Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER EN SCIENCES CHIMIQUE À FINALITÉ APPROFONDIE

Caractérisation des interactions protéine-protéine dans le complexe DPF3-PLK1 impliqué dans la mitose cellulaire

Mignon, Julien

Award date:
2021

Awarding institution:
Universite de Namur

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Jul. 2025

Université de Namur

Faculté des Sciences

Laboratoire de Chimie Physique des Biomolécules

Caractérisation des interactions protéine-protéine dans le complexe DPF3-PLK1 impliqué dans la mitose cellulaire

Mémoire présenté pour l'obtention

du grade académique de Master Chimie «Chimie du Vivant et des Nanomatériaux» : Finalité Approfondie

Julien MIGNON

Janvier 2021

Table des matières

Propos liminaire 1
I. Résumé 2
II. Remerciements 3
III. Liste des abréviations 5

Partie n°1 – Introduction et contextualisation 8
I. La mitose dans le cycle cellulaire et le développement des cellules cancéreuses 9
A. Cycle cellulaire eucaryote 9
B. Phases de la mitose 10
C. Mitose, apoptose et cancer 13
II. Interactions protéine-protéine 15
III. Description des partenaires protéiques 17
A. La sérine/thréonine kinase PLK1 17
B. La protéine nucléaire à doigts de zinc DPF3 21
IV. Hypothèse de la formation du complexe DPF3-PLK1 26
V. Hypothèse de la nature désordonnée de la DPF3 29
VI. Protéines intrinsèquement désordonnées 31
A. Introduction au concept de désordre intrinsèque 31
B. Propriétés biophysiques et méthodes de caractérisation 31
C. Fonctions biologiques cellulaires 33
D. Pathologies associées et voies thérapeutiques 34

Partie n°2 – Objectifs et stratégies 36
I. Objectifs de recherche 37
II. Stratégies de production et de purification 37
III. Stratégies de caractérisation 38

Partie n°3 – Résultats et discussions 41
Section a – Purification et caractérisation des protéines de fusion 42
I. Transformations et surexpressions 43
II. Purifications par chromatographie d'affinité sur colonne d'écoulement par gravité 44
A. Vérification de la spécificité du processus de purification 44
B. Sensibilité à la protéolyse 46
III. Purifications par FPLC 49

IV. Caractérisation biophysique de la protéine de fusion DPF3a-GST 51
A. Spectroscopie d'absorption UV-visible 51
B. Fluorescence intrinsèque du tryptophane (ITF) 53
C. Dichroïsme circulaire (CD) 54
D. Diffusion dynamique de la lumière (DLS) 56
E. Chromatographie d'exclusion stérique (SEC) 57
Section b – Purification et caractérisation de la DPF3a clivée 60
I. Clivages par chromatographie d'affinité au format *batch* 61
II. Clivages par FPLC et comportement singulier par SDS-PAGE 62
III. Identification protéique par spectrométrie de masse (LC-ESI-MS/MS) 65
IV. Comportement singulier en spectroscopie d'absorption 66
V. Investigation de l'hypothèse de contamination par l'ADN 69
A. Traitement à la DNase I 69
B. Électrophorèse sur gel d'agarose (AGE) 71
VI. Investigation de l'hypothèse d'agrégation protéique 72
A. Stabilisation de protéine en solution par le glycérol 72
B. Comportement en spectroscopie d'émission 74
1. Fluorescence intrinsèque du tryptophane (ITF) 74
2. Autofluorescence dans le bleu profond (dbAF) 76
C. Dichroïsme circulaire (CD) 77
D. Diffusion dynamique de la lumière (DLS) 79
E. Chromatographie d'exclusion stérique (SEC) 81
VII. Investigation de l'hypothèse de complexation métal-protéine 82
A. Influence des ions zinciques sur les propriétés d'absorption 82
B. Influence des ions zinciques sur les propriétés d'émission (ITF) 84
Section c – Prédiction de propriétés associées au désordre 86
I. Analyses de composition de séquence 87
II. Acides aminés promoteurs de désordre 90
III. Hydropathie et charge nette moyennes 92
IV. Diagramme de phase de Das-Pappu 94

Partie n°4 – Conclusions et perspectives 98
I. Conclusions 99
II. Perspectives 102

Partie n°5 – Matériels et méthodes 105
I. Référence et composition des tampons 106
II. Transformation de souches compétentes de *E. coli* pour l'expression d'une protéine de fusion marquée à la GST 106
A. DPF3, isoformes et domaines associés 106
B. PLK1 et domaines associés 107
C. Stockage à long terme de souches de *E. coli* efficacement transformées 108
III. Surexpression hétérologue dans *E. coli* d'une protéine de fusion marquée à la GST 108
IV. Lyse par sonication de souches de *E. coli* induites à l'IPTG 109
V. Méthodes de purification et de clivage en phase liquide d'une protéine de fusion marquée à la GST 110
A. Purification sur colonne d'écoulement par gravité 110
B. Purification au format *batch* 111
C. Purification par FPLC sur colonnes pré-packées 112
VI. Dessalage et échange de tampon sur colonne PD-10 d'une protéine purifiée 115
VII. Analyse SDS-PAGE d'échantillons protéiques 115
VIII. Méthodes de caractérisation biophysiques 116
A. Spectrophotométrie UV-visible 116
B. Fluorimétrie (ITF et dbAF) 117
C. Spectrométrie de masse (LC-ESI-MS/MS) 117
D. Diffusion dynamique de la lumière (DLS) 118
E. Dichroïsme circulaire (CD) 118
F. Chromatographie d'exclusion stérique (SEC) 119

Partie n°6 – Bibliographie 121

Propos liminaire

UNIVERSITE DE NAMUR
Faculté des Sciences
Secrétariat du Département de Chimie
Rue de Bruxelles 61 – 5000 NAMUR
Téléphone : +32(0)81 72.54.44 – Téléfax : +32(0)81 72.54.40
E-mail : enseignement.chimie@unamur.be - www.unamur.be/sciences

Caractérisation des interactions protéine-protéine dans le complexe DPF3-PLK1 impliqué dans la mitose cellulaire

MIGNON Julien

I. Résumé

La régulation des phases mitotiques du cycle cellulaire est contrôlée par de nombreux acteurs protéiques en interaction physique (PPI). Dans ce contexte, le complexe DPF3-PLK1 est considéré comme un régulateur essentiel de la mitose de par son implication dans la séparation des centrosomes et l'attache des kinétochores aux microtubules. Par conséquent, la compréhension de l'interaction préférentielle entre ces deux partenaires pourra conduire non seulement à une meilleure appréhension des mécanismes régissant la mitose mais également au développement d'inhibiteurs spécifiques, autrement dit de nouveaux antimitotiques. Ce projet de recherche s'est focalisé sur la caractérisation des propriétés biophysiques du second isoforme de la DPF3 : DPF3a. Les analyses combinées en spectroscopies d'absorption et d'émission, dichroïsme circulaire (CD), diffusion dynamique de la lumière et en chromatographie d'exclusion stérique ont mis en évidence une importante propension de la DPF3a à l'agrégation. En outre, l'émergence d'une autofluorescence intrinsèque dans le bleu profond laisse à penser qu'elle s'assemblerait en fibrilles amyloïdes. Au-delà de cette tendance, les spectroscopies d'absorption et d'émission ont également dévoilé une sensibilité aux ions Zn^{2+} et une signature typiquement désordonnée (*random coil*) a été détectée par CD. La DPF3a appartiendrait à la classe des protéines intrinsèquement désordonnées (IDPs), connues pour leur caractère multifonctionnel en l'absence d'une structure tridimensionnelle définie. En effet, l'analyse de sa structure primaire montre qu'elle en possède les traits caractéristiques : une faible complexité de composition de séquence, un enrichissement en résidus promoteurs de désordre ainsi qu'une faible distribution charge-hydropathie. Étant donné l'implication des IDPs dans de nombreuses pathologies graves et celle des fibrilles amyloïdes dans les maladies neurodégénératives, un meilleur portrait du comportement *in vitro* de la DPF3a peut se révéler clef dans la conception d'agents thérapeutiques capables de cibler les protéines désordonnées présentant une prédisposition à l'amylose.

Mémoire de master en Sciences Chimiques à Finalité Approfondie

Janvier 2021

Promoteur : C. Michaux

Co-promoteur : E. A. Perpète

II. Remerciements

Je souhaite en tout premier lieu remercier très chaleureusement ma promotrice Catherine Michaux ainsi que mon co-promoteur Éric Perpète du Laboratoire de Chimie Physique des Biomolécules (CPB) sans qui mon éveil à la recherche fondamentale n'aurait pas eu la même saveur. Au-delà de l'infrastructure offerte par le laboratoire, je souhaite souligner non seulement toute l'attention, la bienveillance, l'altruisme et l'humanisme dont j'ai pu profiter mais également la rigueur, la richesse et le plaisir des discussions échangées ensemble qu'elles fussent derrière une paillasse ou un verre de Rochefort.

Je n'oublie évidemment pas les autres acteurs essentiels du CPB. C'est pourquoi, j'adresse mes remerciements non seulement à Virgile, l'unique (pour le moment) et vaillant doctorant du laboratoire, pour son accueil, son humour et sa générosité (le « Vautourion » ne vole jamais très loin) mais également à mes compagnons de recherche Amandine, Blinera et Laurelenn grâce à qui le mémoire est devenu une expérience inoubliable et incroyablement épanouissante. Je suis heureux d'avoir pu participer avec vous trois à la fondation d'une trame solide pour le laboratoire qui, j'en formule le souhait, motivera davantage les futurs mémorants à tenter leur aventure au CPB car le jeu en vaut la chandelle.

Cette expérience de mémoire ne serait complète sans mentionner nos chers camarades électrochimistes du Laboratoire de Chimie et d'Électrochimie des Surfaces (CES), à savoir Coco, Géraldine, Louis-Charles, Marie et Maxime, avec qui je garde des souvenirs et anecdotes mémorables de cette vie pendant une année au quatrième étage. Aussi, merci pour le pain d'épices, les pains au chocolat, les barres de Toblerone et l'Amaretto !

Mes remerciements vont également aux autres étudiants en seconde année de master en sciences chimiques dont j'ai partagé, depuis le bachelier, la vie universitaire, les galères, les colères, les gueules de bois, les fous rires, les moments de joie et les réussites. Nous formons un noyau dur et solidaire dont les amitiés, je l'espère, ne s'éroderont pas trop vite avec le temps.

Dans le cadre de ce mémoire, mes remerciements vont également à l'ensemble des collaborateurs avec qui j'ai eu l'occasion de travailler ou de discuter. Au Dr. Denis Mottet et son équipe du GIGA pour son approche biologique et sans qui ce sujet de recherche sur les protéines DPF3 et PLK1 n'existerait probablement pas.

Au Pr. Vladimir N. Uversky pour sa sollicitude, sa connaissance dans le domaine des protéines intrinsèquement désordonnées et pour m'avoir ouvert à cet univers si singulier.

Au Pr. André Matagne de l'ULiège pour l'utilisation du spectromètre de dichroïsme circulaire.

Aux membres de la plateforme de spectrométrie de masse de l'UNamur (MaSUN), en particulier Marc Dieu, Catherine Demazy et Maude Fransolet pour leur grande disponibilité, les nombreuses analyses réalisées par leurs soins et les discussions qui en ont découlé.

À l'Unité de Recherche en Biologie des Microorganismes (URBM) pour les facilités de culture bactérienne ainsi qu'à certains de ses membres, en particulier Aurélie Mayard et Françoise Tilquin pour leur accessibilité, leurs conseils, les échantillons d'ADN et les électrophorèses sur gel d'agarose.

De manière générale, un tout grand merci est adressé à l'ensemble de mes proches et amis, qu'ils soient chimistes, scouts, Floriffoliens, Soyens, de Namur, de Gembloux ou de ma belle région floreffoise, sans qui la vie sur cette bonne vieille Terre serait bien moins festive et palpitante.

