-terminal. (B) Masses ajoutées par le marquage de la protéine Hsp90 N-terminal.

Liste des publications :

- **Crétinon, Y.**, Awad, R., Kerfah, R., Vivès, R.R., and Crublet, E. Development of a new culture medium for methyl specific labeling of proteins in HEK293 cells. *J. Biol NMR*. Publication en rédaction.
- El Masri, R., Seffouh, A., Roelants, C., Seffouh, I., Gout, E., Pérard, J., Dalonneau, F., Nishitsuji, K., Noborn, F., Nikpour, F., Larson, G., **Crétinon, Y.**, Friedel-Arboleas, M., Uchimura, K., Daniel, R., Lortat-Jacob, H., Filhol, O., and Vivès, R.R. (2022). Extracellular endosulfatase Sulf-2 harbors a chondroitin/dermatan sulfate chain that modulates its enzyme activity. *Cell rep.* **38**, 110516.
- El Masri, R., **Crétinon, Y.**, Gout, E., Vivès, R.R. (2020). HS and Inflammation: A Potential Playground for the Sulfs ? *Front Immunol.* **11**, 570. Doi: 10.3389/fimmu.2020.00570.
- Annaval, T., Wild, R., **Crétinon, Y.**, Sadir, R., Vivès, R.R., and Lortat-Jacob, H. (2020). Heparan Sulfate Proteoglycans Biosynthesis and Post Synthesis Mechanisms Combine Few Enzymes and Few Core Proteins to Generate Extensive Structural and Functional Diversity. *Molecules*, **25**, 4215. Doi: 10.3390/molecules25184215.
- Crublet, E., Kerfah, R., Imbert, L., Awad, R., Ayala, I., **Crétinon, Y.**, Gans, P., and Boisbouvier, Methyl Specific labeling approaches for the NMR investigation of challenging proteins. *J. Advanced Publication en rédaction.*

## ***Remerciements***





## IX. Remerciements

Je voudrais remercier, dans un premier temps, les membres de jury d'avoir pris le temps d'évaluer mon travail de thèse.

Cette thèse n'aurait pu voir le jour et aboutir sans de nombreuses personnes qui ont contribué, en m'aidant et en m'épaulant, de près ou de loin, à la réalisation de celle-ci. Je voudrais ici leur exprimer mes sincères remerciements.

En premier lieu, je tiens à exprimer toute ma gratitude à mes directeurs de thèse, Romain Vivès et Elodie Crublet, sans qui ce travail n'aurait été possible. Je sais que cette thèse a été compliquée par les nombreux obstacles rencontrés mais si aujourd'hui je garde un très beau souvenir et plaisir de cette thèse, c'est grâce à vos qualités humaines et la bienveillance que vous avez eue envers moi. De ce fait, je vous remercie pour votre aide, votre implication et votre confiance. Apprendre à vos côtés a été une chance dont je suis pleinement conscient.

Egalement, je tiens à remercier l'ensemble des personnes du groupe SAGAG (présent et passé) et de NMR-bio, sans qui bien des journées auraient été moroses. Je vous remercie donc pour votre gentillesse, vos conseils, votre aide précieuse et pour toutes les choses que vous m'avez apportées et qui m'ont fait grandir ; vous resterez dans ma mémoire et avez toute mon amitié.



## ***Bibliographie***



## X. Bibliographie

- Abiatari, I., Kleeff, J., Li, J., Felix, K., Büchler, M.W., and Friess, H. (2006). Hsulf-1 regulates growth and invasion of pancreatic cancer cells. *J. Clin. Pathol.* **59**, 1052-1058.
- Adhikari, N., Rusch, M., Mariash, A., Li, Q., Selleck, S.B., and Hall, J.L. (2008). Alterations in Heparan Sulfate in the Vessel in Response to Vascular Injury in the Mouse. *J Cardiovasc. Transl. Res.* **1**, 236-240.
- Afratis, N.A., Nikitovic, D., Mulhaupt, H.A.B., Theocharis, A.D., Couchman, J.R., and Karamanos, N.K. (2017a). Syndecans - key regulators of cell signaling and biological functions. *FEBS J.* **284**, 27-41.
- Afratis, N.A., Karamanou, K., Piperigkou, Z., Vynios, D.H., and Theocharis, A.D. (2017b). The role of heparins and nano-heparins as therapeutic tool in breast cancer. *Glycoconj J.* **34**, 299-307.
- Ahmad, M., Hirz, M., Pichler, H., and Schwab, H. (2014). Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production, *Appl. Microbiol. Biotechnol.* **98**, 5301-5317.
- Ai, X., Do, A.-T., Lozynska, O., Kusche-Gullberg, M., Lindahl, U., and Emerson, C.P. (2003). QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling. *J. Cell. Biol.* **162**, 341-351.
- Ai, X., Do, A.-T., Kusche-Gullberg, M., Lindahl, U., Lu, K., and Emerson, C.P. (2006). Substrate Specificity and Domain Functions of Extracellular Heparan Sulfate 6-O-Endosulfatases, QSulf1 and QSulf2. *J. Biol. Chem.* **281**, 4969-4976.
- Ai, X., Kitazawa, T., Do, A.-T., Kusche-Gullberg, M., Labosky, P.A., and Emerson, C.P. (2007). SULF1 and SULF2 regulate heparan sulfate-mediated GDNF signaling for esophageal innervation. *Development* **134**, 3327-3338.
- Aikawa, J., and Esko, J.D. (1999). Molecular Cloning and Expression of a Third Member of the Heparan Sulfate/Heparin GlcNAcN-Deacetylase/ N-Sulfotransferase Family. *J. Biol. Chem.* **274**, 2690-2695.
- Aikawa, J., Grobe, K., Tsujimoto, M., and Esko, J.D. (2001). Multiple Isozymes of Heparan Sulfate/Heparin GlcNAcN-Deacetylase/GlcN N-Sulfotransferase structure and activity of the fourth member NDST4. *J. Biol. Chem.* **276**, 5876-5882.
- Alex, J.M., Rennie, M.L., Engilberge, S., Lehoczki, G., Dorotyya, H., Fizil, Á., Batta, G., and Crowley, P.B. (2019). Calixarene-mediated assembly of a small antifungal protein. *IUCrJ.* **6**, 238-247.
- Alexopoulou, A.N., Mulhaupt, H.A.B., and Couchman, J.R. (2007). Syndecans in wound healing, inflammation and vascular biology. *Int. J. Biochem. Cell Biol.* **39**, 505-528.
- Ali, R., Clark, L., Zahm, J.A., Lemoff, A., Ramesh, K., Rosenbaum, D.M., and Rosen, M.K. (2019). Improved strategy for isoleucine 1H/13C methyl labeling in *Pichia pastoris*. *J. Biomol. NMR.* **73**, 687-697.
- Ambasta, R.K., Ai, X., and Emerson, C.P. (2007). Quail Sulf1 Function Requires Asparagine-linked Glycosylation. *J. Biol. Chem.* **282**, 34492-34499.
- Annaval, T., Wild, R., Créton, Y., Sadir, R., Vivès, R.R., and Lortat-Jacob, H. (2020). Heparan Sulfate Proteoglycans Biosynthesis and Post Synthesis Mechanisms Combine Few Enzymes and Few Core Proteins to Generate Extensive Structural and Functional Diversity. *Molecules*, **25**, 4215
- Ashikari-Hada, S., Habuchi, H., Kariya, Y., Itoh, N., Reddi, A.H., and Kimata, K. (2004). Characterization of Growth Factor-binding Structures in Heparin/Heparan Sulfate Using an Octasaccharide Library. *J. Biol. Chem.* **279**, 12346-12354.
- Assenberg, R., Wan, P.T., Geisse, S., and Mayr, L.M. (2013). Advances in recombinant protein expression for use in pharmaceutical research. *Current Opinion in Structural Biology* **23**, 393-402.
- Asundi, V.K., Erdman, R., Stahl, R.C., and Carey, D.J. (2003). Matrix metalloproteinase-dependent shedding of syndecan-3, a transmembrane heparan sulfate proteoglycan, in Schwann cells. *J. Neurosci.*

Res. **73**, 593-602.

Austin, J.W., Gilchrist, C., Fehlings, M.G. (2012). High molecular weight hyaluronan reduces lipopolysaccharide mediated microglial activation. *J. Neurochem.* **112**, 344-355.

Axelsson, J., Xu, D., Na Kang, B., Nussbacher, J.K., Handel, T.M., Ley, K., Sriramarao, P., and Esko, J.D. (2012). Inactivation of heparan sulfate 2-O-sulfotransferase accentuates neutrophil infiltration during acute inflammation in mice. *Blood* **120**, 1742-1751.

Ayala, I., Sounier, R., Use, N., Gans, P., and Boisbouvier, J. (2009). An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR.* **43**, 111-119.

Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M.J, Gans, P., and Boisbouvier, J. (2012). An optimized isotopic labelling strategy of isoleucine-gamma2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem Commun (Camb)* **48**, 1434-1436.

Backliwal, G., Hildinger, M., Hasija, V., and Wurm, F.M. (2008). High-density transfection with HEK-293 cells allows doubling of transient titers and removes need for a priori DNA complex formation with PEI. *Biotechnol Bioeng.* **99**, 721-7.

Ballabio, A., Parenti, G., Carozzo, R., *et al.* (1987). Isolation and characterization of a steroid sulfatase cDNA clone: genomic deletions in patients with X-chromosome-linked ichthyosis. *Proc Natl Acad Sci USA.* **84**, 4519-4523.

Bame, K.J., Lidholt, K., Lindahl, U., and Esko, J.D. (1991a). Biosynthesis of heparan sulfate. Coordination of polymer-modification reactions in a Chinese hamster ovary cell mutant defective in N-sulfotransferase. *J. Biol. Chem.* **266**, 10287-10293.

Bame, K.J., Reddy, R.V., and Esko, J.D. (1991b). Coupling of N-deacetylation and N-sulfation in a Chinese hamster ovary cell mutant defective in heparan sulfate N-sulfotransferase. *J. Biol. Chem.* **266**, 12461-12468.

Bame, K.J., Zhang, L., David, G., and Esko, J.D. (1994). Sulphated and undersulphated heparan sulphate proteoglycans in a Chinese hamster ovary cell mutant defective in N-sulphotransferase. *Biochem. J.* **303**, 81-87.

Bar-Ilan, A., Balan, V., Tittmann, K., Golbik, R., Vyazmensky, M., Hubner, G., Barak, Z., and Chipman, D.M. (2001). Binding and activation of thiamin diphosphate in acetohydroxyacid synthase. *Biochemistry* **40**, 11946-11954.

Barak, Z., Chipman, D.M., and Gollop, N. (1987). Physiological implications of the specificity of acetohydroxy acid synthase isozymes of enteric bacteria. *J Bacteriol.* **169**, 3750-3756.

Baranyi, L., Doering, C.B., Denning, G., Gautney, R.E., Harris, K.T., Spencer, H.T., ... and Dropulic, B. (2013). Rapid generation of stable cell lines expressing high levels of erythropoietin, factor VIII, and an antihuman CD20 antibody using lentiviral vectors. *Human Gene Therapy Methods* **24**, 214-227.

Bartlett, A.H., and Park, P.W. (2010). Proteoglycans in host-pathogen interactions: molecular mechanisms and therapeutic implications. *Expert Rev. Mol. Med.* **12**, e5.

Bartolini, B., Thelin, M.A., Svensson, L., Ghiselli, G., van Kuppevelt, T.H., Malmström, A., *et al.* (2013). Iduronic Acid in Chondroitin/dermatan Sulfate Affects Directional Migration of Aortic Smooth Muscle Cells. *PLoS One* **8**, e66704.

Bethea, H.N., Xu, D., Liu, J., and Pedersen, L.C. (2008). Redirecting the substrate specificity of heparan sulfate 2-O-sulfotransferase by structurally guided mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18724-18729.

Bernhardt, R.R., and Schachner, M. (2000). Chondroitin Sulfates Affect the Formation of the Segmental Motor Nerves in Zebrafish Embryos. *Dev. Biol.* **221**, 206-219.

Berninsone, P., and Hirschberg, C.B. (1998). Heparan Sulfate/Heparin N-Deacetylase/N-sulfotransferase the N-sulfotransferase activity domain is at the carboxyl half of the holoenzyme. *J. Biol. Chem.* **273**, 25556-25559.

- Bezakova, G., and Ruegg, M.A. (2003). New insights into the roles of agrin. *Nat Rev Mol Cell Biol.* **4**, 295-308.
- Biederbick, A., Licht, A., and Kleene, R. (2003). Serglycin proteoglycan is sorted into zymogen granules of rat pancreatic acinar cells. *Eur. J. Cell. Biol.* **82**, 19-29.
- Bielicki, J., and Hopwood, J.J. (1991). Human liver N-acetylgalactosamine 6-sulphatase. Purification and characterization. *Biochem. J.* **279**, 515-520.
- Bilong, M., Bayat, P., Bourderioux, M., Jérôme, M., Giuliani, A., and Daniel, R. (2021). Mammal hyaluronidase activity on chondroitin sulfate and dermatan sulfate: Mass spectrometry analysis of oligosaccharide products. *Glycobiology* **31**, 751-761.
- Bishop, J.R., Schuksz, M., and Esko, J.D. (2007). Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* **446**, 1030-37.
- Blackhall, F.H., Merry, C.L.R., Lyon, M., Jayson, G.C., Folkman, J., Javaherian, K., and Gallagher, J.T. (2003). Binding of endostatin to endothelial heparan sulphate shows a differential requirement for specific sulphates. *Biochem. J.* **375**, 131-139.
- Boltes, I., Czapinska, H., Kahnert, A., von Bülow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001). 1.3 A structure of arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. *Structure.* **9**, 483-91.
- Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., and Guss, J.M. (1997). Structure of a human lysosomal sulfatase. *Structure* **5**, 277-89.
- Bredrup, C., Knappskog, P.M., Majewski, J., Rødahl, E., and Boman, H. (2005). Congenital stromal dystrophy of the cornea caused by a mutation in the decorin gene. *Invest. Ophthalmol. Vis. Sci.* **46**, 420-426.
- Bruggert, M., Rehm, T., Shanker, S., Georgescu, J., and Holak, T.A. (2003). A novel medium for expression of proteins selectively labeled with 15 N-amino acids in *Spodoptera frugiperda* (Sf9) insect cells. *J Biomol NMR.* **25**, 335-348.
- Buczek-Thomas, J.A., Hsia, E., Rich, C.B., Foster, J.A., and Nugent, M.A. (2008). Inhibition of histone acetyltransferase by glycosaminoglycans. *J. Cell. Biochem.* **105**, 108-120.
- Bullock, S.L., Fletcher, J.M., Beddington, R.S.P., and Wilson, V.A. (1998). Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase. *Genes Dev.* **12**, 1894-1906.
- Bülow, H.E., and Hobert, O. (2004). Differential sulfations and epimerization define heparan sulfate specificity in nervous system development. *Neuron* **41**, 723-736.
- Buraschi, S., Neill, T., and Iozzo, R.V. (2019). Decorin is a devouring proteoglycan: Remodeling of intracellular catabolism via autophagy and mitophagy. *Matrix Biol.* **75-76**, 260-270.
- Busse, M., Feta, A., Presto, J., Wilén, M., Grønning, M., Kjellén, L., and Kusche-Gullberg, M. (2007). Contribution of EXT1, EXT2, and EXTL3 to Heparan Sulfate Chain Elongation. *J. Biol. Chem.* **282**, 32802-32810.
- Byrne, B. (2015). *Pichia pastoris* as an expression host for membrane protein structural biology. *Curr. Opin. Struct. Biol.* **32**, 9-17.
- Cadwallader, A.B., and Yost, H.J. (2006). Combinatorial expression patterns of heparan sulfate sulfotransferases in zebrafish: I. The 3-O-sulfotransferase family. *Dev. Dyn.* **235**, 3423-3431.
- Canet, D., Bouguet-Bonnet, S., and Leclerc, S. (2010). *RMN express : Spin<sup>1/2</sup>, état liquide*. Presses Universitaires de Nancy.
- Capurro, M.I., Xiang, Y.-Y., Lobe, C., and Filmus, J. (2005). Glypican-3 Promotes the Growth of Hepatocellular Carcinoma by Stimulating Canonical Wnt Signaling. *Cancer Res.* **65**, 6245-6254.



- Capurro, M.I., Xu, P., Shi, W., Li, F., Jia, A., and Filmus, J. (2008). Glypican-3 inhibits Hedgehog signaling during development by competing with patched for Hedgehog binding. *Dev. Cell* **14**, 700-711.
- Cardin, A.D., and Weintraub, H.J. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arterioscler. Dallas Tex* **9**, 21-32.
- Carlson, E.D., Gan, R., Hodgman, C.E., and Jewett, M.C. (2012). Cell-free protein synthesis: applications come of age. *Biotechnol Adv.* **30**, 1185-1194.
- Carlsson, P., and Kjellén, L. (2012). Heparin biosynthesis. *Handb. Exp. Pharmacol.* **207**, 23-41.
- Caterson, B., and Melrose, J. (2018). Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* **28**, 182-206.
- Celie, J.W.A.M., Rutjes, N.W.P., Keuning, E.D., Soininen, R., Heljasvaara, R., Pihlajaniemi, T., Dräger, A.M., Zweegman, S., Kessler, F.L., Beelen, R.H.J., *et al.* (2007). Subendothelial heparan sulfate proteoglycans become major L-selectin and monocyte chemoattractant protein-1 ligands upon renal ischemia/reperfusion. *Am. J. Pathol.* **170**, 1865-1878.
- Chang, M.-C., Chen, C.-A., Hsieh, C.-Y., Lee, C.-N., Su, Y.-N., Hu, Y.-H., and Cheng, W.-F. (2009). Mesothelin inhibits paclitaxel-induced apoptosis through the PI3K pathway. *Biochem. J.* **424**, 449-458.
- Chau, B.N., Diaz, R.L., Saunders, M.A., Cheng, C., Chang, A.N., Warrenner, P., Bradshaw, J., Linsley, P.S., and Cleary, M.A. (2009). Identification of SULF2 as a novel transcriptional target of p53 by use of integrated genomic analyses. *Cancer Res.* **69**, 1368-1374.
- Chen, J., and Liu, J. (2005). Characterization of the structure of antithrombin-binding heparan sulfate generated by heparan sulfate 3-O-sulfotransferase 5. *Biochim Biophys Acta, Gen Subj.* **1725**, 190-200.
- Chen, L., and Sanderson, R.D. (2009). Heparanase Regulates Levels of Syndecan-1 in the Nucleus. *PLOS ONE* **4**, e4947.
- Chhabra, M., Doherty, G.G., See, N.W., Gandhi, N.S., and Ferro, V. (2021). From Cancer to COVID-19: A Perspective on Targeting Heparan Sulfate-Protein Interactions. *Chem Rec.* **21**, 3087-3101.
- Choi, S., Choi, Y., Jun, E., Kim, I.S., Kim, S.E., Jung, S.A., and Oh, E.S. (2015). Shed syndecan-2 enhances tumorigenic activities of colon cancer cells. *Oncotarget.* **6**, 3874-3886.
- Cole, C.L., Rushton, G., Jayson, G.C., and Avizienyte, E. (2014). Ovarian cancer cell heparan sulfate 6-O-sulfotransferases regulate an angiogenic program induced by heparin-binding epidermal growth factor (EGF)-like growth factor/EGF receptor signaling. *J. Biol. Chem.* **289**, 10488-10501.
- Conrad, A.H., Zhang, Y., Tasheva, E.S., and Conrad, G.W. (2010). Proteomic analysis of potential keratan sulfate, chondroitin sulfate A, and hyaluronic acid molecular interactions. *Invest. Ophthalmol. Vis. Sci.* **51**, 4500-4515.
- Cosma, M.P., Pepe, S., Annunziata, I., Newbold, R.F., Grompe, M., Parenti, G., and Ballabio, A. (2003). The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* **113**, 445-456.
- Couchman, J.R., Gopal, S., Lim, H.C., Nørgaard, S., and Multhaupt, H.A.B. (2015). Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. *Int. J. Exp. Pathol.*, **96**, 1-10.
- Clark, L., Zahm, J.A., Ali, R., Kukula, M., Bian, L., Patrie, S.M., Gardner, K.H., Rosen, M.K., and Rosenbaum, D.M. (2015). Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. *J. Biomol. NMR.* **62**, 239-245.
- Clark, L.D., Dikiy, I., Chapman, K., Rödström, K.E., Aramini, J., LeVine, M.V., Khelashvili, G., Rasmussen, S.G., Gardner, K.H., and Rosenbaum, D.M. (2017). Ligand modulation of sidechain dynamics in a wild-type human GPCR. *ELife* **6**, e28505.
- Clark, L., Dikiy, I., Rosenbaum, D.M., and Gardner, K.H. (2018). On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies. *J. Biomol. NMR.* **71**, 203-211.

