



**HAL**  
open science

# Caractérisation d'une famille de protéines intrinsèquement désordonnées, les syndécans

Frank Gondelaud

► **To cite this version:**

Frank Gondelaud. Caractérisation d'une famille de protéines intrinsèquement désordonnées, les syndécans. Biochimie, Biologie Moléculaire. Université de Lyon, 2020. Français. NNT : 2020LYSE1160 . tel-03376362

**HAL Id: tel-03376362**

**<https://theses.hal.science/tel-03376362>**

Submitted on 13 Oct 2021

**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.



N°d'ordre NNT : 2020LYSE1160

## **THESE de DOCTORAT DE L'UNIVERSITE DE LYON**

opérée au sein de

**l'Université Claude Bernard Lyon 1**

**Ecole Doctorale N° 205**

**Ecole Doctorale Interdisciplinaire Sciences – Santé**

**Spécialité de doctorat : BIOCHIMIE**

Soutenue publiquement le 16/10/2020, par :

**Frank GONDELAUD**

---

# **Caractérisation d'une famille de protéines intrinsèquement désordonnées, les syndécans**

---

Devant le jury composé de :

Dr Sonia Longhi	Directeur de Recherche CNRS	Université Aix Marseille	Rapporteuse
Pr Herman van Tilbeurgh	Professeur des Universités	Université Paris-Saclay	Rapporteur
Pr Agnès Girard-Egrot	Professeur des Universités	Université Lyon 1	Examinatrice
Dr Martin Blackledge	Directeur de Recherche CEA	Université Grenoble Alpes	Examineur
Dr Nicole Thielens	Directeur de Recherche INSERM	Université Grenoble Alpes	Examinatrice
Dr Lionel Ballut	Maître de Conférences HDR	Université Lyon 1	Examineur
Pr Sylvie Ricard-Blum	Professeur des Universités	Université Lyon 1	Directrice de thèse



## **Université Claude Bernard – LYON 1**

Administrateur provisoire de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Pierre ROLLAND

### **COMPOSANTES SANTE**

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

### **COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIES**

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOSNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI



## LABORATOIRE D'ACCUEIL

### INSTITUT DE CHIMIE ET DE BIOCHIMIE MOLECULAIRES ET SUPRAMOLECULAIRES

UMR 5246

Université Lyon 1 – CNRS – INSA – CPE



### Equipe Assemblages Supramoléculaires Péricellulaires Extracellulaires

Sous la direction du

Professeur Sylvie Ricard-Blum

Bâtiment Raulin

Université Claude Bernard Lyon 1  
43 Boulevard du 11 novembre 1918  
69622 Villeurbanne Cedex

2016 – 2020



Université Claude Bernard  Lyon 1



## LISTE DES ABBREVIATIONS

4-PBA : acide 4-phénylbutyrique	ERM : ezrine – radixine – moésine
6xHis : étiquette 6 histidine	EROS : ensemble-refinement of SAXS
ADAMs : a disintegrin and metalloproteinase	ESI : electrospray ionization
ADN : acide désoxyribonucléique	ESI-MS : electrospray ionization mass spectrometry
ATD : arrival time distribution	FA : funnel A / force d'activation
AUC : analytical ultracentrifugation	FGF : fibroblast growth factor
BLI : interférométrie de biocouche	FITC : isothiocyanate de fluorescéine
BMP : bone morphogenetic protein	FNIII : domaine fibronectine III
CCD : charge coupled device	FNIII <sub>1c</sub> : anastelline
CD : circular dichroism	FPLC : fast protein liquid chromatography
CEM : chain ejection mechanism	FRET : Förster resonance energy transfer
CIGu : chlorure de guanidinium	FTIR : Fourier-transform infrared spectroscopy
CRM : charge residue mechanism	GA : glutaraldéhyde
CS : chondroïtine sulfate	GAGs : glycosaminoglycanes
CSD : distribution des états de charges	Gal : galactose
CSPG4: chondroitin sulfate proteoglycan 4	GalNAc : N-acétylgalactosamine
DEAE : diéthylaminoéthyle	GDF : growth differentiation factor
DLS : dynamic light scattering	GlcNAc : N-acétylglucosamine
DMEM : Dulbecco's Modified Eagle Medium	GlcNS(6S) : glucose N- et 6-sulfaté
DMSO : diméthylsulfoxyde	GlcUA : acide glucuronique
DNase I : désoxyribonucléase I	GPA : glycophorine A
Domaine NA : domaine non sulfaté	GPC : glypican
Domaine NA/NS : domaine de sulfatation intermédiaire	GPI : glycosylphosphatidylinositol
Domaine NS : domaine très sulfaté	HA : acide hyaluronique
dp : degré de polymérisation	HBS : tampon HEPES salin
DS : dermatane sulfate	HDX-MS : hydrogen-deuterium exchange mass spectrometry
DSC : differential scanning calorimetry	HE : oligosaccharides d'héparine
DTSSP : 3,3'-dithiobis(sulfosuccinimidylpropionate)	He : hélium
DTT : dithiothréitol	HEK : human embryonic kidney
ED : ectodomaine des syndécans	HEPES : acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
ED-1-3 : ectodomaine des syndécans 1 à 3	HER2 : human epidermal growth factor receptor 2
ED-4 : isoforme 2 du syndécane-4	HexUA(2S) : acide hexuronique 2-sulfaté
EDC : 1-Ethyl-3-(3-diméthylaminopropyl) carbodiimide	HGF : hepatocyte growth factor
ED-PG : ectodomaine de syndécans portant leurs chaînes de glycosaminoglycanes	HP : héparine
EDTA : acide éthylènediaminetétraacétique	HPLC : high-performance liquid chromatography
EGF : endothelial growth factor	HS : héparane sulfate
EGFR : endothelial growth factor receptor	HSNA : tétrasaccharide d'héparane sulfate non sulfaté
EOM : ensemble optimization method	
ERK : Extracellular signal-regulated kinases	

HSNS : tétrasaccharide d'héparane sulfate sulfaté  
HSPG : protéoglycane à héparane sulfate  
IdoA(2S) : acide iduronique 2-sulfaté  
IdoUA : acide iduronique  
IDPRs / IDRs : intrinsically disordered (protein) regions  
IDPs : intrinsically disordered proteins  
IGF1R : insulin-like growth factor-1 receptor  
IM (ESI-IM-MS) : ion mobility  
IPTG : isopropyl-β-D-1-thiogalactopyranoside  
KS : kératine sulfate  
LRR : leucine-riche repeat  
MALLS : multiangle laser light scattering  
MDM2 : murine double minute 2  
MES : minimal ensemble search  
MG : molten globule  
MMP : matrix metalloproteinase  
MoRE / MoRF : molecular recognition element / feature  
MREMD : multiplexed replica exchange molecular dynamics  
MS/MS : spectrométrie de masse en tandem  
MT-MMP : membrane-type matrix metalloproteinase  
NANA / Neu5Ac : acide N-acétylneuraminique  
Ne : néon  
NIBS : Non-Invasive BackScatter  
Ni-NTA : acide nitrilotriacétique – Ni<sup>2+</sup>  
NO : monoxyde d'azote  
NRP : neuropiline  
NuMA : nuclear mitotic apparatus  
O-GPA : O-glycosylation de la glycophorine A  
PAGE : polyacrylamide gel electrophoresis  
PBS : tampon phosphate salin  
PDZ : PSD-95, Discs-large, Zonula occludens-1  
pHi : pH isoélectrique  
PIP2 : phosphatidylinositol-4,5-bisphosphate

PKCα : protéine kinase C α  
PMG : premolten globule  
PMLOs : proteinaceous membrane-less organelles  
POPC : 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine  
PTMs : post-translational modifications  
PVDF : polyfluorure de vinylidène  
RC : random coil  
RMN : résonance magnétique nucléaire  
RPTP-γ : receptor protein tyrosine phosphatases  
RX : cristallographie aux rayons X  
SAXS : diffusion des rayons X aux petits angles  
SDC : syndécan  
SDCBP : Syndecan-binding protein 1  
SDS : dodécylsulfate de sodium  
SEC : size exclusion chromatography  
SLRPs : small leucine-rich proteoglycans  
s-NHS : N-hydroxysulfosuccinimide  
SPOCK : SPARC/Osteonectin CWCV and Kazal-like domain proteoglycans  
SV2 : synaptic vesicle protein 2  
SVF : sérum de veau foetal  
TCEP : Tris(2-carboxyethyl)phosphine)  
TCQ : tryptophan triplet cysteine quenching  
TFE : trifluoroéthanol  
TGFβ : transforming growth factor β  
ThT : thioflavine T  
TIAM-1 : T-lymphoma invasion and metastasis-inducing protein 1  
TMAO : oxyde de triméthylamine  
TMD : transmembrane domain  
TOF : time of flight  
U.A. : unité arbitraire  
VEGF : vascular endothelial growth factor  
VEGFR-2 : vascular endothelial growth factor receptor 2



## PARAMETRES & CONSTANTES

### PARAMETRES GENERAUX

C : concentration  
E : énergie  
MM : masse moléculaire  
T : température  
 $\lambda$  : longueur d'onde  
 $k_B$  : constante de Boltzmann

### DLS :

D : coefficient de diffusion  
Dh : diamètre hydrodynamique  
Rh : rayon hydrodynamique  
 $\eta$  : viscosité

### INTERACTIONS BIOMOLECULAIRES

$k_a$  : constante cinétique d'association  
 $k_d$  : constante cinétique de dissociation  
 $K_D$  : constante d'équilibre de dissociation  
 $k_{obs}$  : constante de vitesse observée  
 $\Delta\lambda$  : décalage de longueur d'onde (*shift*)

### FILTRATION SUR GEL

$K_{av}$  : coefficient de partition  
 $K_d$  : coefficient de distribution  
 $R_s$  : rayon de Stokes  
 $V_0$  : volume mort  
 $V_e$  : volume d'élution  
 $V_i$  : volume interne  
 $V_t$  : volume total

### RHEOLOGIE

EI : indice d'élongation  
M : mesure de l'agrégation / agrégabilité en stase  
M1 : mesure de l'agrégation / agrégabilité à faible vitesse de cisaillement

### SAXS

$\bar{V}$  : volume partiel spécifique  
Dmax : taille maximale  
 $H_b(S)$  : entropie  
I(s) : intensité de diffusion  
P(s) : facteur de forme  
 $R_0$  : facteur de Flory  
Rflex : paramètre de flexibilité  
Rg : rayon de gyration  
 $R\sigma$  : variance de la distribution de l'ensemble sélectionné par rapport au pool  
s : module du vecteur de diffusion  
S(s) : facteur de structure  
 $s_{min/max}$  : valeur la plus petite / grande de s  
 $\nu$  : exposant de Flory  
V : volume  
 $\Delta\rho$  : contraste  
 $\zeta$  : exposant de Porod  
 $\Theta$  : angle de diffusion

### SPECTROMETRIE DE MASSE

A : aire sous la courbe  
CCS ( $\Omega$ ) : section efficace de collision  
e : charge élémentaire  
E : valeur du champ électrique  
FWHM : largeur à mi-hauteur  
K : mobilité d'un ion  
L : longueur du tube de dérive  
SASA : surface accessible au solvant  
SASAav : SASA moyenne  
SASAcpt : SASA des conformations compactes  
SASAext : SASA des conformations étendues  
t : temps total mesuré  
 $t_0$  : temps de dérive entre la sortie du tube et le détecteur  
 $t_d$  : temps de dérive dans le tube  
z : charge  
 $z_{av}$  : charge moyenne

## REMERCIEMENTS

Je tiens tout d'abord à remercier très chaleureusement les membres de mon jury de thèse qui ont gentiment acceptés d'évaluer ces travaux de recherche notamment les Docteurs Nicole Thielens et Lionel Ballut pour leur accompagnement durant ces 4 années à l'occasion de mes comités de suivi de thèse.

Je tiens vivement à remercier le Professeur Sylvie Ricard-Blum de m'avoir accueilli au sein de l'équipe ASPE, d'abord pendant mon stage de 2<sup>ème</sup> année de Master, dans lequel j'ai eu la mauvaise surprise de tomber sur des protéines désordonnées, puis pendant mes quatre années de thèse où j'ai rapidement dû admettre que l'étude des protéines globulaires était quand même bien moins passionnante. Je tiens particulièrement à remercier Sylvie de m'avoir laissé beaucoup de liberté d'un point de vue expérimental et pour sa disponibilité, mais aussi de m'avoir offert une formation sur les techniques de diffusion aux petits angles organisée par l'EMBO.

Je remercie évidemment tous les membres de l'équipe, toujours présents ou non, d'abord le jeune Docteur Sylvain D. Vallet (allias SDV) pour m'avoir formé à l'ensemble des appareils du laboratoire et aux fameuses BPL (les *bonnes pratiques de laboratoire*) dans une ambiance toujours agréable, même lorsque les noms d'oiseaux peuplaient majoritairement ses phrases. Je le remercie également d'avoir répondu à mes questions sans que j'en écoute la réponse. Je remercie également le Professeur Adriana Erica Miele et le Professeur Bertrand Duclos pour leur aide et pour leur très grande amabilité, Adriana pour sa présence lors de longs shifts de nuit aux synchrotrons, et Bertrand pour son humour très spontané, oui tout ce qui est dissous n'est pas cher.



**Figure 0 : Equipe ASPE (Pr. Sylvie Ricard-Blum).** De droite à gauche : Olivier, Sylvain, Adriana, Sylvie, et moi. Barre d'échelle : 180 cm.

Je tiens chaleureusement à remercier l'ensemble des membres présents dans l'équipe lors de la naissance de ma fille Jade, c'est-à-dire Sylvie, Adriana, Bertrand, Sylvain, Madeline et Olivier, pour l'accueil qui lui a été réservé lors sa venue au laboratoire.

Je tiens également à remercier le Professeur Patrice Gouet, le Docteur Halima Yajjou et le Docteur Christelle Folio de m'avoir très chaleureusement accueilli lors d'un stage de biochimie en 1<sup>ère</sup> année de Master, ce qui m'a permis d'en acquérir les bases essentielles. Je remercie également les Professeurs David Magne et Abdelkarim Abousalham, le Docteur Sébastien Violot et notamment le Docteur Leyre Brizuela-Madrid d'avoir pu me confier des groupes d'étudiants lors de mon année en tant qu'Attaché Temporaire d'Enseignement et de Recherche et d'avoir pu arranger au mieux mon emploi du temps.

Je tiens très sincèrement à remercier l'ensemble de nos collaborateurs pour l'ensemble de leurs réalisations, à savoir :

- Le Professeur Philippe Connes (Laboratoire Interuniversitaire de Biologie de la Motricité), pour son introduction à la physiologie des globules rouges et pour sa grande sympathie, et je remercie aussi tous les membres de la plateforme de biochimie pour leur très chaleureux accueil.
- Le Docteur Fabien Chirot (Institut des Sciences Analytiques - ILM), le Docteur Luke MacAleese (Institut Lumière Matière - ISA), Clothilde Comby-Zerbino (ILM), Mathilde Bouakil (ILM) et Aurélien Le Fèvre (ISA) pour l'ensemble des expériences de mobilité ionique, leur amabilité et d'une façon générale de m'avoir fait découvrir les joies de cette fabuleuse technique.
- Martyna Maszota-Zieleniak et le Docteur Sergey Samsonov (Université de Gdansk, Pologne) pour les expériences de docking entre des oligosaccharides de glycosaminoglycanes et l'anastelline.

J'en profite pour très chaleureusement remercier :

- Les Docteurs Martha Brennich et Aurélien Thureau, scientifiques respectivement de la ligne de lumière BM29 (ESRF) et SWING (Soleil) pour leur aide et leur gentillesse pendant les expériences de SAXS, et particulièrement Aurélien pour sa disponibilité (et de m'avoir laissé prendre pleins de photos).
- Les Docteurs Alexandra Berlioz-Barbier et Frédéric Delolme pour la réalisation des expériences de spectrométrie de masse native ainsi que pour leur aide et leur disponibilité (voire leur pédagogie). Je remercie également le Docteur Elodie Froment pour avoir répondu à toutes mes questions avec enthousiasme.
- Les Docteurs Thierry Granjon et Bastien Doumèche pour la formation et l'utilisation des spectrofluorimètres.

- Le Professeur John Couchman pour ses discussions très enrichissantes sur la biologie des syndécans.
- Le Docteur Cy Jeffries pour son aide et sa disponibilité sur la théorie de la diffusion de la lumière aux petits angles et pour son introduction à la physique des polymères.
- Xavier Robert ainsi que Lionel Ballut pour leur aide dans les expériences préliminaires de docking d'oligosaccharides sur l'anastelline.
- Roland Montserret et Virginie Gueguen-Chaignon pour l'utilisation du Chirscan et pour les expériences préliminaires de SEC-MALLS.
- Alexis Moreno pour son introduction à AutoDock Vina et Clotilde Raynard pour m'avoir gentiment offert de la poly-D-lysine, sacrifices qui n'ont pas été inutiles.

Je souhaite également remercier les membres des équipes GEMBAS (Pr. Loïc Blum), SMITH (Pr. Picot) et MEM2 (Pr. David Magne) pour les nombreux emprunts de produits chimiques ou pour l'utilisation de leurs matériels. Je remercie sincèrement Christine Blondel pour sa gentillesse et sa disponibilité dans la gestion financière de l'équipe.

Je remercie mes amis, dont je ne citerais pas le nom afin d'éviter d'en oublier, et plus particulièrement ceux qui n'ont pas fait fausse route pour se retrouver dans le côté obscur de la biologie (à jouer avec des cellules). Je remercie également chaleureusement Lorris, Thibaut, Joël et Romain (et sa maman), sans raison précise (sauf pour la maman de Romain).

Enfin et surtout, je remercie tendrement mes proches pour leur soutien et leur présence et plus particulièrement Mélody, notamment pour m'avoir plusieurs fois laissé dormir après mes longues nuits de rédaction. En revanche je ne remercie pas ma fille Jade qui a réussi à fermer mon fichier de thèse sur mon ordinateur sans enregistrer, mais je lui dédicace l'ensemble de ces travaux, en espérant qu'elle prenne le temps de les lire plus tard, du moins ces quelques lignes.

# SOMMAIRE

<b>I.</b>	<b>INTRODUCTION GENERALE &amp; OBJECTIFS DES TRAVAUX</b>	<b>22</b>
<b>II.</b>	<b>RAPPELS BIBLIOGRAPHIQUES</b>	<b>26</b>
	<b>II.1 LA MATRICE EXTRACELLULAIRE</b>	27
	<b>II.2 LES GLYCOSAMINOGLYCANES</b>	29
	<b>II.3 LES PROTEOGLYCANES</b>	32
	<b>II.4 LES PROTEOGLYCANES ASSOCIES AUX MEMBRANES</b>	36
	<b>II.4.1 Les protéoglycanes transmembranaires</b>	36
	<b>II.4.2 Les protéoglycanes à ancre glycosylphosphatidylinositol</b>	39
	<b>II.5 LES SYNDECANS</b>	40
	<b>II.5.1 Domaines et structures des syndécans</b>	43
	<i>a) Domaine cytosolique</i>	43
	<i>b) Domaine transmembranaire</i>	47
	<i>c) Domaine extracellulaire (ectodomaine)</i>	49
	<i>d) Modélisation des syndécans entiers</i>	55
	<b>II.5.2 Fonctions des syndécans</b>	57
	<b>II.6 LE DESORDRE INTRINSEQUE DES PROTEINES</b>	61
	<b>II.6.1 Généralités</b>	61
	<b>II.6.2 Caractéristiques du désordre intrinsèque</b>	63
	<i>a) Composition en acides aminés des séquences désordonnées</i>	63
	<i>b) Le concept de structure – fonction revisité</i>	65
	<i>c) Modulation de l'espace conformationnel des protéines désordonnées</i>	66
	<i>d) Transition de phase liquide – liquide</i>	67
	<b>II.6.3 Mécanismes d'interaction des protéines intrinsèquement non structurées</b>	69
	<i>a) Caractéristiques de séquences</i>	69
	<i>b) Fuzzy complexes</i>	71
	<b>II.6.4 Fonctions du désordre intrinsèque</b>	73
	<b>II.6.5 Association du désordre aux pathologies</b>	76
	<b>II.6.6 Prédiction et analyse du désordre intrinsèque</b>	79
	<i>a) Prédiction du désordre intrinsèque et des MoRF</i>	79
	<i>b) Techniques d'analyse des protéines et régions désordonnées</i>	80
<b>III.</b>	<b>MATERIEL &amp; METHODES</b>	<b>86</b>
	<b>III.1 EXPRESSION DES ECTODOMAINES DES SYNDECANS EN SYSTEME PROCARYOTE</b>	87
	<b>III.1.1 Vecteur d'expression</b>	87
	<b>III.1.2 Transformation des bactéries</b>	89
	<b>III.1.3 Mise en culture des bactéries transformées</b>	90
	<b>III.2 EXPRESSION DES ECTODOMAINES DES SYNDECANS EN SYSTEME EUARYOTE</b>	90
	<b>III.2.1 Vecteur d'expression</b>	90
	<b>III.2.2 Transfection des cellules eucaryotes</b>	91
	<b>III.2.3 Culture des cellules eucaryotes et récoltes des milieux</b>	92
	<b>III.3 PURIFICATION DES ECTODOMAINES DES SYNDECANS</b>	93
	<b>III.3.1 Purification des ectodomaines exprimés en système procaryote</b>	93
	<i>a) Purification par chromatographies d'affinité</i>	93
	<i>b) Purification par chromatographie d'exclusion stérique</i>	94
	<i>c) Concentration des ectodomaines exprimés en système procaryote</i>	95
	<b>III.3.2 Purification des ectodomaines exprimés en système eucaryote</b>	95
	<i>a) Purification par chromatographies échangeuse d'anions</i>	95
	<i>b) Purification par chromatographie d'affinité sur résine anti-FLAG</i>	95
	<i>c) Concentration des ectodomaines exprimés en système eucaryote</i>	96

III.4 ANALYSE DES SYNDECANS PURIFIES PAR ELECTROPHORESE ET PAR WESTERN BLOT.....	96
III.4.1 Electrophorèse sur gel de polyacrylamide en conditions dénaturantes.....	96
III.4.2 Western blot.....	97
III.5 CALCUL DES CONCENTRATIONS DES PROTEINES PURIFIEES PAR SPECTROPHOTOMETRIE.....	98
III.6 BIOINFORMATIQUE STRUCTURALE.....	99
III.6.1 Prédiction du désordre intrinsèque.....	99
III.6.2 Prédiction de la composition en structures secondaires.....	99
III.6.3 Prédiction des <i>Molecular Recognition Features</i> (MoRFs).....	99
III.6.4 Modélisation l'extrémité C-terminale d'ED-4 et du dimère d'ED-4.....	99
III.7 CARACTERISATION BIOPHYSIQUE DES ECTODOMAINES DES SYNDECANS.....	100
III.7.1 Dichroïsme circulaire.....	100
III.7.2 Détermination du rayon hydrodynamique des ectodomaines des syndécans.....	101
a) Calculs théoriques du rayon hydrodynamique.....	101
b) Détermination du rayon hydrodynamique par chromatographie d'exclusion stérique.....	102
c) Diffusion dynamique de la lumière.....	103
d) Diffusion dynamique de la lumière couplée à une chromatographie d'exclusion stérique.....	103
III.7.3 Diffusion de la lumière laser multi-angles.....	104
III.7.4 Diffusion des rayons X aux petits angles.....	104
a) Introduction.....	104
b) Rappels théoriques.....	106
c) Déconvolution des intensités de diffusion des systèmes polydisperses (EOM).....	111
d) Collection et traitement des intensités de diffusion obtenues expérimentalement.....	113
III.7.5 Spectrométrie de masse en conditions natives et mobilité ionique.....	117
a) Rappels sur l'ionisation par électronébulisation.....	117
b) Acquisition des données d'ESI-MS native.....	118
c) Acquisition des données de mobilité ionique.....	121
III.8 ETUDE DES INTERACTIONS BIOMOLECULAIRES.....	124
III.8.1 Interférométrie de biocouche.....	124
a) Rappels théoriques.....	124
b) Crible des partenaires extracellulaires des ectodomaines des syndécans.....	125
c) Complexe entre l'anastelline et les ectodomaines des syndécans.....	128
d) Etude des interactions entre les globules rouges et le fibrinogène.....	130
III.8.2 Docking.....	131
a) Docking d'un peptide de l'ectodomaine du syndécan-1 sur l'anastelline.....	131
b) Docking d'oligosaccharides d'héparine et d'héparane sulfate sur l'anastelline.....	132
c) Docking d'un oligosaccharide et de l'acide N-acétylneuraminique sur le fibrinogène.....	133
III.8.3 Détermination du site d'interaction entre l'anastelline et l'ectodomaine du syndécan-1 par spectrométrie de masse en tandem.....	134
III.8.4 Turbidimétrie.....	135
III.8.5 Fluorimétrie.....	135
III.9 DIGESTION DES GLYCOSAMINOGLYCANES ET DES ACIDES SIALIQUES A LA SURFACE DES GLOBULES ROUGES.....	135
III.10 MESURES RHEOLOGIQUES EFFECTUEES SUR LES GLOBULES ROUGES.....	136
III.10.1 Mesure de la déformabilité des globules rouges (ektacytométrie).....	136
III.10.2 Mesure de l'agrégation et de l'agrégabilité des globules rouges.....	137
III.10.3 Analyses statistiques.....	138
<b>IV. RESULTATS.....</b>	<b>139</b>
IV.1 CARACTERISATION STRUCTURALE DES ECTODOMAINES DES SYNDECANS SANS GAGS.....	140
IV.1.1 Expression et purification des ectodomaines de syndécans sans GAGS.....	141
IV.1.2 Détermination de la masse moléculaire des ectodomaines de syndécans.....	146
a) Diffusion de la lumière laser multi-angles couplée à la chromatographie d'exclusion stérique.....	146
b) Spectrométrie de masse en conditions natives.....	147

<b>IV.1.3 Structures secondaires et désordre intrinsèque des ectodomains des syndécans</b>	149
<b>IV.1.4 Détermination des paramètres hydrodynamiques des ectodomains des syndécans</b>	155
<i>a) Calcul du rayon hydrodynamique par chromatographie d'exclusion stérique</i>	156
<i>b) Calcul du rayon hydrodynamique par diffusion dynamique de la lumière</i>	158
<i>c) Calcul du rayon hydrodynamique par diffusion dynamique de la lumière couplée à la chromatographie d'exclusion stérique</i>	160
<b>IV.1.5 Analyses structurales des ectodomains des syndécans</b>	163
<i>a) Nature du désordre intrinsèque</i>	163
<i>b) Flexibilité des ectodomains des syndécans</i>	169
<i>c) Conformations préférentielles adoptées par les ectodomains des syndécans</i>	171
<i>d) Espace conformationnel et stabilité des ectodomains des syndécans</i>	184
<i>e) Mécanisme de dimérisation d'ED-4</i>	203
<b>IV.1.6 Discussion</b>	221
<b>IV.1.7 Perspectives</b>	227
<b>IV.2 CARACTERISATION STRUCTURALE DES ECTODOMAINES DES SYNDECANS AVEC GAGs</b>	256
<b>IV.2.1 Purification des ectodomains portant leurs chaînes de GAGs</b>	257
<b>IV.2.2 Analyses par SEC-SAXS des ectodomains des syndécans avec GAGs</b>	261
<b>IV.2.3 Discussion et Perspectives</b>	269
<b>IV.3 INTERACTIONS DES SYNDECANS</b>	272
<b>IV.3.1 Le réseau d'interactions des syndécans</b>	273
<i>a) Construction de l'interactome des syndécans</i>	273
<i>b) Analyse d'enrichissement de l'interactome</i>	275
<i>c) Les ectodomains de syndécan, sources de matricryptines ?</i>	276
<b>IV.3.2 Identification de partenaires des ectodomains des syndécans par interférométrie de biocouche</b>	309
<i>a) Interaction entre les ectodomains des syndécans et le FGF-2</i>	310
<i>b) Interaction entre les syndécans et l'anastelline</i>	312
<b>IV.3.3 Discussion et Perspectives</b>	340
<b>IV.4 ROLE DES GLYCOSAMINOGLYCANES ET DES ACIDES SIALIQUES PRESENTS A LA SURFACE DES GLOBULES ROUGES SAINS ET DREPANOCYTAIRES DANS LEURS PROPRIETES RHEOLOGIQUES</b>	345
<b>IV.4.1 Introduction</b>	346
<b>IV.4.2 Buts de l'étude</b>	350
<b>IV.4.3 Résultats</b>	351
<b>IV.4.4 Discussion</b>	362
<b>IV.4.5 Perspectives</b>	365
<b>V. CONCLUSIONS &amp; PERSPECTIVES</b>	<b>371</b>
<b>Références Bibliographiques</b>	381
<b>Publications &amp; Communications Orales et par Affiches</b>	412
<b>Annexes</b>	413
<b>Résumé des Travaux</b>	424

## LITE DES FIGURES ET DES TABLEAUX

### LISTE DES FIGURES

---

- Figure 1** : Composition du matrisome humain et murin.
- Figure 2** : Organisation de la matrice extracellulaire, péricellulaire et interstitielle.
- Figure 3** : Composition des glycosaminoglycanes et de leur site d'attachement sur les protéines.
- Figure 4** : Classification des protéoglycanes.
- Figure 5** : Protéoglycanes membranaires.
- Figure 6** : Les protéoglycanes de la famille des hyalectans.
- Figure 7** : Protéoglycanes transmembranaires.
- Figure 8** : Structure cristallographique du glypican-1 humain.
- Figure 9** : Organisation structurale des syndécans.
- Figure 10** : Comparaison de l'organisation en domaines des deux isoformes du syndécan-4
- Figure 11** : Alignement des séquences du domaine cytosolique des quatre syndécans humains
- Figure 12** : Structures cristallographiques de la région C2 du syndécan-1 humain avec la protéine TIAM-1
- Figure 13** : Structures cristallographiques de la région C2 du syndécan-4 cocrystallisée avec la synténine.
- Figure 14** : Structures du domaine cytosolique du SDC-4 déterminées par RMN.
- Figure 15** : Alignement des séquences du domaine transmembranaire des syndécans humains.
- Figure 16** : Domaine transmembranaire de la glycophorine A et du syndécan-2 humain.
- Figure 17** : Localisation du domaine de type mucine au sein de l'ectodomaine du syndécan-3 humain.
- Figure 18** : Alignement des 20 derniers résidus d'acides aminés des ectodomains des syndécans humains.
- Figure 19** : Structure de l'octapeptide de syndécan-1 cocrystallisé avec la xylosyltransférase-1.
- Figure 20** : Modélisation du syndécan-4 entier.
- Figure 21** : Modélisation de la structure tridimensionnelle du syndécan-3 de poulet.
- Figure 22** : Composition en acides aminés des protéines désordonnées.
- Figure 23** : Evolution du concept de structure – fonction.
- Figure 24** : Observation de la formation d'organelles protéiques sans membranes (PMLOs) d'un peptide (K18) de la protéine Tau.
- Figure 25** : Modulation de la transition de phase liquide – liquide par les modifications post-traductionnelles des protéines désordonnées.
- Figure 26** : Exemples d'éléments de reconnaissance moléculaire (MoRFs).
- Figure 27** : La séquence caméléon du domaine C-terminal de p53 se replie différemment en fonction de ses partenaires cellulaires.
- Figure 28** : Mécanismes de la formation des fuzzy complexes.
- Figure 29** : Fonctions majeures des protéines intrinsèquement non structurées ou des régions désordonnées.
- Figure 30** : Structure d'une fibrille amyloïde formée par l' $\alpha$ -synucléine.
- Figure 31** : Procédure et analyse multi-approches des protéines intrinsèquement non structurées.
- Figure 32** : Détermination de la proportion en structures secondaires par dichroïsme circulaire et spectroscopie infrarouge à transformée de Fourier.
- Figure 33** : Analyse de l'hétérogénéité conformationnelle des protéines désordonnées par spectrométrie de masse.



**Figure 34** : Alignement des séquences du syndécan-3 entier humain, de souris et de rat.

**Figure 35** : Carte du vecteur d'expression pET-15b vide montrant ses principales caractéristiques.

**Figure 36** : Carte du vecteur d'expression pCEP-Pu/BM40

**Figure 37** : Environnement expérimental des expériences de SEC-SAXS.

**Figure 38** : Représentation schématique de la ligne de lumière.

**Figure 39** : Représentation normalisée de Kratky pour quatre protéines ayant des degrés de compaction différents.

**Figure 40** : Procédure de déconvolution des intensités de diffusion suivie par l'outil *Ensemble Optimization Method*, EOM.

**Figure 41** : Exemples de cartes de corrélation (*CorMap*).

**Figure 42** : Source Nano-ESI d'un spectromètre de masse à mobilité ionique.

**Figure 43** : Mécanisme d'ionisation des protéines par électronébulisation (ESI).

**Figure 44** : Déconvolution de la distribution des états de charges obtenue par ESI-MS.

**Figure 45** : Spectromètre de masse utilisé lors des expériences d'ESI-MS en conditions natives.

**Figure 46** : Schéma général montrant les principales caractéristiques du spectromètre à mobilité ionique utilisé dans nos expériences.

**Figure 47** : Principe de l'interférométrie de biocouche.

**Figure 48** : Alignement de la séquence de l'anastelline commerciale et de la séquence utilisée pour obtenir la structure tridimensionnelle de l'anastelline.

**Figure 49** : Analyse par électrophorèse en conditions dénaturantes et réductrices de l'expression en système procaryote des ectodomaines des syndécans.

**Figure 50** : Purification par chromatographie d'affinité Ni-NTA des 4 ectodomaines de syndécans.

**Figure 51** : Purification par chromatographie d'affinité anti-FLAG des 4 ectodomaines de syndécans pré-purifiés par chromatographie d'affinité Ni-NTA.

**Figure 52** : Purification par chromatographie d'exclusion stérique des 4 ectodomaines de syndécan préalablement purifiés par chromatographies d'affinité.

**Figure 53** : Analyse par SDS-PAGE et Western blot des 4 ectodomaines de syndécans purifiés.

**Figure 54** : Calcul de la masse moléculaire des ectodomaines des syndécans par SEC-MALLS.

**Figure 55** : Spectre de masse des ectodomaines des syndécans obtenus en conditions natives.

**Figure 56** : Composition en structures secondaires et pourcentage de désordre intrinsèque des EDs de syndécans.

**Figure 57** : Spectres de dichroïsme circulaire des EDs en présence de concentrations croissantes de trifluoroéthanol.

Figure 58 : Evolution de la composition en structures secondaires des 4 EDs en présence de TFE.

**Figure 59** : Spectres de dichroïsme circulaire des EDs en présence de chlorure de guanidinium.

**Figure 60** : Spectres de dichroïsme circulaire des EDs en présence de chaperonnes chimiques.

**Figure 61** : Prédiction des proportions en structures secondaires, en désordre intrinsèque et des MoRFs au sein des EDs des syndécans.

**Figure 62** : Détermination de la masse moléculaire apparente et du rayon hydrodynamique des ectodomaines des syndécans par chromatographie d'exclusion stérique.

**Figure 63** : Détermination du diamètre hydrodynamique des ectodomaines des syndécans par diffusion dynamique de la lumière.

**Figure 64** : Détermination du rayon hydrodynamique des ectodomaines des syndécans par diffusion dynamique de la lumière couplée à la chromatographie d'exclusion stérique.

**Figure 65 :** Déconvolution des intensités de diffusion collectées pour les ectodomains des syndécans lors des expériences de SEC-SAXS.

**Figure 66 :** Rayon de gyration calculé par différentes méthodes d'analyse des données de SEC-SAXS.

**Figure 67 :** Détermination du rayon de gyration des 4 ectodomains des syndécans.

**Figure 68 :** Détermination des dimensions maximales des ectodomains des syndécans à partir de la fonction de distribution de paires  $P(r)$ .

**Figure 69 :** Evaluation du degré de compaction et de la flexibilité des 4 ectodomains des syndécans déterminés par SEC-SAXS.

**Figure 70 :** Déconvolution des données de SEC-SAXS des ectodomains des syndécans par EOM.

**Figure 71 :** Distributions des valeurs de rayons de gyration et de  $D_{max}$  des 4 EDs obtenues par la déconvolution effectuée par EOM avec un modèle *random coil*.

**Figure 72 :** Modèles générés par EOM.

**Figure 73 :** Principe de l'utilisation de l'ESI-MS dans le cadre d'une protéine désordonnée.

**Figure 74 :** Analyses des ectodomains des syndécans par ESI-MS en conditions natives.

**Figure 75 :** Relation entre la charge moyenne et les valeurs de SASA pour des protéines globulaires.

**Figure 76 :** Illustration du degré de compaction des conformations compactes des ectodomains des syndécans.

**Figure 77 :** Distribution des sections efficaces de collision (CCS) des ectodomains des syndécans (EDs) et du mutant ED-4 C150A lors des expériences de mobilité ionique.

**Figure 78 :** Espace conformationnel des ectodomains des syndécans d'après les expériences de mobilité ionique.

**Figure 79 :** Distribution des sections efficaces de collision (CCS) des ectodomains des syndécans (EDs) et du mutant ED-4 C150A.

**Figure 80 :** Distribution des temps d'arrivée (ATD) des états de charges où deux populations de conformations ont été détectées lors des expériences de mobilité ionique.

**Figure 81 :** Mécanisme d'ionisation pendant l'étape d'électronébulisation des protéines selon leur degré de compaction.

**Figure 82 :** Activation des états de charges correspondant aux conformations les plus compactes des ectodomains des syndécans.

**Figure 83 :** Activations des états de charges correspondant aux conformations de la phase de transition entre les conformations compactes et les conformations intermédiaires des ectodomains des syndécans.

**Figure 84 :** Activations des états de charge correspondant aux conformations intermédiaires des ectodomains des syndécans.

**Figure 85 :** Activations des états de charges correspondant aux conformations compactes, aux conformations présentes dans la phase de transitions et aux conformations intermédiaires adoptées par la forme dimérique de l'ectodomaine du syndécane-4.

**Figure 86 :** Bilan des mesures d'activation des états de charges des différentes conformations adoptées par les ectodomains des syndécans.

**Figure 87 :** Bilan des mesures d'activation des états des charges correspondant aux conformations adoptées par la forme dimérique d'ED-4.

**Figure 88 :** Comparaison de l'organisation des deux isoformes du syndécane-4.

**Figure 89 :** La formation d'un pont disulfure permet la dimérisation d'ED-4.

**Figure 90 :** Purification du mutant ED-4 C150A par chromatographies d'affinité.

**Figure 91 :** Purification du mutant ED-4 C150A par chromatographie d'exclusion stérique.

**Figure 92 :** Analyse par SEC-SAXS du mutant ED-4 C150A.

**Figure 93** : Mesure du rayon hydrodynamique du mutant ED-4 C150A par SEC-DLS.

**Figure 94** : Comparaison de la distribution des états de charge d'ED-4 et du mutant C150A.

**Figure 95** : Détection d'une faible quantité de dimère d'ED-4 C150A.

**Figure 96** : Réticulation d'ED-4 C150A par le glutaraldéhyde (GA).

**Figure 97** : Réticulation d'ED-4 C150A par le DTSSP.

**Figure 98** : Réticulation d'ED-4 sauvage par le DTSSP.

**Figure 99** : Détermination de l'arrangement des monomères au sein du dimère d'ED-4 par mobilité ionique.

**Figure 100** : Mécanisme de dimérisation proposé pour ED-4.

**Figure 101** : Prédiction de la structure tridimensionnelle de l'extrémité C-terminale d'ED-4.

**Figure 102** : Docking de l'extrémité C-terminale d'ED-4 sur elle-même.

**Figure 103** : Ajustement des intensités de diffusion théoriques du modèle calculé par BUNCH pour la totalité de la séquence d'ED-4 sous forme dimérique sur les données expérimentales.

**Figure 104** : Modélisation de la totalité de la séquence d'ED-4 sous forme dimérique.

**Figure 105** : CCS calculées pour le modèle du dimère d'ED-4 au sein de la distribution des CCS expérimentales.

**Figure 106** : Prédiction du désordre intrinsèque de l'ectodomaine du syndécan-2 (ED-2) et de l'isoforme-2 du syndécan-4 (ED-4) et localisation de leurs résidus de proline.

**Figure 107** : Structures de la *photoactive yellow protein* résolues par RMN.

**Figure 108** : Bilan des analyses biophysiques des ectodomains des syndécans humains.

**Figure 109** : Purification par chromatographie échangeuse d'anions de l'ectodomaine du syndécan-2 et -4 portant leurs chaînes de glycosaminoglycanes.

**Figure 110** : Purification par chromatographie d'affinité sur résine anti-FLAG de l'ectodomaine du syndécan-2 et -4 portant leurs chaînes de glycosaminoglycanes.

**Figure 111** : Purification de l'ectodomaine du syndécan-3 portant ses chaînes de glycosaminoglycanes.

**Figure 112** : Détermination par spectrophotométrie de la concentration des ectodomains des syndécans avec GAGs.

**Figure 113** : Suivi de l'éluion des espèces moléculaires présentes dans l'échantillon d'ED-2PG par spectrophotométrie lors des expériences de SEC-SAXS.

**Figure 114** : Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-2PG.

**Figure 115** : Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-4PG.

**Figure 116** : Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-4PG.

**Figure 117** : Analyse d'ED-4PG1 par SEC-SAXS en conditions réductrices.

**Figure 118** : Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-3PG.

**Figure 119** : Détermination de la masse moléculaire et du rayon hydrodynamique d'ED-3PG.

**Figure 120** : Représentation schématique de la conformation générale de l'ectodomaine du syndécan-3 portant ou non ses chaînes de GAGs.

**Figure 121** : Interactome des syndécans.

**Figure 122** : Sites de clivage retrouvés à proximité des peptides de synstatine.

**Figure 123** : Expériences d'interactions entre FGF-2 et les EDs des syndécans sans GAGs.

**Figure 124 :** Détermination de la constante  $k_{obs}$  lors de l'interaction entre le facteur de croissance fibroblastique-2 et l'ectodomaine du syndécan-1 et du syndécan-3.

**Figure 125 :** Mécanisme d'interaction entre l'anastelline et le domaine FNIII<sub>3</sub> de la fibronectine.

**Figure 126 :** Interaction entre l'anastelline et les chaînes d'héparane et de chondroïtine sulfate.

**Figure 127 :** Structure des 5 oligosaccharides d'héparine (HE) ou d'héparane sulfate (HSNA, HSNS) utilisés lors des expériences de docking.

**Figure 128 :** Docking des oligosaccharides d'héparine et d'héparane sulfate sur l'anastelline.

**Figure 129 :** Fréquence d'apparition des résidus d'acides aminés de l'anastelline retrouvés dans l'interaction avec les oligosaccharides.

**Figure 130 :** Localisation des résidus d'acides aminés de l'anastelline impliqués dans l'interaction avec les oligosaccharides d'héparine et d'héparane sulfate.

**Figure 131 :** Détection de l'interaction entre l'anastelline et les syndécans.

**Figure 132 :** La fibronectine n'interagit pas avec les ectodomains des syndécans sans leurs chaînes de glycosaminoglycane.

**Figure 133 :** Détermination de la constante  $k_{obs}$  lors de la formation des complexes anastelline – ectodomains des syndécans portant ou non leurs chaînes de GAGs.

**Figure 134 :** Sensorgrammes enregistrés par BLI correspondant à l'interaction entre ED-1 et l'anastelline en présence de CHAPS ou de concentrations croissantes de NaCl.

**Figure 135 :** Influence du pH sur l'interaction entre l'anastelline et l'ectodomaine du syndécan-1.

**Figure 136 :** Mécanisme d'interaction proposé entre l'anastelline et l'ectodomaine du syndécan-1.

**Figure 137 :** Expériences de dichroïsme circulaire sur l'ectodomaine du syndécan-1 et l'anastelline à différents pH.

**Figure 138 :** Localisation des résidus d'histidine et du site de fixation des oligosaccharides d'héparine et d'héparane sulfate sur la structure 3D de l'anastelline.

**Figure 139 :** Rôle du pH dans l'interaction entre l'anastelline et l'héparine.

**Figure 140 :** Analyse par électrophorèse des expériences de réticulation de l'ectodomaine du syndécan-1 et de l'anastelline par le DTSSP.

**Figure 141 :** Localisation des peptides d'anastelline réticulés avec le peptide KDTQLL de l'ectodomaine du syndécan-1.

**Figure 142 :** Docking du peptide de l'ectodomaine du syndécan-1 sur l'anastelline.

**Figure 143 :** Docking du peptide de l'ectodomaine du syndécan-1 sur l'anastelline avec CABS-Dock.

**Figure 144 :** Recherche du complexe ternaire anastelline – ectodomaine du syndécan-1 – oligosaccharide par interférométrie de biocouche.

**Figure 145 :** Variation de l'intensité du signal pendant l'étape d'association du dp8 d'héparine sur le complexe préformé anastelline – ED-1.

**Figure 146 :** Agrégation de l'ectodomaine du syndécan-1 et de l'anastelline.

**Figure 147 :** Ultrastructure du glycocalyx endothélial.

**Figure 148 :** Visualisation du glycocalyx du globule rouge par microscopie électronique.

**Figure 149 :** Structure de l'hémoglobine S et la conséquence de sa polymérisation sur le phénotype du globule rouge.

**Figure 150 :** Propriétés rhéologiques des globules rouges sains et drépanocytaires non traités.

**Figure 151 :** Propriétés rhéologiques des globules rouges sains et drépanocytaires traités par l'héparinase III.

**Figure 152 :** Propriétés rhéologiques des globules rouges sains et drépanocytaires traités par la chondroïtinase ABC.

**Figure 153 :** Propriétés rhéologiques des globules rouges sains et drépanocytaires traités par la hyaluronidase.

**Figure 154 :** Déformabilité des globules rouges sains et drépanocytaires traités par la neuraminidase.

**Figure 155 :** Effets de la digestion des acides sialiques par la neuraminidase sur l'agrégabilité (en dextran) des globules rouges sains et drépanocytaires.

**Figure 156 :** Effets de la digestion des acides sialiques par la neuraminidase sur l'agrégation (en plasma) des globules rouges sains et drépanocytaires.

**Figure 157 :** Sensorgrammes obtenus lors des expériences d'interférométrie de biocouche.

**Figure 158 :** Sensorgrammes alignés obtenus lors des expériences d'interférométrie de biocouche.

**Figure 159 :** Docking entre la O-GPA et le fibrinogène humain.

**Figure 160 :** Observation du glycocalyx de globules rouges sains par microscopie confocale.

## **LISTE DES TABLEAUX**

---

**Tableau I :** Etudes structurales des syndécans.

**Tableau II :** Caractéristiques des ectodomains des syndécans humains exprimés sous forme recombinante en système procaryote.

**Tableau III :** Masse moléculaire théorique (Da) et coefficient d'extinction molaire théorique à 280 nm ( $M^{-1}.cm^{-1}$ ) calculés par ProtParam des ectodomains des syndécans exprimés en système procaryote et eucaryote.

**Tableau IV :** Protéines utilisées pour la calibration de la colonne de filtration sur gel (Superdex S200 Increase 10/300 GL) ainsi que leurs rayons hydrodynamiques respectifs.

**Tableau V :** Conditions expérimentales utilisées pour la collection des données de SEC-SAXS sur les ectodomains des syndécans exprimés en système procaryote.

**Tableau VI :** Conditions expérimentales utilisées pour la collection des données de SEC-SAXS sur les ectodomains des syndécans exprimés en système eucaryote.

**Tableau VII :** Protéines immobilisées lors des cribles d'interactions réalisés par interférométrie de biocouche.

**Tableau VIII :** Conditions expérimentales des expériences d'interférométrie de biocouche.

**Tableau IX :** Formation des complexes binaires et ternaires par BLI.

**Tableau X :** Ligands utilisés lors des expériences de docking et de dynamique moléculaire.

**Tableau XI :** Détails des enzymes utilisées dans les expériences d'hydrolyse des constituants du glycocalyx des globules rouges sains et drépanocytaires.

**Tableau XII :** Bilan des mesures de la masse moléculaire (Da) des ectodomains des syndécans (EDs).

**Tableau XIII :** Détermination de la masse moléculaire apparente et du rayon hydrodynamique ( $R_h$ ) des ectodomains des syndécans par chromatographie d'exclusion stérique.

**Tableau XIV :** Rayons hydrodynamiques des ectodomains des syndécans calculées par chromatographie d'exclusion stérique (SEC), par diffusion dynamique de la lumière couplée (SEC-DLS) ou non (DLS) à une étape de chromatographie d'exclusion stérique.

**Tableau XV :** Comparaison des valeurs de rayon hydrodynamiques mesurés en SEC-DLS et calculés selon des équations empiriques pour chaque degré de compaction.

**Tableau XVI :**  $R_g$  expérimentaux (SEC-SAXS) et théoriques des ectodomains des syndécans calculés à partir de l'équation de Flory.

**Tableau XVII :** Déconvolution des données de SEC-SAXS collectées pour les ectodomains des syndécans par EOM avec les modèles PMG et RC.

**Tableau XVIII** : Détails des distributions de rayon de gyration et de la taille maximale des particules ( $D_{max}$ ) issues de la déconvolution par EOM avec un modèle *random coil*.

**Tableau XIX** : Valeur moyenne de la surface accessible au solvant ( $SASA_{av}$ ) des ectodomaines des syndécans.

**Tableau XX** : Valeur de la surface accessible au solvant des conformations les plus étendues ( $SASA_{ext}$ ) des ectodomaines des syndécans.

**Tableau XXI** : Détermination du degré de compaction des conformations les plus compactes des ectodomaines des syndécans (EDs) à partir des données d'ESI-MS.

**Tableau XXII** : Etats de charges sélectionnés pour l'analyse de la stabilité conformationnelle des EDs des syndécans.

**Tableau XXIII** : Résumé des résultats obtenus sur les ectodomaines des syndécans portant leurs chaînes de GAGs ainsi que leur comparaison aux paramètres des protéines seules.

**Tableau XXIV** : Partenaires criblés lors des expériences d'interférométrie de biocouche avec les ectodomaines des syndécans dépourvus de leurs chaînes de glycosaminoglycanes.

**Tableau XXV** : Caractéristiques des oligosaccharides utilisés comme ligands dans les expériences de docking avec l'anastelline.

**Tableau XXVI** : Analyse des simulations par dynamique moléculaire de 5 poses représentatives pour chacun des oligosaccharides.

**Tableau XXVII** : Décomposition de l'énergie libre d'interaction par résidus d'acides aminés de l'anastelline dans son interaction avec les oligosaccharides d'héparine.

**Tableau XXVIII** : Décomposition de l'énergie libre d'interaction par résidus d'acides aminés de l'anastelline dans son interaction avec les oligosaccharides d'héparane sulfate.

**Tableau XXIX** : Valeurs de la constante  $k_{obs}$  lors de la formation des complexes anastelline – ectodomaines des syndécans portant ou non leurs chaînes de GAGs.

**Tableau XXX** : Qualité de l'ajustement des modèles lors de la formation du complexe anastelline – ectodomaines du syndécan-1 à différents pH.

**Tableau XXXI** : Détermination des paramètres cinétiques de l'interaction entre l'anastelline et l'ectodomaine du syndécan-1.

**Tableau XXXII** : Résultats des expériences d'interférométrie de biocouche.

**Tableau XXXIII** : Effets des digestions enzymatiques sur la déformabilité, l'agrégation et l'agrégabilité des globules rouges sains et drépanocytaires.



# **CHAPITRE I**

## **INTRODUCTION GÉNÉRALE & OBJECTIFS DES TRAVAUX**



Chez les vertébrés, les syndécans sont une famille de quatre protéoglycanes membranaires de type I exprimés dans tous les types cellulaires où au moins un type de syndécan est exprimé, excepté chez les globules rouges (*Afratis et al., 2017*). Ils réalisent leurs fonctions biologiques d'abord en tant que molécules d'adhésion mais aussi en tant que récepteurs ou corécepteurs à la surface cellulaire et participent à plusieurs processus physiopathologiques (*Couchman et al, 2015, Afratis et al., 2017*).

Les syndécans possèdent une organisation structurale identique et sont composés d'un domaine extracellulaire qui porte plusieurs chaînes de glycosaminoglycanes et qui est peu conservé, d'un domaine transmembranaire et d'un court domaine cytosolique qui sont à l'inverse très conservés. Plusieurs structures du domaine cytosolique, notamment du syndécan-4 (*Lee et al., 1998, Shin et al., 2001*), ainsi que celle du domaine transmembranaire du syndécan-2 (*Li et al., 2019*) ont été déterminées. L'ectodomaine des syndécans est prédit comme intrinsèquement non structuré (*Leonova & Galzitskaya, 2015*) et peut être libéré par le processus de shedding dans l'espace extracellulaire, ce processus étant amplifié lorsque les chaînes d'héparane sulfate qu'il porte sont préalablement dégradées par l'héparanase humaine (*Manon-Jensen et al., 2010*). Leur ectodomaine a très peu été étudié et seule la structure d'un octapeptide de l'ectodomaine du syndécan-1 cocrystallisé avec un de ses partenaires a été résolue et n'adopte pas de structure secondaire (*Briggs & Hohenester, 2018*).

Le premier objectif de ces travaux a ainsi été de caractériser l'ectodomaine des quatre syndécans sur le plan structural et conformationnel. Dans un premier temps, nous avons étudié la partie protéique des ectodomains des syndécans ne portant pas leurs chaînes de glycosaminoglycanes en les exprimant sous forme recombinante en système procaryote. En absence de leurs chaînes de glycosaminoglycanes, la partie protéique de l'ectodomaine des syndécans conserve ses fonctions de molécule d'adhésion (*McFall & Rapraeger, 1998, Whiteford et al., 2008, De Rossi et al., 2014, Jung et al., 2016*). Les ectodomains acquièrent également de nouvelles fonctions en jouant le rôle d'effecteur paracrine, autocrine ou de compétiteur des syndécans transmembranaires (*Manon-Jensen et al., 2010*).

Nous avons pour cela employé diverses approches biophysiques. L'étude de leur composition en structures secondaires a été réalisée par dichroïsme circulaire et les informations sur leur taille et sur leur forme ont été obtenues par diffusion dynamique de la lumière (DLS) et par diffusion des rayons X aux petits angles (SAXS). Ces techniques nous ont permis de mesurer leur rayon hydrodynamique, leur rayon de gyration ainsi que la taille maximale des ectodomains en solution.

L'analyse des intensités de diffusion par SAXS nous a également permis de caractériser leur flexibilité et leur hétérogénéité conformationnelle. Les ectodomains des syndécans ont été étudiés en phase gazeuse par spectrométrie de masse en conditions natives et par mobilité ionique nous permettant de définir leur espace conformationnel ainsi que leur degré de compaction par le calcul de leurs surfaces accessibles au solvant (SASA) ainsi que leurs sections efficaces de collision (CCS). Les expériences de mobilité ionique ont été réalisées en collaboration avec le Dr Fabien Chirot (Institut des Sciences Analytiques – ISA, UMR5280 Villeurbanne), le Dr Luke MacAleese (Institut Lumière Matière – ILM, UMR5306, Villeurbanne), Mathilde Bouakil (ILM) et Aurélien Le Fèvre (ISA). Ces travaux ont fait l'objet d'un article soumis pour publication dont je suis le premier auteur. Des expériences préliminaires de diffusion dynamique de la lumière et de SAXS ont également été réalisées sur les ectodomains des syndécans portant leurs chaînes de glycosaminoglycane que nous avons exprimés en système eucaryote.

Le second objectif de nos travaux a été de créer le premier interactome exhaustif des syndécans par la curation manuelle de la littérature et par l'interrogation de plusieurs bases de données, de l'analyser puis de définir l'interactome commun aux quatre syndécans. Les informations du site d'interaction sur les syndécans ont également été apportées lorsqu'elles étaient disponibles. Ces travaux ont fait l'objet d'une revue dont je suis le premier auteur qui a été publiée dans *FEBS Journal*.

Nous avons également cherché à identifier de nouveaux partenaires interagissant directement avec la partie protéique de l'ectodomaine des syndécans par interférométrie de biocouche (BLI). Nous avons caractérisé l'un des complexes identifiés avec un fragment de la fibronectine, l'anastelline, par différentes approches comme la spectrométrie de masse et l'amarrage moléculaire (docking) en collaboration avec le Dr. Sergey Samsonov et son équipe (Université de Gdansk, Pologne).

Finalement, un troisième axe de recherche a porté sur le rôle des glycosaminoglycane et des acides sialiques présents dans le glycocalyx des globules rouges dans leurs propriétés rhéologiques. Nous avons comparé leur rôle à la surface des globules rouges de sujets sains et de patients drépanocytaires par agrégométrie et par ektacytométrie, cette-dernière technique permettant d'étudier la déformabilité des globules rouges. Ces travaux ont été réalisés en collaboration avec le Pr. Philippe Connes et son équipe (Laboratoire Interuniversitaire de Biologie de la Motricité – LIBM, EA7424, Université Lyon 1). Ils ont fait l'objet d'une publication dont je suis le premier auteur dans *American Journal of Hematology*.

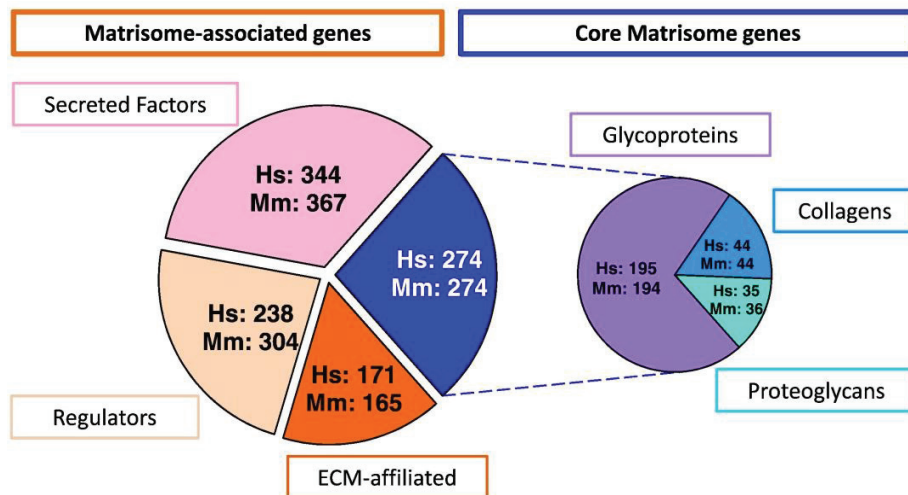


## **CHAPITRE II**

# **RAPPELS BIBLIOGRAPHIQUES**

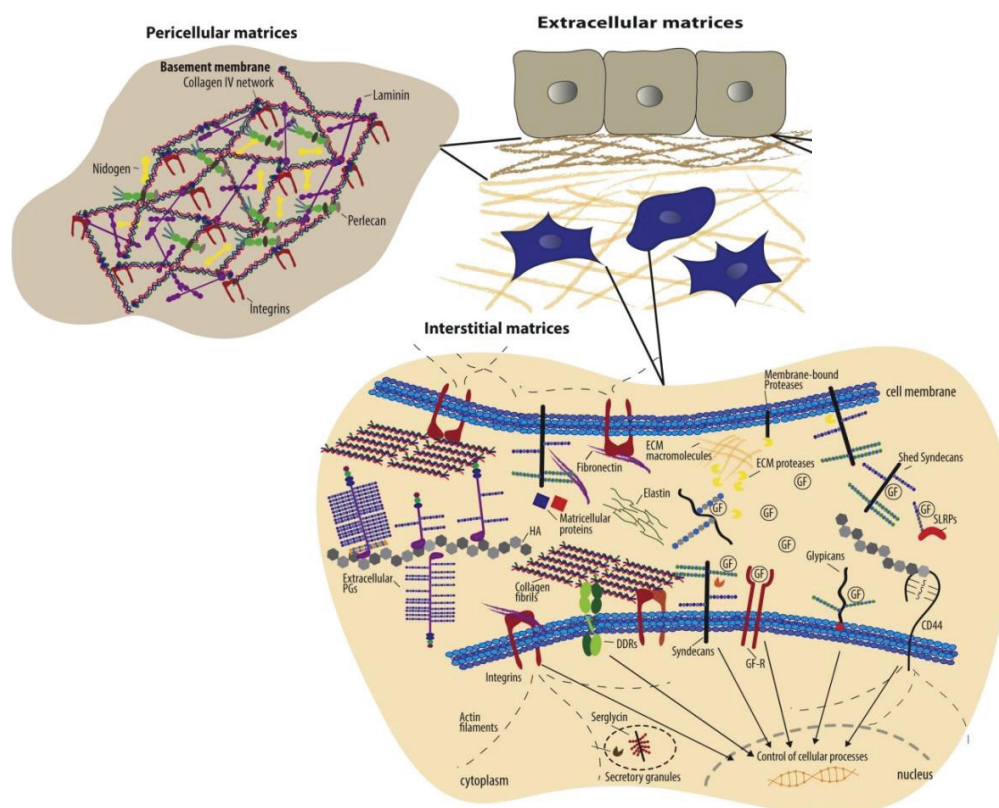
## II.1 La matrice extracellulaire

La matrice extracellulaire est un réseau tridimensionnel dynamique composé d'une grande diversité de molécules (*Theocharis et al., 2016*). D'abord considéré comme un support structural, la matrice extracellulaire régule en réalité un grand nombre de fonctions biologiques comme la croissance, la migration, la différenciation et la survie cellulaire (*Mecham, 2012, Naba et al., 2016*). La composition en biomolécules de la matrice extracellulaire varie selon les tissus mais plusieurs familles de molécules y sont retrouvées, ce qui constitue le matrisome. La matrisome a été défini *in silico* par une approche bioinformatique en se basant sur les caractéristiques des protéines de la matrice extracellulaire, c'est-à-dire la présence d'un peptide signal et de domaines protéiques fréquemment retrouvés dans ces protéines (*Naba et al., 2012, Naba et al., 2016*). La composition du matrisome dans différents tissus sains et pathologiques a été déterminée par protéomique quantitative (*Naba et al., 2016*). Certaines de ces données protéomiques sont disponibles dans la base de données MatrisomeDB (<http://www.pepchem.org/matrisomedb>, *Shao et al. 2020*). Ces techniques ont également permis de déterminer les biomolécules qui sont associées au matrisome (*Naba et al., 2016*). Chez l'Homme, le core matrisome contient environ 300 gènes codant majoritairement pour des glycoprotéines (environ 200) et pour des collagènes et des protéoglycans (**Figure 1**).



**Figure 1 : Composition des matrisomes humain et murin.** Nombre de gènes humains (Hs) et murins (Mm) codant pour des biomolécules systématiquement retrouvées dans les matrices extracellulaires (*core matrisome*) ou qui y sont associées (*matrisome-associated genes*). D'après *Naba et al., (2016)*.

Les protéines associées au matrisome ont été regroupées en trois catégories, les facteurs sécrétés, les régulateurs de la matrice ou les biomolécules qui y sont affiliées (**Figure 1**). Au total, près de 800 protéines sont associées au core matrisome. Les facteurs sécrétés sont représentés par de petites protéines comme les facteurs de croissance ou les cytokines. Les protéines affiliées à la matrice sont notamment des protéoglycanes membranaires ou associés aux membranes comme les syndécans ou les glypicans ainsi que des glycoprotéines et certaines sémaphorines. Enfin, les protéines ayant pour fonctions la régulation des assemblages extracellulaires, de leur réticulation et de leur dégradation sont regroupées dans la catégorie des régulateurs. Ils sont majoritairement composés d'enzymes impliquées dans la réticulation covalente de la matrice extracellulaire comme les enzymes de la famille des lysyl oxydases et des transglutaminases. On retrouve également des métalloprotéases matricielles ou des enzymes modifiant les glycosaminoglycanes comme les sulfatases (*Naba et al., 2016*). Deux types majeurs de matrice extracellulaire ont été décrits : la matrice péricellulaire, qui est à proximité directe des cellules comme les membranes basales, et la matrice interstitielle, plus éloignée (*Theocharis et al., 2016*) (**Figure 2**).

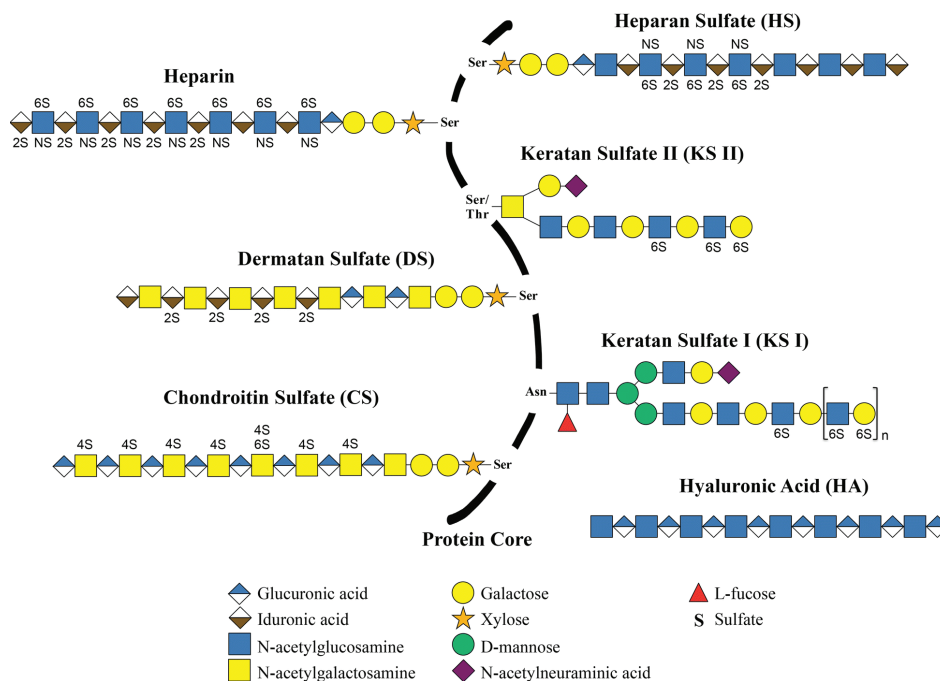


**Figure 2 : Organisation de la matrice extracellulaire, péricellulaire et interstitielle.** D'après *Theocharis et al., (2016)*.

## II.2 Les glycosaminoglycanes

Les glycosaminoglycanes (GAGs) sont des polymères linéaires de sucres composés d'unités disaccharidiques. Six GAGs ont été décrits chez les mammifères et diffèrent par la composition de leur unité disaccharidique et leur taux de sulfatation.

On retrouve les chaînes d'héparane sulfate (HS), de chondroïtine sulfate (CS), de kératane sulfate (KS), de dermatane sulfate (DS), d'acide hyaluronique (HA) et d'héparine (HP) (Bülow & Hobert, 2006, Karamanos et al., 2018). L'unité disaccharidique des chaînes d'héparane sulfate et d'héparine est composée d'un résidu de N-acétylglucosamine et d'un acide glucuronique ou iduronique. L'unité disaccharidique des chaînes de kératane sulfate et d'acide hyaluronique est composée d'un résidu de N-acétylglucosamine et d'acide glucuronique (HA) ou de galactose (KS). Enfin, celle des chaînes de chondroïtine sulfate et de dermatane sulfate est composée de N-acétylgalactosamine et d'acide glucuronique (CS) ou iduronique (DS) (Figure 3).



**Figure 3 : Composition des glycosaminoglycanes et de leur site d'attachement sur les protéines.** Le hyaluronane ne se fixe pas de façon covalente sur les protéines. D'après Merida-de-Barros et al., (2018).

Les chaînes d'HS, HP, CS, KS et de DS sont sulfatées et donc chargées négativement. Elles sont liées de façon covalente à des protéines menant à la formation de protéoglycanes.

L'héparine possède le taux de sulfatation le plus important des GAGs et sa densité de charge négative est la plus importante des biomolécules (*Sasarman et al., 2015*). Les chaînes de CS, DS, HS et d'héparine sont liées aux protéines *via* des résidus de sérine (O-glycosylation) par l'intermédiaire d'un tétrasaccharide composé d'un résidu de xylose, de deux résidus de galactose et d'un résidu d'acide glucuronique sur lequel sont greffés ces GAGs (**Figure 3**). Les chaînes de KS sont soit O-liées soit N-liées aux protéines sur un résidu d'asparagine ou de sérine / thréonine, respectivement. La taille du motif saccharidique reliant les chaînes de KS aux protéines varie et n'est pas restreinte à 4 résidus osidiques (*Sasarman et al., 2015*). A l'inverse, l'acide hyaluronique, qui n'est pas sulfaté, ne se lie pas de façon covalente aux protéines. Il interagit avec des protéoglycanes de la famille des hyalectans (*cf.* ci-dessous) par l'intermédiaire de protéines de liaison pour former des agrégats de très haute masse moléculaire (*Garantziotis & Savani, 2019*). L'unité disaccharidique des chaînes de GAGs sulfatés est répétée en moyenne de 50 à 150 fois (*Bülow & Hovert, 2006*) mais certaines chaînes de GAGs peuvent atteindre une masse moléculaire de 100 kDa (*Xu & Esko, 2014*). L'unité disaccharidique du hyaluronane est en revanche répétée entre 2000 et 25 000 fois ce qui mène à la synthèse de molécules d'HA dont la taille varie de 2 à 25  $\mu\text{m}$  (*Toole, 2004*).

Les chaînes de GAGs sulfatés sont modifiées pendant leur synthèse dans l'appareil de Golgi par des processus d'épimérisation pour les chaînes de DS, HS et HP, ou d'addition de groupements sulfate. Ces fonctions peuvent être ajoutées sur les chaînes d'héparane sulfate par des N-, 2-O, 6-O et 3-O sulfotransférases et l'acide glucuronique peut être épimérisé en acide iduronique par l'activité d'une C5-épimérase (*Habuchi et al., 2004, Karamanos et al., 2018*). Cela confère aux HS une diversité structurale très importante puisque 48 disaccharides différents peuvent théoriquement les composer et 23 ont été décrits jusqu'à présent (*Esko & Selleck, 2002, Li & Kusche-Gulberg, 2016*). Ces chaînes peuvent également être partiellement désulfatées à la surface cellulaire par les deux endosulfatases extracellulaires, Sulf-1 et Sulf-2 (*Nishitsuji, 2018*). Au moins 26 enzymes différentes modulant la structure des chaînes d'HS ont été décrites (*Bishop et al., 2007*). Le type et l'abondance de ces modifications sont assurés par la machinerie enzymatique propre à chaque type cellulaire (*Couchman, 2010*) et peuvent être altérés dans plusieurs conditions pathologiques (*Negrone et al., 2014, Soares da Costa et al., 2017*).



La sulfatation des unités disaccharidiques des chaînes d'héparane sulfate n'est pas distribuée de façon homogène le long de la chaîne et engendre la formation de régions sulfatées (domaines NS ou S) ou non sulfatées (domaines NA) qui sont séparés par des régions de degré de sulfatation intermédiaire (domaines NA/NS) (*Couchman, 2010, Sarrazin et al., 2011, Nishitsuji, 2018*). Les domaines NS sont relativement courts et comprennent généralement de 6 à 16 unités disaccharidiques alors que les domaines NA en comptent environ 18. La majorité des chaînes d'HS est composée de domaines ayant des degrés de sulfatation intermédiaire (*Xu & Esko, 2014*).

Les domaines NS sont des sites d'interaction privilégiés pour une multitude de protéines extracellulaires. Ces interactions sont médiées par des interactions électrostatiques auxquelles viennent parfois s'ajouter des liaisons hydrogène et des interactions hydrophobes (*Couchman et al., 2010, Xu & Esko, 2014, Li & Kusche-Gulberg, 2016*). Les domaines NS des chaînes d'héparane sulfate ressemblent à l'héparine dont l'interactome comprend au moins 435 partenaires où l'on retrouve notamment des protéines de la matrice extracellulaire, des chimiokines, des enzymes, des facteurs de croissance et des pathogènes (*Ori et al., 2011, Karamanos et al., 2018*). Une première étude portant sur la séquence des protéines interagissant avec l'héparine avait permis de déduire l'existence de séquences consensus d'interaction présentes chez ses partenaires et comprenant trois (xBBxBx), quatre (xBBBxxBx) ou huit (xBBBxxBBBxxBBx) résidus basiques (B, x étant n'importe quel résidu) (*Cardin & Weintraub, 1989*). Cependant, la majorité des protéines interagissant avec les chaînes d'HS/HP dont la séquence d'interaction a été décrite ne possède pas ces motifs ce qui montre que la présence de séquences consensus n'est pas un paramètre requis dans l'interaction protéines – HS/HP (*Karamanos et al., 2018*). Les partenaires protéiques doivent cependant posséder une région basique, comprenant généralement entre 4 à 7 résidus d'arginine ou de lysine (*Xu & Esko, 2014*).

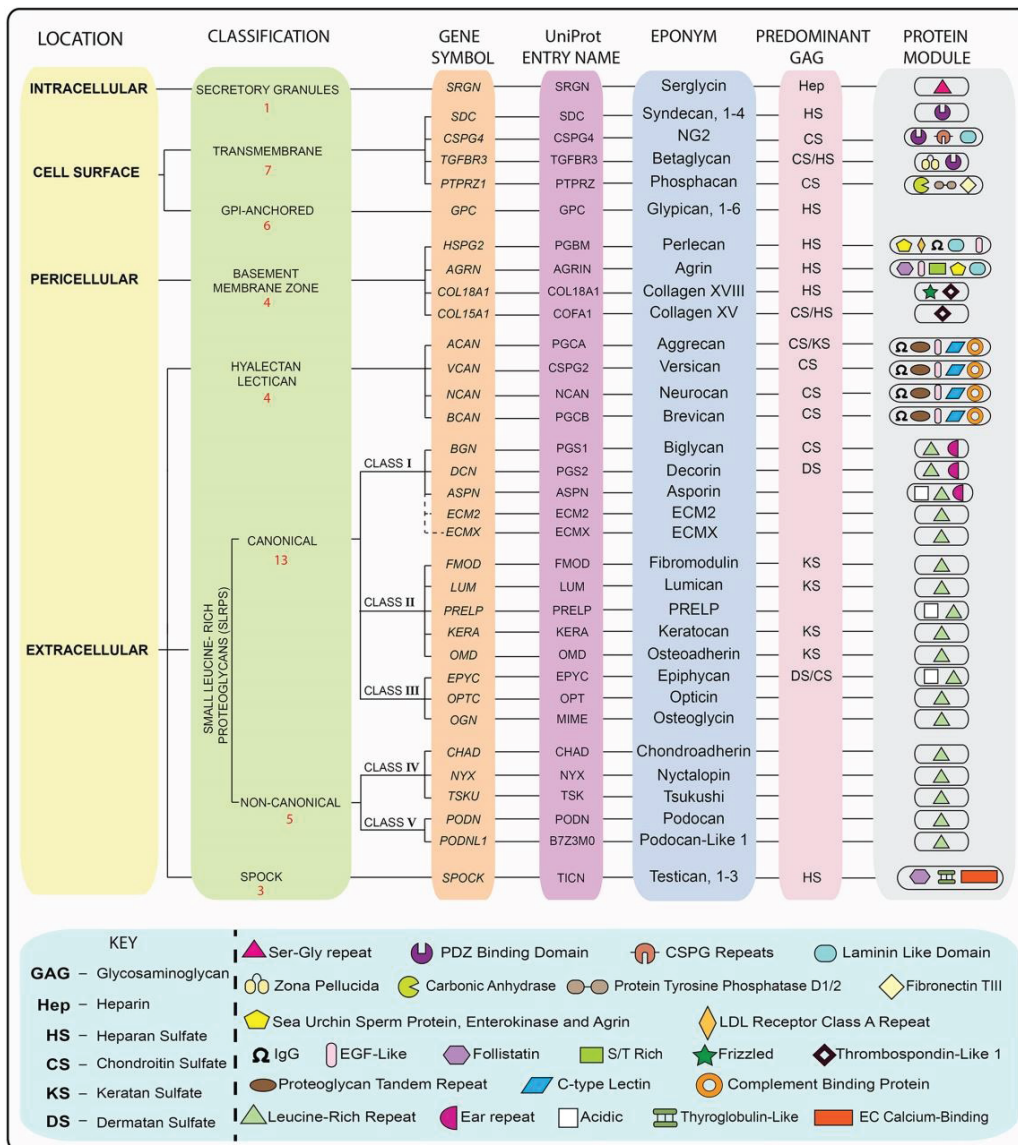
Certaines pathologies impliquent directement les chaînes de GAGs et sont causées par la mutation de gènes codant pour la machinerie enzymatique régulant leur biosynthèse, leur modification ou leur dégradation. La mutation de gènes codant pour le transporteur de sulfate (DTDST, ou diastrophic dysplasia protein) provoque certaines formes d'achondrogenèse ou de dysplasie diastrophique (nanisme). La mutation de la DS épimérase-2 mène à des troubles bipolaires ou des dépressions et celle de la dermatane sulfotransférase (D4ST-1) provoque un syndrome d'Ehlers-Danlos, caractérisé notamment par une hyperlaxité des articulations (*Mizumoto et al., 2013*).

Le syndrome de Larsen, caractérisé par des luxations congénitales des grosses articulations (genou, épaule), et le syndrome de Desbuquois, qui est marqué entre autres par un nanisme sévère et une dysmorphie faciale, sont causés par la mutation du gène codant pour la chondroïtine 6-O-sulfotransférase-1 (*Mizumoto et al., 2013*). Certaines pathologies sont associées à un excès de sulfatation, comme la maladie d'Alzheimer, ou de désulfatation des GAGs, notamment certains glioblastomes (*da Costa et al., 2017*). Finalement, un déficit en enzymes de dégradation des GAGs provoquent des mucopolysaccharidoses engendrant leur accumulation dans les lysosomes (*Fecarotta et al., 2020*).

### II.1.3 Les protéoglycanes

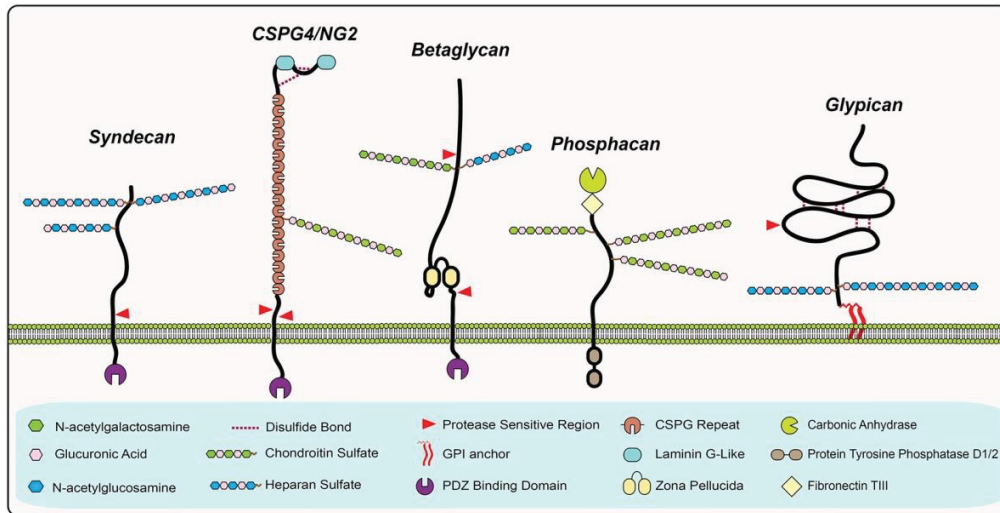
Les protéoglycanes sont des protéines sur lesquelles sont greffées de façon covalente une ou plusieurs chaînes de glycosaminoglycanes. Chez les mammifères, l'épissage alternatif de 43 gènes mène à la formation d'un nombre important de protéoglycanes qui ont été classés en fonction de leur localisation : extracellulaire, péricellulaire, surface cellulaire ou intracellulaire (**Figure 4**) (*Iozzo & Schaefer, 2015*). Les protéoglycanes sont exprimés par toutes les cellules eucaryotes et sont retrouvés dans toutes les matrices extracellulaires. Ils possèdent une très grande diversité fonctionnelle qui leur est notamment conférée par les chaînes de GAGs qu'ils portent (*Karamanos et al., 2018*).

La nature des chaînes de GAGs des protéoglycanes évolue avec la distance séparant la cellule et l'espace extracellulaire. Les protéoglycane à héparane sulfate (HSPG) sont majoritaires au niveau intracellulaire et membranaire. Seule la serglycine est intracellulaire et porte des chaînes d'héparine (*Iozzo & Schaefer, 2015*). Des chaînes d'héparane sulfate, de chondroïtine sulfate et de dermatane sulfate peuvent également y être attachées selon le type cellulaire (*Kolset & Pejler, 2011, Karamanos et al., 2018*). Elle a initialement été identifiée dans les granules des mastocytes où elle assure la fonction de réservoir d'enzymes et d'histamine mais son expression a été observée dans l'ensemble des cellules hématopoïétiques où elle régule la réponse immunitaire (*Karamanos et al., 2018*). La serglycine est également un agent pro-inflammatoire et pro-angiogénique lorsqu'elle est sécrétée dans la matrice extracellulaire par les cellules tumorales et elle est associée à un phénotype cancéreux agressif et pro-invasif (*Manou et al., 2020*). La serglycine est associée à d'autres pathologies comme l'athérosclérose, les maladies rhumatismales ou la dermatomyosite (*Scully et al., 2012*).



**Figure 4 : Classification des protéoglycans.** D'après *Iozzo & Schaefer* (2015).

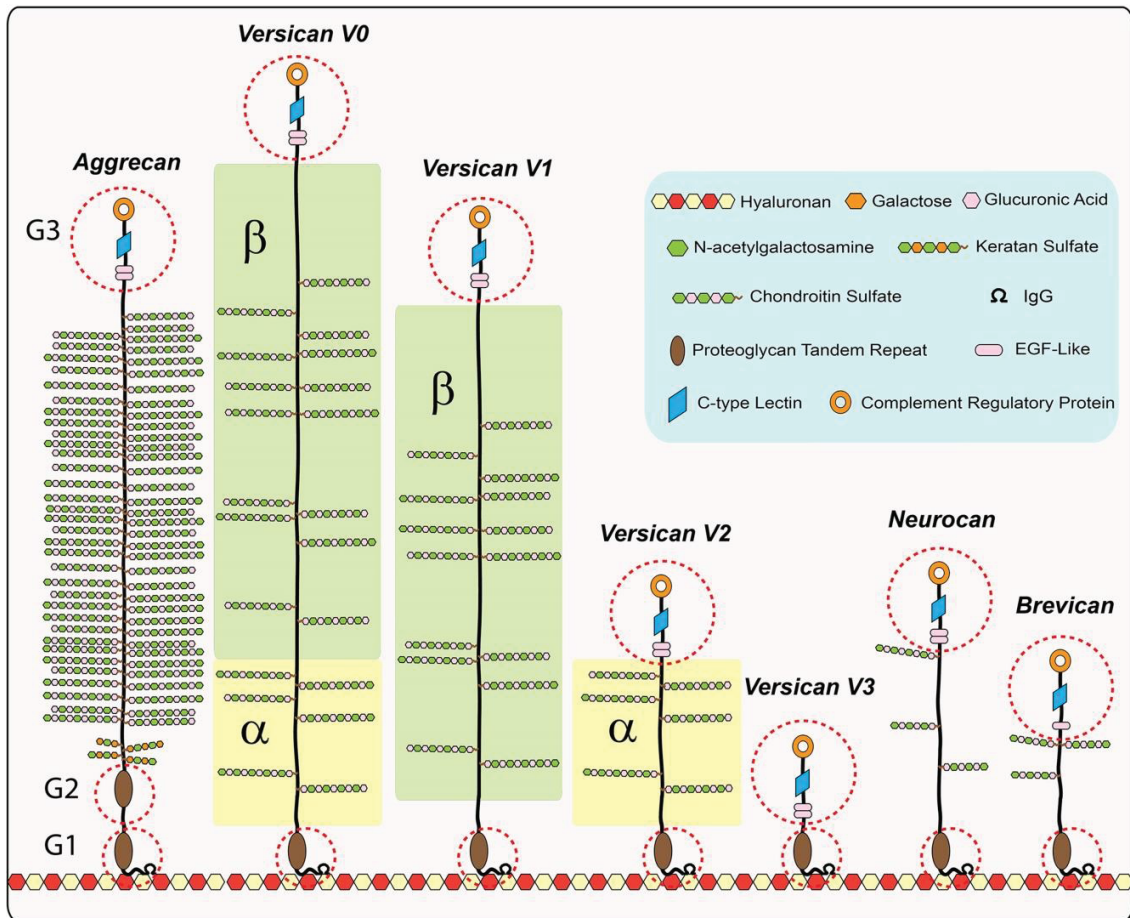
Les protéoglycans de la surface cellulaire sont principalement représentés par la famille des syndécans qui compte quatre membres (SDC-1-4), et par celle des glypicans qui en comprend six (GPC-1-6) (**Figure 5**). Ils seront décrits de façon détaillée dans le paragraphe II.4.



**Figure 5 : Protéoglycanes membranaires.** CSPG4/NG2 : protéoglycane chondroïtine sulfate-4 ou antigène neural/gliale 2. Figure tirée de *Iozzo & Schaefer (2015)*. GPI : glycosylphosphatidylinositol, PDZ : PSD-95, Discs-large, Zonula occludens-1, CSPG : chondroïtine sulfate protéoglycane.

L'abondance des HSPG reste élevée dans la matrice péri-cellulaire notamment dans les membranes basales qui sont des feuillets de matrice extracellulaire faisant la jonction entre un tissu épithélial, mésothélial ou endothélial et le tissu conjonctif sous-jacent (*Randles et al., 2016*). Elle est majoritairement constituée de laminines, de collagène IV, de nidogènes et de protéoglycanes à héparane sulfate (perlécane et agrine) et de collagènes XV et XVIII qui peuvent également porter des chaînes de CS (*Amenta et al., 2005, Couchman & Pataki, 2012*).

Finalement, les protéoglycanes présents dans l'espace extracellulaire sont enrichis en chaînes de chondroïtine sulfate, de kératane sulfate et de dermatane sulfate. La famille des hyalectans regroupe l'aggrécane, le versican, le brevican ainsi que le neurocan qui sont des macromolécules dont la masse moléculaire de la partie protéique peut atteindre 300 kDa. Ce sont des composants clés du cartilage, des vaisseaux sanguins et du système nerveux central (*Howell & Gottschall, 2012, Iozzo & Schaefer, 2015*). Les hyalectans possèdent un domaine globulaire à leurs extrémités N- et C-terminales leur permettant d'interagir avec de nombreux partenaires, notamment l'acide hyaluronique, avec lequel ils forment des agrégats dans la matrice extracellulaire (*Karamanos et al., 2018*) (**Figure 6**).



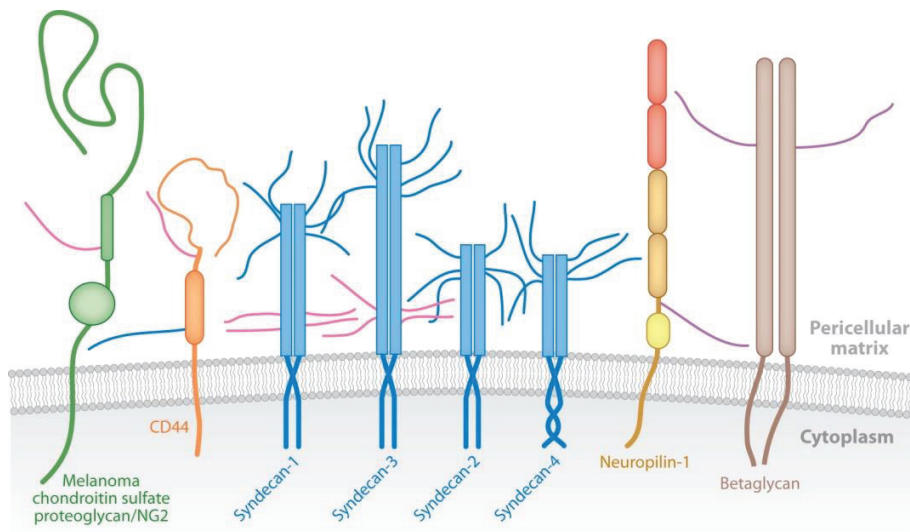
**Figure 6 : Les protéoglycans de la famille des hyalectans.** V0-3 : variants d'épissage. D'après Iozzo & Schaefer (2015).

A l'inverse, les petits protéoglycans riches en leucines (SLRPs) forment une famille de 18 protéines de faible masse moléculaire (36 – 42 kDa) constituées de la répétition de séquences riches en leucine (*leucine-rich repeat*, LRR) comme la décorine, le biglycan, ou le lumican. Ils portent majoritairement des chaînes de chondroïtine sulfate ou de kératane sulfate et sont impliqués dans la régulation de la fibrillogenèse des collagènes, la prolifération cellulaire, l'apoptose, l'immunité innée, la cicatrisation ainsi que dans les processus d'inflammation et de fibrogenèse (Merline et al., 2009, Iozzo & Schaefer, 2015, Zou et al., 2018, Appunni et al., 2019, Pang et al., 2020). Enfin, la famille SPOCK (pour *SPARC/Osteonectin CWCV and Kazal-like domain proteoglycans*) regroupe trois testicains qui sont des protéoglycans à héparane sulfate interagissant avec le calcium grâce à la présence de deux mains EF (Iozzo & Schaefer, 2015). Ils peuvent également porter des chaînes de CS et sont pratiquement exclusivement exprimés dans le système nerveux central où ils sont impliqués dans la régulation neuronale.

## II.4 Les protéoglycanes associés aux membranes

### II.4.1 Les protéoglycanes transmembranaires

Les protéoglycanes transmembranaires possèdent un court domaine intracellulaire (**Figure 7**). Les quatre syndécans ainsi que le protéoglycane à chondroïtine sulfate-4 (CSPG4), appelé également NG2 (Neural/glial antigen 2), sont des protéoglycanes à temps plein (ou *full-time proteoglycans*) ce qui signifie qu'ils portent constamment au moins une chaîne de glycosaminoglycane sur leur partie protéique (*Couchman, 2010*). A l'inverse, le récepteur CD44, la neuropiline-1, le phosphacan et le bétaglycan sont qualifiés de *part-time proteoglycans* et ne portent pas toujours de chaînes de GAGs.



**Figure 7 : Protéoglycanes transmembranaires.** Les chaînes d'héparane sulfate sont en bleu et celles de chondroïtine sulfate sont en rose. Des chaînes d'héparane sulfate ou de chondroïtine sulfate (violet) peuvent être présentes sur la neuropiline-1 et le bétaglycan. NG2 : Neural/glial antigen 2. D'après *Couchman (2010)*.

CSPG4 est un long protéoglycane membranaire de 2322 résidus d'acides aminés sur lesquels sont greffées jusqu'à 3 chaînes de chondroïtine sulfate (*Karamanos et al., 2018*). CSPG4 est peu exprimé dans les tissus sains mais son expression est en revanche très importante dans les mélanomes, ce qui permet de l'utiliser comme marqueur afin d'identifier notamment les exosomes dérivés de cellules malignes par rapport à ceux issus de cellules saines (*Ferrone & Whiteside, 2020*).

A la surface des cellules de mélanomes, CSPG4 agit en tant que support permettant l'organisation d'un complexe de signalisation promouvant la motilité cellulaire. Ce processus est aboli par la mutation de l'unique résidu de cystéine présent dans son domaine transmembranaire (*Yang et al., 2019*).

Le gène codant pour CD44 comprend 20 exons et 10 d'entre eux subissent un processus d'épissage alternatif (*Karamanos et al., 2018*). Deux catégories ont été décrites et sont différenciables par la présence (CD44v) ou l'absence (CD44s) d'une chaîne de glycosaminoglycane. Il s'agit du seul récepteur du hyaluronane ayant une chaîne d'héparane sulfate ou de chondroïtine sulfate (*Chen et al., 2018*). Les deux isoformes de CD44 interagissent avec le hyaluronane et leurs modifications post-traductionnelles leurs permettent également d'interagir avec d'autres protéines comme l'ostéopontine, le collagène IV, la fibronectine, des laminines, la serglycine ou le versican (*Karousou et al., 2016, Chen et al., 2018*). Il agit également comme corécepteur au facteur de croissance épidermique (EGF) et hépatique (HGF) (*Couchman et al., 2010*). Le degré de sulfatation de sa chaîne de GAG module sa capacité à interagir avec la fibrine et la surexpression de CD44 a été associée avec la progression de plusieurs cancers (*Karousou et al., 2016, Karamanos et al., 2018, Zhang et al., 2019*).

La neuropiline-1 (NRP-1) est un protéoglycane transmembranaire de type I d'environ 130 kDa. Deux neuropilines ont été décrites (NRP-1 et NRP-2) mais seule NRP-1 porte une unique chaîne de glycosaminoglycane (HS ou CS) fixée dans la région juxtamembranaire. Son ectodomaine comporte deux domaines N-terminaux interagissant avec les sémaphorines et deux domaines juxtamembranaires impliqués dans l'interaction avec l'héparine et le facteur de croissance de l'endothélium vasculaire (VEGF), dont elle est le corécepteur (*Couchman, 2010, Chaudhary et al., 2014*). Chez l'Homme, NRP-1 est exprimée par les cellules dendritiques plasmacytoïdes, l'endothélium artériel et par certains lymphocytes T régulateurs montrant son implication dans les mécanismes immunitaires (*Chaudhary et al., 2014*). Le type de glycosaminoglycane (HS ou CS) porté par NRP-1 varie selon le type cellulaire à la surface duquel elle est exprimée (*Shintani et al., 2006*). NRP-1 porte des chaînes d'HS à la surface des cellules endothéliales sur lesquelles le VEGF se fixe et sa fixation potentialise la signalisation induite par son récepteur VEGFR-2. A la surface des cellules musculaires lisses, NRP-1 est modifiée par une chaîne de chondroïtine sulfate avec laquelle le VEGF interagit également. Cependant, NRP-1 agit comme un leurre (decoy receptor) sous cette forme et inhibe partiellement la voie de signalisation médiée par VEGFR-2 (*Shintani et al., 2006*).

L'interaction de NRP-1 avec la sémaphorine-3A diminue l'infiltration des lymphocytes T régulateurs au sein des tumeurs solides et sa neutralisation par des anticorps permet d'augmenter leur migration et leur cytotoxicité, faisant de NRP-1 une cible thérapeutique de choix dans certains types de cancers (*Leclerc et al., 2020*).

Le phosphacan (460 kDa) est l'ectodomaine du membre  $\gamma$  des phosphatases membranaires (receptor protein tyrosine phosphatases, RPTP- $\gamma$ ) libéré par le processus de shedding. Il porte jusqu'à 4 chaînes de kératane sulfate ou de chondroïtine sulfate et est majoritairement exprimé dans le cerveau où il est notamment impliqué dans les processus de différenciation neuronale et de remyélinisation (*Melrose, 2019*).

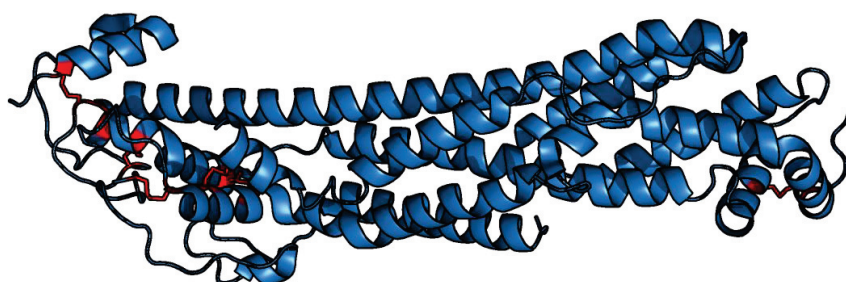
Le bétaglycan est un protéoglycane de 851 résidus d'acides aminés interagissant avec plusieurs membres de la superfamille du TGF- $\beta$  dont le TGF- $\beta$ , l'inhibine A, des activines, des protéines morphogénétiques osseuses (BMPs) ou des facteurs de croissance et de différenciation (GDFs) (*Bilandzic & Stenvers, 2011*). Le bétaglycan contient deux sites d'attachement pour des chaînes d'héparane ou de chondroïtine sulfate qui ne sont pas essentielles à l'interaction avec les membres de la superfamille du TGF- $\beta$ .

D'autres protéoglycanes transmembranaires ont été décrits (*Couchman & Pataki, 2012*) comme la thrombomoduline qui est un protéoglycane à chondroïtine sulfate ayant une fonction anticoagulante et qui est impliquée dans les processus d'inflammation et d'embryogenèse (*Weiler & Isermann, 2003*). Le neuroglycan-C porte également des chaînes de chondroïtine sulfate (jusqu'à 6) et est exprimé uniquement dans le système nerveux central où il intervient dans la migration neuronale et dans la formation des neurites (*Ooshira et al., 2004, Maeda, 2015*). L'agrine est majoritairement retrouvée sécrétée dans les membranes basales mais un variant d'épissage est exprimé sous forme transmembranaire dans les tissus nerveux (*Iozzo & Schaefer, 2015*). Elle peut porter jusqu'à trois chaînes d'héparane sulfate, responsables de la formation de filopodes au niveau des axones et des dendrites des neurones (*McCarthy, 2015, Karamanos et al, 2018*). Enfin, le protéoglycane SV2 (synaptic vesicle protein 2) fait partie d'un groupe de protéines possédant 12 domaines transmembranaires qui sont exprimées dans le système nerveux central et périphérique ainsi que dans le tissu cardiaque. Il peut porter jusqu'à trois chaînes de kératane sulfate et est impliqué dans le stockage et le transport de neurotransmetteurs (*Wan et al., 2010*). Il agit également comme réservoir d'ions calcium et joue un rôle dans la génération de potentiels d'actions au niveau synaptique (*Caterson & Melrose, 2018*).



#### II.4.2 Les protéoglycanes à ancre glycosylphosphatidylinositol

Les six glypicans (GPCs) sont les seuls représentants de la famille des protéoglycanes à ancre glycosylphosphatidylinositol (GPI). Ils possèdent un peu moins de 600 résidus d'acides aminés dont 14 résidus de cystéine conservés qui forment 7 ponts disulfures (*Filmus et al., 2008, Karamanos et al., 2018*) (**Figure 8**). Les glypicans sont des protéoglycanes qui peuvent porter jusqu'à trois chaînes d'héparane sulfate dans la région juxtamembranaire. Ses chaînes d'héparane sulfate lui permettent d'interagir avec divers facteurs de croissance comme Wnts, hedgehog, des facteurs de croissance fibroblastiques et des protéines morphogénétiques osseuses. Hedgehog a également une forte affinité pour la partie protéique du glypican-3. Ils régulent ainsi des processus de prolifération cellulaire, de mobilité et de différenciation cellulaire ainsi que l'angiogenèse (*Li et al., 2018, Karamanos et al., 2018*). Ses chaînes d'HS jouent également un rôle dans la localisation cellulaire des glypicans puisqu'en leur présence, ils sont exprimés au niveau des membranes basolatérales des cellules polarisées mais à la surface de leur pôle apical en leur absence (*Filmus et al., 2008*). Seule la structure cristallographique du glypican-1 a été résolue (*Svensson et al., 2012*) et montre une structure repliée riche en hélices  $\alpha$  (**Figure 8**).



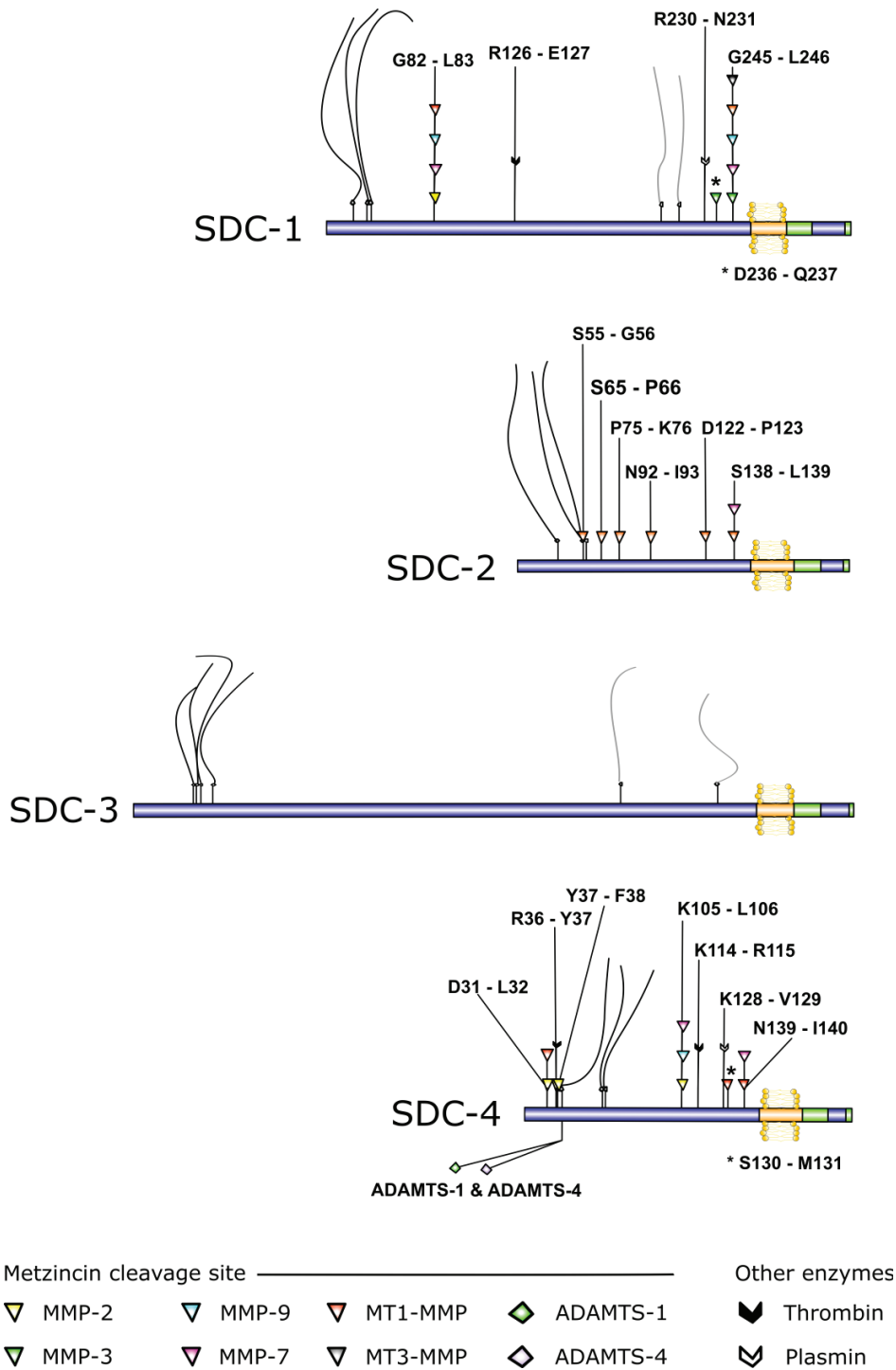
**Figure 8 : Structure cristallographique du glypican-1 humain.** Les résidus de cystéine et les ponts disulfures sont en rouge (PDB : 4AD7). Le glypican-1 (résidus 24-529) a été cristallisé et le résidu de sérine modifié par l'ancre GPI (Ser<sup>530</sup>) n'est pas présent.

Les glypicans -3, -5 et -6 peuvent être libérés dans l'espace extracellulaire par une phospholipase, Notum, qui clive l'ancre GPI. Le glypican-3 est surexprimé dans plusieurs cancers dont celui du foie mais pas dans les tissus sains ce qui en fait une cible thérapeutique de choix (*Li et al., 2018, Guo et al., 2020*). Plusieurs stratégies de traitement contre le cancer du foie impliquant le glypican-3 ont été développées comme l'utilisation d'un anticorps anti-glypican-3 ou d'un vaccin peptidique (*Shimizu et al., 2019*).

## II.5 Les syndécans

Chez les vertébrés, les syndécans sont une famille de quatre protéoglycanes membranaires de type I exprimés dans tous les types cellulaires, excepté les globules rouges (*Afratis et al., 2017, Zandonadi et al., 2019*). Le terme *syndecan* a été proposé par le groupe de Bernfield et vient du grec *syndein* qui signifie "réunir" ou "lier ensemble" (*Saunders et al., 1989*). Le syndécan-1 (syndécan) est ainsi le premier membre à avoir été décrit chez les vertébrés, puis les syndécan-2 (fibroglycan), -3 (N-syndécan) et -4 (amphiglycan ou ryudocan), avant d'être regroupés dans la même famille. Les quatre syndécans sont issus d'un gène ancestral ayant subi deux duplications successives qui sont survenues lors de la divergence Cnidaire – Bilatériens (0,6 – 1 milliard d'années, *Chakravarti & Adams, 2006*) ce qui explique que les syndécans réalisent des fonctions critiques depuis les cnidaires jusqu'aux mammifères. Ces deux duplications ont mené à l'apparition de deux groupes comprenant les syndécans -1 et -3 d'une part et les syndécans -2 et -4 d'autre part.

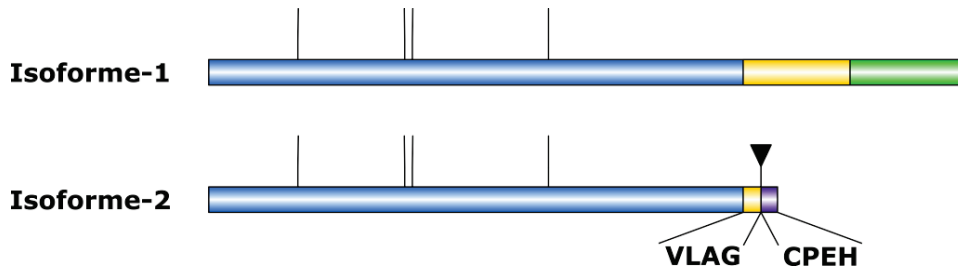
Chez les vertébrés, le syndécan-1 est majoritairement exprimé dans les cellules épithéliales et plasmatiques et le syndécan-2 dans les cellules mésenchymateuses comme les fibroblastes. Le syndécan-3 est presque exclusivement retrouvé dans les tissus neuronaux et musculo-squelettiques et le syndécan-4 est à l'inverse exprimé de façon quasi ubiquitaire (*Afratis et al., 2017*). Les syndécans se composent d'un court domaine cytosolique, d'un domaine transmembranaire et d'un domaine extracellulaire (ectodomaine, ED) (**Figure 9**). Le domaine cytosolique et le domaine transmembranaire sont très conservés entre les quatre membres (60 – 70 % d'identité de séquence) alors que la séquence de leur ectodomaine diverge (10 – 20 % d'identité). De courtes séquences d'attachement des glycosaminoglycanes sont cependant conservées entre les quatre syndécans et contiennent des séquences consensus sérine-glycine à l'extrémité N-terminale de leur ectodomaine (**Figure 9**). Les syndécans -1 et -3 portent en plus des chaînes de chondroïtine sulfate dans la région juxtamembranaire (*Couchman et al., 2015, Iozzo & Schaefer, 2015*). Le syndécan-4 peut également porter des chaînes de CS mais leurs sites d'attachements n'ont pas été décrits (*Deepa et al., 2004*). Les syndécans ne possèdent pas d'activité tyrosine kinase intrinsèque et réalisent leur fonction majoritairement en tant que corécepteurs (*Couchman et al., 2015*).



**Figure 9 : Organisation structurale des syndécans.** Les syndécans (SDCs) sont composés d'un court domaine cytosolique (vert et bleu, extrémité C-terminale), transmembranaire (jaune) et d'un ectodomaine (bleu, extrémité N-terminale) sur lequel sont attachées de façon covalente des chaînes d'héparane sulfate (noir) ou de chondroïtine sulfate (gris). Le SDC-4 pourrait également porter des chaînes de chondroïtine sulfate. Les représentations sont à l'échelle, excepté pour les chaînes de GAGs. Les sites de coupure des enzymes sont spécifiés. Figure adaptée de *Gondelaud & Ricard-Blum (2019)*.

Les syndécans peuvent subir le processus de shedding par l'action de métalloprotéases, qui sont des endopeptidases nécessitant un ion métallique ( $Zn^{+2}$  ou  $Ca^{2+}$ ) pour leur activité catalytique, ce qui mène à la libération de leur ectodomaine dans l'espace extracellulaire (*Gill & Parks, 2008*, revue par *Gondelaud & Ricard-Blum, 2019*). Les syndécans possèdent de nombreux sites de clivage au niveau juxtamembranaire ainsi que le long de leur ectodomaine (*Manon-Jensen et al., 2010, Gondelaud & Ricard-Blum, 2019*) (**Figure 9**). Ce clivage est catalysé au niveau juxtamembranaire par l'action de métalloprotéases matricielles qui sont transmembranaires (MT1- et MT3-MMP) ou non (MMP-2, -3, -7 et -9) (**Figure 9**). L'extrémité N-terminale du syndécan-4 peut être libérée par l'action de deux protéases de la famille ADAMs (A disintegrin and metalloproteinase), ADAMTS-1 et ADAMTS-4. D'autres enzymes permettent également la libération de tout ou partie de leur ectodomaine comme la thrombine ou la plasmine (*Manon-Jensen et al., 2013, Jannaway et al., 2019*). Le shedding des syndécans est stimulé dans plusieurs pathologies comme les cancers (*Ramani et al., 2013, Barbouri et al., 2014, Szatmári et al., 2015*), et les ectodomains des syndécans- 1, -2 et -4 sont des biomarqueurs dans de nombreuses pathologies. Ils peuvent également être utilisés comme prédicteurs de crise cardiaque ou de pneumonie aiguë communautaire (*Bertrand & Bollmann, 2008*). La partie résiduelle transmembranaire et cytosolique des syndécans 1, -2 et -3 résultant du processus de shedding est ensuite clivée dans la membrane par le complexe préséniline –  $\gamma$ -sécrétase (*Haapasalo & Kovacs, 2011*).

Le syndécan-4 possède deux isoformes. La première est membranaire et son organisation en domaines est similaire à celle des autres syndécans. La seconde isoforme est exprimée sous forme d'un ectodomaine soluble (*Xing et al., 2003*). Ce variant d'épissage est exprimé dans les mélanocytes, le cœur fœtal, l'utérus gravide et dans certains carcinomes. La séquence de l'isoforme-2 du syndécan-4 (ED-4) est identique à celle de l'ectodomaine de l'isoforme-1 (SDC-4) à laquelle s'ajoutent 4 résidus du domaine transmembranaire du SDC-4 (VLAG) ainsi que 4 résidus (CPEH) qui sont propres à cette isoforme. Cette seconde isoforme est exprimée dans plusieurs espèces mais seule la séquence humaine possède un résidu de cystéine qui est localisé à son extrémité C-terminale (**Figure 10**).



**Figure 10 : Comparaison de l'organisation en domaines des deux isoformes du syndécan-4.**

L'isoforme-1 possède un domaine cytosolique (vert), transmembranaire (jaune) et un ectodomaine (bleu), où les sites d'attachement des glycosaminoglycanes sont indiqués (lignes verticales). L'isoforme-2 (ED-4) conserve la séquence de l'ectodomaine de l'isoforme-1 et quatre de ses résidus transmembranaires (VLAG) auxquels s'ajoutent quatre résidus qui lui sont propres (CPEH, en violet). Le résidu de cystéine est indiqué par un triangle noir.

## II.5.1 Domaines et structures des syndécans

### a) Domaine cytosolique

Le domaine cytosolique des syndécans compte une trentaine de résidus d'acides aminés. Il est divisé en trois régions qui ont été déterminées en fonction de l'homologie existante entre les quatre syndécans. La région proximale C1, proche du feuillet interne de la membrane plasmique, et la région distale C2, présente à leur extrémité C-terminale, sont conservées aussi bien entre les quatre syndécans qu'entre les espèces (**Figure 11**). Elles sont séparées par la région V, dont la séquence est variable et est propre à chaque syndécans ce qui leur confère leur spécificité fonctionnelle (*Couchman et al., 2015*).

	C1	V	C2
SDC-1	RMKKKDEGSYSL	E <sup>EE</sup> PKQANGGAYQ <sup>K</sup>	PTKQEEFYA
SDC-2	RMRKKDEGSYDL	GERKPSS-AA <sup>YQ</sup> K	APTK-EFYA
SDC-3	RMKKKDEGSYTL	E <sup>EE</sup> PKQAS-VT <sup>YQ</sup> K	PDKQEEFYA
SDC-4	RMKKKDEGSYDL	GKKP-----I <sup>YK</sup> K	APTN-EFYA

**Figure 11 : Alignement des séquences du domaine cytosolique des quatre syndécans humains.** L'alignement des séquences permet d'identifier deux régions conservées (C1 et C2, bleu) et une région variable (V, brun). Les résidus en gras indiquent les deux seuls résidus conservés de la région V.

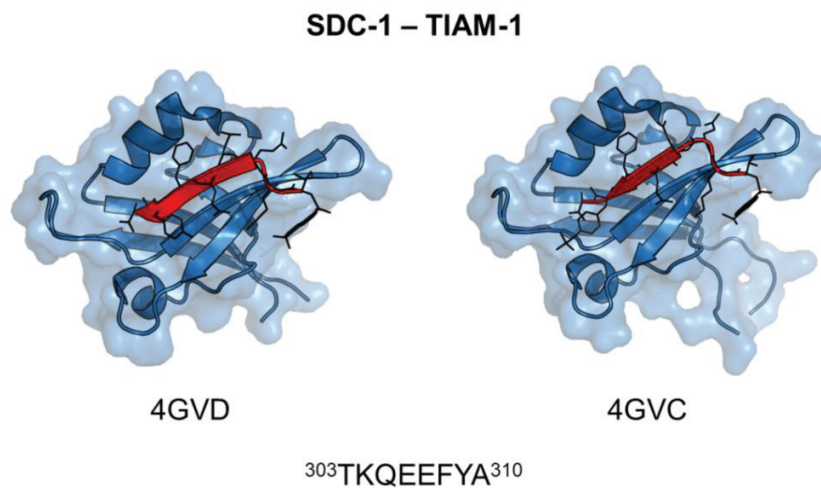
Les résidus N-terminaux de la région C1 comprennent une série de résidus basiques très conservés (KKK pour les SDCs -1, -3 et -4 et RKK pour le SDC-2). Ils sont responsables de l'endocytose du syndécan-1 par un mécanisme nécessitant la phosphorylation de résidus de tyrosine présents dans son domaine cytosolique (*Chen & Williams, 2013*). L'interaction d'un ligand extracellulaire avec les chaînes d'héparane sulfate du syndécan-1 mène à l'activation des kinases ERK (*Extracellular signal-regulated kinases*) et au regroupement de plusieurs molécules de SDC-1 dans des radeaux lipidiques. L'activation de ERK mène à la dissociation du complexe SDC-1 –  $\alpha$ -tubuline et à la phosphorylation de résidus de tyrosine présents dans son domaine cytosolique par la kinase Src (Sarc, de sarcome). Le site phosphorylé du SDC-1 lui permet d'interagir avec la cortactine, qui est une protéine impliquée dans l'organisation du cytosquelette d'actine mais aussi dans des mécanismes d'endocytose (*Chen & Williams, 2013*). La région C1 permet de façon générale l'interaction des SDCs avec des protéines du cytosquelette comme les tubulines, les protéines ERM (ezrine – radixine – moésine) et la cortactine, ainsi que leur internalisation par endocytose (*Couchman et al., 2015, Gondelaud & Ricard-Blum, 2019*).

La séquence de la région variable V n'est pas conservée entre les membres de la famille des syndécans mais est conservée entre les espèces. C'est par exemple le cas de la région V du syndécan-4 dont la séquence est homologue chez les mammifères, les oiseaux et le poisson zèbre (*Whiteford et al., 2008*). Seuls des résidus de tyrosine et de lysine sont conservés entre les syndécans (**Figure 11**).

La région C2 possède le motif conservé EFYA à son extrémité C-terminale qui est responsable de l'interaction des syndécans avec des protéines à domaine PDZ, pour *PSD-95*, *Discs-large*, *Zonula occludens-1*, qui sont les trois premières protéines où a été identifié ce domaine. Plus de 500 protéines possèdent un domaine PDZ dans le protéome humain (*Liu & Fuentes, 2019*). L'interaction entre les syndécans et les protéines à domaine PDZ est notamment régulée par la phosphorylation des résidus de sérine et de tyrosine conservés du domaine C1 et du résidu de tyrosine du motif EFYA (*Cheng et al., 2016*).

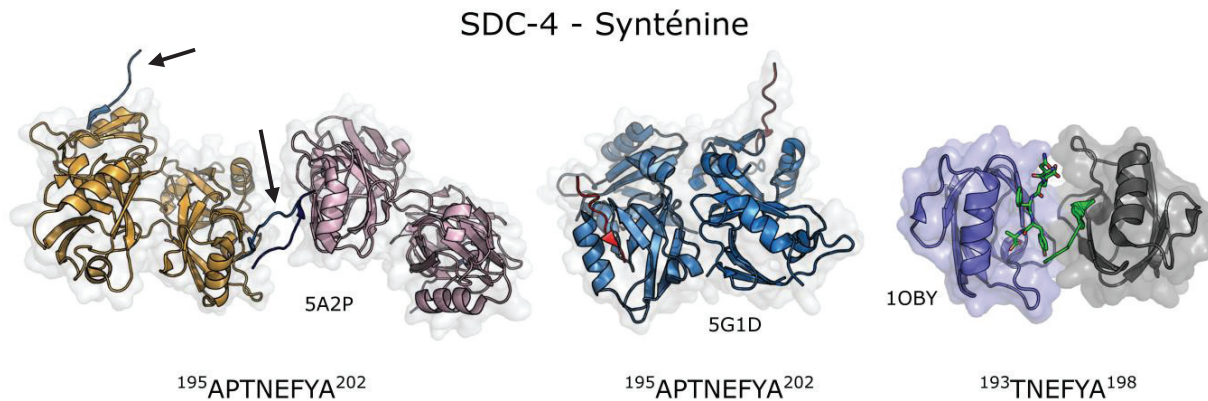
Aucune information structurale n'est disponible pour le domaine cytosolique du SDC-2 et du SDC-3. La totalité du domaine cytoplasmique du SDC-2 a été exprimée, purifiée puis analysée par résonance magnétique nucléaire (RMN) mais aucune information structurale n'a été rapportée par les auteurs de l'étude (*Chae & Lee, 2008*). L'extrémité C-terminale du domaine cytosolique du SDC-1 ( $^{303}\text{TKQEEFYA}^{310}$ ) a été cocrystallisée avec le domaine PDZ de la protéine TIAM-1 (T-lymphoma invasion and metastasis-inducing protein 1, *Liu et al., 2013*).

Le peptide du SDC-1 utilisé dans ces expériences correspond à son domaine C2, c'est-à-dire à son extrémité C-terminale, où le résidu de tyrosine 309 était phosphorylé (PDB : 4GVC) ou non (PDB : 4GVD) (**Figure 12**). Dans ces conditions, le domaine C2 du SDC-1 est replié en brin  $\beta$  et la phosphorylation du résidu de tyrosine 309 n'induit pas de modification majeure de sa structure (**Figure 12**).



**Figure 12 : Structures cristallographiques de la région C2 du syndécan-1 humain avec la protéine TIAM-1.** Le peptide de SDC-1 est en rouge et le domaine PDZ de TIAM-1 en bleu. La séquence du peptide comportant le résidu de tyrosine 309 phosphorylé (PDB : 4GVC) ou non (PDB : 4GVD) est indiquée.

Le domaine cytoplasmique du SDC-4 a été davantage étudié. Plusieurs structures cristallographiques ont été résolues avec la région C2 du SDC-4 et le domaine PDZ de la synténine-1 (ou SDCBP, Syndecan-binding protein 1). Le peptide contient soit les résidus <sup>193</sup>TNEFYA<sup>198</sup> (PDB : 1OBY et 1YBO) soit les résidus <sup>191</sup>APTNEFYA<sup>198</sup> (PDB : 5G1D et 5A2P) présents à l'extrémité C-terminale du SDC-4 humain ou de rat, les deux séquences étant strictement conservées (*Kang et al., 2003, Grembecka et al., 2006, Choi et al., 2016*) (**Figure 13**). Le domaine PDZ de la synténine-1 dimérise avec le domaine C2 du syndécan-4, ce-dernier restant majoritairement non structuré.

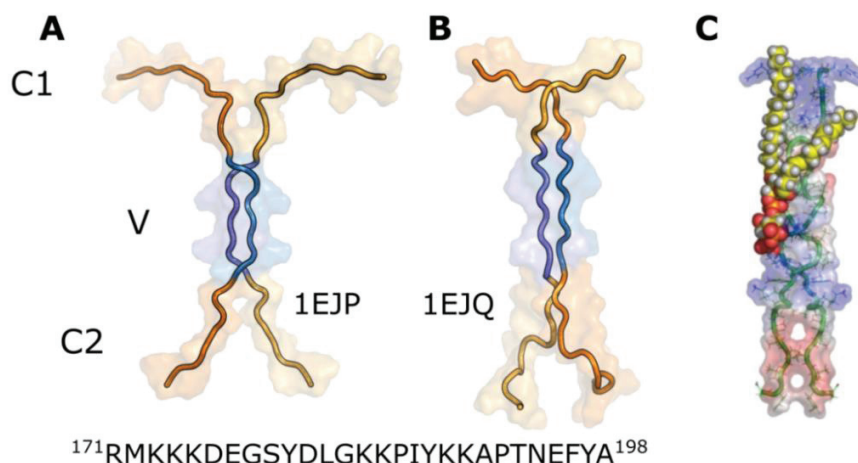


**Figure 13 : Structures cristallographiques de la région C2 du syndécan-4 cocristallisée avec la synténine.** Le peptide de SDC-4 est en bleu (flèche, 52AP), vert (10BY) et rouge (5G1D). Les séquences des peptides de SDC-4 sont indiquées.

La totalité du domaine cytoplasmique du SDC-4 de rat, dont la séquence est strictement identique à celle de l'Homme, a été résolue par RMN soit seul (*Shin et al., 2001*, PDB : 1EJP), soit en présence de phosphatidylinositol-4,5-bisphosphate (PIP2) (*Lee et al., 1998*, PDB : 1EJQ) (**Figure 14**). Dans des conditions physiologiques, le domaine cytosolique du SDC-4 dimérise et ce mécanisme est médié par des interactions ioniques où les deux chaînes sont entrelacées (**Figures 14A** et **14B**). Les deux chaînes forment également un petit feuillet composé de deux brins  $\beta$  au niveau de la région V. L'interaction avec le PIP2 provoque une légère modification structurale menant à la formation d'un dimère plus compact et d'un feuillet  $\beta$  plus long. Les coordonnées atomiques du PIP2 ne sont en revanche pas présentes dans le fichier PDB.

La structure du domaine cytosolique entier du SDC-4 de poisson zèbre a également été résolue par RMN en présence de PIP2 (*Whiteford et al., 2008*). La taille de son domaine cytosolique est identique à celui de l'Homme et du rat et sa séquence diffère de trois résidus d'acides aminés. Les structures du domaine cytosolique de mammifère ou du poisson zèbre sont ainsi très similaires (**Figure 14C**). Les résidus présents dans la région V interagissent aussi bien avec la tête polaire du PIP2 qu'avec ses chaînes acyles. La formation du complexe entre le SDC-4 et le PIP2 sert de site d'attachement à la protéine kinase C alpha (PKC $\alpha$ ) (*Couchman et al., 2015*).





**Figure 14 : Structures du domaine cytosolique du SDC-4 déterminées par RMN.** Structures du domaine cytosolique du SDC-4 humain / murin seul **(A)** ou en présence de PIP2 **(B)** résolues par RMN. Les coordonnées du PIP2 ne sont pas présentes dans le fichier PDB. Les régions C1 et C2 sont en orange et la région variable V est en bleu. **(C)** Structure du domaine cytosolique du SDC-4 du poisson zèbre résolue par RMN en présence de PIP2. Les deux chaînes du SDC-4 sont en vert et le PIP2 est en jaune et rouge en représentation sphère. Le potentiel électrostatique a été simulé par les auteurs (bleu : potentiel positif, rouge : négatif, blanc : neutre) (*Whiteford et al., 2008*).

En résumé, seule la structure de la région C2 du domaine cytosolique du syndécan-1 est connue et se replie en brin  $\beta$  lorsqu'elle interagit avec le domaine PDZ de la protéine TIAM-1. La structure de la totalité du domaine cytosolique du syndécan-4 est connue et dimérise. La région V forme un petit feuillet  $\beta$  alors que les régions C1 et C2 restent non structurées.

#### *b) Domaine transmembranaire*

Le domaine transmembranaire des quatre syndécans est également conservé (**Figure 15**). Il contient le motif GXXXG ainsi qu'un résidu de phénylalanine permettant leur homo- et hétéro-dimérisation (*Dews & MacKenzie, 2007, Kwon et al., 2016*).

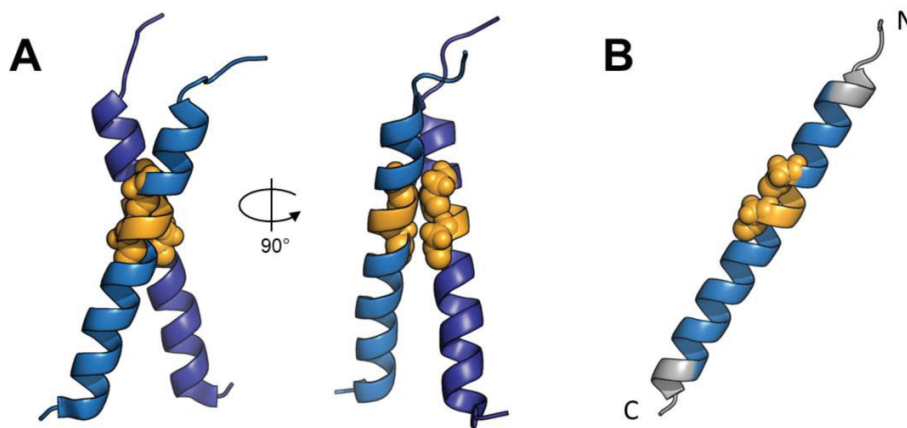
SDC-1	---GVIAG <b>GGLVG</b> LIFAVCLVGFML-	21
SDC-2	VLA <b>AVIAGGVIG</b> FLFAIFLILLLVY	25
SDC-3	--- <u>AVIV</u> <b>GGVVG</b> ALFAAFLVTLII-	21
SDC-4	VLA <b>ALIVGGIVG</b> ILFAVFLILLMY	25
	. : * . * * : : * : * * * : : :	

**Figure 15 : Alignement des séquences du domaine transmembranaire des syndécans humains.** Gras : motif GXXXG, résidus soulignés : motif AXXXG. Alignement réalisé avec Clustal Oméga.

Environ 12 % des protéines membranaires effectuant un unique passage dans la bicouche lipidique possède le motif de dimérisation GXXXG (Teese & Langosch, 2015). Il a initialement été décrit dans le domaine transmembranaire de la glycophorine A puis dans d'autres protéines comme certains transporteurs ABC, l'ATPase, la sous-unité d'intégrine  $\alpha$ IIb ou encore la neuropiline-1 (Teese & Langosch, 2015). Ce motif est également responsable de la formation de dimères de syndécans résistants au dodécylsulfate de sodium (SDS) lors de leur analyse par électrophorèse en conditions dénaturantes (Choi et al., 2005, Choi et al., 2015).

La présence d'un résidu de phénylalanine dans le domaine transmembranaire des syndécans favorise la formation d'homo- et d'hétérodimères (Kwon et al., 2015, Kwon et al., 2016). Le résidu d'alanine du domaine transmembranaire du SDC-3 favorise également leur formation (Jung et al., 2020). La formation d'homo- et d'hétérodimères de syndécans dans des micelles lipidiques et leur analyse par électrophorèse a permis de déterminer l'affinité relative entre les domaines transmembranaires des syndécans (Dews & MacKenzie, 2007). Dans ces analyses, la quantité d'homodimères formée par le syndécan-1 est la plus faible et celle formée par le syndécan-2 est la plus élevée. Les syndécans ont la capacité de former des hétérodimères avec les autres membres. Les hétérodimères de SDC-1 et de SDC-4 n'ont pas été observés mais d'autres études ont finalement permis de conclure positivement sur la formation d'hétérodimères médiée par le domaine transmembranaire du syndécan-1 et du syndécan-4 (Kwon et al., 2016). La formation d'une faible quantité d'homodimère de SDC-1 semble être causée par la présence de deux motifs GXXXG successifs au sein de son domaine transmembranaire (Chen et al., 2020) (**Figure 15**). Ces deux motifs entrent en compétition lors du processus d'homodimérisation et déstabilisent ainsi la formation d'homodimères. Des motifs de dimérisation similaires ont été décrits pour d'autres protéines membranaires comme le motif AXXXA (Kleiger et al., 2002) et AXXXG (Kairys et al., 2004), ce-dernier étant présent dans le domaine transmembranaire des SDC-2, -3 et -4 (**Figure 15**).

L'analyse par dichroïsme circulaire et par RMN du domaine transmembranaire du SDC-4 sous forme dimérique a montré qu'il se repliait en hélice  $\alpha$  asymétrique, c'est-à-dire que les deux monomères ne sont pas associés face à face mais forment un angle (Park et al., 2011). La formation de dimère asymétrique est couramment observée pour les protéines comportant un motif GXXXG dans leur domaine transmembranaire (Teese & Langosch, 2015) (**Figure 16A**). La structure du domaine transmembranaire du SDC-2 a été résolue par RMN et confirme un repliement en hélice  $\alpha$  (Li et al., 2019) (**Figure 16B**).



**Figure 16 : Domaine transmembranaire de la glycophorine A et du syndécan-2 humain.** Structure du domaine transmembranaire **(A)** de la glycophorine A sous forme dimérique (PDB : 2KPF) et **(B)** du syndécan-2 sous forme monomérique (PDB : 6ITH, *Li et al., 2019*). Bleu : chaîne polypeptidique, orange : motif GXXXG. Sphères : résidus de glycine du motif. Les résidus qui ne sont pas retrouvés dans le domaine transmembranaire du SDC-2 sont en gris et les extrémités N- et C-terminales sont annotées.

### *c) Domaine extracellulaire (ectodomaine)*

L'ectodomaine des syndécans est la région la moins caractérisée et possède une masse moléculaire d'environ 24 (SDC-1), 14 (SDC-2 et -4) et 35 kDa (SDC-3) (*Gondelaud & Ricard-Blum, 2019*). La séquence de l'ectodomaine est peu conservée et possède entre 10 et 20 % d'identité (*Iozzo & Schaefer, 2015*).

L'ectodomaine des syndécans peut être divisé en deux régions dont les extrémités sont mal définies (*Rapraeger, 2000, Beauvais & Rapraeger, 2004, Multhaupt et al., 2009*). La première, située dans la moitié N-terminale où sont retrouvées les séquences consensus sérine-glycine, constitue le domaine d'attachement des chaînes d'héparane sulfate. Le regroupement de résidus d'acides aminés acides à proximité de ce motif est nécessaire à la fixation des chaînes de GAGs, et la présence d'un résidu aromatique favorise l'attachement des chaînes d'héparane sulfate (*Zhang & Esko, 1994*). Les ectodomains du SDC-1, -2 et -4 portent trois chaînes d'héparane sulfate proches de leur extrémité N-terminale alors que le SDC-3 en porte quatre. Les chaînes d'héparane sulfate sont responsables de la majorité des interactions réalisées par les syndécans (revue par *Gondelaud & Ricard-Blum, 2019*).

La seconde région concerne la moitié C-terminale des ectodomains des syndécans, appelée domaine de liaison cellulaire (cell binding domain) en raison de son interaction directe avec d'autres protéines, notamment les intégrines. Cette fonction dans l'adhésion cellulaire a initialement été détectée chez le syndécan-4 (McFall & Rapraeger, 1997) puis chez les autres syndécans (McFall & Rapraeger, 1998, Whiteford et al., 2008, De Rossi et al., 2014, Jung et al., 2016). La séquence minimale requise pour cette fonction est la séquence  $^{87}\text{NHIP}^{90}$  du syndécan-4 humain et plus généralement NxIP dans les orthologues du syndécan-4 de mammifères, d'oiseaux, d'amphibiens et du poisson zèbre (Whiteford & Couchman, 2006). La mutation du résidu d'isoleucine diminue particulièrement l'adhésion cellulaire. Une séquence homologue est retrouvée dans l'ectodomaine du syndécan-2 humain ( $^{95}\text{NKIP}^{98}$ ) mais pas dans les autres syndécans, ni dans l'orthologue murin du syndécan-2 (Liyanagea et al., 2020). Une fonction similaire indépendante des intégrines a été retrouvée pour le syndécan-2 murin et nécessite la séquence s'étendant du résidu de proline 124 au résidu de phénylalanine 141 ( $^{124}\text{PAIKSTDVYTEKHSDNLF}^{141}$ ) permettant l'interaction avec le récepteur CD148 (Receptor-type tyrosine-protein phosphatase eta) (Whiteford et al., 2011).

Outres les chaînes de glycosaminoglycane, les ectodomains des syndécans possèdent de potentiels sites de O- et de N-glycosylation, mais aucune preuve expérimentale de leur glycosylation n'a été apportée (Chen et al., 2002, Velleman & Coy, 2020). Par exemple, 33 sites de O-glycosylation ont été prédits sur l'ectodomaine du syndécan-1 (Zhang et al., 2012). L'ectodomaine du syndécan-3 possède la particularité de contenir un domaine de type mucine, enrichi en résidus de proline et de thréonine et dans une moindre mesure de sérine (**Figure 17**). Près de 90 % (47/54) des résidus de thréonine de l'ectodomaine du syndécan-3 humain sont localisés dans ce domaine ainsi qu'environ 70 % (33/48) et 37 % (10/27) des résidus de proline et de sérine, respectivement. Ce domaine pourrait être très fortement O-glycosylé comme les mucines (Gould et al., 1992, Carey et al., 1996, Reily et al., 2019).

Finalement, un site de phosphorylation extracellulaire a été décrit pour l'ectodomaine du syndécan-2 uniquement (Tagliabracci et al., 2015). Le résidu de sérine 115 présent dans la région juxtamembranaire du SDC-2 est phosphorylé par Fam20C, une kinase impliquée dans la phosphorylation des protéines sécrétées comme le facteur de croissance fibroblastique 23 (FGF-23) ou l'ostéopontine (Tibaldi et al., 2020).



**Figure 17 : Localisation du domaine de type mucine au sein de l'ectodomaine du syndécan-3 humain.** Le domaine mucine est en rouge et le reste de l'ectodomaine en vert. La localisation des chaînes d'héparane sulfate (HS) et de chondroïtine sulfate (CS) est indiquée. Les résidus sont numérotés selon la numérotation UniProt KB (O75056).

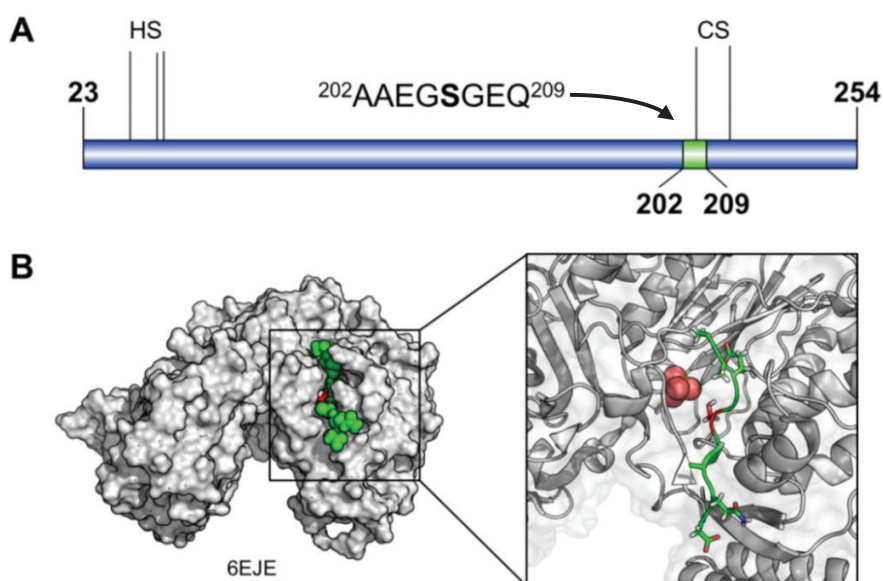
Un mécanisme d'homodimérisation non covalente de l'ectodomaine du syndécan-3 dépendant des résidus <sup>381</sup>ERKE<sup>384</sup> présents à son extrémité C-terminale a été décrit (*Asundi & Carey, 1995*). La mutation des résidus <sup>381</sup>ERKE<sup>384</sup> provoque l'abolition partielle ou totale de la formation de dimères ce qui a mené à l'hypothèse de leur stabilisation par des interactions électrostatiques (*Asundi & Carey, 1995*). La dimérisation des autres ectodomains des syndécans n'a pas été décrite bien qu'une séquence homologue (<sup>248</sup>DRKE<sup>251</sup>) soit présente dans l'ectodomaine du syndécan-1 humain (**Figure 18**). L'ectodomaine du SDC-2 et celui du SDC-4 humain possède une séquence similaire (<sup>141</sup>KRTE<sup>144</sup> et <sup>142</sup>ERTE<sup>146</sup>, respectivement) ne contenant cependant pas le résidu de lysine présent dans la séquence de l'ectodomaine du SDC-1 et du SDC-3 (**Figure 18**). La mutation du résidu de lysine dans la séquence <sup>381</sup>ERKE<sup>384</sup> du SDC-3 mène d'ailleurs à l'abolition de la formation de dimères ce qui montre l'importance de ce résidu (*Asundi & Carey, 1995*).

SDC-2	EEDTNVYTEKHSDSLF <b>KRTE</b> ---	20	(P18827)
SDC-4	SNKVSMSSSTVQGSNIF <b>ERTE</b> ---	20	(P34741)
SDC-1	---VDQGATGASQGLL <b>DRKE</b> VLG	20	(O75056)
SDC-3	---GSSAAQLPQKSIL <b>ERKE</b> VLV	20	(P31431)
	. : . . . . * . *		

**Figure 18 : Alignement des 20 derniers résidus d'acides aminés des ectodomains des syndécans humains.** Les résidus sont numérotés selon la numérotation UniProt KB (codes d'accession à droite) et l'alignement a été réalisé avec Clustal Oméga.

La seule structure atomique contenant un peptide de l'ectodomaine des syndécans est celle obtenue avec l'octapeptide <sup>202</sup>AAEGSGEQ<sup>209</sup> qui est présent dans l'ectodomaine du syndécan-1 et qui a été cocrystallisé avec la xylosyltransférase-1 (*Briggs & Hohenester, 2018*). Le peptide se situe à l'extrémité C-terminale de l'ectodomaine et comprend la séquence consensus sérine-glycine nécessaire à l'attachement d'une chaîne de chondroïtine sulfate (**Figure 19**). Le peptide n'adopte pas de structure secondaire lorsqu'il est en complexe avec la xylosyltransférase-1, qui est une protéine transmembranaire présente à la surface de l'appareil de Golgi qui catalyse le transfert d'un résidu de xylose de l'uridine phosphate (UDP)- $\alpha$ -D-xylose sur un résidu de sérine. Le résidu de xylose fait partie du tétrasaccharide nécessaire à l'élongation des chaînes de glycosaminoglycanes (*Sasarmann et al., 2015, Briggs & Hohenester, 2018*).

En conclusion, le domaine cytosolique du syndécan-4 est le plus caractérisé sur le plan structural. Il dimérise en formant un court feuillet  $\beta$  (**Tableau I**). Seule la structure tridimensionnelle du domaine transmembranaire du syndécan-2 a été résolue (*Li et al., 2019*) et le domaine extracellulaire reste en revanche très peu étudié. Cette absence de données structurales concernant l'ectodomaine des syndécans est la conséquence de leur enrichissement de désordre intrinsèque qui a précédemment été prédit par *Leonova & Galzitskaya (2015)*.



**Figure 19: Structure de l'octapeptide de syndécan-1 cocrystallisé avec la xylosyltransférase-1.** **A)** Localisation et séquence de l'octapeptide (vert) sur l'ectodomaine du syndécan-1 (bleu). La localisation des chaînes d'héparane sulfate (HS) et de chondroïtine sulfate (CS) est indiquée par des traits verticaux. **B)** Structure cristallographique (PDB : 6EJE) de la xylosyltransférase-1 (gris) en complexe avec l'octapeptide de syndécan-1 (vert). Le groupement phosphate est en rouge et en représentation sphère dans l'agrandissement, et le résidu de sérine est en rouge.

Les expériences structurales décrites dans la littérature impliquant au moins un peptide de syndécan sont résumées dans le **Tableau I**.

Syndécans (SDCs)			Partenaires	Méthode & Code PDB
Séquence du peptide utilisé dans les expériences	Séquence du peptide observée dans le fichier PDB	Domaine		
<b>Syndécan-1 (P18827)</b>				
<sup>303</sup> TKQEEFYA <sup>310</sup> (Tyr <sup>309</sup> : phosphorylé)	<i>identique</i>	CD (région C2)	TIAM1 (Q13009)	RX 4GVC
<sup>303</sup> TKQEEFYA <sup>310</sup>	<i>identique</i>	CD (région C2)	TIAM1 (Q13009)	RX 4GVD
<sup>201</sup> PAAEGSGEQDFT <sup>212</sup>	<sup>202</sup> AAEGSGEQ <sup>210</sup>	ED	XYLT1 (Q86Y38)	RX 6EJE
<b>Syndécan-2 (P34741)</b>				
<b>GSRMRKKDEGSYDLGERKPSSAAYQK APTKEFYA</b> (Les résidus en gras sont ceux présents dans le CD du SDC-2, 170-201)	-	CD (entier)	-	RMN (Chae & Lee, 2008)
<b>MEQSTEVLAAVIAGGVIGFLFAIFLILL VYRMRKHHHHHH</b> (Les résidus en gras sont ceux présents dans le TMD du SDC-2, 143-173)	MEQSTEVLAAVIAGGVI GFLFAIFLILLLVYRMRK	TMD	-	RMN 6ITH
<b>Syndécan-3 (O75056) : Pas de structure disponible</b>				

**Tableau I (1/2) : Etudes structurales des syndécans.** CD : domaine cytosolique. TMD : domaine transmembranaire. ED : ectodomaine. Les codes d'accès UniProt KB sont indiqués. RX : cristallographie aux rayons X. RMN : résonance magnétique nucléaire. Non déterminé (-).

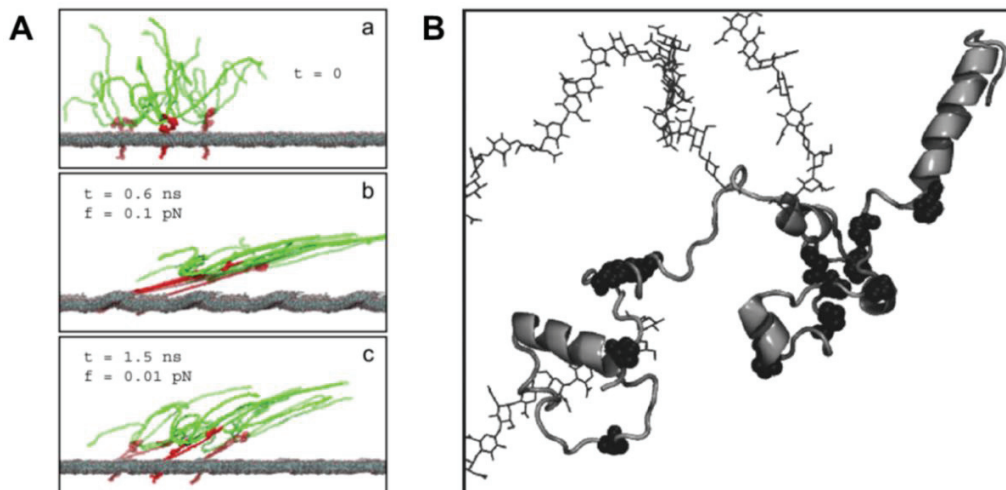
Syndécans (SDCs)			Partenaires	Méthode & Code PDB ou référence
Séquence du peptide utilisé dans les expériences	Séquence du peptide observée dans le fichier PDB	Domaine		
<b>Syndécan-4 (P31431)</b>				
<sup>176</sup> RMKKKDEGSYDLGKKPIYKKAPTNEFY A <sup>202</sup> (rat) (Identiques aux résidus 171-198 du SDC-4 humain)	<i>identique</i>	CD (entier)	-	RMN 1EJP
<sup>170</sup> RMKKKDEGSYDLGKKPIYKKAPTNEFY A <sup>197</sup> de poule et de rat SDC4 (résidus 171-198 du SDC-4 humain)	<i>identique</i>	CD (entier)	PIP <sub>2</sub> <i>Non visible dans le fichier PDB</i>	RMN 1EJQ
<sup>193</sup> TNEFYA <sup>198</sup>	<sup>193</sup> TNEFYA <sup>198</sup>	CD (région C2)	SDCBP (O00560)	RX 1OBY
<sup>193</sup> TNEFYA <sup>198</sup>	<sup>193</sup> TNEFYA <sup>198</sup>	CD (région C2)	SDCBP (O00560)	RX 1YBO
<sup>195</sup> APTNEFYA <sup>202</sup> (rat) (Identiques aux résidus 191-198 du SDC-4 humain)	<i>identique</i>	CD (région C2)	SDCBP (O00560)	RX 5G1D
<sup>195</sup> APTNEFYA <sup>202</sup> (rat)	<i>identique</i>	CD (région C2)	SDCBP (O00560)	RX 5A2P
SDC-4 du poisson zèbre <sup>174</sup> RIKKKDEGSYDLGKTPIYKKAPTTEIYA <sup>201</sup>	<i>identique</i>	CD (entier)	-	RMN ( <i>Whiteford et al., 2008</i> )
<b>MVLAALIVGGVVGILFAVFLILLVYGG</b> KKKKM (Les résidus en gras sont ceux présents dans le TMD du SDC-4 humain, 146-170)	-	TMD	-	RMN ( <i>Park et al., 2011</i> )
SDC-4 de rat (P34901) (Séquence inconnue)	-	CD	Syndesmos (Q9BRJ7)	RMN ( <i>Kim et al., 2015</i> )

**Tableau I (2/2) : Etudes structurales des syndécans.** CD : domaine cytosolique. TMD : domaine transmembranaire. ED : ectodomaine. Les codes d'accès UniProt sont indiqués. RX : cristallographie aux rayons X. RMN : résonance magnétique nucléaire. Non déterminé (-).



#### d) Modélisation des syndécans entiers

Un modèle tridimensionnel du syndécan-4 entier portant des chaînes d'héparane sulfate a été réalisé afin d'étudier le glycocalyx des cellules endothéliales par dynamique moléculaire (Cruz-Chu et al., 2014). La structure du domaine cytosolique du syndécan-4 sous forme dimérique (PDB : 1EJP) a été solvatée et ionisée. Son domaine transmembranaire a été modélisé en utilisant la structure de celui de la glycopherine A qui contient également un motif GXXXG, puis inséré dans une bicouche lipidique composée de phosphatidylcholine (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC). Finalement, l'ectodomaine du syndécan-4 a été modélisé en partant d'une structure toute étendue qui a été soumise au protocole de recuit simulé, c'est-à-dire que le système a été chauffé succinctement puis lentement refroidi afin de trouver un minimum énergétique. Trois chaînes d'héparane sulfate contenant 100 résidus osidiques et différents types de sulfatation (domaines NA et NS) ont été covalamment attachées à l'ectodomaine *via* un pont tétrasaccharidique. L'ectodomaine et les domaines cytosolique et transmembranaire ont été reliés puis solvatés et ionisés dans du NaCl 100 mM (**Figure 20A**). Dans ces expériences, l'ectodomaine du syndécan-4 se replie partiellement en hélices  $\alpha$  et certaines régions restent non structurées, notamment celles portant les chaînes d'héparane sulfate (**Figure 20B**).

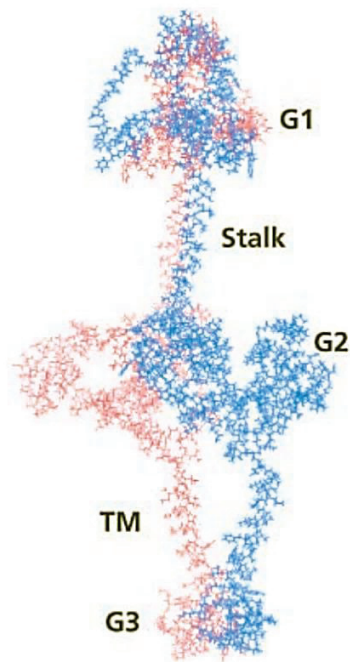


**Figure 20 : Modélisation du syndécan-4 entier.** **A)** Le syndécan-4 entier a été placé dans une bicouche lipidique (gris) et soumis à différentes forces de cisaillement (0, 30,42 et 3,63 MPa). La partie protéique est en rouge et les chaînes d'héparane sulfate sont en vert. **B)** Agrandissement de l'ectodomaine du syndécan-4 montrant son repliement partiel en hélice  $\alpha$  pendant la simulation. Les résidus de proline sont représentés par des sphères noires. Figures adaptées de Cruz-Chu et al., (2014).

Ces expériences ont également montré que des forces de cisaillement élevées sur l'ectodomaine du syndécan-4 influençaient la structure de son domaine cytosolique par un mécanisme de mécanotransduction. Ce mécanisme nécessiterait un mouvement en "ciseau" (scissor-like motion) de la partie protéique du syndécan-4 sous forme dimérique, notamment de son domaine transmembranaire (*Jiang et al., 2019*). L'application de forces de cisaillement engendre ainsi une perturbation du cytosquelette avec lequel son domaine cytosolique interagit. Les chaînes d'héparane sulfate jouent le rôle de "voiles", captant les forces de cisaillement et les transmettent à la partie protéique du syndécan-4. Un moyen pour la cellule de limiter de trop fortes modifications conformationnelles du syndécan-4 serait alors de libérer les chaînes de glycosaminoglycanes qu'il porte. En supprimant moins de 17 % des chaînes, les auteurs ont observé une diminution de 55 % du mouvement en ciseau de la partie protéique du SDC-4 (*Jiang et al., 2019*).

Le syndécan-3 de poulet entier et sans chaînes de glycosaminoglycanes a également été modélisé (*Kirsch et al., 2002*). Les proportions et la localisation des structures secondaires ont été prédites ce qui a permis de définir l'existence de 5 régions : N-terminale, tige (stalk), périmembranaire, transmembranaire et cytosolique. La structure tridimensionnelle de chaque région a été construite avec le programme Insight II (Accelrys) en imposant à chaque domaine la structure secondaire prédite. Les structures ont été minimisées puis assemblées afin de former une unique structure comprenant la totalité de la séquence du syndécan-3 (**Figure 21**).

Trois domaines globulaires ont été observés à l'extrémité N-terminale de la protéine, au niveau juxtamembranaire ainsi que dans le domaine cytosolique. Le domaine globulaire N-terminal et la région juxtamembranaire sont séparés par une région ayant une conformation étendue (stalk) se repliant en brins  $\beta$  et enrichie en résidus de thréonine et de sérine. Les auteurs proposent que l'interaction entre les deux monomères mène à la formation d'un feuillet  $\beta$  parallèle. Deux monomères ont ainsi été disposés de façon arbitraire et l'énergie du système a été minimisée (**Figure 21**). Ce modèle a été réalisé en l'absence de bicouche lipidique ce qui peut être une cause de l'absence d'interaction entre les deux monomères au niveau de leur domaine transmembranaire.



**Figure 21 : Modélisation de la structure tridimensionnelle du syndécan-3 de poulet.** Deux monomères de SDC-3 ont été modélisés (chaînes en bleu et en rouge). G1-3 : domaines globulaires, stalk : région étendue repliée en brins  $\beta$ , TMD : domaine transmembranaire. D'après *Kirsch et al., (2002)*.

### II.5.2 Fonctions des syndécans

Les fonctions des syndécans sont notamment régulées par leurs interactions avec leurs très nombreux partenaires. Au moins 347 partenaires des syndécans ont été identifiés et la majorité se fixe sur leurs chaînes d'héparane sulfate. Très peu d'entre eux interagissent directement sur la partie protéique de leur ectodomaine (*Gondelaud & Ricard-Blum, 2019*). Les syndécans agissent comme récepteurs ou corécepteurs à la surface cellulaire et sont impliqués dans les interactions cellule – matrice régulant ainsi des processus comme l'adhésion, la migration, la différenciation et la prolifération cellulaire (*Couchman et al, 2015, Gopal et al., 2017, Mitsou et al., 2017*). Ils sont également impliqués dans le processus de cicatrisation (*Fears & Woods, 2006, Vuong et al., 2014*), ou de régénération et du remodelage tissulaire (*Chung et al., 2016*). Ils recrutent des facteurs de croissance comme le facteur de croissance fibroblastique 2 (FGF-2) qu'ils présentent à leurs récepteurs (*Kwon et al., 2012*). Ils jouent également un rôle dans le métabolisme lipidique (*Leonova & Galzitskaya, 2015*) et l'angiogenèse (*De Rossi & Whiteford, 2014*), et participent à plusieurs processus physiopathologiques comme la tumorigenèse (*Afratis et al., 2017*). Leur localisation à la surface cellulaire leur permet également d'interagir avec des pathogènes comme des virus (*Cagno et al., 2019, Ozbun, 2019*), des bactéries (*Bartlett & Park, 2015*) et certains parasites (*Fatoux-Ardore et al., 2014, Maciej-Hulme et al., 2018*).

Ils y jouent également le rôle de mécanorécepteurs comme cela a été décrit pour le syndécan-4 à la surface des fibroblastes cardiaques (*Samarel et al., 2013*) et pour les syndécans-1 et -4 dans le glycocalyx endothélial (*Voyvodic et al., 2014, Jiang et al., 2020*). La base de données Pubmed Central (PMC) regroupe à ce jour plus de 12 000 entrées portant sur les syndécans montrant ainsi l'intérêt de ces protéoglycanes ainsi que l'étendue de leurs fonctions (interrogation réalisée avec le terme "syndecan"). De ce fait, les fonctions majeures des syndécans seront décrites dans ce chapitre.

Les chaînes de glycosaminoglycanes que les syndécans portent leurs permettent d'interagir avec des macromolécules extracellulaires comme les collagènes I et V (*Koda et al., 1985*), la fibronectine (*Woods et al., 2000*), les laminines 111 et 332 (*Herndon et al., 1999, Carulli et al., 2012*), la ténascine-C (*Salmivirta et al., 1991*) ou encore la thrombospondine-1 (*Adams & Taylor, 2001*). Leur domaine cytosolique interagit avec le cytosquelette notamment par le biais des protéines à domaine PDZ comme la synténine, la synectine, la synbindine, CASK ou encore TIAM-1 (*Cheng et al., 2016*). Ils interagissent également avec des protéines associées au cytosquelette comme la cortactine, l' $\alpha$  et la  $\beta$  tubuline (*Kinnunen et al., 1998*), la protéine associée au cytosquelette 5 (*Huttlin et al., 2015*), les protéines de la famille ERM (*Granés et al., 2003*), l' $\alpha$  actinine (*Greene et al., 2003*) ou encore la paxilline (*Denhez et al., 2002*).

Certaines interactions sont médiées par la région variable V des syndécans et sont ainsi spécifiques à chacun d'entre eux. Par exemple, la région V du syndécan-1 interagit avec la fascine (*Chakravarti et al., 2005, Adams et al., 2001*) alors que cette région permet l'interaction du syndécan-4 avec le phosphatidylinositol-4,5-bisphosphate et la protéine kinase C  $\alpha$ . Ce complexe ternaire est nécessaire à la formation des adhésions focales à travers l'activation des petites GTPases Rac1 et Rho (*Lambaerts et al., 2009*). Ce processus d'adhésion cellulaire se fait en concert avec certaines intégrines (*Morgan et al., 2007*). Le syndécan-4 interagit avec les intégrines  $\alpha 5\beta 1$  et  $\alpha v\beta 3$  lors de la formation des adhésions focales alors que le syndécan-1 interagit avec les intégrines  $\alpha v\beta 3$  ou  $\beta 5$  dans les contacts focaux ou avec l'intégrine  $\alpha 2\beta 1$  et le collagène de type I (*Roper et al., 2012*). Les quatre syndécans interagissent avec l'intégrine  $\alpha 6\beta 4$  grâce à leur séquence C-terminale EFYA présente dans leur domaine cytosolique.

Les chaînes de GAGs, et notamment celles d'héparane sulfate, permettent aux syndécans de jouer le rôle de réservoirs de facteurs de croissance à la surface cellulaire. Ils interagissent notamment avec les facteurs de croissance fibroblastique 1 et 2 (FGF-1, -2), des chimiokines (CCL-2, -5, -7, -11, -17), des interleukines (IL-8, IL-34), la pléiotrophine ou encore avec le TGF- $\beta$  (revue par *Gondelaud & Ricard-Blum, 2019*).

Ils fonctionnent en collaboration avec les récepteurs des facteurs de croissance auxquels ils présentent les facteurs de croissance ce qui module la croissance, la mobilité et la différenciation cellulaire (*Afratis et al., 2017*).

La présence des chaînes d'héparane sulfate est requise, puisqu'en leur absence, le FGF-1 perd sa capacité à activer son récepteur FGFR-1. Les chaînes d'héparane sulfate, notamment 2-O sulfatées, forment, avec le FGF-1 et son récepteur un complexe ternaire nécessaire à son activation (*Wu et al., 2001*). La fonction de corécepteur pour les facteurs de croissance positionne également les syndécans dans les molécules pro-oncogéniques. Ils agissent par exemple en tant que corécepteurs du facteur de croissance hépatique (HGF) favorisant la survie et la prolifération des cellules de myélome (*Derksen et al., 2002*). Ils sont également impliqués dans le cancer du sein (*Sullivan et al., 2019*), du pancréas (*Yao et al., 2019*), colorectal (*Vicente et al., 2018*), de la prostate (*Hadaschik et al., 2018*) ou encore du poumon (*Tsoyi et al., 2019*).

Les syndécans sont également impliqués dans des processus d'endocytose durant lequel ils sont eux-mêmes internalisés (*Christianson & Berltng, 2014*). Différents mécanismes ont été décrits, soit dépendant des clathrines soit de microdomaines membranaires comme les radeaux lipidiques. Les syndécans interagissent avec plusieurs protéines impliquées dans ces processus comme l' $\alpha$ -tubuline, la cortactine, la cavéoline-2 ou encore la dynamine-2 (*Yoo et al., 2005, Chen & Williams, 2013, Lim et al., 2015*). Les résidus basiques conservés du domaine cytoplasmique des syndécans seraient nécessaires à ce processus (*Chen & Williams, 2013*). Ces interactions permettent soit le recyclage des syndécans, soit l'internalisation de molécules comme le FGF-2 seul ou avec son récepteur FGFR-1 (*Elfenbein et al., 2012*). Les quatre syndécans, et notamment le syndécan-3, sont également responsables de l'internalisation du peptide  $\beta$ -amyloïde, de l' $\alpha$ -synucléine ainsi que des fibrilles formées par la protéine Tau, avec lesquelles ils interagissent *via* leurs chaînes d'héparane sulfate (*Letoha et al., 2019, Hudák et al., 2019*). Ils internalisent également certains virus comme les virus de l'hépatite C et E ou de l'herpès, l'internalisation de ce-dernier ne nécessitant pas la présence des chaînes de GAGs (*Karasneh et al., 2011, Lefèvre et al., 2014*).

Les syndécans subissent le processus de shedding comme environ 1 % des protéines membranaires (*Fitzgerald et al., 2000*) ce qui libère leur ectodomaine dans l'espace extracellulaire. Ils sont clivés de manière constitutive mais ce processus est augmenté lorsqu'ils sont soumis à des *stimuli* extracellulaires comme leurs interactions avec certains facteurs de croissance, des facteurs de virulence bactériens, l'insuline, le stress cellulaire ou dans certains processus comme la cicatrisation et la tumorigenèse (*Manon-Jensen et al., 2010, Manon-Jensen et al., 2013*). L'héparanase, qui dégrade les chaînes d'héparane sulfate, stimule également le shedding des syndécans libérant des ectodomains peu glycosylés.

La dégradation de ces chaînes permet une meilleure accessibilité aux sheddases, c'est-à-dire aux enzymes catalysant le clivage des ectodomains (*Ramani et al., 2012*). Les ectodomains solubles sont impliqués dans le processus d'angiogenèse. L'ectodomaine du syndécan-1 et -4 stimule l'angiogenèse à l'inverse de celui du syndécan-3. Le syndécan-2 stimule l'angiogenèse sous sa forme membranaire mais l'inhibe lorsqu'il est soluble (*Purushothaman et al., 2010, De Rossi & Whiteford, 2013, De Rossi et al., 2014*). Les ectodomains solubles peuvent être de nouveaux internalisés, comme le syndécan-1, mais uniquement s'il porte ses chaînes d'héparane sulfate qui lui servent à cointernaliser le facteur de croissance hépatocytaire. L'ectodomaine du syndécan-1 est transloqué jusqu'au noyau et interagit avec l'histone acétyltransférase dont il diminue l'activité (*Stewart et al., 2015*).