Je tiens à signifier toute ma reconnaissance et ma tendresse envers mes parents, mon petit frère, mes grands-parents ainsi qu'à toute la famille Mignon pour le soutien (et les petits plats) qu'ils m'ont constamment apporté au cours de mes études en chimie ainsi que l'intérêt sincère qu'ils y ont porté même si je n'ai pas toujours été bon vulgarisateur. Je remercie particulièrement mes parents pour m'avoir permis et offert l'opportunité d'entreprendre ces magnifiques et folkloriques études dont la finalité est concrétisée par ce manuscrit.

Pour conclure cette section, j'aimerais tout spécialement exprimer ma gratitude envers mon professeur de chimie du secondaire Luc Daiche pour m'avoir fait découvrir la chimie, me l'avoir faite aimer (et je l'aime encore) et pour m'avoir initié à l'expérimentation ainsi qu'à la méthode scientifique. Si j'en suis ici aujourd'hui, après cinq années d'études dans les sciences chimiques et un premier travail de recherche accouché sur le papier, c'est en grande partie grâce à lui.

Namur, le 17 décembre 2020

Julien Mignon

III. Liste des abréviations

À des fins de confort de lecture, l'ensemble des abréviations détaillées ci-après (sans leurs traductions et/ou origines anglaises) peuvent être retrouvées dans le dépliant attaché en fin de manuscrit.

3D : tridimensionnel
α-syn : alpha-synucléine
ADN : acide désoxyribonucléique
AGE : électrophorèse sur gel d'agarose (*agarose gel electrophoresis*)
APS : persulfate d'ammonium (*ammonium persulfate*)
ARN : acide ribonucléique
ATP : adénosine triphosphate
BAF : facteurs associés à BRG1 (*BRG1-associated factors*)
BAF45C : facteur 45C associé à BRG1 (*BRG1-associated factor 45C*)
BB : tampon de fixation (*binding buffer*)
BRG1 : gène 1 relatif au complexe Brahma (*Brahma-relative gene 1*)
BSA : albumine de sérum bovin (*bovine serum albumin*)
BSB : tampon acétate de stockage de la benzamidine (*benzamidine storage buffer*)
C_2H_2 : doigt de zinc de type Krüppel (*Krüppel-type zinc finger*)
CD : dichroïsme circulaire (*circular dichroism*)
CDK : kinase dépendante de cycline (*cyclin-dependent kinase*)
dbAF : autofluorescence dans le bleu profond (*deep-blue autofluorescence*)
DLS : diffusion dynamique de la lumière (*dynamic light scattering*)
DNase I : désoxyribonucléase I
DO : densité optique
DPF : doubles doigts PHD (*double PHD fingers*)
DPF3 : doubles doigts PHD 3 (*double PHD fingers 3*)
DPR : résidu promoteur de désordre (*disorder-promoting residue*)
DRX : diffraction des rayons X
E. coli : *Escherichia coli*
EB : tampon d'élution de la GST (*elution buffer*)
EMD : émerine
EtBr : bromure d'éthidium
FASP : *filter-aided sample preparation*
FB : tampon d'élution au facteur Xa (*factor Xa buffer*)

FEB : tampon d'élution du facteur Xa (*factor Xa elution buffer*)
FL : *full-length*
FPLC : chromatographie en phase liquide de protéine rapide (*fast protein liquid chromatography*)
FT : *flow-through*
GIGA : Grappe Interdisciplinaire de Génoprotéomique Appliquée
GS4B : Glutathione Sepharose™ 4B
GSH : glutathion réduit
GSSG : glutathion oxydé
GST : glutathion S-transférase
IDP : protéine intrinsèquement désordonnée (*intrinsically disordered protein*)
IDPR : protéine avec des régions intrinsèquement désordonnées (*protein with intrinsically disordered regions*)
IDR : région intrinsèquement désordonnée (*intrinsically disordered region*)
IPTG : isopropyl β-D-1-thiogalactopyranoside
ITC : titrage calorimétrique isotherme (*isothermal titration calorimetry*)
ITF : fluorescence intrinsèque du tryptophane (*intrinsic tryptophan fluorescence*)
IUP : protéine intrinsèquement déstructurée (*intrinsically unstructured protein*)
KIN : domaine kinase
LB Lennox™ : bouillon de culture lysogène Lennox™ (*lysogeny broth Lennox™*)
LB : tampon de lyse (*lysis buffer*)
LC-ESI-MS/MS : spectrométrie de masse en tandem par ionisation en électrospray couplée à une chromatographie en phase liquide (*liquid chromatography-electrospray ionisation-tandem mass spectrometry*)
LMCT : transfert de charge du ligand au métal (*ligand to metal charge transfer*)
Lys C : lysozyme C du blanc d'œuf de poule
MaSUN : Plateforme Technologique de Spectrométrie de Masse de l'Université de Namur
MT : microtubule
MWSM : marqueur de masse moléculaire (*molecular-weight size marker*)
NDP : protéine nativement dénaturée (*natively denatured protein*)
NLS : signal de localisation nucléaire (*nuclear localisation signal*)
NPR : résidu non-promoteur (*non-promoting residue*)
NUP : protéine nativement dépliée (*natively unfolded protein*)
o/n : durant la nuit (*overnight*)

OPN : ostéopontine
OPR : résidu promoteur d'ordre (*order-promoting residue*)
ORDP : protéine ordonnée (*ordered protein*)
PB : tampon phosphate (*phosphate buffer*)
PBD (ou PB) : domaine *polo-box* (*polo-box domain*)
PBS : tampon phosphate salin (*phosphate buffered saline*)
PDB : *Protein Data Bank*
PES : poléthersulfone
PHD : homéodomaine végétal (*plant homeodomain*)
PLK : *polo-like kinase*
PLK1 : *polo-like kinase 1*
PMSF : fluorure de phénylméhtylsulfonyle (*phenylmethylsulfonyl fluoride*)
PPI : interaction protéine-protéine (*protein-protein interaction*)
PTM : modification post-traductionnelle (*post-translational modification*)
RC : cellulose recomposée (*recomposed cellulose*)
RD : résidu désordonné
RMN : résonance magnétique nucléaire
SAXS : diffusion des rayons X aux petits angles (*small-angle X-rays scattering*)
S. japonicum : *Schistosoma japonicum*
SDS : dodécylsulfate de sodium (*sodium dodecyl sulfate*)
SDS-PAGE : électrophorèse au dodécylsulfate de sodium sur gel de polyacrylamide (*sodium dodecyl sulfate-polyacrylamide gel electrophoresis*)
SEC : chromatographie d'exclusion stérique (*steric exclusion chromatography* ou *size-exclusion chromatography*)
SEM : microscopie électronique à balayage (*scanning electron microscopy*)
SWI/SNF : *switch/sucrose non-fermentable*
TCS : taux de composition de séquence
TEMED : tétraméthyléthylènediamine
ThT : thioflavine T
TMC : machinerie de transcription (*transcriptional machinery complex*)
TOF : Tétralogie de Fallot (*Tetralogy of Fallot*)
Tris : tris(hydroxyméthyl)aminométhane
ZNF : doigt de zinc (*zinc finger*)

Partie n°1

—

Introduction et contextualisation

I. La mitose dans le cycle cellulaire et le développement des cellules cancéreuses

A. Cycle cellulaire eucaryote

Le cycle cellulaire correspond aux différentes étapes de la vie d'une cellule au bout desquelles elle procède à sa division en deux cellules-filles identiques. Chez les eucaryotes, ce cycle comprend deux grandes phases successives : l'interphase suivie de la mitose (cf. Fig. 1). L'interphase correspond à l'étape de croissance de la cellule au cours de laquelle est également répliqué le matériel génétique.

Elle est elle-même subdivisée en trois phases intermédiaires, respectivement appelées G1, S et G2. La phase G1 (pour *gap 1*) équivaut à l'intervalle (d'où la dénomination *gap*) entre la mitose précédente et l'amorçage de la réplication de l'acide désoxyribonucléique (abrégé ADN) durant lequel la cellule débute sa croissance. Dès lors que la cellule atteint une taille seuil, elle entre dans la phase S (pour synthèse) durant laquelle le matériel génétique est dupliqué par réplication (ou synthèse) de l'ADN. Les centrosomes le sont également durant la phase S. La synthèse de l'ADN et des centrosomes ainsi que la croissance cellulaire s'achèvent à la phase G2 (pour *gap 2*) durant laquelle sont également produites les protéines nécessaires au bon déroulement de la mitose. Les erreurs de réplication sont également corrigées au cours de la phase G2 avant d'entrer dans la phase suivante.

La phase G2 est directement suivie de la phase M correspondant à la mitose au cours de laquelle les chromosomes dupliqués (chromatides) sont répartis entre les deux cellules-filles séparées en fin du processus de division par cytokinèse. En fin de mitose, le cycle reprend à la phase G1.[1–3] La phase M équivaut bel et bien à la mitose dans le cas de la division cellulaire mais elle fait référence à la méiose dans le cas de la production des gamètes, les cellules sexuelles haploïdes.[4]

La prolifération des cellules est contrôlée par l'action conjointe de signaux extracellulaires et de régulateurs intracellulaires intervenant aux différentes étapes du cycle ou qui en autorisent la poursuite par transition de phase. Cette régulation s'opère via différents modes de modifications post-traductionnelles (abrégées PMTs pour *post-translational modifications*) à l'image des phosphorylations. Ces cascades de phosphorylations sont exécutées par un ensemble coordonné de kinases conservées d'un organisme eucaryote à l'autre. Parmi celles-ci sont retrouvées les familles des kinases dépendantes des cyclines (abrégées CDKs pour *cyclin-dependent kinases*) ou des *polo-like kinases* (abrégées PLKs). À titre exemplatif, l'enzyme CDK1 occupe une place prépondérante dans la transition de l'interphase à la mitose. Son activité catalytique

est maintenue basse durant les phases S et G2 tandis que son activation est par la suite suffisante et nécessaire à l'entrée de la cellule en phase mitotique.[5–7]

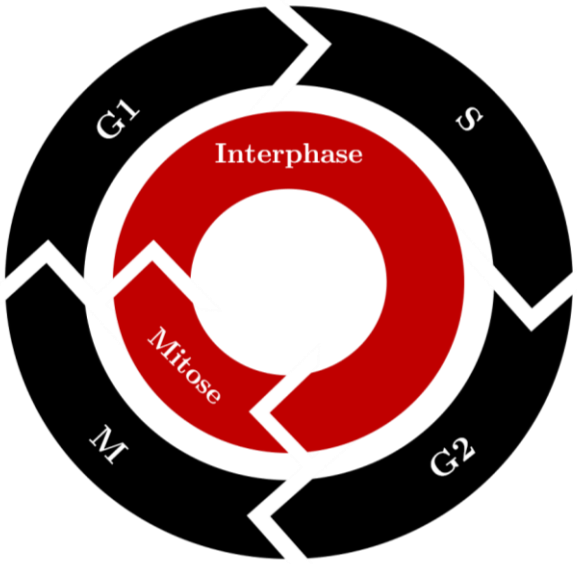

Figure 1 – Représentation schématique des étapes du cycle cellulaire d'une cellule eucaryote.