- Cregg, J.M., Barringer, K.J., Hessler, A.Y., and Madden, K.R. (1985). *Pichia pastoris* as a host system for transformations, *Mol. Cell. Biol.* **5**, 3376-3385.
- Cregg, J.M., Tolstorukov, I, Kusari, A., Sunga, J., Madden, K., and Chappell, T. (2009). Expression in the yeast *Pichia pastoris*, *Methods Enzymol.* **463**, 169-189.
- Csoka, A.B., Frost, G.I., and Stern, R. (2001). The six hyaluronidase-like genes in the human and mouse genomes. *Matrix Biol.* **20**, 499-508.
- Dai, Y., Yang, Y., MacLeod, V., Yue, X., Rapraeger, A.C., Shriver, Z., Venkataraman, G., Sasisekharan, R., and Sanderson, R.D. (2005). HSulf-1 and HSulf-2 Are Potent Inhibitors of Myeloma Tumor Growth in Vivo. *J. Biol. Chem.* **280**, 40066-40073.
- Danesin, C., Agius, E., Escalas, N., Ai, X., Emerson, C., Cochard, P., and Soula, C. (2006). Ventral Neural Progenitors Switch toward an Oligodendroglial Fate in Response to Increased Sonic Hedgehog (Shh) Activity: Involvement of Sulfatase 1 in Modulating Shh Signaling in the Ventral Spinal Cord. *J. Neurosci.* **26**, 5037-5048.
- Dang, X., Zhu, Q., Wang, L., Su, H., Lin, H., Zhou, N., Liang, T., Wang, Z., Huang, S., Ren, Q., *et al.* (2009). Macular corneal dystrophy in a Chinese family related with novel mutations of CHST6. *Mol. Vis.* **15**, 700-705.
- Davami, F., Barkhordari, F., Alebouyeh, M., Adeli, A., and Mahboudi, F. (2011). Combined TGE-SGE expression of novel PAI-1-resistant t-PA in CHO DG44 cells using orbitally shaking disposable bioreactors. *J Microbiol Biotechnol.* **21**, 1299-1305.
- Davami, F., Eghbalpour, F., Barkhordari, F., and Mahboudi, F. (2014). Effect of Peptone Feeding on Transient Gene Expression Process in CHO DG44. *Avicenna J Med Biotechnol.* **6**, 147-155.
- De Cat, B., and David, G. (2001). Developmental roles of the glypicans. *Semin Cell Dev Biol.* **12**, 117-125.
- De Rossi, G., Evans, A.R., Kay, E., Woodfin, A., McKay, T. R., Nourshargh, S., and Whiteford, J.R. (2014). Shed syndecan-2 inhibits angiogenesis. *J. Cell Sci.* **127**, 4788-4799.
- Debarnot, C., Monneau, Y.R., Roig-Zamboni, V., Delauzun, V., Le Narvor, C., Richard, E., Hénault, J., Goulet, A., Fadel, F., Vivès, R.R., *et al.* (2019). Substrate binding mode and catalytic mechanism of human heparan sulfate d-glucuronyl C5 epimerase. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6760-6765.
- Delafosse, L., Xu, P., and Durocher, Y. (2016). Comparative study of polyethylenimines for transient gene expression in mammalian HEK293 and CHO cells. *Journal of Biotechnology* **227**, 103-111.
- Deligny, A., Denys, A., Marcant, A., Melchior, A., Mazurier, J., van Kuppevelt, T.H., and Allain, F. (2010). Synthesis of heparan sulfate with cyclophilin B-binding properties is determined by cell type-specific expression of sulfotransferases. *J Biol Chem.* **285**, 1701-1715.
- Deligny, A., Dierker, T., Dagälv, A., Lundequist, A., Eriksson, I., Nairn, A.V., Moremen, K.W., Merry, C.L.R., and Kjellén, L. (2016). NDST2 (N-Deacetylase/N-Sulfotransferase-2) Enzyme regulates heparan sulfate chain length. *J. Biol. Chem.* **291**, 18600-18607.
- Delmotte, P., Delmotte-Plaquee, J.A. (1953). New Antifungal Substance of Fungal Origin. *Nature* **171**, 344.
- Dhoot, G. K. *et al.* (2001). Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. *Science* **293**, 1663-6.
- Dick, G., Tan, C.L., Alves, J.N., Ehlert, E.M.E., Miller, G.M., Hsieh-Wilson, L.C., Sugahara, K., Oosterhof, A., van Kuppevelt, T.H., Verhaagen, J., *et al.* (2013). Semaphorin 3A binds to the perineuronal nets via chondroitin sulfate type E motifs in rodent brains. *J. Biol. Chem.* **288**, 27384- 27395.
- Dierks, T., Schmidt, B., and von Figura, K. (1997). Conversion of cysteine to formylglycine: a protein modification in the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11963-11968.
- Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M.A., and von Figura, K. (1998a). Posttranslational Formation of Formylglycine in Prokaryotic Sulfatases by Modification of Either Cysteine or Serine. *J. Biol. Chem.* **273**, 25560-25564.

- Dierks, T., Lecca, M.R., Schmidt, B., and von Figura, K. (1998b). Conversion of cysteine to formylglycine in eukaryotic sulfatases occurs by a common mechanism in the endoplasmic reticulum. *FEBS Lett.* **423**, 61-65.
- Dierks, T., Lecca, M.R., Schlotterhose, P., Schmidt, B., and von Figura, K. (1999). Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J.* **18**, 2084-2091.
- DiGabriele, A.D., Lax, I., Chen, D.I., Svahn, C.M., Jaye, M., Schlessinger, J., and Hendrickson, W.A. (1998). Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **393**, 812.
- Dikiy, I., Clark, L.D., Gardner, K.H., and Rosenbaum, D.M. (2019). Chapter Two - Isotopic Labeling of Eukaryotic Membrane Proteins for NMR Studies of Interactions and Dynamics. in: A.J. Wand (Ed.), *Methods Enzymol.*, Academic Press, pp. 37-65.
- Djerbal, L., Lortat-Jacob, H., and Kwok, J. (2017). Chondroitin sulfates and their binding molecules in the central nervous system. *Glycoconj. J.* **34**, 363-376.
- Do, A.-T., Smeds, E., Spillmann, D., and Kusche-Gullberg, M. (2006). Overexpression of heparan sulfate 6-O-sulfotransferases in human embryonic kidney 293 cells results in increased N-acetylglucosaminyl 6-O-sulfation. *J. Biol. Chem.* **281**, 5348-5356.
- Dominguez-Gutierrez, P.R., Kwenda, E.P., Donelan, W., O'Malley, P., Crispen, P.L., and Kusmartsev, S. (2021). Hyal2 Expression in Tumor-Associated Myeloid Cells Mediates Cancer-Related Inflammation in Bladder Cancer. *Cancer Res.* **81**, 648-657.
- Dou, W., Xu, Y., Pagadala, V., Pedersen, L.C., and Liu, J. (2015). Role of Deacetylase Activity of N-Deacetylase/N-Sulfotransferase 1 in Forming N-sulfated Domain in Heparan Sulfate. *J. Biol. Chem.* **290**, 20427-20437.
- Doverskog, M., Han, L., and Häggström, L. (1998). Cystine/cysteine metabolism in cultured Sf9 cells: Influence of cell physiology on biosynthesis, amino acid uptake and growth. *Cytotechnology* **26**, 91-102.
- Drews, M., Doverskog, M., Öhman, L., Chapman, B.E., Jacobsson, U., Kuchel, P.W., *et al.* (2000). Pathways of glutamine metabolism in *Spodoptera frugiperda* (Sf9) insect cells: Evidence for the presence of the nitrogen assimilation system, and a metabolic switch by  $^1\text{H}/^{15}\text{N}$  NMR. *Journal of Biotechnology* **78**, 23-37.
- Dubey, A., Stoyanov, N., Viennet, T., Chhabra, S., Elter, S., Borggräfe, J., Viegas, A., Nowak, R.P., Burdzhiyev, N., Petrov, O., Fischer, R.S., Etzkorn, M., Gelev, V., and Arthanari H. (2021). Local Deuteration Enables NMR Observation of Methyl Groups in Proteins from Eukaryotic and Cell-Free Expression Systems. *Angew. Chem. Int. Ed.* **60**, 13783-13787
- Duhoo, Y., Roche, J., Trinh, T.T.N., Desmyter, A., Gaubert, A., Kellenberger, C., Cambillau, C., Roussel, A., and Leone, P. (2017). Camelid nanobodies used as crystallization chaperones for different constructs of PorM, a component of the type IX secretion system from *Porphyromonas gingivalis*. *Acta Crystallogr F Struct Biol Commun.* **1**, 286-293.
- Edavettal, S.C., Lee, K.A., Negishi, M., Linhardt, R.J., Liu, J., and Pedersen, L.C. (2004). Crystal structure and mutational analysis of heparan sulfate 3-O-sulfotransferase isoform 1. *J. Biol. Chem.* **279**, 25789-25797.
- Egorova-Zachernyuk, T.A., Bosman, G.J.C.G.M., Pistorius, A.M.A., and DeGrip, W.J. (2009). Production of yeastolates for uniform stable isotope labelling in eukaryotic cell culture, *Appl. Microbiol. Biotechnol.* **84**, 575-581.
- Egorova-Zachernyuk, T.A., Bosman, G.J.C.G.M., and DeGrip, W.J. (2011). Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. *Appl. Microbiol. Biotechnol.* **89**, 397-406.
- El Masri, R., Seffouh, A., Lortat-Jacob, H., and Vivès, R.R. (2017). The “in and out” of glucosamine 6-O-sulfation: the 6th sense of heparan sulfate. *Glycoconj. J.* **34**, 285-298.

- El Masri, R., Créton, Y., Gout, E., Vivès, R.R. (2020). HS and Inflammation: A Potential Playground for the Sulfs? *Front Immunol.* **11**, 570.
- El Masri, R., Seffouh, A., Roelants, C., Seffouh, I., Gout, E., Pérard, J., Dalonneau, F., Nishitsuji, K., Noborn, F., Nikpour, F., Larson, G., Créton, Y., Friedel-Arboleas, M., Uchimura, K., Daniel, R., Lortat-Jacob, H., Filhol, O., and Vivès, R.R. (2022). Extracellular endosulfatase Sulf-2 harbors a chondroitin/dermatan sulfate chain that modulates its enzyme activity. *Cell rep.* **38**, 110516.
- Elsheereef1, A.A., Jochums, A., Lavrentieva, A., Stuckenberg, L., Scheper, T., and Solle, D. (2019). High Cell Density Transient Transfection of CHO Cells for TGF-B1 expression. *Eng Life Sci.* **19**, 730-740.
- Emami, S., Fan, Y., Munro, R., Ladizhansky, V., and Brown, L.S. (2013). Yeast-expressed human membrane protein aquaporin-1 yields excellent resolution of solid-state MAS NMR spectra, *J. Biomol. NMR.* **55**, 147-155.
- Esko, J.D., and Zhang, L. (1996). Influence of core protein sequence on glycosaminoglycan assembly. *Curr. Opin. Struct. Biol.* **6**, 663-670.
- Esko, J.D., Selleck, S.B. (2002). Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71**, 435-71.
- Esko, J.D., Kimata, K., and Lindahl, U. (2009). Proteoglycans and Sulfated Glycosaminoglycans. In *Essentials of Glycobiology*. 2nd edition (Varki A, C. R., Esko JD, Freeze H, Hart G, Marth J, ed.), Cold Spring Harbor Laboratory Press.
- Faham, S., and Hileman, R. e. (1996). Heparin structure and interactions with basic fibroblast growth factor. *Science* **271**, 1116.
- Fan, G., Xiao, L., Cheng, L., Wang, X., Sun, B., and Hu, G. (2000). Targeted disruption of NDST-1 gene leads to pulmonary hypoplasia and neonatal respiratory distress in mice. *FEBS Lett.* **467**, 7-11.
- Fan, Y., Shi, L., Ladizhansky, V., and Brown, L.S. (2011). Uniform isotope labeling of a eukaryotic seven-transmembrane helical protein in yeast enables high-resolution solid-state NMR studies in the lipid environment, *J. Biomol. NMR.* **49**, 151-161.
- Fan, Y., Emami, S., Munro, R., Ladizhansky, V., and Brown, L.S. (2015). Isotope labeling of eukaryotic membrane proteins in yeast for solid-state NMR. *Methods Enzymol.* **565**, 193-212.
- Fedarko, N.S., and Conrad, H.E. (1986). A unique heparan sulfate in the nuclei of hepatocytes: structural changes with the growth state of the cells. *J. Cell Biol.* **102**, 587-599.
- Felber, M., Pichler, H., and Ruth, C. (2014). Strains and molecular tools for recombinant protein production in *Pichia pastoris*. *Methods Mol. Biol. Clifton NJ.* **1152**, 87-111.
- Ferreras, C., Rushton, G., Cole, C.L., Babur, M., Telfer, B.A., Kuppevelt, T.H. van, Gardiner, J.M., Williams, K.J., Jayson, G.C., and Avizienyte, E. (2012). Endothelial Heparan Sulfate 6-O-sulfation Levels Regulate Angiogenic Responses of Endothelial Cells to Fibroblast Growth Factor 2 and Vascular Endothelial Growth Factor. *J. Biol. Chem.* **287**, 36132-36146.
- Fey, J., Balleininger, M., Borissenko, L.V., Schmidt, B., von Figura, K., and Dierks, T. (2001). Characterization of posttranslational formylglycine formation by luminal components of the endoplasmic reticulum. *J. Biol. Chem.* **276**, 47021-47028.
- Feyerabend, T.B., Jin-Ping Li, Lindahl, U., and Rodewald, H.-R. (2006). Heparan sulfate C5- epimerase is essential for heparin biosynthesis in mast cells. *Nat. Chem. Biol.* **2**, 195-196.
- Feyzi, E., Lustig, F., Fager, G., Spillmann, D., Lindahl, U., and Salmivirta, M. (1997). Characterization of heparin and heparan sulfate domains binding to the long splice variant of platelet- derived growth factor A chain. *J. Biol. Chem.* **272**, 5518-5524.
- Filmus, J., and Capurro, M. (2014). The role of glypicans in Hedgehog signaling. *Matrix Biol. J. Int. Soc. Matrix Biol.* **35**, 248-252.

- Fischer, M., Kloiber, K., Hausler, J., Ledolter, K., Konrat, R., and Schmid, W. (2007). Synthesis of a <sup>13</sup>C-methyl- group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *Chembiochem.* **8**, 610-612.
- Forsberg, E., Pejler, G., Ringvall, M., Lunderius, C., Tomasini-Johansson, B., Kusche-Gullberg, M., Eriksson, I., Ledin, J., Hellman, L., and Kjellen, L. (1999). Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme. *Nature* **400**, 773-776.
- Fransson, L.A., Cheng, F., Yoshida, K., Heinegård, D., Malmström, A., and Schmidtchen, A. (1993). "Patterns of Epimerization and Sulphation in Dermatan Sulphate Chains," in *Dermatan Sulphate Proteoglycans: Chemistry, Biology, Chemical Pathology*. Editor J. E. Scott (London: Portland Press), 11-25.
- Franzmann, E.J., Schroeder, G.L., Goodwin, W.J., Weed, D.T., Fisher, P., and Lokeshwar, V.B. (2003). Expression of tumor markers hyaluronic acid and hyaluronidase (HYAL1) in head and neck tumors. *Int J Cancer* **106**, 438-45.
- Freeman, S.D., Keino-Masu, K., Masu, M., and Ladher, R.K. (2015). Expression of the heparan sulfate 6-O-endosulfatases, Sulf1 and Sulf2, in the avian and mammalian inner ear suggests a role for sulfation during inner ear development. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **244**, 168-180.
- Frese, M.-A., Miltz, F., Dick, M., Lamanna, W.C., and Dierks, T. (2009). Characterization of the human sulfatase Sulf1 and its high affinity heparin/heparan sulfate interaction domain. *J. Biol. Chem.* **284**, 28033-28044.
- Funderburgh, J.L. (2002). Keratan sulfate biosynthesis. *IUBMB Life* **54**, 187-194.
- Gallagher, J.T. (1996). Heparan sulphate proteoglycans: the control of cell growth. In *Extracellular Matrix* (Cooper, W. D., ed.). pp. 230-245.
- Gallagher, J. (2015). Fell-Muir Lecture: Heparan sulphate and the art of cell regulation: a polymer chain conducts the protein orchestra. *Int. J. Exp. Pathol.* **96**, 203-231.
- Gans, P., Hamelin, O., Sounier, R., Ayala, I., Dura, M.A., Amero, C.D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M.J. and Boisbouvier, J. (2010). Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl.* **49**, 1958-1962.
- Gans, P., Boisbouvier, J., Ayala, I., and Hamelin, O. (2010b). Brevet Procédé de marquage isotopique spécifique de groupements méthyle de val, leu et ile. Commissariat A L'énergie Atomique Et Aux Energies Alternatives, Centre National De La Recherche Scientifique and U. J. Fourier. France.
- Gardner, K.H. and Kay, L.E. (1997). Production and Incorporation of <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H (<sup>1</sup>H- $\delta$ 1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *Journal of the American Chemical Society* **119**, 7599-7600.
- Garrett, D.S., Seok, Y.J. Peterkofsky, A. Clore, G.M., and Gronenborn, A.M. (1997). Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system. *Biochemistry* **36**, 4393-4398.
- Gelis, I., Bonvin, A.M., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C.G. (2007). Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR Cell. **131**, 756-769.
- Gifford, J.L., Ishida, H., and Vogel, H.J. (2011). Fast methionine-based solution structure determination of calcium-calmodulin complexes. *J Biomol NMR.* **50**, 71-81.
- Gómez Toledo, A., Sorrentino, J.T., Sandoval, D.R., Malmström, J., Lewis, N.E., and Esko, J.D. (2021). A Systems View of the Heparan Sulfate Interactome. *J Histochem Cytochem.* **69**, 105-119.
- Gondelaud, F., and Ricard-Blum, S. (2019). Structures and interactions of syndecans. *FEBS J.* **286**, 2994-3007.
- Gondelaud, F., Bouakil, M., Le Fèvre, A., Miele, A.E., Chirot, F., Duclos, B., Liwo, A., and Ricard-Blum, S. (2021). Extended disorder at the cell surface: The conformational landscape of the ectodomains of syndecans. *Matrix Biol Plus.* **12**, 100081.