Bien que les chaînes de glycosaminoglycanes qu'ils portent soient généralement essentielles à leurs fonctions, certaines régions de leur partie protéique permettent d'établir directement des interactions. La partie protéique du syndécan-1 interagit avec les intégrines  $\alpha\beta3$  ou  $\alpha\beta5$  ainsi qu'avec le récepteur IGF1-R (insulin-like growth factor-1 receptor) menant à la formation d'un complexe ternaire impliqué dans la survie et la mobilité cellulaire des cellules épithéliales malignes. L'addition d'un peptide mimant la séquence de l'ectodomaine du syndécan-1 impliqué dans ce complexe mène à la déstabilisation du complexe ternaire et à l'inhibition de la croissance tumorale et de l'angiogenèse (*Beauvais & Rapraeger, 2009*). Ce peptide de syndécan-1 a été nommé synstatine. Une seconde synstatine issue de l'ectodomaine du syndécan-1 a été décrite, déstabilisant le complexe ternaire syndécan-1 – intégrine  $\alpha4\beta1$  – VEGFR-2 (Vascular endothelial growth factor receptor 2). Cette synstatine a pour effet de diminuer l'invasion de cellules de myélome (*Jung et al., 2016*).

Une synstatine a également été décrite pour le syndécan-4 qui interagit avec l'intégrine  $\alpha 3 \beta 1$  et l'EGFR (Epidermal growth factor receptor) dont les fonctions sont globalement identiques aux synstatines du syndécan-1 (*Jung et al, 2016*). Les synstatines sont ainsi de potentiels peptides thérapeutiques bloquant la croissance, la survie et l'invasion tumorale.

Les souris KO (knockout) n'exprimant plus les syndécans -1 et -4 sont viables et ne présentent que de faibles complications post-natales bien que les syndécans soient impliqués dans un très grand nombre de processus biologiques (*Couchman et al., 2015*). L'abolition de l'expression du syndécan-1 altère la cicatrisation alors que celle du syndécan-4 perturbe la migration des fibroblastes et l'angiogenèse. L'abolition de l'expression du syndécan-3 perturbe la migration des neurones étant donné sa forte expression dans le système nerveux et aucune souris dans laquelle l'abolition de l'expression du syndécan-2 a été effectuée n'a pour le moment été décrite (*Couchman et al., 2015*). Les fonctions biologiques de l'isoforme-2 du syndécan-4 n'ont en revanche pas été étudiées.

## **II.6 Le désordre intrinsèque des protéines**

### **II.6.1 Généralités**

L'un des principaux paradigmes en biologie structurale admet que la fonction d'une protéine dépend de sa structure tridimensionnelle, et que cette information structurale est inscrite dans sa séquence en acides aminés, ce dernier point étant notamment étayé par les travaux d'Anfinsen (1961). Plusieurs modèles de repliement des protéines ont été proposés et particulièrement le modèle de nucléation – condensation unissant le mécanisme d'effondrement hydrophobe et le modèle charpente (framework), qui suggère une formation des structures secondaires et une coopérativité de repliement de proche en proche (*Dobson, 2003, Daggett & Fersht, 2003*). La découverte du désordre intrinsèque au sein des protéines s'est faite graduellement, d'abord par le fait que la structure tridimensionnelle des protéines globulaires pouvait fluctuer, notamment lors de leurs interactions avec leurs ligands indiquant une certaine flexibilité (ajustement induit), puis par l'observation de mouvement de domaines à grande échelle. Enfin, l'absence de densité électronique de certaines régions de protéines analysées par cristallographie aux rayons X indiquait la présence de régions n'adoptant pas une unique structure (*Dunker et al., 2001*).

Le terme de protéine nativement dénaturée a ensuite été introduit par le groupe de Mandelkow (1994) lors de l'étude de la protéine Tau qui apparaissait comme dénaturée lors de son analyse par diffusion des rayons X aux petits angles et par dichroïsme circulaire. Le désordre intrinsèque d'une protéine représente ainsi l'absence de structure secondaire et tertiaire stables dans des conditions physiologiques (*Uversky, 2019*).

Le désordre intrinsèque au sein d'une protéine peut concerner qu'une faible partie de sa séquence et lorsqu'au moins 30 résidus d'acides aminés consécutifs sont désordonnés, on parle de régions protéiques intrinsèquement non structurées (IDPRs) ou de régions intrinsèquement non structurées (IDRs). En revanche, lorsque plus de 50 % des résidus d'acides aminés sont désordonnés, on parle de protéines totalement désordonnées ou plus généralement de protéines intrinsèquement non structurées (IDPs). Les protéines intrinsèquement non structurées, ou protéines désordonnées, peuvent être définies comme étant des protéines fonctionnelles ne possédant pas de structures secondaires ou tridimensionnelles stables et sont susceptibles d'adopter un ensemble de conformations (*Uversky, 2019*).

La fréquence du désordre intrinsèque fluctue selon les protéomes (*Xue et al., 2012*). La quantité de résidus désordonnés n'est pas dépendante de la taille des protéomes mais l'abondance des protéines possédant de longues régions désordonnées (IDPRs) ou des protéines désordonnées (IDPs) augmente avec la complexité de l'organisme. La fraction de résidus désordonnés varie de 35 à 50 % dans les organismes eucaryotes, de 12 à 38 % chez les archées et de 18 à 35 % chez les bactéries. La proportion de désordre est plus variable chez les virus, entre 7 et 77 % (*Xue et al., 2012*). Chez les organismes supérieurs, près d'une protéine sur deux contient une proportion significative de désordre (*van der Lee et al., 2014b*).

Plusieurs classifications des protéines désordonnées ont plus récemment été proposées. Elles sont basées sur leurs fonctions, la présence d'éléments fonctionnels (motifs, domaines), leurs conformations, leurs interactions, leur évolution, leurs régulations ou encore sur certaines propriétés comme leur solubilité ou leur capacité à subir des transitions de phase (*van der Lee et al., 2014b*).

Plusieurs bases de données ont été créées afin de recenser les IDRs et les IDPs de tous les protéomes comme DisProt ([www.disprot.org](http://www.disprot.org), *Piovesan et al., 2017*) qui regroupe des IDPs et des annotations manuellement curées de la littérature ou IDEAL (<http://idp1.force.cs.is.nagoya-u.ac.jp/IDEAL>, *Fukuchi et al., 2014*), qui contient des interactions des protéines et régions désordonnées et de leurs partenaires cellulaires.



MobiDB (<https://mobidb.bio.unipd.it>, Piovesan et al., 2018), propose en plus des prédictions de désordre ou de structures secondaires, comme D<sup>2</sup>P<sup>2</sup> (<http://d2p2.pro>, Oates et al., 2013). DisBind (<http://biophy.dzu.edu.cn/DisBind>, Yu et al., 2017) propose une classification des IDPs en fonction de leurs partenaires physiologiques et *The Protein Ensemble Database* (<http://pedb.vib.be>, Varadi & Tompa, 2015) porte davantage sur l'aspect méthodologique en regroupant les protocoles et les modèles générés dans la littérature scientifique sur les protéines désordonnées.

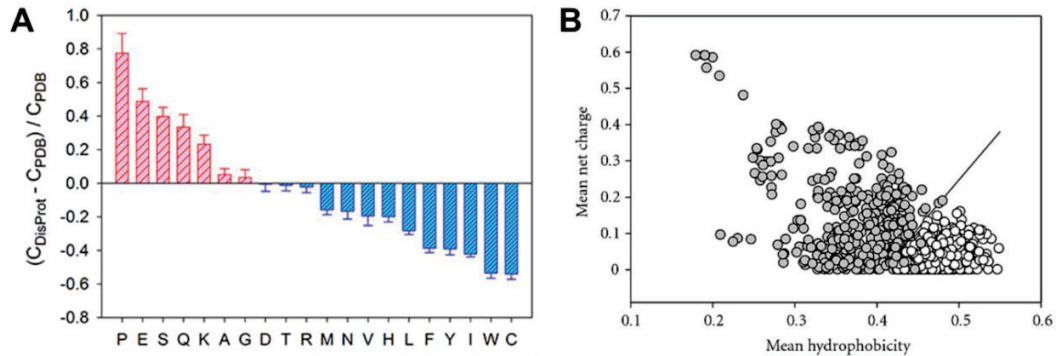
Etant donné l'absence de contraintes structurales, les protéines intrinsèquement non structurées évoluent rapidement. Il est ainsi plus difficile d'identifier des protéines homologues (van der Lee et al., 2014). L'absence d'information structurale et de protéine homologue positionne les IDPs dans le *Dark Proteome* (<http://darkproteome.ws>) dont font partie près de 43 % des protéines eucaryotes (Perdigão & Rosa, 2019). Il est possible d'estimer le potentiel d'obscurité (darkness potential) d'une protéine. Ce potentiel, variant de 0 à 1 où 1 est le plus sombre, est défini à partir des annotations disponibles dans la base de données UniProt KB, et n'est pas uniquement basé sur les informations de structure. Il prend en compte la totalité des annotations comme la présence de domaines, leurs modifications post-traductionnelles ou encore l'existence d'isoformes. Le potentiel d'obscurité est alors le reflet d'une part de la disponibilité des informations structurales, et d'autre part de l'ampleur des études réalisées sur une protéine donnée. Plus une protéine est caractérisée, pour son potentiel tend vers zéro.

## II.6.2 Caractéristiques du désordre intrinsèque

### a) Composition en acides aminés des séquences désordonnées

Les protéines intrinsèquement non structurées possèdent une faible complexité de séquence marquée par l'abondance de certains résidus d'acides aminés. Leur séquence est enrichie en résidus chargés et de même charge, qui tendent à se repousser et à étendre la chaîne polypeptidique. L'analyse de 91 protéines désordonnées a montré que la distribution de leur point isoélectrique était bimodale, centrée sur environ 4 et 10,5, avec une abondance plus élevée de protéines désordonnées acides, notamment grâce leur enrichissement en résidus de glutamate (Uversky et al., 2000). Les IDPs sont appauvries en résidus hydrophobes, en résidus aromatiques ainsi qu'en résidus de cystéine (**Figure 22**). Elles sont à l'inverse enrichies en résidus polaires et chargés, comme les résidus de glutamate et de lysine, en résidus de faible masse moléculaire comme la glycine et l'alanine, et enfin en résidus de proline dont la présence défavorise la formation de structures secondaires (Uversky, 2013, Uversky, 2019).

Les protéines intrinsèquement non structurées possèdent ainsi une hydrophobicité moyenne plus faible que les protéines globulaires et à l'inverse une charge nette plus élevée (**Figure 22B**).

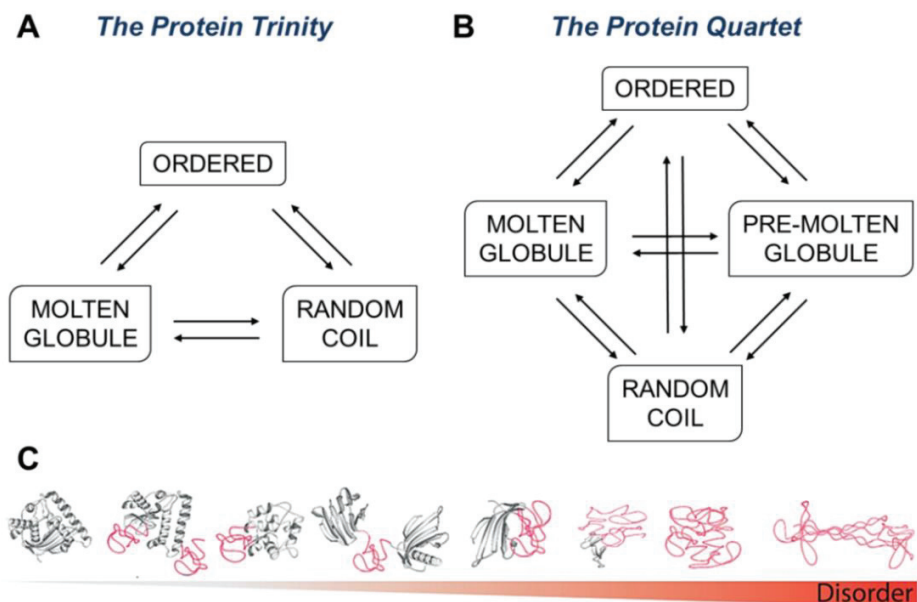


**Figure 22 : Composition en acides aminés des protéines désordonnées. A)** Abondance des 20 acides aminés au sein des protéines désordonnées par rapport aux protéines globulaires. Les valeurs positives et négatives indiquent des résidus enrichis dans les protéines désordonnées par rapport aux protéines globulaires ou un appauvrissement, respectivement. D'après *Uversky* (2019). **B)** Comparaison de la charge nette et de l'hydrophobicité moyenne de 275 protéines globulaires (cercles blancs) et de 91 protéines désordonnées (cercles gris). La droite représente la limite entre les deux catégories de protéines. D'après *Uversky* (2009).

La faible abondance de résidus d'acides aminés hydrophobes et à l'inverse la prépondérance des résidus polaires et chargés favorisent l'interaction des protéines désordonnées avec les solvants aqueux. Les protéines intrinsèquement non structurées possèdent donc une solubilité accrue par rapport aux protéines globulaires. Elles sont également plus stables et moins sujettes à l'agrégation ou à la précipitation (*van der Lee et al., 2014*). Les protéines désordonnées sont très sensibles aux protéases étant donné que les sites de clivages sont fréquemment retrouvés dans des régions flexibles (*Uversky, 2019*). Cette propriété permet de détecter *in vitro* la présence de régions non structurées par protéolyse limitée (*Schramm et al., 2019*). De manière similaire, la présence de régions désordonnées favorise leur dégradation *in vivo* (*van der Lee et al., 2014b*). La demi-vie des protéines eucaryotes possédant un segment désordonné d'un moins 30 résidus d'acides aminés est significativement plus faible que celle des protéines qui en sont dépourvues, quelle que soit la localisation de ce segment au sein de la protéine (*van der Lee et al., 2014b*).

b) Le concept de structure – fonction revisité

Ces caractéristiques de séquence provoquent l'apparition de conformations possédant un faible degré de compaction, la quasi-absence de structures secondaires et une grande flexibilité conformationnelle (Habchi et al., 2014, Uversky, 2019). Afin de décrire le concept de structure – fonction prenant en compte l'état conformationnel des protéines désordonnées, un premier concept incluant les états de random coil, de globule fondu et de l'état natif (ou ordonné) a initialement été proposé et nommé The Protein Trinity (Figure 23A, Dunker et al., 2001). Le concept admet qu'une protéine fonctionnelle est nécessairement dans l'un de ces trois états et que des transitions existent entre chacun de ces états conformationnels.



**Figure 23 : Evolution du concept de structure – fonction.** Le modèle Trinity (A) a été étendu au modèle Quartet (B) puis au continuum de désordre (C). La Figure C est extraite de Habchi et al., (2014).

Il s'est cependant avéré que les protéines désordonnées ne représentaient pas une classe conformationnelle distincte et uniforme, souvent décrite comme des random coil (ou chaînes gaussiennes). Cette observation a notamment été réalisée lors d'expériences de résonance magnétique nucléaire où des structures résiduelles ont été identifiées ainsi que des contacts longues distances suggérant la formation de structures secondaires et tertiaires transitoires (Uversky, 2002).

L'analyse des dimensions hydrodynamiques d'un jeu de protéines a permis de mettre en évidence l'existence de plusieurs classes de désordre différant par leur degré de compaction, à savoir des états de globules fondus, de pré-globules fondus et de random coils (*Uversky, 2002*). Ces états ont ensuite été classés en deux catégories, le désordre dit étendu (*extended disorder*) regroupant les random coils ainsi que les pré-globules fondus, ainsi que le *collapsed disorder* caractérisant l'état de globule fondu (*Uversky, 2019*). Les conformations de pré-globule fondu et de random coil peuvent se distinguer par leur contenu en structures secondaires, plus abondantes dans l'état de pré-globule fondu (*Uversky, 2002*). Le modèle *The Protein Trinity* caractérisant le concept de structure – fonction a ainsi été étendu et le nouveau modèle a été nommé modèle *Quartet* (*Habchi et al., 2014*) qui inclut des transitions conformationnelles entre chaque état (**Figure 23B**). Il est désormais admis que la proportion de désordre intrinsèque au sein des protéines n'est pas strictement réduite à l'un de ces états discrets, mais suit davantage un *continuum* de désordre comprenant une infinité d'états propres à chaque système étudié (*Uversky, 2013, Habchi et al., 2014*) (**Figure 23C**).

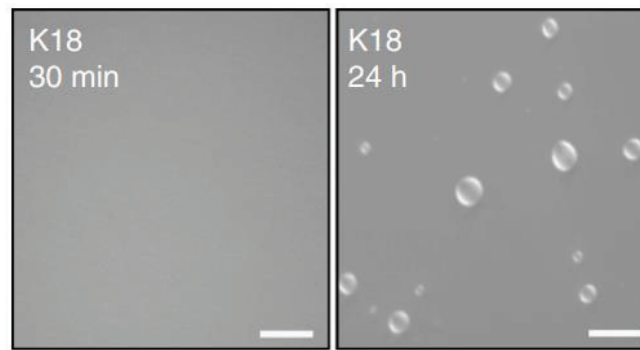
### *c) Modulation de l'espace conformationnel des protéines désordonnées*

La proportion en désordre intrinsèque d'une protéine n'est pas figée et évolue du fait de l'entropie élevée du système et de sa modulation par plusieurs facteurs (*Uversky, 2019*). L'analyse de protéines intrinsèquement non structurées dans différentes conditions physico-chimiques a montré qu'à l'inverse des protéines repliées, les IDPs ont la capacité à se structurer avec l'augmentation de la température et du pH (*Uversky, 2009*). La formation de structures plus compactes avec l'augmentation de la température est la conséquence d'une part de la favorisation des interactions hydrophobes et d'autre part de la diminution de l'énergie de solvation des résidus polaires (*Wuttke et al., 2014*). Cette observation a par exemple été effectuée sur des protéines humaines comme l' $\alpha$ -synucléine et l' $\alpha$ <sub>s</sub>-caséine mais également sur des régions ou des protéines bactériennes, de céréales ou encore virales (*Kumar, 2007, Wuttke et al., 2014, Hamdi et al., 2017, Uversky, 2019*). Des modifications structurales dépendantes du pH ont également été décrites comme pour la prothymosine  $\alpha$  et l' $\alpha$ -synucléine qui se replie partiellement en hélice  $\alpha$  à pH acide. La formation de structures plus compactes est causée par la minimisation de la charge nette présente à pH physiologique diminuant ainsi les répulsions électrostatiques. Un phénomène similaire est observé en présence de contre-ions qui neutralisent alors les nombreuses charges portées par les IDPs (*Uversky, 2009*).

Les protéines intrinsèquement non structurées peuvent également être la cible de modifications post-traductionnelles (PTMs) très diverses (*Darling & Uversky, 2018*). Les modifications post-traductionnelles sont généralement retrouvées dans des courts segments désordonnés au sein des protéines globulaires, du fait de leur accessibilité accrue pour les enzymes catalysant les PTMs. De ce fait, les protéines contenant de multiples sites de modifications post-traductionnelles possèdent un nombre plus important de régions désordonnées (*Huang et al., 2014*). La phosphorylation, l'acétylation, la protéolyse partielle, l'ubiquitinylation, l'acylation et la méthylation sont les modifications post-traductionnelles les plus fréquentes des IDPs (*Habchi et al., 2014*). Elles induisent une modification des états conformationnels adoptés par une protéine ou une région désordonnée. Une même modification post-traductionnelle n'aura cependant pas les mêmes conséquences structurales selon les protéines, pouvant dans certains cas provoquer une transition désordre – ordre ou au contraire déstabiliser des structures secondaires locales préformées. L'interaction des protéines sous leur forme apoprotéique avec leur groupement prosthétique peut également provoquer leur repliement, comme l'interaction du cytochrome C avec son groupement hémique (*Uversky, 2009*). L'ajout d'un groupement fonctionnel sur des régions désordonnées peut moduler leurs interactions avec leurs partenaires, ces modifications pouvant donc jouer le rôle d'interrupteur biologique. Par exemple, l'acétylation et la désacétylation des résidus de lysines de la queue désordonnée des histones régulent l'activité transcriptionnelle (*Bah & Forman-Kay, 2016, Darling & Uversky, 2018*).

#### *d) Transition de phase liquide – liquide*

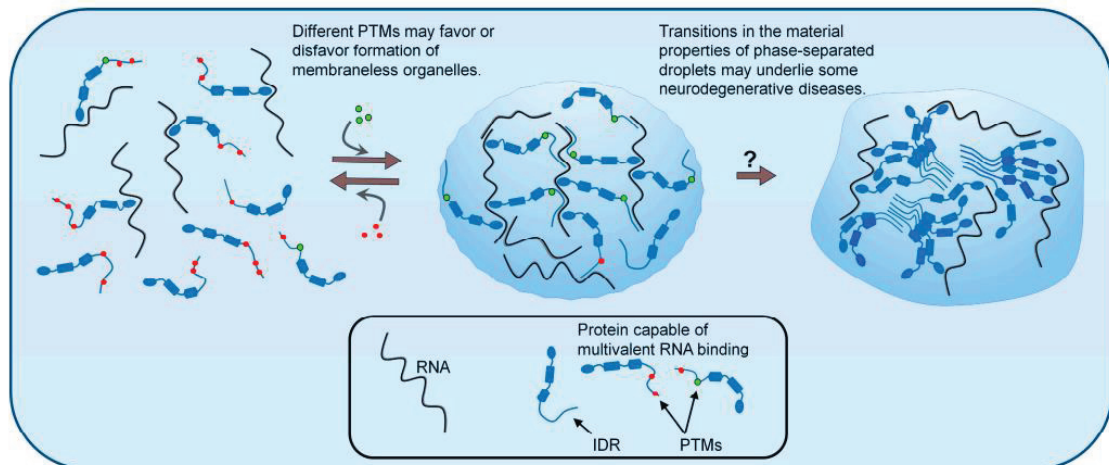
De nombreuses protéines impliquées dans des mécanismes de transition de phase liquide – liquide sont intrinsèquement non structurées ou possèdent des régions désordonnées (*Lin et al., 2017, Uversky, 2017, Owen & Shewmaker, 2019*). Ces mécanismes mènent à la formation d'organelles protéiques sans membranes (PMLOs, proteinaceous membrane-less organelles) (**Figure 24**). Bien que les fonctions et la localisation des PMLOs divergent, ils sont tous enrichis en acides nucléiques et en protéines intrinsèquement non structurées qui permettent leurs assemblages (*Uversky, 2017*). Les PMLOs se forment par un mécanisme de coacervation, ou de transition de phase liquide – liquide, c'est-à-dire un regroupement local d'une grande quantité de molécules, dont la concentration atteint 10 à 300 fois celle de la phase non coacervée, tout en gardant leur état liquide. Ces gouttelettes ont la capacité de fusionner, se déformer et d'échanger des molécules avec la phase non coacervée (*Owen & Shewmaker, 2019*).



**Figure 24 : Observation de la formation d’organelles protéiques sans membranes (PMLOs) d’un peptide (K18) de la protéine Tau.** Les observations ont été réalisées par microscopie à contraste interférentiel après 30 min et 24 h d’incubation à 37°C. Barre : 10 µm. D’après *Ambadipudi et al., (2017)*.

La formation des PMLOs est régulée par des interactions protéine – protéine ou protéine – acide nucléique et peut être modulée par des facteurs externes comme la température, le pH ou la concentration en sel (*Aguzzi & Altmeyer, 2016*). Les transitions de phase liquide – liquide peuvent être gouvernées par des modifications post-traductionnelles des protéines désordonnées (**Figure 25**). La phosphorylation de la protéine désordonnée Tau favorise par exemple la transition de phase liquide – liquide. L’acétylation ou la méthylation de résidus basiques ont tendance à défavoriser la formation de PMLOs (*Owen & Shewmaker, 2019*). Certaines régions désordonnées de protéines virales sont également responsables de la formation d’inclusions cytoplasmiques dans le cytosol de cellules infectées comme le virus de la rougeole (*Zhou et al., 2019, Guseva et al., 2020*). Les protéines impliquées dans les transitions de phase liquide – liquide sont regroupées dans la base de données PhaSePro (<https://phasepro.elte.hu>, *Mészáros et al., 2020*).

Un mécanisme de transition de phase liquide – liquide a également été observé dans l’espace extracellulaire impliquant des mucines (*van der Lee et al., 2014, Meldrum et al., 2018*). Ce mécanisme mène à la formation de mucus recouvrant et protégeant les cellules épithéliales. D’une manière générale, les protéines désordonnées enrichies en résidus de proline, de thréonine et de sérine et qui sont fortement glycosylées participent à la formation de structures ressemblant à des gels (*van der Lee et al., 2014*).

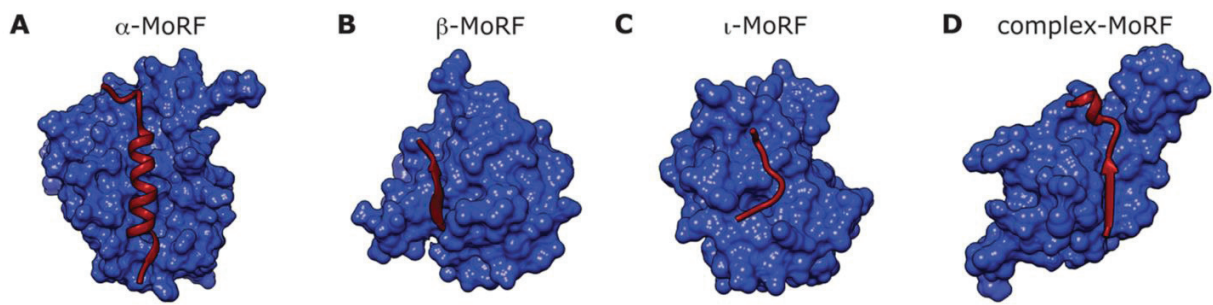


**Figure 25 : Modulation de la transition de phase liquide – liquide par les modifications post-traductionnelles des protéines désordonnées.** Figure extraite d'*Owen & Shewmaker* (2019). IDR : région intrinsèquement non structurée, PTMs : modifications post-traductionnelles.

### II.6.3 Mécanismes d'interaction des protéines désordonnées

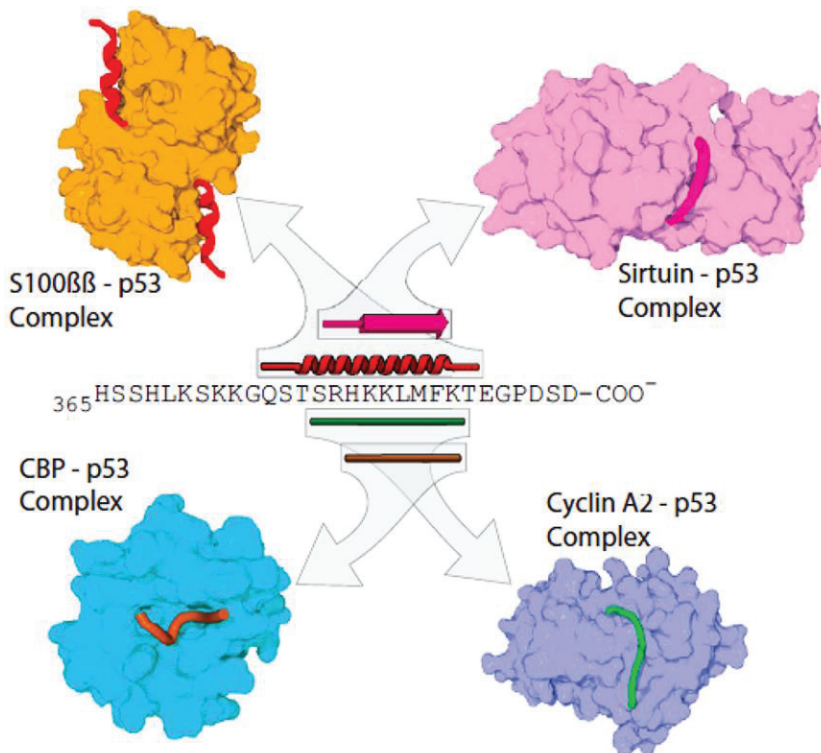
#### a) Caractéristiques de séquences

L'interaction des protéines désordonnées avec leurs partenaires physiologiques peut également moduler leur structure et ainsi diminuer leur hétérogénéité conformationnelle. La plupart des protéines désordonnées ont la capacité à se structurer, du moins localement, lorsqu'elles sont en complexe avec un partenaire (*Uversky, 2009*). Les séquences désordonnées présentes au sein des IDPs ayant la capacité à se structurer lorsqu'elles sont en complexe avec leur partenaire sont appelées MoRE (molecular recognition element) ou MoRF (molecular recognition feature) (**Figure 26**). Elles sont généralement composées de 5 à 25 résidus d'acides aminés alors que les séquences plus longues sont appelées domaine d'interaction désordonné (disordered binding domain) (*Yan et al., 2016*). Ces séquences peuvent se replier en hélice  $\alpha$  ( $\alpha$ -MoRF), en brin  $\beta$  ( $\beta$ -MoRF) ou adopter des structures irrégulières (i-MoRF). Elles sont qualifiées de MoRF complexe (*complex-MoRF*) lorsqu'elles se replient en deux ou plusieurs structures secondaires différentes et d'abondance similaire (*Vacic et al., 2007*). Les  $\beta$ - et i-MoRFs sont enrichis en résidus de proline à l'inverse des  $\alpha$ -MoRFs. Ces derniers sont appauvris en résidus de tyrosine et d'histidine. Les  $\beta$ -MoRFs ne possèdent pas de résidus de tryptophane ni de cystéine (*Vacic et al., 2007*).



**Figure 26 : Exemples d'éléments de reconnaissance moléculaire (MoRFs).** La protéine globulaire est en bleu et les MoRFs en rouge. D'après *van der Lee et al.*, (2014).

Certaines séquences peuvent en outre se replier de façon différente selon la protéine avec laquelle elles interagissent. C'est par exemple le cas de la région C-terminale de la protéine p53 qui est intrinsèquement non structurée (*Laptenko et al.*, 2016). Cette région a la capacité de se replier en hélice  $\alpha$  lorsque la protéine p53 interagit avec la protéine S100 $\beta$ , en brin  $\beta$  lorsqu'elle interagit avec la sirtuine et reste non structurée en complexe avec la cycline A2 ou la *CREB binding protein*, CBP (**Figure 27**). Ces séquences portent le nom de séquences caméléons et sont généralement enrichies en résidus hydrophobes comme la valine et l'isoleucine, ou aromatiques comme la tyrosine et la phénylalanine, ainsi qu'en résidu de glycine (*Habchi et al.*, 2014, *Bahramali et al.*, 2016).



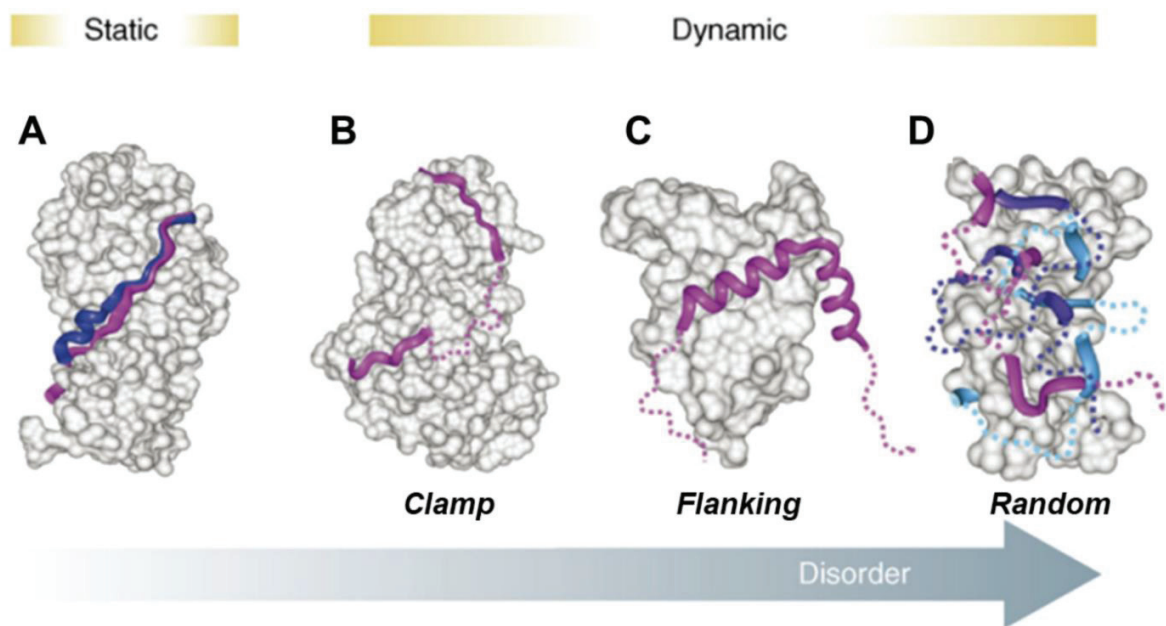
**Figure 27 : La séquence caméléon du domaine C-terminal de p53 se replie différemment en fonction de ses partenaires cellulaires.** D'après *Habchi et al.*, (2014).



### *b) Fuzzy complexes*

Certaines protéines désordonnées restent non structurées lorsqu'elles sont en complexe avec un partenaire et conservent leur flexibilité et leur dynamique. Ce phénomène a été baptisé *fuzziness* et mène à la formation de complexes dynamiques (fuzzy complexes) où l'hétérogénéité conformationnelle de la protéine (ou région) désordonnée est maintenue et est requise pour la fonction (Tompa & Fuxreiter, 2008, Sharma et al., 2015, Olsen et al., 2017). L'interaction entre une région désordonnée ou une protéine intrinsèquement non structurée et un partenaire globulaire peut être décrite selon quatre mécanismes qui diffèrent par l'entropie du système (**Figure 28**). Le mode *statique* admet que la région désordonnée interagissant avec son partenaire est rigide, mais pas nécessairement repliée. Plusieurs conformations peuvent cependant exister, on parle de modèle polymorphique (**Figure 28A**). Les trois autres mécanismes mettent en jeu une région désordonnée (se repliant ou non) gardant une part plus ou moins importante de flexibilité et de dynamique (**Figures 28B – 28D**). Le modèle *clamp* implique la présence d'une région restant désordonnée et flexible séparant deux régions repliées (**Figure 28B**). À l'inverse, le modèle *flanking* nécessite le repliement partiel de la région en interaction avec le partenaire mais également de régions non structurées la flanquant. Ces régions flexibles interagissent également avec le partenaire protéique et leur élimination mène à une diminution de l'affinité entre les deux partenaires (**Figure 28C**). Dans le modèle *random*, la totalité de la région non structurée interagissant avec son partenaire reste désordonnée et flexible (**Figure 28D**) (Tompa & Fuxreiter, 2008, Sharma et al., 2015).

Dans des cas extrêmes, deux protéines ou deux régions désordonnées peuvent interagir sans subir de transition désordre – ordre. Ce type d'interaction entre deux IDPs sans formation de structures secondaires locales a été nommé *extreme fuzziness* (Wang & Wang, 2019). Ce phénomène a été observé pour deux complexes, entre le domaine C-terminal de la protéine 4.1G et la protéine NuMA (nuclear mitotic apparatus) ainsi qu'entre la prothymosine  $\alpha$  et l'histone de liaison H1. Ces deux derniers partenaires présentent une affinité de l'ordre du picomolaire. Ils interagissent par des liaisons électrostatiques du fait de la forte basicité de l'histone H1 et à l'inverse de l'enrichissement en résidus acides de la prothymosine  $\alpha$  (Borgia et al., 2018).



**Figure 28 : Mécanismes de la formation des fuzzy complexes.** **A)** Modèle du désordre statique. Le partenaire protéique est en gris et les deux conformations de la protéine désordonnée sont en bleu et violet. **B-D)** Modèles du désordre dynamique. Une partie (B-C) ou la totalité (D) de la protéine désordonnée reste non structurée et flexible (pointillés). Les différentes couleurs en D réfèrent à la région désordonnée pouvant former de façon très transitoire des interactions stables avec différents sites sur le partenaire protéique, la majorité de la chaîne restant désordonnée et flexible (en pointillés). Figure extraite de Tompa & Fuxreiter (2008).

Plusieurs sites d'interaction différents peuvent exister à la fois sur la protéine désordonnée et également sur le partenaire protéique, ce qui a donné naissance au phénomène d'*allovalence* (*allovalency*). Ce terme est issu du préfixe *allo*, du grec autres / différents, et de *valency* dont l'origine latine signifie pouvoir, ce qui reflète un phénomène de capture du partenaire protéique par la protéine désordonnée (Klein et al., 2003, Levchenko, 2003, Locasale, 2008). La présence de plusieurs sites en compétition augmente l'affinité entre les deux protéines étant donné que le partenaire globulaire est, de façon générale, rapidement recapturé après s'être dissocié (Olsen et al, 2017).

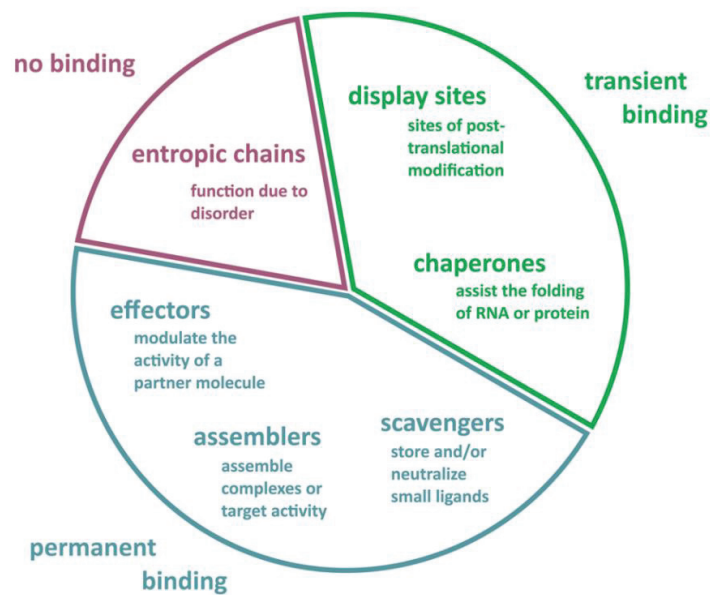
#### II.6.4 Fonctions du désordre intrinsèque

La formation de fuzzy complexes présente plusieurs avantages fonctionnels comme l'interaction d'une même séquence désordonnée avec plusieurs partenaires, parfois simultanément, ce qui permet notamment l'assemblage d'édifices supramoléculaires (*Habchi et al., 2014, Fuxreiter, 2018*). Ce mécanisme permet ainsi de limiter les contraintes de spécificité de séquence. C'est par exemple le cas d'une région riche en proline de la myelin basic protein (MBP) qui est capable d'interagir avec huit protéines différentes possédant un domaine SH3 (*Fuxreiter & Tompa, 2012*). La formation de fuzzy complexes est également retrouvée dans la régulation du cycle cellulaire ou dans l'assemblage de particules virales (*Fuxreiter & Tompa, 2012*). Les fuzzy complexes décrits dans la littérature sont rassemblés dans la base de données FuzDB (<http://protodyn-database.org>, *Miskei et al., 2017*).

Comme évoqué précédemment, la proportion de désordre entre les différents protéomes des trois domaines de la vie et chez les virus croît avec la complexité de l'organisme (*Xue et al., 2012*). Une plus grande quantité de désordre dans les protéines eucaryotes serait la conséquence de plus longs segments reliant les différents domaines d'une protéine par rapport aux protéines procaryotes où les segments de liaison sont plus petits. Les segments eucaryotes sont ainsi plus longs mais aussi en moyenne deux fois plus désordonnés (*Basile et al., 2019*). Les protéines eucaryotes possèdent également davantage de résidus de sérine et de proline et sont en revanche moins riches en résidus d'isoleucine. Du fait de leur complexité, les cellules eucaryotes nécessitent des mécanismes de régulation plus nombreux et plus contrôlés et la synthèse de protéines désordonnées serait ainsi un moyen de répondre à ces nécessités (*Vacic et al., 2007, Schlessinger et al., 2012, Basile et al., 2019*). L'enrichissement en résidus de sérine des protéines désordonnées eucaryotes pourrait être la conséquence d'une régulation plus étroite impliquant des kinases (*Basile et al., 2019*). Une proportion élevée de désordre est cependant observée chez les virus (*Xue et al., 2012*). Cette abondance leur permet notamment d'interagir avec de nombreux partenaires de la cellule hôte afin de les détourner à leur avantage, notamment par la formation de fuzzy complexes (*Communie et al., 2013, Habchi et al., 2014*). Il s'agirait également d'un moyen de limiter les conséquences structurales des fréquentes mutations survenant dans leur génome. La synthèse de protéines désordonnées offre ainsi au virus un tropisme élevé tout en ayant un génome de petite taille.

De par leur capacité à interagir avec une multitude de partenaires, les protéines intrinsèquement non structurées sont des *hubs* au sein des interactomes eucaryotes (Ekman et al., 2006, Haynes et al., 2006). Les protéines enrichies en désordre sont particulièrement représentées dans les processus de régulation, de différenciation, de transcription, de développement ou encore d'apoptose mais sont sous représentées de manière générale dans le métabolisme des biomolécules (Haynes et al., 2006, Habchi et al., 2014). Les IDPs sont ainsi majoritairement impliquées dans des fonctions de régulation et de signalisation mais ne possèdent que très rarement une activité catalytique pour des raisons de contraintes structurales (Vendruscolo, 2010).

Six fonctions générales des protéines et régions intrinsèquement non structurées ont été identifiées (Dunker et al., 2002, van der Lee et al., 2014, Tompa et al., 2015) (**Figure 29**). Les chaînes entropiques (entropic chains) n'interagissent pas ou peu avec d'autres molécules et réalisent leurs fonctions uniquement par le fait qu'elles sont désordonnées et flexibles. C'est par exemple le cas des segments protéiques non structurés reliant différents domaines globulaires. Les chaînes entropiques n'ont pas ou peu de partenaires et l'affinité de ces interactions est très faible, de l'ordre du millimolaire (Tompa et al., 2015). La flexibilité de certaines régions désordonnées leur permet d'être la cible de modifications post-traductionnelles comme évoqué précédemment. Ces modifications peuvent être reconnues par leurs partenaires et engagent les IDPs dans des fonctions de régulation et dans des processus de signalisation notamment. Les modifications post-traductionnelles jouent ainsi le rôle d'interrupteurs, ce qui est causé par l'existence des sites modifiés sur les régions désordonnées (display sites). Les *display sites* présentent également une affinité faible envers leurs partenaires. La multitude de partenaires des protéines désordonnées leur permet également de jouer le rôle de chaperones. Les protéines chaperones nécessitent une faible contrainte de séquence afin d'interagir avec le plus grand nombre de protéines. L'interaction des protéines non ou mal repliées avec les IDPs provoquerait la formation de repliements locaux au sein de la chaîne mal, ou non repliée, ce qui favoriserait son repliement. Les rôles de chaperones et d'interrupteurs (display sites) impliquent des interactions transitoires avec leurs partenaires, caractérisées par une faible affinité, de l'ordre du millimolaire au micromolaire (Tompa et al., 2015).



**Figure 29 : Fonctions majeures des protéines intrinsèquement non structurées ou des régions désordonnées.** D'après *van der Lee et al.*, (2014b).

A l'inverse, les rôles d'effecteurs, d'assembleurs et de piègeurs nécessitent des interactions stables (**Figure 29**). Les protéines désordonnées ayant la fonction d'effecteurs interagissent avec leurs partenaires en subissant une transition désordre – ordre locale modifiant par ce biais l'activité de leur partenaire. Les IDPs peuvent avoir une fonction d'assembleurs pour la formation d'édifices supramoléculaires. Cette fonction est la conséquence de la présence de plusieurs sites d'interactions sur une même protéine désordonnée. Elle joue ainsi le rôle de support augmentant localement la concentration des différents partenaires présents dans l'édifice supramoléculaire favorisant ainsi sa formation. Finalement, les IDPs peuvent posséder la fonction de piègeurs (scavengers) en interagissant et en immobilisant certains ligands de petite taille (**Figure 29**). Ces trois fonctions sont associées à des interactions ayant des affinités plus élevées, de l'ordre du micromolaire au picomolaire (*Tompa et al.*, 2015).

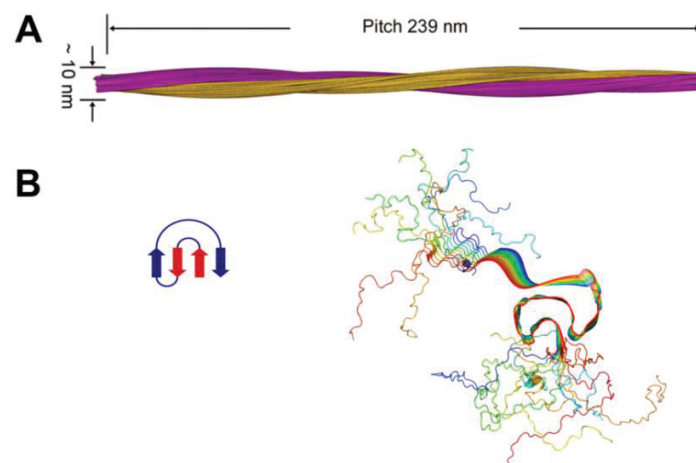
### II.6.5 Association du désordre intrinsèque aux pathologies

Les protéines désordonnées sont fréquemment associées à des pathologies, ce qui a mené au concept D<sup>2</sup> pour *disorder in disorders* (Uversky et al., 2008). Du fait de leur capacité à interagir avec un nombre important de partenaires différents, les protéines intrinsèquement non structurées sont des hubs cellulaires et sont impliquées dans plusieurs fonctions biologiques. Des dysfonctions au sein des protéines hubs mènent ainsi au développement d'un grand nombre de pathologies (Uversky et al., 2008, Midic et al., 2009). La pathogénicité des protéines désordonnées implique dans certains cas la stabilisation de leurs structures secondaires, ou de façon générale de leurs conformations les plus compactes, ce qui est observé dans le processus de fibrillogénèse (Uversky et al., 2008). Les IDPs et les IDRs sont particulièrement associées à certains cancers, la plupart des pathologies neurodégénératives, les diabètes, les maladies cardiovasculaires, les processus amyloïdogéniques et certaines pathologies génétiques (Uversky et al., 2008, Kulkarni & Uversky, 2019).

Dans des conditions physiologiques, les protéines associées aux maladies neurodégénératives sont majoritairement non structurées mais se replient et s'agrègent dans la maladie d'Alzheimer, de Parkinson ou d'Huntington (Martinelli et al., 2019). Ces pathologies sont caractérisées par une perte progressive de neurones accompagnée d'un dépôt d'agrégats protéiques. Outre le rôle du peptide  $\beta$  amyloïde qui est également non structuré sous forme monomérique (Uversky, 2008), la maladie d'Alzheimer est caractérisée par l'agrégation intracellulaire de la protéine Tau, une protéine régulant l'assemblage des microtubules (Naseri et al., 2019). La protéine Tau est intrinsèquement non structurée et est la cible de modifications post-traductionnelles. Elle contient un domaine d'interaction aux microtubules qui est responsable de la formation de fibrilles caractérisées par des feuilletts  $\beta$  croisés (Flitzpatrick et al., 2017). Au sein des agrégats, chacune des six isoformes de la protéine Tau est hyperphosphorylée (Martinelli et al., 2019).

La maladie de Parkinson est une synucléinopathie, c'est-à-dire une pathologie causée par l'agrégation et le dépôt de l' $\alpha$ -synucléine sous forme de corps de Lewy dans le système nerveux central (Martinelli et al., 2019). L' $\alpha$ -synucléine est une protéine intrinsèquement non structurée se repliant partiellement en hélices  $\alpha$  lorsqu'elle interagit avec la membrane et plus particulièrement avec des phospholipides (Meade et al., 2019). Bien que cette protéine ait été extensivement étudiée, son mécanisme d'agrégation n'est pas parfaitement décrit.

L' $\alpha$ -synucléine agrégée est cependant davantage phosphorylée que la forme non pathogénique (*Martinelli et al., 2019*), ce qui pourrait moduler ses états conformationnels (*Uversky et al., 2008*). L' $\alpha$ -synucléine exprimée sous forme recombinante dimérise de façon covalente par des ponts dityrosine, et est également capable de former des tétramères (*Maarschalkerweerd et al., 2015, Meade et al., 2019*). Elle possède une région enrichie en résidus hydrophobes qui est essentielle à son agrégation en fibrilles amyloïdes dans lesquelles elle adopte un motif en clé grecque (*Tuttle et al., 2016, Li et al., 2018*) (**Figure 30**). L' $\alpha$ -synucléine possède la particularité de s'agréger sous différentes formes, appelées polymorphes "tornade" et "bâton", et qui ont la caractéristique commune d'être composé de deux protofilaments s'enroulant (*Li et al., 2018*).



**Figure 30 : Structure d'une fibrille amyloïde formée par l' $\alpha$ -synucléine. A)** Structure d'une fibrille d' $\alpha$ -synucléine obtenue par cryomicroscopie électronique. D'après *Li et al., (2018)*. **B)** Vue en coupe de la fibrille montrant le motif en clé grecque (*Tuttle et al., 2016, PDB : 2N0A*). Le schéma du motif en clé grecque est présenté à gauche.

Les IDPs sont également impliquées dans les processus de tumorigenèse. Environ 80 % protéines associées aux cancers possèdent des régions désordonnées d'au moins 30 résidus (*Uvesky et al., 2008, Midic et al., 2009*). A ce titre, la protéine p53 a été la plus étudiée puisque près de 50 % des cancers humains présentent des mutations de son gène. Elle contient une région non structurée à ses extrémités N- et C-terminales représentant 40 % de sa séquence (*Martinelli et al., 2019*). D'autres protéines enrichies en désordre intrinsèque sont également impliquées dans les cancers comme c-Myc ou BRCA1 (breast cancer 1) (*Uversky et al., 2008, Uversky, 2016*). Les protéines désordonnées sont associées aux maladies cardiovasculaires, squelettiques, osseuses, développementales ou dermatologiques (*Uversky et al., 2008, Uversky et al., 2009, Midic et al., 2009*).

Enfin, plus de 20 % des mutations faux sens provoquant une pathologie ont lieu dans une région intrinsèquement non structurée et 20 % d'entre elles favorisent une transition locale désordre – ordre (*Vacic et al., 2012*). Un exemple concret de la conséquence d'une mutation faux sens d'un résidu d'acide aminé présent dans une région désordonnée a été récemment décrit (*Meyer et al., 2018*). Cette mutation provoque l'apparition d'un motif dileucine dans le domaine cytosolique de trois protéines transmembranaires entraînant leur internalisation ce qui mène à la perte de leurs fonctions (*Meyer et al., 2018*).

Du fait de leurs implications dans un grand nombre de pathologies, les protéines intrinsèquement non structurées semblent être des cibles thérapeutiques de choix. Plusieurs stratégies sont envisagées dans cette optique (*Joshi & Vendruscolo, 2015, Ruan et al., 2018*). La première réside dans la stabilisation des IDPs dans leurs conformations désordonnées ou au contraire, d'induire leur compaction par de petites molécules comme cela a été réalisé pour l' $\alpha$ -synucléine, provoquant l'inhibition de son agrégation en fibrilles amyloïdes (*Chen et al., 2020*). La seconde stratégie vise à l'inhibition de la formation de complexe entre une IDP et son partenaire globulaire, ou entre deux IDPs. Il s'agirait alors d'identifier le ou les sites d'interaction et de les perturber par des petites molécules. Cette stratégie a été envisagée pour le complexe p53 – MDM2 (murine double minute 2) où une région désordonnée de la protéine p53 se replie en hélice  $\alpha$  lorsqu'elle est en complexe avec la protéine MDM2. La formation de ce complexe est inhibée en présence de petites molécules, des cis-imidazolines baptisées Nutlines (dénotées de 1 à 3), qui entrent en compétition avec p53 lors de leur interaction avec MDM2. Cette méthode a également été utilisée avec succès dans l'interaction entre deux régions désordonnées des protéines c-Myc et Max qui sont associées à différents cancers (*Santofimia-Castaño et al., 2020*). Finalement, la dernière stratégie est basée sur une inhibition allostérique de la protéine ou de la région désordonnée. L'objectif est alors de cibler des régions repliées, si elles existent, par des petites molécules dans le but de moduler les conformations adoptées par les régions désordonnées en induisant leur repliement. C'est le cas de la molécule MSI-1436, un aminostérol, qui interagit avec le domaine de régulation de la phosphatase PT1B entraînant sa compaction et son inactivation (*Santofimia-Castaño et al., 2020, Joshi & Vendruscolo, 2015*).



## II.6.6 Prédiction et analyse du désordre intrinsèque

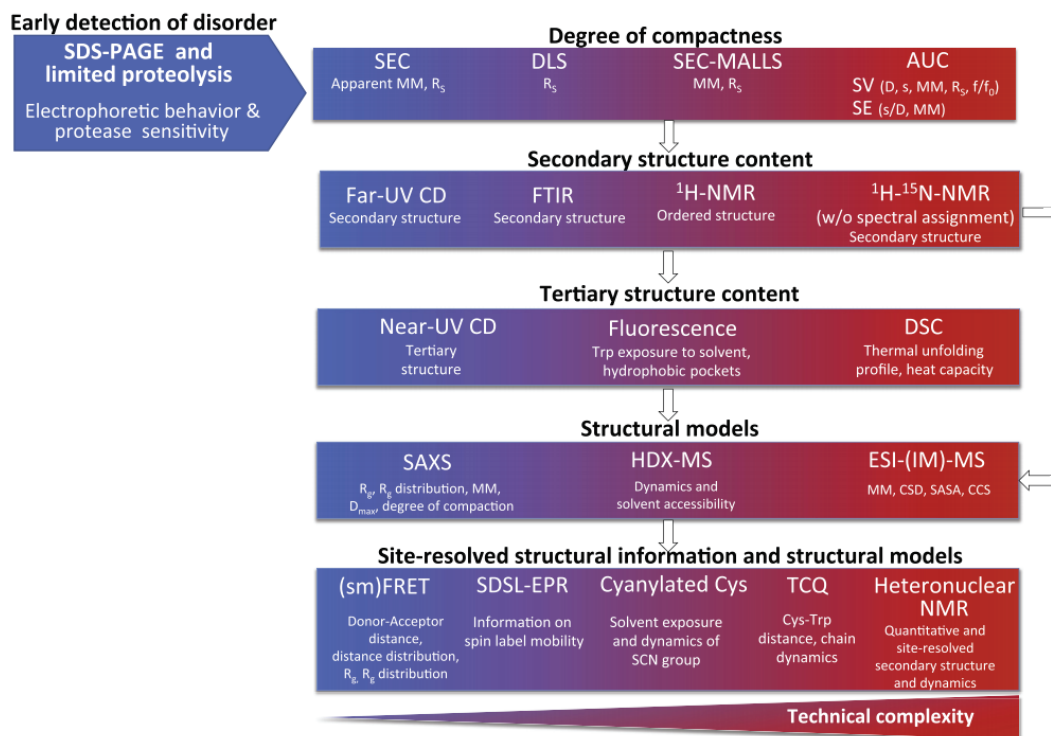
### a) Prédiction du désordre intrinsèque et des MoRFs

Une trentaine de prédicteurs de désordre intrinsèque ont été développés à ce jour (*Lieutaud et al., 2016, Meng et al., 2017*). D'une manière générale, la prédiction du désordre intrinsèque est facilitée par le biais existant dans leur composition en acides aminés. Ils peuvent être classés selon trois catégories : les méthodes basées sur une fonction de score (une formule), l'apprentissage automatique (machine learning) et les méta-prédicteurs. IUPred (*Dosztányi, 2018, Erdős & Dosztányi, 2020*) est l'un des prédicteurs les plus populaires et est basé sur une fonction de score, en l'occurrence un algorithme qui calcule la propension de désordre à partir des propriétés physicochimiques des acides aminés et leur propension à établir des interactions par paires. IUPred sert également de base à la génération de séquences d'IDPs artificielles contenant une proportion de désordre déterminée par l'utilisateur dans l'outil InSiDDE (*In Silico Disorder Design*) (*Schramm et al., 2017*). Des algorithmes d'apprentissage automatique ont également été développés comme DISOPRED (*Jones & Cozzetto, 2015*). Ces prédicteurs utilisent notamment des propriétés de conservation évolutive des séquences peptidiques, leur accessibilité au solvant ainsi que la prédiction de leurs structures secondaires afin de calculer une propension de désordre intrinsèque. Les méta-prédicteurs, comme PrDOS (*Ishida & Kinoshita, 2007*), combinent la prédiction de plusieurs prédicteurs afin de déterminer la proportion de désordre intrinsèque. PrDOS possède l'une des meilleures performances de prédiction (*Meng et al., 2017*).

Comme décrit précédemment, les MoRFs sont des séquences désordonnées (typiquement de 5 à 25 résidus d'acides aminés) subissant une transition désordre – ordre lorsqu'elles interagissent avec leurs partenaires protéiques. Chez les eucaryotes, environ une région désordonnée sur cinq possède une telle séquence (*Katuwawala et al., 2019*). Une quinzaine de prédicteurs de MoRFs sont actuellement disponibles et diffèrent selon leur méthode de prédiction et leur performance. La plupart se basent dans un premier temps dans la recherche des séquences désordonnées dans lesquelles le prédicteur de MoRFs, à proprement parler, va effectuer ses prédictions. La recherche de MoRFs était initialement basée sur l'estimation de la propension qu'a une séquence désordonnée à adopter une structure secondaire (*Katuwawala et al., 2019*). Désormais, la grande majorité des prédicteurs reposent sur des algorithmes d'apprentissage automatique (machine learning) entraînés sur un jeu de données de MoRFs, comme *MoRFChiBiWeb* (*Malhis et al., 2016*) ou *OPAL+* (*Sharma et al., 2019*).

b) Techniques d'analyse des protéines et régions désordonnées

L'hétérogénéité conformationnelle des protéines intrinsèquement non structurées ne leur permet pas d'être étudiées de manière conventionnelle par des méthodes structurales à haute résolution comme la cristallographie aux rayons X. La structure de certains segments désordonnés peut cependant être obtenue par cette méthode notamment en les cocristallisant avec leurs partenaires protéiques mais cela ne représente pas l'ensemble des conformations qu'elles peuvent adopter. Pour cela, plusieurs approches qui peuvent être combinées ont été développées (Na et al., 2018, Schramm et al., 2019). La caractérisation des protéines désordonnées nécessite une analyse multi-approches pouvant faire intervenir à la fois des techniques de résolution atomique en solution, comme la résonance magnétique nucléaire (RMN), ou en phase gazeuse, comme la spectrométrie de masse couplée ou non à la mobilité ionique (ESI-(IM)-MS) (Figure 31).



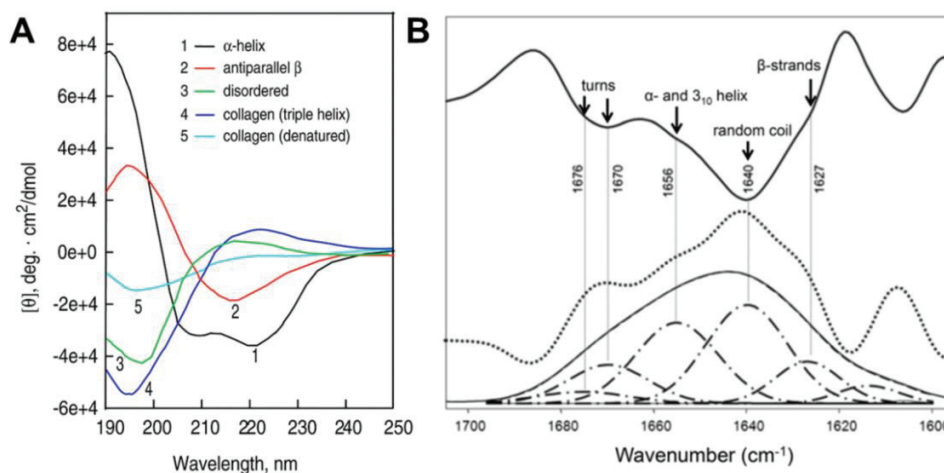
**Figure 31 : Procédure et analyse multi-approches des protéines intrinsèquement non structurées.** Les techniques sont classées de gauche à droite en fonction de leur complexité et de haut en bas selon la procédure d'analyse proposée par Schramm et al., (2019). MALLS : diffusion de la lumière laser multi-angles, AUC : ultracentrifugation analytique, FTIR : spectroscopie infrarouge à transformée de Fourier, DSC : calorimétrie différentielle à balayage, HDX-MS : spectrométrie de masse par échange hydrogène/deutérium, FRET : transfert d'énergie par résonance de type Förster, SDSL-EPR : site-directed spin labeling - electron paramagnetic resonance, TCQ : tryptophan triplet cysteine quenching. D'après Schramm et al., (2019).

D'autres techniques sont fréquemment utilisées afin de déterminer par exemple la composition en structures secondaires par dichroïsme circulaire (CD), ou les propriétés hydrodynamiques par diffusion dynamique de la lumière (DLS) ou par chromatographie d'exclusion stérique analytique (SEC). La détection de désordre intrinsèque peut être réalisée en utilisant des techniques courantes et faciles à mettre en œuvre. Lors de leur analyse par électrophorèse en conditions dénaturantes (SDS-PAGE), les protéines enrichies en désordre migrent anormalement et possèdent une masse moléculaire apparente 1,2 à 1,8 fois plus élevée que celle déduite de leur séquence en acides aminés. Les IDPs possèdent également une masse moléculaire apparente plus importante lors de leur analyse par chromatographie d'exclusion stérique lorsque la calibration de la résine a été effectuée avec des standards globulaires. Ces trois observations sont une conséquence des conformations étendues des protéines désordonnées, les IDPs occupant un volume hydrodynamique plus important qu'une protéine globulaire ayant le même nombre de résidus d'acides aminés. Enfin, les IDPs sont également plus sensibles aux protéases.

Les propriétés hydrodynamiques des IDPs peuvent être déterminées par la mesure de leur rayon hydrodynamique par diffusion dynamique de la lumière ou par chromatographie d'exclusion stérique. Les techniques de diffusion étant très sensibles aux particules de haute masse moléculaire, l'analyse par DLS peut être couplée à une colonne de filtration sur gel (SEC-DLS). Le rayon hydrodynamique des IDPs peut également être mesuré par ultracentrifugation analytique. Cette technique est suffisamment sensible pour distinguer les conformations globulaires, les pré-globules fondus et les conformations random coils (*Schramm et al., 2019*). Le rayon hydrodynamique mesuré peut être comparé à des valeurs théoriques correspondant à une conformation globulaire, globule fondu, pré-globule fondu ou totalement désordonnée (*Uversky, 2002*). Cette comparaison permet d'évaluer l'état de compaction de la protéine étudiée. La diffusion de la lumière laser multi-angles et l'ultracentrifugation analytique permettent en plus de déterminer l'état d'oligomérisation des protéines par la mesure de leur masse moléculaire.

Les proportions en structures secondaires peuvent être déterminées par dichroïsme circulaire qui ne nécessite pas une concentration importante de protéine (*Greenfield, 2006, Chemes et al., 2012*) (**Figure 32A**). Le calcul des proportions de structures secondaire et de désordre intrinsèque nécessite des jeux de données contenant des protéines de références afin de déconvoluer les spectres expérimentaux.

Plusieurs jeux de données ont été décrits mais leur contenu en protéines dépourvues de structures secondaires est faible voire inexistant, ce qui peut biaiser les résultats de la déconvolution des spectres de protéines intrinsèquement non structurées. Le jeu de donnée « 7 » (*Sreerama et al., 2000*) contient 5 protéines chimiquement dénaturées et se rapproche ainsi des caractéristiques des IDPs. Il est accessible depuis le serveur Dichroweb (<http://dichroweb.cryst.bbk.ac.uk>, *Whitmore & Wallace, 2008*). Le dichroïsme circulaire est très utile afin de cribler des conditions expérimentales influençant les proportions en structures secondaires, comme la température, le pH, ou l'effet d'osmolytes. La spectroscopie infrarouge à transformée de Fourier (FTIR) permet également d'obtenir des proportions en structures secondaires par la déconvolution de la bande amide I ( $1700\text{-}1600\text{ cm}^{-1}$ ). Les structures random coil sont identifiables aux alentours de  $1640\text{ cm}^{-1}$  alors que les hélices  $\alpha$  et les brins  $\beta$  le sont aux alentours de  $1656$  et  $1627\text{ cm}^{-1}$ , respectivement (**Figure 32B**).



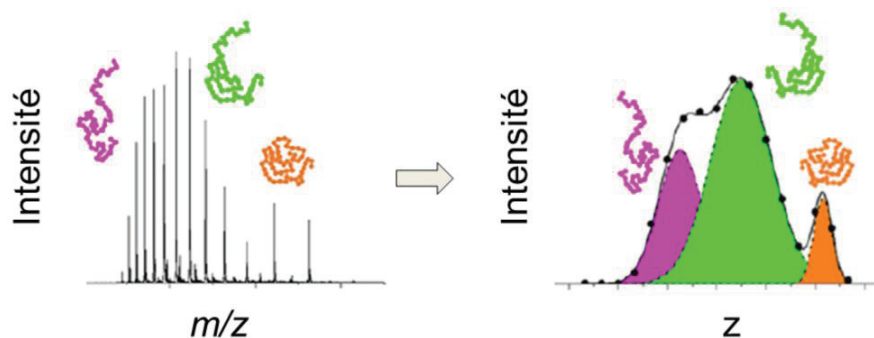
**Figure 32 : Détermination de la proportion en structures secondaires par dichroïsme circulaire (A) et spectroscopie infrarouge à transformée de Fourier (B).** Les spectres de dichroïsme circulaire et de spectroscopie infrarouge sont extraits de *Greenfield (2006)* et de *Schramm et al., (2019)*, respectivement.

La diffusion des rayons X aux petits angles (SAXS) est également une technique fréquemment employée dans la caractérisation des protéines désordonnées (*Receveur-Bréchet & Durand, 2012, Kikhney & Svergun, 2015, Cordeiro et al., 2017, Schramm et al., 2019*). Le calcul de la valeur du rayon de gyration ( $R_g$ ) permet de déterminer le degré de compaction des IDPs, notamment en la comparant à des valeurs théoriques calculées pour des conformations étendues ou plus ou moins repliées selon l'équation de Flory.

La dimension maximale de la protéine ( $D_{max}$ ), qui dans le cas des IDPs correspond à la taille de la conformation la plus étendue, peut également être déterminée ainsi que la masse moléculaire de la protéine étudiée et donc son état d'oligomérisation. Le SAXS permet également de déterminer la flexibilité de la protéine. De plus, le calcul du rapport  $R_g/R_h$  informe du degré de compaction globale de la protéine. Ce rapport est égal à 0,778 pour une sphère et augmente de façon inversement proportionnelle avec le degré de compaction. Il est égal à 2,36 pour une conformation apparentée à un bâton rigide (*Brewer & Striegel, 2011*).

Les intensités de diffusion étant additives, la courbe de diffusion résultante des expériences de SAXS représente ainsi une moyenne pondérée de la participation de chaque conformation adoptée par une IDP. Il est possible de caractériser l'hétérogénéité conformationnelle des protéines désordonnées par déconvolution des intensités de diffusion. Pour cela, plusieurs méthodes ont été décrites (*Bonomi et al., 2017*) comme EROS (*ensemble-refinement of SAXS, Różycki et al., 2011*), ENSEMBLE (*Choy & Forman-Kay, 2001*), MES (*minimal ensemble search, Pelikan et al., 2009*) ou EOM (*ensemble optimization method, Bernadó et al., 2007, Tria et al., 2015*). Brièvement, EOM génère un grand nombre de conformations ( $\geq 10\ 000$ ) à partir de la séquence en acides aminés de la protéine étudiée, puis sélectionne par un algorithme génétique un sous-ensemble de conformations de façon à ce que la somme des intensités de diffusion théorique des conformations sélectionnées reproduise les intensités de diffusion expérimentales. EOM procure ainsi la distribution du  $R_g$  et du  $D_{max}$  de l'ensemble sélectionné ce qui reflète l'espace conformationnel de la protéine désordonnée étudiée.

La spectrométrie de masse en conditions natives (ESI-MS), couplée ou non à la mobilité ionique (IM-MS), permet également d'évaluer l'hétérogénéité conformationnelle des protéines désordonnées (*Santambrogio et al., 2018, Stuchfield et al., 2018, Mitra et al., 2019*). Les protéines sont transférées en phase gazeuse et ionisées par électronebulisation. Plus les conformations sont étendues, plus elles acquièrent un nombre important de charges. L'analyse de la distribution des états de charge, c'est-à-dire du graphique représentant l'intensité de chaque ion en fonction de son rapport masse sur charge ( $m/z$ ), permet de distinguer des sous-populations de conformations qui diffèrent par la quantité de charges qu'elles ont acquises (**Figure 33**). Il est également possible de quantifier l'abondance relative de chaque sous-population et de calculer leur surface accessible au solvant (SASA), ce qui permet d'estimer le degré de compaction de chaque population de conformations.



**Figure 33 : Analyse de l'hétérogénéité conformationnelle des protéines désordonnées par spectrométrie de masse.** La déconvolution de la distribution des états de charge (à gauche) est réalisée en modifiant la représentation de  $I(m/z) = f(m/z)$  en  $I(z) = f(z)$  puis en ajustant des courbes Gaussiennes sous chaque populations observées dans cette nouvelle représentation (à droite). Figure adaptée de *Santambrogio et al.*, (2018).

Dans les expériences de mobilité ionique, un tube de dérive, c'est-à-dire un tube contenant du gaz, typiquement de l'hélium, est inséré entre la source d'ionisation et les analyseurs du spectromètre. Il permet de séparer différentes sous populations de conformations peuplant l'espace conformationnel d'une protéine désordonnée en fonction de leur degré de compaction et de leur charge. Plus une conformation est étendue, plus la probabilité de collision avec un atome d'hélium est importante, ralentissant ainsi son transfert et augmentant de ce fait son temps de dérive. Les facteurs de conformation et de charge sont regroupés dans le calcul de la section efficace de collision (CCS) dont il est possible de calculer des valeurs théoriques pour une conformation globulaire (une sphère) ou une conformation étendue (en bâton). Ces deux extrêmes forment une gamme théorique de valeurs de CCS et la comparaison des valeurs expérimentales de CCS à cette gamme permet de quantifier l'espace conformationnel exploré par les protéines désordonnées (*Beveridge et al.*, 2014, *Beveridge et al.*, 2015, *Stuchfield & Barran*, 2018).

Parmi l'ensemble des techniques utilisées pour la caractérisation des protéines désordonnées, la spectroscopie RMN est la seule pouvant rendre compte de leur hétérogénéité conformationnelle et de leur dynamique avec une résolution atomique (*Brutscher et al.*, 2015, *Gibbs et al.*, 2017, *Schneider et al.*, 2018). Plusieurs paramètres sont mesurables comme les déplacements chimiques et les couplages dipolaires qui permettent de détecter la présence de repliements locaux comme des structures secondaires et de détecter des interactions longue distance (*Jensen et al.*, 2013).

Les mesures de relaxation du spin permettent en plus de caractériser la dynamique des protéines désordonnées avec une résolution atomique (*Salvi et al., 2017*). La spectroscopie RMN est également utilisée afin d'étudier des complexes impliquant des protéines désordonnées et permet de caractériser la cinétique d'interaction entre les deux partenaires et d'étudier des complexes de faible affinité avec une haute résolution (*Schneider et al., 2019*). La spectroscopie RMN permet enfin d'étudier les protéines désordonnées dans des conditions mimant l'encombrement stérique retrouvé *in vivo* (*Adamski et al., 2019*).

Plusieurs techniques sont généralement employées conjointement dans la caractérisation des protéines désordonnées. Le SAXS et les techniques de spectrométrie de masse sont fréquemment utilisées pour les caractériser (*Saikusa et al., 2013, Boon et al., 2018, Han et al., 2018*) ainsi que le SAXS et la spectroscopie RMN (*Sibille & Bernadó, 2012, Aznauryan et al., 2016, Schramm et al., 2019*) pour lequel certains programmes ont été développés afin de modéliser des ensembles de conformations en utilisant comme contraintes des données de SAXS et de RMN tels que les programmes ENSEMBLE (*Choy & Forman-Kay, 2001*) ou ASTEROID (*Jensen et al., 2008*). Des ensembles de conformations peuvent également être générés par dynamique moléculaire (*Ithuralde et al., 2016, Bonomi et al., 2017*). Ces simulations sont généralement combinées aux méthodes expérimentales afin de caractériser au mieux l'espace conformationnel des IDPs (*Chan-Yao-Chong et al., 2019a, Chan-Yao-Chong et al., 2019b*). Par exemple, des modèles du propeptide de la lysyl oxydase, une protéine désordonnée, ont été générés par dynamique moléculaire avec UNRES (*Czaplewski et al., 2018*) en utilisant comme contraintes des données expérimentales obtenues lors d'expériences de SAXS (*Vallet et al., 2018*).

# **CHAPITRE III**

**MATÉRIELS**

**&**

**MÉTHODES**



### III.1 Expression des ectodomains des syndécans en système procaryote

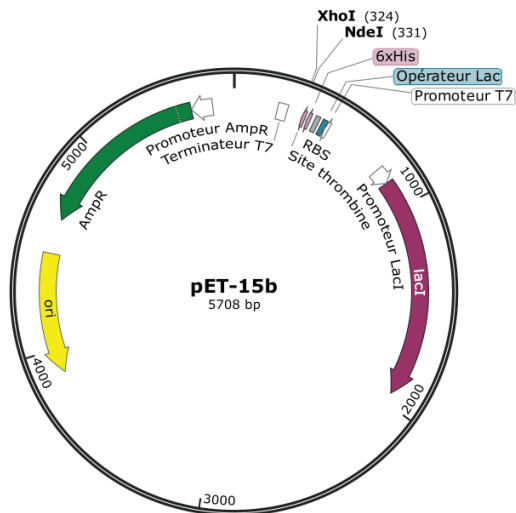
#### III.1.1 Vecteur d'expression

L'ADN optimisé pour une expression en système procaryote et codant pour les ectodomains des syndécans-1, -2, -3 (ED-1 – 3), de l'isoforme 2 du syndécane-4 (ED-4) et de son mutant (ED-4 C150A) a été synthétisé et cloné dans le vecteur d'expression pET-15b par Genscript ([www.genscript.com](http://www.genscript.com)). Les bornes des ectodomains ont été choisies selon les informations disponibles dans la base de données UniProt KB ([www.uniprot.org](http://www.uniprot.org), *The Uniprot Consortium, 2018*), excepté pour l'ectodomaine du syndécane-3 pour lequel aucune information concernant la longueur du peptide signal n'est disponible. Etant donné que les séquences en acides aminés des syndécans sont très conservées entre les différentes espèces, la séquence du syndécane-3 humain a été alignée avec celle du SDC-3 de souris et de rat, dont la taille du peptide signal a été déterminée (Code d'accès UniProt KB : Q64519 et P33671 pour la souris et le rat, respectivement) ce qui a permis de définir le résidu d'alanine 47 comme résidu N-terminal d'ED-3 (**Figure 34**).

L'ADN des ectodomains des syndécans a été cloné dans le vecteur pET-15b en utilisant les sites de restriction *NdeI* et *XhoI* situés en aval de la séquence codant pour une étiquette 6-Histidines (6xHis) (**Figure 35**). Une séquence d'ADN codant pour le peptide FLAG a été insérée à l'extrémité 3' des ADNs codant pour les ectodomains des syndécans afin d'obtenir une protéine recombinante possédant une étiquette 6xHis à l'extrémité N-terminale et une étiquette FLAG (DYKDDDDK) à l'extrémité C-terminale. L'ADN synthétisé et cloné par Genscript a été repris en tampon Tris-EDTA (Tris-HCl 10 mM, EDTA 1 mM, pH 8,0) à une concentration de 40 µg/ml puis stocké à -80°C.

Humain	<u>MKPGPPHRAGAAHGAGAGAGAAAGPGARGLLLPPLLLLLLAGRAAGAQRWRSENFERPVD</u>	60
Souris	-----AQRWRNENFERPVD	14
Rat	-----AQRWRNENFERPVD	14
	*****	
Humain	LEGSDDDSFPDDELDDLYSGSGSGYFEQESGIETAMRFS PDVALAVSTTPAVLPTTNIQ	120
Souris	LEGSDDDSFPDDELDDLYSGSGSGYFEQESGLETAMRFIPDMALAAPTAPAMLPTTVIQ	74
Rat	LEGSDDDSFPDDELDDLYSGSGSGYFEQESGLETAMRFIPDIALAAPTAPAMLPTTVIQ	74
	*****:***** **:*** . *:**:**** **	
Humain	PVGTPEEELPSERPTLEPATSPLVVTE---VPEEPSQRATTVSTTMATTAATSTGDPPTVA	177
Souris	PVDTPFEELLSEHPSPEPVTSPPLVTEVTEVVEESSQKATTISTTTSTAATTTGAPTMA	134
Rat	PVDTPFEELLSEHPGPEPVTSPPLVTEVTEVVEEPSQRATTISTTTSTAATTTGAPTMA	134
	** .***** **: * ** .*** :*** * ** **:***:*** :*****: ** **:	
Humain	TVPATVATATPSTPAAPPFTATTAVIRTTGVRLLPLPLTTVATARATTPPEAPSPPTTAA	237
Souris	TAPATAATTAPSTPEAPPATATVADVRTTGIQGLMLPLPLTTAATAKITTPAAPSPPTTVA	194
Rat	TAPATAATTAPSTPAAPPATATTADIRTTGIQGLLPLPLTTAATAKATTPAVPSPTTVT	194
	*.***. **: :**** ** ** . * :****: : *****.***: ** .*****.:	
Humain	VLDTEAPT PRLVSTATSRPRALPRPATTQEPDIPERSTLPLGTTAPGPTEVAQTPTPETF	297
Souris	TLDTEAPT PRLVNTATSRPRALPRPVTTQEPDVAERSTLPLGTTAPGPTEMAQTPTPESL	254
Rat	TLDTEAPT PRLVNTATSRPRALPRPVTTQEPDVAERSTLPLGTTAPGPTEVAQTPTPESL	254
	.***** .***** .*****: : *****:*****:*****::	
Humain	LTTIRDEPEVPVSGGPGDFELPEEETTQPDANEVAVGGAAAKASSPPGTLPKGARPG	357
Souris	LTTIQDEPEVPVSGGPGDFELQE-ETTQPDANEVAVEGAAAKPSPLGTLPKGARPG	313
Rat	LTTTQDEPEVPVSGGPGDFELQE-ETTQPDANEVAVEGAAAKPSPLGTLPKGARPG	313
	*** :***** * ***** ***** * * *****	
Humain	PGLLDNAIDSGSSAAQLPQKSILERKEVLVAVIVGGVVGALFAAFLVTL LIYRMKKKDEG	417
Souris	PGLHDNAIDSGSSAAQLPQKSILERKEVLVAVIVGGVVGALFAAFLVTL LIYRMKKKDEG	373
Rat	LGLHDNAIDSGSSAAQLPQKSILERKEVLVAVIVGGVVGALFAAFLVTL LIYRMKKKDEG	373
	** *****	
Humain	SYTLEEPKQASVTYQKPKQEEFYA	442
Souris	SYTLEEPKQASVTYQKPKQEEFYA	398
Rat	SYTLEEPKQASVTYQKPKQEEFYA	398
	*****	

**Figure 34 : Alignement des séquences du syndécan-3 entier humain, de souris et de rat.** L'alignement a été réalisé par *Clustal Omega* (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). La séquence déduite du peptide signal du SDC-3 humain est soulignée. Un astérisque (\*) définit un résidu d'acide aminé conservé, les deux points (:) et le point (.) indiquent que les résidus ont des propriétés physico-chimiques plus ou moins similaires, respectivement. L'absence de symboles indique que les résidus ont des propriétés totalement différentes et le tiret (-) indique des séquences non alignables (*gap*).



**Figure 35 : Carte du vecteur d'expression pET-15b vide montrant ses principales caractéristiques** dont les deux sites de restriction utilisés pour le clonage (*NdeI* et *XhoI*). Carte réalisée avec *SnapGene Viewer* ([www.snapgene.com](http://www.snapgene.com)).

### III.1.2 Transformation des bactéries

20 µl d'un stock en glycérol de souche d'expression *Escherichia coli* BL21 (DE3) ont été resuspendus dans environ 5 ml de milieu de culture LB-Lennox (Sigma, L3022) et la culture a été incubée pendant la nuit à 37°C sous agitation (environ 150 rpm). La culture a ensuite été centrifugée à 4°C pendant 10 min à 6000 x g et le surnageant a été éliminé. Le culot de bactéries a été repris par 10 ml de CaCl<sub>2</sub> 50 mM en eau stérile puis laissé 1 h dans la glace afin de les rendre compétentes. La suspension a ensuite été centrifugée à 4°C pendant 10 min à 6000 x g et le surnageant a été éliminé. Le culot de bactéries compétentes a été repris par 1 ml de CaCl<sub>2</sub> 50 mM et stocké à 4°C pour une utilisation extemporanée.

80 ng d'ADN plasmidique ont été ajoutés à 50 µl de suspension de bactéries compétentes et le mélange a été laissé entre 30 et 40 min dans la glace. La suspension a été soumise à un choc thermique à 42°C au bain-marie pendant 1 min. 250 µl de LB-Lennox ont été ajoutés et la suspension a été incubée pendant 1 h à 37°C sous agitation (environ 250 rpm). Après incubation, 50 µl de suspension ont été prélevés et étalés sur des boîtes de Pétri préalablement préparées par 25 ml de LB-Lennox contenant de l'agar (L2897, Sigma) auxquelles a été ajoutée de l'ampicilline à une concentration finale de 100 µg/ml, et les boîtes de culture ont été placées à l'étuve (37°C) pendant la nuit.

### III.1.3 Mise en culture des bactéries transformées

Une colonie isolée s'étant développée sur une boîte ampicilline a été prélevée et resuspendue dans environ 50 ml de LB-Lennox, puis la préculture a été incubée pendant la nuit à 37°C sous agitation (environ 150 rpm). Le lendemain, la densité cellulaire a été évaluée par spectrophotométrie à 600 nm ( $A_{600}$ ), généralement comprise entre 2 et 2,5. La préculture a ensuite été diluée dans 1 L de LB-Lennox contenant 100 µg/ml d'ampicilline afin d'obtenir une  $A_{600}$  d'environ 0,1 puis incubée à 37°C sous agitation (environ 180 rpm) jusqu'à ce que l'absorbance à 600 nm atteigne environ 0,8. Puis l'expression des protéines recombinantes a été induite par l'addition d'isopropyl-β-D-1-thiogalactopyranoside (IPTG) 1 mM pendant 4 h. Après induction, la culture bactérienne a été centrifugée à 4°C pendant 10 min à 6000 x g, le surnageant a été éliminé et le culot de bactéries a été stocké à -20°C. La totalité des expériences a été effectuée en milieu stérile. Les caractéristiques des protéines recombinantes exprimées en système procaryote sont détaillées dans le **Tableau II**.

Protéine recombinante	Code d'accès UniProt KB	1 <sup>er</sup> et dernier résidu d'acides aminés (UniProt KB)	Nombre de résidus d'acides aminés dans la protéine recombinante
ED-1	P18827	23 - 254	260
ED-2	P34741	19 - 144	154
ED-3	O75056	47 - 387	373
ED-4	P31431-2	19 - 153	163
ED-4 C150A	-	19 - 153	163

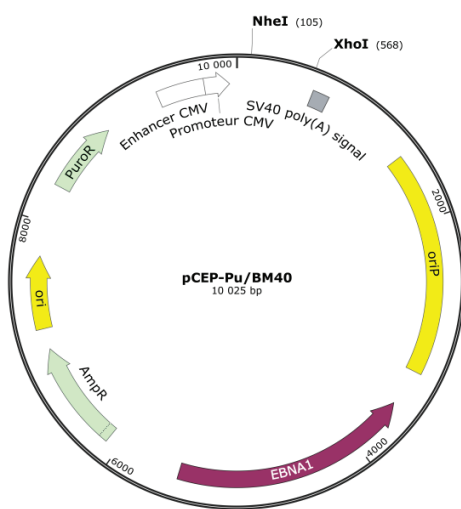
**Tableau II : Caractéristiques des ectodomaines des syndécans humains exprimés sous forme recombinante en système procaryote.**

### III.2 Expression des ectodomaines des syndécans en système eucaryote

#### III.2.1 Vecteur d'expression

Afin d'exprimer des protéines recombinantes dont les séquences sont identiques à celles des protéines exprimées en système procaryote, les séquences d'ADN codant pour une étiquette 6xHis et un site de clivage à la thrombine ont été incorporées à l'extrémité 5' de l'ADN codant pour les ectodomaines des syndécans. De ce fait, les protéines recombinantes posséderont également une étiquette 6xHis à l'extrémité N-terminale et une étiquette FLAG à l'extrémité C-terminale.

Le clonage a été effectué par Genscript dans le vecteur d'expression eucaryote pCEP-Pu/BM40 en utilisant les sites de restriction *NheI* et *XhoI* (**Figure 36**). L'utilité majeure de ce vecteur repose sur l'expression d'une protéine contenant la protéine d'intérêt couplée au peptide signal de la protéine BM40 (ou ostéonectine, une protéine extracellulaire). Le peptide BM40 permet la sécrétion de la protéine exprimée et sera ensuite clivé, laissant cependant une cicatrice menant à la synthèse de 4 résidus (APLA) à l'extrémité N-terminale de la protéine, qui précèdera donc l'étiquette 6xHis. L'ADN a été repris en tampon Tris-EDTA (Tris-HCl 10 mM, EDTA 1 mM, pH 8,0) à une concentration de 40 µg/ml puis stocké à -80°C.



**Figure 36 : Carte du vecteur d'expression pCEP-Pu/BM40** montrant ses principales caractéristiques dont les deux sites de restriction utilisés pour le clonage (*NheI* et *XhoI*).

### III.2.2 Transfection des cellules eucaryotes

Des cellules embryonnaires de rein humain (HEK 293 EBNA) stockées dans l'azote liquide en milieu de culture eucaryote (DMEM, D5769, Sigma) contenant du sérum de veau foetal (SVF, A3160402, ThermoFisher) et du DMSO en rapport 5:4:1 (v/v, environ 1,5 ml au total) et en quantité suffisante pour recouvrir 75 cm<sup>2</sup> de surface à confluence cellulaire ont été centrifugées à température ambiante pendant 5 min à 500 x g en présence d'environ 10 ml de DMEM contenant 10 % de SVF et 50 µg/ml de gentamycine (G1272, Sigma). Le surnageant a été éliminé et le culot de cellules a été repris par 15 ml du même milieu. La suspension a ensuite été transvasée dans un flacon de culture cellulaire possédant une surface de 75 cm<sup>2</sup> (T75) et les cellules ont été incubées à 37°C, 5 % CO<sub>2</sub> jusqu'à ce qu'elles soient à confluence (environ 2 jours). Le milieu a ensuite été éliminé et les cellules ont été remises en suspension à l'aide d'un râteau en présence d'environ 5 ml de PBS (D8537, Sigma).

La suspension a ensuite centrifugée à température ambiante pendant 5 min à 500 x g puis le PBS a été éliminé et les cellules ont été remises en suspension dans du DMEM contenant 10 % de SVF et 50 µg/ml de gentamycine à une concentration cellulaire équivalent à environ 70 % de confluence. 1 ml de suspension a été distribué dans plusieurs puits d'une plaque 12 puits (équivalent à environ 3,8 cm<sup>2</sup> par puits) et la plaque a été incubée pendant la nuit à 37°C, 5 % CO<sub>2</sub> afin d'atteindre environ 80 % de confluence.

Ces cellules ont été transfectées par la lipofectamine 2000 (11668030, Invitrogen) 8 % (v/v) diluée dans du DMEM contenant 50 µg/ml de gentamycine et 800 ng d'ADN plasmidique. Le mélange a été incubé pendant 20 min à température ambiante (volume total de 200 µl). Le milieu de culture dans la plaque 12 puits a été renouvelé par du DMEM contenant 10 % de SVF et 50 µg/ml de gentamycine et la solution de transfection a été ajoutée. Les cellules ont ensuite été incubées à 37°C, 5 % CO<sub>2</sub> jusqu'à atteindre la confluence cellulaire (généralement en 24 h). A ce stade, la puromycine (P9620, Sigma) a été ajoutée à une concentration de 0,25 µg/ml après avoir renouvelé le milieu et les cellules ont de nouveau été incubées à 37°C, 5 % CO<sub>2</sub>. La concentration en puromycine a ensuite été augmentée à 0,5 µg/ml et a été conservée à cette concentration tout le long de la croissance cellulaire et du passage des cellules. Arrivée à l'état de confluence, typiquement au bout de 2 jours d'incubation, les cellules ont été transvasées dans des flacons de culture ayant sur une surface 3 fois supérieure à la précédente, donc chronologiquement en T25 (surface de 25 cm<sup>2</sup>), T75 (75 cm<sup>2</sup>) puis T225 (225 cm<sup>2</sup>) jusqu'à obtenir 9 flacons de culture de 225 cm<sup>2</sup> (environ 50 ml de milieu de culture par T225).

### **III.2.3 Culture des cellules eucaryotes et récoltes des milieux**

Le SVF a été supprimé du milieu ainsi que la puromycine dans les 9 flacons de T225 soit une surface totale de 2025 cm<sup>2</sup> afin de faciliter les purifications des protéoglycanes. Le milieu de culture ne contenait alors que du DMEM et 50 µg/ml de gentamycine. La récolte de ces milieux s'est effectuée tous les 2 jours. Le milieu récolté a été centrifugé à 4°C pendant 30 min à 14 000 x g, puis le surnageant a été stocké à -80°C. Entre 6 et 8 récoltes ont généralement été effectuées soit un volume total d'environ 3 L de milieu par production. Toutes les expériences de culture cellulaires ont été effectuées en milieu stérile (poste de sécurité microbiologique).

### III.3 Purification des ectodomains des syndécans

#### III.3.1 Purification des ectodomains exprimés en système procaryote

##### a) Purification par chromatographies d'affinité

Le culot de culture de bactéries a été décongelé et resuspendu par 10 ml de tampon HBS (HEPES 10 mM, NaCl 150 mM pH 7,4) contenant du MgCl<sub>2</sub> 2,5 mM, du CaCl<sub>2</sub> 2,5 mM, de la désoxyribonucléase I (DNase I) 100 µg/ml (D5025, Sigma), du CHAPS 1 mM (Roche) et une tablette de cocktail d'inhibiteurs de protéases sans EDTA (Complete EDTA-free, Roche). La suspension a été lysée mécaniquement (*bead beater*, BioSpec) 8 fois pendant 30 s en présence de billes de verre de 100 µm de diamètre (G4649, Sigma). Le lysat a été récupéré et les billes de verre ont été lavées par environ 5 ml d'une solution de tampon HBS contenant du CHAPS 1 mM. La totalité du lysat (lysat initial et lavages) a ensuite été centrifugé à 4°C pendant 10 min à 10 000 x g, le surnageant a été récupéré et de nouveau centrifugé dans les mêmes conditions afin d'éliminer tous types de débris membranaires ou de grosses particules. Le lysat clarifié (environ 15 ml) a été chargé sur 1,5 ml de résine Ni-NTA (Qiagen) préalablement équilibrée par au moins 50 ml de tampon HBS à un débit de 20 ml/h. Le lysat clarifié a été passé deux fois sur la résine Ni-NTA et la fraction non retenue a été conservée pour de futures analyses. La résine Ni-NTA a été lavée par environ 40 ml de tampon HBS contenant de l'imidazole 10 mM puis 20 mM. Les protéines ont été éluées par environ 5 ml de tampon HBS contenant successivement de l'imidazole à une concentration finale de 100, 300 et 500 mM. La résine a ensuite été régénérée en passant environ 10 ml de tampon HBS contenant de l'imidazole 2 M puis rééquilibrée par au moins 50 ml de tampon HBS.

Après analyse, les fractions contenant les ectodomains des syndécans pré-purifiés sur résine Ni-NTA ont été décongelées et aliquotées par fraction de 1 ml. Chaque fraction a été diluée 10 fois en tampon HBS contenant 1/20<sup>ème</sup> d'une tablette de cocktail d'inhibiteurs sans EDTA en vue d'une purification par chromatographie d'affinité anti-FLAG. Ce fractionnement est nécessaire du fait de la plus faible capacité de la résine anti-FLAG à capturer les protéines recombinantes possédant une étiquette FLAG (typiquement 0,6 à 1 mg de protéine capturée par ml de résine anti-FLAG contre environ 20 à 50 mg/ml de résine Ni-NTA). 6 ml de suspension de résine anti-FLAG (A2220, Sigma) en glycérol (en rapport 1/1, soit environ 3 ml de résine final) ont été équilibrés par au moins 60 ml de tampon HBS puis l'aliquot dilué de protéines pré-purifiées sur résine Ni-NTA a été passé environ 2 fois sur la résine anti-FLAG.

La résine a ensuite été lavée par 20 ml de tampon HBS puis les protéines ont été éluées par 10 ml de peptide FLAG (synthétisé par Proteogenix, pureté > 95%) 200 µg/ml en tampon HBS. La résine a ensuite été régénérée par environ 7 ml d'une solution de glycine 100 mM, pH 3,5 puis rééquilibrée par au moins 60 ml de tampon HBS. Toutes les étapes de purification par chromatographie d'affinité sur résines Ni-NTA et anti-FLAG ont été effectuées à 4°C et à un débit de 20 ml/h. Les tampons utilisés pour chaque purification ont préalablement été filtrés sur membrane d'ester de cellulose (pores de 0,22 µm, Millipore) et stockés à 4°C.

#### *b) Purification par chromatographie d'exclusion stérique*

Une ou plusieurs fractions (selon la quantité de protéine présente) purifiées sur résine anti-FLAG ont été regroupées puis concentrées par centrifugation. Un dispositif de concentration d'une capacité de 15 ml et d'une masse moléculaire limite de 10 kDa (Amicon Ultra-15 10K, Millipore) a préalablement été lavé par pipetage par environ 5 ml d'eau distillée afin d'éliminer le glycérol de stockage, puis lavé par centrifugation avec 10 ml d'eau distillée et enfin équilibré par 10 ml de tampon HBS par centrifugation à 4°C à 2500 x g pendant 10 min. Les fractions d'éluion ont été concentrées dans les mêmes conditions afin d'obtenir un volume final d'environ 2 ml. Le rétentat a ensuite été de nouveau concentré à 4°C par centrifugation et en cycle de 2 min à 8500 x g, en utilisant un dispositif plus petit (0,5 ml, Amicon Ultra-0.5 10K, Millipore), préalablement lavé et équilibré dans les mêmes conditions, jusqu'à obtenir un volume de rétentat de 0,5 ou de 1 ml selon la concentration en protéine.

La purification par chromatographie d'exclusion stérique a été effectuée sur une résine semi-préparative de 24 ml (Superdex S200 Increase 10/300 GL, GE Healthcare) installée sur un système de purification Äkta (FPLC, *Äkta purification system*, GE Healthcare). La résine a été équilibrée par 2 volumes de colonne d'eau distillée et 1 volume de colonne de tampon HBS à l'envers afin d'éliminer de potentiels contaminants conservés après de précédents lavages, puis équilibrée à l'endroit par au moins 5 volumes de colonne de tampon HBS à un débit de 0,5 ml/min et à 10°C. 500 µl de la solution de protéine concentrée ont été injectés par une boucle de 500 µl en tampon HBS à un débit de 0,5 ml/min et à une température de 10°C. L'éluion des protéines a été suivie par spectrophotométrie à 280, 230 et 215 nm et les fractions d'éluions ont été récupérées par fraction de 50 ou 100 µl. La résine a ensuite été lavée selon les instructions du fournisseur et stockée entre 4 et 10°C. Les tampons utilisés pour chaque purification ont préalablement été filtrés sur membrane d'ester de cellulose (pores de 0,22 µm, Millipore) et stockés à 4°C.



### *c) Concentration des ectodomaines exprimés en système procaryote*

Les fractions d'élution en sortie de chromatographie d'exclusion stérique contenant les ectodomaines des syndécans purifiés ont été concentrées à 4°C par centrifugation et en cycle de 2 min à 8500 x g (sur Amicon Ultra-0.5 10K, Millipore), préalablement lavé et équilibré dans les mêmes conditions que décrites précédemment, jusqu'à obtenir un volume d'environ 500 µl ou une concentration souhaitée. Le rétentat a ensuite été centrifugé à 18 000 x g pendant 20 min à 4°C et le surnageant a délicatement été pipeté, transféré dans un nouveau tube à faible adsorption de protéines (Axygen) puis stocké à -20°C pour de futures expériences.

## **III.3.2 Purification des ectodomaines exprimés en système eucaryote**

### *a) Purification par chromatographie échangeuse d'anions*

10 ml de résine échangeuse d'anions (diéthylaminoéthyle, DEAE) ont été prélevés et lavées à 4°C à un débit de 20 ml/h par 30 ml d'une solution de NaOH 100 mM. La résine a ensuite été équilibrée par au moins 100 ml de tampon acétate de sodium 100 mM pH 4,5 contenant du NaCl 150 mM et de l'urée 5 M (tampon d'équilibration). Après décongélation du milieu de culture, de l'urée a été ajouté à une concentration finale de 5 M et de l'acétate de sodium à une concentration finale de 100 mM, puis le pH a été ajusté à pH 4,5 par de l'acide acétique. Le milieu contenant les ectodomaines des syndécans exprimés en système eucaryote a été passé une seule fois sur la résine DEAE à 20 ml/h et à 4°C. La résine a ensuite été lavée par au moins 50 ml de tampon d'équilibration puis le tampon a été échangé par 30 ml de tampon HBS (pH 7,4). Les protéoglycanes ont été élués par paliers de 20 ml par du tampon d'élution contenant du NaCl 0,5, 1, 1,5 et 2 M en HBS. Les fractions d'élution ont été congelées à -20°C pour de futures analyses et la résine a été rééquilibrée par au moins 100 ml de tampon d'équilibration.

### *b) Purification par chromatographie d'affinité sur résine anti-FLAG*

Après analyse, les fractions d'élutions récoltées après chromatographie échangeuse d'anions contenant les ectodomaines des syndécans exprimés en système eucaryote ont été décongelées et purifiées par chromatographie d'affinité sur résine anti-FLAG. Le même protocole que pour la purification des ectodomaines des syndécans exprimé en système procaryote a été suivi à l'exception de l'utilisation d'une concentration en NaCl de 300 mM dans le tampon HBS dans toutes les étapes de purification. L'étape d'élution a également été réalisée par du peptide FLAG 200 µg/ml en HBS (NaCl 0,3 M).

### *c) Concentration des ectodomains exprimés en système eucaryote*

Les fractions éluées lors de la purification par chromatographie d'affinité sur résine anti-FLAG ont été regroupées puis concentrées par centrifugation. Un dispositif de concentration d'une capacité de 15 ml et d'une masse moléculaire limite de 30 kDa (Amicon Ultra-15 30K, Millipore) a préalablement été lavé dans les mêmes conditions que décrits précédemment puis équilibré à 4°C par 10 ml de tampon HBS par centrifugation à 2500 x g pendant 10 min. Les fractions d'élution ont été concentrées afin d'obtenir un volume de rétentat d'environ 2 ml dans les mêmes conditions de centrifugation. Si nécessaire, le rétentat a été de nouveau concentré par centrifugation en cycle de 2 min à 8500 x g en utilisant un dispositif plus petit (0,5 ml, Amicon Ultra-0.5 30K, Millipore), préalablement lavé et équilibré dans les mêmes conditions. Le rétentat a ensuite été centrifugé à 18 000 x g pendant 20 min à 4°C et le surnageant a été délicatement pipeté, transféré dans un nouveau tube à faible adsorption et stocké à -20°C pour de futures expériences.

## **III.4 Analyse des syndécans purifiés par électrophorèse et par Western blot**

### **III.4.1 Electrophorèse sur gel de polyacrylamide en conditions dénaturantes**

Les protéines ont été analysées par électrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) suivant le protocole de *Laemmli* (1970) selon lequel elles sont séparées sur un gel discontinu composé d'un gel de concentration à 4 % de polyacrylamide (pH 6,8) et d'un gel de séparation dont la concentration en polyacrylamide variait 5 à 15 % (pH 8,8) dans nos expériences. L'ajout de dodécylsulfate de sodium (SDS) permet de dénaturer les protéines et de les charger négativement de façon homogène (1,4 g de SDS / g de protéine) ce qui permet de les séparer en fonction de leur masse moléculaire apparente. Les protéines exprimées en système procaryote et purifiées ont été analysées par SDS-PAGE avec une concentration en polyacrylamide de 8 à 15 % dans le tampon de migration Tris-HCl 62,5 mM, glycine 192 mM, SDS 0,1 % (*m/v*), pH 8,3). Les échantillons ont été dilués 2 fois en tampon non réducteur (Tris-HCl 125 mM pH 6,8, glycérol 20 % (*v/v*), SDS 4 % (*m/v*), bleu de bromophénol 0,1 % (*m/v*)) ou réducteur lorsque spécifié, puis placés au bain sec pendant 4 min à 95°C avant d'être déposés sur gel.

La migration a été effectuée à puissance constante (entre 3 et 5 W par gel). Les protéines ont été colorées toute la nuit par une solution de bleu de Coomassie (InstantBlue, ISB1L, Sigma) sous agitation (environ 300 rpm), puis le gel a été lavé plusieurs fois en eau distillée jusqu'à obtenir un fond incolore.

Les protéines exprimées en système eucaryote ont été analysées de façon similaire mais elles ont préalablement été précipitées avant leur dépôt. Pour cela, 100 ou 200 µl de solution ont été précipités par 4 volumes d'acétone froid (-20°C) puis placés à -20°C pendant une heure. La solution a ensuite été centrifugée à 14 000 x g, pendant 10 min à température ambiante. Le surnageant a été éliminé et les tubes ont été placés sous hotte pendant 5 min afin de faciliter l'évaporation de l'acétone résiduel. Le culot a ensuite été resuspendu en tampon non réducteur (Tris-HCl 62,5 mM pH 6,8, glycérol 10 % (v/v), SDS 2 % (m/v), bleu de bromophénol 0,05 % (m/v)) puis la solution a été placée au bain sec pendant 4 min à 95°C avant d'être déposée sur gel. Le détail de chaque analyse par SDS-PAGE est indiqué dans les figures. Lorsqu'une coloration au nitrate d'argent a été réalisée, le gel de polyacrylamide a été incubé toute la nuit dans une solution de fixation contenant de l'éthanol 50 % (v/v) et de l'acide acétique 10 % (v/v) diluée dans de l'eau distillée. Les étapes suivantes ont été réalisées dans le respect du protocole du fournisseur (ProteoSilver, PROTSIL1, Sigma-Aldrich).

### III.4.2 Western blot

Après la séparation par électrophorèse, les protéines exprimées en système procaryote ont été transférées sur une membrane de polyfluorure de vinylidène (PVDF ayant des pores de diamètre 0,45 µm, Millipore) ou sur une membrane de Nylon (ayant des pores de diamètre 0,45 µm, Invitrogen) pour celles exprimées en système eucaryote. La membrane de PVDF a préalablement été incubée pendant quelques minutes dans de l'éthanol pur puis dans le tampon de transfert (Tris 25 mM, glycine 192 mM, éthanol 20 % (v/v), pH 8,3). La membrane de Nylon a directement été incubée dans le tampon de transfert. Le transfert a été effectué pendant 1 h (Nylon) ou 1h30 (PVDF) à intensité constante (250 mA) dans le tampon de transfert. La membrane a ensuite été incubée dans 25 ml de solution de saturation (lait écrémé 5 % (m/v) / Tween-20 0,05 % (v/v) dans du tampon phosphate salin) pendant 1h30 à 37°C sur roue (10 rpm). La membrane a été lavée trois fois par 5 ml de PBS/Tween-20 0,1 % (v/v) puis incubée pendant la nuit dans 5 ml d'une solution de PBS/Tween-20 0,1 % (v/v) contenant 76 ng/ml d'anticorps primaire monoclonal anti-FLAG (F3165, Sigma-Aldrich) à 4°C et sur roue (10 rpm).

La membrane a été lavée 3 fois par 5 ml de PBS/Tween-20 0,1 % (v/v) puis incubée 1 h à température ambiante sur roue (10 rpm) dans 5 ml de PBS/Tween-20 0,1 % (v/v) contenant 40 ng/ml d'anticorps secondaire couplé à la peroxydase du raifort (172-1011, Bio-Rad). La membrane a été lavée 3 fois en PBS/Tween-20 0,1 % (v/v) puis 3 fois en PBS seul à température ambiante.

La révélation des protéines immunoréactives a été effectuée par l'incubation de la membrane avec 2,5 ml de deux solutions contenant le peroxyde d'hydrogène d'une part et le luminol (+ amplificateur) d'autre part (SuperSignal West Pico Plus, Thermo Scientific) pendant 5 min sur roue (10 rpm) et à température ambiante. L'excès de solution a été éliminé et le signal a été enregistré pendant 5 à 60 s par un ChemiDoc XRS+ (Bio-Rad).

### III.5. Calcul des concentrations des protéines purifiées par spectrophotométrie

Le serveur ProtParam (<https://web.expasy.org/protparam>) a été utilisé afin d'obtenir les masses moléculaires et le coefficient d'extinction molaire à 280 nm théorique des protéines exprimées en systèmes procaryote (ED-*x*) et eucaryote (ED-*x* PG) en utilisant leurs séquences en acides aminés (**Tableau III**).

Protéines recombinantes	Masse moléculaire théorique (Da)	Coefficients d'extinction molaire théorique à 280 nm ( $M^{-1}.cm^{-1}$ )
<b>Protéines exprimées en système procaryote</b> (sans glycosaminoglycanes)		
ED-1	27 160,92	6 990
ED-2	17 214,51	7 450
ED-3	38 830,73	9 970
ED-4	17 959,42	2 980
ED-4 C150A	17 790,22	2 980
<b>Protéoglycanes exprimés en système eucaryote</b> (avec glycosaminoglycanes)		
ED-1 PG	27 513,36	6 990
ED-2 PG	17 566,94	7 450
ED-3 PG	41 313,55	9 970
ED-4 PG	18 311,85	2 980

**Tableau III : Masse moléculaire théorique (Da) et coefficient d'extinction molaire théorique à 280 nm ( $M^{-1}.cm^{-1}$ ) calculés par ProtParam des ectodomaines des syndécans exprimés en systèmes procaryote et eucaryote.** Concernant les protéoglycanes (exprimés en cellules eucaryotes), la masse moléculaire représente celle de la protéine seule.

## III.6 Bioinformatique structurale

### III.6.1 Prédiction du désordre intrinsèque

La propension de désordre intrinsèque a été prédite avec le prédicteur metaPrDOS (*Ishida & Kinoshita, 2008*) en utilisant les séquences en acides aminés des protéines recombinantes. Les paramètres de prédiction ont été laissés par défaut (taux de faux-positifs 5 %, limite ordre/désordre 0,5).

### III.6.2 Prédiction de la composition en structures secondaires

La composition en structures secondaire a été prédite en utilisant les séquences en acides aminés des protéines recombinantes avec le prédicteur PSIPRED 4.0 (*Buchan & Jones, 2019*). Les paramètres de prédiction ont été laissés par défaut.

### III.6.3 Prédiction des *Molecular Recognition Features* (MoRFs)

La taille et la localisation des MoRFs sur les ectodomains des syndécans ont été prédites en utilisant les séquences en acides aminés des protéines recombinantes avec le prédicteur MoRFchibi System (*Malhis et al., 2016*).

### III.6.4 Modélisation l'extrémité C-terminale d'ED-4 et du dimère d'ED-4

La structure tridimensionnelle de l'extrémité C-terminale de l'isoforme-2 du syndécan-4 (40 derniers résidus, <sup>116</sup>KRISPVEESEDVSNKVSMSSTVQGSNIFERTEVLGCEP<sup>155</sup>) a été prédite par dynamique moléculaire (MREMD, *multiplexed replica exchange molecular dynamics*) en utilisant le serveur UNRES (*Czaplewski et al., 2018*). Seule la séquence en acides aminés a été utilisée comme données initiales et l'ensemble des paramètres a été laissé par défaut. Cinq clusters sont générés et une structure représentative pour chaque cluster est donnée. Ces 5 structures représentatives sont classées selon leur probabilité, comprise entre 0 et 1 où 1 représente la structure la plus probable. Le docking par dynamique moléculaire de la structure représentative la plus probable a été réalisé avec UNRES-Dock sans contrainte et en utilisant comme ligand et comme récepteur la même structure. Tous les paramètres ont été laissés par défaut. La modélisation de la totalité de la séquence d'ED-4 a été réalisée localement avec BUNCH (*Petoukhov & Svergun, 2005*). Dans un premier temps, les intensités de diffusion théoriques de la structure représentative des 40 résidus C-terminaux d'ED-4 sous forme monomérique et ne comprenant pas l'étiquette FLAG ont été calculées par CRY SOL (*Svergun et al., 1995*) avec un nombre d'harmoniques de 15, puis elles ont été utilisées dans le programme *pre\_bunch* afin de générer une structure initiale d'ED-4 sous forme monomérique.

Le calcul de BUNCH a été configuré en imposant une symétrie P2 (dimère). Le facteur de pénalité des boucles a été supprimé (valeur de 0 au lieu de 1) étant donné que les séquences sont prédites comme désordonnées. Les résidus de glycine 149, de cystéine 150 et de proline 151 ont été contraints de se situer à moins de 6 Å de leur homologue présent dans la seconde chaîne dans la structure finale. Les données expérimentales de SEC-SAXS collectées précédemment pour le dimère d'ED-4 ont été utilisées avec une valeur de  $s_{\max} \leq 0,4 \text{ \AA}^{-1}$ . Les autres paramètres ont été laissés par défaut. Le squelette peptidique et les chaînes latérales ont été ajoutés par SABBAC (*Maupetit et al., 2006*) et l'énergie du modèle tout atome a été minimisée avec le champs de force FF2 de UNRES (*Czaplewski et al., 2018*). La section efficace de collision de ce modèle a été calculée avec IMPACT selon les méthodes de trajectoire et de projection (*Marklund et al., 2015*).

### III.7 Caractérisation biophysique des ectodomains des syndécans

#### III.7.1 Dichroïsme circulaire

Les expériences de dichroïsme circulaire ont été réalisées sur un spectropolarimètre Chirascan (Applied Photophysics, SFR Biosciences UMS3444/US8, Protein Science Facility, Lyon, France). Les spectres d'ED-1 (150 µg/ml), ED-2 (100 µg/ml), ED-3 (100 µg/ml) et ED-4 (150 µg/ml) en tampon phosphate de potassium ( $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$  10 mM, pH 7,5 filtré sur membrane d'acétate de cellulose 0,2 µm) ont été enregistrés à 20°C dans une cuve en quartz de trajet optique 0,1 cm sur la gamme 260 – 190 nm avec un pas de 0,5 nm, un temps de réponse de 1 s et une bande passante de 0,5 nm. 5 spectres ont été enregistrés par échantillon puis moyennés et le signal du tampon seul a été soustrait. Le signal a été enregistré en millidegrés (mdeg) et converti en ellipticité molaire moyenne par résidu ( $\text{deg.cm}^2.\text{dmol}^{-1}.\text{résidu}^{-1}$ ).

Les expériences de dichroïsme circulaire ont également été réalisées en présence de concentrations croissantes de 2,2,2-trifluoroéthanol (TFE) de 10 à 60 % (v/v) afin de favoriser la formation d'hélices  $\alpha$ . Ces expériences ont également été effectuées en présence de chlorure de guanidinium (ClGu) 3 et 5 M afin d'une part de déstabiliser les structures secondaires existantes et d'autre part de favoriser la formation d'hélices de type polyproline II. Les spectres en présence de ClGu ont été enregistrés sur la gamme 260 – 190 nm mais les données enregistrées sur la gamme 215 – 190 nm n'ont pas été utilisées du fait de la forte absorption de la lumière polarisée circulairement des ions chlorures.

Une série d'expériences de dichroïsme circulaire a été réalisée sur les ectodomains des syndécans (100 µg/ml) en présence des chaperonnes chimiques (oxyde de triméthylamine 0,5 M – TMAO et d'acide 4-phénylbutyrique – 4-PBA 10 mM) dans les mêmes conditions expérimentales que pour les expériences précédentes.

Les expériences portant sur l'effet du pH sur la structure secondaire de l'anastelline et de l'ectodomaine du syndécan-1 ont été réalisées comme suit : les spectres d'ED-1 (6,7 µM) et de l'anastelline (20 µM) en tampon phosphate de potassium 10 mM pH 7,5 ou pH 6,2 ou en tampon acétate de sodium 1 mM pH 5,0 ont été enregistrés à 20°C dans une cuve en quartz de trajet optique 0,1 cm sur la gamme 260 – 190 nm avec un pas de 0,5 nm, un temps de réponse de 1,5 s et une bande passante de 2 nm. 5 spectres ont été enregistrés par échantillon puis moyennés et le signal du tampon seul a été soustrait. Le signal a été enregistré en millidegrés (mdeg) et converti en ellipticité molaire moyenne par résidu ( $\text{deg.cm}^2.\text{dmol}^{-1}.\text{residue}^{-1}$ ).

Les spectres ont été déconvolués en utilisant le serveur Dichroweb (<http://dichroweb.cryst.bbk.ac.uk>, *Whitmore & Wallace, 2008*) avec l'algorithme CDSSTR (*Sreerama and Woody, 2000*) et le jeu de données « 7 » (*Sreerama et al., 2000*). Seul ce jeu de données possède 5 protéines dénaturées chimiquement et se rapproche ainsi de la proportion en structures secondaires des protéines intrinsèquement non structurées.

### III.7.2 Détermination du rayon hydrodynamique des ectodomains des syndécans

#### a) Calculs théoriques du rayon hydrodynamique

Les rayons hydrodynamiques ( $R_h$ ) théoriques des ectodomains des syndécans ont été calculés selon les équations décrites dans la littérature (*Uversky, 2002*). Ces équations empiriques ont été établies à partir d'un jeu de données expérimentales sur les différents états de compaction de protéines globulaires à l'état natif (N) ou à l'état de globule fondu (*molten globule*, MG) et de protéine intrinsèquement non structurées extrêmement dépliées (*coil*, C) ou plus compactes (*premolten globule*, PMG), et sont indiqués ci-dessous, par ordre de degré de compaction :

$$\log R_h(N) = (0,493 \pm 0,008) \times \log(M) - (0,551 \pm 0,032)$$

$$\log R_h(MG) = (0,334 \pm 0,021) \times \log(M) - (0,053 \pm 0,094)$$

$$\log R_h(PMG) = (0,403 \pm 0,012) \times \log(M) - (0,239 \pm 0,055)$$

$$\log R_h(C) = (0,493 \pm 0,008) \times \log(M) - (0,551 \pm 0,032)$$

où la masse moléculaire ( $M$ ) est exprimée en Da et le  $R_h$  en Å.

*b) Détermination du rayon hydrodynamique par chromatographie d'exclusion stérique*

La colonne de filtration sur gel de 24 ml (Superdex S200 Increase 10/300 GL) a tout d'abord été calibrée en tampon HBS en utilisant un mélange de protéines de masse moléculaire et de rayon hydrodynamique (ou rayon de Stokes,  $R_s$ ) connus, excepté pour la conalbumine. 100  $\mu$ l d'un mélange de protéines (kits de calibration GE28-4038-41 et GE28-4038-42, GE Healthcare) diluées à 1,5 mg/ml en HBS ont été injectés par une boucle d'injection de 100  $\mu$ l à un débit de 0,5 ml/min à 10°C (**Tableau IV**). Le volume mort ( $V_0$ ) et le volume total ( $V_t$ ) de la colonne ont été déterminés par l'injection de 100  $\mu$ l de bleu de dextran (environ  $2 \times 10^6$  Da) à 3 mg/ml en tampon HBS et de 100  $\mu$ l d'acétone (environ 58 Da) 1,5 % (v/v) en tampon HBS, respectivement. Le volume interne ( $V_i$ ) a été calculé de la façon suivante :  $V_i = V_t - V_0$  où  $V_i$  représente le volume dans les billes. Les coefficients de distribution ( $K_d$ ) et de partition ( $K_{av}$ ) ont été calculés selon les équations suivantes :

$$K_d = \frac{V_e - V_0}{V_i} \qquad K_{av} = \frac{V_e - V_0}{V_r - V_0}$$

où  $V_e$  correspond au volume d'éluion de la protéine étudiée et  $V_r$  le volume de la phase mobile additionnée au volume de la matrice (soit le volume total de la colonne). La courbe de calibration de la colonne a été obtenue en traçant le coefficient de distribution ( $K_d$ ) en fonction du logarithme (base 10) du rayon hydrodynamique.

<b>Protéines injectées</b>	<b>Masse moléculaire (kDa)</b>	<b>Rayon hydrodynamique (Å)</b>
Ribonucléase A (RNase A)	13,7	16,4
Anhydrase carbonique	29,0	21,0
Ovalbumine	44,0	30,5
Conalbumine	75,0	-
Apoferritine	443,0	61,0
Thyroglobuline	669,0	85,0

**Tableau IV : Protéines utilisées pour la calibration de la colonne de filtration sur gel (Superdex S200 Increase 10/300 GL) ainsi que leurs rayons hydrodynamiques respectifs.** Le domaine de fractionnement de la colonne utilisée s'étend de 10 à 600 kDa pour une protéine globulaire et de 1 à 100 kDa pour du dextran ainsi qu'une limite d'exclusion de 1300 kDa (pour une protéine globulaire). La valeur du rayon hydrodynamique de la conalbumine n'a pas été reportée (-).



Les ectodomains des syndécans ont été injectés à une concentration de 1 mg/ml dans les mêmes conditions que lors de la calibration de la colonne (en tampon HBS à un débit de 0,5 ml/min à 10°C) afin de déterminer leurs rayons hydrodynamiques apparents à partir de l'équation extraite de la courbe de calibration.

#### *c) Diffusion dynamique de la lumière*

Les expériences de diffusion dynamique de la lumière (DLS, *dynamic light scattering*) ont été réalisées avec un Zetasizer Nano ZS (Malvern Instruments, SFR Biosciences UMS3444/US8, Protein Science Facility, Lyon, France) équipé d'un laser He-Ne ( $\lambda = 633$  nm) en collectant la lumière rétrodiffusée à un angle  $\Theta$  de 173° (appelé NIBS, pour *Non-Invasive BackScatter*) après 60 s d'équilibration. La position et l'atténuation du laser ont été déterminées automatiquement par l'appareil. Au moins 11 mesures de 10 s chacune ont été enregistrées à 20°C dans des cuvettes à usage unique (UVette, Eppendorf). Les mesures de diffusion ont été réalisées en tampon HBS seul pour ED-1 (0,68 – 2,72 mg/ml), ED-2 (1,78 mg/ml), ED-3 (3,43 mg/ml) et en HBS  $\pm$  DTT 2 mM pour ED-4 (1,5 mg/ml). Le rayon hydrodynamique a été calculé à partir de la distribution en volume. Les paramètres de mesures ont été laissés par défaut (indice de réfraction des protéines et du solvant de 1,45 et 1,33, respectivement, viscosité du solvant : 1,0031 cP).

#### *d) Diffusion dynamique de la lumière couplée à une chromatographie d'exclusion stérique*

Les ectodomains des syndécans ont été injectés sur une colonne d'exclusion stérique couplée directement à un appareil de DLS afin d'obtenir des mesures de rayons hydrodynamiques précises et de s'affranchir de la diffusion de contaminants ou de poussières. Les intensités de diffusions ont été collectées par un Wyatt QELS+ (ligne de lumière SWING, synchrotron SOLEIL, Gif-sur-Yvette, France) en fonction du temps d'élution. 50  $\mu$ l d'ED-1 (7,7 mg/ml) et d'ED-2 (5,0 mg/ml) ont été injectés sur une colonne de filtration sur gel de 3 ml (Superdex S200 Increase 5/150 GL, GE Healthcare) à un débit de 0,2 ml/min à 20°C et en HBS, et 50  $\mu$ l d'ED-3 (6,5 mg/ml) et d'ED-4 (5,2 mg/ml) ont été injectés sur une colonne de 24 ml (Superdex S200 Increase 10/300 GL, GE Healthcare) à un débit de 0,5 ml/min, 20°C en HBS. Les analyses des données ont été effectuées sur le logiciel Astra avec les paramètres laissés par défaut (index de réfraction : 1,331, viscosité du solvant : 0,992 cP).

### III.7.3 Diffusion de la lumière laser multi-angles

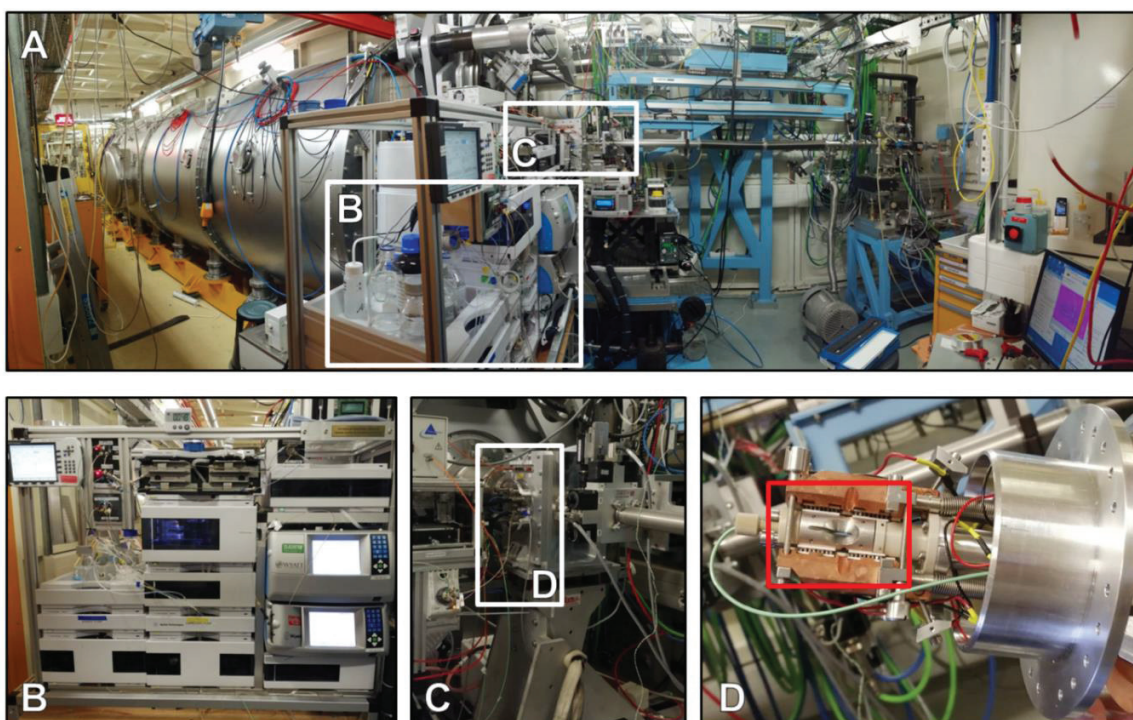
Les mesures de diffusion de la lumière multi-angles couplée à une colonne de chromatographie d'exclusion stérique (SEC-MALLS, *multi-angle laser light scattering*) ont été réalisées sur ligne de lumière SWING (synchrotron SOLEIL, Gif-sur-Yvette, France) avec un détecteur Wyatt Dawn Heleos (pour les mesures d'intensités diffusées) équipé d'un réfractomètre Optilab rEX et d'un spectrophotomètre afin de mesurer les masses moléculaires des protéines injectées. 50 µl d'ED-1 (7,7 mg/ml) et d'ED-2 (5,0 mg/ml) ont été injectés sur une colonne de filtration sur gel de 3 ml (Superdex S200 Increase 5/150 GL, GE Healthcare) à un débit de 0,2 ml/min à 20°C et en HBS, et 50 µl d'ED-3 (6,5 mg/ml) et d'ED-4 (5,2 mg/ml) ont été injectés sur une colonne de 24 ml (Superdex S200 Increase 10/300 GL, GE Healthcare) à un débit de 0,5 ml/min, 20°C en HBS. Les analyses des données ont été effectuées sur le logiciel Astra avec les paramètres laissés par défaut.

### III.7.4 Diffusion des rayons X aux petits angles

#### a) Introduction

Des expériences de diffusion des rayons X aux petits angles couplées à une colonne de chromatographie d'exclusion stérique (SEC-SAXS, *small angle X-ray scattering*) ont été réalisées dans le but de déterminer le rayon de gyration ( $R_g$ ) et la taille maximale ( $D_{max}$ ) des ectodomains des syndécans ainsi que d'obtenir des informations concernant leur degré de compaction et leur flexibilité. Le SAXS est basé sur la diffusion des rayons X par les électrons des molécules présentes dans un échantillon en solution. La génération du faisceau de rayons X est généralement réalisée dans des synchrotrons. Lorsqu'une particule chargée est accélérée, elle émet un rayonnement électromagnétique. Dans le cas des synchrotrons, des électrons sont accélérés successivement dans un accélérateur linéaire puis circulaire jusqu'à une vitesse proche de celle de la lumière ( $> 99,99\%$ ) et sont stockés dans un anneau de stockage où règne un vide poussé ( $10^{-10}$  torr). Cet anneau n'est pas parfaitement circulaire, il s'agit d'une succession de courts segments liés par des sections incurvées où sont situés des aimants de courbures. Ils permettent de dévier le paquet d'électrons relativistes, qui perdent une partie de leur énergie sous forme de photons, c'est le rayonnement synchrotron. Ces photons sont émis dans le domaine des rayons X durs ( $E > 10$  keV,  $\lambda \approx 1$  Å) et de façon tangentielle à l'anneau de stockage. Le rayonnement synchrotron est traité par un jeu de miroirs, collimateurs et monochromateurs dans la cabine d'optique afin de le conditionner avant d'illuminer l'échantillon dans la cabine d'expérience (**Figure 37**).

En fin de ligne se trouve la cabine de contrôle. Les rayons X illuminant l'échantillon sont majoritairement transmis, absorbés et réémis (fluorescence) ou diffusés en perdant (diffusion inélastique) ou non de l'énergie (diffusion élastique), la diffusion élastique étant le phénomène majoritaire. Cependant, il est estimé qu'un seul photon sur 1 million est diffusé par une protéine de 14 kDa concentrée à 1 mg/ml. Les photons transmis sont stoppés en aval de l'échantillon et ceux qui ont été diffusés sont enregistrés sur une certaine gamme angulaire grâce à un détecteur 2D plongé dans un vide poussé afin de s'abstenir de la diffusion des molécules présentes dans l'air (**Figure 37**).



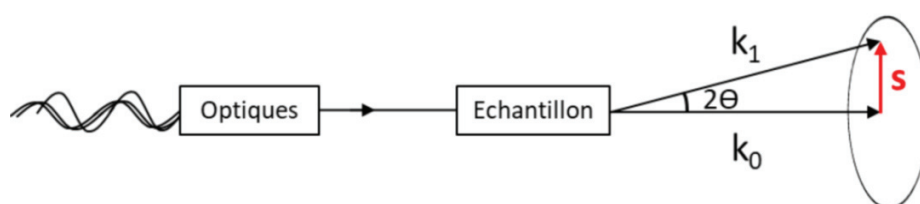
**Figure 37 : Environnement expérimental des expériences de SEC-SAXS.** **A)** Cabine d'expérience de la ligne SWING (SOLEIL). Sur la droite se trouve un fin cylindre où se propagent les rayons X conditionnés dans la cabine optique, et sur la gauche se trouve le tunnel de détection où règne un vide poussé contenant le détecteur qui peut s'y mouvoir. L'échantillon peut être injecté sur une colonne de chromatographie d'exclusion stérique grâce à un équipement de chromatographie en phase liquide haute performance (HPLC) couplé à un appareil de diffusion de la lumière laser multi-angles équipé d'un appareil de diffusion dynamique de la lumière (MALLS-DLS), agrandi en **B)**. L'agrandissement en **C)** montre l'environnement de l'échantillon où on distingue à l'arrière le tunnel de détection. L'échantillon circule dans un capillaire en quartz, encadré et agrandi en **D)**, qui sera illuminé par les rayons X dans ce plan et où des mesures par spectrophotométrie (jusqu'à 8 longueurs d'onde) peuvent être réalisées dans le plan perpendiculaire. Photographies prises avec l'aimable accord d'Aurélien Thureau, scientifique de la ligne de lumière SWING (synchrotron SOLEIL, Gif-Sur-Yvette, France).

b) *Rappels théoriques*

Cette section a été réalisée d'après *Receveur-Bréchet & Durand (2012)*, *Jeffries & Svergun (2015)* et *Jeffries et al., (2016)*. Les intensités de diffusion  $I(s)$  sont collectées en fonction de l'angle de diffusion  $2\Theta$ , représenté par le vecteur de diffusion  $s$  (équation 1) :

$$s = \frac{4\pi \sin \theta}{\lambda} \quad (\text{Equation 1})$$

où  $\Theta$  représente la moitié de l'angle entre le rayon incident et le rayon diffusé et  $\lambda$  la longueur d'onde du faisceau de rayons X incident (**Figure 38**).



**Figure 38 : Représentation schématique de la ligne de lumière.** Les optiques (miroirs, monochromateurs, collimateurs) permettent d'obtenir une onde monochromatique focalisée qui est majoritairement transmise ( $k_0$ ) ou diffusée par l'échantillon ( $k_1$ ). Le vecteur de diffusion  $s$  dépend de la gamme angulaire ( $2\Theta$ ) où  $\vec{s} = \vec{k}_1 - \vec{k}_0$ .

Les expériences de SAXS visent ainsi à obtenir les intensités de diffusion  $I(s)$  propres à l'échantillon en fonction du vecteur de diffusion  $s$ . Les protéines étant en solution, les données de diffusion du tampon et du capillaire sont également collectées et doivent être soustraites par la mesure du tampon seul, qui doit exactement correspondre à celui de l'échantillon. Après soustraction, et considérant que les distances intramoléculaires au sein d'une même particule sont conservées dans le temps et dans l'espace, les intensités de diffusion sont moyennées pour chaque angle permettant la réduction des données bidimensionnelles en 1D, c'est-à-dire le graphique représentant  $I(s)$  en fonction de  $s$ . En considérant une solution monodisperse (*i.e.* pure) et idéale (*i.e.* sans interaction ou répulsion entre les particules la composant), les intensités de diffusion peuvent être décrites comme suivant (équation 2) :

$$I(s) = \sum_i^n [(\Delta\rho_i V_i)^2 P_i(s)] S(s) \quad (\text{Equation 2})$$

où  $\Delta\rho$  et  $V$  représentent respectivement le contraste et le volume de chaque particule  $i$  et  $P(s)$  et  $S(s)$  sont les facteurs de formes et de structures, respectivement.

Le contraste ( $\Delta\rho$ ) représente la différence de densité de longueur de diffusion  $\rho$ , ou en d'autres termes, le nombre moyen d'électrons par unité de volume, entre la protéine et le tampon. Cette différence est très faible (environ 420 électrons / nm<sup>3</sup> pour les protéines et 334 pour l'eau). Un moyen simple d'augmenter le contraste est d'alors d'utiliser un tampon dont la composition est la plus simple possible et d'utiliser une concentration en protéine plus importante. Le facteur de forme  $P(s)$  est directement corrélé à la forme globale de la protéine et reflète ainsi la façon dont les intensités de diffusion décroissent en fonction de  $s$ . Enfin, le facteur de structure  $S(s)$  décrit l'influence sur la diffusion des particules étant à proximité, c'est donc un paramètre reflétant des distances entre les composants de l'échantillon et non pas un paramètre interne à ces derniers.

Dans le cas d'une solution monodisperse et suffisamment diluée, le facteur de structure  $S(s)$  est réduit à l'unité, et les intensités de diffusion de  $N$  particules identiques peuvent être décrites comme suivant (équation 3) :

$$I(s) = N(\Delta\rho V)^2 P(s) \quad (\text{Equation 3})$$

Cette dernière équation montre que l'intensité de diffusion, et donc la qualité des données sur une large gamme angulaire (*i.e.* rapport signal / bruit), dépend directement de la concentration en protéine utilisée (ce qui est reflété par le nombre de molécules  $N$  et le contraste  $\Delta\rho$ ). Le facteur de forme  $P(s)$  reflète la forme globale des molécules dans l'espace réel et caractérise ainsi la façon dont les intensités décroissent en fonction de  $s$ . Les intensités de diffusion des protéines globulaires décroissent en  $s^{-4}$ , c'est-à-dire très rapidement par rapport à des protéines multidomaines ( $s^{-3}$ ) ou des chaînes Gaussiennes ( $s^{-2}$ ). Les intensités de diffusion des protéines très allongées (de type bâton rigide) décroissent en  $s^{-1}$ . Ainsi, plus la conformation globale d'une protéine est allongée, moins les intensités décroissent rapidement en fonction de  $s$  et réciproquement.

L'analyse des intensités de diffusion aux petits angles permet de déterminer le rayon de gyration ( $R_g$ ) grâce à l'approximation de Guinier (équation 4) :

$$\ln I(s) = \ln(I_0) - \frac{(sR_g)^2}{3} \quad (\text{Equation 4})$$

où  $I(0)$  est l'intensité extrapolée à angle zéro.

La pente de la droite dans la représentation graphique  $\ln I(s) = f(s^2)$  équivaut ainsi à  $-R_g/3$  d'où la valeur de  $R_g$  est extraite. Pour respecter cette linéarité, le rayon de gyration doit être déterminé à des petits angles tel que  $s \leq 1,3 / R_g$ .

Le rayon de gyration est un paramètre global directement affecté par la taille et la forme des protéines. Ainsi, une protéine en conformation étendue possédera un rayon de gyration plus important qu'une protéine globulaire de même masse moléculaire. L'approximation de Guinier permet également de déterminer l'intensité à angle zéro utilisée dans la détermination de la masse moléculaire des protéines étudiées. Il est possible d'estimer la valeur du rayon de gyration pour différents degrés de compaction à partir de l'équation de Flory (équation 5) :

$$Rg = R_0 N^\nu \quad (\text{Equation 5})$$

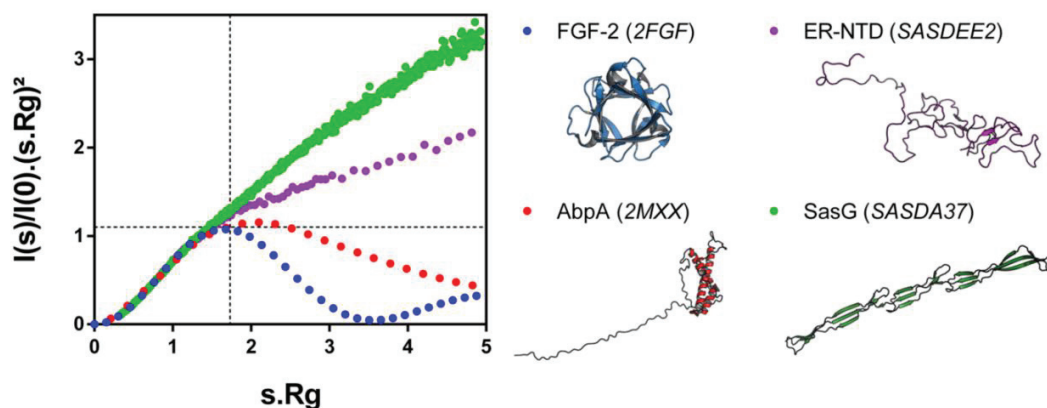
où  $R_0$  et  $\nu$  sont des facteurs définis pour chaque degré de compaction.  $R_0 = \sqrt{3/5} \times 4,75$  et  $\nu = 0,29$  pour des protéines globulaires (Receveur-Bréchet & Durand, 2012),  $R_0 = 2,54 \pm 0,01$  et  $\nu = 0,522 \pm 0,01$  pour des protéines intrinsèquement non structurées et  $R_0 = 1,927 \pm 0,27$  et  $\nu = 0,598 \pm 0,028$  pour des conformations très étendues comme des protéines chimiquement dénaturées (Bernadó & Svergun, 2012). La valeur de  $R_0$  décroît avec la diminution du degré de compaction et indique, qu'en moyenne, les conformations retrouvées dans les protéines intrinsèquement non structurées sont plus compactes que celles retrouvées dans les protéines globulaires dénaturées chimiquement. La valeur de  $\nu$  peut être retrouvée expérimentalement en calculant l'exposant de Porod  $\zeta$ . L'exposant de Porod reflète la manière dont décroissent les intensités de diffusion (de  $s^{-4}$  à  $s^{-1}$ , comme décrit précédemment). L'inverse de l'exposant de Porod ( $1/\zeta$ ) devrait être proche, sinon égal, au facteur  $\nu$  retrouvé dans l'équation de Flory, ce qui permet d'évaluer le degré de compaction d'une protéine avec précision à partir des données de SAXS.

La manière dont décroissent les intensités de diffusion et donc le degré de compaction peut être observée graphiquement en utilisant la représentation de Porod-Debye  $s^4 I(s) = f(s^4)$  pour les protéines globulaires, la représentation de Sibyls  $s^3 I(s) = f(s^3)$  pour les protéines multidomaines et la représentation de Kratky-Debye  $s^2 I(s) = f(s^2)$  pour les protéines intrinsèquement non structurées. Si les intensités décroissent en  $s^{-4}$ ,  $s^{-3}$  ou  $s^{-2}$ , un plateau devrait apparaître sur les représentations de Porod-Debye, de Sibyls et de Kratky-Debye, respectivement. Enfin, la forme globale des protéines peut être évaluée en traçant la représentation de Kratky normalisée par l'intensité à angle zéro et le rayon de gyration selon la relation suivante :

$$\frac{I(s)}{I(0)} (sRg)^2 = f(sRg)$$

Dans cette représentation, une courbe en cloche dont les coordonnées maximales sont (1,732 ; 1,1) reflète un degré de compaction élevé telles que les protéines globulaires, alors que des courbes où ce maximum est de moins en moins marqué jusqu'à évoluer quasi linéairement reflètent des formes globales plus allongées. Une courbe atteignant un plateau aux alentours de 2-3 reflète des conformations très flexibles, ce qui n'est pas nécessairement le cas pour des protéines ayant une représentation normalisée de Kratky croissant linéairement. Cette représentation permet également de comparer sur un même graphique plusieurs protéines de taille ou de conformation différentes.

Afin d'illustrer ces propos, la représentation normalisée de Kratky de quatre protéines possédant différents degrés de compactions a été tracée (**Figure 39**). Les protéines plus ou moins globulaires, ici le facteur de croissance fibroblastique 2 (FGF-2) et de l'*amylase-binding protein A* (AbpA) possèdent une courbe en cloche plus ou moins marquée. En revanche, une courbe correspondant à une protéine désordonnée et flexible, ici le domaine N-terminal du récepteur alpha des œstrogènes (ER-NTD) croît graduellement en atteignant un plateau aux alentours de 2-3. Cependant, la courbe issue d'une protéine dont la conformation globale est un bâton rigide, ici la protéine de surface G de *Staphylococcus aureus* (SasG) croît d'autant plus et est reconnaissable par une certaine linéarité (**Figure 39**).



**Figure 39 : Représentation normalisée de Kratky pour quatre protéines ayant des degrés de compaction différents.** Les données de diffusion pour le facteur de croissance fibroblastique 2 (FGF-2) et de l'*amylase-binding protein A* (AbpA) ont été simulées par *CRY SOL* à partir des fichiers PDB. Les données de diffusion du domaine N-terminal du récepteur alpha des œstrogènes (ER-NTD) et de de la protéine de surface G de *Staphylococcus aureus* (SasG) ont été obtenues depuis la base de données SASDB ([www.sasbdb.org](http://www.sasbdb.org)). Les codes d'accès sont mentionnés. Les lignes en pointillés possèdent les coordonnées (1,732 ; 1,1) et indiquent la position du point maximal de la courbe en cloche pour une protéine globulaire.

Le passage de l'espace réciproque à l'espace réel est effectué par la transformation de Fourier sur les intensités de diffusion, ce qui permet de calculer la fonction de distribution de paires  $P(r)$ . La fonction  $P(r)$  est un histogramme représentant la distribution des distances  $r$  entre tous les points d'une particule. Brièvement, elle est calculée par la fonction d'autocorrélation  $\gamma(r)$  avec  $P(r) = r^2 \cdot \gamma(r)$ . La fonction  $\gamma(r)$  est une fonction strictement décroissante et est définie comme  $\gamma_0(r) = f(r)$  où  $\gamma_0(r)$  est compris entre 0 et 1.  $\gamma_0(r)$  représente la probabilité de trouver un point dans une particule se trouvant à une distance  $r$  d'un autre point. Ainsi, pour  $r = 0$ ,  $\gamma_0(r) = 1$ , c'est-à-dire que la probabilité est maximale de trouver ce point au même endroit (ce qui paraît trivial). Ce qui est plus intéressant est de rechercher pour quelle distance  $r$  la fonction  $\gamma_0(r) = 0$ . Dans la fonction  $P(r)$  associée, ce point correspond à la distance  $r$  (non nulle) qui n'est retrouvée qu'une seule fois. Cette distance est nommée distance maximale des particules ( $D_{\max}$ ) et reflète ainsi la taille maximale de la protéine en solution.

Concernant des solutions polydisperses (présence d'agrégats, d'oligomères ou une protéine intrinsèquement non structurée), cette distance correspond au  $D_{\max}$  de l'espèce la plus allongée et en quantité suffisante pour pouvoir être détectée. La fonction de distribution de paires est notamment utilisée à des fins de modélisations sur des solutions de protéines monodisperses et idéales. Etant donné que, par définition, les protéines intrinsèquement non structurées sont des systèmes polydisperses, ces outils de modélisation ne seront pas détaillés davantage. De par leur flexibilité accrue, les protéines intrinsèquement non structurées ne possèdent pas une unique conformation en solution (l'équation 3 ne s'applique plus). Les intensités de diffusion de chaque conformères s'additionnent alors selon l'équation suivante (équation 6) :

$$I(s) = \sum_{k=1}^K V_k I_k(s) \quad (\text{Equation 6})$$

où  $V_k$  et  $I_k(s)$  sont la fraction volumique et les intensités de diffusion pour chaque conformère  $k$ , respectivement. La fraction volumique est dépendante de la quantité de conformères présents dans une conformation donnée ( $C_k$ ) et de leur volume partiel spécifique ( $\overline{V}_k$ ) selon l'équation suivante (équation 7) :

$$V_k = C_k \overline{V}_k \quad (\text{Equation 7})$$

Bien que la valeur du volume partiel spécifique diffère légèrement selon les protéines, il est admis qu'elle est égale à  $0,7425 \text{ cm}^3 \cdot \text{g}^{-1}$ , ce qui indique que seule la quantité relative des différentes populations de conformations ( $C_k$ ) est responsable des intensités de diffusion collectées sur un tel système.



### c) Déconvolution des intensités de diffusion des systèmes polydisperses (EOM)

Il est possible de déconvoluer les données de diffusion collectées sur des protéines seules (sans modifications post-traductionnelles) afin d'avoir une vue globale de l'espace conformationnel des protéines intrinsèquement non structurées étudiées. Plusieurs approches ont été développées dont l'outil EOM (*Ensemble Optimization Method*) qui a été utilisé dans nos expériences portant sur la caractérisation des ectodomains des syndécans exprimés en système procaryote. Trois étapes régissent le calcul réalisé par EOM (**Figure 40**).

Dans un premier temps, un large *pool* de conformations est généré de façon aléatoire à partir de la séquence en acides aminés de la protéine. Ce pool doit contenir au moins 10 000 conformères afin que la distribution de la distance de bout en bout des polymères générés ne soit pas biaisée aux dépens de conformations compactes ou étendues, et suive une distribution Gaussienne afin de visiter l'espace conformationnel qui lui est théoriquement offert. Il est possible de générer différents *pool* initiaux dont les distributions des angles diédraux correspondent à des conformations retrouvées dans les protéines intrinsèquement non structurées, où de générer des conformations plus étendues de type *random coil*. Des conformations plus compactes peuvent également être générées artificiellement.

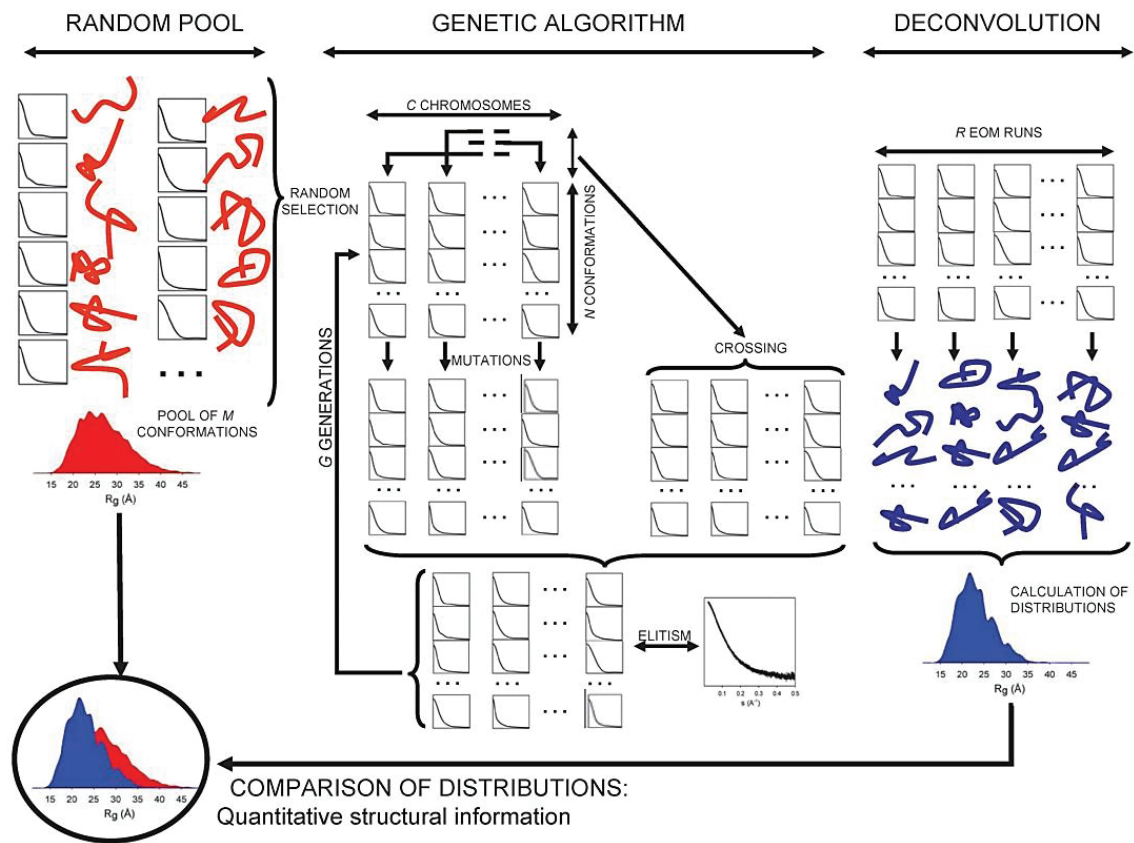
Les intensités de diffusion théoriques sont ensuite calculées par *CRY SOL* pour chaque conformation présente dans le *pool*. Enfin, une sous-sélection de conformations issues du *pool* (appelé *ensemble*) est sélectionnée de telle sorte que les intensités de diffusion additionnées (selon l'équation 6) de cet ensemble soient ajustées aux données de diffusion expérimentales. L'ajustement de la somme des intensités de l'ensemble sélectionné par rapport aux intensités expérimentales est réalisé par le test du  $\chi^2$ . La sélection de ces *ensembles* est réalisée par un algorithme génétique dont l'objectif est de réduire la valeur du  $\chi^2$ . L'ensemble de conformations s'ajustant le mieux aux données expérimentales ( $\chi^2$  le plus faible) est déconvolué afin d'extraire la distribution du rayon de gyration et du Dmax. La comparaison de ces distributions par rapport à celles du *pool* initial de conformations permet d'évaluer l'espace conformationnel de la protéine étudiée ainsi que sa flexibilité (**Figure 40**).

Des paramètres quantitatifs (Rflex et R $\sigma$ ) permettent également d'évaluer la flexibilité des conformères de l'*ensemble* sélectionné. La valeur de Rflex est obtenue en calculant l'entropie du système  $H_b(S)$  qui est comprise entre -1 et 0. Une entropie élevée caractérisant une large distribution de paramètres structuraux mène à une valeur d' $H_b(S)$  proche de -1.

A l'inverse, une distribution plus restreinte possédera une valeur d' $H_b(S)$  proche de 0. Le paramètre  $R_{flex}$  est défini selon l'équation suivante (équation 8) :

$$R_{flex} = -H_b(S) \quad (\text{Equation 8})$$

Les systèmes flexibles possèdent une valeur de  $R_{flex}$  proche de 1, annotée sous forme de pourcentage (100 %). La valeur de  $R\sigma$  permet de calculer la variance de la distribution de l'*ensemble* sélectionné par rapport à celle du *pool*. Une valeur de  $R\sigma > 1$  devrait être retrouvée dans le cas où la valeur de  $R_{flex}$  de l'*ensemble* sélectionné est supérieure à celle du *pool*, et réciproquement. Le calcul de  $R\sigma$  permet ainsi de vérifier la qualité de la déconvolution (Tria et al., 2015).



**Figure 40 : Procédure de déconvolution des intensités de diffusion suivie par l'outil *Ensemble Optimization Method, EOM*.** Figure extraite de Bernadó & Svegún (2012).

EOM génère également la structure 3D des conformères faisant partie de l'*ensemble* sélectionné. Ces conformères sont des modèles de basse résolution où chaque résidu d'acide aminé est représenté par une sphère. Ces modèles permettent de visualiser de façon générale l'étendue et la flexibilité du système.

*d) Collection et traitement des données de diffusion obtenues expérimentalement*

Les intensités de diffusion ont été collectées sur la ligne SWING (synchrotron SOLEIL). La calibration a été réalisée sur l'eau et les protéines ont été centrifugées à 18 000 x g pendant 20 minutes à 4°C avant leur injection sur une colonne de chromatographie d'exclusion stérique précédant leur illumination aux rayons X (**Tableaux Vet VI**).

<b>Acquisition des données de diffusion</b>				
<b>Instrument</b>	Ligne de lumière SWING (Synchrotron SOLEIL)			
<b>Détecteur</b>	Dectris EIGER 4M			
<b>Longueur d'onde (Å)</b>	1,03324			
<b>Energie (keV)</b>	12			
<b>Distance cellule - détecteur (m)</b>	2,0			
<b>Gamme enregistrée (<math>s, \text{Å}^{-1}</math>)</b>	0,0041 – 0,5597			
<b>Expériences de SEC-SAXS réalisées sur les ectodomains des syndécans sans GAGs</b>				
<b>Echantillons</b>	<b>ED-1</b>	<b>ED-2</b>	<b>ED-3</b>	<b>ED-4</b>
<b>Colonne de chromatographie d'exclusion stérique</b>	Superdex S200 Increase 5/150 GL (GE Healthcare)			
<b>Volume injecté (µl)</b>	50			
<b>Concentration en protéine (mg.ml<sup>-1</sup>)</b>	7,7	5,0	14,6	8,3
<b>Débit (ml.min<sup>-1</sup>)</b>	0,2			
<b>Température (°C)</b>	20			
<b>Exposition</b>	Continue : 1 acquisition.s <sup>-1</sup> (990 ms d'exposition + 10 ms de temps mort)			
<b>Tampon</b>	HBS (HEPES 10 mM, NaCl 150mM, pH 7,4)			
<b>Guinier :</b>				
<b><math>s_{\text{max}} \cdot R_g</math></b>	1,29	0,998	1,28	0,985
<b>nombre de points</b>	23	34	24	33

**Tableau V : Conditions expérimentales utilisées pour la collection des données de SEC-SAXS sur les ectodomains des syndécans exprimés en système procaryote.**

Expériences de SEC-SAXS réalisées sur les ectodomains des syndécans avec GAGs			
Echantillons	ED-2 PG	ED-3 PG	ED-4 PG
Colonne de chromatographie d'exclusion stérique	BioSec 5-1000 (Agilent)	BioSec 3-300 (Agilent)	BioSec 5-1000 puis 3-300 (Agilent)
Volume injecté (µl)	50		
Concentration en protéine (µM)	<i>N.D.</i>	124	<i>N.D.</i>
Débit (ml.min <sup>-1</sup> )	0,3		
Température (°C)	20		
Exposition	Continue : 1 acquisition.s <sup>-1</sup> (990 ms d'exposition + 10 ms de temps mort)		
Tampon	HEPES 10 mM, NaCl 300 mM pH 7,4		
Guinier : s <sub>max</sub> .Rg nombre de points	1,29 10	<i>N.D.</i>	1,26 14

**Tableau VI : Conditions expérimentales utilisées pour la collection des données de SEC-SAXS sur les ectodomains des syndécans exprimés en système eucaryote.** Les paramètres d'acquisition des données sont identiques au Tableau V. La détermination du rayon de gyration d'ED-4 PG selon l'approximation de Guinier a été réalisée sur l'expérience où la colonne BioSec 3-300 a été utilisée. *N.D.* : non déterminé.

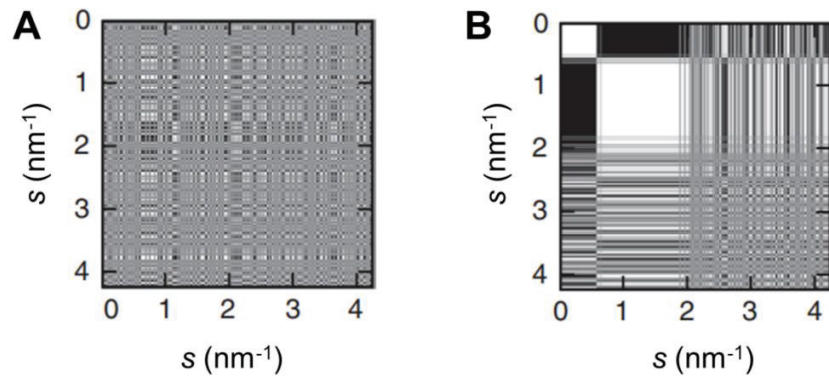
La colonne de chromatographie d'exclusion stérique BioSec 5-1000 (Agilent) de 5 ml possédant une gamme de fractionnement s'étendant de 50 à 750 kDa a été utilisée pour analyser l'ectodomaine du syndécan-2 exprimé en système eucaryote (ED-2PG) puis une Biosec 3-300 (Agilent) de 5 ml possédant une gamme de fractionnement comprise entre 5 kDa et 1 250 kDa a été utilisée pour analyser ED-3PG. ED-4PG a été analysé avec les deux colonnes. Ces colonnes sont disponibles sur la ligne lumière SWING (Synchrotron SOLEIL) et possède un support en silice recouvert d'une couche hydrophile.

La réduction des données de diffusion a été réalisée avec le programme Foxtrot (logiciel de la ligne de lumière SWING). Les données de diffusion du tampon seul ont été enregistrées dans le volume mort des colonnes de chromatographie d'exclusion stérique puis moyennées et soustraites aux données collectées pour chaque échantillon. La présence d'une faible quantité d'espèce de hautes masses moléculaire éluées à un volume d'éluion proche de celui des ectodomains des syndécans a nécessité un traitement supplémentaire des données de diffusion. Ce traitement a été effectué avec le module *US-SOMO HPLC SAXS* (Brookes et al., 2016).

Brièvement, ce module permet de convertir les intensités de diffusion  $I(s) = f(s)$  en intensité de diffusion en fonction du temps  $I(t) = f(t)$  et ce pour chaque valeur du vecteur de diffusion  $s$  (donc pour chaque angle). Il en résulte un graphique ressemblant à un chromatogramme où sont superposées toutes les intensités mesurées en fonction du temps et pour chaque valeur de  $s$ . Il est alors possible de visualiser si une ou plusieurs espèces sont à l'origine des intensités de diffusion mesurées en ajustant des courbes Gaussiennes sous ces courbes. Typiquement, les données  $I(t) = f(t)$  pour un seul angle sont d'abord sélectionnées (entre 0,01 et 0,02 Å<sup>-1</sup> afin d'avoir un signal intense et peu bruité) et une ou plusieurs courbes Gaussiennes sont ajustées (la qualité de l'ajustement est indiquée par la valeur du  $\chi^2$ ). Ces courbes Gaussiennes sont ensuite étendues à la totalité de la gamme angulaire. Finalement, les données propres à chaque courbe Gaussienne sont reconverties en intensité de diffusion  $I(s) = f(s)$  et la participation d'autres espèces à la diffusion de la protéine étudiée est ainsi éliminée.

Les données  $I(s) = f(s)$  propres à la protéine étudiée ont été comparées avec le test *CorMap* (Franke et al., 2015) présent dans la suite *PRIMUSqt* (Franke et al., 2017). Le test *CorMap* (pour *correlation map*) est un test statistique très restrictif. A l'inverse du  $\chi^2$  qui évalue la similarité entre deux courbes de façon globale, le test *CorMap* est très sensible aux différences locales et ne nécessite pas d'indiquer les erreurs sur la mesure, contrairement au  $\chi^2$ . Le test *CorMap* calcule les probabilités de similarité (des corrélations) entre deux courbes. Si cette probabilité (indiquée par une *p-value*) est inférieure à une valeur seuil ( $\alpha = 0,01$ ), les deux courbes sont considérées comme statistiquement différentes. Ce test peut être visualisé sur une carte où les valeurs de corrélations de deux jeux de données sont représentées avec un code binaire *oui / non*, imagé par l'alternance d'aires noires et blanches (**Figure 41**).

Deux jeux de données  $I(s) = f(s)$  similaires sont caractérisés par une carte où les valeurs de corrélations sont distribuées de façon homogène (**Figure 41A**), et à l'inverse, ils sont différents si ces corrélations forment de larges aires blanches ou noires (**Figure 41B**). La localisation de ces aires indique la gamme angulaire où les deux courbes diffèrent de façon significative.



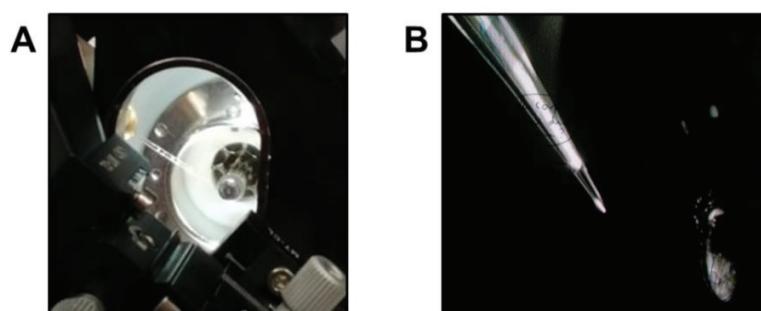
**Figure 41 : Exemples de cartes de corrélation (*CorMap*)** correspondant à des courbes de diffusion similaires **(A)** ou différentes **(B)** où une déviation importante est observée dans la gamme angulaire s'étalant de 0 à environ  $2 \text{ nm}^{-1}$ . D'après *Franke et al.*, (2015).

Au moins 29 acquisitions ont été moyennées pour chaque ectodomaine des syndécans après l'élimination des courbes non superposables d'après le test *CorMap*. Les quelques points différent de l'approximation de Guinier (*i.e.* équation 4) aux petits angles ont été supprimés puis la courbe résultante a été soumise au programme *SHANUM* (*Konarev & Svergun, 2015*). Cet outil a été développé afin de définir la gamme angulaire où les données de diffusion contiennent de l'information structurale afin de s'affranchir notamment des données bruitées collectées aux grands angles, dépourvues d'information structurale. Il en résulte une courbe de diffusion définitive utilisée pour les futures analyses. Les données de diffusion ont été déconvoluées avec EOM. La séquence en acides aminés des protéines recombinantes a été utilisée afin de générer deux *pools* initiaux contenant 10 000 conformères ayant une distribution de leurs angles diédraux cohérente avec des conformations plus ou moins compactes annotées respectivement comme *pré-globule fondu* (PMG) ou *random coil* (RC) dans ce manuscrit. Le nombre d'harmoniques a été laissé par défaut (15) et 100 cycles ont été réalisés par l'algorithme génétique.