B. Phases de la mitose

La mitose, à l'instar de l'interphase, est en réalité un processus multi-étapes. Elle se compose dans l'ordre de la prophase, la prométaphase, la métaphase, l'anaphase, la télophase et de la cytokinèse (cf. Fig. 2). La prophase se caractérise par la condensation des chromosomes en fibres de chromatine ; ils deviennent alors plus courts mais plus épais. Ces fibres sont formées d'une succession de nucléosomes constitués d'octamères d'histone autour desquelles s'enroule l'ADN. Les histones sont des protéines monomères fortement basiques capables de s'oligomériser à partir de l'assemblage des sous-unités H2A, H2B, H3 et H4 (cf. Fig. 3). Il existe également des histones H1 faisant office de *linker* entre les nucléosomes.[8] Cet état condensé rendant l'ADN inaccessible, le processus de transcription est désactivé durant toute la mitose.[9]

D'autres phénomènes participent voire exacerbent ce processus de compaction dont notamment des PTMs sur les histones. Ces PTMs telles que des acétylations, méthylations ou phosphorylations s'opèrent majoritairement sur les queues des histones (en position N-terminale) dont l'accessibilité assure une modulation dynamique de la structure chromatinienne et permet le recrutement d'autres partenaires d'interaction.[10]

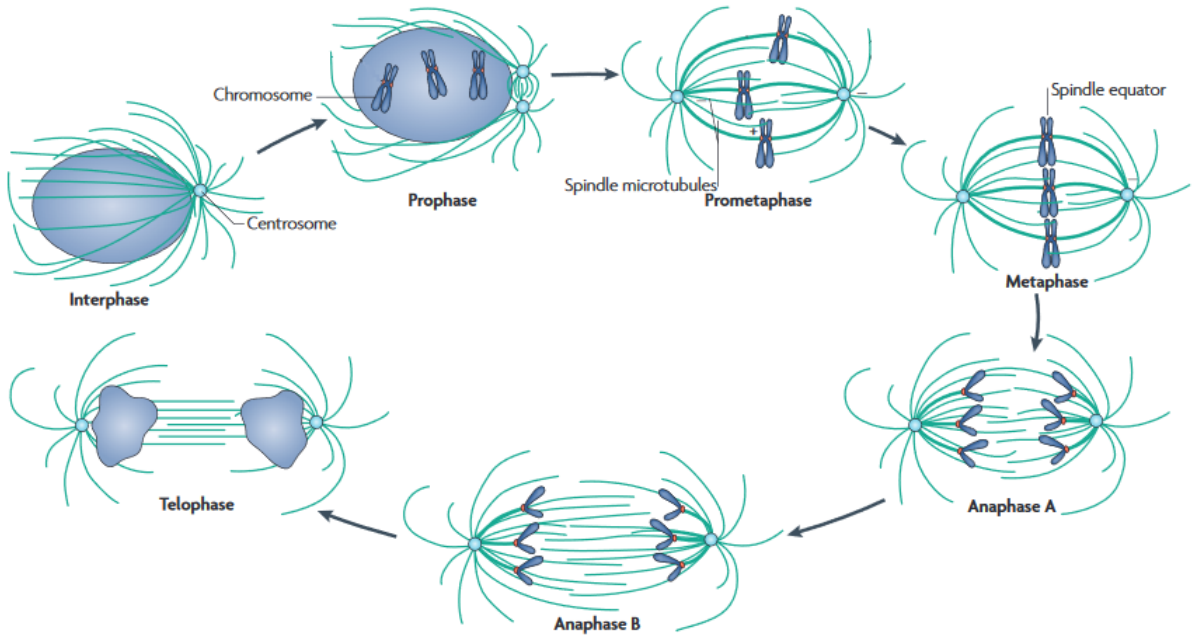

Figure 2 – Représentation schématique des phases de la mitose d'une cellule eucaryote.[15]

L'entrée en prophase s'accompagne d'une forte augmentation de motifs méthylés et phosphorylés vis-à-vis des modifications de la chromatine durant l'interphase. Parallèlement, une désacétylation conséquente des histones est également observée. À l'interphase, la chromatine peu condensée dont les histones sont hyperacétylées (ou hypométhylées) est appelée eurochromatine. À la prophase, la chromatine condensée dont les histones sont hypoacétylées (ou hyperméthylées) est appelée hétérochromatine.[11–13]

Les motifs acétylés sont associés à une structure chromatinienne plus décondensée via la réduction des interactions ADN-histone. En effet, les acétylations s'opèrent sur les résidus Lys des queues des histones, ce qui entraîne la neutralisation de l'amine terminale chargée des lysines et réduit l'affinité entre ces dernières et l'ADN. Conséquemment, ces mécanismes d'acétylation ou de désacétylation apparaissent essentiels à l'activation ou à la désactivation locale des gènes pour la transcription (cf. Fig. 4).[14] La formation du fuseau mitotique (*spindle*) par polymérisation des microtubules (abrégés MTs) est également initiée durant la prophase.

L'entrée en prométaphase est marquée par la libération dans le cytosol des chromosomes par rupture de l'enveloppe nucléaire. La prométaphase se poursuit avec l'attache des chromosomes au fuseau mitotique. Chaque chromatide-sœur est alors ancrée au polymère de MT par l'intermédiaire d'un kinétochore, un assemblage complexe de protéines. Par un mécanisme dynamique de polymérisation-

dépolymérisation des MTs, les chromosomes migrent jusqu'au centre du fuseau (entre ses deux pôles) et sont alignés au niveau de la plaque équatoriale (*spindle equator*).

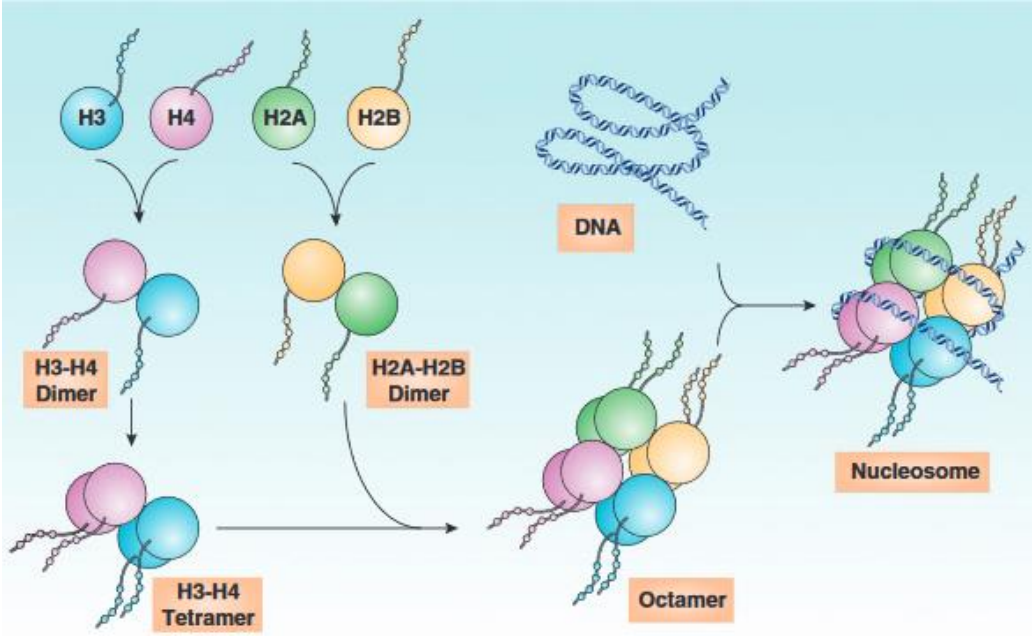

Figure 3 – Représentation schématique des structures oligomériques des histones et de l'assemblage des nucléosomes.[8]

La formation de cette dernière correspond au commencement de la métaphase au cours de laquelle la cohésion des chromatides-sœurs est fragilisée par leur séparation du centromère.[9,15] La dissociation des chromatides est notamment causée par hydrolyse de la cohésine par la séparase à la suite de l'inactivation de la CDK1. Cet exemple démontre à nouveau l'importance des PTMs dans la régulation de la mitose.[9,16,17]

L'anaphase se distingue en deux phénomènes de séparation s'opérant consécutivement (ou simultanément dans certaines cellules) : l'anaphase A fait référence à la séparation des chromatides-sœurs du centromère et de leur migration vers le pôle auquel elles font face. L'augmentation de la distance entre les pôles du fuseau ainsi que leur séparation sont achevées durant l'anaphase B. Ces mouvements sont rendus possibles par dépolymérisation progressive du fuseau mitotique qui finit par disparaître complètement. Les pôles séparés deviennent deux nouveaux centrosomes indépendants et fonctionnels.

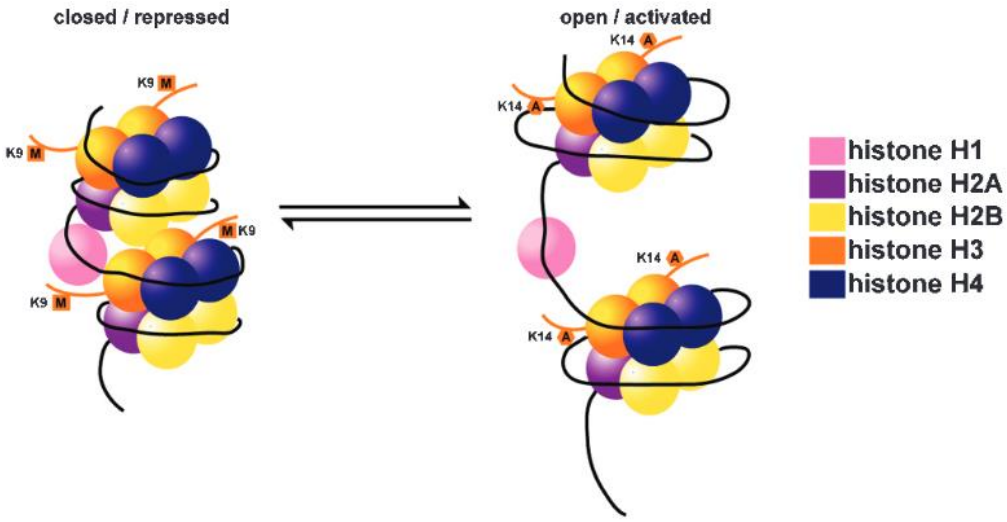

Figure 4 – Représentation schématique des états de condensation de la chromatine. La chromatine condensée (fermée) ou hétérochromatine inactive la transcription des gènes tandis que la chromatine décondensée (ouverte) ou eurochromatine l'active.[14]

En fin de mitose, durant la télophase, la cellule se prépare à la cytokinèse et à la prochaine interphase. Les enveloppes nucléaires se reforment autour des deux nouveaux noyaux dont les chromosomes se décondensent. L'hétérochromatine repasse à l'état d'eurochromatine, ce qui réactive les mécanismes de transcription. Dès lors que les compartiments nucléaires sont nouvellement formés, la cytokinèse finit de séparer les deux cellules-filles par scission du corps central (*midbody*), un assemblage de MTs.[9,15]

C. Mitose, apoptose et cancer

Le processus de la mitose est complexe et régulé par un grand nombre de gènes. Pour garantir la croissance de cellules saines, ces dernières ont développé des points de contrôle aux différentes étapes du cycle cellulaire capables d'en arrêter la progression en réponse à la détection d'erreurs. Il s'agit de voies moléculaires régulées par l'expression de protéines essentielles à l'image des kinases. À titre exemplatif, ces points de contrôle préviennent l'entrée en phase S des cellules présentant de l'ADN endommagé ou le passage à l'anaphase si tous les kinétochores ne sont pas attachés au fuseau mitotique. Ces arrêts du cycle cellulaire permettent soit d'en réparer les défauts soit d'induire la mort cellulaire si les dommages sont trop sévères. L'arrêt brusque de la mitose entraînant la dégradation de la cellule est connu sous le nom de catastrophe mitotique.[18,19]