- Gopal, S. (2020). Syndecans in inflammation at a glance. *Front. Immunol.* **11**, 227.
- Gossert, A.D., Hinniger, A., Gutmann, S., Jahnke, W., Strauss, A., and Fernández, C. (2011). A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. *J. Biomol. NMR.* **51**, 449-456.
- Gossert, A.D. and Jahnke, W. (2012). Isotope labeling in insect cells. *Adv Exp Med Biol.* **992**, 179-196.
- Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C., and Kay, L.E. (1999). A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta$  1) methyl-protonated  $^{15}\text{N}$ -,  $^{13}\text{C}$ -,  $^2\text{H}$ -labeled proteins. *J Biomol NMR.* **13**, 369-374.
- Grant, D.S., Yenisey, C., Rose, R.W., Tootell, M., Santra, M. and Iozzo, R.V. (2002). Decorin suppresses tumor cell-mediated angiogenesis. *Oncogene.* **21**, 4765-4777.
- Grigorieva, E., Eshchenko, T., Rykova, V.I., Chernakov, A., Zabarovsky, E., and Sidorov, S.V. (2008). Decreased expression of human D-glucuronyl C5-epimerase in breast cancer. *Int. J. Cancer* **122**, 1172-1176.
- Grigorieva, E.V. (2011). D-Glucuronyl C5-epimerase suppresses small-cell lung cancer cell proliferation in vitro and tumour growth in vivo. *Br. J. Cancer* **9**.
- Grootjans, J.J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Durr, J., and David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. *Proc Natl Acad Sci U.S.A.* **94**, 13683-13688.
- Gross, J.D., Gelev, V.M., and Wagner, G. (2003). A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. *J Biomol NMR.* **25**, 235-242.
- Gubbiotti, M.A., Vallet, S.D., Ricard-Blum, S., and Iozzo, R.V. (2016). Decorin interacting network: A comprehensive analysis of decorin-binding partners and their versatile functions. *Matrix Biol.* **55**, 7-21.
- Gubbiotti, M.A., Neill, T., and Iozzo, R.V. (2017). A Current View of Perlecan in Physiology and Pathology: A Mosaic of Functions. *Matrix Biol.* **57-58**, 285-298.
- Gulberti, S., Lattard, V., Fondeur, M., Jacquinet, J.C., Mulliert, G., Netter, P., Magdalou, J., Ouzzine, M., and Fournel-Gigleux, S. (2005). Phosphorylation and sulfation of oligosaccharide substrates critically influence the activity of human beta1,4-galactosyltransferase 7 (GalT-I) and beta1,3-glucuronosyltransferase I (GlcAT-I) involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J Biol Chem.* **280**, 1417-1425.
- Gustafsson, R., Stachtea, X., Maccarana, M., Grotting, E., Eklund, E., Malmström, A., *et al.* (2014). Dermatan Sulfate Epimerase 1 Deficient Mice as a Model for Human Abdominal wall Defects. *Birth Defects Res. A: Clin. Mol. Teratology* **100**, 712-720.
- Habuchi, H., Habuchi, O., and Kimata, K. (1995). Purification and characterization of heparan sulfate 6-sulfotransferase from the culture medium of chinese hamster ovary cells. *J. Biol. Chem.* **270**, 4172-4179.
- Habuchi, O. (2000). Diversity and functions of glycosaminoglycan sulfotransferases. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1474**, 115-127.
- Habuchi, H., Tanaka, M., Habuchi, O., Yoshida, K., Suzuki, H., Ban, K., and Kimata, K. (2000). The occurrence of three isoforms of heparan sulfate 6-O-sulfotransferase having different specificities for hexuronic acid adjacent to the targeted N-sulfoglucosamine. *J. Biol. Chem.* **275**, 2859-2868.
- Habuchi, H., Miyake, G., Nogami, K., Kuroiwa, A., Matsuda, Y., Kusche-Gullberg, M., Habuchi, O., Tanaka, M., and Kimata, K. (2003). Biosynthesis of heparan sulphate with diverse structures and functions: two alternatively spliced forms of human heparan sulphate 6-O-sulphotransferase-2 having different expression patterns and properties. *Biochem. J.* **371**, 131-142.
- Habuchi, H., Nagai, N., Sugaya, N., Atsumi, F., Stevens, R.L., and Kimata, K. (2007). Mice deficient in heparan sulfate 6-O-sulfotransferase-1 exhibit defective heparan sulfate biosynthesis, abnormal placentation, and late embryonic lethality. *J. Biol. Chem.* **282**, 15578-15588.

- Hagner-McWhirter, A., Lindahl, U., and Li, J. (2000). Biosynthesis of heparin/heparan sulphate: mechanism of epimerization of glucuronyl C-5. *Biochem. J.* **347**, 69-75.
- Hagner-McWhirter, Å., Li, J.-P., Oscarson, S., and Lindahl, U. (2004). Irreversible Glucuronyl C5-epimerization in the Biosynthesis of Heparan Sulfate. *J. Biol. Chem.* **279**, 14631-14638.
- Hajduk, P.J., Augeri, D.J., Mack, J., Mendoza, R., Yang, J., Betz, S.F., and Fesik, S.W. (2000). NMR-Based Screening of Proteins Containing <sup>13</sup>C-Labeled Methyl Groups. *J. Am. Chem. Soc.* **122**, 7898-7904.
- Hamatsu, J., O'Donovan, D., Tanaka, T., Shirai, T., Hourai, Y., Mikawa, T., *et al.* (2013). High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *Journal of the American Chemical Society* **135**, 1688-1691.
- Hammond, E., Khurana, A., Shridhar, V., and Dredge, K. (2014). The Role of Heparanase and Sulfatases in the Modification of Heparan Sulfate Proteoglycans within the Tumor Microenvironment and Opportunities for Novel Cancer Therapeutics. *Front. Oncol.* **4**.
- Hansen, A.P., Petros, A.M., Mazar, A.P., Pederson, T.M., Rueter, A., and Fesik, S.W. (1992). A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. *Biochemistry* **31**, 12713-12718.
- Hansen, S.B., Laursen, N.S., Andersen, G.R., and Andersen, K.R. (2017). Introducing site-specific cysteines into nanobodies for mercury labelling allows de novo phasing of their crystal structures. *Acta Crystallographica Section D Structural Biology*, **73** (10).
- Hanson, S.R., Best, M.D., and Wong, C.-H. (2004). Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed Engl.* **43**, 5736-5763.
- Harder, A., Möller, A.-K., Milz, F., Neuhaus, P., Walhorn, V., Dierks, T., and Anselmetti, D. (2015). Catch bond interaction between cell-surface sulfatase Sulf1 and glycosaminoglycans. *Biophys. J.* **108**, 1709-1717.
- Hassing, H.C., Mooij, H., Guo, S., Monia, B.P., Chen, K., Kulik, W., Dallinga-Thie, G.M., Nieuwdorp, M., Stroes, E.S.G., and Williams, K.J. (2012). Inhibition of hepatic Sulf2 in vivo: a novel strategy to correct diabetic dyslipidemia. *Hepatology*, **55**, 1746-1753.
- Hayano, S., Kurosaka, H., Yanagita, T., Kalus, I., Milz, F., Ishihara, Y., Islam, M.N., Kawanabe, N., Saito, M., Kamioka, H., *et al.* (2012). Roles of heparan sulfate sulfation in dentinogenesis. *J. Biol. Chem.* **287**, 12217-12229.
- Henrotin, Y., Mathy, M., Sanchez, C., and Lambert, C. (2010). Chondroitin Sulfate in the Treatment of Osteoarthritis: From in Vitro Studies to Clinical Recommendations. *Ther. Adv. Musculoskelet. Dis.* **2**, 335-348.
- Hernandez-Guzman, F.G., Higashiyama, T., Pangborn, W., Osawa, Y., and Ghosh, D. (2003). Structure of human estrone sulfatase suggests functional roles of membrane association. *J. Biol. Chem.* **278**, 22989-97.
- Hirano, K., Sasaki, N., Ichimiya, T., Miura, T., Van Kuppevelt, T.H., and Nishihara, S. (2012). 3-O-sulfatéd heparan sulfate recognized by the antibody HS4C3 contribute to the differentiation of mouse embryonic stem cells via Fas signaling. *PLoS One* **7**, e43440.
- Holmborn, K., Ledin, J., Smeds, E., Eriksson, I., Kusche-Gullberg, M., and Kjellén, L. (2004). Heparan Sulfate Synthesized by Mouse Embryonic Stem Cells Deficient in NDST1 and NDST2 Is 6- O-sulfatéd but Contains No N-sulfatéd Groups. *J. Biol. Chem.* **279**, 42355-42358.
- Holst, C.R., Bou-Reslan, H., Gore, B.B., Wong, K., Grant, D., Chalasani, S., Carano, R.A., Frantz, G.D., Tessier-Lavigne, M., Bolon, B., *et al.* (2007). Secreted sulfatases Sulf1 and Sulf2 have overlapping yet essential roles in mouse neonatal survival. *PLoS One* **2**, e575.
- Hudak, A., Kusz, E., Domonkos, I., Josvay, K., Kodamullil, A.T., Szilak, L., Hofmann-Apitius, M., and Letoha, T. (2019). Contribution of syndecans to cellular uptake and fibrillation of a-synuclein and tau. *Sci. Rep.* **9**, 16543.

- Hwa, V., and Salyers, A.A. (1992). Analysis of two chondroitin sulfate utilization mutants of *Bacteroides thetaiotaomicron* that differ in their abilities to compete with the wild type in the gastrointestinal tracts of germfree mice. *Appl. Environ. Microbiol.* **58**, 869 - 876.
- Hwang, H.Y., Olson, S.K., Esko, J.D., and Horvitz, R.H. (2003). *Caenorhabditis elegans* Early Embryogenesis and Vulval Morphogenesis Require Chondroitin Biosynthesis. *Nature* **423**, 439-443.
- Ihrcke, N.S., Parker, W., Reissner, K.J. and Platt, J.L. (1998). Regulation of platelet heparanase during inflammation: role of pH and proteinases. *J Cell Physiol.* **175**, 255-267.
- Ikeya, T., Terauchi, T., Guntert, P., and Kainosho, M. (2006). Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA. *Magn Reson Chem.* **44**, 152- 157.
- Ikeya, T., Takeda, M., Yshida, H., Terauchi, T., Jee, J.G., Kainosho, M., and Guntert, P. (2009). Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL- FLYA system. *J Biomol NMR.* **44**, 261-272.
- Imberty, A., Lortat-Jacob, H., and Pérez, S. (2007). Structural view of glycosaminoglycan-protein interactions. *Carbohydr. Res.* **342**, 430-439.
- Iozzo, R.V., and Murdoch, A.D. (1996). Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **10**, 598-614.
- Iozzo, R.V. (1998). Matrix proteoglycans: from molecular design to cellular function. *Annu Rev Biochem.* **67**, 609-652.
- Iozzo, R.V., and San Antonio, J.D. (2001). Heparan sulfate proteoglycans: heavy hitters in the angiogenesis arena. *J Clin Invest.* **108**, 349-355.
- Iozzo, R.V., and Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol. J. Int. Soc. Matrix Biol.* **42**, 11-55.
- Isaacson, R.L., Simpson, P.J., Liu, M., Cota, E., Zhang, X., Freemont, P., and Matthews, S. (2007). A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc.* **129**, 15428-15429.
- Ishihara, M. (1994). Structural requirements in heparin for binding and activation o-1 and FGF- 4 are different from that for FGF-2. *Glycobiology* **4**, 817-824.
- Jarvis, D.L., Bohlmeier, D.A., Liao, Y.F., Lomax, K.K., Merkle, R.K., Weinkauf, C., and Moremen, K.W. (1997). Isolation and characterization of a class II alpha-mannosidase cDNA from lepidopteran insect cells. *Glycobiology* **7**, 113-127.
- Jayson, G.C., Hansen, S.U., Miller, G.J., Cole, C.L., Rushton, G., Avizienyte, E., and Gardiner, J.M. (2015). Synthetic heparan sulfate dodecasaccharides reveal single sulfation site interconverts CXCL8 and CXCL12 chemokine biology †Electronic supplementary information (ESI) available. See DOI: 10.1039/c5cc05222j. *Chem. Commun. Camb. Engl.* **51**, 13846-13849.
- Jedrzejas, M.J., and Stern, R. (2005). Structures of vertebrate hyaluronidases and their unique enzymatic mechanism of hydrolysis. *Proteins Struct. Funct. Bioinforma.* **61**, 227-238.
- Jemth, P., Kreuger, J., Kusche-Gullberg, M., Sturiale, L., Giménez-Gallego, G., and Lindahl, U. (2002). Biosynthetic Oligosaccharide Libraries for Identification of Protein-binding Heparan Sulfate motifs exploring the structural diversity by screening for fibroblast growth factor (FGF) 1 and FGF2 binding. *J. Biol. Chem.* **277**, 30567-30573.
- Jemth, P., Smeds, E., Do, A.-T., Habuchi, H., Kimata, K., Lindahl, U., and Kusche-Gullberg, M. (2003). Oligosaccharide library-based assessment of heparan sulfate 6-O-sulfotransferase substrate specificity. *J. Biol. Chem.* **278**, 24371-24376.
- Jiang, D., Liang, J., and Noble, P.W. (2007). Hyaluronan in tissue injury and repair. *Annu Rev Cell Dev Biol.* **23**, 435-61.



- Johansson, F.K., Göransson, H., and Westermark, B. (2005). Expression analysis of genes involved in brain tumor progression driven by retroviral insertional mutagenesis in mice. *Oncogene* **24**, 3896-3905.
- Jong-Seon, L., Ju-Hwan, K., Young-Seok, S., Jung-Bo, Y., Yong-Il, K., Hye-Jin, K., and Ki-Hwan, L. (2013). The influence of serum substituents on serum-free Vero cell conditioned culture media manufactured from Dulbecco's modified Eagle medium in mouse embryo culture. *Obstet Gynecol Sci.* **56**, 320-9.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A., and Guntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52-57.
- Kainosho, M., and Güntert, P. (2009). SAIL—stereo-array isotope labeling. *Q. Rev. Biophys.* **42**, 247-300.
- Kalia, M., Chandra, V., Rahman, S.A., Sehgal, D., and Jameel, S. (2009). Heparan sulfate proteoglycans are required for cellular binding of the hepatitis E virus ORF2 capsid protein and for viral infection. *J. Virol.* **83**, 12714-12724.
- Kalus, I., Salmen, B., Viebahn, C., von Figura, K., Schmitz, D., D'Hooge, R., and Dierks, T. (2009). Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity. *J. Cell. Mol. Med.* **13**, 4505-4521.
- Kalus, I., Rohn, S., Puvirajesinghe, T.M., Guimond, S.E., Eyckerman-Kölln, P.J., ten Dam, G., van Kuppevelt, T.H., Turnbull, J.E., and Dierks, T. (2015). Sulf1 and Sulf2 Differentially Modulate Heparan Sulfate Proteoglycan Sulfation during Postnatal Cerebellum Development: Evidence for Neuroprotective and Neurite Outgrowth Promoting Functions. *PLoS ONE* **10**.
- Kamimura, K., Fujise, M., Villa, F., Izumi, S., Habuchi, H., Kimata, K., and Nakato, H. (2001). Drosophila Heparan Sulfate 6-O-sulfotransferase (dHS6ST) Gene structure, expression, and function in the formation of the tracheal system. *J. Biol. Chem.* **276**, 17014-17021.
- Kato, K., Matsunaga, C., Odaka, A., Yamato, S., Takaha, W., Shimada, I., and Arata, Y. (1991). <sup>13</sup>C NMR-study of switch variant antidansyl anti- bodies—antigen-binding and domain domain interactions. *Biochemistry* **30**, 6604-6610.
- Kato, K., Yamaguchi, Y., and Arata, Y. (2010). Stable-isotope-assisted NMR approaches to glycoproteins using immunoglobulin G as a model system. *Prog Nucl Mag Res Spectrosc.* **56**, 346-359.
- Kato, K., Yanaka, S., Yagi, H. (2018). Technical Basis for Nuclear Magnetic Resonance Approach for Glycoproteins. *The Nuclear Magnetic Resonance Soci (Ed.), Exp. Approaches NMR Spectrosc., Springer Singapore pp.* 415-438.
- Katta, K., Imran, T., Busse-Wicher, M., Grønning, M., Czajkowski, S., and Kusche-Gullberg, M. (2015). Reduced Expression of EXTL2, a Member of the Exostosin (EXT) Family of Glycosyltransferases, in Human Embryonic Kidney 293 Cells Results in Longer Heparan Sulfate Chains. *J. Biol. Chem.* **290**, 13168-13177.
- Kameyama, H., Uchimura, K., Yamashita, T., Kuwabara, K., Mizuguchi, M., Hung, S.-C., Okuhira, K., Masuda, T., Kosugi, T., Ohgita, T., *et al.* (2019). The Accumulation of Heparan Sulfate S- Domains in Kidney Transthyretin Deposits Accelerates Fibril Formation and Promotes Cytotoxicity. *Am. J. Pathol.* **189**, 308-319.
- Kamimura, K., Koyama, T., Habuchi, H., Ueda, R., Masu, M., Kimata, K., and Nakato, H. (2006). Specific and flexible roles of heparan sulfate modifications in Drosophila FGF signaling. *J. Cell Biol.* **174**, 773-778.
- Kaneiwa, T., Mizumoto, S., Sugahara, K., and Yamada, S. (2010). Identification of human hyaluronidase-4 as a novel chondroitin sulfate hydrolase that preferentially cleaves the galactosaminidic linkage in the trisulfated tetrasaccharide sequence. *Glycobiology* **20**, 300-309.