### III.7.5 Spectrométrie de masse en conditions natives et mobilité ionique

#### a) Rappels sur l'ionisation par électronébulisation

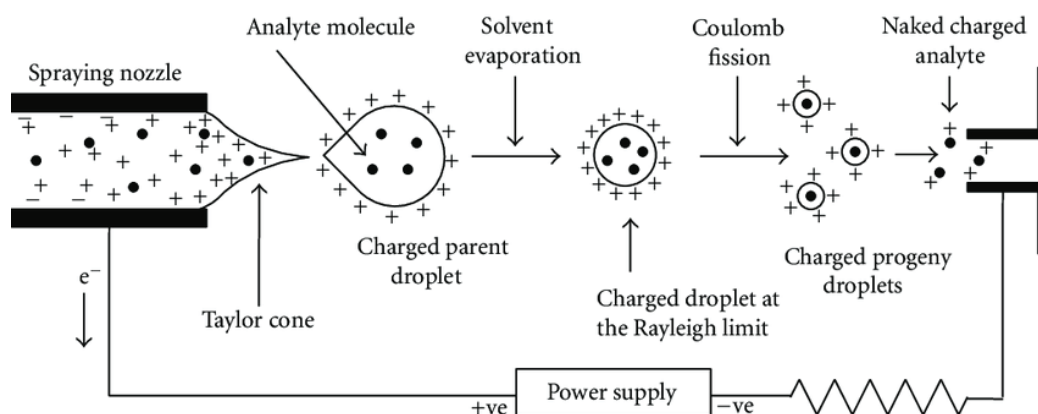
Des expériences d'ESI-MS (*electrospray ionization – mass spectrometry*) et de mobilité ionique (ESI-IM-MS ou IM-MS) ont été réalisées sur les ectodomains des syndécans exprimés en système procaryote afin de caractériser leur hétérogénéité conformationnelle. Tout comme l'APCI (ionisation chimique à pression atmosphérique) et la CSI (*cold spray ionization*), l'ionisation par électronébulisation (ESI) permet l'analyse d'échantillon liquide à pression atmosphérique, mais possède cependant la gamme de masse analysable la plus étendue (jusqu'à environ 250 kDa). Cette source d'ionisation est commune aux appareils d'ESI-MS classique et de mobilité ionique utilisés dans nos expériences. Une source nano-ESI permet d'utiliser un plus faible volume d'échantillon (**Figure 42A**) où la solution de protéine est placée dans une aiguille avec un filament en nickel et rapprochée de l'entrée du spectromètre (**Figure 42B**).



**Figure 42 : Source Nano-ESI d'un spectromètre de masse à mobilité ionique. A)** Vue générale de la source Nano-ESI du spectromètre à mobilité ionique utilisée pour nos expériences. On distingue l'aiguille dans laquelle se trouve la solution de protéine. **B)** Agrandissement de l'aiguille contenant l'échantillon où on observe le filament en nickel. A droite se trouve l'orifice d'entrée du spectromètre. Photographies prises avec l'aimable accord de Fabien Chirot, responsable de l'appareil de mobilité ionique (Institut des Sciences Analytiques).

Sous l'effet d'un champ électrique (3 – 5 kV avec une source ESI, et environ 800 V avec une source Nano-ESI), la solution à analyser va se déformer à l'extrémité de l'aiguille, formant un cône de Taylor. Des gouttes vont être formées et les protéines qui y sont présentes vont s'ioniser en capturant des protons du solvant (**Figure 43**). Le solvant s'évapore progressivement jusqu'à que la limite de Rayleigh soit atteinte engendrant une explosion Coulombienne. Il en résulte des gouttelettes chargées où les molécules de solvant résiduelles vont être éliminées grâce à un gaz de séchage, typiquement du diazote, ce qui mène à la formation d'analytes chargés et désolvatés (**Figure 43**).

Ce type d'ionisation mène à la génération d'ions multichargés. Le solvant doit être volatil et ne pas perturber la structure tertiaire ou quaternaire des protéines. On parle alors de mode *natif*. Une solution d'acétate d'ammonium 50 mM pH 7,4 est la plus souvent utilisée lors de ces expériences.



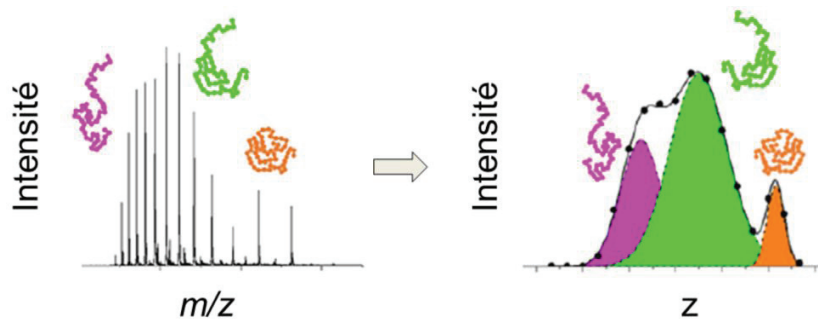
**Figure 43 : Mécanisme d'ionisation des protéines par électroébulisaison (ESI).** Figure extraite de *Banerjee & Mazumdar (2012)*.

#### *b) Acquisition des données d'ESI-MS native*

L'objectif des expériences d'ESI-MS en conditions natives est d'obtenir la distribution des états de charges (CSD) des protéines analysées, c'est-à-dire un graphique représentant l'intensité de chaque ion en fonction de son rapport masse sur charge ( $m/z$ ). Dans le cas des protéines intrinsèquement non structurées, le degré d'ionisation par électroébulisaison est corrélé à la surface des conformères présents en solution au moment de l'analyse. Les conformations les plus étendues acquerront davantage de charges que des conformations plus compactes (*Testa et al., 2013, Schramm et al., 2019*).

La distribution des états de charges des protéines globulaires est généralement restreinte et unimodale. A l'inverse, celle des protéines intrinsèquement non structurées est multimodale où plusieurs enveloppes peuvent être détectées, correspondant à la présence de sous-populations de conformations. La distribution des états de charge peut être déconvoluée afin de définir l'abondance relative de chaque sous-population en représentant l'intensité de chaque état de charge  $z$  en fonction de l'état de charge,  $I(z) = f(z)$ . Des courbes Gaussiennes sont ensuite ajustées sous chaque population observées dans cette nouvelle représentation (*Dobo & Kaltashov, 2001, Santambrogio et al., 2018*) (**Figure 44**).





**Figure 44 : Déconvolution de la distribution des états de charges obtenue par ESI-MS.** La déconvolution des CSDs est réalisée en modifiant la représentation de  $I(m/z) = f(m/z)$  (à gauche) en  $I(z) = f(z)$  puis en ajustant des courbes Gaussiennes sous chaque population observée dans cette nouvelle représentation (à droite). Adaptée de *Santambrogio et al.*, (2018).

Cette représentation permet également de déterminer la proportion ( $P_i$  en pourcentage) de chaque sous-population détectée en calculant le rapport de l'aire d'une courbe Gaussienne pour une population ( $A_i$ ) sur l'aire totale, c'est-à-dire la somme des aires des courbes Gaussiennes ajustées (équation 9) :

$$P_i = \frac{A_i}{\sum A} \times 100 \quad (\text{Equation 9})$$

La déconvolution de la distribution des états de charge permet d'évaluer l'hétérogénéité conformationnelle des protéines intrinsèquement non structurée mais uniquement de façon relative en observant par exemple qu'une population est plus chargée qu'une autre ce qui reflète des conformations plus étendues. Un moyen d'obtenir des valeurs numériques caractérisant le degré de compaction de ces populations est de calculer la surface accessible au solvant (SASA) exprimée en  $\text{\AA}^2$ . Le calcul de cette valeur est basé sur la charge moyenne acquise par les conformations ( $z_{av}$ ), l'intensité de chaque état de charge  $z$  et des équations empiriques issus de jeux de données. La charge moyenne en prenant en compte la totalité du CSD ou d'une population précise est calculée selon l'équation suivante (équation 10) :

$$z_{av} = \sum_{i=A}^B z_i I_i / I_{tot} \quad (\text{Equation 10})$$

où  $A$  et  $B$  représentent l'état de charge sélectionné le plus petit et le plus grand, respectivement, et  $I_i$  et  $I_{tot}$  sont les intensités d'un état de charge  $i$  donné et la somme des intensités des états de charges de la gamme sélectionnée, respectivement.

La valeur de SASA peut être calculée sur la totalité de la gamme des états de charge afin d'obtenir une valeur de moyenne de SASA ( $SASA_{av}$ ) reflétant l'ensemble des conformations adoptées par une protéine intrinsèquement non structurée (*Li et al., 2016*) selon l'équation suivante (équation 11) :

$$\ln(z_{av}) = 0,913 \times \ln(SASA_{av}) - 6,013 \quad (\text{Equation 11})$$

La valeur de SASA peut également être calculée uniquement pour les conformations les plus étendues ( $SASA_{ext}$ ) en sélectionnant les états de charge appartenant à la population la plus ionisée (*Testa et al., 2011*) selon l'équation suivante (équation 12) :

$$\ln(z_{av}) = 0,9024 \times \ln(SASA_{ext}) - 5,9013 \quad (\text{Equation 12})$$

Les expériences de spectrométrie de masse en conditions natives ont été effectuées sur un spectromètre Impact II (Bruker) équipé avec une source ESI Apollo II (Bruker) au Centre Commun de Spectrométrie de Masse de l'Institut (ICBMS, UMR5246) par Alexandra Berlioz-Barbier, responsable de la plateforme (**Figure 45**).



**Figure 45 : Spectromètre de masse utilisé lors des expériences d'ESI-MS en conditions natives.** L'ionisation des protéines a été réalisée par électroébulisation (ESI classique). Les rapports  $m/z$  ont été calculés grâce à un analyseur à temps de vol (TOF) visible sur la gauche (la *cheminée* bleue). L'analyse peut être couplée à une HPLC (à droite) du moment que le tampon reste volatil. Photographie prise avec l'aimable accord d'Alexandra Berlioz-Barbier, responsable du centre Commun de Spectrométrie de Masse de l'Institut (ICBMS, UMR5246)

Le tampon HBS a préalablement été échangé contre une solution d'acétate d'ammonium 50 mM pH 7,4 par chromatographie d'exclusion stérique (Superdex S200 Increase 10/300 GL, GE Healthcare) puis les fractions d'élutions contenant les ectodomaines des syndécans exprimés en système procaryote ont été concentrées comme décrit précédemment. 200  $\mu$ l d'ED-1 (600  $\mu$ g/ml), ED-2 (590  $\mu$ g/ml), ED-3 (420  $\mu$ g/ml), et ED-4 (180  $\mu$ g/ml) ont été infusés à un débit de 10  $\mu$ l/min par un pousse-seringue. Ce spectromètre est équipé d'un analyseur quadripolaire et d'un analyseur à temps de vol (TOF) en série.

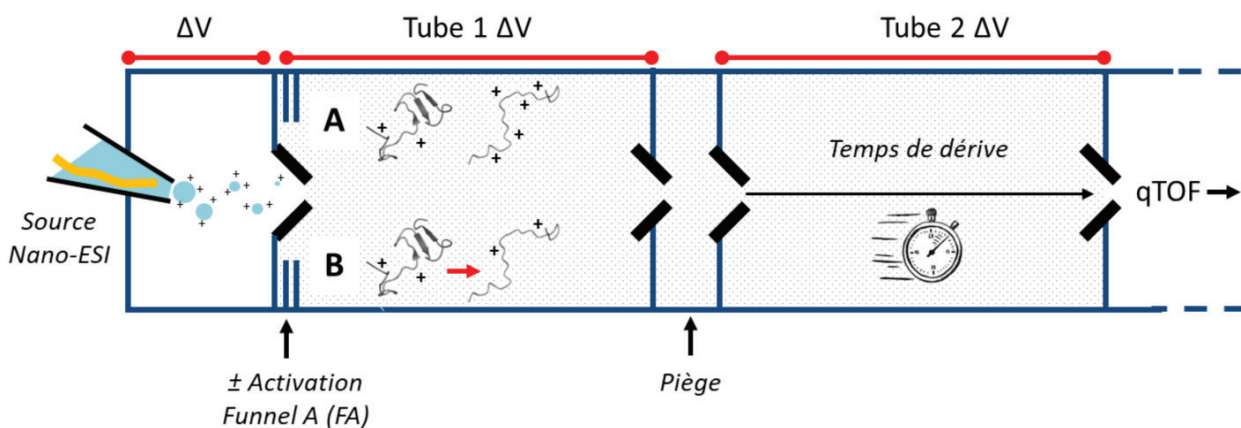
Dans ces expériences, l'analyseur quadripolaire n'a été utilisé que pour guider et focaliser les analytes après leur ionisation étant donné qu'aucune étape de sélection ne doit être appliquée. Le TOF a en revanche permis de calculer les rapports masse sur charge ( $m/z$ ) afin d'obtenir la distribution des états de charges des ectodomains des syndécans. Les paramètres de l'instrument ont été fixés comme suit : température de la source ESI : 200 °C, tension du capillaire : 3 kV, pression du gaz de nébulisation : 41 Psi, fréquence d'acquisition : 2 Hz. L'ensemble des expériences a été réalisé dans des conditions expérimentales identiques afin de pouvoir comparer les ectodomains des syndécans entre eux. Le CSD des protéines a été déconvolué en ajustant des courbes Gaussiennes comme décrit précédemment avec le logiciel Origin (OriginLab) avec le minimum de courbes possible et jusqu'à ce que la qualité de l'ajustement atteigne au moins  $R^2 = 0,99$ . L'abondance relative des sous-populations détectées et la valeur de  $SASA_{av}$  et de  $SASA_{ext}$  ont été calculées comme décrit précédemment (équations 9 – 12).

### *c) Acquisition des données de mobilité ionique*

Ces expériences ont été réalisées en collaboration avec Fabien Chirot (Institut des Sciences Analytiques - ILM, Villeurbanne) et Luke MacAleese (Institut Lumière Matière - ISA, Villeurbanne) avec l'aide de Mathilde Bouakil (ILM) et d'Aurélien Le Fèvre (ISA) sur un spectromètre à mobilité ionique au sein de l'ISA (*Simon et al., 2015*). L'équipement nécessaire aux expériences de mobilité ionique est assez similaire à celui utilisé lors des expériences d'ESI-MS. Dans nos expériences, le spectromètre (Maxis Impact, Bruker) se composait notamment d'une source Nano-ESI (voir Figure 42), de deux analyseurs (quadripolaire et à temps de vol) et d'un détecteur. Deux tubes de dérivation de 79 cm de long chacun étaient positionnés en série entre la source et l'analyseur quadripolaire et contenaient de l'hélium à une pression d'environ 4 Torr ( $\approx 500$  Pa). Une cinquantaine de microlitres d'ectodomains des syndécans exprimés en système procaryote et dilués à une concentration de 10  $\mu$ M dans une solution d'acétate d'ammonium 50 mM pH 7,4 ont été placés dans l'aiguille de la source Nano-ESI.

Après leur nébulisation, les protéines ionisées et désolvatées sont guidées dans le premier tube de dérivation par un champ électrique appliqué au niveau de l'entrée du premier tube (*funnel A* – FA, **Figure 46**). La tension appliquée sur le FA est modulable et est fixée à la valeur la plus faible possible (20 V) afin de limiter l'énergie cinétique des ions et d'éviter ainsi des forces de collisions trop importantes entre les protéines et les atomes d'hélium (FA20). Elle peut être augmentée pour favoriser au contraire ces collisions, provoquant le dépliement des protéines, on parle alors d'*activation*.

La totalité des ions est ensuite piégée juste avant le second tube, ce-dernier permettant de mesurer leur temps de dérive. Les ions sont relâchés au même moment du piège et les protéines voyagent ensuite grâce à l'application d'un champ électrique à travers ce second tube (**Figure 46**). Leur temps de dérive est dépendant de leur charge et de leur conformation du fait de leur collision avec les atomes d'hélium. Les protéines ayant des conformations compactes ont une probabilité plus faible de collisionner que celles ayant des conformations plus étendues. D'un point de vue statistique, une molécule d'analyte collisionne avec un seul atome d'hélium à la fois.



**Figure 46 : Schéma général montrant les principales caractéristiques du spectromètre à mobilité ionique utilisé dans nos expériences. A)** Après leur ionisation (source Nano-ESI), le mélange d'ions peut être transféré sans activation (FA20), piégé puis transféré jusqu'au deuxième tube où leur temps de dérive est mesuré. **B)** Les ions peuvent également être activés ce qui engendre des conformations plus étendues (flèche rouge), piégés, puis transférés de la même façon jusqu'au second tube de dérive.

On peut ainsi obtenir la distribution des temps d'arrivée (ATD) de l'ion analysé, c'est-à-dire l'intensité de l'ion en fonction de son temps de dérive (en ms), où sa distribution suit une loi gaussienne dont la courbe est plus ou moins large selon l'hétérogénéité conformationnelle des espèces portant la même charge et selon l'ampleur de sa diffusion dans le tube de dérive. Afin de prendre en compte les paramètres d'ionisation et de conformations de chaque analyte, la valeur de la section efficace de collision (CCS,  $\Omega$ ) a été calculée pour chaque état de charge ( $z$ ) selon l'équation suivante (équation 13) :

$$\Omega = \frac{3}{16} \times \frac{ze}{N} \times \sqrt{\frac{1}{m} + \frac{1}{M}} \times \sqrt{\frac{2\pi}{k_B T}} \times \frac{1}{K} \quad (\text{Equation 13})$$

où  $K$  est la mobilité de l'ion dans le tube,  $e$  la charge élémentaire,  $N$  le nombre d'atomes d'hélium dans le tube,  $m$  et  $M$  la masse de l'hélium et de l'ion, respectivement,  $k_B$  la constante de Boltzmann et  $T$  la température dans le tube. La mobilité  $K$  de l'ion est le seul paramètre inconnu dans le calcul de la CCS, et est défini comme (équation 14) :

$$t = t_0 + \frac{L}{K} \times \frac{1}{E} \quad (\text{Equation 14})$$

où  $t$  est le temps total entre l'injection des ions dans le tube de dérive et leur détection, et  $t_0$  le temps de dérive des ions entre la sortie du tube de dérive et le détecteur.  $L$  représente la longueur du tube de dérive et  $E$  le champ électrique appliqué dans le tube. Cette équation est du type  $y = ax + b$ . En modulant uniquement la valeur du champ électrique  $E$ , il est alors possible de déterminer la valeur de la mobilité  $K$  de l'ion en traçant la représentation linéaire :

$$t = f\left(\frac{1}{E}\right)$$

où la valeur de  $K$  est déduite du coefficient directeur de la droite. Dans nos expériences, 5 tensions différentes ont été appliquées (250, 350, 400, 450 et 550 V) afin de déterminer précisément la valeur des sections efficaces de collision de chaque ion. Afin d'éviter des biais expérimentaux liés à l'accumulation d'énergie, l'application des différentes tensions dans le tube n'a pas été croissante mais aléatoire.

Les mesures de stabilité des conformations les plus compactes ont été réalisées dans les mêmes conditions que les expériences précédentes, et 6 valeurs d'activation (FA) ont été appliquées (20, 50, 80, 120, 150 et 190 V). La tension dans le tube 2 a été fixée à 550 V et n'a pas été modulée, en conséquence les CCS n'ont pas été calculées dans cette série d'expériences. La stabilité conformationnelle des conformations les plus compactes a été déterminée en calculant la variation (en %) du temps de dérive mesuré à FA190 ( $t_{d190}$ ) par rapport à celle à FA20 ( $t_{d20}$ ) selon (équation 15) :

$$\text{Variation (\%)} = \frac{t_{d190} - t_{d20}}{t_{d20}} \times 100 \quad (\text{Equation 15})$$

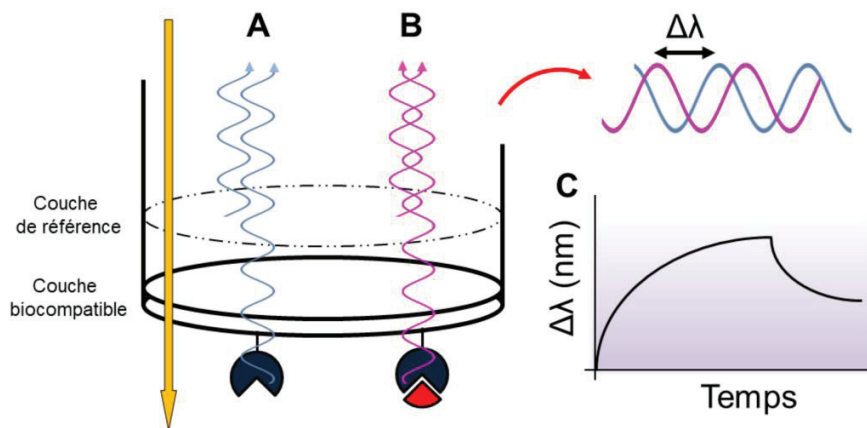
La variation de temps de dérive est inversement proportionnelle à la stabilité d'une conformation.

## III.8 Etude des interactions biomoléculaires

### III.8.1 Interférométrie de biocouche

#### a) Rappels théoriques

L'interférométrie de biocouche (BLI) est une technique de détection et de caractérisation des interactions biomoléculaires en temps réel et sans marquage où l'un des partenaires est immobilisé à la surface d'un biocapteur (*ligand*) et l'autre est en solution (*analyte*, *Shah & Duncan, 2014*). La technique est basée sur le phénomène d'interférométrie, où une source lumineuse (lumière blanche) est divisée en un premier faisceau illuminant une couche biocompatible à l'extrémité du biocapteur sur laquelle est immobilisé le ligand, et un second faisceau illuminant une référence interne (**Figure 47**). Les deux faisceaux sont réfléchis et interfèrent de façon constructive ou destructive à différentes longueurs d'onde et le motif d'interférence est détecté par une caméra CCD (*charge coupled device*). Lorsque l'analyte interagit avec le ligand, le motif d'interférence du faisceau réfléchi est décalé vers des longueurs d'ondes plus importantes (*blue shift*) du fait de l'augmentation de l'épaisseur du trajet optique. Ce décalage (*shift*) est enregistré au cours du temps et est présenté sous forme d'un sensorgramme (**Figure 47**).



**Figure 47 : Principe de l'interférométrie de biocouche.** En absence de variation d'épaisseur à la surface du capteur, le phénomène d'interférence ne se produit pas (**A**), mais les ondes interfèrent lorsque l'épaisseur du trajet optique varie, comme dans lors de l'interaction entre un ligand (bleu) et un analyte (rouge) (**B**). Ce décalage de longueur d'onde ( $\Delta\lambda$ ) est mesuré et tracé en fonction du temps afin d'obtenir un sensorgramme (**C**). Flèche jaune : lumière incidente, ondes violette et grise : faisceaux réfléchis.

Seules les molécules interagissant avec la surface du biocapteur provoquent une modification du motif d'interférence. Une modification conformationnelle importante du ligand à la surface du biocapteur peut également altérer l'épaisseur de la biocouche et provoquer une variation du motif d'interférence, induisant une modification du décalage de longueur d'onde (*shift*) enregistré. L'utilisation de la BLI permet de suivre les interactions en temps réel et d'en déduire les constantes cinétiques d'association ( $k_a$ , en  $M^{-1}.s^{-1}$ ) et de dissociation ( $k_d$ , en  $s^{-1}$ ) à partir desquelles est calculée la constante d'équilibre de dissociation ( $K_D$ , en M) reflétant l'affinité de l'analyte envers son ligand. La détermination avec précision de ces paramètres nécessitent d'une part l'utilisation de différentes concentrations en analyte et d'autre part qu'au moins 10 % des complexes se dissocient lors de la phase de dissociation. Dans le cas contraire, il est possible d'extraire la valeur de  $k_{obs}$  (en  $s^{-1}$ ) qui est une constante de vitesse observée reflétant le phénomène dynamique d'association - dissociation lors de la formation du complexe. C'est une constante de vitesse reflétant les changements de vitesse d'association en fonction de la diminution de la concentration en ligand libre (*Wilson et al., 2010*). La valeur de  $k_{obs}$  ne peut être déduite que de la phase d'association.

#### *b) Crible des partenaires extracellulaires des ectodomains des syndécans*

Les expériences d'interférométrie de biocouche ont été réalisées sur un système Octet RED96 (FortéBio, Molecular Devices). Les protéines criblées ont été immobilisées de façon covalente par couplage amine sur un biocapteur AR2G (*Amine Reactive Second-Generation*, FortéBio, Molecular Devices) en ajustant le pH de la solution des ligands ( $pH < pHi$ ) afin de favoriser leur immobilisation (**Tableau VII**). Le couplage amine implique la formation d'esters de N-hydroxysulfosuccinimide (s-NHS) réagissant d'une part avec les fonctions carboxyliques présentes à la surface du capteur et d'autre part avec les fonctions amines primaires présentes dans le ligand à immobiliser. Le biocapteur AR2G a été activé de façon identique pour chaque protéine criblée et cette étape a été réalisée dans une plaque 96 puits noir avec 200  $\mu$ l de solution. Le biocapteur a été équilibré au moins 10 minutes dans de l'eau distillée afin d'éliminer le saccharose utilisé pour son stockage. Il a ensuite été trempé dans de l'eau distillé pendant 60 s, puis dans une solution contenant du s-NHS 20 mM et du 1-Ethyl-3-(3-diméthylaminopropyl) carbodiimide (EDC) 40 mM dilués dans de l'eau distillée pendant 300 s. Le capteur a ensuite été trempé dans la solution de protéine criblée pendant 300 s (**Tableau VII**) puis les groupements activés n'ayant pas réagi ont été neutralisés dans une solution d'éthanolamine 1 M pH 8,5 diluée en eau distillée pendant 300 s.

Enfin, les capteurs activés ont été trempés successivement dans deux puits contenant du tampon HBS-P20 0,05 % (v/v) pH 7,4 pendant 180 s chacun (le P20 étant un détergent), et stockés ainsi jusqu'aux expériences d'interactions réalisées moins d'une dizaine de minutes après l'immobilisation des protéines criblées. Toutes les protéines criblées sont d'origine humaine (**Tableau VII**).

Ligands	Référence	Concentration (µg/ml)	Tampon d'immobilisation
α-synucléine	SP480 Boston Biochem	100	Acétate de sodium 10 mM pH 3,0
Anastelline (Fragment C-terminal du 1 <sup>er</sup> domaine FNIII de la fibronectine)	F3542 Sigma-Aldrich	100	Maléate de sodium 10 mM pH 6,2
Collagène IV	C7521 Sigma-Aldrich	250	Maléate de sodium 10 mM pH 5,0
Domaine NCI du collagène VIII	Exprimé au laboratoire	100	Acétate de sodium 10 mM pH 4,0
Ectodomaine du collagène XVII	Exprimé au laboratoire	250	Maléate de sodium 10 mM pH 5,0
Ectodomaine du collagène XXIII	Exprimé au laboratoire	100	Maléate de sodium 10 mM pH 5,0
Ectodomaine du récepteur à domaine discoïdine de type 1 / DDR-1	2396-DR R&D Systems	50	Acétate de sodium 10 mM pH 5,0
Ectodomaine du récepteur à domaine discoïdine de type 2 / DDR-2	2538-DR R&D Systems	20	Acétate de sodium 10 mM pH 5,0
Endostatine (Fragment C-terminal du collagène XVIII)	Exprimée au laboratoire	128	Maléate de sodium 10 mM pH 6,2
Fibronectine cellulaire	F2518 Sigma-Aldrich	100	Maléate de sodium 10 mM pH 6,2
Facteur de croissance fibroblastique 2 / FGF-2	C60240 Promokine	50	Maléate de sodium 10 mM pH 6,2
<i>Leukocyte-associated immunoglobulin-like receptor 1</i> LAIR-1	2664-LR R&D Systems	50,25	Acétate de sodium 10 mM pH 4,0
<i>Lysyl oxidase like-2</i> LOX-L2	2639-A0 R&D Systems	257	Acétate de sodium 10 mM pH 4,0

**Tableau VII (1/2) : Protéines immobilisées lors des cribles d'interactions réalisés par interférométrie de biocouche.**



Ligands	Référence	Concentration (µg/ml)	Tampon d'immobilisation
Ectodomaine de MARCO / <i>Macrophage receptor with collagenous structure</i>	7586-MA R&D Systems	62,5	Maléate de sodium 10 mM pH 6,2
Ectodomaine de la Neuropiline 1 / NRP1	Exprimé au laboratoire	50	Acétate de sodium 10 mM pH 5,0
Propeptide de la lysyl oxydase PPLOX	Exprimé au laboratoire	100	Maléate de sodium 10 mM pH 6,2
Ectodomaine du <i>Tumor endothelium marker 8</i> / TEM-8	H00084168-P01 Abnova	17	HBS-P20 0,05 % (v/v)
Facteur de Nécrose Tumorale alpha / TNF-α	210-TA R&D Systems	50	Acétate de sodium 10 mM pH 5,0
Facteur de croissance de l'endothélium vasculaire / VEGF	V7259 Sigma-Aldrich	12,5	Maléate de sodium 10 mM pH 6,2
Ectodomaine du récepteur de type 2 au facteur de croissance de l'endothélium vasculaire / VEGFR-2	357-KD R&D Systems	250	Maléate de sodium 10 mM pH 6,2

**Tableau VII (2/2) : Protéines immobilisées lors des cribles d'interactions réalisés par interférométrie de biocouche.**

Les expériences d'interaction ont été réalisées en trempant les capteurs sur lesquels ont été immobilisées les protéines cibles dans deux puits consécutifs contenant 200 µl de tampon HBS-P20 0,05 % (v/v) pH 7,4 pendant 60 s puis pendant 300 s dans 200 µl de solutions d'ectodomains des syndécans exprimés en système procaryote purifiés. Les ectodomains des syndécans ont systématiquement été utilisés en tant qu'analytes afin de conserver leur caractère flexible. Ils ont été dilués à une concentration de 500 nM dans le tampon HBS-P20 0,05 % (v/v). L'étape de dissociation a été effectuée pendant 300 s dans le tampon HBS-P20 0,05 % (v/v). Puis le capteur a été stocké dans le même tampon. La totalité des expériences a été réalisée à 25 °C et sous agitation (1000 rpm).

Un capteur de référence a systématiquement été activé dans les mêmes conditions que décrit précédemment mais en absence de ligand. Le sensorgramme de référence obtenu par immersion dans l'analyte a été soustrait à celui correspondant à l'expérience effectuée avec le ligand pour évaluer la fixation non spécifique de l'analyte à la surface du capteur.

*c) Complexe entre l'anastelline et les ectodomains des syndécans*

La totalité des expériences a été réalisée en immobilisant l'anastelline sur des capteurs AR2G et en utilisant les ectodomains syndécans portant leurs chaînes de glycosaminoglycanes (d'origine commerciale), les ectodomains des syndécans exprimés en système procaryote (sans chaînes de glycosaminoglycanes) ou les glycosaminoglycanes comme analytes (**Tableau VIII**). L'activation des capteurs AR2G a été réalisée dans les mêmes conditions que précédemment.

Les capteurs activés ont été trempés pendant 300 s dans une solution d'anastelline (F3542, Sigma-Aldrich) diluée à une concentration de 100 µg/ml en tampon maléate de sodium 10 mM pH 6,2 puis dans deux puits contenant le tampon HBS-P20 0,05 % (v/v) pH 7,4 pendant 60 s chacun avant les étapes d'interactions (**Tableau VIII**). L'étape de dissociation a systématiquement été effectuée dans ce même tampon pendant 600 s. La totalité des expériences a été réalisée à 25 °C et sous agitation (1000 rpm).

Analytes	Référence	Présence des GAGs	Concentration (µM)
<b>Crible général des ectodomains des syndécans portant leurs chaînes de GAGs (SDCs) ou non (EDs) et des GAGs seuls</b>			
Ectodomaine du syndécan-1 + GAGs SDC-1	2780 R&D Systems	+	0,3
Ectodomaine du syndécan-2 + GAGs SDC-2	2965 R&D Systems	+	0,3
Ectodomaine du syndécan-3 + GAGs SDC-3	3539 R&D Systems	+	0,3
Ectodomaine du syndécan-4 + GAGs SDC-4	2918 R&D Systems	+	0,3
Héparane sulfate / HS (Muqueuse intestinale de porc)	GAG-HS01 Iduron	+	5
Chondroïtine sulfate / CS (Cartilage bovin)	C6737 Sigma-Aldrich	+	5
Ectodomaine du syndécan-1 ED-1	Exprimés au laboratoire	-	5
Ectodomaine du syndécan-2 ED-2		-	5
Ectodomaine du syndécan-3 ED-3		-	5
Ectodomaine du syndécan-4 ED-4		-	5

**Tableau VIII (1/2) : Conditions expérimentales des expériences d'interférométrie de biocouche.** Le tampon HBS-P20 0,05 % (v/v) pH 7,4 a été utilisé pendant la phase d'association.

Analytes	Concentration de l'analyte ( $\mu\text{M}$ )	Tampon Phase d'association
<b>A) Influence de la force ionique</b>		
ED-1	5	HBS-P20 0,05 % (v/v) pH 7,4 + <b>CHAPS 3 mM</b>
ED-1	5	HBS-P20 0,05 % (v/v) <b>(NaCl 0,15 M)</b> pH 7,4
ED-1	5	HBS-P20 0,05 % (v/v) <b>(NaCl 0,3 M)</b> pH 7,4
ED-1	5	HBS-P20 0,05 % (v/v) <b>(NaCl 0,6 M)</b> pH 7,4
ED-1	5	HBS-P20 0,05 % (v/v) <b>(NaCl 1,0 M)</b> pH 7,4
<b>B) Influence du pH</b>		
ED-1	5	HBS-P20 0,05 % (v/v) <b>pH 7,4</b>
ED-1	5	Maléate de sodium 10 mM - P20 0,05 % (v/v) <b>pH 6,2</b>
ED-1	5	Acétate de sodium 10 mM - P20 0,05 % (v/v) <b>pH 5,0</b>
<b>C) Interaction entre un oligosaccharide d'héparine (dp8) et l'anastelline</b>		
Dp8 d'héparine (Iduron, HO08)	5	HBS-P20 0,05 % (v/v) <b>pH 7,4</b>
[ HexUA(2S) – GlcNS(6S)– (IdoUA(2S) – GlcNS(6S)x3 ]	5	Acétate de sodium 10 mM - P20 0,05 % (v/v) <b>pH 5,0</b>

**Tableau VIII (2/2) : Conditions expérimentales des expériences d'interférométrie de biocouche.** Caractérisation de l'interaction entre l'anastelline avec **A-B)** l'ectodomaine du syndécan-1 (ED-1) et **C)** un octasaccharide d'héparine de degré de polymérisation 8 (dp8). ED-1 a été exprimé en système procaryote et est dépourvu de chaînes de glycosaminoglycanes. HexUA(2S), GlcNS(6S) et IdoUA(2S) représentent un résidu d'acide hexuronique 2-sulfaté, de glucose N- et 6-sulfaté ou d'acide iduronique 2-sulfaté, respectivement.

La recherche du complexe ternaire entre l'anastelline, l'ectodomaine du syndécan-1 (ED-1) et les chaînes d'héparane sulfate (HS – Iduron, HS-GAG01) ou de l'octasaccharide d'héparine (dp8 – Iduron, HO08) a été réalisée en immobilisant l'anastelline dans les mêmes conditions que précédemment. Les concentrations d'analytes (ED-1 et HS ou dp8), l'ordre des trempages ainsi que les détails des expériences sont résumées dans le **Tableau IX**. La totalité des expériences a été réalisée dans du tampon HBS-P20 0,05 % (v/v), à 25 °C et sous agitation (1000 rpm).

	<b>Trempage 1</b> Chargement <b>ED-1</b> (5µM)	<b>Trempage 2</b> Ligne de base (Deux à la suite)	<b>Trempage 3</b> Association <b>HS</b> ou <b>dp8</b> (5 µM)	<b>Trempage 4</b> Dissociation libre en tampon
<b>Complexe ternaire</b> <b>Anastelline – ED-1 –</b> <b>HS ou dp8</b>	900 s	HBS-P20 0,05 % (v/v) 480 s au total	300 s	HBS-P20 0,05 % (v/v) 300 s
<b>Complexe binaire</b> <b>Anastelline – ED-1</b>	900 s		HBS-P20 0,05 % (v/v), 300 s	
<b>Complexe binaire</b> <b>Anastelline – HS ou</b> <b>dp8</b>	HBS-P20 0,05 % (v/v), 900 s		300 s	

**Tableau IX : Formation des complexes binaires et ternaires par BLI.** Les trempages, numérotés de 1 à 4, des capteurs sur lesquels l'anastelline a été immobilisée ont été réalisés de façon séquentielle. Le temps de contact pour chaque étape est indiqué. La dissociation des chaînes d'héparane sulfate (HS) ou de l'octasaccharide d'héparine (dp8) a été réalisée dans le même puits que la deuxième étape de ligne de base dans le tampon HBS-P20 0,05 % (v/v) du trempage 2. La totalité des expériences a été réalisée dans du tampon HBS-P20 0,05 % (v/v) y compris le tampon dans lequel les analytes ont été dilués.

La saturation du ou des sites d'interaction des chaînes d'héparane sulfate sur l'anastelline a été réalisée en trempant une série de capteurs où a été immobilisée l'anastelline dans les mêmes conditions que précédemment. Puis les capteurs ont été trempés dans une solution d'ED-1 diluée à 5 µM en tampon HBS-P20 0,05 % (v/v) pH 7,4 pendant 0 s (contrôle), 90, 300, 600, 1200 et 3 600 s. Deux lignes de bases ont été réalisées (360 s) en trempant les capteurs dans du tampon HBS-P20 0,05 % (v/v) puis les capteurs ont été trempés dans une solution d'octasaccharide d'héparine (dp8) pendant 300 s avant que la dissociation soit effectuée en tampon comme décrit précédemment.

#### *d) Etude des interactions entre les globules rouges et le fibrinogène*

5 ml de sang d'un sujet sain ont été collectés dans des tubes EDTA, héparine et citrate puis centrifugés à température ambiante pendant 10 min à 500 x g. Le plasma a été éliminé et le culot de globules rouges a été resuspendu dans un tampon HEPES 10 mM, NaCl 133 mM, KCl 4,5 mM à pH 7,4 (tampon de course) puis la suspension a été lavée à température ambiante par trois cycles de 5 min de centrifugation à 500 x g. Le culot de globules rouges a ensuite été resuspendu dans le même tampon et la suspension a été laissée 10 minutes sur roue (10 rpm) à température ambiante.

Une partie de la suspension de globules rouges (hématocrite de 40 %) a été soumise à la digestion par la neuraminidase (N2876, Sigma-Aldrich) à une concentration de 400 mU/ml pendant 60 min au bain-marie (37°C). Les globules ont ensuite été lavés par centrifugation comme décrit précédemment et le culot de globules rouges a été resuspendu dans le tampon de course à un hématocrite de 40 %.

Les expériences de BLI ont été effectuées sur un système Octet RED96 (FortéBio, Molecular Devices). Les capteurs AR2G ont été activés de façon similaire au protocole précédent puis trempés dans une solution de poly-D-lysine (Corning) à une concentration de 100 µg/ml en présence de NaCl 1 M afin de l'immobiliser de façon covalente à la surface des capteurs. Les capteurs ont ensuite été trempés dans le tampon de course pendant 360 s puis la capture des globules rouges (traités ou non à la neuraminidase) s'est effectuée en immergeant les capteurs pendant 20 min dans une suspension de globules rouges (1x10<sup>6</sup> globules rouges / µl). Une ligne de base a été effectuée pendant 360 s dans le même tampon puis les capteurs ont été trempés pendant 180 s dans une solution de fibrinogène 100 µg/ml (F3879, Sigma-Aldrich) dans le même tampon et contenant du CaCl<sub>2</sub> 1 mM. Les complexes ont été dissociés en tampon seul pendant 300 s. La totalité des expériences a été réalisée à 25 °C et sous agitation (1000 rpm).

### III.8.2 Docking

#### *a) Docking d'un peptide de l'ectodomaine du syndécan-1 sur l'anastelline*

Les procédures de docking ont été effectuées sur le serveur spécialisé dans le docking protéine - peptide CABS-Dock (Kurcinski et al., 2015, Kurcinski et al., 2020). L'anastelline a été utilisée comme récepteur et la structure de la séquence de 26 résidus présents dans l'ectodomaine du syndécan-1 (<sup>54</sup>DITLSQQTPSTWKDTQLLTAIPTSPE<sup>79</sup>) a été générée de par le serveur à partir d'un jeu de conformations de peptides, puis utilisée comme ligand. La première structure résolue par résonance magnétique nucléaire de l'anastelline (PDB : 1Q38, 35 structures au total) a été modifiée de façon à ce que sa séquence en acides aminés soit identique à la séquence de la protéine commerciale (**Figure 48**). A l'extrémité N-terminale de la séquence de la protéine commerciale, les résidus <sup>1</sup>MRGS<sup>4</sup> ont été supprimés et un résidu de méthionine a été ajouté. Les résidus présents à l'extrémité C-terminale <sup>78</sup>TPGSRSHHHHH<sup>89</sup> et contenant l'étiquette 6xHis ont été supprimés. La structure modifiée de l'anastelline a été utilisée comme récepteur et les paramètres du serveur CABS-Dock ont été laissés par défaut.

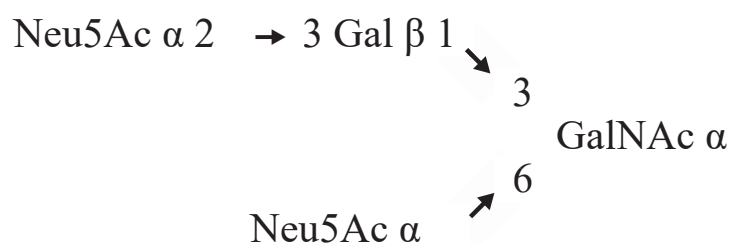


Noms des oligosaccharides	Séquence des oligosaccharides	Source / création des oligosaccharides
<b><i>Oligosaccharides d'héparine</i></b>		
<b>HE dp4</b>	[ GlcNS(6S)-IdoA(2S) ] x 2	<i>Samsonov et al., 2015</i> <i>Sattelle et al., 2013</i>
<b>HE dp6</b>	[ GlcNS(6S)-IdoA(2S) ] x 3	
<b>HE dp8</b>	[ GlcNS(6S)-IdoA(2S) ] x 4	
<b><i>Oligosaccharides d'héparane sulfate</i></b>		
<b>HSNA (dp4)</b>	GlcNAc-GlcUA-GlcNAc-IdoUA	GLYCAM web server <a href="http://glycam.org">http://glycam.org</a>
<b>HSNS (dp4)</b>	GlcNS(6S)-GlcUA-GlcNS(6S)-IdoUA(2S)	

**Tableau X : Ligands utilisés lors des expériences de docking et de dynamique moléculaire.**

*c) Docking d'un oligosaccharide et de l'acide N-acétylneuraminique sur le fibrinogène*

Les expériences de docking ont été réalisées localement avec la suite AutoDock Vina (version 4.2.6, *Trott & Olson, 2010*). Les atomes d'hydrogène des groupements polaires des résidus du fibrinogène (PDB : 3GHG) ont été ajoutés et les molécules d'eau ont été supprimées. Les structures de l'acide N-acétylneuraminique (NANA) et de la O-glycosylation de la glycophorine A (O-GPA) ont été créés par l'outil *CarbohydrateBuilder* (*Glycam.org, Complex Carbohydrate Research Center, The University of Georgia, USA*) selon la séquence suivante (*Aoki, 2017*) :



où les résidus Neu5Ac, Gal et GalNAc correspondent aux résidus de N-acétylneuraminique, de galactose et de N-acétylgalactosamine, respectivement.

Le fibrinogène étant un homodimère très étendu ( $\approx 45$  nm de long), un seul monomère (composé de trois chaînes,  $\alpha, \beta, \gamma$ ) et l'interface entre les deux monomères ont été utilisés durant ces expériences. Le monomère a été divisé en 5 grilles (cubes de 90 Å de côté, espacement de 1 Å) où les grilles se chevauchaient sur environ la moitié de leur longueur. 40 modèles ont été générés par grille (200 au total) avec une exhaustivité de 32.

### III.8.3 Détermination du site d'interaction entre l'anastelline et l'ectodomaine du syndécan-1 par spectrométrie de masse en tandem

L'anastelline et l'ectodomaine du syndécan-1 exprimé en système procaryote ont été réticulés par du 3,3'-dithiobis(sulfosuccinimidylpropionate), DTSSP, puis les complexes obtenus ont été analysés par spectrométrie de masse en tandem (MS/MS) sur la plateforme *Protein Science Facility* (Institut de Biologie et Chimie des Protéines, Lyon).

L'ectodomaine du syndécan-1 (ED-1), dilué à une concentration de 20  $\mu\text{M}$  (environ 540  $\mu\text{g/ml}$ ), et l'anastelline (F3542, Sigma-Aldrich), diluée à une concentration de 60  $\mu\text{M}$  (environ 504  $\mu\text{g/ml}$ ), ont été incubés à température ambiante pendant 30 min en tampon HBS en présence de DTSSP 4 mM (21578, Pierce), dont la solution mère (20 mM) a été préparée extemporanément. La neutralisation du DTSSP a été effectuée par du Tris-HCl pH 7,4 à une concentration finale de 300 mM pendant 15 min à température ambiante et les complexes réticulés ont été conservés à - 20 °C avant leur analyse. Les analyses par spectrométrie de masse ont été effectuées sur des complexes avec et sans étapes de réduction par du Tris(2-carboxyethyl)phosphine), TCEP, 5 mM et d'alkylation par l'iodoacétamide 10 mM. Les analyses ont été précédées d'une digestion des complexes par la chymotrypsine (pendant 5 h à 25°C) puis la trypsine (toute la nuit à 37°C) et l'échantillon a été dessalé sur microcolonne C18 (Pierce).

1  $\mu\text{l}$  de l'échantillon digéré a été injecté sur un spectromètre de masse Q Exactive HF (Thermo Scientific) suivant un gradient de 60 min en mode TOP20 HCD (les 20 ions les plus intenses sont automatiquement sélectionnés pour analyse par MS-MS). Les fichiers de données ont ensuite été traités avec deux approches. L'identification des peptides a été réalisée par *Proteome Discoverer 2.4* en utilisant le moteur de recherche SEQUEST HT et en restreignant la recherche à la banque de données contenant les deux séquences d'ED-1 et de l'anastelline.

Les modifications potentielles induites par le DTSSP ont été prises en compte : une différence de masse de +174 Da est détectée si la liaison est inter- ou intra-protéines, de +192 Da si l'agent réticulant est hydrolysé à une extrémité (*dead end*) et de +145 Da après réduction du pont disulfure du DTSSP et alkylation par l'iodoacétamide. La validation des peptides identifiés a été faite par une approche « Fixed Value » basée sur les scores SEQUEST. La recherche des sites d'interactions a été réalisée grâce deux logiciels dédiés (MeroX, v2.0 et pLink, v2.3.9).



### III.8.4 Turbidimétrie

L'anastelline (Sigma, F3542) a été incubée à une concentration de 60  $\mu\text{M}$  avec l'ectodomaine du syndécan-1 (20  $\mu\text{M}$ ) pendant 112 h à 37°C sous agitation (environ 200 rpm) dans du tampon HBS (pH 7,4) ou d'acétate de sodium 10 mM (pH 5,0). Les mesures de turbidimétrie ont été effectuées au bout de 65, 72 et 112 h d'incubation sur un spectrophotomètre Cary 60 UV-VIS (Agilent Technologies) équipé d'un adaptateur ayant un trajet optique de 1 mm. Les mesures ont été effectuées sur la gamme 700 – 200 nm à température ambiante.

### III.8.5 Fluorimétrie

Les expériences de fluorimétrie ont été réalisées sur un spectrofluorimètre Jasco FP-8500 selon le protocole décrit par *Nilsson* (2004). 9 spectres ont tout d'abord été enregistrés sur 1 ml de solution de thioflavine T 50  $\mu\text{M}$  (Sigma, T3516) en PBS filtré sur membrane d'acétate de cellulose 0,2  $\mu\text{m}$ , puis moyennés. 30  $\mu\text{l}$  d'échantillons d'ED-1, d'anastelline ou du complexe incubés à 37°C à pH 5,0 ont été ensuite directement ajoutés à la cuve contenant la solution de thioflavine T et 9 spectres ont de nouveau été enregistrés puis moyennés. La longueur d'onde d'excitation a été fixée à 440 nm (bande passante de 5 nm) et l'émission de fluorescence a été enregistrée sur la gamme 460 – 600 nm (bande passante de 10 nm) avec un pas de 0,2 nm. Les mesures ont été effectuées dans une cuve de 1 cm de trajet optique à température ambiante.

## III.9 Digestion des glycosaminoglycanes et des acides sialiques à la surface des globules rouges

Les échantillons de sang ont été collectés à partir de sujets sains (génotype AA) et de patients atteints de drépanocytose (génotype SS) dans des tubes EDTA. Le consentement individuel de chaque donneur a été obtenu et le protocole a été approuvé par le comité d'éthique des Hospices Civils de Lyon – CPP Est (L14-127) en accord avec les directives de la déclaration d'Helsinki. Les globules rouges ont été lavés par 3 séries de centrifugation à 4°C pendant 10 minutes à 500 x g en tampon phosphate salin pH 7,2 (PBS, Sigma-Aldrich, D8537). Le culot de globules rouges a ensuite été resuspendu et les cellules ont été incubées en PBS pendant 60 minutes à 37°C au bain-marie et à un hémocrite de 40 %, en présence de différentes concentrations d'enzymes (**Tableau XI**).

Détails des enzymes utilisées	Concentrations finales (mU/ml)
<b>Héparinase III</b> Sigma-Aldrich (H8891) Source : <i>Flavobacterium heparinum</i>	67,5
	150
	300
	400
<b>Chondroïtinase ABC</b> Sigma-Aldrich (C2905) Source : <i>Proteus vulgaris</i>	50
	250
	500
<b>Hyaluronidase</b> Sigma-Aldrich (H3631) Source : Testicule bovin	100
	250
	500
<b>Neuraminidase</b> Sigma-Aldrich (N2876) Source : <i>Clostridium perfringens</i>	400
	600

**Tableau XI : Détails des enzymes utilisées dans les expériences d'hydrolyse des constituants du glycocalyx des globules rouges sains et drépanocytaires.**

Après la digestion enzymatique, les globules rouges ont été centrifugés à température ambiante pendant 5 minutes à 500 x g et le surnageant a été recueilli. Les cellules ont ensuite été lavées en PBS par 2 séries de centrifugation dans les mêmes conditions. Les globules rouges ont été resuspendus à un hématoците final de 40 % soit dans leur plasma autologue, ce qui permet de mesurer leur agrégation, soit dans une solution de dextran 3 % (*m/v*) en PBS (Sigma-Aldrich, 31390) de poids moléculaire moyen de 70 kDa. La resuspension des globules rouges dans une solution de dextran permet de mesurer leur agrégabilité, c'est-à-dire leur capacité intrinsèque à s'agréger. Enfin, la suspension a été oxygénée pendant 10 minutes comme recommandé (*Baskurt et al., 2009*) avant les différentes séries de mesures.

### III.10 Mesures rhéologiques effectuées sur les globules rouges

#### III.10.1 Mesure de la déformabilité des globules rouges (ektacytométrie)

Les mesures de déformabilité ont uniquement été réalisées sur les globules rouges resuspendus dans leur plasma autologue avec l'ektacytomètre LORRCA MaxSis (*Laser-assisted Optical Rotational Cell Analyser*, Mechatronics, Pays-Bas) à 37°C.

L'appareil se compose de deux cylindres concentriques (système Couette) séparés d'environ 300  $\mu\text{m}$  (entrefer) où la suspension de globules rouges a été déposée. La mise en mouvement du cylindre interne permet de soumettre les globules rouges à des forces de cisaillement selon l'équation suivante :

$$\text{Force de cisaillement } (\tau; \text{Pa}) = \text{Viscosité } (\eta; \text{Pa.s}) \times \text{Vitesse de cisaillement } (\gamma; \text{s}^{-1})$$

Il n'est cependant pas possible de déposer la suspension de globule rouge telle quelle, car cela induirait une vitesse de rotation trop importante du fait de la faible viscosité du sang (entre 4,8 et 5,3 cP pour des vitesses de cisaillement supérieure à 200  $\text{s}^{-1}$ ). 7  $\mu\text{l}$  de suspension de globules rouges ont alors été mélangés avec une solution de polyvinylpyrrolidone de viscosité plus importante (27 cP à 37°C). La vitesse de rotation étant connue, il est alors possible d'imposer aux globules rouges des forces de cisaillement déterminées. Un laser de longueur d'onde de 670 nm est ensuite émis du cylindre externe et cible une caméra CCD devant laquelle l'échantillon circule. Le rayon est diffracté par les globules rouges de l'échantillon. Le spectre de diffraction enregistré par la caméra CCD, est retransmis sur l'écran d'ordinateur et est de forme elliptique sur lequel le logiciel adapte une ellipse théorique de dimensions connues (en pixels). Pour une même force de cisaillement, au moins 50 spectres sont mesurés et 50 ellipses adaptées afin de mesurer la déformabilité moyenne des globules rouges, exprimée en indice d'élongation, EI :

$$EI = \frac{A - B}{A + B}$$

A et B étant les axes majeur et mineur, respectivement, de l'ellipse théorique. La résultante est un graphique représentant l'évolution de l'indice d'élongation en fonction des forces de cisaillement appliquées.

### III.10.2 Mesure de l'agrégation et de l'agrégabilité des globules rouges

La mesure de l'agrégation est effectuée en plasma et reflète la tendance qu'ont les globules rouges à interagir ensemble pour former des rouleaux. L'agrégation peut être provoquée par des facteurs plasmatiques. L'agrégabilité, en revanche, caractérise la capacité intrinsèque des globules rouges à former des agrégats, et n'est donc pas due à des facteurs externes.

Ces mesures ont été effectuées avec l'appareil Myrenne (GmbH, Allemagne). 10  $\mu$ l de globules rouges, préalablement resuspendus soit dans leur plasma autologue soit dans une solution de dextran 3% ( $m/v$ ) de 70 kDa, ont été déposés entre le cône présent dans l'appareil et une lame de microscope. Le cône est en rotation, imposant ainsi à l'échantillon une vitesse de cisaillement déterminée ( $600 \text{ s}^{-1}$ ) pendant 10 secondes, ce qui désagrège de potentiels agrégats de globules rouges. Puis un rayon lumineux émis par une diode électroluminescente infrarouge traverse le système, d'environ 1,5 mm d'épaisseur. L'intensité lumineuse est enregistrée et sa valeur dépendra du nombre et de la taille des agrégats formés pendant une période de 5 secondes une fois le cisaillement stoppé. Deux programmes existent : soit le signal est enregistré alors que le cône n'est plus en rotation pendant la mesure (fonction M) et l'agrégation/agrégabilité sera donc mesurée en stase, soit le signal est enregistré alors que le cône tourne encore et à faible vitesse (fonction M1), ce qui permet de mesurer l'agrégation/agrégabilité des globules rouges à faible vitesse de cisaillement ( $3 \text{ s}^{-1}$ ). La fonction M1 donne des valeurs généralement plus importantes que l'indice M étant donné que la faible rotation du cône potentialise la rencontre des globules rouges et donc l'agrégation/agrégabilité. La transmittance de la lumière est mesurée et exprimée sous forme d'une valeur adimensionnelle en unité arbitraire. Plus la valeur est élevée, plus l'agrégation/agrégabilité est importante.

### III.10.3 Analyses statistiques

La déformabilité et l'agrégation/agrégabilité des sangs prélevés de sujets sains ou de patients drépanocytaires sans digestion enzymatique ont été comparées avec le test de Mann-Whitney.

L'effet des digestions enzymatiques sur les propriétés rhéologiques des globules rouges a été évalué en utilisant l'analyse de la variance (ANOVA) de Friedman pour mesures répétées suivie du test post hoc de Dunn. Une *p-value* < 0,05 a été considérée comme statistiquement significative et la légende suivante a été utilisée : \*  $p < 0,05$ ; \*\*  $p < 0,01$ ; \*\*\*  $p < 0,001$ ; \*\*\*\*  $p < 0,0001$ .

# **CHAPITRE IV**

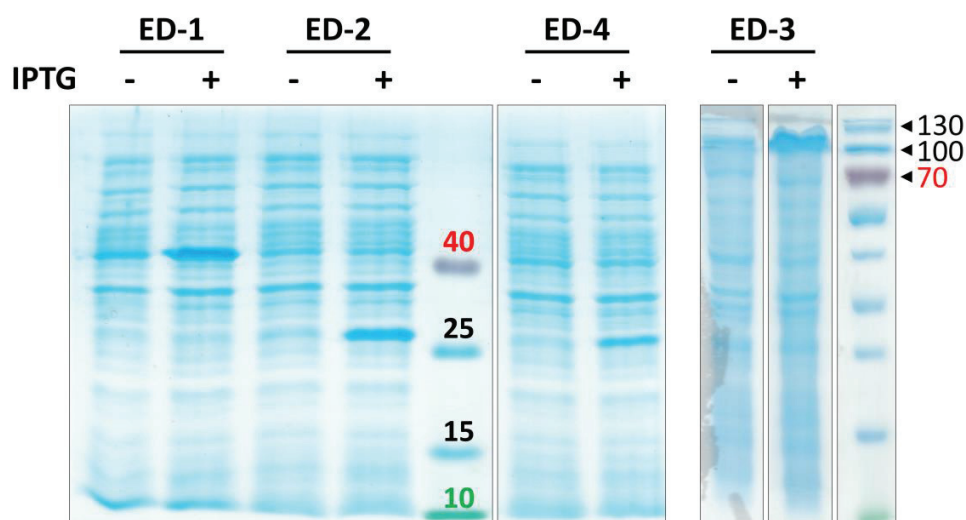
## **RÉSULTATS**

## **IV.1**

# **CARACTÉRISATION STRUCTURALE DES ECTODOMAINES DES SYNDÉCANS SANS GAGS**

#### IV.1.1 Expression et purification des ectodomaines de syndécans sans GAGs

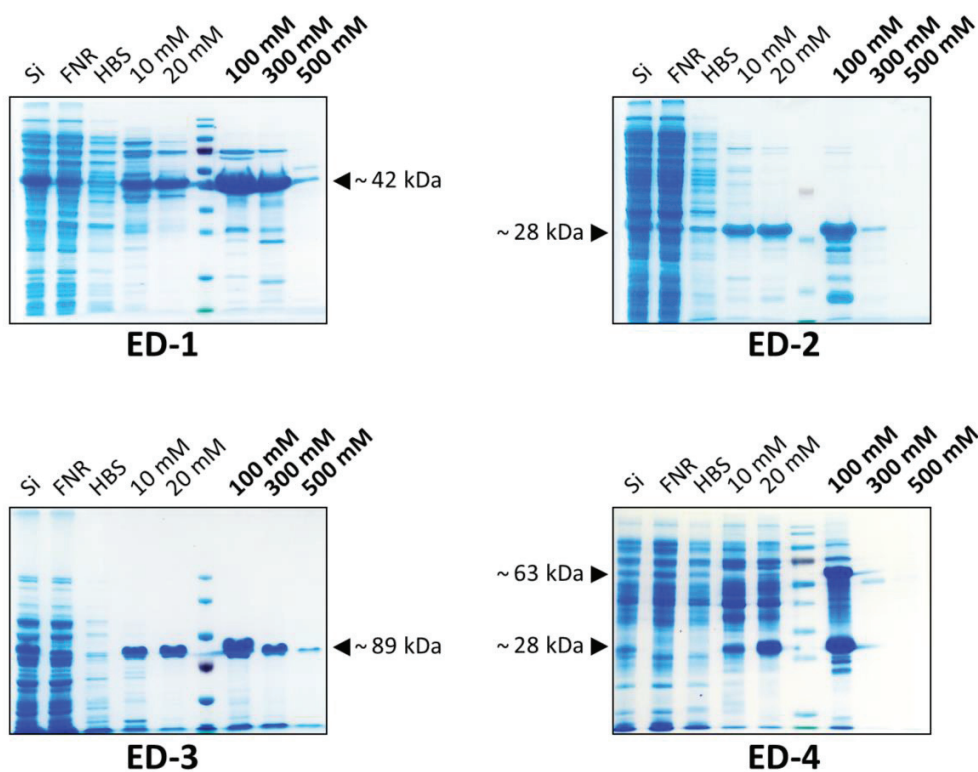
Après les étapes de transformation et de mise en culture des clones, l'expression des ectodomaines de syndécans sous forme recombinante (EDs) a été suivie par électrophorèse sur gel de polyacrylamide en conditions dénaturantes et réductrices (SDS-PAGE) avant et après induction par l'IPTG (**Figure 49**).



**Figure 49 : Analyse par électrophorèse en conditions dénaturantes et réductrices de l'expression en système procaryote des ectodomaines des syndécans.** 5  $\mu$ l de culture d'*E. coli* BL21 (DE3) non induite (-) ou induite par l'IPTG 1 mM pendant 3 h (+) ont été dilués dans 15  $\mu$ l de tampon de charge réducteur ( $\beta$ -mercaptoéthanol 5 % (v/v)) puis la totalité a été déposée sur gel de polyacrylamide 15 %. 5  $\mu$ l de marqueurs de masses moléculaires Spectra Low Range (26628, ThermoFisher pour ED-1, -2 et -4) ou PageRuler (26616, ThermoFisher pour ED-3) ont été déposés. Les protéines ont été colorées par du bleu de Coomassie InstantBlue (ISB1L, Sigma-Aldrich). Les masses moléculaires sont en kDa.

L'analyse par électrophorèse montre que les ectodomaines des syndécans sont exprimés et que 3 heures d'induction sont nécessaires. Cependant, les masses moléculaires apparentes des ectodomaines des syndécans diffèrent des masses théoriques calculées à partir de la séquence en acides aminés. ED-1 migre à une masse moléculaire d'environ 40 kDa, ED-2 et ED-4 à environ 25 kDa et ED-3 n'a pas suffisamment migré dans le gel afin de déterminer sa masse moléculaire apparente précisément, mais semble être aux alentours de 100 kDa.

Les masses moléculaires théoriques attendues sont de 27,2, 17,2, 38,8 et 18,0 kDa pour ED-1 à ED-4, respectivement. Les culots de bactéries ont ensuite été lysés mécaniquement et les lysats ont été clarifiés par centrifugation avant d'être purifiés par chromatographie d'affinité sur une résine Ni-NTA en tampon HEPES salin (HBS) (**Figure 50**). L'éluition des EDs a été effectuée par compétition en effectuant 3 paliers successifs avec des concentrations d'imidazole de 100, 300 et 500 mM. Les ectodomaines des syndécans ont été retenus sur la résine Ni-NTA, bien que l'interaction avec la résine ne soit pas optimale car une faible quantité d'ectodomaine de syndécans est présente dans les fractions de lavage par l'imidazole 10 et 20 mM. Les 4 ectodomaines ont été élués par l'imidazole 100 mM et en plus faible quantité à la concentration de 300 mM.

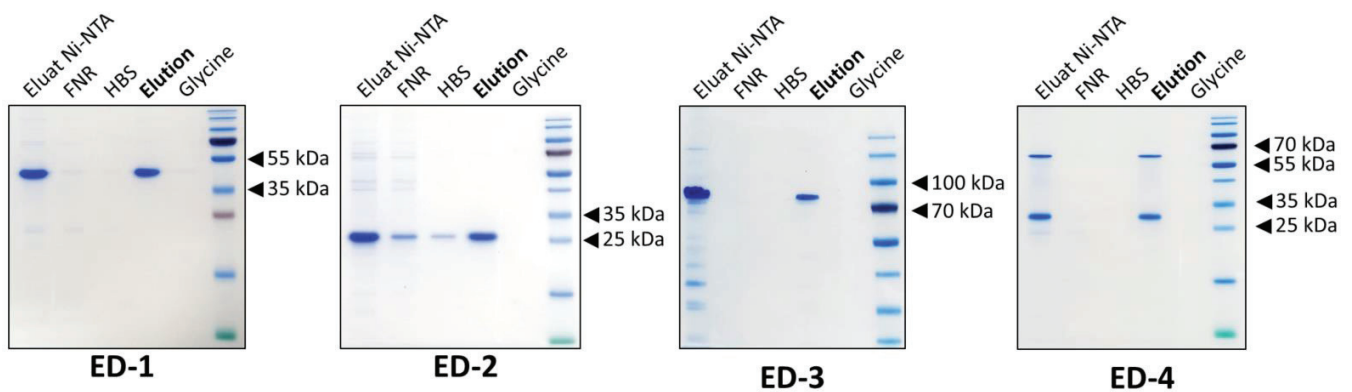


**Figure 50 : Purification par chromatographie d'affinité Ni-NTA des 4 ectodomaines de syndécans.** Entre 5 et 10  $\mu$ l de solution initiale (*Si*), de fraction non retenue (*FNR*) ou de lavage en HBS seul (*HBS*) ou en présence d'imidazole 10 ou 20 mM (*10 mM*, *20 mM*) et 15  $\mu$ l de fractions éluées par 100, 300 ou 500 mM d'imidazole ont été déposés sur un gel de polyacrylamide 15 % (ED-1, -2 et -4) ou 8 % (ED-3). 5  $\mu$ l de marqueurs de masses moléculaires PageRuler (26616, ThermoFisher pour ED-1, -3 et -4) et Spectra Low Range (26628, ThermoFisher pour ED-2) ont été déposés en conditions non réductrices. Les protéines ont été colorées par du bleu de Coomassie InstantBlue (ISB1L, Sigma-Aldrich).



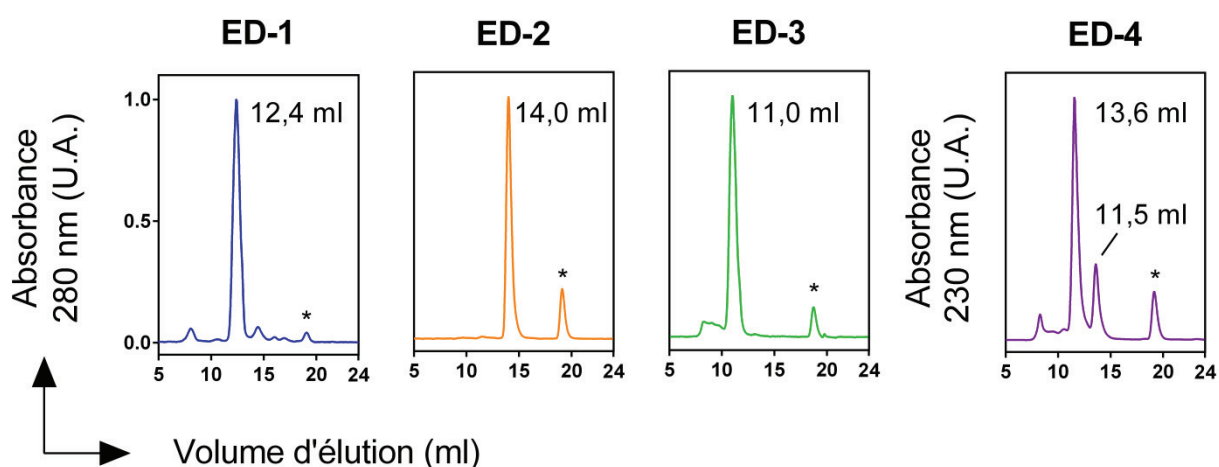
Cette première purification permet l'élimination de la majorité des contaminants sur un support ayant une grande capacité de fixation des protéines possédant une étiquette poly-histidines. L'analyse par SDS-PAGE en conditions non réductrices montre la présence de deux bandes pour ED-4, migrant à environ 28 et 64 kDa, soit un rapport d'environ 2, suggérant une dimérisation médiée par un pont disulfure. L'utilisation d'un gel 8 % en polyacrylamide a également permis d'estimer la masse moléculaire apparente d'ED-3 à environ 89 kDa (**Figure 50**).

Une seconde étape de purification a été réalisée sur une résine portant un anticorps anti-FLAG pour éliminer les contaminants dans les fractions éluées lors de la purification sur résine Ni-NTA. Etant donné que la capacité de la résine anti-FLAG est de l'ordre de 0,6 à 1 mg de protéine par ml de résine (contre 20 à 50 mg/ml pour les résines Ni-NTA), les éluats issus de la purification Ni-NTA ont été dilués entre 4 et 8 fois en tampon afin de ne pas saturer la résine anti-FLAG. L'avantage de l'utilisation de cette résine réside dans la haute spécificité de l'anticorps couplé à l'agarose vis-à-vis du peptide FLAG présent à l'extrémité C-terminale des EDs recombinants. Les 4 ectodomaines ont été élués par compétition avec une solution de peptide FLAG à 200 µg/ml en tampon HBS. (**Figure 51**).



**Figure 51 : Purification par chromatographie d'affinité anti-FLAG des 4 ectodomaines de syndécans pré-purifiés par chromatographie d'affinité Ni-NTA.** 15 µl d'ED pré-purifié (Eluat Ni-NTA), de fraction non retenue (FNR), de lavage en HBS (HBS), d'éluat (Elution) et du lavage par une solution de glycine 100 mM pH 3,5 (glycine) ont été déposés sur gel 15 % (ED-1, -2 et -4) ou 8 % (ED-3). Les protéines ont été colorées par du bleu de Coomassie InstantBlue (ISB1L, Sigma-Aldrich).

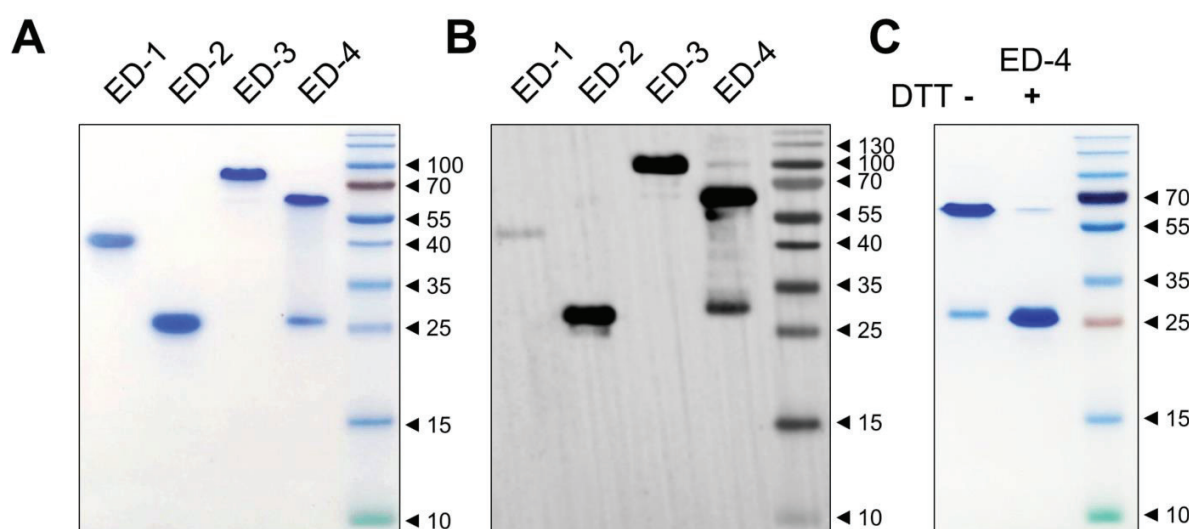
Les EDs ayant subi deux étapes de purification par chromatographie d'affinité sont purs d'après l'analyse par électrophorèse. Afin d'éliminer le peptide FLAG présent dans les éluats et d'éliminer de potentiels agrégats, une troisième étape de purification est effectuée par chromatographie d'exclusion stérique. Les éluats issus de la purification sur la résine anti-FLAG sont tout d'abord concentrés par centrifugation jusqu'à obtenir 500 µl de solution, puis cette solution est injectée sur une colonne semi-préparative de 24 ml dont le support est de l'agarose réticulé et du dextran (**Figure 52**). Cette colonne possède un domaine de fractionnement s'étendant de 10 à 600 kDa.



**Figure 52 : Purification par chromatographie d'exclusion stérique des 4 ectodomaines de syndécan préalablement purifiés par chromatographies d'affinité.** 500 µl de solution d'EDs pré-purifiés par chromatographies d'affinité sur des résines Ni-NTA puis anti-FLAG ont été injectés sur une colonne d'exclusion stérique (Superdex S200 Increase 10/300 GL, GE Healthcare) équilibrée en tampon HBS à 10°C et à un débit de 0,5 ml/min. Les volumes approximatifs d'élution des 4 EDs sont indiqués (\* : élution du peptide FLAG). L'élution d'ED-4 a été suivie à 230 nm du fait de sa trop faible absorbance à 280 nm. U.A. : unité arbitraire.

Les fractions contenant les EDs de syndécans ont été ensuite concentrées par centrifugation. Puis les rétentats ont été centrifugés à 4°C pendant 20 minutes à 18 000 x g avant d'être stockés à -20°C. Concernant ED-4, les éluats contenant le monomère et le dimère ont été regroupés et concentrés ensemble.

Les EDs de syndécans ont été analysés par SDS-PAGE et Western blot avec un anticorps anti-FLAG (**Figures 53A** et **53B**). Etant donné qu'ED-4 est le seul ectodomaine possédant un résidu de cystéine dans sa séquence, une dimérisation médiée par un pont disulfure est envisagée. Afin de valider cette hypothèse, un aliquot d'ED-4 purifié a été analysé par SDS-PAGE en conditions réductrices (**Figure 53C**), ce qui a provoqué une diminution de l'intensité de la bande ayant une masse d'environ 63 kDa et une augmentation de celle ayant une masse de 28 kDa, confirmant la dimérisation d'ED-4 par un pont disulfure.



**Figure 53 : Analyse par SDS-PAGE et Western blot des 4 ectodomains de syndécans purifiés.** 2  $\mu$ g de protéines ont été déposés sur gel de polyacrylamide 15 %. Les protéines ont été colorées par du bleu de Coomassie **(A)** ou transférées sur membrane de PVDF et détectées avec un anticorps anti-FLAG **(B)**. **(C)** 2  $\mu$ g d'ED-4 ont été incubés ou non avec du DTT 2 mM ( $\pm$  DTT) puis déposés sur gel de polyacrylamide 15 % et colorés par du bleu de Coomassie.

Afin de vérifier leur identité, le séquençage des protéines a été réalisé par spectrométrie de masse en tandem précédée d'une séparation par chromatographie en phase inverse (Plateforme *Protein Science Facility*, Institut de Biologie et Chimie des Protéines, Lyon). Le taux de couverture des séquences atteignait 76 %, 22 %, 43 % et 81 % pour ED-1 à ED-4, respectivement, permettant de confirmer l'identité des protéines, bien que le taux de couverture d'ED-2 soit assez faible (**Annexe 1**).

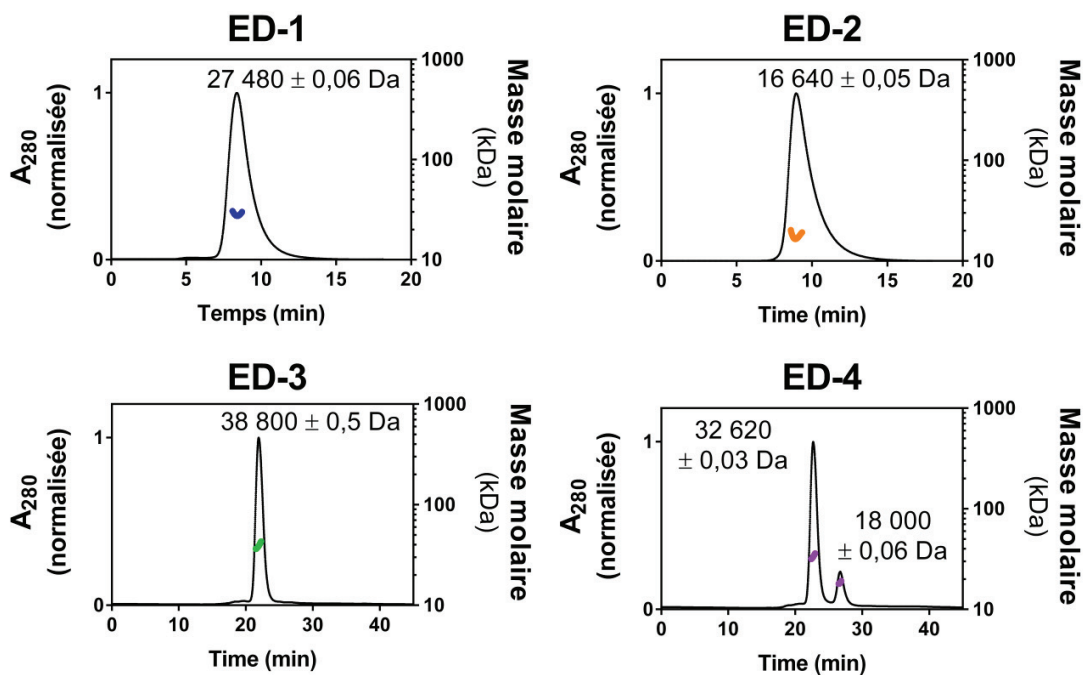
#### IV.1.2 Détermination de la masse moléculaire des ectodomaines de syndécans

Les masses moléculaires apparentes observées en électrophorèse en conditions dénaturantes sont environ 1,5 à 2 fois plus importantes que celles calculées à partir de la séquence des ectodomaines de syndécans. Ce phénomène de migration anormale peut être causé par une présence importante de désordre intrinsèque ce qui affecte leur mobilité électrophorétique (*Uversky, 2010*). Afin de déterminer leur masse réelle, les 4 protéines ont été analysées par diffusion de la lumière laser multi-angles (MALLS) et spectrométrie de masse en conditions natives (ESI-MS).

##### *a) Diffusion de la lumière laser multi-angles couplée à la chromatographie d'exclusion stérique (SEC-MALLS)*

La masse moléculaire des protéines exprimées en système procaryote a été mesurée en solution par diffusion de la lumière laser multi-angles. Les techniques de diffusion de la lumière étant extrêmement sensibles aux particules de haut poids moléculaire (agrégats, poussières), l'analyse s'est effectuée après passage des ectodomaines des syndécans sur une colonne de chromatographie d'exclusion stérique en amont de l'appareillage (SEC-MALLS). Les résultats des expériences de SEC-MALLS sont présentés dans la **Figure 54**.

Les masses moléculaires des ectodomaines des syndécans 1 - 3 mesurées par SEC-MALLS sont de 27,5, 16,6, et 38,6 kDa, respectivement, en accord avec les masses calculées à partir des séquences en acides aminés (27,2, 17,2 et 38,8 kDa, respectivement). La masse du monomère d'ED-4 mesurée en SEC-MALLS est de 18,0 kDa et identique à celle calculée (18,0 kDa). En revanche, la masse mesurée du dimère présumé par SEC-MALLS est d'environ 32,6 kDa, la masse théorique étant de 35,9 kDa (**Tableau XII**).

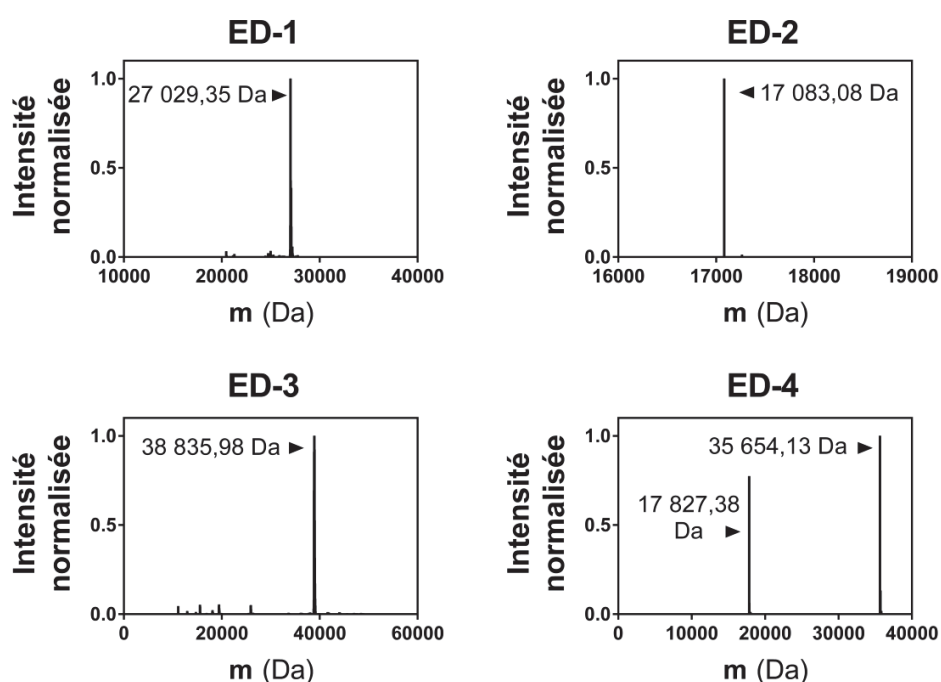


**Figure 54 : Calcul de la masse moléculaire des ectodomaines des syndécans par SEC-MALLS.** Pour chaque ectodomaines de syndécans (EDs), le chromatogramme (en noir) et la valeur de la masse moléculaire (en couleur) sont présentés. Les masses moléculaires obtenues après analyse sont indiquées ( $\pm$  erreur).

*b) Spectrométrie de masse en conditions natives (ESI-MS)*

La masse moléculaire des ectodomaines des syndécans a également été mesurée par spectrométrie de masse en conditions natives. Le terme de *conditions natives* regroupe toutes les conditions expérimentales permettant de conserver d'une part l'intégrité de la protéine et d'autre part sa structure tridimensionnelle. Il est toutefois nécessaire que le tampon utilisé soit volatil. Dans cette série d'expériences, le tampon HBS a été échangé contre une solution d'acétate d'ammonium 50 mM (pH 7,4). Les protéines sont en phase gazeuse mais les mesures effectuées en ESI-MS sont beaucoup plus précises (**Figure 55**).

Les masses moléculaires des ectodomaines des syndécans calculées par ESI-MS sont en accord avec les calculs théoriques et les mesures de SEC-MALLS avec cependant un léger écart de masse d'environ 132 Da pour ED-1, ED-2 et ED-4. Cette masse correspond exactement à celle d'un résidu de méthionine engageant sa fonction carboxylique dans une liaison peptidique. Il est alors très probable qu'ED-1, ED-2 et ED-4 aient perdu leur résidu de méthionine N-terminal. Concernant ED-3, l'écart de masse est de moins de 5 Da, suggérant que ce résidu de méthionine est toujours présent (**Tableau XII**).



**Figure 55 : Spectre de masse des ectodomaines des syndécans obtenus en conditions natives.** Les masses moléculaires obtenues après analyse sont indiquées.

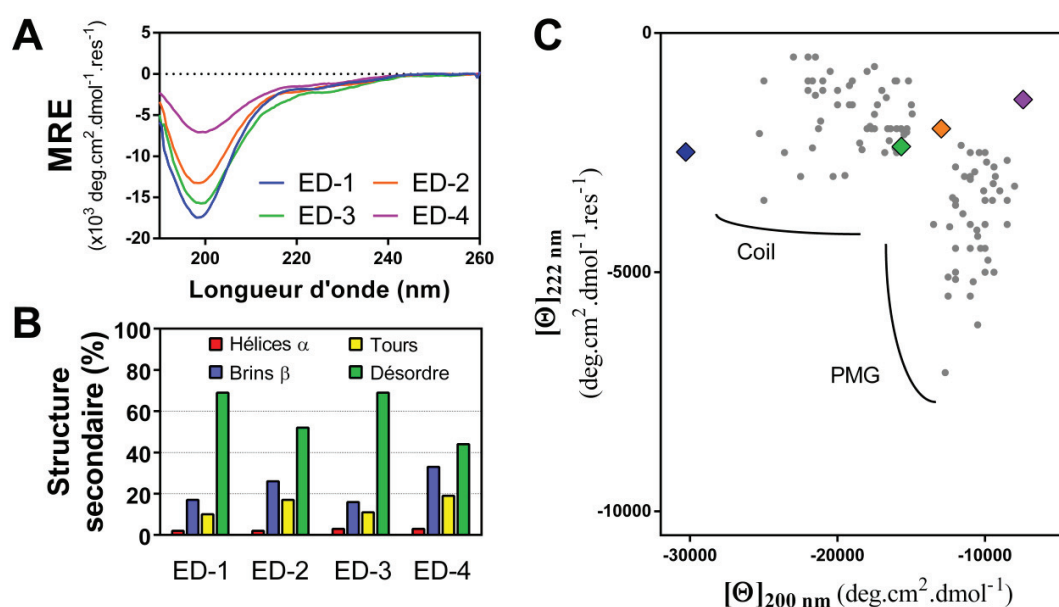
L'ensemble des analyses et prédictions subséquentes nécessitant la séquence en acides aminés des ectodomaines des syndécans-1, -2 et -4 sera donc effectué en tenant compte de la perte de leur résidu de méthionine en position N-terminale. Le bilan des mesures de masses moléculaires est présenté dans le **Tableau XII**.

	Masse théorique	Masse calculée (SEC-MALLS)	Masse calculée (ESI-MS)	Ecart de masse (Da)
<b>ED-1</b>	27 161	27 480 ± 0,06	27 029,35	+ 131,65
<b>ED-2</b>	17 215	16 640 ± 0,05	17 083,08	+ 131,92
<b>ED-3</b>	38 831	38 800 ± 0,5	38 835,98	- 4,98
<b>ED-4</b>	17 959	18 000 ± 0,06	17 827,38	+ 131,62
	35 918	32 620 ± 0,03	35 654,13	+ 263,87

**Tableau XII : Bilan des mesures de la masse moléculaire (Da) des ectodomaines des syndécans (EDs).** L'écart de masse correspond à la différence entre la masse théorique (calculée à partir de la séquence) et la masse mesurée par ESI-MS.

### IV.1.3 Structures secondaires et désordre intrinsèque des ectodomaines des syndécans

Des expériences de dichroïsme circulaire ont été menées afin de déterminer la proportion en structures secondaires et en désordre intrinsèque des ectodomaines des syndécans (**Figure 56**). Le tampon HBS a préalablement été échangé par du tampon phosphate de potassium 10 mM (pH 7,4) par chromatographie d'exclusion stérique et les fractions d'élutions ont été concentrées par centrifugation.



**Figure 56 : Composition en structures secondaires et pourcentage de désordre intrinsèque des EDs de syndécans. A)** Spectres de dichroïsme circulaire des 4 EDs de syndécans. MRE : Ellipticité moyenne par résidu. **B)** Déconvolution des spectres réalisé avec l'algorithme CDSSTR et le jeu de données « 7 » de Dichroweb. **C)** Détermination de la nature du désordre intrinsèque des EDs. ED-1 : bleu, ED-2 : orange, ED-3 : vert, ED-4 : violet. Le jeu de données expérimentales (gris) de protéines adoptant des conformations de pré-globules fondus (PMG) ou très étendues (Coil) est celui d'Uversky (2002).

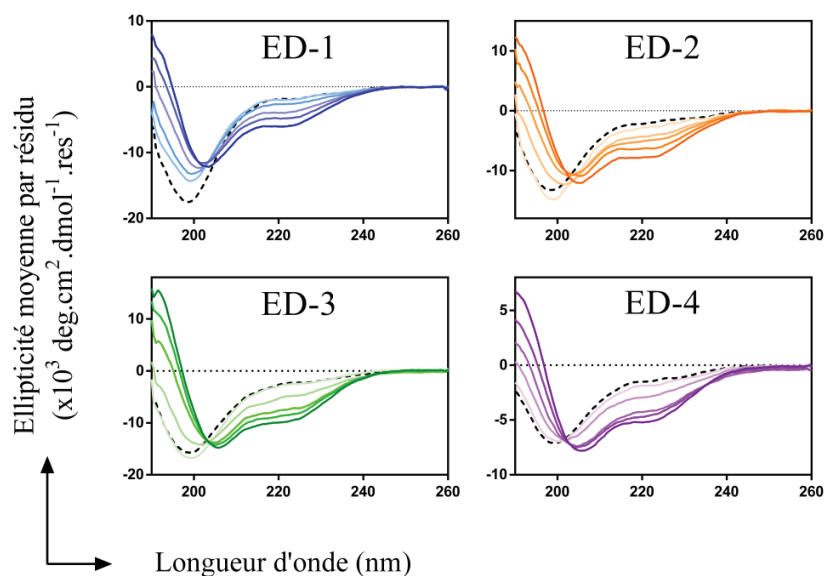
La présence d'un minimum à 199 nm sur les spectres de dichroïsme circulaire montre la présence de désordre intrinsèque dans les 4 ectodomaines de syndécans (**Figure 56A**). Ces spectres ont été déconvolués en utilisant un jeu de données comprenant 5 protéines chimiquement dénaturées proches des protéines intrinsèquement non structurées (jeu de données « 7 »).

D'après cette déconvolution, ED-1 et ED-3 sont les plus désordonnés (69 %) alors que les deux ectodomaines les plus petits, ED-2 et ED-4 sont les moins désordonnés (52 % et 44 %, respectivement). Les structures secondaires résiduelles sont essentiellement composées de brins  $\beta$ , qui sont majoritairement retrouvés dans ED-2 et ED-4 (26 % et 33 %, respectivement) et il y en a 16 % et 17 % dans ED-1 et ED-3, respectivement (**Figure 56B**).

Comme décrit précédemment, la famille des protéines désordonnées se subdivise en deux groupes majeurs, définis selon leur contenance en structures secondaires résiduelles et ainsi leur degré de compaction. Les protéines intrinsèquement non structurés dites *Coil* sont plus étendues et possèdent ainsi moins de structures secondaires résiduelles que les protéines à l'état de pré-globule fondu (PMG). Les travaux d'Uversky (2002) ont permis de démontrer qu'il était possible de différencier ces deux groupes à partir des résultats expérimentaux obtenus par dichroïsme circulaire, en traçant la valeur de l'ellipticité moyenne par résidu à 222 nm (caractéristique des hélices  $\alpha$ ) en fonction de celle à 200 nm (caractéristique du désordre intrinsèque). Cette approche montre qu'ED-1 et ED-3 appartiennent au groupe des IDPs dites *coils* alors qu'ED-4 se retrouve plutôt dans le groupe des IDPs plus compactes (PMG). En revanche, ED-2 se situe à l'interface entre la famille des *coils* et des globules pré-fondus (**Figure 56C**).

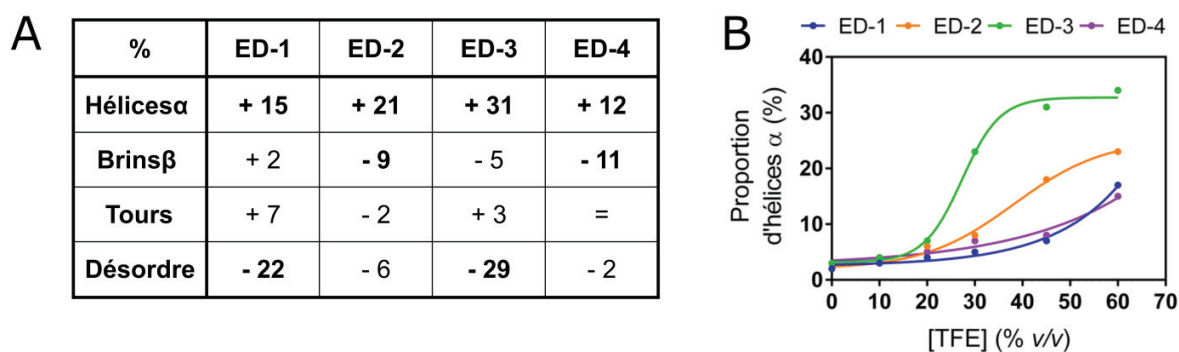
Des mesures de dichroïsme circulaire ont ensuite été réalisées en présence de concentrations croissantes de trifluoroéthanol (TFE) qui favorise la formation des hélices  $\alpha$  (Culik *et al.*, 2014). Le chlorure de guanidinium (ClGu) est un agent chaotrope qui déstabilise les hélices  $\alpha$  et les brins  $\beta$ , mais favorise la formation de structure très étendue, notamment les hélices de type polyproline II (Kelly *et al.*, 2001, Chemes *et al.*, 2012). L'utilisation de ces deux composés permet donc d'évaluer le degré de plasticité structurale des ectodomaines, à savoir leur capacité à former des hélices  $\alpha$  ou polyproline II. Les 4 ectodomaines de syndécans ont d'abord été incubés en absence ou en présence de TFE (10 – 60 %, v/v) et les spectres ont été analysés de la même façon que précédemment (**Figure 57**).





**Figure 57 : Spectres de dichroïsme circulaire des EDs en présence de concentrations croissantes de trifluoroéthanol.** Les ectodomaines des syndécans (EDs) ont été incubés sans (courbe noire en pointillé) ou avec des concentrations croissantes de TFE (10 - 60%  $v/v$ ) (du plus clair au plus foncé).

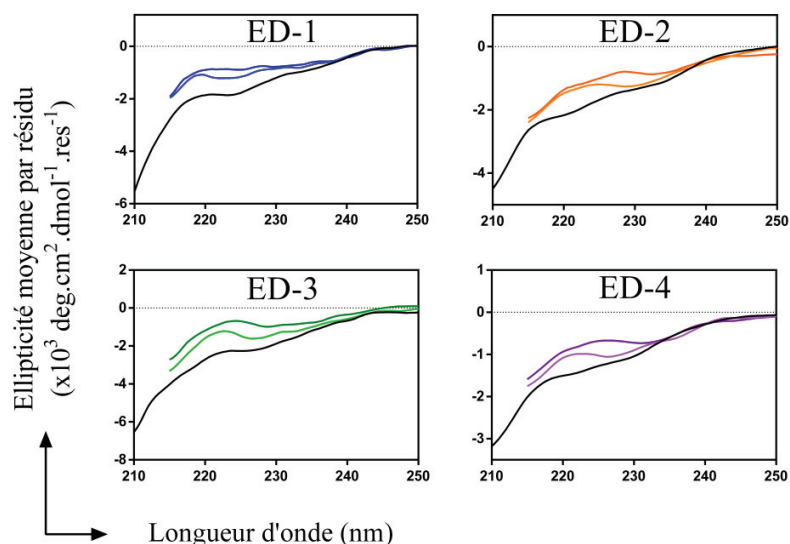
L'incubation des EDs en présence de TFE provoque une diminution de l'intensité du signal à 200 nm ainsi qu'une apparition progressive de deux minima de plus en plus négatifs à 204 nm (ED-1) et 207 nm (ED-2 - 4), valeurs proches de celles caractérisant les hélices  $\alpha$  (208 et 222 nm) (**Figure 57**). Ces observations mènent à la conclusion que les 4 ectodomaines possèdent la capacité de former des hélices  $\alpha$  en présence de TFE. D'après la déconvolution des spectres, environ un tiers de la séquence d'ED-3 (34 %), un quart d'ED-2 (23 %) et environ 15 % d'ED-1 et d'ED-4 forment des hélices  $\alpha$  après l'ajout de 60 % ( $v/v$ ) de TFE. Ces transitions sont accompagnées d'une perte de désordre pour ED-1 (-22 %) et ED-3 (-29 %), alors que les hélices  $\alpha$  se forment majoritairement aux dépens des brins  $\beta$  pour ED-2 (-9 %) et ED-4 (-11 %) (**Figure 58A**). La formation d'hélices  $\alpha$  semble être un processus coopératif pour ED-3 (**Figure 58B**).



**Figure 58 : Evolution de la composition en structures secondaires des 4 EDs en présence de TFE. A)** Variations en structures secondaires des 4 EDs en présence de trifluoroéthanol (60 %  $v/v$ ). Les variations les plus importantes sont en gras (= : pas de variation). **B)** Evolution de la proportion en hélices  $\alpha$  au sein des EDs en fonction de la concentration en TFE.

Une autre série d'expériences a été réalisée en présence de chlorure de guanidinium afin de promouvoir la formation d'hélices polyproline II (**Figure 59**). Bien que les ectodomaines des syndécans soient enrichis en résidu de proline, leur présence n'est toutefois pas obligatoire dans la formation de ce type de structure. Les hélices polyproline II sont difficilement visibles sur des spectres de dichroïsme circulaire et sont caractérisées par un minimum à 198 nm, donc un signal semblable au désordre, ainsi qu'un maximum de très faible intensité à 218 nm. Comme indiqué ci-dessus, le chlorure de guanidinium déstabilise les structures secondaires existantes afin de faciliter l'observation des hélices polyproline II dont il favorise l'apparition (*Chemes et al., 2012*). Les 4 ectodomaines ont été incubés en présence de ClGu 3 et 5 M, et les spectres ont été enregistrés sur une gamme spectrale réduite étant donné que les ions chlorures absorbent très fortement aux longueurs d'ondes inférieures à 215 nm.

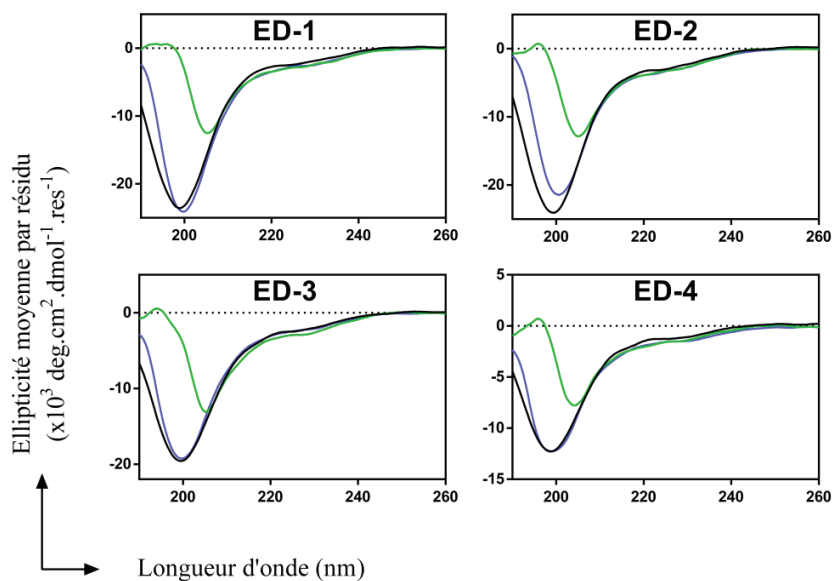
Une augmentation de l'intensité du signal est détectée en présence de chlorure de guanidinium de 215 à environ 235 nm ce qui peut suggérer la formation d'hélices de type polyproline II (**Figure 59**). En revanche, le signal demeure négatif et les variations par rapport aux spectres obtenus en absence de chlorure de guanidinium restent faibles. Les algorithmes de déconvolution ne permettent pas de calculer le pourcentage d'hélice de type polyproline II mais les 4 ectodomaines de syndécans ne peuvent vraisemblablement pas, ou très peu, former d'hélice de type polyproline II.



**Figure 59 : Spectres de dichroïsme circulaire des EDs en présence de chlorure de guanidinium.** Les ectodomaines des syndécans (EDs) ont été incubés sans (courbe noire) ou avec du chlorure de guanidinium 3 et 5 M (du plus clair au plus foncé).

Finalement, une dernière série d'expériences de dichroïsme circulaire a été réalisée en incubant les EDs en présence de deux chaperones chimiques, l'oxyde de triméthylamine (TMAO), qui stabilise les conformations intermédiaires de repliement (*Jethva & Udgaonkar, 2018*), ainsi que le 4-phénylbutyrate (4-PBA), un osmolyte hydrophobe qui limite ainsi l'exposition de régions apolaires des protéines non ou mal repliées, diminuant ainsi leur agrégation. Les ectodomaines ont été incubés soit en présence de TMAO 0,5 M soit en présence de 4-PBA à une concentration finale de 100  $\mu$ M et les spectres enregistrés dans les mêmes conditions que précédemment (**Figure 60**).

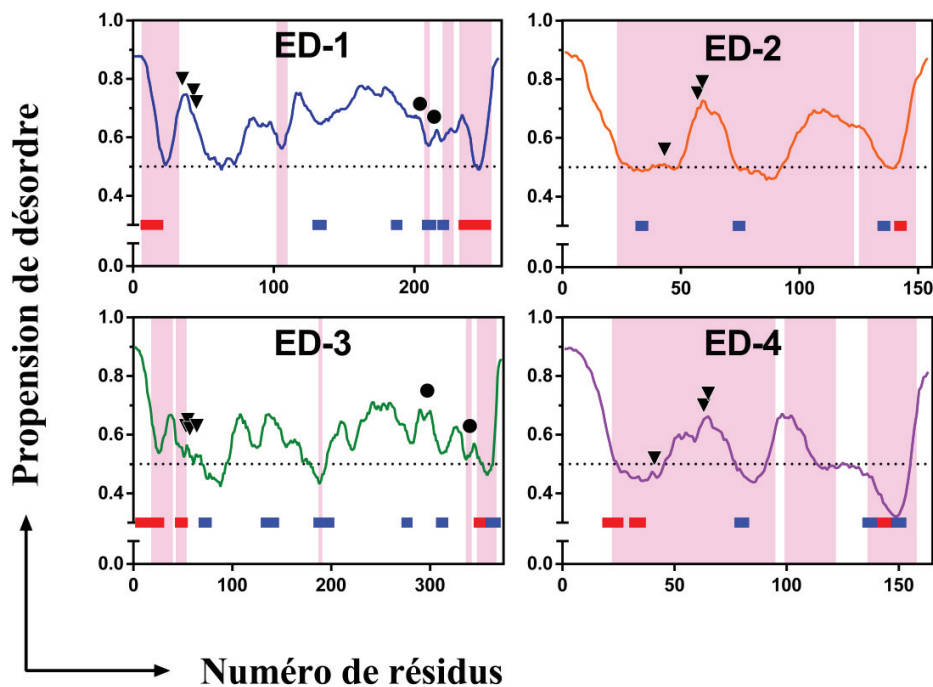
L'incubation avec du 4-PBA n'a pas significativement affecté les spectres des ectodomaines de syndécans qui restent ainsi majoritairement intrinsèquement non structurés. Cette observation est en accord avec le fait que le 4-PBA interagit de façon privilégiée avec les domaines hydrophobes des protéines et que par définition, les protéines intrinsèquement non structurées en sont dépourvues. En revanche, l'ajout du TMAO engendre une diminution de la quantité de désordre intrinsèque comme en atteste la diminution du minimum à 200 nm. Cette diminution est accompagnée d'un léger décalage du spectre vers des longueurs d'ondes plus importantes, entre 204 et 206 nm, ainsi que d'un maximum positif entre 194 et 196 nm. En conséquence, le TMAO semble diminuer la quantité de désordre intrinsèque (**Figure 60**).



**Figure 60 : Spectres de dichroïsme circulaire des EDs en présence de chaperonnes chimiques.** Les ectodomaines des syndécans ont été incubés sans (courbe noire) ou en présence de TMAO 0,5 M (vert) ou de 4-PBA 100  $\mu$ M (bleu).

La présence et la localisation des structures secondaires et des régions intrinsèquement non structurées dans les ectodomaines des syndécans ont été prédites *in silico* en utilisant respectivement le métaprédicteur *metaPrDOS* et le prédicteur *PSIPRED 4.0*. La présence de régions pouvant se structurer au contact d'un partenaire (MoRFs) a été prédite en utilisant le prédicteur *MoRFchibi System* (**Figure 61**). En accord avec les expériences de dichroïsme circulaire et des prédictions effectuées sur les séquences des syndécans entiers (*Leonova & Galzitskaya, 2015*), les ectodomaines de syndécans sont prédits comme étant majoritairement non structurés.

Peu de structures secondaires sont prédites : les hélices  $\alpha$  sont situées principalement aux extrémités N- et C-terminales des ectodomaines des syndécans alors que les brins  $\beta$  sont distribués de façon homogène le long de la séquence d'ED-2 et d'ED-3 et plutôt du côté C-terminal d'ED-1 et d'ED-4. Les sites d'attachements des glycosaminoglycanes sont principalement retrouvés dans des régions prédites comme désordonnées. Seules de courts MoRFs sont prédits principalement aux extrémités N- et C-terminales d'ED-1 et ED-3. En revanche, ED-2 et ED-4 possèdent des MoRFs sur presque la totalité de leur séquence (**Figure 61**). Une similarité est observée dans les prédictions d'une part entre ED-1 et ED-3, et d'autre part entre ED-2 et ED-4, tant au niveau de la prédiction du désordre que de celle des MoRFs. Ceci est cohérent avec l'homologie de séquence des couples SDC-1 / SDC-3 d'une part et SDC-2 / SDC-4 d'autre part (*Chakravarti & Adams, 2006*).



**Figure 61 : Prédiction des proportions en structures secondaires, en désordre intrinsèque et des MoRFs au sein des EDs des syndécans.** La propension de désordre des résidus d'acides aminés est représentée par une courbe colorée (valeur > 0,5). La prédiction des MoRFs est sous forme d'aire colorée en rose. La prédiction et la position des hélices  $\alpha$  et des brins  $\beta$  est représentée par des zones colorées en rouge et bleu, respectivement. Les positions des sites d'attachement des chaînes d'héparane sulfate (triangles noirs) et de chondroïtine sulfate (cercles noirs) sont indiquées.

#### IV.1.4 Détermination des paramètres hydrodynamiques des ectodomaines des syndécans

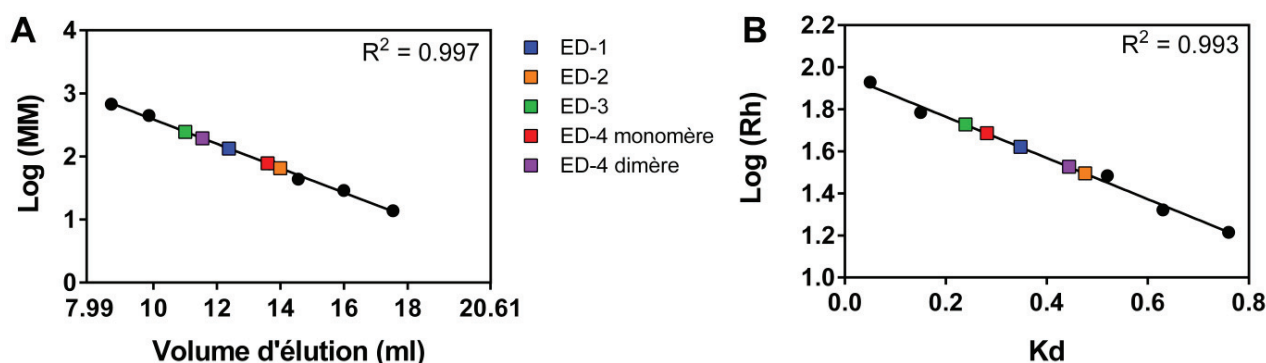
La mesure du rayon hydrodynamique ( $R_h$ ) des protéines est un moyen de déterminer la forme globale d'une protéine ainsi que son taux de compaction. Les paramètres hydrodynamiques des ectodomaines des syndécans ont été déterminés expérimentalement par diffusion dynamique de la lumière couplée (SEC-DLS) ou non (DLS) à une colonne de chromatographie d'exclusion stérique et par chromatographie d'exclusion stérique à l'aide de protéines standards dont le rayon hydrodynamique est connu.

*a) Calcul du rayon hydrodynamique par chromatographie d'exclusion stérique*

L'utilisation des colonnes de chromatographie d'exclusion stérique permet une détermination de la masse moléculaire apparente et du rayon hydrodynamique des protéines étudiées (**Figure 62** et **Tableau XIII**). L'inconvénient majeur réside dans le fait que le calcul de la masse est dépendant du degré de compaction des standards utilisés pour la calibration de la colonne. Dans cette expérience, 6 protéines globulaires ont été utilisées pour sa calibration. Ces résultats sont décrits dans la **Figure 62** et le **Tableau XIII**.

La masse moléculaire apparente a été déterminée à partir des standards globulaires en traçant le logarithme de la masse moléculaire (MM) en fonction du volume d'élution. La masse moléculaire apparente calculée à partir de la calibration de la colonne de chromatographie d'exclusion stérique est bien supérieure à la masse réelle mesurée par ESI-MS, ce qui est attendu du fait de l'enrichissement en désordre intrinsèque des EDs, menant à un volume hydrodynamique plus important qu'une protéine globulaire ayant le même nombre de résidus d'acides aminés. Dans nos expériences cette différence est d'autant plus marquée que la protéine possède une masse moléculaire élevée (d'environ 4 à 6 fois).

Le rayon hydrodynamique a été déterminé en utilisant le coefficient de distribution ( $K_d$ ) calculé pour cette colonne au lieu du coefficient de partition ( $K_{av}$ ), ce-dernier étant dépendant du volume approximatif de la colonne indiqué pas le fournisseur (24 ml). A l'inverse, le  $K_d$  dépend des paramètres propres à la colonne (volume mort et volume total) qui ont été déterminés expérimentalement. D'après ces expériences, ED-3 possède le rayon hydrodynamique le plus élevé (64,4 Å), celui d'ED-1 est de 46,3 Å alors que les Rh d'ED-2 (31,3 Å) et la forme monomérique d'ED-4 (33,6 Å) sont équivalents, en accord avec leur masse moléculaire réelle. Le dimère d'ED-4 a un volume d'élution compris entre celui d'ED-1 et d'ED-3, et il en est de même pour son rayon hydrodynamique (48,7 Å) (**Tableau XIII**).



**Figure 62 : Détermination de la masse moléculaire apparente et du rayon hydrodynamique des ectodomaines des syndécans par chromatographie d'exclusion stérique.** **A)** Détermination de la masse moléculaire apparente des EDs des syndécans. L'axe des abscisses représente le domaine de fractionnement de la colonne (Superdex S200 10/300 Increase, GE Healthcare), de 7,99 ml (volume mort) à 20,61 ml (volume total). **B)** Calcul du rayon hydrodynamique. Pour les deux graphes, 6 protéines globulaires ont été injectées lors de la calibration (●), un point est masqué par les données d'ED-2 et d'ED-4 monomère. La légende est commune aux deux graphiques et pour chaque graphe et le coefficient de détermination  $R^2$  est indiqué.

Ectodomaines des syndécans	Masse moléculaire calculée ESI-MS (Da)	Masse moléculaire apparente SEC (Da)	Rapport : masse apparente/réelle	Rh SEC (Å)
<b>ED-1</b>	27 029,4	133 256	4,9	41,8
<b>ED-2</b>	17 083,1	64 963	3,8	31,3
<b>ED-3</b>	38 836,0	245 550	6,3	53,5
<b>ED-4 monomère</b>	17 827,4	77 684	4,4	33,6
<b>ED-4 dimère</b>	35 654,1	193 285	5,4	48,6

**Tableau XIII : Détermination de la masse moléculaire apparente et du rayon hydrodynamique (Rh) des ectodomaines des syndécans par chromatographie d'exclusion stérique.** Le rapport masse apparente/réelle correspond au rapport de la masse moléculaire apparente calculée par chromatographie d'exclusion stérique (SEC) et de celle calculée en spectrométrie de masse (ESI-MS)

*b) Calcul du rayon hydrodynamique par diffusion dynamique de la lumière*

Les rayons hydrodynamiques ont été également mesurés par diffusion dynamique de la lumière (DLS). D'après la loi de Stokes-Einstein, le coefficient de diffusion (D) d'une particule dans une solution de viscosité ( $\eta$ ) connue et à une température (T) donnée dépend directement de son rayon hydrodynamique ( $R_h$ ), selon l'équation suivante :

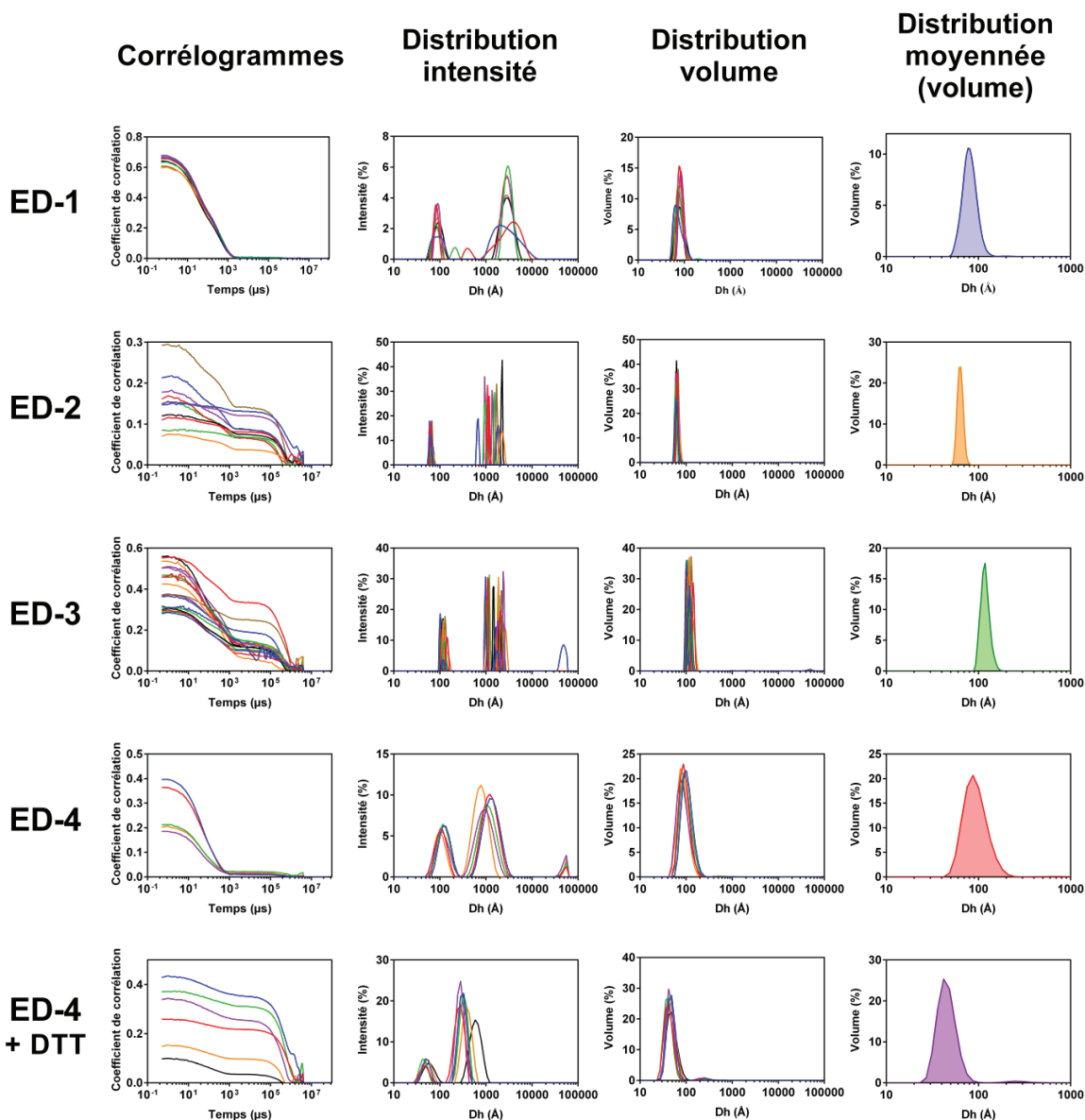
$$D = \frac{kT}{6\pi\eta R_h}$$

où k est la constante de Boltzmann. Dans cette équation, le rayon hydrodynamique correspond à celui d'une sphère qui diffuserait de la même manière que la molécule étudiée. Ainsi pour une masse moléculaire donnée, une protéine désordonnée possède un rayon hydrodynamique plus important que celui d'une protéine globulaire. Les ectodomaines des syndécans purifiés ont été directement analysés dans des cuves lors des mesures. Afin d'obtenir des données concernant uniquement la forme monomérique d'ED-4, un agent réducteur (DTT) a été ajouté à la solution (**Figure 63**).

Les corrélogrammes enregistrés lors des différentes mesures ne sont pas superposés. De plus, ceux d'ED-2 et d'ED-3 présentent deux points d'inflexion et celui d'ED-4 est stable pendant un temps important, caractéristiques d'une solution très polydisperse, dominée par des espèces de grandes tailles comme des agrégats ou des poussières. L'intensité de diffusion en DLS étant proportionnelle au carré de la masse des particules en solution (et du diamètre de la particule élevé à la puissance 6 dans le cadre de la diffusion de Rayleigh), une faible quantité d'agrégat peut dominer la distribution des intensités de diffusion. Une espèce de faible diamètre hydrodynamique ( $D_h$ ) est cependant systématiquement retrouvée pour chaque ectodomaine des syndécans dans la distribution des intensités diffusées (**Figure 63**).

Cette distribution est ensuite transformée en distribution en volume par l'application de la théorie de Mie, valable pour des particules dont la taille est supérieure au dixième de la longueur d'onde de l'onde électromagnétique illuminant l'échantillon. Alors que la distribution en intensité traite de la quantité absolue de lumière diffusée par les particules de différentes tailles, la distribution en volume indique au combien une particule d'une certaine taille diffuse ainsi que sa proportion vis-à-vis des autres particules présentes dans l'échantillon.





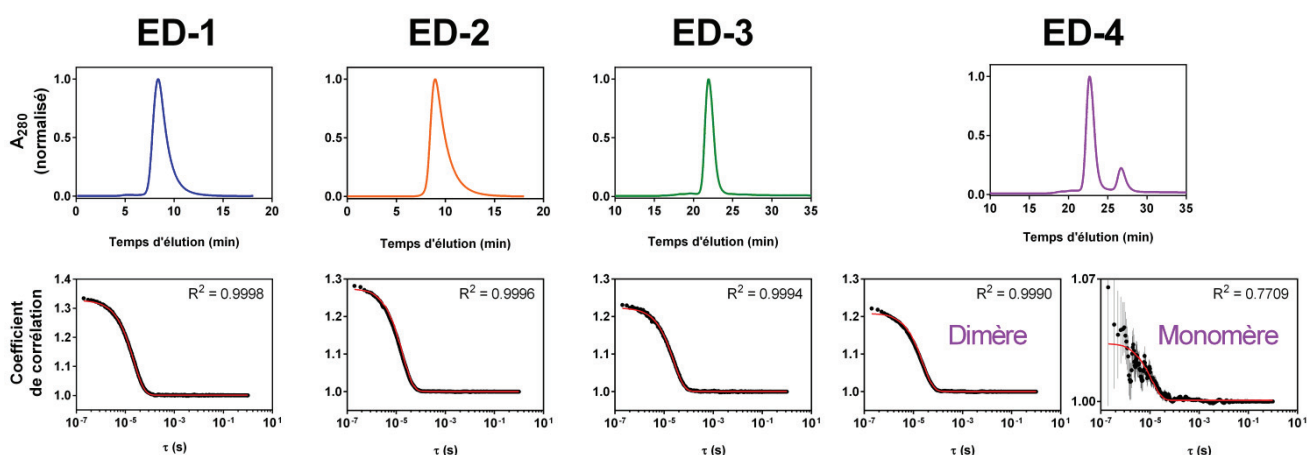
**Figure 63 : Détermination du diamètre hydrodynamique des ectodomaines des syndécans par diffusion dynamique de la lumière.** Corrélogrammes associés aux différentes mesures réalisées sur le même échantillon qui ont été utilisés pour déterminer les distributions en intensité et en volume par l'application de la théorie de Mie. Les distributions en volume ont été moyennées avec  $n = 7$  (ED-1), 11 (ED-2), 18 (ED-3), 5 (ED-4) et 6 (ED-4 + DTT), afin de déterminer le diamètre hydrodynamique ( $D_h$ ) des ectodomaines des syndécans (EDs). ED-4 a également été analysé en présence de DTT afin d'obtenir le diamètre hydrodynamique de sa forme monomérique.

Pour chaque ectodomaines de syndécans, un seul pic centré sur son diamètre hydrodynamique ( $D_h$ ) est observé dans la distribution en volume. En moyennant plusieurs spectres, il a été possible de déduire le diamètre hydrodynamique des EDs des syndécans, et donc leur rayon hydrodynamique, qui est de  $40,7 \pm 7,2 \text{ \AA}$ ,  $31,8 \pm 2,3 \text{ \AA}$  et  $59,4 \pm 6,7 \text{ \AA}$  pour ED-1, -2 et -3, et de  $47,8 \pm 14,3 \text{ \AA}$  et  $23,0 \pm 5,4 \text{ \AA}$  pour la forme dimérique et monomérique de ED-4, respectivement (**Tableau XIV**).

*c) Calcul du rayon hydrodynamique par diffusion dynamique de la lumière couplée à la chromatographie d'exclusion stérique*

Comme indiqué précédemment, la mauvaise superposition des corrélogrammes et leurs formes suggèrent une solution polydisperse et entraîne ainsi une estimation du rayon hydrodynamique peu précise, caractérisée par des écart-types importants. Des mesures de DLS couplées à une colonne d'exclusion stérique (SEC-DLS) ont donc été effectuées (**Figure 64**). Les spectres UV des fractions éluées indiquent la présence d'une faible quantité de contaminants ou d'agrégats ayant une masse moléculaire supérieure à celle des ectodomaines des syndécans et ces espèces ont été efficacement séparées par la colonne d'exclusion stérique. De ce fait, les corrélogrammes et leurs ajustements respectifs sont de meilleure qualité, excepté pour la forme monomérique d'ED-4 où le signal n'était pas assez intense mais dont il a été possible d'extraire une valeur de rayon hydrodynamique de  $21,6 \pm 4 \text{ \AA}$ .

Ces mesures ont permis de déterminer les  $R_h$  d'ED-1 ( $44,9 \pm 0,8 \text{ \AA}$ ), ED-2 ( $34,1 \pm 0,7 \text{ \AA}$ ), ED-3 ( $50,1 \pm 1,3 \text{ \AA}$ ) et de la forme dimérique d'ED-4 ( $45,4 \pm 1,2 \text{ \AA}$ ). La valeur des écart-types sont nettement inférieurs lorsque les valeurs de  $R_h$  sont calculées sur des fractions homogènes (**Tableau XIV**).



**Figure 64 : Détermination du rayon hydrodynamique des ectodomains des syndécans par diffusion dynamique de la lumière couplée à la chromatographie d'exclusion stérique.** Le chromatogramme de l'analyse par chromatographie d'exclusion stérique des ectodomains des syndécans ainsi que les corrélogrammes sont présentés.  $R^2$  : coefficient de détermination.

	Rayons hydrodynamiques ( $\text{\AA} \pm \text{écart-type}$ )		
	SEC	DLS	SEC-DLS
<b>ED-1</b>	41,8	$40,7 \pm 7,2$	$44,9 \pm 0,8$
<b>ED-2</b>	31,3	$31,8 \pm 2,3$	$34,1 \pm 0,7$
<b>ED-3</b>	53,5	$59,4 \pm 6,7$	$50,1 \pm 1,3$
<b>ED-4 monomère</b>	33,6	$23,0 \pm 5,4$	$21,6 \pm 4,0$
<b>ED-4 dimère</b>	48,6	$47,8 \pm 14,3$	$45,4 \pm 1,2$

**Tableau XIV : Rayons hydrodynamiques des ectodomains des syndécans calculées par chromatographie d'exclusion stérique (SEC), par diffusion dynamique de la lumière couplée (SEC-DLS) ou non (DLS) à une étape de chromatographie d'exclusion stérique.**

Le rayon hydrodynamique d'une protéine peut également être calculé selon des équations empiriques déterminées à partir d'un jeu de données expérimental et décrites dans les travaux d'Uversky (2002). Selon ces équations, rappelées dans la section *Matériel & Méthodes*, le  $R_h$  est directement proportionnel d'une part à la forme globale de la protéine et d'autre part à sa masse moléculaire.

Quatre équations reflètent le degré de compaction d'une protéine. L'état natif (N) correspond à une protéine globulaire parfaitement repliée alors que l'état de globule fondu (MG) s'apparente à une protéine peu compacte dont les éléments de structures secondaires sont formés. Les états de pré-globule fondu (PMG) et de *coil* (C) reflètent une protéine intrinsèquement non structurée possédant respectivement plus ou moins de repliements locaux. Ces rayons hydrodynamiques théoriques peuvent être utilisés pour comparer les valeurs de rayon hydrodynamique expérimentales à des valeurs de Rh théoriques correspondant à l'un des degrés de compaction décrits ci-dessus. Les Rh théoriques des 4 ectodomains des syndécans sont présentés dans le **Tableau XV**.

Ectodomains des syndécans	Rh (Å) SEC-DLS	Rh théoriques (Å ± écart type)			
		N	MG	PMG	C
<b>ED-1</b>	<b>44,9 ± 0,8</b>	24,0 ± 3,5	29,3 ± 16,8	36,3 ± 12,5	<b>43,6 ± 9,5</b>
<b>ED-2</b>	<b>34,1 ± 0,7</b>	20,4 ± 2,9	25,0 ± 14,1	30,2 ± 10,2	<b>34,7 ± 7,4</b>
<b>ED-3</b>	<b>50,1 ± 1,3</b>	27,3 ± 4,1	33,1 ± 19,3	42,1 ± 14,8	<b>52,1 ± 11,6</b>
<b>ED-4 (monomère)</b>	<b>21,6 ± 4,0</b>	<b>20,7 ± 3,0</b>	25,4 ± 14,3	30,7 ± 10,4	35,5 ± 7,6
<b>ED-4 (dimère)</b>	<b>45,4 ± 1,2</b>	26,5 ± 3,9	32,2 ± 18,7	<b>40,7 ± 14,2</b>	<b>50,0 ± 11,0</b>

**Tableau XV : Comparaison des valeurs de rayon hydrodynamiques mesurés en SEC-DLS et calculés selon des équations empiriques pour chaque degré de compaction.** Pour le calcul du Rh du dimère d'ED-4 dans des conformations étendues (PMG et *coil*), l'utilisation des équations implique une dimérisation par une extrémité de la protéine et donc une très courte zone de chevauchement des monomères. Pour chaque ectodomains les valeurs les plus proches des valeurs expérimentales sont en gras.

La comparaison des Rh expérimentaux (déterminés par SEC-DLS) et théoriques montre que les conformations globales d'ED-1, ED-2, ED-3 sont très étendues, compatibles avec des *coils*. En revanche, le Rh expérimental de la forme monomérique d'ED-4 montre une compaction similaire de celle des protéines globulaires (**Tableau XV**). A l'inverse, le rayon hydrodynamique expérimental de sa forme dimérique est compris entre le Rh d'un état de pré-globule fondu et de celui d'un *coil*. La valeur expérimentale du rayon hydrodynamique de la forme dimérique d'ED-4 est 2,1 fois plus importante que celle de la forme monomérique ce qui suggère une conformation étendue des deux monomères formant le dimère.

L'ensemble de ces expériences montrent que les 4 ectodomains des syndécans sont intrinsèquement non structurés. Les prédictions *in silico* et les expériences de dichroïsme circulaire indiquent un enrichissement en désordre qui peut être modulé en présence d'osmolytes (TFE, TMAO), mettant en avant leur plasticité structurale. Les propriétés hydrodynamiques d'ED-1, -2 et -3 indiquent un faible degré de compaction et le dimère d'ED-4 semble adopter une conformation étendue. Afin de poursuivre la caractérisation de cette famille de protéines intrinsèquement non structurées, différentes techniques ont été utilisées (SAXS, ESI-(IM)-MS) pour définir la nature du désordre intrinsèque et définir l'espace conformationnel exploré par ces protéines.

#### **IV.1.5 Analyses structurales des ectodomains des syndécans**

##### *a) Nature du désordre intrinsèque*

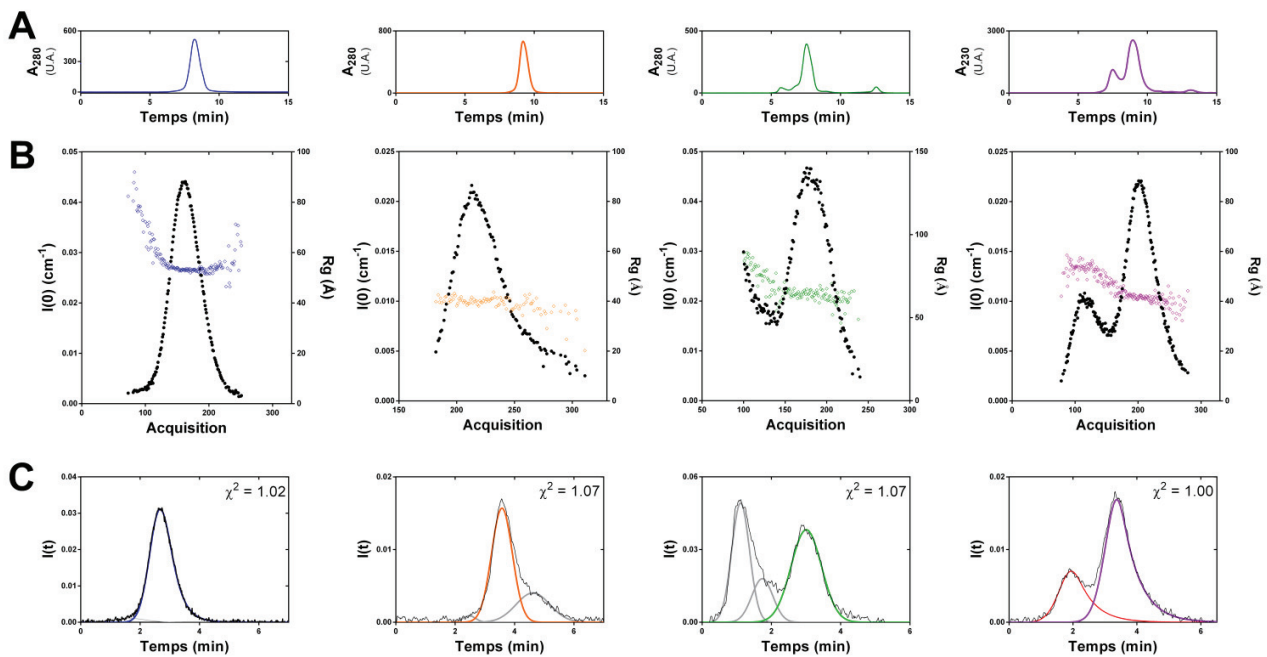
Comme rappelé précédemment, le désordre intrinsèque des protéines peut être classé selon leur degré de compaction, bien qu'il puisse être considéré comme un *continuum* (Uversky, 2013, Uversky, 2019). Le but de ces expériences a été de déterminer si les ectodomains de syndécane se comportaient soit comme des pré-globules fondus (PMG), qui possèdent des structures secondaires résiduelles, soit comme des *coils*, caractérisés par des conformations très étendues.

Pour cela, des expériences de diffusion des rayons X aux petits angles (SAXS) ont été réalisées. Le SAXS est l'une des techniques les plus adaptées à l'étude structurale des protéines intrinsèquement non structurées et l'analyse des intensités de diffusion permet de définir leurs paramètres structuraux à savoir leur rayon de gyration ( $R_g$ ) et leur dimension maximale ( $D_{max}$ ). La manière dont les intensités de diffusion décroissent reflète directement leur degré de compaction. Finalement, étant donné qu'une unique structure n'est pas à même de représenter la diversité conformationnelle de ce type de protéine, aucune modélisation *ab initio* n'a été menée. En revanche, les intensités de diffusion peuvent être déconvoluées par des algorithmes adaptés afin de déterminer l'espace conformationnel de ces protéines. La difficulté majeure réside dans la préparation d'un échantillon de protéine pur et monodisperse, s'approchant le plus possible d'une solution idéale, sans interaction ni répulsion des protéines en solution. Ce prérequis est d'autant plus important pour les protéines intrinsèquement non structurées.

Afin de s'approcher au mieux de ces conditions *sine qua non*, les 4 ectodomaines des syndécans ayant préalablement subi 3 étapes de purifications ont été injectés sur une colonne d'exclusion stérique de 3 ml en amont de l'illumination aux rayons X lors des expériences de SAXS, afin d'éviter notamment la participation de potentiels agrégats aux intensités de diffusion collectées (**Figure 65**). Une colonne de 24 ml aurait également pu être utilisée, mais aurait dilué de façon trop importante les échantillons, menant à des signaux bruités, notamment aux grands angles. Malgré les précédentes étapes de purification, différentes espèces en quantité très minoritaire sont éluées avant les ectodomaines solubles comme le montrent les différents chromatogrammes (**Figure 65A**). C'est notamment le cas pour ED-3, où la diffusion des rayons X de ces espèces vient perturber celle de la protéine soluble, résultant ainsi une estimation peu fiable du rayon de gyration (**Figure 65B**). Les deux formes monomérique et dimérique d'ED-4 possèdent des volumes d'élution proches sur cette colonne menant à un problème similaire. Le traitement des données le plus simple aurait été de sélectionner différentes acquisitions et les moyenner lorsqu'une seule et unique espèce est retrouvée, généralement en sélectionnant des acquisitions au milieu et la fin des fractions éluées. Cependant, le but est d'obtenir des données les moins bruitées et ce, sur une couverture angulaire la plus grande afin d'extraire le maximum d'information structurale sur les ectodomaines des syndécans.

Les intensités de diffusion ont ainsi été déconvoluées en utilisant le module US-SOMO HPLC-SAXS (*Brookes et al., 2016*). Comme décrit dans la partie *Matériel & Méthodes*, les intensités de diffusion  $I(s) = f(s)$  ont été converties en intensité diffusée en fonction du temps,  $I(t) = f(t)$  pour chaque angle. Il en résulte un graphique similaire à celui d'un chromatogramme, c'est-à-dire une intensité en fonction du temps. Pour des raisons de clarté, seules les données pour un unique angle sont représentées dans la **Figure 65C**.

Dans le cas de la transformation en  $I(t) = f(t)$ , les intensités de diffusion sont suivies en fonction du temps et pour chaque angle, et sont directement proportionnelles à la quantité mais aussi à la taille de l'espèce moléculaire. Ainsi les grosses particules qui contaminent la solution d'ED-3 ont une forte influence même si elles sont très minoritaires. Elles contribuent ainsi grandement aux intensités diffusées par la forme soluble d'ED-3. C'est dans cette optique que la déconvolution des intensités de diffusion est la plus utile et pertinente. En ajustant des courbes Gaussiennes, on peut extraire uniquement les intensités diffusées par une seule espèce (**Figure 65C**). C'est le cas ici pour la solution d'ED-2 qui contient un petit fragment de dégradation, celle d'ED-3 qui est contaminée par deux espèces de plus grande taille, et enfin celle d'ED-4 qui contient les intensités diffusées des formes monomérique et dimérique.

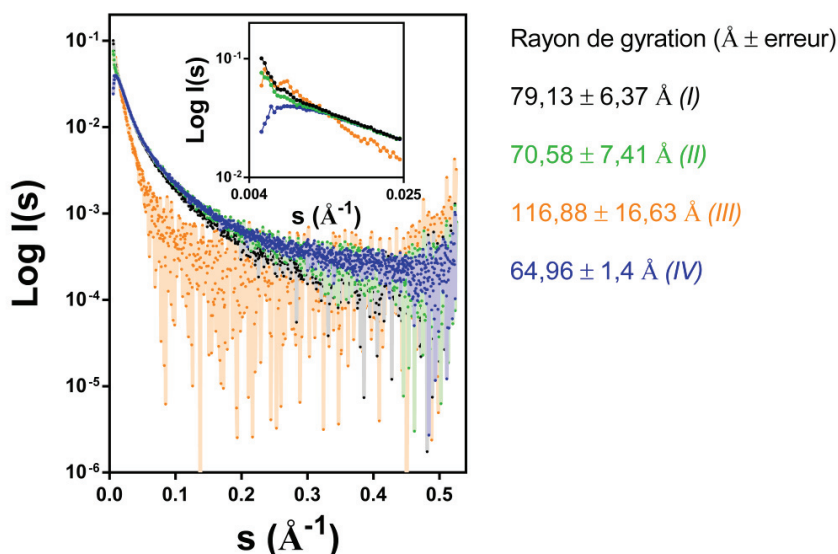


**Figure 65 : Déconvolution des intensités de diffusion collectées pour les ectodomains des syndécans lors des expériences de SEC-SAXS. A)** Chromatogramme enregistré lors de la chromatographie d'exclusion stérique. **B)** Intensité de diffusion et estimation du rayon de gyration en fonction des acquisitions séquentielles. **C)** Déconvolution des intensités de diffusion en fonction du temps par *US-SOMO*. Les données issues d'un unique angle sont représentées pour des raisons de clarté (entre 0,01 et 0,02 Å<sup>-1</sup>). Les courbes colorées correspondent aux formes solubles et monomériques des EDs. Les formes monomérique et dimérique d'ED-4 sont en violet et rouge, respectivement. La qualité de l'ajustement est représentée par le  $\chi^2$ .

Au moins une quarantaine d'acquisitions issues de cette déconvolution ont ensuite sélectionnées et comparées entre elles par un test de corrélation, le *CorMap* (Franke et al., 2015). A l'inverse du  $\chi^2$  qui évalue la similarité entre deux courbes de façon globale, le test *CorMap* est très sensible aux différences locales et est ainsi très restrictif. Environ 30 acquisitions ont été sélectionnées puis moyennées. Cette courbe de diffusion moyennée a ensuite été soumise au programme *SHANUM* (Konarev & Svergun, 2015) qui détermine la gamme angulaire contenant toutes les informations structurales de la protéine, et élimine notamment des données trop bruitées et inutilisables, typiquement aux grands angles. Les quelques points aux très petits angles qui diffèrent de l'approximation de Guinier ont été supprimés conduisant à une courbe de diffusion finale qui sera utilisée pour l'analyse.

Afin de montrer l'utilité de ce processus, les données acquises pour ED-3 présentée ci-dessous ont été analysées de différentes manières (**Figure 66**). Le 1<sup>er</sup> traitement (I) a été effectué en sélectionnant la totalité des acquisitions issues de l'élution d'ED-3, le 2<sup>ème</sup> (II) en sélectionnant manuellement des acquisitions spécifiques à une seule espèce en utilisant le programme *CHROMIXS* (Panjkovich & Svergun, 2017) et le 3<sup>ème</sup> (III) en utilisant la courbe de diffusion automatiquement traitée et générée par le programme Foxtrot de la ligne de lumière Swing (Synchrotron SOLEIL). Dans cet exemple, le programme a automatiquement sélectionné le contaminant élué en amont d'ED-3, du fait de sa forte diffusion. La courbe représentant les données traitées par *US-SOMO* (IV), comparées entre elles par *CorMap* et moyennées (avant le traitement par *SHANUM* et l'élimination de quelques points aux petits angles) est également représentée (**Figure 66**).

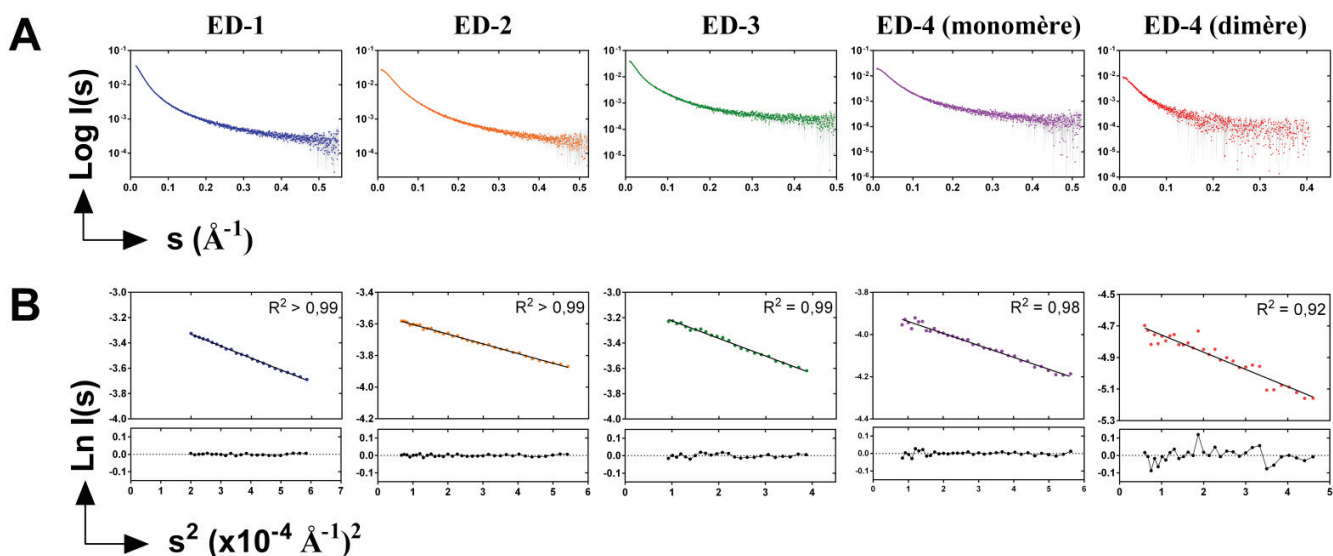
La première différence repose sur l'estimation du rayon de gyration qui diffère d'au moins 5 Å. Le deuxième point est que la totalité de la courbe est affectée selon la méthode, ce qui est particulièrement visible aux grands angles (0,3 – 0,5 Å<sup>-1</sup>). Ces analyses simples montrent l'utilité de la déconvolution effectuée par *US-SOMO* dans le but d'obtenir des paramètres structuraux fiables.



**Figure 66 : Rayon de gyration calculé par différentes méthodes d'analyse des données de SEC-SAXS.** Les intensités collectées pour ED-3 ont été traitées en sélectionnant arbitrairement toutes (noir, I) ou certaines (vert, II) acquisitions d'ED-3. La courbe générée par Foxtrot est en orange (III) et la courbe finale traitée comme détaillé ci-dessus est en bleu (IV). Insert : Agrandissement de la région de Guinier. Les rayons de gyration ont été déterminés par *AutoRg* (Primus) avec  $s.Rg < 1,3$ . Les techniques sont indiquées en chiffres romains (I à IV) et rappelées dans le corps du texte.



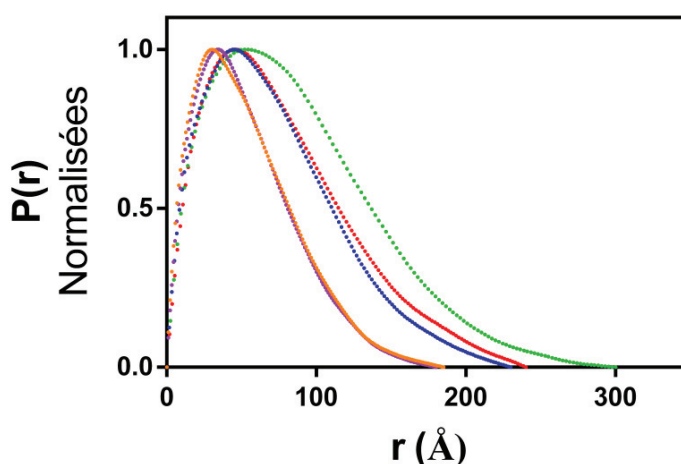
Les courbes finales pour chaque ectodomaines des syndécans ainsi que pour les deux formes d'ED-4 sont présentées dans la **Figure 67A**. L'analyse des intensités de diffusion aux petits angles selon l'approximation de Guinier a permis de déterminer le rayon de gyration d'ED-1 ( $53,3 \pm 0,30 \text{ \AA}$ ), d'ED-2 ( $42,9 \pm 0,33 \text{ \AA}$ ), d'ED-3 ( $65,0 \pm 0,62 \text{ \AA}$ ), de la forme monomérique d'ED-4 ( $41,53 \pm 0,67 \text{ \AA}$ ) ainsi que de sa forme dimérique ( $59,25 \pm 2,27 \text{ \AA}$ ), ce-dernier étant compris entre celui d'ED-1 et d'ED-3 (**Figure 67B** et **Tableau XVI**). ED-2 et la forme monomérique d'ED-4 possèdent des Rg similaires, ce qui est cohérent avec leur nombre d'acides aminés très proches (154 et 163 résidus, respectivement), ce qui montre qu'ils ont une forme globale similaire.



**Figure 67 : Détermination du rayon de gyration des 4 ectodomaines des syndécans. A)** Courbes de diffusions obtenues lors des expériences de SEC-SAXS et après traitement des données. **B)** Détermination du Rg aux petits angles selon l'approximation de Guinier. La qualité des régressions linéaires est indiquée par le coefficient de détermination R<sup>2</sup> ainsi que par les valeurs résiduelles respectives.

La dimension maximale des protéines (Dmax) a été déduite de la fonction de distribution de paires P(r) (**Figure 68**). Ces dimensions sont proportionnelles au nombre de résidus des 4 EDs. ED-3 possède le Dmax le plus élevé (300 Å) qui est supérieur à celui d'ED-1 (230 Å). Comme attendu, ED-2 et la forme monomérique d'ED-4 possèdent des Dmax très similaires (185 et 180 Å, respectivement). Enfin, le dimère d'ED-4 possède un Dmax de 240 Å, compris entre celui d'ED-1 et d'ED-3, confortant l'hypothèse d'une conformation étendue proposée à l'issue de l'analyse des données hydrodynamiques des ectodomaines des syndécans (**Tableau XVI**).

Ces valeurs de  $D_{max}$  reflètent la conformation la plus étendue des ectodomains des syndécans au moment de l'analyse mais dépendent également de l'abondance des différentes conformations. Une population de conformation en faible abondance sera ainsi difficilement observable. Il est donc plus raisonnable de considérer que ces  $D_{max}$  reflètent les dimensions de la population la plus abondante des conformations étendues mais n'empêche pas la présence de conformations beaucoup étendues ou au contraire plus compactes. Il est intéressant de constater une forte similitude dans les fonctions de distribution de paires d'ED-2 et de la forme monomérique d'ED-4.



**Figure 68 : Détermination des dimensions maximales des ectodomains des syndécans à partir de la fonction de distribution de paires  $P(r)$ .** Les  $P(r)$  ont été calculées par *GNOM*. La qualité des  $P(r)$  a systématiquement été évaluée avec le test *CorMap*. ED-1, -2 et -3 sont en bleu, orange et vert, respectivement, alors que les formes monomériques et dimériques d'ED-4 sont en violet et rouge, respectivement.

Les valeurs des rayons de gyration ont ensuite été comparées à celles des  $R_g$  théoriques calculés pour des protéines globulaires, des états de pré-globules fondus (PMG) ou bien extrêmement étendus (*random coil*). Ces calculs reposent sur l'équation de Flory reliant la valeur du rayon de gyration théorique au nombre de résidus d'acides aminés ainsi qu'au degré de compaction (voir la section *Matériel & Méthodes*). Les rayons de gyration expérimentaux sont très proches de ceux calculés pour des protéines ayant une conformation très étendue (*random coil*) (**Tableau XVI**). Ces observations sont confortées par le calcul du facteur de forme (ou *shape factor*) qui est le rapport de la valeur du rayon de gyration et de celle du rayon hydrodynamique, mesurée précédemment par SEC-DLS. Ce rapport est égal à environ 0,778 pour une sphère parfaite et augmente graduellement avec l'allongement du polymère, jusqu'à une valeur d'environ 2,36 pour une forme en bâton rigide (*Brewer & Striegel, 2011*).

La valeur de ce facteur pour les ectodomains des syndécans est supérieure à l'unité et est compris entre 1,19 (ED-1) et 1,92 pour la forme monomérique d'ED-4, confortant l'idée de conformations étendues pour ces protéines.

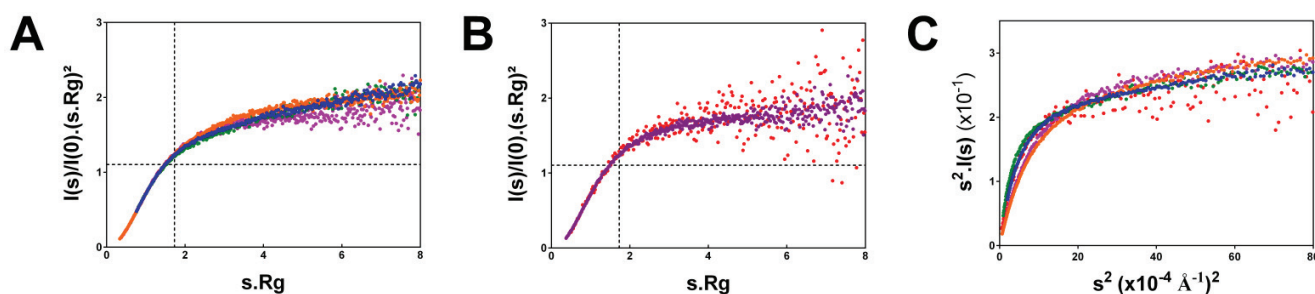
	ED-1	ED-2	ED-3	ED-4 (monomère)	ED-4 (dimère)
<b>Rg (Å)</b> (± erreur)	<b>53,30 ± 0,3</b>	<b>42,90 ± 0,33</b>	<b>65,00 ± 0,62</b>	<b>41,53 ± 0,67</b>	<b>59,25 ± 2,27</b>
<b>Rg théoriques (Å, équation de Flory)</b>					
<b>Protéine globulaire</b>	18,5	15,9	20,5	16,1	19,7
<b>PMG</b> (± erreur)	46,3 ± 0,6	35,2 ± 0,4	55,9 ± 0,8	36,3 ± 0,5	52,1 ± 0,7
<b>Random coil</b> (± erreur)	54,2 ± 11,8	39,6 ± 7,8	67,4 ± 15,7	40,9 ± 8,2	62,2 ± 14,1
<b>Forme globale</b>					
<b>Shape factor Rg/Rh</b>	1,19	1,26	1,30	1,92	1,31

**Tableau XVI: Rg expérimentaux (SEC-SAXS) et théoriques des ectodomains des syndécans calculés à partir de l'équation de Flory.** La forme globale des protéines a été évaluée à partir du facteur de forme Rg/Rh, la valeur des rayons hydrodynamiques ayant été mesurée par SEC-DLS.

*b) Flexibilité des ectodomains des syndécans*

Le degré de compaction et la flexibilité d'une protéine peuvent être également appréciés sur la représentation normalisée de Kratky  $I(s)/I(0) \times (s.Rg)^2 = f(s.Rg)$ . Une protéine globulaire présente une courbe en cloche dont le maximum atteint environ 1,1 à  $s.Rg = \sqrt{3}$ . Cette forme de courbe caractéristique s'estompe graduellement avec la diminution du degré de compaction. Elle atteint un plateau aux alentours de 2-3 pour des protéines flexibles comme des protéines globulaires dénaturées ainsi que les protéines intrinsèquement non structurées, et forme une courbe quasi-linéaire dans le cas de conformation en bâton rigide (voir section *Matériel & Méthodes*).

Les représentations normalisées de Kratky des 4 ectodomaines des syndécans sont très semblables. Les courbes croissent graduellement et s'infléchissent jusqu'à former un plateau aux alentours de 2 caractérisant les protéines intrinsèquement non structurées et flexibles (**Figure 69A**). Le dimère d'ED-4 conserve les mêmes caractéristiques que la forme monomérique puisque les courbes semblent être superposables bien que les données du dimère soient plus bruitées (**Figure 69B**). Comme attendu pour ce type de protéines, les intensités de diffusions décroissent en  $s^{-2}$  comme en atteste la présence d'un plateau dans la représentation de Kratky-Debye, alors que des courbes représentant des niveaux de compaction plus élevés tendent à décroître en  $s^{-3}$  pour les protéines multidomaines, ou  $s^{-4}$  pour les protéines globulaires (**Figure 69C**).



**Figure 69 : Evaluation du degré de compaction et de la flexibilité des 4 ectodomaines des syndécans déterminés par SEC-SAXS.** Comparaison des représentations normalisées de Kratky des 4 EDs (**A**) et des formes monomérique et dimérique d'ED-4 (**B**). **C** Représentation de Kratky-Debye. Légende : ED-1 (bleu), ED-2 (orange), ED-3 (vert), ED-4 monomère (violet), ED-4 dimère (rouge).

Le calcul de l'exposant de Porod permet de déterminer avec précision la façon dont les intensités de diffusion décroissent. Comme observé précédemment dans la représentation de Kratky-Debye, ces exposants sont proches de 2 et sont égaux à 1,7 pour ED-1, ED-3 et la forme monomérique d'ED-4, 1,8 pour ED-2 et 1,9 pour la forme dimérique d'ED-4. Une fois de plus, ces valeurs sont directement reliées à la forme globale de la protéine étant donné que la valeur de l'inverse de l'exposant de Porod doit être proche, sinon égale, au facteur exponentiel (ou exposant  $\nu$ ) retrouvé dans l'équation de Flory. L'inverse de l'exposant de Porod pour les ectodomaines des syndécans est compris entre 0,526 et 0,588 soit des valeurs très proches de l'exposant de Flory pour des conformations extrêmement étendues (0,598).

### c) Conformations préférentielles adoptées par les ectodomains des syndécans

Etant donné que les ectodomains des syndécans sont des protéines intrinsèquement non structurées et sont très flexibles, il est difficile d'utiliser les outils courant de modélisation 3D qui réduiraient l'hétérogénéité des conformations à une unique structure moyenne. Cependant, il est possible d'étudier l'espace conformationnel de ces protéines par déconvolution des données de SEC-SAXS en utilisant notamment l'outil *Ensemble Optimization Method (EOM)* (Bernado et al., 2007, Tria et al., 2015).

Trois étapes clés régissent le calcul réalisé par EOM. Dans un premier temps, un large *pool* de conformations, typiquement 10 000, est créé à partir de la séquence en acides aminés de la protéine afin de définir l'espace conformationnel qui lui est théoriquement offert et ce, sans restriction. Il est possible de générer différents *pool* initiaux dont les distributions des angles diédraux sont cohérentes avec des conformations retrouvées dans les protéines intrinsèquement non structurées ou de générer des conformations beaucoup plus étendues de type *random coil*. Des conformations *compactes* peuvent également être générées artificiellement. Une courbe de diffusion théorique est ensuite calculée par *CRY SOL* (Svergun et al., 1995) pour chaque conformation présente dans le *pool*. Enfin, une sous-sélection de conformations issues du *pool* (appelé *ensemble*) est sélectionnée de telle sorte que les intensités de diffusion résultantes de cet *ensemble* soient identiques aux données de diffusion expérimentales (voir section *Matériel & Méthodes*). L'analyse des distributions du rayon de gyration et du  $D_{max}$  de cet *ensemble* sélectionné par rapport à celles du *pool* initial de conformations permet d'évaluer l'espace conformationnel de la protéine étudiée ainsi que sa flexibilité. Les paramètres  $R_{flex}$  (0 – 100 %), qui caractérise un système flexible, et  $R_{\sigma}$  qui permet notamment d'évaluer la qualité de la déconvolution, sont déduits.

Deux *pools* initiaux composés de conformations étendues (*random coil*, RC) et un *pool* de conformations plus compactes (PMG) ont été générés lors de la déconvolution (**Tableau XVII** et **Figure 70**). La courbe de diffusion théorique résultante de l'ensemble sélectionné pour ces deux modèles a été comparée à la courbe de diffusion expérimentale de chaque ectodomains des syndécans par le test *CorMap*. Une valeur de *CorMap* inférieure à 0,01 indique un mauvais ajustement entre les deux jeux de données comparés.

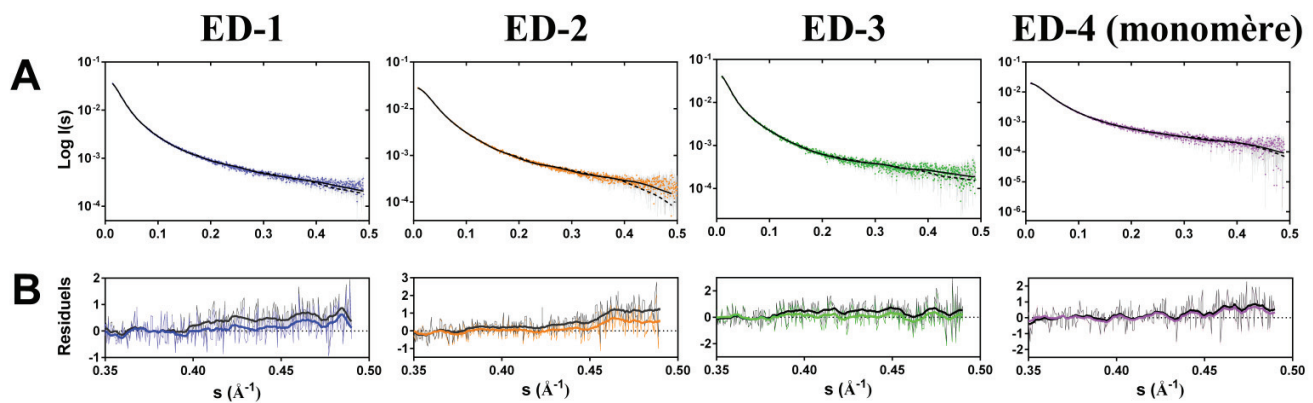
Le modèle PMG, qui contient des conformations plus compactes que le modèle RC, dévie systématiquement de la courbe expérimentale, notamment aux grands angles ( $> 4 \text{ \AA}^{-1}$ ) comme en attestent les valeurs résiduelles (**Tableau XVII** et **Figures 70A** et **70B**).

A l'inverse et pour chacun des EDs, le modèle *Random coil* reproduit parfaitement la courbe de diffusion dans toute la gamme angulaire. C'est donc ce modèle qui a été utilisé pour la suite de l'analyse de la déconvolution par EOM.

	ED-1	ED-2	ED-3	ED-4 (monomère)
<b>Modèle pré-globule fondu (PMG)</b>				
$\chi^2$	0,904	0,947	0,765	0,636
<b>CorMap (<i>p-value</i>)</b>	0,0002	0,000	0,008	0,004
<b>Modèle random-coil (RC)</b>				
$\chi^2$	0,616	0,632	0,663	0,597
<b>CorMap (<i>p-value</i>)</b>	0,224	0,227	0,0645	0,0618
<b>Rflex ensemble (%)</b>	82,71	86,17	81,18	84,02
<b>Rflex pool (%)</b>	84,31	85,60	84,59	83,69
<b>R<math>\sigma</math></b>	0,92	2,47	0,85	1,01

**Tableau XVII : Déconvolution des données de SEC-SAXS collectées pour les ectodomains des syndécans par EOM avec les modèles PMG et RC.** Pour chaque modèle, la qualité de l'ajustement est indiquée ( $\chi^2$  et la *p-value* issus du test *CorMap*) entre la courbe expérimentale et la courbe théorique de l'ensemble sélectionné, soit du pool de conformères très étendus (*Random-coil*) ou plus compacts (pré-globule fondu). Les paramètres Rflex et R $\sigma$  sont issus du modèle *Random coil*.

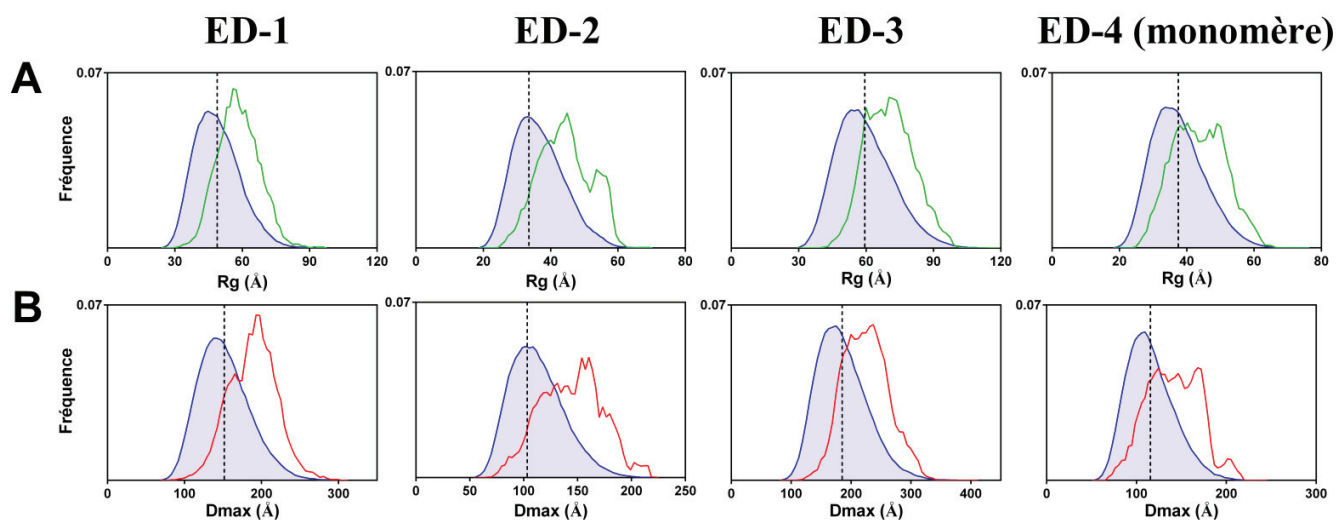
Le paramètre R $\sigma$  permet d'évaluer la qualité de la déconvolution. Si la valeur de Rflex de l'ensemble sélectionné est inférieure au Rflex du pool, R $\sigma$  doit être inférieur à 1 et réciproquement, ce qui est le cas pour chacune des déconvolutions. Pour chaque ectodomaine, le paramètre Rflex, qui est directement proportionnel à l'entropie du système, est très élevé (81 – 86 %) montrant l'aspect très flexible et l'hétérogénéité conformationnelle de ces protéines. Il est également relativement similaire à celui du *pool* initial ce qui est cohérent avec la nature très étendue des conformations de type *random coil*.