Certains de ces gènes entrent aussi dans la régulation de l'apoptose qui est un mécanisme de mort cellulaire programmée par lequel un organisme eucaryote peut se débarrasser de cellules défectueuses ou indésirables sans entraîner de réponse

inflammatoire. La mitose et l'apoptose sont deux mécanismes cellulaires essentiels du cycle de vie d'une cellule et sont, par ailleurs, étroitement liés pour assurer l'homéostasie de l'organisme. Bien que les deux phénomènes soient opposés en termes de finalité, ils présentent des propriétés structurales et morphologiques communes. La chromatine existe sous sa forme la plus compactée et l'enveloppe nucléaire est, dans les deux cas, perdue. Malgré ces similarités, ces deux mécanismes se différencient en bien des points. L'ADN existe sous une forme chromatinienne condensée durant la mitose au bout de laquelle une cellule-mère se divise en deux cellules-filles saines. Par contre, l'ADN apoptotique est progressivement dégradé en fragments et la cellule est phagocytée par les cellules voisines ou des macrophages.[20,21]

L'apoptose n'est pas le seul mécanisme de mort cellulaire. Il existe également la nécrose, une mort cellulaire non-programmée provoquée par des dégâts de nature mécanique ou chimique sur la cellule, ou encore la macroautophagie par laquelle les organites sont digérés par l'action conjointe de l'autophagosome et des lysosomes. Le phénomène de catastrophe mitotique peut induire non seulement l'activation de l'apoptose mais également l'autophagie ou la provocation d'une nécrose cellulaire.[22,23]

L'accumulation de mutations, au fil des divisions cellulaires, est à l'origine de l'apparition de cellules cancéreuses. Ces anomalies conduisent à une suractivation des gènes pro-prolifération ainsi qu'à un silençage (inactivation) des gènes anti-prolifération et des points de contrôle du cycle cellulaire. Les cancers, dont les cellules échappent à l'apoptose, sont par conséquent caractérisés par une croissance cellulaire anormale et incontrôlée de par cette multiplication d'erreurs au cours de la mitose. Parmi ces erreurs mitotiques de nature tumorigène sont rencontrées une mauvaise séparation des chromosomes ainsi qu'une défaillance de la cytokinèse entraînant la formation de cellules tétraploïdes, autrement dit possédant le double de chromosomes d'une cellule diploïde saine.[24,25]

Le ciblage des erreurs mitotiques ou des mécanismes de contrôle du cycle cellulaire par des agents thérapeutiques spécifiques est exploité dans les traitements de lutte contre le cancer. Ces antimitotiques possèdent plusieurs modes d'action. Le premier s'effectue par perturbation de la polymérisation des MTs conduisant à une réponse du point de contrôle de l'assemblage du fuseau mitotique, bloquant le passage de la métaphase à l'anaphase. Les cellules demeurent dans un état d'arrêt de la mitose prolongé, ce qui est suivi par une induction de la mort cellulaire (cf. Fig. 5). Plus précisément, ce genre d'antimitotiques agit soit par déstabilisation des MTs soit par leur stabilisation en entravant les mécanismes de dépolymérisation en fin de mitose. La

seconde voie d'action s'établit par inhibition des protéines régulatrices de la mitose telles que les CDKs et les PLKs. En effet, ces kinases sont surexprimées dans les cellules tumorales et leur inhibition permet d'entraver la progression de la division cellulaire et de provoquer une réponse apoptotique.[18,26,27]

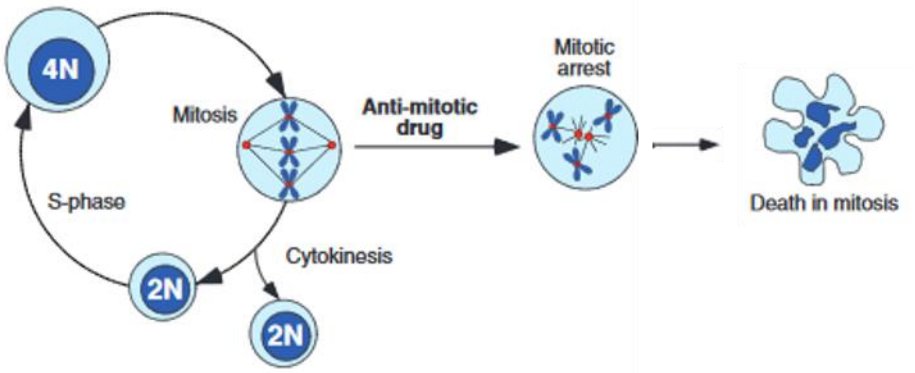

Figure 5 – Représentation schématique du mode d'action général d'un antimitotique conduisant à l'arrêt de la mitose suivi de la mort cellulaire.[26]

Comme cela a été évoqué, la mitose est régulée par de nombreux points de contrôles protéiques. Ces protéines agissent rarement de manière isolée et s'assemblent en des systèmes multiprotéiques complexes et fonctionnels par l'intermédiaire d'interactions protéine-protéine (abrégées PPIs pour *protein-protein interactions*). Ainsi, une interruption ou perturbation de ces PPIs essentielles peut entraîner l'émergence de comportements cellulaires anormaux et, par conséquent, de tumeurs. En outre, les protéines associées au cancer sont réputées pour agir comme des *hubs* (centres) au sein des réseaux de PPIs et sont, de fait, d'autant plus sensibles aux modifications d'interaction entre les partenaires composant ces complexes.[28,29]

II. Interactions protéine-protéine

Pour assurer leurs fonctions biologiques, les protéines sont souvent impliquées dans des machineries moléculaires dynamiques au sein desquelles elles interagissent. Ces interactions, connues sous la dénomination de PPIs, sont définies comme les contacts physiques spécifiques entre deux ou plusieurs protéines au sein d'une cellule ou d'un organisme vivant. Les complexes protéiques s'assemblent et se désorganisent en fonction du contexte biologique, ce qui explique la nature dynamique de ces PPIs. Leur occurrence dépend du type de cellule, de l'étape de développement, des conditions du milieu ou encore des PTMs sur l'un ou l'autre partenaire. Les PPIs ne sont pas à confondre avec d'autres types d'interactions entre biomolécules telles que des

interactions protéine-ADN, protéine-ARN (ARN pour acide ribonucléique) ou encore protéine-ligand ni avec le phénomène d'agrégation pour lequel l'assemblage protéique ne conduit pas à une structure macromoléculaire fonctionnelle.[30]

L'ensemble des PPIs d'une cellule ou d'un organisme constituent un réseau interconnecté d'une grande complexité qui est désigné par l'appellation d'interactome. Ce dernier est primordial au bon fonctionnement de la cellule de par son implication dans la globalité des processus physiologiques. À titre exemplatif, il est estimé qu'il existe entre 130000 et 650000 types de PPIs au sein de l'interactome humain.[28,31]

De par le rôle crucial joué par les PPIs dans la cellule, celles-ci sont particulièrement sensibles aux mutations ou dysfonctionnements des voies biologiques dont elles sont les régulateurs. C'est pourquoi, elles sont associées à l'apparition de nombreuses pathologies sévères telles que les cancers, les maladies infectieuses, cardiovasculaires ou neurodégénératives. Toutefois, bien qu'elles se présentent comme de nouvelles cibles thérapeutiques prometteuses, la conception d'inhibiteurs anti-PPIs demeure particulièrement exigeante de par la grande complexité de ce type de machinerie moléculaire.[31,32]

Jusqu'à présent, le ciblage des PPIs est proposé selon quatre mécanismes (cf. Fig. 6) : l'inhibition orthostérique, la stabilisation orthostérique, l'inhibition allostérique et la stabilisation allostérique. L'ensemble de ces stratégies font appel à des modulateurs des PPIs de nature inhibitrice ou stabilisatrice. Il en existe actuellement trois catégories, ayant chacune ses avantages et inconvénients : les petites molécules, les peptides et les anticorps. La modulation orthostérique s'établit par liaison du modulateur à l'interface des PPIs.

Elle se distingue de la modulation allostérique pour laquelle le modulateur lie l'un des partenaires protéiques sur un autre site que celui associés aux PPIs. Dans le cas du mécanisme d'inhibition, la formation du complexe fonctionnel est prévenue. Certains modulateurs sont, au contraire, capables d'améliorer la stabilité du complexe par modification conformationnelle (allostérie) ou par augmentation des forces de cohésion entre les domaines en interaction (orthostérie).[31]

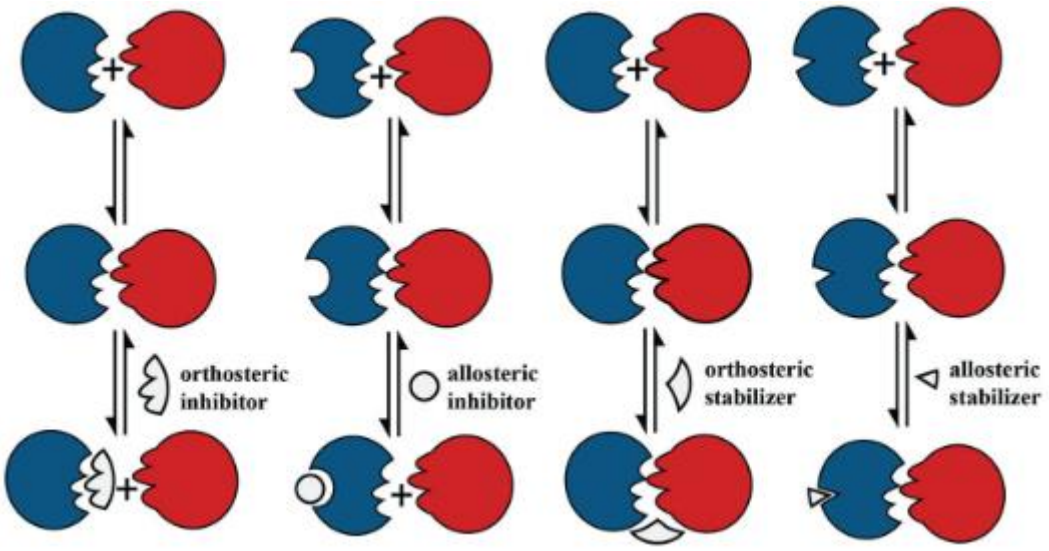

Figure 6 – Représentation schématique des mécanismes orthostériques et allostériques d'inhibition et de stabilisation des interactions protéine-protéine.[31]

III. Description des partenaires protéiques

A. La sérine/thréonine kinase PLK1

Le premier partenaire protéique considéré dans le complexe étudié est la protéine PLK1 dont l'acronyme signifie *polo-like kinase protein 1*. Historiquement, la désignation de *polo-like* provient de la détection d'un mutant du gène *polo* identifié en 1988 chez la drosophile et codant pour un homologue de sérine/thréonine kinase dont le domaine catalytique est situé en position N-terminale. Des mutations sur ledit gène *polo* résultent en des altérations du fuseau mitotique.[33]

Cette famille de kinases comporte cinq membres caractérisés jusqu'à aujourd'hui dans les cellules humaines, à savoir PLK1, PLK2, PLK3, PLK4 et PLK5. Alors que les PLKs sont évolutivement conservées des champignons (comme la levure de boulanger) jusqu'aux organismes mammaliens (comme la souris ou l'Homme), il est intéressant de noter qu'il n'existe aucun homologue ou orthologue de PLK chez les bactéries, les archées ou les végétaux à l'exception de quelques espèces d'algues.[34,35]

La famille des PLKs joue des rôles cruciaux aux différentes étapes de la division cellulaire, notamment au niveau de la réplication de l'ADN, de la maturation des centrosomes ou encore de la formation du fuseau mitotique (cf. Fig. 7). Plus particulièrement, la PLK1 intervient dans la progression de la mitose au cours de ses différentes phase. La PLK2 est localisée au niveau des centrosomes et participe à l'entrée de la cellule dans la phase S du cycle cellulaire, autrement dit l'étape de synthèse de l'ADN. La PLK3 est également essentielle à la transition de la phase G1 à la phase S et participe activement à la réplication de l'ADN. La duplication des

centrioles est régulée par les PLK2 et PLK4. Les protéines PLK1, PLK2 et PLK3 participent de concert au mécanisme de cytokinèse.[34,36]