- Karagöz, G.E., Duarte, A.M.S., Ippel, H., Uetrecht, C., Sinnige, T., van Rosmalen, M., Hausmann, J., Heck, A.J.R., Boelens, R., and Rüdiger, S.G.D. (2011). N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *Proc Natl Acad Sci.* **108**, 580-5.
- Karamanos, N.K., Piperigkou, Z., Theocharis, A.D., Watanabe, H., Franchi, M., Baud, S., Brézillon, S., Götte, M., Passi, A., Vigetti, D., Ricard-Blum, S., Sanderson, R.D., Neill, T., and Iozzo, R.V. (2018). Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics. *Chem Rev.* **118**, 9152-9232.
- Kawasaki, T., Kitsukawa, T., Bekku, Y., Matsuda, Y., Sanbo, M., Yagi, T., and Fujisawa, H. (1999). A requirement for neuropilin-1 in embryonic vessel formation. *Development* **126**, 4895-4902.
- Kay, L.E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-8979.
- Kelley, L., Mezulis, S., Yates, C. *et al.* (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858.
- Kerfah, R., Plevin, M.J., Sounier, R., Gans, P., and Boisbouvier, J. (2015). Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113-122.
- Khaing, Z.Z., Milman, B.D., Vanscoy, J.E., Seidlits, S.K., Grill, R.J., and Schmidt, C.E (2011). High molecular weight hyaluronic acid limits astrocyte activation and scar formation after spinal cord injury. *J. Neural Eng.* **8**, 046033.
- Khurana, A., McKean, H., Kim, H., Kim, S.-H., Mcguire, J., Roberts, L.R., Goetz, M.P., and Shridhar, V. (2012a). Silencing of HSulf-2 expression in MCF10DCIS.com cells attenuate ductal carcinoma in situ progression to invasive ductal carcinoma in vivo. *Breast Cancer Res. BCR* **14**, R43.
- Khurana, A., Jung-Beom, D., He, X., Kim, S.-H., Busby, R.C., Lorenzon, L., Villa, M., Baldi, A., Molina, J., Goetz, M.P., *et al.* (2013a). Matrix detachment and proteasomal inhibitors diminish Sulf-2 expression in breast cancer cell lines and mouse xenografts. *Clin. Exp. Metastasis* **30**, 407-415.
- Kim, C.W., Goldberger, O.A., Gallo, R.L., and Bernfield, M. (1994). Members of the syndecan family of heparan sulfate proteoglycans are expressed in distinct cell-, tissue-, and development-specific patterns. *Mol Biol Cell.* **5**, 797-805.
- Kim, S. H., Schneider, B. L., and Reitzer, L. (2010). Genetics and regulation of the major enzymes of alanine synthesis in *Escherichia coli*. *J Bacteriol.* **192**, 5304-5311.
- Kim, D.Y., Chaudhry, M.A., Kennard, M.L., Jardon, M.A., Braasch, K., Dionne, B., Butler, M., and Piret, J.M. (2013). Fed-batch CHO cell t-PA production and feed glutamine replacement to reduce ammonia production. *Biotechnol. Prog.* **29**, 165-175.
- Kinnunen, T., Kaksonen, M., Saarinen, J., Kalkkinen, N., Peng, H.B., and Rauvala, H. (1998). Cortactin- Src kinase signaling pathway is involved in N-syndecan-dependent neurite outgrowth. *J Biol Chem.* **273**, 10702- 10708.
- Kinnunen, T., Huang, Z., Townsend, J., Gatdula, M.M., Brown, J.R., Esko, J.D., and Turnbull, J.E. (2005). Heparan 2-O-sulfotransferase, hst-2, is essential for normal cell migration in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1507-1512.
- Kitagawa, H., Tsutsumi, K., Tone, Y., and Sugahara, K. (1997). Developmental Regulation of the Sulfation Profile of Chondroitin Sulfate Chains in the Chicken Embryo Brain. *J. Biol. Chem.* **272**, 31377-31381.
- Kleinschmit, A., Koyama, T., Dejima, K., Hayashi, Y., Kamimura, K., and Nakato, H. (2010). *Drosophila* heparan sulfate 6-O endosulfatase regulates Wingless morphogen gradient formation. *Dev. Biol.* **345**, 204-214.
- Kleist, A.B., Peterson, F., Tyler, R.C., Gustavsson, M., Handel, T.M., and Volkman, B.F. (2019). Solution NMR spectroscopy of GPCRs: Residue-specific labeling strategies with a focus on <sup>13</sup>C-methyl methionine labeling of the atypical chemokine receptor ACKR3. *Methods Cell Biol.*, Elsevier. pp. 259-288.

- Kofuku, Y., Ueda, T., Okude, J., Shiraishi, Y., Kondo, K., Maeda, K., Tsujishita, H., and Shimada, I. (2012). Efficacy of the  $\beta_2$ -adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nat. Commun.* **3**, 1045.
- Kofuku, Y., Ueda, T., Okude, J., Shiraishi, Y., Kondo, K., Mizumura, T., Suzuki, S., and Shimada, I. (2014). Functional Dynamics of Deuterated  $\beta_2$ -Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angew. Chem. Int. Ed.* **53**, 13376-13379.
- Kofuku, Y., Yokomizo, T., Imai, S., Shiraishi, Y., Natsume, M., Itoh, H., Inoue, M., Nakata, K., Igarashi, S., Yamaguchi, H., Mizukoshi, T., Suzuki, E., Ueda, T., and Shimada, I. (2018). Deuteration and selective labeling of alanine methyl groups of  $\beta_2$ -adrenergic receptor expressed in a baculovirus-insect cell expression system. *J. Biomol. NMR.* **71**, 185-192.
- Koike, T., Izumikawa, T., Sato, B., and Kitagawa, H. (2014). Identification of Phosphatase That Dephosphorylates Xylose in the Glycosaminoglycan-Protein Linkage Region of Proteoglycans. *J. Biol. Chem.* **289**, 6695-6708.
- Kokenyesi, R., and Bernfield, M. (1994). Core protein structure and sequence determine the site and presence of heparan sulfate and chondroitin sulfate on syndecan-1. *J Biol Chem.* **269**, 12304-12309.
- Kolodkin, A.L., Levenson, D.V., Rowe, E.G., Tai, Y.T., Giger, R.J., and Ginty, D.D. (1997). Neuropilin is a semaphorin III receptor. *Cell.* **90**, 753-762.
- Kolset, S.O., and Gallagher, J.T. (1990). Proteoglycans in haemopoietic cells. *Biochim. Biophys. Acta.* **1032**, 191-211.
- Kolset, S.O., Prydz, K. and Pejler, G. (2004). Intracellular proteoglycans. *Biochem J.* **379**, 217-227.
- Kolset, S.O. and Tveit, H. (2008). Serglycin-structure and biology. *Cell Mol Life Sci.* **65**, 1073-1085.
- Kolset, S.O., and Pejler, G. (2011). Serglycin: A Structural and Functional Chameleon with Wide Impact on Immune Cells. *J. Immunol.* **187**, 4927-4933.
- Kosho, T., Mizumoto, S., Watanabe, T., Yoshizawa, T., Miyake, N., and Yamada, S. (2019). Recent Advances in the Pathophysiology of Musculocontractural Ehlers-Danlos Syndrome. *Genes* **11**, 43.
- Kreuger, J., Salmivirta, M., Sturiale, L., Giménez-Gallego, G., and Lindahl, U. (2001). Sequence analysis of heparan sulfate epitopes with graded affinities for fibroblast growth factors 1 and 2. *J. Biol. Chem.* **276**, 30744-30752.
- Kreuger, J., and Kjellén, L. (2012). Heparan sulfate biosynthesis: regulation and variability. *J. Histochem. Cytochem.* **60**, 898-907.
- Kumagai, S., Ishibashi, K., Kataoka, M., Oguro, T., Kiko, Y., Yanagida, T., Aikawa, K., and Kojima, Y. (2016). Impact of Sulfatase-2 on cancer progression and prognosis in patients with renal cell carcinoma. *Cancer Sci.* **107**, 1632-1641.
- Kusche, M., Bäckström, G., Riesenfeld, J., Petitou, M., Choay, J., and Lindahl, U. (1988). Biosynthesis of heparin. O-sulfation of the antithrombin-binding region. *J. Biol. Chem.* **263**, 15474- 15484.
- Kusche-Gullberg, M., and Kjellén, L. (2003). Sulfotransferases in Glycosaminoglycan Biosynthesis. *Curr. Opin. Struct. Biol.* **13**, 605-611.
- Kuss, E. (1994). The fetoplacental unit of primates. *Exp Clin Endocrinol.* **102**, 135-165.
- Kwon, M.J., Jang, B., Yi, J.Y., Han, I.O., and Oh, E.S. (2012). Syndecans play dual roles as cell adhesion receptors and docking receptors. *FEBS Lett.* **586**, 2207- 2211.
- Laabs, T., Carulli, D., Geller, H.M., and Fawcett, J.W. (2005). Chondroitin sulfate proteoglycans in neural development and regeneration. *Curr. Opin. Neurobiol.* **15**, 116-120.
- Lai, J., Chien, J., Staub, J., Avula, R., Greene, E.L., Matthews, T.A., Smith, D.I., Kaufmann, S.H., Roberts, L.R., and Shridhar, V. (2003). Loss of HSulf-1 Up-regulates Heparin-binding Growth Factor Signaling in Cancer. *J. Biol. Chem.* **278**, 23107-23117.

- Lai, J.-P., Chien, J.R., Moser, D.R., Staub, J.K., Aderca, I., Montoya, D.P., Matthews, T.A., Nagorney, D.M., Cunningham, J.M., Smith, D.I., *et al.* (2004a). hSulf1 Sulfatase promotes apoptosis of hepatocellular cancer cells by decreasing heparin-binding growth factor signaling. *Gastroenterology* **126**, 231-248.
- Lai, J.-P., Chien, J., Strome, S.E., Staub, J., Montoya, D.P., Greene, E.L., Smith, D.I., Roberts, L.R., and Shridhar, V. (2004b). HSulf-1 modulates HGF-mediated tumor cell invasion and signaling in head and neck squamous carcinoma. *Oncogene* **23**, 1439-1447.
- Lai, J.-P., Yu, C., Moser, C.D., Aderca, I., Han, T., Garvey, T.D., Murphy, L.M., Garrity-Park, M.M., Shridhar, V., Adjei, A.A., *et al.* (2006). SULF1 inhibits tumor growth and potentiates the effects of histone deacetylase inhibitors in hepatocellular carcinoma. *Gastroenterology* **130**, 2130-2144.
- Lamanna, W.C., Baldwin, R.J., Padva, M., Kalus, I., Ten Dam, G., van Kuppevelt, T.H., Gallagher, J.T., von Figura, K., Dierks, T., and Merry, C.L.R. (2006). Heparan sulfate 6-O-endosulfatases: discrete in vivo activities and functional co-operativity. *Biochem. J.* **400**, 63-73.
- Langsdorf, A., Schumacher, V., Shi, X., Tran, T., Zaia, J., Jain, S., Taglienti, M., Kreidberg, J.A., Fine, A., and Ai, X. (2011). Expression regulation and function of heparan sulfate 6-O-endosulfatases in the spermatogonial stem cell niche. *Glycobiology* **21**, 152-161.
- Lanzi, C., and Cassinelli, G. (2018). Heparan Sulfate Mimetics in Cancer Therapy: The Challenge to Define Structural Determinants and the Relevance of Targets for Optimal Activity. *Molecules* **23**, 2915.
- Ledin, J., Ringvall, M., Thuveson, M., Eriksson, I., Wilén, M., Kusche-Gullberg, M., Forsberg, E., and Kjellén, L. (2006). Enzymatically active N-deacetylase/N-sulfotransferase-2 is present in liver but does not contribute to heparan sulfate N-sulfation. *J. Biol. Chem.* **281**, 35727-35734.
- Lee, H.S., Qi, Y., and Im, W. (2015). Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *SCIENTIFIC REPORTS* **5**, 8926
- Lehninger, A. L. (1975). Chapter 22: Biosynthesis of amino acids, nucleotides, and related molecules. In *Principles of Biochemistry* (pp. 833-880). New York, NY: Worth Publishers Inc.
- LeMaster, D.M., and Richards, F.M. (1982). Preparative-scale isolation of isotopically labeled amino acids. *Analytical Biochemistry* **122**, 238-247.
- Lemire, J.M., Chan, C.K., Bressler, S., Miller, J., LeBaron, R.G., and Wight, T.N. (2007). Interleukin-18 selectively decreases the synthesis of versican by arterial smooth muscle cells. *J. Cell. Biochem.* **101**, 753-766.
- Leonova, E.I., and Galzitskaia, O.V. (2013). Comparative characteristics of the structure and function for syndecan-1 from animal organisms. *Mol. Biol. (Mosk.)* **47**, 505-512.
- Leonova, E.I., and Galzitskaya, O.V. (2015). Role of Syndecans in Lipid Metabolism and Human Diseases. *Adv. Exp. Med. Biol.* **855**, 241-258.
- Letoha, T., Hudak, A., Kusz, E., Pettko-Szandtner, A., Domonkos, I., Josvay, K., Hofmann-Apitius, M., and Szilak, L. (2019). Contribution of syndecans to cellular internalization and fibrillation of amyloid-b(1-42). *Sci. Rep.* **9**, 1393.
- Levy-Adam, F., Feld, S., Cohen-Kaplan, V., Shteingauz, A., Gross, M., Arvatz, G., Naroditsky, I., Ilan, N., Doweck, I., and Vlodaysky, I. (2010). Heparanase 2 Interacts with Heparan Sulfate with High Affinity and Inhibits Heparanase Activity. *J. Biol. Chem.* **285**, 28010-28019.
- Li, J.-P., Gong, F., Hagner-McWhirter, Å., Forsberg, E., Åbrink, M., Kisilevsky, R., Zhang, X., and Lindahl, U. (2003). Targeted Disruption of a Murine Glucuronyl C5-epimerase Gene Results in Heparan Sulfate Lacking 1-Iduronic Acid and in Neonatal Lethality. *J. Biol. Chem.* **278**, 28363- 28366.
- Li, J.-P., and Kusche-Gullberg, M. (2016). Chapter Six - Heparan Sulfate: Biosynthesis, Structure, and Function. In *International Review of Cell and Molecular Biology*, K.W. Jeon, ed. (Academic Press), pp. 215-273.

- Li, J., Kleeff, J., Abiatari, I., Kayed, H., Giese, N.A., Felix, K., Giese, T., Büchler, M.W., and Friess, H. (2005). Enhanced levels of HSulf-1 interfere with heparin-binding growth factor signaling in pancreatic cancer. *Mol. Cancer* **4**, 14.
- Li, J., Mo, M.-L., Chen, Z., Yang, J., Li, Q.-S., Wang, D.-J., Zhang, H., Ye, Y.-J., Li, H.-L., Zhang, F., *et al.* (2011). HSulf-1 inhibits cell proliferation and invasion in human gastric cancer. *Cancer Sci.* **102**, 1815-1821.
- Li, Y., Laue, K., Temtamy, S., Aglan, M., Kotan, L.D., Yigit, G., *et al.* (2010). Temtamy Preaxial Brachydactyly Syndrome Is Caused by Loss-Of-Function Mutations in Chondroitin Synthase 1, a Potential Target of BMP Signaling. *Am. J. Hum. Genet.* **87**, 757-767.
- Lichtenecker, R.J., Weinhaupl, K., Schmid, W., and Konrat, R. (2013). alpha-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *J Biomol NMR.* **57**, 327-331.
- Lichtenthaler, S.F., Lemberg, M.K., Fluhrer, R. (2018). Proteolytic ectodomain shedding of membrane proteins in mammals—hardware, concepts, and recent developments. *EMBO J.* **37**, e99456.
- Lidholt, K., Eriksson, I., and Kjellén, L. (1995). Heparin proteoglycans synthesized by mouse mastocytoma contain chondroitin sulphate. *Biochem. J.* **311**, 233-238.
- Lindahl, U., Jacobsson, I., Höök, M., Backström, G., and Feingold, D.S. (1976). Biosynthesis of heparin. Loss of C-5 hydrogen during conversion of d-glucuronic to l-iduronic acid residues. *Biochem. Biophys. Res. Commun.* **70**, 492-499.
- Lindahl, B., Eriksson, L., and Lindahl, U. (1995). Structure of heparan sulphate from human brain, with special regard to Alzheimer's disease. *Biochem. J.* **306**, 177-184.
- Lindahl, B., Westling, C., Giménez-Gallego, G., Lindahl, U., and Salmivirta, M. (1999). Common Binding Sites for B-Amyloid Fibrils and Fibroblast Growth Factor-2 in Heparan Sulfate from Human Cerebral Cortex. *J. Biol. Chem.* **274**, 30631-30635.
- Lindahl, U., and Li, J. (2009). Interactions between heparan sulfate and proteins—design and functional implications. *Int. Rev. Cell Mol. Biol.* **276**, 105-159.
- Liu, J., Shworak, N.W., Sinay, P., Schwartz, J.J., Zhang, L., Fritze, L.M.S., and Rosenberg, R.D. (1999). Expression of Heparan Sulfate d-Glucosaminyl 3-O-sulfotransferase Isoforms Reveals Novel Substrate Specificities. *J. Biol. Chem.* **274**, 5185-5192.
- Liu, P., Khurana, A., Rattan, R., He, X., Kalloger, S., Dowdy, S., Gilks, B., and Shridhar, V. (2009). Regulation of HSulf-1 Expression by Variant Hepatic Nuclear Factor 1 in Ovarian Cancer. *Cancer Res.* **69**, 4843-4850.
- Liu, H., Fu, X., Ji, W., Liu, K., Bao, L., Yan, Y., Wu, M., Yang, J., and Su, C. (2013). Human sulfatase-1 inhibits the migration and proliferation of SMMC-7721 hepatocellular carcinoma cells by downregulating the growth factor signaling. *Hepatol. Res. Off. J. Jpn. Soc. Hepatol.* **43**, 516-525.
- Liu, C., Sheng, J., Krahn, J.M., Perera, L., Xu, Y., Hsieh, P.-H., Dou, W., Liu, J., and Pedersen, L.C. (2014). Molecular Mechanism of Substrate Specificity for Heparan Sulfate 2-O-sulfotransferase. *J. Biol. Chem.* **289**, 13407-13418.
- Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol.* **285**, 2177-2198.
- Lokeshwar, V.B., Obek, C., Soloway, M.S., and Block, N.L. (1997). Tumor-associated hyaluronic acid: a new sensitive and specific urine marker for bladder cancer. *Cancer Res.* **57**, 773-7.
- London, R.E., Wingad, B.D. and Mueller, G.A. (2008). Dependence of amino acid side chain <sup>13</sup>C shifts on dihedral angle: application to conformational analysis. *J Am Chem Soc.* **130**, 11097-11105.
- Longley, R.L., Woods, A., Fleetwood, A., Cowling, G.J., Gallagher, J.T., and Couchman, J.R. (1999). Control of morphology, cytoskeleton and migration by syndecan-4. *J Cell Sci.* **112**, 3421-3431.
- Lui, N.S., van Zante, A., Rosen, S.D., Jablons, D.M., and Lemjabbar-Alaoui, H. (2012). SULF2 expression by immunohistochemistry and overall survival in oesophageal cancer: a cohort study. *BMJ Open* **2**, e001624.