**Figure 70 : Déconvolution des données de SEC-SAXS des ectodomaines des syndécans par EOM. A)** Ajustements de la courbe de diffusion théorique pour des conformations très étendues (*random coil*, trait noir plein) et plus compactes (pré-globule fondu, trait noir en pointillé) aux données de diffusion expérimentales (en couleur). **B)** Valeurs résiduelles des modèles *random coil* (en couleur) et pré-globule fondu PMG (noir). Les résiduels ont été lissés pour ces deux modèles pour plus de clarté (courbes en gras), la légende des couleurs restant identique.

Les distributions des ensembles sélectionnés pour chacun des EDs sont décalées vers des valeurs plus importantes que celles du *pool* (**Figures 71A et 71B**). Les distributions d'ED-1 et d'ED-3 sont assez homogènes et centrées sur des valeurs de rayon de gyration de 56 et 70 Å, respectivement, similaires à celles observées lors de l'analyse des données de SEC-SAXS (53,3 et 65,0 Å, respectivement). Les valeurs de Dmax sont environ centrées respectivement sur 193 et 225 Å, ce qui est en revanche bien inférieur aux valeurs précédemment déduites de la fonction de distribution de paires (230 et 300 Å, respectivement). Cette différence n'est pas inattendue étant donné que si une population de conformations possède une valeur de Dmax très élevée mais en quantité extrêmement minoritaire, elle ne sera pas visible sur la P(r). C'est pourquoi certaines conformations atteignent des dimensions maximales d'environ 299 et 337 Å pour ED-1 et ED-3, lors de la déconvolution par EOM (**Tableau XVIII**).

Concernant ED-2 et la forme monomérique d'ED-4, les distributions des paramètres sont plus complexes. La population majoritaire d'ED-2 possède un rayon de gyration d'environ 45 Å, proche de celui observé précédemment de 43 Å, alors qu'ED-4 possède deux populations peu définies centrées sur environ 40 et 49 Å, différentes de la valeur précédemment observée de 42 Å pour la seconde population. Les distributions du Dmax sont centrées sur environ 157 Å pour ED-2 et environ sur 146 et 201 Å pour ED-4 (**Tableau XVIII**).



**Figure 71 : Distributions des valeurs de rayons de gyration et de Dmax des 4 EDs obtenues par la déconvolution effectuée par EOM avec un modèle *random coil*.** Distribution des rayons de gyration **(A)** et des Dmax **(B)** des *ensembles* (courbes vertes et rouges, respectivement) et des *pools* respectifs (courbe bleue pleine) pour chaque EDs. La ligne verticale noire en pointillés indique la moyenne des valeurs de Rg et Dmax du *pool*.

Les gammes de Rg et de Dmax explorées par les conformères présents dans les ensembles sélectionnés par EOM sont assez étendues. Les rayons de gyration varient d'un facteur 2,6 à 3,0 et il en est de même pour les Dmax (de 3 à 3,6). D'une façon générale, on peut ainsi considérer que la taille globale des conformères des 4 ectodomaines des syndécans varie du simple au triple. Cette observation confirme d'une part l'aspect très flexible de ces protéines et d'autre part, elle suggère fortement que ces protéines ont la capacité intrinsèque à adopter des populations compactes ou extrêmement étendues.

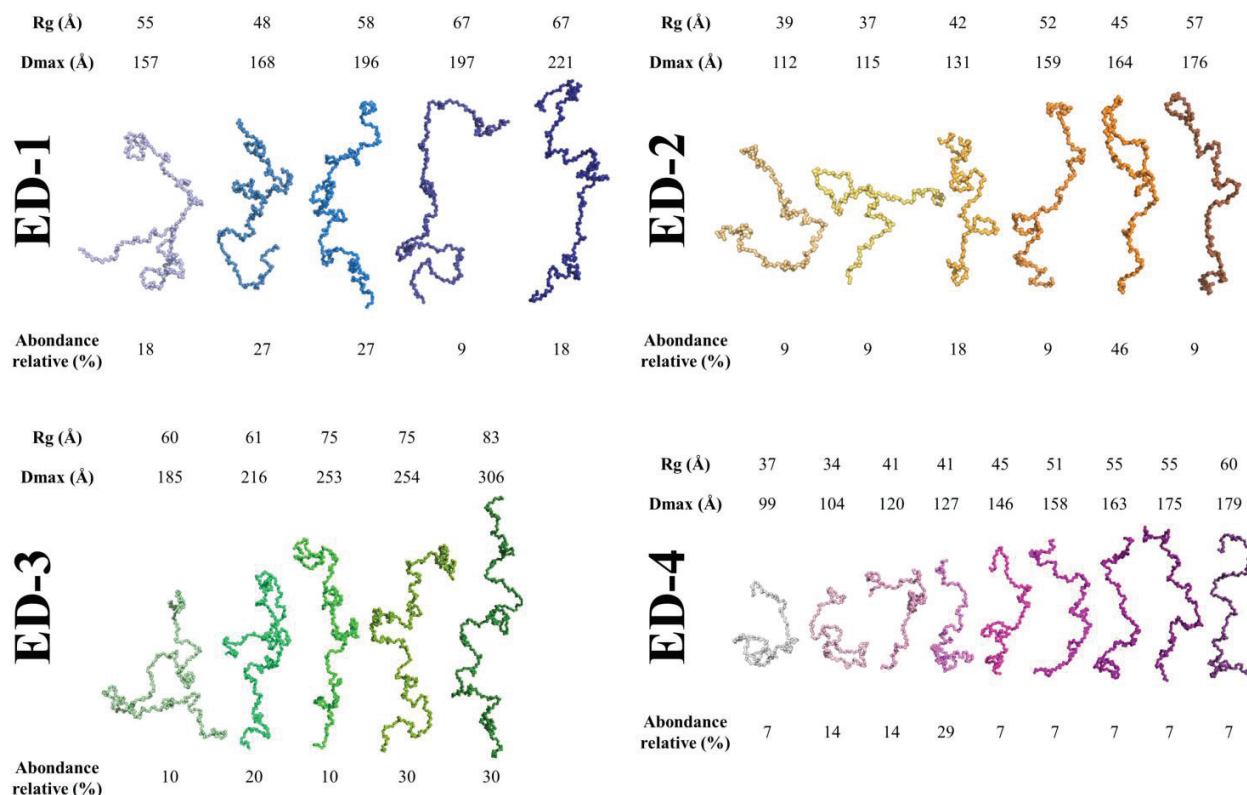
Les valeurs du rayon de gyration des conformères les plus compacts, qui sont d'environ 30 Å pour ED-1, 40 Å pour ED-3 et 25 Å pour ED-2 et la forme monomérique d'ED-4, restent cependant 1,3 à 2 fois plus importantes que celles calculées avec l'équation de Flory pour des protéines globulaires ayant le même nombre de résidus d'acides aminés que les EDs. Aussi, les valeurs de Rg issus d'EOM des populations les plus étendues sont nettement plus importantes que celles calculées avec cette même équation.



	ED-1	ED-2	ED-3	ED-4 (monomère)
<b>Rayon de gyration (Å)</b>				
<b>Expérimental (± erreur)</b>	53,30 ± 0,3	42,90 ± 0,33	65,00 ± 0,62	41,53 ± 0,67
<i>Distribution (EOM)</i>				
<b>Population majoritaire</b>	56	45	70	40 et 49
<b>Ensemble sélectionné</b>	30 - 96	25 - 69	40 - 103	25 - 66
<b>Moyenne pondérée</b>	59,0	44,7	71,2	44,2
<b>Dimension maximale (Å)</b>				
<b>Expérimental P(r)</b>	230	185	300	180
<i>Distribution (EOM)</i>				
<b>Population majoritaire</b>	193	157	225	146 - 201
<b>Ensemble sélectionné</b>	83 - 299	67 - 222	113 - 337	69 - 216
<b>Moyenne pondérée</b>	191,0	145,7	227,8	142,7

**Tableau XVIII : Détails des distributions de rayon de gyration et de la taille maximale des particules (Dmax) issues de la déconvolution par EOM avec un modèle *random coil*.** Les valeurs des paramètres structuraux de l'ensemble sélectionné représentent la gamme de Rg ou de Dmax où la fréquence d'apparition des conformations est non nulle. La moyenne des paramètres structuraux est pondérée par la fréquence d'apparition d'un conformère donné. Les valeurs expérimentales du Rg et du Dmax sont indiquées à titre de comparaison.

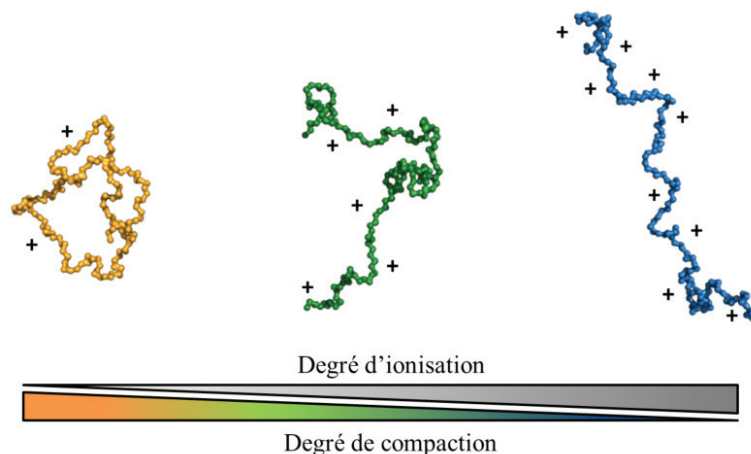
A titre d'illustration, un petit nombre de conformères représentant l'espace conformationnel des EDs des syndécans a été modélisé par EOM. L'addition des intensités de diffusion théoriques de 5 conformères pour ED-1 et ED-3, 6 pour ED-2 et 9 pour la forme monomérique d'ED-4 dans une proportion donnée reproduit ainsi les données expérimentales (**Figure 72**).



**Figure 72 : Modèles générés par EOM.** Les paramètres Rg et Dmax ainsi que l'abondance relative de chaque conformère sont indiqués.

Les résultats des expériences de SEC-SAXS indiquent que les ectodomains des syndécans ont des caractéristiques similaires à celles des polymères très étendus et la déconvolution réalisée par EOM montre qu'ils pourraient adopter des conformations plus compactes ou très étendues dont la taille varie d'un facteur 3. Ces populations sont difficilement détectables en SAXS du fait de leur faible abondance. L'objectif des expériences suivantes a donc été de vérifier la présence de ces populations et de déterminer approximativement l'abondance relative de chaque sous-population de conformations. Des expériences de spectrométrie de masse (ESI-MS) en conditions natives ont été réalisées dans ce but. Comme rappelé précédemment, le terme de *conditions natives* regroupe des conditions dans lesquelles les protéines ne se fragmentent pas et conservent leur structure intacte ainsi que leurs interactions non covalentes (on parle d'ionisation douce). Il est également nécessaire que le tampon ne dénature pas la protéine et soit volatil (*Santambrogio et al., 2018*).

Dans ces expériences, le tampon HBS a été échangé contre une solution d'acétate d'ammonium 50 mM pH 7,4 par chromatographie d'exclusion stérique puis les protéines ont été transférées en phase gazeuse par électronebulisation. Leur degré d'ionisation est directement relié à leur structure tridimensionnelle, notamment leur surface, et n'est pas dépendant de la nature des résidus d'acides aminés exposés à leur surface (*Kaltashov & Mohimen, 2005*). Pour les protéines intrinsèquement non structurées, une conformation étendue acquerra donc davantage de charges qu'une conformation plus compacte (**Figure 73**).



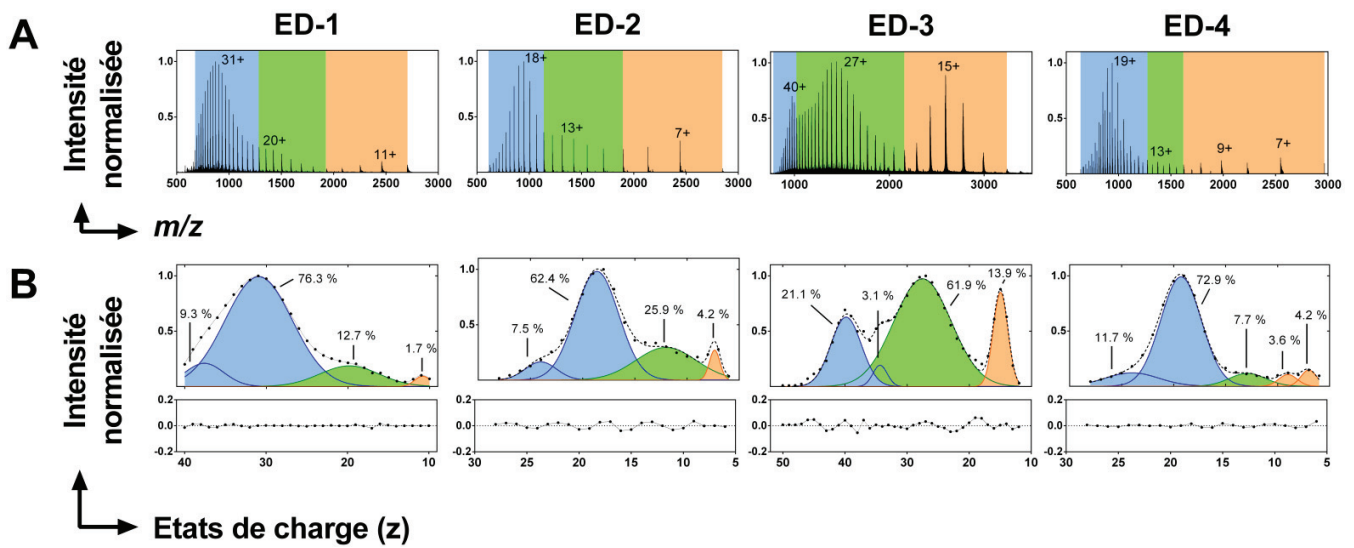
**Figure 73 : Principe de l'utilisation de l'ESI-MS dans le cadre d'une protéine désordonnée.**

Le degré d'ionisation est inversement proportionnel au degré de compaction de la protéine étudiée ce qui permet de discriminer d'éventuelles sous-populations de conformations.

Le résultat de ces expériences est représenté sous forme d'une distribution des états de charge (CSD) c'est-à-dire un graphique représentant l'intensité de chaque ion détecté en fonction de son rapport masse sur charge ( $m/z$ ). Une distribution unimodale est observée pour une protéine globulaire étant donné que l'ensemble des conformations qu'elle peut adopter est restreint. En revanche, l'hétérogénéité de cette distribution augmente avec la présence de désordre intrinsèque, menant à des distributions bi- ou multimodales. L'analyse du CSD permet ainsi de détecter la présence de sous-populations de conformations qui diffèrent par leur état de compaction (*Testa et al., 2011, Testa et al., 2013, Schramm et al., 2019*).

Une distribution multimodale a systématiquement été détectée pour chacun des 4 ectodomaines des syndécans ce qui est cohérent avec leur enrichissement en désordre intrinsèque (**Figure 74A**). Trois populations d'abondances relatives différentes ont été détectées sur la distribution des états de charge suggérant que les ectodomaines des syndécans peuvent adopter des conformations préférentielles. Une population d'ions très chargés, reflétant des conformations étendues, est majoritaire pour ED-1 (centrée sur 31+), ED-2 (centrée sur 18+) et la forme monomérique d'ED-4 (centrée sur 19+). Les conformations les plus étendues pour ED-3 (centrées sur 40+) ne sont en revanche pas les plus abondantes. Une faible abondance d'ions peu chargés a également été détectée, reflétant des conformations compactes, centrée sur 11+ pour ED-1 et 7+ pour ED-2. Concernant ED-4, une plus large gamme de faibles états de charge s'étale de 7+ à environ 11+. ED-3 possède en revanche la proportion de conformations compactes la plus élevée, centrée sur 15+ (**Figure 74A**). Une troisième population possédant des états d'ionisation intermédiaires a été observée. Cette population d'abondance moyenne est centrée à 13+ pour ED-1 et est majoritaire pour ED-3, centrée sur 27+. ED-2 et la forme monomérique d'ED-4 partagent une population intermédiaire d'intensité moyenne aux alentours de 13+.

Ces différentes populations ont été quantifiées par déconvolution des CSDs. Elle a été réalisée en représentant l'intensité normalisée en fonction du nombre de charges (**Figure 74B**). Sous chaque population, une courbe Gaussienne a été ajustée et son aire a été calculée. Le rapport de l'aire d'une population par l'aire totale de l'ensemble des populations donne ainsi une estimation des proportions relatives de chaque population (voir section *Matériel & Méthodes*). Les conformations les plus étendues sont les plus abondantes pour ED-1 ( $\approx 86\%$ ), ED-2 ( $\approx 70\%$ ) et ED-4 ( $\approx 85\%$ ). En revanche, cette population ne représente qu'environ 21 % pour ED-3, dont le CSD est dominé par une population intermédiaire représentant environ 62 % des conformations totales. Les populations les plus compactes sont en revanche minoritaires pour ED-1 et ED-2 (2 – 4 %), et légèrement plus importante pour ED-4 ( $\approx 8\%$ ). Une proportion plus importante de conformations compactes est en revanche détectée pour ED-3 ( $\approx 14\%$ ).



**Figure 74 : Analyses des ectodomaines des syndécans par ESI-MS en conditions natives. A)** Distribution des états de charge (CSD) des 4 EDs. Les populations sont colorées en bleu (conformations étendues), vert (conformations intermédiaires) et orange (conformations compactes). Les états de charge sur lesquels sont centrées ces populations sont indiqués. **B)** Déconvolution des CSDs. Les proportions (%) de chaque population colorées comme en A) sont indiquées.

Pour ED-4, le CSD est peuplé de pics intermédiaires intenses et le calcul de leur charge à partir de la masse de la protéine et du rapport  $m/z$  mène à des espèces possédant des demi-charges. Cette observation trahit la présence de dimère, ce qui est attendu pour ED-4. Des pics intermédiaires de faible intensité sont également détectés pour ED-3 et le dimère possède entre 33 et 73 charges. L'état de charge 45+ possède l'intensité la plus élevée et correspond à environ 10 % de l'intensité de l'état de charge 27+ du monomère. Une observation similaire est réalisée pour ED-1, mais l'intensité des pics intermédiaires s'étendant de 29+ à environ 49+ est très faible, où l'état de charge 45+ ne représente que 3 % de l'intensité de l'état de charge 31+ de la forme monomérique.

A partir des états de charges détectés ainsi que de leur intensité, il est possible d'estimer la surface accessible au solvant (SASA) des ectodomaines des syndécans. Cette surface, exprimée en  $\text{\AA}^2$ , est directement proportionnelle à la charge moyenne acquise durant le processus d'ionisation ( $z_{av}$ ). La valeur de SASA peut être déduite de la totalité de la distribution des états de charge afin d'obtenir une valeur moyenne ( $SASA_{av}$ ) (Li et al., 2015).

Il est également possible de calculer la charge moyenne acquise par les conformations les plus étendues ou les plus compactes afin de déterminer leur valeur de SASA ( $SASA_{ext}$  et  $SASA_{cpt}$ , respectivement) (Testa *et al.*, 2011; Kaltashov & Mohimen, 2001). Afin de comparer les ectodomains des syndécans entre eux, la valeur moyenne de SASA a été normalisée par leur masse moléculaire respective (**Tableau XIX**).

	ED-1	ED-2	ED-3	ED-4 (monomère)
<b>Charge moyenne</b> ( $z_{av}$ )	29,62+	16,93+	28,68+	18,57+
<b>SASA moyenne</b> ( $\text{Å}^2$ )	29 653,0	16 074,2	28 624,4	17 780,2
<b>SASA normalisé</b> ( $\text{Å}^2/\text{Da}$ )	1,10	0,94	0,74	1,00

**Tableau XIX: Valeur moyenne de la surface accessible au solvant ( $SASA_{av}$ ) des ectodomains des syndécans.** Les valeurs de  $SASA_{av}$  ont été normalisées par la masse moléculaire (en Da) des protéines. Les valeurs ont été calculées à partir de l'équation déterminée par Li *et al.*, (2015).

La charge moyenne acquise par ED-2 (16,93+) est similaire à celle de la forme monomérique d'ED-4 (18,57+) ce qui est cohérent avec leur nombre de résidus d'acides aminés. Les valeurs de leur SASA calculées sur l'ensemble du CSD sont alors proches ( $\approx 16\ 070$  et  $17\ 780\ \text{Å}^2$ , respectivement) et les valeurs de leurs SASA normalisés sont similaires ( $0,94$  et  $1,00\ \text{Å}^2/\text{Da}$  respectivement). La charge moyenne acquise par ED-1 est également similaire à celle d'ED-3 (29,62+ et 28,68+, respectivement) ce qui est une conséquence de l'enrichissement en conformations intermédiaires détectées dans le CSD d'ED-3, ces dernières étant moins chargées que les conformations les plus étendues. Le calcul de leur  $SASA_{av}$  mène à des valeurs proches ( $\approx 29\ 650$  et  $28\ 620\ \text{Å}^2$ , respectivement). ED-3 possède cependant environ 30 % de résidus supplémentaires par rapport à ED-1 ce qui mène à un  $SASA_{av}$  normalisé plus faible pour ED-3 ( $0,8\ \text{Å}^2/\text{Da}$ ) par rapport à ED-1 ( $1,10\ \text{Å}^2/\text{Da}$ ) (**Tableau XIX**). En moyenne, ED-3 adopte ainsi des conformations plus compactes que celles d'ED-1 en phase gazeuse. Cette observation est cohérente avec la proportion importante de conformations compactes et intermédiaires détectées pour ED-3 lors des expériences d'ESI-MS. Afin de déterminer la surface accessible au solvant des conformations les plus étendues, les conformations ayant acquises le plus de charges ont été sélectionnées et la charge moyenne de cette population a été calculée dans le but de calculer  $SASA_{ext}$  (**Tableau XX**).

	ED-1	ED-2	ED-3	ED-4 (monomère)
<b>Charge moyenne</b> ( $z_{av}$ )	31,00+	19,18+	40,71+	19,72+
<b>SASA<sub>ext</sub></b> ( $\text{Å}^2$ )	31 103,9	18 268,6	42 071,3	18 843,5
<b>SASA<sub>ext</sub> normalisé</b> ( $\text{Å}^2/\text{Da}$ )	1,15	1,07	1,08	1,06

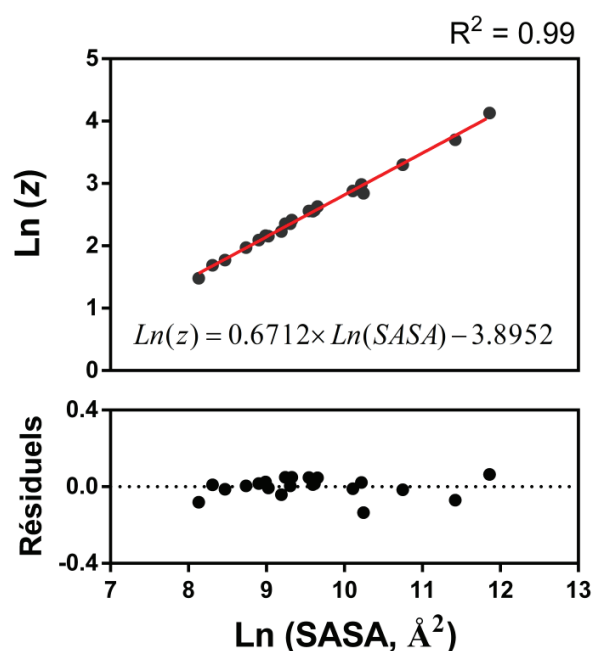
**Tableau XX : Valeur de la surface accessible au solvant des conformations les plus étendues (SASA<sub>ext</sub>) des ectodomaines des syndécans.** Les valeurs de SASA<sub>ext</sub> ont été normalisées par la masse moléculaire (en Da) des protéines. Les valeurs ont été calculées à partir de Testa et *al.*, (2011).

La charge moyenne acquise pour les conformations étendues des ectodomaines des syndécans est proportionnelle à leur taille et est égale à environ 19+ pour ED-2 et ED-4, 31+ pour ED-1 et 41+ pour ED-3. La valeur de SASA<sub>ext</sub> calculée est similaire pour ED-2 et ED-4 ( $\approx 18\,270$  et  $18\,843 \text{ Å}^2$ , respectivement) et est plus importante pour ED-3 ( $\approx 42\,070 \text{ Å}^2$ ) que pour ED-1 ( $\approx 31\,104 \text{ Å}^2$ ). Cela suggère qu'ED-3 conserve sa capacité à adopter des conformations très étendues bien qu'il puisse adopter des conformations plus compactes.

L'analyse de la distribution des états de charge ne permet pas d'évaluer le degré de compaction des conformations les plus compactes. En effet, le degré d'ionisation de chaque population permet déterminer relativement l'état de compaction des différentes conformations les unes par rapport aux autres, mais pas de manière absolue. Il est cependant possible de quantifier approximativement le degré de compaction des conformations les plus compactes en comparant les états d'ionisation de protéines globulaires possédant le même nombre de résidus d'acides aminés que les ectodomaines des syndécans.

Comme rappelé précédemment, la quantité de charge acquise dépend uniquement de la surface accessible. Ainsi, deux protéines de même taille et de degré de compaction similaire acquerront le même nombre de charges pendant le processus d'ionisation. Les valeurs de la charge moyenne et de SASA d'une vingtaine de protéines globulaires ont été déterminées expérimentalement (*Kaltashov & Mohimen, 2001*). Ce jeu de données inclut des protéines possédant le même nombre de résidus d'acides aminés, ou un nombre très proche, que les ectodomaines des syndécans.

Dans un premier temps, le jeu de données expérimentales obtenues par *Kaltashov & Mohimen* (2001) sur des protéines globulaires a été tracé afin de déterminer la relation proportionnelle entre la charge moyenne des conformations et la valeur de SASA (**Figure 75**). La valeur de la charge moyenne des conformations compactes des EDs des syndécans a ensuite été calculée et comparée à celle des protéines globulaires (**Tableau XXI**). Si ces valeurs sont très proches, alors leurs conformations le sont également. Enfin, la valeur de SASA pour les conformations compactes ( $SASA_{cpt}$ ) des ectodomaines des syndécans a été calculée à partir de l'équation de la courbe (**Figure 75**).



**Figure 75 : Relation entre la charge moyenne et les valeurs de SASA pour des protéines globulaires.** Le jeu de données est issu de *Kaltashov & Mohimen* (2001). La qualité de la régression linéaire ( $R^2$ ) et l'équation de la droite sont indiquées sur le graphique ainsi que les valeurs résiduelles.

La régression linéaire effectuée sur le jeu de données expérimentales montre une parfaite corrélation entre la charge moyenne et la valeur de SASA pour des conformations globulaires. Parmi le jeu de données, l'anhydrase carbonique 1, la myoglobine, l'ovalbumine et la  $\beta$ -lactoglobuline A possèdent un nombre de résidus en acides aminés identique ou très proche de celui des 4 EDs des syndécans (**Tableau XXI**). Le degré d'ionisation de ces protéines globulaires est similaire (à  $\pm 1$  charge pour ED-2 et ED-4 et  $\pm 2$  charges pour ED-1 et ED-3) à celui des conformations compactes des EDs des syndécans.

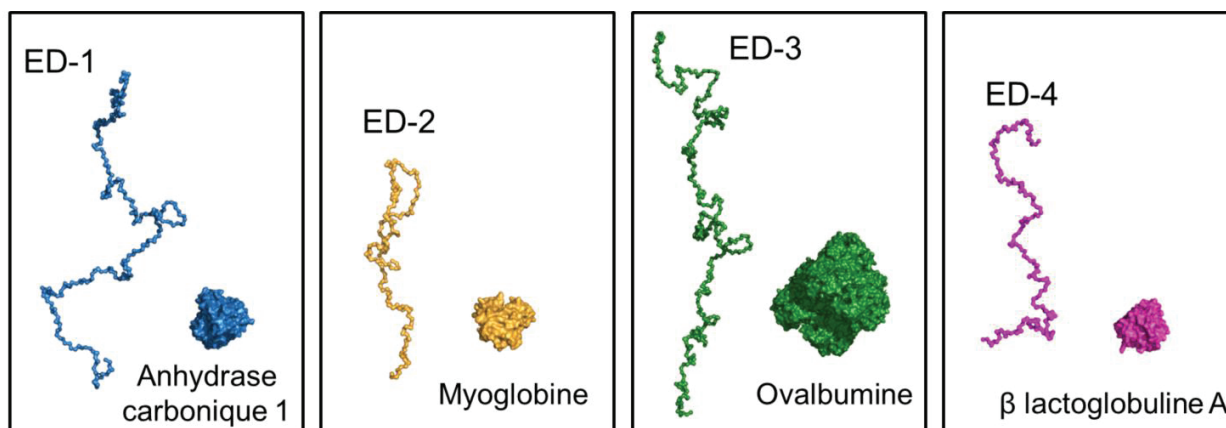


Cette observation suggère que les conformations les plus compactes des ectodomaines des syndécans correspondent à des conformations repliées, proches de celles des protéines globulaires. La valeur normalisée de  $SASA_{cpt}$  est proche pour les 4 EDs des syndécans suggérant que l'état de compaction de cette population plus compacte est similaire pour les quatre ectodomaines des syndécans.

		ED-1	ED-2	ED-3	ED-4 (monomère)
<b>Charge moyenne</b>	EDs	11,69	7,81	15,24	8,35
	Protéine globulaire	10,54	8,70	13,89	8,65
	Différence	+ 1,15	- 0,89	+ 1,35	- 0,30
<b><math>SASA_{cpt}</math> (<math>\text{\AA}^2</math>)</b>	EDs	12 912,8	7 090,0	19 178,1	7 832,4
	Protéine globulaire	10 997	8 017	15 586	8 321
<b><math>SASA_{cpt}</math> normalisé des EDs (<math>\text{\AA}^2/\text{Da}</math>)</b>		0,48	0,42	0,49	0,44
<b>Nombre de résidus d'acides aminés</b>	EDs	260	154	373	163
	Protéine globulaire	260 (Anhydrase carbonique 1)	153 (Myoglobine)	386 (Ovalbumine)	163 ( $\beta$ -lacto-globuline A)

**Tableau XXI : Détermination du degré de compaction des conformations les plus compactes des ectodomaines des syndécans (EDs) à partir des données d'ESI-MS.** Le jeu de données expérimentales est issu de *Kaltashov & Mohimen* (2001).

A titre d'illustration, la **Figure 76** présente les structures tridimensionnelles des protéines globulaires possédant un nombre de résidus similaires et des valeurs de SASA proches de celles des EDs des syndécans, et les conformères les plus étendus proposés par EOM.



**Figure 76 : Illustration du degré de compaction des conformations compactes des ectodomaines des syndécans.** Le conformère issu d'EOM ayant le Dmax le plus élevé est représenté à côté d'une protéine globulaire possédant une valeur de SASA similaire à celle des EDs. ED-1 est comparé à l'anhydrase carbonique 1 (PDB : 2CAB), ED-2 à la myoglobine (PDB : 1WLA), ED-3 à l'ovalbumine (PDB : 1OVA) et ED-4 à la  $\beta$ -lactoglobuline A (PDB : 1BSY).

#### *d) Espace conformationnel et stabilité des ectodomaines des syndécans*

Les expériences de SEC-SAXS et d'ESI-MS ont mis en évidence le caractère très flexible des ectodomaines des syndécans ce qui leur permet d'adopter des conformations très étendues de type *random coil* et certaines plus compactes. Ces résultats indiquent que les ectodomaines des syndécans possèdent un espace conformationnel étendu mais ils ne permettent pas d'en déterminer ses limites. Les conformations les plus compactes et les plus étendues sont séparées par une population intermédiaire de faible abondance pour ED-1, ED-2 et ED-4 et majoritaire pour ED-3. La présence de cette population et la flexibilité des EDs permet de suggérer qu'il existe un équilibre en faveur des conformations étendues.

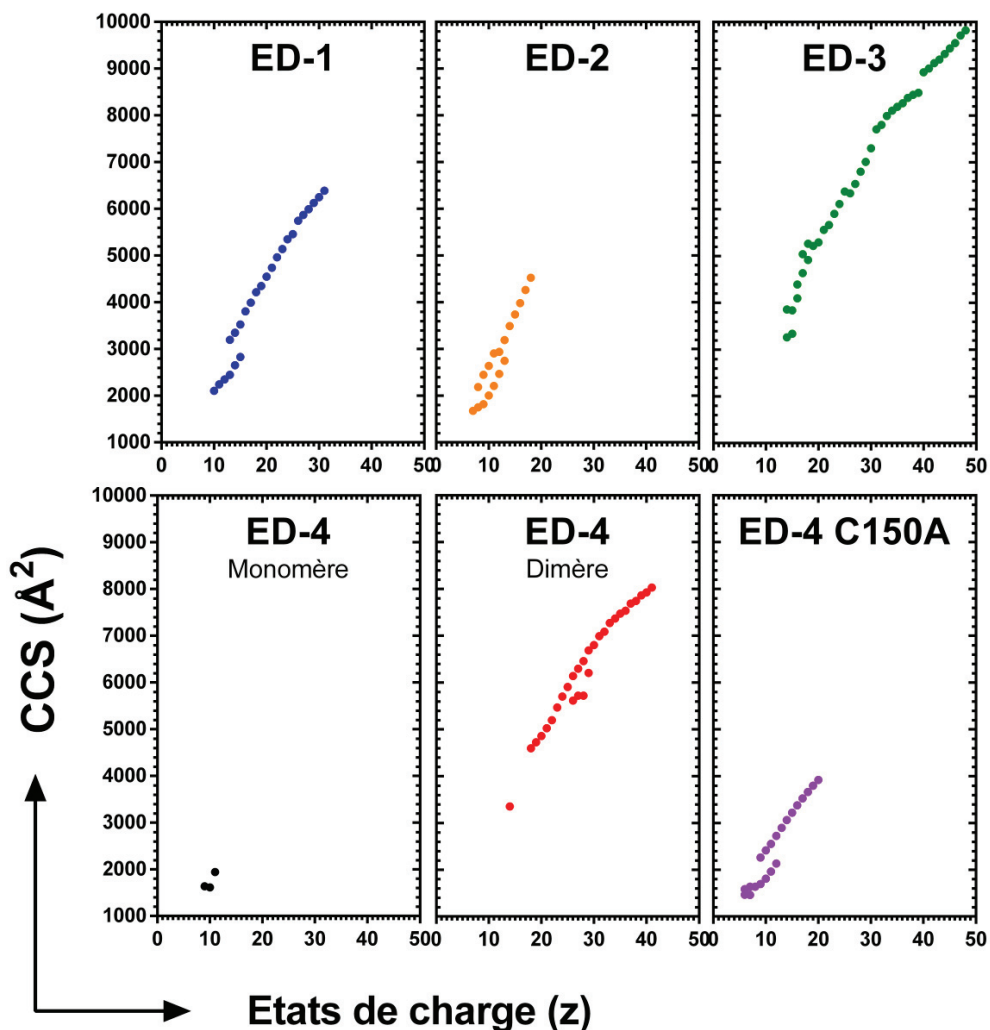
L'objectif des expériences suivantes a été de déterminer d'une part les limites de l'espace conformationnel des EDs des syndécans et d'autre part d'étudier la stabilité de leurs conformations les plus compactes afin de proposer un modèle général permettant de décrire la totalité des conformations des syndécans.

Des expériences de mobilité ionique (ESI-IM-MS) ont été réalisées en collaboration avec Fabien Chirot (Institut des Sciences Analytiques, Villeurbanne) et Luke MacAleese (Institut Lumière Matière, Villeurbanne) avec l'aide de Mathilde Bouakil (ILM) et d'Aurélien Le Fèvre (ISA).

Le spectromètre de masse utilisé est composé d'une source nano-ESI qui permet de transférer les protéines en phase gazeuse dans des conditions très douces, et d'un analyseur à temps de vol permettant de mesurer les rapports masse/charge de chaque ion. Un tube contenant de l'hélium est placé en aval de la source Nano-ESI. Après avoir été transférées en phase gazeuse, les protéines vont plus ou moins collisionner contre les atomes d'hélium, augmentant ainsi leur temps de dérive (ou *drift time*) vers le détecteur. Le temps de dérive dans le tube dépend d'une part de la conformation de la protéine et d'autre part de sa charge. Plus une protéine adoptera une conformation étendue, plus la probabilité qu'elle collisionne avec l'hélium sera importante ce qui allongera son temps de dérive. Une espèce plus chargée aura un temps de dérive plus court qu'une autre espèce moins chargée possédant la même conformation étant donné qu'elle sera plus sensible au champ électrique. Ces deux paramètres sont regroupés dans le calcul de la section efficace de collision (ou *collisional cross section*, CCS) qui reflète la géométrie globale d'une conformation.

Après le transfert en phase gazeuse, il est nécessaire d'induire un champ électrique afin de faire voyager les ions dans le tube de dérive. Une faible tension de 20 V (FA20, pour *force d'activation*) a été appliquée afin que les collisions avec les atomes d'hélium n'engendrent pas de modifications importantes de leurs conformations (**Figure 77**).

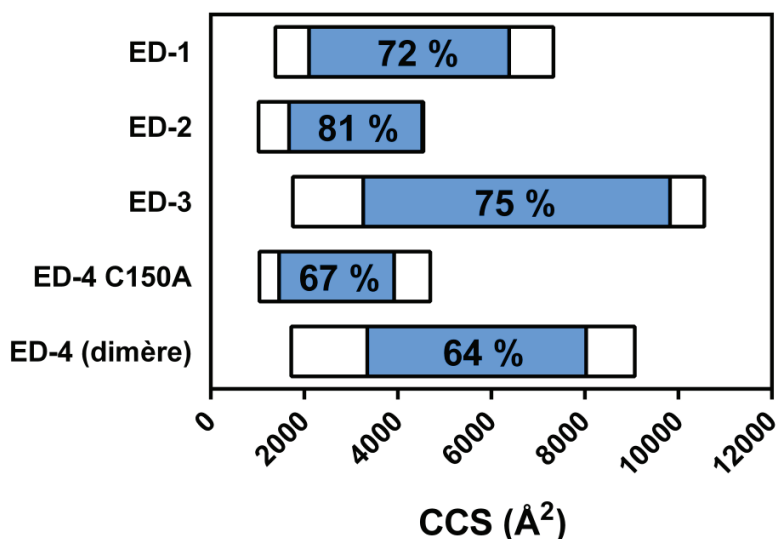
Une gamme de section efficace de collision (CCS) s'étalant de 2102 à 6384 Å<sup>2</sup> a été observée pour ED-1 (260 résidus d'acides aminés), de 1676 à 4522 Å<sup>2</sup> pour ED-2 (154 résidus) et de 3261 à 9823 Å<sup>2</sup> pour ED-3 (373 résidus), ce qui est cohérent avec leur nombre de résidus d'acides aminés. Les états de charges correspondant à la forme monomérique d'ED-4 ont été difficilement observables où seuls les ions portant de 9 à 11 charges ont été détectés. Afin d'obtenir des données propres au monomère, le résidu de cystéine impliqué dans sa dimérisation a été muté pour un résidu d'alanine (ED-4 C150A). ED-4 C150A (163 résidus d'acides aminés) possède une gamme de CCS allant de 1466 à 3920 Å<sup>2</sup> alors que le dimère (326 résidus) possède une gamme de CCS environ deux fois plus importante, comprise entre 3351 et 8028 Å<sup>2</sup>. La valeur des CCS augmente avec l'état de charge mais pas de façon linéaire suggérant la présence de plusieurs populations de conformations, les états de charge les plus faibles représentant les conformations les plus compactes et réciproquement (**Figure 77**).



**Figure 77 : Distribution des sections efficaces de collision (CCS) des ectodomaines des syndécans (EDs) et du mutant ED-4 C150A lors des expériences de mobilité ionique.** Les valeurs des CCS ont été mesurées avec une force d'activation de 20 V en faisant varier la tension du tube de dérive de 250 à 550 V. Les graphes sont à la même échelle afin de faciliter leur comparaison.

Le mutant ED-4 C150A a permis d'obtenir les données propres à la forme monomérique d'ED-4. Ces données seront utilisées pour les futures analyses. L'étendue de l'espace conformationnel des ectodomaines des syndécans a été évaluée en calculant les limites théoriques des CCS des EDs et en la comparant aux valeurs expérimentales (**Figure 78**). Les limites théoriques de la distribution des CCS ont été calculées en supposant que les conformations les plus compactes des ectodomaines des syndécans étaient globulaires et sphériques, et que les plus étendues étaient similaires à des conformation en bâton rigide (*Beveridge et al., 2014*).

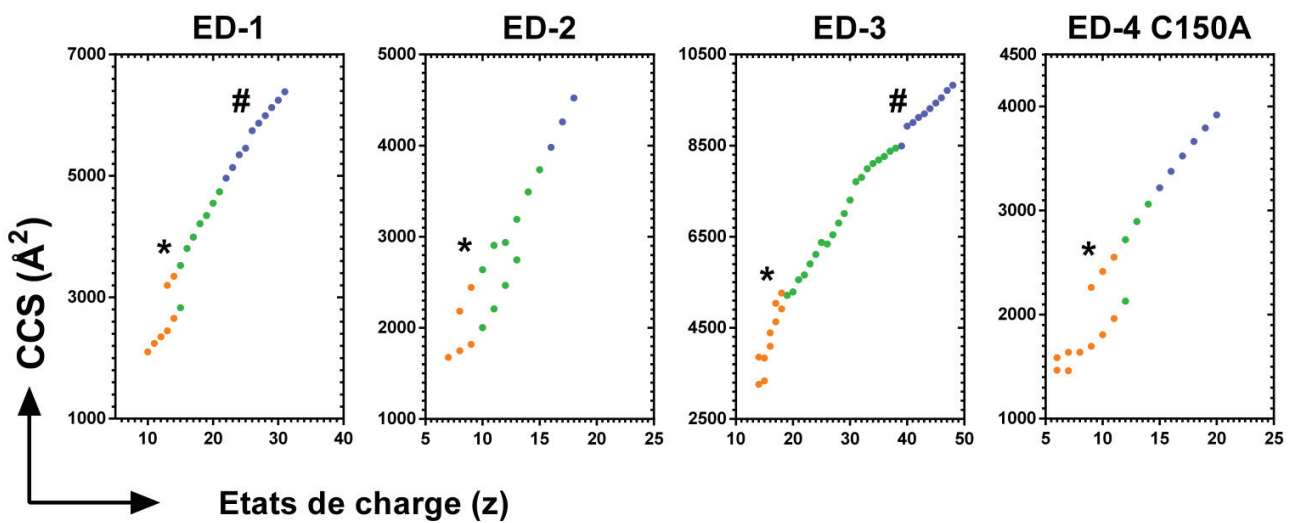
Les 4 ectodomaines des syndécans explorent plus de la moitié de leur espace conformationnel théorique (**Figure 78**). La couverture des CCS expérimentales par rapport aux limites théoriques est similaire pour ED-1 et ED-3 (72 et 75 % respectivement). La gamme de CCS observée expérimentalement pour ces deux protéines est plus ou moins centrée par rapport aux limites théoriques, confortant les précédentes observations suggérant que les EDs ont la capacité d'adopter des conformations très étendues (valeurs de CCS élevées) ou plus compactes (valeurs de CCS faibles). ED-2 possède la couverture la plus élevée (81 %) et atteint la limite théorique maximale (4547 Å<sup>2</sup>) avec une valeur de CCS expérimentale de 4522 Å<sup>2</sup>. La forme monomérique d'ED-4, étudiée avec le mutant ED-4 C150A, explore une gamme de conformations plus réduite qu'ED-2 (67 %). Elle est similaire à la forme dimérique d'ED-4 (64 %) ce qui suggère que le dimère d'ED-4 conserve une flexibilité importante, en accord avec les observations réalisées en SEC-SAXS.



**Figure 78 : Espace conformationnel des ectodomaines des syndécans d'après les expériences de mobilité ionique.** Boîtes blanches et bleues superposées : limites théoriques et expérimentales, respectivement, des sections efficaces de collision (CCS). Le pourcentage exploré par les ectodomaines des syndécans (EDs) est indiqué.

Afin de visualiser les différents degrés de compaction sur la distribution des CCS expérimentales, la même classification des états de charge réalisée lors des analyses d'ESI-MS native a été appliquée pour les ectodomaines des syndécans (**Figure 79**).

Dans cette représentation, une rupture de linéarité est observée dans la gamme de CCS correspondant au passage des conformations intermédiaires aux conformations étendues pour ED-1 (entre 25 et 26+) et qui est plus marquée pour ED-3 (entre 39 et 40+) mais pas pour ED-2 et ED-4 C150A où une linéarité dans l'évolution des valeurs de CCS est observée entre ces deux populations de conformations. Ces observations suggèrent que les conformations d'ED-2 et de la forme monomérique d'ED-4 gagnent de façon graduelle une part de plus en plus importante de désordre intrinsèque, alors que la rupture de linéarité observée pour ED-1 et ED-3 montre la formation d'une population étendue et distincte des conformations intermédiaires (**Figure 79**).

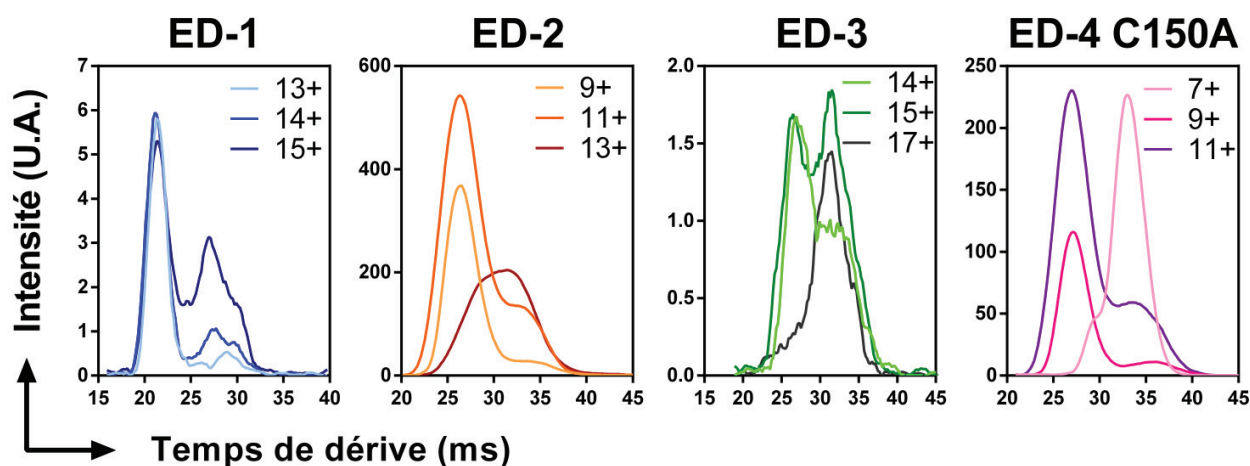


**Figure 79 : Distribution des sections efficaces de collision (CCS) des ectodomaines des syndécans (EDs) et du mutant ED-4 C150A.** Les états de charges ont été classifiés de la même façon que lors de l'analyse ESI-MS native où les conformations les plus compactes sont en orange, les conformations intermédiaires sont en vert et les plus étendues sont en bleu. La rupture dans la linéarité des valeurs de CCS entre les conformations compactes et intermédiaires est indiquée par une étoile (\*) et par un dièse (#) pour la rupture observée entre les conformations intermédiaires et étendues.

Une rupture de linéarité des valeurs de CCS est également observée pour les faibles états de charge entre les conformations compactes et les conformations intermédiaires pour ED-1 (13 – 15+), ED-2 (8 – 13+) et ED-4 C150A (6 – 12+) et dans une moindre mesure pour ED-3 (14 – 18+).

A l'inverse de ce qui a été observé lors du passage des conformations intermédiaires à étendues, deux populations coexistantes ont été détectées à ces états de charge menant au calcul de deux valeurs de CCS (**Figure 79**). La présence d'une large gamme d'états de charge pour ED-4 C150A (6 – 12+) possédant deux populations coexistantes est également en accord avec les mesures d'ESI-MS native où une gamme étendue propre aux conformations compactes a été observée (de 6 à 11+, voir *Figure 74*).

L'observation de deux populations dans la distribution des temps d'arrivée (ATD) des ions suggère que les populations compactes et les populations intermédiaires coexistent dans une certaine phase de transition entre ces deux types de populations, mais pas entre les populations intermédiaires et les plus étendues (**Figure 80**). L'ATD de ces états de charge montrent une inversion graduelle entre l'espèce possédant le temps de dérive le plus faible, donc la plus compacte, et celle arrivant plus tardivement ayant donc une conformation plus étendue, excepté pour ED-4 C150A (**Figure 80**). Cette inversion se fait donc aux dépens des conformations compactes et en faveur des conformations intermédiaires pour ED-1, ED-2 et ED-3. Concernant ED-4 C150A, le temps d'arrivée plus court des ions 9 et 11+ par rapport au 7+ est probablement causé par leur sensibilité plus importante au champ électrique. En revanche, les ions 9 et 11+ ont un temps d'arrivée similaire ( $\approx 28$  ms) indiquant que la conformation portant 11 charges est plus étendue que celle en portant 9 ce qui augmente son temps de dérive.



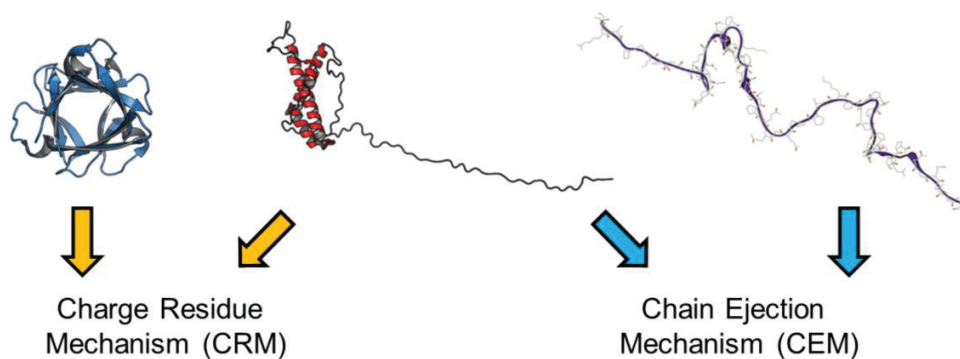
**Figure 80 : Distribution des temps d'arrivée (ATD) des états de charges où deux populations de conformations ont été détectées lors des expériences de mobilité ionique.** Pour des raisons de clarté, seuls trois états de charges par ectodomains des syndécans (EDs) sont présentés. Les distributions globales ont été obtenues à une tension dans le tube de dérive de 550 V pour ED-1 et ED-3 et de 450 V pour ED-2 et ED-4 C150A.

Ces premières expériences de mobilité ionique ont permis de conforter les précédentes observations réalisées en ESI-MS native sur la présence de plusieurs sous-populations adoptées par les EDs. La distribution des CCS d'ED-2 et de la forme monomérique d'ED-4 (*i.e.* ED-4 C150A) suggère que les conformations intermédiaires s'enrichissent progressivement en désordre intrinsèque comme en atteste l'évolution linéaire des valeurs de CCS avec l'augmentation de l'état de charge. Pour ED-2 et ED-4 C150A, les populations étendues et intermédiaires pourraient être ainsi regroupées sous unique population majeure. La distribution des CCS d'ED-1 et d'ED-3 manifeste une rupture de linéarité qui est plus marquée pour ED-3, ce qui suggère la formation d'une population très étendue conformationnellement distincte de la population intermédiaire.

Ces mesures ont également permis d'évaluer l'espace conformationnel des EDs des syndécans en se basant sur un modèle simple, dont les limites sont basées sur le calcul de la CCS d'une protéine en conformation compacte (sphérique) ou très étendue (bâton rigide).

Une phase de transition est observée entre les conformations les plus compactes et les conformations intermédiaires pour les quatre ectodomains des syndécans. Elle est caractérisée par la coexistence de deux populations dont les valeurs des CCS ont été calculées. La présence de deux populations pour un même état de charge peut également être la conséquence de deux mécanismes d'ionisation différents pour une même conformation (*Beveridge et al., 2015*). En effet, les protéines globulaires et les conformations compactes des protéines désordonnées suivent généralement le mécanisme CRM (*charge residue mechanism*) alors que les protéines étendues suivent le mécanisme CEM (*chain ejection mechanism*). Un ensemble de conformations ayant le même état de charge et possédant à la fois une région globulaire et une région désordonnée pourra suivre l'un de ses deux mécanismes, ce qui mène à la formation d'états de charge intermédiaires comprenant plusieurs conformations ayant le même état de charge (**Figure 81**).





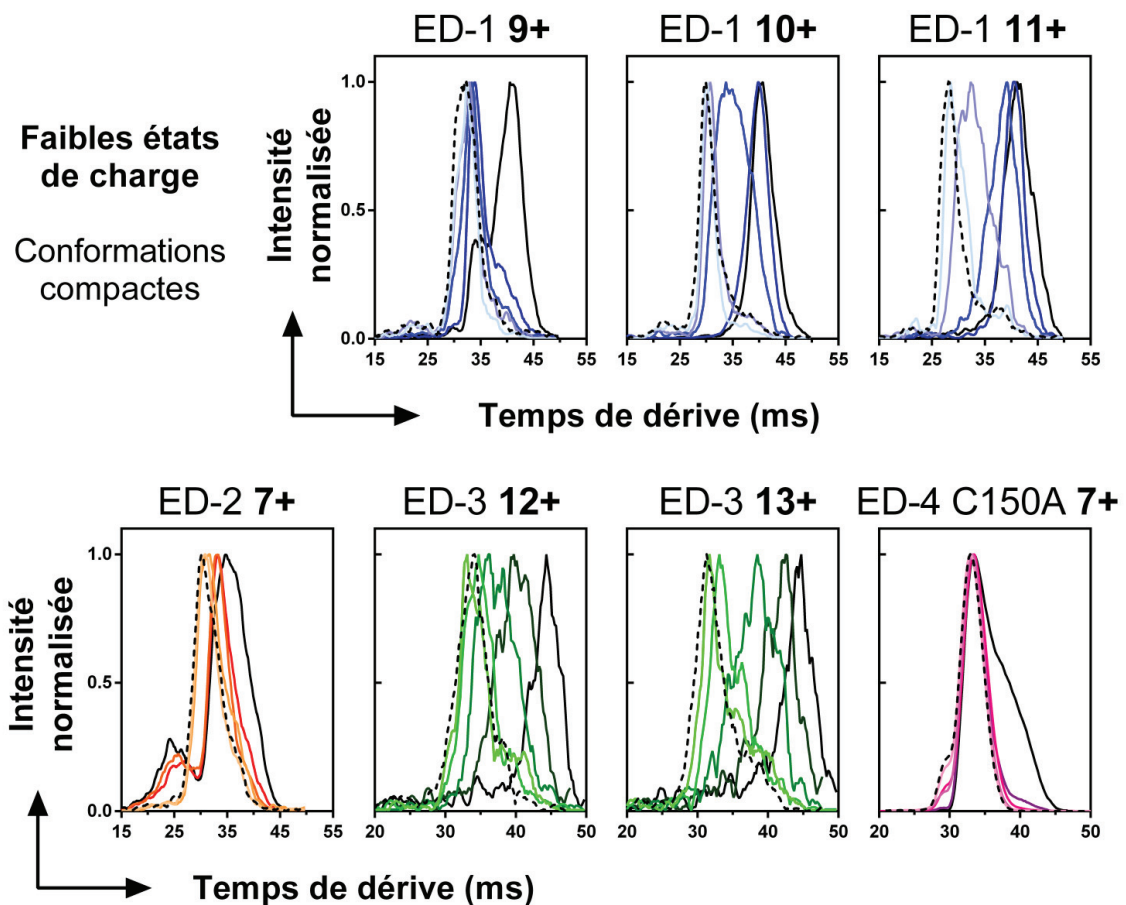
**Figure 81 : Mécanisme d'ionisation pendant l'étape d'électronébulisation des protéines selon leur degré de compaction.** Les protéines globulaires (ici FGF-2, PDB : 2FGF) suivent le mécanisme CRM et les protéines intrinsèquement désordonnées (ici IB5, d'après *Boze et al., 2010*) suivent davantage le mécanisme CEM. Les protéines ou les conformations ont un degré de compaction intermédiaire peuvent suivre l'un des deux mécanismes menant à l'apparition de deux conformations, et donc le calcul de deux valeurs de CCS pour un unique état de charge (la protéine Abpa, PDB : 2MXX, est donnée à titre d'illustration).

Le but des expériences suivantes a été de déterminer par des expériences de mobilité ionique la stabilité globale des conformations les plus compactes des quatre EDs des syndécans. Les états de charges les plus faibles des quatre EDs, reflétant leurs conformations les plus compactes, ont été activés, c'est-à-dire que des énergies croissantes ont été appliquées en modulant l'intensité du champ électrique (*i.e.* la force d'activation) à l'entrée du tube de dérive. Six valeurs de force d'activation ont été appliquées, de 20 V (peu ou pas d'activation) à 190 V (FA20 – 190).

La force des collisions entre les protéines et les atomes d'hélium augmente avec la force d'activation ce qui déstabilise les conformations et tend à les déplier. L'évolution du degré de compaction des différentes conformations pour chaque état de charge a été étudiée avec leur temps de dérive. Des états de charge propres à la phase de transition et aux conformations intermédiaires ont également été sélectionnés dans le même but (**Tableau XXII**).

Conformations	ED-1	ED-2	ED-3	ED-4 C150A
<b>Compactes</b> (Figure 82)	9+ – 11+	7+	12+ – 13+	7+
<b>Phase de transition</b> (Figure 83)	13+ – 15+	8+ – 10+	16+ – 19+	10+ – 13+
<b>Intermédiaires</b> (Figure 84)	16+ – 17+	11+	20+ – 33+	14 – 15+

**Tableau XXII : Etats de charges sélectionnés pour l'analyse de la stabilité conformationnelle des conformations les plus compactes, de celles retrouvées dans la phase de transition (entre les conformations compactes et intermédiaires) et des conformations intermédiaires des ectodomaines des syndécans.**



**Figure 82 : Activation des états de charges correspondant aux conformations les plus compactes des ectodomaines des syndécans.** Les faibles états de charge des EDs ont été sélectionnés et soumis à des forces d'activation de 20 (courbe noire en pointillés noirs), 50, 80, 120, 150 (nuances de couleur) et 190 V (courbe noire pleine). Les intensités ont été normalisées.

Le temps de dérive de l'ion 9+ d'ED-1 n'a pas été significativement affecté par l'activation de FA20 à FA150 mais est sensiblement augmenté à FA190 alors que celui des ions 10 et 11+ est augmenté dès l'application d'une force d'activation de 120 V et de 80 V, respectivement. D'une façon globale, plus l'état de charge est petit, plus l'énergie d'activation à appliquer pour observer des variations de temps de dérive doit être importante (**Figure 82**). L'activation des faibles états de charges 12+ et 13+ d'ED-3 mène également à une variation importante du temps de dérive dès l'application d'une FA de 120 V et 80 V, respectivement.

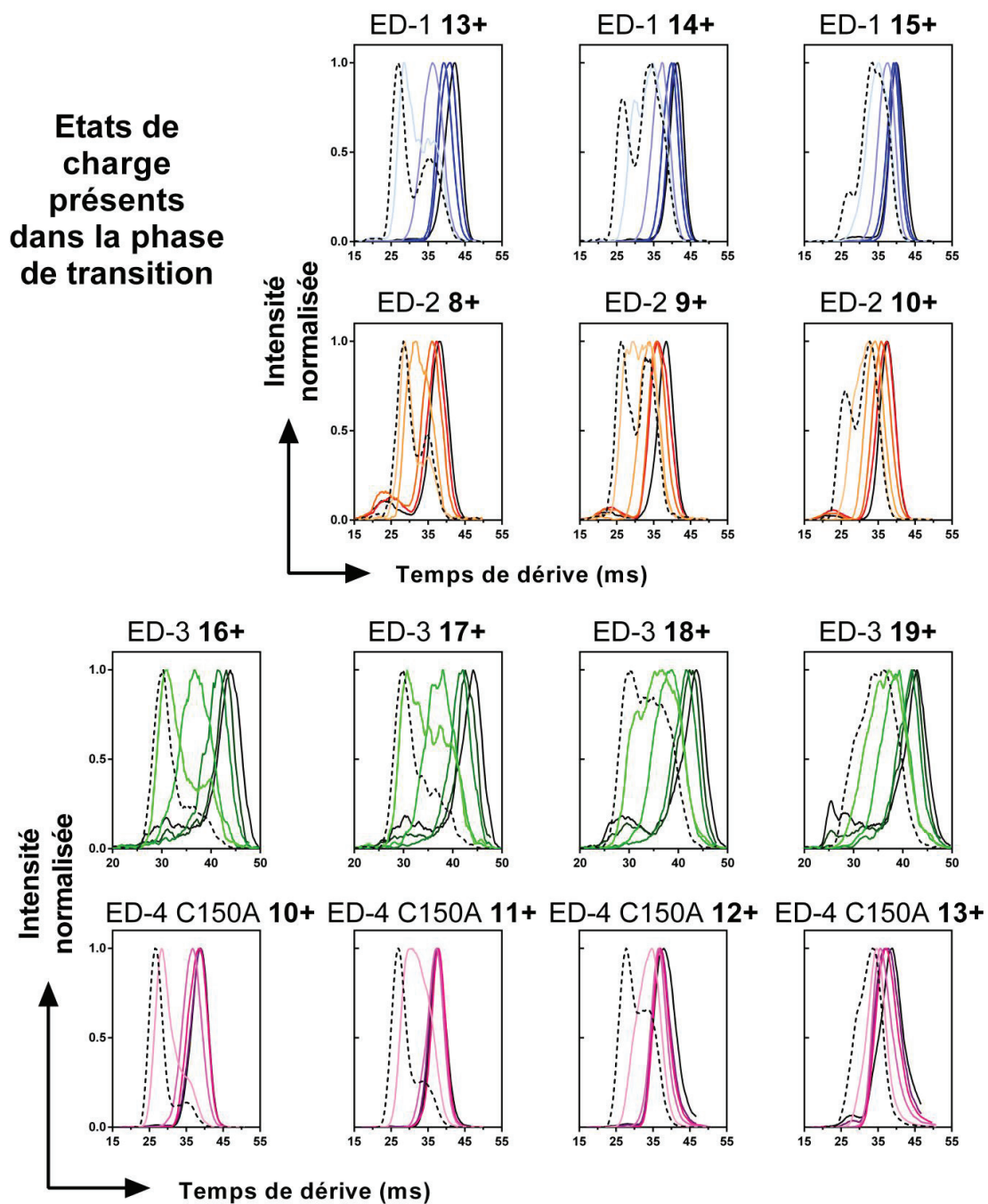
A l'inverse, le faible état de charge (7+) d'ED-2 et d'ED-4 C150A est difficilement activable. Le temps de dérive de l'ion 7+ d'ED-2 est légèrement augmenté lors de l'application des FA 50 V à 150 V et significativement accru à FA 190V. Une deuxième population possédant un temps de dérive plus faible (25 ms) est observée lors des FA 120 – 190 V, ce qui peut indiquer une fragmentation d'ED-2. Des observations similaires ont été réalisées pour l'ion 7+ d'ED-4 C150A où une variation de temps de dérive n'a été observée qu'à FA 190V (**Figure 82**). En accord avec les précédentes observations, une seconde population possédant un temps de dérive d'environ 28 ms et dont l'abondance décroît avec l'augmentation de la FA est également détectée. Ces observations suggèrent que la forme monomérique d'ED-4 possède deux conformations compactes distinctes, la première dont le temps de dérive est le plus faible (environ 25 ms) étant facilement activable, mais pas la deuxième dont le temps de dérive est d'environ 33 ms et qui n'est pas activable. La présence de deux types de conformations compactes pour ED-4 est en accord avec la large gamme de faibles états de charge détectés lors des expériences d'ESI-MS native (*voir Figure 74*).

Cette première série d'expérience montre que les conformations les plus compactes d'ED-2 et de la forme monomérique d'ED-4 sont très stables et qu'une énergie importante est donc nécessaire afin de les déplier. Cela suggère que, dans ces conditions, le passage des conformations compactes à intermédiaires n'est pas favorisé. A l'inverse, cette transition semble être favorisée pour ED-1 et ED-3 où les états de charges les plus faibles sont facilement activables et sont donc plus instables.

La même série d'expérience a été réalisée sur les états de charges correspondant à la phase de transition entre les conformations compactes et intermédiaires des ectodomains des syndécans (**Tableau XXII** et **Figure 83**). Pour les 4 EDs des syndécans, chaque état de charge possède deux populations de conformations dans les conditions les plus douces (FA20) où une inversion en termes d'abondance a été observée avec la quantité de charge caractérisant cette phase de transition. Comme décrit précédemment, ce phénomène peut également être une conséquence de l'existence de deux mécanismes d'ionisation. Une faible activation (FA50 - 80) a été suffisante pour activer les conformations appartenant à cette phase où une augmentation du temps de dérive a été observée. Un équilibre partiel en termes d'abondance entre les conformations compactes et les conformations intermédiaires a été observé pour ED-1 (14+), ED-2 (9+), ED-3 (18+) et ED-4 C150A (entre 12 et 13+) (**Figure 83**).

Pour ED-1, ED-2 et ED-3, cet étage de charge est exactement situé à l'interface entre les conformations compactes et intermédiaires définies lors des expériences d'ESI-MS native alors que pour la forme monomérique d'ED-4, cette interface a été située à l'état de charge 11+ (*voir Figure 74*).

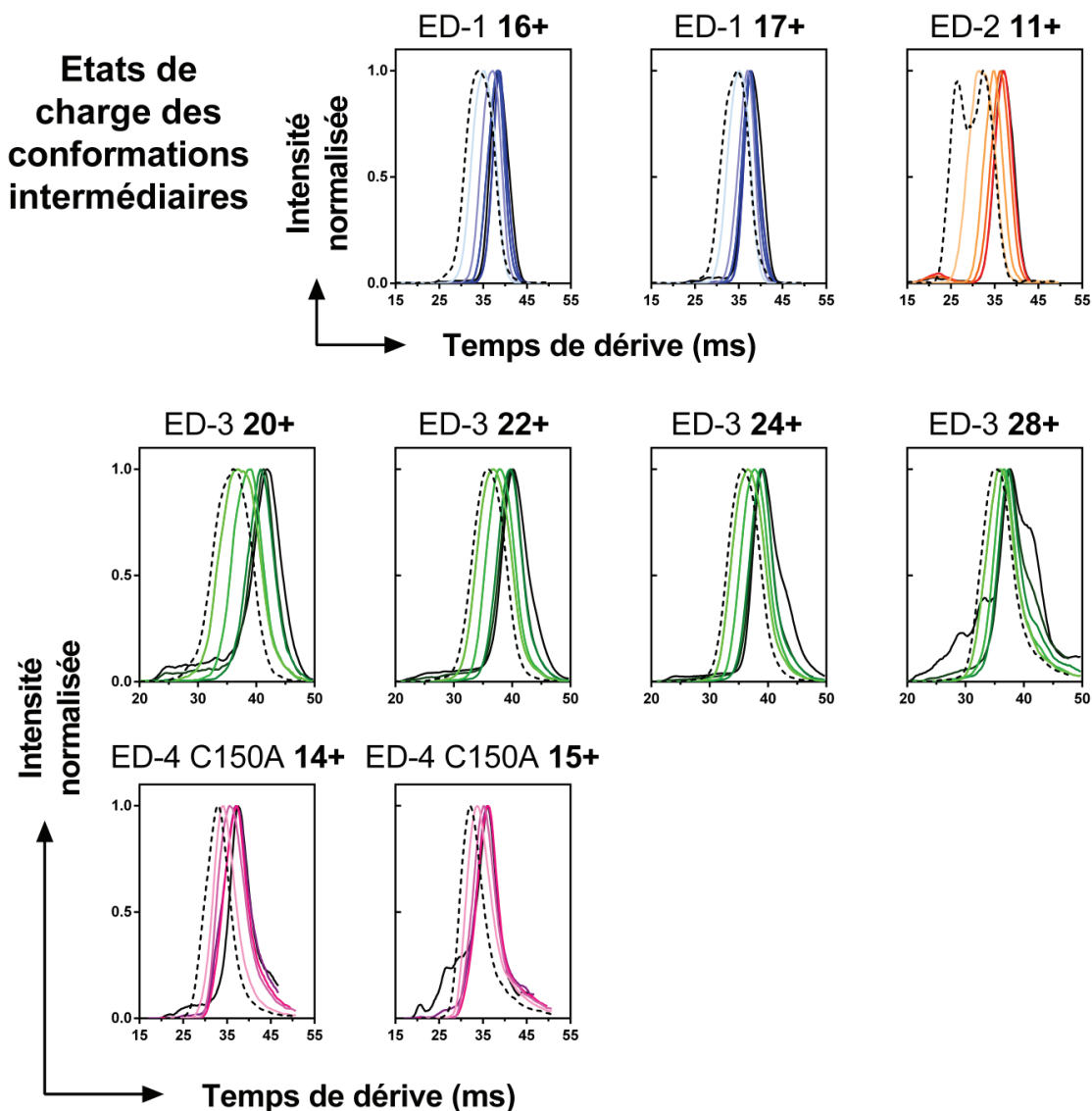
A partir d'une force d'activation de 120 V, le temps de dérive varie de moins en moins, et plus l'état de charge augmente, moins l'énergie d'activation nécessaire pour déstabiliser les conformations compactes est importante (**Figure 83**). Ces observations suggèrent que les conformations retrouvées dans cette phase de transition sont instables et que, passé cette phase, les conformations sont déjà très étendues.



**Figure 83 : Activations des états de charges correspondant aux conformations de la phase de transition entre les conformations compactes et les conformations intermédiaires des ectodomaines des syndécans.** Les états de charge présents dans la phase de transition des EDs ont été sélectionnés et soumis à des forces d'activation de 20 (pointillés noirs), 50, 80, 120, 150 (nuances de couleur) et 190 V (noir). Les intensités ont été normalisées.

Finalement, les mêmes expériences ont été menées sur des états de charges plus importants afin de caractériser les conformations intermédiaires entre les conformations les plus compactes et les plus étendues (**Figure 84**). Dans des conditions douces (FA20), une unique population de conformations intermédiaires est présente pour ED-1, ED-3 et ED-4 C150A alors qu'ED-2 possède toujours deux populations et en abondance équivalente, similaire à l'état de charge 9+. D'une façon globale, les caractéristiques observées sur les états de charge présents dans la phase de transition sont retrouvées. Pour ED-1, ED-3 et ED-4 C150A, le temps de dérive évolue dès l'application d'une force d'activation de 50 V suggérant des conformations instables. Cependant, les variations sont beaucoup plus faibles lorsque des FA plus importantes sont appliquées ce qui indique que les conformations sont déjà étendues et qu'il est difficile de les étendre davantage. L'application de forces d'activation élevées (FA150 – 190 V) mène à une fragmentation des protéines, ce qui se traduit par la formation d'espèces possédant un temps de dérive plus court, notamment pour ED-3 et ED-4 C150A (**Figure 84**).

Pour ED-2, les mêmes caractéristiques que celles observées sur la phase de transition sont retrouvées où les conformations sont facilement activables. Une FA de 50 V a été suffisante pour totalement perdre les conformations les plus compactes laissant une unique population de conformations (**Figure 84**).



**Figure 84 : Activations des états de charge correspondant aux conformations intermédiaires des ectodomaines des syndécans.** Les états de charge caractérisant les conformations intermédiaires des EDs ont été sélectionnés et soumis à des forces d'activation de 20 (pointillés noirs), 50, 80, 120, 150 (couleurs) et 190 V (noir). Les intensités ont été normalisées. L'ensemble des états de charge n'est pas représenté pour des raisons de clarté.

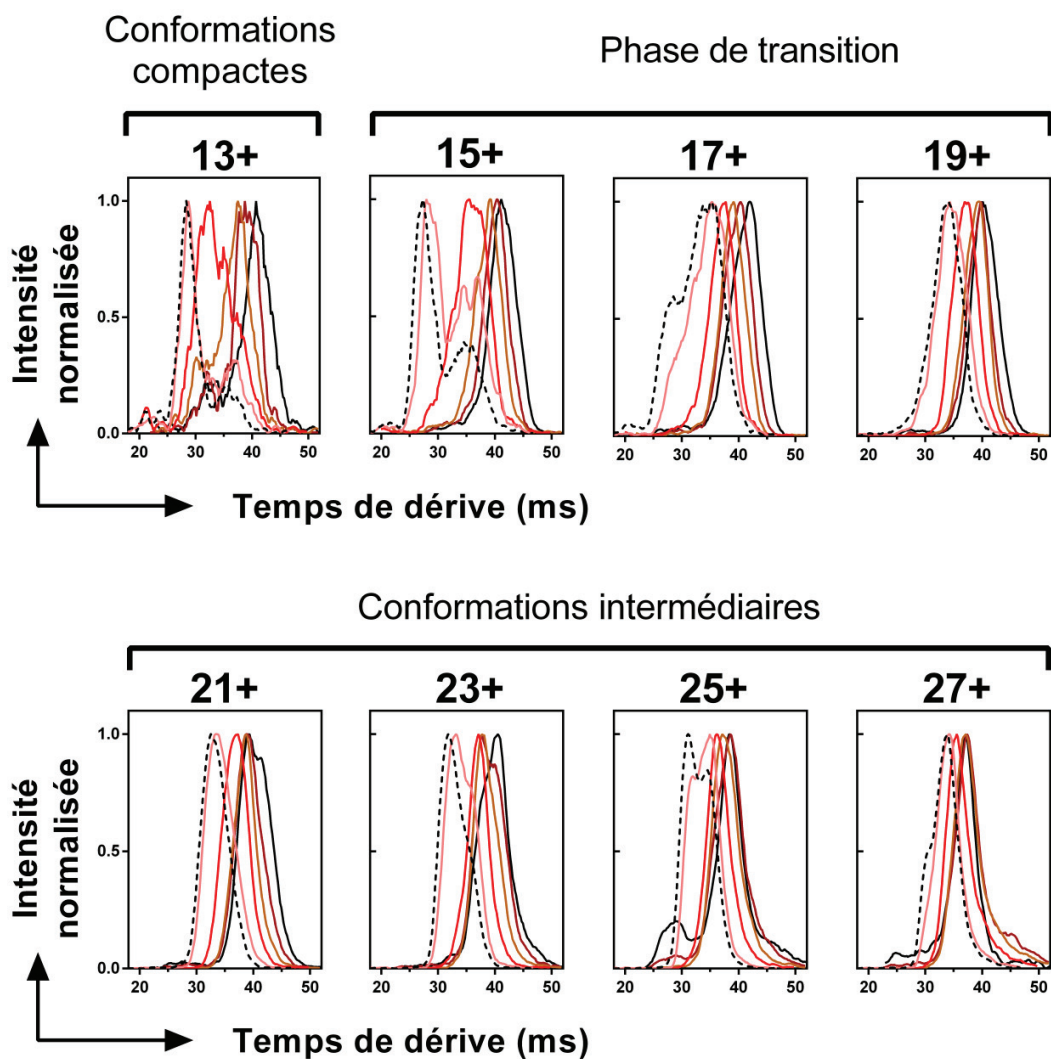
L'ensemble de ces expériences a également été réalisée sur la forme dimérique d'ED-4 (**Figure 85**). Afin de ne pas utiliser de données comprenant la participation de sa forme monomérique, seuls des états de charge impairs ont été sélectionnés. L'état de charge 13+ a été sélectionné afin d'étudier les conformations les plus compactes du dimère alors que les gammes d'états de charge 15 – 19+ et 21 – 27+ ont été sélectionnées afin d'étudier ses conformations présentes dans la phase de transition et ses conformations intermédiaires, respectivement.

Il a été délicat d'obtenir des signaux d'intensité suffisante pour les conformations les plus compactes du dimère d'ED-4 comme en témoignent les courbes bruitées de la distribution des temps d'arrivée de l'état de charge 13+ (**Figure 85**). En conditions douces (FA20), une espèce possédant un temps de dérive d'environ 28 ms est majoritaire avec une espèce de faible abondance (36 ms) et de conformation plus étendue. Contrairement aux formes monomériques des EDs des syndécans, cet état de charge est facilement activable puisqu'une augmentation du temps de dérive est observée dès l'application d'une force d'activation de 80 V. La faible abondance de cette population et son instabilité suggèrent que la formation de conformations compactes de dimère d'ED-4 n'est pas favorisée.

La phase de transition (15 – 19+) comprend également deux populations dont l'interversion s'effectue graduellement et où les deux populations seraient en abondance équivalentes à environ 16+. A la fin de cette phase, une seule population est détectée. L'activation de ces états de charge permet la transition des conformations les plus compactes en conformations intermédiaires puisqu'une seule espèce persiste dès une FA de 120 V pour l'état de charge 15+ et 80 V pour 17+ (**Figure 85**).

Dans des conditions douces (FA20), une nouvelle transition est observée dans la distribution des temps d'arrivée des états de charges intermédiaires (21 – 27+), s'effectuant entre les états de charge 23+ et 27+. Cette transition semble être déjà visible lors de l'activation à 190 V de l'état de charge 21+ mais l'absence du calcul de CCS ne permet pas de vérifier cette observation. Les activations appliquées sur les états de charges 23 – 27+ montrent la disparition progressive de l'espèce possédant le temps de dérive le plus faible aux dépens de celle possédant un temps de dérive plus important, donc une conformation globale plus étendue (**Figure 85**). Cette transition serait vraisemblablement celle s'effectuant entre les conformations intermédiaires et les conformations les plus étendues.



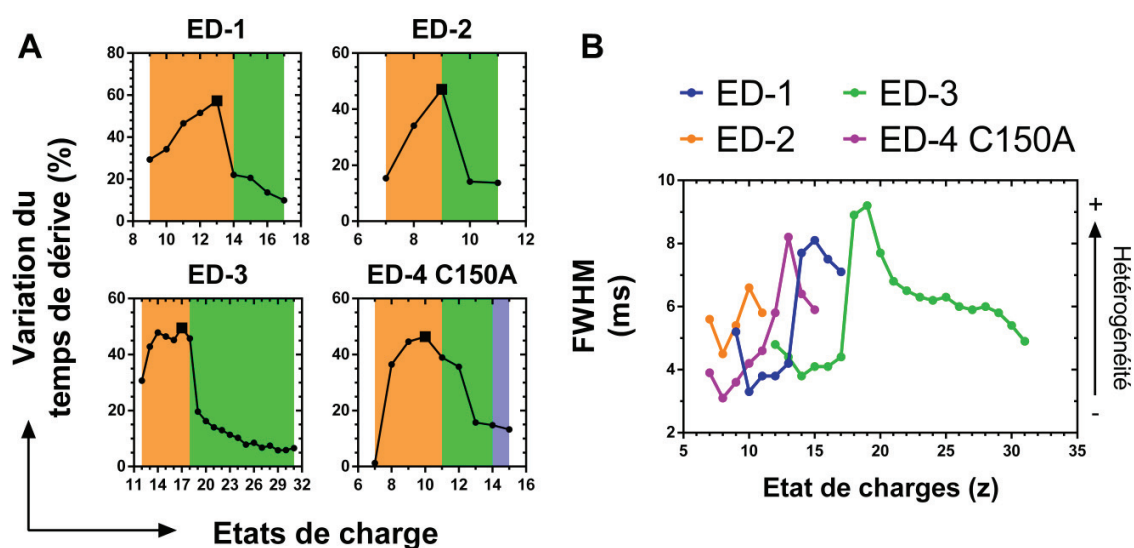


**Figure 85 : Activations des états de charges correspondant aux conformations compactes, aux conformations présentes dans la phase de transitions et aux conformations intermédiaires adoptées par la forme dimérique de l'ectodomaine du syndécan-4.** Les états de charge caractérisant les conformations compactes et intermédiaires ainsi que ceux présents dans la phase de transition ont été sélectionnés et soumis à des forces d'activation de 20 (pointillés noirs), 50, 80, 120, 150 (couleurs) et 190 V (noir). Les intensités ont été normalisées.

A des fins de comparaison et pour les analyses suivantes, la stabilité globale des différentes conformations a été définie comme étant la variation du temps de dérive (en %) mesurée sur les distributions globales lorsque l'activation a été la plus importante (190 V) par rapport à celle appliquée à 20 V (variations FA190 vs FA20). Une importante variation est donc corrélée à une faible stabilité conformationnelle et réciproquement. Elle est représentée en fonction des états de charge ce qui permet de les comparer au degré de compaction des différentes conformations définies précédemment (**Figure 86A**). La représentation des valeurs expérimentales du temps de dérive en fonction de la force d'activation, et ce pour chaque état de charge et pour chaque EDs, est présentée en **Annexe 2**, ainsi qu'une représentation 3D de l'état de charge en fonction de la section efficace de collision qui résume à la fois l'étendue de l'espace conformationnel des EDs des syndécans, les différentes populations observées ainsi que de l'hétérogénéité de chaque état de charge (**Annexe 3**).

Les courbes de stabilité des 4 EDs ont une forme similaire en cloche (**Figure 86A**). D'une façon générale, ces observations suggèrent que les conformations les plus compactes (*i.e.* les faibles états de charge) sont les plus stables étant donné leur faible variation de temps de dérive. Cette stabilité est accrue pour ED-2 et ED-4 C150A qui restent peu (15 %) ou insensibles à l'activation (1,2 %), respectivement. A l'inverse, les conformations compactes d'ED-1 et d'ED-3 sont plus facilement déstabilisées ( $\approx 30$  % de variation). La variation du temps de dérive augmente graduellement jusqu'à atteindre un maximum à l'état de charge 13+, 9+, 17+ et 10+ pour ED-1 – 4 (C150A). Ces états de charge sont les plus facilement activables suggérant que ces conformations sont très instables. Au-delà de ce maximum, la variation du temps de dérive diminue (et donc la stabilité croît) ce qui peut être expliqué par le fait que les conformations sont déjà étendues et que leur activation mène alors qu'à de faibles variations du temps de dérive.

Les états de charges correspondant aux conformations les moins stables coïncident à  $\pm 1$  charge avec la frontière estimée en ESI-MS native entre les conformations les plus compactes et les conformations intermédiaires (**Figure 86A**, voir *Figure 74*). Cette observation suggère que la transition entre ces deux conformations est favorisée en faveur des conformations les plus étendues.

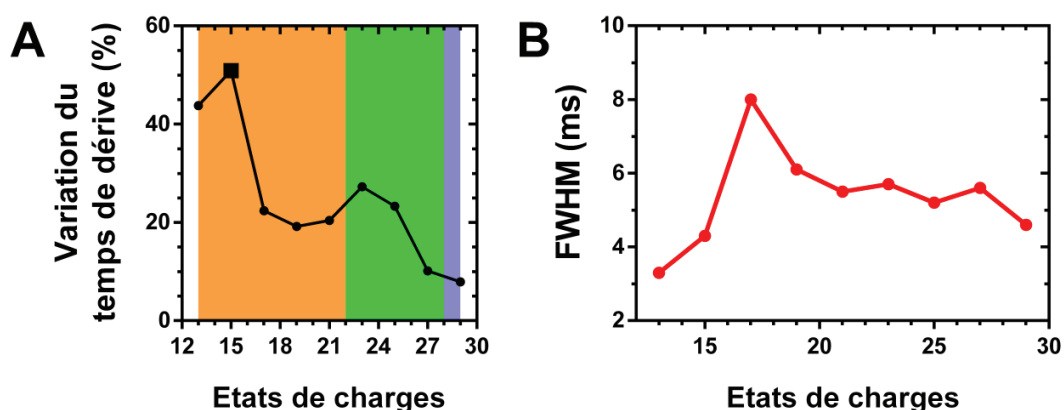


**Figure 86 : Bilan des mesures d'activation des états de charges des différentes conformations adoptées par les ectodomains des syndécans. A)** Variation (%) du temps de dérive lorsque l'activation a été appliquée à 190 V par rapport à celle appliquée à 20 V pour chaque état de charge. Lorsque deux populations sont présentes pour un même état de charge, les données concernant la population majoritaire ont été sélectionnées. Le point représenté par un carré (■) correspond à l'état de charge dont les conformations sont les plus instables. Les aires colorées en orange, vert et bleu correspondent aux conformations compactes, intermédiaires et étendues, respectivement. **B)** Evolution de la largeur à mi-hauteur (FWHM, *full width at half maximum*) des populations observées sur les distributions globales en fonction de l'état de charge.

La largeur à mi-hauteur (FWHM) de chaque état de charge observé sur les distributions globales a également été mesurée et représentée en fonction de l'état de charge (**Figure 86B**). La mesure de la FWHM permet d'évaluer l'hétérogénéité de chaque ensemble de conformation. Plus cette valeur est élevée (et donc le pic large), plus la population contient un nombre important conformations différentes et réciproquement.

Les courbes sont assez similaires à celles observées précédemment et présentent une région où la FWHM est assez faible pour les faibles états de charges, avant de rapidement croître puis de diminuer à nouveau. Cela suggère que les conformations compactes et intermédiaires possèdent quantité de conformations restreinte. A l'inverse, les conformations appartenant à la phase de transition possèdent une valeur de FWHM élevée, indiquant une polydispersité accrue et reflétant la coexistence des plusieurs populations de conformations en accord avec les précédentes observations.

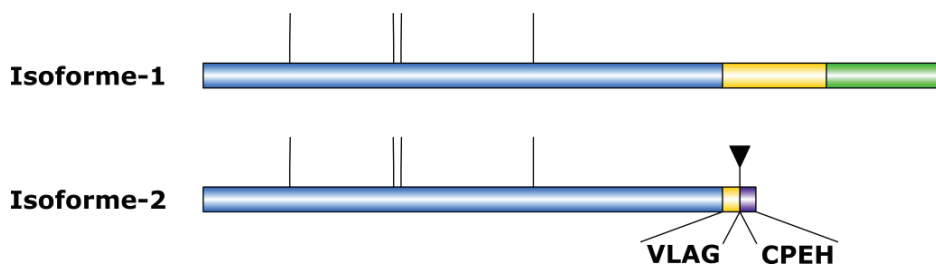
Les données obtenues sur la forme dimérique d'ED-4 ont été représentées de la même façon (**Figure 87**). La courbe représentant la variation du temps de dérive de chaque conformation montre deux maxima à 15+ et 23+ alors qu'un seul n'avait été observé pour la forme monomérique d'ED-4 (**Figure 87A**). Les variations du temps de dérive calculées sur les faibles états de charge (13 et 15+) sont très importantes (45 – 50 %) suggérant que les conformations compactes du dimère d'ED-4 ne sont pas stables et donc que leur formation n'est pas favorisée. Cette observation est confortée par le fait que l'état de charge le moins stable ne se situe pas à la frontière entre les conformations compactes et intermédiaires mais bien avant. On retrouve cependant la présence d'un deuxième maximum (23+) précédant une zone où les temps de dérive varient de moins en moins (de 26+ à 29+) suggérant que les conformations étendues du dimère sont très favorisées. Ces conclusions sont confortées par le calcul de la FWHM sur les distributions globales des conformations (**Figure 87B**). L'état de charge reflétant les conformations compactes (13+) possède une FWHM d'environ 3 ms indiquant un nombre restreint de conformations que peut adopter le dimère, à l'inverse des états de charges plus élevées où la FWHM est plus importante (5 – 8 ms) reflétant une plus grande hétérogénéité conformationnelle.



**Figure 87 : Bilan des mesures d'activation des états des charges correspondant aux conformations adoptées par la forme dimérique d'ED-4. A)** Variation (%) du temps de dérive lorsque l'activation a été appliquée à 190 V par rapport à celle appliquée à 20 V pour chaque état de charge. Lorsque deux populations étaient présentes pour un même état de charge, les données concernant la population majoritaire ont été sélectionnées. Le point représenté par un carré (■) correspond à l'état de charge dont les conformations sont les plus instables. Les aires colorées en orange, vert et bleu correspondent aux conformations compactes, intermédiaires et étendues, respectivement, basé sur la distribution des états de charge de la forme monomérique. **B)** Evolution de la largeur à mi-hauteur (FWHM) des populations observées sur les distributions globales en fonction de l'état de charge.

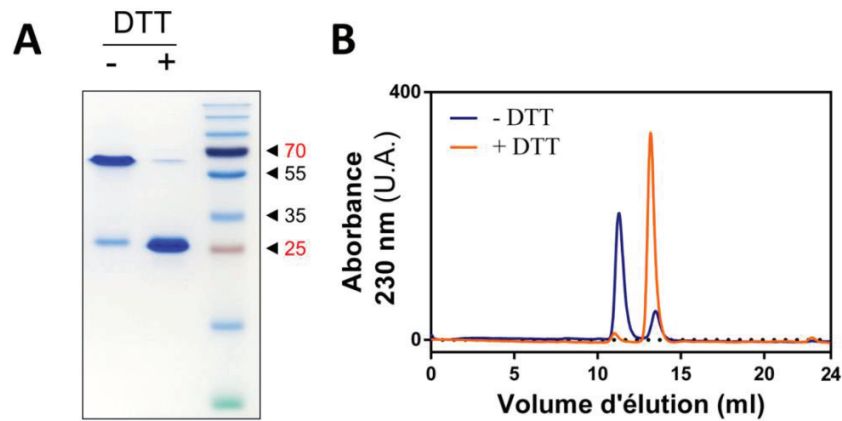
e) Mécanisme de dimérisation d'ED-4

Deux variants d'épissage du syndécan-4 ont été décrits (Xing et al., 2003). La première isoforme partage les mêmes caractéristiques que les autres syndécans et est composée d'un domaine cytosolique, transmembranaire et d'un ectodomaine qui peut être libéré par le processus de *shedding*. La seconde isoforme est en revanche directement exprimée sous forme d'un ectodomaine soluble, possédant la même séquence en acides aminés que l'ectodomaine de l'isoforme-1 auxquels s'ajoutent quatre résidus hydrophobes (VLAG) qui sont également retrouvés dans le domaine transmembranaire de l'isoforme-1, et quatre résidus supplémentaires (CPEH) propres à la seconde isoforme (Figure 88).



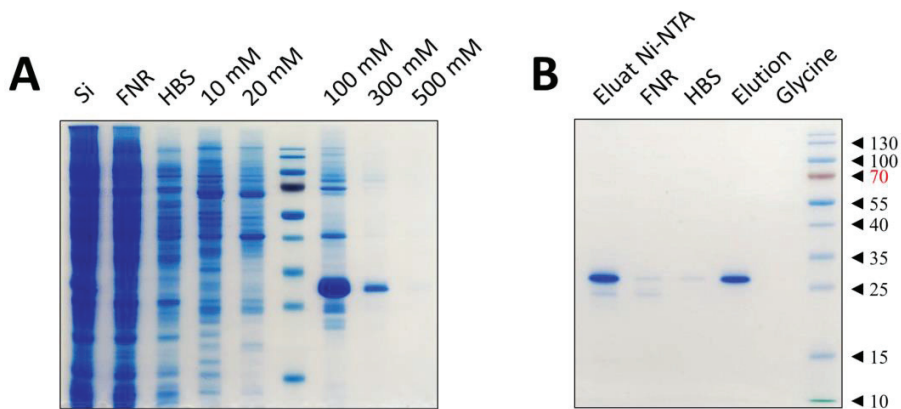
**Figure 88 : Comparaison de l'organisation des deux isoformes du syndécan-4.** L'isoforme-1 possède un domaine cytosolique (vert), transmembranaire (jaune) et un ectodomaine (bleu), où les sites d'attachement des glycosaminoglycanes sont indiqués (lignes verticales). L'isoforme-2 (ED-4) conserve la séquence de l'isoforme-1, de quatre de ses résidus transmembranaires (VLAG) auxquels s'ajoutent quatre résidus qui lui sont propres (CPEH, en violet). Le résidu de cystéine présent uniquement dans l'isoforme-2 est indiqué par un triangle noir.

La présence d'un résidu de cystéine suggère que la dimérisation pourrait être médiée par un pont disulfure. L'analyse d'ED-4 en condition réductrice par SDS-PAGE ainsi que par chromatographie d'exclusion stérique a provoqué une diminution très importante de la quantité de dimère, montrant le rôle majeur de cet unique résidu de cystéine dans le mécanisme de dimérisation (Figure 89).



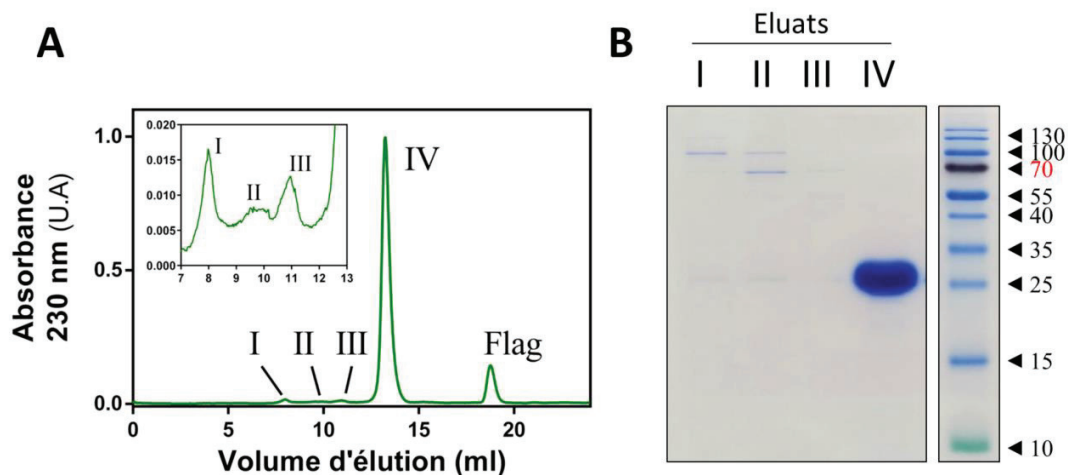
**Figure 89 : La formation d'un pont disulfure permet la dimérisation d'ED-4.** **A)** 2  $\mu\text{g}$  d'ED-4 ont été incubés pendant 20 minutes à température ambiante avec ou sans dithiothréitol ( $\pm$  DTT) puis analysés par SDS-PAGE (gel ayant une concentration en acrylamide de 15 %). Les protéines ont été colorées par du bleu de Coomassie. **B)** 100  $\mu\text{l}$  d'ED-4 à 1 mg/ml  $\pm$  DTT ont été injectés sur colonne de chromatographie d'exclusion stérique (Superdex S200 Increase 10/300 GL, GE Healthcare) équilibrée en HBS (- DTT) ou en HBS + DTT 2 mM (+ DTT).

Afin de vérifier cette observation, le résidu de cystéine 150 a été muté pour un résidu d'alanine (ED-4 C150A). ED-4 C150A a été exprimé en système procaryote et purifié par chromatographie d'affinité Ni-NTA puis anti-FLAG (**Figure 90**).



**Figure 90 : Purification du mutant ED-4 C150A par chromatographies d'affinité.** Analyses par SDS-PAGE des purifications par chromatographies d'affinité sur résine Ni-NTA (**A**), et résine anti-FLAG (**B**) du mutant ED-4 C150A. 15  $\mu\text{l}$  de solution initiale (Si), de fraction non retenue (FNR), de lavages par de l'HBS et par de l'imidazole 10 et 20 mM (HBS, 10 mM, 20 mM), de fractions éluées par de l'imidazole 100, 300 et 500 mM (100, 300, 500 mM) et de lavage glycine (Glycine) ont été déposés sur gel ayant une concentration en acrylamide de 15 %. Les protéines ont été colorées par du bleu de Coomassie.

Les fractions éluées de la purification anti-FLAG ont été concentrées et injectées sur une colonne de chromatographie d'exclusion stérique, de façon similaire aux autres ectodomains des syndécans (**Figure 91**). L'espèce majoritaire a été éluée à 13,2 ml et son analyse par SDS-PAGE a indiqué une masse moléculaire apparente d'environ 28 kDa, identique à la masse moléculaire de la forme monomérique d'ED-4 non muté.

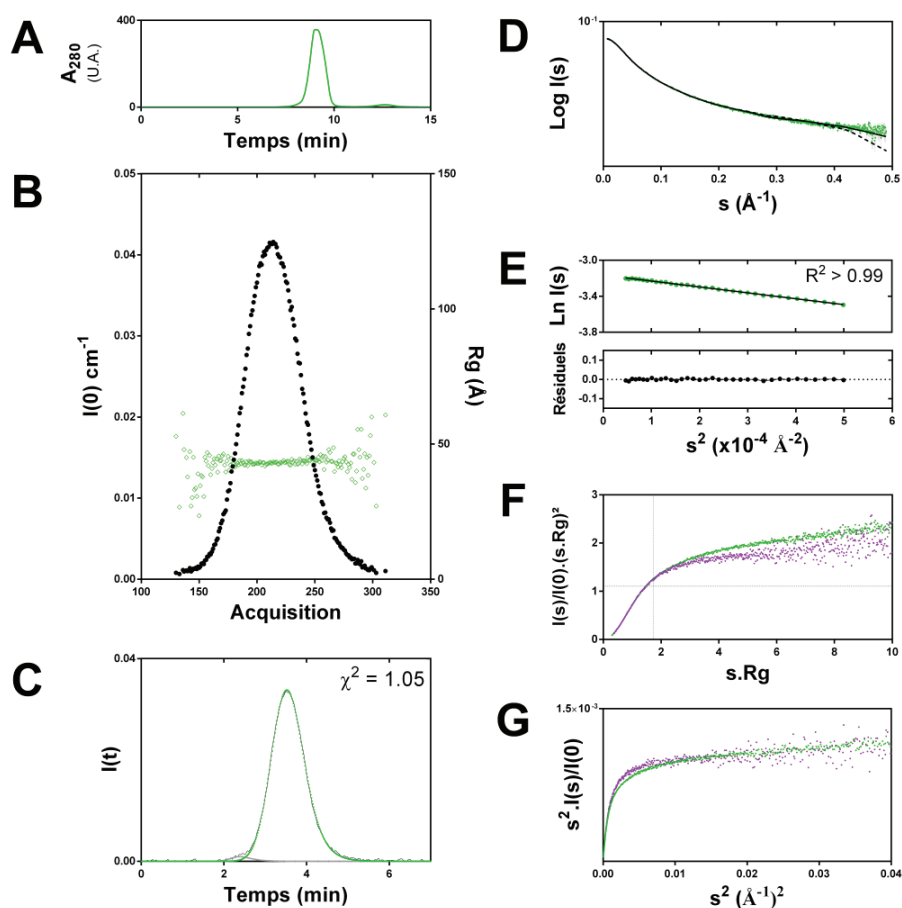


**Figure 91 : Purification du mutant ED-4 C150A par chromatographie d'exclusion stérique.**

**A)** Chromatogramme de la purification par chromatographie d'exclusion stérique (Superdex S200 Increase 10/300 GL, GE Healthcare). Les différentes espèces éluées sont annotées de I à IV. Flag: élution du peptide FLAG. **B)** 10  $\mu$ l des fractions éluées de la purification par chromatographie d'exclusion stérique (pics I à IV) ont été analysés par SDS-PAGE (gel de polyacrylamide 15 %). Les protéines ont été colorées par le bleu de Coomassie.

Des analyses biophysiques ont également été réalisées sur ED-4 C150A dans le but de comparer les formes monomériques sauvage et mutée d'ED-4. Des expériences de SEC-SAXS ont permis de déterminer le rayon de gyration du mutant ( $44,5 \pm 0,26 \text{ \AA}$ ), valeur similaire à celle de la forme monomérique d'ED-4 sauvage ( $41,53 \text{ \AA} \pm 0,67 \text{ \AA}$ ) (**Figure 92**). La représentation normalisée de Kratky et de Kratky-Debye montrent que le mutant ED-4 C150A est une protéine désordonnée et flexible. Bien que les intensités de diffusion soient très similaires à celles de la forme monomérique d'ED-4, elles semblent cependant légèrement diverger aux grands angles, ne décroissant pas strictement en  $s^{-2}$  mais plutôt entre  $s^{-2}$  et  $s^{-1}$ , ce qui peut suggérer la présence de régions très étendues (**Figures 92F et 92G**).

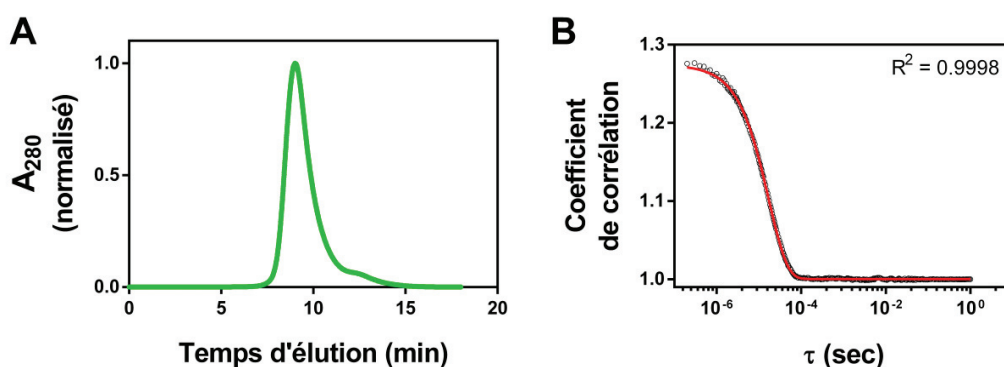
La déconvolution de ces données par EOM a permis d'attribuer une conformation très étendue de type *random coil* à ED-4 C150A plutôt qu'une conformation en pré-globule fondu (**Figure 92D**). Cependant, le test *CorMap* a permis d'identifier une déviation statistiquement significative aux grands angles entre les intensités de diffusion expérimentales et la courbe théorique générée par EOM ( $p\text{-value} = 6,2 \cdot 10^{-5} <$  au seuil de 0,01). Les résultats de la déconvolution n'ont donc pas été analysés davantage.



**Figure 92 : Analyse par SEC-SAXS du mutant ED-4 C150A.** **A)** Chromatogramme du mutant ED-4 C150A injecté sur une colonne Superdex S200 Increase 5/150 GL (3ml). **B)** Intensités de diffusion (noir) et estimation du rayon de gyration en fonction des acquisitions (vert). **C)** Déconvolution des intensités de diffusion en fonction du temps par *US-SOMO*. Vert : ED-4 C150A, gris : espèce de taille plus importante. Les données issues d'un angle unique sont représentées pour des raisons de clarté. La qualité de l'ajustement est indiquée ( $\chi^2$ ). **D)** Courbe de diffusion finale, avec les ajustements d'EOM pour un modèle *random coil* (courbe pleine) et pré-globule fondu (pointillé). **E)** Détermination du rayon de gyration selon l'approximation de Guinier. La qualité de l'ajustement est indiquée ( $\chi^2$ ). **F-G)** Représentations normalisées de Kratky et de Kratky-Debye, respectivement. ED-4 C150A : vert, monomère d'ED-4 sauvage : violet.



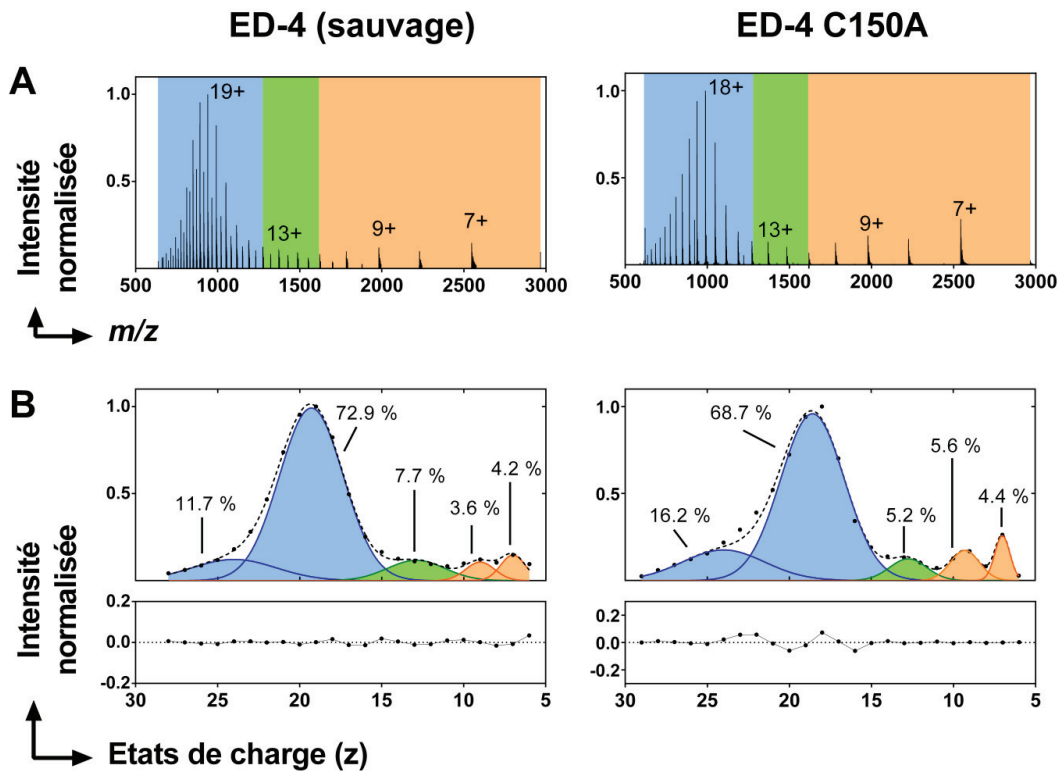
Le rayon hydrodynamique d'ED-4 C150A a été estimé à  $35,7 \pm 0,8 \text{ \AA}$  par SEC-DLS (**Figure 93**). Cette valeur est supérieure à celle mesurée par SEC-DLS sur la forme monomérique d'ED-4 ( $21,6 \pm 4,0 \text{ \AA}$ ), bien que les signaux de cette dernière mesure soient bruités. La valeur du rayon hydrodynamique d'ED-4 C150A déterminée par chromatographie d'exclusion stérique est de  $35,8 \text{ \AA}$ , quasiment identique à celle mesurée par SEC-DLS. Ces observations suggèreraient que la forme monomérique d'ED-4 possède un rayon hydrodynamique d'environ  $36 \text{ \AA}$  au lieu d'environ  $22 \text{ \AA}$  menant à un rapport  $R_g/R_h$  de 1,16 au lieu de 1,92, ce qui reste cohérent avec une conformation étendue.



**Figure 93 : Mesure du rayon hydrodynamique du mutant ED-4 C150A par SEC-DLS. A)** Chromatogramme du mutant ED-4 C150A sur une colonne Superdex S200 5/150 GL (3ml). **B)** Corrélogramme obtenus lors des expériences de SEC-DLS. La qualité de l'ajustement (courbe rouge) est indiquée par le coefficient de détermination ( $R^2$ ).

Des expériences d'ESI-MS en conditions natives ont également été effectuées sur le mutant afin de comparer la distribution des états de charge avec celle de la forme sauvage (**Figure 94**). La distribution des états de charge d'ED-4 C150A est très similaire à celle de la forme sauvage, indiquant que les deux protéines explorent le même espace conformationnel, composé de conformations étendues, intermédiaires et compactes (**Figure 94A**). Des pics de  $m/z$  intermédiaires sont visibles sur la distribution des états de charge d'ED-4 sauvage, ce qui reflète la présence du dimère. Ces pics sont de faible intensité étant donné que l'analyse a été effectuée sur un échantillon dont la majeure partie d'ED-4 était sous forme monomérique. En revanche, lorsqu'ED-4 C150A est analysé, ces pics sont absents ce qui est en accord avec le mécanisme de dimérisation médié par un pont disulfure (**Figure 94A**). La déconvolution des CSDs montre également que, de façon globale, les proportions des différentes populations de conformations sont similaires dans les formes sauvage et mutée (**Figure 94B**).

Les aires calculées pour chacune des conformations de la forme sauvage et du mutant sont du même ordre de grandeur, les conformations les plus étendues étant les plus abondantes (75 et 85 %, respectivement) et les conformations compactes les moins représentées (8 et 10 %).

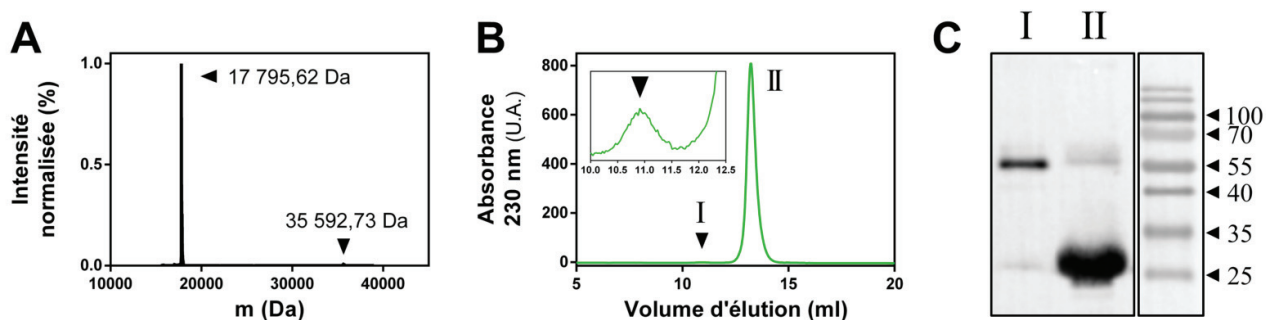


**Figure 94 : Comparaison de la distribution des états de charge d'ED-4 et du mutant C150A.**

**A)** Distribution des états de charge (CSD). Les différentes populations sont colorées en bleu (étendues), vert (intermédiaires) et orange (compactes). Les états de charge sur lesquels sont centrées ces populations sont indiqués. **B)** Déconvolution des CSDs. Les proportions (%) de chaque population de conformations sont indiquées ainsi que les valeurs résiduelles. La légende est la même qu'en A).

L'analyse ESI-MS du mutant C150A a cependant révélé qu'une seconde espèce moléculaire en quantité très minoritaire ( $\approx 1$  %) était présente (**Figure 95A**). Cette espèce possède une masse d'environ 35 593 Da, soit deux fois celle d'ED-4 C150A, mesurée également dans cette expérience (environ 17 796 Da, masse très proche de la masse théorique calculée à partir de la séquence en acides aminés, 17 790 Da), ce qui suggérerait que le mutant d'ED-4 conserve une faible capacité à dimériser.

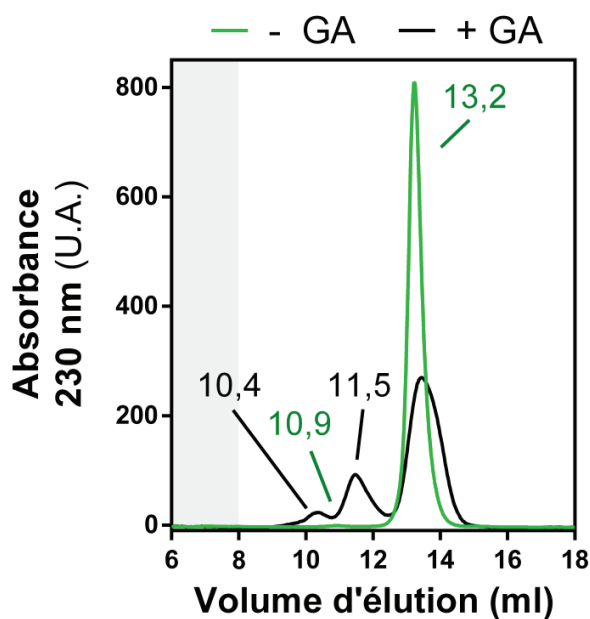
Afin de vérifier l'identité de cette espèce de plus haute masse moléculaire, les fractions d'éluion en sortie de colonne de chromatographie d'exclusion stérique ont été regroupées et concentrées. Puis un échantillon du rétentat a été réinjecté sur la même colonne afin de favoriser la détection du dimère (**Figure 95B**). D'après le chromatogramme correspondant, une nouvelle espèce en très faible quantité se forme et est retrouvée à un volume d'éluion d'environ 10,9 ml. L'analyse par Western blot avec un anticorps anti-FLAG a montré qu'il s'agissait bien d'ED-4 C150A et non pas d'un contaminant (**Figure 95C**). Sa masse estimée sur gel est d'environ 57 kDa, soit environ deux fois la masse estimée sur gel du monomère d'ED-4 C150A ( $\approx 28$  kDa). Une très faible quantité de dimère est également détectée par Western blot dans la fraction d'éluion correspondant à la forme monomérique (*pic II* sur les **Figures 95B** et **95C**).



**Figure 95 : Détection d'une faible quantité de dimère d'ED-4 C150A.** **A)** Analyse par spectrométrie de masse en conditions natives (ESI-MS) d'ED-4 C150A purifié. **B)** 100  $\mu$ l d'ED-4 C150A à 1 mg/ml ont été injectés sur une colonne Superdex S200 Increase 10/300 GL, GE Healthcare. Deux espèces sont détectées (I et II). Insert : Agrandissement du chromatogramme centré sur le dimère d'ED-4 C150A. **C)** 15  $\mu$ l des fractions éluées I et II issues de la purification par chromatographie d'exclusion stérique effectuée en *B*) ont été analysés par Western blot avec un anticorps anti-FLAG avec une détection par chimiluminescence.

La forme dimérique, bien qu'étant en quantité très faible, est résistante au SDS étant donné qu'elle est détectée lors de l'analyse par Western blot (**Figure 95C**). Seules des interactions fortement hydrophobes pourraient *a priori* être à l'origine de la conservation du dimère lors des analyses par électrophorèse. Comme décrit précédemment, les résidus présents à l'extrémité C-terminale de l'isoforme-2 d'ED-4 sont majoritairement hydrophobes (VLAGCPEH), puisqu'ils sont retrouvés en partie dans le domaine transmembranaire de l'isoforme-1 (*voir Figure 88*).

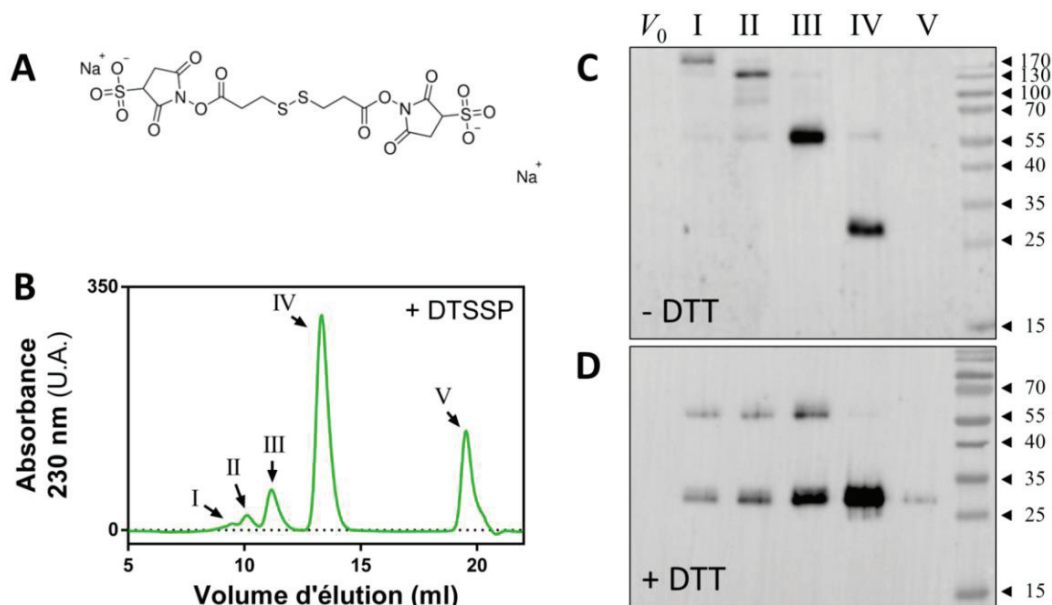
La dimérisation d'ED-4 pourrait alors être médiée dans un premier temps par des interactions hydrophobes entre les deux monomères avant qu'ils soient covalentement liés par un pont disulfure dans la forme sauvage d'ED-4, ce qui expliquerait la faible abondance de dimère chez le mutant dépourvu de résidu de cystéine. Des expériences de réticulation sur le mutant ED-4 C150A ont alors été réalisées afin de déterminer la capacité du mutant C150A à multimériser. La réticulation d'ED-4 C150A a été réalisée en présence de glutaraldéhyde pendant 10 minutes à température ambiante (**Figure 96**). Les conditions ont été suffisamment douces pour éviter la formation d'agrégats comme en témoigne l'absence d'espèces éluées dans le volume mort de la colonne d'exclusion stérique (8 ml). Deux espèces se forment aux dépens de la forme monomérique (13,2 ml, masse moléculaire apparente 92,5 kDa), et sont éluées à environ 10,4 (322 kDa) et 11,5 ml (197 kDa), la 1ère étant relativement proche du volume d'éluion du dimère non réticulé (10,9 ml, 257 kDa). Cette première expérience montre que le mutant C150A peut former des multimères.



**Figure 96 : Réticulation d'ED-4 C150A par le glutaraldéhyde (GA).** 100  $\mu$ l d'ED-4 C150A (1 mg/ml) non réticulé (vert) ou réticulé par du glutaraldéhyde 20 mM (noir) ont été injectés sur une colonne de chromatographie d'exclusion stérique (Superdex S200 Increase 10/300 GL, GE Healthcare). Les volumes d'éluion du mutant avec (noir) ou sans glutaraldéhyde (vert) sont indiqués. Le volume mort de la colonne ( $\leq$  8 ml) est représenté comme une aire grisée.

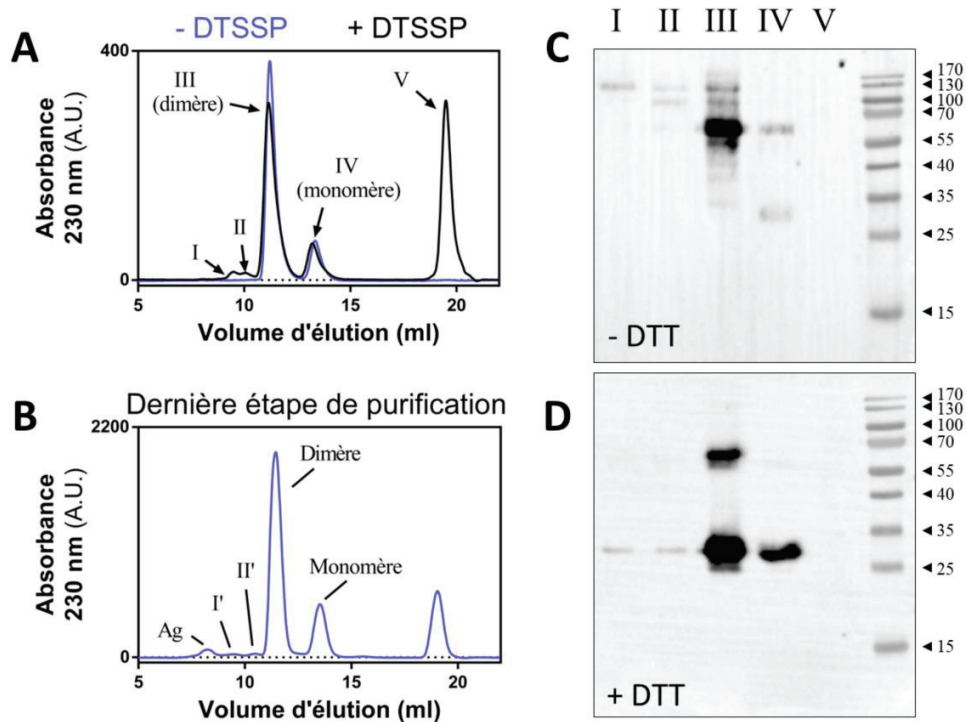
Une deuxième série d'expériences de réticulation a été réalisée en utilisant le 3,3'-dithiobis(sulfosuccinimidylpropionate), DTSSP, qui possède un pont disulfure (**Figure 97A**). Si les monomères interagissent par des interactions hydrophobes pour initier la dimérisation, la réduction des oligomères réticulés ne devrait pas mener à la seule apparition d'une bande correspondant à celle du monomère, mais aussi à celle du dimère, puisque les deux chaînes restent liées par des interactions hydrophobes.

L'analyse du chromatogramme d'ED-4 C150A réticulé par le DTSSP montre la formation de trois espèces éluées respectivement à 9,2, 10,1 et 11,2 ml (*pics I à III*), la majorité restant non réticulée (*pic IV*) et la dernière espèce éluée dans le volume total de la colonne correspondant DTSSP (*pic V*) (**Figure 97B**). Ces espèces ont été analysées par Western blot avec un anticorps anti-FLAG (**Figure 97C**). Ils migrent avec une masse moléculaire apparente de 170 kDa (*pic I*), 130 kDa (*pic II*), 57 kDa (*pic III*) et 28 kDa (*pic IV*), soit des rapports de masse de 2 (*pic III*), 4,6 (*pic II*) et 6 fois (*I*) celle du monomère. Ces fractions d'éluion ont ensuite été réduites par du DTT et analysées suivant le même protocole (**Figure 97D**). Excepté pour la fraction correspondant au monomère (*IV*), deux bandes sont systématiquement retrouvées, migrant à la masse moléculaire attendue pour la forme dimérique d'une part et la forme monomérique d'autre part (environ 57 et 28 kDa, respectivement). Cette observation suggère que les monomères interagissent *via* des interactions hydrophobes.



**Figure 97 : Réticulation d'ED-4 C150A par le DTSSP. A)** Structure du DTSSP. **B)** ED-4 C150A (1 mg/ml) a été incubé pendant 30 min à 10°C avec du DTSSP, puis 100 µl ont été injectés sur une colonne de chromatographie d'exclusion stérique (Superdex S200 Increase 10/300 GE Healthcare). 15 µl d'éluats, notés I à IV, ont ensuite été analysés par Western blot avec un anticorps anti-FLAG soit **(C)** sans réduction (- DTT) soit **(D)** après réduction par du DTT 50 mM pendant 20 minutes à température ambiante. Les bandes immunoréactives ont été détectées par chimiluminescence. La bande immunoréactive présente dans la fraction V est vraisemblablement une erreur de manipulation.

Une autre série d'expériences a été conduite sur la forme sauvage d'ED-4 dans les mêmes conditions (**Figure 98**). La réticulation d'ED-4 sauvage mène à une faible formation de deux espèces (*pics I et II* élués respectivement à 9,5 et 10 ml) ayant des volumes hydrodynamiques supérieurs à ceux du dimère et du monomère (*pics III et IV*, respectivement), l'espèce *V* correspondant au DTSSP (**Figure 98A**). Les espèces réticulées (*I et II*) semblent se former aux dépens du dimère (*III*) mais pas aux dépens du monomère (*IV*). Cette observation pourrait indiquer un mécanisme de formation d'oligomères nécessitant au préalable la formation d'un dimère. Ces espèces possèdent des volumes d'éluions semblables aux espèces éluées lors de la dernière étape de purification d'ED-4 par chromatographie d'exclusion stérique (**Figure 98B**). Cependant, il est également possible que ces oligomères se forment artificiellement lors de l'étape de concentration des protéines.



**Figure 98 : Réticulation d'ED-4 sauvage par le DTSSP. A)** 100  $\mu$ l d'ED-4 sauvage (1 mg/ml) non réticulé (bleu) ou réticulé par du DTSSP (noir) ont été injectés sur une colonne de chromatographie d'exclusion stérique (Superdex S200 Increase 10/300 GL). **B)** Chromatogramme de la dernière étape de purification d'ED-4 par chromatographie d'exclusion stérique effectuée sur la même colonne. Ag : agrégats. **C-D)** Analyse par Western blot avec un anticorps anti-FLAG des fractions réticulées par le DTSSP. 15  $\mu$ l d'éluats issus de la colonne Superdex S200 Increase 10/300 GL de *A*), notés I à V, ont été analysés soit **(C)** sans réduction (- DTT) soit **(D)** après réduction par du DTT 50 mM pendant 20 minutes à température ambiante. Les protéines ont été révélées par chimiluminescence.

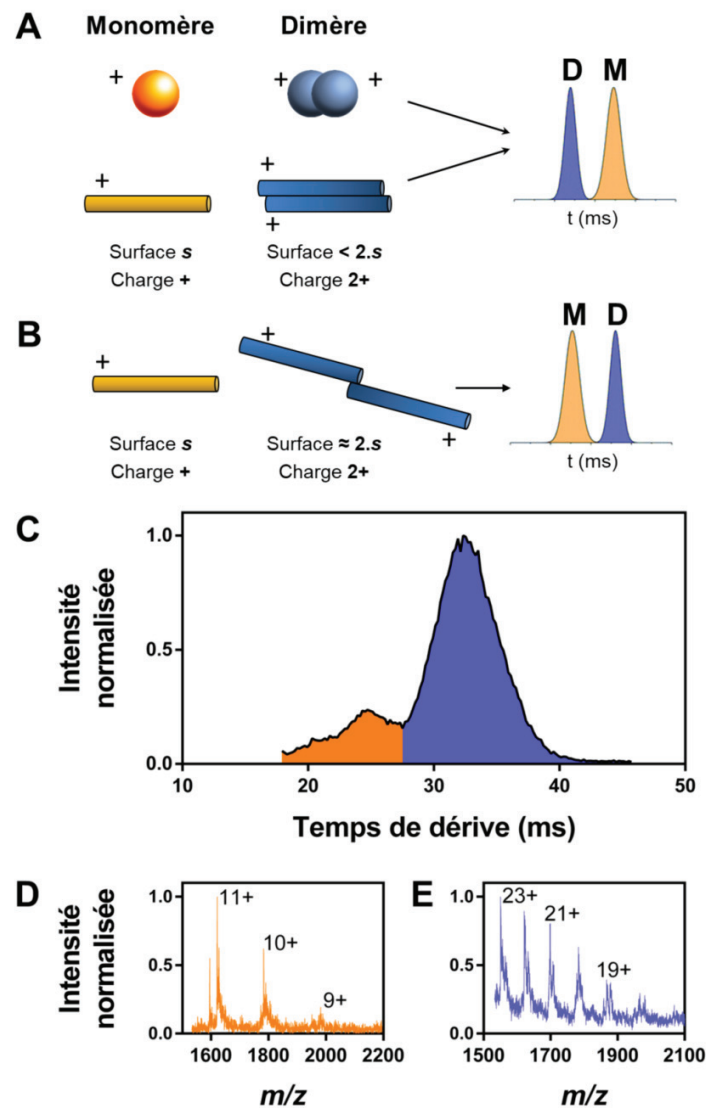
Comme observé pour ED-4 C150A, la réduction des espèces issues de la réticulation de la forme sauvage d'ED-4 conduit à la disparition des bandes de masses moléculaires les plus élevées et à l'apparition d'une bande correspondant à la forme monomérique. Cependant, seul le dimère (*pic III*) conserve à la fois une bande correspondant au monomère et au dimère. L'absence d'une bande correspondant au dimère lors de la réduction des espèces à hautes masses moléculaires (*pics I et II*) est probablement due à une trop faible quantité de protéine déposée rendant ainsi leur détection difficile (**Figure 98D**).

Ces expériences montrent qu'en absence du résidu de cystéine, le mutant conserve sa capacité à dimériser mais en très faible quantité. Cette faible abondance de dimère peut être causée par des interactions hydrophobes à l'extrémité C-terminale de la protéine mais elles ne permettent pas de stabiliser le dimère en absence d'un pont disulfure. La valeur du rayon de gyration du dimère de la forme sauvage ( $59,25 \pm 2,27 \text{ \AA}$ ) est bien supérieure à celui du monomère ( $41,53 \pm 0,67 \text{ \AA}$ ) et est similaire au Rg théorique calculé pour une protéine possédant deux fois plus de résidus ( $62,2 \pm 14,1 \text{ \AA}$ ). De plus, le rayon hydrodynamique du dimère est plus important que celui du monomère de la forme sauvage d'ED-4 ( $45,4 \pm 1,2$  contre  $21,6 \pm 4,0 \text{ \AA}$  ou  $35,7 \pm 0,8 \text{ \AA}$  selon les méthodes d'analyse). Ces mesures suggèrent que les deux monomères au sein du dimère adoptent une conformation dans laquelle les deux monomères interagissent par leur extrémité C-terminale plutôt qu'un arrangement latéral. Afin de vérifier cette hypothèse, des expériences de mobilité ionique ont été réalisées sur la forme sauvage d'ED-4 (**Figure 99**).

Si l'on considère un dimère formé par l'arrangement latéral des deux monomères, la surface du dimère est inférieure au double de la surface d'un monomère seul car une zone d'interaction importante existe entre les deux monomères. La surface étant proche de celle d'un monomère et la quantité de charges étant doublée, une telle conformation devrait avoir un temps de dérive inférieur à celui de la forme monomérique (**Figures 99A et 99B**). A l'inverse, une conformation où les deux monomères interagissent par l'une de leur extrémité possède une surface environ deux fois plus importante que celle d'un monomère seul. Dans ce cas, le dimère devrait être davantage ralenti par des collisions au sein du tube de dérive et ainsi avoir un temps de dérive supérieur à celui du monomère.

Lors de ces expériences, deux populations possédant des temps de dérive d'environ 25 et 32 ms sont effectivement retrouvées sur la distribution globale du fait de la présence des formes monomérique et dimérique d'ED-4 (**Figure 99C**).

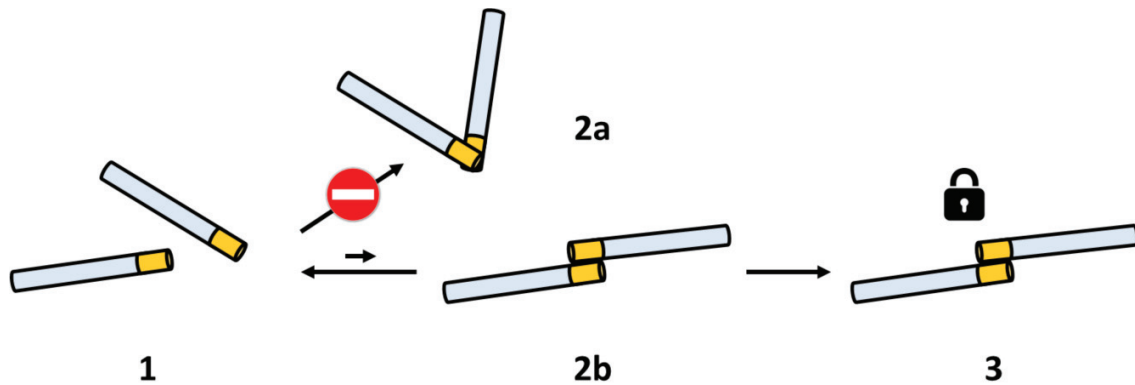
La déconvolution de ces espèces a montré que le premier pic (25 ms) pouvait être attribué à la forme monomérique d'ED-4 alors que le second pic (32 ms) correspond à sa forme dimérique (**Figures 99D et 99E**). La forme monomérique possède un temps de dérive inférieur à celui du dimère, ce-dernier étant davantage ralenti par ses collisions dans le tube de dérive. Cette observation montre que les deux monomères d'ED-4 s'assemblent de façon étendue au sein du dimère, en accord avec les expériences de SEC-SAXS et de SEC-DLS.



**Figure 99 : Détermination de l'arrangement des monomères au sein du dimère d'ED-4 par mobilité ionique. A-B)** Schémas de la distribution globale théorique des formes monomérique (M, orange) et dimérique (D, bleu) pour des conformations compactes (sphère) ou étendues (bâton). **C)** Distribution globale obtenue expérimentalement pour la forme sauvage d'ED-4. **D-E)** Déconvolution des spectres de masse des deux espèces observées en *C*). Certains états de charge sont mentionnés. La légende des couleurs est commune à l'ensemble de la figure.



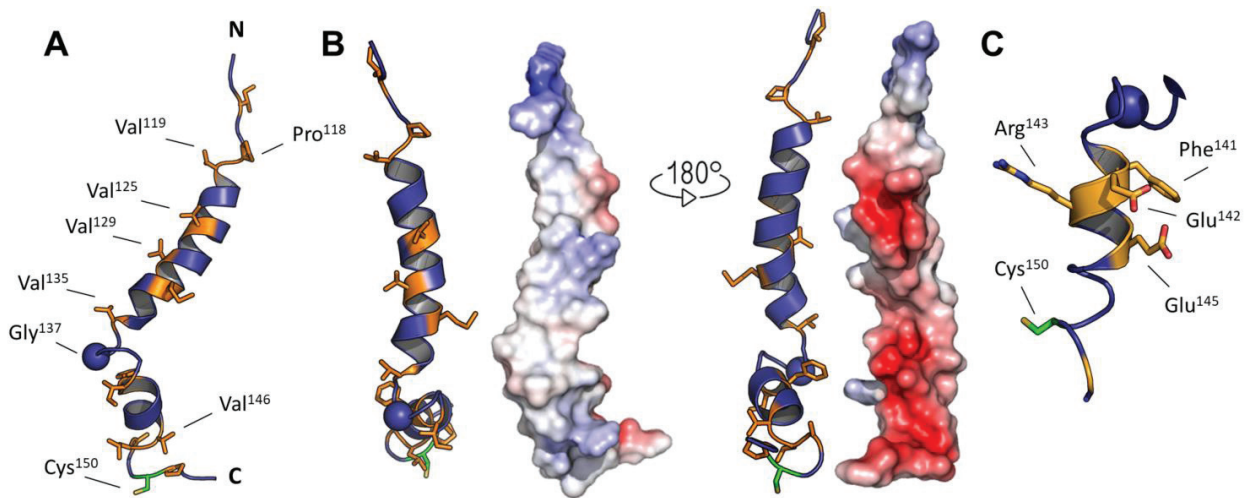
Il est probable que les interactions non covalentes entre les deux monomères permettent de les positionner spécifiquement. La seconde étape reposerait sur la stabilisation du dimère par la formation d'un pont disulfure. Ce mécanisme de dimérisation d'ED-4 est schématisé dans la figure ci-dessous (**Figure 100**).



**Figure 100 : Mécanisme de dimérisation proposé pour ED-4.** Les deux monomères d'ED-4 (étape 1) peuvent s'associer uniquement de façon étendue (étapes 2a/2b) *via* des résidus hydrophobes présents à leur extrémité C-terminale (en jaune). Le dimère est ensuite stabilisé par la formation d'un pont disulfure (étape 3).

La structure tridimensionnelle des 40 résidus C-terminaux d'ED-4 (Lys116 - His155) a ensuite été prédite étant donné que ces résidus ne sont pas prédits comme désordonnés par metaPrDOS (*voir Figure 61*). La prédiction de la structure 3D a été réalisée en utilisant le serveur *UNRES* (*Czaplewski et al., 2018*), sans inclure l'étiquette FLAG, présente à cette extrémité dans la protéine recombinante (**Figure 101**). Cinq structures ont été modélisées ayant une probabilité respective de 0,595, 0,215, 0,099, 0,080 et 0,011. Le modèle ayant la meilleure probabilité (0,595) a été utilisé pour les analyses suivantes. Le peptide adopte de façon majoritaire une structuration en hélice  $\alpha$ . Une première hélice  $\alpha$  de 17 résidus (Val119 – Val135) est délimitée par des résidus de proline 118 et de glycine 137. La seconde hélice  $\alpha$ , plus courte (Ile140 – Val146), précède la séquence propre à ED-4 (VLAGCPEH) qui est enrichie en résidus hydrophobes et qui possède le résidu de cystéine (**Figure 101A**).

Sur l'ensemble du peptide, les résidus hydrophobes sont majoritairement représentés par les résidus de valine ( $\approx 39$  % des résidus hydrophobes). Les résidus hydrophobes forment une face plutôt apolaire qui pourrait favoriser l'établissement d'interactions hydrophobes au sein du dimère, comme en atteste le potentiel électrostatique neutre (**Figure 101B**). L'autre face est cependant enrichie en résidus acides, notamment en glutamates, qui représentent 17,5 % des résidus du peptide.



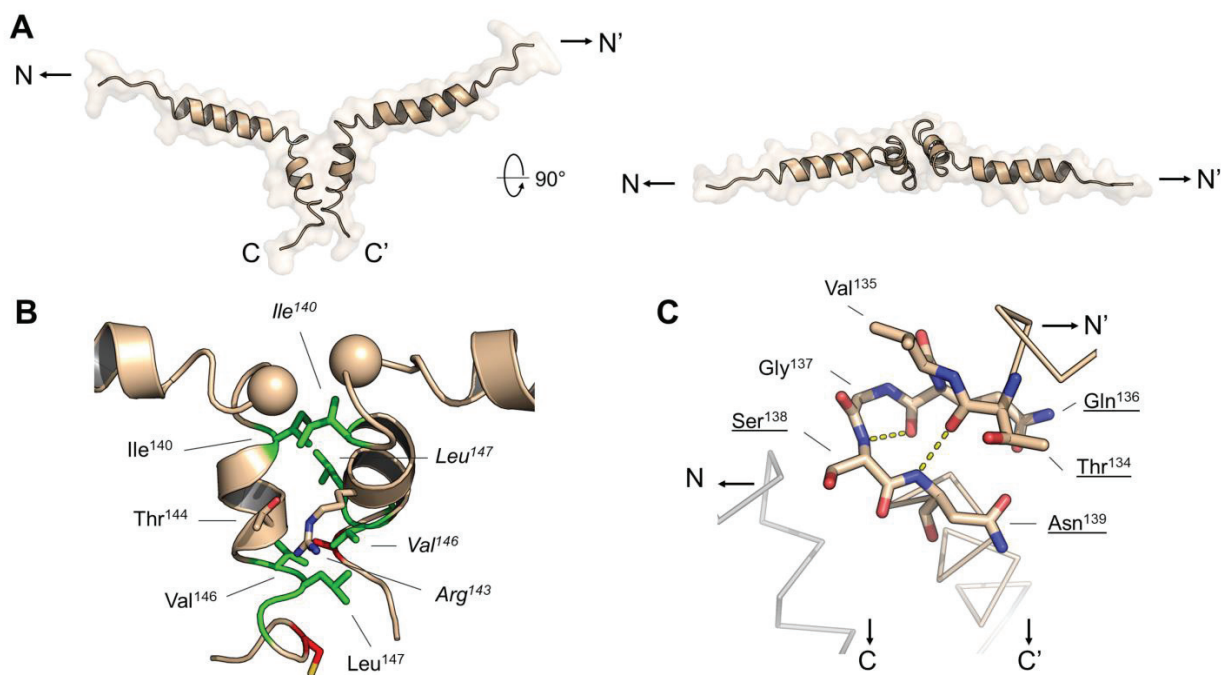
**Figure 101 : Prédiction de la structure tridimensionnelle de l'extrémité C-terminale d'ED-4.**

**A)** Structure 3D prédite par UNRES (Czaplewski *et al.*, 2018) des 40 derniers résidus d'ED-4 ( $^{116}$ KRISPVEESEDVSNKVSMSSTVQGSNIFERTEVLAGCPEH $^{155}$ ). Les résidus de valine, les résidus flanquant les deux hélices  $\alpha$  ainsi que les extrémités N- et C-terminales sont annotés. **B)** Potentiel électrostatique (bleu : basique, rouge : acide) du peptide et localisation des résidus hydrophobes (orange) sur la structure prédite. **C)** Agrandissement de la seconde hélice portant les résidus de phénylalanine, d'arginine et de glutamate situés en amont du résidu de cystéine (en vert).

Concernant la plus courte hélice présente en amont du résidu de cystéine et formée par les résidus  $^{140}$ IFERTEV $^{146}$ , deux mécanismes supplémentaires pourraient favoriser la formation du dimère en solution. Le premier repose sur l'établissement d'interactions de type  $\pi$ - $\pi$  du fait de la présence d'un résidu de phénylalanine, présent à proximité du résidu de cystéine. Au sein du domaine transmembranaire des syndécans, certains résidus de phénylalanine favorisent d'ailleurs la formation d'homo et d'hétérodimères de syndécans (Kwon *et al.*, 2016). Le second repose sur la présence des résidus de glutamate et d'arginine et donc sur une potentielle formation de ponts salins entre les monomères, notamment du fait de leur orientation (**Figure 101C**).

Afin de mimer la structure du dimère, le modèle prédit par UNRES des 40 derniers résidus d'ED-4 a été docké sur lui-même en utilisant UNRES-Dock (Czaplewski *et al.*, 2018) où aucune contrainte n'a été imposée (**Figure 102**).

Cinq poses de l'extrémité C-terminale d'ED-4 interagissant avec lui-même ont été générées par UNRES-Dock et possèdent une probabilité de 0,373, 0,261, 0,251 et 0,102. La pose ayant la meilleure probabilité (0,373) a été utilisée pour les analyses suivantes (**Figure 102**). Les deux monomères forment une structure globale en 'V' en interagissent *via* l'hélice  $\alpha$  la plus courte ( $^{140}\text{IFERTEV}^{146}$ ) et par des interactions hydrophobes impliquant les résidus d'isoleucine 140, de valine 146 et de leucine 147. La formation de liaisons hydrogène entre le résidu d'arginine 143 du premier monomère avec le résidu de thréonine 144 du second pourrait également stabiliser le dimère (**Figures 102A** et **102B**). Dans cette configuration, les résidus de cystéine restent éloignés (environ 11 Å) alors que la distance séparant les carbones  $\alpha$  de deux résidus de cystéine formant un pont disulfure est en moyenne de 5,63 Å (*Schmidt et al., 2006*). La flexibilité de cette région pourrait cependant leur permettre de se retrouver à une distance suffisamment proche pour pouvoir en former un.

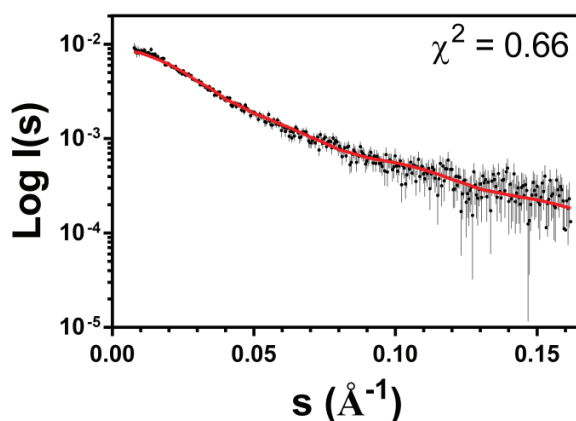


**Figure 102 : Docking de l'extrémité C-terminale d'ED-4 sur elle-même. A)** Pose la plus probable calculée par UNRES-Dock de la région C-terminale d'ED-4 (40 résidus d'acides aminés) sur elle-même. Les extrémités N- et C-terminales des peptides sont indiquées. **B)** Agrandissement de la zone d'interaction entre les deux peptides. Vert : résidus hydrophobes, beige : résidus formant des liaisons hydrogène (Arg<sup>143</sup> et Thr<sup>144</sup>), rouge : résidus de cystéine 150 (non annotés). Sphères : résidus de glycine 137. **C)** Agrandissement de la boucle présente entre les deux hélices  $\alpha$ . Les résidus soulignés sont ceux impliqués dans la formation de deux liaisons hydrogène (pointillés jaunes).

Le segment <sup>134</sup>TVQGSN<sup>139</sup> séparant les deux hélices  $\alpha$  forme une petite boucle stabilisée par deux liaisons hydrogène impliquant les résidus de thréonine 134 et d'asparagine 139 d'une part, et les résidus de glutamine 136 et de sérine 138 d'autre part (**Figure 102C**). Cette boucle semble essentielle afin de placer les deux chaînes de façon opposée entraînant une organisation étendue des deux monomères au sein du dimère d'ED-4.

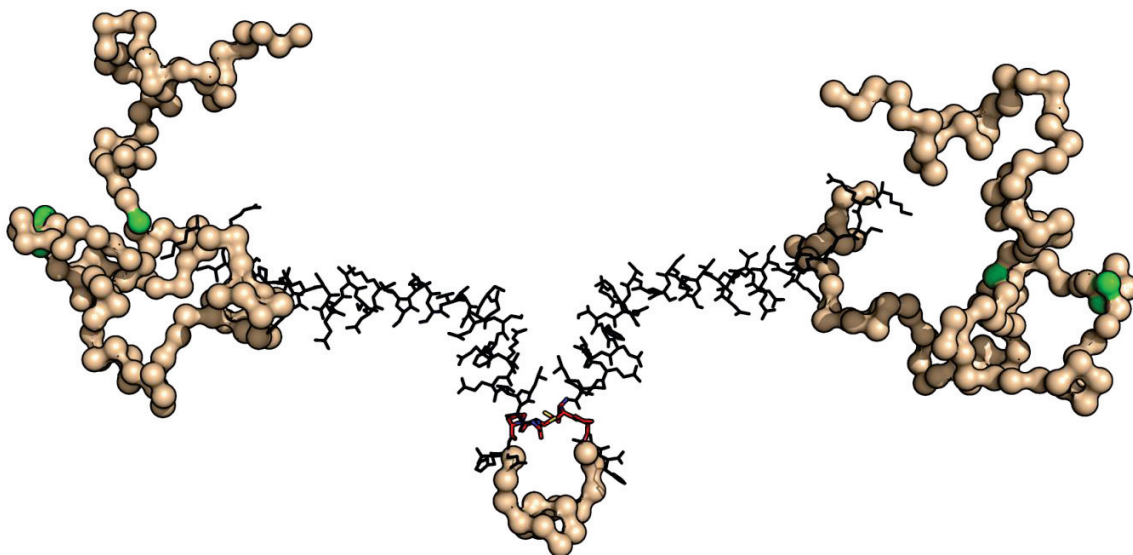
La formation d'une région repliée aussi longue (environ 40 résidus sur 163, soit près de 25 % de la séquence d'ED-4) implique la compaction, du moins partielle, d'ED-4, ce qui impacterait son rayon de gyration. Afin de vérifier si la formation d'une telle structure était compatible avec les données expérimentales de SEC-SAXS, la totalité de la séquence d'ED-4 a été modélisée. Le but n'est pas de proposer une structure figée qui réduirait l'hétérogénéité conformationnelle d'ED-4 mais de vérifier que le rayon de gyration d'une telle structure soit du même ordre de grandeur que celui calculé expérimentalement pour la forme dimérique d'ED-4. Les données de SEC-SAXS sont utilisées en tant que contraintes. Ainsi, si le rayon de gyration du dimère d'ED-4 possédant son extrémité C-terminale repliée est trop faible par rapport au rayon de gyration expérimental, les segments manquants (*i.e.* le reste de la chaîne non modélisée précédemment) devraient être extrêmement étendus pour combler la différence de rayon de gyration entre la valeur expérimentale et la valeur obtenue sur le dimère d'ED-4 possédant son extrémité C-terminale repliée.

La modélisation de la totalité d'ED-4 sous forme dimérique a été réalisée avec BUNCH (*Petoukov & Svergun, 2005*) en utilisant les données de SEC-SAXS collectées pour sa forme dimérique ainsi que le modèle de son extrémité C-terminale sous forme monomérique et en contraignant les résidus de glycine 149, de cystéine 150 et de proline 151 à se situer à moins de 6 Å de leur homologue dans la deuxième chaîne afin de mimer leur rapprochement qui serait normalement causé par la présence du pont disulfure. Seules les intensités aux petits angles ont été utilisées par BUNCH dans la modélisation (jusqu'à  $\approx 0,16 \text{ \AA}^{-1}$ ) mais l'ajustement entre les données expérimentales et le modèle est bon d'après la valeur du  $\chi^2$  (0,66) (**Figure 103**).



**Figure 103 : Ajustement des intensités de diffusion théoriques du modèle calculé par BUNCH pour la totalité de la séquence d'ED-4 sous forme dimérique sur les données expérimentales.** Données expérimentales : noir, ajustement : rouge.

Concernant l'extrémité repliée, la contrainte de distance imposée aux résidus de glycine 149, de cystéine 150 et de proline 151 a permis le rapprochement entre les deux monomères et dans une orientation similaire à la pose précédemment obtenue par docking moléculaire. Le reste de la chaîne adopte une conformation aléatoire et n'est pas étirée, ce qui suggère qu'un degré de compaction similaire pourrait effectivement être retrouvé pour le dimère d'ED-4 (**Figure 104**).



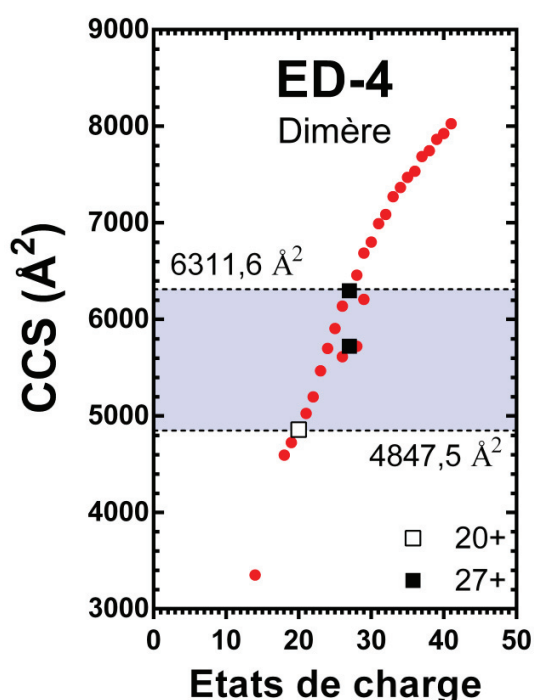
**Figure 104 : Modélisation de la totalité de la séquence d'ED-4 sous forme dimérique.**

Modèle calculé par BUNCH (*Petoukov & Svergun, 2005*) de la totalité de la séquence d'ED-4 sous forme dimérique. Noir (lignes) : domaine replié dont la structure a été modélisée précédemment par docking. Beige : reste de la chaîne où chaque résidu est représenté comme une sphère centrée sur son carbone  $\alpha$ . Le reste de la chaîne comprend les étiquettes 6xHis et FLAG aux extrémités N- et C-terminales, respectivement. Les résidus de glycine 149, de cystéine 150 et de proline 151 contraints à se rapprocher à moins de 6 Å sont en rouge et les résidus fictifs (*dummy residues*) portant théoriquement les chaînes de glycosaminoglycanes (absentes dans les expériences de SEC-SAXS) sont en vert (sérines 39, 61 et 63).

Un second moyen de vérifier si une telle conformation globale est cohérente avec nos données expérimentales a été de calculer la valeur de la section efficace de collision théorique avec IMPACT (*Marklund et al., 2015*). Le squelette peptidique et les chaînes latérales ont d'abord été reconstruits avec SABBAC (*Maupetit et al., 2006*) puis l'énergie du système a été minimisée, au cas où des clashes stériques seraient apparus lors de l'ajout des chaînes latérales.

IMPACT propose deux valeurs de CCS pour le modèle tout atome de la forme dimérique d'ED-4 selon deux méthodes de calcul, et ces valeurs sont égales à environ 6 300 Å<sup>2</sup> pour la méthode la plus complexe (dite de trajectoire) et environ 4 850 Å<sup>2</sup> pour la méthode la plus simple, dite de projection (**Figure 105**). Ces valeurs calculées sont dans la gamme de CCS mesurées expérimentalement lors des expériences de mobilité ionique (de 3 351 à 8 828 Å<sup>2</sup>). La valeur la plus faible (4 850 Å<sup>2</sup>) et la plus élevée (6 300 Å<sup>2</sup>) correspondraient à une espèce chargée respectivement 20 et 27 fois. A noter également que deux valeurs de CCS avaient été mesurées expérimentalement pour l'état de charge 27+, ce qui pouvait suggérer soit la coexistence de deux populations, soit l'existence de deux phénomènes d'ionisation distincts (**Figure 105**).

Les états de charge 20 et 27+ correspondent à des conformations plutôt compactes et intermédiaires, respectivement, en accord avec le repliement partiel de la protéine à son extrémité C-terminale.



**Figure 105 : CCS calculées pour le modèle du dimère d'ED-4 au sein de la distribution des CCS expérimentales.** Les valeurs de CCS ont été calculées par IMPACT et la gamme les séparant est colorée en bleu.

Le modèle proposé par BUNCH ainsi que le calcul théorique de la section efficace de collision montrent ainsi qu'une conformation possédant un domaine replié est compatible avec les données expérimentales obtenues par SEC-SAXS et par mobilité ionique. Ce modèle ne représente pas en revanche l'hétérogénéité conformationnelle du dimère du syndécan-4.

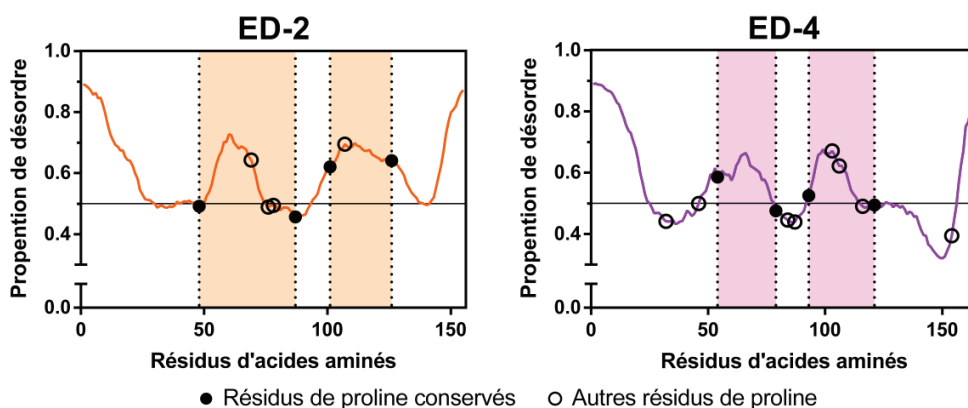
#### IV.1.6 Discussion

Ces travaux représentent la première caractérisation biophysique des ectodomaines des syndécans, à l'exception de la résolution de la structure cristallographique d'un octapeptide de l'ectodomaine du syndécan-1 humain en complexe avec un de ses partenaires (*Briggs & Hohenester, 2018*) et de prédictions *in silico* de désordre intrinsèque (*Leonova & Galzitskaya, 2015*). L'analyse des ectodomaines des syndécans a permis de montrer qu'ils étaient intrinsèquement non structurés et qu'ils possèdent une grande hétérogénéité conformationnelle composée majoritairement de conformations très étendues.

L'ensemble de ces travaux a permis de caractériser l'isoforme 2 du syndécan-4 (ED-4), dont l'unique preuve d'existence réside dans l'analyse du transcriptome de l'utérus, du cœur fœtal et dans certains cancers de la peau (*Xing et al., 2003*). Elle est la seule à posséder un résidu de cystéine dans sa séquence ce qui lui permet de dimériser *via* un pont disulfure. Seule l'isoforme-2 du syndécan-4 de l'Homme possède un résidu de cystéine, la séquence de ses homologues dans d'autres espèces étant soit raccourcie au résidu de glycine le précédent, soit de taille plus importante mais où ce résidu est substitué par un autre résidu d'acide aminé. L'analyse de son mécanisme de dimérisation par différentes techniques suggère l'implication de résidus hydrophobes présents à son extrémité C-terminale. Les deux monomères s'assemblent en conformation étendue ce qui permet de conserver un très faible degré de compaction et une très grande flexibilité. Son extrémité C-terminale est prédite comme structurée et la prédiction de sa structure tridimensionnelle montre un repliement en hélices  $\alpha$  ce qui est en accord avec des simulations par dynamique moléculaire où l'ectodomaine du syndécan-4 se repliait partiellement en hélice  $\alpha$  tout en gardant une part importante de désordre intrinsèque (*Cruz-Chu et al., 2014*). La proportion en hélice  $\alpha$  mesurée expérimentalement par dichroïsme circulaire devrait en revanche être bien plus importante puisqu'environ 40 résidus sur 163 (soit environ 25 % des résidus) sont repliés en hélice  $\alpha$  dans la structure prédite. Cette séquence est cependant prédite comme pouvant se replier (MoRFs) ce qui expliquerait ces résultats. Le shedding du syndécan-4 est causé par les métalloprotéases MMP-7 et MT1-MMP dont le clivage protéolytique s'effectue entre les résidus d'asparagine 139 et d'isoleucine 140 (*Manon-Jensen et al., 2013*). Ces résidus sont retrouvés dans le court segment non replié séparant les deux hélices  $\alpha$  présentes dans la structure 3D prédite, ce qui est cohérent avec sa sensibilité aux protéases.

En revanche, le clivage opéré entre les résidus de sérine 130 et de méthionine 131 par la MT1-MMP se situe dans la plus longue hélice  $\alpha$  (Val119 – Val135), ce qui suggère que l'extrémité C-terminale d'ED-4 n'est pas systématiquement repliée, ce qui conforte l'existence d'une séquence MoRF à cette extrémité, d'autant plus que les résidus de sérine 130 et de méthionine 131 possèdent une propension prédite de désordre d'environ 0,49 c'est-à-dire à la limite du seuil ordre / désordre (0,5). Ce dimère soluble pourrait servir de leurre et d'antagoniste dans le milieu extracellulaire du fait de la conservation de ses interactions avec ses partenaires, notamment grâce à ses chaînes de glycosaminoglycanes. A l'inverse, il pourrait avoir un rôle de réservoir où il concentrerait des facteurs de croissance dans l'espace péricellulaire.

Les analyses de diffusion de la lumière aux petits angles ont mis en évidence le caractère très flexible des EDs ce qui leur permet d'adopter de façon préférentielle des conformations très étendues, de type *random coil*. Cette flexibilité leur offre également une plasticité structurale où la formation d'hélice  $\alpha$  est induite par l'incubation avec du trifluoroéthanol malgré le fait que les EDs soient enrichi en résidu de proline (entre 6 et 13 %). L'alignement des séquences d'ED-1 et d'ED-3 d'une part et d'ED-2 et d'ED-4 d'autre part met d'ailleurs en évidence la conservation de certains résidus de proline. Ces résidus conservés chez ED-2 et ED-4 semblent encadrer des régions prédites comme étant désordonnées (**Figure 106**). Ces régions pourraient ainsi définir des domaines d'interaction et subir une transition désordre – ordre lors de leur interaction avec leurs partenaires, en accord avec les prédictions des MoRFs.

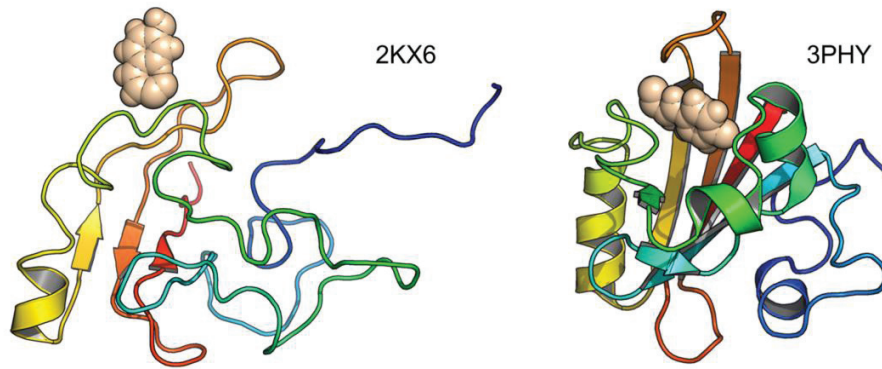


**Figure 106 : Prédiction du désordre intrinsèque de l'ectodomaine du syndécan-2 (ED-2) et de l'isoforme-2 du syndécan-4 (ED-4) et localisation de leurs résidus de proline.** La courbe en couleur représente la prédiction de désordre (seuil : 0,5, ligne noire). Les résidus de proline conservés après alignement entre les deux séquences sont représentés par des points noirs et encadrent des régions (en couleur, délimitées par des lignes verticales en pointillé) prédites comme non structurées. Les autres résidus de proline sont représentés par des cercles.



L'analyse de la distribution des états de charge obtenus lors des expériences de spectrométrie de masse a permis de définir trois populations de conformations globales possédant différents degrés de compaction. En accord avec la déconvolution des données de SAXS par EOM, les populations de conformations les plus étendues sont majoritaires pour ED-1, -2 et -4 ce qui est vraisemblablement la conséquence de leur déplétion en résidus hydrophobes et de leur enrichissement en résidus de proline (6 – 9 %), en résidus acides (16 – 23 %) et en résidus polaires, notamment de sérine et de thréonine (28 – 34 %), ces derniers favorisant leur interaction avec les solvants aqueux. A l'inverse, les conformations intermédiaires d'ED-3 dominent la distribution des états de charges en ESI-MS. Pour ED-3, l'inadéquation entre les expériences de SAXS dont l'analyse était en faveur de la présence de conformations très étendues et celles d'ESI-MS où les conformations intermédiaires étaient majoritaires pourrait être expliquée par l'absence de solvant aqueux lors des expériences de spectrométrie de masse. Aussi, les interactions hydrophobes sont conservées en phase gazeuse (*Frański et al., 2008; Barylyuk et al., 2011*). Elles pourraient ainsi être la cause de la forte abondance des populations intermédiaires chez ED-3, dont la séquence est enrichie en résidus apolaires ( $\approx 44\%$ ) contre 31 à 35 % pour les autres EDs. A l'inverse, la présence d'un domaine de type mucine enrichi en résidus polaires, notamment de thréonine, favoriserait son interaction avec le solvant et donc le dépliement de la chaîne en solution.