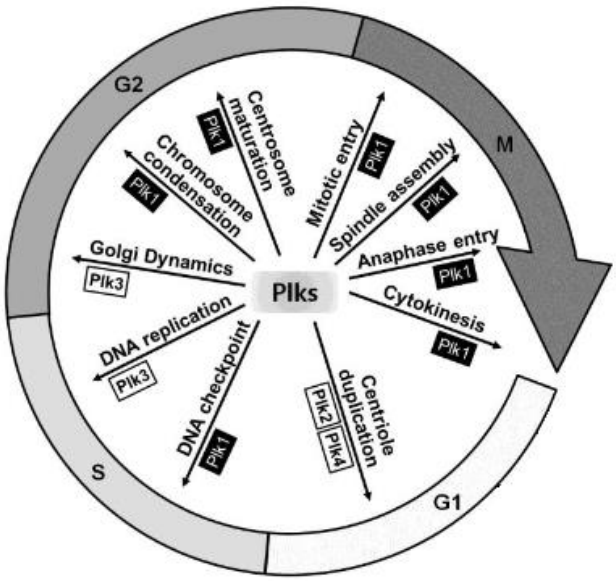

Figure 7 – Représentation schématique de la diversité des fonctions cellulaires des PLKs.[36]

La PLK1 (entrée UniProt : P53350) possède une séquence longue de 603 a.a. ainsi qu'une masse moléculaire moyenne de 68.25 kDa. À partir de l'extrémité N-terminale jusqu'à la C-terminale, sa séquence (cf. Fig. 8) comporte un domaine catalytique kinase (abrégé KIN) responsable de sa fonction phosphorylante, un signal de localisation nucléaire (abrégé NLS pour *nuclear localisation signal*), qui permet l'adressage de la protéine au noyau, ainsi que deux domaines *polo-box* (abrégés PBD ou PB pour *polo-box domain*) consécutifs, respectivement PBD1 (ou PB1) et PBD2 (ou PB2). Ces domaines sont hautement conservés chez l'ensemble des membres de la famille des PLKs. Appartenant à la classe enzymatique des kinases, la PLK1 est responsable de la phosphorylation d'autres protéines préalablement phosphorylées sur un site spécifique reconnu par les domaines PBD, communs aux PLKs. Pour être activée, la PLK1 requiert une phosphorylation préalable de la Thr^{210} au niveau de son domaine KIN, ce qui induit un changement conformationnel rendant les PBD disponibles à la reconnaissance du substrat phosphorylé. La partie kinase de la PLK1 peut alors procéder à une nouvelle phosphorylation sélective du substrat sur un résidu Ser ou Thr.[34,37–39]

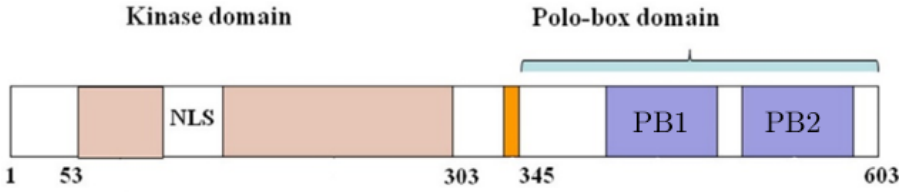

Figure 8 – Représentation schématique de l'organisation des domaines de la séquence de la PLK1 humaine de l'extrémité N-terminale à C-terminale.[38]

La PLK1 est une protéine bien connue dans la littérature avec plus de 3381 références lui étant associées dans la base de données SciFinder depuis 1994, 2293 dans PubMed ainsi que 2261 dans Scopus également depuis 1994. Elle est également de structure tridimensionnelle (abrégée 3D) résolue (cf. Fig. 9) exclusivement par diffraction (ou cristallographie) des rayons X (abrégée DRX) avec 139 structures associées dans la *Protein Data Bank* (abrégée PDB) dont 113 pour la PLK1 humaine. Toutefois, il peut être souligné qu'il n'existe actuellement aucune structure cristallographique complète de la PLK1 pour laquelle les domaines KIN et PBD ont été cristallisés séparément.

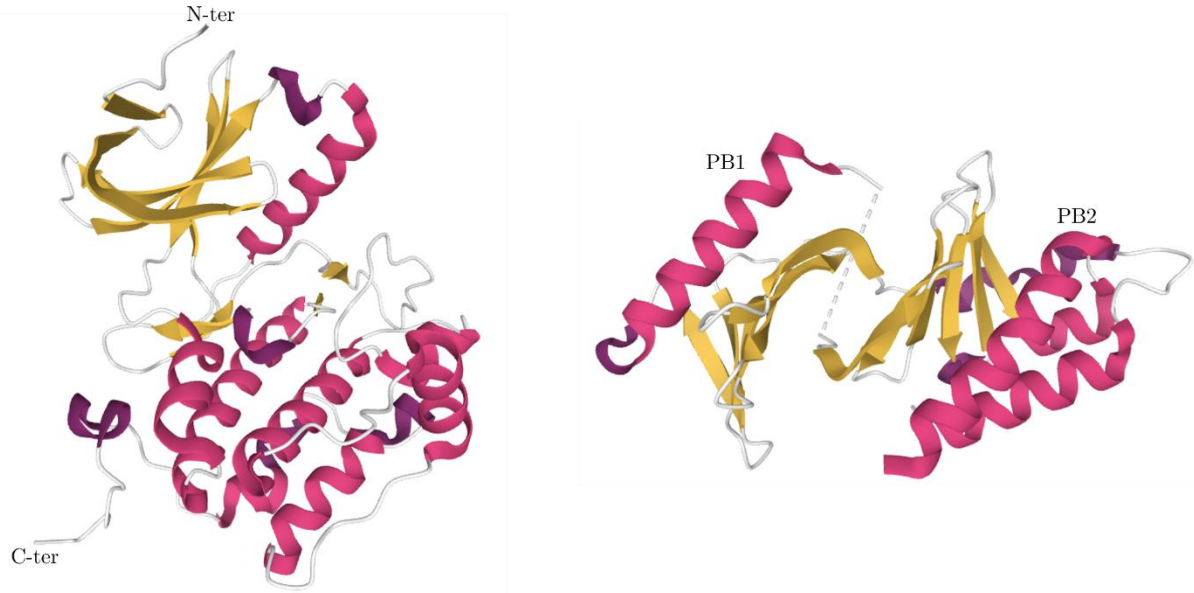

Figure 9 – Structures cristallographiques (DRX) des domaines kinase (à gauche ; entrée PDB : 2OU7) et polo-box PB1 et PB2 (à droite ; entrée PDB : 1Q4O) de la PLK1 humaine.[38]

La PLK1 est impliquée dans toutes les phases de la mitose (cf. Fig. 10). Cette nature pléiotropique s'exprime par la localisation de la PLK1 dans des structures mitotiques typiques durant l'entrée de la cellule en phase M, la maturation, l'alignement et la séparation des chromosomes, l'attache des kinétochores au fuseau

mitotique, l'inactivation de points de contrôle ainsi que durant la cytokinèse au sein du corps central. C'est pourquoi, une dérégulation de la PLK1 conduit la cellule à demeurer dans une phase pro-métaphasique pour laquelle les chromosomes sont désorganisés, mal séparés et partiellement attachés aux kinétochores. Le DAPI (pour 4',6-diamino-2-phénylindole) est un fluorophore liant l'ADN qui permet de suivre le déplacement des chromosomes au cours de la mitose (cf. Fig. 10).[40,41]

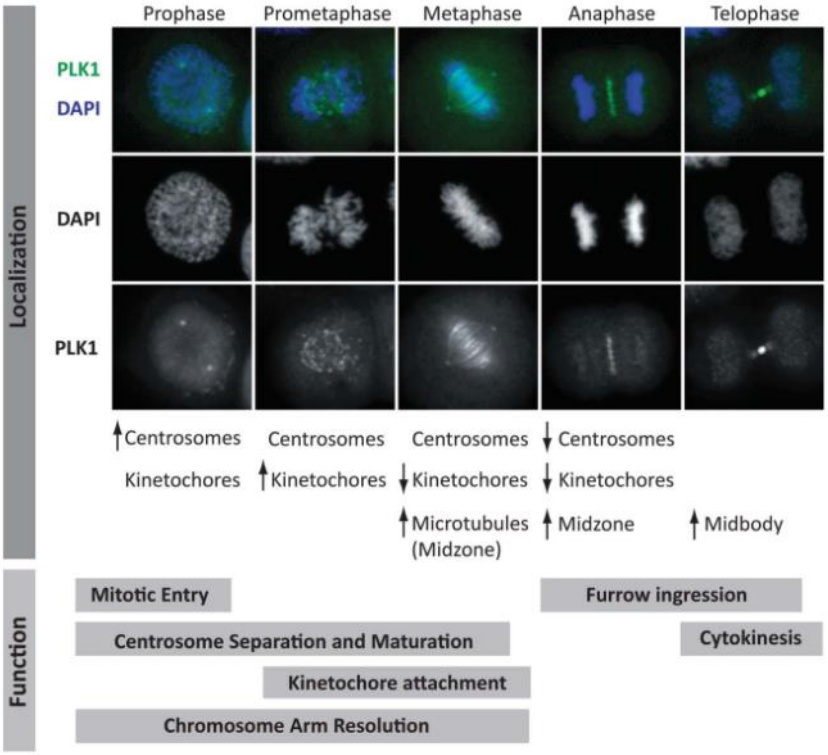

Figure 10 – Localisation et fonction de la PLK1 (en vert) humaine au cours des phases de la mitose (de la prophase à la télophase). Le DAPI (en bleu) est un fluorophore liant l'ADN. La localisation de la PLK1 dans différentes structures mitotiques est indiquée par un système de flèches rendant compte d'une association augmentée ou diminuée.[40]

La PLK1 a déjà été identifiée non seulement comme une cible potentielle dans le traitement contre le cancer mais également comme un marqueur pertinent de diagnostic de par son caractère surexprimé et suractivé dans différentes formes de cancer humain. Par inhibition de la PLK1, aussi bien vis-à-vis de son activité enzymatique par blocage du site actif que vis-à-vis de ses partenaires d'interaction par modulation des PPIs, les cellules tumorales pourraient être sélectivement éliminées. En effet, de par la prépondérance de la PLK1 dans la mitose, son inhibition provoquerait l'activation des points de contrôle du cycle cellulaire, conduisant à une catastrophe

mitotique et, *in fine*, à la réponse apoptotique. Les stratégies d'inhibition actuelles se concentrent davantage sur le ciblage des PBD que le domaine KIN étant donné qu'ils sont les médiateurs de l'interaction de la PLK1 avec son substrat. En outre, le silençage de l'oncogène associé à la PLK1 (*plk1*) peut atténuer la résistance médicamenteuse des cellules cancéreuses, ce qui augmenterait leur sensibilité aux traitements par chiomiothérapie.[38,42–44]

B. La protéine nucléaire à doigts de zinc DPF3

Le second partenaire considéré dans le complexe étudié est la protéine DPF3 dont l'acronyme signifie double doigts PHD 3 (*double PHD fingers 3*) de par son tandem de doigts de zinc PHD à l'extrémité C-terminale. Historiquement, la DPF3 est connue sous la dénomination CER-D4, équivalant à son gène codant initialement nommé *cer-d4*.[45] Cette appellation provient de l'identification en 1996 dudit gène particulièrement surexprimé dans les neurones du cervelet (*cerebellum* en Anglais) ainsi que de son appartenance à la famille génique *d4*.[46]

En effet, la DPF3 appartient à la plus large famille protéique D4 mise en évidence dès 1992 par l'intermédiaire de son premier membre NEURO-D4, autrement appelé DPF1, détecté dans le cortex cérébral du rat.[47] La famille de gène *d4* trouve son origine dans la découverte d'un nouveau motif en doigt de zinc (abrégé ZNF pour *zinc finger*) unique riche en cystéines et histidines, baptisé le domaine D4, situé à l'extrémité C-terminale de la NEURO-D4.