- Luis, A.S., Jin, C., Pereira, G.V. *et al.* (2021). A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature* **598**, 332-337.
- Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, K. (1998). Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* **37**, 3654-64.
- Lum, D.H., Tan, J., Rosen, S.D., and Werb, Z. (2007). Gene Trap Disruption of the Mouse Heparan Sulfate 6-O-Endosulfatase Gene, *Sulf2*. *Mol. Cell. Biol.* **27**, 678-688.
- Lyon, M., Deakin, J.A., Mizuno, K., Nakamura, T., and Gallagher, J.T. (1994). Interaction of hepatocyte growth factor with heparan sulfate. Elucidation of the major heparan sulfate structural determinants. *J. Biol. Chem.* **269**, 11216-11223.
- Lyon, M., Rushton, G., and Gallagher, J.T. (1997). The Interaction of the transforming Growth Factor- $\beta$ s with Heparin/Heparan Sulfate Is Isoform-specific. *J. Biol. Chem.* **272**, 18000-18006.
- Lyon, M., Deakin, J.A., Rahmoune, H., Fernig, D.G., Nakamura, T., and Gallagher, J.T. (1998). Hepatocyte Growth Factor/Scatter Factor Binds with High Affinity to Dermatan Sulfate. *J. Biol. Chem.* **273**, 271-278.
- Ma, H.-Y., Zhang, F., Li, J., Mo, M.-L., Chen, Z., Liu, L., Zhou, H.-M., and Sheng, Q. (2011). HSulf-1 suppresses cell growth and down-regulates Hedgehog signaling in human gastric cancer cells. *Oncol. Lett.* **2**, 1291-1295.
- Maccarana, M., Casu, B., and Lindahl, U. (1993). Minimal sequence in heparin/heparan sulfate required for binding of basic fibroblast growth factor. *J. Biol. Chem.* **268**, 23898-23905.
- Maccarana, M., Sakura, Y., Tawada, A., Yoshida, K., and Lindahl, U. (1996). *J. Biol. Chem.* **271**, 17804-10.
- Maccarana, M., Kalamajski, S., Kongsgaard, M., Magnusson, S.P., Oldberg, A., and Malmström, A. (2009). Dermatan Sulfate Epimerase 1-deficient Mice Have Reduced Content and Changed Distribution of Iduronic Acids in Dermatan Sulfate and an Altered Collagen Structure in Skin. *Mol. Cell. Biol.* **29**, 5517-5528.
- Macek, P., Kerfah, R., Boeri Erba, E., Crublet, E., Moriscot, C., Schoehn, G., Amero, C., and Boisbouvier, J. (2017). Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2. *Sci Adv.* **3**, e1601601.
- Magwenyane, A.M., Mhlongo, N.N., Lawal, M.M., Amoako, D.G., Somboro, A.M., Sosibo, S.C., Shunmugam, L., Khan, R.B., and Kumalo, H.M. (2020). Understanding the Hsp90 N-terminal Dynamics: Structural and Molecular Insights into the Therapeutic Activities of Anticancer Inhibitors Radicicol (RD) and Radicicol Derivative (NVP-YUA922). *Molecules* **25**, 1785.
- Mahmoud, S., Ibrahim, M., Hago, A., Huang, Y., Wei, Y., Zhang, J., Zhang, Q., Xiao, Y., Wang, J., Adam, M., *et al.* (2016). Overexpression of sulfatase-1 in murine hepatocarcinoma Hca-F cell line downregulates mesothelin and leads to reduction in lymphatic metastasis, both in vitro and in vivo. *Oncotarget* **7**, 75052-75063.
- Malavaki, C., Mizumoto, S., Karamanos, N., and Sugahara, K. (2008). Recent Advances in the Structural Study of Functional Chondroitin Sulfate and Dermatan Sulfate in Health and Disease. *Connect. Tissue Res.* **49**, 133-139.
- Malmström, A., Bartolini, B., Thelin, M. A., Pacheco, B., and Maccarana, M. (2012). Iduronic Acid in Chondroitin/Dermatan Sulfate. *J. Histochem. Cytochem.* **60**, 916-925.
- Manon-Jensen, T., Itoh, Y., and Couchman, J.R. (2010). Proteoglycans in health and disease: the multiple roles of syndecan shedding. *FEBS J.* **277**, 3876-3889.
- Manou, D., Karamanos, N.K., and Theocharis, A.D. (2020). Tumorigenic functions of serglycin: Regulatory roles in epithelial to mesenchymal transition and oncogenic signaling. *Semin Cancer Biol.* **62**, 108-115.
- Markley, J.L., Putter, I., and Jardetzky, O. (1968). High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science* **161**, 1249-1251.

- Marques, C., Reis C.A., Vivès, R.R., and Magalhães, A. (2021). Heparan Sulfate Biosynthesis and Sulfation Profiles as Modulators of Cancer Signalling and Progression. *Front. In oncology In press*.
- Marques, C., Poças, J., Faria-Ramos, I., Gomes, C., Reis, C., Vivès, R., and Magalhães, A. (2021). Exostosin-like 2 Abrogation Promotes Heparan Sulfate Biosynthesis and Switches Cancer Cell Signalling Towards an Invasive Phenotype. *FASEB Journal*. <https://doi.org/10.1096/fasebj.2021.35.S1.01973>. – **SUBMITTED**
- Mas, G., Crublet, E., Hamelin, O., Gans, P., and Boisbouvier, J. (2013). Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. *J Biomol NMR*. **57**, 251-262.
- Mas, G., Guan, J.Y., Crublet, E., Debled, E.C., Moriscot, C., Gans, P., Schoehn, G., Macek, P., Schanda, P., and Boisbouvier, J. (2018). Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. *Sci Adv*. **4**, 4196.
- Matalon, R., Arbogast, B., and Dorfman, A. (1974). Deficiency of chondroitin sulfate N-acetylgalactosamine 4-sulfatase in Maroteaux-Lamy syndrome. *Biochem. Biophys. Res. Commun*. **61**, 1450-1457.
- McAtee, C.O., Barycki, J.J., and Simpson, M.A. (2014). Emerging roles for hyaluronidase in cancer metastasis and therapy. *Adv. Cancer Res*. **123**, 1-34.
- McCaldon, P. and Argos, P. (1988). Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**, 99-122.
- McCormick, C., Duncan, G., Goutsos, K.T., and Tufaro, F. (2000). The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proc. Natl. Acad. Sci*. **97**, 668-673.
- McKeehan, W.L., Wu, X.C., and Kan, M. (1999). Requirement for anticoagulant heparan sulfate in the fibroblast growth factor receptor complex. *J Biol Chem*. **274**, 21511-21514.
- McKenzie, E., Tyson, K., Stamps, A., Smith, P., Turner, P., Barry, R., Hircock, M., Patel, S., Barry, E., Stubberfield, C., *et al.* (2000). Cloning and expression profiling of Hpa2, a novel mammalian heparanase family member. *Biochem. Biophys. Res. Commun*. **276**, 1170-1177.
- Meen, A.J., Øynebråten, I., Reine, T.M., Duelli, A., Svennevig, K., Pejler, G., Jenssen, T., and Kolset, S.O. (2011). Serglycin is a major proteoglycan in polarized human endothelial cells and is implicated in the secretion of the chemokine GROalpha/CXCL1. *J. Biol. Chem*. **286**, 2636-2647.
- Meola, A., Deville, C., Jeffers, S.A., Guardado-Calvo, P., Vasiliauskaite, I., Sizun, C., Girard-Blanc, C., Malosse, C., van Heijenoort, C., Chamot-Rooke, J., Krey, T., Guittet, E., Pêtres, S., Rey, F.A., and Bontems, F. (2014). Robust and low cost uniform (15)N-labeling of proteins expressed in *Drosophila* S2 cells and *Spodoptera frugiperda* Sf9 cells for NMR applications. *J. Struct. Biol*. **188**, 71-78.
- Mikami, T., and Kitagawa, H. (2013). Biosynthesis and Function of Chondroitin Sulfate. *Biochim. Biophys. Acta (Bba)-Gen. Subjects* **1830**, 4719-4733.
- Miller, S., Janin, J., Lesk, A.M., and Chothia, C. (1987). Interior and surface of monomeric proteins. *J Mol Biol*. **196**, 641-656.
- Milz, F., Harder, A., Neuhaus, P., Breitzkreuz-Korff, O., Walhorn, V., Lübke, T., Anselmetti, D., and Dierks, T. (2013). Cooperation of binding sites at the hydrophilic domain of cell-surface sulfatase Sulf1 allows for dynamic interaction of the enzyme with its substrate heparan sulfate. *Biochim. Biophys. Acta BBA - Gen. Subj*. **1830**, 5287-5298.
- Mitsuhashi, J. (1982). Determination of essential amino acids for insect cell lines. In K. Maramorosch & J. Mitsuhashi (Eds.), *Invertebrate cell culture applications* (pp. 9-52). New York, NY: Academic Press Inc.
- Miyamoto, K., Asada, K., Fukutomi, T., Okochi, E., Yagi, Y., Hasegawa, T., Asahara, T., Sugimura, T., and Ushijima, T. (2003). Methylation-associated silencing of heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2 (3-OST-2) in human breast, colon, lung and pancreatic cancers. *Oncogene* **22**, 274-280.

- Miyanoiri, Y., Takeda, M., Okuma, K., Ono, A.M., Terauchi, T., and Kainosho, M. (2013). Differential isotope-labeling for Leu and Val residues in a protein by *E. coli* cellular expression using stereo-specifically methyl labeled amino acids. *J Biomol NMR*. **57**, 237-249.
- Miyanoiri, Y., Ishida, Y., Takeda, M., Terauchi, T., Inouye, M., and Kainosho, M. (2016). Highly efficient residue-selective labeling with isotope-labeled Ile, Leu, and Val using a new auxotrophic *E. coli* strain. *J. Biomol. NMR*. **65**, 109-119.
- Miyazawa-Onami, M., Takeuchi, K., Takano, T., Sugiki, T., Shimada, I., and Takahashi, H. (2013). Perdeuteration and methyl-selective <sup>1</sup>H, <sup>13</sup>C-labeling by using a *Kluyveromyces lactis* expression system. *J. Biomol. NMR*. **57**, 297-304.
- Mizuguchi, S., Uyama, T., Kitagawa, H., Nomura, K.H., Dejima, K., Gengyo-Ando, K., *et al.* (2003). Chondroitin Proteoglycans Are Involved in Cell Division of *Caenorhabditis elegans*. *Nature* **423**, 443-448.
- Mizumoto, S., Mikami, T., Yasunaga, D., Kobayashi, N., Yamauchi, H., Miyake, A., *et al.* (2009). Chondroitin 4-O-sulfotransferase-1 Is Required for Somitic Muscle Development and Motor Axon Guidance in Zebrafish. *Biochem. J.* **419**, 387-399.
- Mizumoto, S., Ikegawa, S., and Sugahara, K. (2013). Human Genetic Disorders Caused by Mutations in Genes Encoding Biosynthetic Enzymes for Sulfated Glycosaminoglycans. *J. Biol. Chem.* **288**, 10953-10961.
- Mizumoto, S., Yamada, S., and Sugahara, K. (2015). Molecular Interactions between Chondroitin-Dermatan Sulfate and Growth Factors/receptors/ matrix Proteins. *Curr. Opin. Struct. Biol.* **34**, 35-42.
- Mizumoto, S., Kosho, T., Yamada, S., and Sugahara, K. (2017). Pathophysiological Significance of Dermatan Sulfate Proteoglycans Revealed by Human Genetic Disorders. *Pharmaceuticals* **10**, 34.
- Mizumoto, S. (2018). Defects in Biosynthesis of Glycosaminoglycans Cause Hereditary Bone, Skin, Heart, Immune, and Neurological Disorders. *Trends Glycosci. Glycotechnol.* **30**, E67-E89.
- Murphy, K.J., Merry, C.L., Lyon, M., Thompson, J.E., Roberts, I.S., and Gallagher, J.T. (2004). *J. Biol. Chem.* **279**, 27239-45.
- Monneau, Y., Arenzana-Seisdedos, F., and Lortat-Jacob, H. (2016). The sweet spot: how GAGs help chemokines guide migrating cells. *J. Leukoc. Biol.* **99**, 935-953.
- Moon, A.F., Edavettal, S.C., Krahn, J.M., Munoz, E.M., Negishi, M., Linhardt, R.J., Liu, J., and Pedersen, L.C. (2004). Structural analysis of the sulfotransferase (3-O-sulfotransferase isoform 3) involved in the biosynthesis of an entry receptor for herpes simplex virus 1. *J. Biol. Chem.* **279**, 45185-45193.
- Moon, A.F., Xu, Y., Woody, S.M., Krahn, J.M., Linhardt, R.J., Liu, J., and Pedersen, L.C. (2012). Dissecting the substrate recognition of 3-O-sulfotransferase for the biosynthesis of anticoagulant heparin. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5265-5270.
- Morimoto-Tomita, M., Uchimura, K., Werb, Z., Hemmerich, S., and Rosen, S.D. (2002). Cloning and Characterization of Two Extracellular Heparin-degrading Endosulfatases in Mice and Humans. *J. Biol. Chem.* **277**, 49175-49185.
- Mosser, M., Chevalot, I., Olmos, E., Blanchard, F., Kapel, R., Oriol, E., Marc, I., and Marc, A. (2013). Combination of yeast hydrolysates to improve CHO cell growth and IgG production, *Cytotechnology* **65**, 629-641.
- Moussay, E., Palissot, V., Vallar, L., Poirel, H.A., Wenner, T., El Khoury, V., Aouali, N., Van Moer, K., Leners, B., Bernardin, F., *et al.* (2010). Determination of genes and microRNAs involved in the resistance to fludarabine in vivo in chronic lymphocytic leukemia. *Mol. Cancer* **9**, 115.
- Muchmore, S.W., Sattler, M., Liang, H., Meadows, R.P., Harlan, J.E., Yoon, H.S., Nettesheim, D., Chang, B.S., Thompson, C.B., Wong, S.L., Ng, S.L., and Fesik, S.W. (1996). X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **381**, 335-341.
- Mulloy, B. (2005). The specificity of interactions between proteins and sulfated polysaccharides. *An Acad Bras Cienc.* **77**, 651-664.



- Murphy, J., Desai, C., Giaretti, W., Kendall, F., and Nicolini, C. (1977). Experimental results on mammalian cells growing in vitro in deuterated medium for neutron-scattering studies. *J. Cell Sci.* **25**, 87-94.
- Myette, J.R., Shriver, Z., Claycamp, C., McLean, M.W., Venkataraman, G., and Sasisekharan, R. (2003). The heparin/heparan sulfate 2-O-sulfatase from *Flavobacterium heparinum*. Molecular cloning, recombinant expression, and biochemical characterization. *J. Biol. Chem.* **278**, 12157-12166.
- Myette, J.R., Soundararajan, V., Shriver, Z., Raman, R., and Sasisekharan, R. (2009a). Heparin/heparan sulfate N-sulfamidase from *Flavobacterium heparinum*: structural and biochemical investigation of catalytic nitrogen-sulfur bond cleavage. *J. Biol. Chem.* **284**, 35177-35188.
- Myette, J.R., Soundararajan, V., Behr, J., Shriver, Z., Raman, R., and Sasisekharan, R. (2009b). Heparin/heparan sulfate 6-O-sulfatase from *Flavobacterium heparinum*: integrated structural and biochemical investigation of enzyme active site and substrate specificity. *J. Biol. Chem.* **284**, 35189-35200.
- Nadanaka, S., Zhou, S., Kagiya, S., Shoji, N., Sugahara, K., Sugihara, K., Asano, M., and Kitagawa, H. (2013a). EXTL2, a Member of the EXT Family of Tumor Suppressors, Controls Glycosaminoglycan Biosynthesis in a Xylose Kinase-dependent Manner. *J. Biol. Chem.* **288**, 9321-9333.
- Nadanaka, S., Kagiya, S., and Kitagawa, H. (2013b). Roles of EXTL2, a member of the EXT family of tumour suppressors, in liver injury and regeneration processes. *Biochem. J.* **454**, 133-145.
- Nagai, N., Habuchi, H., Esko, J.D., and Kimata, K. (2004). Stem domains of heparan sulfate 6-O-sulfotransferase are required for Golgi localization, oligomer formation and enzyme activity. *J. Cell Sci.* **117**, 3331-3341.
- Nagamine, S., Keino-Masu, K., Shiomi, K., and Masu, M. (2010). Proteolytic cleavage of the rat heparan sulfate 6-O-endosulfatase SulFP2 by furin-type proprotein convertases. *Biochem. Biophys. Res. Commun.* **391**, 107-112.
- Narita, K., Staub, J., Chien, J., Meyer, K., Bauer, M., Friedl, A., Ramakrishnan, S., and Shridhar, V. (2006). HSulf-1 Inhibits Angiogenesis and Tumorigenesis In vivo. *Cancer Res.* **66**, 6025-6032.
- Narita, K., Chien, J., Mullany, S.A., Staub, J., Qian, X., Lingle, W.L., and Shridhar, V. (2007). Loss of HSulf-1 Expression Enhances Autocrine Signaling Mediated by Amphiregulin in Breast Cancer. *J. Biol. Chem.* **282**, 14413-14420.
- Nawroth, R., van Zante, A., Cervantes, S., McManus, M., Hebrok, M., and Rosen, S.D. (2007). Extracellular sulfatases, elements of the Wnt signaling pathway, positively regulate growth and tumorigenicity of human pancreatic cancer cells. *PLoS One* **2**, e392.
- Neill, T., Schaefer, L., and Iozzo, R.V. (2015). Decoding the Matrix: Instructive Roles of Proteoglycan Receptors. *Biochemistry* **54**, 4583-4598.
- Neugebauer, J.M., Cadwallader, A.B., Amack, J.D., Bisgrove, B.W., and Yost, H.J. (2013). Differential roles for 3-OSTs in the regulation of cilia length and motility. *Development* **140**, 3892-3902.
- Ndeh, D., Baslé, A., Strahl, H., Yates, E.A., McClurg, U.L., Henrissat, B., Terrapon, N., and Cartmell, A. (2020). Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus. *Nat Commun.* **11**, 646.
- Ning, L., Xu, Z., Furuya, N., Nonaka, R., Yamada, Y., and Arikawa-Hirasawa, E. (2015). Perlecan Inhibits Autophagy to Maintain Muscle Homeostasis in Mouse Soleus Muscle. *Matrix Biol.* **48**, 26-35.
- Nygaard, R., Zou, Y., Dror, R.O., Mildorf, T.J., Arlow, D.H., Manglik, A., Pan, A.C., Liu, C.W., Fung, J.J., Bokoch, M.P., Thian, F.S., Kobilka, T.S., Shaw, D.E., Mueller, L., Prosser, R.S., and Kobilka, B.K. (2013). The dynamic process of  $\beta(2)$ -adrenergic receptor activation. *Cell.* **152**, 532-542.
- O'Reilly, D.R., Miller, L.K., and Luckow, V.A. (1994). *Baculovirus expression vectors: a laboratory manual.* Oxford University Press, Oxford.