L'observation d'une telle hétérogénéité conformationnelle coexistante est un phénomène commun chez les protéines désordonnées et est facilement observable lors d'expériences de spectrométrie de masse ou de résonance magnétique nucléaire (*Salvi et al., 2017, Mitra, 2019, Santambrogio et al., 2019*). Certaines protéines désordonnées ont la capacité à se replier dans des conformations similaires à celles des protéines globulaires et de conserver une conformation compacte lorsqu'elles sont non structurées comme la *photoactive yellow protein* (PYP, *Düx et al., 1998, Ramachandran et al., 2011*) (**Figure 107**). Cela a également déjà été observé pour des protéines ou des domaines repliés comme le domaine FNIII<sub>10</sub> de la fibronectine (*Gee et al., 2013*).



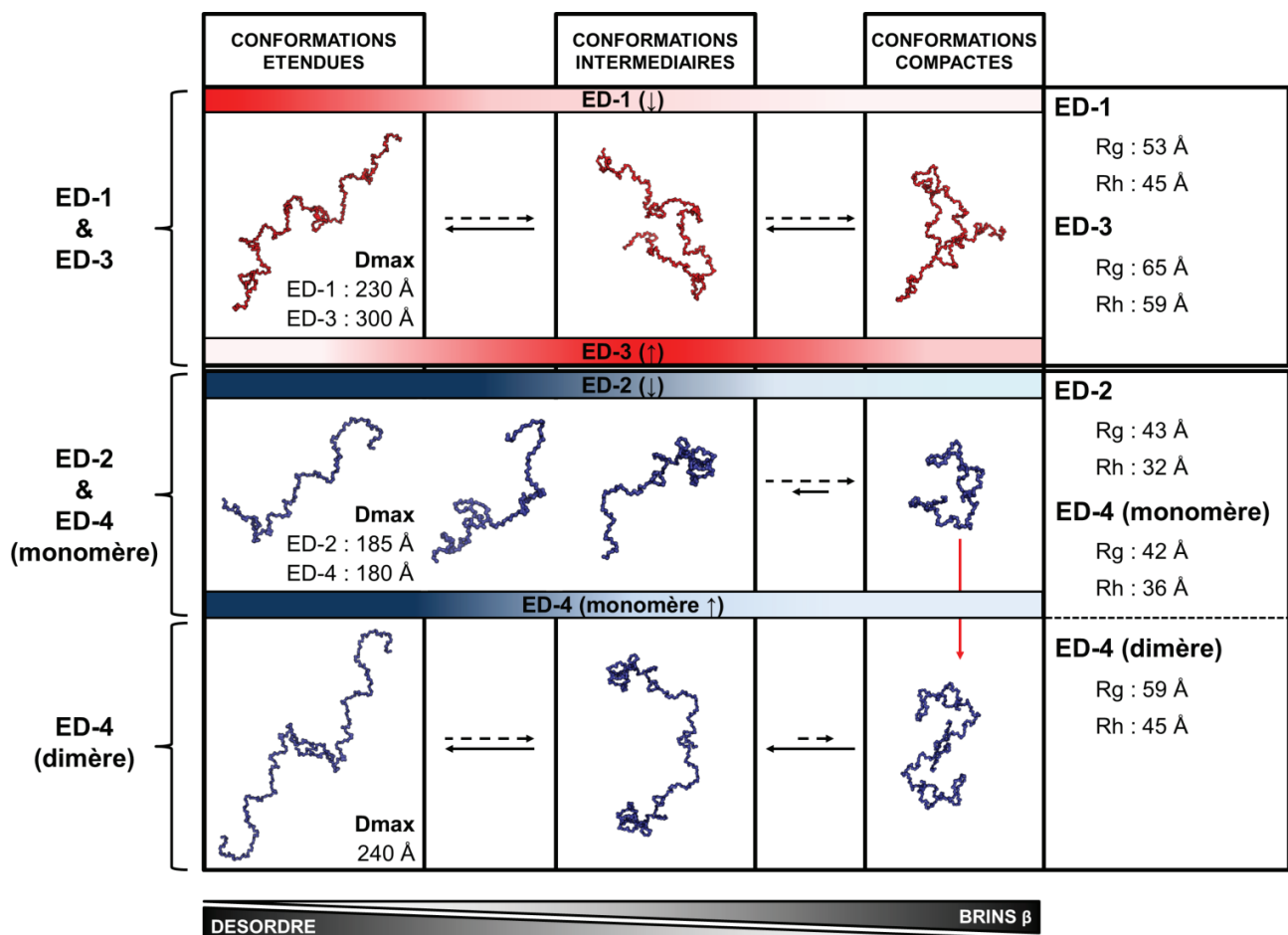
**Figure 107 : Structures de la *photoactive yellow protein* résolues par RMN.** Les protéines sont colorées de leur extrémité N- (bleu) à C-terminale (rouge). Le chromophore (acide p-coumarique) est en beige en représentation sphère. Les codes PDB sont indiqués.

Etant donné que les syndécans font la jonction entre la matrice extracellulaire et la membrane plasmique des cellules et qu'ils sont notamment impliqués dans des mécanisme de mécanotransduction (*Jiang et al., 2019, Effenbein & Simons, 2013*), des forces extérieures, comme le stress mécanique ou des forces de cisaillement, pourraient induire la perte de leurs structures secondaires menant à des conformations plus étendues et exposant certaines régions d'interaction avec les intégrines avec lesquelles ils interagissent (*Roper et al., 2012*).

Lors des expériences de mobilité ionique, une rupture de linéarité des valeurs de CCS est observée lors de la transition des conformations les plus compactes vers les conformations intermédiaires pour les quatre ectodomains des syndécans. Dans ce cas, deux populations ont été observées dans la distribution globale des temps de dérive suggérant la coexistence des populations compactes et des populations intermédiaires, ces-dernières gagnants graduellement en abondance. La coexistence de plusieurs populations pour un même état de charge peut également être la conséquence de deux mécanismes d'ionisation pour une même conformation générant des conformères ionisés plus ou moins étendus (*Beveridge et al., 2015*). D'après les mesures effectuées en mobilité ionique, la transition des conformations compactes qui sont quasi globulaires vers des conformations intermédiaires est favorisée pour ED-1 et ED-3 mais pas pour ED-2 et ED-4 où les conformations les plus compactes étaient très stables. Cette stabilité apparente peut être la conséquence d'une proportion environ deux fois plus importante en brins  $\beta$  dans ED-2 et ED-4 par rapport à celle retrouvée dans ED-1 et ED-3.

De précédentes expériences ont mis en évidence la capacité de l'ectodomaine du syndécan-3 à dimériser *via* des interactions électrostatiques grâce aux résidus <sup>381</sup>ERKE<sup>384</sup> présents à son extrémité C-terminale, au niveau juxtamembranaire (Asundi & Carey, 1995). Dans nos expériences, une très faible quantité de dimère a été détectée lors des expériences de spectrométrie de masse. Une espèce moléculaire est également élue à un volume d'élution supérieur à celui de la forme monomérique d'ED-3 par chromatographie d'exclusion stérique. Lors des expériences de SEC-SAXS, cette espèce était également visible sur le chromatogramme et diffusait de façon assez importante nécessitant la déconvolution des intensités de diffusion par *US-SOMO*. Bien que très préliminaire, l'analyse de ces données a montré que le dimère d'ED-3 posséderait un rayon de gyration de 84 Å et un Dmax identique à la forme monomérique de 300 Å. Il possède une conformation plus compacte comme en atteste la représentation de Kratky normalisée et celle de Sibyls (**Annexe 4**). Cette observation est cohérente avec le repliement partiel observé lors de la modélisation d'un dimère de syndécan-3 où les deux monomères s'assemblent latéralement (Kirsch *et al.*, 2002). Cet assemblage ne modifie pas significativement son volume hydrodynamique entraînant son élution à un volume proche de celui de la forme monomérique lors de son analyse par chromatographie d'exclusion stérique. L'ectodomaine du syndécan-1 possède une séquence homologue (<sup>248</sup>DRKE<sup>251</sup>) ce qui pourrait également provoquer sa dimérisation, mais aucune formation significative de dimère n'a été détectée dans nos expériences de spectrométrie de masse.

L'ensemble des résultats de nos expériences visant à caractériser la partie protéique des ectodomains des syndécans est résumé dans la **Figure 108**.



**Figure 108 : Bilan des analyses biophysiques des ectodomaines des syndécans humains.** Les EDs sont regroupés par deux (ED-1 / ED-3 et ED-2 / ED-4 monomérique) du fait de leur ressemblance. Pour chaque EDs, la valeur du Dmax et du rayon de gyration (Rg) déterminées par SEC-SAXS et du rayon hydrodynamique (Rh) déterminée par SEC-DLS est indiquée. La composition en structures secondaires (notamment en brins  $\beta$ ) et en désordre a été déterminée par dichroïsme circulaire. Les conformations représentées sont issues de l'analyse EOM reflétant les conformations étendues, intermédiaires et compactes. Les barres horizontales colorées en dégradé de rouge (ED-1 et -3) ou de bleu (ED-2 / ED-4 monomère) indiquent l'abondance relative de chaque conformation (forte : couleur intense, faible : blanc). La transition d'une conformation à l'autre est indiquée par une flèche pleine (données expérimentales) ou en pointillés (transition non étudiée). La flèche verticale rouge indique la dimérisation d'ED-4.

#### IV.1.7 Perspectives

L'ensemble des paramètres mesurés par les différentes méthodes utilisées dans ces travaux (proportion en structures secondaires, rayons de gyration et hydrodynamique, taille maximale et distribution des sections efficaces de collision, stabilité conformationnelle, flexibilité) pourrait être utilisé en tant que contraintes afin de générer un ensemble de structures tridimensionnelles par modélisation et simulations par dynamique moléculaire. Il serait également intéressant de comparer l'ensemble de ces résultats par l'utilisation de technique à haute résolution comme la résonance magnétique nucléaire, du moins pour les ectodomains les plus petits. Il serait intéressant de mimer le phénomène d'encombrement stérique retrouvé *in vivo* en présence de polymères comme le polyéthylène glycol ou le dextran afin d'étudier leur influence sur les différentes conformations des syndécans.

La formation des conformations les plus compactes ou très étendues joue certainement un rôle dans la signalisation médiée par les syndécans étant donné leur fonction majeure de corécepteurs. Il est cependant difficile d'envisager de figer de telles conformations *in vivo* afin d'étudier leur fonction. Des expériences visant à déterminer si les ectodomains des syndécans peuvent subir le phénomène de transition de phase liquide – liquide peuvent également être envisagées.

Afin de trancher sur les conformations les plus abondantes adoptées par l'ectodomaine du syndécan-3, des expériences de microscopie à force atomique pourraient être envisagées afin d'appliquer des forces d'extensions progressives sur la protéine et de potentiellement détecter des régions partiellement repliées, ou des expériences de transfert d'énergie par résonance de type Förster (FRET). Finalement, le rôle du dimère formé par l'isoforme-2 du syndécan-4 (ED-4) pourrait être élucidé par la comparaison de son répertoire d'interaction avec celui de sa forme monomérique.

Une partie des travaux portant sur la caractérisation structurale des ectodomains des syndécans a été soumise pour publication sous la forme présentée ci-après. De nouvelles modélisations de l'extrémité C-terminale d'ED-4 sous forme dimérique sont également en cours en collaboration avec le Pr. Adam Liwo (Université de Gdansk, Pologne). Ces modèles, plus pertinents, seront ajoutés au manuscrit afin d'apporter des éléments structuraux expliquant la conformation étendue du dimère d'ED-4.

## The ectodomains of the four syndecans: a disordered family with distinct conformations

Frank Gondelaud<sup>1</sup>, Mathilde Bouakil<sup>2†</sup>, Aurélien Le Fèvre<sup>3†</sup>, Adriana Miele<sup>1</sup>, Fabien Chirot<sup>3</sup>,  
Bertrand Duclos<sup>1</sup>, Sylvie Ricard-Blum<sup>1\*</sup>

<sup>1</sup>Univ Lyon, University Claude Bernard Lyon 1, CNRS, INSA Lyon, CPE, Institute of Molecular and Supramolecular Chemistry and Biochemistry (ICBMS), UMR 5246, F-69622 Villeurbanne cedex, France.

<sup>2</sup>Univ Lyon, University Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, UMR 5306, Cité Lyonnaise de l'Environnement et de l'Analyse, 5 rue de la Doua, 69100 Villeurbanne, France.

<sup>3</sup>Univ Lyon, University Claude Bernard Lyon 1, CNRS, Institut des Sciences Analytiques, UMR 5280, 5 Rue de la Doua, 69100 Villeurbanne, France.

† Both authors contributed equally to this work

\*Corresponding author

Professor Sylvie RICARD-BLUM

ICBMS, UMR 5246

University Claude Bernard Lyon 1,

43 Boulevard du 11 novembre 1918

69622 Villeurbanne Cedex, France

Phone: +33(0)4 72 44 82 32

E-mail: [sylvie.ricard-blum@univ-lyon1.fr](mailto:sylvie.ricard-blum@univ-lyon1.fr)

### ABBREVIATIONS

CD: circular dichroism, CCS: collision cross section, CSD: charge state distribution, DLS: dynamic light scattering, DTT: dithiothreitol, ED: ectodomain, ESI-MS: electrospray ionization - mass spectrometry, GAG: glycosaminoglycan, IDP: intrinsically disordered protein, ESI-IM-MS: electrospray ionization - ion mobility - mass spectrometry, MoRF: molecular recognition feature, PAGE: polyacrylamide gel electrophoresis, PMG: pre-molten globule, RC: random-coil, SASA: solvent accessible surface area, SAXS: small angle X-ray scattering, SDC: syndecan, SDS: sodium dodecyl sulfate, SEC: size exclusion chromatography, TFE: trifluoroethanol.

## ABSTRACT

Syndecans are membrane proteoglycans regulating extracellular matrix assembly, cell adhesion and signaling, involved in many diseases such as cancer. Their ectodomains can be shed from the cell surface, and act in the extracellular matrix as paracrine and autocrine effectors or as competitors of full-length syndecans. We report here the first biophysical characterization of the ectodomains of the 4 human syndecans (ED-1-4). Their size, shape, flexibility, compactness and ability to undergo disorder-to-order transition were determined using circular dichroism, dynamic light scattering, and small-angle X-ray scattering. We found that although the four ectodomains contain  $\beta$ -strands they are enriched in intrinsic disorder, and behave like random-coil intrinsically disordered proteins (IDPs). They adopt distinct conformation ensembles identified by electrospray ionization- and ion mobility-mass spectrometry, including compact conformations, which are more abundant in the ectodomain of syndecan-3. In addition, ED-4 forms head-to-tail, disulfide-bonded, dimers.

The four mammalian syndecans (SDC-1 - 4) are membrane proteoglycans, which regulate numerous biological processes including cell signaling<sup>1,2</sup>, cell adhesion, migration, and interactions with their microenvironment<sup>3</sup>, host-pathogen interactions<sup>4</sup>, tumorigenesis<sup>2</sup>, angiogenesis<sup>5</sup>, lipid metabolism<sup>6</sup>, growth factor recruitment and presentation to their receptors<sup>7</sup>, extracellular matrix assembly<sup>8,9</sup>, and tissue regeneration<sup>10</sup>. The four syndecans exert their functions *via* a large interaction network we recently drafted, which comprises 347 partners<sup>11</sup>. The syndecans are made of an ectodomain (ED), a transmembrane domain and a cytosolic domain containing a variable region flanked by two constant regions. Glycosaminoglycan (GAG) chains are covalently attached to the ectodomains of syndecans through conserved serine-glycine motifs. All syndecans contain heparan sulfate chains at their N-terminus, but SDC-1 and -3 also carry chondroitin sulfate chains in the juxtamembrane region<sup>1,2</sup>. Their shedding by various proteinases leads to the release of their soluble ectodomains in the extracellular milieu, where they can act as paracrine or autocrine effectors and competitive inhibitors of cell surface proteoglycans<sup>12,13</sup>. These ectodomains play a role in wound repair and modulate the behavior of cancer cells and their microenvironment<sup>1</sup>. Shed syndecan-2 inhibits angiogenesis<sup>14</sup>, and enhances tumorigenic activities of cancer cells<sup>15</sup> whereas shed syndecan-1 translocates to the nucleus of cells delivering growth factors and inhibiting histone acetylation<sup>16</sup>. It is also involved in chemotherapy resistance via the epidermal growth factor receptor in cancer<sup>17</sup>, and inhibits early stages of liver fibrogenesis by binding to transforming growth factor  $\beta$ 1 and interfering with its action<sup>18</sup>. Shed syndecan-4 promotes immune cell recruitment, extracellular matrix remodeling and mitigates cardiac dysfunction in response to lipopolysaccharide<sup>19</sup>. Shed syndecan-1 is elevated under inflammatory conditions, and the shedding of syndecan-4 is also associated with inflammation<sup>20</sup>.

Very few structures are available for syndecans as recently reviewed<sup>11</sup>. The structures of the transmembrane domain of SDC-2<sup>21</sup> and SDC-4 (PDB 1EJQ, unpublished data) and of the cytoplasmic domain of SDC-1 and -4 alone<sup>22,23</sup> or in complex with other proteins<sup>24,25</sup> have been determined but no structural data are available for the ectodomains of syndecans, which have been predicted to lack any ordered structure<sup>26</sup>.

We report here the extensive characterization of the recombinant ectodomains of the four human syndecans (ED 1-4) expressed in *E. coli*. The natural isoform 2 of SDC-4, lacking the transmembrane and cytosolic domains, and mimicking the proteolytic ectodomain shedding<sup>27</sup>, was used to study the ectodomain of SDC-4, and is referred to as ED-4 throughout this study. Their secondary structure, shape, size (maximum dimension, radius of gyration and hydrodynamic radius), and flexibility were determined by circular dichroism (CD), dynamic light scattering (DLS), and size exclusion chromatography - small angle X-ray scattering (SEC-SAXS). Their conformational space was determined by native mass spectrometry (ESI-MS) and ion-mobility mass spectrometry (ESI-IM-MS), which gives insights into the conformational heterogeneity of disordered proteins<sup>28</sup>.

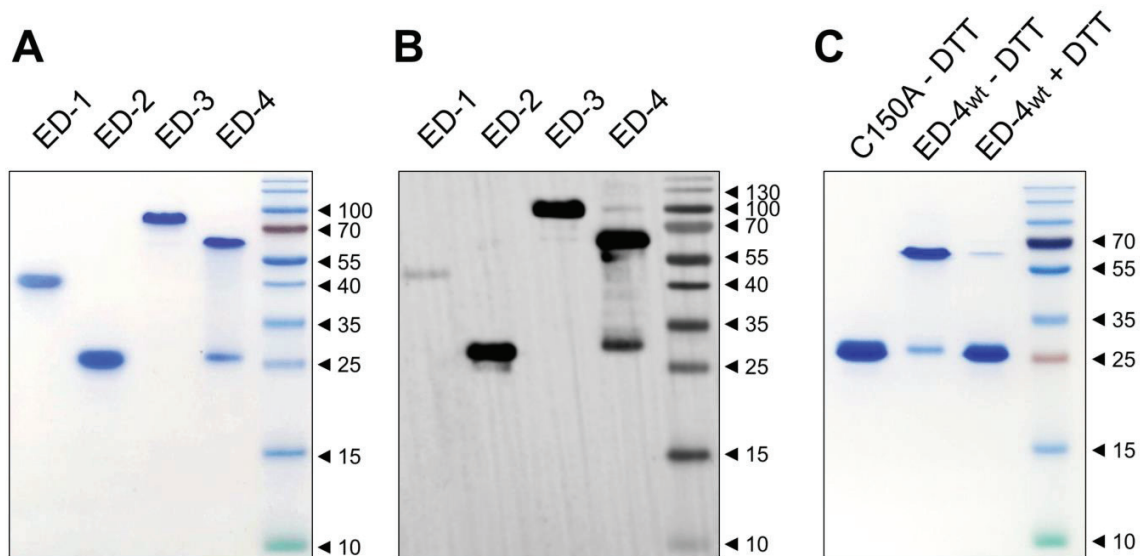
## RESULTS

### Expression and purification of the ectodomains of the four human syndecans

The ectodomains of the four human syndecans were expressed under a recombinant form in *E. coli* with a 6xHis tag at the N-terminus and a FLAG tag at the C-terminus as described in the online Methods section (**Supplementary Table S1**). The ectodomains were purified by two affinity chromatography steps using Ni-NTA agarose and anti-FLAG agarose for CD experiments. An additional step of purification by size exclusion chromatography (SEC) was carried out for SAXS, MALLS, DLS, ESI-MS and ESI-IM-MS experiments. The ED-1-4 expressed as recombinant proteins in *E. coli* do not bear their GAG chains but some syndecan partners bind not their protein cores and not to their GAG chains<sup>11</sup>, and recombinant syndecan EDs expressed in bacteria have been shown to be biologically active (e.g. by supporting cell adhesion)<sup>29</sup>.

A single band was detected by SDS-PAGE (**Figure 1A**) and Western blot (**Figure 1B**) for ED-1, ED-2 and ED-3. In contrast, two major bands migrating with an apparent molecular weight of 28 kDa and 63 kDa were observed for ED-4 (**Figure 1A and 1B**). The reduction of ED-4 by dithiothreitol (DTT) led to a marked decrease in the 63-kDa band, and an increase in the 28-kDa band suggesting that the dimerization may occur *via* disulfide bonding. The unique cysteine residue of syndecan-4 was mutated into alanine (C150A), and the mutant migrated as a single band confirming that the dimer contains a disulfide bond (**Figure 1C**). The apparent molecular weight of the syndecan EDs calculated by SDS-PAGE and size-exclusion chromatography were higher than their theoretical ones (**Supplementary Table S2**). This is consistent with the presence of intrinsic disorder predicted using metaPrDOS (**Supplementary Figure S1**). This aberrant migration has been previously observed but has been attributed to the high proline content of the ectodomains<sup>30</sup>. The molecular masses of the ectodomains determined by ESI-MS and SEC-MALLS (**Supplementary Figure S2**) were in agreement with the theoretical ones, with ED-1, ED-2 and ED-3 being monomers, and ED-4 existing as both a monomer and a dimer (**Supplementary Table S2**). It is known that full-length syndecans dimerize via a conserved GxxxG motif located in their transmembrane domains leading to the formation of SDS-resistant homo- and hetero-oligomers<sup>31</sup>, but this is the first report that the ectodomain of syndecan-4 is able to covalently dimerize.





**Figure 1: SDS-PAGE and Western blot analysis of the recombinant ectodomains of human syndecans.** 2  $\mu$ g of proteins purified by affinity chromatography and gel filtration were analyzed in non-reducing conditions by SDS-PAGE and stained by Coomassie blue (A), and by Western blot with an anti-FLAG antibody and chemiluminescence detection (B). The wild-type ectodomain of syndecan-4 ( $\pm$  2 mM DTT), and the C150A mutant were analyzed by SDS-PAGE (C).

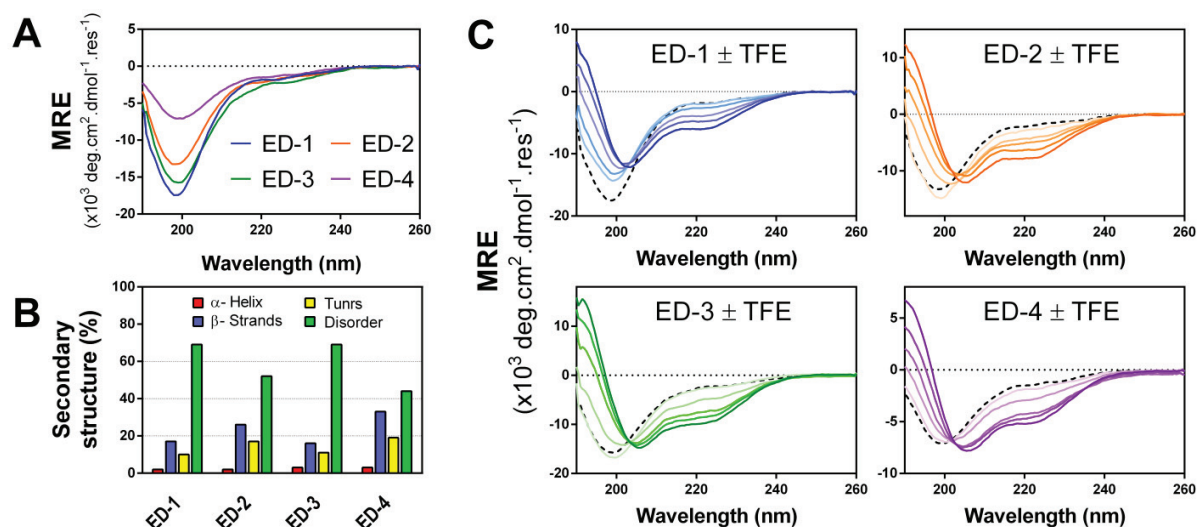
### The ectodomains of syndecans contain $\beta$ -strands and are enriched in intrinsic disorder

The ectodomains were analyzed by circular dichroism to determine their amount of secondary structure and intrinsic disorder by the deconvolution of the spectra as described in the online Methods section (Figure 2A). ED-1 and ED-3 were the most disordered (69 % of random coil for both), whereas ED-2 and ED-4 contained 52 % and 44 % of random coil respectively (Figure 2B). The content in  $\alpha$ -helix of the four ectodomains was very low (<3 %), but they contained a significant amount of  $\beta$ -strands, which were found in higher amount in ED-2 and ED-4 (26 % and 33 % respectively) than in ED-1 and ED-3 (16 % and 17 % respectively) (Figure 2B). According to the secondary structure prediction using PSIPRED 4.0,  $\beta$ -strands were located in the C-terminal half of ED-1 and ED-4, and scattered along the sequence of ED-2 and ED-3 (Supplementary Figure S1A).

To determine if the ectodomains were able to undergo a disorder-to-order transition, the Molecular Recognition Features (MoRFs), short disordered sequences able to undergo this transition upon binding to a partner, were identified computationally as described in the online Methods section. Two major MoRFs were predicted at the N- and C-termini of ED-1 and ED-3 and a short one in their central part (Supplementary Figure S1). In contrast, almost the entire sequence of ED-2 was predicted to contain MoRFs, except at the N- and C-termini, and three major MoRFs were predicted in ED-4 (Supplementary Figure S1). To check experimentally if the ectodomains were able to undergo a disorder-to-order transition, CD experiments were performed in presence of increased concentrations of 2,2,2-trifluoroethanol (TFE) (Figure 2C).

The increase in  $\alpha$ -helix, in ED-1 and ED-3 (+15 and +31 % respectively according to the deconvolution of CD spectra) in presence of 60 % TFE was associated with a decrease in disorder (-22% and -29 % respectively), suggesting that both ED-1 and ED-3 might undergo a disorder-to-order transition. In contrast, the increase in  $\alpha$ -helix in ED-2 and ED-4 (+21% and +12 % respectively) was associated with a decrease in  $\beta$ -strands (-9% and -11 % respectively) and a low or negligible decrease in disorder (-6% and -2 %, respectively). The shape of the curve suggests that the formation of  $\alpha$ -helix in syndecan-3 could be a cooperative process (Supplementary Figure S3).

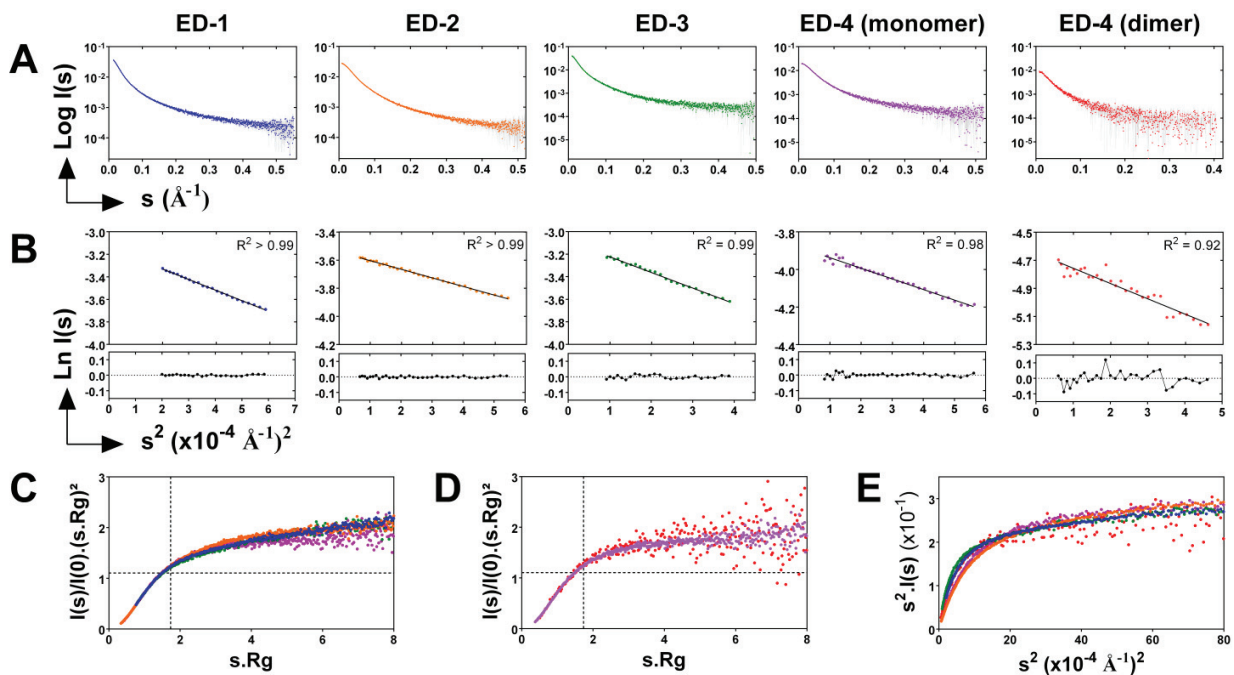
CD spectra of the ectodomains were collected in presence of 3 M and 5 M guanidinium chloride, which destabilizes folded proteins but stabilizes polyproline II helix<sup>32</sup>. A small increase in signal in the 220-230 nm region was observed for the four ectodomains in presence of guanidinium chloride suggesting that they were not able to form a significant amount of polyproline II helix (Supplementary Figure S1B). CD spectra were also collected in presence of two chemical chaperones in order to determine if they were able to stabilize or favor any conformation of the ectodomains. The stabilizing osmolyte chaperone trimethylamine N-oxide, which shifts the folding–unfolding equilibrium toward the folded state, induced changes in the CD signal, which could be related to a decrease in disorder. The hydrophobic chaperone 4-phenylbutyric acid, which prevents aggregation, did not induce significant changes (Supplementary Figure S1C).



**Figure 2: Circular dichroism experiments.** A) CD spectra of the syndecan ectodomains (MRE: mean residue molar ellipticity), B) Percentage of secondary structures and random coil calculated from the deconvolution of CD spectra using Dichroweb as described in the online Methods. C) CD spectra of the syndecan ectodomains recorded in presence of increased concentrations of trifluoroethanol (TFE) from 0 % (black dotted line) to 60 % *v/v* (the darkest color).

### Size and shape of the syndecan ectodomains

The four syndecan ectodomains were analyzed by small-angle X-ray scattering (SEC-SAXS) (**Figure 3** and **Supplementary Figure S4**). ED-2 and the monomeric form of ED-4 had a radius of gyration ( $R_g$ ) of  $42.9 \pm 0.33 \text{ \AA}$  and  $41.53 \pm 0.67 \text{ \AA}$  respectively, and a maximum intramolecular distance  $D_{max}$  of  $185 \text{ \AA}$  and  $180 \text{ \AA}$ , respectively (**Table 1** and **Supplementary Figure S5**). ED-3 had the largest  $R_g$  ( $65.0 \pm 0.62 \text{ \AA}$ ) and  $D_{max}$  ( $300 \text{ \AA}$ ) whereas the  $R_g$  of ED-1 was  $52.7 \pm 1.43 \text{ \AA}$ , and its  $D_{max}$  was  $230 \text{ \AA}$ . The  $R_g$  and  $D_{max}$  of the ED-4 dimer were far larger than those of the ED-4 monomer (**Table 1** and **Supplementary Figure S5**). The  $D_{max}$  values of the syndecan ectodomains ranged from  $180$  to  $300 \text{ \AA}$ , indicating that they are extended proteins. The shape of the normalized Kratky-plots of the four EDs (**Figure 3C**) were consistent with those of intrinsically disordered proteins<sup>33</sup>, and the dimer of ED-4 was disordered too (**Figure 3D**). The flexibility of the ectodomains was then assessed by the calculation of the Porod exponents and by the Kratky-Debye plots. The Porod exponents of syndecan EDs were similar, being  $1.7$  for ED-1, ED-3 and the monomer of ED-4,  $1.8$  for ED-2, and  $1.9$  for the dimer of ED-4 indicating that the ectodomains and the dimer of ED-4 are flexible since the Porod exponent has a value of  $2$  for fully flexible systems<sup>34</sup>. This was reflected in the plateau observed in the Kratky-Debye plots (**Figure 3E**). The syndecan ectodomains are thus disordered, flexible, and extended proteins.



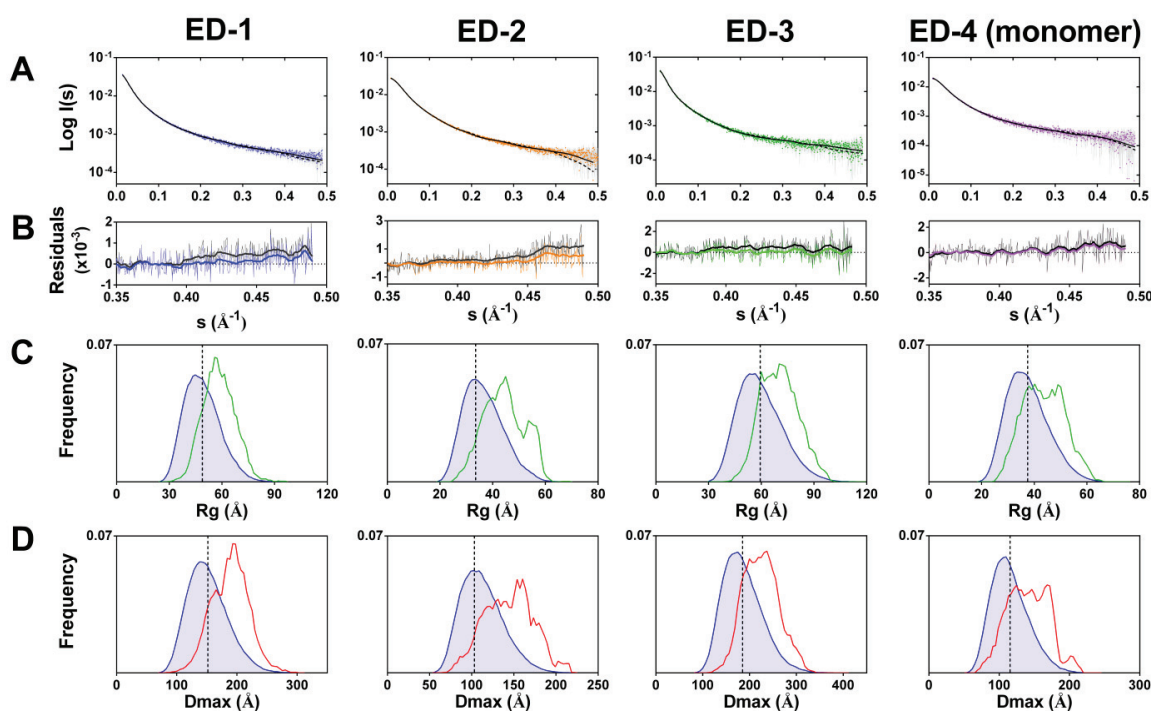
**Figure 3: SEC-SAXS analysis of the syndecan ectodomains.** **A)** SEC-SAXS scattering intensities after signal deconvolution and treatment. **B)** Determination of the radius of gyration according to the Guinier approximation at low angle with residual plots and coefficient of determination ( $R^2$ ). Normalized Kratky plots of **C)** the four syndecan ectodomains (EDs), and **D)** the monomeric and dimeric forms of the ectodomain of syndecan-4, **E)** Scaled Kratky-Debye plots of the syndecan ectodomains (blue: ED-1, orange: ED-2, green: ED-3, purple: ED-4 monomer, red: ED-4 dimer).

Then we determined if the ectodomains behaved like random coil (RC), or pre-molten globule (PMG) disordered proteins<sup>35</sup> using several approaches. The theoretical Rg values of the ectodomains were calculated using the Flory's equation for a globular protein and for PMG and RC protein models containing the same number of amino acid residues than the ectodomains, and compared to the experimental values. Experimental Rg values of ED-1-4 were higher than the theoretical values calculated for both the globular protein and the PMG protein model but close to those calculated for the RC protein model (**Table 1**). The hydrodynamic radius (Rh) of the ectodomains, determined by size-exclusion chromatography-dynamic light scattering (SEC-DLS), ranged from 21.6 Å (ED-4) to 50.1 Å (ED-3) (**Table 1**). The Rh of the ED-4 dimer was two-fold larger than that of the monomer. The Rg/Rh ratio of the ectodomains, that provides information about a protein shape and compactness, ranged from 1.19 (ED-1) to 1.92 (ED-4 monomer). These values were far higher than the value reported for globular proteins (0.778) but close to those reported for prolate ellipsoids (1.36 - 2.24)<sup>36</sup> (**Table 1**)

	ED-1	ED-2	ED-3	ED-4 (monomer)	ED-4 (dimer)
<b>SAXS</b>					
<b>Rg (Å)</b> Guinier approximation Mean ± error)	53.30 ± 0.3	42.90 ± 0.33	65.00 ± 0.62	41.53 ± 0.67	59.25 ± 2.27
<b>Dmax (Å)</b> (Pair function distribution)	230	185	300	180	240
<b>Theoretical Rg calculated using Flory's equation</b>					
<b>Folded protein</b>	18.5	15.9	20.5	16.1	19.7
<b>Pre Molten Globule</b> (Mean ± error)	46.3 ± 0.6	35.2 ± 0.4	55.9 ± 0.8	36.3 ± 0.5	52.1 ± 0.7
<b>Random coil</b> (Mean ± error)	54.2 ± 11.8	39.6 ± 7.8	67.4 ± 15.7	40.9 ± 8.2	62.2 ± 14.1
<b>Hydrodynamic properties (SEC-DLS)</b>					
<b>Rh (Å)</b> (Mean ± error)	44.9 ± 0.8	34.1 ± 0.7	50.1 ± 1.3	21.6 ± 4.0	45.4 ± 1.2
<b>Shape factor Rg/Rh</b>	1.19	1.26	1.30	1.92	1.31

**Table 1: Size, shape and compactness of the syndecan ectodomains.** The radii of gyration were calculated experimentally from SEC-SAXS data by Guinier approximation and theoretically using Flory's equation (see online Methods section). Their hydrodynamic radii (Rh) were determined by SEC-DLS, and the shape factors (Rg/Rh) were calculated using experimental Rg values.

SAXS scattering intensities were analyzed using the Ensemble Optimization Method (EOM) in order to confirm their behavior as random coil IDPs, and to explore their conformational landscape. Two different pools of conformers covering the available conformational space of PMG and RC protein models were generated. Both pools fitted SAXS data in the low and median angles regions but the PMG model failed to correctly fit SAXS data at high angle ( $> 4 \text{ \AA}$ ) values (**Figure 4A, 4B** and **Table 2**). The random coil model was thus used for further analysis, and for the calculation of the  $R_{flex}$  and  $R_g$  parameters. The values of  $R_{flex}$  for the selected ensembles of the four ectodomains (81 - 86 %) were very close to those of the pools of random coil conformers (84 - 86 %) (**Table 2**) confirming that the four ectodomains behave as random coil IDPs in solution. The distributions of  $R_g$  and  $D_{max}$  of the selected ensembles fitting experimental SAXS data were shifted to higher values than those of the random coil ensemble. The distributions of  $R_g$  and  $D_{max}$  of the selected ensembles were roughly centered on  $R_g \sim 56 \text{ \AA}$  and  $D_{max} \sim 193 \text{ \AA}$  for ED-1 and on  $R_g \sim 70 \text{ \AA}$  and  $D_{max} \sim 225 \text{ \AA}$  for ED-3. The multimodal distribution of both  $R_g$  and  $D_{max}$  for ED-2 and ED-4 might indicate the existence of several subpopulations of conformations for these two ectodomains (**Figure 4C** and **4D**). Selected models of these conformation ensembles of ED1-4 are presented in **Supplementary Figure S6**.



**Figure 4: Ensemble Optimization Method analysis of SAXS data.** **A)** Random-coil (RC, black solid line) and pre-molten globule (PMG) models (black dotted line) fitting to SAXS experimental data (colored curves). **B)** Residual plot from PMG (black) and RC models (colored) fits. Smoothed curves are represented as thick lines, with the same color legend as in A. **C-D)** The  $R_g$  and  $D_{max}$  distributions from EOM analysis: the blue line shows the distribution of the initial pool of random coil conformers, the green and red lines represent the  $R_g$  and  $D_{max}$  distributions of the selected ensembles of conformers that fit the SAXS data. The dotted vertical black line represents the average  $R_g$  and  $D_{max}$  of the random coil initial pool.

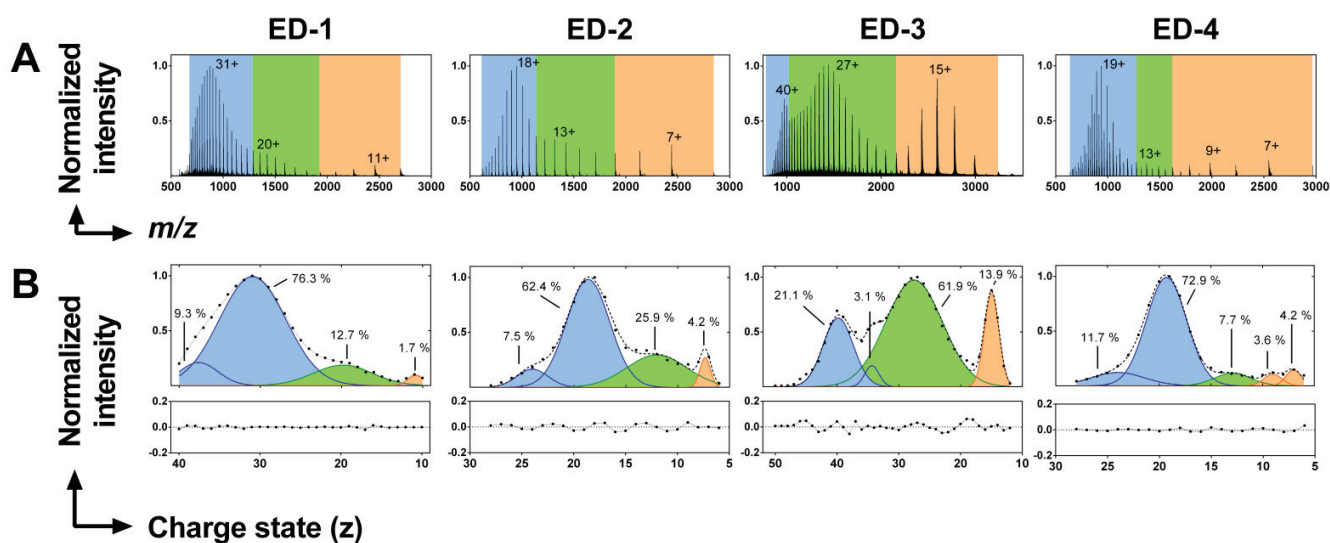
EOM analysis	ED-1	ED-2	ED-3	ED-4 (monomer)
<b>Pre-molten globule (PMG) model</b>				
Reduced $\chi^2$	0.904	0.947	0.765	0.636
CorMap p-value	0.0002	0.000	0.008	0.004
<b>Random-coil (RC) model</b>				
Reduced $\chi^2$	0.616	0.632	0.663	0.597
CorMap p-value	0.224	0.227	0.0645	0.0618
Rflex ensemble (%)	82.71	86.17	81.18	84.02
Rflex pool (%)	84.31	85.60	84.59	83.69
R $\sigma$	0.92	2.47	0.85	1.01

**Table 2: Ensemble Optimization Method (EOM) analysis of SAXS data.** Fitting assessment between the experimental and theoretical scattering intensities generated by EOM for the pre-molten globule and the random coil ensembles. The CorMap test is a goodness-of-fit test<sup>51</sup>.

### The conformational space of the syndecan ectodomains

Electrospray ionization-mass spectrometry (ESI-MS) in native conditions provides information on the variable degrees of compactness of IDPs. The charge state distributions (CSDs) determined by ESI-MS are related to the global compactness of proteins upon their transfer from solution to gas phase, the lowest charge states corresponding to the most compact conformations<sup>37</sup>. The CSDs of the four ectodomains were trimodal, showing that the ectodomains of syndecans could adopt preferential conformations (**Figure 5A**). Gaussian functions were fitted to the envelopes of the CSDs to determine the relative proportions of the three major conformation ensembles. The most compact conformations (*i.e.* the lowest charge states) were the minor ones for all the ectodomains. They represented 2 % for ED-1, 4% for ED-2, 8% for ED-4, and 14 % for ED-3. The most extended conformations (*i.e.* the highest charge states) were the most abundant for ED-1 (86 %), ED-2 (70 %) and ED-4 (85 %) but accounted for only 24 % for ED-3, in which the population with intermediate charge states predominated (**Figure 5B**). ED-1, -2 and -4 primarily adopted fully extended conformations in gaseous phase, whereas ED-3 existed as partially extended conformations with a low but significant contribution of more compact states.

The average solvent-accessible surface area (SASA) of the ectodomains and the SASA of their most extended and most compact conformations were calculated (**Supplementary Table S3**). The average SASA values of ED-2 and of the monomer of ED-4 were in the same range (16074 Å<sup>2</sup> and 17780 Å<sup>2</sup> respectively), as were those of ED-1 and ED-3 (29 653 Å<sup>2</sup> and 28624 Å<sup>2</sup> respectively) although the sequence of ED-3 is 30 % longer than ED-1. The average SASA values being related to the extent of protein ionization, this is consistent with the fact that ED-3 might adopt more compact conformations than ED-1 as mentioned above. However, the SASA values of the extended conformations of the ectodomains were directly proportional to their size, ranging from 18269 Å and 18844 Å for ED-2 and the monomer of ED-4, respectively, to 31104 Å and 42071 Å for ED-1 and ED-3, respectively.



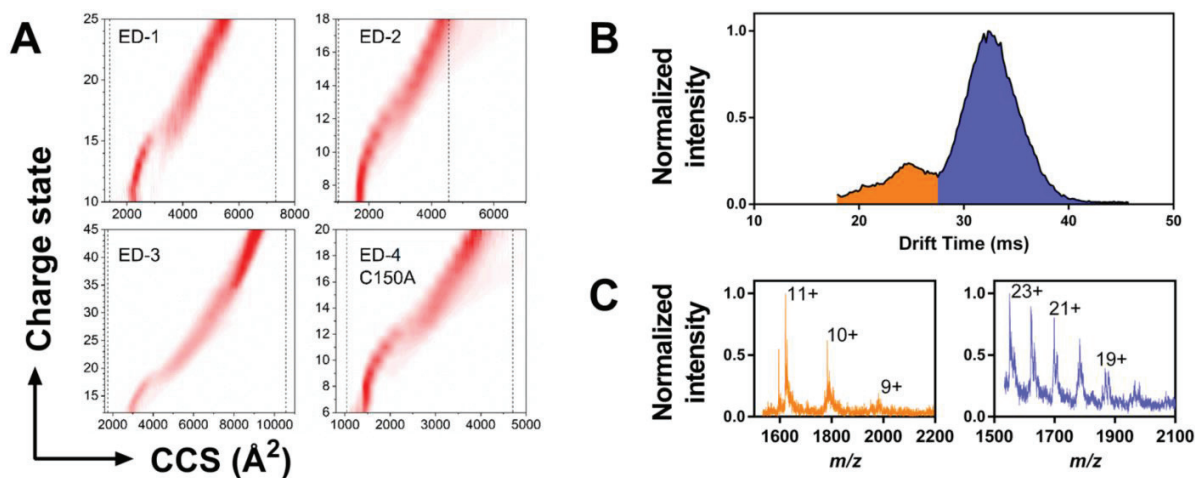
**Figure 5: Analysis of the syndecan ectodomains by native ESI-MS. A)** Charge state distribution (CSD) of the syndecan ectodomains (EDs). Blue: extended conformations, green: intermediate conformations, orange: more compact conformations, **B)** Deconvolution of the CSDs by Gaussian fitting, and their respective residual plots.

The average charges of the most compact conformations of the ectodomains were very close ( $\pm 1$  proton) to that reported for globular proteins with a similar number of amino acid residues<sup>38</sup>. They were thus calculated using this globular protein data set<sup>38</sup>. The SASA of the most compact conformation of ED-1 (260 residues) was 12913 Å<sup>2</sup> compared to 10 997 Å<sup>2</sup> for carbonic anhydrase I (PDB: 2CAB, 260 residues), 7204 Å<sup>2</sup> for ED-2 (154 residues) compared to 8 017 Å<sup>2</sup> for myoglobin (PDB: 1WLA, 153 residues), 19178 Å<sup>2</sup> for ED-3 (373 residues) compared to 15 586 Å<sup>2</sup> for ovalbumin (PDB: 1OVA, 386 residues), and 7832 Å<sup>2</sup> for ED-4 (163 residues) compared to 8 321 Å<sup>2</sup> for  $\beta$ -lactoglobulin A (PDB: 1BSY, 162 residues). The SASA values of the most compact conformations of ED1-4 were in the same range that those calculated for globular proteins with a similar number of amino acid residues, suggesting that the syndecan ectodomains are able to adopt compact conformations resembling those of globular proteins.

Ion-mobility mass spectrometry (ESI-IM-MS) experiments were performed to get further insights on the conformational heterogeneity of the syndecan ectodomains. The C150A mutant of ED-4, unable to dimerize, was used as the monomeric form of ED-4. The dimer of ED-4 was studied by selecting charge states with odd values of wild-type ED-4. The collision cross section (CCS) distributions of the different ectodomains increased as a function of charge state as expected due to Coulomb repulsion (**Figure 6A**). The boundaries of the accessible CCS space of the ectodomains were roughly estimated using a simple model based on the volume of the protein amino acids arranged in a spherical (the most compact) or cylindrical (the most extended) configuration<sup>39</sup>. ED-1 explored roughly the half of its accessible CCS space (54 %) whereas ED-2, -3 and monomeric ED-4 (C150A) explored most 73 %, 70 % and 68 %, of their CCS space respectively) (**Figure 6A**).

This, in combination with the broad range of charge states detected, is in agreement with the fact that all ectodomains are highly flexible, and exist in elongated conformations. The predominance of high charge states with relatively large CCS values suggests these elongated structures represent a majority of the sampled conformational space although low charge states with low CCS values (**Figure 6A**) indicates the existence of a compact conformational ensemble.

A transition from a compact conformational ensemble to a more extended one occurred in all ectodomains as the charge state  $z$  increased, yielding a bimodal CCS distribution. Beveridge *et al.*<sup>40</sup> proposed that such transition may be related to the fact that compact and elongated conformations follow distinct desolvation mechanisms in electrospray ionization process. The transition between the compact and extended states should occur when the radius of an ion of charge state  $z$  reaches the maximum size of a solvent droplet bearing a charge  $z$ , the so-called Rayleigh limit. This transition occurred at  $z = 8-13$  for ED-2 and ED-4, at  $z = 11-15$  for ED-1, and at  $z = 16-18$  for ED-3. The charge state where the transition in the CCS distributions arose for each ED corresponded to 80 to 85 % of the corresponding Rayleigh limit estimated according to<sup>41</sup>. No transition was observed in the CCS distribution of the dimer of ED-4. These results are consistent with the existence of two different conformational ensembles of the syndecan ectodomains, each following a single but different desolvation mechanism, and a transitional state where both desolvation mechanisms occur<sup>40</sup>.



**Figure 6: Ion mobility-mass spectrometry (ESI-IM-MS) experiments.** A) Normalized CCS distributions of the syndecans ectodomains (EDs) as a function of charge state. The dashed lines indicate the limits of the CCS space as estimated in each case from the model of Beveridge *et al.*<sup>39</sup>. B) Global distribution of ED-4, containing both the monomeric (orange) and dimeric (blue) forms. Their respective extracted charge state distributions are indicated in (C).



Two peaks can be distinguished in the global ESI-IM-MS arrival time distribution recorded for ED-4 (**Figure 6B**). The mass spectrum extracted for the first peak at a drift time of 25 ms displays a charge state distribution consistent with 9-11 positive charges and the mass of ED-4 monomer. Conversely, the mass spectrum extracted for the second at a drift time of 32 ms rather corresponds to 18-23 positive charges, in agreement with the mass of ED-4 dimer (**Figure 6C**). The drift time of a molecular ion being proportional to the ratio of its CCS to its charge  $z$ , and the drift time of the dimer being higher than that of the monomer with a half charge, this indicates that the CCS of the ED-4 dimer is at least twice that of the monomer. This is consistent with a head-to-tail association of two ED-4 monomers retaining a relatively extended conformation. This is supported by the values of the hydrodynamic radius for the monomer and the dimer of ED-4 (21 Å and 45 Å respectively), and of their radius of gyration (41.5 Å and 59.2 Å respectively).

## DISCUSSION

We report here the first characterization of size, shape, secondary structure, and conformational landscape of the ectodomains of the four human syndecans, that play major roles as paracrine or autocrine effectors or as competitors of full-length syndecans and are involved in diseases such as angiogenesis, cancer and fibrosis. We have shown that they are enriched in intrinsic disorder, providing experimental support to *in silico* predictions<sup>26</sup>. The juxtamembrane region of the four syndecan ectodomains are predicted to be highly disordered, which is consistent with this region being highly sensitive to proteolytic cleavage, including the ED shedding sites<sup>11,13</sup>. Further cleavage sites by metalloproteinases exist along the sequence of ED-1, ED-2, and ED-4<sup>11</sup>. The NxIP motif, required for cell adhesion, and found in the sequence of human ED-2 and ED-4, is also intrinsically disordered. The ectodomains contain from ~15% (ED-1 and ED-3) to ~30 % (ED2 and ED-4) of  $\beta$ -strands, but few, if any  $\alpha$  helices. However, they are able to form  $\alpha$  helices in presence of trifluoroethanol. This is in agreement with the partial folding into  $\alpha$ -helix of the ectodomain of syndecan-4 during molecular dynamic simulations carried out at the glycocalyx level<sup>42</sup>.

Full-length syndecans form homo-and heterodimers<sup>31</sup> but our study is the first one to report that the isoform-2 of syndecan-4 (ED-4) dimerizes via a disulfide bridge mediated by a unique cysteine residue located at its C-terminus. Indeed, the mutation of this cysteine residue into alanine (C150A) nearly abolishes the dimerization process. The dimer of ED-4 remains fully disordered and flexible as shown by SAXS and MS experiments. Ion-mobility MS data suggests that ED-4 dimerizes in a head-to-tail fashion, which is consistent with the values of the radius of gyration and of the hydrodynamic radius calculated for the dimer. No dimer of ED-1 and ED-2 has been detected, and minute amount of a dimer of ED-3 has been identified during ESI-MS experiments, which is consistent with the self-association of syndecan 3 mediated by a motif in the transmembrane domain and the ectodomain flanking region<sup>43</sup>. It should be noted that the ectodomains 1, 2 and 3 do not contain any cysteine residue.

SEC-SAXS data show that the syndecan ectodomains are highly flexible and extended in solution, and behave like random coils rather than pre-molten globule IDPs. The four ectodomains display three major ensembles of conformations with various degrees of compactness in the gas phase as shown by ESI-MS and ESI-IM-MS. ED-1 explores a reduced part of its conformational space (54 %) compared to the three other ectodomains (~70%), suggesting that it might adopt a restricted number of conformations. The most compact conformations are the less abundant for the four EDs. The most extended conformations represent 70%-85% of the conformations for ED-2, ED-1 and the monomeric form of ED-4. In contrast, ED-3 exists mostly as partially extended conformations with a low but significant contribution of more compact states, although it contains the same amount of disorder but is longer than ED-1. ED-3 is able to form more compact conformations than the other ectodomains as shown by its SASA and the intensity of low and intermediate charge states by ESI-MS. In contrast to the other EDs, ED-3 contains a "mucin-like" region of ED-3 (amino acid residues 115-302), enriched in threonine, serine, and proline residues, and putatively carrying O-linked oligosaccharides. This domain is predicted to be intrinsically disordered as expected according to its amino acid composition. However, it domain contains several short hydrophobic sequences, which might favor compact conformations of ED-3 in combination with additional hydrophobic sequences located at the N-terminus of ED-3.

The maximum intramolecular distance measured by SEC-SAXS for the ectodomains varies from ~ 180 Å (ED-4 and ED-2) to 300 Å (ED-3) with an intermediate value for ED-1 (230 Å). The ectodomains of syndecans 1, 2, and 4 have a size similar to extended, ligand-bound integrins such as integrin  $\alpha$ IIb $\beta$ 3, which extends to 18-20 nm in length upon binding to fibrin for example<sup>44</sup>. The conformations of the ectodomains may be modulated upon binding to their 347 syndecan partners<sup>11</sup>. Some of their partners interact directly with the syndecan ectodomains, and their docking to the core protein ectodomain might stabilize particular conformations of the ectodomains. This will be investigated by studying complexes formed by the syndecan EDs and their ligands. On the other hand, the disorder of syndecan EDs might lead to the formation of fuzzy complexes with their partners. This may be the case for the highly disordered synstatin sequences of human ED-1, namely residues 93-120, which form a ternary complex with insulin-like growth factor-1 receptor and  $\alpha$ v $\beta$ 3 integrin<sup>29</sup>, and residues 210-240, which form a ternary complex with vascular endothelial growth factor receptor 2 and  $\alpha$ 4 $\beta$ 1 integrin<sup>45</sup>. Mechanical forces could also regulate the conformations of the syndecan ectodomains because conditional disorder may be initiated by such forces<sup>35</sup>, and because the syndecans, which link the extracellular matrix to the actin cytoskeleton, act as mechanosensors as reported for syndecan 4 in cardiac fibroblasts<sup>46</sup>, and for syndecans 1 and 4 in the endothelial glycocalyx<sup>42,47</sup>. The glycosaminoglycan chains covalently attached to the syndecan ectodomains can also modulate their conformational space, and this regulation process might be fine-tuned by the number, length or number/type of sulfate groups which vary depending on the cell type, and the physiopathological context. The mutual impact of syndecan-1 protein core and its GAG chains is difficult to analyze<sup>48</sup>. Ongoing experiments with the EDs expressed in eukaryotic cells and bearing their GAG chains will provide new insights on the influence of GAG chains, which are themselves flexible, on the conformations and flexibility of syndecan EDs.

The ultimate goal will be to determine how ligand binding to the protein core and/or the GAG chains of the syndecan ectodomains influence dimerization of the transmembrane domain, and cell signaling. Another crucial challenge will be to determine if the conformational ensembles of the syndecan ectodomains we have identified by SAXS and MS are all biologically active whatever the biological context, if only some of them are able to fulfill biological functions, or if the different conformational ensembles are associated with specific functions in a particular context. The study of cell-matrix interactions, and cell signaling in presence of the chemical chaperone trimethylamine N-oxide, which induces a decrease in disorder in the syndecan ectodomains, should provide new insights into the biological role of syndecan conformations.

## ON LINE METHODS

### Expression of the recombinant ectodomains of human syndecans

The constructs for human ectodomains of syndecans (ED-1 to 3), the isoform 2 of syndecan-4 (ED-4), and the ED-4 C150A mutant were synthesized and subcloned in the pET-15b vector by GenScript (<https://www.genscript.com>). They were then expressed as recombinant proteins containing an N-terminal 6xHis tag and a C-terminal FLAG tag in *Escherichia coli* BL21 (DE3) cells (**Supplementary Table S1**) and purified by two affinity chromatography steps followed by size exclusion chromatography for some experiments as detailed below.

*Escherichia coli* BL21 (DE3) cells were transformed with 80 ng of plasmid, and colonies were grown at 37°C on ampicillin-containing LB-agar plates. Cells were suspended in LB broth Lennox (Sigma-Aldrich, L3022) with ampicillin 100 µg/ml until the absorbance at 600 nm reached 0.6 - 0.8. The expression of the EDs was induced by 1 mM isopropyl-β-D-thiogalactopyranoside for 4 h at 37 °C. After centrifugation at 6 000 x g for 10 min, 4°C, the pellets were resuspended in 10 mM HEPES, 150 mM NaCl pH 7.4 (HEPES buffered saline - HBS) containing 2.5 mM MgCl<sub>2</sub>, 2.5 mM CaCl<sub>2</sub>, 100 µg/ml DNase I (Sigma, D5025), 1 mM CHAPS (Roche), and a protease inhibitors cocktail (Complete EDTA-free, Roche) and lysed for 8 x 30 s with 100 µm-diameter glass beads in a bead beater (BioSpec). The lysates were then centrifuged twice at 10 000 x g for 10 min and the supernatant was loaded on Ni-NTA agarose resin (Qiagen). The purification steps were carried out at 4°C with a flow rate of 20 ml/h. After washing with HBS containing 10 mM and 20 mM imidazole, the recombinant syndecan EDs were eluted with 100 and 300 mM imidazole in HBS. Eluates were loaded on anti-FLAG M2 agarose resin (Sigma-Aldrich, A2220) and the ectodomains were eluted by the FLAG peptide at 200 µg/ml in HBS. The purified ectodomains were then concentrated with Amicon devices (MWCO 10 kDa). For SEC-SAXS, SEC-MALLS, SEC-DLS, ESI-MS and ESI-IM-MS experiments, the ectodomains were further purified by size exclusion chromatography on a Superdex S200 Increase 10/300 GL column (GE Healthcare) at 10°C, with a flow rate of 0.5 ml/min. The eluted fractions were concentrated in Amicon devices (MWCO 10 kDa), centrifuged 20 min at 18 000 x g and stored at -20°C until use.

### **SDS-PAGE and Western blot**

The purified ectodomains were analyzed by SDS-PAGE on a 15 % acrylamide gel and stained overnight with InstantBlue (ISB1L, Sigma-Aldrich). For Western blotting, proteins were transferred on a polyvinylidene fluoride membrane (IPVH20200, Millipore) for 2 h at 250 mA and the membrane was saturated with phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 11.9 mM phosphate, Biosolve) containing 5 % (w/v) fat milk and 0.05 % (v/v) Tween-20 for 90 min at 37°C. The membrane was incubated overnight at 4°C with a monoclonal anti-FLAG antibody (F3165, Sigma-Aldrich, 76 ng/ml in blocking buffer), washed 4 times with the blocking solution and incubated with a secondary antibody conjugated to horseradish peroxidase (40 ng/ml, 172-1011, Bio-Rad) for 1 h at room temperature. Finally, the membrane was washed twice in PBS containing 0.05 % (v/v) Tween-20, and twice in PBS alone. The immunoreactive bands were detected by chemiluminescence with the Amersham ECL Prime kit (GE Healthcare), and the signal recorded with a ChemiDoc XRS+ system (Biorad) after 5 to 60 s of exposure.

### **Circular dichroism (CD)**

Far-UV CD spectra were recorded in a quartz cuvette at 20°C with a path length of 0.1 cm in a Chirascan spectrometer (Applied Photophysics, SFR BioSciences UMS3444/US8, Protein Science Facility, Lyon, France), from 260 to 190 nm with a step of 0.5 nm, a response time of 1 s and a bandwidth of 0.5 nm. The purified ectodomains were diluted in 10 mM potassium phosphate pH 7.5 at 150 µg/ml for ED-1 and ED-4 and 100 µg/ml for ED-2 and ED-3. Five spectra were collected and averaged for each sample and the buffer signal was subtracted. CD signal was recorded in millidegrees (mdeg) and converted to mean residue molar ellipticity per residue ( $\text{deg.cm}^2.\text{dmol}^{-1}.\text{residue}^{-1}$ ). Some experiments were carried out in presence of increasing concentrations (10, 20, 30, 45 and 60 % (v/v)) of trifluoroethanol (TFE). Spectra were deconvolved using the Dichroweb server (<http://dichroweb.cryst.bbk.ac.uk>), and the CDSSTR algorithm with the set 7 as a reference set containing five chemically denatured proteins<sup>49</sup>.

### **Size Exclusion Chromatography - Small Angle X-ray Scattering (SEC-SAXS)**

Scattering intensities were collected at SOLEIL (SWING beamline, Gif-sur-Yvette, France) with the scattering intensity  $I$  as a function of the modulus of the scattering vector  $s$  with  $s = 4\pi \cdot \sin(\theta)/\lambda$  where  $\theta$  is the half of the scattering angle and  $\lambda$  the X-ray wavelength (**Table 3**). The calibration was made with water and the protein samples were centrifuged at 18 000 x g for 20 min at 4°C before injection on the size exclusion chromatography column. The concentrations of the ectodomains were determined by the measurement of absorbance at 280 nm using the theoretical molar extinction coefficient calculated by ProtParam (<https://web.expasy.org/protparam/protpar-ref.html>). Data reduction and frames subtraction were done manually with the beamline software FOXTROT. Buffer blank frames were taken in the dead volume of the column.

<b>Data acquisition</b>				
Instrument	SWING beamline (SOLEIL, France)			
Detector	Dectris EIGER 4M			
Wavelength (Å)	1.03324			
Energy (keV)	12			
Sample to detector distance (m)	2.0			
s-range (Å <sup>-1</sup> )	0.0041 – 0.5597			
<b>SEC-SAXS experiments</b>				
	ED-1	ED-2	ED-3	ED-4
Column	Superdex S200 Increase 5/150 GL (GE Healthcare)			
Injection volume (µl)	50			
Protein concentration (mg.ml <sup>-1</sup> )	7.7	5.0	14.6	8.3
Flow rate (ml.min <sup>-1</sup> )	0.2			
Temperature (°C)	20			
Exposure	Continuous: 1 frame/s (990 ms exposure and 10 ms dead time)			
Buffer	HBS (10 mM HEPES, 150 mM NaCl pH 7.4)			
s <sub>min</sub> (Å <sup>-1</sup> )	0.0114020	0.00821002	0.00961379	0.00912193
I(0) (cm <sup>-1</sup> )	0.0430 ± 1.8 x10 <sup>-4</sup>	0.02892 ± 0.94 x10 <sup>-4</sup>	0.04588 ± 3.24 x10 <sup>-4</sup>	0.02070 ± 1.42 x10 <sup>-4</sup>
Guinier analysis s <sub>max</sub> .Rg / number of points	1.29 / 23	0.998 / 34	1.28 / 24	0.985 / 33

**Table 3:** Experimental setup for the collection of SEC-SAXS scattering intensities.

Gaussian decompositions were made with the UltraScan solution modeler (US-SOMO) HPLC-SAXS module<sup>50</sup> after conversion from buffer-subtracted  $I(s)$  vs  $s$  chromatograms to  $I(t)$  vs  $t$  (t, time). Scattering curves corresponding to the syndecan EDs were then back-converted to  $I(s)$  vs  $s$  curves. These frames (at least 29) were averaged with PRIMUSqt to create the final scattering data file after curves comparison with the CorMap test implemented in the program<sup>51</sup>. Data points differing from the Guinier fitting at low angles were deleted and the useful data range was determined with SHANUM<sup>51</sup>. The radius of gyration was estimated at low angles according to the Guinier approximation using the equation  $I(s) = I(0) \cdot \exp(-s^2 \cdot Rg^2/3)$  with  $s \cdot Rg < 1.0$  or 1.3 depending of the number of points.

The theoretical radius of gyration (Rg) was calculated with the Flory's equation  $Rg = R_0 N^\nu$  where N is the number of amino acid residues,  $R_0$  a constant and  $\nu$  an exponential scaling factor.  $R_0 = 2.54 \pm 0.01$  and  $\nu = 0.522 \pm 0.01$  for intrinsically disordered proteins,  $R_0 = 1.927 \pm 0.27$  and  $\nu = 0.598 \pm 0.028$  for more extended proteins such as chemically denatured proteins<sup>52</sup>, while  $R_0 = \sqrt{(3/5)} \times 4.75$  and  $\nu = 0.29$  for folded proteins<sup>53</sup>. The flexibility of the proteins was assessed with the dimensionless Kratky plot  $(s \cdot Rg)^2 \cdot I(s)/I(0)$  vs  $s \cdot Rg$  and the Kratky-Debye plot  $s^2 I(s)$  vs  $s^2$ .

### Ensemble Optimization Method (EOM)

Intrinsically disordered proteins do not adopt a single conformation in solution and each of the conformations contribute to the collected scattering intensities. This ensemble of conformation is described according to the following equation:

$$I(s) = \sum_{k=1}^K v_k I_k(s)$$

where  $v_k$  and  $I_k(s)$  are the volume fraction and the scattering intensity of the  $k$ -th conformers respectively<sup>54</sup>. SEC-SAXS data were processed with the Ensemble Optimization Method (EOM 2.0)<sup>55</sup>, which computes a large pool of possible conformations from protein sequence in order to cover its conformational space. Gaussian distributions of  $R_g$  and  $D_{max}$  of the generated conformers were automatically calculated as well as their respective theoretical scattering profile using *CRY SOL*<sup>51</sup>. A genetic algorithm was run to select an ensemble of conformations which best fits the experimental scattering intensities. The best ensemble of conformations was described as a distribution of their radii of gyration ( $R_g$ ) and their maximum particle size ( $D_{max}$ ). The flexibility was evaluated by  $R_{flex}$  (100 % corresponding to the highest flexibility), with the pool representing a reference for flexibility<sup>55</sup>.  $R\sigma$  corresponds to the ratio of the standard deviations for the distributions of the selected ensemble and of the pool respectively<sup>55</sup>. Two pools of 10,000 conformers corresponding to the pre-molten globule and random-coil models were generated with specific dihedral angle distribution<sup>55</sup>. The number of harmonics was set to 15 and 100 cycles were performed on the generated pool by the genetic algorithm. EOM fitting to the experimental scattering intensities was assessed with the CorMap test implemented in PRIMUSqt<sup>51</sup>.

### Multi Angle Laser Light Scattering and Dynamic Light Scattering

SEC-MALLS-DLS experiments were carried out at SOLEIL (SWING beamline, Gif-sur-Yvette, France) with a Wyatt Dawn Heleos light scattering instrument equipped with an Optilab rEX refractometer and a spectrophotometer for molecular mass measurements. The hydrodynamic radii ( $R_h$ ) were measured with a Wyatt QELS+. For ED-1 and ED-2, the measurements were carried out on line after X-ray exposure and injection on a Superdex S200 Increase 5/150 GL column (GE Healthcare) as described above. 50  $\mu$ l of ED-3 (6.5 mg/ml) and ED-4 (5.2 mg/ml) were injected on a Superdex S200 Increase 10/300 GL column (GE Healthcare) at a flow rate of 0.5 ml/min and at 20°C in HBS. Data analysis was carried out with the Astra software and default parameters (refractive index: 1.331, viscosity: 0.992 cP).

### Native electrospray ionization mass spectrometry (ESI-MS)

ESI-MS experiments were carried out in pseudo-native conditions, *i.e.* by minimizing potential sources of denaturation in solution and during the desolvation/ionization process<sup>56</sup>. A Superdex S200 Increase 10/300 GL column (GE Healthcare) was used to exchange proteins buffer from HBS to 50 mM ammonium acetate pH 7.4 and then the ectodomain solutions were concentrated as described above to 600, 590, 420 and 180  $\mu$ g/ml for ED-1, ED-2, ED-3 and ED-4, respectively. 200  $\mu$ l of syndecans EDs in 50 mM ammonium acetate pH 7.4 were directly injected at a flow rate of 10  $\mu$ l/min in an Impact II mass spectrometer (Bruker) equipped with an Apollo II ionization source (Bruker, ICBMS UMR5246, Villeurbanne, France).

The instrument parameters were set as follows: ESI-source temperature 200°C, capillary voltage 3 kV, nebulization gas pressure 41 Psi allowing the detection of species between 500 and 3,100 Da, and acquisition frequency 2 Hz. The average solvent-accessible surface areas (SASA) were calculated according to Li *et al.*<sup>57</sup>. Briefly, the average charge  $z_{av}$  was calculated according to:

$$z_{av} = \sum_{i=A}^B z_i I_i / I_{tot}$$

where A and B are the lowest and highest charge states of the distribution respectively.  $I$  and  $I_{tot}$  are the peak intensity at a given charge state, and the sum of the intensities of the charge state distribution of the conformers respectively.  $z_{av}$  is directly proportional to SASA for extended conformers according to:  $\ln(z_{av}) = 0.9024 \times \ln(SASA, \text{Å}^2) - 5.9013$ <sup>58</sup>. SASA from globular proteins were extracted from<sup>38</sup>, and calculated for the syndecan EDs according to  $\ln(z_{av}) = 0.6712 \times \ln(SASA, \text{Å}^2) - 3.8952$ . ESI-MS data were deconvolved as described in<sup>37</sup> using Origin software (OriginLab).  $I(m/z)$  vs  $m/z$  graphs were transformed into  $I(z)$  vs  $z$  and intensities were normalized. Gaussian fitting was performed until the fit reached at least  $R^2 = 0.99$ .

### Electrospray ionization - ion mobility - mass spectrometry (ESI-IM-MS)

In ion mobility - mass spectrometry, electrosprayed ions are driven by a weak electric field through a cell filled with a buffer gas. Ions with higher charge state generally fly faster. Moreover, the velocity of the ions is affected by the strength of their interaction with the buffer gas, depending of their collision cross section (CCS). Since helium was used as a buffer gas, this CCS is, in good approximation, a measure of the geometrical area of an ion. Ions with a compact structure are less prone to collision with the gas, and then travel faster through the cell. Ion populations can then be separated as a function of the compactness of their structure in an arrival time distribution. The arrival time distributions were converted into CCS distributions to ease the comparison between different charge states. Due to Coulomb repulsion between charges, CCS generally increases with charge state. Ion mobility-mass spectrometry experiments were performed using a home-built dual-drift tube instrument previously described in details<sup>59</sup>. IM-MS separation was achieved by injecting ion bunches in a 79-cm long drift tube filled with 4 Torr helium at room temperature. The drift voltage across the tube varied from 200 to 600 V. Mass spectra were recorded as a function of the arrival time of the ions using a time-of-flight mass spectrometer (Maxis Impact Bruker, Bremen, Germany). This allowed the extraction of either arrival time distributions (ATDs) for ions at the mass-to-charge ratios ( $m/z$ ) of interest, or of mass spectra for species within a selected arrival time range. CCS values were derived from drift time measurements at different drift voltages, based on the Mason-Schamp relation<sup>60</sup>.

## ACKNOWLEDGEMENTS

This work was supported by a grant from the Fondation pour la Recherche Médicale n°DBI20141231336 to SRB, and by an ANR grant n°17-CE29-0013-01 to ALF and FC. SEC-SAXS experiments were carried out on the BM29 BioSAXS beamline (ESRF, Grenoble, France, MX-1841 and MX-1920), and SWING beamline (SOLEIL, Université Paris-Saclay, France, project 20170906). We thank Martha Brennich (BM29, ESRF) and Aurélien Thureau (SWING, Soleil) for their skillful assistance and for their kind and expert suggestions on SAXS data processing, and Cy Jeffries (EMBL, Hamburg, Germany) for sharing with us his expertise in SAXS theory. We are also grateful to Alexandra Berlioz-Barbier (ICBMS, UMR 5246, Villeurbanne, France) for her help in native ESI-MS experiments. We thank John Couchman (University of Copenhagen, Denmark) for sharing with us his expertise of syndecans, and for stimulating discussions. We thank Lionel Ballut (UMR 5086, Lyon, France), Philippe Dugourd and Luke MacAleese (Institut Lumière Matière, UMR 5306, Villeurbanne, France) for their critical reading of the manuscript.

## AUTHOR CONTRIBUTION

FG expressed, purified and analyzed the ectodomains of syndecans by SDS-PAGE and Western blot, performed, analyzed and interpreted CD, SEC-SAXS, and SEC-MALLS-DLS experiments, analyzed and interpreted native ESI-MS experiments and drafted the manuscript and the figures. MB and ALF performed ESI-IM-MS experiments, interpreted the results, and drafted the corresponding sections of the manuscript. AEM contributed to the first SAXS experiments. FC interpreted ESI-IM-MS data, and wrote the corresponding sections of the manuscript. FG and BD designed the wild type and mutant syndecan constructs and set up the experimental conditions of their expression in bacteria. SRB designed and supervised the study, analyzed and interpreted the data, and wrote the manuscript. All the authors revised the manuscript.

**Supplementary information** available as a .pdf file.

## REFERENCES

1. Couchman, J. R., Gopal, S., Lim, H. C., Nørgaard, S. & Multhaupt, H. A. B. Fell-Muir Lecture: Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. *Int J Exp Pathol* **96**, 1–10 (2015).
2. Afratis, N. A. *et al.* Syndecans - key regulators of cell signaling and biological functions. *FEBS J.* **284**, 27–41 (2017).
3. Mitsou, I., Multhaupt, H. A. B. & Couchman, J. R. Proteoglycans, ion channels and cell-matrix adhesion. *Biochem. J.* **474**, 1965–1979 (2017).
4. Bartlett, A. H. & Park, P. W. Proteoglycans in host-pathogen interactions: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med* **12**, e5 (2010).
5. De Rossi, G. & Whiteford, J. R. Syndecans in angiogenesis and endothelial cell biology. *Biochem. Soc. Trans.* **42**, 1643–1646 (2014).
6. Leonova, E. I. & Galzitskaya, O. V. Role of Syndecans in Lipid Metabolism and Human Diseases. *Adv. Exp. Med. Biol.* **855**, 241–258 (2015).
7. Kwon, M.-J., Jang, B., Yi, J. Y., Han, I.-O. & Oh, E. S. Syndecans play dual roles as cell adhesion receptors and docking receptors. *FEBS Lett.* **586**, 2207–2211 (2012).
8. Klass, C. M., Couchman, J. R. & Woods, A. Control of extracellular matrix assembly by syndecan-2 proteoglycan. *J. Cell. Sci.* **113** ( Pt 3), 493–506 (2000).



9. Yang, N. & Friedl, A. Syndecan-1-Induced ECM Fiber Alignment Requires Integrin  $\alpha\beta 3$  and Syndecan-1 Ectodomain and Heparan Sulfate Chains. *PLoS ONE* **11**, e0150132 (2016).
10. Chung, H., Mulhaupt, H. A. B., Oh, E.-S. & Couchman, J. R. Minireview: Syndecans and their crucial roles during tissue regeneration. *FEBS Lett.* **590**, 2408–2417 (2016).
11. Gondelaud, F. & Ricard-Blum, S. Structures and interactions of syndecans. *FEBS J.* **286**, 2994–3007 (2019).
12. Asundi, V. K., Erdman, R., Stahl, R. C. & Carey, D. J. Matrix metalloproteinase-dependent shedding of syndecan-3, a transmembrane heparan sulfate proteoglycan, in Schwann cells. *J. Neurosci. Res.* **73**, 593–602 (2003).
13. Manon-Jensen, T., Itoh, Y. & Couchman, J. R. Proteoglycans in health and disease: the multiple roles of syndecan shedding. *FEBS J.* **277**, 3876–3889 (2010).
14. De Rossi, G. *et al.* Shed syndecan-2 inhibits angiogenesis. *J. Cell. Sci.* **127**, 4788–4799 (2014).
15. Choi, S. *et al.* Shed syndecan-2 enhances tumorigenic activities of colon cancer cells. *Oncotarget* **6**, 3874–3886 (2015).
16. Stewart, M. D., Ramani, V. C. & Sanderson, R. D. Shed syndecan-1 translocates to the nucleus of cells delivering growth factors and inhibiting histone acetylation: a novel mechanism of tumor-host cross-talk. *J. Biol. Chem.* **290**, 941–949 (2015).
17. Wang, X. *et al.* Shed Syndecan-1 is involved in chemotherapy resistance via the EGFR pathway in colorectal cancer. *Br. J. Cancer* **111**, 1965–1976 (2014).
18. Regős, E. *et al.* Syndecan-1 inhibits early stages of liver fibrogenesis by interfering with TGF $\beta$ 1 action and upregulating MMP14. *Matrix Biol.* **68–69**, 474–489 (2018).
19. Strand, M. E. *et al.* Shedding of syndecan-4 promotes immune cell recruitment and mitigates cardiac dysfunction after lipopolysaccharide challenge in mice. *J. Mol. Cell. Cardiol.* **88**, 133–144 (2015).
20. Gopal, S. Syndecans in Inflammation at a Glance. *Front Immunol* **11**, 227 (2020).
21. Li, Q., Ng, H. Q. & Kang, C. Secondary structure and topology of the transmembrane domain of Syndecan-2 in detergent micelles. *FEBS Lett.* **593**, 554–561 (2019).
22. Lee, D., Oh, E. S., Woods, A., Couchman, J. R. & Lee, W. Solution structure of a syndecan-4 cytoplasmic domain and its interaction with phosphatidylinositol 4,5-bisphosphate. *J. Biol. Chem.* **273**, 13022–13029 (1998).
23. Shin, J. *et al.* Solution structure of the dimeric cytoplasmic domain of syndecan-4. *Biochemistry* **40**, 8471–8478 (2001).
24. Liu, X., Shepherd, T. R., Murray, A. M., Xu, Z. & Fuentes, E. J. The structure of the Tiam1 PDZ domain/ phospho-syndecan1 complex reveals a ligand conformation that modulates protein dynamics. *Structure* **21**, 342–354 (2013).
25. Kang, B. S., Cooper, D. R., Devedjiev, Y., Derewenda, U. & Derewenda, Z. S. Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. *Structure* **11**, 845–853 (2003).
26. Leonova, E. I. & Galzitskaya, O. V. Cell communication using intrinsically disordered proteins: what can syndecans say? *J. Biomol. Struct. Dyn.* **33**, 1037–1050 (2015).
27. Xing, Y., Xu, Q. & Lee, C. Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS Lett.* **555**, 572–578 (2003).
28. Stuchfield, D. & Barran, P. Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Curr Opin Chem Biol* **42**, 177–185 (2018).
29. Rapraeger, A. C. Synstatin: a selective inhibitor of the syndecan-1-coupled IGF1R- $\alpha\beta 3$  integrin complex in tumorigenesis and angiogenesis. *FEBS J.* **280**, 2207–2215 (2013).
30. Park, P. W. Isolation and functional analysis of syndecans. *Methods Cell Biol.* **143**, 317–333 (2018).
31. Choi, Y. *et al.* Trans-regulation of Syndecan Functions by Hetero-oligomerization. *J. Biol. Chem.* **290**, 16943–16953 (2015).
32. Chemes, L. B., Alonso, L. G., Noval, M. G. & de Prat-Gay, G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol. Biol.* **895**, 387–404 (2012).
33. Burger, V. M., Arenas, D. J. & Stultz, C. M. A Structure-free Method for Quantifying Conformational Flexibility in proteins. *Sci Rep* **6**, 29040 (2016).
34. Brosey, C. A. & Tainer, J. A. Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Curr. Opin. Struct. Biol.* **58**, 197–213 (2019).
35. Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* **7**, (2019).
36. Brewer, A. K. & Striegel, A. M. Characterizing the size, shape, and compactness of a polydisperse prolate ellipsoidal particle via quadruple-detector hydrodynamic chromatography. *Analyst* **136**, 515–519 (2011).

37. Santambrogio, C., Natalello, A., Brocca, S., Ponzini, E. & Grandori, R. Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. *Proteomics* **19**, e1800060 (2019).
38. Kaltashov, I. A. & Mohimen, A. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal. Chem.* **77**, 5370–5379 (2005).
39. Beveridge, R. *et al.* A mass-spectrometry-based framework to define the extent of disorder in proteins. *Anal. Chem.* **86**, 10979–10991 (2014).
40. Beveridge, R. *et al.* Relating gas phase to solution conformations: Lessons from disordered proteins. *Proteomics* **15**, 2872–2883 (2015).
41. Mora, null *et al.* Electrochemical processes in electrospray ionization mass spectrometry. *J Mass Spectrom* **35**, 939–952 (2000).
42. Jiang, X. Z., Luo, K. H. & Ventikos, Y. Principal mode of Syndecan-4 mechanotransduction for the endothelial glycocalyx is a scissor-like dimer motion. *Acta Physiol (Oxf)* **228**, e13376 (2020).
43. Asundi, V. K. & Carey, D. J. Self-association of N-syndecan (syndecan-3) core protein is mediated by a novel structural motif in the transmembrane domain and ectodomain flanking region. *J. Biol. Chem.* **270**, 26404–26410 (1995).
44. Dai, A. *et al.* The Structure of a Full-length Membrane-embedded Integrin Bound to a Physiological Ligand. *J. Biol. Chem.* **290**, 27168–27175 (2015).
45. Jung, O. *et al.* Heparanase-induced shedding of syndecan-1/CD138 in myeloma and endothelial cells activates VEGFR2 and an invasive phenotype: prevention by novel synstatins. *Oncogenesis* **5**, e202 (2016).
46. Samarel, A. M. Syndecan-4: a component of the mechanosensory apparatus of cardiac fibroblasts. *J. Mol. Cell. Cardiol.* **56**, 19–21 (2013).
47. Voyvodic, P. L. *et al.* Loss of syndecan-1 induces a pro-inflammatory phenotype in endothelial cells with a dysregulated response to atheroprotective flow. *J. Biol. Chem.* **289**, 9547–9559 (2014).
48. Eriksson, A. S. & Spillmann, D. The mutual impact of syndecan-1 and its glycosaminoglycan chains—a multivariable puzzle. *J. Histochem. Cytochem.* **60**, 936–942 (2012).
49. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392–400 (2008).
50. Brookes, E., Vachette, P., Rocco, M. & Pérez, J. US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *J Appl Crystallogr* **49**, 1827–1841 (2016).
51. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212–1225 (2017).
52. Bernadó, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst* **8**, 151–167 (2012).
53. Receveur-Brechot, V. & Durand, D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13**, 55–75 (2012).
54. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).
55. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).
56. Leney, A. C. & Heck, A. J. R. Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2017).
57. Li, J. *et al.* Conformational effects in protein electrospray-ionization mass spectrometry. *Mass Spectrom Rev* **35**, 111–122 (2016).
58. Testa, L. *et al.* Electrospray ionization-mass spectrometry conformational analysis of isolated domains of an intrinsically disordered protein. *Biotechnol J* **6**, 96–100 (2011).
59. Simon, A.-L. *et al.* Tandem ion mobility spectrometry coupled to laser excitation. *Rev Sci Instrum* **86**, 094101 (2015).
60. Revercomb, H. E. & Mason, E. A. Theory of plasma chromatography/gaseous electrophoresis. Review. *Anal. Chem.* **47**, 970–983 (1975).

**SUPPLEMENTARY TABLES AND FIGURES**

<b>Recombinant protein</b>	<b>UniProt KB Accession number</b>	<b>Residues (UniProt KB)</b>	<b>Number of amino acid residues including tags</b>
<b>ED-1</b>	P18827	23 - 254	260
<b>ED-2</b>	P34741	19 - 144	154
<b>ED-3</b>	O75056	47 - 387	373
<b>ED-4</b>	P31431-2	19 - 153	163
<b>ED-4 C150A</b>	-	19 - 153	163

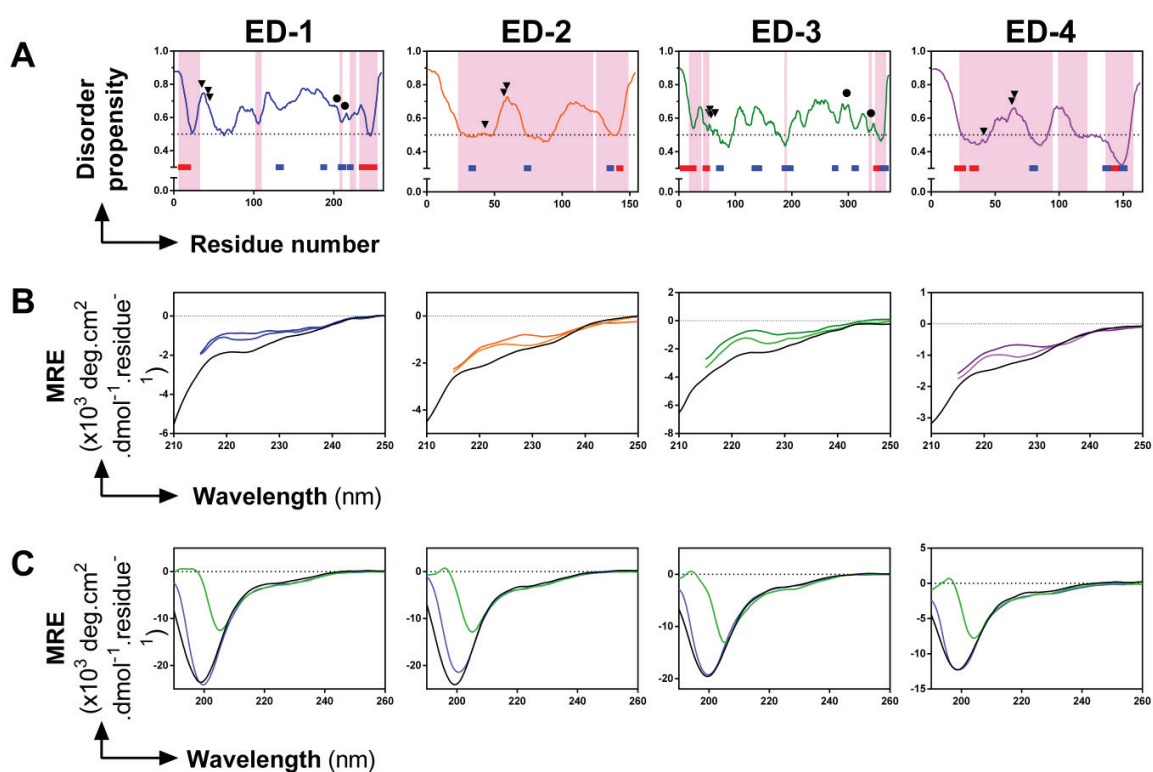
**Supplementary Table S1:** Characteristics of the recombinant ectodomains (EDs) of human syndecans expressed with a 6xHis tag at the N-terminus, and a FLAG tag at the C-terminus.

<b>Syndecans ectodomains</b>	<b>Molecular weight of tagged recombinant EDs (Da)</b>				
	<b>Theoretical</b>	<b>SDS-PAGE</b>	<b>SEC</b>	<b>ESI-MS</b>	<b>SEC-MALLS</b>
<b>ED-1</b>	27 030	42 300	172 000	27 029.4	27 480 ± 60
<b>ED-2</b>	17 083	27 700	72 000	17 083.1	16 640 ± 50
<b>ED-3</b>	38 831	89 000	389 000	38 836.0	38 800 ± 500
<b>ED-4</b>	17 828	27 500	78 000	17 827.4	18 000 ± 60
		63 300	195 000	35 654.1	32 620 ± 30
<b>ED-4 C150A</b>	17 790	27 300	91 000	17 795.6	17 592 ± 50

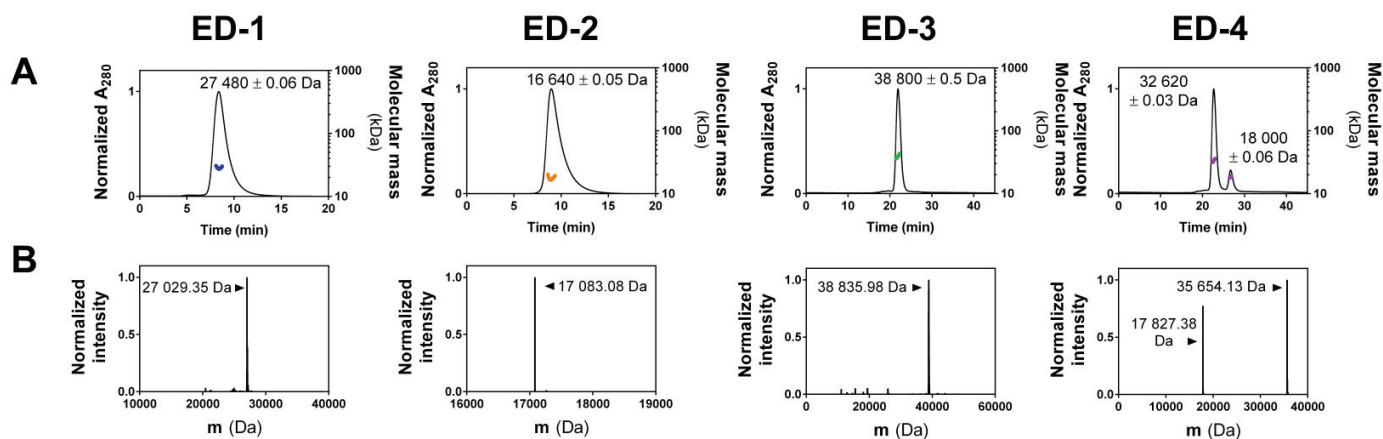
**Supplementary Table S2:** Apparent molecular weights of the recombinant tagged ectodomains of human syndecans calculated by SDS-PAGE and size-exclusion chromatography using globular proteins as standards, and molecular masses determined by ESI-MS and SEC-MALLS. The proteins were expressed with a N-terminal 6xHis tag and a C-terminal FLAG tag.

<b>Conformations</b>	<b>Solvent-accessible surface area (SASA) (<math>\text{\AA}^2</math>)</b>			
	<b>ED-1</b>	<b>ED-2</b>	<b>ED-3</b>	<b>ED-4</b>
<b>Average</b>	29 653.0	16 074.2	28 624.4	17 780.2
<b>Extended</b>	31 103.9	18 268.6	42 071.3	18 843.5
<b>Compact</b>	12 912.8	7 090.0	19 178.1	7 832.4

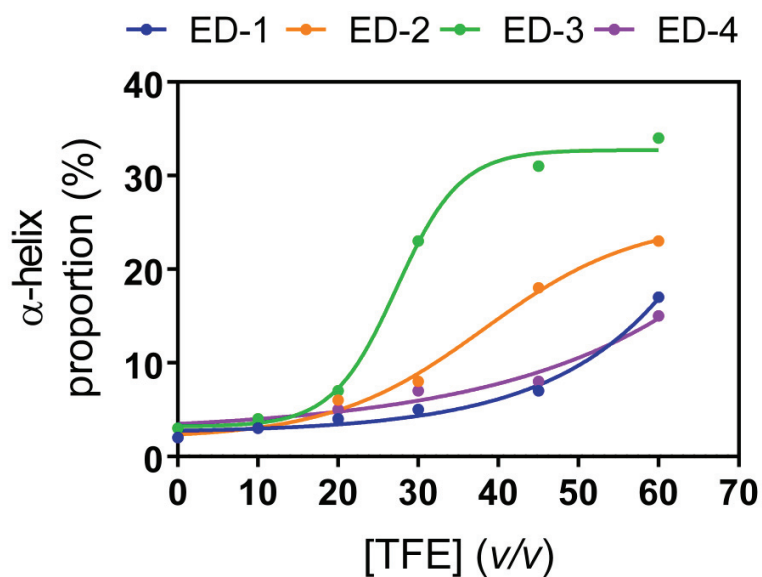
**Supplementary Table S3:** Solvent-accessible surface area (SASA) of the syndecan ectodomains calculated from native ESI-MS experiments. The average SASA was calculated using the entire charge state distribution whereas the extended and compact SASAs were calculated for the most and less charged envelopes, respectively. ED-4 SASAs refer to its monomeric form.



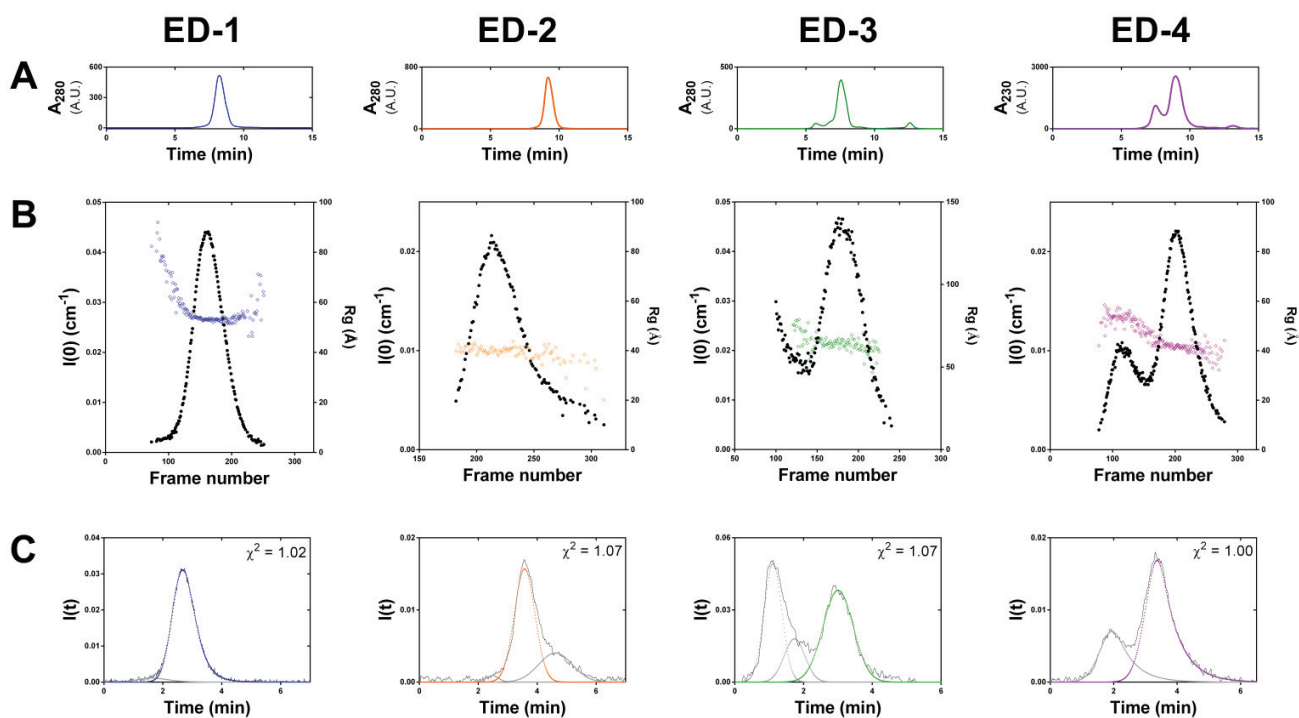
**Supplementary Figure S1. Intrinsic disorder, molecular recognition features (MoRFs) and secondary structures of the four ectodomains of human syndecans.** **A)** The disorder propensity of the tagged recombinant ectodomains was predicted with the metaPrDOS predictor (Ishida & Kinoshita, 2008, *Bioinformatics* 24:1344-8) (false positive rate: 5%, propensity cut-off: 0.5). The disordered residues are displayed as a curve, and the disorder cut-off (0.5) is visualized by black dotted line. Molecular recognition features, represented as pink areas, were predicted with MoRFchibi SYSTEM based on the propensity of amino acid residues to undergo a disorder-to-order transition upon binding to a globular protein (Malhis et al., 2016 *Nucleic Acids Res* 44: W488-93). The secondary structure was predicted using PSIPRED 4.0 with default parameters (Buchan and Jones, 2019 *Nucleic Acids Res.* 47: W402-W407). Red and blue rectangles: predicted  $\alpha$ -helix and  $\beta$ -sheets respectively. Heparan sulfate and chondroitin sulfate chains attachment sites on the ectodomains are represented as black triangles and circles respectively. The amino acid residues are numbered using the sequence of the recombinant syndecan ectodomains, including the 6 Histidine tag at the N-terminus, and the FLAG tag at the C-terminus. **B)** Circular dichroism (CD) spectra of the syndecan ectodomains in absence (black line) and in presence of 3 and 5 M (the darkest color) guanidinium chloride. The spectra were recorded between 260 and 215 nm in presence of guanidinium chloride because of the high absorbance of chloride ions. **C)** CD spectra of the syndecan ectodomains without (black spectra) or with 0.5 M trimethylamine N-oxide (green spectra) or 100  $\mu\text{M}$  4-phenylbutyric acid (blue spectra) in 10 mM potassium phosphate pH 7.4. The deconvolution of CD spectra using Dichroweb (Whitmore and Wallace, 2008 *Biopolymers* 89: 392-400) was performed as described in the online Methods section.



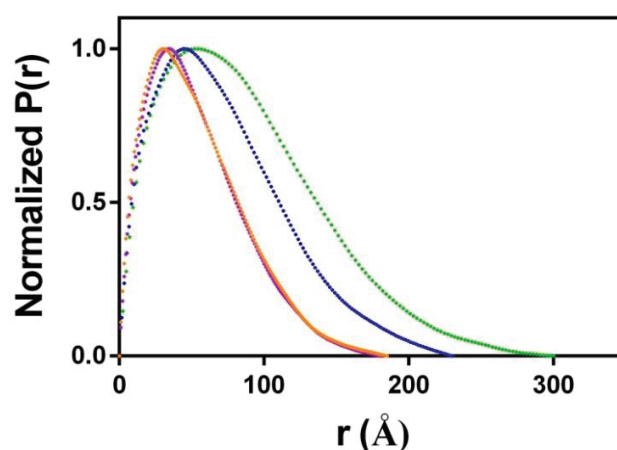
**Supplementary Figure S2: Molecular mass determination of the syndecan ectodomains by SEC-MALLS and ESI-MS.** **A)** Analysis of the four ectodomains of syndecans (EDs) by SEC-MALLS. UV spectra (black lines) and molecular mass estimation (colored curves) are represented. **B)** ESI-MS spectra of the four syndecan ectodomains. Measured masses (Da) are indicated for both techniques.



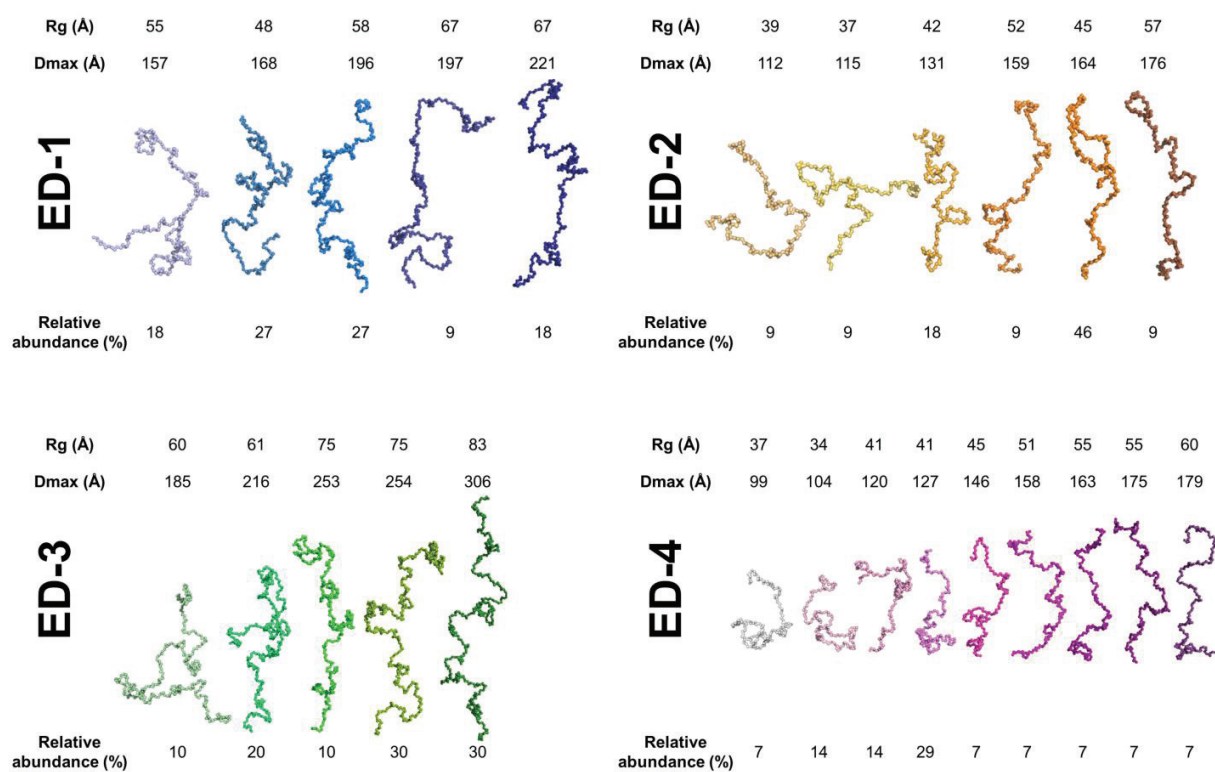
**Supplementary Figure S3: Formation of  $\alpha$ -helix in the ectodomains of syndecans (ED) in presence of increasing concentrations of trifluoroethanol (TFE).** The respective proportions of  $\alpha$ -helix were extracted from the deconvolution of CD spectra as detailed in the online Methods section. A sigmoidal fitting was performed for each dataset (goodness-of-fit ( $R^2$ )  $\geq 0.99$  for ED-1, ED-2 and ED-3 and 0.97 for ED-4).



**Supplementary Figure S4: SEC-SAXS data processing.** **A)** SEC elution profiles of the four ectodomains of syndecans (EDs). **B)**  $I(0)/R_g$  as a function of the recorded frames. Black dots:  $I(0)$  intensities, colored diamonds: estimation of the radius of gyration. **C)** Deconvolution of the scattering intensities using the US-SOMO HPLC SAXS module (Brookes *et al.*, 2016, *J Appl Crystallogr.* 49:1827-41). The colored Gaussian curves correspond to the monomeric form of the ectodomains. Molecular species of ED-4 eluted at around 2 min correspond to the dimeric form of this ectodomain.



**Supplementary Figure S5: Pair-distance distribution function  $P(r)$  calculated from SAXS intensities using GNOM implemented in the ATSAS package.** Curves were normalized and adjusted until obtaining a smooth decay to zero. The quality of  $P(r)$  was assessed with the CorMap test. ED-1: blue, ED2: orange, ED-3: green, ED-4: purple. The curves obtained for ED-2 and ED-4 are similar, and appear to be superimposed.



**Supplementary Figure S6: Models composing the best selected ensemble from EOM deconvolution using random-coil parameters.** The radius of gyration (Rg), the maximum particle size (Dmax) and the relative abundance (%) for each conformer ensemble are indicated for the 4 syndecan ectodomains (EDs).





## **IV.2**

# **CARACTÉRISATION STRUCTURALE DES ECTODOMAINES DES SYNDÉCANS AVEC GAGs**

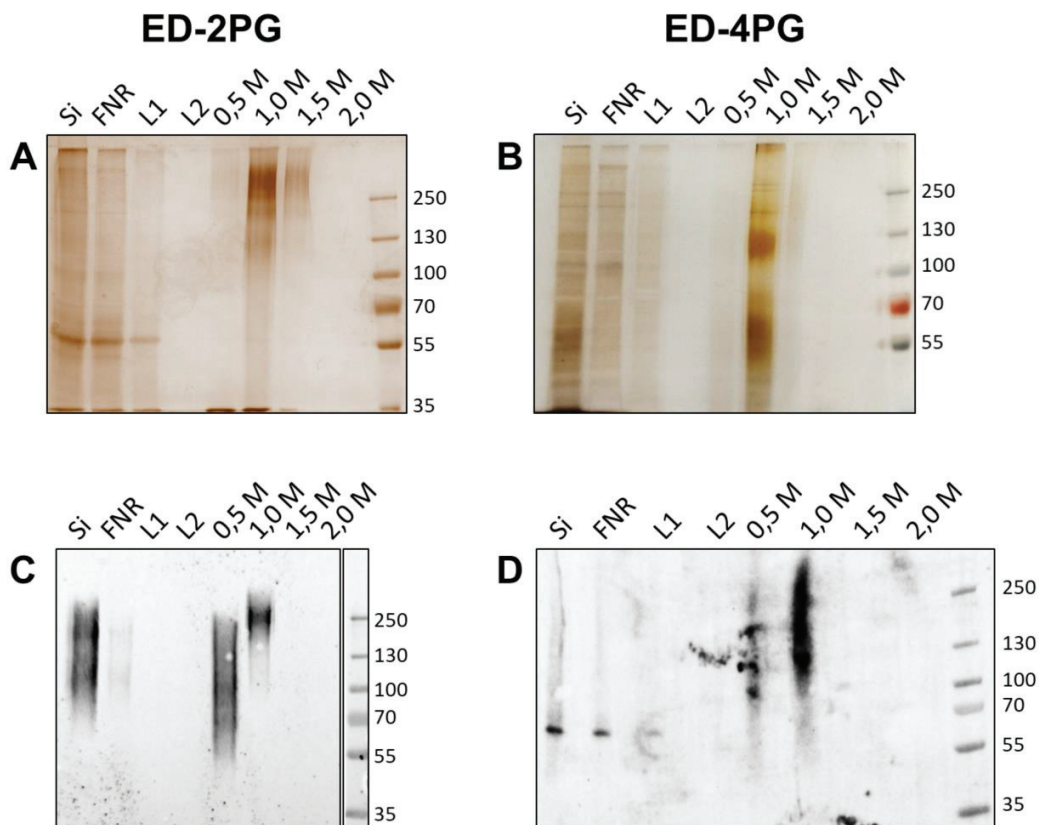
### IV.2.1 Purification des ectodomains portant leurs chaînes de GAGs

Les ectodomains des syndécans humains ont été exprimés en cellules embryonnaires de rein humain (HEK 293). Ils possèdent une étiquette 6xHis et une étiquette FLAG à leurs extrémités N- et C-terminales, respectivement. Etant donné la présence des chaînes de glycosaminoglycanes, la première étape de purification a été effectuée sur une résine échangeuse d'anions (DEAE). Les premiers essais de purifications ont été réalisés sur les ectodomains les plus petits, c'est-à-dire ED-2 et ED-4 (nommés respectivement ED-2PG et ED-4PG) avec un protocole similaire à celui détaillé par *Park* (2018). De l'urée a été dissoute dans le milieu récolté à une concentration finale de 5 M ainsi que de l'acétate de sodium 100 mM et le pH de la solution a été ajusté à 4,5. Le milieu a été purifié par chromatographie échangeuse d'anions puis le tampon a été échangé contre du tampon HBS pH 7,4. ED-2PG et ED-4PG ont été élués par différents paliers contenant des concentrations croissantes de 0,5 à 2 M NaCl en HBS (**Figure 109**).

L'analyse par électrophorèse des fractions obtenues au cours de la purification d'ED-2PG montre la présence de bandes diffuses majoritairement dans les fractions éluées par du NaCl 1,0 et 1,5 M (**Figure 109A**). L'analyse des fractions issus de la purification d'ED-4PG montre que les molécules éluées de la résine échangeuse d'anions sont concentrées dans la fraction éluee par du NaCl 1,0 M (**Figure 109B**).

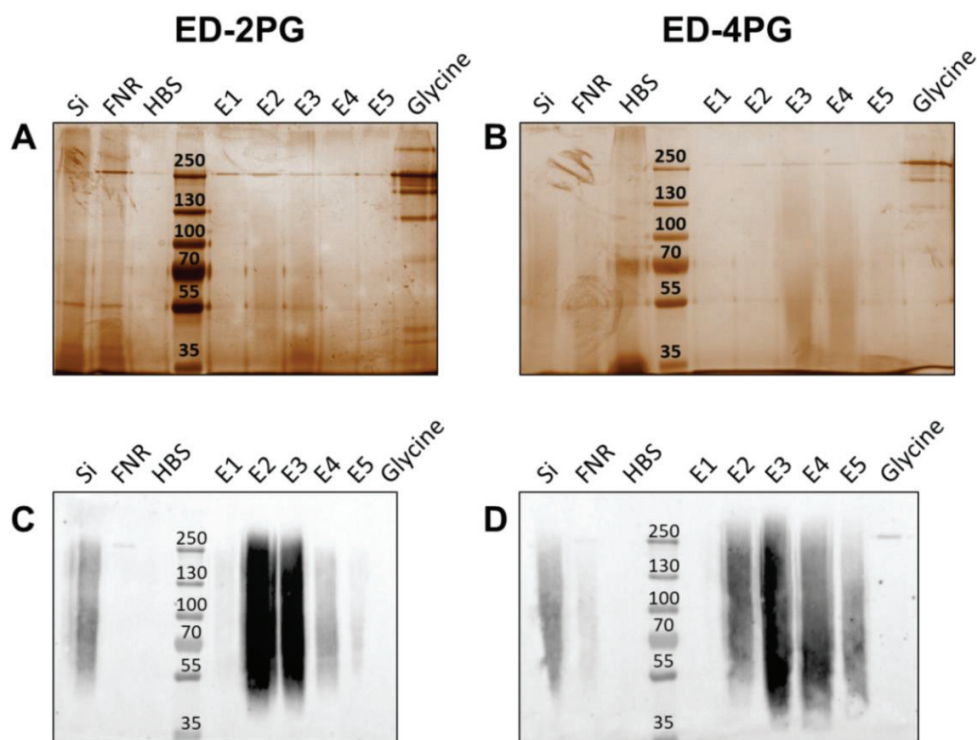
L'analyse par Western blot montre qu'ED-2PG est sécrété par les cellules HEK293 et qu'il est capturé par la résine échangeuse d'anions car il n'est pas détecté dans la fraction non retenue (**Figure 109C**). ED-2PG est élué par du NaCl 0,5 M et migre sous forme de bandes diffuses dont la masse moléculaire apparente varie de 55 à 250 kDa. ED-2PG est également élué par du NaCl 1,0 M et migre sous la forme d'une bande diffuse d'environ 250 kDa (**Figure 109C**).

En ce qui concerne ED-4PG, l'analyse par Western blot montre la présence d'une seule bande migrant avec une masse moléculaire apparente de 55 kDa dans le milieu de culture, dans la fraction non retenue et dans le premier lavage (**Figure 109D**). L'essentiel du matériel est élué par du NaCl 1,0 M sous forme d'une bande diffuse dont la masse moléculaire varie de 100 à 250 kDa. Une partie des molécules est également éluee par du NaCl 0,5 M (**Figure 109D**).



**Figure 109 : Purification par chromatographie échangeuse d’anions de l’ectodomaine du syndécan-2 et -4 portant leurs chaînes de glycosaminoglycanes.** Analyse par électrophorèse en conditions dénaturantes des fractions issues de la purification d’ED-2PG (**A**) et d’ED-4PG (**B**). *Si*: solution initiale, *FNR*: fraction non retenue, lavage en acétate de sodium 100 mM pH 4,5 (*L1*) puis en HBS (*L2*), *0,5 – 2 M*: fractions éluées par 4 concentrations de NaCl (0,5, 1,0, 1,5 et 2 M). Les protéines ont été colorées par du nitrate d’argent. Analyse par Western blot avec un anticorps anti-FLAG des fractions issues de la purification d’ED-2PG (**C**) et d’ED-4PG (**D**). Les bandes immunoréactives ont été détectées par chimiluminescence. 200 µl de chaque fraction ont été précipités par l’acétone et la totalité du précipité dissous dans le tampon de charge a été déposé sur un gel de polyacrylamide 8 %. 5 µl de marqueurs de masse moléculaire ont été déposés (PageRuler Plus, 26619, ThermoFisher).

Une seconde étape de purification a été réalisée par chromatographie d’affinité sur résine anti-FLAG pour la fraction éluée par le NaCl 0,5 M pour ED-2PG, et par le NaCl 1,0 M pour ED-4PG (**Figure 110**).

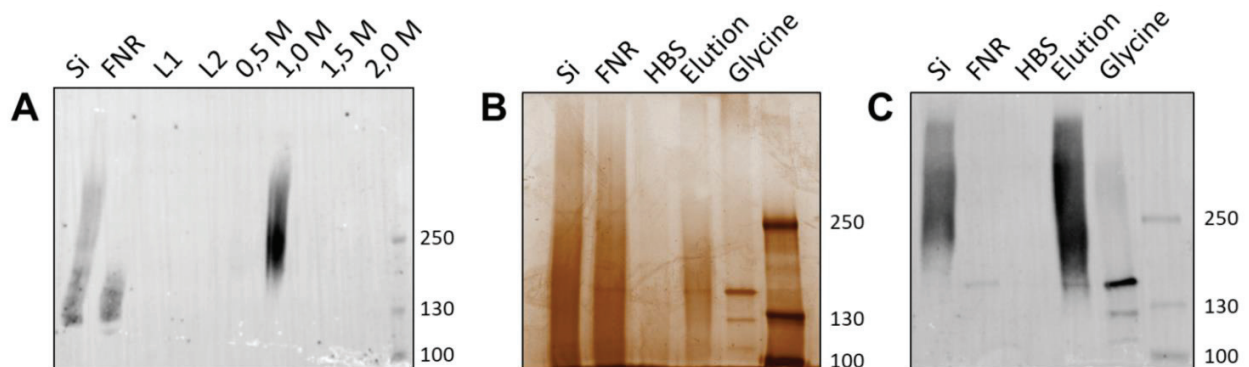


**Figure 110 : Purification par chromatographie d'affinité sur résine anti-FLAG de l'ectodomaine du syndécan-2 et -4 portant leurs chaînes de glycosaminoglycanes.** Analyse par électrophorèse en conditions dénaturantes des fractions issues de la purification d'ED-2PG (**A**) et d'ED-4PG (**B**). *Si*: solution initiale, *FNR*: fraction non retenue, *HBS*: lavage par du tampon HBS, *E1-5*: fractions éluées lors de l'incubation avec du peptide FLAG, *Glycine*: lavage par une solution de glycine 100 mM pH 3,5. Les protéines ont été colorées par du nitrate d'argent. Analyse par Western blot avec un anticorps anti-FLAG des fractions issues de la purification d'ED-2PG (**C**) et d'ED-4PG (**D**). Les bandes immunoréactives ont été détectées par chimiluminescence. 200  $\mu$ l de de chaque fraction ont été précipités par l'acétone et la totalité du précipité dissous dans le tampon de charge a été déposé sur un gel de polyacrylamide 8 %. 5  $\mu$ l de marqueurs de masse moléculaire ont été déposés (PageRuler Plus, 26619, ThermoFisher).

L'analyse par électrophorèse suivie d'une coloration au nitrate d'argent montre qu'ils sont majoritairement élués dans les fractions 2 et 3 pour ED-2PG, et 3 et 4 pour ED-4PG. L'analyse par Western blot confirme ces résultats et montre que la deuxième et la cinquième fraction contiennent une quantité non négligeable d'ED-4PG (**Figures 110C** et **110D**).

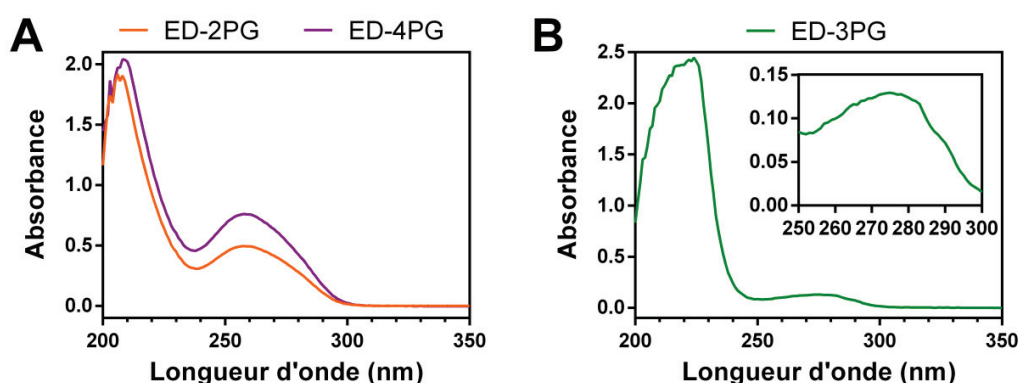
Une espèce de masse moléculaire apparente d'environ 250 kDa et migrant comme une bande définie est présente majoritairement dans la fraction éluée par la glycine et en faible quantité dans les fractions d'éluion d'ED-2PG et d'ED-4PG (**Figures 110A et 110B**). Elle est reconnue par l'anticorps anti-FLAG uniquement pour ED-4PG dans la fraction éluée par la glycine lors de la purification d'ED-4PG (**Figure 110D**).

Le même protocole a été suivi pour l'expression et la purification de l'ectodomaine du syndécan-3 (ED-3PG). ED-3PG a dans un premier temps été purifié par chromatographie échangeuse d'anions (**Figure 111A**). La totalité d'ED-3PG est éluée à une concentration en NaCl de 1,0 M. Cette fraction a été ensuite purifiée par chromatographie d'affinité sur une résine anti-FLAG (**Figures 111B et 111C**). ED-3PG est élué sous forme d'une bande diffuse reconnue par l'anticorps anti-FLAG et de masse moléculaire apparente d'au moins 150 kDa. Trois espèces ayant une masse moléculaire d'environ 110, 130 et 155 kDa sont majoritairement éluées lors de la régénération de la résine anti-FLAG (lavage glycine) et une faible quantité est également retrouvé co-éluée avec ED-3PG.



**Figure 111 : Purification de l'ectodomaine du syndécan-3 portant ses chaînes de glycosaminoglycanes. A)** Analyse par Western blot avec un anticorps anti-FLAG des fractions issues de la purification d'ED-3 PG par chromatographie échangeuse d'anions. *Si*: solution initiale, *FNR*: fraction non retenue, lavage en acétate de sodium 100 mM pH 4,5 (*L1*) puis en HBS (*L2*), *0,5 – 2 M*: fractions éluées par 4 paliers successifs contenant des concentrations croissances de NaCl (0,5 - 2 M). Les bandes immunoréactives ont été détectées par chimiluminescence. Analyse par électrophorèse en conditions dénaturantes suivie d'une coloration au nitrate d'argent (**B**) et par Western blot avec un anticorps anti-FLAG (**C**) des fractions issues de la purification d'ED-3 PG par chromatographie d'affinité sur résine anti-FLAG. *Si*: solution initiale, *FNR*: fraction non retenue, *HBS*: lavage par du tampon HBS, *Elution*: fraction éluée lors de l'incubation avec du peptide FLAG, *Glycine*: lavage par une solution de glycine 100 mM pH 3,5. Pour chaque analyse, 200 µl de chaque fraction ont été précipités par l'acétone et la totalité du précipité dissous dans du tampon de charge a été déposé sur un gel de polyacrylamide 5 %. Piste de droite : marqueurs de masse moléculaire (kDa).

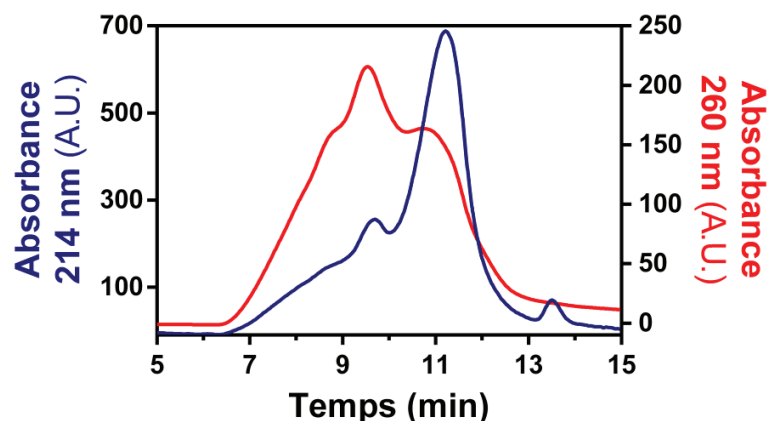
Les fractions d'élutions ont été concentrées par centrifugation et l'absorbance du rétentat a été mesurée par spectrophotométrie afin de déterminer les concentrations d'ED-2, -3 et -4PG (**Figure 112**). La fraction concentrée d'ED-2PG et d'ED-4PG absorbe fortement à 260 nm (**Figure 112A**). Une forte absorbance à 260 nm trahit la présence d'acides nucléiques qui ont pu être co-purifiés notamment lors de leur passage sur la résine échangeuse d'anions. La concentration d'ED-2PG et d'ED-4PG n'a ainsi pas pu être déterminée précisément. La fraction concentrée d'ED-3PG n'est apparemment pas contaminée par des acides nucléiques et sa concentration a pu être estimée à 124  $\mu\text{M}$  (**Figure 112B**).



**Figure 112 : Détermination par spectrophotométrie de la concentration des ectodomaines des syndécans avec GAGs.** Spectres d'absorbance du rétentat après concentration d'ED-2PG (orange) et d'ED-4PG (violet) (**A**) et d'ED-3PG (**B**). L'agrandissement montre la longueur d'onde maximale d'absorbance entre 250 et 300 nm. Les valeurs d'absorbance ont été enregistrées avec un trajet optique de 1 mm dans du tampon HBS contenant du NaCl 0,3 M.

#### IV.2.2 Analyses par SEC-SAXS des ectodomaines des syndécans avec GAGs

ED-2, -3 et -4PG ont été analysés par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique ainsi qu'à un appareil de diffusion de la lumière laser multi-angles équipé d'un appareil de diffusion dynamique de la lumière (expériences de SEC-SAXS-MALLS-DLS sur la ligne de lumière SWING, Synchrotron SOLEIL). Dans un premiers temps, une colonne BioSec 5-1000 (Agilent) de 5 ml, disponible sur la ligne de lumière SWING, a été utilisée pour analyser ED-2PG et ED-4PG. Cette colonne possède une gamme de fractionnement s'étendant de 50 à 750 kDa pour une protéine globulaire. L'identification des EDs des syndécans portant leurs chaînes de GAGs parmi les autres espèces moléculaires s'est faite en suivant leur élution par spectrophotométrie à 214 nm (**Figure 113**).



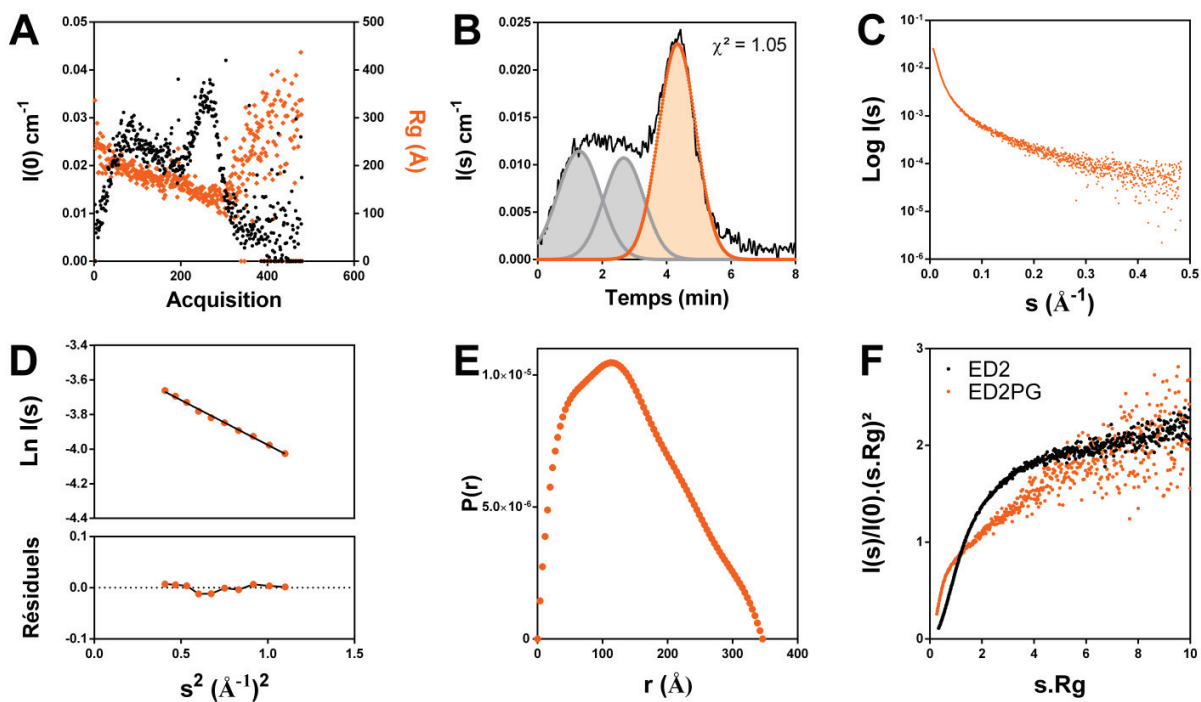
**Figure 113 : Suivi de l'élution des espèces moléculaires présentes dans l'échantillon d'ED-2PG par spectrophotométrie lors des expériences de SEC-SAXS.** Colonne BioSec 5-1000.

L'analyse de l'échantillon d'ED-2PG par chromatographie d'exclusion stérique montre une importante hétérogénéité. Une espèce est éluée à environ 9,5 min et absorbe à 260 nm alors qu'ED-2PG est distinguable par sa forte absorbance à 214 nm du fait de la présence des chaînes de glycosaminoglycanes (**Figure 113**). Le graphique représentant l'intensité de diffusion et l'estimation du rayon de gyration en fonction des acquisitions montre également que la solution n'est pas homogène (**Figure 114A**). Plusieurs espèces ayant des volumes hydrodynamiques plus importants sont éluées avant ED-2PG et possèdent un rayon de gyration d'environ 200 Å. Les intensités de diffusion ont alors été déconvoluées par le module *US-SOMO* bien que les différentes espèces soient peu séparées lors de la chromatographie d'exclusion stérique (**Figure 114B**). Les intensités de diffusion propres à ED-2PG ont pu être extraites (**Figure 114C**). ED-2PG possède un rayon de gyration de  $127,17 \pm 1,29$  Å bien que la gamme linéaire dans la représentation de Guinier soit limitée à seulement 10 points ce qui limite la précision de la mesure. La distance maximale ( $D_{max}$ ) d'ED-2PG a été déduite de la fonction de distribution de paires et est égale à environ 350 Å, ce qui correspond au double du  $D_{max}$  de la protéine seule, estimé à 185 Å (**Figures 114D et 114E**).

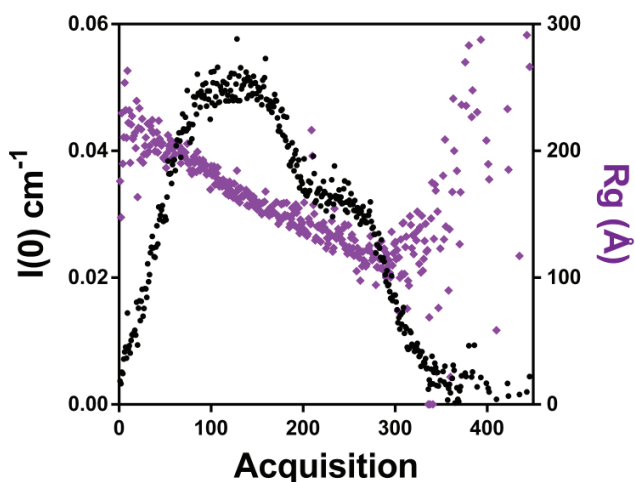
La représentation normalisée de Kratky d'ED-2PG est caractéristique d'un système flexible ce qui est cohérent avec la présence de désordre intrinsèque et des chaînes de glycosaminoglycanes (**Figure 114F**). Les intensités décroissent environ en  $s^{-2}$  et l'exposant de Porod est égal à 1,7, en accord avec un système flexible.

Concernant ED-4PG, le graphique représentant l'intensité de diffusion et l'estimation du rayon de gyration en fonction des acquisitions montre que les différentes espèces ne sont pas suffisamment séparées (**Figure 115**). Les intensités de diffusion n'ont donc pas été déconvoluées.



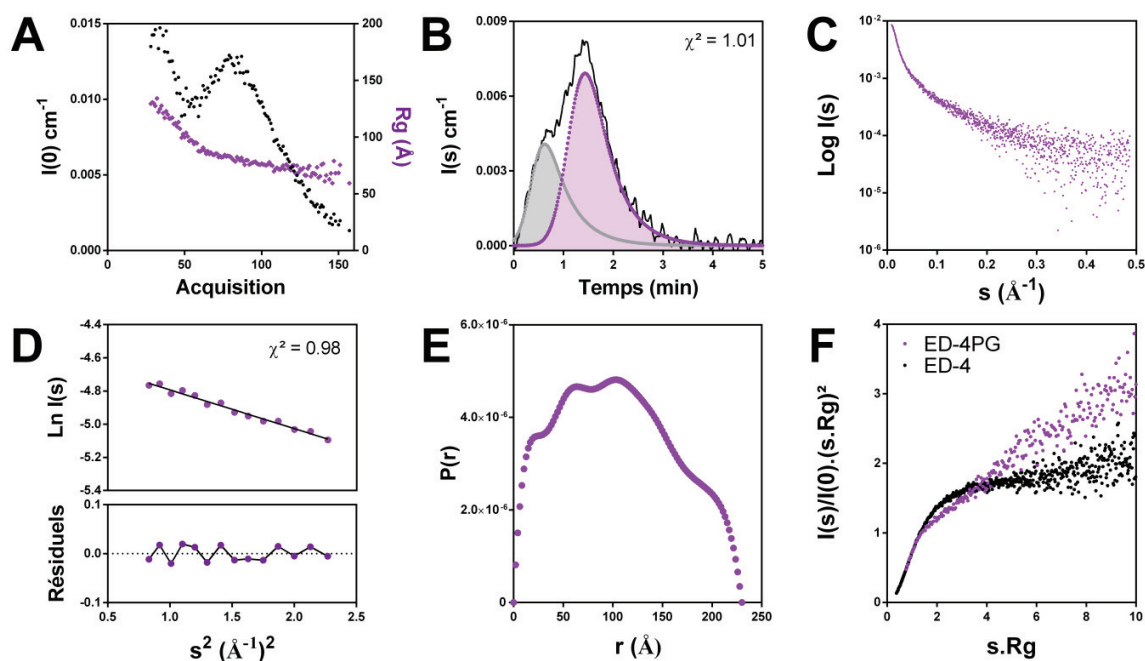


**Figure 114: Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-2PG. A)** Intensité de diffusion et estimation du rayon de gyration en fonction des acquisitions. **B)** Déconvolution des intensités de diffusion en fonction du temps par *US-SOMO*. Les données issues d'un unique angle sont représentées pour des raisons de clarté. La courbe en orange correspond à ED-2PG. La qualité de l'ajustement est indiquée par le  $\chi^2$ . **C)** Intensités de diffusion finales. **D)** Estimation du rayon de gyration selon l'approximation de Guinier. **E)** Fonction de distribution de paires,  $P(r)$ . **F)** Représentation normalisée de Kratky pour ED-2PG et d'ED-2 (sans chaînes de glycosaminoglycanes).



**Figure 115: Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-4PG.** Le graphique montre l'intensité de diffusion et l'estimation du rayon de gyration en fonction des acquisitions. Colonne BioSec 5-1000.

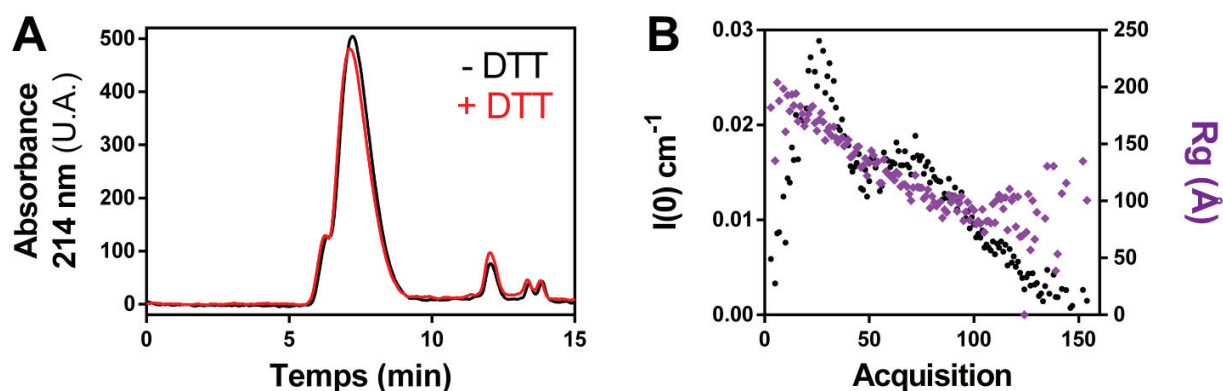
Une seconde colonne de chromatographie d'exclusion stérique (Biosec 3-300 de 5 ml, Agilent) ayant une gamme de fractionnement plus importante (5 kDa et 1 250 kDa) a été utilisée afin de séparer les ectodomains des syndécans 3 et 4 (ED-3PG et ED-4PG) des contaminants de plus haute masse moléculaire. Les différentes espèces présentes dans la solution d'ED-4PG restent peu séparées mais suffisamment pour qu'une déconvolution par *US-SOMO* puisse être réalisée (**Figures 116A** et **116B**) et qu'une courbe de diffusion propre à ED4PG soit extraite (**Figure 116C**).



**Figure 116 : Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-4PG. A)** Intensité de diffusion et estimation du rayon de gyration en fonction des acquisitions. **B)** Déconvolution des intensités de diffusion en fonction du temps par *US-SOMO*. Les données issues d'un unique angle sont représentées pour des raisons de clarté. La courbe en violet correspond à ED-4PG. La qualité de l'ajustement est représentée par le  $\chi^2$ . **C)** Intensités de diffusion finales. **D)** Estimation du rayon de gyration selon l'approximation de Guinier. **E)** Fonction de distribution de paires  $P(r)$ . **F)** Représentation normalisée de Kratky pour ED-4PG et la forme monomérique d'ED-4 exprimé en système procaryote.

ED-4PG possède un rayon de gyration de  $83,59 \pm 2,4 \text{ \AA}$  (**Figure 116D**) et un  $D_{\text{max}}$  d'environ  $230 \text{ \AA}$  d'après la fonction de distribution de paires (**Figure 116E**). La représentation normalisée de Kratky montre un système flexible comme pour ED-2PG (**Figure 116F**) et plus flexible que la forme monomérique d'ED-4 dépourvu de ses chaînes de glycosaminoglycanes comme en atteste la valeur de l'exposant de Porod qui est d'environ 1,3 et celui de la protéine seule précédemment estimée à 1,7.

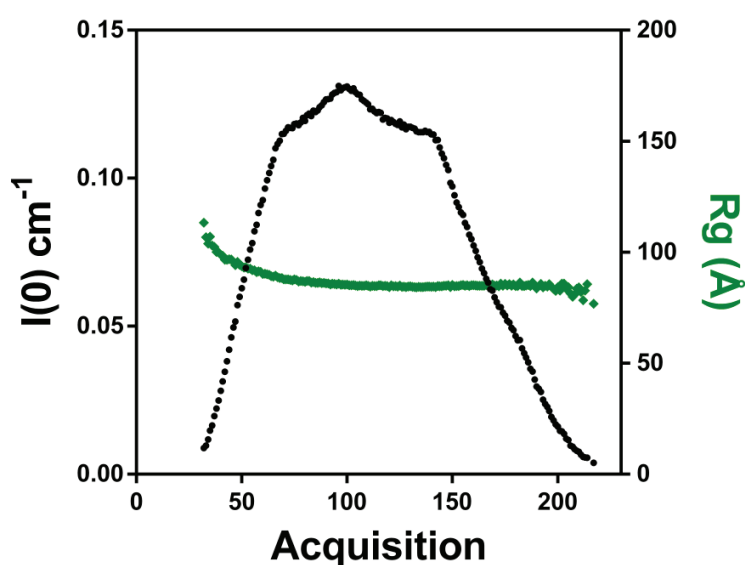
ED-4PG a ensuite été incubé en présence de DTT 2 mM pendant 30 min et à température ambiante avant son analyse afin de vérifier s'il avait la capacité à dimériser. Le chromatogramme ne montre pas de différence avec celui d'ED-4PG en conditions non réductrices (**Figure 117A**) et son rayon de gyration n'a pas été affecté (**Figure 117B**) ce qui suggère qu'ED-4PG ne dimérise pas dans ces conditions.



**Figure 117: Analyse d'ED-4PG1 par SEC-SAXS en conditions réductrices. A)** Chromatogramme de la séparation par chromatographie d'exclusion stérique en conditions non réductrices (-DTT) ou réductrices (+DTT). **B)** Intensité de diffusion et estimation du rayon de gyration en fonction des acquisitions d'ED-4PG réduit.

ED-2PG et ED-4PG n'ont pas été séparés de manière optimale des contaminants par chromatographie d'exclusion stérique ce qui n'a pas permis de mesurer leur masse moléculaire et leur rayon hydrodynamique d'autant plus que des phénomènes de dilution et de ré-homogénéisation des espèces peuvent se produire entre le capillaire exposé aux rayons X et l'appareil de MALLS.

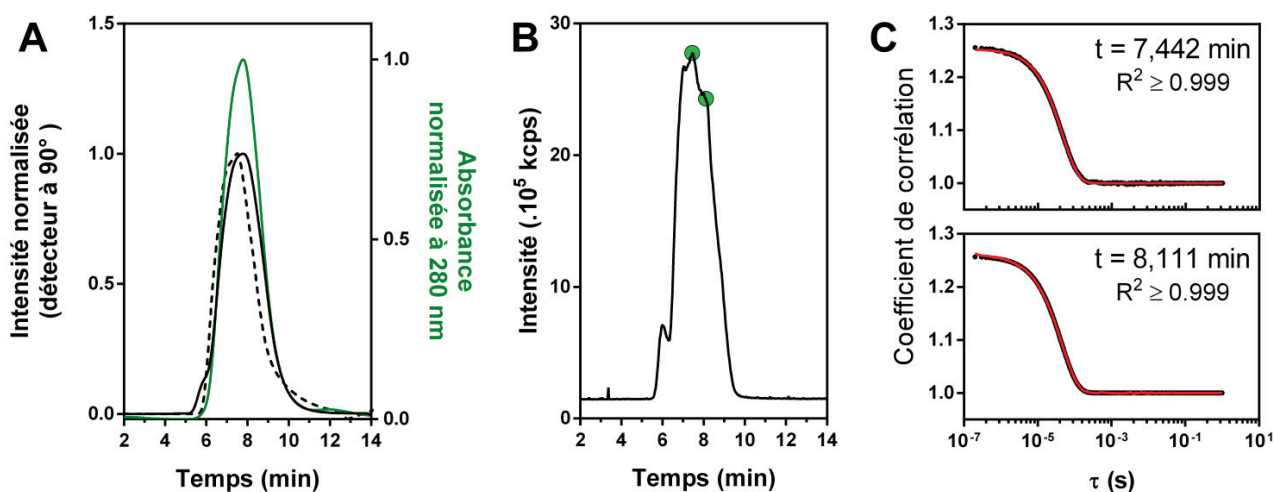
L'ectodomaine du syndécan-3 portant ses chaînes de GAGs, ED-3PG, a été analysé dans les mêmes conditions (**Figure 118**). D'après les précédentes analyses par spectrophotométrie, ED-3PG ne semblait pas être contaminé par de potentiels acides nucléiques et l'analyse par électrophorèse montrait la présence d'une faible quantité de contaminants ayant une masse moléculaire apparente d'environ 155 kDa (*cf.* Figures 111B et 111C). La représentation de l'intensité de diffusion à angle zéro et l'estimation du Rg en fonction des acquisitions montre cependant la présence de plusieurs espèces coéluees (**Figure 118**). Le rayon de gyration estimé est cependant stable sur presque la totalité des acquisitions et est d'environ 85 Å. Etant donné la polydispersité de l'échantillon, les intensités de diffusion n'ont pas été déconvoluées. Cette polydispersité pourrait être la conséquence de la présence de différentes chaînes de glycosaminoglycanes (héparane sulfate et chondroïtine sulfate) qui ne sont pas nécessairement de même longueur ou de même taux de sulfatation.



**Figure 118: Analyse par diffusion des rayons X aux petits angles couplée à la chromatographie d'exclusion stérique d'ED-3PG.** Le graphique montre l'intensité de diffusion et l'estimation du rayon de gyration en fonction des acquisitions.

Les espèces présentes dans l'échantillon d'ED-3PG possèdent un rayon de gyration similaire, excepté pour la première espèce éluée entre environ la 40<sup>ème</sup> et la 50<sup>ème</sup> acquisition selon l'estimation du Rg (**Figure 118**). La détermination de la masse moléculaire par MALLS et du rayon hydrodynamique par DLS a été effectuée en ne prenant en compte que la seconde partie des pics afin de s'abstenir de la contribution de la première espèce éluée (**Figure 119**).

Le contaminant est élué au bout d'environ 6 min et diffuse relativement peu (**Figure 119A**). La masse moléculaire d'ED-3PG estimée par SEC-MALLS est comprise entre 95 et 100 kDa, bien supérieure à celle de la protéine seule (38,8 kDa) (**Figure 119B**). Cette observation suggère que la masse moléculaire des chaînes de glycosaminoglycanes est d'environ 60 kDa. Son rayon hydrodynamique a été calculé au sommet du pic le plus intense (7,442 min) ainsi qu'au niveau du dernier pic différenciable (8,111 min) pour ne pas inclure la contribution du contaminant. Les corrélogrammes associés montrent la présence d'une unique espèce ayant un rayon hydrodynamique de  $94,2 \pm 2,0 \text{ \AA}$  et de  $90,9 \pm 1,9 \text{ \AA}$  pour les temps d'éluéon d'environ 7,4 et 8,1 min, respectivement, c'est-à-dire près de deux fois plus que la protéine seule dont le Rh a été estimé à  $50,1 \text{ \AA}$  (**Figures 119B** et **119C**). L'espèce contaminante possède une masse moléculaire d'environ 1 million de dalton.



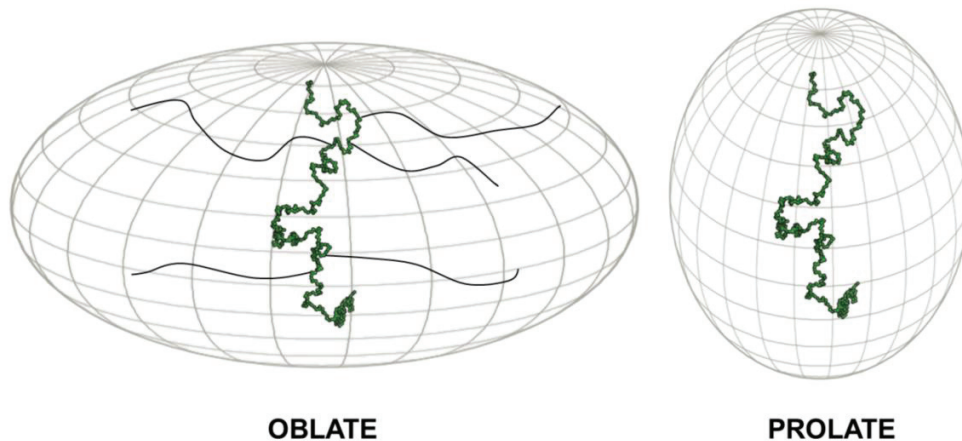
**Figure 119 : Détermination de la masse moléculaire et du rayon hydrodynamique d'ED-3PG.** **A)** Expérience de SEC-MALLS. Les intensités diffusées à  $90^\circ$  sont représentées par une courbe noire et la différence de l'indice de réfraction est en pointillés. **B)** Analyse d'ED-3PG par SEC-DLS. Deux valeurs du rayon hydrodynamique ont été calculées aux temps d'éluéon de 7,442 et 8,111 min (cercles verts). kcps : milliers de coups par seconde (*i.e.* nombre moyen de photons détectés). **C)** Corréllogrammes des mesures par SEC-DLS réalisées en *B)* aux temps 7,442 et 8,111 min. La qualité des ajustements est indiquée ( $R^2$ ).

Ces expériences préliminaires montrent que les ectodomaines des syndécans portant leurs chaînes de glycosaminoglycanes sont bien plus volumineux que la partie protéique seule (**Tableau XXIII**), notamment pour ED-2PG dont le rayon de gyration (127 Å) est triplé par rapport à la partie protéique seule (43 Å) et son Dmax est quasiment doublé (350 et 185 Å, respectivement). Le rayon de gyration d'ED-4PG correspond au double de la forme monomérique d'ED-4 (83,6 et 41,5 Å, respectivement) et est supérieur à celui de sa forme dimérique (59,3 Å) mais son Dmax est en revanche similaire à celui du dimère d'ED-4 sans chaînes de GAGs (230 et 240 Å, respectivement). Le rayon de gyration d'ED-3PG est également plus important que sa partie protéique seule mais la différence est plus faible que pour ED-2PG et ED-4PG.

	ED-2	ED-2PG	ED-3	ED-3PG	ED-4	ED-4PG
<b>Masse moléculaire apparente (kDa)</b>	28	<b>55 - 250</b>	89	<b>150 - &gt;250</b>	28 63	<b>100 - 250</b>
<b>Rayon de gyration (Å)</b>	42,9	<b>127</b>	65,0	<b>≈ 85</b>	41,5 59,3	<b>83,6</b>
<b>Dmax (Å)</b>	185	<b>350</b>	300	<i>ND</i>	180 240	<b>230</b>
<b>Masse moléculaire (SEC-MALLS, kDa)</b>	16,6	<i>ND</i>	38,8	<b>≈ 100</b>	18,0 32,6	<i>ND</i>
<b>Rayon Hydrodynamique (Å)</b>	34,1	<i>ND</i>	50,1	<b>≈ 92,5</b>	21,6 45,4	<i>ND</i>

**Tableau XXIII : Résumé des résultats obtenus sur les ectodomaines des syndécans portant leurs chaînes de GAGs ainsi que leur comparaison aux paramètres des protéines seules.** La masse moléculaire apparente a été déterminée par électrophorèse sur gel de polyacrylamide. *ND*: non déterminé.

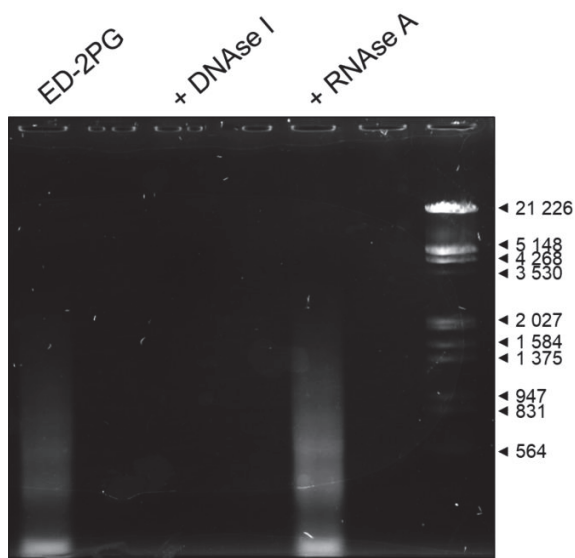
Le rayon hydrodynamique d'ED-3PG est presque deux fois plus important que celui de l'ectodomaine du syndécan-3 sans chaînes de GAGs (92,5 et 50,1 Å, respectivement) menant à un rapport Rg/Rh d'environ 0,9 ce qui caractérise une conformation générale en ellipsoïde oblate à l'inverse de la protéine seule dont la conformation est plutôt prolate (**Figure 120**).



**Figure 120 : Représentation schématique de la conformation générale de l'ectodomaine du syndécan-3 portant ou non ses chaînes de GAGs.** Selon la présence des chaînes de glycosaminoglycanes (traits noirs), l'ectodomaine du syndécan-3 (vert) adopte une conformation en sphéroïde oblate (+ GAG) ou prolate (protéine seule).

#### IV.2.3 Discussion et Perspectives

Les ectodomains des syndécans portant leurs chaînes de glycosaminoglycanes ont été exprimés en système eucaryote et leur purification sur résine échangeuse d'anions a permis d'en isoler la grande majorité. La purification par chromatographie d'affinité sur résine anti-FLAG n'a en revanche pas permis d'éliminer la totalité des contaminants. Une espèce en particulier absorbe intensément à 260 nm et possède une masse moléculaire d'environ 1 MDa suggérant une contamination par des acides nucléiques qui ont pu être purifiés lors de l'étape de chromatographie échangeuse d'anions, mais ils n'auraient pas dû en revanche être retenus par la résine anti-FLAG. Ce contaminant pourra être identifié sur gel d'agarose suivie d'une révélation au bromure d'éthidium ou lors de digestion par la désoxyribonucléase ou la ribonucléase qui devrait provoquer sa disparition s'il s'agit d'acides nucléiques. Dans le cas où cette espèce serait une protéine, son identification pourra être réalisée par spectrométrie de masse en tandem à partir d'une bande isolée d'un gel d'électrophorèse. Ces hypothèses ont été vérifiées récemment. Pour cela, un échantillon d'ED-2PG absorbant fortement à 260 nm a été incubé en présence de DNase I ou de RNase A puis analysé par électrophorèse sur gel d'agarose (Figure ci-après). Cette expérience montre qu'ED-2PG est fortement contaminé par de l'ADN et ces résultats permettront de préparer des échantillons purs lors de prochaines expériences.



**Digestion d'un échantillon d'ED-2PG par la DNase I et la RNase A.** 30  $\mu$ l d'un échantillon d'ED-2PG ont été incubés seul (*ED-2PG*) ou en présence de 100  $\mu$ g/ml de DNase I (+ *DNase I*, Sigma, D5025) ou de RNase A (+ *RNase A*, Sigma, R6513) pendant une heure à 37°C puis analysés par électrophorèse sur gel d'agarose 0,5 %. Les acides nucléiques ont été révélés par le bromure d'éthidium. Piste de droite : marqueurs de masses moléculaires (en paires de bases, Lambda DNA/EcoRI+HindIII Marker 3, SM0191, ThermoFisher).

La contamination par de l'ADN ou de l'ARN a fréquemment été observée lors de la purification de glycosaminoglycanes ou de protéoglycanes de différents tissus (*Lyon & Gallagher, 1991, van Gemst et al., 2016*). En revanche, les protocoles d'extraction utilisés par les auteurs nécessitaient la lyse des cellules, ce qui n'est pas le cas dans nos expériences où les protéoglycanes sont sécrétés dans le milieu de culture.

L'incubation des échantillons de protéoglycanes ou de glycosaminoglycanes pré-purifiés avec des nucléases permet cependant d'éliminer la totalité des acides nucléiques. *Lyon & Gallagher* (1991) ont également montrés que les protéoglycanes pouvaient être purifiés des acides nucléiques par ajout d'acide trichloroacétique où les protéoglycanes restaient solubles. La présence de ce contaminant a généré des difficultés à analyser les ectodomains par SAXS étant donné que leurs volumes d'élution étaient proches lors de la séparation par chromatographie d'exclusion stérique. La seconde colonne utilisée (BioSec 3-300, Agilent) a mené à une meilleure séparation mais qui ne restait cependant pas optimale. L'ensemble des résultats obtenus lors des expériences de SEC-SAXS-MALLS-DLS doivent ainsi être vérifiés sur des échantillons plus homogènes.

La masse apparente moyenne des ectodomains portant leurs chaînes de glycosaminoglycanes a été estimée sur gel de polyacrylamide et varie de 55 à 250 kDa pour ED-2PG, de 130 à plus de 250 kDa pour ED-3PG et de 100 à 250 kDa pour ED-4PG. L'analyse par électrophorèse de l'ectodomaine de l'isoforme-1 du syndécans-4 exprimé en cellules de glande mammaire de souris mène également à l'apparition d'une bande diffuse de 37 à 250 kDa ce qui est très similaire à nos observations (*Park, 2018*). D'après ces premières analyses, ED-2PG possède le rayon de gyration le plus important des trois ectodomains étudiées.



Cela est inattendu étant donné que sa taille est similaire à celle d'ED-4PG. La taille maximale d'ED-2PG ( $D_{max}$ ) est également bien supérieure à celle d'ED-4PG (350 et 230 Å, respectivement). L'analyse d'ED-4PG en conditions réductrices suggère qu'il ne dimérise pas dans ces conditions. L'encombrement stérique généré par les chaînes de GAGs pourrait défavoriser l'interaction entre les extrémités C-terminales d'ED-4PG et donc empêcher la formation du pont disulfure bien que les chaînes de GAGs soient greffées à son extrémité N-terminale.

L'ectodomaine du syndécan-3 portant ses chaînes de GAGs est également plus volumineux que la protéine seule, ce qui est attendu. Son rayon de gyration est 30 % plus important mais son  $D_{max}$  n'a pas pu être calculé étant donné que les intensités de diffusion n'ont pas été déconvoluées. L'analyse par SEC-MALLS-DLS a cependant permis d'estimer sa masse moléculaire valant environ 100 kDa ce qui suggère que la masse des chaînes de GAGs représente environ 60 kDa. Étant donné que le syndécan-3 peut porter jusqu'à 6 chaînes, cela correspondrait à une masse moyenne de 10 kDa par chaîne soit une trentaine de répétition de l'unité disaccharidique, ce qui est relativement peu par rapport aux 50 – 150 unités généralement admises (*Bülow & Hobert, 2006*).

L'ectodomaine du syndécan-1 portant ses chaînes de GAGs (ED-1PG) a également été exprimé et purifié dans les mêmes conditions que les autres ectodomains, mais aucune analyse biophysique n'a été effectuée jusqu'à présent en raison d'une trop grande quantité de contaminant.

Lors de futures expériences de SAXS, le débit de passage de l'échantillon dans le capillaire illuminé par les rayons X pourrait être modulé. Des analyses en absence de flux ou à débit élevé pourraient donner des résultats différents, notamment sur la taille maximale des particules. Cela permettrait de calculer des paramètres sur la taille des syndécans soumis à des forces de cisaillement faibles voire nulles, comme ceux exposés à la surface des cellules telles que les fibroblastes, ou élevées, comme ceux exposés à la surface des cellules endothéliales.

## **IV.3**

### **INTERACTIONS DES SYNDÉCANS**

### IV.3.1 Le réseau d'interactions des syndécans

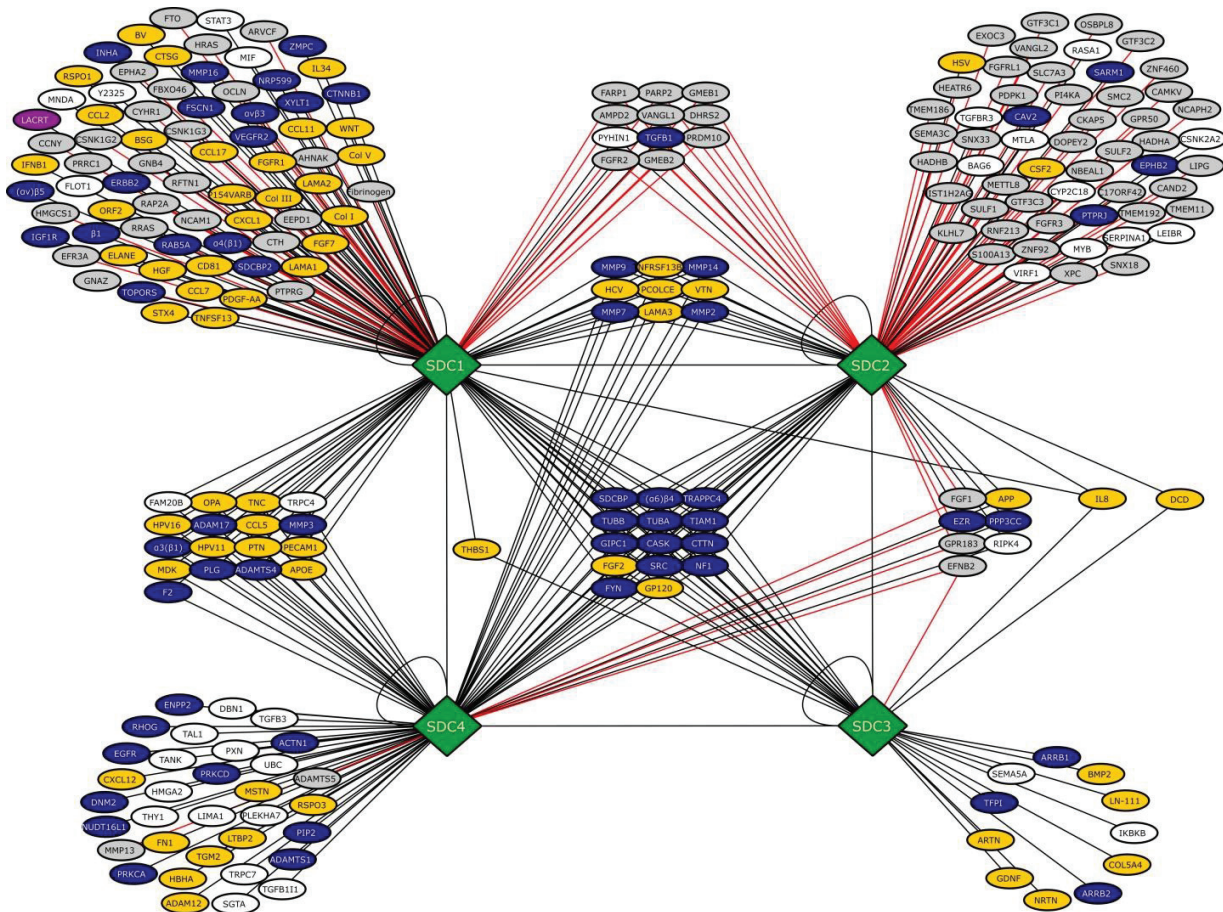
#### a) Construction de l'interactome des syndécans

L'interactome des syndécans a été construit d'une part par la curation manuelle de la littérature scientifique, dont font partie les données de deux précédentes revues (*Roper et al., 2012 ; Stepp et al., 2015*), et d'autre part, par l'interrogation des bases de données *IntAct* (*Orchard et al., 2014, [www.ebi.ac.uk/intact](http://www.ebi.ac.uk/intact)*), *MatrixDB* (*Clerc et al., 2019, <http://matrixdb.univ-lyon1.fr>*) et *Adhesome* ([www.adhesome.org/](http://www.adhesome.org/)) qui regroupe uniquement les interactions protéines-protéines des complexes d'adhésion focale dans les cellules de mammifères, et du jeu de données *Bioplex 2.0* (*Huttlin et al., 2017, <https://bioplex.hms.harvard.edu>*) qui sont des données d'interaction obtenues par chromatographie d'affinité couplée à la spectrométrie de masse. Cette méthode conduit à la purification de complexes mais ne permet pas de déterminer si les interactions existant entre les protéines constituant les complexes sont directes ou indirectes.