Il s'avère que le domaine D4 est en réalité un tandem composé de deux doigts de zinc de type PHD (pour *plant homeodomain*) étroitement liés sous la forme d'une paire de deux petits feuillets β antiparallèles, chélatant quatre ions Zn^{2+} (deux par PHD) par l'intermédiaire de résidus Cys et His.[48–51] Ce tandem caractéristique (cf. Fig. 11)[52] a été identifié originellement dans des protéines végétales contenant des homéodomaines capables de reconnaître des régions spécifiques de certains gènes à des fins régulatrices.[53]

La famille D4 se compose d'un ensemble de trois protéines nucléaires à doigts de zinc : la NEURO-D4 ou DPF1, l'UBI-D4 ou protéine Requiem ou DPF2 et la CER-D4 ou DPF3 ; les appellations NEURO-D4, UBI-D4 et CER-D4 ne sont plus usitées. La DPF2 a été identifiée pour la première fois en 1994 et devait ses autres appellations de par son expression ubiquiste dans les différents tissus ainsi que son implication dans le mécanisme d'apoptose.[54] Ces trois protéines sont caractérisées par une excellente conservation de leurs domaines structuraux dont la composition est homologue entre

leurs séquences (cf. Fig. 12a). À partir de l'extrémité N-terminale sont similairement retrouvés un domaine 2/3 unique à la famille D4, renfermant un signal de localisation nucléaire NLS pour l'adressage de la protéine au noyau, ainsi qu'un doigt de zinc de type Krüppel ; la fonction propre au domaine 2/3 demeure encore inconnue. Au niveau de l'extrémité C-terminale se situe la paire de doigts de zinc PHD (respectivement PHD-1 et PHD-2) arrangée en tandem.[55,56]

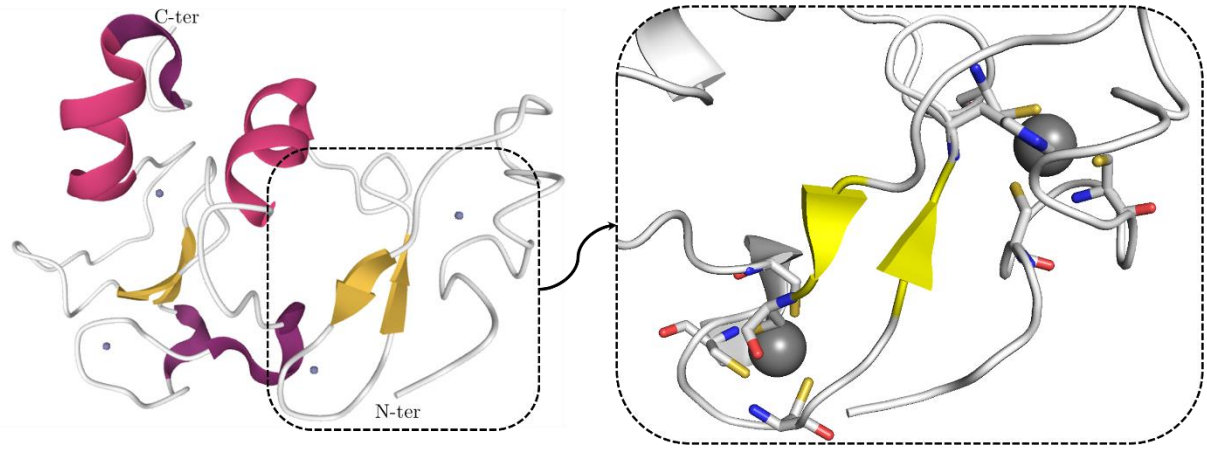

Figure 11 – Structure cristallographique (DRX) type du tandem de doigts de zinc PHD de la famille D4. Il s'agit de celui de la DPF2 (entrée PDB : 5B79). Les feuillets β antiparallèles sont représentés en jaune tandis que les ions Zn^{2+} complexés sont représentés par des sphères grises.[52]

La DPF3 (entrée UniProt : Q92784) existe sous deux isoformes fonctionnels, respectivement DPF3b (entrée UniProt : Q92784-1) pour le premier et DPF3a (Q92784-2) pour le second. Ils se différencient à la fois par leur composition en acides aminés et leur longueur de séquence au niveau de l'extrémité C-terminale. La DPF3b est l'isoforme majoritaire, souvent simplifié en DPF3, long de 378 résidus et d'une masse moléculaire moyenne de 43.08 kDa. La DPF3a correspond à l'isoforme le moins exprimé, long de 357 résidus et d'une masse moléculaire moyenne de 40.24 kDa. En considérant leur organisation de séquence (cf. Fig. 12b), le domaine 2/3, le NLS ainsi que le doigt de zinc Cys_2-His_2, abrégé C_2H_2 (l'atome de Zn^{2+} est complexé par deux cystéines et deux histidines), de type Krüppel sont parfaitement conservés. Alors que la fin de la séquence de la DPF3b est caractérisée par le tandem de doigts de zinc PHD typique de la famille D4 (PHD-1 et 2), la DPF3a ne possède qu'un seul domaine PHD tronqué, appelé PHD-1/2, suivi d'un domaine C-terminal dont la fonction n'est pas encore connue à l'instar du domaine 2/3 et dudit PHD-1/2.[39,57,58]

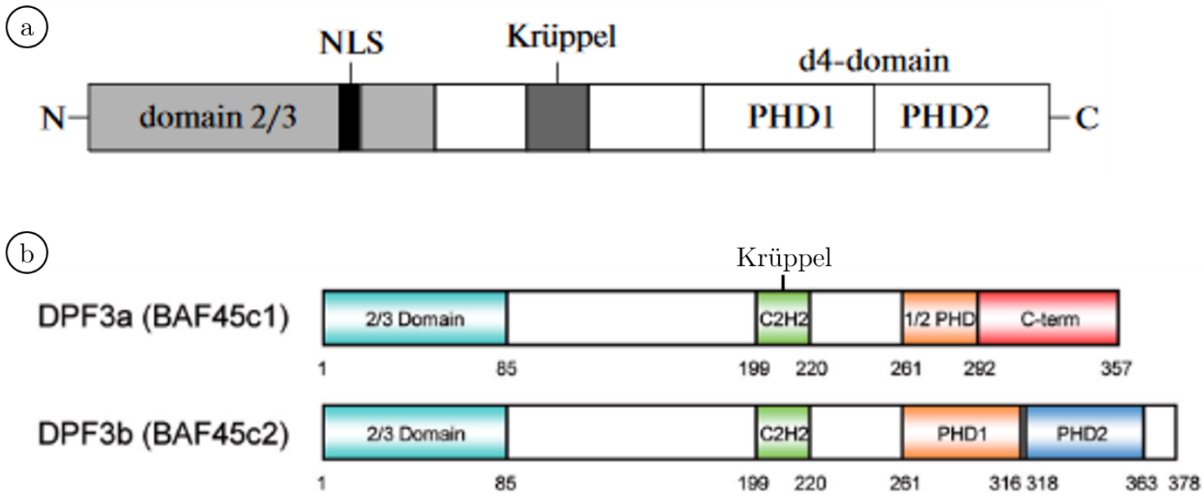

Figure 12 – (a) Représentation schématique de l'organisation typique des domaines de la famille D4[56] *et (b) de la séquence des isoformes de la DPF3 humaine, DPF3a et DPF3b, de l'extrémité N-terminale à C-terminale.*[66]

Contrairement à la PLK1, la DPF3 apparaît comme une protéine relativement nouvelle dans la littérature scientifique. En guise d'illustration, seules 78 références sont trouvées dans la base de données SciFinder depuis 2001, 36 dans Scopus et 30 entrées dans PubMed également depuis 2001 ; sa première dénomination peu usitée CER-D4 renvoie à seulement 10 références dans SciFinder depuis 1996 et à 7 dans Scopus depuis 2001. Cette rareté d'articles n'est pas exclusive à la DPF3 et s'étend aussi aux autres membres de la famille protéique D4.

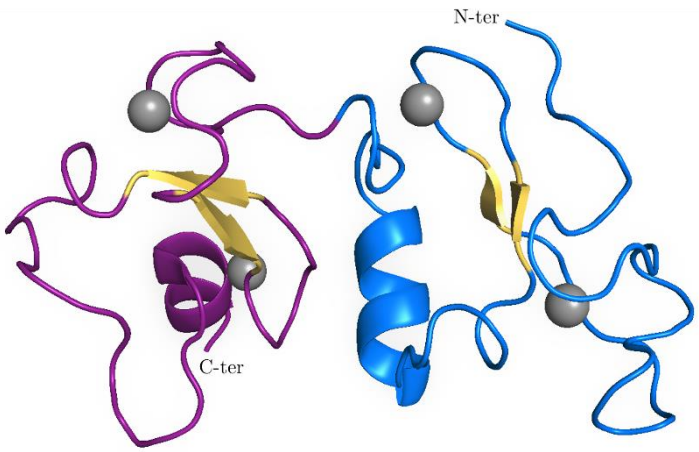

Figure 13 – Structure cristallographique (DRX) du tandem de doigts de zinc PHD-1 et 2 de la DPF3b humaine (entrée PDB : 5I3L). Le domaine PHD-1 est représenté en bleu, le PHD-2 en violet tandis que les feuillets β antiparallèles le sont en jaune et les ions Zn^{2+} complexés sont représentés par des sphères grises.[58]

Bien que cette famille ait été découverte il y a 28 ans et la DPF3 il y a 24 ans, ces protéines demeurent fort peu connues. Aucune étude biophysique complète les concernant n'existe à ce jour en dehors de quelques structures cristallographiques des doigts de zinc des protéines DPF2 ou DPF3b. En effet, la seule donnée cristallographique connue de la DPF3b est son tandem de doigts PHD-1 et 2. (cf. Fig. 13). Il n'existe aucune donnée relative à la DPF1 ni à la DPF3a dans la PDB.

La DPF3 est récemment connue sous une autre dénomination rendant compte de son appartenance au complexe BAF : BAF45C pour *BRG1-associated factor 45C*. Le complexe BAF (cf. Fig. 14) est une structure comportant jusqu'à 15 sous-unités codées par 29 gènes distincts et qui correspond, chez l'humain et d'autres organismes mammaliens, à l'analogue du complexe protéique SWI/SNF (pour *switch/sucrose non-fermentable*) responsable du remodelage de la chromatine. La dénomination SWI/SNF est un nom générique qui est souvent retrouvé en tant que synonyme de BAF dans la littérature scientifique consacrée à l'étude de ce complexe chez l'Homme. La protéine BRG1 (pour *Brahma-related gene 1*) est la sous-unité catalytique centrale du complexe.[59,60]

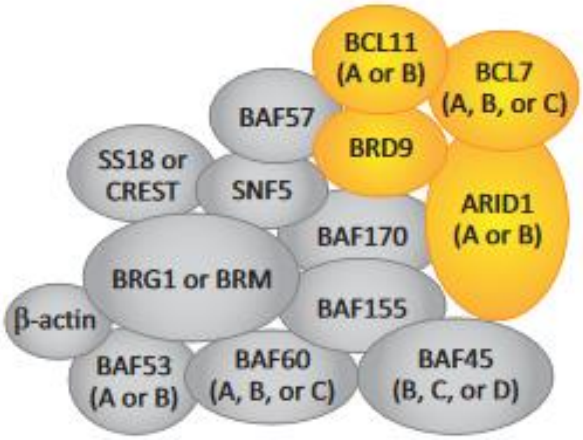

Figure 14 – Représentation schématique du complexe BAF et de ses sous-unités constitutives.[62]