- Öhman, L., Ljunggren, J., and Häggström, L. (1995). Induction of a metabolic switch in insect cells by substrate-limited fed batch cultures. *Applied Microbiology and Biotechnology* **43**, 1006-1013.
- Öhman, L., Alarcon, M., Ljunggren, J., Ramqvist, A.-K., and Häggström, L. (1996). Glutamine is not an essential amino acid for Sf-9 insect cells. *Biotechnology Letters* **18**, 765-770.
- Olson, S.K., Bishop, J.R., Yates, J.R., Oegema, K., and Esko, J. D. (2006). Identification of Novel Chondroitin Proteoglycans in *Caenorhabditis elegans*: Embryonic Cell Division Depends on CPG-1 and CPG-2. *J. Cell Biol.* **173**, 985-994.
- Olsson, L., Stigson, M., Perris, R., Sorrell, J.M., and Lofberg, J. (1996). Distribution of keratan sulphate and chondroitin sulphate in wild type and white mutant axolotl embryos during neural crest cell migration. *Pigment Cell Res.* **9**, 5-17.
- Ono, K., Hattori, H., Takeshita, S., Kurita, A., and Ishihara, M. (1999). Structural features in heparin that interact with VEGF165 and modulate its biological activity. *Glycobiology* **9**, 705-711.
- Opitz, C., Isogai, S., and Grzesiek, S. (2015). An economic approach to efficient isotope labeling in insect cells using homemade <sup>15</sup>N-, <sup>13</sup>C- and <sup>2</sup>H-labeled yeast extracts. *J. Biomol. NMR.* **62**, 373-385.
- Ori, A., Wilkinson, M.C., and Fernig, D.G. (2011). A Systems Biology Approach for the Investigation of the Heparin/Heparan Sulfate Interactome. *J. Biol. Chem.* **286**, 19892-19904.
- Otsuki, S., Hanson, S.R., Miyaki, S., Grogan, S.P., Kinoshita, M., Asahara, H., Wong, C.-H., and Lotz, M.K. (2010). Extracellular sulfatases support cartilage homeostasis by regulating BMP and FGF signaling pathways. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10202-10207.
- Otsuki, S., Alvarez-Garcia, O., Lotz, M.K., and Neo, M. (2019). Role of heparan sulfate 6-O endosulfatases in intervertebral disc homeostasis. *Histol. Histopathol.* 18107.
- Pallerla, S.R., Lawrence, R., Lewejohann, L., Pan, Y., Fischer, T., Schlomann, U., Zhang, X., Esko, J.D., and Grobe, K. (2008). Altered heparan sulfate structure in mice with deleted NDST3 gene function. *J. Biol. Chem.* **283**, 16885-16894.
- Patel, V.N., Likar, K.M., Zisman-Rozen, S., Cowherd, S.N., Lassiter, K.S., Sher, I., Yates, E.A., Turnbull, J.E., Ron, D., and Hoffman, M.P. (2008). Specific Heparan Sulfate Structures Modulate FGF10-mediated Submandibular Gland Epithelial Morphogenesis and Differentiation. *J. Biol. Chem.* **283**, 9308-9317.
- Patel, V.N., Lombaert, I.M., Cowherd, S.N., Shworak, N.W., Xu, Y., Liu, J., and Hoffman, M.P. (2014). Hs3st3-Modified Heparan Sulfate Controls KIT+ Progenitor Expansion by Regulating 3-O-sulfotransferases. *Dev Cell.* **29**, 662-673.
- Pejler, G., Abrink, M., and Wernersson, S. (2009). Serglycin proteoglycan: regulating the storage and activities of hematopoietic proteases. *Biofactors.* **35**, 61-68.
- Pellecchia, M., Meininger, D., Dong, Q., Chang, E., Jack, R., and Sem, D.S. (2002). NMR-based structural characterization of large protein-ligand interactions. *J. Biomol. NMR,* **22**, 165-173.
- Pempe, E.H., Burch, T.C., Law, C.J., and Liu, J. (2012). Substrate specificity of 6-O-endosulfatase (Sulf-2) and its implications in synthesizing anticoagulant heparan sulfate. *Glycobiology* **22**, 1353- 1362.
- Peterson, S.M., Iskenderian, A., Cook, L., Romashko, A., Tobin, K., Jones, M., Norton, A., Gómez- Yafal, A., Heartlein, M.W., Concino, M.F., *et al.* (2010). Human Sulfatase 2 inhibits in vivo tumor growth of MDA-MB-231 human breast cancer xenografts. *BMC Cancer* **10**, 427.
- Petitou, M., Casu, B., and Lindahl, U. (2003). 1976-1983, a critical period in the history of heparin: the discovery of the antithrombin binding site. *Biochimie* **85**, 83-89.
- Pham, P.L. Perret, S., Cass, B., Carpentier, E., St-Laurent, G., Bisson, L., Kamen, A., Durocher, Y. (2005). Transient gene expression in HEK293 cells: Peptone addition post transfection improves recombinant protein synthesis. *Biotechnol Bioeng.* **90**, 332-44.

- Phillips, J.J., Huillard, E., Robinson, A.E., Ward, A., Lum, D.H., Polley, M.-Y., Rosen, S.D., Rowitch, D.H., and Werb, Z. (2012). Heparan sulfate sulfatase SULF2 regulates PDGFR $\alpha$  signaling and growth in human and mouse malignant glioma. *J. Clin. Invest.* **122**, 911-922.
- Pichert, A., Schlorke, D., Franz, S., and Arnhold, J. (2012). Functional aspects of the interaction between interleukin-8 and sulfated glycosaminoglycans. *Biomatter* **2**, 142-148.
- Pickford, A.R., and O'Leary, J.M. (2004). Isotopic labeling of recombinant proteins from the methylotrophic yeast *Pichia pastoris*. *Methods Mol. Biol. Clifton NJ.* **278**, 17-33.
- Pikas, D.S., Eriksson, I., and Kjellén, L. (2000). Overexpression of different isoforms of glucosaminyl N-deacetylase/N-sulfotransferase results in distinct heparan sulfate N-sulfation patterns. *Biochemistry* **39**, 4552-4558.
- Pinhal, M.A.S., Smith, B., Olson, S., Aikawa, J., Kimata, K., and Esko, J.D. (2001). Enzyme interactions in heparan sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact in vivo. *Proc. Natl. Acad. Sci.* **98**, 12984-12989.
- Poluzzi, C., Casulli, J., Goyal, A., Mercer, T.J., Neill, T., and Iozzo, R.V. (2014). Endorepellin evokes autophagy in endothelial cells. *J Biol Chem.* **289**, 16114-16128.
- Pönighaus, C., Ambrosius, M., Casanova, J.C., Prante, C., Kuhn, J., Esko, J.D., *et al.* (2007). Human Xylosyltransferase II Is Involved in the Biosynthesis of the Uniform Tetrasaccharide Linkage Region in Chondroitin Sulfate and Heparan Sulfate Proteoglycans. *J Biol Chem.* **282**, 5201-6.
- Pomin, V.H., Piquet, A.A., Pereira, M.S., and Mourão, P.A. (2012). *Carbohydr. Polym.* **90**, 839-846.
- Pomin, V.H. (2014). *Anal. Chem.* **86**, 65-94.
- Pozzi, A., Yurchenco, P.D., and Iozzo, R.V. (2017). The Nature and Biology of Basement Membranes. *Matrix Biol.* **57-58**, 1-11.
- Pratt, T., Conway, C.D., Tian, N.M.M.-L., Price, D.J., and Mason, J.O. (2006). Heparan sulphation patterns generated by specific heparan sulfotransferase enzymes direct distinct aspects of retinal axon guidance at the optic chiasm. *J. Neurosci.* **26**, 6911-6923.
- Préchoux, A., Halimi, C., Simorre, J.-P., Lortat-Jacob, H., and Laguri, C. (2015). C5-Epimerase and 2-O-sulfotransferase Associate in Vitro to Generate Contiguous Epimerized and 2-O-sulfated Heparan Sulfate Domains. *ACS Chem. Biol.* **10**, 1064-1071.
- Presto, J., Thuveson, M., Carlsson, P., Busse, M., Wilén, M., Eriksson, I., Kusche-Gullberg, M., and Kjellén, L. (2008). Heparan sulfate biosynthesis enzymes EXT1 and EXT2 affect NDST1 expression and heparan sulfate sulfation. *Proc. Natl. Acad. Sci.* **105**, 4751-4756.
- Proudfoot, A.E.I. (2006). The biological relevance of chemokine-proteoglycan interactions. *Biochem. Soc. Trans.* **34**, 422-426.
- Prydz, K. (2015). Determinants of Glycosaminoglycan (GAG) Structure. *Biomolecules.* **5**, 2003-2022.
- Puxbaum, V., Mattanovich, D., and Gasser, B. (2015). Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* **99**, 2925-2938.
- Pye, D.A., Vives, R.R., Turnbull, J.E., Hyde, P., and Gallagher, J.T. (1998). Heparan sulfate oligosaccharides require 6-O-sulfation for promotion of basic fibroblast growth factor mitogenic activity. *J Biol Chem.* **273**, 22936-42.
- Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J., and Wuthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell.* **59**, 573-580.
- Qin, Y., Ke, J., Gu, X., Fang, J., Wang, W., Cong, Q., Li, J., Tan, J., Brunzelle, J.S., Zhang, C., *et al.* (2015). Structural and Functional Study of d-Glucuronyl C5-epimerase. *J. Biol. Chem.* **290**, 4620-4630.
- Quantock, A.J., Young, R.D., and Akama, T.O. (2010). *Cell. Mol. Life Sci.* **67**, 891-906

- Raman, R., Myette, J.R., Shriver, Z., Pojasek, K., Venkataraman, G., and Sasisekharan, R. (2003). The Heparin/Heparan Sulfate 2-O-sulfatase from *Flavobacterium heparinum* a structural and biochemical study of the enzyme active site and saccharide substrate specificity. *J. Biol. Chem.* **278**, 12167-12174.
- Rapraeger, A.C., Krufka, A., and Olwin, B.B. (1991). Requirement of heparan sulfate for bFGF- mediated fibroblast growth and myoblast differentiation. *Science* **252**, 1705-1708.
- Reijmers, R.M., Vondenhoff, M.F.R., Roozendaal, R., Kuil, A., Li, J.-P., Spaargaren, M., Pals, S.T., and Mebius, R.E. (2010). Impaired lymphoid organ development in mice lacking the heparan sulfate modifying enzyme glucuronyl C5-Epimerase. *J. Immunol.* **184**, 3656-3664.
- Reijmers, R.M., Groen, R.W.J., Kuil, A., Weijer, K., Kimberley, F.C., Medema, J.P., Kuppevelt, T.H. van, Li, J.-P., Spaargaren, M., and Pals, S.T. (2011). Disruption of heparan sulfate proteoglycan conformation perturbs B-cell maturation and APRIL-mediated plasma cell survival. *Blood* **117**, 6162- 6171.
- Religa, T.L., Sprangers, R., and Kay, L.E. (2010). Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* **328**, 98-102.
- Rickard, S.M., Mummery, R.S., Mulloy, B., and Rider, C.C. (2003). The binding of human glial cell line-derived neurotrophic factor to heparin and heparan sulfate: importance of 2-O-sulfat  groups and effect on its interaction with its receptor, GFRalpha1. *Glycobiology* **13**, 419-426.
- Richardson, T.P., Trinkaus-Randall, V., and Nugent, M.A. (2001). Regulation of heparan sulfate proteoglycan nuclear localization by fibronectin. *J. Cell Sci.* **114**, 1613-23.
- Robinson, C.J., Mulloy, B., Gallagher, J.T., and Stringer, S.E. (2006). VEGF165-binding Sites within Heparan Sulfate Encompass Two Highly Sulfated Domains and Can Be Liberated by K5 Lyase. *J. Biol. Chem.* **281**, 1731-1740.
- Rod n, L. (1980). "Structure and Metabolism of Connective Tissue Proteoglycans," in *The Biochemistry Of Glycoproteins And Proteoglycans*. Editor W. J. Lennarz (New York: Plenum Press), 267-371.
- Rodr guez, E., and Krishna, N.R. (2001). An economical method for (15)N/(13)C isotopic labeling of proteins expressed in *Pichia pastoris*, *J. Biochem.* **130**, 19-22.
- Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T., and Kay, L.E. (1996). Selective methyl group protonation of perdeuterated proteins. *J Mol Biol.* **263**, 627-636.
- Rosen, S.D., and Lemjabbar-Alaoui, H. (2010). SULF-2: an extracellular modulator of cell signaling and a cancer target candidate. *Expert Opin. Ther. Targets* **14**, 935-949.
- Roughley, P.J. (2006). The structure and function of cartilage proteoglycans. *Eur Cell Mater.* **12**, 92-101.
- Roy, A.B. (1975). L-ascorbic acid 2-sulphate. A substrate for mammalian arylsulphatases. *Biochim. Biophys. Acta* **377**, 356-363.
- Roy, S., El Hadri, A., Richard, S., Denis, F., Holte, K., Duffner, J., Yu, F., Galcheva-Gargova, Z., Capila, I., Schultes, B., *et al.* (2014b). Synthesis and biological evaluation of a unique heparin mimetic hexasaccharide for structure-activity relationship studies. *J. Med. Chem.* **57**, 4511-4520.
- Ruijter, M.H., Wagemans, T., Ijlst, L., Lund, A.M., Orchard, P.J., Schaefer, G.B., Wijburg, F.A., and van Vlies, N. (2012). Heparan sulfate and dermatan sulfate derived disaccharides are sensitive markers for newborn screening for mucopolysaccharidoses types I, II and III. *Mol Genet Metab.* **107**, 705-10.
- Sadir, R., Baleux, F., Grosdidier, A., Imberty, A., and Lortat-Jacob, H. (2001). Characterization of the Stromal Cell-derived Factor-1 -Heparin Complex. *J. Biol. Chem.* **276**, 8288-8296.
- Sadir, R., Imberty, A., Baleux, F., and Lortat-Jacob, H. (2004). Heparan Sulfate/Heparin Oligosaccharides Protect Stromal Cell-derived Factor-1 (SDF-1)/CXCL12 against Proteolysis Induced by CD26/Dipeptidyl Peptidase IV. *J. Biol. Chem.* **279**, 43854-43860.
- Salmivirta, M., and Jalkanen, M. (1995). Syndecan family of cell surface proteoglycans: developmentally regulated receptors for extracellular effector molecules. *Experientia.* **51**, 863-872.