Selon les données expérimentales disponibles dans la littérature, des informations portant sur le site de fixation des partenaires sur les syndécans ont été ajoutées, notamment s'il s'agissait d'interactions avec leur partie protéique, et si possible le domaine précis, ou si l'interaction s'effectuait *via* les chaînes d'héparane et / ou de chondroïtine sulfate portées par l'ectodomaine des syndécans. L'interactome des syndécans et son analyse ont été publiés sous forme de revue dans *FEBS journal* (*Gondelaud & Ricard-Blum, 2019*). Il s'agit de la première version exhaustive de l'interactome des syndécans (**Figure 121**). Au total, 351 partenaires des syndécans ont été identifiés, dont la grande majorité interagissent avec leurs chaînes de GAGs (71 %, 67 %, 91 % et 68 % des partenaires pour les syndécans-1 à -4, respectivement). 131 partenaires ont été décrits pour le SDC-1 et 103 pour le SDC-2, pour lequel la majorité a été détectée par méthode haut-débit. Le SDC-4 possède 84 partenaires et le SDC-3 seulement 33. Les rares partenaires extracellulaires interagissant avec la partie protéique des syndécans sont des récepteurs membranaires comme les intégrines  $\alpha\beta3$ ,  $\alpha4\beta1$  ou  $\alpha6\beta4$  ou par exemple les récepteurs IGF1-R ou VEGFR-2, ainsi que des enzymes les clivant, comme les métalloprotéases matricielles.

Parmi les 351 partenaires, seules 14 protéines interagissent avec les quatre syndécans ce qui constitue leur interactome consensus, associé aux fonctions communes des syndécans. Il est constitué de protéines kinases, de protéines impliquées dans l'organisation du cytosquelette et du transport vésiculaire ainsi que du facteur de croissance fibroblastique 2 et de la glycoprotéine gp120 qui est exposée à la surface du virus de l'immunodéficience humaine.

De plus, 12 des 14 protéines constituant l'interactome consensus sont intracellulaires ce qui est cohérent avec la diversité des fonctions extracellulaires des quatre syndécans et la conservation de leurs fonctions intracellulaires (**Figure 121**). Les syndécans -1 et -4 sont les plus étudiés et ont donc un nombre de partenaires élevé par rapport au nombre de partenaires du syndécane-3 (33) qui est en plus exprimé dans un nombre réduit de type cellulaire.



**Figure 121 : Interactome des syndécans.** Les syndécans (SDCs) sont représentés en tant que losanges verts et leurs partenaires (ellipses) sont colorés en fonction de leur site de fixation sur les SDCs (chaînes de glycosaminoglycane : jaune, partie protéique : bleu, site hybride : violet, absence de données : blanc). Les partenaires du jeu de données *Bioplex 2.0* et de la base de données *Adhesome* sont en gris avec des liens rouges. Les interactions directes sont représentées par des liens noirs. D'après *Gondelaud & Ricard-Blum, (2019)*.

### *b) Analyse d'enrichissement de l'interactome*

L'analyse d'enrichissement de l'interactome des syndécans a été effectuée en utilisant l'outil FunRich (*Functional Enrichment analysis tool*, Pathan et al., 2015). Cette analyse a permis de mettre en évidence certaines fonctions biologiques ou voies de signalisation associées à l'interactome. L'analyse des fonctions biologiques des partenaires des syndécans a mis en avant leur implication dans des voies ou processus biologiques liés directement à ces protéoglycanes puisque 74, 65, 81 et 76 % des partenaires des SDCs - 1 à -4 y sont associés.

Presque la moitié (43 %) des protéines interagissant avec le syndécan-1 sont impliquées dans la communication cellulaire ou la transduction du signal et 77 % d'entre elles sont impliquées dans des processus cellulaires médiés par les intégrines  $\alpha\beta3$ ,  $\alpha\beta5$ ,  $\alpha3\beta1$ ,  $\alpha4\beta1$  et  $\alpha6\beta4$ . Enfin 43 % de ses partenaires sont associés aux exosomes.

Il ne ressort en revanche pas de caractéristique notable de l'analyse des partenaires spécifiques du syndécan-2, bien que 25 % d'entre eux soient associés aux lysosomes.

La plupart (75 %) des partenaires spécifiques du syndécan-3 sont extracellulaires et 25 % sont présents dans les membranes basales. 67 % de ses partenaires sont impliqués, au même titre que le syndécan-1, dans la communication cellulaire ou la transduction du signal. 58 % de ses partenaires sont retrouvés impliqués dans le guidage axonal, le syndécan-3 étant principalement exprimé dans le système nerveux.

71 % des partenaires du syndécan-4 sont impliqués dans des processus cellulaires médiés par les intégrines  $\alpha3\beta1$  et  $\alpha6\beta4$ . 39 % et 45 % des partenaires du syndécan-4 sont impliqués dans la communication cellulaire et la transduction du signal, respectivement. Enfin, 42 % sont extracellulaires et 39 % sont associés aux exosomes.

D'une façon générale, il ressort de ces analyses d'enrichissement des fonctions propres à des récepteurs et des corécepteurs, notamment celles de transduction du signal et d'interactions avec des intégrines, ce qui est attendu. Les annotations disponibles sur les partenaires des syndécans ont cependant montré que 68 % des partenaires du SDC-1 appartiennent aux voies de signalisation dans lesquelles interviennent les membres de la famille des glypicans et qu'environ 20 % de ces partenaires sont impliqués dans le métabolisme du fait de son interaction avec divers enzymes qui y sont associées comme l'AMP désaminase-2, qui intervient dans le métabolisme des nucléotides puriques, l'hydroxyméthylglutaryl-CoA synthase qui est impliquée dans la biosynthèse du cholestérol ou encore la xylosyltransférase-1 qui catalyse le transfert d'un résidu de xylose sur un résidu de sérine, nécessaire à la fixation d'une chaîne de GAG.

Le SDC-2 serait également impliqué dans la voie des nectines, qui sont ces protéines engagées dans l'adhésion cellulaire (43 % de ses partenaires). Bien que cette proportion soit importante, cette voie partage avec de nombreuses autres les partenaires du syndécan-2 comme la voie de régulation de l'apoptose mTOR (44 %), ou du suppresseur de tumeur Lkb1 (46 %). Les nouvelles pistes proposées pour le syndécan-3 sont biaisées par le peu de partenaires (seulement 11) propres à ce protéoglycane. On notera cependant qu'il interagit spécifiquement avec les arrestines et IKBKB (*Inhibitor of nuclear factor kappa-B kinase subunit  $\beta$* ) qui sont impliquées dans des voies régulées par les récepteurs de type Toll et donc dans l'immunité. Au même titre que le syndécan-1, les partenaires du syndécan-4 sont impliqués dans le métabolisme (environ 22 %) et environ 15 % dans la croissance cellulaire.

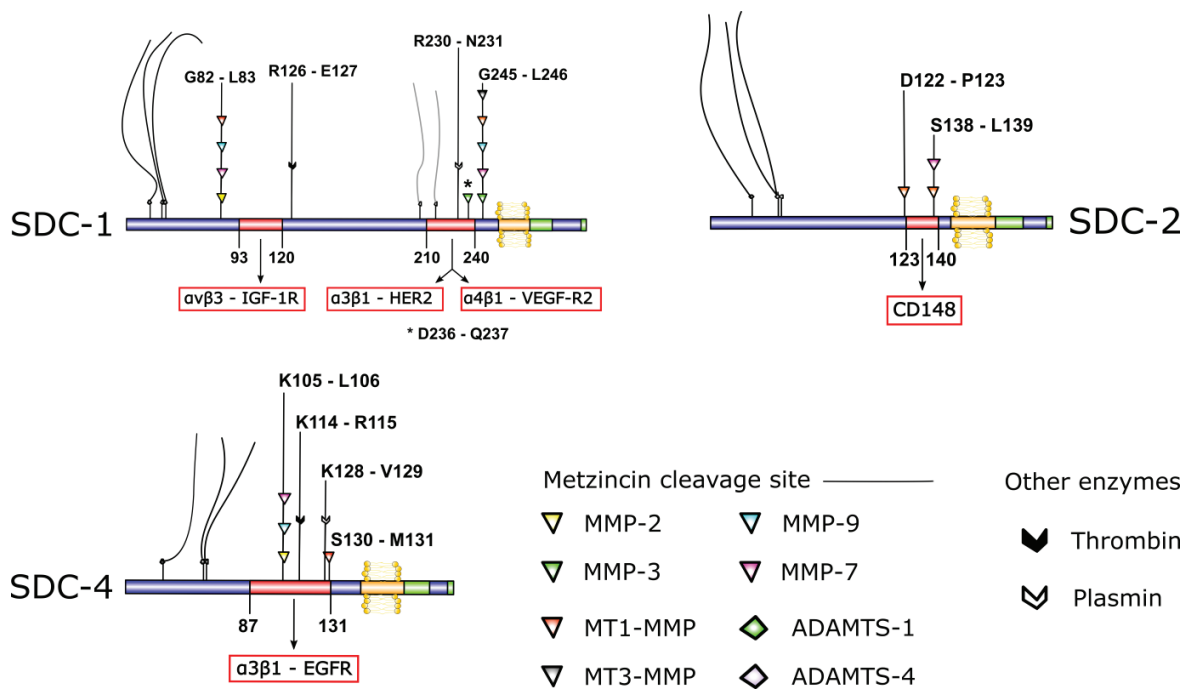
### c) Les ectodomains de syndécan, sources de matricryptines ?

Les syndécans-1 et -4 ont la capacité de former des complexes ternaires en s'associant généralement avec une intégrine et le récepteur d'un facteur de croissance à la surface cellulaire. Ce mécanisme ne nécessite pas la présence des chaînes de GAGs et seul l'ectodomaine est requis indiquant une interaction directe protéine - protéine entre les syndécans et leurs partenaires. Comme rappelé précédemment, l'interaction du syndécan-1 avec l'intégrine  $\alpha\beta3$  ou  $\alpha\beta5$  ainsi que le récepteur IGF1R (*insulin-like growth factor-1 receptor*) a été retrouvé à la surface des cellules épithéliales malignes ou durant des processus d'angiogenèse à la surface des cellules endothéliales (Rapraeger, 2013). La séquence minimale d'interaction au sein de l'ectodomaine du syndécan-1 a été identifiée et se compose des résidus de 93 à 120. Le syndécan-1 forme également un complexe ternaire avec l'intégrine  $\alpha3\beta1$  et le récepteur HER2 (*human epidermal growth factor receptor 2*). La formation de ce complexe stimule la mobilité et la survie de cellules de carcinomes. Le syndécan-1 interagit avec ses partenaires *via* les résidus 210 – 240. Ce même site interagit avec l'intégrine  $\alpha4\beta1$  (également appelé VLA-4, pour *very late antigen-4*) et VEGFR-2 (*Vascular endothelial growth factor receptor 2*) lorsque l'ectodomaine du syndécan-1 est libéré par le processus de shedding.

Des mécanismes similaires ont été retrouvés pour le syndécan-4, qui interagit avec l'intégrine  $\alpha3\beta1$  et EGFR (*Epidermal growth factor receptor*) *via* la séquence 87 – 131 (Jung *et al*, 2016), mais aussi avec l'intégrine  $\alpha5\beta1$  et CD90 (ou glycoprotéine Thy-1 qui est impliquée dans la synaptogenèse) dont la séquence d'interaction n'a pas été identifiée (Fiore *et al*, 2014).

Les fonctions biologiques et mécanismes cellulaires sous-jacents à la formation des complexes ternaires formés par le syndécan-1 et le syndécan-4 peuvent être inhibés par compétition par un peptide, nommé synstatine, mimant les régions des syndécans interagissant avec les autres membres du complexe.

Les synstatines pourraient être libérées par les métalloprotéases susceptibles de catalyser l'hydrolyse de liaisons peptidiques dans l'ectodomaine des syndécans. Le peptide 93 – 120 de l'ectodomaine du syndécan-1 pourrait être libéré par l'action des MMPs -2, -7 et -9 et de la MT1-MMP à son extrémité N-terminale (Gly 82 – Leu 83) et de la thrombine à son extrémité C-terminale (Arg 126 – Glu 127) (**Figure 122**). Le peptide 210 – 240 pourrait être libéré à son extrémité C-terminale (Gly 245 – Leu 246) par l'action de différentes sheddases : les MMPs -3, -7 et -9, la MT1-MMP et la MT3-MMP, mais aucun site n'a en revanche été décrit pour l'extrémité N-terminale du peptide. Le peptide 123 – 140 de l'ectodomaine du syndécan-2 est flanqué par des sites Asp 122 – Pro 123 et Ser 138 – Leu 139 aux extrémités N- et C-terminales, respectivement, qui sont reconnus par la MT1-MMP (**Figure 122**). Enfin, les sites de clivages des MMPs -2, -7 et -9 (Lys 105 – Leu 106) et de la MT1-MMP (Asn 139 – Ile 140) couvrent partiellement le peptide 87 – 131 de l'ectodomaine du syndécan-4.



**Figure 122 : Sites de clivage retrouvés à proximité des peptides de synstatine.** Ectodomaine : bleu. Domaine transmembranaire : jaune, domaine cytosolique conservé et variable : vert et bleu, respectivement. Rouge : synstatine. Chaînes d'héparane et de chondroïtine sulfate : lignes ondulées noires et grises, respectivement.



## Structures and interactions of syndecans

Frank Gondelaud and Sylvie Ricard-Blum

ICBMS, UMR 5246 CNRS – University Lyon 1, Univ Lyon, Villeurbanne, France

### Keywords

chondroitin sulfate; extracellular matrix; heparan sulfate; interaction networks; intrinsic disorder; proteoglycans; structure; syndecans

### Correspondence

S. Ricard-Blum, Institute of Molecular and Supramolecular of Chemistry and Biochemistry, ICBMS, UMR 5246, University Lyon 1, 43 Boulevard du 11 novembre 1918, F-69622 Villeurbanne Cedex, France  
Tel: +33(0)4 72 44 82 32  
E-mail: sylvie.ricard-blum@univ-lyon1.fr

(Received 24 November 2018, revised 4 February 2019, accepted 29 March 2019)

doi:10.1111/febs.14828

The four syndecans identified in mammals are membrane proteoglycans that play major roles in regulating cell behavior, cell signaling, and cell-matrix interactions. The membrane forms of these syndecans function as receptors and co-receptors. Their ectodomains, which are proteolytically released in the extracellular matrix by shedding, also regulate various biological processes. Apart from the cytoplasmic domain of syndecan-4, the 3D structures of syndecans are poorly characterized, which hinders our understanding of the molecular mechanisms underlying syndecan functions that are mediated by numerous interactions. This mini-review summarizes the structural data that are available for syndecans and provides a comprehensive syndecan interactome, which comprises three hundred and fifty-one partners, including those identified by the high-throughput method affinity purification-mass spectrometry. It also gives a perspective on future studies of syndecan structures and interactions, which are required to further elucidate the molecular recognition processes that mediate the biological roles of the membrane and shed forms of syndecans.

## Introduction

Proteoglycans are comprised of a core protein and of at least one glycosaminoglycan (GAG) chain covalently attached to the protein part. They exhibit a high structural diversity [1]. Forty-three proteoglycan-encoding genes have been identified and in mammals four of them encode the core protein of syndecans 1–4 (SDCs 1–4), which are single-pass, type I transmembrane proteoglycans [2,3]. In addition, syndecan-4 exists as a soluble protein isoform lacking the transmembrane and cytoplasmic domains and generated by alternative splicing [4]. The ectodomains of syndecan protein cores bear heparan sulfate chains and are often classified as heparan sulfate proteoglycans although syndecans 1 and 3 also bear two chondroitin sulfate chains [5]. Syndecans 1 and 3 are the largest members

of the syndecan family (33 and 41 kDa respectively), whereas syndecans 2 and 4 are smaller (23 and 22 kDa respectively) [6]. Syndecan-1 is found in leucocytes and is enriched in many epithelia, syndecan-2 is present in endothelia and fibroblasts, syndecan-3 in neural tissues, and syndecan-4 in most tissues [3].

Syndecans play numerous biological roles, which have been previously reviewed [3,7–14]. Their major roles are thus only briefly mentioned in this paragraph, which highlights the most recent findings. They act as receptors and co-receptors [15] and regulate cell behavior and cell signaling [3,7,16] such as Wnt signaling [17]. Syndecan-4 for example mediates cell-extracellular matrix interaction and cell-cell adhesion [18]. Syndecans regulate transient receptor potential canonical

### Abbreviations

AP-MS, affinity purification-mass spectrometry; C1 and C2, conserved regions 1 and 2; CD, cytoplasmic domain; CS, chondroitin sulfate; ED, ectodomain; ESI-MS, electrospray ionization mass spectrometry; GAG, glycosaminoglycan; HS, heparan sulfate; MMP, matrix metalloproteinase; NMR, nuclear magnetic resonance; SAXS, small-angle X-ray scattering; SDC, syndecan; TMD, transmembrane domain; TRPC, transient receptor potential canonical; V, variable region.



(TRPC) channels to control cytoplasmic calcium equilibria and cell behavior [19,20]. They play a role in lipid metabolism [21], tissue regeneration [8], angiogenesis [9] and are involved in the pathogenesis of numerous diseases, including fibrosis and heart disease [10,22], host-pathogen interactions [23], chronic inflammatory and autoimmune diseases [11], cartilage breakdown and synovial inflammation [24]. Syndecan-1 has been recently shown to regulate homeostasis of T $\gamma$  $\delta$ 17 cells and to modulate psoriasis-like skin inflammation [25], whereas syndecan-4 modulates epithelial gut barrier function and epithelial regeneration in intestinal inflammation [26]. Syndecans also play key roles in cancer [12,27–31]. Syndecan-1, induced in the lung microenvironment after mammary carcinoma cell dissemination, promotes outgrowth of metastases [32], and controls lung tumorigenesis by regulating miRNAs packaged in exosomes [33]. Although syndecans are membrane (co)receptors they may have other functions in the nucleus. Nuclear translocation of syndecan-1 inhibits the proliferation of fibrosarcoma cells, activates the TGF- $\beta$  pathway, and alters the activity of several transcription factors [34].

The molecular recognition processes underlying the numerous biological roles of the membrane and shed forms of syndecans rely on their ability to interact with ECM, membrane and intracellular partners but they have not been fully deciphered at the molecular level so far. The partners of syndecans listed in previous reviews [3,35,36] were mostly identified by low-throughput methods. We have added in this mini-review the partners of syndecans found in the adhesome dataset (<http://www.adhesome.org/>) and those recently identified by a high-throughput method (affinity purification-mass spectrometry, AP-MS) [37,38] to generate the first comprehensive syndecan interactome. We also review here the syndecan binding sites and the few 3D structures available in the Protein Data Bank [39] for syndecan domains or peptides, either alone or complexed to a partner. The ectodomains of syndecans are enriched in intrinsic disorder, and the study of their conformational flexibility will benefit from recent advances in the characterization of intrinsically disordered proteins by mass spectrometry [40–43] and computational modeling [44,45].

## Domains and structures of syndecans

### Full-length syndecans

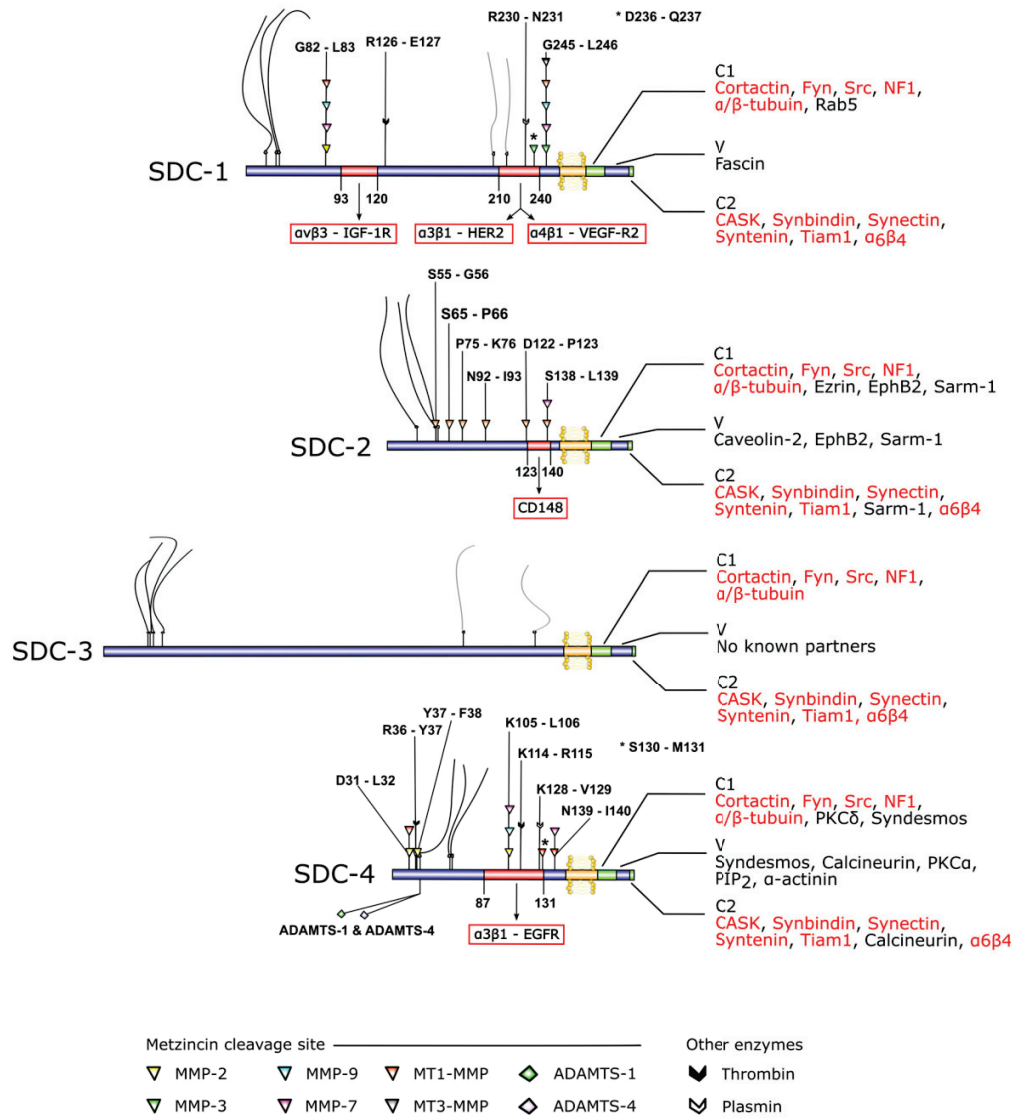
The core proteins of syndecans are composed of three domains, an extracellular domain (ectodomain, ED), a transmembrane domain (TMD) and a cytoplasmic

domain (CD) (Fig. 1). Few experimental structural data are available for syndecans 1, 2 and 4 and none for syndecan 3 (Table 1 and Fig. 2). This might be due to the fact that syndecans are enriched in intrinsic disorder [13,46]. Indeed, human syndecans are predicted to contain 86% (SDC1), 62% (SDC2), 80% (SDC3) and 55% (SDC4) of disordered amino acid residues [46].

Although no experimental 3D structure of full-length syndecans exists, models have been built for syndecans 3 and 4. The molecular model of chick syndecan-3 without glycosaminoglycan chains was generated using secondary structure predictions and the three-dimensional models of each domain built by homology [47]. Then, these models were joined and the entire structure was minimized to generate the model of full-length syndecan-3. The N-terminal, the juxtamembrane and the cytoplasmic domains are globular, whereas an extended stalk containing  $\beta$ -strands connect the juxtamembrane and the N-terminal regions [47]. Syndecan-4 was modeled as part of the endothelial cell glycocalyx by splitting the proteoglycan into three parts [48] and including GAG chains. The ectodomain of syndecan-4 with three HS chains made of various combinations of N-sulfated and non-sulfated sequences, the transmembrane dimer modeled from the dimer of glycoporphin A embedded into a lipid bilayer, and the 3D structure of the cytoplasmic domain were connected to build the model of full-length syndecan-4. There is thus a single model of a full-length syndecan encompassing GAG chains but models of the syndecans 1, 2 and 3 with GAG chains could be generated using the computational modeling approaches successfully used for syndecan-4 [48].

### Ectodomains of syndecans

The ectodomains of syndecans are released in the extracellular milieu by limited proteolysis. They contain cleavage sites for metalloproteinases of the metzincin family, namely matrix metalloproteinases (MMPs) [49], disintegrin and metalloproteinase domain-containing proteins (ADAMs) [50], ADAMs with thrombospondin domains (ADAMTSs) [51], and also for plasmin and thrombin in syndecans 1 and 4 (Fig. 1) [6,52]. Further cleavage sites of syndecans have been predicted by Bertrand and Bollman [53] but no information of the prediction method(s) used is available, and the existence of these sites has to be experimentally confirmed. Several enzymes act as sheddases, which release the ectodomains of syndecans 1–4 under a soluble form [6,52] (Fig. 1). Although there are evidences that syndecan-3 is shed by a member of



**Fig. 1.** Domain organization of human syndecans with proteolytic cleavage sites, GAG attachment sites, partners binding to the intracellular part of SDCs and receptor binding sites. Core protein: dark blue bars, transmembrane domain: yellow bars, conserved cytoplasmic regions C1 and C2: green bars, variable cytoplasmic region: blue bars, binding regions of receptors: red bars. Heparan sulfate (HS) and chondroitin sulfate (CS): curly lines in black and grey respectively. Proteolytic cleavage sites: colored triangles for metzincins, white and black arrowheads for plasmin and thrombin respectively. Binding partners of the cytoplasmic domain are indicated and those common to the four syndecans are in red. (GAG: glycosaminoglycan).

the MMP family [54], and that numerous cleavage sites by MMPs have been predicted in its sequence [53], no specific protease has been experimentally

identified so far as a syndecan-3 sheddase. The shedding of syndecans is regulated by external stimuli (i.e. molecular and biological context) [6]. Cancer [55],

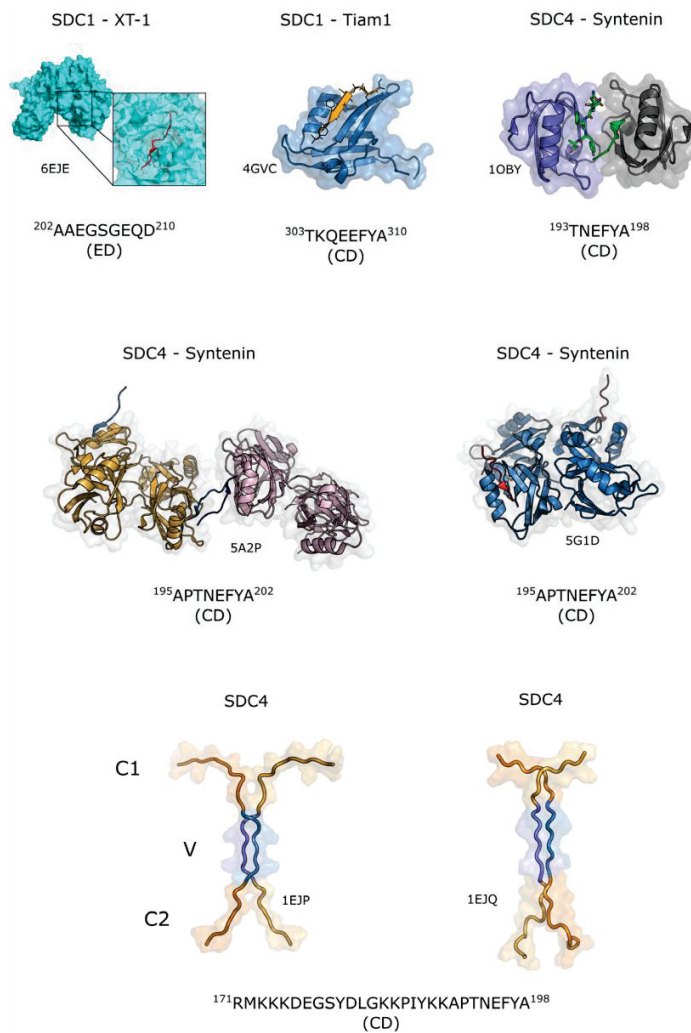
**Table 1.** Structural studies of syndecans. 3D structures deposited in the Protein Data Bank (PDB, <https://www.rcsb.org/>), UniprotKB release 2018.10) and in the Biological Magnetic Resonance Data Bank (BMRB, <http://www.bmrwisc.edu/>). The sequence of peptides used for structural studies and those found in the PDB file are indicated. CD, cytoplasmic domain; ED, ectodomain; NMR, nuclear magnetic resonance; SDCBP, syndecan-binding protein-1 (syntenin-1); TMD, transmembrane domain; X-ray, X-ray crystallography.

Syndecans (SDCs)				Method & PDB/BMRB entries	References (PMID or DOI)
Peptide sequence used	Peptide sequence observed in the structure	Domain	Partners		
<b>Syndecan-1 (P18827)</b>					
<sup>303</sup> TKQEEFYA <sup>310</sup> (Tyr <sup>309</sup> : phosphorylated)	Identical	CD (part of C2 domain)	TIAM1 (Q13009)	X-ray 4GVC	23395182
<sup>303</sup> TKQEEFYA <sup>310</sup>	Identical	CD (part of C2 domain)	TIAM1 (Q13009)	X-ray 4GVD	23395182
<sup>201</sup> PAAEGSGEQDFT <sup>212</sup>	<sup>202</sup> AAEGSGEQ <sup>210</sup>	ED	XYLT1 (Q86Y38)	X-ray 6EJE	29681470
<b>Syndecan-2 (P34741)</b>					
<b>GSRMRKKDEGSYDLGERKPSSAAYQKAPTKEFYA</b> – (the residues in bold are those found in the TMD of SDC2, 146–170)	–	CD (full-length)	–	NMR	10.5012/bkcs.2008.29.12.2449
<b>Syndecan-3 (O75056)</b> No 3D structure available in the Protein Data Bank					
<b>Syndecan-4 (P31431)</b>					
<sup>176</sup> RMKKKDEGSYDLGKKPIYKKAPTNEFYA <sup>202</sup> (rat) (Identical to residues 171–198 of human SDC4)	Identical	CD (full-length)	–	NMR 1EJP/4591	11456484
<sup>170</sup> RMKKKDEGSYDLGKKPIYKKAPTNEFYA <sup>197</sup> chicken, and rat SDC4 (residues 171–198 in human SDC4)	Identical	CD (full-length)	PIP <sub>2</sub> Not visible in the PDB file	NMR 1EJQ/4592	9582338
<sup>193</sup> TNEFYA <sup>198</sup>	1st chain: <sup>194</sup> NEFYA <sup>198</sup> 2nd chain: <sup>193</sup> TNEFYA <sup>198</sup>	CD (part of C2 domain)	SDCBP (O0J560)	X-ray 1OBY	12842047
<sup>193</sup> TNEFYA <sup>198</sup>	1st chain: <sup>193</sup> TNEFYA 198 <sup>2nd</sup> chain: <sup>194</sup> NEFYA <sup>198</sup>	CD (part of C2 domain)	SDCBP (O0J560)	X-ray 1YBO	16533050
<sup>195</sup> APTNEFYA <sup>202</sup> (rat) (Identical to residues 191–198 of human SDC4)	Identical	CD (part of C2 domain)	SDCBP (O0J560)	X-ray 5G1D	27830760
<sup>195</sup> APTNEFYA <sup>202</sup> (rat) (Identical to residues 191–198 of human SDC4)	Identical	CD (part of C2 domain)	C-terminal truncated SDCBP (O0J560)	X-ray 5A2P	27830760
<b>Zebrafish syndecan-4</b>					
<sup>174</sup> RIKKKDEGSYDLGKTPI YKKAPTEIYA <sup>201</sup> of zebrafish SDC4	Identical	CD (full-length)	–	NMR	18753130
<b>MVLAALIVGGVVGILFAVFLILLVYGGKKKM</b> (the residues in bold are those found in the TMD of human SDC4, 146–170)	–	TMD (with N- and C-terminal extra residues)	–	NMR	10.1016/j.procbio.2011.02.006
Rat syndecan-4 (P34901) (exact sequence unknown)	–	CD	NUDT16L1 (Q9BRJ7)	NMR	26100207

traumatic brain injury and heparanase enhance syndecan-1 shedding [56,57], whereas chemicals, and increased oxidative stress enhance syndecan-4 shedding in dysfunctional endothelial cells [58]. In contrast, resveratrol suppresses the shedding of syndecan-4 [59]

and the inhibition of syndecan-1 shedding induced by ischemia/reperfusion may protect against ischemic acute kidney injury [60].

The ectodomains of syndecans have a number of biological functions [61]. They may act as competitive



**Fig. 2.** 3D structures of syndecan domains and syndecans complexes. Upper panel from left to right: X-ray structures of the complex formed by the AAEGSGEQD sequence located in the ectodomain (ED) of syndecan-1 (SDC1) with xylosyltransferase 1 (6EJE), the complex formed by the TKQEEFYA sequence located in the cytoplasmic domain (CD) of syndecan-1 with the PDZ domain of Tiam-1 (4GVC), and the complex of the C-terminal sequence TNEFYA of the cytoplasmic domain of syndecan-4 (SDC4) with the PDZ domain of syntenin. Middle panel: X-ray structures of the complex formed by the APTNEFYA sequence located in the cytoplasmic domain of syndecan-4 with syntenin-1 (5G1D) or with the C-terminal truncated syntenin-1 (5A2P). Lower panel: NMR structures of the full-length cytoplasmic domain of syndecan-4 with (1EJP) or without (1EJQ) phosphatidylinositol-2 phosphate (PIP<sub>2</sub>). C1 and C2 regions: orange, V region: blue.

inhibitors and also as agonists of the membrane forms of syndecans [53]. The ectodomains of syndecans 1–4 bearing GAG chains suppress osteoclast differentiation [62], and the ectodomain of syndecan-4 promotes tubule-interstitial fibrosis [58]. Shed syndecan-1 can be translocated to the nucleus of tumor cells, where it inhibits the activity of histone acetyltransferase and delivers growth factors such as hepatocyte growth factor [63]. In addition, soluble ectodomains of syndecans serve as biomarkers for several diseases such as inflammatory diseases, sepsis, leukemia, celiac disease or systemic sclerosis (ectodomain of syndecan-1), colon cancer (shed syndecan-2), and atopic dermatitis (shed

syndecan-4) as reviewed in [53]. The soluble form of syndecan-4 is a predictor of heart failure [64–66] and the increase in soluble syndecan-1 level in cerebrospinal fluid is correlated with the severity of neuromyelitis optica, an inflammatory demyelinating disease of the central nervous system [67].

The ectodomains of syndecans regulate various biological processes both outside and inside the cells, and this might be due, at least partly, to their structural plasticity. Indeed, the comparison of the sequences of the ectodomains of syndecans 1–4 in 25 animals has shown that they are enriched in glycine, alanine, glutamic acid and proline residues and depleted in

order-promoting amino acid residues [13]. The high amount of intrinsic disorder might explain why no experimental 3D structure is available for the ectodomains of syndecans in absence of binding partners. The only X-ray structure available for a syndecan ectodomain comprises a dodecapeptide of syndecan-1 in complex with human xylosyltransferase 1 [68] (Table 1 and Fig. 2). A 3D model of the syndecan-4 ectodomain has been built and shown to partially fold into  $\alpha$ -helices at its N- and C-termini, whereas the disordered region containing the attachment sites of HS chains remains flexible [48].

In conclusion, no structural data are available for the entire ectodomains of syndecans 1–4. Since they contain a significant amount of disorder, which would likely prevent their crystallization, low-resolution structural techniques such as small-angle X-ray scattering (SAXS) should be favored. Intrinsically disordered proteins can fold upon binding to their partners, and crystallography studies of the syndecan ectodomains in presence of partners interacting with their protein part would give new insights on their ability to fold upon binding, differently or not depending on their partners, and on the extent of their folding. Mass spectrometry (MS) approaches such as native electrospray ionization (ESI)-MS and native ion mobility MS would be useful to evaluate the conformational space of the syndecan ectodomains, dissect their conformational ensembles and capture compact conformations from their charge-state distributions [40,42,43,69]. In addition, 3D models of the ectodomains of syndecans could be generated using computational modeling approaches used for intrinsically disordered proteins by coarse-grained molecular dynamics simulations restrained by SAXS data [45] or by approaches based on the dock-and-coalesce mechanism of intrinsically disordered protein binding (IDP-LZerD [70]). The 3D structures and/or models of syndecans would be used for docking experiments with their partners to study the molecular recognition processes associated with their biological roles.

Last, to get a global structure of syndecans it will be necessary to investigate the influence, if any, of GAGs chains on the conformations of the syndecan ectodomains. A single model has been built for a syndecan ectodomain with three HS chains [48] and no experimental data are available. Structural analyses by SAXS of the ectodomains of syndecans 1–4, expressed in eukaryotic cells allowing the synthesis of GAG chains, would provide new insights on the role of GAGs on the flexibility, folding and conformation of the syndecan ectodomains when bound to partners. Another challenge would be to determine the effect on

the syndecan core proteins of GAG-modifying enzymes such as sulfatases and heparanase.

### Transmembrane domains of syndecans

The transmembrane domains of syndecans modulate intracellular signaling by regulating the oligomeric status of the cytoplasmic domain [71]. Their self-association, which is crucial for their functions is driven by the consensus sequence GXXXG [72]. A conserved phenylalanine residue of the transmembrane domain regulates heteromeric interactions of syndecans [73], which form homodimers and heterotrimers [72,73].

No data are available either for transmembrane domains of syndecans -1, -2 and -3, whereas nuclear magnetic resonance (NMR) and circular dichroism experiments have been carried out on the transmembrane domain of syndecan-4. It forms an asymmetrical  $\alpha$ -helical dimer in micelles [74] but the corresponding structure is not available in the Protein Data Bank (Table 1). The 3D structure or models of syndecan transmembrane homodimers and heterodimers are required to decipher the molecular mechanisms of the dimerization. This process, which is of major importance for the regulation of syndecan functions [75,76], has not been fully characterized so far although several amino acids critical for the dimerization have been identified [73].

### Cytoplasmic domains of syndecans

The cytoplasmic domains of syndecans are comprised of two constant, conserved, regions, C1 and C2, flanking a variable region V containing two conserved amino acid residues (Tyr and Lys), which promotes syndecan-specific signaling [3,7].

No structural data are available for the cytoplasmic domains of syndecans 1 and 3. The cytoplasmic domain of syndecan-2 has been analyzed by NMR [77] but no model of the domain was generated and the data are available neither in the PDB nor in the Biological Magnetic Resonance Data Bank (BRMB). Several structures of the cytoplasmic domain of syndecan-4, either alone or in complex with a partner, have been determined by NMR (Table 1). The cytoplasmic domain of syndecan-4 is a compact intertwined dimer with a symmetric clamp shape in the variable region and two parallel  $\beta$ -strands forming a cavity, whereas the C1 region appears to be disordered within the dimeric cytoplasmic domain [78,79]. This structure is conserved in the cytoplasmic domain of zebrafish syndecan-4 although two of its amino acid residues are different [80,81]. The X-ray structure of the syntenin-

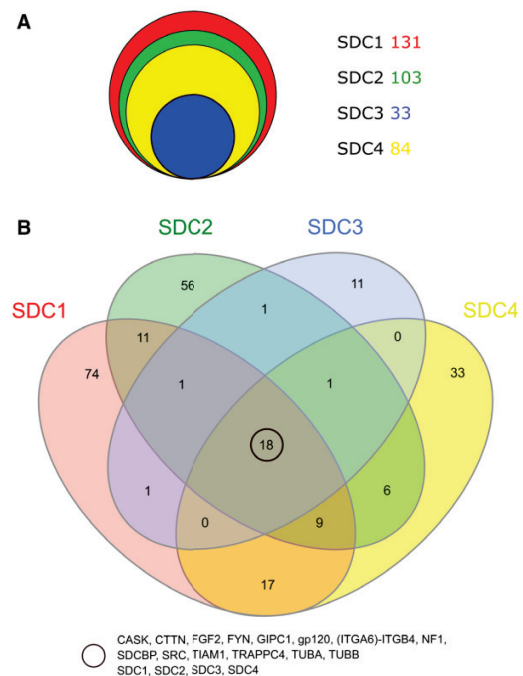
1/syndecan-4 complex has shown that syntenin-1 forms a symmetrical pair of dimers anchored by a syndecan-4 dimer. The cytoplasmic domain of syndecan-4 wraps the syntenin-1 dimers [82]. The binding of syndesmos to rat syndecan-4 induces changes in the C1 and V regions, and the amino acid residues of rat syndecan-4 directly involved in the interaction are Lys177, Lys178, Lys179, Gly182, and Ser183 in the C1 region, and Lys193 and Lys194 in the variable region [83]. The cytoplasmic domain of syndecan-4 also undergoes a shape change upon binding to phosphatidylinositol 4,5 biphosphate [78]. Two 3D structures of the complexes formed by regions of the cytoplasmic domains of syndecans 1 and 4 encompassing the EFYA tetrapeptide and PDZ protein domains are also available (Table 1 and Fig. 2). In contrast, the TKEFYA sequence of syndecan-2 does not bind to the PDZ-like domain of synbindin although NMR titration experiments showed that the cytoplasmic domain, comprised of 32-amino acid residues, interacts with it [84]. No 3D structure of the complex or of the cytoplasmic domain of syndecan-2 is reported in the article or is available in the Protein Data Bank. Only the solution structure of the PDZ domain of synbindin (2JSN) is deposited in the PDB [84]. Whether the cytoplasmic domains of syndecans 1, 2 and 3 are able to form twisted clamp dimers remains to be determined. However, their 3D structures are needed to decipher the molecular mechanisms of signaling mediated by the interactions between the syndecans and their partners as previously done for syndecan-4-syntenin-1 and syndecan-4-phosphatidylinositol 4,5 biphosphate interactions [78,82].

## The syndecan interaction networks

### The global interaction network of syndecans

We report here the first comprehensive interaction network of syndecans 1–4 by combining interaction data from previous reviews [35,36], manual curation of the literature, queries of interaction databases, namely MatrixDB [85,86] and IntAct [87], the large scale BioPlex 2.0 interaction dataset [38] obtained by high-throughput affinity purification mass spectrometry (AP-MS), and from the adhesome (<http://www.adhesome.org/>) (Figs 3 and 4). The entire dataset of syndecan interactions, including the names of their binding partners, the location of their binding sites on syndecans (i.e. core protein and/or glycosaminoglycan chains) when available, and the references reporting experimental evidences of these interactions, is provided in Table S1.

The four syndecans have 351 partners in total. Syndecans 1, 2 and 4 have a high number of binding partners

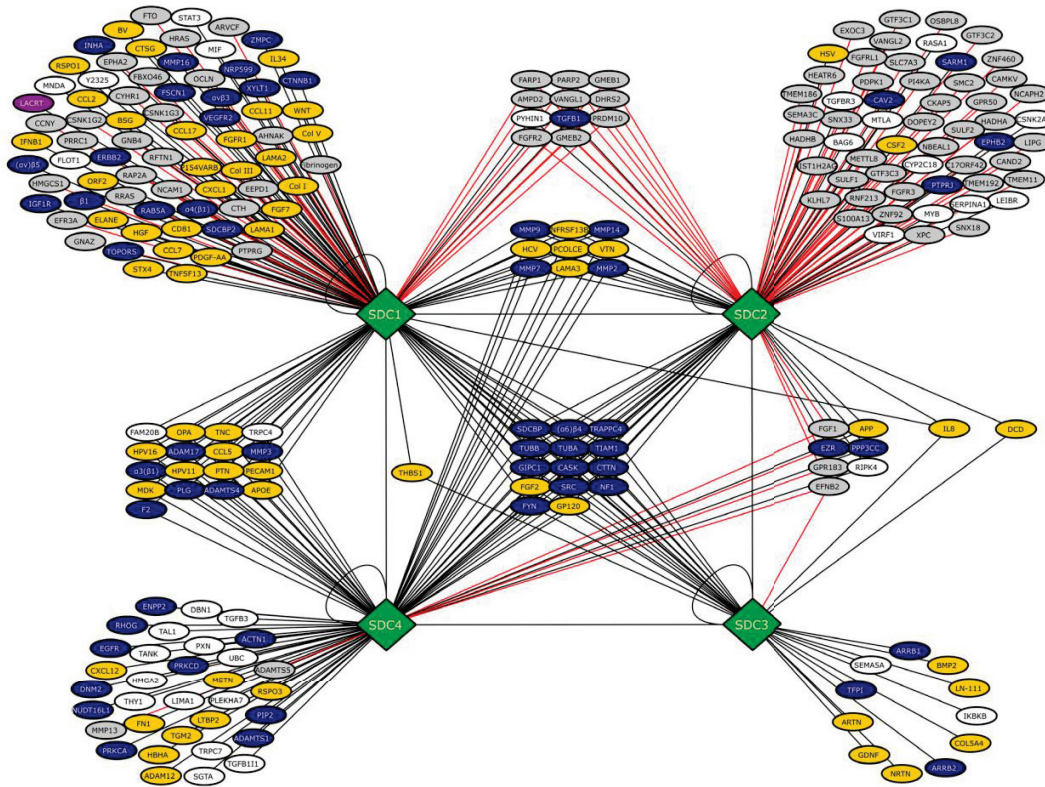


**Fig. 3.** Partners of syndecans 1–4. A: total number of partners of syndecans 1–4, B: number of specific partners of syndecans 1–4 and number of partners shared by two, three or four syndecans. The names of the partners common to four syndecans are listed below the Venn diagram built with InteractiVenn (<http://www.interactiVenn.net/>).

(131, 103 and 84 respectively), whereas syndecan-3 has only 33 partners (Fig. 3A and Table S1). Syndecan-3 has a more restricted distribution than the other syndecans, which might partly explain the lower number of its partners. There might be also a research bias. Indeed, syndecans 1 and 4 are the most studied syndecans with 670 and 304 publications in PubMed respectively (queries made with syndecan-1 and syndecan-4 in the title of the publication), whereas there are 123 and 56 publications on syndecans 2 and 3 respectively.

### The consensus interaction network of syndecans

The partners specific to each syndecan and those shared by two, three and four syndecans are listed in Table S2. The four syndecans share 18 binding proteins corresponding to the syndecan consensus interactome, including the four syndecans, which are able to form homo- and hetero-dimers (Fig. 3B and Table S2). In addition to the four syndecans, the consensus interactome comprises 14 proteins, namely



**Fig. 4.** The interaction network of syndecans 1–4. SDCs are represented as green diamonds. Black lines: direct interactions. Red lines: interactions detected by affinity purification-mass spectrometry (Bioplex 2.0 dataset) and listed in the adhesome (<http://www.adhesome.org/>), which may be either direct or indirect. Partners binding to syndecan GAG chains: yellow, partners binding to syndecan protein core: blue, lacritin binding to a hybrid site made of GAG saccharides and protein core: purple, partners without information on binding site: white, partners of the BioPlex 2.0 dataset and adhesome: grey.

protein kinases (the peripheral plasma membrane protein CASK, the tyrosine-protein kinase Fyn, and the proto-oncogene tyrosine-protein kinase Src), proteins involved in actin cytoskeletal organization (cortactin, syntenin-1 and neurofibromin) and in microtubule formation (tubulins  $\alpha$  and  $\beta$ ), a protein playing a role in vesicular transport from endoplasmic reticulum to Golgi (synbindin or TRAPPC4), and proteins involved in signaling (synectin or PDZ domain-containing protein GIPC1 involved in G protein-coupled receptor signaling pathway and TIAM1 or T-lymphoma invasion and metastasis-inducing protein 1). TIAM1 connects extracellular signals to cytoskeletal activities like integrin  $\alpha 6 \beta 4$ , another partner common to the four syndecans, which mediates bidirectional signaling between the cytoskeleton and the extracellular matrix. In addition, all syndecans bind to fibroblast growth

factor-2 (FGF2) and to gp120, a glycoprotein from the human immunodeficiency virus. Most partners interacting with the four syndecans are intracellular or membrane proteins involved in signaling and the organization of actin cytoskeleton.

**The specific networks of syndecans 1–4**

The cellular location and the molecular functions of the partners specific to each syndecan together with the biological processes and the biological pathways they are involved in have been analyzed using the Functional Enrichment analysis tool FunRich [88] we have used to analyze the interaction network of the extracellular matrix with the amyloid  $\beta$ -42 peptide [89]. The annotations used by the FunRich tool, discussed below, are put into quotation marks.

The 74 proteins interacting only with syndecan-1 (Fig. 4 and Table S2) are mostly membrane and extracellular proteins involved in 'cell communication' (43%) and 'cell transduction' (43%). 77% of them are involved in 'integrin family cell surface interactions' and 68% in 'glypican pathway'. Syndecan-1 is the only syndecan to interact with fibrillar collagens I, III and V, and it binds to more pathogens and pathogen proteins than the other syndecans (Fig. 4 and Table S1). It is shed by the metalloproteinase virulence factor ZmpC of *Streptococcus pneumoniae* and plays a role in bacterial infections [90].

No significant feature is observed for the 56 specific partners of syndecan-2 (Fig. 4 and Table S2), except that 25% of them are associated with lysosomes. 75% of the 11 specific partners of syndecan-3 (Fig. 4 and Table S2) are extracellular, 25% of them are 'extracellular matrix constituents' and 25% are found in basement membranes. Furthermore, 67% are involved in 'cell communication' and 67% in 'cell transduction', and 58% of them are associated with axon guidance. Syndecan-3 interacts with the extracellular sulfatase Sulf-1 and Sulf2, which have endoglucosamine-6-sulfatase activity, and modulate heparan sulfate sulfation. The 33 specific partners of syndecan-4 (Fig. 4 and Table S2) are extracellular (42%) or associated with exosomes (39%). 39% are involved in 'cell communication', 45% in 'cell transduction', and 71% in 'integrin family cell surface interactions'.

In conclusion, a number of the partners of syndecans 1, 2, and 4 are associated with signal transduction and cell communication, and interact with integrins for the partners of syndecans 1 and 4. A significant percentage of the partners of syndecans 1 and 4 (43% and 39% respectively) are associated with exosomes. The exosomes are involved in cell-cell communication [91], and play crucial roles in cancer development [92]. They may thus be associated with syndecans 1 and 4 in these processes.

### Syndecan binding sites

According to the location of their binding sites on syndecans (Table S1), the vast majority of syndecan partners interacts with their GAG chains (71% for SDC1, 67% for SDC2, 91% for SDC3 and 68% for SDC4). Lacritin, a partner of syndecan-1, requires the combination of HS and CS saccharides and of the <sup>48</sup>GAGAL<sup>52</sup> sequence of the syndecan ectodomain to bind to syndecan-1 [93]. Structural differences of HS and CS chains dictate different unbinding pathways for syndecan complexes as shown by single-molecule force spectroscopy used to rupture SDC1 and SDC4

complexes with monoclonal antibodies. The unbinding pathways comprises two energy barriers for syndecan-1, whereas syndecan-4 unbinds by crossing over one energy barrier [94]. Furthermore, the same protein may bind either to the protein core or to the GAG chains depending on the syndecan such as ADAMTS-4, which interacts with heparan sulfate chains of SDC1 but binds to the core protein of SDC4 [95,96]. Integrins, metalloproteinases and few other proteins (VEGFR2, HER2, and IGF1R for syndecan-1, TGF- $\beta$  and PTRJ receptor for syndecan-2, and TFPI for syndecan-3) directly interact with syndecan core proteins (Fig. 4 and Table S1). In conclusion, most interactions established by syndecans in the extracellular matrix or at the cell surface are mediated by their GAG chains. Only few interactions occurring in the extracellular matrix or at the cell surface are mediated by the syndecan core proteins, most of them being receptors and proteinases.

### Ternary complexes formed by syndecans

Besides the binary interactions included in the global interaction network, the syndecans are engaged in ternary complexes such as those formed by the ectodomain of syndecan-1 with  $\alpha 3\beta 1$  integrin and HER2 (receptor tyrosine-protein kinase erbB-2), and with either  $\alpha \nu \beta 3$  integrin or  $\alpha \nu \beta 5$  integrin and IGF1R [97]. The shed ectodomain of syndecan-1 also forms a ternary complex with VEGFR2 and  $\alpha 4\beta 1$  integrin [98]. In addition, the ectodomain of syndecan-4 forms a complex with the integrin  $\alpha 3\beta 1$  and EGFR [99], and the formation of a ternary complex comprising syndecan-4, integrin  $\alpha 5\beta 1$  and Thy-1 (CD90) has been reported [100]. The structural characterization of these ternary complexes would provide new clues of the regulation at the molecular level of cell signaling by syndecans.

### Conclusion and perspectives

Syndecans -1 and -4 are the most frequently studied syndecans, both at the structural and interaction levels. In addition, they share the highest number of partners among the syndecans. Less data are available for syndecans 2 and 3, although fifty-two new partners of syndecan-2 have been recently identified by affinity purification-mass spectrometry [37,38]. Syndecan-3 has been even less characterized, and the enzyme(s) releasing its shed ectodomain remain(s) to be identified. Regarding the structures of full-length syndecans, the major challenge is to collect experimental structural data and/or to generate reliable 3D models for the individual domains of syndecans, and then to connect



them by integrating sequence-to-conformation relationships of disordered regions tethered to folded domains [101]. Native ESI-MS will be also useful to characterize the conformational transitions of the membrane forms of syndecans [102]. In parallel, coarse-grain models of heparan and chondroitin sulfate chains of syndecans should be generated by previously developed methods [103,104]. GAG models could also be generated from conformational maps using the bioinformatics pipeline we have recently developed [105]. It is freely available on the web site of the extracellular matrix interaction database (MatrixDB, <http://matrixdb.univ-lyon1.fr/>), which has been recently updated with a focus on GAGs and GAG interactions [86].

The comprehensive interaction network of syndecans 1–4 combines low- and high-throughput interaction data, direct and potentially indirect interactions, and warrants more in-depth analyses. Few data are available on GAG sequences of syndecans interacting with their protein partners. The identification of GAG chemical groups involved in the binding of syndecan partners, their spatial distribution along the GAG sequences and a 3D model of these sequences would help to decipher the molecular mechanism(s) of actions of syndecans and to understand how their biological effects are triggered by GAG-protein interactions in the extracellular matrix and at the cell surface. The global interactome of syndecans will provide new clues on how the syndecans are connected, and how they regulate biological processes in a concerted and/or in a specific manner. It will be used for identifying protein-protein interactions, which are crucial for a specific biological process, and could be targeted for a therapeutic purpose. Furthermore, it provides a scaffold to be contextualized with *i*) transcriptomics and quantitative proteomics data to build tissue- or cell-specific syndecan networks, *ii*) binding sites to determine the interactions occurring simultaneously and those that compete with each other, and *iii*) the amount of predicted intrinsic disorder to identify the zones of structural flexibility in the network, which are prone to rewiring of interactions depending on the biological context (e.g. changes in growth factor concentration, and upregulation or downregulation of gene expression). Kinetics parameters (association and dissociation rates) will be added when available to identify transient and stable interactions, whereas the integration of equilibrium dissociation constants will discriminate weak from strong interactions. The integration of all these data will generate a dynamic syndecan interaction network mimicking the *in vivo* situation. The ultimate goal is to build a 3D syndecan interactome.

## Acknowledgements

We acknowledge Sylvain D. Vallet (ICBMS, UMR 5246, Villeurbanne, France) for his valuable help in handling the BioPlex 2.0 and the adhesome datasets, in drafting Fig. 4 and for his critical reading of the revised manuscript. MatrixDB, the database used among other sources, to build the syndecan interactome, is supported by the Fondation pour la Recherche Médicale (grant DBI20141231336 to SRB), and by the Institut Français de Bioinformatique (ANR-11-INS-0013, grant Glyco-matrixDB to SRB).

## Conflicts of interest

The authors declare no conflict of interest.

## Author contributions

FG and SRB performed literature searches and database queries and designed the tables. FG drafted the manuscript and the figures, and revised the Tables and the figures. SRB wrote the final version of the manuscript, revised the manuscript and contributed to the figure design.

## References

- 1 Karamanos NK, Piperigkou Z, Theocharis AD, Watanabe H, Franchi M, Baud S, Brézillon S, Götte M, Passi A, Vigetti D *et al.* (2018) Proteoglycan chemical diversity drives multifunctional cell regulation and therapeutics. *Chem Rev* **118**, 9152–9232.
- 2 Iozzo RV & Schaefer L (2015) Proteoglycan form and function: a comprehensive nomenclature of proteoglycans. *Matrix Biol* **42**, 11–55.
- 3 Couchman JR, Gopal S, Lim HC, Nørgaard S & Multhaupt HAB (2015) Fell-Muir Lecture: Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. *Int J Exp Pathol* **96**, 1–10.
- 4 Xing Y, Xu Q & Lee C (2003) Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS Lett* **555**, 572–578.
- 5 Couchman JR (2010) Transmembrane signaling proteoglycans. *Annu Rev Cell Dev Biol* **26**, 89–114.
- 6 Manon-Jensen T, Itoh Y & Couchman JR (2010) Proteoglycans in health and disease: the multiple roles of syndecan shedding. *FEBS J* **277**, 3876–3889.
- 7 Afratis NA, Nikitovic D, Multhaupt HAB, Theocharis AD, Couchman JR & Karamanos NK (2017) Syndecans – key regulators of cell signaling and biological functions. *FEBS J* **284**, 27–41.

- 8 Chung H, Multhaupt HAB, Oh E-S & Couchman JR (2016) Minireview: Syndecans and their crucial roles during tissue regeneration. *FEBS Lett* **590**, 2408–2417.
- 9 De Rossi G & Whiteford JR (2014) Syndecans in angiogenesis and endothelial cell biology. *Biochem Soc Trans* **42**, 1643–1646.
- 10 Lunde IG, Herum KM, Carlson CC & Christensen G (2016) Syndecans in heart fibrosis. *Cell Tissue Res* **365**, 539–552.
- 11 Agere SA, Kim EY, Akhtar N & Ahmed S (2018) Syndecans in chronic inflammatory and autoimmune diseases: pathological insights and therapeutic opportunities. *J Cell Physiol* **233**, 6346–6358.
- 12 Cheng B, Montmasson M, Terradot L & Rousselle P (2016) Syndecans as cell surface receptors in cancer biology. A focus on their interaction with PDZ domain proteins. *Front Pharmacol* **7**, 10.
- 13 Leonova EI & Galzitskaya OV (2015) Cell communication using intrinsically disordered proteins: what can syndecans say? *J Biomol Struct Dyn* **33**, 1037–1050.
- 14 Gharbaran R (2016) Insights into the molecular roles of heparan sulfate proteoglycans (HSPGs-syndecans) in autocrine and paracrine growth factor signaling in the pathogenesis of Hodgkin's lymphoma. *Tumour Biol* **37**, 11573–11588.
- 15 Choi Y, Chung H, Jung H, Couchman JR & Oh E-S (2011) Syndecans as cell surface receptors: unique structure equates with functional diversity. *Matrix Biol* **30**, 93–99.
- 16 Elfénbein A & Simons M (2013) Syndecan-4 signaling at a glance. *J Cell Sci* **126**, 3799–3804.
- 17 Pataki CA, Couchman JR & Brábek J (2015) Wnt signaling cascades and the roles of Syndecan proteoglycans. *J Histochem Cytochem* **63**, 465–480.
- 18 Gopal S, Multhaupt HAB, Pocock R & Couchman JR (2017) Cell-extracellular matrix and cell-cell adhesion are linked by syndecan-4. *Matrix Biol* **60–61**, 57–69.
- 19 Gopal S, Søgaard P, Multhaupt HAB, Pataki C, Okina E, Xian X, Pedersen ME, Stevens T, Griesbeck O, Park PW *et al.* (2015) Transmembrane proteoglycans control stretch-activated channels to set cytosolic calcium levels. *J Cell Biol* **210**, 1199–1211.
- 20 Mitsou I, Multhaupt HAB & Couchman JR (2017) Proteoglycans, ion channels and cell-matrix adhesion. *Biochem J* **474**, 1965–1979.
- 21 Leonova EI & Galzitskaya OV (2015) Role of Syndecans in lipid metabolism and human diseases. *Adv Exp Med Biol* **855**, 241–258.
- 22 Christensen G, Herum KM & Lunde IG (2018) Sweet, yet underappreciated: proteoglycans and extracellular matrix remodeling in heart disease. *Matrix Biol* **2019** 75–76, 286–299.
- 23 García B, Merayo-Llodes J, Martín C, Alcalde I, Quirós LM & Vazquez F (2016) Surface proteoglycans as mediators in bacterial pathogens infections. *Front Microbiol* **7**, 220.
- 24 Pap T & Bertrand J (2013) Syndecans in cartilage breakdown and synovial inflammation. *Nat Rev Rheumatol* **9**, 43–55.
- 25 Jaiswal AK, Sadasivam M, Archer NK, Miller RJ, Dillen CA, Ravipati A, Park PW, Chakravarti S, Miller LS & Hamad ARA (2018) Syndecan-1 regulates psoriasiform dermatitis by controlling homeostasis of IL-17-producing  $\gamma\delta$  T cells. *J Immunol* **201**, 1651–1661.
- 26 Fröhling M, Tepas P, Intemann J, Sambale M, Sherwood J, Paruzel P, Tiemeyer N-M, Nowacki TM, Brückner M, Mennigen R *et al.* (2018) Syndecan-4 modulates epithelial gut barrier function and epithelial regeneration in experimental colitis. *Inflamm Bowel Dis* **24**, 2579–2589.
- 27 Yoneda A, Lendorf ME, Couchman JR & Multhaupt HAB (2012) Breast and ovarian cancers: a survey and possible roles for the cell surface heparan sulfate proteoglycans. *J Histochem Cytochem* **60**, 9–21.
- 28 Theocharis AD, Skandalis SS, Neill T, Multhaupt HAB, Hubo M, Frey H, Gopal S, Gomes A, Afratis N, Lim HC *et al.* (2015) Insights into the key roles of proteoglycans in breast cancer biology and translational medicine. *Biochim Biophys Acta* **1855**, 276–300.
- 29 Couchman JR, Multhaupt H & Sanderson RD (2016) Recent insights into cell surface heparan sulphate proteoglycans and cancer. *F1000Res* **5**, 1541.
- 30 Multhaupt HAB, Leitinger B, Gullberg D & Couchman JR (2016) Extracellular matrix component signaling in cancer. *Adv Drug Deliv Rev* **97**, 28–40.
- 31 Theocharis AD & Karamanos NK (2019) Proteoglycans remodeling in cancer: underlying molecular mechanisms. *Matrix Biol* **75–76**, 220–259.
- 32 Chute C, Yang X, Meyer K, Yang N, O'Neil K, Kasza I, Eliceiri K, Alexander C & Friedl A (2018) Syndecan-1 induction in lung microenvironment supports the establishment of breast tumor metastases. *Breast Cancer Res* **20**, 66.
- 33 Parimon T, Brauer R, Schlesinger SY, Xie T, Jiang D, Ge L, Huang Y, Birkland TP, Parks WC, Habel DM *et al.* (2018) Syndecan-1 controls lung tumorigenesis by regulating miRNAs packaged in exosomes. *Am J Pathol* **188**, 1094–1103.
- 34 Szatmári T, Mundt F, Kumar-Singh A, Möbus L, Ötvös R, Hjerpe A & Dobra K (2017) Molecular targets and signaling pathways regulated by nuclear translocation of syndecan-1. *BMC Cell Biol* **18**, 34.
- 35 Roper JA, Williamson RC & Bass MD (2012) Syndecan and integrin interactomes: large complexes in small spaces. *Curr Opin Struct Biol* **22**, 583–590.
- 36 Stepp MA, Pal-Ghosh S, Tadvalkar G & Pajoohesh-Ganji A (2015) Syndecan-1 and its expanding list of contacts. *Adv Wound Care (New Rochelle)* **4**, 235–249.

- 37 Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K *et al.* (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440.
- 38 Schweppe DK, Huttlin EL, Harper JW & Gygi SP (2018) BioPlex display: an interactive suite for large-scale AP-MS protein-protein interaction data. *J Proteome Res* **17**, 722–726.
- 39 Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Dalenberg K, Duarte JM, Dutta S *et al.* (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res* **47**, D464–D474.
- 40 Testa L, Brocca S, Santambrogio C, D'Urzo A, Habchi J, Longhi S, Uversky VN & Grandori R (2013) Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. *Intrinsically Disord Proteins* **1**, e25068.
- 41 Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017) Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr Opin Struct Biol* **42**, 15–23.
- 42 Natalello A, Santambrogio C & Grandori R (2017) Are charge-state distributions a reliable tool describing molecular ensembles of intrinsically disordered proteins by native MS? *J Am Soc Mass Spectrom* **28**, 21–28.
- 43 Santambrogio C, Natalello A, Brocca S, Ponzini E & Grandori R (2018) Conformational characterization and classification of intrinsically disordered proteins by native mass spectrometry and charge-state distribution analysis. *Proteomics* **19**, e1800060.
- 44 Cragnell C, Rieloff E & Skepö M (2018) Utilizing coarse-grained modeling and Monte Carlo simulations to evaluate the conformational ensemble of intrinsically disordered proteins and regions. *J Mol Biol* **430**, 2478–2492.
- 45 Vallet SD, Miele AE, Uciechowska-Kaczmarzyk U, Liwo A, Duclos B, Samsonov SA & Ricard-Blum S (2018) Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. *Sci Rep* **8**, 11768.
- 46 Peysselon F, Xue B, Uversky VN & Ricard-Blum S (2011) Intrinsic disorder of the extracellular matrix. *Mol BioSyst* **7**, 3353–3365.
- 47 Kirsch T, Koyama E, Liu M, Golub EE & Pacifici M (2002) Syndecan-3 is a selective regulator of chondrocyte proliferation. *J Biol Chem* **277**, 42171–42177.
- 48 Cruz-Chu ER, Malafeev A, Pajarskas T, Pivkin IV & Koumoutsakos P (2014) Structure and response to flow of the glycocalyx layer. *Biophys J* **106**, 232–243.
- 49 Apte SS & Parks WC (2015) Metalloproteinases: a parade of functions in matrix biology and an outlook for the future. *Matrix Biol* **44–46**, 1–6.
- 50 Weber S & Saftig P (2012) Ectodomain shedding and ADAMs in development. *Development* **139**, 3693–3709.
- 51 Dubail J & Apte SS (2015) Insights on ADAMTS proteases and ADAMTS-like proteins from mammalian genetics. *Matrix Biol* **44–46**, 24–37.
- 52 Nam EJ & Park PW (2012) Shedding of cell membrane-bound proteoglycans. *Methods Mol Biol* **836**, 291–305.
- 53 Bertrand J & Bollmann M (2018) Soluble syndecans: biomarkers for diseases and therapeutic options. *Br J Pharmacol* **176**, 67–81.
- 54 Asundi VK, Erdman R, Stahl RC & Carey DJ (2003) Matrix metalloproteinase-dependent shedding of syndecan-3, a transmembrane heparan sulfate proteoglycan, in Schwann cells. *J Neurosci Res* **73**, 593–602.
- 55 Choi S, Lee H, Choi JR & Oh ES (2010) Shedding: towards a new paradigm of syndecan function in cancer. *BMB Rep* **43**, 305–310.
- 56 Sanderson RD, Elkin M, Rapraeger AC, Ilan N & Vlodavsky I (2017) Heparanase regulation of cancer, autophagy and inflammation: new mechanisms and targets for therapy. *FEBS J* **284**, 42–55.
- 57 Gonzalez Rodriguez E, Cardenas JC, Cox CS, Kitagawa RS, Stensballe J, Holcomb JB, Johansson PI & Wade CE (2018) Traumatic brain injury is associated with increased syndecan-1 shedding in severely injured patients. *Scand J Trauma Resusc Emerg Med* **26**, 102.
- 58 Lipphardt M, Song JW, Ratliff BB, Dihazi H, Müller GA & Goligorsky MS (2018) Endothelial dysfunction is a superinducer of syndecan-4: fibrogenic role of its ectodomain. *Am J Physiol Heart Circ Physiol* **314**, H484–H496.
- 59 Wu H, Chen Z, Chen J-Z, Xie J & Xu B (2018) Resveratrol improves tube formation in AGE-induced late endothelial progenitor cells by suppressing Syndecan-4 shedding. *Oxid Med Cell Longev* **2018**, 9045976.
- 60 Lu Z, Song N, Shen B, Xu X, Fang Y, Shi Y, Ning Y, Hu J, Dai Y, Ding X *et al.* (2018) Syndecan-1 shedding inhibition to protect against ischemic acute kidney injury through HGF target signaling pathway. *Transplantation* **102**, e331–e344.
- 61 De Rossi G & Whiteford JR (2013) Novel insight into the biological functions of syndecan ectodomain core proteins. *BioFactors* **39**, 374–382.
- 62 Kim J-M, Lee K, Kim MY, Shin H-I & Jeong D (2018) Suppressive effect of syndecan ectodomains and N-desulfated heparins on osteoclastogenesis via direct binding to macrophage-colony stimulating factor. *Cell Death Dis* **9**, 1119.

- 63 Stewart MD, Ramani VC & Sanderson RD (2015) Shed syndecan-1 translocates to the nucleus of cells delivering growth factors and inhibiting histone acetylation: a novel mechanism of tumor-host cross-talk. *J Biol Chem* **290**, 941–949.
- 64 Takahashi R, Negishi K, Watanabe A, Arai M, Naganuma F, Ohyama Y & Kurabayashi M (2011) Serum syndecan-4 is a novel biomarker for patients with chronic heart failure. *J Cardiol* **57**, 325–332.
- 65 Bielecka-Dabrowa A, von Haehling S, Aronow WS, Ahmed MI, Rysz J & Banach M (2013) Heart failure biomarkers in patients with dilated cardiomyopathy. *Int J Cardiol* **168**, 2404–2410.
- 66 Bielecka-Dabrowa A, Gluba-Brzózka A, Michalska-Kasiczak M, Misztal M, Rysz J & Banach M (2015) The multi-biomarker approach for heart failure in patients with hypertension. *Int J Mol Sci* **16**, 10715–10733.
- 67 Pei S, Zheng D, Wang Z, Hu X, Pan S & Wang H (2018) Elevated soluble syndecan-1 levels in neuromyelitis optica are associated with disease severity. *Cytokine* **111**, 140–145.
- 68 Briggs DC & Hohenester E (2018) Structural basis for the initiation of glycosaminoglycan biosynthesis by human xylosyltransferase 1. *Structure* **26**, 801–809.e3.
- 69 Konijnenberg A, Butterer A & Sobott F (2013) Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim Biophys Acta* **1834**, 1239–1256.
- 70 Peterson LX, Roy A, Christoffer C, Terashi G & Kihara D (2017) Modeling disordered protein interactions from biophysical principles. *PLoS Comput Biol* **13**, e1005485.
- 71 Jang B, Jung H, Hong H & Oh E-S (2018) Syndecan transmembrane domain modulates intracellular signaling by regulating the oligomeric status of the cytoplasmic domain. *Cell Signal* **52**, 121–126.
- 72 Dews IC & Mackenzie KR (2007) Transmembrane domains of the syndecan family of growth factor coreceptors display a hierarchy of homotypic and heterotypic interactions. *Proc Natl Acad Sci USA* **104**, 20782–20787.
- 73 Kwon M-J, Park J, Jang S, Eom C-Y & Oh E-S (2016) The conserved phenylalanine in the transmembrane domain enhances heteromeric interactions of Syndecans. *J Biol Chem* **291**, 872–881.
- 74 Park T-J, Lee M-H, Kim J-S & Kim Y (2011) Recombinant expression, purification, and characterization of transmembrane domain of syndecan-4. *Process Biochem* **46**, 1166–1171.
- 75 Choi S, Lee E, Kwon S, Park H, Yi JY, Kim S, Han I-O, Yun Y & Oh E-S (2005) Transmembrane domain-induced oligomerization is crucial for the functions of syndecan-2 and syndecan-4. *J Biol Chem* **280**, 42573–42579.
- 76 Choi Y, Kwon M-J, Lim Y, Yun J-H, Lee W & Oh E-S (2015) Trans-regulation of Syndecan functions by hetero-oligomerization. *J Biol Chem* **290**, 16943–16953.
- 77 Chae YK & Weontae L (2008) Recombinant expression and purification of cytoplasmic domain of Syndecan-2 proteoglycan. *Bull Korean Chem Soc* **29**, 2449–2452.
- 78 Lee D, Oh ES, Woods A, Couchman JR & Lee W (1998) Solution structure of a syndecan-4 cytoplasmic domain and its interaction with phosphatidylinositol 4,5-bisphosphate. *J Biol Chem* **273**, 13022–13029.
- 79 Shin J, Lee W, Lee D, Koo BK, Han I, Lim Y, Woods A, Couchman JR & Oh ES (2001) Solution structure of the dimeric cytoplasmic domain of syndecan-4. *Biochemistry* **40**, 8471–8478.
- 80 Whiteford JR, Ko S, Lee W & Couchman JR (2008) Structural and cell adhesion properties of zebrafish syndecan-4 are shared with higher vertebrates. *J Biol Chem* **283**, 29322–29330.
- 81 Multhaupt H a. B, Yoneda A, Whiteford JR, Oh E-S, Lee W & Couchman JR (2009) Syndecan signaling: when, where and why? *J Physiol Pharmacol* **60** (Suppl. 4), 31–38.
- 82 Choi Y, Yun J-H, Yoo J, Lee I, Kim H, Son H-N, Kim I-S, Yoon HS, Zimmermann P, Couchman JR *et al.* (2016) New structural insight of C-terminal region of Syntenin-1, enhancing the molecular dimerization and inhibitory function related on Syndecan-4 signaling. *Sci Rep* **6**, 36818.
- 83 Kim H, Yoo J, Lee I, Kang YJ, Cho H-S & Lee W (2015) Crystal structure of syndesmos and its interaction with Syndecan-4 proteoglycan. *Biochem Biophys Res Commun* **463**, 762–767.
- 84 Fan S, Feng Y, Wei Z, Xia B & Gong W (2009) Solution structure of synbindin atypical PDZ domain and interaction with syndecan-2. *Protein Pept Lett* **16**, 189–195.
- 85 Launay G, Salza R, Multedo D, Thierry-Mieg N & Ricard-Blum S (2015) MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res* **43**, D321–D327.
- 86 Clerc O, Deniaud M, Vallet SD, Naba A, Rivet A, Perez S, Thierry-Mieg N & Ricard-Blum S (2018) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res* **47**, D376–D381.
- 87 Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363.
- 88 Pathan M, Keerthikumar S, Ang C-S, Gangoda L, Quek CYJ, Williamson NA, Mouradov D, Sieber OM,

- Simpson RJ, Salim A *et al.* (2015) FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* **15**, 2597–2601.
- 89 Salza R, Lethias C & Ricard-Blum S (2017) The multimerization state of the Amyloid- $\beta$ 42 amyloid peptide governs its interaction network with the extracellular matrix. *J Alzheimers Dis* **56**, 991–1005.
- 90 Aquino RS, Teng YH-F & Park PW (2018) Glycobiology of syndecan-1 in bacterial infections. *Biochem Soc Trans* **46**, 371–377.
- 91 De Toro J, Herschlik L, Waldner C & Mongini C (2015) Emerging roles of exosomes in normal and pathological conditions: new insights for diagnosis and therapeutic applications. *Front Immunol* **6**, 203.
- 92 Xu R, Rai A, Chen M, Suwakulsiri W, Greening DW & Simpson RJ (2018) Extracellular vesicles in cancer – implications for future improvements in cancer care. *Nat Rev Clin Oncol* **15**, 617–638.
- 93 Zhang Y, Wang N, Raab RW, McKown RL, Irwin JA, Kwon I, van Kuppevelt TH & Laurie GW (2013) Targeting of heparanase-modified syndecan-1 by prosecretory mitogen lacritin requires conserved core GAGAL plus heparan and chondroitin sulfate as a novel hybrid binding site that enhances selectivity. *J Biol Chem* **288**, 12090–12101.
- 94 Herman K, Lekka M & Ptak A (2018) Unbinding kinetics of Syndecans by single-molecule force spectroscopy. *J Phys Chem Lett* **9**, 1509–1515.
- 95 Gao G, Plaas A, Thompson VP, Jin S, Zuo F & Sandy JD (2004) ADAMTS4 (aggrecanase-1) activation on the cell surface involves C-terminal cleavage by glycosylphosphatidyl inositol-anchored membrane type 4-matrix metalloproteinase and binding of the activated proteinase to chondroitin sulfate and heparan sulfate on syndecan-1. *J Biol Chem* **279**, 10042–10051.
- 96 Rodríguez-Manzanera JC, Carpizo D, Plaza-Calonge Mdel C, Torres-Collado AX, Thai SN-M, Simons M, Horowitz A & Iruela-Arispe ML (2009) Cleavage of syndecan-4 by ADAMTS1 provokes defects in adhesion. *Int J Biochem Cell Biol* **41**, 800–810.
- 97 Beauvais DM & Rapraeger AC (2010) Syndecan-1 couples the insulin-like growth factor-1 receptor to inside-out integrin activation. *J Cell Sci* **123**, 3796–3807.
- 98 Jung O, Trapp-Stamborski V, Purushothaman A, Jin H, Wang H, Sanderson RD & Rapraeger AC (2016) Heparanase-induced shedding of syndecan-1/CD138 in myeloma and endothelial cells activates VEGFR2 and an invasive phenotype: prevention by novel synstatins. *Oncogenesis* **5**, e202.
- 99 Wang H, Jin H & Rapraeger AC (2015) Syndecan-1 and Syndecan-4 capture epidermal growth factor receptor family members and the  $\alpha$ 3 $\beta$ 1 integrin via binding sites in their ectodomains: NOVEL SYNSTATINS PREVENT KINASE CAPTURE AND INHIBIT  $\alpha$ 6 $\beta$ 4-INTEGRIN-DEPENDENT EPITHELIAL CELL MOTILITY. *J Biol Chem* **290**, 26103–26113.
- 100 Fiore VF, Ju L, Chen Y, Zhu C & Barker TH (2014) Dynamic catch of a Thy-1- $\alpha$ 5 $\beta$ 1+syndecan-4 trimolecular complex. *Nat Commun* **5**, 4886.
- 101 Mittal A, Holehouse AS, Cohan MC & Pappu RV (2018) Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins. *J Mol Biol* **430**, 2403–2421.
- 102 van Dyck JF, Konijnenberg A & Sobott F (2017) Native mass spectrometry for the characterization of structure and interactions of membrane proteins. *Methods Mol Biol* **1635**, 205–232.
- 103 Kolesnikov AL, Budkov YA & Nogovitsyn EA (2014) Coarse-grained model of glycosaminoglycans in aqueous salt solutions. A field-theoretical approach. *J Phys Chem B* **118**, 13037–13049.
- 104 Samsonov SA, Bichmann L & Pisabarro MT (2015) Coarse-grained model of glycosaminoglycans. *J Chem Inf Model* **55**, 114–124.
- 105 Clerc O, Mariethoz J, Rivet A, Lisacek F, Pérez S & Ricard-Blum S (2018) A pipeline to translate glycosaminoglycan sequences into 3D models. Application to the exploration of glycosaminoglycan conformational space. *Glycobiology* **29**, 36–44.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table S1.** List of the partners of syndecans.

**Table S2.** List of partners specific to each syndecans and of their shared partners.

**Supplementary Table 1**

**Gondelaud and Ricard-Blum, Structures and interactions of syndecans**

**LIST OF SYNDECANS 1-4 PARTNERS**

<b>SDC species</b>	<b>Partners</b>	<b>Gene name</b>	<b>Label on Figure 4</b>	<b>Binding domain on SDCs</b>	<b>Reference (PMID or DOI)</b>
<b>SYNDECAN-1</b>					
RAT	$\alpha 3(\beta 1)$ integrin	ITGA3	$\alpha 3(\beta 1)$	ED: core protein	26350464
MOUSE	$\alpha 4(\beta 1)$ integrin	ITGA4	$\alpha 4(\beta 1)$	ED: core protein	26926788
HUMAN	$\alpha v\beta 3$ integrin	ITGAV, ITGB3	$\alpha v\beta 3$	ED: core protein	19255147
HUMAN	$\beta 1$ subunit	ITGB1	$\beta 1$	CD: core protein	18957427
HUMAN	$(\alpha 6)\beta 4$ integrin	ITGB4	$(\alpha 6)\beta 4$	CD: core protein	20181947
HUMAN	$(\alpha v)\beta 5$ integrin	ITGB5	$(\alpha v)\beta 5$	ED: core protein	19255147
HUMAN	A Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	ADAM17	ED	19875451
HUMAN	A disintegrin and metalloproteinase with thrombospondin motifs 4	ADAMTS4	ADAMTS4	ED: HS and CS	14701864
HUMAN	Neuroblast differentiation-associated protein AHNAK	AHNAK	AHNAK	-	26186194
HUMAN	AMP deaminase 2	AMPD2	AMPD2	-	26186194
HUMAN	Apolipoprotein E	APOE	APOE	ED: HS	16945147
HUMAN	Armadillo repeat protein deleted in velo-cardio-facial syndrome	ARVCF	ARVCF	-	26186194
HUMAN	Basigin	BSG	BSG	ED: HS	17267519
HUMAN	Baculovirus	-	BV	ED: HS	23926339
HUMAN	CASK*	CASK	CASK	CD: core protein	9660869
HUMAN	Eotaxin	CCL11	CCL11	ED: HS	15843578
MOUSE	CCL-17	CCL17	CCL17	ED: HS	15843578
HUMAN	CCL-2	CCL2	CCL2	ED: HS	29310785
MOUSE	CCL-5	CCL5	CCL5	ED: HS	<a href="https://doi.org/10.1093/glycob/cwg083">https://doi.org/10.1093/glycob/cwg083</a>
MOUSE	CCL-7	CCL7	CCL7	ED: HS	15843578
HUMAN	Cyclin-Y	CCNY	CCNY	-	26186194

MOUSE	Cortactin *	CTTN	CTTN	CD: core protein	9553134
RAT	Tetraspanin-28	CD81	CD81	ED: HS	27930836
HUMAN	Collagen I	COL1A1, COL1A2	Col I	ED: HS	3159726
MOUSE	Collagen III (chicken)	COL3A1	Col III	ED: HS	3159726
MOUSE	Collagen V	COL5A1, COL5A2	Col V	ED: HS	3159726
HUMAN	Casein kinase I isoform gamma-2	CSNK1G2	CSNK1G2	-	26186194
HUMAN	Casein kinase I isoform gamma-3	CSNK1G3	CSNK1G3	-	26186194
HUMAN	Cystathionine gamma-lyase	CTH	CTH	-	26186194
MOUSE	$\beta$ -catenin	CTNNB1	CTNNB1	CD: core protein	19255147
MOUSE	Cathepsin-G	CTSG	CTSG	ED: HS	9565572
MOUSE	Growth-regulated alpha protein	CXCL1	CXCL1	ED: HS	12464176
HUMAN	Cysteine and histidine-rich protein 1	CYHR1	CYHR1	-	26186194
HUMAN	Dehydrogenase/reductase SDR family member 2, mitochondrial	DHRS2	DHRS2	-	26186194
HUMAN	Endonuclease/exonuclease/p hosphatase family domain- containing protein 1	EEPD1	EEPD1	-	26186194
HUMAN	Protein EFR3 homolog A	EFR3A	EFR3A	-	26186194
RAT	Elastase	ELANE	ELANE	ED: HS	9565572
HUMAN	Ephrin type-A receptor 2	EPHA2	EPHA2	-	26186194
MOUSE	Receptor tyrosine-protein kinase erbB-2	ERBB2	ERBB2	ED: core protein	26350464
HUMAN	Thrombin	F2	F2	ED: core protein	23384311
MONKEY	Glycosaminoglycan xylosylkinase	FAM20B	FAM20B	-	25331875
HUMAN	FERM, ARHGEF and pleckstrin domain-containing protein 1	FARP1	FARP1	-	26186194
HUMAN	F-box only protein 46	FBXO46	FBXO46	-	26186194
MOUSE	Fibroblast growth factor 2	FGF2	FGF2	ED: HS	9462693
MOUSE	Fibroblast growth factor 7	FGF7	FGF7	ED: HS	29310785
RAT	Fibroblast growth factor receptor 1	FGFR1	FGFR1	ED: HS	11431373
RAT	Fibroblast growth factor receptor 2	FGFR2	FGFR2	ED: HS	11431373
HUMAN	Fibrinogen	FGA,FGB,FGG	Fibrinogen	-	29905671
HUMAN	Flotillin-1	FLOT1	FLOT1	-	29162604

HUMAN	Fascin	FSCN1	FSCN1	CD: core protein	11257118
HUMAN	Alpha-ketoglutarate-dependent dioxygenase FTO	FTO	FTO	-	26186194
HUMAN	Fyn	FYN	FYN	CD: core protein	9553134
HUMAN	Synectin *	GIPC1	GIPC1	CD: core protein	19581409
HUMAN	Glucocorticoid modulatory element-binding protein 1	GMEB1	GMEB1	-	26186194
HUMAN	Glucocorticoid modulatory element-binding protein 2	GMEB2	GMEB2	-	26186194
HUMAN	Guanine nucleotide-binding protein G(z) subunit alpha	GNAZ	GNAZ	-	26186194
HUMAN	Guanine nucleotide-binding protein subunit beta-4	GNB4	GNB4	-	26186194
HUMAN	Surface protein gp120	gp120	GP120	ED: HS	16157597
HUMAN	Hepatitis C virus	-	HCV	ED: HS	23576506
HUMAN	Hepatocyte growth factor	HGF	HGF	ED: HS	11830493
HUMAN	Hydroxymethylglutaryl-CoA synthase,	HMGCS1	HMGCS1	-	26186194
MOUSE	Papillomavirus 11	-	HPV11	ED: HS	14645569
HUMAN	Papillomavirus 16	-	HPV16	ED: HS	26289843
HUMAN	GTPase HRas	HRAS	HRAS		26186194
HUMAN	Interferon- $\beta$	IFNB1	IFNB1	ED: HS	29310785
HUMAN	Insulin-like growth factor 1 receptor	IGF1R	IGF1R	ED: core protein	20971705
RAT	Interleukin-34	IL34	IL34	ED: CS	25662098
HUMAN	Interleukin-8	IL8	IL8	ED: HS	12902511
HUMAN	Immune inhibitor A metalloprotease	InhA	INHA	ED: core protein	16926147
MOUSE	Lacritin	LACRT	LACRT	ED: HS & CS & core protein	16982797 23504321
HUMAN	Laminin a1 chain	LAMA1	LAMA1	ED: HS	9786856
MOUSE	Laminin a2 chain	LAMA2	LAMA2	ED: HS	9786856
HUMAN	Laminin $\alpha$ 3 LG45 Domain	LAMA3	LAMA3	ED: HS	22351752
HUMAN	Midkine	MDK	MDK	ED: HS & CS	15226297
HUMAN	Macrophage migration inhibitory factor	MIF	MIF	-	26852939
HUMAN	Matrix metalloproteinase-14	MMP14	MMP14	ED: core protein	12904296
HUMAN	Matrix metalloproteinase-16	MMP16	MMP16	ED: core protein	12904296
HUMAN	Matrix metalloproteinase-2	MMP2	MMP2	ED: core protein	23384311



HUMAN	Matrix metalloproteinase-3	MMP3	MMP3	ED: core protein	23384311
RAT	Matrix metalloproteinase-7	MMP7	MMP7	ED: core protein	12464176
HUMAN	Matrix metalloproteinase-9	MMP9	MMP9	ED: core protein	16513763
MOUSE	Myeloid cell nuclear differentiation antigen	MNDA	MNDA	-	25665578
HUMAN	Neural cell adhesion molecule 1	NCAM1	NCAM1	-	26186194
MOUSE	Neurofibromin	NF1	NF1	TMD & CD: core protein	11356864
HUMAN	Nrp599	Nrp599	NRP599	ED: core protein	16926147
HUMAN	Occludin	OCLN	OCLN	-	26186194
HUMAN	Opa	Opa	OPA	ED: HS	11207564
MOUSE	pORF2	ORF2	ORF2	ED: HS	19812150
HUMAN	VAR2CSA	p154varB	P154VARB	ED: CS	27556547
HUMAN	Poly [ADP-ribose] polymerase 2	PARP2	PARP2	-	26186194
MOUSE	Procollagen C-endopeptidase enhancer 1	PCOLCE	PCOLCE	ED: HS	25286301
HUMAN	PDGF-AA	PDGFA, PDGFA	PDGF-AA	ED: HS	29310785
HUMAN	Platelet endothelial cell adhesion molecule	PECAM1	PECAM1	ED: HS	24497640
HUMAN	Plasmin	PLG	PLG	ED: core protein	23384311
HUMAN	PR domain zinc finger protein 10	PRDM10	PRDM10	-	26186194
HUMAN	Protein PRRC1 (Proline-rich and coiled-coil-containing protein 1)	PRRC1	PRRC1	-	26186194
HUMAN	Pleiotrophin	PTN	PTN	ED: HS & CS	15226297
HUMAN	Receptor-type tyrosine-protein phosphatase gamma	PTPRG	PTPRG	-	26186194
RAT	Pyrin and HIN domain-containing protein 1	PYHIN1	PYHIN1	-	25665578
MOUSE	Rab5	RAB5A	RAB5A	CD: core protein	18957427
HUMAN	Ras-related protein Rap-2a	RAP2A	RAP2A	-	26186194
HUMAN	Raftlin	RFTN1	RFTN1	-	26186194
HUMAN	Ras-related protein R-Ras2	RRAS	RRAS	-	26186194
HUMAN	R-spondin-1	RSPO1	RSPO1	ED: HS	29212806
HUMAN	Syndecan-1	SDC1	SDC1	TMD: core protein	18093920

HUMAN	Syndecan-2	SDC2	SDC2	TMD: core protein	18093920
HUMAN	Syndecan-3	SDC3	SDC3	TMD: core protein	18093920
HUMAN	Syndecan-4	SDC4	SDC4	TMD: core protein	18093920
MOUSE	Syntenin-1	SDCBP	SDCBP	CD: core protein	9391086
RAT	Syntenin-2	SDCBP2	SDCBP2	CD: core protein	10911369
HUMAN	c-Src	SRC	SRC	CD: core protein	9553134
HUMAN	Syntaxin-4	STX4	STX4	ED: HS	27463539
HUMAN	Signal transducer and activator of transcription 3	STAT3	STAT3	-	19282476
HUMAN	Transforming growth factor beta-1	TGFB1	TGFB1	-	29454902
MOUSE	Thrombospondin-1	THBS1	THBS1	ED: HS	11257118
HUMAN	T-lymphoma invasion and metastasis-inducing protein 1*	TIAM1	TIAM1	CD: core protein	20361982
MOUSE	Tenascin-C	TNC	TNC	ED: HS	1708391
HUMAN	Tumor necrosis factor receptor superfamily member 13B	TNFRSF13B	TNFRSF13B	ED: HS	16357320
HUMAN	Tumor necrosis factor ligand superfamily member 13	TNFSF13	TNFSF13	ED: HS	16357320
HUMAN	E3 ubiquitin-protein ligase Topors	TOPORS	TOPORS	CD: core protein	22912899
HUMAN	Synbindin *	TRAPPC4	TRAPPC4	CD: core protein	11018053
HUMAN	Short transient receptor potential channel 4	TRPC4	TRPC4	-	26391658
HUMAN	$\alpha$ -tubulin *	TUBA	TUBA	CD: core protein	9553134
HUMAN	$\beta$ -tubulin *	TUBB	TUBB	CD: core protein	9553134
HUMAN	Vang-like protein 1	VANGL1	VANGL1	-	26186194
HUMAN	Vascular endothelial growth factor receptor 2	VEGFR2	VEGFR2	ED: core protein	26926788
HUMAN	Vitronectin	VTN	VTN	ED: HS	14681059
HUMAN	Wnts	WNT	WNT	ED: HS	29212806
MOUSE	Xylotransferase-1	XYLT1	XYLT1	ED: core protein	29681470
HUMAN	y2325	y2325	Y2325	-	20711500
RAT	Zinc metalloprotease ZmpC	zmpc	ZMPC	ED: core protein	17098735

SYNDECAN-2					
HUMAN	( $\alpha$ 6) $\beta$ 4 integrin	ITGB4	( $\alpha$ 6) $\beta$ 4	CD: core protein	20181947
HUMAN	AMP deaminase 2	AMPD2	AMPD2	-	26186194
MOUSE	Amyloid- $\beta$ protein 42	APP	APP	ED: HS	21460438
HUMAN	Large proline-rich protein BAG6	BAG6	BAG6	-	21988832
HUMAN	Transcription elongation factor, mitochondrial	C17ORF42	C17ORF42	-	26186194
HUMAN	CaM kinase-like vesicle-associated protein	CAMKV	CAMKV	-	26186194
HUMAN	Cullin-associated NEDD8-dissociated protein 2	CAND2	CAND2	-	26186194
HUMAN	CASK	CASK	CASK	CD: core protein	9660869
RAT	Caveolin-2	CAV2	CAV2	CD: core protein	25623282
RAT	Cortactin *	CTTN	CTTN	CD: core protein	9553134
HUMAN	Cytoskeleton-associated protein 5	CKAP5	CKAP5	-	26186194
HUMAN	Granulocyte-macrophage colony-stimulating factor	CSF2	CSF2	ED: HS	10734053
HUMAN	Casein kinase II subunit alpha'	CSNK2A2	CSNK2A2	-	23455922
MOUSE	Cytochrome P450 2C18	CYP2C18	CYP2C18	-	21988832
RAT	Survival-promoting peptide	DCD	DCD	ED: GAG	18599487
HUMAN	Dehydrogenase/reductase SDR family member 2, mitochondrial	DHRS2	DHRS2	-	26186194
HUMAN	Protein dopey-2	DOPEY2	DOPEY2	-	26186194
HUMAN	Ephrin-B2a	EFNB2	EFNB2	-	26186194
HUMAN	Ephrin type-B receptor 2	EPHB2	EPHB2	CD: core protein	11580899
HUMAN	Exocyst complex component 3	EXOC3	EXOC3	-	26186194
HUMAN	Ezrin	EZR	EZR	CD: core protein	12860416
HUMAN	FERM, ARHGEF and pleckstrin domain-containing protein 1	FARP1	FARP1	-	26186194
HUMAN	Fibroblast growth factor 1	FGF1	FGF1	-	26186194
HUMAN	Fibroblast growth factor 2	FGF2	FGF2	ED: HS	20382221
HUMAN	Fibroblast growth factor receptor 2	FGFR2	FGFR2	-	26186194
HUMAN	Fibroblast growth factor receptor 3	FGFR3	FGFR3	-	26186194
HUMAN	Fibroblast growth factor receptor-like 1	FGFRL1	FGFRL1	-	26186194

MOUSE	Fyn *	FYN	FYN	ED: core protein	9553134
HUMAN	Synectin *	GIPC1	GIPC1	CD: core protein	19581409
HUMAN	Glucocorticoid modulatory element-binding protein 1	GMEB1	GMEB1	-	26186194
HUMAN	Glucocorticoid modulatory element-binding protein 2	GMEB2	GMEB2	-	26186194
HUMAN	Surface protein gp120	gp120	GP120	ED: HS	16157597
HUMAN	G-protein coupled receptor 183	GPR183	GPR183	-	26186194
HUMAN	Melatonin-related receptor	GPR50	GPR50	-	26186194
HUMAN	General transcription factor 3C polypeptide 1	GTF3C1	GTF3C1	-	26186194
HUMAN	General transcription factor 3C polypeptide 2	GTF3C2	GTF3C2	-	26186194
HUMAN	General transcription factor 3C polypeptide 3	GTF3C3	GTF3C3	-	26186194
HUMAN	Trifunctional enzyme subunit alpha, mitochondrial	HADHA	HADHA	-	26186194
HUMAN	Trifunctional enzyme subunit beta, mitochondrial	HADHB	HADHB	-	26186194
HUMAN	HEAT repeat-containing protein 6	HEATR6	HEATR6	-	26186194
HUMAN	Histone H2A type 1	HIST1H2A G	HIST1H2A G	-	26186194
HUMAN	Hepatitis C virus	-	HCV	ED: HS	23576506
HUMAN	Herpes simplex virus	-	HSV	ED: HS	17553929
HUMAN	Interleukin-8	IL8	IL8	ED: HS	14527339
HUMAN	Kelch-like protein 7	KLHL7	KLHL7		26186194
HUMAN	Laminin $\alpha$ 3 chain	LAMA3	LAMA3	ED: HS	22351752
HUMAN	<i>L. eishmania braziliensis</i>	-	LEIBR	-	24478075
HUMAN	Endothelial lipase	LIPG	LIPG	-	26186194
HUMAN	Methyltransferase-like protein 8	METTL8	METTL8	-	26186194
HUMAN	Matrix metalloproteinase-14	MMP14	MMP14	ED: core protein	28972070
HUMAN	Matrix metalloproteinase-2	MMP2	MMP2	ED: core protein	23384311
HUMAN	Matrix metalloproteinase-7	MMP7	MMP7	ED: core protein	22227189
HUMAN	Matrix metalloproteinase-9	MMP9	MMP9	ED: core protein	23384311
HUMAN	mtlA	mtlA	MTLA	-	20711500
HUMAN	MYB	MYB	MYB	-	20195357
HUMAN	Neurobeachin-like protein 1	NBEAL1	NBEAL1	-	26186194
HUMAN	Condensin-2 complex subunit H2	NCAPH2	NCAPH2	-	26186194

HUMAN	Neurofibromin	NF1	NF1	TMD & CD: core protein	11356864
HUMAN	Oxysterol-binding protein-related protein 8	OSBPL8	OSBPL8	-	26186194
HUMAN	Poly [ADP-ribose] polymerase 2	PARP2	PARP2	-	26186194
HUMAN	Procollagen C-endopeptidase enhancer 1	PCOLCE	PCOLCE	ED: HS	25286301
HUMAN	3-phosphoinositide-dependent protein kinase 1	PDPK1	PDPK1	-	26186194
HUMAN	Phosphatidylinositol 4-kinase alpha	PI4KA	PI4KA	-	26186194
HUMAN	Calcineurin	PPP3CC	PPP3CC	CD: core protein	22164265
HUMAN	PR domain zinc finger protein 10	PRDM10	PRDM10	-	26186194
MOUSE	Receptor-type tyrosine-protein phosphatase eta	PTPRJ	PTPRJ	ED: core protein	25179601
HUMAN	Pyrin and HIN domain-containing protein 1	PYHIN1	PYHIN1	-	25665578
HUMAN	Ras GTPase-activating protein 1	RASA1	RASA1	-	22471946
MOUSE	Receptor-interacting serine/threonine-protein kinase 4	RIPK4	RIPK4	-	26972000
HUMAN	E3 ubiquitin-protein ligase RNF213	RNF213	RNF213	-	26186194
HUMAN	Protein S100-A13	S100A13	S100A13	-	26186194
HUMAN	Sterile alpha and TIR motif-containing protein 1	SARM1	SARM1	CD: core protein	21555464
HUMAN	Syndecan-1	SDC1	SDC1	TMD: core protein	18093920
HUMAN	Syndecan-2	SDC2	SDC2	TMD: core protein	18093920
HUMAN	Syndecan-3	SDC3	SDC3	TMD: core protein	18093920
HUMAN	Syndecan-4	SDC4	SDC4	TMD: core protein	18093920
HUMAN	Syntenin	SDCBP	SDCBP	CD: core protein	9391086
HUMAN	Semaphorin-3C	SEMA3C	SEMA3C	-	26186194
MOUSE	$\alpha$ 1-antitrypsin	SERPINA1	SERPINA1	-	21988832
HUMAN	Cationic amino acid transporter 3	SLC7A3	SLC7A3	-	26186194
HUMAN	Structural maintenance of chromosomes protein 2	SMC2	SMC2	-	26186194
HUMAN	Sorting nexin-18	SNX18	SNX18	-	26186194
HUMAN	Sorting nexin-33	SNX33	SNX33	-	26186194

HUMAN	c-Src *	SRC	SRC	CD: core protein	9553134
HUMAN	Extracellular sulfatase Sulf-1	SULF1	SULF1	-	26186194
HUMAN	Extracellular sulfatase Sulf-2	SULF2	SULF2	-	26186194
HUMAN	Transforming growth factor beta-1	TGFB1	TGFB1	ED: core protein	14976204
	Transforming growth factor beta receptor type 3	TGFBR3	TGFBR3	-	14976204
HUMAN	T-lymphoma invasion and metastasis-inducing protein 1*	TIAM1	TIAM1	CD: core protein	20361982
HUMAN	Transmembrane protein 11, mitochondrial	TMEM11	TMEM11	-	26186194
HUMAN	Transmembrane protein 186	TMEM186	TMEM186	-	26186194
HUMAN	Transmembrane protein 192	TMEM192	TMEM192	-	26186194
HUMAN	Tumor necrosis factor receptor superfamily member 13B	TNFRSF13B	TNFRSF13B	ED: HS	16357320
HUMAN	Synbindin *	TRAPPC4	TRAPPC4	CD: core protein	11018053
RAT	$\alpha$ -tubulin *	TUBA	TUBA	CD: core protein	9553134
HUMAN	$\beta$ -tubulin *	TUBB	TUBB	CD: core protein	9553134
HUMAN	Vang-like protein 1	VANGL1	VANGL1	-	26186194
HUMAN	Vang-like protein 2	VANGL2	VANGL2	-	26186194
RAT	vIRF	vIRF-1	VIRF1	-	25544563
RAT	Vitronectin	VTN	VTN	ED: HS	14681059
HUMAN	DNA repair protein complementing XP-C cells	XPC	XPC	-	26186194
HUMAN	Zinc finger protein 460	ZNF460	ZNF460	-	26186194
HUMAN	Zinc finger protein 92	ZNF92	ZNF92	-	26186194
<b>SYNDECAN-3</b>					
HUMAN	( $\alpha$ 6) $\beta$ 4 integrin	ITGB4	( $\alpha$ 6) $\beta$ 4	CD: core protein	20181947
CHICKEN	$\beta$ -arrestin-1	ARRB1	ARRB1	CD: core protein	17620599
RAT	$\beta$ -arrestin-2	ARRB2	ARRB2	CD: core protein	17620599
RAT	Artemin	ARTN	ARTN	ED: HS	21200028
RAT	Bone morphogenetic protein 2	BMP2	BMP2	ED: HS	16226436
RAT	CASK*	CASK	CASK	CD: core protein	9660869

HUMAN	Cortactin *	CTTN	CTTN	CD: core protein	9553134
RAT	collagen V $\alpha$ 4 chain	COL5A4	COL5A4	ED: HS	11751872
MOUSE	Survival-promoting peptide	DCD	DCD	ED: GAG	18599487
HUMAN	Ephrin-B2a	EFNB2	EFNB2	-	26186194
HUMAN	Fibroblast growth factor 2	FGF2	FGF2	ED: HS	8344959
RAT	Fyn *	FYN	FYN	ED: core protein	9553134
RAT	Glial cell line-derived neurotrophic factor	GDNF	GDNF	ED: HS	21200028
HUMAN	Synectin *	GIPC1	GIPC1	CD: core protein	19581409
HUMAN	Surface protein gp120	gp120	GP120	ED: HS	18040049
HUMAN	Inhibitor of nuclear factor kappa-B kinase subunit beta	IKBKB	IKBKB	-	AfCS 31405
HUMAN	Interleukin-8	IL8	IL8	ED: HS	16052590
HUMAN	Laminin-111	LAMA1, LAMB1, LAMC1	LN-111	ED: HS	9949192
HUMAN	Neurofibromin	NF1	NF1	TMD & CD: core protein	11356864
RAT	Neurturin	NRTN	NRTN	ED: HS	21200028
HUMAN	Syndecan-1	SDC1	SDC1	TMD: core protein	18093920
HUMAN	Syndecan-2	SDC2	SDC2	TMD: core protein	18093920
HUMAN	Syndecan-3	SDC3	SDC3	TMD: core protein	18093920
HUMAN	Syndecan-4	SDC4	SDC4	TMD: core protein	18093920
HUMAN	Syntenin	SDCBP	SDCBP	CD: core protein	9391086
HUMAN	Semaphorin-5A	SEMA5A	SEMA5A	-	15603739
RAT	c-Src *	SRC	SRC	CD: core protein	9553134
RAT	Tissue factor pathway inhibitor	TFPI	TFPI	ED: core protein	25617766
RAT	Thrombospondin-1	THBS1	THBS1	ED: HS	9949192
HUMAN	T-lymphoma invasion and metastasis-inducing protein 1	TIAM1	TIAM1	CD: core protein	20361982
RAT	Synbindin *	TRAPPC4	TRAPPC4	CD: core protein	11018053
HUMAN	$\alpha$ -tubulin *	TUBA	TUBA	CD: core protein	9553134

HUMAN	$\beta$ -tubulin *	TUBB	TUBB	CD: core protein	9553134
<b>SYNDECAN-4</b>					
RAT	$\alpha$ 3( $\beta$ 1) integrin	ITGA3	$\alpha$ 3( $\beta$ 1)	ED: core protein	26350464
RAT	( $\alpha$ 6) $\beta$ 4 integrin	ITGB4	( $\alpha$ 6) $\beta$ 4	CD: core protein	26350464
HUMAN	$\alpha$ -actinin	ACTN1	ACTN1	CD: core protein	12493766
HUMAN	A Disintegrin and metalloproteinase domain-containing protein 12	ADAM12	ADAM12	ED: HS	10831617
HUMAN	A Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	ADAM17	-	19875451
HUMAN	A disintegrin and metalloproteinase with thrombospondin motifs 1	ADAMTS1	ADAMTS1	ED: core protein	18775505
HUMAN	A disintegrin and metalloproteinase with thrombospondin motifs 4	ADAMTS4	ADAMTS4	ED: core protein	18775505
HUMAN	A disintegrin and metalloproteinase with thrombospondin motifs 5	ADAMTS5	ADAMTS5	-	19684582
HUMAN	Apolipoprotein E	APOE	APOE	ED: HS	24751902
HUMAN	Amyloid- $\beta$ protein 42	APP	APP	ED: HS	28106549
HUMAN	CASK*	CASK	CASK	CD: core protein	9660869
HUMAN	CCL-5	CCL5	CCL5	ED: HS	<a href="https://doi.org/10.1093/glycob/cwg083">https://doi.org/10.1093/glycob/cwg083</a>
RAT	Cortactin *	CTTN	CTTN	CD: core protein	9553134
RAT	Stromal cell-derived factor 1	CXCL12	CXCL12	ED: HS	16513763
HUMAN	Drebrin	DBN1	DBN1		26496610
HUMAN	Dynamin II	DNM2	DNM2	CD: core protein	15694365
HUMAN	Ephrin-B2a	EFNB2	EFNB2	-	26186194
HUMAN	Epidermal growth factor receptor	EGFR	EGFR	ED: core protein	26350464
HUMAN	Autotaxin- $\beta$	ENPP2	ENPP2	ED: core protein	30237860
HUMAN	Ezrin	EZR	EZR	-	10704377
RAT	Thrombin	F2	F2	ED: core protein	23384311



HUMAN	Glycosaminoglycan xylosylkinase	FAM20B	FAM20B	-	25331875
HUMAN	Fibroblast growth factor 1	FGF1	FGF1	-	26186194
MOUSE	Fibroblast growth factor 2	FGF2	FGF2	ED: HS	15226297
RAT	Fibronectin	FN1	FN1	ED: HS	10640397
HUMAN	Fyn	FYN	FYN	ED: core protein	9553134
HUMAN	Synectin *	GIPC1	GIPC1	CD: core protein	19581409
HUMAN	Surface protein gp120	gp120	GP120	ED: HS	16157597
HUMAN	G-protein coupled receptor 183	GPR183	GPR183	-	26186194
MOUSE	HBHA	hbhA	HBHA	ED: HS	27279134
HUMAN	Hepatitis C virus	-	HCV	ED: HS	24751902
HUMAN	High mobility group protein HMGI-C	HMGA2	HMGA2	-	26045162
HUMAN	Papillomavirus 11	-	HPV11	ED: HS	14645569
HUMAN	Papillomavirus 16	-	HPV16	ED: HS	14645569
HUMAN	Laminin $\alpha$ 3 chain	LAMA3	LAMA3	ED: HS	22351752
MOUSE	LIM domain and actin-binding protein 1	LIMA1	LIMA1	-	26496610
HUMAN	Latent-transforming growth factor beta-binding protein 2	LTBP2	LTBP2	ED: HS	20382221
HUMAN	Midkine	MDK	MDK	ED: HS & CS	15226297
HUMAN	Matrix metalloproteinase-3	MMP13	MMP13	-	20073988
HUMAN	Matrix metalloproteinase-14	MMP14	MMP14	ED: core protein	23384311
HUMAN	Matrix metalloproteinase-2	MMP2	MMP2	ED: core protein	23384311
HUMAN	Matrix metalloproteinase-3	MMP3	MMP3	ED: core protein	23384311
MOUSE	Matrix metalloproteinase-7	MMP7	MMP7	ED: core protein	23384311
HUMAN	Matrix metalloproteinase-9	MMP9	MMP9	ED: core protein	16513763
HUMAN	Myostatin	MSTN	MSTN	ED: HS	<a href="https://doi.org/10.1016/j.nmd.2009.06.026">https://doi.org/10.1016/j.nmd.2009.06.026</a>
MOUSE	Neurofibromin	NF1	NF1	TMD & CD: core protein	11356864
HUMAN	Syndesmos	NUDT16L1	NUDT16L1	CD: core protein	10633082
HUMAN	Opa	Opa	OPA	ED: HS	11207564
HUMAN	Procollagen C-endopeptidase	PCOLCE	PCOLCE	ED: HS	25286301

	enhancer 1				
HUMAN	Platelet endothelial cell adhesion molecule	PECAM1	PECAM1	ED: HS	24497640
HUMAN	Phosphatidylinositol 4,5-bisphosphate		PIP <sub>2</sub>	CD: core protein	16310216
HUMAN	Pleckstrin homology domain-containing family A member 7	PLEKHA7	PLEKHA7	-	28877994
MOUSE	Plasmin	PLG	PLG	ED: core protein	23384311
RAT	Calcineurin	PPP3CC	PPP3CC	CD: core protein	22164265
HUMAN	Protein kinase C alpha type	PRKCA	PRKCA	CD: core protein	16310216
HUMAN	Protein kinase C delta type	PRKCD	PRKCD	CD: core protein	11916978
HUMAN	Pleiotrophin	PTN	PTN	ED: HS & CS	15226297
HUMAN	Paxillin	PXN	PXN	-	11805099
HUMAN	Rho-related GTP-binding protein RhoG	RHOG	RHOG	CD: core protein	19581409
HUMAN	Receptor-interacting serine/threonine-protein kinase 4	RIPK4	RIPK4	-	26972000
HUMAN	R-spondin-3	RSPO3	RSPO3	ED: HS	21397842
HUMAN	Syndecan-1	SDC1	SDC1	TMD: core protein	18093920
HUMAN	Syndecan-2	SDC2	SDC2	TMD: core protein	18093920
HUMAN	Syndecan-3	SDC3	SDC3	TMD: core protein	18093920
HUMAN	Syndecan-4	SDC4	SDC4	TMD: core protein	18093920
CHICKEN	Syntenin	SDCBP	SDCBP	CD: core protein	9391086
MOUSE	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	SGTA	SGTA	-	25416956
HUMAN	c-Src *	SRC	SRC	CD: core protein	9553134
MOUSE	T-cell acute lymphocytic leukemia protein 1	TAL1	TAL1	-	19346495
HUMAN	TRAF family member-associated NF-kappa-B activator	TANK	TANK	-	21988832
HUMAN	Transforming growth factor beta-1-induced transcript 1	TGFB111	TGFB111	-	11805099

	protein				
HUMAN	Transforming growth factor beta-3 proprotein	TGFB3	TGFB3	-	21988832
HUMAN	Transglutaminase-2	TGM2	TGM2	ED: HS	19398782
HUMAN	Thy-1 membrane glycoprotein	THY1	THY1	-	25216363
HUMAN	T-lymphoma invasion and metastasis-inducing protein 1	TIAM1	TIAM1	CD: core protein	20361982
HUMAN	Tenascin-C	TNC	TNC	ED: HS	17901052
HUMAN	Tumor necrosis factor receptor superfamily member 13B	TNFRSF13B	TNFRSF13B	ED: HS	16357320
HUMAN	Synbindin*	TRAPPC4	TRAPPC4	CD: core protein	11018053
MOUSE	Short transient receptor potential channel 4	TRPC4	TRPC4	-	26391658
HUMAN	Short transient receptor potential channel 7	TRPC7	TRPC7	-	26391658
HUMAN	$\alpha$ -tubulin *	TUBA	TUBA	CD: core protein	9553134
HUMAN	$\beta$ -tubulin *	TUBB	TUBB	CD: core protein	9553134
HUMAN	Polyubiquitin-C	UBC	UBC	-	22871113
HUMAN	Vitronectin	VTN	VTN	ED: HS	14681059

**Supplementary Table 1: List of the partners of syndecans (SDCs)** identified by manual curation of the literature, by querying two interaction databases, MatrixDB (*Launay et al., 2015, Clerc et al., in press*) and IntAct database (*Orchard et al., 2014*), the large-scale Bioplex 2.0 interaction dataset (*Huttlin et al., 2015, Schweppe et al., 2018*) and the adhesome (<http://www.adhesome.org/>). AfCS: Yeast 2-Hybrid: Interaction Data ([http://www.signaling-gateway.org/data/Y2H/cgi-bin/y2h\\_int.cgi?id=31405](http://www.signaling-gateway.org/data/Y2H/cgi-bin/y2h_int.cgi?id=31405)), source used for the interaction between syndecan-3 and the inhibitor of kappaB kinase beta (IKBKB). CD: cytoplasmic domain, CS: chondroitin sulfate, ED: ectodomain, GAG: glycosaminoglycan, HS: heparan sulfate, TMD: transmembrane domain. \*Binding sequences found in the four syndecans. For integrins: the subunits encoded by the genes in brackets have not been experimentally demonstrated to be directly involved in syndecan binding, but have been mentioned by the authors in the publications. (-: undetermined binding site).

## Supplementary Table 2

### Gondelaud and Ricard-Blum, Structures and interactions of syndecans

#### SYNDECAN SPECIFIC PARTNERS

##### SDC-1 SPECIFIC PARTNERS: 74

Gene names

AHNAK	CYHR1	IGF1R	PRRC1
ARVCF	EEPDI	IL34	PTPRG
BSG	EFR3A	INHA	RAB5A
BV	ELANE	ITGA4, (ITGB1)	RAP2A
CCL11	EPHA2	(ITGAV), ITGB3	RFTN1
CCL17	ERBB2	(ITGAV), ITGB5	RRAS
CCL2	FBXO46	ITGB1	RSPO1
CCL7	FGA, FGB, FGG	LACRT	SDCBP2
CCNY	FGF7	LAMA1	STX4
CD81	FGFR1	LAMA2	STAT3
COL1A1, COL1A2	FLOT1	MIF	TNFSF13
COL3A1	FSCN1	MMP16	TOPORS
COL5A1, COL5A2	FTO	MNDA	VEGFR2
CSNK1G2	GNAZ	NCAM1	WNT
CSNK1G3	GNB4	NRP599	XYLT1
CTH	HGF	OCLN	Y2325
CTNNB1	HMGCS1	ORF2	ZMPC
CTSG	HRAS	P154VARB	
CXCL1	IFNB1	PDGF-AA	

##### SDC-2 SPECIFIC PARTNERS: 56

Gene names

BAG6	GPR50	MYB	SMC2
C17ORF42	GTF3C1	NBEAL1	SNX18
CAMKV	GTF3C2	NCAPH2	SNX33
CAND2	GTF3C3	OSBPL8	SULF1
CAV2	HADHA	PDPK1	SULF2
CKAP5	HADHB	PI4KA	TGFBR3
CSF2	HEATR6	PTPRJ	TMEM11
CSNK2A2	HIST1H2AG	RASA1	TMEM186
CYP2C18	HSV	RNF213	TMEM192
DOPEY2	KLHL7	S100A13	VANGL2
EPHB2	LEIBR	SARM1	VIRF1
EXOC3	LIPG	SEMA3C	XPC
FGFR3	METTL8	SERPINA1	ZNF460
FGFRL1	MTLA	SLC7A3	ZNF92

##### SDC-3 SPECIFIC PARTNERS: 11

Gene names

ARRB1	ARTN	COL5A4
ARRB2	BMP2	GDNF

IKBKB  
LAMA1, LAMB1, LAMC1

NRTN  
SEMA5A

TFPI

#### SDC-4 SPECIFIC PARTNERS: 33

Gene names

ACTN1	FN1	PLEKHA7	TGFB1I1
ADAM12	HBHA	PRKCA	TGFB3
ADAMTS1	HMGA2	PRKCD	TGM2
ADAMTS5	LIMA1	PXN	THY1
CXCL12	LTBP2	RHOG	TRPC7
DBN1	MMP13	RSPO3	UBC
DNM2	MSTN	SGTA	
EGFR	NUDT16L1	TAL1	
ENPP2	PIP2	TANK	

---

#### SYNDECAN COMMON PARTNERS

##### SDC1 and SDC2: 11

Gene names

AMPD2	FGFR2	PARP2	TGFB1
DHRS2	GMEB1	PRDM10	VANGL1
FARP1	GMEB2	PYHIN1	

##### SDC1 and SDC3: 1

Gene name

THBS1

##### SDC1 and SDC4: 17

Gene names

ADAM17	FAM20B	MMP3	TNC
ADAMTS4	HPV11	OPA	TRPC4
APOE	HPV16	PECAM1	
CCL5	ITGA3, (ITGB1)	PLG	
F2	MDK	PTN	

##### SDC2 and SDC3: 1

Gene name

DCD

##### SDC2 and SDC4: 6

Gene names

APP	FGF1	PPP3CC
EZR	GPR183	RIPK4

**SDC1 and SDC2 and SDC3: 1**

Gene name

IL8

**SDC1 and SDC2 and SDC4: 9**

Gene names

HCV	MMP2	PCOLCE
LAMA3	MMP7	TNFRSF13B
MMP14	MMP9	VTN

**SDC2 and SDC3 and SDC4: 1**

Gene name

EFNB2

**SDC1 and SDC2 and SDC3 and SDC4: 18**

Gene names

CASK	GP120	SDC3	TRAPPC4
CTTN	(ITGA6), ITGB4	SDC4	TUBA
FGF2	NF1	SDCBP	TUBB
FYN	SDC1	SRC	
GIPC1	SDC2	TIAM1	

**Supplementary Table 2: List of partners specific to each syndecans and of their shared partners.** For integrins: the subunits encoded by the genes in brackets have not been experimentally demonstrated to be directly involved in syndecan binding, but have been mentioned by the authors in the publications.

### IV.3.2 Identification de partenaires des ectodomains des syndécans par interférométrie de biocouche

La grande majorité des partenaires physiologiques des syndécans interagissent avec leurs chaînes d'héparane sulfate. Seules quelques protéines interagissent directement avec leur partie protéique, majoritairement sur leur domaine cytosolique. Les syndécans peuvent être clivés au niveau juxtamembranaire et les chaînes d'héparane sulfate dégradées par l'héparanase, libérant ainsi dans l'espace extracellulaire leur ectodomaine déglycosylé. La recherche de nouveaux partenaires a donc été effectuée en utilisant les ectodomains des syndécans dépourvus de leurs chaînes de GAGs (**Tableau XXIV**). Ces expériences ont été réalisées par interférométrie de biocouche où les ectodomains des syndécans ont été utilisés en tant qu'analytes afin de conserver leur flexibilité. Seuls deux nouveaux partenaires ont été identifiés, le facteur fibroblastique 2 (FGF-2) et l'anastelline.

Catégorie	Nom	Espèce
<b>Facteurs de croissance</b>	<i>Fibroblast Growth Factor 2</i> <i>Vascular endothelial growth factor</i> <i>Tumor Necrosis Factor</i>	Humain
<b>Récepteurs</b>	<i>Discoidin domain-containing receptor 1</i> <i>Discoidin domain-containing receptor 2</i> <i>Macrophage receptor MARCO</i> <i>Leukocyte-associated immunoglobulin-like receptor 1</i> <i>Anthrax toxin receptor 1</i>	Humain
<b>Protéines matricielles ou associés</b>	Collagène IV Collagène XVII Collagène XXIII Domaine NCI du collagène VIII Fibronectine cellulaire Fibronectine plasmatique Neuropiline-1	Humain

**Tableau XXIV (1/2) : Partenaires criblés lors des expériences d'interférométrie de biocouche avec les ectodomains des syndécans dépourvus de leurs chaînes de glycosaminoglycanes.**

Catégorie	Nom	Espèce
Enzymes ou associé	<i>Lysyl oxidase-like 2</i> Propeptide de la lysyl oxydase	Humain
Matricryptines	Anastelline Endostatine	Humain
Autres molécules extracellulaires	$\alpha$ -synucleine	Humain

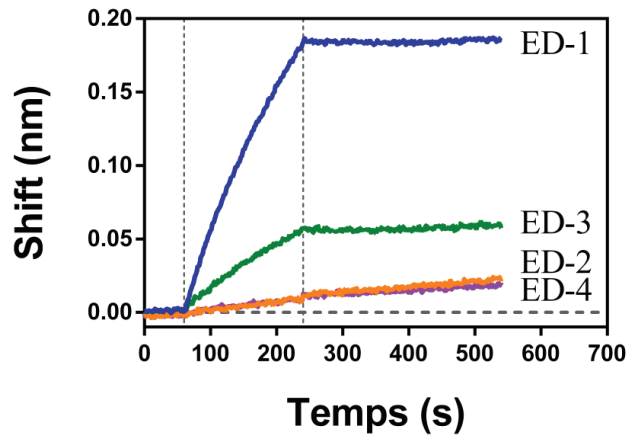
**Tableau XXIV (2/2) : Partenaires criblés lors des expériences d'interférométrie de biocouche avec les ectodomains des syndécans dépourvus de leurs chaînes de glycosaminoglycanes.**

*a) Interaction entre les ectodomains des syndécans et le FGF-2*

Le facteur de croissance fibroblastique 2 (FGF-2) fait partie de l'interactome consensus des 4 syndécans, et interagit avec ces-derniers *via* leurs chaînes d'héparanes sulfate. En particulier, les N-sulfatations des résidus de glucosamine et les 2-O-sulfatations des résidus d'acide iduronique des chaînes d'HS sont des prérequis à l'interaction HS - FGF-2 (*van Wijk & van Kuppevelt, 2014*). Cette interaction n'est également possible que si la chaîne d'HS est longue d'au moins 5 résidus. En l'absence de ces oligosaccharides, FGF-2 perd sa capacité à interagir et à activer son récepteur FGFR-1 (*Ashikari-Hada et al., 2009*).

Cependant, le rôle de la partie protéique de l'ectodomaine des syndécans dans cette interaction n'a pas été étudié. Pour cela, le FGF-2 a été immobilisé par couplage amine et les ectodomains dépourvus de leurs chaînes de GAGs ont été utilisés en tant qu'analyte (**Figure 123**). Une interaction a été détectée pour les 4 ectodomains des syndécans. Le *shift* maximal observé en fin d'étape d'association est le plus important pour ED-1 (0,18 nm) et en moindre mesure pour ED-3 (0,06 nm). En revanche, ED-2 ou ED-4 n'interagissent que faiblement avec le FGF-2 (0,01 nm).

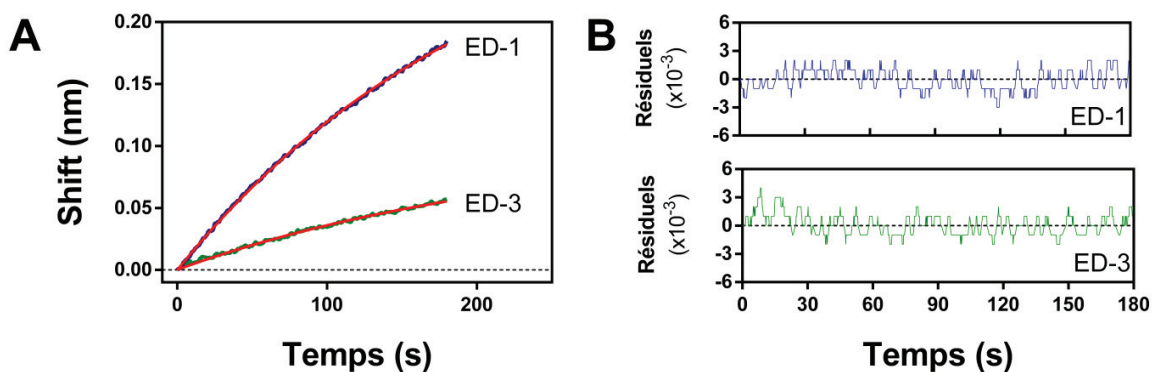




**Figure 123 : Expériences d'interactions entre FGF-2 et les EDs des syndécans sans GAGs.** Les sensorgrammes lissés pour chaque ectodomaines de syndécans (EDs) sont présentés. Le FGF-2 a été immobilisé et les EDs sont en solution à une concentration de 500 nM (ED-1 : bleu, ED-2 : orange, ED-3 : vert, ED-4 : violet). Les courbes verticales en pointillé séparent les différentes étapes de l'expérience.

Ces complexes sont très stables étant donné qu'aucune dissociation n'est observée. Une unique concentration en analyte a été utilisée dans ces expériences rendant délicat le calcul des constantes cinétiques. Cependant, il est possible d'extraire la valeur de  $k_{obs}$  (exprimée en  $s^{-1}$ ) qui est une constante de vitesse observée caractérisant l'interaction entre les deux partenaires. C'est un paramètre général reflétant le phénomène dynamique d'association - dissociation lors de la formation du complexe. La valeur de  $k_{obs}$  peut être déduite de la phase d'association uniquement, ce qui permet d'éviter la problématique de quasi-absence de dissociation du complexe observée dans nos expériences (**Figure 123**). Le modèle le plus simple pouvant décrire ces données a été utilisé, où un seul site d'interaction existe (modèle 1 : 1) (**Figures 124A** et **124B**). Il décrit parfaitement les sensorgrammes expérimentaux comme en atteste les valeurs de  $\chi^2$  (0,0002) et  $R^2$  (1,0) pour chacun des deux ectodomains. La valeur de  $k_{obs}$  est légèrement plus élevée pour ED-1 ( $4,7 \cdot 10^{-3} s^{-1}$ ) que pour ED-3 ( $4,3 \cdot 10^{-3} s^{-1}$ ).

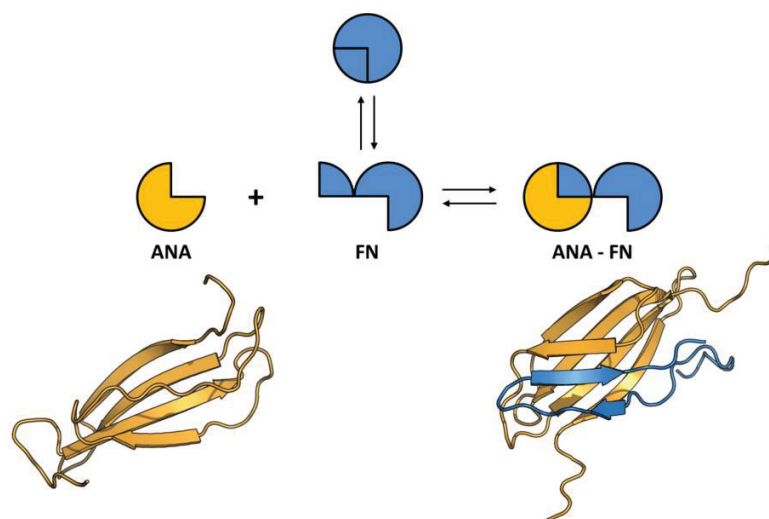
Ces expériences montrent que, malgré l'absence des chaînes d'héparane sulfate qui sont des sites d'interaction privilégiés de FGF-2, les ectodomains des syndécans 1 et 3 notamment ont toujours la capacité d'interagir avec le FGF-2. Ces interactions peuvent vraisemblablement être médiées par des interactions électrostatiques étant donné que le FGF-2 est une protéine très basique (d'où son nom de *facteur de croissance fibroblastique basique*) possédant un point isoélectrique d'environ 11,2, les ectodomains de syndécan possédant quant à eux des points isoélectriques assez acides compris entre 4,4 et 4,7.



**Figure 124 : Détermination de la constante  $k_{obs}$  lors de l'interaction entre le facteur de croissance fibroblastique-2 et l'ectodomaine du syndécan-1 et du syndécan-3. A)** Les sensorgrammes pour l'ectodomaine du syndécan-1 (ED-1, bleu) et du syndécan-3 (ED-3, vert) sont présentés avec leur ajustement respectif (en rouge) réalisé sur la phase d'association. **B)** Résiduels des ajustements avec un modèle 1 : 1.

*b) Interaction entre les syndécans et l'anastelline*

L'anastelline (FNIII<sub>1D</sub>) est une matricryptine, c'est-à-dire un fragment bioactif de 8,4 kDa possédant des propriétés anti-angiogéniques, anti-tumorales et anti-métastatiques (*Cho et al., 2015, Ricard-Blum & Vallet, 2017*). Elle est issue de l'extrémité C-terminale du premier domaine de type fibronectine III (FNIII) de la fibronectine, une protéine dimérique d'environ 440 kDa synthétisée sous une forme soluble et libérée dans le plasma ou bien sous une forme insoluble déposée dans la matrice extracellulaire. L'anastelline est une protéine riche en feuillets  $\beta$  mais est cependant dépourvue de deux brins  $\beta$  présents normalement à l'extrémité N-terminale du domaine FNIII (*Briknarová et al., 2003*). L'anastelline interagit avec elle-même et avec des domaines FNIII<sub>1-2</sub>, FNIII<sub>3</sub> et FNIII<sub>11</sub> de la fibronectine, ce qui mène à la formation d'agrégats insolubles de fibronectine appelée superfibronectine (*Ohashi et al., 2009*). L'interaction entre l'anastelline et la fibronectine nécessite un dépliement partiel du module FNIII présent dans la fibronectine intacte sur laquelle l'anastelline interagit. Puis, elle acquiert ses deux brins N-terminaux (brins  $\beta$  A et B) aux dépens du domaine FNIII par un mécanisme d'échange de brins (*Ohashi et al., 2009, Stine et al., 2017*) (**Figure 125**).



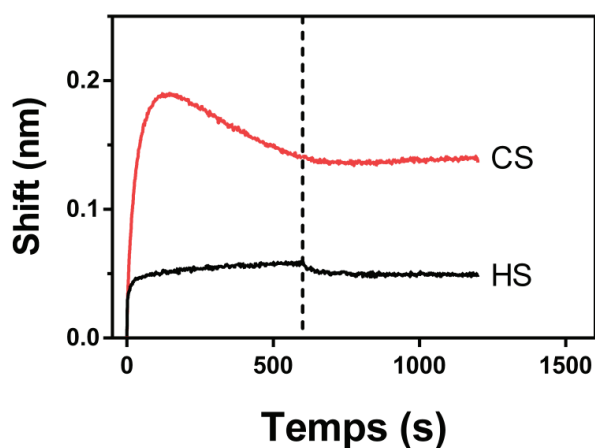
**Figure 125 : Mécanisme d'interaction entre l'anastelline et le domaine FNIII<sub>3</sub> de la fibronectine.** L'anastelline (ANA, jaune) interagit avec le domaine FNIII<sub>3</sub> (bleu) lorsqu'il est partiellement désordonné. Cette interaction est accompagnée d'un mécanisme d'échange de brins  $\beta$  (*domain swapping*) afin de reconstituer un module FNIII intact. Les structures PDB de l'anastelline seule (1Q38) ou en complexe avec un fragment du 1<sup>er</sup> domaine FNIII de la fibronectine (5J6Z) sont présentées. D'après *Stine et al.* (2017).

L'anastelline interagit également avec le fibrinogène dont elle provoque l'agrégation, ce qui mène à la formation de superfibrinogène. Le superfibrinogène possède des activités anti-tumorales et anti-angiogéniques, c'est-à-dire des propriétés similaires à l'anastelline seule (*Yi & Ruoslahti, 2001*). Un deuxième fragment issu du dixième module FNIII de la fibronectine mène également à la formation de superfibronectine (*Gee et al., 2013*). Il a également été montré que des cellules en culture adhèrent et s'étalent sur l'anastelline, et que ce mécanisme est dépendant d'intégrines et de protéoglycane à héparane sulfate (*Mercurius & Morla, 2001*). De ce fait, les syndécans pourraient être des partenaires de l'anastelline.

### 1. COMPLEXE ANASTELLINE – CHAINES D'HEPARANE SULFATE

Les chaînes d'héparane sulfate et l'héparine avaient déjà été proposées comme partenaires de l'anastelline qui est une protéine basique ayant un point isoélectrique de 9,9 (*Briknarová et al., 2003*) et cette hypothèse a été vérifiée (*Gui et al., 2006*). Nous avons effectivement retrouvé expérimentalement cette interaction en immobilisant l'anastelline sur un capteur et en le mettant en contact avec une solution contenant des chaînes d'héparane sulfate de muqueuse intestinale porcine ou de chondroïtine sulfate de cartilage bovin (**Figure 126**).

Une faible dissociation des complexes est observée lors de la phase de dissociation. Les complexes semblent être stables bien que le sensorgramme correspondant aux chaînes de CS a une forme inhabituelle. Etant donné que le signal observé en BLI est dépendant de l'épaisseur fixée à la surface du capteur, la forme de la courbe pourrait indiquer 1) une modification de la conformation des chaînes de CS, qui se déposeraient à la surface de l'anastelline, 2) une dissociation des chaînes de CS après avoir interagit avec l'anastelline, ou 3) un changement de conformation de l'anastelline.



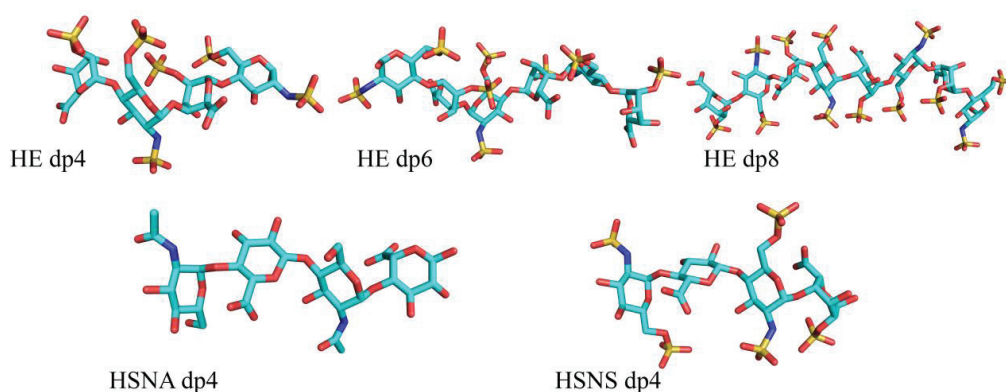
**Figure 126 : Interaction entre l'anastelline et les chaînes d'héparane et de chondroïtine sulfate.** Sensorgrammes obtenus par contact entre l'anastelline immobilisée et des chaînes d'héparane sulfate (HS) ou de chondroïtine sulfate (CS) à une concentration de 5  $\mu$ M. Les sensorgrammes contrôles ont été soustraits.

Les résidus d'acides aminés de l'anastelline responsables de son interaction avec l'héparine ont été déterminés de façon indirecte (*Gui et al., 2006*). Les auteurs de l'étude ont utilisé des peptides d'anastelline de 18 résidus et ont observé leurs effets sur la croissance cellulaire, qui est un processus régulé entre autres par l'interaction de la fibronectine avec les chaînes d'héparane sulfate portées par des protéoglycanes. Ils ont déduit que l'héparine interagissait avec l'anastelline par les résidus basiques Arg<sup>644</sup>, Arg<sup>646</sup> et Lys<sup>648</sup> ainsi que par le résidu de Trp<sup>645</sup> (selon la numérotation UniProt KB de la fibronectine, P02751) présents à sa surface. Afin de vérifier ces observations indirectes et de déterminer le ou les sites d'interactions de l'héparine sur l'anastelline, des expériences de docking ont été réalisées en collaboration avec Martyna Maszota-Zieleniak et Sergey Samsonov (Université de Gdansk, Pologne).

L'anastelline a été utilisée comme récepteur sans spécifier de site d'interaction et 5 oligosaccharides d'héparine ou d'héparane sulfate ayant des degrés de polymérisation (dp) et de sulfatation différents ont été utilisés (**Tableau XXV** et **Figure 127**). Les trois oligosaccharides d'héparine varient uniquement en taille (dp4, dp6 et dp8) et deux dp4 d'héparane sulfate ont été utilisés, le premier étant non sulfaté (HSNA) et le second très sulfaté (HSNS) reflétant les régions appauvries (domaine NA) ou enrichies en sulfatation (domaine NS) retrouvées le long des chaînes d'héparane sulfate (*Couchman, 2010*).

Noms des oligosaccharides	Séquence des oligosaccharides	Degré de sulfatation
<i>Oligosaccharides d'héparine</i>		
<b>HE dp4</b>	[ GlcNS(6S)-IdoA(2S) ] x 2	6
<b>HE dp6</b>	[ GlcNS(6S)-IdoA(2S) ] x 3	9
<b>HE dp8</b>	[ GlcNS(6S)-IdoA(2S) ] x 4	12
<i>Oligosaccharides d'héparane sulfate</i>		
<b>HSNA (dp4)</b>	GlcNAc-GlcUA-GlcNAc-IdoUA	0
<b>HSNS (dp4)</b>	GlcNS(6S)-GlcUA-GlcNS(6S)-IdoUA(2S)	5

**Tableau XXV : Caractéristiques des oligosaccharides utilisés comme ligands dans les expériences de docking avec l'anastelline.** GlcNS(6S) : glucose N- et 6-sulfaté, IdoA(2S) : acide iduronique 2-sulfaté, GlcNAc : N-acétylglucosamine, GlcUA : acide glucuronique, IdoUA : acide iduronique.



**Figure 127 : Structure des 5 oligosaccharides d'héparine (HE) ou d'héparane sulfate (HSNA, HSNS) utilisés lors des expériences de docking.**

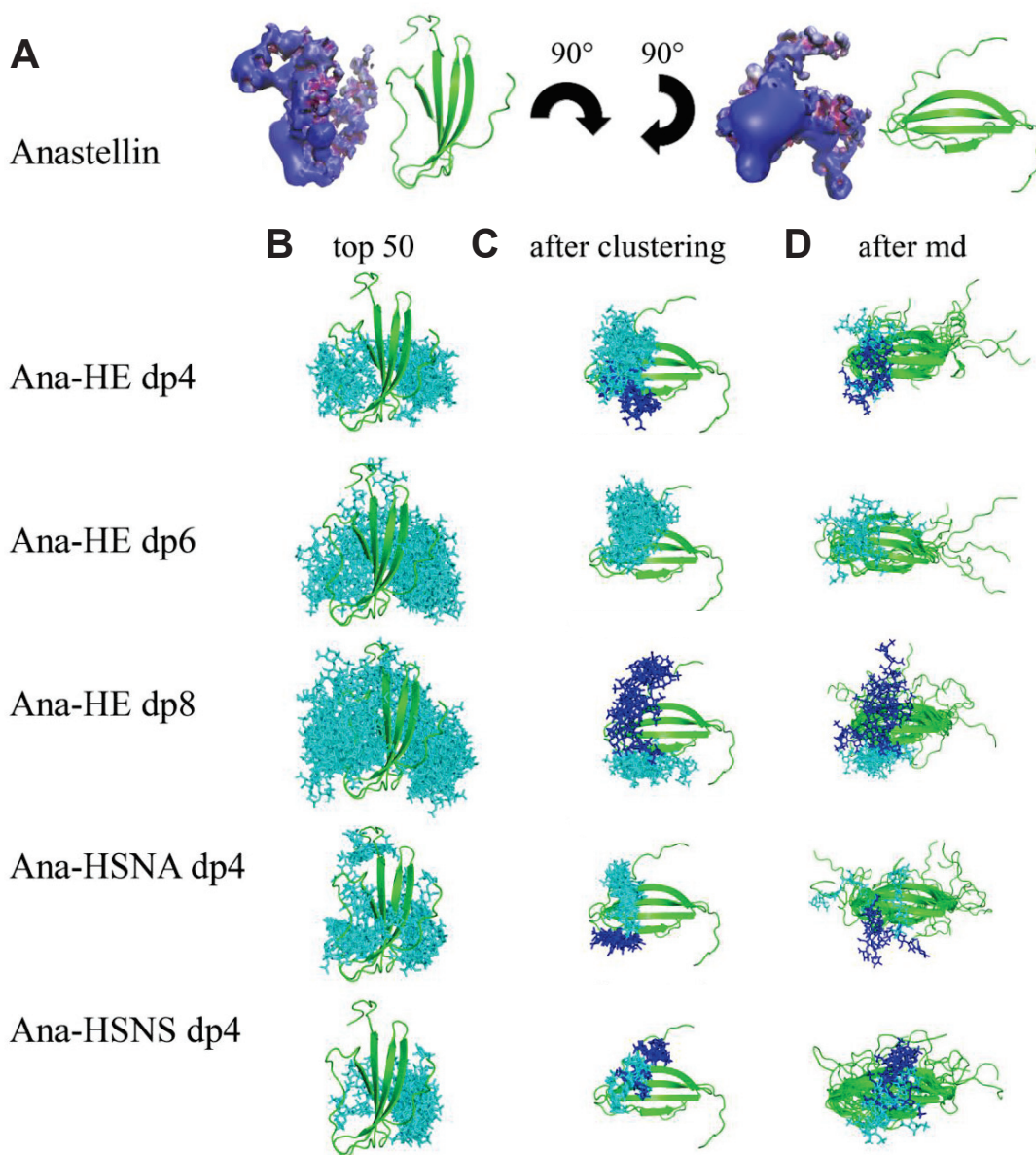
Le potentiel électrostatique de l'anastelline est fortement positif notamment sur une de ses faces qui regroupe plusieurs résidus d'acides aminés basiques (**Figure 128A**). Au total, 300 poses ont été générées et les 50 plus favorables (**Figure 128B**) ont été regroupées par clusters selon leurs énergies d'interaction (**Figure 128C**). Les 5 oligosaccharides utilisés dans ces expériences interagissent avec la face basique de l'anastelline. La distribution des poses des oligosaccharides d'héparine est cependant plus étendue que celle des tétrasaccharides d'héparane sulfate, notamment par rapport à celle de HSNS (**Figure 128B**). Le regroupement des 50 meilleures poses mène à la formation de deux clusters pour chacun des oligosaccharides, excepté pour le dp6 d'héparine (**Figure 128C**).

Il permet de mettre de nouveau en évidence une distribution plus étendue des poses pour les oligosaccharides d'héparine que pour ceux d'héparane sulfate ce qui suggère une plus grande spécificité d'interaction entre les oligosaccharides d'héparane sulfate avec l'anastelline, notamment pour le plus sulfaté (HSNS). Pour chacun des clusters, 5 structures représentatives ont été sélectionnées et une simulation par dynamique moléculaire de 25 ns a été effectuée (**Figure 128D**). Les complexes sont très stables étant donné qu'aucune dissociation entre les oligosaccharides et l'anastelline n'a été observée pendant la simulation. Les valeurs d'énergie d'interaction ont été déduites (**Tableau XXVI**).

Le dp4 d'héparane sulfate non sulfaté (HSNA) possède l'énergie d'interaction la plus faible ( $- 17,9 \pm 3,7 \text{ kcal.mol}^{-1}$ ) alors que celle du dp4 d'héparane sulfate très sulfaté (HSNS) est similaire à celle du dp4 d'héparine ( $- 46,3 \pm 14,0$  et  $- 44,9 \pm 13,7 \text{ kcal.mol}^{-1}$ , respectivement). Cela suggère que la sulfatation présente sur l'acide iduronique du dp4 d'héparine n'est pas requise pour l'interaction avec l'anastelline. Plus le degré de polymérisation de l'oligosaccharide d'héparine est élevé, plus l'interaction avec l'anastelline est favorisée. La différence d'énergie d'interaction du dp4 et du dp6 est faible (environ  $- 45$  et  $- 49 \text{ kcal.mol}^{-1}$ , respectivement) mais plus importante entre le dp6 et le dp8 (environ  $- 49$  et  $- 66 \text{ kcal.mol}^{-1}$ , respectivement). D'une manière générale, plus l'oligosaccharide est chargé, plus l'interaction est favorisée (**Tableau XXVI**).

Oligosaccharides	$\Delta G$ moyen $\pm$ écart-type / cluster ( $\text{kcal.mol}^{-1}$ )	$\Delta G$ moyen $\pm$ écart-type ( $\text{kcal.mol}^{-1}$ )	Degré de sulfatation
<i>Oligosaccharides d'héparine</i>			
<b>HE dp4</b>	Cluster 1 : $- 52,9 \pm 13,2$	<b><math>- 44,9 \pm 13,7</math></b>	6
	Cluster 2 : $- 37,0 \pm 9,7$		
<b>HE dp6</b>	Cluster 1 : $- 49,3 \pm 15,3$	<b><math>- 49,3 \pm 15,3</math></b>	9
<b>HE dp8</b>	Cluster 1 : $- 73,2 \pm 17,3$	<b><math>- 66,4 \pm 14,9</math></b>	12
	Cluster 2 : $- 59,6 \pm 9,1$		
<i>Oligosaccharides d'héparane sulfate</i>			
<b>HSNA dp4</b>	Cluster 1 : $- 18,2 \pm 4,9$	<b><math>- 17,9 \pm 3,7</math></b>	0
	Cluster 2 : $- 17,6 \pm 2,7$		
<b>HSNS dp4</b>	Cluster 1 : $- 40,8 \pm 13,9$	<b><math>- 46,3 \pm 14,0</math></b>	5
	Cluster 2 : $- 51,8 \pm 13,2$		

**Tableau XXVI : Analyse des simulations par dynamique moléculaire de 5 poses représentatives pour chacun des oligosaccharides.** Les valeurs indiquent l'énergie libre d'interaction ( $\Delta G$ ) entre les oligosaccharides et l'anastelline. Le degré de sulfatation des oligosaccharides est indiqué.



**Figure 128 : Docking des oligosaccharides d’héparine et d’héparane sulfate sur l’anastelline. A)** Potentiel électrostatique de surface de l’anastelline (positif : bleu, négatif : rouge). **B)** 50 meilleures poses des oligosaccharides (cyan) sur l’anastelline (Ana, vert). **C)** Regroupement (*clustering*) des 50 meilleures poses en clusters (cluster 1 : cyan, cluster 2 : bleu foncé) sur l’anastelline. **D)** Poses des oligosaccharides sur l’anastelline après simulations par dynamique moléculaire (md). La légende est identique à C). HE : oligosaccharides d’héparine, HSNS et HSNA : oligosaccharides d’héparane sulfate sulfatés ou non, dp : degré de polymérisation. Travaux de Martyna Maszota-Zieleniak et Sergey Samsonov (Université de Gdansk, Pologne).

Il a également été possible de déterminer la participation de chaque résidu d'acides aminés de l'anastelline dans l'interaction avec les oligosaccharides d'héparine (**Tableau XXVII**) et d'héparane sulfate (**Tableau XXVIII**). Les résidus d'arginine 646, 653 et 694 sont systématiquement retrouvés pour les 5 complexes et sont les plus impliqués dans l'interaction, comme en témoigne leur énergie libre d'interaction.

Oligosaccharides d'héparine	Résidus d'acides aminés de l'anastelline	$\Delta G$ (kcal.mol <sup>-1</sup> )
HE dp4	K640	-0,7
	R644	-5,5
	<b>R646</b>	<b>-6,2</b>
	K648	-3,4
	<b>R653</b>	<b>-10,2</b>
	W654	-4,4
	K655	-4,0
	K672	-1,4
	<b>R694</b>	<b>-4,6</b>
HE dp6	<b>R646</b>	<b>-10,8</b>
	K648	-3,5
	<b>R653</b>	<b>-10,8</b>
	W654	-4,8
	<b>R694</b>	<b>-7,8</b>
HE dp8	R644	-4,7
	<b>R646</b>	<b>-11,5</b>
	K648	-6,1
	<b>R653</b>	<b>-8,1</b>
	K669	-7,1
	K672	-5,7
	<b>R694</b>	<b>-8,4</b>

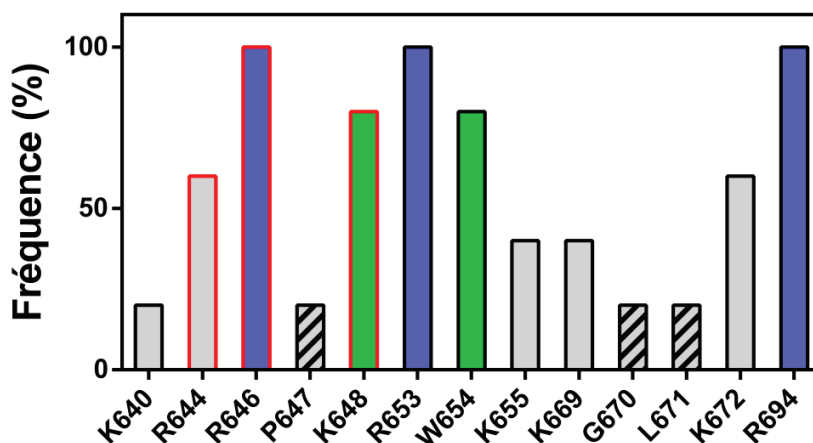
**Tableau XXVII : Décomposition de l'énergie libre d'interaction par résidus d'acides aminés de l'anastelline dans son interaction avec les oligosaccharides d'héparine.** Les résidus en gras sont ceux retrouvés dans tous les complexes. Les numéros des résidus d'acides aminés correspondent à la numérotation UniProt KB de la fibronectine.



Oligosaccharides d'héparane sulfate	Résidus d'acides aminés de l'anastelline	$\Delta G$ (kcal.mol <sup>-1</sup> )
<b>HSNA dp4</b>	<b>R646</b>	<b>-1,9</b>
	P647	-1,1
	<b>R653</b>	<b>-1,9</b>
	W654	-1,6
	K655	-2,1
	K669	-1,1
	G670	-0,7
	L671	-1,1
	K672	-1,5
	<b>R694</b>	<b>-1,7</b>
<b>HSNS dp6</b>	R644	-5,3
	<b>R646</b>	<b>-6,6</b>
	K648	-1,5
	<b>R653</b>	<b>-7,7</b>
	W654	-3,9
	<b>R694</b>	<b>-6,1</b>

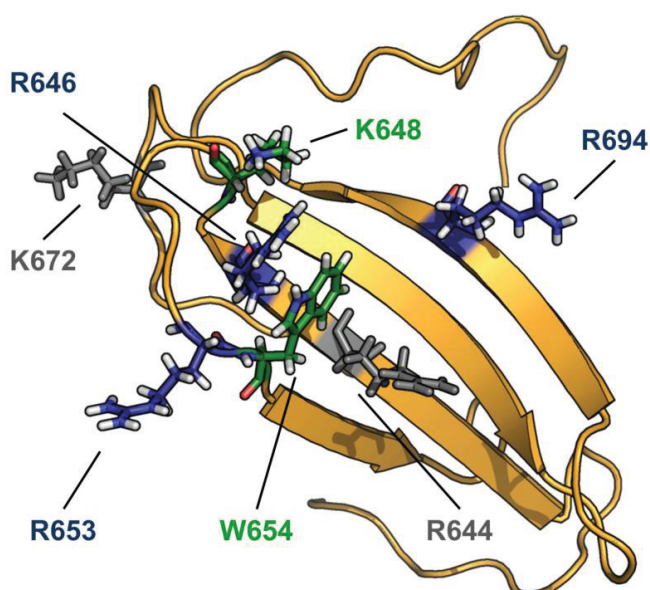
**Tableau XXVIII : Décomposition de l'énergie libre d'interaction par résidus d'acides aminés de l'anastelline dans son interaction avec les oligosaccharides d'héparane sulfate.** Les résidus en gras sont ceux retrouvés dans tous les complexes. Les numéros des résidus d'acides aminés correspondent à la numérotation UniProt KB de la fibronectine.

L'analyse des résidus d'acides aminés de l'anastelline interagissant avec les oligosaccharides montre que les résidus de lysine 648 et de tryptophane 654 sont également représentés et en moindre mesure les résidus d'arginine 644 et de lysine 672 (**Figure 129**). Excepté le résidu de tryptophane, les résidus apolaires (Pro 647, Gly 670 et Leu 671) sont uniquement impliqués dans l'interaction avec le dp4 d'héparane sulfate non sulfaté (HSNA). L'ensemble de ces résidus diffère partiellement des précédentes observations (*Gui et al., 2006*). En effet, seuls les résidus d'arginine 644 et 646 ainsi que de lysine 648 avaient été identifiés (**Figure 129**). Le résidu de tryptophane 645 n'est en revanche pas retrouvé à l'inverse du résidu de tryptophane 654 dans nos expériences.



**Figure 129 : Fréquence d'apparition des résidus d'acides aminés de l'anastelline interagissant avec les oligosaccharides.** Les résidus d'acides aminés en bleu sont systématiquement retrouvés dans les expériences de docking avec les 5 oligosaccharides (100 %), ceux en vert le sont fréquemment (80 %) et les autres sont en gris. Les résidus hachurés sont uniquement retrouvés dans l'interaction avec le dp4 d'héparane sulfate non sulfaté (HSNA). Les résidus où les bords sont en rouge sont ceux retrouvés aussi bien dans nos expériences de docking que dans les travaux de *Gui et al.*, (2006).

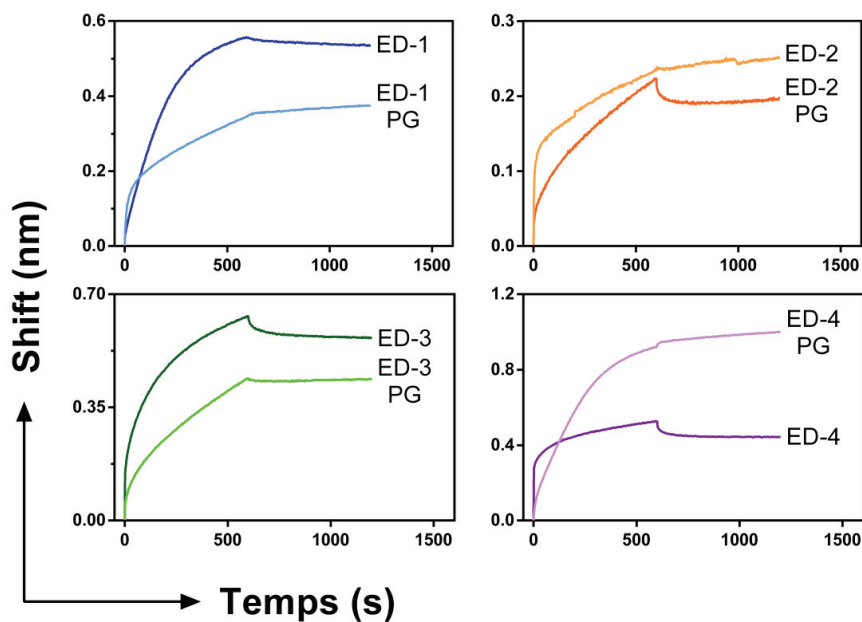
Les résidus d'acides aminés de l'anastelline étant impliqués dans l'interaction avec les oligosaccharides sont situés à son extrémité N-terminale, notamment dans le premier et deuxième brin  $\beta$ , à l'exception du résidu d'arginine 694 (**Figure 130**). L'ensemble des résidus forment cependant une face très basique responsable de l'interaction avec les oligosaccharides d'héparine et d'héparane sulfate.



**Figure 130 : Localisation des résidus d'acides aminés de l'anastelline impliqués dans l'interaction avec les oligosaccharides d'héparine et d'héparane sulfate.** Les résidus en bleu, vert et gris sont ceux retrouvés respectivement dans 100, 80 et 60 % des expériences de docking avec les différents oligosaccharides.

## 2. COMPLEXES ANASTELLINE – SYNDECANS

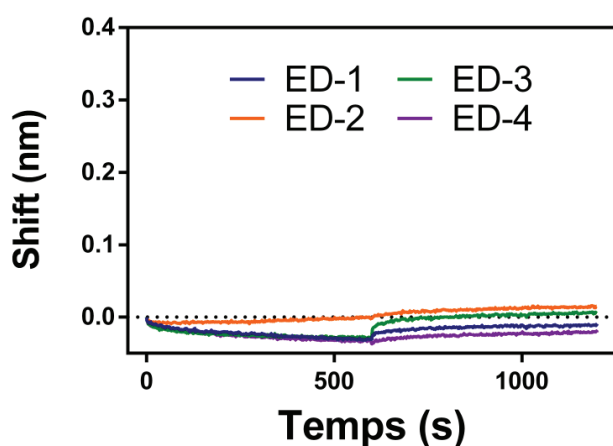
Les syndécans ont été proposés comme corécepteurs de l'anastelline mais aucune preuve expérimentale n'a été apportée (Gui *et al.*, 2006). Afin de vérifier cette hypothèse, des expériences d'interactions ont été réalisées par BLI. L'anastelline a été immobilisée de façon covalente et les ectodomaines des syndécans-1 à -4 humains portant leurs chaînes de glycosaminoglycans d'origine commerciale (R&D systems) ont été utilisés en tant qu'analytes (**Figure 131**). Le capteur où a été immobilisé l'anastelline a également été mis en contact avec les ectodomaines des syndécans dépourvus de leurs chaînes de GAGs (**Figure 131**).



**Figure 131 : Détection de l'interaction entre l'anastelline et les syndécans.** Sensorgrammes obtenus par contact entre l'anastelline immobilisée et les syndécans portant (ED-PG) ou non (ED) leurs chaînes de glycosaminoglycans à une concentration de 300 nM et 5  $\mu$ M, respectivement. ED-4PG correspond à l'isoforme 1 du SDC-4 et ne contient pas de pont disulfure à l'inverse de l'isoforme 2 (ED-4). Les sensorgrammes des expériences contrôles ont été soustraits.

Ces expériences montrent que l'anastelline interagit avec les quatre syndécans portant leurs chaînes de GAGs, notamment le syndécane-4. D'une façon similaire aux interactions entre les EDs et le FGF-2 réalisées précédemment, les complexes sont très stables puisque quasiment aucune dissociation n'est observée en tampon. La même série d'expériences réalisée en utilisant les ectodomaines des syndécans sans GAGs montre les mêmes résultats. La partie protéique des 4 ectodomaines des syndécans a la capacité d'interagir avec l'anastelline et forme des complexes stables, notamment pour ED-1 et ED-3, alors qu'une très faible dissociation est observée pour ED-2 et ED-4 (**Figure 131**).

La fibronectine, contenant le domaine FNIII<sub>1</sub> à l'origine de l'anastelline, a également été immobilisée et le capteur a été mis en contact avec les ectodomaines des syndécans sans GAGs mais aucune interaction n'a été détectée (**Figure 132**). Cette observation permet de suggérer qu'un site cryptique est masqué dans la fibronectine intacte et se retrouve exposé lors de la libération de l'anastelline.