La régulation de l'expression d'une importante proportion de gènes chez les organismes eucaryotes est rendue possible par la modification de la structure de la chromatine. En effet, l'état de condensation de la chromatine définit l'accessibilité d'une sélection de gènes à leurs facteurs de transcriptions associés. Les transitions sont assurées non seulement par des modifications épigénétiques sur les histones mais également par des perturbateurs des interactions ADN-histone tels que les complexes

de remodelage de la chromatine à l'image de BAF. Ce type de complexe est dépendant de l'adénosine triphosphate (abrégée ATP).[61,62]

Au sein du complexe BAF, les deux isoformes de la DPF3 (ou BAF45C) sont impliqués conjointement comme co-facteurs transcriptionnels essentiels de la transactivation de l'hétérodimère de Nf-κB RelA/p50, un facteur de transcription et médiateur des réponses inflammatoires. Les protéines DPF1 et DPF2 participent également à ce mécanisme d'activation.[63]

La DPF3b agit en tant que régulateur épigénétique de la transcription capable de recruter le complexe BAF en reconnaissant spécifiquement, par l'intermédiaire de son tandem PHD, des motifs acétylés ou méthylés sur les histones (cf. Fig. 15) ; ce mécanisme est suivi par la transcription du gène cible par la machinerie de transcription (abrégée TMC pour *transcriptional machinery complex*). Plus précisément, les deux doigts de zinc PHD peuvent se lier à des lysines méthylées ou acétylées des histones H3 et H4. Ce comportement a été répertorié au sein d'une structure cristallographique d'un peptide modifié des histones et du domaine PHD-1 et 2 de la DPF3b.[64] De par l'absence d'un tandem PHD complet chez la DPF3a, cette dernière est incapable de reconnaître des méthylations ou acétylations sur les histones et, par conséquent, de lier ces dernières.[57]

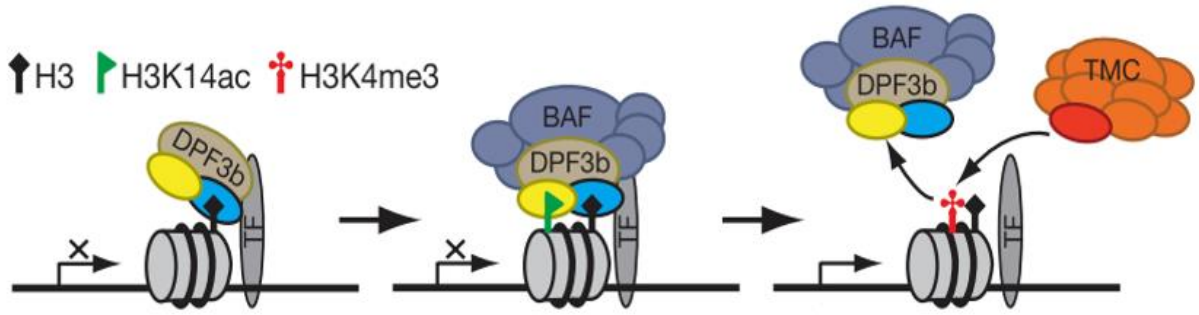

Figure 15 – Représentation schématique de la modulation de liaison du tandem PHD-1 (en jaune) et 2 (en bleu) de la DPF3b humaine à l'histone H3 pour le recrutement du complexe BAF et la transcription génique. Le motifs reconnus par les doigts PHD sont la Lys^{40} acétylée (K14ac) et la Lys^{4} triméthylée (K4me3).[64]

D'un point de vue pathologique, la DPF3 est impliquée dans la différenciation et le développement des cellules musculaires (myocytes) du squelette et du cœur. Ainsi, sa surexpression a été identifiée comme un marqueur caractéristique chez les personnes souffrant d'hypertrophies cardiaques à l'image de la Tétralogie de Fallot (abrégée TOF). Il s'agit d'une malformation cardiaque présentant une hypertrophie du

ventricule droit ainsi qu'une sténose (rétrécissement) de la valve pulmonaire, ce qui conduit à une mauvaise oxygénation du sang se manifestant par une cyanose.[57,65–67]

L'expression génique de la DPF3 semble également intervenir dans la spermatogénèse ainsi que dans la détermination de la morphologie du sperme dont des variants ou altérations de la qualité et de la composition peuvent conduire à une infertilité masculine.[68,69] La DPF3 est associée à des risques plus élevés d'infertilité masculine idiopathique et d'azoospermie.[70,71]

La DPF3 est également retrouvée surexprimée dans la maladie de Hirschsprung, une malformation gastrointestinale,[72] des anomalies dentaires et des lèvres (fente labio-palatine ou bec-de-lièvre)[73] ou encore chez les malades atteints de leucémie lymphoïde chronique, une forme de cancer des cellules de la moelle osseuse.[74] L'implication de la DPF3 dans d'autres formes de cancer a déjà été répertoriée. Des variations génétiques de la DPF3 sont identifiées comme un indicateur de plus grands risques de cancer du sein. Par ailleurs, son silençage ou sa dérégulation promeut la prolifération et la motilité des tumeurs du sein.[75,76] De manière plus spécifique, au sein du complexe BAF, la régulation négative de la DPF3a et de la DPF1 joue un rôle crucial dans le maintien du caractère souche des cellules initiatrices de tumeur dans les glioblastomes, une forme de tumeur cérébrale.[77]

A titre informatif, il est à noter que la DPF3 ou son gène codant associé *DPF3* (aussi écrit *dpf3*) ne sont pas à confondre avec la protéine antithrombotique DPf3 récemment découverte dans le ver de terre *Pheretima guillelmi*.[78]

IV. Hypothèse de la formation du complexe DPF3-PLK1

Les protéines PLK1 et DPF3 s'assembleraient en un complexe fonctionnel intervenant dans la régulation de la mitose, notamment au niveau de la séparation des chromosomes et de l'ancrage du fuseau mitotique aux kinétochores (cf. Fig. 16). Ces mécanismes s'établiraient par la liaison de la DPF3 aux MTs, proches des centrosomes ou de la plaque équatoriale, par l'intermédiaire de lysines modifiées (méthylées) reconnues par l'un des domaines de la DPF3. Cette reconnaissance ne serait rendue possible que via une DPF3 préalablement phosphorylée par la PLK1, d'où la formation du complexe protéique. Ainsi, le ciblage de ce complexe ferait office de nouvelle stratégie antimitotique prometteuse dans la lutte contre le cancer.

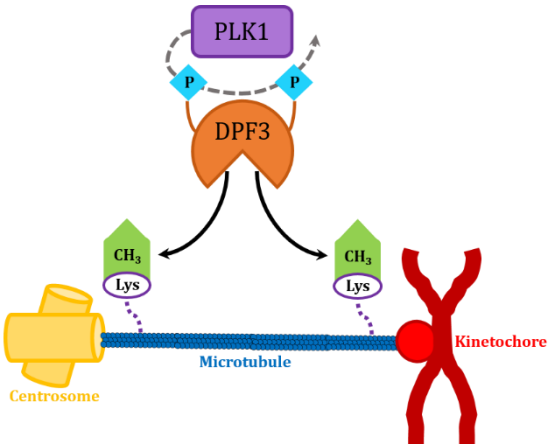

Figure 16 – Représentation schématique des mécanismes de régulation de la mitose par le complexe hypothétique DPF3-PLK1.

La formation du complexe DPF3-PLK1 ainsi que son mode d'action (ou de liaison) sont supportés par un ensemble de résultats préliminaires obtenus en interne par l'équipe du Dr. Denis Mottet de la Grappe Interdisciplinaire de Génoprotéomique Appliquée (GIGA) de l'Université de Liège (ULiège). Des tests par *microtubule binding assay* montrent qu'il existe une interaction entre la DPF3 et les MTs. L'imagerie par immunofluorescence de la DPF3 (cf. Fig. 17) donne à voir que cette dernière est localisée dans des structures mitotiques enrichies en MTs : elle est retrouvée, à la prophase, dans les centrosomes qui sont responsables de l'organisation des MTs ; en début d'anaphase, au niveau de la plaque équatoriale où les MTs sont attachés aux kinétochores ; dans le corps central au cours de la cytokinèse.

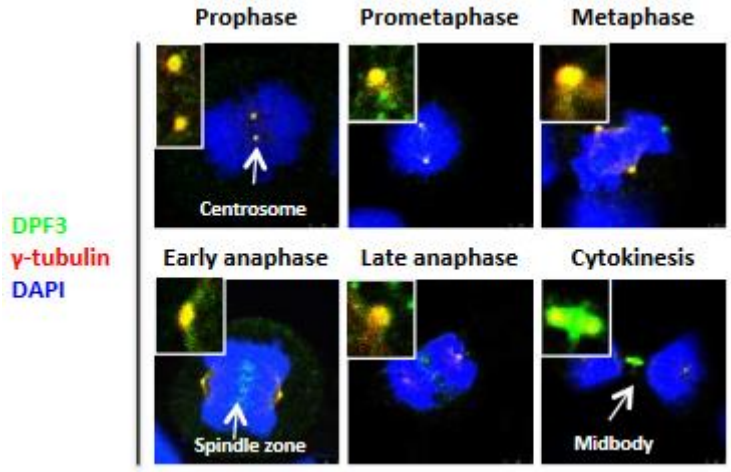

Figure 17 – Imagerie par immunofluorescence de la DPF3 (en vert) au cours des phases de la mitose. Le DAPI (en bleu) permet de suivre le déplacement des chromosomes par liaison à l'ADN.

En faveur de la formation du complexe DPF3-PLK1, ces deux partenaires sont non seulement activement impliqués dans la mitose mais également colocalisés dans certaines structures caractéristiques (cf. Fig. 18) dont la plaque équatoriale durant l'anaphase et le corps central pendant la cytokinèse. Cette colocalisation est révélée par imagerie d'immunofluorescence.

Le caractère essentiel de la DPF3 en tant que régulateur clef de la mitose a été mis en évidence par l'émergence de défauts, d'erreurs mitotiques à la suite de la déplétion de la DPF3 (cf. Fig. 19). Il apparaît des aberrations d'alignement et de ségrégation des chromosomes au cours de la métaphase et de l'anaphase, ce qui conduit à l'activation de points de contrôle mitotiques et, subséquemment, à une réponse cellulaire apoptotique.

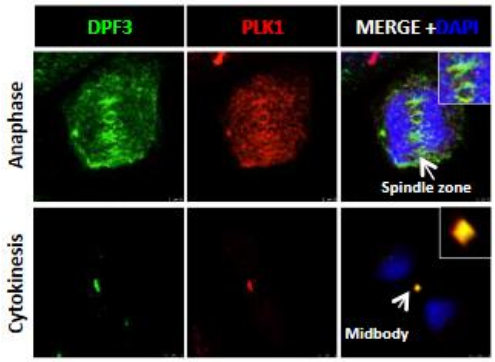

Figure 18 – Imagerie par immunofluorescence de la colocalisation de la DPF3 (en vert) et de la PLK1 (en rouge) au cours de l'anaphase et de la cytokinèse.

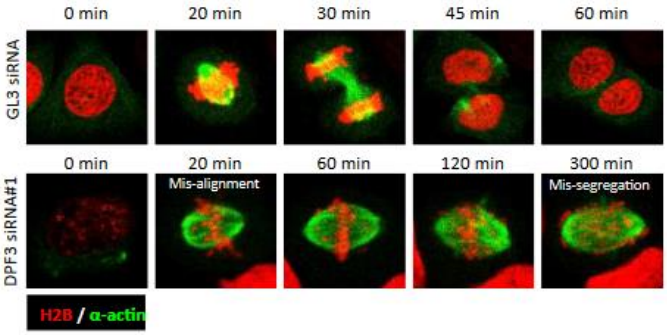

Figure 19 – Imagerie par immunofluorescence des défauts mitotiques apparaissant à la suite de la déplétion de la DPF3 (bas) vis-à-vis d'une cellule non-déplétée en DPF3 (haut).