- Samson, S.C., Ferrer, T., Jou, C.J., Sachse, F.B., Shankaran, S.S., Shaw, R.M., Chi, N.C., Tristani-Firouzi, M., and Yost, H.J. (2013). 3-OST-7 regulates BMP-dependent cardiac contraction. *PLoS Biol.* **11**, e1001727.
- Sanderson, R. D., Yang, Y., Suva, L. J. and Kelly, T. (2004). Heparan sulfate proteoglycans and heparanase--partners in osteolytic tumor growth and metastasis. *Matrix Biol.* **23**, 341-352.
- Sarrazin, S., Lamanna, W.C., and Esko, J.D. (2011). Heparan Sulfate Proteoglycans. *Cold Spring Harb. Perspect. Biol.* **3**.
- Sastry, M., Xu, L., Georgiev, I.S., Bewley, C.A., Nabel, G.J., and Kwong, P.D. (2011). Mammalian production of an isotopically enriched outer domain of the HIV-1 gp120 glycoprotein for NMR spectroscopy. *J Biomol NMR.* **50**, 197-207.
- Sattler, M., and Fesik, S.W. (1996). Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **4**, 1245-9.
- Saunders, S., Paine-Saunders, S., and Lander, A.D. (1997). Expression of the cell surface proteoglycan glypican-5 is developmentally regulated in kidney, limb, and brain. *Dev Biol.* **190**, 78-93.
- Schaefer, L., Tredup, C., Gubbiotti, M.A., and Iozzo, R.V. (2017). Proteoglycan Neofunctions: Regulation of Inflammation and Autophagy in Cancer Biology. *FEBS J.* **284**, 10-26.
- Schanda, P., Kupce, E., and Brutscher, B. (2005). SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR.* **33**, 199-211.
- Schmaus, A., Klusmeier, S., Rothley, M., Dimmler, A., Sipos, B., Faller, G., Thiele, W., Allgayer, H., Hohenberger, P., Post, S., et al. (2014). Accumulation of small hyaluronan oligosaccharides in tumour interstitial fluid correlates with lymphatic invasion and lymph node metastasis. *British J Cancer* **111**, 559-67.
- Schick, B.P., Gradowski, J.F., and San Antonio, J.D. (2001). Synthesis, secretion, and subcellular localization of serglycin proteoglycan in human endothelial cells. *Blood* **97**, 449-458.
- Schmidt, B., Selmer, T., Ingendoh, A., and von Figura, K. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **82**, 271-278.
- Schulte, T.W., Akinaga, S., Soga, S., Sullivan, W., Stensgard, B., Toft, D., and Neckers, L.M. (1998). Antibiotic radicicol binds to the N-terminal domain of Hsp90 and shares important biologic activities with geldanamycin. *Cell Stress Chaperones* **3**, 100-108.
- Sedita, J., Izvolsky, K., and Cardoso, W.V. (2004). Differential expression of heparan sulfate 6-O-sulfotransferase isoforms in the mouse embryo suggests distinctive roles during organogenesis. *Dev. Dyn.* **231**, 782-794.
- Seffouh, A., Milz, F., Przybylski, C., Laguri, C., Oosterhof, A., Bourcier, S., Sadir, R., Dutkowski, E., Daniel, R., van Kuppevelt, T.H., et al. (2013). HSulf sulfatases catalyze processive and oriented 6-O- desulfation of heparan sulfate that differentially regulates fibroblast growth factor activity. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **27**, 2431-2439.
- Seffouh, I., Przybylski, C., Seffouh, A., El Masri, R., Vivès, R.R., Gonnet, F., and Daniel, R. (2019). Mass spectrometry analysis of the human endosulfatase Hsulf-2. *Biochem. Biophys. Rep.* **18**.
- Shao, H., Lee, S., Gae-Scott, S., Nakata, C., Chen, S., Hamad, A.R., et al. (2012). *Extracellular matrix lumican promotes bacterial phagocytosis, and Lum-/- mice show increased Pseudomonas aeruginosa lung infection severity.* *J Biol Chem.* **287**, 35860-35872.
- Shaw, J.P., Johnson, Z., Borlat, F., Zwahlen, C., Kungl, A., Roulin, K., Harrenga, A., Wells, T.N.C., and Proudfoot, A.E.I. (2004). The X-ray structure of RANTES: heparin-derived disaccharides allows the rational design of chemokine inhibitors. *Struct. Lond. Engl.* **12**, 2081-2093.
- Sheng, J., Liu, R., Xu, Y., and Liu, J. (2011). The Dominating Role of N-Deacetylase/N- Sulfotransferase 1 in Forming Domain Structures in Heparan Sulfate. *J. Biol. Chem.* **286**, 19768- 19776.

- Shi, X., and Jarvis, D.L. (2018). Protein N-Glycosylation in the Baculovirus-Insect Cell System, *Curr. Drug Targets*. **8** (2007) 1116-1125.S.
- Shirsat, N., Avesh, M., English, N.J., Glennon, B., and Al-Rubeai, M. (2013). Application of statistical techniques for elucidating flow cytometric data of batch and fed-batch cultures. *Biotechnol. Appl. Biochem.* **60**, 536-545.
- Shworak, N.W., Fritze, L.M.S., Liu, J., Butler, L.D., and Rosenberg, R.D. (1996). Cell-free synthesis of anticoagulant heparan sulfate reveals a limiting converting activity that modifies an excess precursor pool. *J Biol Chem.* **271**, 27063-27071.
- Shworak, N.W., Liu, J., Fritze, L.M.S., Schwartz, J.J., Zhang, L.J., Logeart, D., and Rosenberg, R.D. (1997). Molecular cloning and expression of mouse and human cDNAs encoding heparan sulfate D-glucosaminyl 3-O- sulfotransferase. *J Biol Chem.* **272**, 28008-28019.
- Shworak, N.W., Liu, J., Petros, L.M., Zhang, L., Kobayashi, M., Copeland, N.G., Jenkins, N.A., and Rosenberg, R.D. (1999). Multiple Isoforms of Heparan Sulfate d-Glucosaminyl 3-O-sulfotransferase isolation, characterization, and expression of human cDNAs and identification of distinct genomic loci. *J. Biol. Chem.* **274**, 5170-5184.
- Silbert, J.E., and Sugumaran, G. (1995). Intracellular membranes in the synthesis, transport, and metabolism of proteoglycans. *Biochim. Biophys. Acta BBA - Rev. Biomembr.* **1241**, 371-384.
- Simizu, S., Ishida, K., Wierzba, M. K. and Osada, H. (2004). Secretion of heparanase protein is regulated by glycosylation in human tumor cell lines. *J Biol Chem.* **279**, 2697-2703
- Sitarska, A., Skora, L., Klopp, J., Roest, S., Fernandez, C., Shrestha, B., and Gossert, A.D. (2015). Affordable uniform isotope labeling with 2 H, 13 C and 15 N in insect cells. *J Biomol NMR.* **62**, 191-197.
- Skliris, A., Labropoulou, V.T., Papachristou, D.J., Aletras, A., Karamanos, N.K., Theocharis, A.D. (2013). Cell-surface seryglycin promotes adhesion of myeloma cells to collagen type I and affects the expression of matrix metalloproteinases. *FEBS J.* **280**, 2342-52.
- Skora, L., Binesh Shrestha B., and Gossert. A.D. (2015). Isotope labeling of proteins in insect cells. *Methods Enzymol.* **565**, 245-88.
- Smeds, E., Habuchi, H., Do, A.-T., Hjertson, E., Grundberg, H., Kimata, K., Lindahl, U., and Kusche-Gullberg, M. (2003). Substrate specificities of mouse heparan sulphate glucosaminyl 6-O-sulphotransferases. *Biochem. J.* **372**, 371-380.
- Soares da Costa, D., Reis, R.L., and Pashkuleva, I. (2017). Sulfation of Glycosaminoglycans and Its Implications in Human Health and Disorders. *Annu Rev Biomed Eng.* **21**, 19:1-26.
- Solà, A., Maaheimo, H., Ylönen, K., Ferrer, P., and Szyperski, T. (2004). Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* **271**, 2462-2470.
- Solt, A.S., Bostock, M.J., Shrestha, B., Kumar, P., Warne, T., Tate, C.G., and Nietlispach, D. (2017). Insight into partial agonism by observing multiple equilibria for ligand-bound and Gs-mimetic nanobody-bound  $\beta$ 1-adrenergic receptor. *Nat. Commun.* **8**, 1795.
- Sommarin, Y., Wendel, M., Shen, Z., Hellman, U., and Heinegård, D. (1998). *J. Biol. Chem.* **273**, 16723-16739.
- Spillmann, D., Witt, D., and Lindahl, U. (1998). Defining the Interleukin-8-binding Domain of Heparan Sulfate. *J. Biol. Chem.* **273**, 15487-15493.
- Sprangers, R. and Kay, L.E. (2007a). Probing supramolecular structure from measurement of methyl (1)H-(13)C residual dipolar couplings. *J Am Chem Soc.* **129**, 12668-12669.
- Sprangers, R. and Kay, L.E. (2007b). Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature.* **445**, 618-622.
- Sprangers, R., Velyvis, A., and Kay, L.E. (2007c). Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods.* **4**, 697-703.

- Staples, G.O., Shi, X., and Zaia, J. (2011). Glycomics analysis of mammalian heparan sulfates modified by the human extracellular sulfatase HSulf2. *PLoS One* **6**, e16689.
- Staub, J., Chien, J., Pan, Y., Qian, X., Narita, K., Aletti, G., Scheerer, M., Roberts, L.R., Molina, J., and Shridhar, V. (2007). Epigenetic silencing of HSulf-1 in ovarian cancer: implications in chemoresistance. *Oncogene* **26**, 4969-4978.
- Stein, C., Hille, A., *et al.* (1989). Cloning and expression of human steroid sulfatase. Membrane topology, glycosylation, and subcellular distribution in BHK-21 cells. *J Biol Chem.* **264**, 13865-13872.
- Steinmetz, A., Vyazmensky, M., Meyer, D., Barak, Z.E., Golbik, R., Chipman, D.M., and Tittmann, K. (2010). Valine 375 and phenylalanine 109 confer affinity and specificity for pyruvate as donor substrate in acetohydroxy acid synthase isozyme II from *Escherichia coli*. *Biochemistry* **49**, 5188-5199.
- Stoffregen, M.C., Schwer, M.M., Renschler, F.A., and Wiesner, S. (2012). Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces. *Structure* **20**, 573-581.
- Strand, M.E., Aronsen, J.M., Braathen, B., Sjaastad, I., Kvaløy, H., Tønnessen, T., Christensen, G., and Lunde, I.G. (2015). Shedding of syndecan-4 promotes immune cell recruitment and mitigates cardiac dysfunction after lipopolysaccharide challenge in mice. *J. Mol. Cell. Cardiol.* **88**, 133-144.
- Strauss, A., Bitsch, F., Cutting, B., Fendrich, G., Graff, P., Liebetanz, J., Zurini, M., and Jahnke, W. (2003). Amino-acid-type selective isotope labeling of proteins expressed in baculovirus-infected insect cells useful for NMR studies. *J Biomol NMR.* **26**, 367-372.
- Strauss, A., Bitsch, F., Fendrich, G., Graff, P., Knecht, R., Meyhack, B., and Jahnke, W. (2005). Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells, *J. Biomol. NMR.* **31**, 343-349.
- Stressler, T., Seitzl, I., Kuhn, A., and Fischer, L. (2016). Detection, production, and application of microbial arylsulfatases. *Appl. Microbiol. Biotechnol.* **100**, 9053-9067.
- Struve, J., Maher, P.C., Li, Y.Q., Kinney, S., Fehlings, M.G., Kuntz, I.V.C., and Sherman, L.S. (2005). Disruption of the hyaluronan-based extracellular matrix in spinal cord promotes astrocyte proliferation. *Glia* **52**, 16-24.
- Sugahara, K., and Schwartz, N. B. (1979). Defect in 3'-phosphoadenosine 5'- phosphosulfate Formation in Brachymorphic Mice. *Proc. Natl. Acad. Sci.* **76**, 6615-6618.
- Sugahara, K., and Kojima, T. (1996). Specificity studies of bacterial sulfatases by means of structurally defined sulfated oligosaccharides isolated from shark cartilage chondroitin sulfate D. *Eur. J. Biochem.* **239**, 865- 870.
- Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K., and Kitagawa, H. (2003). Recent Advances in the Structural Biology of Chondroitin Sulfate and Dermatan Sulfate. *Curr. Opin. Struct. Biol.* **13**, 612-620.
- Sugahara, K., and Mikami, T. (2007). Chondroitin/dermatan Sulfate in the central Nervous System. *Curr. Opin. Struct. Biol.* **17**, 536-545.
- Sugaya, N., Habuchi, H., Nagai, N., Ashikari-Hada, S., and Kimata, K. (2008). 6-O-sulfation of heparan sulfate differentially regulates various fibroblast growth factor-dependent signalings in culture. *J. Biol. Chem.* **283**, 10366-10376.
- Sugiki, T., Ichikawa, O., Miyazawa-Onami, M., Shimada, I., and Takahashi, H. (2012). Isotopic labeling of heterologous proteins in the yeast *Pichia pastoris* and *Kluyveromyces lactis*. *Methods Mol. Biol. Clifton NJ.* **831**, 19-36.
- Suzuki, R., Sakakura, M., Mori, M., Fujii, M., Akashi, S., and Takahashi, H. (2018). Methyl-selective isotope labeling using  $\alpha$ -ketoisovalerate for the yeast *Pichia pastoris* recombinant protein expression system, *J. Biomol. NMR.* **71**, 213-223.

- Taghizadeh, E., Kalantar, S.M., Mahdian, R., Sheikhha, M.H., Farashahi-Yazd, E., Ghasemi, S., and Shahbazi, Z. (2015). SULF 1 gene polymorphism, rs6990375 is in significant association with fetus failure in IVF technique. *Iran. J. Reprod. Med.* **13**, 215-220.
- Takemae, H., Ueda, R., Okubo, R., Nakato, H., Izumi, S., Saigo, K., *et al.* (2003). Proteoglycan UDP-Galactose:β-Xylose β1,4-Galactosyltransferase I Is Essential for Viability in *Drosophila Melanogaster*. *J. Biol. Chem.* **278**, 15571-15578.
- Takemura, M., Noborn, F., Nilsson, J., Bowden, N., Nakato, E., Baker, S., *et al.* (2020). Chondroitin Sulfate Proteoglycan Windpipe Modulates Hedgehog Signaling in *Drosophila*. *Mol. Biol. Cell* **31**, 813-824.
- Tamura, K., Fukuda, Y., Ishizaki, M., Masuda, Y., Yamanaka, N., and Ferrans, V.J. (1995). Abnormalities in elastic fibers and other connective-tissue components of floppy mitral valve. *Am Heart J.* **129**, 1149-1158.
- Tang, R., and Rosen, S.D. (2009). Functional Consequences of the Subdomain Organization of the Sulfs. *J. Biol. Chem.* **284**, 21505-21514.
- Tecle, E., Diaz-Balzac, C.A., and Bulow, H.E. (2013). Distinct 3-O-sulfated heparan sulfate modification patterns are required for kal-1-dependent neurite branching in a context-dependent manner in *Caenorhabditis elegans*. *G3: Genes, Genomes, Genet.* **3**, 541-552.
- Teng, Y.H.-F., Aquino, R.S., and Park, P.W. (2012). Molecular functions of syndecan-1 in disease. *Matrix Biol. J. Int. Soc. Matrix Biol.* **31**, 3-16.
- Thacker, B.E., Xu, D., Lawrence, R., and Esko, J.D. (2013). Heparan sulfate 3-O-sulfation: A rare modification in search of a function. *Matrix Biol.* **35**, 60-72.
- Thacker, B.E., Xu, D., Lawrence, R., and Esko, J.D. (2014). Heparan sulfate 3-O-sulfation: A rare modification in search of a function. *Matrix Biol. J. Int. Soc. Matrix Biol.* **35**, 60-72.
- Thacker, B.E., Seamen, E., Lawrence, R., Parker, M.W., Xu, Y., Liu, J., Vander Kooi, C.W., and Esko, J.D. (2016). Expanding the 3-O-sulfated Proteome--Enhanced Binding of Neuropilin-1 to 3-O-sulfated Heparan Sulfate Modulates Its Activity. *ACS Chem Biol.* **11**, 971-80.
- Thelin, M. A., Bartolini, B., Axelsson, J., Gustafsson, R., Tykesson, E., Pera, E., *et al.* (2013). Biological Functions of Iduronic Acid in Chondroitin/dermatan Sulfate. *FEBS J.* **280**, 2431-2446.
- Theocharis, A.D., Seidel, C., Borset, M., Dobra, K., Baykov, V., Labropoulou, V., Kanakis, I., Dalas, E., Karamanos, N.K., Sundan, A., *et al.* (2006). Serglycin Constitutively Secreted by Myeloma Plasma Cells Is a Potent Inhibitor of Bone Mineralization in Vitro. *J. Biol. Chem.* **281**, 35116-35128.
- Theocharis, A.D., and Karamanos, N.K. (2017). Proteoglycans Remodeling in Cancer: Underlying Molecular Mechanisms. *Matrix Biol.* doi: 10.1016/j.matbio.2017.10.008.
- Thompson, L.D., Pantoliano, M.W. and Springer, B.A. (1994). Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain. *Biochemistry* **33**, 3831-3840.
- Thoms, S., Max, K.E.A., Wunderlich, M., Jacso, T., Lilie, H., Reif, B., Heinemann, U., and Schmid, F.X. (2009). Dimer Formation of a Stabilized Gβ1 Variant: A Structural and Energetic Analysis. *J. Mol. Biol.* **391**, 918-932
- Tian, J., Ling, L., Shboul, M., Lee, H., O'Connor, B., Merriman, B., *et al.* (2010). Loss of CHSY1, a Secreted FRINGE Enzyme, Causes Syndromic Brachydactyly in Humans via Increased NOTCH Signaling. *Am. J. Hum. Genet.* **87**, 768-778.
- Tomatsu, S., Fujii, T., Fukushi, M., Oguma, T., Shimada, T., Maeda, M., Kida, K., Shibata, Y., Futatsumori, H., Montañó, A.M., Mason, R.W., Yamaguchi, S., Suzuki, Y., and Orii, T. (2013). Newborn screening and diagnosis of mucopolysaccharidoses. *Mol Genet Metab.* **110**, 42-53.
- Torchia, D.A., Sparks, S.W., and Bax, A. (1988). NMR signal assignments of amide protons in the alpha-helical domains of staphylococcal nuclease. *Biochemistry* **27**, 5135-5141.



- Traenkle, B., Kaiser, P.D., and Rothbauer, U. (2016). Nanobody platform for determination of protein structure, application of peptides and proteins. Doi : 10.1002/9780470027318.a9548
- Tsai, T.-T., Ho, N.Y.-J., Fang, H.-C., Lai, P.-L., Niu, C.-C., Chen, L.-H., Chen, W.-J., and Pang, J.-H.S. (2015). Increased sulfatase 1 gene expression in degenerative intervertebral disc cells. *J. Orthop. Res. Off. Publ. Orthop. Res. Soc.* **33**, 312-317.
- Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E., and Kay, L.E. (2003). Cross-correlated relaxation enhanced  $^1\text{H}$ [bond] $^{13}\text{C}$  NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420-10428.
- Tugarinov, V. and Kay, L.E. (2004). An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR.* **28**, 165-172.
- Tugarinov, V. and Kay, L.E. (2004). Stereospecific NMR assignments of prochiral methyls, rotameric states and dynamics of valine residues in malate synthase G. *J. Am. Chem. Soc.* **126**, 9827-9836.
- Tugarinov, V., Choy, W.Y., Orekhov, V.Y., and Kay, L.E. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc Natl Acad Sci USA.* **102**, 622-627.
- Tugarinov, V., Sprangers, R., and Kay, L.E. (2007). Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 1743-1750.
- Uchański, T., Pardon, E., and Steyaert, J. (2020). Nanobodies to study protein conformational states. *Curr Opin Struct Biol.* **60**, 117-123.
- Uchimura, K., Morimoto-Tomita, M., Bistrup, A., Li, J., Lyon, M., Gallagher, J., Werb, Z., and Rosen, S.D. (2006). HSulf-2, an extracellular endoglucosamine-6-sulfatase, selectively mobilizes heparin-bound growth factors and chemokines: effects on VEGF, FGF-1, and SDF-1. *BMC Biochem.* **7**, 2.
- Ulmer, J.E., Vilén, E.M., Namburi, R.B., Benjdia, A., Beneteau, J., Malleron, A., Bonnaffé, D., Driguez, P.A., Descroix, K., Lassalle, G., Le Narvor, C., Sandström, C., Spillmann D., and Berteau O. (2014). Characterization of Glycosaminoglycan (GAG) Sulfatases from the Human Gut Symbiont *Bacteroides thetaiotaomicron* Reveals the First GAG-specific Bacterial Endosulfatase. *J. Biol. Chem.* **289**, 24289-24303.
- Vallet, S.D., Clerc, O., and Ricard-Blum, S. (2021). Glycosaminoglycan-Protein Interactions: The First Draft of the Glycosaminoglycan Interactome. *J. Histochem Cytochem.* **69**, 93-104.
- Vallet, S.D., Berthollier, C., Ricard-Blum, S. (2022). L'interaction glycosaminoglycane 2.0. *Am J Physiol Cell Physiol.* **322**, C1271-C1278.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, 439-444.
- Varki, et al., (2008). *Essentials of Glycobiology*, 2nd edition, Cold Spring Harbor Press, pp. 236-237.
- Velyvis, A., Yang, Y.R., Schachman, H.K., and Kay, L.E. (2007). A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc Natl Acad Sci USA.* **104**, 8815-8820.
- Velyvis, A., Ruschak, A.M., and Kay, L.E. (2012). An economical method for production of  $(2)\text{H}$ ,  $(13)\text{CH}_3$ -threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One* **7**, 43725.
- Venters, R.A., Huang, C.C., Farmer, B.T., Trolard, R., Spicer, L.D., and Fierke, C.A. (1995). High-level  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$  labeling of proteins for NMR studies. *J. Biomol. NMR.* **5**, 339-344.
- Vivès, R.R., Seffouh, A., and Lortat-Jacob, H. (2014). Post-Synthetic regulation of HS structure: the yin and yang of the Sulfs in cancer. *Front. Oncol.* **3**.