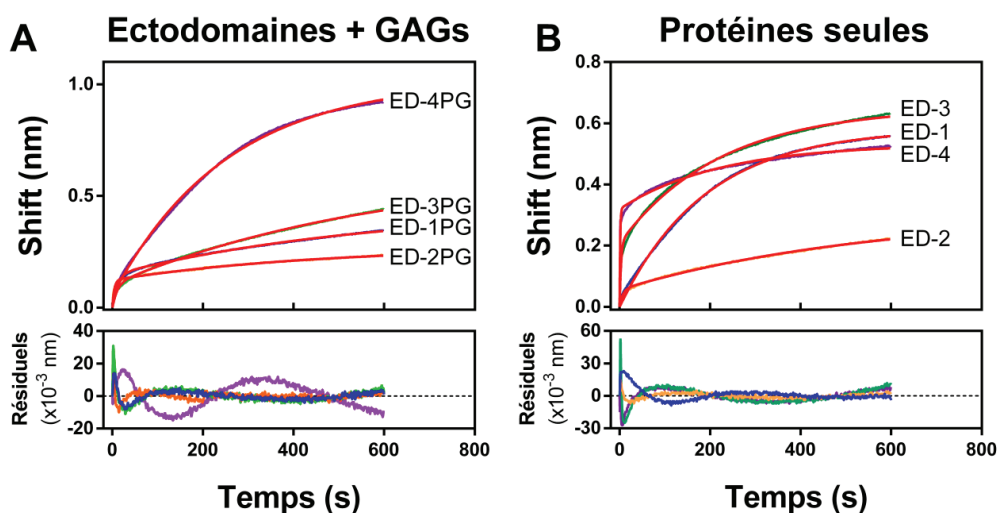
**Figure 132 : La fibronectine n'interagit pas avec les ectodomaines des syndécans sans leurs chaînes de glycosaminoglycane.** Sensorgrammes obtenus par contact entre la fibronectine immobilisée et les syndécans dépourvus de leurs chaînes de glycosaminoglycane à une concentration de 5  $\mu$ M, respectivement. Les sensorgrammes des expériences contrôles ont été soustraits.

L'utilisation d'une unique concentration en analyte dans ces expériences et l'absence de dissociation empêchent le calcul des constantes cinétiques et de la constante d'équilibre de dissociation. Seule la valeur de  $k_{obs}$  a été déduite de la phase d'association. Trois modèles ont été utilisés dans l'analyse de la phase d'association des partenaires. Le plus simple d'entre eux, le modèle 1 : 1, définit un unique site d'interaction à la surface du ligand (anastelline) et de l'analyte (EDs  $\pm$  GAGs). Le modèle 2 : 1 (ou ligand hétérogène) suppose l'existence de deux sites distincts sur le ligand et d'un seul sur l'analyte. Le modèle 1 : 2 (ou analyte bivalent) définit à l'inverse l'existence de deux sites d'interaction sur l'analyte (**Tableau XXIX**).

Le modèle de ligand hétérogène décrit le mieux l'interaction entre l'anastelline et les syndécans portant ou non leurs chaînes de GAGs. Cette observation suppose l'existence de deux sites distincts sur l'anastelline et d'un seul sur les ectodomaines des syndécans. Concernant ED-1, les valeurs de  $k_{obs}$  pour ces deux sites sont identiques, suggérant ainsi qu'il n'existe pas de site préférentiel sur l'anastelline, à l'inverse des autres ectodomaines des syndécans. Les sensorgrammes possédant les meilleurs ajustements sont présentés dans la **Figure 133**.

Analytes	Modèles utilisés pour la détermination de $k_{obs}$ ( $s^{-1}$ )		
	Simple 1 : 1	Ligand hétérogène 2 : 1	Analyte bivalent 1 : 2
<b>Ectodomaines des syndécans portant leurs chaînes de GAGs (protéoglycanes)</b>			
ED-1 PG	$9,4 \cdot 10^{-3}$ (0,451/0,81)	<b><math>1,7 \cdot 10^{-3} / 9,0 \cdot 10^{-2}</math> (0,004/1,00)</b>	$1,4 \cdot 10^{-3}$ (0,883/0,85)
ED-2 PG	$1,8 \cdot 10^{-2}$ (0,311/0,54)	<b><math>2,2 \cdot 10^{-3} / 2,5 \cdot 10^{-1}</math> (0,002/1,00)</b>	$4,6 \cdot 10^{-3}$ (0,567/0,65)
ED-3 PG	$4,7 \cdot 10^{-3}$ (0,339/0,94)	<b><math>9,6 \cdot 10^{-2} / 1,6 \cdot 10^{-3}</math> (0,009/1,00)</b>	$9,4 \cdot 10^{-4}$ (0,419/0,96)
ED-4 PG	$4,6 \cdot 10^{-3}$ (0,148/1,00)	<b><math>3,9 \cdot 10^{-1} / 4,0 \cdot 10^{-3}</math> (0,039/1,00)</b>	$9,8 \cdot 10^{-4}$ (0,618/0,99)
<b>Ectodomaines des syndécans sans GAGs (protéines seules)</b>			
ED-1	<b><math>5,3 \cdot 10^{-3}</math> (0,016/1,00)</b>	<b><math>5,3 \cdot 10^{-3} / 5,3 \cdot 10^{-3}</math> (0,016/1,00)</b>	<b><math>2,0 \cdot 10^{-3}</math> (0,021/1,00)</b>
ED-2	$5,2 \cdot 10^{-3}$ (0,106/0,93)	<b><math>1,5 \cdot 10^{-1} / 1,8 \cdot 10^{-3}</math> (0,003/1,00)</b>	$3,2 \cdot 10^{-3}$ (0,094/0,95)
ED-3	$1,0 \cdot 10^{-2}$ (1,057/0,87)	<b><math>3,2 \cdot 10^{-1} / 4,5 \cdot 10^{-3}</math> (0,027/1,00)</b>	$8,2 \cdot 10^{-3}$ (0,412/0,96)
ED-4	$4,4 \cdot 10^{-2}$ (1,231/0,38)	<b><math>8,2 \cdot 10^{-1} / 4,5 \cdot 10^{-3}</math> (0,019/0,99)</b>	$2,3 \cdot 10^{-1}$ (0,297/0,86)

**Tableau XXIX : Valeurs de la constante  $k_{obs}$  lors de la formation des complexes anastelline – ectodomaines des syndécans portant ou non leurs chaînes de GAGs.** Le biocapteur sur lequel a été immobilisée l’anastelline a été trempé dans des solutions d’ectodomaines des syndécans portant leurs chaînes de glycosaminoglycanes (ED-PG) ou non (ED). Les valeurs de  $k_{obs}$  sont indiquées pour les trois modèles testés ainsi que la qualité des ajustements ( $\chi^2 / R^2$ ) où les plus pertinents sont en gras.

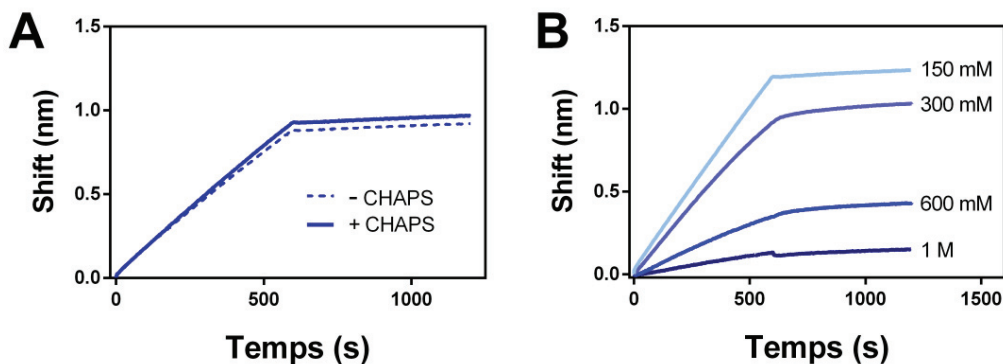


**Figure 133 : Détermination de la constante  $k_{obs}$  lors de la formation des complexes anastelline – ectodomaines des syndécans portant ou non leurs chaînes de GAGs.** Le modèle de ligand hétérogène représente le mieux les sensorgrammes de l’interaction entre l’anastelline et les ectodomaines des syndécans portant leurs chaînes de glycosaminoglycanes (A) ainsi que ceux de l’interaction entre l’anastelline et les ectodomaines des syndécans sans GAGs (B). Les courbes expérimentales (courbes colorées), les ajustements (en rouge) et les résiduels respectifs sont présentés.

### 3. CARACTERISATION DE L'INTERACTION ENTRE L'ANASTELLINE ET L'ECTODOMAINE DU SYNDECAN-1

L'objectif des expériences suivantes a été de déterminer quels facteurs étaient déterminants dans la stabilité de l'interaction entre l'ectodomaine du syndécan-1 dépourvu de ses chaînes de GAGs et l'anastelline. L'anastelline interagit avec la fibronectine d'une part *via* des liaisons hydrogènes impliquées dans la formation de feuillets  $\beta$ , mais aussi par des interactions hydrophobes. Ces dernières sont induites par l'exposition du cœur hydrophobe de l'anastelline, causée par la perte de deux brins- $\beta$  le masquant dans le module FNIII originel. Ces interactions permettent la polymérisation de la fibronectine en superfibronectine. La formation de ces agrégats est cependant inhibée lorsque l'anastelline est incubée en présence de CHAPS, un détergent, qui vient masquer le cœur hydrophobe de l'anastelline (Briknarová *et al.*, 2003).

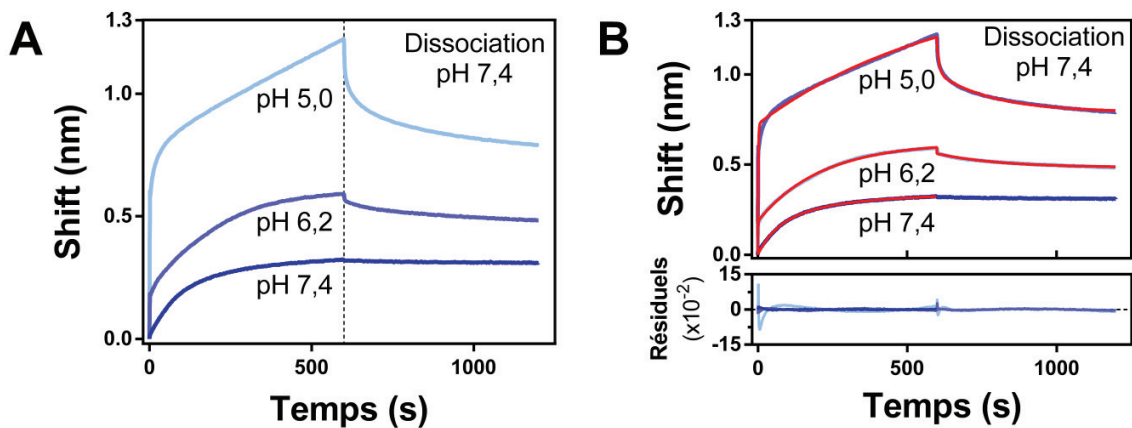
Afin de déterminer quels types d'interactions régissent la formation du complexe ED-1 – anastelline, une série d'expériences de BLI a été menée en immobilisant l'anastelline et en la mettant en contact avec ED-1 soit en présence de CHAPS, soit en présence de concentrations croissantes en NaCl, afin de déterminer respectivement si des interactions hydrophobes ou ioniques / hydrogènes étaient à l'origine de la formation du complexe (Figure 134). Le CHAPS n'a pas significativement altéré l'interaction entre l'anastelline et ED-1 (Figure 134A). En revanche, lorsque les expériences sont réalisées en présence de NaCl (de 0,15 M à 1 M dans la solution d'ED-1), une diminution graduelle de l'interaction est observée (Figure 134B). Environ un quart (23 %) du signal est perdu en présence de NaCl 0,3 M et plus de deux-tiers (71 %) en présence de NaCl 0,6 M. L'interaction reste cependant détectable en présence d'une concentration de 1 M en NaCl bien que 89 % du signal ait été perdu.



**Figure 134 : Sensorgrammes enregistrés par BLI correspondant à l'interaction entre ED-1 et l'anastelline en présence de CHAPS ou de concentrations croissantes de NaCl.** L'anastelline a été immobilisée et ED-1 (5  $\mu$ M) a été utilisé en tant qu'analyte (A) en présence (trait plein) ou en absence (pointillé) de CHAPS 3 mM ou (B) en présence de concentrations croissantes en NaCl.

Ces observations suggèrent que l'interaction entre l'anastelline et ED-1 est très certainement médiée par des interactions électrostatiques et / ou des liaisons hydrogènes. Il est intéressant également de noter que la forme globale des sensorgrammes diffère selon les différentes expériences. Par exemple, la phase d'association d'ED-1 est quasi-linéaire en présence de NaCl 150 mM (**Figure 134B**) en comparaison de cette même phase lors de la première série d'expériences (*voir Figure 131*). Les shifts observés diffèrent également. Il semblerait que le lot d'anastelline commerciale influe sur l'interaction entre l'anastelline et ED-1.

Des séries d'expériences de BLI ont été réalisées en suivant le même protocole mais en trempant le capteur où a été immobilisée l'anastelline dans des solutions d'ED-1 à différents pH (pH 7,4, 6,2 ou 5,0). L'étape de dissociation a été réalisée dans un tampon à pH 7,4 (**Figure 135**). Les pH acides favorisent fortement l'interaction entre l'anastelline et ED-1 d'un facteur environ 2 à pH 6,2 et 4 à pH 5,0 d'après les valeurs des shifts (**Figure 135A**). En revanche, les complexes se formant à pH acide se dissocient partiellement à pH 7,4, alors que le complexe formé à pH 7,4 reste très stable comme observé précédemment. Le modèle du ligand hétérogène décrit le mieux l'interaction entre l'anastelline et l'ectodomaine du syndécan-1 quel que soit le pH de la solution d'ED-1 (**Figure 135B** et **Tableau XXX**).



**Figure 135 : Influence du pH sur l'interaction entre l'anastelline et l'ectodomaine du syndécan-1 dépourvu de ses chaînes de GAGs. A)** Sensorgrammes obtenus avec de l'anastelline immobilisée mise en contact avec une solution d'ectodomaine de syndécan-1 dépourvu de chaînes de GAGs (5  $\mu$ M) dont le pH a été ajusté à 7,4, 6,2, ou 5,0. La phase de dissociation a été effectuée à pH 7,4. **B)** Détermination des paramètres cinétiques avec un modèle 2 : 1 (ligand hétérogène). Les ajustements sont en rouge et les données résiduelles sont en nuances de bleu, identique à (A).

pH de la solution d'ED-1	Modèles et ajustements ( $\chi^2 / R^2$ )		
	Simple 1 : 1	Ligand hétérogène 2 : 1	Analyte bivalent 1 : 2
pH 7,4	0,021 / 0,993	<b>0,002 / 0,999</b>	0,004 / 0,999
pH 6,2	0,985 / 0,870	<b>0,006 / 0,999</b>	0,647 / 0,921
pH 5,0	8,701 / 0,315	<b>0,127 / 0,990</b>	4,355 / 0,796

**Tableau XXX : Qualité de l'ajustement des modèles lors de la formation du complexe anastelline – ectodomaine du syndécan-1 à différents pH.** La qualité des ajustements ( $\chi^2 / R^2$ ) est indiquée pour chaque modèle et pour chaque condition expérimentale. Seule la phase d'association a été prise en compte lors de l'évaluation de la qualité des ajustements pour les trois modèles lorsqu'aucune dissociation du complexe anastelline – ectodomaine du syndécan-1 (ED-1) n'a été observée (expérience à pH 7,4). La totalité du sensorgramme a été utilisé pour les autres conditions expérimentales étant donné que le complexe se dissocie partiellement. Le modèle ayant le meilleur ajustement est en gras.

Etant donné que le complexe se dissocie partiellement à pH acide, il a été possible d'extraire des paramètres cinétiques (**Figure 135B** et **Tableau XXXI**). Bien que ces résultats semblent refléter l'allure des sensorgrammes, ces valeurs restent approximatives étant donné qu'une seule concentration en analyte a été utilisée dans ces expériences qui n'ont été réalisées qu'une seule fois.

pH	Paramètres cinétiques			
	$k_{obs}$ ( $s^{-1}$ )	$k_a$ ( $M^{-1}.s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$K_D$ (M)
7,4	$1,6.10^{-2}$	-	-	-
6,2	$4,7.10^{-3}$ 1,3	9,0 $2,5.10^5$	$4,7.10^{-3}$ $4,7.10^{-3}$	$5,2.10^{-6}$ $1,9.10^{-8}$
5,0	$1,6.10^{-3}$ $6,3.10^{-1}$	<i>ND</i> $1,3.10^5$	$1,3.10^{-1}$ $5,3.10^{-3}$	<i>ND</i> $4,2.10^{-8}$

**Tableau XXXI : Détermination des paramètres cinétiques de l'interaction entre l'anastelline et l'ectodomaine du syndécan-1.** Le modèle du ligand hétérogène (2 : 1) a été utilisé. *ND* : valeur non déterminée. Les valeurs non calculables sont indiquées par des "-".  $k_a$  et  $k_d$  : constantes cinétiques d'association et de dissociation, respectivement.  $K_D$  : constante d'équilibre de dissociation.

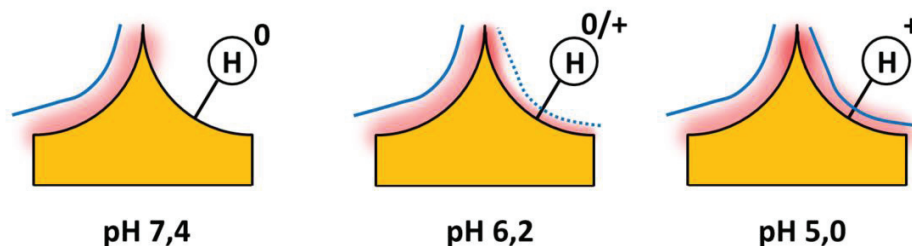


La comparaison des  $k_{obs}$  entre les différentes conditions montre que la formation du complexe anastelline – ED-1 est favorisée à pH acide ( $4,7 \cdot 10^{-3} \text{ s}^{-1} / 1,3 \text{ s}^{-1}$  et  $1,6 \cdot 10^{-3} \text{ s}^{-1} / 6,3 \cdot 10^{-1} \text{ s}^{-1}$  à pH 6,2 et 5,0 respectivement, contre  $1,6 \cdot 10^{-2} \text{ s}^{-1}$  à pH 7,4). La détermination des valeurs de constantes cinétiques d'association ( $k_a$ ) et de dissociation ( $k_d$ ) a permis de calculer la constante d'équilibre de dissociation ( $K_D$ ) reflétant l'affinité des deux partenaires, mais uniquement pour l'un des deux sites de l'anastelline car une valeur de  $k_a$  n'a pas pu être calculée lors de l'analyse. L'affinité est élevée et est environ deux fois plus importante à pH 6,2 qu'à pH 5,0 (19 et 42 nM, respectivement).

Bien que la valeur des paramètres cinétiques de la formation du complexe soit approximative du fait de l'utilisation d'une seule concentration en analyte, un mécanisme d'interaction est proposé. La diminution du pH pourrait provoquer l'apparition d'un second site d'interaction sur l'anastelline. A pH 6,2, l'étape d'association présente une phase très rapide puis une seconde phase plus lente (qui pourrait caractériser la fixation d'ED-1 sur un 2<sup>ème</sup> site). A pH 5,0, deux sites sont observables sur le sensorgramme par la présence de deux phases d'association. ED-1 s'associe très rapidement avec l'anastelline et pourrait saturer alors le premier site puis interagir avec un second site. Le modèle 2 : 1 reflète alors parfaitement ce phénomène (*voir Figure 135*).

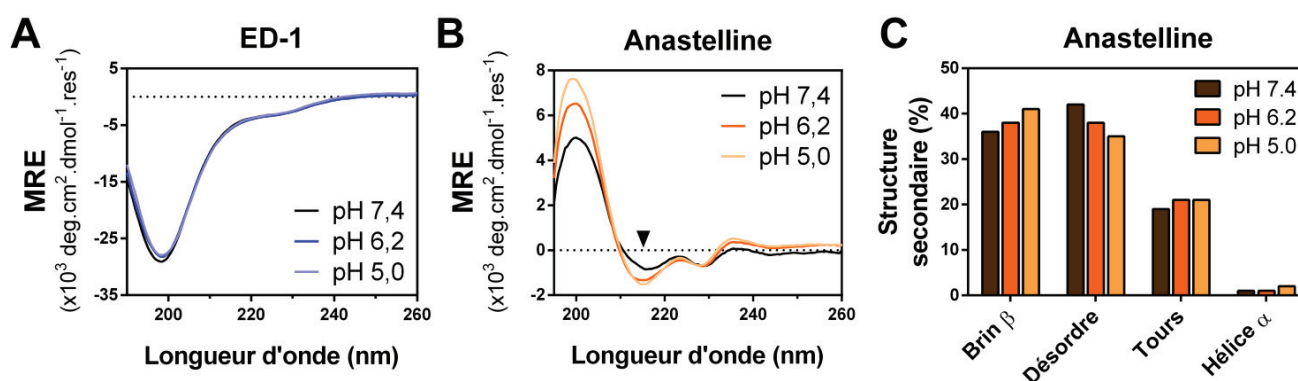
L'étape de dissociation des complexes formés aux pH acides s'effectue à pH 7,4 où une dissociation du complexe est observée. Cette dissociation pourrait être causée par la perte partielle du second site d'interaction, qui s'était formé à pH acide. Néanmoins, les sensorgrammes associés aux expériences réalisées aux pH acides ne regagnent pas la ligne de base, ce qui indique que tous les complexes ne se dissocient pas.

Comme rappelé précédemment, l'anastelline est une protéine très basique à l'inverse d'ED-1. Une explication probable serait donc l'apparition d'un nouveau patch basique sur l'anastelline, formant ainsi un second site favorable à l'interaction avec ED-1, qui est une protéine acide. Seule la fonction imidazole portée par les résidus d'histidine possède un pKa compris entre 5,0 et 7,4 (pKa 6,0), ce qui indique qu'à pH 6,2, environ 50 % des fonctions sont sous forme imidazolium, donc chargées positivement, et environ 91 % à pH 5,0. La formation d'un nouveau patch positif causée par la diminution du pH pourrait être une explication cohérente avec les observations expérimentales. Ce mécanisme proposé est présenté en **Figure 136**.



**Figure 136 : Mécanisme d'interaction proposé entre l'anastelline et l'ectodomaine du syndécan-1.** A pH 7,4, l'anastelline (orange) présente principalement un unique site d'interaction (rouge) pour ED-1 (trait bleu) et un potentiel second site comportant un ou plusieurs résidus d'histidine non chargé ( $H^0$ ). A pH 6,2, environ 50 % des résidus d'histidine sont chargés ( $H^{0/+}$ ) créant un patch basique favorisant l'interaction avec ED-1 (trait en pointillé). A pH acide, les résidus d'histidine sont chargés ( $H^+$ ), favorisant l'interaction entre l'anastelline et ED-1.

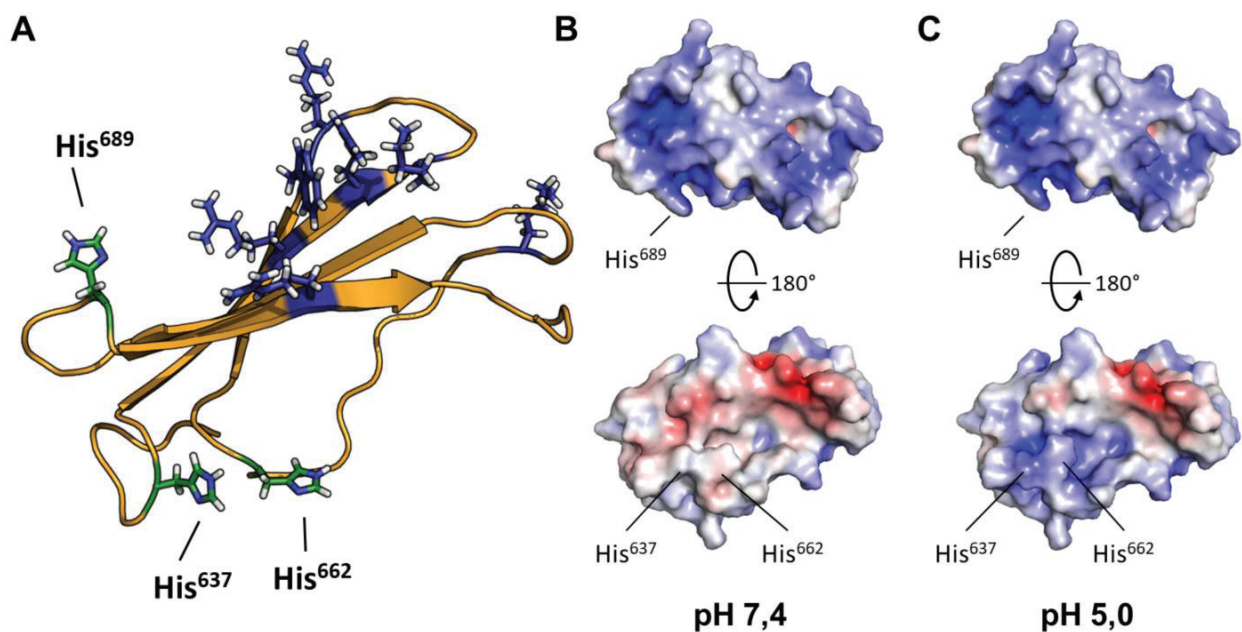
Une modification conformationnelle à pH acide de l'anastelline ou de l'ectodomaine du syndécan-1 pourrait également favoriser la formation du complexe. Afin de vérifier cette possibilité, des expériences de dichroïsme circulaire ont été réalisées (**Figure 137**). ED-1 reste intrinsèquement non structuré quel que soit le pH (**Figure 137A**). Concernant l'anastelline, les pics à 200 et à 215 nm gagnent graduellement en intensité et sont caractéristiques des brins  $\beta$  (**Figure 137B**). La déconvolution des spectres des expériences menées sur l'anastelline a confirmé cette observation, où l'anastelline gagne une très faible proportion de brin  $\beta$  (4 %, soit moins de 4 résidus) à pH 5,0 au profit du désordre intrinsèque (- 7 %, soit environ 5 résidus, **Figure 137C**).



**Figure 137 : Expériences de dichroïsme circulaire sur l'ectodomaine du syndécan-1 et l'anastelline à différents pH.** Spectres soustraits, moyennés et lissés de l'ectodomaine du syndécan-1 (**A**) et de l'anastelline (**B**) aux pH 7,4, 6,2 et 5,0. MRE : ellipticité moyenne par résidu. La pointe de flèche en **B**) est située à 215 nm. **C**) Proportion en structures secondaires (%) après déconvolution des spectres enregistrés sur l'anastelline par Dichroweb avec l'algorithme CDSSTR et le jeu de données 7.

Les expériences suivantes ont été réalisées dans le but d'apporter des éléments soutenant l'hypothèse de la formation d'un second site basique à la surface de l'anastelline, notamment par la protonation de ses résidus d'histidine. Afin de visualiser leur localisation à la surface de l'anastelline, sa structure tridimensionnelle (PDB : 1Q38, *Briknarová et al., 2003*) a été modifiée afin de concorder avec la séquence de la protéine recombinante disponible commercialement et utilisée dans nos expériences (voir section *Matériel & Méthodes*).

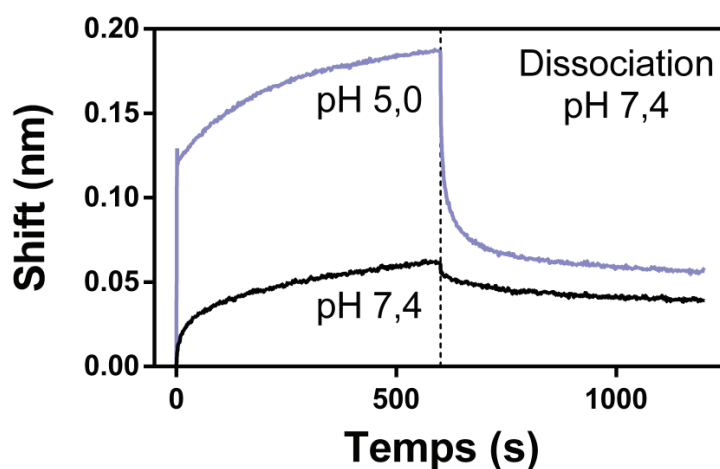
Cette-dernière ne possédant pas d'étiquette, ces modifications ont principalement porté sur l'élimination de l'étiquette 6xHis présente à son extrémité C-terminale. Les sites d'interaction des oligosaccharides d'héparine et d'héparane sulfate précédemment observés lors des expériences de docking sont également représentés et le potentiel électrostatique de la protéine a été généré en utilisant la structure telle quelle, ou en protonant les résidus d'histidine afin de mimer ses charges à pH 5 (**Figure 138**).



**Figure 138 : Localisation des résidus d'histidine et du site de fixation des oligosaccharides d'héparine et d'héparane sulfate sur la structure 3D de l'anastelline. A)** Structure 3D de l'anastelline possédant la même séquence que la protéine commerciale. Les résidus d'histidine et le site d'interaction à l'héparine sont en vert et bleu, respectivement. **B)** Potentiel électrostatique de surface de l'anastelline mimant un pH de 7,4 ou **C)** un pH de 5,0. La numérotation des résidus d'histidine correspond à celle de la fibronectine entière (UniProt : P02751).

La séquence de l'anastelline comporte trois résidus d'histidine (637, 662 et 689). Deux d'entre eux (637 et 662) sont localisés dans des régions flexibles (extrémité N-terminale et région désordonnée, respectivement), mais sont retrouvés à proximité (**Figure 138A**). D'après le potentiel électrostatique de l'anastelline mimé pour un pH de 5, la protonation de ces deux résidus pourrait entraîner la formation d'un nouveau patch basique qui n'est pas présent à pH neutre (**Figures 138B et 138C**). Le résidu d'His<sup>689</sup> est en revanche situé à proximité du site de fixation des oligosaccharides d'héparine et d'héparane sulfate. La protonation de ce résidu, mais aussi des résidus d'histidine 632 et 667, pourrait favoriser l'interaction entre l'anastelline et l'héparine.

La formation du complexe entre des chaînes d'HS ou ED-1 et l'anastelline semble mettre en jeu dans les deux cas des interactions électrostatiques. De ce fait, l'apparition d'une nouvelle région basique à la surface de l'anastelline devrait également promouvoir la formation de complexes entre l'anastelline et des oligosaccharides sulfatés. Cette hypothèse a été testée par BLI en immobilisant l'anastelline et la mettant en contact avec une solution d'octasaccharide d'héparine (dp8) dont le pH a été ajusté à 7,4 ou 5,0, puis en laissant le complexe se dissocier à pH 7,4 (**Figure 139**).



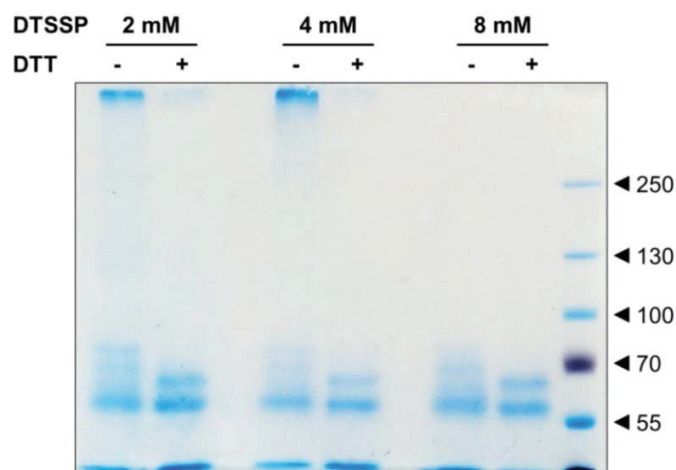
**Figure 139 : Rôle du pH dans l'interaction entre l'anastelline et l'héparine.** Sensorgrammes obtenus lorsque l'anastelline immobilisée sur un capteur a été trempée dans une solution d'octasaccharide d'héparine (séquence [ HexUA(2S) – GlcNS(6S)– (IdoUA(2S) – GlcNS(6S)x3 ] où HexUA(2S), GlcNS(6S) et IdoUA(2S) représentent un résidu d'acide hexuronique 2-sulfaté, de glucose N- et 6-sulfaté ou d'acide iduronique 2-sulfaté, respectivement). La solution d'octasaccharide d'héparine a été diluée à une concentration de 5  $\mu$ M à pH 7,4 (noir) ou pH 5,0 (bleu). La phase de dissociation a été effectuée à pH 7,4.

Les résultats de cette expérience sont assez analogues à ceux obtenus entre l'anastelline et ED-1 à différents pH. La valeur du signal observé à pH 5,0 est augmentée d'un facteur 3 par rapport à pH 7,4. Le complexe se dissocie d'avantage à pH 5,0 où le signal diminue à des valeurs semblables à celles obtenues lorsque l'expérience est réalisée à pH 7,4 (**Figure 139**). Cette observation pourrait être la conséquence de la protonation des résidus d'histidine ce qui favoriserait l'interaction de l'octasaccharide d'héparine avec l'anastelline.

Dans le but de localiser le ou les sites d'interaction de l'ectodomaine du syndécan-1 et de l'anastelline, des expériences de spectrométrie de masse en tandem ont été réalisées. ED-1 et l'anastelline ont été covalamment réticulés par le 3,3'-dithiobis(sulfosuccinimidylpropionate), DTSSP, qui est un agent réticulant possédant deux fonctions sulfo-NHS susceptibles de réagir avec les fonctions amines primaires. Il contient également un pont disulfure qui peut être réduit. Trois conditions de réticulation ont préalablement été testées où ED-1 (20  $\mu$ M) et l'anastelline (60  $\mu$ M) ont été incubés à température ambiante pendant 30 min en présence de DTSSP 2, 4 et 8 mM. L'analyse de ces expériences de réticulations a été effectuée par électrophorèse en conditions dénaturantes avec ou sans réduction préalable avec du DTSSP (**Figure 140**).

En absence de DTT, environ 3 espèces distinguables ayant une masse moléculaire comprise entre 55 et 70 kDa sont détectées ainsi qu'une espèce migrant très peu dans le gel de séparation lorsque les protéines ont été incubées en présence de DTSSP 2 et 4 mM uniquement. Ces espèces ne correspondent pas aux formes solubles d'ED-1, dont la masse moléculaire apparente sur gel est d'environ 40 kDa et de l'anastelline, dont la masse moléculaire est de 8,4 kDa. Ces protéines sont visibles dans le front de migration du gel. La réduction des espèces a mené à la disparition de l'espèce de haute masse moléculaire et seules deux bandes sont observées entre 55 et 70 kDa quel que soit les conditions de réticulation (**Figure 140**). Cela suggère que l'espèce de haute masse moléculaire et celle de masse moléculaire apparente d'environ 70 kDa sont des produits de la réticulation par le DTSSP.

Une nouvelle expérience dont la réticulation a été effectuée en présence de DTSSP 4 mM a été réalisée pour l'analyse par spectrométrie de masse afin de favoriser la détection des peptides réticulés.

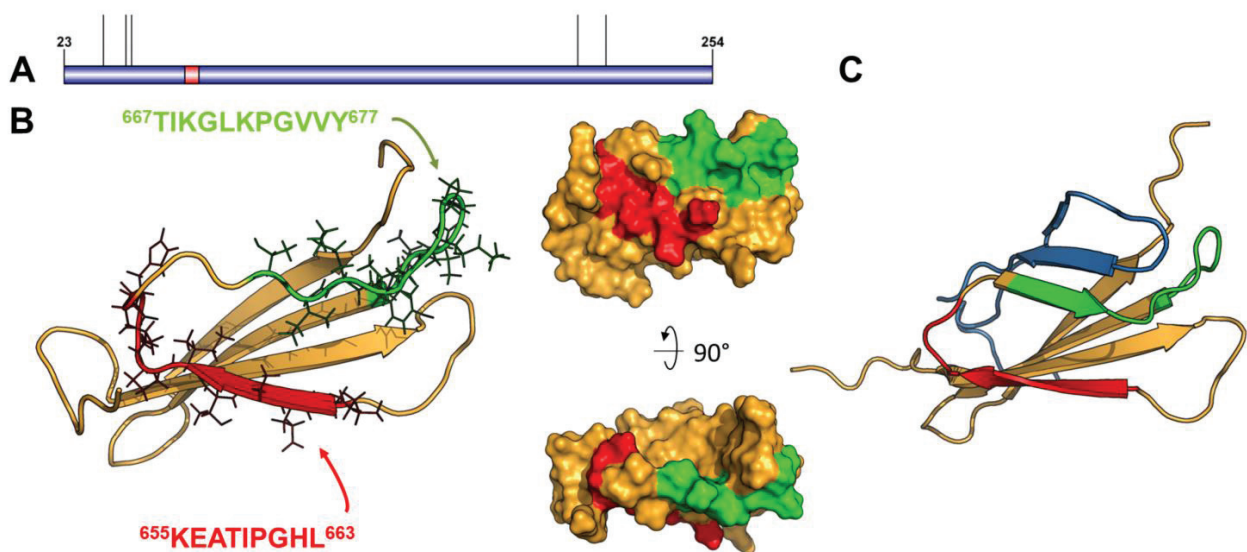


**Figure 140 : Analyse par électrophorèse des expériences de réticulation de l'ectodomaine du syndécan-1 et de l'anastelline par le DTSSP.** ED-1 (20  $\mu$ M) et l'anastelline (60  $\mu$ M) ont été incubés pendant 30 min à température ambiante en présence de 2, 4 ou 8 mM de DTSSP (volume réactionnel : 25  $\mu$ l). La réticulation a été stoppée par l'ajout de Tris-HCl pH 7,4 0,3 M. 15  $\mu$ l de tampon de charge non réducteur concentré 4 fois ont été ajoutés puis la solution (60  $\mu$ l au total) a été divisée en deux et du DTT à une concentration finale de 50 mM a été ajouté dans l'une d'entre elles (+ DTT) et la solution a été incubée 10 min à température ambiante. La seconde solution a été laissée telle quelle (- DTT). 20  $\mu$ l de chaque solution ont été déposés sur gel de polyacrylamide 7 % et les protéines ont été colorées toute la nuit par de l'InstantBlue.

Lors de l'analyse de spectrométrie de masse, les protéines réticulées ont été digérées par la chymotrypsine puis par la trypsin, afin de détecter des peptides d'ED-1 et d'anastelline réticulés. Les modifications engendrées par la réticulation par le DTSSP provoquent un écart de masse de + 174 Da dans le cas d'une liaison inter- ou intramoléculaire. Deux peptides modifiés ont été détectés lors de cette expérience. Le peptide  $^{66}\text{KDTQLL}^{71}$  d'ED-1 est réticulé avec le peptide  $^{655}\text{KEATIPGHL}^{663}$  de l'anastelline par leur résidu de lysine en position N-terminale. Le même peptide d'ED-1 est retrouvé réticulé soit avec le peptide  $^{667}\text{TIKGLKPGVVY}^{677}$  de l'anastelline, soit avec le peptide  $^{672}\text{KPGVVY}^{677}$ , selon le résidu de lysine impliqué dans la réticulation. Il n'est en revanche pas possible de quantifier les peptides réticulés. Afin de réduire le DTSSP et de l'empêcher de s'oxyder à nouveau, la même expérience a été réalisée après incubation des protéines réticulées par un agent réducteur (TCEP) suivie d'une alkylation par l'iodoacétamide. La réticulation par le DTSSP puis sa réduction et son alkylation induisent un écart de masse de + 145 Da. Les résidus de lysine 655, 669 et 672 de l'anastelline ont été systématiquement retrouvés modifiés avec cet écart de masse, confortant ainsi les résultats précédemment obtenus.

La séquence d'ED-1 est largement retrouvée pendant son séquençage (77 % et 83 % de couverture en condition non réductrice et réductrice, respectivement). En revanche, le résidu de lysine 66 d'ED-1 ne l'est jamais, ce qui peut également supporter l'implication du peptide dans l'interaction avec l'anastelline. Ces analyses sont présentées en **Annexe 5**. Le peptide de l'ectodomaine du syndécan-1 est localisé à son extrémité N-terminale de l'ectodomaine (**Figure 141A**). Les deux peptides d'anastelline sont retrouvés à proximité l'un de l'autre sur sa structure tridimensionnelle (**Figure 141B**). Cette observation conforte les analyses des expériences de BLI où le modèle de ligand hétérogène était le plus approprié pour décrire l'allure des sensorgrammes.

Le premier peptide (<sup>655</sup>KEATIPGHL<sup>663</sup>) est retrouvé au sein d'un brin- $\beta$  dans la structure 3D de l'anastelline. Ce peptide contient un résidu d'histidine (His<sup>662</sup>) dont l'existence avait été suggérée précédemment. Le second (<sup>667</sup>TIKGLKPGVVY<sup>677</sup>) se retrouve en revanche dans une région désordonnée de l'anastelline qui se replie en brin  $\beta$  lorsqu'elle est en interaction avec le domaine FNIII<sub>3</sub> de la fibronectine (**Figures 141B** et **141C**). Un mécanisme d'interaction entre ED-1 et l'anastelline pourrait impliquer un mécanisme de *domain swapping*, d'autant plus qu'ED-1 contient environ 16 % de brin  $\beta$ , comme l'ont indiqué les précédentes expériences de dichroïsme circulaire.



**Figure 141 : Localisation des peptides d'anastelline réticulés avec le peptide KDTQLL de l'ectodomaine du syndécan-1. A)** Localisation du peptide d'ED-1 réticulé (en rouge) sur l'ectodomaine entier (en bleu). Les chaînes de glycosaminoglycanes, absentes dans nos analyses, sont représentées par des traits verticaux. **B)** Structure 3D de l'anastelline (PDB : 1Q38) où sont indiquées la localisation et la séquence des peptides réticulés détectés par spectrométrie de masse. **C)** Structure 3D de l'anastelline en complexe avec un fragment du domaine FNIII<sub>3</sub> de la fibronectine (PDB : 5J6Z). La légende est identique à *B*) et le fragment de FNIII<sub>3</sub> est en bleu.

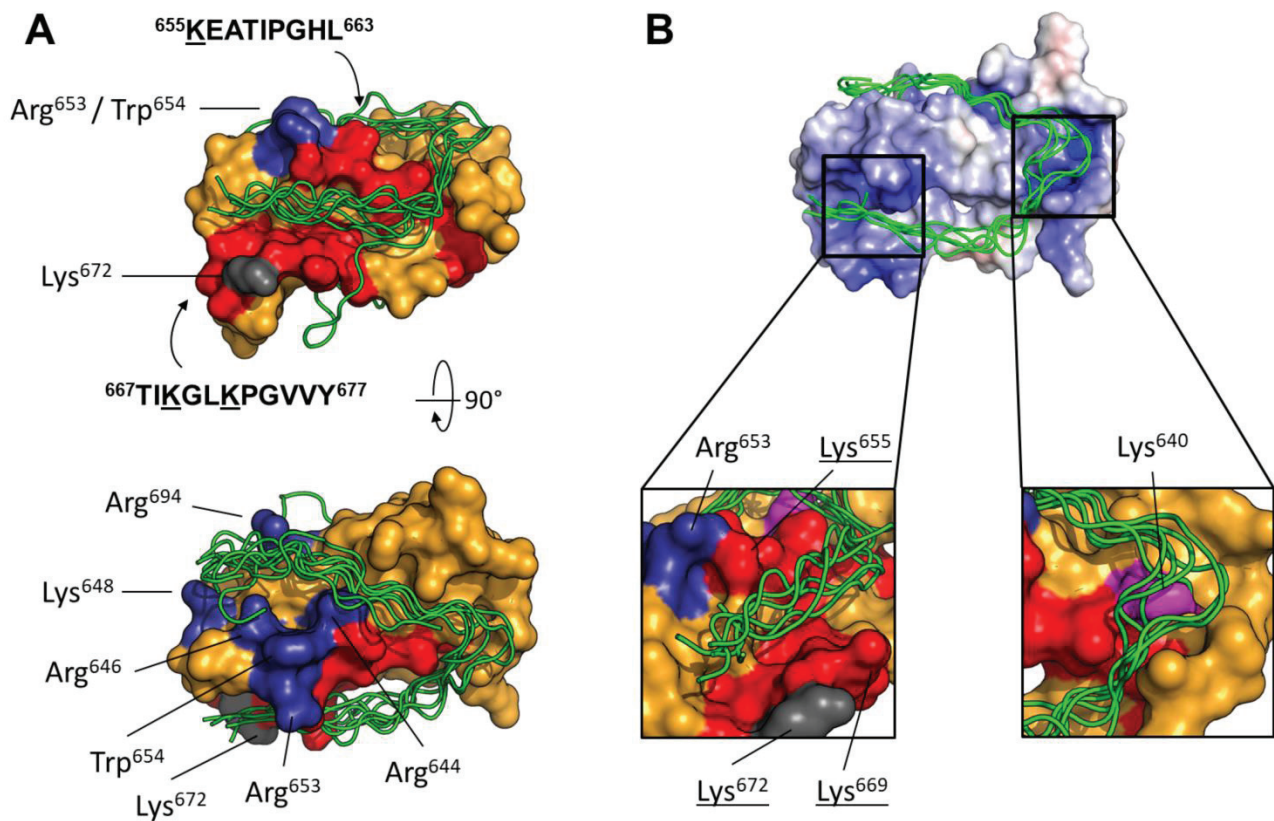
Les expériences de BLI en présence de concentrations croissantes de NaCl ont permis de conclure que des interactions électrostatiques et / ou des liaisons hydrogène étaient responsables de l'interaction entre ED-1 et l'anastelline. Si l'interaction est médiée par des liaisons électrostatiques entre des résidus basiques de l'anastelline d'une part et acides d'ED-1 d'autre part, il serait nécessaire que des résidus acides (glutamate ou aspartate) soient présents à proximité du peptide <sup>66</sup>KDTQLL<sup>71</sup> d'ED-1. Cette séquence possède le résidu d'aspartate 67, mais en revanche, les résidus acides les plus proches dans sa séquence sont les résidus d'aspartate 54 et de glutamate 79. Autrement dit, seuls 3 résidus acides sont présents sur une séquence de 26 résidus. En revanche, seul le résidu de lysine 66 est basique ce qui mène à un point isoélectrique théorique de 4,0 pour cette séquence de 26 résidus.

Des expériences de docking ont été réalisées entre l'anastelline, dont l'état de protonation est cohérent avec celui retrouvé à pH neutre, et un peptide d'ED-1 de 26 résidus d'acides aminés contenant le site détecté par spectrométrie de masse (<sup>66</sup>KDTQLL<sup>71</sup>) et les résidus acides Asp 54 et de Glu 79 aux extrémités N- et C-terminales. 9 poses ont été générées par le serveur CABS-Dock (*Kurcinski et al., 2019*), utilisé pour le docking protéine - peptide (**Figure 142**).

Parmi les 9 solutions, 8 interagissent avec l'anastelline dans un sillon entouré des deux peptides identifiés comme réticulés par spectrométrie de masse (**Figure 142A**). 6 poses sont superposables et interagissent en plus avec le site basique identifié dans l'interaction avec les oligosaccharides d'héparine ou d'héparane sulfate, et finalement 5 des 6 poses sont orientées de la même façon, c'est-à-dire que l'extrémité N-terminale du peptide d'ED-1 interagit dans le sillon formé par les deux peptides identifiés par spectrométrie de masse alors que leur extrémité C-terminale est à proximité des résidus interagissant avec les oligosaccharides d'héparine ou d'héparane sulfate (**Figure 142B**).

L'extrémité C-terminale de ces 5 poses est à proximité du résidu d'arginine 653 et des trois résidus de lysine 655, 669 et 672 qui ont été identifiés comme étant modifiés par le DTSSP dans les expériences de spectrométrie de masse (**Figure 142B**). Le résidu de lysine 66 du peptide réticulé d'ED-1 (<sup>66</sup>KDTQLL<sup>71</sup>) n'est pas retrouvé à proximité des résidus de lysine 655, 669 et 672 de l'anastelline mais est cependant à une distance comprise entre 12,5 et 15 Å de ces résidus ce qui est similaire à la taille du bras espaceur du DTSSP utilisé dans les expériences de spectrométrie de masse (12 Å).

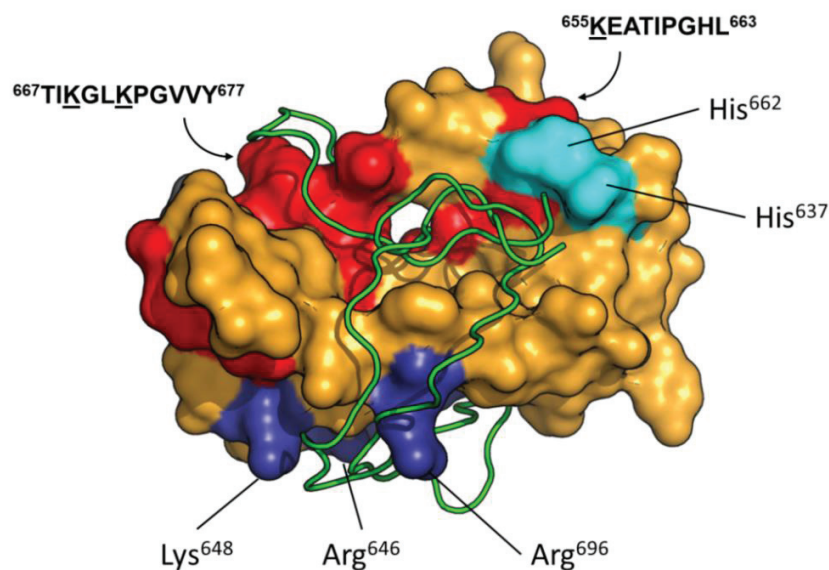




**Figure 142 : Docking du peptide de l'ectodomain du syndécan-1 sur l'anastelline. A)** Superposition des 9 poses du peptide (vert) sur l'anastelline (orange). Les peptides d'anastelline détectés par spectrométrie de masse sont en rouge sur la structure et sont détaillés où les résidus de lysine réticulés sont soulignés. Les 7 résidus retrouvés le plus fréquemment dans le docking des oligosaccharides avec l'anastelline sont indiqués sur la structure et colorés en bleu. **B)** Visualisation des 5 poses juxtaposables (vert) à la surface de l'anastelline dont le potentiel électrostatique a été simulé par APBS. Les deux zones basiques sont agrandies, les résidus basiques sont annotés et les résidus de lysine réticulés sont soulignés. Violet : résidu de lysine 640 non réticulé retrouvé à proximité des poses du peptide d'ED-1.

Les trois autres poses non superposables, dont celle étant la plus probable selon le serveur CABS-Dock, ont la caractéristique commune de se retrouver dans une poche contenant les résidus d'histidine 637 et 662, le résidu 662 étant dans un des peptides d'anastelline. Deux des trois poses sont retrouvées à proximité du site d'interaction avec les oligosaccharides d'héparine ou d'héparane sulfate précédemment déterminé (**Figure 143**).

Ces expériences de docking sont compatibles avec le fait qu'ED-1 soit d'une part à proximité des peptides d'anastelline détectés en spectrométrie de masse (8 poses / 9) ainsi que du site basique d'interaction des oligosaccharides d'héparine ou d'héparane sulfate déterminé précédemment (8 poses / 9), et d'autre part que certaines poses (3 / 9) soient à proximité des résidus d'histidine dont la protonation favoriserait l'interaction d'ED-1 sur ce site. En revanche, le résidu de lysine 66 réticulé dans le peptide d'ED-1 est à proximité du résidu de lysine 640 de l'anastelline qui n'a pas été retrouvé réticulé dans les expériences de spectrométrie de masse.



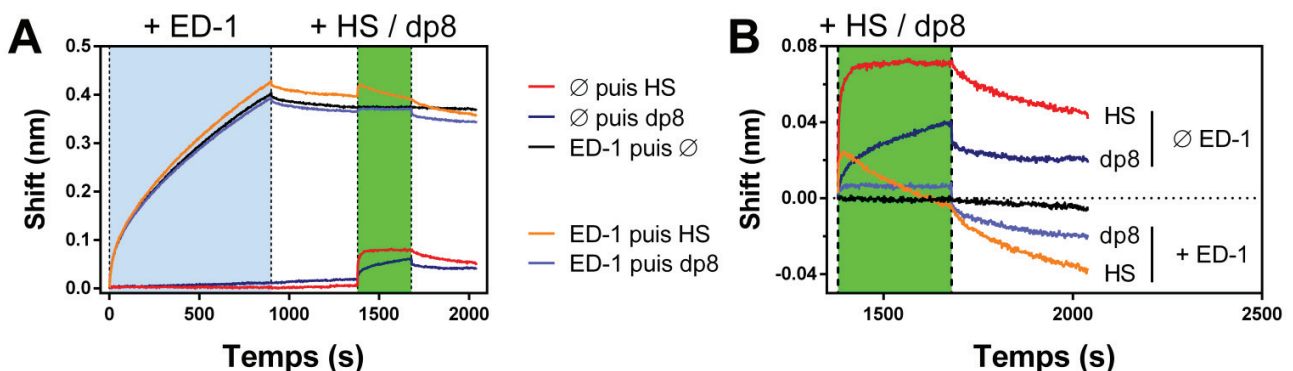
**Figure 143 : Docking du peptide de l'ectodomaine du syndécan-1 sur l'anastelline avec CABS-Dock.** Les 3 poses non superposables du peptide d'ED-1 (vert) sur l'anastelline (orange) sont représentées. Rouge : localisation et la séquence des peptides d'anastelline détectés par spectrométrie de masse où les résidus de lysine réticulés sont soulignés. Les 3 résidus interagissant avec les oligosaccharides d'héparine ou d'héparane sulfate sont en bleu foncé et les deux résidus d'histidine composant la poche sont en bleu cyan.

#### 4. COMPLEXE TERNAIRE ANASTELLINE – ED-1 – HEPARANE SULFATE

Les expériences de docking d'oligosaccharides d'héparine et d'héparane sulfate sur l'anastelline ont permis d'identifier au moins 7 résidus basiques impliqués dans la formation de ces complexes. Un certain nombre de ces résidus est également retrouvé à proximité du peptide d'ED-1 qui a été utilisé dans les expériences de docking avec l'anastelline. Ces observations suggèrent que les chaînes d'héparane sulfate portées par l'ectodomaine du syndécan-1 ainsi que sa partie protéique pourraient être en compétition dans leur interaction avec l'anastelline.

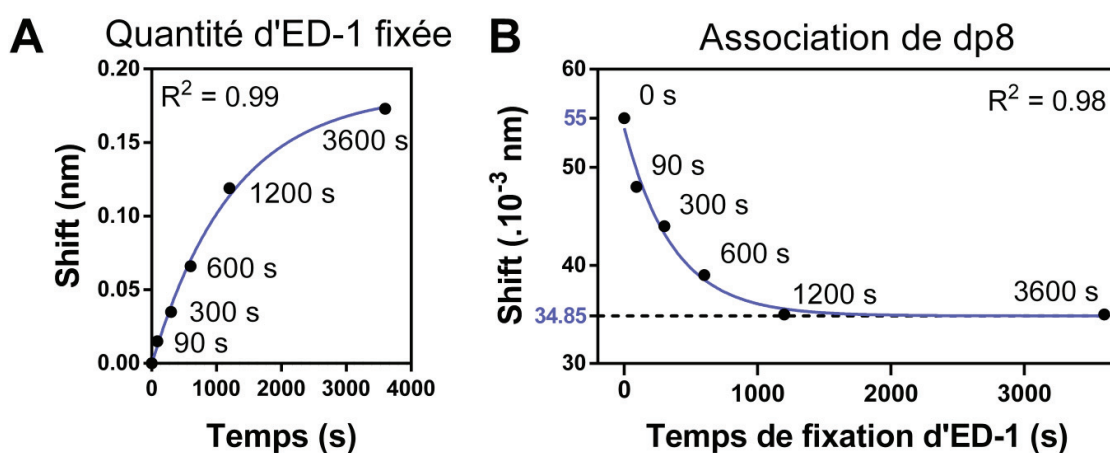
Il est probable que les chaînes d'héparane sulfate et de chondroïtine sulfate soient privilégiées dans l'interaction avec l'anastelline étant donné d'une part leur volume et leur charge, et d'autre part le fait qu'elles masquent probablement la partie protéique des syndécans.

Afin de vérifier cette éventualité, des expériences de BLI ont été réalisées à pH 7,4 en mettant un contact de façon séquentielle un capteur où a été immobilisée l'anastelline d'abord dans une solution d'ED-1 puis dans une solution de chaînes d'héparane sulfate. Une solution d'octasaccharide d'héparine (dp8) a également été utilisée dans le cas où les chaînes d'héparane sulfate provoquaient un encombrement stérique trop important (**Figure 144**). Une interaction est comme attendue détectée dans les expériences contrôles, lorsque le capteur où a été immobilisée l'anastelline est mis en contact avec du tampon (sans ED-1) puis dans les solutions d'oligosaccharides (HS, *courbe rouge* ou dp8, *courbe bleu foncé* sur la **Figure 144A**) ou lorsqu'il est uniquement mis en contact avec ED-1 (*courbe noire* sur la **Figure 144A**). L'intensité du signal est en revanche quasiment nulle lorsque le complexe anastelline – ED-1 est préformé puis trempé dans une solution de HS ou de dp8 (**Figure 144B**). Ces observations suggèrent que les oligosaccharides et ED-1 sont en compétition pour l'interaction avec l'anastelline.



**Figure 144 : Recherche du complexe ternaire anastelline – ectodomaine du syndécan-1 – oligosaccharide par interférométrie de biocouche. A)** Sensorgrammes de l'interaction entre l'anastelline, l'ectodomaine du syndécan-1 (ED-1, 5  $\mu$ M) et les chaînes d'héparane sulfate (HS) ou d'octasaccharide d'héparine (dp8, 5  $\mu$ M chacun). L'anastelline a été immobilisée puis mise en contact de façon séquentielle dans ED-1 seul (noir), dans ED-1 puis HS (orange), dans ED-1 puis dp8 (bleu clair), dans HS seul (rouge) ou dans dp8 seul (bleu foncé). Les aires colorées en bleu ou en vert réfèrent aux étapes de mise en contact avec ED-1 ou des oligosaccharides, respectivement. **B)** Sensorgrammes de la même expérience alignés à l'étape d'association des oligosaccharides sur les capteurs ayant été préalablement trempés (+ ED-1) ou non ( $\emptyset$  ED-1) dans une solution d'ED-1. La légende est indiquée et est commune aux deux figures.

Une série d'expériences similaires a été réalisée en mettant en contact un capteur sur lequel a été immobilisé l'anastelline avec une solution d'ED-1 pendant des durées variables, de 0 s à 1 h, puis dans une solution d'octasaccharide d'héparine (**Figure 145**). L'intensité du signal lorsque l'anastelline est mise en contact avec du tampon (sans ED-1) puis avec l'octasaccharide d'héparine est assez faible (0,055 nm) ce qui est probablement causé par sa petite taille. Lorsque l'anastelline est mise en contact avec ED-1 pendant 90 à 3600 s, l'intensité du signal lors de la phase d'association avec l'octasaccharide diminue graduellement jusqu'à atteindre un plateau d'intensité égale à environ 0,035 nm (**Figure 145B**). La fixation simultanée d'oligosaccharides et d'ED-1 sur l'anastelline semble ainsi être délicate étant donné leur probable fixation sur le même site à la surface de l'anastelline ou sur des sites très proches. Cette cofixation n'est cependant pas à exclure.

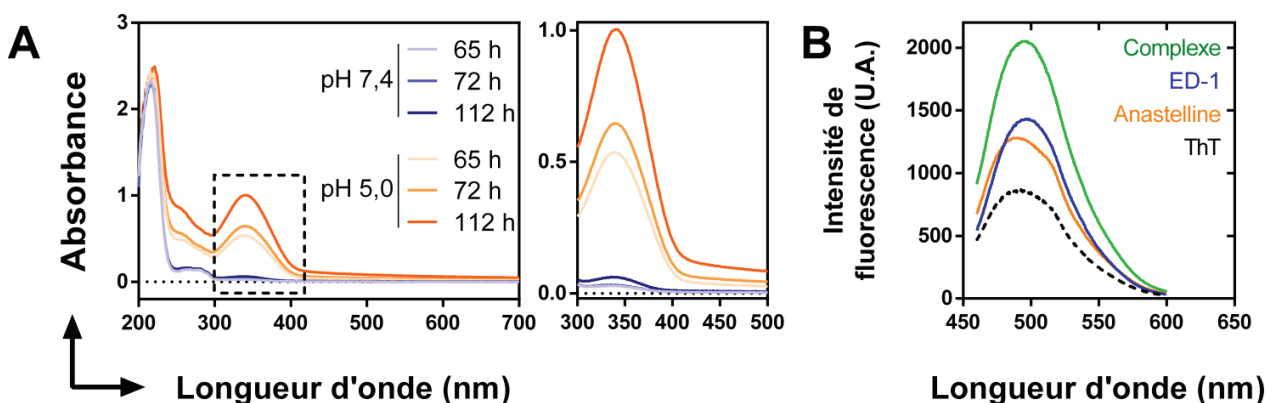


**Figure 145 : Variation de l'intensité du signal pendant l'étape d'association du dp8 d'héparine sur le complexe préformé anastelline – ED-1.** **A)** Quantité d'ED-1 fixée sur l'anastelline. Le capteur où a été immobilisée l'anastelline a été mis en contact pendant des durées variables (0 à 3600 s) avec ED-1 (5  $\mu$ M) avant d'être trempé en tampon HBS P20 0,05 % (v/v) pendant 480 s puis **(B)** dans une solution d'octasaccharide d'héparine (dp8, 5  $\mu$ M) pendant 600 s.

## 5. AGREGATION DU COMPLEXE ANASTELLINE – ECTODOMAINE DU SYNDECAN-1

L'interaction de l'anastelline avec la fibronectine et le fibrinogène mène à leur agrégation respectivement en superfibronectine et en superfibrinogène (*Yi & Ruoslahti, 2001, Ohashi et al., 2009, Stine et al., 2017*). Afin de vérifier si un phénomène similaire avait lieu lors de son interaction avec l'ectodomaine du syndécan-1, des expériences de turbidimétrie ont été réalisées.

L'anastelline et ED-1 ont été incubés à 37°C à pH 7,4 et à pH 5,0 et l'absorbance a été mesurée après 65, 72 et 112 heures d'incubation (**Figure 146A**). Une importante augmentation de l'absorbance a été observée dans la gamme de longueur d'onde de 300 à 400 nm lorsque le complexe est incubé à pH 5,0 et en moindre mesure à pH 7,4 (**Figure 146A**). La longueur d'onde maximale est de 330 nm qui est typique de la formation d'agrégats (*Malher et al., 2008*). Afin de vérifier si l'anastelline et ED-1 avait la propriété de former des fibrilles amyloïdes, le complexe précédemment incubé pendant 112 h à pH 5,0 a été analysé par fluorimétrie en présence de thioflavine T, une sonde spécifique des fibrilles amyloïdes (*Nilsson, 2004, Xue et al., 2016*). Une augmentation de l'intensité de fluorescence des protéines incubées seules ou en complexe a été observée dans la gamme 480 - 500 nm en comparaison de la thioflavine seule, ce qui indique une probable formation de fibrilles amyloïdes (**Figure 146B**). Cependant, cette augmentation d'intensité de fluorescence est assez faible ce qui peut suggérer que la majorité des agrégats formés sont amorphes dans ces conditions. Un tel signal peut refléter davantage un enrichissement en brins  $\beta$  plutôt que la formation de fibrilles amyloïdes (*Jozawa et al., 2016*).



**Figure 146 : Agrégation de l'ectodomaine du syndécan-1 et de l'anastelline. A)** Spectres UV-visible résultants de l'incubation d'ED-1 avec l'anastelline à pH 7,4 (bleu) et à pH 5,0 (orange). La zone en pointillés est agrandie. **B)** Spectres de fluorescence du complexe ED-1 – anastelline (vert), de l'anastelline seule (orange) ou d'ED-1 seul (bleu) incubé pendant 112 h à pH 5,0. La thioflavine T (ThT) seule est en pointillés noirs. U.A. : unité arbitraire.

### IV.3.3 Discussion et Perspectives

Les syndécans sont exprimés à la surface de tous les types cellulaires, à l'exception des érythrocytes, et sont impliqués dans de très nombreuses fonctions biologiques (*Manon-Jensen et al., 2010*). Leur interactome comprend au moins 347 partenaires mais il existe un biais d'interprétation dans l'analyse de l'interactome des syndécans étant donné le nombre limité de partenaires décrits pour le syndécan-3. Il s'agit également du seul syndécan dont aucun complexe ternaire n'a été décrit. Malgré leur large interactome, les modèles murins dont l'expression des syndécans a été abolie ne montrent pas de troubles du développement et sont viables et aucune mutation engendrant des pathologies chez l'Homme n'est à ce jour connue (*Couchman et al., 2015*). La conservation des fonctions associées aux syndécans malgré l'abolition de l'un de ses membres est certainement une conséquence du fait que la grande majorité de leurs partenaires interagissent avec leurs chaînes de glycosaminoglycanes, les autres membres de la famille des syndécans ou d'autres protéoglycanes pouvant alors prendre le relais.

Cependant, certaines fonctions sont propres à chaque membre de la famille des syndécans. Le syndécan-1 est majoritairement exprimé dans les cellules épithéliales et est ainsi impliqué dans le processus de cicatrisation (*Elenius et al., 2004*) alors que le syndécan-2 est majoritairement exprimé dans les cellules mésenchymateuses, notamment les fibroblastes, et est impliqué dans la réorganisation du cytosquelette d'actine (*Mytilinaïou et al., 2017*). Le syndécan-3 est en revanche majoritairement exprimé dans le système nerveux et régule le guidage axonal (*Kantor et al., 2004*). Le syndécan-4 est plus largement exprimé et régule notamment la formation des adhésions focales (*Afratis et al., 2017*). La spécificité fonctionnelle des syndécans est vraisemblablement dictée par les modifications post-synthétiques des chaînes de glycosaminoglycanes qu'ils portent sur leur ectodomaine selon le type cellulaire, engendrant des fonctions différentes (*Karamanos et al., 2018*). Le domaine cytosolique des syndécans est conservé ce qui leur permet d'interagir de façon consensus avec 14 protéines dont 12 intracellulaires. Il en résulte des fonctions communes dans l'organisation du cytosquelette d'actine, le transport vésiculaire et dans diverses voies de signalisation (*Gondelaud & Ricard-Blum, 2019*). Seules deux protéines extracellulaires sont présentes dans l'interactome consensus des quatre syndécans, le FGF-2 et la glycoprotéine virale gp120, qui interagissent *via* les chaînes de GAGs des quatre syndécans. D'après nos expériences, le FGF-2 interagit également avec la partie protéique de l'ectodomaine des syndécans -1 et -3 et en moindre mesure avec celui du syndécan-2 et du syndécan-4 ce qui signifie qu'il peut interagir à la fois avec les chaînes de GAGs et la partie protéique des SDCs.

Les constantes cinétiques d'association et de dissociation n'ont pas été déterminées car les complexes restent très stables suggérant que les syndécans pourraient avoir le rôle de réservoir de FGF-2 à la surface cellulaire. Un gradient d'affinité pourrait apparaître entre la partie protéique et les chaînes de GAGs portées par les syndécans.

A ce titre, le pistage d'une seule molécule de FGF-2 modifié par une nanoparticule d'or dans la matrice péricellulaire de fibroblastes en culture avait montré qu'il passait la majorité de son temps dans des régions restreintes mais pouvait facilement diffuser, notamment grâce aux chaînes d'héparane sulfate (*Duchesne et al., 2012*). La partie protéique des syndécans pourrait ainsi être en partie responsable de la capture et la rétention des molécules de FGF-2 à la surface cellulaire, du moins si la présence des chaînes de GAGs ne perturbe pas cette interaction.

Nos expériences d'interférométrie de biocouche ont également permis d'ajouter un nouveau partenaire consensus aux quatre syndécans, l'anastelline, qui est un fragment de la fibronectine. L'anastelline interagit avec les chaînes d'héparane sulfate des syndécans, comme cela avait été proposé plus tôt (*Mercurius & Morla, 2001*), mais aussi avec leurs chaînes de chondroïtine sulfate. Il est généralement admis que seuls les syndécans -1 et -3 portent des chaînes de CS et de façon juxtamembranaire (*Couchman, 2010*), cependant il semblerait que certaines chaînes d'HS peuvent être substituées par des chaînes de CS pour le syndécan-4 (*Deepa et al., 2004*), ce qui pourrait expliquer la valeur importante du signal observée dans nos expériences entre l'anastelline et l'ectodomaine du syndécan-4 portant ses chaînes de GAGs. Les expériences de docking d'oligosaccharides d'héparine et d'héparane sulfate ont permis de déterminer précisément leur site d'interaction à la surface de l'anastelline. Ces expériences ont également révélées une plus grande spécificité des tétrasaccharides d'héparane sulfate très sulfaté (HSNS) par rapport aux tétrasaccharides d'héparine ou d'héparane sulfate non sulfatés (HSNA).

Ces premières séries d'expérience ont ainsi permis d'identifier le (co)-récepteur de l'anastelline à la surface des cellules dont la formation du complexe favorise l'adhésion cellulaire (*Mercurius & Morla, 2001*). De façon similaire au FGF-2, l'anastelline interagit avec les chaînes de GAGs mais aussi avec la partie protéique des syndécans. Les complexes entre l'anastelline et les EDs sont très stables et les interactions électrostatiques et / ou hydrogènes semblent en être responsables.

Les expériences de BLI ont également montré que cette interaction était favorisée aux pHs acides et l'hypothèse la plus probable serait la formation d'une poche basique à la surface de l'anastelline du fait de la protonation des résidus d'histidine s'y trouvant.

La valeur de pKa théorique pour les résidus d'histidine 637 et 662 est de 6,17 et 6,45, respectivement (au lieu de 6,0 pour l'histidine libre), d'après le serveur *DelPhiPKa* (*Pahari et al., 2018*), ce qui indique qu'une faible diminution du pH serait à même de pouvoir protoner ces résidus et donc de former un nouveau site basique. La formation d'un nouveau site d'interaction est cohérente avec la forme générale des sensorgrammes obtenus dans des conditions acides ainsi que leur analyse avec le modèle du ligand hétérogène. Dans des conditions physiologiques où le pH est environ neutre, l'interaction avec l'anastelline promeut l'adhésion cellulaire. Favoriser la formation de ce complexe dans les microenvironnements tumoraux où le pH est plus acide (*Kato et al., 2013*) pourrait être une réponse cellulaire au phénotype invasif de certains types de cancers. Cependant, le shedding des syndécans y est également augmenté (*Ramani et al., 2013*) limitant ainsi les effets anti-invasifs de la formation de ce complexe.

Les syndécans solubles (*i.e.* ayant subi le processus de shedding) pourraient servir de vecteur de l'anastelline. A pH acide, l'interaction entre les deux partenaires est importante et le relargage de l'anastelline pourrait alors être effectué dans des environnements dont le pH est moins acide. Des interactions dépendantes du pH et régulées par des résidus d'histidine ont déjà été observées (*Watanabe et al., 2019*). C'est par exemple le cas de la protéine TSG-6 (*Tumor necrosis factor-inducible gene 6 protein*) dont l'interaction avec l'acide hyaluronique est favorisée à pH acide (pH 5,7) du fait de la protonation de deux résidus d'histidine, dont un à proximité du site d'interaction entre les deux partenaires (*Blundell et al., 2006*). Un exemple de transport et de relargage de protéine dépendant du pH est celui de l'isomérase des ponts disulfures ERp44, présente dans des compartiments en aval du réticulum endoplasmique, dont le compartiment intermédiaire (ERGIC, pour *Endoplasmic Reticulum-Golgi Intermediate Compartment*) et le cis-Golgi dans lequel règne un pH légèrement acide (pH 6,7). ERp44 y adopte une conformation ouverte dans laquelle 5 résidus d'histidine protonés favorisent son interaction avec différentes protéines. ERp44 est une sorte de point de contrôle, rapatriant certaines protéines immatures, comme des immunoglobulines M, vers le réticulum endoplasmique où le pH y est presque neutre (pH 7,2), ce qui entraîne la déprotonation des résidus d'histidine, menant au relargage des protéines avec lesquelles elle interagit (*Watanabe et al., 2019*).

La poche contenant des résidus d'histidine au sein de l'anastelline se retrouve à proximité d'un des peptides détectés par spectrométrie de masse. Ce-dernier contient également un résidu d'histidine qui pourrait avoir un rôle à jouer dans cette interaction. Un moyen de vérifier ces observations serait de modifier sélectivement les résidus d'histidine afin d'annuler leur charge à pH acide, comme avec du 4-hydroxynonanal.



Il serait également envisageable de générer des mutants de l'anastelline sur ces résidus. Dans le cas d'une implication de ces résidus d'histidine, les sensorgrammes obtenus à pH acide devraient être similaires à ceux enregistrés à pH 7,4 du fait de l'abolition de la création de ce nouveau site basique.

La formation d'un complexe ternaire entre l'anastelline, l'ectodomaine du syndécan-1 et les chaînes d'héparane sulfate ne semble pas être favorisé comme en attestent les expériences de BLI. La fonction biologique de l'interaction entre l'anastelline et les syndécans semble être connue, où la formation du complexe favorise l'adhésion cellulaire (*Mercurius & Morla, 2001*). En revanche, nos expériences montrent une interaction directe entre l'anastelline et la partie protéique d'ED-1 dont la fonction reste à déterminer. Cependant, les expériences préliminaires de turbidimétrie et de fluorimétrie sur le complexe anastelline – ED-1 ont montré la formation d'agrégats, notamment à pH 5,0 où le complexe pourrait former des fibrilles de type  $\beta$ -amyloïde. Ces observations pourraient être expliquées par la formation du néosite basique sur l'anastelline. L'anastelline pourrait interagir avec ED-1 *via* deux sites différents, favorisant la polymérisation du complexe. Les conditions optimales de la formation de ces fibrilles restent cependant à déterminer étant donné la faible augmentation d'intensité de fluorescence observée par rapport à la thioflavine T seule. Cependant, la formation des fibrilles amyloïdes est potentialisée à pH acide ce qui concorde avec nos expériences préliminaires (*Su & Chang, 2001*). La présence des glycosaminoglycanes sur ED-1 pourrait être un facteur limitant son agrégation (amyloïde ou amorphe), ou ils pourraient au contraire la favoriser, ce qui a déjà été observé. En effet, l'héparine a déjà été utilisée afin de promouvoir la formation de fibrilles *in vitro* (*Iannuzzi et al., 2015, Quittot et al., 2017*).

Finalement, il serait possible d'ajouter le cholestérol à l'interactome consensus des syndécans. En effet, le motif GXXXG présent dans leur domaine transmembranaire est également retrouvé dans celui de la protéine précurseur amyloïde et participe à son interaction avec le cholestérol (*Barett et al., 2012*). Certaines séquences consensus d'interaction avec le cholestérol ont été décrites (*Fantini & Barrantes, 2013*) et sont retrouvées au sein du domaine transmembranaire des syndécans. Ces observations vont de pair avec leur localisation préférentielle dans les radeaux lipidiques (*Woods & Couchman, 2001, Tkachenko & Simons, 2002, Couchman, 2003*) qui sont des microdomaines enrichis en cholestérol et en sphingolipides (*Kusumi et al., 2020*).



## **IV.4**

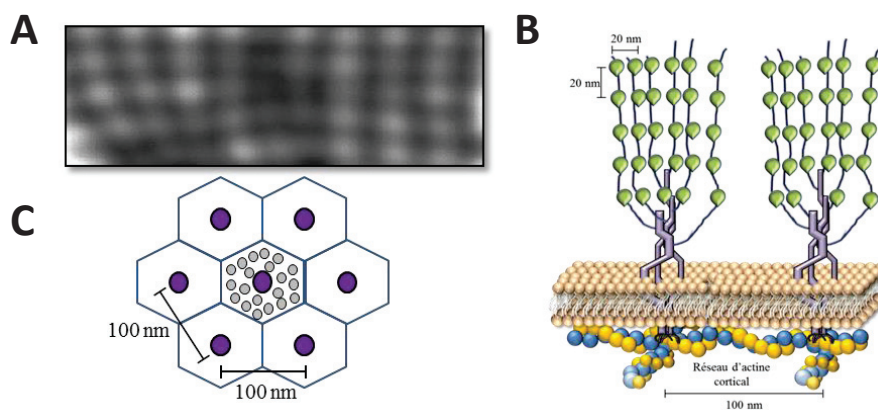
# **RÔLE DES GLYCOSAMINOGLYCANES ET DES ACIDES SIALIQUES PRÉSENTS A LA SURFACE DES GLOBULES ROUGES SAINS ET DRÉPANOCYTAIRES DANS LEURS PROPRIÉTÉS RHÉOLOGIQUES**

#### IV.4.1 Introduction

Le glycocalyx est une fine couche de glycoconjugués et de polysaccharides recouvrant la surface des cellules eucaryotes. Le glycocalyx des cellules endothéliales est de loin le plus étudié. Il est situé au niveau de leur pôle apical et sa structure est apparentée à un gel contribuant à environ 2 % du volume plasmatique (*Kundra & Goswami, 2019*). Son épaisseur est directement corrélée à la taille du vaisseau sanguin et est estimée à environ 1  $\mu\text{m}$  dans les veinules, 2 à 3  $\mu\text{m}$  dans les petites artères et de 4,5  $\mu\text{m}$  à 11  $\mu\text{m}$  dans la carotide ou l'aorte (*Reitsma et al., 2007, Ebong et al., 2011, Yen et al., 2012*). Il est cependant important de noter que l'estimation de son épaisseur dépend de la technique utilisée pour sa visualisation (*Ebong et al., 2011*). Il est majoritairement composé de protéoglycanes qui portent de façon covalente sur leur partie protéique une ou plusieurs chaînes de glycosaminoglycanes. Six GAGs ont été décrits chez les mammifères et diffèrent par leur unité disaccharidique, leurs modifications post-synthétique ou encore leur taille (*Karamanos et al., 2018*). Au sein du glycocalyx, les chaînes d'héparane sulfate et de chondroïtine sulfate sont les plus abondantes (*Gao & Lipowsky, 2010*). En effet, la taille du glycocalyx présent à la surface des veinules est réduite de 43 % après digestion des chaînes d'HS et de 34 % après digestion des chaînes de CS. En revanche, les chaînes de dermatane sulfate ou de kératane sulfate sont beaucoup moins représentées. L'acide hyaluronique peut être piégé dans le glycocalyx notamment par le biais de ses récepteurs que sont HARE (*Hyaluronan receptor for endocytosis*), LYVE-1 (*Lymphatic vessel endothelial receptor 1*), RHAMM (*Receptor for hyaluronan mediated motility*), la layiline ou bien le plus caractérisé d'entre eux, CD44 (*Tarbell et al., 2014*). L'épaisseur du glycocalyx des veinules est diminuée de 12 % après digestion de l'acide hyaluronique.

Au total, environ 90 % de l'épaisseur du glycocalyx est due à la présence des chaînes de GAGs (*Tarbell et al., 2014*). Comme décrit précédemment, les chaînes d'héparane sulfate présentent une distribution des groupements sulfates hétérogène, où les régions hypersulfatées sont des sites privilégiés d'interaction avec des protéines sériques comme l'albumine, diverses enzymes, des cytokines ou encore des facteurs de croissances (*Kundra & Goswami, 2019*). On retrouve également des glycoprotéines membranaires, en particulier les intégrines, notamment  $\alpha\text{v}\beta\text{3}$ , des sélectines E et P ainsi que des membres de la superfamille des immunoglobulines (ICAM-1 et -2, PECAM-1) (*Reitsma et al., 2007*).

Les composants du glycocalyx forment un réseau tridimensionnel dynamique et structuré (**Figure 147**). L'analyse par microscopie électronique du glycocalyx des cellules endothéliales de mésentère de grenouille a montré que ses composants s'assemblent en fibres d'environ 10 à 12 nm de diamètre où se forment périodiquement des *clusters* séparés de 20 nm. Cette structure forme une maille quasi-hexagonale espacée de 100 nm au niveau de la membrane et est ancrée au cytosquelette d'actine sous-jacent (*Squire et al., 2002*). Plus récemment, le glycocalyx de cellules endothéliales provenant du cerveau de souris a été étudié par microscopie optique à super-résolution (*Fan et al., 2019*). En accord avec les précédentes observations, les chaînes d'HS sont presque perpendiculaires par rapport à la surface de la cellule (adoptant un angle de 78 à 85° par rapport à la surface de la cellule) alors que l'acide hyaluronique reste en surface et forme un réseau hexagonal.

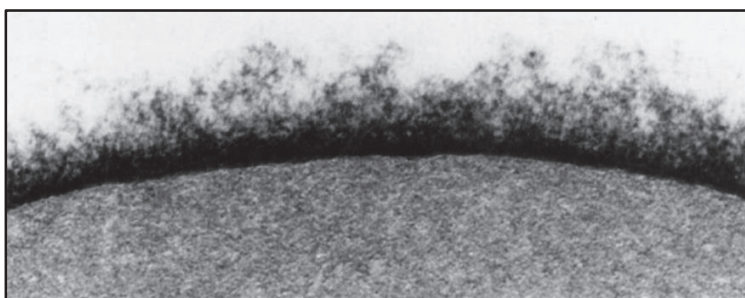


**Figure 147 : Ultrastructure du glycocalyx endothélial.** **A)** Micrographie du glycocalyx endothélial de capillaire de mésentère de grenouille (Adapté de *Squire et al., 2002*). La barre d'échelle n'a pas été indiquée. **B)** Modèle créé à partir des micrographies d'une vue latérale du glycocalyx. Violet : composants du glycocalyx, vert : regroupement périodique de molécules sur le glycocalyx. **C)** Modèle d'une vue lumineuse du glycocalyx, assemblé en maille hexagonale. Violet : regroupement de protéines sur la membrane, gris : composants du glycocalyx. Les schémas ne sont pas à l'échelle.

Une fonction particulièrement importante du glycocalyx endothélial réside dans la mécanoréception des contraintes de cisaillement provoquées par l'écoulement du sang, et leur transmission au cytosquelette. Cette fonction est principalement achevée par les chaînes d'HS portées par le glypican-1 et par l'acide hyaluronique, mais pas par les chaînes de CS (*Thi et al., 2004, Zeng et al., 2018*). La présence de cette structure dense et très électronégative à la surface des cellules endothéliales permet de limiter les interactions avec d'autres types cellulaires ou fragments circulants comme les érythrocytes, les leucocytes ou les plaquettes.

Il joue un rôle de filtre, restant perméable aux molécules de masse moléculaire inférieure à environ 70 kDa (*Alphonsus & Rodseth, 2014, Yilmaz et al., 2019*). Plusieurs conditions physiopathologiques mènent à la dégradation du glycocalyx. L'hyperglycémie provoque la réduction de près de 50 % de son épaisseur (*Nieuwdorp et al., 2006*). Il est également dégradé dans des conditions d'inflammation, de sepsis ou encore chez les patients présentant des troubles cardiovasculaires (*Tarbell & Cancer, 2016, Yilmaz et al., 2019*). Aussi, le *shedding* des différents constituants du glycocalyx entraîne l'exposition de récepteurs d'adhérence vasculaires aux leucocytes, leur permettant d'interagir avec l'endothélium (*Lipowsky, 2012*). A l'inverse, l'exposition de cellules endothéliales à des forces de cisaillement importantes augmente l'épaisseur du glycocalyx (*Gouverneur et al., 2006, Zeng & Tarbell, 2014, Liu et al., 2016*).

Le glycocalyx des globules rouges de sujets sains est en revanche très fin, sa taille variant de 1,7 à 10 nm, soit 1000 fois plus fin que celui recouvrant l'endothélium vasculaire (*Kahane et al., 1978, Donath & Pastushenko, 1979, Linss et al., 1991*) (**Figure 148**). La purification des glycanes d'érythrocytes humains et leur immunomarquage avec un anticorps monoclonal anti-HS a mis en évidence la présence d'une faible quantité de chaînes d'héparane sulfate à la surface des globules rouges (environ 2 000 par cellules), valeur très faible en comparaison de la quantité d'autres molécules comme la protéine Bande 3 ou la Glycophorine A (GPA, environ  $10^6$  molécules par cellule). Ces chaînes d'héparane sulfate sont en outre des sites de fixation au pathogène *Plasmodium falciparum* sur les érythrocytes (*Vogt et al., 2004*). L'extraction des GAGs de globules rouges de rat a permis de mettre en évidence la présence de chaînes de chondroïtine et de dermatane sulfate, majoritairement sous forme 4-O sulfatées dans le glycocalyx des globules rouges. Ils sont présents à une concentration de 0,5 µg pour  $10^{10}$  érythrocytes soit environ  $10^{11}$  copies par cellules (*Srikanth et al., 2012*). Ce type de sulfatation est d'ailleurs diminué dans le cas de régime hypercholestérolémiant (*Kiran et al., 2015*). Enfin, la présence du récepteur CD44 à la surface des globules rouges permet de suggérer la présence d'acide hyaluronique à leur surface et ainsi l'établissement d'interactions érythrocyte-érythrocyte ou érythrocyte-endothélium (*Kerfoot et al., 2008*). La particularité du glycocalyx des globules rouges réside dans leur caractère extrêmement anionique, du fait de la présence d'un nombre important d'acides sialiques présents à leur surface, dont l'écrasante majorité est représentée par l'acide N-acétylneuraminique (Neu5Ac – *Bulai et al., 2003*). 75 % des acides sialiques du globule rouge sont d'ailleurs portés par chacune des 16 O-glycosylations et de l'unique N-glycosylation de la glycophorine A (*Aoki, 2017*), protéine membranaire très abondante dont le nombre varie de  $96 \cdot 10^3$  et  $1 \cdot 10^6$  par cellules (*Burton & Bruce, 2011, Bryk & Wisniewski, 2017*).



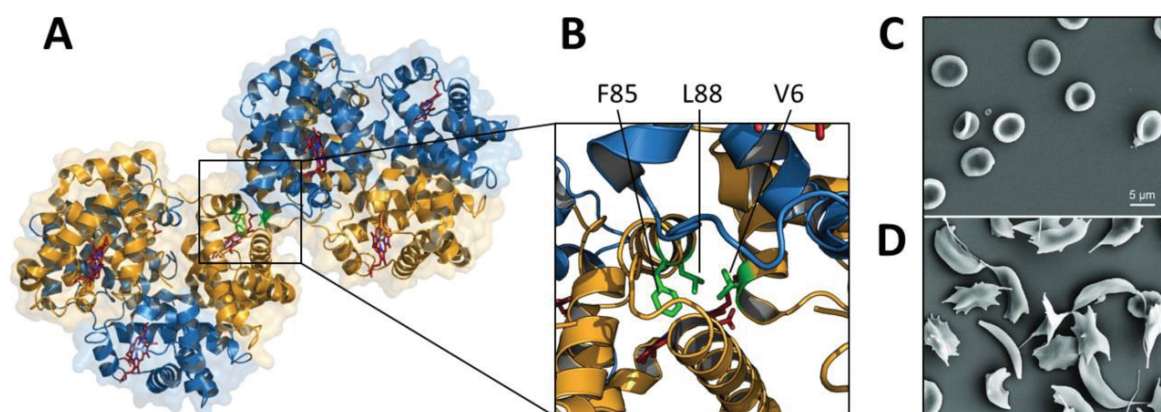
**Figure 148 : Visualisation du glycocalyx du globule rouge par microscopie électronique.**

L'épaisseur du glycocalyx est estimée à 1,4 nm et les filaments d'oligosaccharides ont un diamètre évalué entre 12 et 25 Å. Cette visualisation est la seule disponible à ce jour. La barre d'échelle n'a pas été indiquée. D'après *Roseman*, (2001).

Environ 270 millions de molécules d'hémoglobines sont présentes dans chaque globule rouge (*D'Alessandro et al., 2017*). L'hémoglobine (HbA) est une protéine hétérotétramérique composée de deux chaînes d' $\alpha$ -globine et deux chaînes de  $\beta$ -globine portant chacune un groupement héminique permettant la fixation du dioxygène. Plusieurs hémoglobinopathies ont été décrites, notamment les thalassémies et la drépanocytose. Cette dernière, appelée anémie falciforme, est une pathologie causée par la mutation d'un seul nucléotide dans le gène codant pour la  $\beta$ -globine, substituant le résidu de glutamate 6 par un résidu de valine (Glu6Val) provoquant l'apparition de l'hémoglobine S (HbS) (**Figure 149**).

Dans des conditions de faible oxygénation, l'exposition de ce résidu de valine et son interaction avec les résidus de phénylalanine 85 et de leucine 88 présents dans la chaîne- $\beta$  d'une molécule d'hémoglobine voisine mène à la polymérisation de l'HbS dans le cytoplasme des globules rouges. Il en résulte la formation de longues fibres d'HbS qui déforment le globule rouge en forme de faucille, d'où le nom de la pathologie (*Connes et al., 2016*) (**Figure 149**). Cette modification structurale fragilise le globule rouge et entraîne une perte de sa déformabilité, propriété fondamentale de ce type cellulaire pour la réalisation de ses fonctions. Ces modifications provoquent l'apparition fréquente d'évènements vaso-occlusifs et d'anémie hémolytique sévère chez le patient drépanocytaire (*Ballas et al., 2010, Connes et al., 2016*).

Actuellement, seuls des traitements périodiques par l'hydroxyurée, des transfusions sanguines et dans les cas les plus graves la greffe de moelle osseuse, permettent l'amélioration des conditions de vie des patients. Chez les patients drépanocytaires, l'épaisseur du glycocalyx endothélial est d'ailleurs réduit (*van Beers et al., 2008*).



**Figure 149 : Structure de l'hémoglobine S et la conséquence de sa polymérisation sur le phénotype du globule rouge.** **A)** Structure cristallographique de deux molécules d'hémoglobine S en interaction (PDB : 2HBS). Les chaînes  $\alpha$  et  $\beta$  sont en bleu et orange, respectivement. Rouge : groupements hémiques. **B)** Agrandissement de la zone d'interaction entre deux molécules d'HbS où l'on distingue le résidu de valine 6 d'une première molécule en interaction avec le *patch* hydrophobe formé par les résidus de phénylalanine 85 et de leucine 88 portés par l'une des deux chaînes  $\beta$  de la molécule d'HbS voisine. Les résidus de valine 6, phénylalanine 85 et leucine 88 sont en vert. Visualisation par microscopie électronique à transmission de globules rouges drépanocytaires oxygénés **(C)** et désoxygénés **(D)**. L'échelle est commune aux deux images. Les micrographies sont issues d'*Abay et al.*, (2019).

#### IV.4.2 Buts de l'étude

Le but de notre étude a été de déterminer le ou les rôles des chaînes de glycosaminoglycanes sur les propriétés rhéologiques des globules rouges. La capacité que les globules rouges ont à se déformer (*i.e.* leur déformabilité), leur propension à s'agréger en plasma (*i.e.* leur agrégation) ou en dextran (*i.e.* leur agrégabilité) ont été mesurées sur des globules rouges traités par des enzymes dépolymérisant les chaînes de GAGs et libérant les acides sialiques, et comparées à celles des globules rouges non traités. L'agrégabilité représente la propriété intrinsèque des globules rouges à s'agréger, alors que leur agrégation est liée à la présence de facteurs externes, notamment des facteurs plasmatiques ou cellulaires.

Ces deux paramètres ont été mesurés en stase (fonction M) ou à faible vitesse de cisaillement (M1) comme décrit dans la section *Matériel & Méthodes*. Ces expériences ont été réalisées sur des globules rouges de sujets sains ainsi que sur des globules rouges de patients drépanocytaires afin de comparer les rôles respectifs des GAGs et des acides sialiques sur leurs propriétés rhéologiques.

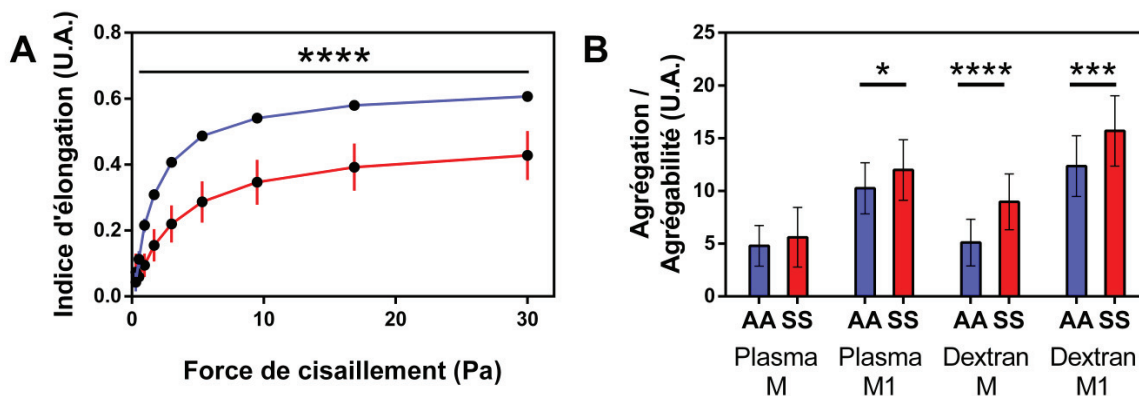


Ce travail a été effectué en collaboration avec le Professeur Philippe Connes et son équipe (Laboratoire Interuniversitaire de Biologie de la Motricité – LIBM, EA7424, Université Lyon 1).

#### IV.4.3 Résultats

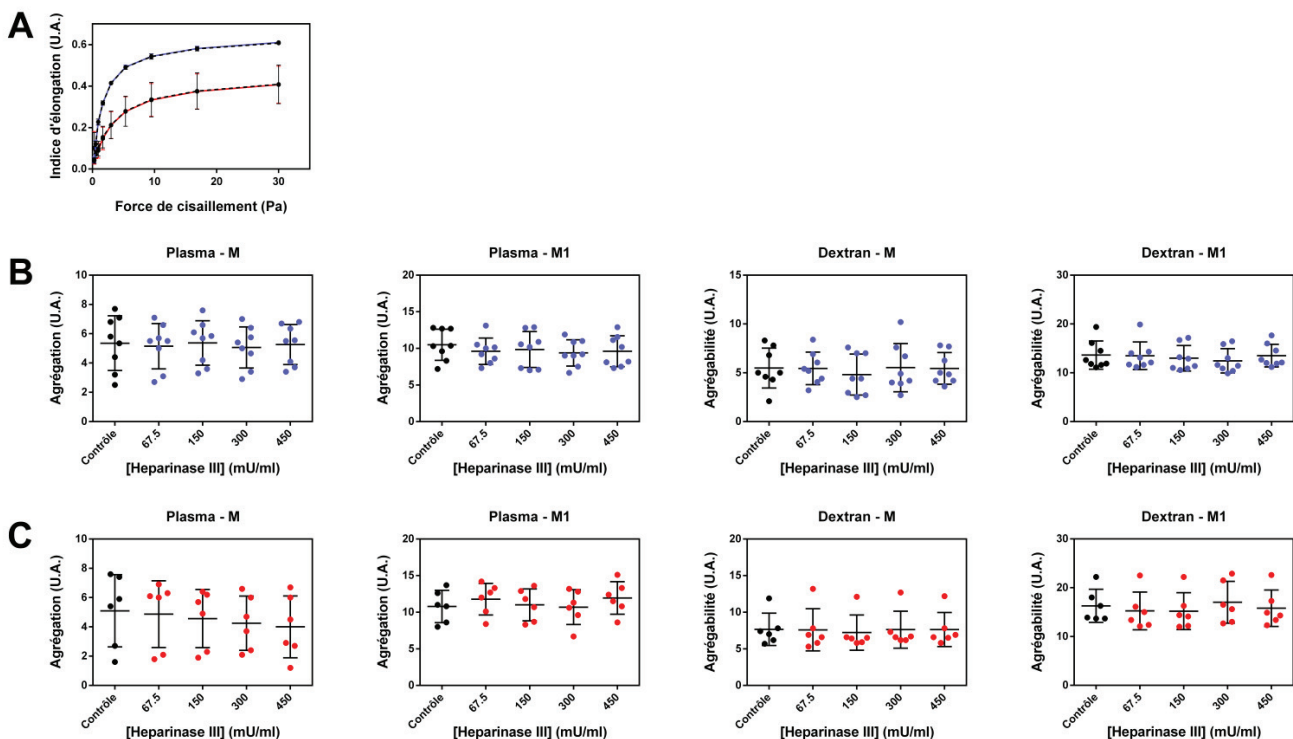
Les propriétés rhéologiques des globules rouges ont été mesurées avant leur incubation avec des enzymes dégradant leurs glycosaminoglycanes et leurs acides sialiques afin de mettre en évidence des différences dans leur déformabilité ainsi que dans leur agrégation et leur agrégabilité. Comme observé précédemment (*Connes et al., 2016*), la déformabilité des globules rouges sains est significativement plus élevée que celles des drépanocytaires (**Figure 150A**). Leur agrégation (mesurée en plasma) et leur agrégabilité (mesurée en dextran) sont inférieures à celles des globules rouges drépanocytaires, excepté lorsque la mesure d'agrégation est réalisée en stase (**Figure 150B**).

De plus, les différences d'agrégabilité (+ 43,1 % et + 21,3 % pour les fonctions M et M1 en dextran) entre les globules rouges drépanocytaires et sains sont plus importantes que les différences de leur agrégation (+ 14,6 % et + 14,5 % pour les fonctions M et M1 en plasma autologue). Cette observation permet de suggérer que des facteurs intrinsèques aux globules rouges drépanocytaires, plutôt que des facteurs externes, sont principalement responsables de leur capacité à interagir avec d'autres globules rouges.



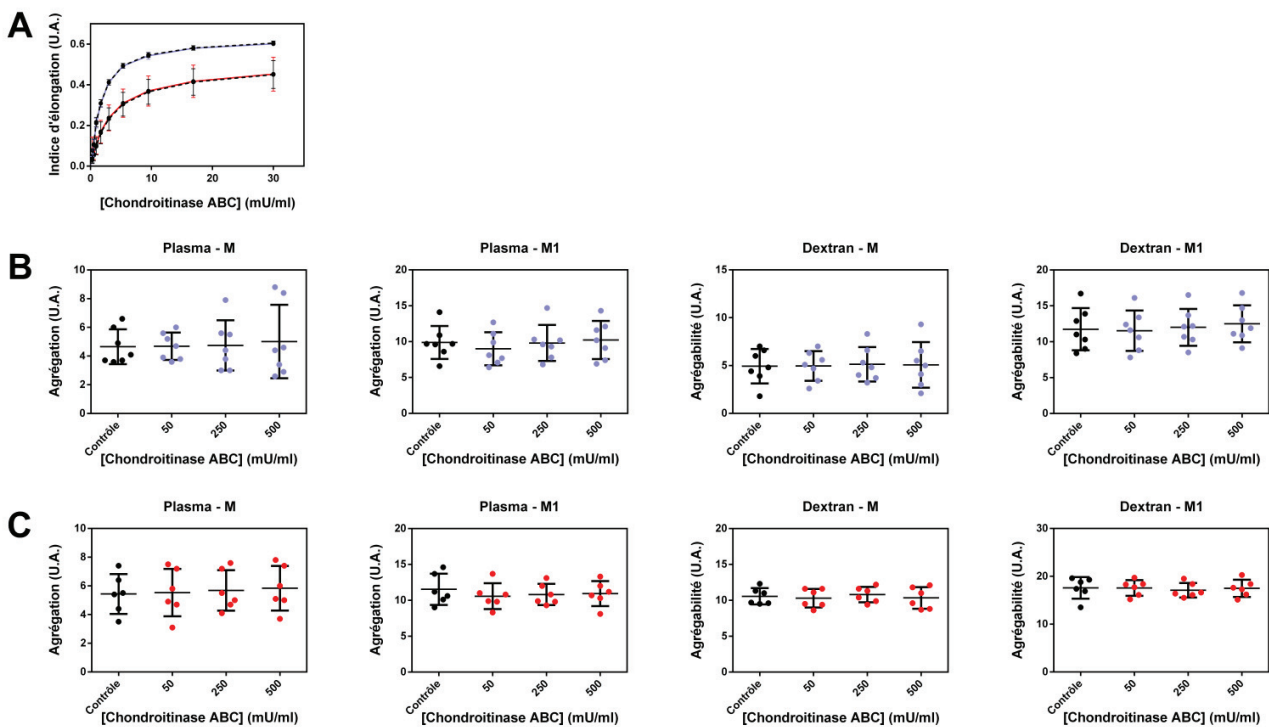
**Figure 150 : Propriétés rhéologiques des globules rouges sains et drépanocytaires non traités. A)** Mesure de la déformabilité des globules rouges sains (bleu, n = 24) et drépanocytaires (rouge, n = 24). **B)** Agrégation (en plasma) et Agrégabilité (en dextran) des globules rouges sains (bleu, n = 23) et drépanocytaires (rouge, n = 23), mesurées en stase (fonction M) ou à faible vitesse de cisaillement (fonction M1).

L'incubation des globules rouges sains et drépanocytaires pendant 60 minutes avec de l'héparinase III, qui dépolymérise spécifiquement les chaînes d'HS en clivant la liaison 1-4 entre un résidu de glucosamine N-sulfaté ou N-acétylé et l'acide glucuronique, n'a pas provoqué de variations significatives de la déformabilité ou de l'agrégation / agrégabilité des globules rouges (**Figure 151A**). Pour des raisons de clarté, seules les valeurs obtenues en absence et en présence de la concentration en enzyme la plus élevée (450 mU/ml) sont présentées. Les valeurs obtenues en présence des autres concentrations en enzymes (67,5, 150 et 300 mU/ml) sont présentées en **Annexe 6**. Aucune modification de l'agrégabilité ou de l'agrégation des globules rouges sains ou drépanocytaires n'a été observée (**Figures 151B et 151C**).



**Figure 151 : Propriétés rhéologiques des globules rouges sains et drépanocytaires traités par l'héparinase III. A)** Déformabilité des globules rouges sains (bleu, n = 8) et drépanocytaires (rouge, n = 8). Les courbes noires en pointillé représentent les valeurs obtenues avant traitement par l'héparinase III pour chacun des deux groupes. Agrégation (en plasma) et agrégabilité (en dextran) des globules rouges sains (**B**, n = 8) et drépanocytaires (**C**, n = 6).

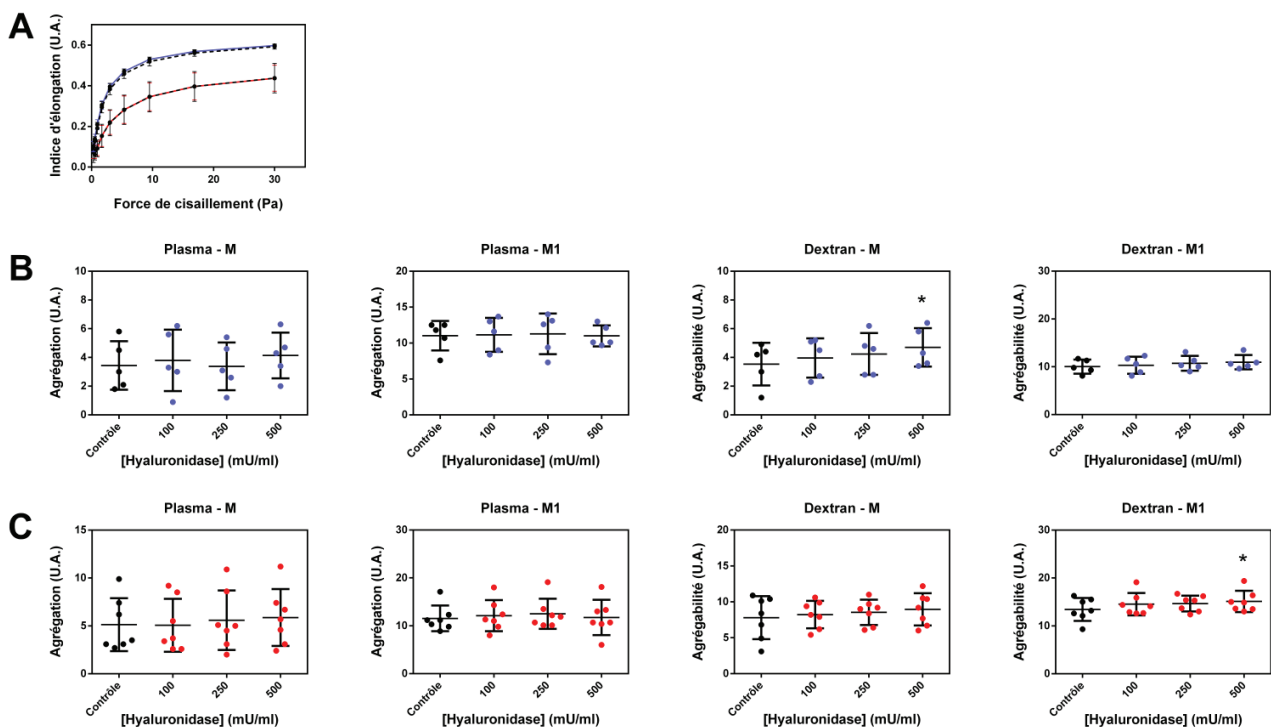
Le traitement des globules rouges pendant 60 minutes en présence de chondroïtinase ABC, qui dépolymérise les chaînes de chondroïtine sulfate de type A et C et les chaînes de dermatane sulfate (également appelée chondroïtine sulfate B), n'a pas induit de variations significatives des paramètres rhéologiques des globules rouges sains et drépanocytaires (**Figure 152**). Comme pour le traitement à l'héparinase III, seules les valeurs obtenues en absence et en présence de la concentration en enzyme la plus élevée (500 mU/ml) sont présentées. Les valeurs obtenues en présence des autres concentrations en enzymes (50 et 250 mU/ml) sont présentées en **Annexe 7**.



**Figure 152 : Propriétés rhéologiques des globules rouges sains et drépanocytaires traités par la chondroïtinase ABC.** **A)** Déformabilité des globules rouges sains (bleu, n = 8) et drépanocytaires (rouge, n = 6). Les courbes noires en pointillés représentent les valeurs obtenues avant traitement par la chondroïtinase ABC pour chacun des deux groupes. Agrégation (en plasma) et agrégabilité (en dextran) des globules rouges sains (**B**, n = 7) et drépanocytaires (**C**, n = 6).

La digestion de l'acide hyaluronique pendant 60 minutes n'a pas significativement affecté la déformabilité des globules rouges sains, bien que les *p-values* soient proches du seuil de significativité dans la faible gamme de force de cisaillement ( $0,05 < p < 1$  pour des forces de cisaillement de 0,95 à 9,49 Pa). La déformabilité des globules rouges drépanocytaires n'a, en revanche, pas du tout été affectée (**Figure 153A**). Seules les valeurs obtenues en absence et en présence de la concentration en enzyme la plus élevée (500 mU/ml) sont présentées. Les valeurs obtenues en présence des autres concentrations en hyaluronidase (100 et 250 mU/ml) sont présentées en **Annexe 8**.

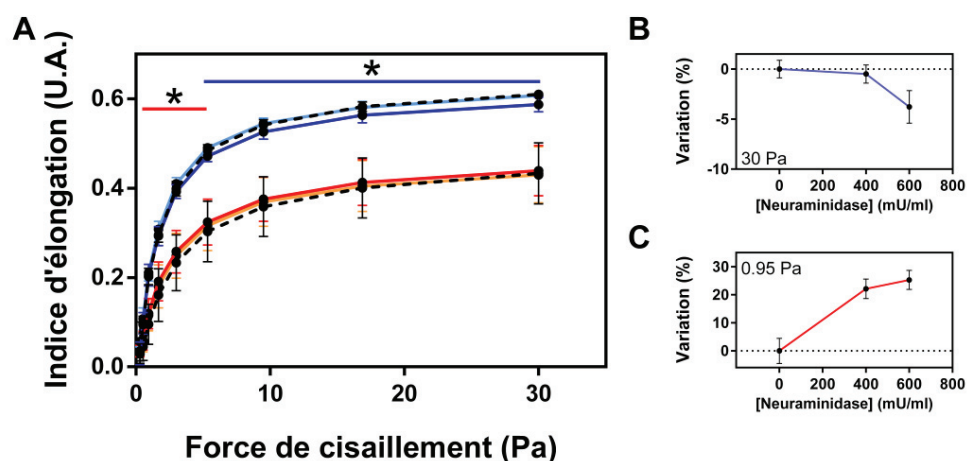
Cependant, une augmentation significative de l'agrégabilité (mesurée en dextran) est détectée en stase pour les globules rouges sains (fonction M, + 22 %) et à faible vitesse de cisaillement (fonction M1, + 12,5 %) pour les globules rouges drépanocytaires (**Figures 153B et 153C**). En revanche, la digestion de l'acide hyaluronique n'a aucun effet sur l'agrégation mesurée en plasma des deux types de globules rouges.



**Figure 153 : Propriétés rhéologiques des globules rouges sains et drépanocytaires traités par la hyaluronidase.** **A)** Déformabilité des globules rouges sains (bleu,  $n = 6$ ) et drépanocytaires (rouge,  $n = 7$ ). Les courbes noires en pointillés représentent les valeurs obtenues avant traitement par la hyaluronidase pour chacun des deux groupes. Agrégation (en plasma) et agrégabilité (en dextran) des globules rouges sains (**B**,  $n = 5$ ) et drépanocytaires (**C**,  $n = 7$ ).

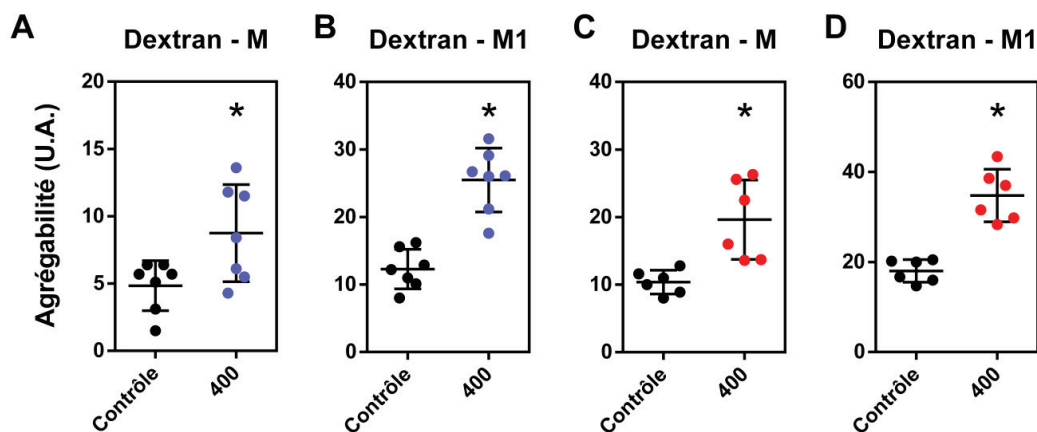
Une diminution faible mais significative (- 3,8 %) de la déformabilité des globules rouges sains est observée dans une gamme de force de cisaillement allant de 5,33 à 30 Pa après leur traitement pendant 60 minutes par la neuraminidase à une concentration de 600 mU/ml (**Figure 154A**).

Au contraire, la digestion des acides sialiques n'a pas affecté la déformabilité des globules rouges drépanocytaires dans cette gamme, mais à des forces de cisaillement plus faibles (de 0,95 à 5,33 Pa) avec la même concentration en enzyme (+ 25,3 % à 0,95 Pa). De plus, l'incubation des globules rouges drépanocytaires à une plus faible concentration enzymatique (400 mU/ml) a également induit une augmentation de la déformabilité dans la même gamme de force de cisaillement. Les globules rouges sains n'ont pas été affectés avec cette même concentration (**Figures 154B et 154C**).



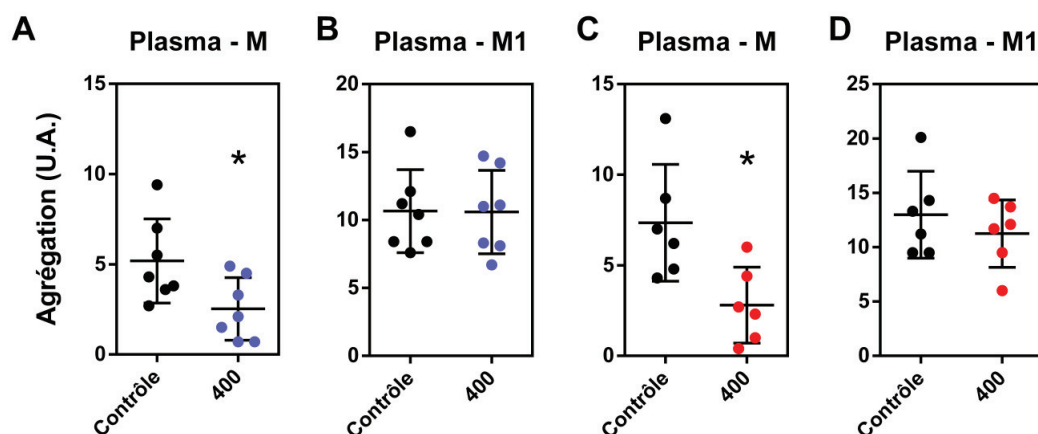
**Figure 154 : Déformabilité des globules rouges sains et drépanocytaires traités par la neuraminidase. A)** Mesures de déformabilité des globules rouges sains (bleu, n = 7) et drépanocytaires (rouge, n = 6) traités par la neuraminidase et non traités (courbes noires en pointillés pour les deux groupes). Concentration en neuraminidase de 400 (couleur claire) et 600 mU/ml (couleur foncée). Variation de la déformabilité des globules rouges sains à 30 Pa (**B**) et des globules rouges drépanocytaires à 0,95 Pa (**C**).

L'agrégabilité des globules rouges sains et drépanocytaires augmente significativement en stase (fonction M, + 81 % et + 89 %, respectivement) et à faible force de cisaillement (+ 107 % et + 93 %, respectivement) lorsqu'ils sont resuspendus en dextran après leur incubation pendant 60 minutes avec la neuraminidase à une concentration de 400 mU/ml (**Figure 155**). Les mêmes variations ont été observées avec une concentration de 600 mU/ml mais une lyse partielle des globules rouges a été observée à cette concentration.



**Figure 155 : Effets de la digestion des acides sialiques par la neuraminidase sur l'agrégabilité (en dextran) des globules rouges sains et drépanocytaires.** Mesures d'agrégabilité des globules rouges sains (A et B, n = 7) et drépanocytaires (C et D, n = 6) en stase (fonction M) et faible vitesse de cisaillement (fonction M1) après leur traitement par la neuraminidase (400 mU/ml) pendant 60 minutes.

Lorsque les globules rouges, sains ou drépanocytaires, sont resuspendus dans leur plasma autologue, leur agrégation diminue significativement en stase (**Figures 156A et 156C**) mais pas à faible vitesse de cisaillement (**Figures 156B et 156D**). L'ampleur de la diminution de leur agrégation est similaire pour les globules rouges sains (- 51 %) ou drépanocytaires (- 62 %). Etant donné que la digestion des acides sialiques augmente l'agrégabilité mais diminue l'agrégation des globules rouges, il est possible qu'un facteur plasmatique susceptible d'interagir avec les acides sialiques soit impliqué.



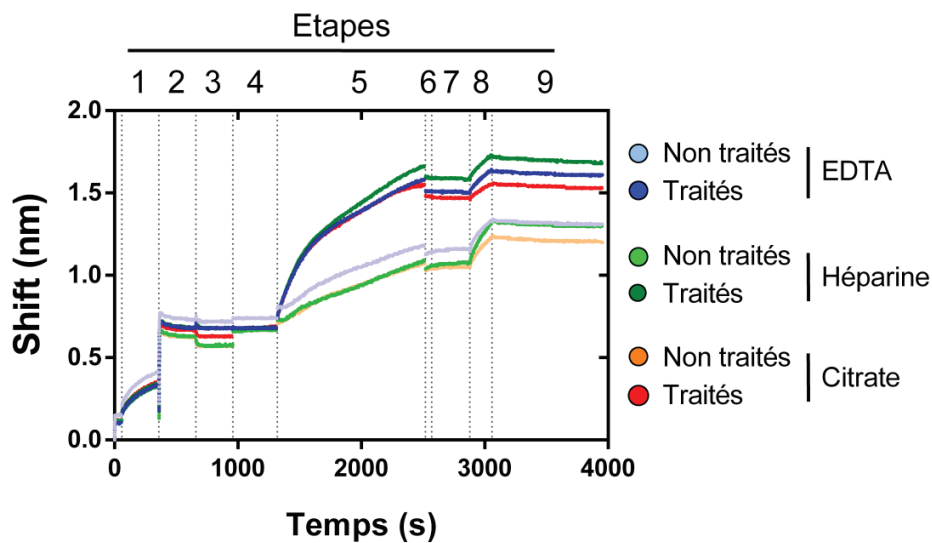
**Figure 156 : Effets de la digestion des acides sialiques par la neuraminidase sur l'agrégation (en plasma) des globules rouges sains et drépanocytaires.** Mesures d'agrégabilité des globules rouges sains (A et B, n = 7) et drépanocytaires (C et D, n = 6) en stase (fonction M) et faible vitesse de cisaillement (fonction M1) après leur traitement par la neuraminidase (400 mU/ml) pendant 60 minutes.

De précédentes études ont montré que le fibrinogène est présent à la surface des globules rouges (*Carvalho et al. 2011*) et qu'il induit leur agrégation (*Simmonds et al., 2013, Guedes et al., 2019*). L'analyse par microscopie à force atomique de globules rouges capturés sur poly-L-lysine et traités à la neuraminidase a mis en évidence une diminution du nombre d'interactions détectées avec le fibrinogène par rapport à des globules rouges non traités (*Carvalho et al. 2011*). Il a été montré que le fibrinogène interagit moins avec les globules rouges âgés, qui présentent une quantité d'acide sialique à leur surface moins importante que celle des globules rouges nouvellement formés (*Bratosin et al., 1995, Jakubowska-Solarska & Solski, 2000*). Ces observations suggèrent que le fibrinogène pourrait interagir avec les globules rouges *via* les acides sialiques exposés à leur surface.

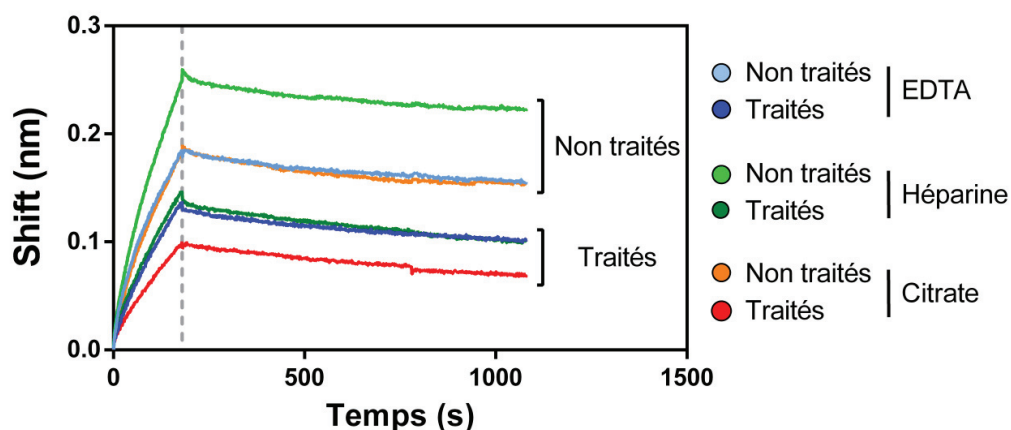
Nous avons donc effectués des expériences d'interférométrie de biocouche afin de déterminer si le fibrinogène interagissait avec les acides sialiques présents à la surface des globules rouges. Etant donné que le fibrinogène interagit avec l'héparine (*Fredenburgh et al., 2013*), dont la structure est similaire à celles des chaînes d'héparane sulfate et avec le zinc et le calcium (*Marguerie et al., 1979, Marx, 1988*), les globules rouges ont été collectés en tubes EDTA, héparine et citrate, afin d'observer d'éventuelles variations d'interactions lors des expériences. Les globules rouges sains, traités ou non à la neuraminidase, ont été capturés sur un capteur où de la poly-D-lysine avait été préalablement immobilisée de façon covalente. Le fibrinogène a ensuite été mis en contact avec le capteur (**Figure 157**).

Les globules rouges ont été efficacement capturés à la surface du capteur modifié par la poly-D-lysine. Un *shift* plus important est cependant détecté lors de l'immobilisation des globules rouges traités à la neuraminidase par rapport aux non traités (**Figure 157** et **Tableau XXXII**). Pour plus de lisibilité, les courbes ont été alignées lors du début de la phase d'association (**Figure 158**).

Quelle que soit la méthode de collecte des globules rouges effectuée, les globules rouges non traités à la neuraminidase interagissent en plus grande quantité avec le fibrinogène que les globules rouges traités. Cependant, la méthode de collecte semble influencer l'interaction où des *shifts* plus importants ont été mesurés lorsque le sang a été collecté en tube héparine



**Figure 157 : Sensorgrammes obtenus lors des expériences d'interférométrie de biocouche.** Des globules rouges ont été collectés en tube EDTA (bleu), héparine (vert) ou citrate (orange) et ont été traités ou non à la neuraminidase. Ils ont ensuite été capturés sur un capteur sur lequel a été préalablement immobilisée de la poly-D-lysine, puis mis en contact avec une solution de fibrinogène (100  $\mu\text{g/ml}$ ). Etapes : 1) activation des fonctions acides carboxyliques du capteur par une solution de s-NHS - EDC, 2) Immobilisation de la poly-D-lysine, 3) Blocage des groupements résiduels activés par une solution d'éthanolamine 1 M pH 8,5, 4) lavage par du tampon HBS, 5) Capture des globules rouges, 6) et 7) ligne de base en tampon HBS, 8) Mise en contact avec le fibrinogène, 9) Dissociation en tampon HBS. Les courbes sont alignées au temps zéro.



**Figure 158 : Sensorgrammes alignés obtenus lors des expériences d'interférométrie de biocouche.** Les courbes sont alignées au début de la phase d'association. La ligne grise en pointillés indique la fin de la phase d'association et le début de la phase de dissociation.



Etant donné que la quantité de globules rouges capturés à la surface des capteurs varie entre les différentes expériences, une normalisation a été effectuée. La valeur du *shift* correspondant à la fixation du fibrinogène sur les globules rouges a été normalisée en la divisant par la valeur du *shift* correspondant à la capture des globules rouges sur le capteur. (**Tableau XXXII**). Les valeurs d'interaction normalisées des globules rouges non traités ont ensuite été divisées par celles de globules rouges traités à des fins de comparaison. Si la valeur est supérieure à l'unité, alors on peut déduire que le fibrinogène interagit davantage avec les globules rouges non traités. Cette valeur est supérieure à 1 dans nos expériences (de 2,9 à 4,5) indiquant la fixation d'une plus grande quantité de fibrinogène sur les globules rouges non traités. Les globules rouges traités conservent leur capacité à interagir avec le fibrinogène quelle que soit les conditions de prélèvement, indiquant que l'interaction globule rouge – fibrinogène n'est pas uniquement dépendante des acides sialiques. Il est cependant important de noter que les *shifts* observés en BLI ne sont pas directement proportionnels au nombre de cellules capturées et qu'il peut donc exister un biais dans l'analyse.

Collecte du sang / Traitement à la neuraminidase	Shift (nm)			Rapport des <i>shifts</i> normalisés : Non traités / traités
	Capture des globules rouges	Fixation du fibrinogène	Fixation du fibrinogène normalisée	
<b>EDTA</b> / Non traités	0,39	0,19	0,49	2,9
<b>EDTA</b> / Traités	0,83	0,14	0,17	
<b>Héparine</b> / Non traités	0,36	0,26	0,72	4,5
<b>Héparine</b> / Traités	0,92	0,15	0,16	
<b>Citrate</b> / Non traités	0,37	0,19	0,51	3,9
<b>Citrate</b> / Traités	0,80	0,10	0,13	

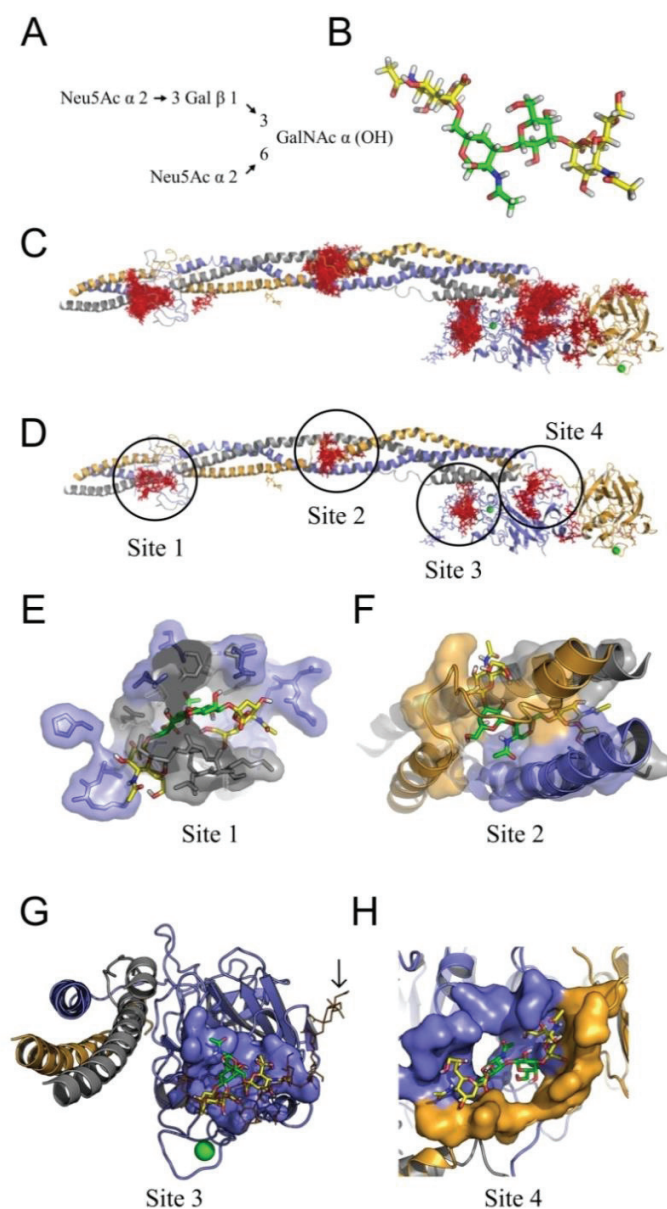
**Tableau XXXII : Résultats des expériences d'interférométrie de biocouche.** Les *shifts* (nm) correspondent aux valeurs obtenues à la fin de l'étape de capture des globules rouges sur les capteurs modifiés par la poly-D-lysine et de l'étape d'interaction avec le fibrinogène, respectivement. Les *shifts* correspondant à la fixation du fibrinogène ont été normalisés par le *shift* correspondant à la capture des globules rouges. Le rapport globules rouges non traités/traités correspond au *shift* normalisé des globules rouges non traités sur celui des traités.

Afin d'apporter des preuves supplémentaires de l'interaction entre le fibrinogène et les acides sialiques, des expériences de docking ont été réalisées. Elles permettront également de localiser le ou les sites d'interactions privilégiés entre ces deux molécules. L'acide N-acétylneuraminique (NANA) est l'acide sialique le plus commun à la surface des globules rouges (*Bulai et al., 2013*). Comme rappelé précédemment, environ 75 % des acides sialiques sont attachés à la glycophorine A (*Aoki et al., 2017*). Chacune des 16 O-glycosylations de la GPA (O-GPA) contiennent deux résidus de NANA à leur extrémité (**Figures 159A** et **159B**), alors qu'un seul résidu de NANA est porté par l'unique N-glycosylation de la GPA. Ainsi, seule la O-glycosylation de la GPA a été utilisée (**Figure 159**) en tant que ligand pour les expériences de docking. Ces expériences ont également été réalisées en utilisant uniquement NANA et sont présentées en **Annexe 9**. Le fibrinogène est un dimère de trimère très étendu d'environ 45 nm de longueur. Etant donné sa taille, seul un monomère composé de trois chaînes ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) a été utilisé en tant que récepteur dans nos expériences de docking. Il a été divisé en 5 zones chevauchantes incluant l'interface entre les deux monomères.

La O-glycosylation de la GPA ou le résidu de NANA seul interagissent de façon prépondérante avec 7 sites à la surface du fibrinogène humain (**Figure 159C**). Leur énergie d'interaction varie de - 6,2 à -8,5 kcal/mol pour la O-GPA, et de - 4,4 à - 6,2 kcal/mol pour le résidu de NANA. NANA, seul ou présent dans la O-GPA, interagit principalement avec ces 7 sites mais 4 d'entre eux regroupent 187 des 200 poses (93,5 %) pour la O-GPA, et 160 des 200 poses (80 %) pour NANA seul (**Figure 159D** et **Annexe 9**). Deux sites possèdent les énergies d'interaction les plus favorables. Le premier, dont l'énergie d'interaction est de - 8,3 kcal/mol (site 1), est situé à l'interface des deux monomères du fibrinogène, dans une poche formée par les extrémités N-terminales des chaînes  $\alpha$  et  $\beta$  (**Figure 159E**). Le second (site 2) est dans une poche sur le domaine *coiled-coil*, formée par l'interruption de l'hélice de la chaîne  $\gamma$ . La O-glycosylation de la GPA se retrouve enroulée autour de l'hélice formée par la chaîne  $\alpha$  et maintenue par les chaînes  $\beta$  et  $\gamma$ . La meilleure pose possède une énergie d'interaction de - 8,5 kcal/mol, correspondant à une valeur calculée de constante d'équilibre de dissociation ( $K_D$ ) de 1,0  $\mu$ M (**Figure 159F**). Cette valeur de  $K_D$  calculée est très proche de la valeur expérimentale de l'interaction entre les globules rouges et le fibrinogène (1,3  $\mu$ M, *Lominadze & Dean, 2002*).

Les sites 2 et 3 sont localisés sur les parties globulaires du fibrinogène, formées par les chaînes  $\beta$  et  $\gamma$ . Le troisième site est proche des glycosylations propres au fibrinogène (**Figure 159G**). Le quatrième site est localisé entre les chaînes  $\beta$  et  $\gamma$  à la jonction des têtes globulaires des deux chaînes et de la région *coiled-coil* (**Figure 159H**).

Les énergies d'interaction les plus favorables sont similaires (-7,2 et - 7,5 kcal/mol, respectivement). Les résidus situés à moins de 5 Å de la O-glycosylation de la GPA sont listés dans l'**Annexe 10**. On retrouve particulièrement dans le site 2 les résidus Gly<sup>96</sup> et Asp<sup>97</sup> appartenant à la séquence <sup>95</sup>RGD<sup>97</sup>, qui est une séquence reconnue par certaines intégrines, suggérant leur implication dans l'interaction entre les globules rouges et le fibrinogène. Cette hypothèse a été proposée (Carvalho et al., 2011, Lominadze & Dean, 2002). D'autant plus que la sous-unité d'intégrine β3 est sialylée (Chiodelli et al., 2012).



**Figure 159 : Docking entre la O-GPA et le fibrinogène humain. (A-B)**

Structures de la O-glycosylation (O-GPA) de la glycophorine A. **C)** Superposition des 200 poses calculées par AutoDock Vina (rouge) de la O-GPA sur le monomère de fibrinogène (chaînes α : grise, β : bleue, γ : orange). Ion calcium : sphère verte. **D)** Seules les 5 meilleures poses ayant les énergies les plus faibles pour chaque zone testée sont représentées mettant en avant 4 sites préférentiels d'interaction. La légende est identique à C). **E-H)** Agrandissement des 4 sites préférentiels. La meilleure solution pour chacun des 4 sites est en vert avec les deux résidus de NANA en jaune. La couleur des chaînes est la même qu'en C) et D). La flèche en G) pointe la glycosylation du fibrinogène.

#### IV.4.4 Discussion

Les globules rouges drépanocytaires sont caractérisés par une déformabilité réduite par rapport à celle des globules rouges sains et forment également des agrégats plus robustes (Connes *et al.*, 2016). Notre étude est la première à montrer que l'agrégation et l'agrégabilité des globules rouges drépanocytaires sont supérieures à celles des globules rouges sains, et que des facteurs cellulaires et plasmatiques peuvent être impliqués dans ces processus. L'agrégation anormale des globules rouges participe à la physiopathologie de la drépanocytose (Lamarre *et al.*, 2012, Lapouméroulie *et al.*, 2019). Cependant, le rôle du glycocalyx dans les propriétés rhéologiques des globules rouges drépanocytaires n'a pas été étudié. Le but de cette étude a été de déterminer et de comparer le rôle de ses composants dans la déformabilité, l'agrégation et l'agrégabilité des globules rouges sains et drépanocytaires. Les principaux résultats du rôle du glycocalyx de globules rouges dans la modulation de leurs propriétés rhéologiques sont résumés dans la table suivante (**Tableau XXXIII**).

Enzymes	Globules rouges sains			Globules rouges drépanocytaires		
	Déformabilité	Agrégation	Agrégabilité	Déformabilité	Agrégation	Agrégabilité
<b>Héparinase III</b>	~	~	~	~	~	~
<b>Chondroïtinase ABC</b>	~	~	~	~	~	~
<b>Hyaluronidase</b>	(↗)	~	↗ (*) + 22,0 %	~	~	↗ (*) + 12,5 %
<b>Neuraminidase</b>	↘ (*) - 3,8 % à 30 Pa	↘ (*) - 51,2 %	↗ (*) M : + 80,5 % M1 : + 107,3 %	↗ (*) + 25,3 % à 0,95 Pa	↘ (*) - 61,9 %	↗ (*) M : + 88,9 % M1 : + 93,1 %

**Tableau XXXIII : Effets des digestions enzymatiques sur la déformabilité, l'agrégation et l'agrégabilité des globules rouges sains et drépanocytaires.** Le seuil de significativité est indiqué par une étoile (\*). Le symbole (↗) indique une augmentation non significative sur le plan statistique et les vaguelettes (~) indiquent l'absence d'effets. Les fonctions M et M1 indiquent que les mesures d'agrégation / agrégabilité ont été réalisées en stase où à faible vitesse de cisaillement, respectivement.

La digestion enzymatique des chaînes d'héparane sulfate et de chondroïtine sulfate n'a pas affecté de façon significative les propriétés rhéologiques des globules rouges. La digestion de l'acide hyaluronique a cependant augmenté l'agrégabilité des globules rouges sains (en stase) et drépanocytaires (à faible vitesse de cisaillement). L'acide hyaluronique est une très grosse molécule dont la taille étendue peut atteindre 25  $\mu\text{m}$  (Toole, 2004). Son volume hydrodynamique peut donc créer un réseau tridimensionnel de taille importante à la surface des globules rouges, de façon similaire au réseau qu'il forme à la surface des cellules endothéliales. Sa digestion favoriserait l'accessibilité du dextran à la surface des globules rouges, provoquant une agrégation accrue. Ces résultats sont en accord avec de précédentes expériences où des mesures rhéologiques ont été effectuées sur des globules rouges incubés avec de l'acide hyaluronique exogène, ce qui a conduit à une diminution de la déformabilité et l'agrégabilité des globules rouges (Luquita et al., 2010). L'absence d'effets majeurs des GAGs sur les propriétés rhéologiques des globules rouges peut s'expliquer par leur faible quantité à la surface des cellules. En effet, le glycocalyx des globules rouges est environ 1 000 fois plus fin que celui des cellules endothéliales, estimé entre 1,7 et 10 nm (Kahane et al., 1978, Donath & Pastushenko, 1979, Linss et al., 1991).

En revanche, la digestion des acides sialiques provoque des modifications importantes des paramètres rhéologiques. L'augmentation de l'agrégabilité des globules rouges après le traitement à la neuraminidase concorde avec le fait que les acides sialiques exercent un puissant effet antiagrégant causé par leur répulsion électrostatique, ce qui a déjà été montré (Jan & Chien, 1973). L'accessibilité du dextran à la membrane des globules rouges est accrue, provoquant leur agrégation (Lee et al., 2018). Après traitement, les globules rouges drépanocytaires s'agrègent davantage que les globules rouges sains en dextran. Il a été montré que les globules rouges drépanocytaires possédaient une quantité réduite d'acide sialique à leur surface comparé aux sains (Aminoff et al., 1980). Ainsi, dans les mêmes conditions et après traitement par la neuraminidase, une quantité plus faible d'acide sialique à la surface des globules rouges drépanocytaires pourrait subsister par rapport aux globules rouges sains, permettant ainsi une meilleure accessibilité du dextran à la surface des globules rouges drépanocytaires.

En revanche, l'agrégation est diminuée pour les deux types de globules rouges resuspendus en plasma après digestion des acides sialiques. Cette observation est en accord avec une précédente étude réalisée sur des sangs sains (Nash et al., 1987).

Cela suggère le rôle d'un facteur plasmatique pouvant interagir avec les acides sialiques des globules rouges. Les expériences de BLI ont révélées que l'interaction entre les globules rouges sains et le fibrinogène est réduite après traitement par la neuraminidase.

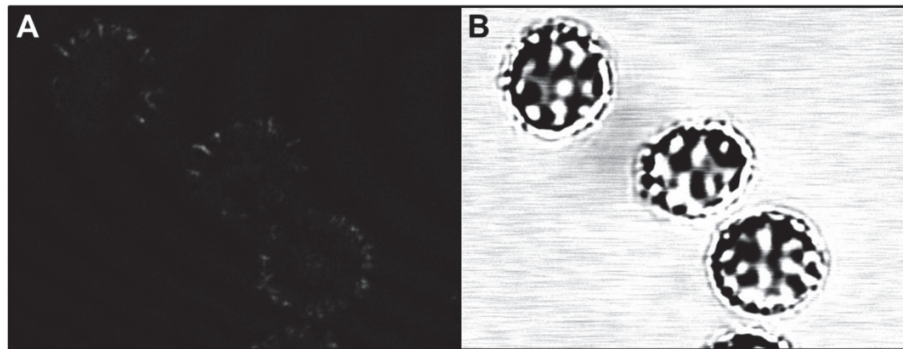
Bien que l'interaction entre le fibrinogène et les globules rouges ait déjà été rapportée, le mécanisme moléculaire de cette interaction est inconnu. CD47 (*Oliveira et al., 2012*) ou des intégrines RGD-dépendantes (*Carvalho et al., 2011, Lominadze & Dean, 2002*) ont été proposés en tant que récepteurs à la surface des globules rouges. Nous avons montré par docking que l'acide N-acétylneuraminique, seul ou a sein des O-glycosylations de la glycophorine A, se fixe sur le fibrinogène, notamment sur quatre sites préférentiels. L'un des deux sites les plus favorables contient les résidus 96 à 207 de la chaîne  $\alpha$  du fibrinogène, les résidus 135 à 139 de la chaîne  $\beta$  et les résidus 67 à 80 de la chaîne  $\gamma$ . Le deuxième site le plus favorable contient les résidus Gly<sup>96</sup> et Asp<sup>97</sup> à moins de 5 Å de l'acide sialique. Ces résidus sont intégrés dans la séquence <sup>95</sup>RGD<sup>97</sup>, soutenant l'hypothèse de l'implication d'une intégrine, et plus particulièrement de la sous-unité  $\beta 3$  d'intégrine précédemment proposé (*Carvalho et al., 2018*). De plus, l'utilisation du peptide antagoniste RGDS inhibe l'interaction du fibrinogène à la surface des globules rouges (*Lominadze & Dean, 2002*) et la sous-unité  $\beta 3$  est sialylée à la surface des cellules endothéliales (*Chiodelli et al., 2012*). Il est ainsi possible de suggérer que de cette sous-unité sialylée présente à la surface des globules rouges interagit avec le fibrinogène, de la même manière que pourrait le faire la glycophorine A. L'énergie d'interaction (- 8,5 kcal/mol) de ce site correspond une affinité de 1,0  $\mu$ M, ce qui est similaire à la valeur expérimentale reportée de l'interaction entre fibrinogène et globules rouges (1,3  $\mu$ M, *Lominadze & Dean, 2002*).

La digestion des acides sialiques affecte également la déformabilité des globules rouges sains et drépanocytaires. La déformabilité des globules rouges sains a diminué dans des conditions de force de cisaillement moyennes à élevées (5,33 à 30 Pa), ce qui concorde avec des résultats précédemment obtenus (*Yang et al., 2014*). Il a été montré qu'une réduction de la déformabilité au-dessus de 3 Pa est corrélée à un rapport surface / volume diminué, une perte de la flexibilité membranaire ou une augmentation de la viscosité interne (*Renoux et al, 2019*). Au contraire, la déformabilité des globules rouges drépanocytaires augmente après le traitement par la neuraminidase mais seulement à de faibles forces de cisaillement (0,95 à 5,33 Pa). Une faible concentration en neuraminidase est suffisante pour affectée leur déformabilité alors qu'aucun effet n'est détecté pour les globules rouges sains.

#### IV.4.5 Perspectives

Cette étude est la première à montrer que l'agrégation et l'agrégabilité sont plus importantes chez les globules rouges drépanocytaires que chez les globules rouges sains, et que des facteurs plasmatiques y sont impliqués. Les GAGs ne jouent pas de rôle majeur dans ces propriétés rhéologiques bien que l'acide hyaluronique exerce un faible effet antiagrégant. Une faible augmentation de la déformabilité des globules rouges sains est cependant observable lorsqu'il est digéré. Augmenter le nombre d'expériences serait un moyen de trancher sur la question du rôle de l'acide hyaluronique dans la déformabilité des globules rouges, d'autant plus qu'un de ces récepteurs, CD44, est directement lié au cytosquelette *via* les protéines de la famille ERM (ezrine, radixine, moésine) (*Thorne et al., 2004*).

La structure et la composition du glycocalyx des globules rouges n'ont été que peu étudiées. Une unique observation par microscopie électronique est à ce jour disponible (*Roseman, 2004*, voir section *Introduction*), ainsi qu'une détection des chaînes d'héparane sulfate par immunofluorescence (*Vogt et al., 2004*). Aucune information concernant la composition et les dimensions du glycocalyx des globules rouges drépanocytaires n'est disponible. Il est cependant possible que la taille de leur glycocalyx soit réduite étant donné qu'il possède une quantité d'acide sialique moins importante à leur surface (*Aminoff et al., 1980*), mais aussi parce que la taille du glycocalyx endothélial et celui des globules rouges sont corrélées, et il a été montré que le volume du glycocalyx des cellules endothéliales de patients drépanocytaires est réduit (*van Beers et al., 2008*). L'utilisation de techniques plus récentes, comme la microscopie optique à super-résolution, pourrait permettre de déterminer plus précisément la taille du glycocalyx des globules rouges sains et drépanocytaires. Nous avons effectués des essais de visualisation par microscopie confocale du glycocalyx de globules rouges sains en les incubant avec la lectine de *Maackia amurensis* (M2757, CliniSciences) conjuguée à une sonde fluorescente (isothiocyanate de fluorescéine, FITC) qui reconnaît les acides sialiques. Ces essais ont été peu concluants étant donné d'une part des faibles intensités de fluorescence et d'autre part de la déformation des globules rouges en échinocytes, c'est-à-dire des hématies présentant de nombreux spicules à leur surface (**Figure 160**).



**Figure 160 : Observation du glycocalyx de globules rouges sains par microscopie confocale.** Le sang a été prélevé en tube EDTA et lavé dans les mêmes conditions que décrites précédemment et le culot de globule rouge a ensuite été resuspendu dans du PBS (hématocrite  $\approx 50\%$ ). Une suspension de globules rouges (500 cellules /  $\mu\text{l}$ ) a été incubée dans l'obscurité pendant 20 min sous agitation (400 rpm) et à  $37^\circ\text{C}$  en présence de la lectine de *Maackia amurensis* (200  $\mu\text{g/ml}$ ), puis lavée trois fois dans les mêmes conditions. Environ 10  $\mu\text{l}$  ont été déposés sur une lame de verre et observés **(A)** par microscopie confocale ( $\lambda_{\text{excitation}}$  : 492 nm,  $\lambda_{\text{émission}}$  : 517 nm) et simultanément par **(B)** microscopie optique avec un zoom optique X40. Le diamètre d'un globule rouge est d'environ 7  $\mu\text{m}$ .

Le glycocalyx endothélial joue un rôle de mécanorécepteur. Cette fonction est notamment réalisée par les chaînes d'héparane sulfate portées par le glypican-1 (*Thi et al., 2004, Zeng et al., 2018*). Les forces appliquées sur l'endothélium mènent à la production et à la libération de monoxyde d'azote (NO) par les cellules endothéliales et la digestion des chaînes d'héparane sulfate, du hyaluronane ou des acides sialiques inhibe la production par la NO synthase (*Pahakis et al., 2007*). Le monoxyde d'azote joue un rôle crucial dans la maintenance de l'homéostasie vasculaire (*Tousoulis et al., 2012*). La NO synthase est également présente dans les globules rouges et son inhibition mène à une diminution de leur déformabilité (*Bor-Kucukatay et al., 2003, Cortese-Krott & Kelm, 2014*). Cependant, la digestion des chaînes d'héparane sulfate et de l'acide hyaluronique n'a pas provoqué de variation significative de la déformabilité dans nos expériences ce qui suggère que ces molécules ne sont probablement pas impliquées dans le phénomène de mécanoréception chez les globules rouges.

A l'inverse, chez les globules rouges sains, la digestion des acides sialiques entraîne une diminution de la déformabilité, qui s'effectue à partir d'une valeur seuil de 5,33 Pa dans nos expériences. Un mécanisme semblable aux cellules endothéliales pourrait donc exister pour les globules rouges sains.



Les acides sialiques présents à leur surface pourraient jouer un rôle de mécanorécepteurs, provoquant une augmentation de la synthèse de NO par les globules rouges et menant à une déformabilité accrue de ces cellules, notamment dans des conditions de forces de cisaillement importantes. D'ailleurs, une production de NO est accrue par les globules rouges subissant des forces de cisaillement (*Ulker et al., 2011*). Afin de confirmer cette hypothèse, il pourrait être envisagé de doser la quantité de NO présente dans des globules rouges soumis à des forces de cisaillement et dont les acides sialiques auraient été préalablement digérés. Etant donné qu'ils sont majoritairement portés par la glycophorine A, les globules rouges n'exprimant pas cette protéine devraient avoir une déformabilité plus faible si cette hypothèse est valable.



## CORRESPONDENCE



## Sialic acids rather than glycosaminoglycans affect normal and sickle red blood cell rheology by binding to four major sites on fibrinogen

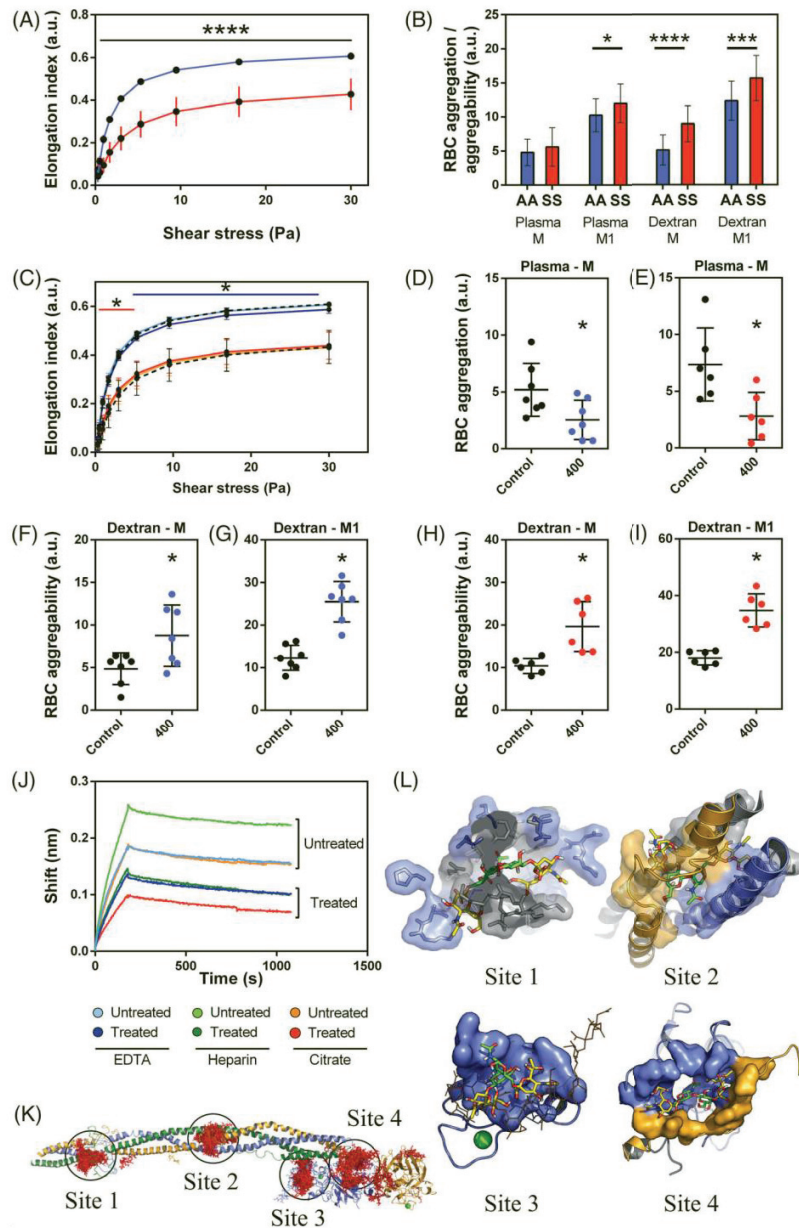
To the Editor:

Sickle cell anemia (SCA) is characterized by impaired red blood cell (RBC) rheology. The loss of RBC deformability plays a key role in the occurrence of painful vaso-occlusive events in SCA.<sup>1</sup> In addition, the strength of RBC aggregates correlates with the occurrence of acute chest syndrome, and RBC aggregation increases during acute vaso-occlusive crisis in adults with SCA.<sup>1</sup> However, the molecular mechanisms leading to abnormal RBC aggregation in SCA are currently unknown. The glycocalyx, covering the cell surface, contains glycoproteins with sialic acids (SA) and proteoglycans made of glycosaminoglycans (GAGs) covalently attached to core proteins.<sup>2</sup> The glycocalyx allows the adhesion of circulating cells to the endothelium, but its role in the modulation of the RBC rheological properties in healthy or SCA individuals has not been investigated so far. The aim of this study was to find the role of GAGs and SA in the deformability and aggregability of RBCs from healthy subjects and SCA patients. Also, the study was to confirm the binding of fibrinogen to SA displayed at the RBC surface, and to identify the binding sites of SA on fibrinogen.

For RBCs deformability, aggregability and aggregation were measured after enzymatic treatment with GAG- and SA-degrading enzymes. Blood was collected on EDTA from healthy subjects and SCA patients in clinical steady state. Informed consent was obtained for every individual. The protocol was approved by the "Hospices Civils de Lyon-CPP Est" Ethics Committee (L14-127) and was carried out in accordance with the declaration of Helsinki. Blood cells were washed three times (500 g, 4°C, 10 minutes) in phosphate buffered saline (PBS). Healthy and sickle RBCs were incubated for 60 minutes in PBS at 37°C at a hematocrit of 40% with four enzymes from Sigma-Aldrich: heparinase III (H8891), chondroitinase ABC (C2905), hyaluronidase (H3631), and neuraminidase (N2876). The RBCs were then centrifuged at 500g for 5 minutes, washed twice with PBS and resuspended at a final hematocrit of 40% in autologous plasma or in 3% (w/v) dextran (70 kDa, Sigma-Aldrich, 31 390) in PBS. Treated RBCs were oxygenated for at least 10 minutes before measurements.<sup>3</sup> The RBC aggregation was measured with a Myrenne aggregometer (GmbH, Roetgen) in the stasis (M) and low shear (M1) modes.<sup>3</sup> Measurements in autologous plasma reflected RBC aggregation while the measurements made in dextran reflected RBC aggregability (ie, the cell intrinsic properties to form RBC aggregates). So, RBC deformability (elongation index EI) was measured by ektacytometry at 9 shear stresses (from 0.3 to 30 Pa).<sup>3</sup>

Red blood cell deformability was lower in SCA than in healthy individuals (Figure 1A). Both RBC aggregation, and aggregability were higher in SCA than in healthy subjects (Figure 1B). The difference in aggregability between sickle and healthy RBCs was higher than the difference in aggregation: +43.1% for M and + 21.3% for M1 in dextran vs +14.6% for M and + 14.5% for M1 in plasma. The treatment of healthy and sickle RBCs with heparinase III (heparan sulfate digestion) or chondroitinase ABC (digestion of chondroitin sulfate A, B and C) did not affect RBC deformability, aggregation or aggregability (data not shown). The digestion of hyaluronan (HA) had no impact on healthy and sickle RBC deformability (data not shown) or on RBC aggregation in both groups (data not shown). But, it induced a significant increase in aggregability of healthy RBCs (M mode, from 3.54 ± 1.5 to 4.7 ± 1.3 a.u. - arbitrary unit) and of sickle RBCs (M1 mode, from 13.4 ± 2.4 to 15.1 ± 2.2 a.u.). Note, SA digestion by neuraminidase at 600 mU/mL decreased the deformability of healthy RBCs from mid (5.33 Pa) to high (30 Pa) shear stresses but increased sickle RBC deformability at low to mid shear stresses (0.95 to 5.33 Pa) (Figure 1C). The use of a lower neuraminidase concentration (400 mU/mL) did not affect deformability in healthy RBCs but increased it in sickle RBCs (Figure 1C). Red blood cell aggregation decreased in both groups after SA digestion (Figure 1D,E). Aggregability increased in both healthy and sickle RBCs after incubation with 400 mU/mL of neuraminidase (Figure 1F-I). The increased aggregability and decreased aggregation after SA digestion suggest that a plasma factor could bind to SA and facilitate RBC aggregation. Fibrinogen, which promotes RBC aggregation,<sup>4</sup> and binds to RBCs via possible contribution of SA,<sup>5</sup> is a likely candidate.

Bio-layer interferometry assays were carried out in an OctetRed96 (Molecular Devices) to investigate fibrinogen binding to healthy RBCs. Blood was collected in EDTA, heparinized and citrated tubes because fibrinogen binds to heparin, zinc and calcium, and centrifuged for 10 minutes at 500g. Red blood cells were washed three times in 10 mM Hepes, 133 mM NaCl and 4.5 mM KCl pH 7.4 (HBS), and resuspended at 40% hematocrit in HBS. Poly-D-lysine (100 µg/mL in 1 M NaCl) was covalently bound to amine reactive sensors (AR2G, Molecular Devices). The coated sensors were dipped in 1 × 10<sup>6</sup> RBCs/µl HBS for 20 minutes. The sensors with captured RBCs were dipped in fibrinogen (F3879, Sigma, 100 µg/mL in HBS + 1 mM CaCl<sub>2</sub>) for 180 seconds and then in HBS for 360 seconds to dissociate RBC-fibrinogen complexes. The binding levels of fibrinogen



**FIGURE 1** (A) Deformability of healthy (blue,  $n = 24$ ) and sickle RBCs (red,  $n = 24$ ). (B) Aggregation and aggregability of healthy (blue,  $n = 23$ ) and sickle RBCs (red,  $n = 23$ ) (M and M1 modes). Statistical analyses were performed with a Mann-Whitney unpaired  $t$  test ( $*P < .05$ ;  $***P < .001$ ;  $****P < .0001$ ). (C) Deformability of untreated healthy and sickle RBCs (black dotted lines), and of treated healthy (blue,  $n = 7$ ) and sickle (red,  $n = 6$ ) RBCs. Bright and dark colors: 400 and 600 mU/mL of neuraminidase respectively. (D) Aggregation of healthy (blue,  $n = 7$ ) and (E) sickle RBCs (red,  $n = 6$ ) in autologous plasma after treatment with 400 mU/mL of neuraminidase (M mode). (F-I) Aggregability of healthy (blue,  $n = 7$ ) and sickle RBCs (red,  $n = 6$ ) in dextran after treatment with 400 mU/mL of neuraminidase (M and M1 modes). A one-way Friedman ANOVA with repeated measurements followed by Dunn post hoc test was used to assess the effects of enzymes on RBC rheological properties ( $*P < .05$ ). (J) Bio-layer interferometry binding assays: binding of soluble fibrinogen to neuraminidase-treated and untreated healthy RBCs collected on EDTA, heparin and citrate captured on AR2G sensors. (K) Docking of the glycoprotein O-glycosylation motif (O-GPA) containing two NANA residues on a fibrinogen monomer (green, blue and orange:  $\alpha$ ,  $\beta$ , and  $\gamma$  chains respectively). The 200 superimposed solutions are in red and calcium ions in green. (L) The four binding sites of O-GPA on fibrinogen: the best solution for each site is in green and NANA residues in yellow (chains are color-coded as in K)


to RBCs were normalized to the levels of captured RBCs. Fibrinogen bound to 2.9 to 4.5-fold more to untreated than to treated RBCs depending on collection conditions (Figure 1J). The binding sites of SA on fibrinogen were identified by docking using the 3D structure of human fibrinogen (PDB 3GHG), and a single N-acetylneuraminic acid (NANA), the most common SA at the RBC surface, or the O-glycan of glycophorin-A (O-GPA) including two NANA residues because 75% of the RBC SA are attached to GPA.<sup>6</sup> The NANA and O-GPA structures were created with CarbohydrateBuilder (Glycam.org). Fibrinogen is a homodimer but a single monomer ( $\alpha$ ,  $\beta$  and  $\gamma$ -chains) was used for docking because monomers associate in a head-to-tail manner. The monomer was divided in five overlapping grids (cubes with sides of 90 Å each, spacing 1 Å) including the interface between the monomers. Autodock Vina (Autodock suite 4.2.6) was run with an exhaustiveness of 32, and 40 models were generated per grid (200 models in total). Both the single NANA residue and O-GPA were docked on the same seven sites on human fibrinogen (Figure 1K), with binding energies from  $-4.4$  to  $-6.2$  kcal/mol for NANA and from  $-6.2$  to  $-8.5$  kcal/mol for O-GPA. Four out of the seven sites clustered 187 poses for O-GPA and 160 poses for NANA over the 200 computationally determined. The most favorable binding sites for O-GPA were located at the interface between both fibrinogen monomers in a pocket formed by the N-termini of  $\alpha$  and  $\beta$  chains (Figure 1L, site 1) or in a pocket of the coiled-coil domain. This is where O-GPA is wrapped around the helix formed by the  $\alpha$  chain and maintained by the  $\beta$  and  $\gamma$  chains (Figure 1L, site 2) with binding energies of  $-8.3$  and  $-8.5$  kcal/mol, respectively, corresponding to a calculated  $K_d$  of  $1.0 \mu\text{M}$  ( $K_d = \exp(\Delta G/RT)$ ,  $T = 310 \text{ K}$ ). The second site encompasses the RGD tripeptide (95-97) involved in the binding of fibrinogen to RBCs, supporting the involvement of the sialylated RGD-dependent  $\alpha\text{v}\beta 3$  integrin in fibrinogen binding to RBCs.<sup>5,7</sup> The third and fourth binding sites, with binding energies of  $-7.2$  and  $-7.5$  kcal/mol, respectively, were located on  $\beta$  (mainly) and  $\gamma$ -globular heads of fibrinogen.

This is the first report showing that both RBC aggregation and aggregability are increased in sickle RBCs, suggesting that both plasma and cellular factors could be involved in this process. Sulfated GAGs, present in low amount on the RBC surface, do not play a significant role in the deformability and aggregability of healthy or sickle RBCs, but the digestion of a non-sulfated GAG, hyaluronan, increases the aggregability of healthy and sickle RBCs. In contrast, SA play a major role in RBC aggregation, likely through antagonist effects, namely strong electrostatic repulsion vs fibrinogen binding. The increased aggregation in sickle RBCs suggests that the decreased electrostatic repulsion caused by the lower amount of SA at their surface compared to healthy RBCs could play a prominent role compared to the pro-aggregant effect of fibrinogen.

#### ACKNOWLEDGEMENTS

This work was supported by a grant from the Fondation pour la Recherche Médicale (grant n°DBI20141231336 to SRB), by the Groupement de Recherche (GDR) GagoSciences [CNRS, GDR 3739,

Structure, Fonction et Régulation des Glycosaminoglycanes to SRB] and funding by the Institut Universitaire de France to PC. We are grateful to Stéphanie Baud (MEDyC, UMR CNRS 7369, University of Reims Champagne-Ardenne, Reims, France) for her advice on docking experiments.

Frank Gondelaud<sup>1</sup>, Philippe Connes<sup>2,3</sup>, Elie Nader<sup>2,3</sup>,  
Céline Renoux<sup>2,3,4</sup>, Romain Fort<sup>2,3,5</sup>, Alexandra Gauthier<sup>2,3,6</sup>,  
Philippe Joly<sup>2,3,4</sup>, Sylvie Ricard-Blum<sup>1</sup> 

<sup>1</sup>Univ Lyon, University Claude Bernard Lyon 1, CNRS, INSA Lyon, CPE, Institute of Molecular and Supramolecular Chemistry and Biochemistry (ICBMS), Villeurbanne, France

<sup>2</sup>Univ Lyon, University Claude Bernard Lyon 1 EA 7424, Laboratoire Interuniversitaire de Biologie de la Motricité (LIBM), Equipe "Biologie Vasculaire et du Globule Rouge", Villeurbanne, France

<sup>3</sup>Laboratoire d'Excellence GR-Ex, Paris, France

<sup>4</sup>Laboratoire de Biochimie et de Biologie Moléculaire, UF de biochimie des pathologies érythrocytaires, Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, Lyon, France

<sup>5</sup>Département de Médecine Interne, Hôpital Edouard Herriot, Hospices Civils de Lyon, Lyon, France

<sup>6</sup>Institut d'Hématologie et d'Oncologie Pédiatrique, Hospices Civils de Lyon, Lyon, France

#### Correspondence

Sylvie Ricard-Blum, Univ Lyon, University Claude Bernard Lyon 1, CNRS, INSA Lyon, CPE Institute of Molecular and Supramolecular Chemistry and Biochemistry (ICBMS) Villeurbanne 69622, France.  
Email: sylvie.ricard-blum@univ-lyon1.fr

#### ORCID

Sylvie Ricard-Blum  <https://orcid.org/0000-0001-9263-1851>

#### REFERENCES

- Connes P, Renoux C, Romana M, et al. Blood rheological abnormalities in sickle cell anemia. *Clin Hemorheol Microcirc.* 2018;68:165-172.
- Gondelaud F, Ricard-Blum S. Structures and interactions of syndecans. *FEBS J.* 2019;286:2994-3007.
- Baskurt OK, Uyuklu M, Ulker P, et al. Comparison of three instruments for measuring red blood cell aggregation. *Clin Hemorheol Microcirc.* 2009;43:283-298.
- Guedes AF, Moreira C, Nogueira JB, Santos NC, Carvalho FA. Fibrinogen-erythrocyte binding and hemorheology measurements in the assessment of essential arterial hypertension patients. *Nanoscale.* 2019;11:2757-2766.
- Carvalho FA, de Oliveira S, Freitas T, Gonçalves S, Santos NC. Variations on fibrinogen-erythrocyte interactions during cell aging. *PLoS One.* 2011;6:e18167.
- Aoki T. A Comprehensive Review of Our Current Understanding of Red Blood Cell (RBC) Glycoproteins. *Membranes.* 2017;7:56.
- Carvalho FA, Guedes AF, Duval C, et al. The 95RGD97 sequence on the A $\alpha$  chain of fibrinogen is essential for binding to its erythrocyte receptor. *Int J Nanomedicine.* 2018;13:1985-1992.

# **CHAPITRE V**

**CONCLUSIONS**

**&**

**PERSPECTIVES**

Les syndécans sont des protéoglycanes transmembranaires, excepté pour l'isoforme 2 du syndécan-4, et subissent le processus de shedding qui est favorisé lorsque les chaînes de glycosaminoglycanes sont digérées par l'héparanase. Les ectodomaines déglycosylés sont biologiquement actifs car ils conservent leur séquence médiant l'adhésion cellulaire qui a été décrite dans les quatre ectodomaines (McFall & Rapraeger, 1998, Whiteford et al., 2008, De Rossi et al., 2014, Jung et al., 2016). Nos travaux ont principalement porté sur les ectodomaines des syndécans exprimés sous forme recombinante sans leurs chaînes de glycosaminoglycanes. La partie protéique des ectodomaines conservent néanmoins certains rôles biologiques. L'inhibition de la différenciation des ostéoclastes par les quatre syndécans est conservée en absence des chaînes de GAGs mais de façon moins efficace qu'en leur présence (Kim et al., 2018). En absence des chaînes de GAGs, l'ectodomaine du syndécan-1 interagit avec les intégrines  $\alpha\beta3$ ,  $\alpha3\beta1$  et  $\alpha4\beta1$  qu'il active (Beauvais & Rapraeger, 2010, Jung et al., 2016). L'ectodomaine du syndécan-4 interagit également les intégrines  $\alpha3\beta1$  et  $\alpha5\beta1$  et les active (Fiore et al., 2014, Wang et al., 2015) et l'ectodomaine du syndécan-2 conserve son interaction avec le récepteur CD148 (*Receptor-type tyrosine-protein phosphatase eta*) (De Rossi et al., 2014). L'élimination des chaînes de GAGs de l'ectodomaine du syndécan-1 lui permet d'interagir avec la lacritine. Cette interaction se produit uniquement lorsque ses chaînes d'héparane sulfate ont été dégradées (Zhang et al., 2012). La présence des chaînes de GAGs modulent ainsi les fonctions biologiques des syndécans, ce qui est cohérent avec la modulation des fonctions d'autres protéines désordonnées par des modifications post-traductionnelles (Bah & Forman-Kay, 2016, Darling & Uversky, 2018).

### ***Caractérisation biophysique de la partie protéique des ectodomaines des syndécans humains***

Les études structurales des syndécans étaient jusqu'à présent limitées à celles portant sur leur domaine cytosolique et plus particulièrement celui du syndécan-4 grâce aux nombreux travaux du Pr. J. Couchman et de ses collaborateurs (Lee et al., 1998, Shin et al., 2001, Whiteford et al., 2008, Choi et al., 2016). Seule la structure du domaine transmembranaire du syndécan-2 a été résolue par résonance magnétique nucléaire (Li et al., 2019) mais les données structurales des ectodomaines des syndécans, qui représente la grande majorité de leur séquence, étaient limitées à la structure d'un octapeptide de l'ectodomaine du syndécan-1 cocristallisé avec un partenaire protéique où le peptide de syndécan n'est pas structuré (Briggs & Hohenester, 2018), ainsi qu'à la prédiction *in silico* du désordre intrinsèque (Leonova & Galzitskaya, 2015).

Afin d'étudier les ectodomaines des syndécans, nous les avons dans un premier temps exprimé sous forme recombinante sans leurs chaînes de glycosaminoglycanes en utilisant un système d'expression procaryote.

En accord avec les prédictions de désordre de *Leonova & Galzitskaya* (2015), l'ensemble de nos expériences montre que les ectodomaines des syndécans dépourvus de leurs chaînes de glycosaminoglycanes sont des protéines intrinsèquement non structurées et qu'ils possèdent des brins  $\beta$  dont la proportion est plus importante dans les deux plus petits ectodomaines (-2 et -4) que chez les plus grands (-1 et -3).

Les techniques de diffusion de la lumière (SEC-SAXS et SEC-DLS) ont mis en évidence leur nature très étendue de type *random coil* et leur flexibilité. En accord avec ces observations, les expériences de spectrométrie de masse en conditions natives et les expériences de mobilité ionique ont montré que les conformations les plus abondantes sont les plus étendues pour l'ectodomaine du syndécan-1, -2 et -4 mais pas pour l'ectodomaine du syndécan-3 dont les conformations plus compactes sont plus abondantes. Cette inadéquation observée avec les résultats obtenus par SEC-SAXS où le rayon de gyration d'ED-3 est important et cohérent avec des conformations très étendues peut s'expliquer par le fait que les mesures de SEC-SAXS sont effectuées en solution et celles de spectrométrie de masse en phase gazeuse. Les conformations très étendues sont favorisées par de très nombreuses interactions s'effectuant entre le solvant aqueux et les résidus d'acides aminés polaires comme c'est le cas lors des expériences de SEC-SAXS. L'absence de solvant aqueux a pu provoquer une compaction partielle de l'ectodomaine du syndécan-3, ce qui a pu être favorisé par des interactions hydrophobes, ces dernières étant conservées en phase gazeuse (*Frański et al., 2008, Barylyuk et al., 2011*), d'autant plus qu'il est enrichi en résidus hydrophobes par rapport à l'ectodomaine des autres syndécans.

Chez les quatre ectodomaines, la mesure de la surface accessible au solvant a révélé que leurs conformations les plus compactes étaient proches de conformations globulaires. Cette observation a été globalement vérifiée par la mesure de leurs sections efficaces de collision par mobilité ionique et leurs comparaisons à des valeurs théoriques calculées pour des conformations globulaires (sphère) ou très allongées (bâton). L'absence de calcul des valeurs de CCS plus faibles peut être la conséquence d'une proportion très faible de conformations compactes rendant leur détection délicate. Les valeurs de CCS expérimentales les plus faibles, reflétant les conformations les plus compactes, restent ainsi supérieures à celles d'une sphère parfaite. Cette observation a cependant été faite sur des protéines désordonnées et sur des protéines globulaires, le modèle étant trop restrictif (*Stuchfield & Barran, 2018*).

Néanmoins, les ectodomains des syndécans semblent pouvoir se replier dans des conformations proches de celles des protéines globulaires bien que la formation des conformations étendues soit favorisée. Ces conformations compactes pourraient être enrichies en brins  $\beta$  étant donné qu'ils ont été détectés par dichroïsme circulaire. Cela pourrait être la cause de la plus grande stabilité des conformations compactes d'ED-2 et d'ED-4 observée lors de l'activation des faibles états de charge effectuée en mobilité ionique par rapport à la stabilité d'ED-1 et d'ED-3. Le rôle biologique des conformations les plus compactes pourra être étudié en trouvant des conditions expérimentales favorisant leur compaction, comme par exemple l'ajout d'osmolytes à de fortes concentrations ou des polymères. Le repliement de protéines ou de régions désordonnées dans ces conditions n'est cependant pas systématique (Fonin et al., 2018). Certaines IDRs / IDPs se replient, du moins partiellement, en présence d'osmolytes et sont *foldable*, certaines sont insensibles (*non-foldable*) et d'autres gagnent au contraire davantage de désordre intrinsèque (*unfoldable*) (Fonin et al., 2018).

L'hétérogénéité conformationnelle, la dynamique et la présence de structures secondaires pourront être étudiées par résonance magnétique nucléaire (Salvi et al., 2017). Des expériences de RMN pourront également être réalisées dans des conditions mimant l'environnement cellulaire, comme l'encombrement stérique, en présence de polymères comme le dextran ou de polyéthylène glycol (Adamski et al., 2019). Des expériences de diffusion de la lumière (DLS et SAXS) et de dichroïsme circulaire pourront être réalisées dans les mêmes conditions expérimentales afin de détecter des modifications du degré de compaction des ectodomains des syndécans et leur potentiel enrichissement en structures secondaires.

Les données expérimentales acquises au cours de mes travaux pourront être utilisées comme contraintes pour la modélisation d'ensembles conformationnels dans des simulations par dynamique moléculaire. Nous avons utilisé le serveur UNRES (Czaplewski et al., 2018) en utilisant comme contraintes les données de SAXS afin de construire des modèles gros grains des syndécans et pour étudier leur comportement par dynamique moléculaire comme cela a été effectué pour une autre protéine désordonnée étudiée au laboratoire, le propeptide de la lysyl oxydase (Vallet et al., 2018). Cependant, il n'a pas été possible d'obtenir des modèles des syndécans pour l'instant mais nous collaborons avec le laboratoire d'Adam Liwo (Université de Gdansk, Pologne) qui a développé le programme.



Des simulations identiques à celles effectuées sur le syndécan-4 entier pourraient également être envisagées (*Cruz-Chu et al., 2014*) mais en utilisant la structure du domaine transmembranaire du syndécan-2 qui a récemment été résolue (*Li et al., 2019*) à la place de celui de la glycophorine A utilisé pour la modélisation du syndécan-4 entier par *Cruz-Chu et al. (2014)*.

### ***Caractérisation biophysique des ectodomaines des syndécans humains portant leurs chaînes de glycosaminoglycanes***

Nous avons ensuite exprimé et purifié les ectodomaines des syndécans -2, -3 et -4 en système eucaryote afin que les chaînes de glycosaminoglycanes soient synthétisées et attachées à la partie protéique des ectodomaines. L'étude structurale des ectodomaines des syndécans portant leurs chaînes de glycosaminoglycanes a été plus complexe, d'abord parce qu'il s'agit d'un système très polydispersé ce qui est causé d'une part par la nature désordonnée des syndécans et d'autre part par la présence des chaînes de glycosaminoglycanes dont la nature, la taille et la composition en modifications post-synthétiques peuvent varier d'une chaîne à l'autre au sein d'un même syndécan. Plusieurs espèces moléculaires ont effectivement été observées lors de l'analyse de l'ectodomaine du syndécan-3 par SEC-SAXS, chacune possédant un rayon de gyration similaire. L'analyse des ectodomaines des syndécans -2 et -4 purifiés a montré qu'ils étaient contaminés *a priori* par des acides nucléiques qui sont des contaminants fréquents des glycosaminoglycanes ou des protéoglycanes purifiés (*Lyon & Gallagher, 1991, van Gemst et al., 2016*). La digestion des acides nucléiques par des nucléases devraient permettre d'obtenir des protéoglycanes purifiés.

La caractérisation des ectodomaines portant leurs chaînes de GAGs et d'une façon générale des molécules hybrides lors des expériences de SAXS ou de DLS est limitée à des mesures globales comme celles de leur rayon de gyration et de leur taille maximale dans le cas d'expériences de SAXS, ou de leur rayon hydrodynamique lors d'expériences de DLS. Dans nos expériences de SEC-MALLS, la masse moléculaire des chaînes de GAGs calculée pour ED-3 est d'environ 60 kDa soit environ le double de la masse moléculaire de la protéine seule (38 kDa) alors que d'une manière générale, les chaînes d'héparane sulfate ont une masse moléculaire comprise entre 30 et 50 kDa (*Shriver et al., 2012, Meneghetti et al., 2015*). Dans un tel système, il n'est pas possible de déconvoluer les intensités de diffusion collectées lors d'expériences de SAXS avec les outils existants.

En revanche, il serait intéressant d'effectuer des mesures de dichroïsme circulaire afin de déterminer si les chaînes de glycosaminoglycanes ont un rôle sur la composition en structures secondaires dans les ectodomaines des syndécans. La mesure de la flexibilité de la partie protéique en SAXS est également biaisée étant donné que les chaînes de glycosaminoglycanes composent une grande partie de la masse moléculaire et du volume des ectodomaines des syndécans. Nos données préliminaires montrent que l'ectodomaine portant des chaînes de GAGs est plus flexible que l'ectodomaine seul mais il n'est pas possible de déterminer si la flexibilité de la partie protéique des ectodomaines des syndécans a été affectée.

Il serait possible de caractériser la dynamique des ectodomaines des syndécans par transfert d'énergie par résonance de type Förster (FRET) (*Schramm et al., 2019*). Des fluorophores doivent pour cela être covalamment fixés aux extrémités N- et C-terminales des ectodomaines des syndécans, l'efficacité du transfert d'énergie entre les deux fluorophores étant directement reliée à la distance moyenne qui les sépare. Cette approche permettra d'estimer le degré d'extension des conformations des syndécans. Ces expériences pourraient être réalisées en présence ou en absence des chaînes de glycosaminoglycanes afin de comparer directement l'effet des GAGs sur la dynamique de la partie protéine des ectodomaines des syndécans.

### ***Interactions des syndécans***

Les protéines intrinsèquement non structurées sont caractérisées par leur large interactome grâce à leur plasticité structurale leur offrant la capacité d'interagir avec un grand nombre de protéines (*Ekman et al., 2006, Haynes et al., 2006 Uversky et al., 2008*). L'analyse de l'interactome des syndécans a montré que la très grande majorité de leurs partenaires extracellulaire interagissait avec leurs chaînes de glycosaminoglycanes, principalement celles d'héparane sulfate et que leur interactome consensus se réduisait majoritairement à des protéines intracellulaires (*Gondelaud & Ricard-Blum, 2019*). Outre certaines métalloprotéases, les rares protéines interagissant directement avec la partie protéique des ectodomaines sont majoritairement des récepteurs de facteurs de croissance ou des intégrines avec lesquels ils forment des complexes ternaires (*Beauvais & Rapraeger, 2010, Fiore et al., 2014, Wang et al., 2015, Jung et al., 2016*). Leur flexibilité structurale peut également être la cause de la formation de tels complexes où les syndécans pourraient localement se replier, notamment les ectodomaines des syndécans-2 et -4 dont la quasi-totalité de la séquence a été prédite comme MoRFs.

Peu d'études d'interactions ont été réalisées avec l'ectodomaine des syndécans dépourvus de chaînes de GAGs, il serait intéressant d'identifier des interactions supplémentaires des ectodomains des syndécans avec les protéines du matrisome cœur ou celles qui sont affiliées au matrisome en utilisant une méthode de criblage à haut débit. Il a également été décrit que le syndécan-4 favorisait la réticulation des collagènes par la lysyl oxydase (*Herum et al., 2015*). Il serait ainsi intéressant de rechercher l'interaction entre le syndécan-4 et la lysyl oxydase par interférométrie de biocouche.

La séquence 210-240 de l'ectodomaine du syndécan-1 peut former un complexe ternaire soit avec l'intégrine  $\alpha 3\beta 1$  et le récepteur du facteur de croissance épidermique 2 (HER2), soit avec l'intégrine  $\alpha 4\beta 1$  et le récepteur du facteur de croissance de l'endothélium vasculaire 2 (VEGFR-2). La structure adoptée par cette séquence n'est pas connue mais il est possible qu'elle se replie de manière différente selon le complexe dans lequel elle est impliquée, et qu'elle soit une séquence caméléon.

Ces séquences sont généralement enrichies en résidus hydrophobes ou aromatiques comme les résidus de valine, d'isoleucine, de tyrosine et de phénylalanine ainsi qu'en résidus de glycine (*Habchi et al., 2014, Bahramali et al., 2016*) ce qui est globalement le cas pour la séquence 210-240 d'ED-1 puisque les résidus de valine sont les plus abondants (13 %), que 2 des 3 résidus de phénylalanine de son ectodomaine y sont présents et qu'elle est enrichie en résidus d'alanine (10 %).

La cocristallisation de ce fragment en présence des domaines repliés du récepteur et de l'intégrine avec lesquels il interagit serait un moyen de déterminer s'il se replie au contact de ces partenaires d'autant plus que 19 des 31 résidus de cette séquence sont prédits comme MoRF d'après les prédictions réalisées avec MoRFchibi SYSTEM (*Malhis et al., 2016*). De manière similaire, 9 des 28 résidus de la séquence 93 – 120 responsable de l'interaction de l'ectodomaine du syndécan-1 avec l'intégrine  $\alpha v\beta 3$  et IGF1R sont prédits comme MoRF. Les séquences 123 – 140 et 87 – 131 présentes respectivement dans l'ectodomaine du syndécan-2 et -4 sont également prédites comme MoRFs et sont engagés dans la formation de complexes (*voir Figure 122*) (*De Rossi et al., 2014, Wang et al., 2015*). La séquence <sup>87</sup>NHIP<sup>90</sup> du syndécan-4, responsable de l'adhésion cellulaire, se situe dans une région dont la prédiction de désordre est ambiguë et est proche du seuil désordre / ordre, tout comme la séquence <sup>95</sup>NKIP<sup>98</sup> retrouvée dans l'ectodomaine du syndécan-2. De plus, ces deux séquences sont prédites comme MoRF ce qui suggère qu'elles pourraient se replier lorsqu'elles sont en complexe.

Le faible nombre de protéines décrites interagissant directement avec la partie protéique des ectodomains des syndécans peut vraisemblablement être expliqué par le fait que plusieurs de leurs fonctions biologiques en tant que corécepteurs sont dues à leurs chaînes de GAGs. Nos expériences ont cependant permis d'identifier deux nouveaux partenaires de la partie protéique de l'ectodomaine des quatre syndécans, le facteur de croissance fibroblastique 2 et l'anastelline, un fragment bioactif de la fibronectine. La formation de ces complexes présente le point commun d'être extrêmement stable d'après les expériences d'interférométrie de biocouche.

L'interaction du FGF-2 avec les chaînes d'héparane sulfate des syndécans a été décrite et elle est nécessaire pour activer le récepteur FGFR-1 (*Wu et al., 2001, Afratis et al., 2017*). L'interaction directe de FGF-2 avec l'ectodomaine des syndécans et plus particulièrement avec ED-1 et ED-3 pourrait concentrer le FGF-2 même en l'absence des chaînes de GAGs leur permettant de jouer le rôle d'effecteur, qui est une fonction nécessitant une interaction permanente entre une protéine désordonnée et son partenaire (*van der Lee et al., 2014*).

Il a également été décrit que le syndécan-4 était responsable de l'internalisation du FGF-2 et de FGFR-1 dans des cellules endothéliales de rat (*Elfenbein et al., 2012*) ce qui nécessite vraisemblablement la formation d'un complexe stable. Dans ces conditions, le syndécan-4 pourrait ainsi jouer le rôle d'effecteur et de piègeur pour internaliser FGF-2 et FGFR-1, bien que le rôle de piègeur caractérise davantage l'interaction stable entre une protéine désordonnée et un petit ligand (*van der Lee et al., 2014*).

Dans nos expériences d'interférométrie de biocouche, l'anastelline forme des complexes très stables avec les chaînes d'héparane sulfate, de chondroïtine sulfate et la partie protéique des syndécans. Les simulations par dynamique moléculaire effectuées sur les complexes formés par l'anastelline et des tétrasaccharides d'héparane sulfate ou d'héparine ont confirmé que les complexes étaient très stables puisqu'aucune dissociation n'a été observée. Les ectodomains des syndécans et particulièrement ED-1, -3 et -4, pourraient également jouer le rôle d'effecteur ou de piègeur en capturant l'anastelline présente dans l'espace extracellulaire.

*Mercurius & Morla* (2003) ont observé une diminution de l'adhésion cellulaire de 25 à 40 % lorsque les cellules musculaires lisses d'aorte de rat étaient traités par de l'héparinase qui digère les chaînes d'héparane sulfate, mais pas lorsque les cellules étaient traités par de la chondroïtinase, qui digère les chaînes de chondroïtine sulfate. La partie protéique des ectodomains des syndécans pourrait contribuer à l'adhésion cellulaire ce qui expliquerait l'adhésion cellulaire résiduelle.

Le rôle des chaînes de chondroïtine sulfate reste cependant à déterminer. Dans nos expériences préliminaires, l'anastelline et l'ectodomaine du syndécan-1 s'agrègent lorsqu'ils sont en complexe. L'agrégation d'autres protéines par l'anastelline comme la fibronectine ou le fibrinogène a été décrite menant à la formation de superfibronectine, qui ressemble aux fibrilles de fibronectine matricielle se formant de manière physiologique et qui possèdent des propriétés d'adhésion plus importantes que celles de la fibronectine (*Morla et al., 1994*), et de superfibrinogène qui possède des activités anti-tumorales et anti-angiogéniques (*Ohashi et al., 2009, Yi & Ruoslahti, 2001*). Nos premières observations sur la formation d'agrégats d'ectodomaine du syndécan-1 en présence d'anastelline et les mesures effectuées par interférométrie de biocouche suggèrent que l'anastelline pourrait également polymériser les ectodomains des syndécans dont l'existence physiologique reste à démontrer.

Les expériences d'interférométrie de biocouche ont également montré que l'interaction entre l'ectodomaine du syndécan-1 et l'anastelline était favorisée à pH acide, probablement par la formation d'une nouvelle région basique à la surface de l'anastelline causée par la protonation de ses résidus d'histidine.

ED-1 pourrait capturer l'anastelline dans des environnements légèrement acides et la relarguer lorsque le pH est légèrement plus alcalin. Des interactions dépendantes du pH ont déjà été observées. C'est le cas de l'isomérase des ponts disulfures ERp44 qui est impliquée dans la voie de sécrétion des protéines. La protéine ERp44 joue le rôle de navette en interagissant avec des protéines comme l'adiponectine au niveau du cis-Golgi, où le pH est légèrement acide, et en les libérant dans le réticulum endoplasmique, où le pH est quasiment neutre. Cette fonction est réalisée par des changements conformationnels dépendants du pH impliquant la protonation / déprotonation de cinq résidus d'histidine (*Watanabe et al., 2017*).

L'hypothèse de la formation de la région basique sur l'anastelline pourra être vérifiée notamment par la génération de mutants sur les résidus d'histidine concernés. Il serait également intéressant de déterminer si les ectodomains des syndécans seuls ou en complexe avec l'anastelline forment des fibrilles  $\beta$ -amyloïde, notamment ED-3 qui est presque exclusivement exprimé dans les tissus neuronaux (*Afratis et al., 2017*). Le cas échéant, il serait également intéressant de comparer les cinétiques de fibrillogenèse des ectodomains des syndécans -1 et -3 d'une part et des ectodomains des syndécans -2 et -4 d'autre part. En effet, la présence d'une proportion plus importante de brins  $\beta$  dans ces deux derniers pourrait favoriser la formation de fibrilles par rapport aux ectodomains des syndécans -1 et -3.

### ***Dimérisation des ectodomaines des syndécans***

Nos expériences ont permis d'observer pour la première fois la formation de dimères covalents formés par l'isoforme 2 du syndécan-4 qui est un variant d'épissage présentant la particularité de posséder un unique résidu de cystéine. L'analyse du dimère par SEC-SAXS, SEC-DLS et mobilité ionique a montré que les monomères se disposaient de façon étendue et que le dimère était très flexible. L'ectodomaine du syndécan-3 dimérise également (*Asundi & Carey, 1995*) et le dimère a été détecté en faible quantité dans nos expériences de spectrométrie de masse. Les observations préliminaires effectuées à partir des données de SEC-SAXS suggèrent que les deux monomères d'ED-3 s'associent latéralement provoquant leur repliement partiel, ce qui est visible par la forme de la courbe dans la représentation normalisée de Kratky et par la diminution des intensités de diffusion en  $s^{-3}$  plutôt qu'en  $s^{-2}$  (**Annexe 4**). Ce repliement partiel est en accord avec un modèle du dimère précédemment établi et qui implique la formation de deux régions globulaires séparées par une région très étendue (*Kirsch et al., 2002*). Dans leur modèle du syndécan-3 entier sous forme dimérique, la séquence ERKE nécessaire à sa dimérisation se trouverait entre le domaine transmembranaire et le domaine globulaire G2 (*voir Figure 21*). L'ectodomaine du syndécan-1 posséderait également la possibilité de dimériser mais le dimère n'a été observé qu'en très faible abondance dans nos expériences ce qui n'a pas permis de le caractériser. Le mécanisme de dimérisation des ectodomaines des syndécans -1 et -3 n'implique pas la formation de ponts disulfures contrairement à celui de l'isoforme-2 du syndécan-4 (ED-4).

**REFERENCES**  
**BIBLIOGRAPHIQUES**

- Abay, A., Simionato, G., Chachanidze, R., Bogdanova, A., Hertz, L., Bianchi, P., van den Akker, E., von Lindern, M., Leonetti, M., Minetti, G., et al. (2019). Glutaraldehyde - A Subtle Tool in the Investigation of Healthy and Pathologic Red Blood Cells. *Front Physiol* *10*, 514.
- Adams, J.C., Kureishy, N., and Taylor, A.L. (2001). A role for syndecan-1 in coupling fascin spike formation by thrombospondin-1. *J. Cell Biol.* *152*, 1169–1182.
- Adamski, W., Salvi, N., Maurin, D., Magnat, J., Milles, S., Jensen, M.R., Abyzov, A., Moreau, C.J., and Blackledge, M. (2019). A Unified Description of Intrinsically Disordered Protein Dynamics under Physiological Conditions Using NMR Spectroscopy. *J. Am. Chem. Soc.* *141*, 17817–17829.
- Afratis, N.A., Nikitovic, D., Multhaupt, H.A.B., Theocharis, A.D., Couchman, J.R., and Karamanos, N.K. (2017). Syndecans - key regulators of cell signaling and biological functions. *FEBS J.* *284*, 27–41.
- Aguzzi, A., and Altmeyer, M. (2016). Phase Separation: Linking Cellular Compartmentalization to Disease. *Trends Cell Biol.* *26*, 547–558.
- Aleksis, R., Oleskovs, F., Jaudzems, K., Pahnke, J., and Biverstål, H. (2017). Structural studies of amyloid- $\beta$  peptides: Unlocking the mechanism of aggregation and the associated toxicity. *Biochimie* *140*, 176–192.
- Alphonsus, C.S., and Rodseth, R.N. (2014). The endothelial glycocalyx: a review of the vascular barrier. *Anaesthesia* *69*, 777–784.
- Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. (2017). Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat Commun* *8*, 275.
- Amenta, P.S., Scivoletti, N.A., Newman, M.D., Sciancalepore, J.P., Li, D., and Myers, J.C. (2005). Proteoglycan-collagen XV in human tissues is seen linking banded collagen fibers subjacent to the basement membrane. *J. Histochem. Cytochem.* *53*, 165–176.
- Aminoff, D., Anderson, J., Dabich, L., and Gathmann, W.D. (1980). Sialic acid content of erythrocytes in normal individuals and patients with certain hematologic disorders. *Am. J. Hematol.* *9*, 381–389.
- Anfinsen, C.B., Haber, E., Sela, M., and White, F.H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* *47*, 1309–1314.
- Aoki, T. (2017). A Comprehensive Review of Our Current Understanding of Red Blood Cell (RBC) Glycoproteins. *Membranes (Basel)* *7*.
- Appunni, S., Anand, V., Khandelwal, M., Gupta, N., Rubens, M., and Sharma, A. (2019). Small Leucine Rich Proteoglycans (decorin, biglycan and lumican) in cancer. *Clin. Chim. Acta* *491*, 1–7.
- Ashikari-Hada, S., Habuchi, H., Sugaya, N., Kobayashi, T., and Kimata, K. (2009). Specific inhibition of FGF-2 signaling with 2-O-sulfated octasaccharides of heparan sulfate. *Glycobiology* *19*, 644–654.
- Asundi, V.K., and Carey, D.J. (1995). Self-association of N-syndecan (syndecan-3) core protein is mediated by a novel structural motif in the transmembrane domain and ectodomain flanking region. *J. Biol. Chem.* *270*, 26404–26410.
- Aznauryan, M., Delgado, L., Soranno, A., Nettels, D., Huang, J.-R., Labhardt, A.M., Grzesiek, S., and Schuler, B. (2016). Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. U.S.A.* *113*, E5389–E5398.



- Bah, A., and Forman-Kay, J.D. (2016). Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* *291*, 6696–6705.
- Bahramali, G., Goliaei, B., Minuchehr, Z., and Salari, A. (2016a). Chameleon sequences in neurodegenerative diseases. *Biochem. Biophys. Res. Commun.* *472*, 209–216.
- Ballas, S.K., Lieff, S., Benjamin, L.J., Dampier, C.D., Heeney, M.M., Hoppe, C., Johnson, C.S., Rogers, Z.R., Smith-Whitley, K., Wang, W.C., et al. (2010). Definitions of the phenotypic manifestations of sickle cell disease. *Am. J. Hematol.* *85*, 6–13.
- Banerjee, S., and Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* *2012*, 282574.
- Barbouri, D., Afratis, N., Gialeli, C., Vynios, D.H., Theocharis, A.D., and Karamanos, N.K. (2014). Syndecans as modulators and potential pharmacological targets in cancer progression. *Front Oncol* *4*, 4.
- Barrett, P.J., Song, Y., Van Horn, W.D., Hustedt, E.J., Schafer, J.M., Hadziselimovic, A., Beel, A.J., and Sanders, C.R. (2012). The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* *336*, 1168–1171.
- Bartlett, A.H., and Park, P.W. (2010). Proteoglycans in host-pathogen interactions: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med* *12*, e5.
- Barylyuk, K., Balabin, R.M., Grünstein, D., Kikkeri, R., Frankevich, V., Seeberger, P.H., and Zenobi, R. (2011). What happens to hydrophobic interactions during transfer from the solution to the gas phase? The case of electrospray-based soft ionization methods. *J. Am. Soc. Mass Spectrom.* *22*, 1167–1177.
- Basile, W., Salvatore, M., Bassot, C., and Elofsson, A. (2019). Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput. Biol.* *15*, e1007186.
- Baskurt, O.K., Boynard, M., Cokelet, G.C., Connes, P., Cooke, B.M., Forconi, S., Liao, F., Hardeman, M.R., Jung, F., Meiselman, H.J., et al. (2009). New guidelines for hemorheological laboratory techniques. *Clin. Hemorheol. Microcirc.* *42*, 75–97.
- Beauvais, D.M., and Rapraeger, A.C. (2004). Syndecans in tumor cell adhesion and signaling. *Reprod. Biol. Endocrinol.* *2*, 3.
- Beauvais, D.M., and Rapraeger, A.C. (2010). Syndecan-1 couples the insulin-like growth factor-1 receptor to inside-out integrin activation. *J. Cell. Sci.* *123*, 3796–3807.
- van Beers, E.J., Nieuwdorp, M., Duits, A.J., Evers, L.M., Schnog, J.-J.B., Biemond, B.J., and CURAMA Study Group (2008). Sickle cell patients are characterized by a reduced glycocalyx volume. *Haematologica* *93*, 307–308.
- Bernadó, P., and Svergun, D.I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst* *8*, 151–167.
- Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* *129*, 5656–5664.
- Bertrand, J., and Bollmann, M. (2019). Soluble syndecans: biomarkers for diseases and therapeutic options. *Br. J. Pharmacol.* *176*, 67–81.

- Beveridge, R., Covill, S., Pacholarz, K.J., Kalapothakis, J.M.D., MacPhee, C.E., and Barran, P.E. (2014). A mass-spectrometry-based framework to define the extent of disorder in proteins. *Anal. Chem.* *86*, 10979–10991.
- Beveridge, R., Phillips, A.S., Denbigh, L., Saleem, H.M., MacPhee, C.E., and Barran, P.E. (2015). Relating gas phase to solution conformations: Lessons from disordered proteins. *Proteomics* *15*, 2872–2883.
- Bilandzic, M., and Stenvers, K.L. (2011). Betaglycan: a multifunctional accessory. *Mol. Cell. Endocrinol.* *339*, 180–189.
- Bishop, J.R., Schuksz, M., and Esko, J.D. (2007). Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* *446*, 1030–1037.
- Blundell, C.D., Mahoney, D.J., Cordell, M.R., Almond, A., Kahmann, J.D., Perczel, A., Taylor, J.D., Campbell, I.D., and Day, A.J. (2007). Determining the molecular basis for the pH-dependent interaction between the link module of human TSG-6 and hyaluronan. *J. Biol. Chem.* *282*, 12976–12988.
- Bonomi, M., Heller, G.T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* *42*, 106–116.
- Boon, P.L.S., Saw, W.G., Lim, X.X., Raghuvamsi, P.V., Huber, R.G., Marzinek, J.K., Holdbrook, D.A., Anand, G.S., Grüber, G., and Bond, P.J. (2018). Partial Intrinsic Disorder Governs the Dengue Capsid Protein Conformational Ensemble. *ACS Chem. Biol.* *13*, 1621–1630.
- Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature* *555*, 61–66.
- Bor-Kucukatay, M., Wenby, R.B., Meiselman, H.J., and Baskurt, O.K. (2003). Effects of nitric oxide on red blood cell deformability. *Am. J. Physiol. Heart Circ. Physiol.* *284*, H1577–1584.
- Bratosin, D., Mazurier, J., Debray, H., Lecocq, M., Boilly, B., Alonso, C., Moisei, M., Motas, C., and Montreuil, J. (1995). Flow cytofluorimetric analysis of young and senescent human erythrocytes probed with lectins. Evidence that sialic acids control their life span. *Glycoconj. J.* *12*, 258–267.
- Brewer, A.K., and Striegel, A.M. (2011). Characterizing the size, shape, and compactness of a polydisperse prolate ellipsoidal particle via quadruple-detector hydrodynamic chromatography. *Analyst* *136*, 515–519.
- Briggs, D.C., and Hohenester, E. (2018). Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure* *26*, 801–809.e3.
- Briknarová, K., Akerman, M.E., Hoyt, D.W., Ruoslahti, E., and Ely, K.R. (2003). Anastellin, an FN3 fragment with fibronectin polymerization activity, resembles amyloid fibril precursors. *J. Mol. Biol.* *332*, 205–215.
- Brookes, E., Vachette, P., Rocco, M., and Pérez, J. (2016). US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *J Appl Crystallogr* *49*, 1827–1841.
- Brutscher, B., Felli, I.C., Gil-Caballero, S., Hošek, T., Kümmerle, R., Piai, A., Pierattelli, R., and Solyom, Z. (2015). NMR Methods for the Study of Intrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv. Exp. Med. Biol.* *870*, 49–122.

- Bryk, A.H., and Wiśniewski, J.R. (2017). Quantitative Analysis of Human Red Blood Cell Proteome. *J. Proteome Res.* *16*, 2752–2761.
- Buchan, D.W.A., and Jones, D.T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* *47*, W402–W407.
- Bulai, T., Bratosin, D., Pons, A., Montreuil, J., and Zanetta, J.P. (2003). Diversity of the human erythrocyte membrane sialic acids in relation with blood groups. *FEBS Lett.* *534*, 185–189.
- Bülow, H.E., and Hobert, O. (2006). The molecular diversity of glycosaminoglycans shapes animal development. *Annu. Rev. Cell Dev. Biol.* *22*, 375–407.
- Burton, N.M., and Bruce, L.J. (2011). Modelling the structure of the red cell membrane. *Biochem. Cell Biol.* *89*, 200–215.
- Cagno, V., Tseligka, E.D., Jones, S.T., and Tapparel, C. (2019). Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? *Viruses* *11*.
- Cardin, A.D., and Weintraub, H.J. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* *9*, 21–32.
- Carey, D.J., Conner, K., Asundi, V.K., O'Mahony, D.J., Stahl, R.C., Showalter, L., Cizmeci-Smith, G., Hartman, J., and Rothblum, L.I. (1997). cDNA cloning, genomic organization, and in vivo expression of rat N-syndecan. *J. Biol. Chem.* *272*, 2873–2879.
- Carulli, S., Beck, K., Dayan, G., Boulesteix, S., Lortat-Jacob, H., and Rousselle, P. (2012). Cell surface proteoglycans syndecan-1 and -4 bind overlapping but distinct sites in laminin  $\alpha 3$  LG45 protein domain. *J. Biol. Chem.* *287*, 12204–12216.
- Carvalho, F.A., de Oliveira, S., Freitas, T., Gonçalves, S., and Santos, N.C. (2011). Variations on fibrinogen-erythrocyte interactions during cell aging. *PLoS ONE* *6*, e18167.
- Carvalho, F.A., Guedes, A.F., Duval, C., Macrae, F.L., Swithenbank, L., Farrell, D.H., Ariëns, R.A., and Santos, N.C. (2018). The 95RGD97 sequence on the A $\alpha$  chain of fibrinogen is essential for binding to its erythrocyte receptor. *Int J Nanomedicine* *13*, 1985–1992.
- Castillo, G.M., Lukito, W., Wight, T.N., and Snow, A.D. (1999). The sulfate moieties of glycosaminoglycans are critical for the enhancement of beta-amyloid protein fibril formation. *J. Neurochem.* *72*, 1681–1687.
- Caterson, B., and Melrose, J. (2018). Keratan sulfate, a complex glycosaminoglycan with unique functional capability. *Glycobiology* *28*, 182–206.
- Chae, I.K., Lee, W. (2008). Recombinant Expression and Purification of Cytoplasmic Domain of Syndecan-2 Proteoglycan. *Bulletin of the Korean Chemical Society* *29*, 2449–2452.
- Chakravarti, R., and Adams, J.C. (2006). Comparative genomics of the syndecans defines an ancestral genomic context associated with matrilins in vertebrates. *BMC Genomics* *7*, 83.
- Chakravarti, R., Sapountzi, V., and Adams, J.C. (2005). Functional role of syndecan-1 cytoplasmic V region in lamellipodial spreading, actin bundling, and cell migration. *Mol. Biol. Cell* *16*, 3678–3691.
- Chan-Yao-Chong, M., Deville, C., Pinet, L., van Heijenoort, C., Durand, D., and Ha-Duong, T. (2019a). Structural Characterization of N-WASP Domain V Using MD Simulations with NMR and SAXS Data. *Biophys. J.* *116*, 1216–1227.

- Chan-Yao-Chong, M., Durand, D., and Ha-Duong, T. (2019b). Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles. *J Chem Inf Model* *59*, 1743–1758.
- Chaudhary, B., Khaled, Y.S., Ammori, B.J., and Elkord, E. (2014). Neuropilin 1: function and therapeutic potential in cancer. *Cancer Immunol. Immunother.* *63*, 81–99.
- Chemes, L.B., Alonso, L.G., Noval, M.G., and de Prat-Gay, G. (2012). Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol. Biol.* *895*, 387–404.
- Chen, K., and Williams, K.J. (2013). Molecular mediators for raft-dependent endocytosis of syndecan-1, a highly conserved, multifunctional receptor. *J. Biol. Chem.* *288*, 13988–13999.
- Chen, C., Zhao, S., Karnad, A., and Freeman, J.W. (2018). The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol* *11*, 64.
- Chen, J., Wang, F., He, C., and Luo, S.-Z. (2020a). Multiple Dimerizing Motifs Modulate the Dimerization of the Syndecan Transmembrane Domains. *BioRxiv* 2019.12.31.891929.
- Chen, J., Liu, X., and Chen, J. (2020b). Targeting Intrinsically Disordered Proteins through Dynamic Interactions. *Biomolecules* *10*.
- Chen, L., Couchman, J.R., Smith, J., and Woods, A. (2002). Molecular characterization of chicken syndecan-2 proteoglycan. *Biochem. J.* *366*, 481–490.
- Cheng, B., Montmasson, M., Terradot, L., and Rousselle, P. (2016). Syndecans as Cell Surface Receptors in Cancer Biology. A Focus on their Interaction with PDZ Domain Proteins. *Front Pharmacol* *7*, 10.
- Chiodelli, P., Urbinati, C., Mitola, S., Tanghetti, E., and Rusnati, M. (2012). Sialic acid associated with  $\alpha\text{v}\beta\text{3}$  integrin mediates HIV-1 Tat protein interaction and endothelial cell proangiogenic activation. *J. Biol. Chem.* *287*, 20456–20466.
- Cho, C., Kelsh-Lasher, R., Ambesi, A., and McKeown-Longo, P.J. (2015). Cryptic activity within the Type III1 domain of fibronectin regulates tissue inflammation and angiogenesis. *Curr Top Pept Protein Res* *16*, 37–47.
- Choi, S., Lee, E., Kwon, S., Park, H., Yi, J.Y., Kim, S., Han, I.-O., Yun, Y., and Oh, E.-S. (2005). Transmembrane domain-induced oligomerization is crucial for the functions of syndecan-2 and syndecan-4. *J. Biol. Chem.* *280*, 42573–42579.
- Choi, Y., Kwon, M.-J., Lim, Y., Yun, J.-H., Lee, W., and Oh, E.-S. (2015). Trans-regulation of Syndecan Functions by Hetero-oligomerization. *J. Biol. Chem.* *290*, 16943–16953.
- Choi, Y., Yun, J.-H., Yoo, J., Lee, I., Kim, H., Son, H.-N., Kim, I.-S., Yoon, H.S., Zimmermann, P., Couchman, J.R., et al. (2016). New structural insight of C-terminal region of Syntenin-1, enhancing the molecular dimerization and inhibitory function related on Syndecan-4 signaling. *Sci Rep* *6*, 36818.
- Choy, W.Y., and Forman-Kay, J.D. (2001). Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* *308*, 1011–1032.
- Christianson, H.C., and Belting, M. (2014). Heparan sulfate proteoglycan as a cell-surface endocytosis receptor. *Matrix Biol.* *35*, 51–55.

- Chung, H., Mulhaupt, H.A.B., Oh, E.-S., and Couchman, J.R. (2016). Minireview: Syndecans and their crucial roles during tissue regeneration. *FEBS Lett.* *590*, 2408–2417.
- Clerc, O., Deniaud, M., Vallet, S.D., Naba, A., Rivet, A., Perez, S., Thierry-Mieg, N., and Ricard-Blum, S. (2019). MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res.* *47*, D376–D381.
- Communie, G., Habchi, J., Yabukarski, F., Blocquel, D., Schneider, R., Tarbouriech, N., Papageorgiou, N., Ruigrok, R.W.H., Jamin, M., Jensen, M.R., et al. (2013). Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog.* *9*, e1003631.
- Connes, P., Alexy, T., Detterich, J., Romana, M., Hardy-Dessources, M.-D., and Ballas, S.K. (2016). The role of blood rheology in sickle cell disease. *Blood Rev.* *30*, 111–118.
- Cordeiro, T.N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N., and Bernadó, P. (2017). Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr. Opin. Struct. Biol.* *42*, 15–23.
- Cortese-Krott, M.M., and Kelm, M. (2014). Endothelial nitric oxide synthase in red blood cells: key to a new erythrocrine function? *Redox Biol.* *2*, 251–258.
- Couchman, J.R. (2003). Syndecans: proteoglycan regulators of cell-surface microdomains? *Nat. Rev. Mol. Cell Biol.* *4*, 926–937.
- Couchman, J.R. (2010). Transmembrane signaling proteoglycans. *Annu. Rev. Cell Dev. Biol.* *26*, 89–114.
- Couchman, J.R., and Pataki, C.A. (2012). An introduction to proteoglycans and their localization. *J. Histochem. Cytochem.* *60*, 885–897.
- Couchman, J.R., Gopal, S., Lim, H.C., Nørgaard, S., and Mulhaupt, H.A.B. (2015). Fell-Muir Lecture: Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. *Int J Exp Pathol* *96*, 1–10.
- Cruz-Chu, E.R., Malafeev, A., Pajarskas, T., Pivkin, I.V., and Koumoutsakos, P. (2014). Structure and response to flow of the glycocalyx layer. *Biophys. J.* *106*, 232–243.
- Culik, R.M., Abaskharon, R.M., Pazos, I.M., and Gai, F. (2014). Experimental validation of the role of trifluoroethanol as a nanocrowder. *J Phys Chem B* *118*, 11455–11461.
- Czaplewski, C., Oldziej, S., Liwo, A., and Scheraga, H.A. (2004). Prediction of the structures of proteins with the UNRES force field, including dynamic formation and breaking of disulfide bonds. *Protein Eng. Des. Sel.* *17*, 29–36.
- Czaplewski, C., Karczynska, A., Sieradzan, A.K., and Liwo, A. (2018). UNRES server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics. *Nucleic Acids Res.* *46*, W304–W309.
- Daggett, V., and Fersht, A.R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* *28*, 18–25.
- D’Alessandro, A., Dzieciatkowska, M., Nemkov, T., and Hansen, K.C. (2017). Red blood cell proteomics update: is there more to discover? *Blood Transfus* *15*, 182–187.
- Darling, A.L., and Uversky, V.N. (2018). Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet* *9*, 158.

- De Oliveira, S., Vitorino de Almeida, V., Calado, A., Rosário, H.S., and Saldanha, C. (2012). Integrin-associated protein (CD47) is a putative mediator for soluble fibrinogen interaction with human red blood cells membrane. *Biochim. Biophys. Acta* *1818*, 481–490.
- De Rossi, G., and Whiteford, J.R. (2014). Syndecans in angiogenesis and endothelial cell biology. *Biochem. Soc. Trans.* *42*, 1643–1646.
- 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.
- Deepa, S.S., Yamada, S., Zako, M., Goldberger, O., and Sugahara, K. (2004). Chondroitin sulfate chains on syndecan-1 and syndecan-4 from normal murine mammary gland epithelial cells are structurally and functionally distinct and cooperate with heparan sulfate chains to bind growth factors. A novel function to control binding of midkine, pleiotrophin, and basic fibroblast growth factor. *J. Biol. Chem.* *279*, 37368–37376.
- Denhez, F., Wilcox-Adelman, S.A., Baciú, P.C., Saoncella, S., Lee, S., French, B., Neveu, W., and Goetinck, P.F. (2002). Syndesmos, a syndecan-4 cytoplasmic domain interactor, binds to the focal adhesion adaptor proteins paxillin and Hic-5. *J. Biol. Chem.* *277*, 12270–12274.
- Derksen, P.W.B., Keehnen, R.M.J., Evers, L.M., van Oers, M.H.J., Spaargaren, M., and Pals, S.T. (2002). Cell surface proteoglycan syndecan-1 mediates hepatocyte growth factor binding and promotes Met signaling in multiple myeloma. *Blood* *99*, 1405–1410.
- Dews, I.C., and Mackenzie, K.R. (2007). Transmembrane domains of the syndecan family of growth factor coreceptors display a hierarchy of homotypic and heterotypic interactions. *Proc. Natl. Acad. Sci. U.S.A.* *104*, 20782–20787.
- Dharmadana, D., Reynolds, N.P., Dekiwadia, C., Conn, C.E., and Valéry, C. (2018). Heparin assisted assembly of somatostatin amyloid nanofibrils results in disordered precipitates by hindrance of protofilaments interactions. *Nanoscale* *10*, 18195–18204.
- Dobo, A., and Kaltashov, I.A. (2001). Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS. *Anal. Chem.* *73*, 4763–4773.
- Dobson, C.M. (2003). Protein folding and misfolding. *Nature* *426*, 884–890.
- Donath, E., and Pastushenko, V. (1979). Electrophoretical study of cell surface properties. The influence of the surface coat on the electric potential distribution and on general electrokinetic properties of animal cells. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry* *104*, 543–554.
- Dosztányi, Z. (2018). Prediction of protein disorder based on IUPred. *Protein Sci.* *27*, 331–340.
- Duchesne, L., Ochteau, V., Bearon, R.N., Beckett, A., Prior, I.A., Lounis, B., and Fernig, D.G. (2012). Transport of fibroblast growth factor 2 in the pericellular matrix is controlled by the spatial distribution of its binding sites in heparan sulfate. *PLoS Biol.* *10*, e1001361.
- Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* *19*, 26–59.
- Dunker, A.K., Brown, C.J., and Obradovic, Z. (2002). Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* *62*, 25–49.

- Düx, P., Rubinstenn, G., Vuister, G.W., Boelens, R., Mulder, F.A., Hård, K., Hoff, W.D., Kroon, A.R., Crielaard, W., Hellingwerf, K.J., et al. (1998). Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry* *37*, 12689–12699.
- Ebong, E.E., Macaluso, F.P., Spray, D.C., and Tarbell, J.M. (2011). Imaging the endothelial glycocalyx in vitro by rapid freezing/freeze substitution transmission electron microscopy. *Arterioscler. Thromb. Vasc. Biol.* *31*, 1908–1915.
- Edwards, D.R., Handsley, M.M., and Pennington, C.J. (2008). The ADAM metalloproteinases. *Mol. Aspects Med.* *29*, 258–289.
- Ekman, D., Light, S., Björklund, A.K., and Elofsson, A. (2006). What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol.* *7*, R45.
- Elenius, V., Götte, M., Reizes, O., Elenius, K., and Bernfield, M. (2004). Inhibition by the soluble syndecan-1 ectodomains delays wound repair in mice overexpressing syndecan-1. *J. Biol. Chem.* *279*, 41928–41935.
- Elfenbein, A., and Simons, M. (2013). Syndecan-4 signaling at a glance. *J. Cell. Sci.* *126*, 3799–3804.
- Elfenbein, A., Lanahan, A., Zhou, T.X., Yamasaki, A., Tkachenko, E., Matsuda, M., and Simons, M. (2012). Syndecan 4 regulates FGFR1 signaling in endothelial cells by directing macropinocytosis. *Sci Signal* *5*, ra36.
- Erdős, G., and Dosztányi, Z. (2020). Analyzing Protein Disorder with IUPred2A. *Curr Protoc Bioinformatics* *70*, e99.
- Esko, J.D., and Selleck, S.B. (2002). Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* *71*, 435–471.
- Ester, M., Kriegel, H.-P., and Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *6*.
- Fan, J., Sun, Y., Xia, Y., Tarbell, J.M., and Fu, B.M. (2019). Endothelial surface glycocalyx (ESG) components and ultra-structure revealed by stochastic optical reconstruction microscopy (STORM). *Biorheology* *56*, 77–88.
- Fantini, J., and Barrantes, F.J. (2013). How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front Physiol* *4*, 31.
- Fatoux-Ardore, M., Peysselon, F., Weiss, A., Bastien, P., Pratlong, F., and Ricard-Blum, S. (2014). Large-scale investigation of *Leishmania* interaction networks with host extracellular matrix by surface plasmon resonance imaging. *Infect. Immun.* *82*, 594–606.
- Fears, C.Y., and Woods, A. (2006). The role of syndecans in disease and wound healing. *Matrix Biol.* *25*, 443–456.
- Fecarotta, S., Tarallo, A., Damiano, C., Minopoli, N., and Parenti, G. (2020). Pathogenesis of Mucopolysaccharidoses, an Update. *Int J Mol Sci* *21*.
- Ferrone, S., and Whiteside, T.L. (2020). Targeting CSPG4 for isolation of melanoma cell-derived exosomes from body fluids. *HNO* *68*, 100–105.

- Filmus, J., Capurro, M., and Rast, J. (2008). Glypicans. *Genome Biol.* *9*, 224.
- Fiore, V.F., Ju, L., Chen, Y., Zhu, C., and Barker, T.H. (2014). Dynamic catch of a Thy-1- $\alpha 5\beta 1$ +syndecan-4 trimolecular complex. *Nat Commun* *5*, 4886.
- Fitzgerald, M.L., Wang, Z., Park, P.W., Murphy, G., and Bernfield, M. (2000). Shedding of syndecan-1 and -4 ectodomains is regulated by multiple signaling pathways and mediated by a TIMP-3-sensitive metalloproteinase. *J. Cell Biol.* *148*, 811–824.
- Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., Goedert, M., and Scheres, S.H.W. (2017). Cryo-EM structures of Tau filaments from Alzheimer's disease brain. *Nature* *547*, 185–190.
- Fonin, A.V., Darling, A.L., Kuznetsova, I.M., Turoverov, K.K., and Uversky, V.N. (2018). Intrinsically disordered proteins in crowded milieu: when chaos prevails within the cellular gumbo. *Cell. Mol. Life Sci.* *75*, 3907–3929.
- Franke, D., Jeffries, C.M., and Svergun, D.I. (2015). Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* *12*, 419–422.
- Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* *50*, 1212–1225.
- Frański, R., Gierczyk, B., Schroeder, G., Frańska, M., and Wyrwas, B. (2008). Do hydrophobic interactions exist in the gas phase? *Rapid Commun. Mass Spectrom.* *22*, 1339–1343.
- Frantz, C., Stewart, K.M., and Weaver, V.M. (2010). The extracellular matrix at a glance. *J. Cell. Sci.* *123*, 4195–4200.
- Fredenburgh, J.C., Leslie, B.A., Stafford, A.R., Lim, T., Chan, H.H., and Weitz, J.I. (2013). Zn<sup>2+</sup> mediates high affinity binding of heparin to the  $\alpha C$  domain of fibrinogen. *J. Biol. Chem.* *288*, 29394–29402.
- Fukuchi, S., Amemiya, T., Sakamoto, S., Nobe, Y., Hosoda, K., Kado, Y., Murakami, S.D., Koike, R., Hiroaki, H., and Ota, M. (2014). IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners. *Nucleic Acids Res.* *42*, D320–325.
- Fuxreiter, M. (2018). Fuzziness in Protein Interactions-A Historical Perspective. *J. Mol. Biol.* *430*, 2278–2287.
- Fuxreiter, M., and Tompa, P. (2012). Fuzzy complexes: a more stochastic view of protein function. *Adv. Exp. Med. Biol.* *725*, 1–14.
- Gao, L., and Lipowsky, H.H. (2010). Composition of the endothelial glycocalyx and its relation to its thickness and diffusion of small solutes. *Microvasc. Res.* *80*, 394–401.
- Garantziotis, S., and Savani, R.C. (2019). Hyaluronan biology: A complex balancing act of structure, function, location and context. *Matrix Biol.* *78–79*, 1–10.
- Gee, E.P.S., Yüksel, D., Stultz, C.M., and Ingber, D.E. (2013). SLLISWD sequence in the 10FNIII domain initiates fibronectin fibrillogenesis. *J. Biol. Chem.* *288*, 21329–21340.



- van Gemst, J.J., Loeven, M.A., de Graaf, M.J.J., Berden, J.H.M., Rabelink, T.J., Smit, C.H., and van der Vlag, J. (2016). RNA Contaminates Glycosaminoglycans Extracted from Cells and Tissues. *PLoS ONE* *11*, e0167336.
- Gibbs, E.B., Cook, E.C., and Showalter, S.A. (2017). Application of NMR to studies of intrinsically disordered proteins. *Arch. Biochem. Biophys.* *628*, 57–70.
- Gill, S.E., and Parks, W.C. (2008). Metalloproteinases and their inhibitors: regulators of wound healing. *Int. J. Biochem. Cell Biol.* *40*, 1334–1347.
- Gondelaud, F., and Ricard-Blum, S. (2019). Structures and interactions of syndecans. *FEBS J.* *286*, 2994–3007.
- Gopal, S., Mülthaupt, H.A.B., Pocock, R., and Couchman, J.R. (2017). Cell-extracellular matrix and cell-cell adhesion are linked by syndecan-4. *Matrix Biol.* *60–61*, 57–69.
- Gould, S.E., Upholt, W.B., and Kosher, R.A. (1992). Syndecan 3: a member of the syndecan family of membrane-intercalated proteoglycans that is expressed in high amounts at the onset of chicken limb cartilage differentiation. *Proc. Natl. Acad. Sci. U.S.A.* *89*, 3271–3275.
- Gouverneur, M., Spaan, J.A.E., Pannekoek, H., Fontijn, R.D., and Vink, H. (2006). Fluid shear stress stimulates incorporation of hyaluronan into endothelial cell glycocalyx. *Am. J. Physiol. Heart Circ. Physiol.* *290*, H458-452.
- Granés, F., Berndt, C., Roy, C., Mangeat, P., Reina, M., and Vilaró, S. (2003). Identification of a novel Ezrin-binding site in syndecan-2 cytoplasmic domain. *FEBS Lett.* *547*, 212–216.
- Greene, D.K., Tumova, S., Couchman, J.R., and Woods, A. (2003). Syndecan-4 associates with alpha-actinin. *J. Biol. Chem.* *278*, 7617–7623.
- Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* *1*, 2876–2890.
- Grembecka, J., Cierpicki, T., Devedjiev, Y., Derewenda, U., Kang, B.S., Bushweller, J.H., and Derewenda, Z.S. (2006). The binding of the PDZ tandem of syntenin to target proteins. *Biochemistry* *45*, 3674–3683.
- Guedes, A.F., Moreira, C., Nogueira, J.B., Santos, N.C., and Carvalho, F.A. (2019). Fibrinogen-erythrocyte binding and hemorheology measurements in the assessment of essential arterial hypertension patients. *Nanoscale* *11*, 2757–2766.
- Gui, L., Wojciechowski, K., Gildner, C.D., Nedelkovska, H., and Hocking, D.C. (2006). Identification of the heparin-binding determinants within fibronectin repeat III1: role in cell spreading and growth. *J. Biol. Chem.* *281*, 34816–34825.
- Guo, M., Zhang, H., Zheng, J., and Liu, Y. (2020). Glypican-3: A New Target for Diagnosis and Treatment of Hepatocellular Carcinoma. *J Cancer* *11*, 2008–2021.
- Guseva, S., Milles, S., Jensen, M.R., Salvi, N., Kleman, J.-P., Maurin, D., Ruigrok, R.W.H., and Blackledge, M. (2020). Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci Adv* *6*, eaaz7095.
- Haapasalo, A., and Kovacs, D.M. (2011). The many substrates of presenilin/ $\gamma$ -secretase. *J. Alzheimers Dis.* *25*, 3–28.

- Habchi, J., Tompa, P., Longhi, S., and Uversky, V.N. (2014). Introducing protein intrinsic disorder. *Chem. Rev.* *114*, 6561–6588.
- Habuchi, H., Habuchi, O., and Kimata, K. (2004). Sulfation pattern in glycosaminoglycan: does it have a code? *Glycoconj. J.* *21*, 47–52.
- Hamdi, K., Salladini, E., O'Brien, D.P., Brier, S., Chenal, A., Yacoubi, I., and Longhi, S. (2017). Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. *Sci Rep* *7*, 15544.
- Han, J.Y., Choi, T.S., and Kim, H.I. (2018). Molecular Role of Ca<sup>2+</sup> and Hard Divalent Metal Cations on Accelerated Fibrillation and Interfibrillar Aggregation of  $\alpha$ -Synuclein. *Sci Rep* *8*, 1895.
- Hascall, V., and Esko, J.D. (2015). Hyaluronan. In *Essentials of Glycobiology*, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, et al., eds. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press), p.
- Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLOS Computational Biology* *2*, e100.
- Herndon, M.E., Stipp, C.S., and Lander, A.D. (1999). Interactions of neural glycosaminoglycans and proteoglycans with protein ligands: assessment of selectivity, heterogeneity and the participation of core proteins in binding. *Glycobiology* *9*, 143–155.
- Herum, K.M., Lunde, I.G., Skrbic, B., Louch, W.E., Hasic, A., Boye, S., Unger, A., Brorson, S.-H., Sjaastad, I., Tønnessen, T., et al. (2015). Syndecan-4 is a key determinant of collagen cross-linking and passive myocardial stiffness in the pressure-overloaded heart. *Cardiovasc. Res.* *106*, 217–226.
- Howell, M.D., and Gottschall, P.E. (2012). Lectican proteoglycans, their cleaving metalloproteinases, and plasticity in the central nervous system extracellular microenvironment. *Neuroscience* *217*, 6–18.
- Huang, Q., Chang, J., Cheung, M.K., Nong, W., Li, L., Lee, M.-T., and Kwan, H.S. (2014). Human proteins with target sites of multiple post-translational modification types are more prone to be involved in disease. *J. Proteome Res.* *13*, 2735–2748.
- Hudák, A., Kusz, E., Domonkos, I., Jósvay, K., Kodamullil, A.T., Szilák, L., Hofmann-Apitius, M., and Letoha, T. (2019). Contribution of syndecans to cellular uptake and fibrillation of  $\alpha$ -synuclein and tau. *Sci Rep* *9*, 16543.
- Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* *162*, 425–440.
- Iannuzzi, C., Irace, G., and Sirangelo, I. (2015). The effect of glycosaminoglycans (GAGs) on amyloid aggregation and toxicity. *Molecules* *20*, 2510–2528.
- Iozzo, R.V., and Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol.* *42*, 11–55.
- Ishida, T., and Kinoshita, K. (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* *24*, 1344–1348.

- Ithuralde, R.E., Roitberg, A.E., and Turjanski, A.G. (2016). Structured and Unstructured Binding of an Intrinsically Disordered Protein as Revealed by Atomistic Simulations. *J. Am. Chem. Soc.* *138*, 8742–8751.
- Jakubowska-Solarska, B., and Solski, J. (2000). Sialic acids of young and old red blood cells in healthy subjects. *Med. Sci. Monit.* *6*, 871–874.
- Jan, K.-M., and Chien, S. (1973). Role of Surface Electric Charge in Red Blood Cell Interactions. *J Gen Physiol* *67*, 638–654.
- Jannaway, M., Yang, X., Meegan, J.E., Coleman, D.C., and Yuan, S.Y. (2019). Thrombin-cleaved syndecan-3/-4 ectodomain fragments mediate endothelial barrier dysfunction. *PLoS ONE* *14*, e0214737.
- Jeffries, C.M., and Svergun, D.I. (2015). High-throughput studies of protein shapes and interactions by synchrotron small-angle X-ray scattering. *Methods Mol. Biol.* *1261*, 277–301.
- Jeffries, C.M., Graewert, M.A., Blanchet, C.E., Langley, D.B., Whitten, A.E., and Svergun, D.I. (2016). Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat Protoc* *11*, 2122–2153.
- Jensen, M.R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R.W.H., and Blackledge, M. (2008). Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* *130*, 8055–8061.
- Jensen, M.R., Ruigrok, R.W.H., and Blackledge, M. (2013). Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* *23*, 426–435.
- Jethva, P.N., and Udgaonkar, J.B. (2018). The Osmolyte TMAO Modulates Protein Folding Cooperativity by Altering Global Protein Stability. *Biochemistry* *57*, 5851–5863.
- Jiang, X.Z., Luo, K.H., and Ventikos, Y. (2020). Principal mode of Syndecan-4 mechanotransduction for the endothelial glycocalyx is a scissor-like dimer motion. *Acta Physiol (Oxf)* *228*, e13376.
- Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* *31*, 857–863.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* *79*, 926–935.
- Joshi, P., and Vendruscolo, M. (2015). Druggability of Intrinsically Disordered Proteins. *Adv. Exp. Med. Biol.* *870*, 383–400.
- Jozawa, H., Kabir, M.G., Zako, T., Maeda, M., Chiba, K., and Kuroda, Y. (2016). Amorphous protein aggregation monitored using fluorescence self-quenching. *FEBS Lett.* *590*, 3501–3509.
- Jung, H., Han, M., Jang, B., Park, E., and Oh, E.-S. (2020). The oligomerization mediated by the alanine 397 residue in the transmembrane domain is crucial to syndecan-3 functions. *Cell. Signal.* *69*, 109544.
- Kahane, I., Polliack, A., Rachmilewitz, E.A., Bayer, E.A., and Skutelsky, E. (1978). Distribution of sialic acids on the red blood cell membrane in beta thalassaemia. *Nature* *271*, 674–675.
- Kairys, V., Gilson, M.K., and Luy, B. (2004). Structural model for an AxxxG-mediated dimer of surfactant-associated protein C. *Eur. J. Biochem.* *271*, 2086–2092.

- Kang, B.S., Cooper, D.R., Devedjiev, Y., Derewenda, U., and Derewenda, Z.S. (2003). Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. *Structure* *11*, 845–853.
- Kantor, D.B., Chivatakarn, O., Peer, K.L., Oster, S.F., Inatani, M., Hansen, M.J., Flanagan, J.G., Yamaguchi, Y., Sretavan, D.W., Giger, R.J., et al. (2004). Semaphorin 5A is a bifunctional axon guidance cue regulated by heparan and chondroitin sulfate proteoglycans. *Neuron* *44*, 961–975.
- Karamanos, N.K., Piperigkou, Z., Theocharis, A.D., Watanabe, H., Franchi, M., Baud, S., Brézillon, S., Götte, M., Passi, A., Vigetti, D., et al. (2018). Proteoglycan Chemical Diversity Drives Multifunctional Cell Regulation and Therapeutics. *Chem. Rev.* *118*, 9152–9232.
- Karasneh, G.A., Ali, M., and Shukla, D. (2011). An important role for syndecan-1 in herpes simplex virus type-1 induced cell-to-cell fusion and virus spread. *PLoS ONE* *6*, e25252.
- Karousou, E., Misra, S., Ghatak, S., Dobra, K., Götte, M., Vigetti, D., Passi, A., Karamanos, N.K., and Skandalis, S.S. (2017). Roles and targeting of the HAS/hyaluronan/CD44 molecular system in cancer. *Matrix Biol.* *59*, 3–22.
- Kato, Y., Ozawa, S., Miyamoto, C., Maehata, Y., Suzuki, A., Maeda, T., and Baba, Y. (2013). Acidic extracellular microenvironment and cancer. *Cancer Cell Int.* *13*, 89.
- Katuwawala, A., Peng, Z., Yang, J., and Kurgan, L. (2019). Computational Prediction of MoRFs, Short Disorder-to-order Transitioning Protein Binding Regions. *Comput Struct Biotechnol J* *17*, 454–462.
- Kelly, M.A., Chellgren, B.W., Rucker, A.L., Troutman, J.M., Fried, M.G., Miller, A.F., and Creamer, T.P. (2001). Host-guest study of left-handed polyproline II helix formation. *Biochemistry* *40*, 14376–14383.
- Kerfoot, S.M., McRae, K., Lam, F., McAvoy, E.F., Clark, S., Brain, M., Lalor, P.F., Adams, D.H., and Kubes, P. (2008). A novel mechanism of erythrocyte capture from circulation in humans. *Exp. Hematol.* *36*, 111–118.
- Kessenbrock, K., Plaks, V., and Werb, Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* *141*, 52–67.
- Kikhney, A.G., and Svergun, D.I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* *589*, 2570–2577.
- Kim, H., Yoo, J., Lee, I., Kang, Y.J., Cho, H.-S., and Lee, W. (2015). Crystal structure of syndesmos and its interaction with Syndecan-4 proteoglycan. *Biochem. Biophys. Res. Commun.* *463*, 762–767.
- Kim, J.-M., Lee, K., Kim, M.Y., Shin, H.-I., and Jeong, D. (2018). Suppressive effect of syndecan ectodomains and N-desulfated heparins on osteoclastogenesis via direct binding to macrophage-colony stimulating factor. *Cell Death Dis* *9*, 1119.
- Kinnunen, A., Kinnunen, T., Kaksonen, M., Nolo, R., Panula, P., and Rauvala, H. (1998). N-syndecan and HB-GAM (heparin-binding growth-associated molecule) associate with early axonal tracts in the rat brain. *Eur. J. Neurosci.* *10*, 635–648.
- Kiran, G., Srikanth, C.B., Salimath, P.V., and Nandini, C.D. (2015). Diet-induced hypercholesterolemia imparts structure-function changes to erythrocyte chondroitin sulphate/dermatan sulphate. *J. Biochem.* *158*, 217–224.
- Kirsch, T., Koyama, E., Liu, M., Golub, E.E., and Pacifici, M. (2002). Syndecan-3 is a selective regulator of chondrocyte proliferation. *J. Biol. Chem.* *277*, 42171–42177.

- Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comput Chem* *29*, 622–655.
- Kleiger, G., Grothe, R., Mallick, P., and Eisenberg, D. (2002). GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* *41*, 5990–5997.
- Klein, P., Pawson, T., and Tyers, M. (2003). Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single receptor site. *Curr. Biol.* *13*, 1669–1678.
- Koda, J.E., Rapraeger, A., and Bernfield, M. (1985). Heparan sulfate proteoglycans from mouse mammary epithelial cells. Cell surface proteoglycan as a receptor for interstitial collagens. *J. Biol. Chem.* *260*, 8157–8162.
- Kolset, S.O., and Pejler, G. (2011). Serglycin: a structural and functional chameleon with wide impact on immune cells. *J. Immunol.* *187*, 4927–4933.
- Konarev, P.V., and Svergun, D.I. (2015). A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCrJ* *2*, 352–360.
- Kulkarni, P., and Uversky, V.N. (2019). Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules* *9*.
- Kumar, N., Shukla, S., Kumar, S., Suryawanshi, A., Chaudhry, U., Ramachandran, S., and Maiti, S. (2008). Intrinsically disordered protein from a pathogenic mesophile *Mycobacterium tuberculosis* adopts structured conformation at high temperature. *Proteins* *71*, 1123–1133.
- Kundra, P., and Goswami, S. (2019). Endothelial glycocalyx: Role in body fluid homeostasis and fluid management. *Indian J Anaesth* *63*, 6–14.
- Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A., and Kmiecik, S. (2015). CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* *43*, W419-424.
- Kurcinski, M., Badaczewska-Dawid, A., Kolinski, M., Kolinski, A., and Kmiecik, S. (2020). Flexible docking of peptides to proteins using CABS-dock. *Protein Sci.* *29*, 211–222.
- Kusumi, A., Fujiwara, T.K., Tsunoyama, T.A., Kasai, R.S., Liu, A.-A., Hirose, K.M., Kinoshita, M., Matsumori, N., Komura, N., Ando, H., et al. (2020). Defining raft domains in the plasma membrane. *Traffic* *21*, 106–137.
- 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.
- Kwon, M.-J., Choi, Y., Yun, J.-H., Lee, W., Han, I.-O., and Oh, E.-S. (2015). A unique phenylalanine in the transmembrane domain strengthens homodimerization of the syndecan-2 transmembrane domain and functionally regulates syndecan-2. *J. Biol. Chem.* *290*, 5772–5782.
- Kwon, M.-J., Park, J., Jang, S., Eom, C.-Y., and Oh, E.-S. (2016). The Conserved Phenylalanine in the Transmembrane Domain Enhances Heteromeric Interactions of Syndecans. *J. Biol. Chem.* *291*, 872–881.
- Lamarre, Y., Romana, M., Waltz, X., Lalanne-Mistrih, M.-L., Tressières, B., Divialle-Doumdo, L., Hardy-Dessources, M.-D., Vent-Schmidt, J., Petras, M., Broquere, C., et al. (2012). Hemorheological risk

- factors of acute chest syndrome and painful vaso-occlusive crisis in children with sickle cell disease. *Haematologica* *97*, 1641–1647.
- Lambaerts, K., Wilcox-Adelman, S.A., and Zimmermann, P. (2009). The signalling mechanisms of syndecan heparan sulphate proteoglycans. *Curr Opin Cell Biol* *21*, 662–669.
- Lapoumeroulie, C., Connes, P., El Hoss, S., Hierso, R., Charlot, K., Lemonne, N., Elion, J., Le Van Kim, C., Romana, M., and Hardy-Dessources, M.-D. (2019). New insights into red cell rheology and adhesion in patients with sickle cell anaemia during vaso-occlusive crises. *Br. J. Haematol.* *185*, 991–994.
- Laptenko, O., Tong, D.R., Manfredi, J., and Prives, C. (2016). The Tail That Wags the Dog: How the Disordered C-Terminal Domain Controls the Transcriptional Activities of the p53 Tumor-Suppressor Protein. *Trends Biochem. Sci.* *41*, 1022–1034.
- Leclerc, M., Voilin, E., Gros, G., Cognac, S., de Montpréville, V., Validire, P., Bismuth, G., and Mami-Chouaib, F. (2019). Regulation of antitumour CD8 T-cell immunity and checkpoint blockade immunotherapy by Neuropilin-1. *Nat Commun* *10*, 3345.
- Lee, D., Oh, E.S., Woods, A., Couchman, J.R., and Lee, W. (1998). Solution structure of a syndecan-4 cytoplasmic domain and its interaction with phosphatidylinositol 4,5-bisphosphate. *J. Biol. Chem.* *273*, 13022–13029.
- Lee, K., Shirshin, E., Rovnyagina, N., Yaya, F., Boujja, Z., Priezhev, A., and Wagner, C. (2018). Dextran adsorption onto red blood cells revisited: single cell quantification by laser tweezers combined with microfluidics. *Biomed Opt Express* *9*, 2755–2764.
- Lee, Y., Lee, W., Chang, H.-H., Kim, S.-W., Kim, J., and Bae, J.-S. (2018). Testican-1, as a novel diagnosis of sepsis. *J. Cell. Biochem.* *119*, 4216–4223.
- Lefèvre, M., Felmler, D.J., Parnot, M., Baumert, T.F., and Schuster, C. (2014). Syndecan 4 is involved in mediating HCV entry through interaction with lipoviral particle-associated apolipoprotein E. *PLoS ONE* *9*, e95550.
- Leonova, E.I., and Galzitskaya, O.V. (2015). Cell communication using intrinsically disordered proteins: what can syndecans say? *J. Biomol. Struct. Dyn.* *33*, 1037–1050.
- Letoha, T., Hudák, A., Kusz, E., Pettkó-Szandtner, A., Domonkos, I., Jósavay, K., Hofmann-Apitius, M., and Szilák, L. (2019). Contribution of syndecans to cellular internalization and fibrillation of amyloid- $\beta$ (1-42). *Sci Rep* *9*, 1393.
- Levchenko, A. (2003). Allovalency: a case of molecular entanglement. *Curr. Biol.* *13*, R876-878.
- Li, J.-P., and Kusche-Gullberg, M. (2016). Heparan Sulfate: Biosynthesis, Structure, and Function. *Int Rev Cell Mol Biol* *325*, 215–273.
- Li, J., Santambrogio, C., Brocca, S., Rossetti, G., Carloni, P., and Grandori, R. (2016). Conformational effects in protein electrospray-ionization mass spectrometry. *Mass Spectrom Rev* *35*, 111–122.
- Li, N., Gao, W., Zhang, Y.-F., and Ho, M. (2018). Glypicans as Cancer Therapeutic Targets. *Trends Cancer* *4*, 741–754.
- Li, Q., Ng, H.Q., and Kang, C. (2019). Secondary structure and topology of the transmembrane domain of Syndecan-2 in detergent micelles. *FEBS Lett.* *593*, 554–561.

Li, Y., Zhao, C., Luo, F., Liu, Z., Gui, X., Luo, Z., Zhang, X., Li, D., Liu, C., and Li, X. (2018). Amyloid fibril structure of  $\alpha$ -synuclein determined by cryo-electron microscopy. *Cell Research* *28*, 897–903.

Lieutaud, P., Ferron, F., Uversky, A.V., Kurgan, L., Uversky, V.N., and Longhi, S. (2016). How disordered is my protein and what is its disorder for? A guide through the “dark side” of the protein universe. *Intrinsically Disord Proteins* *4*, e1259708.

Lim, H.C., Multhaupt, H.A.B., and Couchman, J.R. (2015). Cell surface heparan sulfate proteoglycans control adhesion and invasion of breast carcinoma cells. *Mol. Cancer* *14*, 15.

Lin, Y., Currie, S.L., and Rosen, M.K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J. Biol. Chem.* *292*, 19110–19120.

Linss, W., Pilgrim, C., and Feuerstein, H. (1991). [How thick is the glycocalyx of human erythrocytes?]. *Acta Histochem.* *91*, 101–104.

Lipowsky, H.H. (2012). The endothelial glycocalyx as a barrier to leukocyte adhesion and its mediation by extracellular proteases. *Ann Biomed Eng* *40*, 840–848.

Liu, X., and Fuentes, E.J. (2019). Emerging Themes in PDZ Domain Signaling: Structure, Function, and Inhibition. *Int Rev Cell Mol Biol* *343*, 129–218.

Liu, J.-X., Yan, Z.-P., Zhang, Y.-Y., Wu, J., Liu, X.-H., and Zeng, Y. (2016). Hemodynamic shear stress regulates the transcriptional expression of heparan sulfate proteoglycans in human umbilical vein endothelial cell. *Cell. Mol. Biol. (Noisy-Le-Grand)* *62*, 28–34.

Liu, X., Shepherd, T.R., Murray, A.M., Xu, Z., and Fuentes, E.J. (2013). The structure of the Tiam1 PDZ domain/ phospho-syndecan1 complex reveals a ligand conformation that modulates protein dynamics. *Structure* *21*, 342–354.

Liyanage, D.S., Omeka, W.K.M., Tharuka, M.D.N., Jung, S., Lee, S., and Lee, J. (2020). Molecular insights and immune responses of big belly seahorse syndecan-2 (CD362): Involvement of ectodomain in regulating cell survival, proliferation, and wound healing. *Fish Shellfish Immunol.* *98*, 457–465.

Locasale, J.W. (2008). Allovalency revisited: an analysis of multisite phosphorylation and substrate rebinding. *J Chem Phys* *128*, 115106.

Lominadze, D., and Dean, W.L. (2002). Involvement of fibrinogen specific binding in erythrocyte aggregation. *FEBS Lett.* *517*, 41–44.

Luquita, A., Urli, L., Svetaz, M.J., Gennaro, A.M., Giorgetti, M.E., Pistone, G., Volpintesta, R., Palatnik, S., and Rasia, M. (2010). In vitro and ex vivo effect of hyaluronic acid on erythrocyte flow properties. *J. Biomed. Sci.* *17*, 8.

Lyon, M., and Gallagher, J.T. (1991). Purification and partial characterization of the major cell-associated heparan sulphate proteoglycan of rat liver. *Biochem. J.* *273(Pt 2)*, 415–422.

van Maarschalkerweerd, A., Pedersen, M.N., Peterson, H., Nilsson, M., Nguyen, T., Skamris, T., Rand, K., Vetri, V., Langkilde, A.E., and Vestergaard, B. (2015). Formation of covalent di-tyrosine dimers in recombinant  $\alpha$ -synuclein. *Intrinsically Disord Proteins* *3*, e1071302.

Maciej-Hulme, M.L., Skidmore, M.A., and Price, H.P. (2018). The role of heparan sulfate in host macrophage infection by *Leishmania* species. *Biochem. Soc. Trans.* *46*, 789–796.

- Maeda, N. (2015). Proteoglycans and neuronal migration in the cerebral cortex during development and disease. *Front Neurosci* *9*, 98.
- Mahler, H.-C., Friess, W., Grauschopf, U., and Kiese, S. (2009). Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci* *98*, 2909–2934.
- Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* *11*, 3696–3713.
- Malhis, N., Jacobson, M., and Gsponer, J. (2016). MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.* *44*, W488–493.
- 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.
- Manon-Jensen, T., Multhaupt, H.A.B., and Couchman, J.R. (2013). Mapping of matrix metalloproteinase cleavage sites on syndecan-1 and syndecan-4 ectodomains. *FEBS J.* *280*, 2320–2331.
- 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.
- Marguerie, G., Benabid, Y., and Suscillon, M. (1979). The binding of calcium to fibrinogen: influence on the clotting process. *Biochim. Biophys. Acta* *579*, 134–141.
- Marklund, E.G., Degiacomi, M.T., Robinson, C.V., Baldwin, A.J., and Benesch, J.L.P. (2015). Collision cross sections for structural proteomics. *Structure* *23*, 791–799.
- Martinelli, A.H.S., Lopes, F.C., John, E.B.O., Carlini, C.R., and Ligabue-Braun, R. (2019). Modulation of Disordered Proteins with a Focus on Neurodegenerative Diseases and Other Pathologies. *Int J Mol Sci* *20*.
- Marx, G. (1988). Zinc binding to fibrinogen and fibrin. *Arch. Biochem. Biophys.* *266*, 285–288.
- Maupetit, J., Gautier, R., and Tufféry, P. (2006). SABBAC: online Structural Alphabet-based protein Backbone reconstruction from Alpha-Carbon trace. *Nucleic Acids Res.* *34*, W147–151.
- McCarthy, K.J. (2015). The Basement Membrane Proteoglycans Perlecan and Agrin: Something Old, Something New. *Curr Top Membr* *76*, 255–303.
- McFall, A.J., and Rapraeger, A.C. (1997). Identification of an adhesion site within the syndecan-4 extracellular protein domain. *J. Biol. Chem.* *272*, 12901–12904.
- McFall, A.J., and Rapraeger, A.C. (1998). Characterization of the high affinity cell-binding domain in the cell surface proteoglycan syndecan-4. *J. Biol. Chem.* *273*, 28270–28276.
- McGuffin, L.J., Adiyaman, R., Maghrabi, A.H.A., Shuid, A.N., Brackenridge, D.A., Nealon, J.O., and Philomina, L.S. (2019). IntFOLD: an integrated web resource for high performance protein structure and function prediction. *Nucleic Acids Res.* *47*, W408–W413.
- Meade, R.M., Fairlie, D.P., and Mason, J.M. (2019). Alpha-synuclein structure and Parkinson's disease - lessons and emerging principles. *Mol Neurodegener* *14*, 29.
- Mecham, R.P. (2012). Overview of extracellular matrix. *Curr Protoc Cell Biol Chapter 10*, Unit 10.1.



- Meldrum, O.W., Yakubov, G.E., Bonilla, M.R., Deshmukh, O., McGuckin, M.A., and Gidley, M.J. (2018). Mucin gel assembly is controlled by a collective action of non-mucin proteins, disulfide bridges, Ca<sup>2+</sup>-mediated links, and hydrogen bonding. *Sci Rep* *8*, 5802.
- Melrose, J. (2019). Keratan sulfate (KS)-proteoglycans and neuronal regulation in health and disease: the importance of KS-glycodynamics and interactive capability with neuroregulatory ligands. *J. Neurochem.* *149*, 170–194.
- Meneghetti, M.C.Z., Hughes, A.J., Rudd, T.R., Nader, H.B., Powell, A.K., Yates, E.A., and Lima, M.A. (2015). Heparan sulfate and heparin interactions with proteins. *J R Soc Interface* *12*, 0589.
- Meng, F., Uversky, V.N., and Kurgan, L. (2017). Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell. Mol. Life Sci.* *74*, 3069–3090.
- Mercurius, K.O., and Morla, A.O. (2001). Cell adhesion and signaling on the fibronectin 1st type III repeat; requisite roles for cell surface proteoglycans and integrins. *BMC Cell Biol.* *2*, 18.
- Merline, R., Schaefer, R.M., and Schaefer, L. (2009). The matricellular functions of small leucine-rich proteoglycans (SLRPs). *J Cell Commun Signal* *3*, 323–335.
- Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O.P., Kovács, M., et al. (2020). PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.* *48*, D360–D367.
- Meyer, K., Kirchner, M., Uyar, B., Cheng, J.-Y., Russo, G., Hernandez-Miranda, L.R., Szymborska, A., Zaubner, H., Rudolph, I.-M., Willnow, T.E., et al. (2018). Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs. *Cell* *175*, 239–253.e17.
- Midic, U., Oldfield, C.J., Dunker, A.K., Obradovic, Z., and Uversky, V.N. (2009). Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *BMC Genomics* *10 Suppl 1*, S12.
- Miskei, M., Antal, C., and Fuxreiter, M. (2017). FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res.* *45*, D228–D235.
- Mitra, G. (2019). Application of native mass spectrometry in studying intrinsically disordered proteins: A special focus on neurodegenerative diseases. *Biochim Biophys Acta Proteins Proteom* *1867*, 140260.
- Mitsou, I., Mulhaupt, H.A.B., and Couchman, J.R. (2017). Proteoglycans, ion channels and cell-matrix adhesion. *Biochem. J.* *474*, 1965–1979.
- 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.
- Morgan, M.R., Humphries, M.J., and Bass, M.D. (2007). Synergistic control of cell adhesion by integrins and syndecans. *Nat. Rev. Mol. Cell Biol.* *8*, 957–969.
- Morla, A., Zhang, Z., and Ruoslahti, E. (1994). Superfibronectin is a functionally distinct form of fibronectin. *Nature* *367*, 193–196.
- Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* *19*, 1639–1662.

- Müller, U.C., Deller, T., and Korte, M. (2017). Not just amyloid: physiological functions of the amyloid precursor protein family. *Nat. Rev. Neurosci.* *18*, 281–298.
- Multhaupt, H. a. B., Yoneda, A., Whiteford, J.R., Oh, E.-S., Lee, W., and Couchman, J.R. (2009). Syndecan signaling: when, where and why? *J. Physiol. Pharmacol.* *60 Suppl 4*, 31–38.
- Mytilinaiou, M., Nikitovic, D., Berdiaki, A., Kostouras, A., Papoutsidakis, A., Tsatsakis, A.M., and Tzanakakis, G.N. (2017). Emerging roles of syndecan 2 in epithelial and mesenchymal cancer progression. *IUBMB Life* *69*, 824–833.
- Na, J.-H., Lee, W.-K., and Yu, Y.G. (2018). How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. *Int J Mol Sci* *19*.
- Naba, A., Clauser, K.R., Hoersch, S., Liu, H., Carr, S.A., and Hynes, R.O. (2012). The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell Proteomics* *11*, M111.014647.
- Naba, A., Clauser, K.R., Ding, H., Whittaker, C.A., Carr, S.A., and Hynes, R.O. (2016). The extracellular matrix: Tools and insights for the “omics” era. *Matrix Biol.* *49*, 10–24.
- Naseri, N.N., Wang, H., Guo, J., Sharma, M., and Luo, W. (2019). The complexity of tau in Alzheimer’s disease. *Neurosci. Lett.* *705*, 183–194.
- Nash, G.B., Wenby, R.B., Sowemimo-Coker, S.O., and Meiselman, H.J. (1987). Influence of cellular properties on red cell aggregation. *Clinical Hemorheology and Microcirculation* *7*, 93–108.
- Negrioni, E., Henault, E., Chevalier, F., Gilbert-Sirieix, M., Van Kuppevelt, T.H., Papy-Garcia, D., Uzan, G., and Albanese, P. (2014). Glycosaminoglycan modifications in Duchenne muscular dystrophy: specific remodeling of chondroitin sulfate/dermatan sulfate. *J. Neuropathol. Exp. Neurol.* *73*, 789–797.
- Nieuwdorp, M., van Haefen, T.W., Gouverneur, M.C.L.G., Mooij, H.L., van Lieshout, M.H.P., Levi, M., Meijers, J.C.M., Holleman, F., Hoekstra, J.B.L., Vink, H., et al. (2006). Loss of endothelial glycocalyx during acute hyperglycemia coincides with endothelial dysfunction and coagulation activation in vivo. *Diabetes* *55*, 480–486.
- Nilsson, M.R. (2004). Techniques to study amyloid fibril formation in vitro. *Methods* *34*, 151–160.
- Nishitsuji, K. (2018). Heparan sulfate S-domains and extracellular sulfatases (Sulfs): their possible roles in protein aggregation diseases. *Glycoconj. J.* *35*, 387–396.
- Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B., Dosztányi, Z., Uversky, V.N., Obradovic, Z., Kurgan, L., et al. (2013). D<sup>2</sup>P<sup>2</sup>: database of disordered protein predictions. *Nucleic Acids Res.* *41*, D508-516.
- Ohashi, T., Augustus, A.M., and Erickson, H.P. (2009). Transient opening of fibronectin type III (FNIII) domains: the interaction of the third FNIII domain of FN with anastellin. *Biochemistry* *48*, 4189–4197.
- Olsen, J.G., Teilum, K., and Kragelund, B.B. (2017). Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* *74*, 3175–3183.
- Oohira, A., Shuo, T., Tokita, Y., Nakanishi, K., and Aono, S. (2004). Neuroglycan C, a brain-specific part-time proteoglycan, with a particular multidomain structure. *Glycoconj. J.* *21*, 53–57.

- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., et al. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* *42*, D358–363.
- 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.
- Owen, I., and Shewmaker, F. (2019). The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins. *Int J Mol Sci* *20*.
- Ozbun, M.A. (2019). Extracellular events impacting human papillomavirus infections: Epithelial wounding to cell signaling involved in virus entry. *Papillomavirus Res* *7*, 188–192.
- Pahakis, M.Y., Kosky, J.R., Dull, R.O., and Tarbell, J.M. (2007). The role of endothelial glycocalyx components in mechanotransduction of fluid shear stress. *Biochem. Biophys. Res. Commun.* *355*, 228–233.
- Pahari, S., Sun, L., Basu, S., and Alexov, E. (2018). DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate. *Proteins* *86*, 1277–1283.
- Pang, X., Dong, N., and Zheng, Z. (2019). Small Leucine-Rich Proteoglycans in Skin Wound Healing. *Front Pharmacol* *10*, 1649.
- Panjikovich, A., and Svergun, D.I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* *34*, 1944–1946.
- Park, P.W. (2018). Isolation and functional analysis of syndecans. *Methods Cell Biol.* *143*, 317–333.
- Park, T.-J., Lee, M.-H., Kim, J.-S., and Kim, Y. (2011). Recombinant expression, purification, and characterization of transmembrane domain of syndecan-4. *Process Biochemistry* *46*, 1166–1171.
- Pathan, M., Keerthikumar, S., Ang, C.-S., Gangoda, L., Quek, C.Y.J., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., et al. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics* *15*, 2597–2601.
- Pelikan, M., Hura, G.L., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* *28*, 174–189.
- Perdigão, N., Rosa, A.C., and O'Donoghue, S.I. (2017). The Dark Proteome Database. *BioData Min* *10*, 24.
- Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* *89*, 1237–1250.
- Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C.J., Aspromonte, M.C., Davey, N.E., Davidović, R., Dosztányi, Z., et al. (2017). DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* *45*, D219–D227.
- Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Micetic, I., Camilloni, C., Davey, N., Dosztányi, Z., Mészáros, B., Monzon, A.M., et al. (2018). MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* *46*, D471–D476.
- Purushothaman, A., Uyama, T., Kobayashi, F., Yamada, S., Sugahara, K., Rapraeger, A.C., and Sanderson, R.D. (2010). Heparanase-enhanced shedding of syndecan-1 by myeloma cells promotes endothelial invasion and angiogenesis. *Blood* *115*, 2449–2457.

- Quittot, N., Sebastiao, M., and Bourgault, S. (2017). Modulation of amyloid assembly by glycosaminoglycans: from mechanism to biological significance. *Biochem. Cell Biol.* *95*, 329–337.
- Ramachandran, P.L., Lovett, J.E., Carl, P.J., Cammarata, M., Lee, J.H., Jung, Y.O., Ihee, H., Timmel, C.R., and van Thor, J.J. (2011). The short-lived signaling state of the photoactive yellow protein photoreceptor revealed by combined structural probes. *J. Am. Chem. Soc.* *133*, 9395–9404.
- Ramani, V.C., Purushothaman, A., Stewart, M.D., Thompson, C.A., Vlodaysky, I., Au, J.L.-S., and Sanderson, R.D. (2013). The heparanase/syndecan-1 axis in cancer: mechanisms and therapies. *FEBS J.* *280*, 2294–2306.
- Randles, M.J., Humphries, M.J., and Lennon, R. (2017). Proteomic definitions of basement membrane composition in health and disease. *Matrix Biol.* *57–58*, 12–28.
- Rapraeger, A.C. (2000). Syndecan-regulated receptor signaling. *J. Cell Biol.* *149*, 995–998.
- Rapraeger, A.C. (2013). Synstatin: a selective inhibitor of the syndecan-1-coupled IGF1R- $\alpha$ v $\beta$ 3 integrin complex in tumorigenesis and angiogenesis. *FEBS J.* *280*, 2207–2215.
- Receveur-Brechot, V., and Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* *13*, 55–75.
- Reily, C., Stewart, T.J., Renfrow, M.B., and Novak, J. (2019). Glycosylation in health and disease. *Nat Rev Nephrol* *15*, 346–366.
- Reitsma, S., Slaaf, D.W., Vink, H., van Zandvoort, M.A.M.J., and oude Egbrink, M.G.A. (2007). The endothelial glycocalyx: composition, functions, and visualization. *Pflugers Arch.* *454*, 345–359.
- Renoux, C., Faivre, M., Bessaa, A., Da Costa, L., Joly, P., Gauthier, A., and Connes, P. (2019). Impact of surface-area-to-volume ratio, internal viscosity and membrane viscoelasticity on red blood cell deformability measured in isotonic condition. *Sci Rep* *9*, 6771.
- Ricard-Blum, S., and Vallet, S.D. (2019). Fragments generated upon extracellular matrix remodeling: Biological regulators and potential drugs. *Matrix Biol.* *75–76*, 170–189.
- Rivera, S., García-González, L., Khrestchatsky, M., and Baranger, K. (2019). Metalloproteinases and their tissue inhibitors in Alzheimer's disease and other neurodegenerative disorders. *Cell. Mol. Life Sci.* *76*, 3167–3191.
- Roper, J.A., Williamson, R.C., and Bass, M.D. (2012). Syndecan and integrin interactomes: large complexes in small spaces. *Curr. Opin. Struct. Biol.* *22*, 583–590.
- Roseman, S. (2001). Reflections on glycobiology. *J. Biol. Chem.* *276*, 41527–41542.
- Rosenberg, G.A. (2009). Matrix metalloproteinases and their multiple roles in neurodegenerative diseases. *Lancet Neurol* *8*, 205–216.
- Różycki, B., Kim, Y.C., and Hummer, G. (2011). SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* *19*, 109–116.
- Ruan, H., Sun, Q., Zhang, W., Liu, Y., and Lai, L. (2019). Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discov. Today* *24*, 217–227.
- Saikusa, K., Kuwabara, N., Kokabu, Y., Inoue, Y., Sato, M., Iwasaki, H., Shimizu, T., Ikeguchi, M., and Akashi, S. (2013). Characterisation of an intrinsically disordered protein complex of Swi5-Sfr1 by ion mobility mass spectrometry and small-angle X-ray scattering. *Analyst* *138*, 1441–1449.

- Salmivirta, M., Elenius, K., Vainio, S., Hofer, U., Chiquet-Ehrismann, R., Thesleff, I., and Jalkanen, M. (1991). Syndecan from embryonic tooth mesenchyme binds tenascin. *J. Biol. Chem.* *266*, 7733–7739.
- Salvi, N., Abyzov, A., and Blackledge, M. (2017). Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Prog Nucl Magn Reson Spectrosc* *102–103*, 43–60.
- Salzano, G., Giachin, G., and Legname, G. (2019). Structural Consequences of Copper Binding to the Prion Protein. *Cells* *8*.
- Samarel, A.M. (2013). Syndecan-4: a component of the mechanosensory apparatus of cardiac fibroblasts. *J. Mol. Cell. Cardiol.* *56*, 19–21.
- Samsonov, S.A., Bichmann, L., and Pisabarro, M.T. (2015). Coarse-grained model of glycosaminoglycans. *J Chem Inf Model* *55*, 114–124.
- Santambrogio, C., Natalello, A., Brocca, S., Ponzini, E., and Grandori, R. (2019). Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. *Proteomics* *19*, e1800060.
- Santofimia-Castaño, P., Rizzuti, B., Xia, Y., Abian, O., Peng, L., Velázquez-Campoy, A., Neira, J.L., and Iovanna, J. (2020). Targeting intrinsically disordered proteins involved in cancer. *Cell. Mol. Life Sci.* *77*, 1695–1707.
- Sarrazin, S., Lamanna, W.C., and Esko, J.D. (2011). Heparan sulfate proteoglycans. *Cold Spring Harb Perspect Biol* *3*.
- Sasarman, F., Maftai, C., Campeau, P.M., Brunel-Guitton, C., Mitchell, G.A., and Allard, P. (2016). Biosynthesis of glycosaminoglycans: associated disorders and biochemical tests. *J. Inherit. Metab. Dis.* *39*, 173–188.
- Sattelle, B.M., Shakeri, J., and Almond, A. (2013). Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* *14*, 1149–1159.
- Saunders, S., Jalkanen, M., O'Farrell, S., and Bernfield, M. (1989). Molecular cloning of syndecan, an integral membrane proteoglycan. *J. Cell Biol.* *108*, 1547–1556.
- Sayyad, M.R., Puchalapalli, M., Vergara, N.G., Wangenstein, S.M., Moore, M., Mu, L., Edwards, C., Anderson, A., Kall, S., Sullivan, M., et al. (2019). Syndecan-1 facilitates breast cancer metastasis to the brain. *Breast Cancer Res. Treat.* *178*, 35–49.
- Schlessinger, A., Schaefer, C., Vicedo, E., Schmidberger, M., Punta, M., and Rost, B. (2011). Protein disorder--a breakthrough invention of evolution? *Curr. Opin. Struct. Biol.* *21*, 412–418.
- Schmidt, B., Ho, L., and Hogg, P.J. (2006). Allosteric disulfide bonds. *Biochemistry* *45*, 7429–7433.
- Schneider, R., Blackledge, M., and Jensen, M.R. (2019). Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Curr. Opin. Struct. Biol.* *54*, 10–18.
- Schramm, A., Lieutaud, P., Gianni, S., Longhi, S., and Bignon, C. (2017). InSiDDe: A Server for Designing Artificial Disordered Proteins. *Int J Mol Sci* *19*.

- Schweers, O., Schönbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J. Biol. Chem.* *269*, 24290–24297.
- Scully, O.J., Chua, P.-J., Harve, K.S., Bay, B.-H., and Yip, G.W. (2012). Serglycin in health and diseases. *Anat Rec (Hoboken)* *295*, 1415–1420.
- Shah, N.B., and Duncan, T.M. (2014). Bio-layer interferometry for measuring kinetics of protein-protein interactions and allosteric ligand effects. *J Vis Exp* e51383.
- Shao, X., Taha, I.N., Clauser, K.R., Gao, Y. (Tom), and Naba, A. (2020). MatrisomeDB: the ECM-protein knowledge database. *Nucleic Acids Res* *48*, D1136–D1144.
- Sharma, R., Raduly, Z., Miskei, M., and Fuxreiter, M. (2015). Fuzzy complexes: Specific binding without complete folding. *FEBS Lett.* *589*, 2533–2542.
- Sharma, R., Sharma, A., Raicar, G., Tsunoda, T., and Patil, A. (2019). OPAL+: Length-Specific MoRF Prediction in Intrinsically Disordered Protein Sequences. *Proteomics* *19*, e1800058.
- Shimizu, Y., Suzuki, T., Yoshikawa, T., Endo, I., and Nakatsura, T. (2019). Next-Generation Cancer Immunotherapy Targeting Glypican-3. *Front Oncol* *9*, 248.
- Shin, J., Lee, W., Lee, D., Koo, B.K., Han, I., Lim, Y., Woods, A., Couchman, J.R., and Oh, E.S. (2001). Solution structure of the dimeric cytoplasmic domain of syndecan-4. *Biochemistry* *40*, 8471–8478.
- Shintani, Y., Takashima, S., Asano, Y., Kato, H., Liao, Y., Yamazaki, S., Tsukamoto, O., Seguchi, O., Yamamoto, H., Fukushima, T., et al. (2006). Glycosaminoglycan modification of neuropilin-1 modulates VEGFR2 signaling. *EMBO J.* *25*, 3045–3055.
- Shriver, Z., Capila, I., Venkataraman, G., and Sasisekharan, R. (2012). Heparin and Heparan Sulfate: Analyzing Structure and Microheterogeneity. *Handb Exp Pharmacol* 159–176.
- Sibille, N., and Bernadó, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.* *40*, 955–962.
- Simmonds, M.J., Meiselman, H.J., and Baskurt, O.K. (2013). Blood rheology and aging. *J Geriatr Cardiol* *10*, 291–301.
- Simon, A.-L., Chirot, F., Choi, C.M., Clavier, C., Barbaire, M., Maurelli, J., Dagany, X., MacAleese, L., and Dugourd, P. (2015). Tandem ion mobility spectrometry coupled to laser excitation. *Rev Sci Instrum* *86*, 094101.
- 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* *19*, 1–26.
- Sottile, J., and Chandler, J. (2005). Fibronectin matrix turnover occurs through a caveolin-1-dependent process. *Mol. Biol. Cell* *16*, 757–768.
- Squire, J.M., Chew, M., Nneji, G., Neal, C., Barry, J., and Michel, C. (2001). Quasi-periodic substructure in the microvessel endothelial glycocalyx: a possible explanation for molecular filtering? *J. Struct. Biol.* *136*, 239–255.
- Sreerama, N., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* *287*, 252–260.

- Sreerama, N., Venyaminov, S.Y., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* *287*, 243–251.
- Srikanth, C.B., Salimath, P.V., and Nandini, C.D. (2012). Erythrocytes express chondroitin sulphate/dermatan sulphate, which undergoes quantitative changes during diabetes and mediate erythrocyte adhesion to extracellular matrix components. *Biochimie* *94*, 1347–1355.
- Stepp, M.A., Pal-Ghosh, S., Tadvalkar, G., and Pajooresh-Ganji, A. (2015). Syndecan-1 and Its Expanding List of Contacts. *Adv Wound Care (New Rochelle)* *4*, 235–249.
- Stewart, M.D., Ramani, V.C., and Sanderson, R.D. (2015). Shed syndecan-1 translocates to the nucleus of cells delivering growth factors and inhibiting histone acetylation: a novel mechanism of tumor-host cross-talk. *J. Biol. Chem.* *290*, 941–949.
- Stine, J.M., Ahl, G.J.H., Schlenker, C., Rusnac, D.-V., and Briknarová, K. (2017). The Interaction between the Third Type III Domain from Fibronectin and Anastellin Involves  $\beta$ -Strand Exchange. *Biochemistry* *56*, 4667–4675.
- Stuchfield, D., and Barran, P. (2018). Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Curr Opin Chem Biol* *42*, 177–185.
- Stuchfield, D., France, A.P., Migas, L.G., Thalhammer, A., Bremer, A., Bellina, B., and Barran, P.E. (2018). The Use of Mass Spectrometry to Examine IDPs: Unique Insights and Caveats. *Meth. Enzymol.* *611*, 459–502.
- Su, Y., and Chang, P.T. (2001). Acidic pH promotes the formation of toxic fibrils from beta-amyloid peptide. *Brain Res.* *893*, 287–291.
- Svensson, G., Awad, W., Håkansson, M., Mani, K., and Logan, D.T. (2012). Crystal structure of N-glycosylated human glypican-1 core protein: structure of two loops evolutionarily conserved in vertebrate glypican-1. *J. Biol. Chem.* *287*, 14040–14051.
- Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRY SOL— a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* *28*, 768–773.
- Szarvas, T., Sevcenco, S., Módos, O., Keresztes, D., Nyirády, P., Kubik, A., Romics, M., Kovalszky, I., Reis, H., Hadaschik, B., et al. (2018). Circulating syndecan-1 is associated with chemotherapy-resistance in castration-resistant prostate cancer. *Urol. Oncol.* *36*, 312.e9-312.e15.
- Szatmári, T., Ötvös, R., Hjerpe, A., and Dobra, K. (2015). Syndecan-1 in Cancer: Implications for Cell Signaling, Differentiation, and Prognostication. *Dis. Markers* *2015*, 796052.
- Tagliabracci, V.S., Wiley, S.E., Guo, X., Kinch, L.N., Durrant, E., Wen, J., Xiao, J., Cui, J., Nguyen, K.B., Engel, J.L., et al. (2015). A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* *161*, 1619–1632.
- Tarbell, J.M., and Cancel, L.M. (2016). The glycocalyx and its significance in human medicine. *J. Intern. Med.* *280*, 97–113.
- Tarbell, J.M., Simon, S.I., and Curry, F.-R.E. (2014). Mechanosensing at the Vascular Interface. *Annu Rev Biomed Eng* *16*, 505–532.

- Teese, M.G., and Langosch, D. (2015). Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* *54*, 5125–5135.
- Testa, L., Brocca, S., and Grandori, R. (2011). Charge-surface correlation in electrospray ionization of folded and unfolded proteins. *Anal. Chem.* *83*, 6459–6463.
- Testa, L., Brocca, S., Santambrogio, C., D'Urzo, A., Habchi, J., Longhi, S., Uversky, V.N., and Grandori, R. (2013a). Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. *Intrinsically Disord Proteins* *1*, e25068.
- Testa, L., Brocca, S., Santambrogio, C., D'Urzo, A., Habchi, J., Longhi, S., Uversky, V.N., and Grandori, R. (2013b). Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. *Intrinsically Disord Proteins* *1*, e25068.
- Theocharis, A.D., Skandalis, S.S., Gialeli, C., and Karamanos, N.K. (2016). Extracellular matrix structure. *Adv. Drug Deliv. Rev.* *97*, 4–27.
- Thi, M.M., Tarbell, J.M., Weinbaum, S., and Spray, D.C. (2004). The role of the glycocalyx in reorganization of the actin cytoskeleton under fluid shear stress: a “bumper-car” model. *Proc. Natl. Acad. Sci. U.S.A.* *101*, 16483–16488.
- Thorne, R.F., Legg, J.W., and Isacke, C.M. (2004). The role of the CD44 transmembrane and cytoplasmic domains in co-ordinating adhesive and signalling events. *J. Cell. Sci.* *117*, 373–380.
- Tibaldi, E., Brocca, A., Sticca, A., Gola, E., Pizzi, M., Bordin, L., Pagano, M.A., Mazzorana, M., Donà, G., Violi, P., et al. (2020). Fam20C-mediated phosphorylation of osteopontin is critical for its secretion but dispensable for its action as a cytokine in the activation of hepatic stellate cells in liver fibrogenesis. *FASEB J.* *34*, 1122–1135.
- Tkachenko, E., Rhodes, J.M., and Simons, M. (2005). Syndecans: new kids on the signaling block. *Circ. Res.* *96*, 488–500.
- Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* *33*, 2–8.
- Tompa, P., Schad, E., Tantos, A., and Kalmar, L. (2015). Intrinsically disordered proteins: emerging interaction specialists. *Curr. Opin. Struct. Biol.* *35*, 49–59.
- Toole, B.P. (2004). Hyaluronan: from extracellular glue to pericellular cue. *Nat. Rev. Cancer* *4*, 528–539.
- Tousoulis, D., Kampoli, A.-M., Tentolouris, C., Papageorgiou, N., and Stefanadis, C. (2012). The role of nitric oxide on endothelial function. *Curr Vasc Pharmacol* *10*, 4–18.
- Tria, G., Mertens, H.D.T., Kachala, M., and Svergun, D.I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* *2*, 207–217.
- Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* *31*, 455–461.
- Tsoyi, K., Osorio, J.C., Chu, S.G., Fernandez, I.E., De Frias, S.P., Sholl, L., Cui, Y., Tellez, C.S., Siegfried, J.M., Belinsky, S.A., et al. (2019). Lung Adenocarcinoma Syndecan-2 Potentiates Cell Invasiveness. *Am. J. Respir. Cell Mol. Biol.* *60*, 659–666.



- Tuttle, M.D., Comellas, G., Nieuwkoop, A.J., Covell, D.J., Berthold, D.A., Kloepper, K.D., Courtney, J.M., Kim, J.K., Barclay, A.M., Kendall, A., et al. (2016). Solid-state NMR structure of a pathogenic fibril of full-length human  $\alpha$ -synuclein. *Nat. Struct. Mol. Biol.* *23*, 409–415.
- Ulker, P., Yaras, N., Yalcin, O., Celik-Ozenci, C., Johnson, P.C., Meiselman, H.J., and Baskurt, O.K. (2011). Shear stress activation of nitric oxide synthase and increased nitric oxide levels in human red blood cells. *Nitric Oxide* *24*, 184–191.
- UniProt Consortium, T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* *46*, 2699.
- Uversky, V.N. (2002). What does it mean to be natively unfolded? *Eur. J. Biochem.* *269*, 2–12.
- Uversky, V.N. (2009). Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* *28*, 305–325.
- Uversky, V.N. (2010). The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J. Biomed. Biotechnol.* *2010*, 568068.
- Uversky, V.N. (2013). A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* *22*, 693–724.
- Uversky, V.N. (2017). Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* *44*, 18–30.
- Uversky, V.N. (2019). Protein intrinsic disorder and structure-function continuum. *Prog Mol Biol Transl Sci* *166*, 1–17.
- Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* *41*, 415–427.
- Uversky, V.N., Oldfield, C.J., and Dunker, A.K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* *37*, 215–246.
- Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* *6*, 2351–2366.
- Vacic, V., Markwick, P.R.L., Oldfield, C.J., Zhao, X., Haynes, C., Uversky, V.N., and Iakoucheva, L.M. (2012). Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. *PLoS Comput Biol* *8*.
- Vallet, S.D., Miele, A.E., Uciechowska-Kaczmarzyk, U., Liwo, A., Duclos, B., Samsonov, S.A., and Ricard-Blum, S. (2018). Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. *Sci Rep* *8*, 11768.
- van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sánchez de Groot, N., Huynen, M.A., Matouschek, A., Fuxreiter, M., and Babu, M.M. (2014a). Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution. *Cell Rep* *8*, 1832–1844.
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014b). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* *114*, 6589–6631.

- Varadi, M., and Tompa, P. (2015). The Protein Ensemble Database. *Adv. Exp. Med. Biol.* *870*, 335–349.
- Velleman, S.G., and Coy, C.S. (2020). Research Note: Effect of selection for body weight on the adipogenic conversion of turkey myogenic satellite cells by Syndecan-4 and its covalently attached N-glycosylation chains. *Poult. Sci.* *99*, 1209–1215.
- Vendruscolo, M. (2010). Enzymatic activity in disordered states of proteins. *Curr Opin Chem Biol* *14*, 671–675.
- Vicente, C.M., da Silva, D.A., Sartorio, P.V., Silva, T.D., Saad, S.S., Nader, H.B., Forones, N.M., and Toma, L. (2018). Heparan Sulfate Proteoglycans in Human Colorectal Cancer. *Anal Cell Pathol (Amst)* *2018*, 8389595.
- Vogt, A.M., Winter, G., Wahlgren, M., and Spillmann, D. (2004). Heparan sulphate identified on human erythrocytes: a Plasmodium falciparum receptor. *Biochem. J.* *381*, 593–597.
- Voyvodic, P.L., Min, D., Liu, R., Williams, E., Chitalia, V., Dunn, A.K., and Baker, A.B. (2014). Loss of syndecan-1 induces a pro-inflammatory phenotype in endothelial cells with a dysregulated response to atheroprotective flow. *J. Biol. Chem.* *289*, 9547–9559.
- Vuong, T.T., Reine, T.M., Sudworth, A., Jenssen, T.G., and Kolset, S.O. (2015). Syndecan-4 is a major syndecan in primary human endothelial cells in vitro, modulated by inflammatory stimuli and involved in wound healing. *J. Histochem. Cytochem.* *63*, 280–292.
- Wan, Q.-F., Zhou, Z.-Y., Thakur, P., Vila, A., Sherry, D.M., Janz, R., and Heidelberger, R. (2010). SV2 acts via presynaptic calcium to regulate neurotransmitter release. *Neuron* *66*, 884–895.
- Wang, W., and Wang, D. (2019). Extreme Fuzziness: Direct Interactions between Two IDPs. *Biomolecules* *9*.
- Wang, H., Jin, H., and Rapraeger, A.C. (2015). Syndecan-1 and Syndecan-4 Capture Epidermal Growth Factor Receptor Family Members and the  $\alpha 3\beta 1$  Integrin Via Binding Sites in Their Ectodomains: novel synstatins prevent kinase capture and inhibit  $\alpha 6\beta 4$ -integrin-dependent epithelial cell motility. *J. Biol. Chem.* *290*, 26103–26113.
- Wang, J., Wang, W., Kollman, P.A., and Case, D.A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* *25*, 247–260.
- Watanabe, H., Yoshida, C., Ooishi, A., Nakai, Y., Ueda, M., Isobe, Y., and Honda, S. (2019). Histidine-Mediated Intramolecular Electrostatic Repulsion for Controlling pH-Dependent Protein-Protein Interaction. *ACS Chem. Biol.* *14*, 2729–2736.
- Weiler, H., and Isermann, B.H. (2003). Thrombomodulin. *J. Thromb. Haemost.* *1*, 1515–1524.
- Whiteford, J.R., and Couchman, J.R. (2006). A conserved NXIP motif is required for cell adhesion properties of the syndecan-4 ectodomain. *J. Biol. Chem.* *281*, 32156–32163.
- Whiteford, J.R., Ko, S., Lee, W., and Couchman, J.R. (2008). Structural and cell adhesion properties of zebrafish syndecan-4 are shared with higher vertebrates. *J. Biol. Chem.* *283*, 29322–29330.
- Whiteford, J.R., Xian, X., Chaussade, C., Vanhaesebroeck, B., Nourshargh, S., and Couchman, J.R. (2011). Syndecan-2 is a novel ligand for the protein tyrosine phosphatase receptor CD148. *Mol. Biol. Cell* *22*, 3609–3624.

- Whitmore, L., and Wallace, B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* *89*, 392–400.
- van Wijk, X.M.R., and van Kuppevelt, T.H. (2014). Heparan sulfate in angiogenesis: a target for therapy. *Angiogenesis* *17*, 443–462.
- Wilsie, L.C., Gonzales, A.M., and Orlando, R.A. (2006). Syndecan-1 mediates internalization of apoE-VLDL through a low density lipoprotein receptor-related protein (LRP)-independent, non-clathrin-mediated pathway. *Lipids Health Dis* *5*, 23.
- Wilson, J.L., Scott, I.M., and McMurry, J.L. (2010). Optical biosensing: Kinetics of protein A-IGG binding using biolayer interferometry. *Biochem Mol Biol Educ* *38*, 400–407.
- Woods, A. (2001). Syndecans: transmembrane modulators of adhesion and matrix assembly. *J. Clin. Invest.* *107*, 935–941.
- 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.
- Woods, A., and Couchman, J.R. (2001). Proteoglycan isolation and analysis. *Curr Protoc Cell Biol Chapter 10*, Unit 10.7.
- Woods, A., Longley, R.L., Tumova, S., and Couchman, J.R. (2000). Syndecan-4 binding to the high affinity heparin-binding domain of fibronectin drives focal adhesion formation in fibroblasts. *Arch. Biochem. Biophys.* *374*, 66–72.
- Wu, X., Kan, M., Wang, F., Jin, C., Yu, C., and McKeenan, W.L. (2001). A rare premalignant prostate tumor epithelial cell syndecan-1 forms a fibroblast growth factor-binding complex with progression-promoting ectopic fibroblast growth factor receptor 1. *Cancer Res.* *61*, 5295–5302.
- Wuttke, R., Hofmann, H., Nettels, D., Borgia, M.B., Mittal, J., Best, R.B., and Schuler, B. (2014). Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* *111*, 5213–5218.
- Xing, Y., Xu, Q., and Lee, C. (2003). Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains. *FEBS Lett.* *555*, 572–578.
- Xu, D., and Esko, J.D. (2014). Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* *83*, 129–157.
- Xue, B., Dunker, A.K., and Uversky, V.N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* *30*, 137–149.
- Xue, C., Lin, T.Y., Chang, D., and Guo, Z. (2017). Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *R Soc Open Sci* *4*, 160696.
- Yan, J., Dunker, A.K., Uversky, V.N., and Kurgan, L. (2016). Molecular recognition features (MoRFs) in three domains of life. *Mol Biosyst* *12*, 697–710.
- Yang, J., Price, M.A., Wanshura, L.E.C., He, J., Yi, M., Welch, D.R., Li, G., Conner, S., Sachs, J., Turley, E.A., et al. (2019). Chondroitin sulfate proteoglycan 4 enhanced melanoma motility and growth requires a cysteine in the core protein transmembrane domain. *Melanoma Res.* *29*, 365–375.

- Yang, Y., Koo, S., Heng, L.T., Meiselman, H.J., and Neu, B. (2014). Non-adsorbing macromolecules promote endothelial adhesion of erythrocytes with reduced sialic acids. *Biochim. Biophys. Acta* *1840*, 288–293.
- Yao, W., Rose, J.L., Wang, W., Seth, S., Jiang, H., Taguchi, A., Liu, J., Yan, L., Kapoor, A., Hou, P., et al. (2019). Syndecan 1 is a critical mediator of macropinocytosis in pancreatic cancer. *Nature* *568*, 410–414.
- Yen, W.-Y., Cai, B., Zeng, M., Tarbell, J.M., and Fu, B.M. (2012). Quantification of the endothelial surface glycocalyx on rat and mouse blood vessels. *Microvasc. Res.* *83*, 337–346.
- Yi, M., and Ruoslahti, E. (2001). A fibronectin fragment inhibits tumor growth, angiogenesis, and metastasis. *Proc. Natl. Acad. Sci. U.S.A.* *98*, 620–624.
- Yilmaz, O., Afsar, B., Ortiz, A., and Kanbay, M. (2019). The role of endothelial glycocalyx in health and disease. *Clin Kidney J* *12*, 611–619.
- Yoo, J., Jeong, M.-J., Cho, H.J., Oh, E.S., and Han, M.Y. (2005). Dynamin II interacts with syndecan-4, a regulator of focal adhesion and stress-fiber formation. *Biochem. Biophys. Res. Commun.* *328*, 424–431.
- Yu, J.-F., Dou, X.-H., Sha, Y.-J., Wang, C.-L., Wang, H.-B., Chen, Y.-T., Zhang, F., Zhou, Y., and Wang, J.-H. (2017). DisBind: A database of classified functional binding sites in disordered and structured regions of intrinsically disordered proteins. *BMC Bioinformatics* *18*, 206.
- Zandonadi, F.S., Castañeda Santa Cruz, E., and Korvala, J. (2019). New SDC function prediction based on protein-protein interaction using bioinformatics tools. *Comput Biol Chem* *83*, 107087.
- Zeng, Y., and Tarbell, J.M. (2014). The adaptive remodeling of endothelial glycocalyx in response to fluid shear stress. *PLoS ONE* *9*, e86249.
- Zeng, Y., Zhang, X.F., Fu, B.M., and Tarbell, J.M. (2018). The Role of Endothelial Surface Glycocalyx in Mechanosensing and Transduction. *Adv. Exp. Med. Biol.* *1097*, 1–27.
- 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, H., Brown, R.L., Wei, Y., Zhao, P., Liu, S., Liu, X., Deng, Y., Hu, X., Zhang, J., Gao, X.D., et al. (2019). CD44 splice isoform switching determines breast cancer stem cell state. *Genes Dev.* *33*, 166–179.
- Zhang, Y., McKown, R.L., Raab, R.W., Rapraeger, A.C., and Laurie, G.W. (2011). Focus on molecules: syndecan-1. *Exp. Eye Res.* *93*, 329–330.
- Zhou, Y., Su, J.M., Samuel, C.E., and Ma, D. (2019). Measles Virus Forms Inclusion Bodies with Properties of Liquid Organelles. *J. Virol.* *93*.
- Zimmermann, N., Saiga, H., Houthuys, E., Moura-Alves, P., Koehler, A., Bandermann, S., Dorhoi, A., and Kaufmann, S.H.E. (2016). Syndecans promote mycobacterial internalization by lung epithelial cells. *Cell. Microbiol.* *18*, 1846–1856.
- Zou, W., Wan, J., Li, M., Xing, J., Chen, Q., Zhang, Z., and Gong, Y. (2019). Small leucine rich proteoglycans in host immunity and renal diseases. *J Cell Commun Signal* *13*, 463–471.



## PUBLICATIONS & COMMUNICATIONS ORALES ET PAR AFFICHES

### LISTE DES PUBLICATIONS

- Ricard-Blum S, **Gondelaud F.** Shuttling from the extracellular matrix to the nucleus. *Biol Aujourd'hui* (2016) 210 : 37-44.
- **Gondelaud F.**, Ricard-Blum S., Structures and interactions of syndecans. *FEBS Journal* (2019) 286 : 2994-3007.
- **Gondelaud F.**, Connes P., Nader E., Renoux C., Fort R., Gauthier A., Joly P., Ricard-Blum S. Sialic acids rather than glycosaminoglycans affect normal and sickle red blood cell rheology by binding to four major sites on fibrinogen. *Am J Hematol* (2020). doi : 10.1002/ajh.25718.
- **Gondelaud F.**, Bouakil M., Le Fèvre A., Miele AE., Chirot F., Duclos B., Ricard-Blum S. The ectodomains of the four syndecans: a disordered family with distinct conformations. [*Soumis pour publication*]

### LISTE DES COMMUNICATIONS ORALES ET PAR AFFICHES

- **Gondelaud F.**, Connes P., Nader E., Renoux C., Ricard-Blum S. The role of glycosaminoglycans and sialic acid in the rheology of red blood cells. GDR-GAG (GDR 3739) *Groupement de Recherche sur les glycosaminoglycans*. 10 – 11 octobre 2016, Paris. (Communication par affiche).
- **Gondelaud F.**, Miele AE., Duclos B., Ricard-Blum S. Structural characterization of the ectodomains of human syndecans. GDR-GAG (GDR 3739) *Groupement de Recherche sur les glycosaminoglycans*. 24 - 25 juillet 2017, Lille. (Communication par affiche).
- **Gondelaud F.**, Miele AE., Duclos B., Ricard-Blum S. Structural characterization of the human ectodomains of syndecans. *Journée scientifique de l'ICBMS*, Villeurbanne, 12 Juillet 2018. (Communication orale).
- **Gondelaud F.**, Miele AE., Duclos B., Ricard-Blum S. Structural characterization of the human ectodomains of syndecans. *Matrix Biology Europe 2018*, Manchester, UK. 21 – 24 July 2018. (Communications orale et par affiche).
- **Gondelaud F.**, Miele AE., Duclos B., Ricard-Blum S. Structural characterization of the human ectodomains of syndecans. *Journée Scientifique de l'EDISS*, Villeurbanne, 11 Octobre 2018. (Communication par affiche).

# ANNEXES

## Chapitre IV.1 Caractérisation structurale des ectodomaines des syndécans sans GAGs

**Annexe 1 :** Séquençage des ectodomaines des syndécans humains recombinants exprimés en système procaryote par spectrométrie de masse en tandem (MS-MS).

**Annexe 2 :** Expériences de mobilité ionique visant à déterminer la stabilité des conformations compactes et intermédiaires des ectodomaines des syndécans (EDs).

**Annexe 3 :** Distribution des sections efficaces de collision (CCS) des ectodomaines des syndécans 1-3 et du mutant ED-4 C150A.

**Annexe 4 :** Analyse par SEC-SAXS du dimère de l'ectodomaine du syndécan-3.

## Chapitre IV.3 Interactions des syndécans

**Annexe 5 :** Identification du site d'interaction entre l'ectodomaine du syndécan-1 et l'anastelline par spectrométrie de masse en tandem.

## Chapitre IV.4 Rôle des glycosaminoglycanes et des acides sialiques présents à la surface des globules rouges sains et drépanocytaires dans leurs propriétés rhéologiques

**Annexe 6 :** Indice d'élongation ( $\pm$  écart type) des globules rouges sains et drépanocytaires traités par l'héparinase III.

**Annexe 7 :** Indice d'élongation ( $\pm$  écart type) des globules rouges sains et drépanocytaires traités par la chondroïtinase ABC.

**Annexe 8 :** Indice d'élongation ( $\pm$  écart type) des globules rouges sains et drépanocytaires traités par la hyaluronidase.

**Annexe 9 :** Expériences de docking entre le fibrinogène et l'acide N-acétylneuraminique seul (NANA).

**Annexe 10 :** Résidus du fibrinogène présents à moins de 5 Å de la O-glycosylation de la glycophorine A.

**Annexes du chapitre IV.1**  
**Caractérisation structurale des ectodomains des syndécans sans GAGs**

**ED-1 (Couverture: 76 %)**  
*MGSSHHHHHSSGLVPRGSHMQIVATNLPPEQDGS***GDSDNFSGS**  
**GAGALQDITLSQQTPTWKDTQLLTAIPTSP***PEPTGLEATAASTSTL*  
**PAGEGPK***EGEAVLPEVEPGLTAREQE**ATPRPRE***TTQLP***THLAST*  
**TTATTAQE***PATSHPHRDMQPGHHETSTPAGPSQADLHTPHTEDGGP*  
*SATE***RAAEDGASSQLPAAEGSGEQDFTFETS***GENTAVVAVEPDRRN*  
**QSPVDQ***GATGASQGLLDRKE**VLGDYKDDDDK*

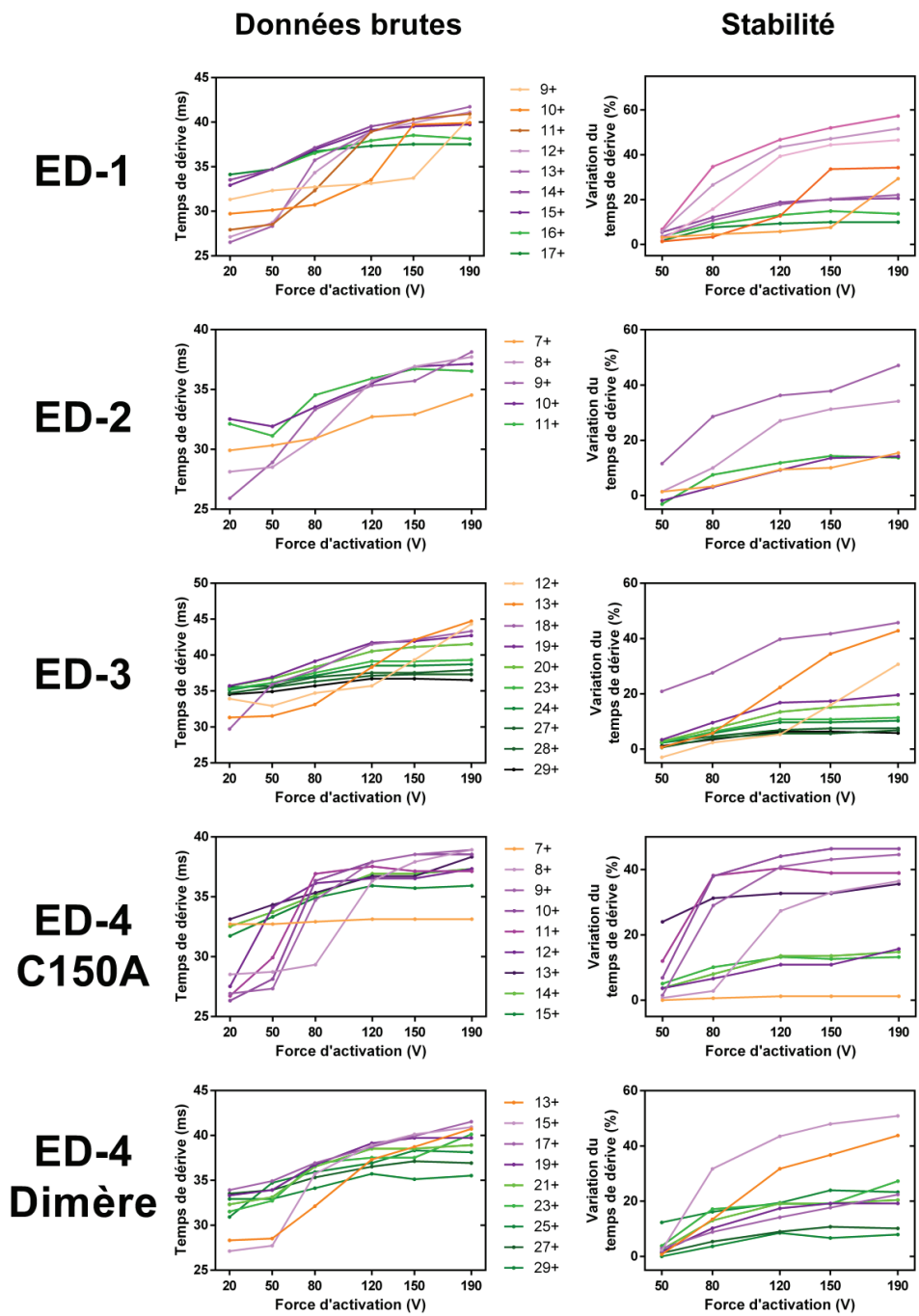
**ED-2 (Couverture: 22 %)**  
*MGSSHHHHHSSGLVPRGSHMESRAELTSDKMYLDNSSIEEASGV*  
*YPIDDDDYASASGSGADEDVESPELTTSRPLPKILL***TS***AAPKVETT*  
**TL***NIQNKIPAQTKSPEETDKEKVHLSDSERKMDPAEEDTNVY***TEKH**  
**SDSLFK***RTEDYKDDDDK*

**ED-3 (Couverture: 43 %)**  
*MGSSHHHHHSSGLVPRGSMQRWRSENFERPVDLEGGDDDSFPD*  
*DELDDLY***SGSGSGYFEQESGIETAMETRFSPDVALAVSTTPAVLPT**  
**TNIQ***PVGT***PFEELP***SERPTLEPATSP**LVVTEVP**EEPSQRATT**VSTT*  
*METATTAATSTGDPTVATVPATVATATPSTPAAPPFTATTAVIRTT*  
*GVRRLPLPL***TTVATARAT***TP***PEAPSPPTTA***AVL**DTEAPT**PRLVSTA*  
*TSRPRALPRPATTQEPDIPERSTLPL***GTTAPGPTEVAQTPTPETFL**  
**TTIRDEPEVPVSGG***PSGDFELPEEETTQPDTANEVVAVGGAAAKAS*  
*SPPGTLPGARPGPGLLDNAIDSGSSAAQLPQKSILERKEVLVDYK*  
**DDDDK**

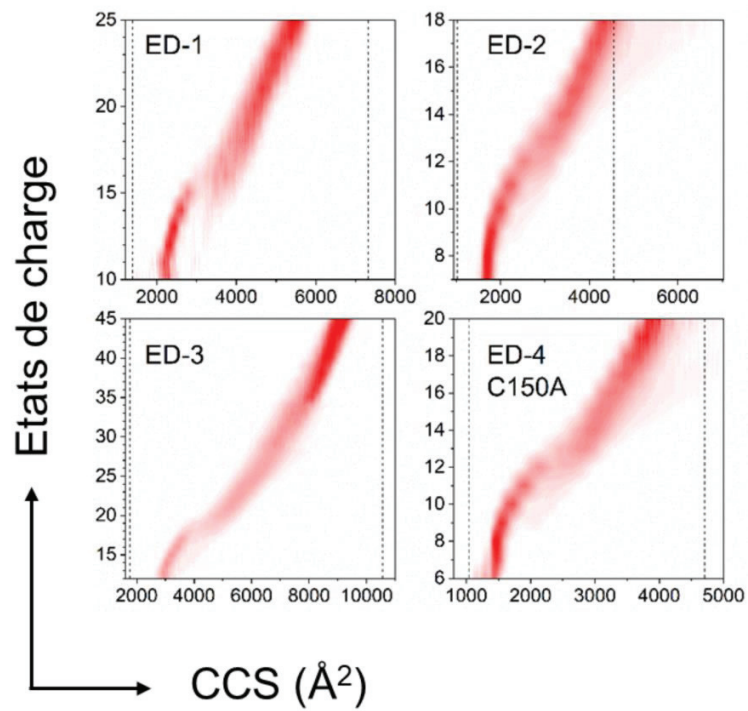
**ED-4 (Couverture: 81 %)**  
*MGSSHHHHHSSGLVPRGSHMESIRETEVIDPQDLLEGRYFSGALP*  
**DDEDVVG***PGQESDDFELSGSGDLDDLEDSMIGPEVVHPLVPLDNHI*  
**PERAGSGSQVPTEPKKLEENEV***IPKRISPVEESEDVSNKVSMSSTV*  
**QGSNIF***ERTE**VLGCPEHDYKDDDDK*

**Annexe 1 : Séquençage des ectodomains des syndécans humains recombinants exprimés en système procaryote par spectrométrie de masse en tandem (MS-MS).** Les peptides identifiés sont en rouge et la séquence des étiquettes 6xHis et FLAG est en italique.

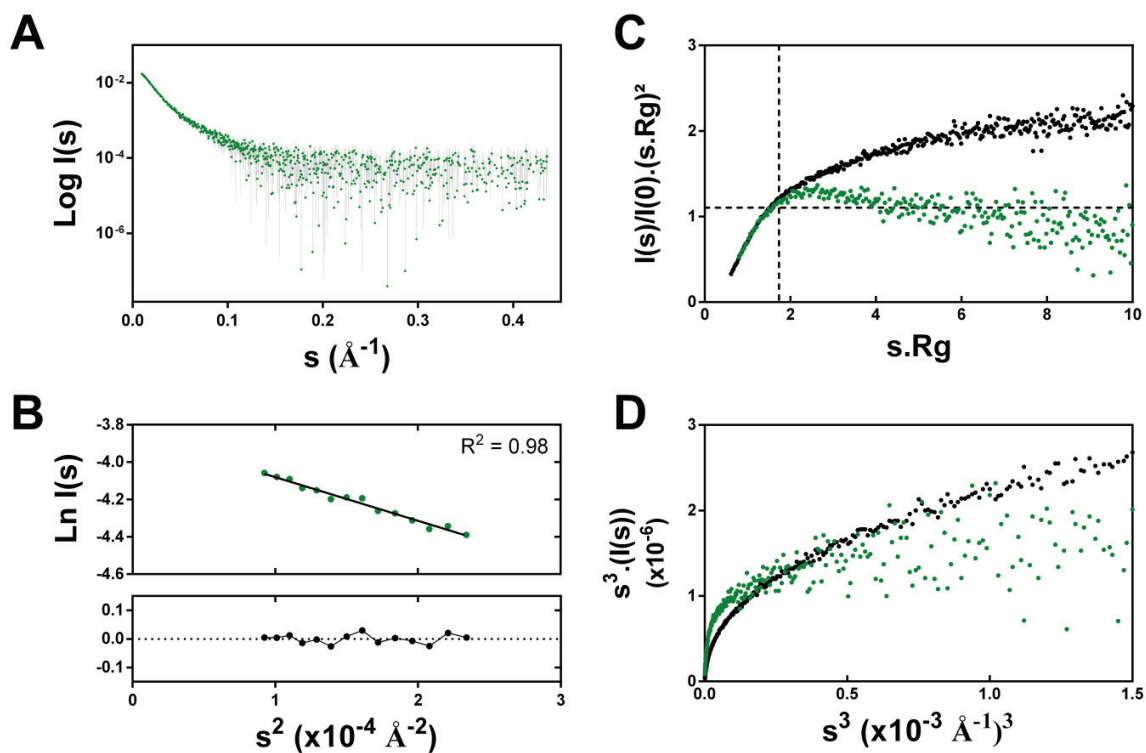




**Annexe 2 : Expériences de mobilité ionique visant à déterminer la stabilité des conformations compactes et intermédiaires des ectodomains des syndécans (EDs).** Pour chaque protéine, les valeurs expérimentales brutes du temps de dérive de chaque état de charge sont présentées (gauche) ainsi que la variation du temps de dérive à 190 V par rapport à celui mesurée à 20 V reflétant la stabilité des conformations (droite). La légende est commune à tous les graphes.

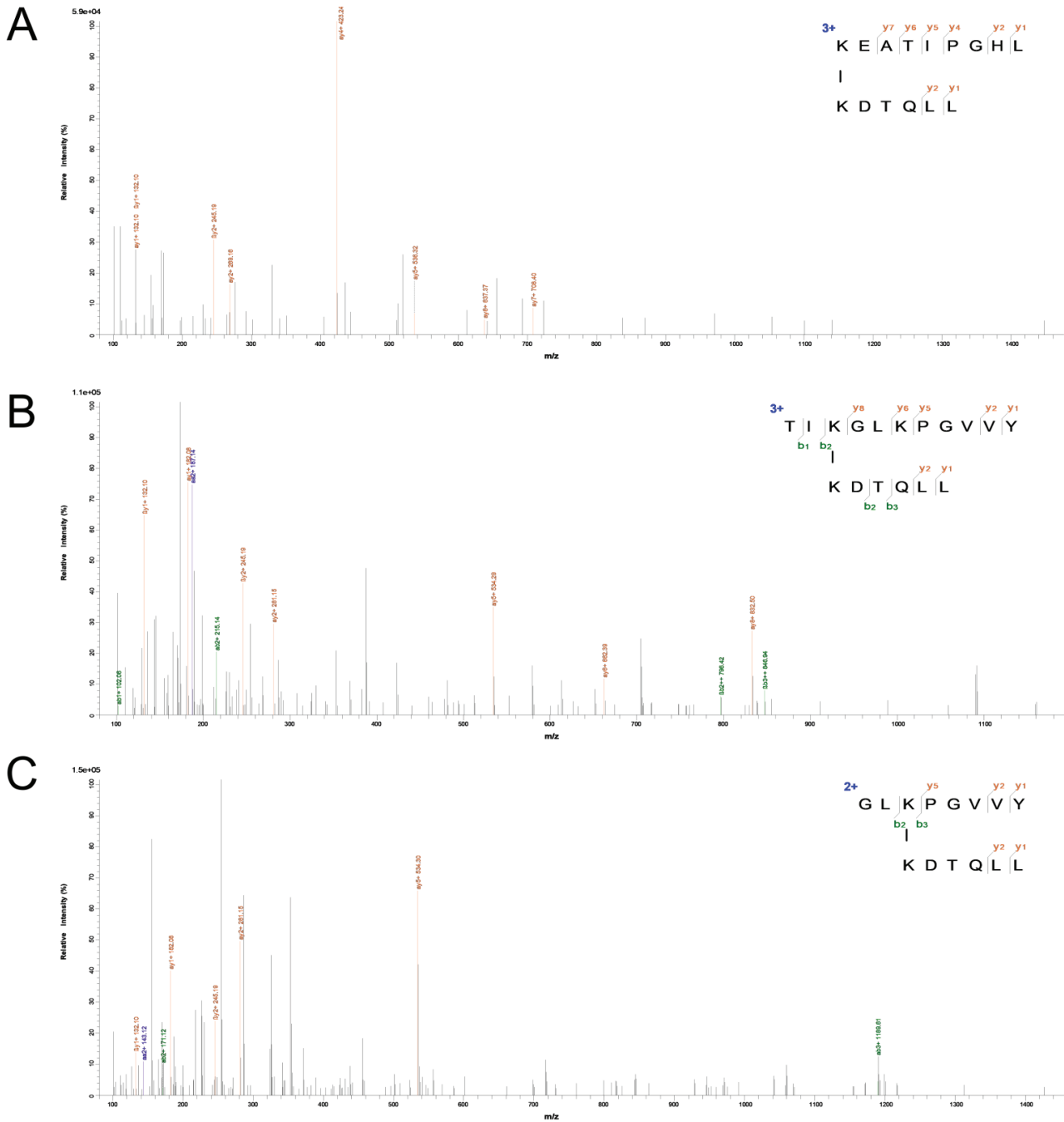


**Annexe 3 : Distribution des sections efficaces de collision (CCS) des ectodomaines des syndécans 1-3 et du mutant ED-4 C150A.** Représentation de la CCS pour chaque état de charge (largeur des bandes) et de leur intensité (dégradé de rouge). ED-4 C150A a été utilisé afin de mimer la forme monomérique d'ED-4. Les lignes verticales en pointillés indiquent les limites théoriques de leur CCS d'après *Beveridge et al.*, (2014).



**Annexe 4 : Analyse par SEC-SAXS du dimère de l'ectodomaine du syndécan-3.** **A)** Intensités de diffusion du dimère d'ED-3 après déconvolution par US-SOMO. **B)** Calcul du rayon de gyration selon l'approximation de Guinier, la qualité de l'ajustement est annotée ( $R^2$ ). Comparaison de la représentation normalisée de Kratky (**C**) et de Sibyls (**D**) entre la forme monomérique (noir) et dimérique (vert).

## Annexe du Chapitre IV.3 Interactions des syndécans



**Annexe 5 : Identification du site d'interaction entre l'ectodomaine du syndécan-1 et l'anastelline par spectrométrie de masse en tandem.** Le peptide d'ED-1 (KDTQLL) est retrouvé réticulé avec l'anastelline sur **A)** le site KEATIPGHL et / ou **B-C)** le site TIKGLKPGVVY ou GLKPGVVY selon le résidu de lysine impliqué dans la réticulation, respectivement.

### Annexes du Chapitre IV.4

#### Rôle des glycosaminoglycanes et des acides sialiques présents à la surface des globules rouges sains et drépanocytaires dans leurs propriétés rhéologiques

<b>GLOBULES ROUGES SAINS (n=8)</b>					
<b>Forces de cisaillement (Pa)</b>	<b>Contrôle</b>	<b>67,5 mU/ml</b>	<b>150 mU/ml</b>	<b>300 mU/ml</b>	<b>450 mU/ml</b>
<b>0,30</b>	0,04 ± 0,01	0,05 ± 0,01	0,04 ± 0,01	0,04 ± 0,01	0,04 ± 0,01
<b>0,53</b>	0,12 ± 0,01	0,12 ± 0,01	0,12 ± 0,01	0,12 ± 0,02	0,12 ± 0,02
<b>0,95</b>	0,23 ± 0,01	0,23 ± 0,02	0,23 ± 0,01	0,23 ± 0,01	0,23 ± 0,01
<b>1,69</b>	0,32 ± 0,01	0,32 ± 0,01	0,32 ± 0,01	0,32 ± 0,01	0,32 ± 0,01
<b>3,00</b>	0,41 ± 0,01	0,41 ± 0,01	0,41 ± 0,01	0,41 ± 0,01	0,41 ± 0,01
<b>5,33</b>	0,49 ± 0,01	0,49 ± 0,01	0,49 ± 0,01	0,49 ± 0,01	0,49 ± 0,01
<b>9,49</b>	0,54 ± 0,01	0,54 ± 0,01	0,54 ± 0,01	0,54 ± 0,01	0,54 ± 0,01
<b>16,87</b>	0,58 ± 0,01	0,58 ± 0,01	0,58 ± 0,01	0,58 ± 0,01	0,58 ± 0,01
<b>30,00</b>	0,61 ± 0,01	0,61 ± 0,01	0,61 ± 0,01	0,61 ± 0,01	0,61 ± 0,01

<b>GLOBULES ROUGES DREPANOCYTAIRES (n=8)</b>					
<b>Forces de cisaillement (Pa)</b>	<b>Contrôle</b>	<b>67,5 mU/ml</b>	<b>150 mU/ml</b>	<b>300 mU/ml</b>	<b>450 mU/ml</b>
<b>0,30</b>	0,10 ± 0,07	0,10 ± 0,08	0,10 ± 0,08	0,09 ± 0,07	0,10 ± 0,08
<b>0,53</b>	0,07 ± 0,02	0,08 ± 0,02	0,08 ± 0,02	0,07 ± 0,02	0,07 ± 0,02
<b>0,95</b>	0,10 ± 0,04	0,10 ± 0,04	0,09 ± 0,04	0,10 ± 0,04	0,09 ± 0,04
<b>1,69</b>	0,15 ± 0,05	0,15 ± 0,05	0,15 ± 0,06	0,15 ± 0,05	0,15 ± 0,05
<b>3,00</b>	0,21 ± 0,07	0,21 ± 0,06	0,21 ± 0,07	0,21 ± 0,07	0,21 ± 0,06
<b>5,33</b>	0,28 ± 0,07	0,27 ± 0,07	0,27 ± 0,08	0,28 ± 0,07	0,28 ± 0,07
<b>9,49</b>	0,34 ± 0,08	0,33 ± 0,08	0,33 ± 0,09	0,33 ± 0,08	0,33 ± 0,08
<b>16,87</b>	0,38 ± 0,09	0,37 ± 0,08	0,37 ± 0,09	0,37 ± 0,09	0,37 ± 0,08
<b>30,00</b>	0,41 ± 0,09	0,40 ± 0,09	0,40 ± 0,09	0,41 ± 0,09	0,41 ± 0,09

**Annexe 6 : Indice d'élongation (± écart type) des globules rouges sains et drépanocytaires traités par l'héparinase III.**

<b>GLOBULES ROUGES SAINS (n=8)</b>				
<b>Forces de cisaillement (Pa)</b>	<b>Contrôle</b>	<b>50 mU/ml</b>	<b>250 mU/ml</b>	<b>500 mU/ml</b>
<b>0,30</b>	0,03 ± 0,02	0,03 ± 0,02	0,031 ± 0,02	0,03 ± 0,02
<b>0,53</b>	0,11 ± 0,03	0,10 ± 0,02	0,11 ± 0,02	0,10 ± 0,03
<b>0,95</b>	0,22 ± 0,02	0,21 ± 0,01	0,21 ± 0,02	0,21 ± 0,02
<b>1,69</b>	0,31 ± 0,02	0,31 ± 0,01	0,31 ± 0,02	0,31 ± 0,02
<b>3,00</b>	0,41 ± 0,01	0,41 ± 0,01	0,41 ± 0,01	0,41 ± 0,01
<b>5,33</b>	0,49 ± 0,01	0,49 ± 0,01	0,49 ± 0,01	0,49 ± 0,01
<b>9,49</b>	0,55 ± 0,01	0,55 ± 0,01	0,55 ± 0,01	0,54 ± 0,02
<b>16,87</b>	0,58 ± 0,01	0,58 ± 0,01	0,58 ± 0,01	0,58 ± 0,02
<b>30,00</b>	0,60 ± 0,01	0,60 ± 0,01	0,60 ± 0,01	0,60 ± 0,01

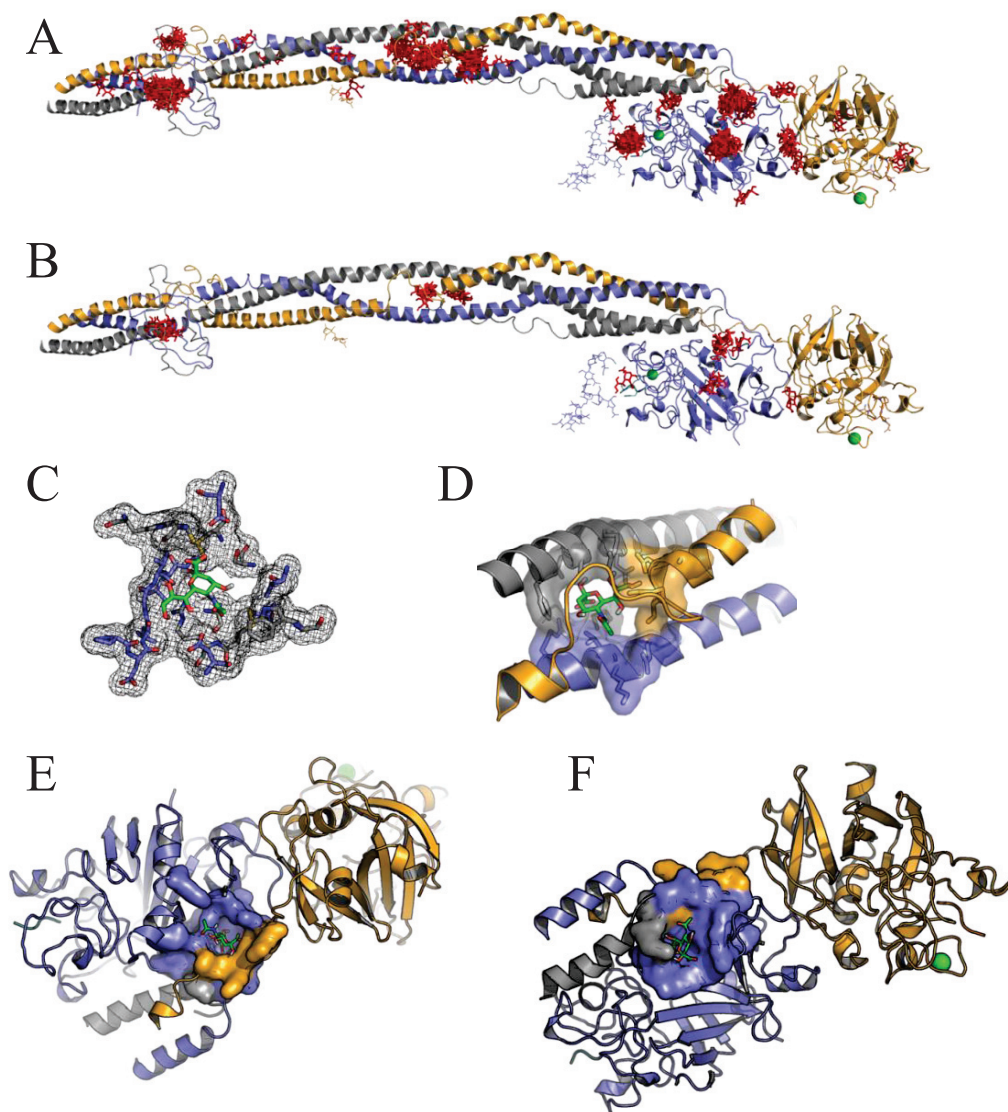
<b>GLOBULES ROUGES DREPANOCYTAIRES (n=6)</b>				
<b>Forces de cisaillement (Pa)</b>	<b>Contrôle</b>	<b>50 mU/ml</b>	<b>250 mU/ml</b>	<b>500 mU/ml</b>
<b>0,30</b>	0,05 ± 0,03	0,07 ± 0,05	0,05 ± 0,04	0,08 ± 0,07
<b>0,53</b>	0,05 ± 0,03	0,05 ± 0,03	0,06 ± 0,03	0,06 ± 0,03
<b>0,95</b>	0,01 ± 0,04	0,01 ± 0,05	0,10 ± 0,04	0,1 ± 0,04
<b>1,69</b>	0,16 ± 0,05	0,16 ± 0,05	0,17 ± 0,06	0,17 ± 0,06
<b>3,00</b>	0,23 ± 0,05	0,23 ± 0,07	0,23 ± 0,07	0,24 ± 0,06
<b>5,33</b>	0,30 ± 0,06	0,3 ± 0,07	0,31 ± 0,07	0,31 ± 0,07
<b>9,49</b>	0,37 ± 0,03	0,3 ± 0,08	0,37 ± 0,08	0,37 ± 0,07
<b>16,87</b>	0,41 ± 0,07	0,41 ± 0,08	0,41 ± 0,08	0,42 ± 0,08
<b>30,00</b>	0,45 ± 0,07	0,44 ± 0,09	0,45 ± 0,08	0,45 ± 0,08

**Annexe 7 : Indice d'élongation (± écart type) des globules rouges sains et drépanocytaires traités par la chondroïtinase ABC.**

<b>GLOBULES ROUGES SAINS (n=6)</b>				
<b>Forces de cisaillement (Pa)</b>	<b>Contrôle</b>	<b>100 mU/ml</b>	<b>250 mU/ml</b>	<b>500 mU/ml</b>
<b>0,30</b>	0,10 ± 0,01	0,09 ± 0,02	0,10 ± 0,01	0,10 ± 0,03
<b>0,53</b>	0,13 ± 0,02	0,12 ± 0,02	0,13 ± 0,02	0,14 ± 0,02
<b>0,95</b>	0,19 ± 0,03	0,20 ± 0,02	0,19 ± 0,02	0,21 ± 0,02
<b>1,69</b>	0,29 ± 0,03	0,30 ± 0,02	0,29 ± 0,02	0,31 ± 0,02
<b>3,00</b>	0,38 ± 0,03	0,38 ± 0,02	0,39 ± 0,02	0,4 ± 0,01
<b>5,33</b>	0,46 ± 0,02	0,46 ± 0,01	0,46 ± 0,01	0,47 ± 0,01
<b>9,49</b>	0,52 ± 0,02	0,52 ± 0,01	0,52 ± 0,01	0,53 ± 0,01
<b>16,87</b>	0,56 ± 0,02	0,56 ± 0,01	0,56 ± 0,01	0,57 ± 0,01
<b>30,00</b>	0,59 ± 0,01	0,59 ± 0,01	0,59 ± 0,01	0,60 ± 0,01

<b>GLOBULES ROUGES DREPANOCYTAIRES (n=7)</b>				
<b>Forces de cisaillement (Pa)</b>	<b>Contrôle</b>	<b>100 mU/ml</b>	<b>250 mU/ml</b>	<b>500 mU/ml</b>
<b>0,30</b>	0,0745 ± 0,05	0,10 ± 0,06	0,07 ± 0,05	0,08 ± 0,04
<b>0,53</b>	0,06 ± 0,03	0,07 ± 0,02	0,07 ± 0,02	0,08 ± 0,03
<b>0,95</b>	0,10 ± 0,04	0,10 ± 0,04	0,09 ± 0,03	0,10 ± 0,04
<b>1,69</b>	0,16 ± 0,06	0,16 ± 0,05	0,16 ± 0,05	0,16 ± 0,06
<b>3,00</b>	0,23 ± 0,07	0,23 ± 0,07	0,23 ± 0,06	0,23 ± 0,07
<b>5,33</b>	0,30 ± 0,08	0,30 ± 0,07	0,30 ± 0,07	0,30 ± 0,08
<b>9,49</b>	0,36 ± 0,08	0,37 ± 0,07	0,36 ± 0,08	0,36 ± 0,08
<b>16,87</b>	0,41 ± 0,08	0,42 ± 0,07	0,41 ± 0,08	0,41 ± 0,08
<b>30,00</b>	0,45 ± 0,08	0,46 ± 0,06	0,45 ± 0,07	0,45 ± 0,08

**Annexe 8 : Indice d'élongation (± écart type) des globules rouges sains et drépanocytaires traités par la hyaluronidase.**



**Annexe 9 : Expériences de docking entre le fibrinogène et l'acide N-acétylneuraminique seul (NANA).** **A)** Superposition des 200 poses (rouge) de NANA sur le monomère de fibrinogène (chaînes  $\alpha$  : grise,  $\beta$  : bleue,  $\gamma$  : orange). Ion calcium : sphère verte. **B)** Seules les 5 meilleures poses ayant les énergies les plus faibles pour chaque zone testée sont représentées mettant en avant 4 sites préférentiels d'interaction. La légende est identique à A). **C-F)** Agrandissement sur les 4 sites préférentiels de la meilleure solution. NANA : vert. La couleur des chaînes est la même qu'en A) et B).



Sites	Résidus	Gamme des énergies d'interaction (kcal/mol)
Site 1	<p><b>Chaîne α :</b>            Chaîne G : A<sup>27</sup>C<sup>28</sup>K<sup>29</sup>W<sup>33</sup>S<sup>47</sup>R<sup>50</sup>            Chaîne J : A<sup>27</sup>C<sup>28</sup>K<sup>29</sup>D<sup>30</sup>S<sup>31</sup>W<sup>33</sup>S<sup>47</sup>C<sup>49</sup>R<sup>50</sup></p> <p><b>Chaîne β :</b>            Chain H : D<sup>61</sup>P<sup>70</sup>V<sup>74</sup>C<sup>76</sup>            Chain K : K<sup>58</sup>A<sup>59</sup>P<sup>60</sup>D<sup>61</sup>L<sup>66</sup>P<sup>70</sup>D<sup>71</sup>L<sup>72</sup>G<sup>73</sup>V<sup>74</sup>C<sup>76</sup></p>	- 6,4 / - 8,3
Site 2	<p><b>Chaîne α :</b>            Chaîne J : G<sup>96</sup>D<sup>97</sup>F<sup>98</sup>S<sup>100</sup>A<sup>101</sup>N<sup>102</sup>R<sup>104</sup>D<sup>105</sup>N<sup>106,205</sup>M<sup>207</sup></p> <p><b>Chaîne β :</b>            Chaîne K : W<sup>125</sup>Q<sup>129</sup>Q<sup>131</sup>V<sup>132</sup>K<sup>133</sup>N<sup>135</sup>E<sup>136</sup>V<sup>138</sup>V<sup>139</sup></p> <p><b>Chaîne γ :</b>            Chaîne L : T<sup>67</sup>Y<sup>68</sup>N<sup>69</sup>P<sup>70</sup>D<sup>71</sup>E<sup>72</sup>S<sup>73</sup>M<sup>78,79</sup>D<sup>80</sup></p>	- 6,2 / - 8,5
Site 3	<p><b>Chaîne β :</b>            Chaîne K : Q<sup>359</sup>L<sup>360</sup>M<sup>361</sup>N<sup>364</sup>E<sup>397</sup>T<sup>423</sup>W<sup>424</sup>G<sup>430</sup>T<sup>431</sup>D<sup>432</sup>M<sup>438</sup>S<sup>443</sup>W<sup>444</sup>            NAG NDG NDG GAL SIA            Chaîne T : G<sup>1</sup>H<sup>2</sup>R<sup>3</sup>P<sup>4</sup></p>	- 6,2 / - 7,2
Site 4	<p><b>Chaîne β :</b>            Chaîne K : L<sup>226</sup>K<sup>234</sup>P<sup>235</sup>Y<sup>236</sup>R<sup>237</sup>R<sup>264</sup>D<sup>267</sup>P<sup>268</sup>Q<sup>271</sup>G<sup>272</sup>F<sup>273</sup>G<sup>274</sup>N<sup>275</sup>V<sup>276</sup>K<sup>298</sup></p> <p><b>Chaîne γ :</b>            Chaîne L : Q<sup>136</sup>E<sup>137</sup>P<sup>138</sup>C<sup>139</sup>D<sup>141</sup>V<sup>143</sup>Q<sup>144</sup>I<sup>145</sup>D<sup>147</sup>F<sup>168</sup></p>	- 6,3 / - 7,5

**Annexe 10 : Résidus du fibrinogène présents à moins de 5 Å de la O-glycosylation de la glycophorine A.** La meilleure solution a été utilisée afin de déterminer les résidus les plus proches. La nomenclature est la même que celle présente dans la structure tridimensionnelle du fibrinogène (PDB : 3GHG). NAG : N-acétyl-D-glucosamine, NDG : 2-(acétylamino)-2-deoxy- $\alpha$ -D-glucopyranose, GAL :  $\beta$ -D-galactose, SIA : acide sialique. Les gammes d'énergies d'interaction indiquent la meilleure et la pire énergie pour chaque site.

## SUMMARY

### **Characterization of an intrinsically disordered proteins family, the syndecans**

Syndecans are a family of four transmembrane proteoglycans which are expressed at the cell surface of virtually all cell types and involved in many biological functions. Their ectodomains can be shed from the cell surface and they have not been structurally characterized so far. Using several biophysical approaches, we found that the core protein of the human syndecans ectodomains is enriched in intrinsic disorder and contains a small proportion of  $\beta$ -strands. Syndecans ectodomains behave like random coil proteins and are highly extended and flexible, allowing them to adopt compact conformations. Additionally, we have identified two novel partners by biolayer interferometry, the fibroblast growth factor 2 and anastellin, a bioactive fragment of fibronectin, and we have characterized the complexes.

Moreover, we have constructed and analyzed the interaction network of syndecans. Their interactome revealed that they interact with more than 347 partners, mainly thanks to their heparan sulfate chains and we defined their consensus interactome.

Finally, a second research axis dealt with the role of the glycosaminoglycans and the sialic acids present at the surface of red blood cells in their rheological properties. We found that sialic acid are implicated in the aggregation and the deformability of healthy and sickle red blood cells. Particularly, we highlighted the role of fibrinogen in their aggregation.

**Key Words:** syndecans, intrinsic disorder, biophysics, biomolecular interactions.

## RÉSUMÉ DES TRAVAUX

Les syndécans sont une famille de quatre protéoglycanes transmembranaires exprimés de façon quasi-ubiquitaire à la surface cellulaire où ils possèdent un nombre très important de fonctions biologiques. Leurs ectodomains peuvent être libérés dans l'espace extracellulaire et aucune caractérisation biophysique et structurale n'a été effectuée jusqu'à présent. Nous avons montré à l'aide de différentes méthodes biophysiques que la partie protéique des ectodomains des syndécans humains contient une faible proportion de brins  $\beta$  et est enrichie en désordre intrinsèque. Les ectodomains se comportent comme des random coils ayant des conformations très étendues et très flexibles, susceptibles d'adopter des conformations compactes. Leur espace conformationnel pourrait être modulé par leurs interactions avec d'autres protéines et nous avons identifié par interférométrie de biocouche deux nouveaux partenaires des ectodomains des syndécans, le facteur de croissance fibroblastique 2 et l'anastelline, et nous avons caractérisé les complexes formés.

Nous avons ensuite construit et analysé l'interactome des syndécans ce qui a permis de recenser 347 partenaires avec lesquels ils interagissent majoritairement *via* leurs chaînes d'héparane sulfate, et de définir leur interactome consensus.

Un second axe de recherche portant sur le rôle des glycosaminoglycanes et des acides sialiques situés à la surface des globules rouges sur leurs propriétés rhéologiques a montré le rôle des acides sialiques dans la déformabilité et l'agrégation des globules rouges de sujets sains et de patients atteints de drépanocytose. Nos expériences montrent également le rôle du fibrinogène dans l'agrégation des globules rouges.

**Mots – Clefs :** syndécans, désordre intrinsèque, biophysique, interactions biomoléculaires.