- Vlodavsky, I., Abboud-Jarrous, G., Elkin, M., Naggi, A., Casu, B., Sasisekharan, R. and Ilan, N. (2006). The impact of heparanase and heparin on cancer metastasis and angiogenesis. *Pathophysiol Haemost Thromb.* **35**, 116-127.
- Vlodavsky, I., Gross-Cohen, M., Weissmann, M., Ilan, N., and Sanderson, R.D. (2018). Opposing Functions of Heparanase-1 and Heparanase-2 in Cancer Progression. *Trends Biochem. Sci.* **43**, 18-31.
- von Bülow, R., Schmidt, B., Dierks, T., von Figura, K., and Uson, I. (2001). Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis. *J. Mol. Biol.* **305**, 269-77.
- Waldow, A., Schmidt, B., Dierks, T., von Bülow, R., and von Figura, K. (1999). Amino acid residues forming the active site of arylsulfatase A. Role in catalytic activity and substrate binding. *J. Biol. Chem.* **274**, 12284-8.
- Walton, W.J., Kasprzak, A.J., Hare, J.T., and Logan, T.M. (2006). An economic approach to isotopic enrichment of glycoproteins expressed from Sf9 insect cells. *J Biomol NMR.* **36**, 225-233.
- Wang, L., Brown, J.R., Varki, A., and Esko, J.D. (2002). Heparins anti-inflammatory effects require glucosamine 6-O-sulfation and are mediated by blockade of L- and P-selectins. *J. Clin. Invest.* **110**, 127-136.
- Wang, L., Xie, L., Wang, J., Shen, J., and Liu, B. (2013). Correlation between the methylation of SULF2 and WRN promoter and the irinotecan chemosensitivity in gastric cancer. *BMC Gastroenterol.* **13**, 173.
- Wang, W., Han, W., Cai, X., Zheng, X., Sugahara, K., and Li, F. (2015). Cloning and characterization of a novel chondroitin sulfate/dermatan sulfate 4-O-endosulfatase from a marine bacterium. *J. Biol. Chem.* **290**, 7823-7832.
- Wang, S., Ai, X., Freeman, S.D., Pownall, M.E., Lu, Q., Kessler, D.S., and Emerson, C.P. (2004). QSulf1, a heparan sulfate 6-O-endosulfatase, inhibits fibroblast growth factor signaling in mesoderm induction and angiogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4833-4838.
- Wang, S., Guan, J., Zhang, Q., Chen, X., and Li, F. (2019). Identification and Signature Sequences of Bacterial Hexuronate-2-O-sulfatases. *Front. Microbiol.* **10**, 704.
- Watanabe, Y., Takeuchi, K., Higa Onaga, S., Sato, M., Tsujita, M., Abe, M., et al. (2010). Chondroitin Sulfate N-Acetylgalactosaminyltransferase-1 Is Required for normal Cartilage Development. *Biochem. J.* **432**, 47-55.
- Wei, G., Bai, X., Bame, K.J., Koshy, T.I., Spear, P.G., and Esko, J.D. (2000). Location of the glucuronosyltransferase domain in the heparan sulfate copolymerase (EXT1) by analysis of Chinese hamster ovary cell mutants. *J. Biol. Chem.* **275**, 27733-27740.
- Weigel, P.H., Hascall, V.C., and Tammi, M. (1997). Hyaluronan Synthases. *J. Biol. Chem.* **272**, 13997-14000.
- Weigelt, J., van Dongen, M., Uppenberg, J., Schultz J., and Wikstrom M. (2002). Site-selective screening by NMR spectroscopy with labeled amino acid pairs. *J. Am. Chem. Soc.* **124**, 2446-2447.
- Weininger, U., Liu, Z., McIntyre, D.D. Vogel, H.J., and Akke, M. (2012). Specific  $^{12}\text{C}\beta\text{D}(2)^{12}\text{C}\gamma\text{D}(2)^{13}\text{C}\epsilon\text{HD}(2)$  isotopomer labeling of methionine to characterize protein dynamics by  $^1\text{H}$  and  $^{13}\text{C}$  NMR relaxation dispersion. *J. Am. Chem. Soc.* **134**, 18562-18565.
- Weiss, R.J., Esko, J.D., and Tor, Y. (2017). Targeting heparin and heparan sulfate protein interactions. *Org Biomol Chem.* **15**, 5656-5668.
- Weissmann, M., Bhattacharya, U., Feld, S., Hammond, E., Ilan, N., and Vlodavsky, I. (2018). The heparanase inhibitor PG545 is a potent anti-lymphoma drug: Mode of action. *Matrix Biol. J. Int. Soc. Matrix Biol.*
- Wen, J., Xiao, J., Rahdar, M., Choudhury, B.P., Cui, J., Taylor, G.S., et al. (2014). Xylose Phosphorylation Functions as a Molecular Switch to Regulate Proteoglycan Biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15723-15728.

- Werner, K., Lehner, I., Dhiman, H.K., Richter, C., Glaubitz, C., Schwalbe, H., Klein-Seetharaman, J., and Khorana, H.G. (2007). Combined solid state and solution NMR studies of alpha,epsilon-15N labeled bovine rhodopsin. *J. Biomol. NMR.* **37**, 03-312.
- Werner, K., Richter, C., Klein-Seetharaman, J., and Schwalbe, H. (2008). Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminal inus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J. Biomol. NMR.* **40**, 49-53.
- Weyers, A., Yang, B., Solakyildirim, K., Yee, V., Li, L., Zhang, F., and Linhardt, R.J. (2013). Isolation of bovine corneal keratan sulfate and its growth factor and morphogen binding. *FEBS J.* **280**, 2285-2293.
- Whitelock, J.M., Melrose, J., and Iozzo, R.V. (2008). Diverse Cell Signaling Events Modulated by Perlecan. *Biochemistry* **47**, 11174-11183.
- Wight, T.N. (2002). Versican: a versatile extracellular matrix proteoglycan in cell biology. *Curr Opin Cell Biol.* **14**, 617-623.
- Wilson, D.G., Phamluong, K., Lin, W.Y., Barck, K., Carano, R.A.D., Diehl, L., et al. (2012). Chondroitin Sulfate Synthase 1 (Chsy1) Is Required for Bone Development and Digit Patterning. *Dev. Biol.* **363**, 413-425.
- Wolny, P.M., Banerji, S., Gounou, C., Brisson, A.R., Day, A.J., Jackson, D.G., and Richter, R.P. (2010). Analysis of CD44-hyaluronan interactions in an artificial membrane system: insights into the distinct binding properties of high and low molecular weight hyaluronan. *J Biol Chem.* **285**, 30170-80.
- Wong, D.T., Fuller, R.W., and Molloy, B.B. (1973). Inhibition of amino acid transaminases by L-cycloserine. *Advances in Enzyme Regulation* **11**, 139-154.
- Woods, A. and Couchman, J.R. (1994). Syndecan 4 heparan sulfate proteoglycan is a selectively enriched and widespread focal adhesion component. *Mol. Biol. Cell.* **5**, 183-192.
- Wood, M.J., and Komives, E.A. (1999). Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation, *J. Biomol. NMR.* **13**, 149-159.
- Wu, L., Viola, C.M., Brzozowski, A.M., and Davies, G.J. (2015). Structural characterization of human heparanase reveals insights into substrate recognition. *Nat. Struct. Mol. Biol.* **22**, 1016-1022.
- Wu, X., and Rapoport, T.A. (2021). Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). Doi : <https://doi.org/10.1101/2021.08.09.455715>.
- Wuyts, W., and Van Hul, W. (2000). Molecular basis of multiple exostoses: mutations in the EXT1 and EXT2 genes. *Hum. Mutat.* **15**, 220-227.
- Xia, G., Chen, J., Tiwari, V., Ju, W., Li, J.-P., Malmström, A., Shukla, D., and Liu, J. (2002). Heparan Sulfate 3-O-sulfotransferase isoform 5 generates both an antithrombin-binding Site and an entry receptor for herpes simplex virus, Type 1. *J. Biol. Chem.* **277**, 37912-37919.
- Xu, D., Moon, A.F., Song, D., Pedersen, L.C., and Liu, J. (2008). Engineering sulfotransferases to modify heparan sulfate. *Nat. Chem. Biol.* **4**, 200-202.
- Xu, R., Ori, A., Rudd, T.R., Uniewicz, K.A., Ahmed, Y.A., Guimond, S.E., Skidmore, M.A., Siligardi, G., Yates, E.A., and Fernig, D.G. (2012). Diversification of the structural determinants of fibroblast growth factor-heparin Interactions implications for being specificity. *J. Biol. Chem.* **287**, 40061-40073.
- Xu, D., and Esko, J.D. (2014). Demystifying heparan sulfate-protein interactions. *Annu Rev Biochem.* **83**, 129-57.
- Xu, G., Ji, W., Su, Y., Xu, Y., Yan, Y., Shen, S., Li, X., Sun, B., Qian, H., Chen, L., et al. (2014). Sulfatase 1 (hSulf-1) reverses basic fibroblast growth factor-stimulated signaling and inhibits growth of hepatocellular carcinoma in animal model. *Oncotarget* **5**, 5029-5039.
- Yagi, H., Nakamura, M., Yokoyama, J., Zhang, Y., Yamaguchi, T., Kondo, S., Kobayashi, J., Kato, T., Park, E.Y., Nakazawa, S., Hashii, N., Kawasaki, N., and Kato, K. (2015). Stable isotope labeling of glycoprotein expressed in silkworms using immunoglobulin G as a test molecule, *J. Biomol. NMR.* **62**, 157-167.

- Yamada, S., and Sugahara, K. (2008). Potential Therapeutic Application of Chondroitin Sulfate/dermatan Sulfate. *Curr. Drug Discov. Technol.* **5**, 289-301.
- Yamaguchi, Y., Nishimura, M., Nagano, M., Yagi, H., Sasakawa, H., Uchida, K., Shitara, K., and Kato, K. (2006). Glycoform-dependent conformational alteration of the Fc region of human immunoglobulin G1 as revealed by NMR spectroscopy. *Biochimica et Biophysica Acta.* **1760**, 693-700.
- Yamaguchi, Y., Takahashi, N., and Kato, K. (2007). Molecular interactions: antibody structures. In: Kamerling JP (ed) *Comprehensive glycoscience*. Elsevier, Oxford 745-763.
- Yamaguchi, Y., and Kato, K. (2010). Dynamics and interactions of glycoconjugates probed by stable-isotope-assisted NMR spectroscopy. *Methods Enzymol.* **478**, 305-322.
- Yamazaki, T., Lee, L., Arrowsmith, C.H, Muhandiram, D.R., and Kay, L.E. (1994). A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H Labeled Proteins with High Sensitivity. *Journal American Chemistry Society* **116**, 11655-11666.
- Yamazaki, T., Weontae, L., Muhandiram, R., and Kay, L.E. (1994). NMR Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly <sup>13</sup>C,<sup>15</sup>N-Labeled Proteins: Application to <sup>13</sup>C.alpha. Carbons. *Journal American Chemistry Society* **116**, 8266-8278.
- Yan, D., and Lin, X. (2009). Shaping Morphogen Gradients by Proteoglycans. *Cold Spring Harb. Perspect. Biol.* **1**, a002493.
- Yanaka, S, Yagi, H., Yogo, R., Yagi-Utsumi, M., and Kato, K. (2018). Stable isotope labeling approaches for NMR characterization of glycoproteins using eukaryotic expression systems. *J. Biomol. NMR.* **71**, 193-202.
- Yang, Y., Yaccoby, S., Liu, W., Langford, J.K., Pumphrey, C.Y., Theus, A., Epstein, J., and Sanderson, R.D. (2002). Soluble syndecan-1 promotes growth of myeloma tumors in vivo. *Blood* **100**, 610-617.
- Yang, N., and Friedl, A. (2016). Syndecan-1-induced ECM fiber alignment requires integrin avb3 and syndecan-1 ectodomain and heparan sulfate chains. *PLoS ONE* **11**, e0150132.
- Yang, C., Cao, M., Liu, H., He, Y., Xu, J., Du, Y., Liu, Y., Wang, W., Cui, L., Hu, J., et al. (2012). The high and low molecular weight forms of hyaluronan have distinct effects on CD44 clustering. *J Biol Chem.* **287**, 43094-107.
- Yayon, A., Klagsbrun, M., Esko, J.D., Leder, P., and Ornitz, D.M. (1991). Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell* **64**, 841-848.
- Yen, P.H., Allen, E., Marsh, B., et al. (1987). Cloning and expression of steroid sulfatase cDNA and the frequent occurrence of deletions in STS deficiency: implications for X-Y interchange. *Cell.* **49**, 443- 454.
- Yoneyama, H., Hori, H., Lim, S.J., Murata, T., Ando, T., Isogai, E., and Katsumata, R. (2011). Isolation of a mutant auxotrophic for L-alanine and identification of three major aminotransferases that synthesize L-alanine in *Escherichia coli*. *Biosci Biotechnol Biochem.* **75**, 930-938.
- Yue, X., Li, X., Nguyen, H.T., Chin, D.R., Sullivan, D.E., and Lasky, J.A. (2008). Transforming Growth Factor-β1 Induces Heparan Sulfate 6-O-Endosulfatase 1 Expression in Vitro and in Vivo. *J. Biol. Chem.* **283**, 20397-20407.
- Yue, X., Lu, J., Auduong, L., Sides, M.D., and Lasky, J.A. (2013). Overexpression of Sulf2 in idiopathic pulmonary fibrosis. *Glycobiology* **23**, 709-719.
- Yue, X. (2017). Epithelial Deletion of Sulf2 Exacerbates Bleomycin-Induced Lung Injury, Inflammation, and Mortality. *Am. J. Respir. Cell Mol. Biol.* **57**, 560-569.
- Yurt, R.W., Leid, R.W., and Austen, K.F. (1977). Native heparin from rat peritoneal mast cells. *J Biol Chem.* **252**, 518-521.
- Zhang, L., and Esko, J.D. (1994). Amino acid determinants that drive heparan sulfate assembly in a proteoglycan. *J Biol Chem.* **269**, 19295-19299.

- Zhang, L., Beeler, D.L., Lawrence, R., Lech, M., Liu, J., Davis, J.C., Shriver, Z., Sasisekharan, R., and Rosenberg, R.D. (2001). 6-O-sulfotransferase-1 represents a critical enzyme in the anticoagulant heparan sulfate biosynthetic pathway. *J. Biol. Chem.* **276**, 42311-42321.
- Zhang, L., Yang, M., Yang, D., Cavey, G., Davidson, P., and Gibson, G. (2010). Molecular Interactions of MMP-13 C-Terminal Domain with Chondrocyte Proteins. *Connect. Tissue Res.* **51**, 230-239.
- Zhang, H., Newman, D.R., and Sannes, P.L. (2012a). HSULF-1 inhibits ERK and AKT signaling and decreases cell viability in vitro in human lung epithelial cells. *Respir. Res.* **13**, 69.
- Zhang, S., Condac, E., Qiu, H., Jiang, J., Gutierrez-Sanchez, G., Bergmann, C., Handel, T., and Wang, L. (2012b). Heparin-induced leukocytosis requires 6-O-sulfation and is caused by blockade of selectin- and CXCL12 protein-mediated leukocyte trafficking in mice. *J. Biol. Chem.* **287**, 5542- 5553.
- Zhang, H., Wood, T., Young, S.P., and Millington, D.S. (2015). A straightforward, quantitative ultra-performance liquid chromatography-tandem mass spectrometric method for heparan sulfate, dermatan sulfate and chondroitin sulfate in urine: an improved clinical screening test for the mucopolysaccharidoses. *Mol Genet Metab.* **114**, 123-8.
- Zhang, M., Yu, X.W., Xu, Y., Jouhten, P., Swapna, G.V.T., Glaser, R.W., Hunt, J.F., Montelione, G.T., Maaheimo, H., and Szyperski, T. (2017). <sup>13</sup>C metabolic flux profiling of *Pichia pastoris* grown in aerobic batch cultures on glucose revealed high relative anabolic use of TCA cycle and limited incorporation of provided precursors of branched-chain amino acids. *FEBS J.* **284**, 3100-3113.
- Zhao, W., Sala-Newby, G.B., and Dhoot, G.K. (2006). Sulf1 expression pattern and its role in cartilage and joint development. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **235**, 3327-3335.
- Zhu, C., HE, L., Zhou, X., Nie, X., and Gu, Y. (2016). Sulfatase 2 promotes breast cancer progression through regulating some tumor-related factors. *Oncol. Rep.* **35**, 1318-1328.
- Zlotnik, A., and Yoshie, O. (2012). The Chemokine Superfamily Revisited. *Immunity* **36**, 705-716.
- Zong, F., Fthenou, E., Wolmer, N., Hollósi, P., Kovalszky, I., Szilák, L., Mogler, C., Nilsonne, G., Tzanakakis, G., and Dobra, K. (2009). Syndecan-1 and FGF-2, but Not FGF Receptor-1, Share a Common Transport Route and Co-Localize with Heparanase in the Nuclei of Mesenchymal Tumor Cells. *PLoS ONE* **4**.