Aussi, étant donné que le second isoforme de la DPF3 est dépourvu de toute capacité de liaison aux histones par l'absence d'un tandem de doigts PHD complet en position C-terminale, il est possible que la DPF3a occupe une fonction autre que le remodelage de la chromatine. Cette suggestion est soutenue par la localisation de chacun des deux isoformes dans des compartiments cellulaires distincts (cf. Fig. 20). Les résultats les plus récents de l'équipe du Dr. Mottet rendent compte que la DPF3b est retrouvée, à l'instar des autres sous-unités du complexe BAF, au niveau de la chromatine, ce qui apparaît cohérent en regard de sa fonction de facteur de transcription épigénétique reconnue dans la littérature. À l'inverse, la DPF3a est majoritairement retrouvée dans le cytoplasme suggérant qu'elle soit adressée hors du noyau et qu'elle pourrait interagir préférentiellement avec les MTs.

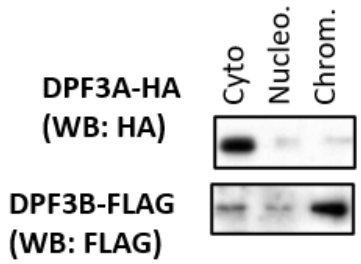

Figure 20 – Localisation des isoformes de la DPF3 dans le cytoplasme (Cyto), le noyau (Nucleo) et la chromatine (Chrom).

Toutefois, il est possible que cette localisation soit due à sa libération à la suite de la rupture de l'enveloppe nucléaire durant la prométaphase. Si le niveau d'expression de la DPF3a demeure prédominant dans le cytoplasme de la cellule en interphase, elle se révèle être un candidat plus adéquat que son isoforme DPF3b pour le rôle de séparateur des centrosomes et de connecteur kinétochorien. Ainsi, la dénomination choisie du complexe DPF3-PLK1 pourrait à l'avenir être affinée en complexe DPF3a-PLK1 ; cela reste à prouver et à vérifier.

Complémentairement aux résultats présentés, un motif consensuel de phosphorylation par les PLKs a été déterminé et se compose d'une série de six acides aminés dans l'ordre suivant : D/E-X-S/T-Φ-X-D/E où D/E est un acide aminé acide (un aspartate ou un glutamate), X est un acide aminé quelconque, S/T correspond au site de phosphorylation (une sérine ou une thréonine) et Φ est un acide aminé hydrophobe.[37] L'étude des séquences de la DPF3b et de la DPF3a révèle que chacune possède un motif de phosphorylation commun et spécifique aux PLKs entre le domaine 2/3 et le doigt de zinc C_2H_2 : E-E-S^{138}-I-Q-E. Par ailleurs, la phosphorylation de la DPF3a a déjà été répertoriée dans la littérature sur la Ser^{348} (appartenant au domaine C-terminal) par la protéine kinase CK2 (anciennement caséine kinase 2), provoquant la libération du répresseur transcriptionnel HEY par PPIs avec la DPF3a. L'absence dudit répresseur entraîne la transcription et la traduction d'une série de protéines impliquées dans des hypertrophies cardiaques.[66]

V. Hypothèse de la nature désordonnée de la DPF3

La DPF3 est soupçonnée d'appartenir à la classe particulière des protéines intrinsèquement désordonnées (abrégées IDPs pour *intrinsically disordered proteins*). En effet, une série de comportements ou de particularités déjà observés et documentés

tendent en ce sens. L'implication de la DPF3, régulée négativement ou surexprimée, dans une grande diversité de pathologies rend compte d'une certaine variété de fonctions et de partenaires d'interaction. Or les IDPs sont réputées pour leur multifonctionnalité ainsi que leur rôle de *hub* des réseaux de PPIs. De plus, l'absence d'informations sur la structure 3D de la DPF3, en dehors du tandem de doigts PHD pour la DPF3b, est peut-être révélatrice d'une conformation désordonnée.

En outre, une étude statistique récente rend compte de la distribution de trois variants protéiques au sein de différentes familles de protéines fonctionnelles dans la cellule. Les trois variants sont distingués selon leur niveau de désordre dans la séquence. Les protéines ordonnées (abrégées ORDPs) possèdent moins de 30 % de résidus désordonnés (abrégés RDs) et aucune région en position C- ou N-terminale de plus de 30 RDs consécutifs ainsi qu'aucun segment de plus de 40 RDs désordonnés consécutifs en dehors de ces extrémités. Les protéines avec des régions intrinsèquement désordonnées (abrégées IDPRs) possèdent également moins de 30 % de RDs mais avec au moins une région de plus de 30 RDs en position C- ou N-terminale ou un segment de plus de 40 RDs en dehors de ces extrémités. Le dernier variant concerne les IDPs pour lesquelles la structure primaire doit contenir plus de 30 % de RDs. Le pourcentage de désordre est déterminé pour une sélection de 20030 protéines du protéome humain permettant leur répartition dans les variants.

Il ressort qu'aucune famille protéique ne semble enrichie significativement en IDPRs tandis que deux le sont en IDPs, à savoir les protéines liant les acides nucléiques et les facteurs de transcription. La DPF3 est considérée comme un facteur de transcription épigénétique ou un régulateur de la transcription par l'intermédiaire du complexe BAF, dont elle est une sous-unité, participant au remodelage de la chromatine par interactions avec les histones. Il n'est pas non plus impossible que la DPF3 puisse interagir directement avec l'ADN.

Du point de vue des processus biologiques et des fonctions moléculaires, les résultats de cette même étude montrent que le variant IDP est surreprésenté dans les mécanismes de développement cellulaire, d'interaction et de liaison à la chromatine et aux acides nucléiques.[79] Une seconde étude récente corrobore ces résultats en décrivant que le niveau de désordre dans les protéines nucléaires et du cytosquelette est significativement plus élevé qu'au sein d'autres compartiments cellulaires.[80]

VI. Protéines intrinsèquement désordonnées

A. Introduction au concept de désordre intrinsèque

Le concept de protéine intrinsèquement désordonnée (IDP) est également désigné dans la littérature sous d'autres appellations synonymes parmi lesquelles sont plus fréquemment rencontrées : protéine nativement dépliée (abrégée NUP pour *natively unfolded protein*), protéine nativement dénaturée (abrégée NDP pour *natively denatured protein*) et protéine intrinsèquement déstructurée (abrégée IUP pour *intrinsically unstructured protein*). Bien que certaines IDPs aient déjà été identifiées ou prédites comme complètement désordonnées, la plupart d'entre elles sont des protéines dites hybrides. Les protéines hybrides contiennent à la fois des motifs ou domaines structurants et des régions où se manifeste le désordre. Il s'agit de régions intrinsèquement désordonnées, abrégées IDRs (pour *intrinsically disordered regions*), octroyant à la protéine une plus grande flexibilité (dynamique conformationnelle) ainsi qu'une multifonctionnalité dans la cellule en fonction du contexte biologique. En effet, les IDPs n'adoptent pas de structure globulaire ou fibreuse bien définie lorsqu'elles sont isolées *in vitro* ou en conditions physiologiques à pH neutre. Cette particularité provient de leur faible complexité de composition de séquence, appauvrie en résidus hydrophobes.[81–83]

Le concept de désordre intrinsèque apparaît dans le domaine de la biochimie dans le courant des années 1990. Cela renverse le paradigme protéique classique selon lequel une séquence correspond à une structure définie qui équivaut à une fonction donnée. Autrement dit, une structure tertiaire ordonnée ou l'existence de structures secondaires, définies par la séquence en acides aminés, ne sont plus suffisantes et nécessaires à la définition d'une ou plusieurs fonctions protéiques. Les IDPs sont présentes dans tous les protéomes connus et leur abondance tend à croître proportionnellement avec la complexité de leur organisme d'expression ; elles sont plus prépondérantes chez les eucaryotes que les bactéries ou les archées par exemple.[84,85]

B. Propriétés biophysiques et méthodes de caractérisation

Bien que les protéines globulaires soient caractérisées par une structure 3D définie, elles sont également réputées pour subir des transitions conformationnelles par modification de leur état de repliement. En outre, à la lueur du concept de désordre intrinsèque, les protéines peuvent adopter une certaine diversité organisationnelle sous leur forme native, autrement dit fonctionnelle. Pour décrire ces états d'organisation protéique et leurs transitions relatives, plusieurs modèles ont été proposés. Le modèle

de trinité protéique (*protein trinity* ; cf. Fig. 21a), établi par Dunker en 2001, donne à voir que les protéines existent selon trois états conformationnels interdépendants : ordonné (*ordered*), effondré-désordonné (*collapsed-disordered*) ou étendu-désordonné (*extended-disordered*). Dans la forme native, la fonction dérive alors de l'un de ces trois états ou d'une transition entre ces derniers.[86,87]

Le second modèle du quatuor protéique ou de la fonction protéique (*protein function quartet* ; cf. Fig. 21b), développé par Uversky en 2002, est une extension du modèle originel de Dunker en ce sens qu'un quatrième état conformationnel intermédiaire (celui de globule préfondu) est ajouté. Il vient que les protéines, sous leur forme native, peuvent adopter un état d'organisation parmi quatre liés entre eux : ordonné ou globule (*ordered*), globule préfondu (*pre-molten globule*), globule fondu (*molten globule*) ou enroulement aléatoire (*random coil*). À nouveau, la fonction est définie par l'un de ces états ou des transitions existant entre ces derniers.[88] Par analogie avec le modèle de trinité protéique, l'état effondré-désordonné ou de désordre effondré s'apparente à celui de globule fondu tandis que la conformation étendue-désordonnée ou de désordre étendu correspond à un enroulement aléatoire.

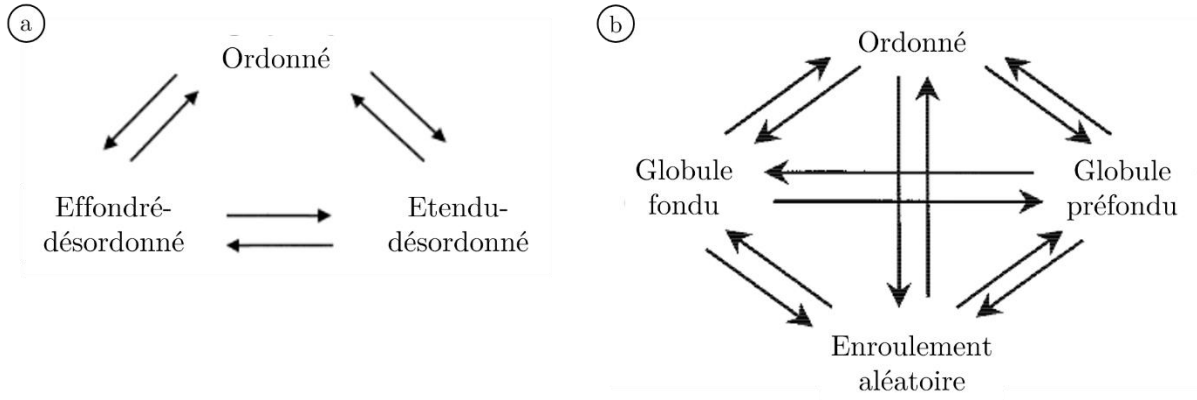

Figure 21 – (a) Représentations schématiques du modèle structure-fonction de la trinité protéique[87] *et (b) du quatuor protéique.*[88]

Les IDPs peuvent subir ce genre de transition désordre-ordre par liaison à un partenaire d'interaction tel qu'une autre protéine, un acide nucléique ou une petite molécule. Aussi, l'influence du désordre sur la structure, et par conséquent la fonction des IDPs, est de moins en moins négligée étant donné qu'il apparaît tout aussi essentiel. En effet, les protéines désordonnées s'apparentant à un état de globule fondu existent au sein de la cellule et y sont parfaitement fonctionnelles.[85,89] Il est estimé que le protéome des organismes eucaryotes comporte plus de 30 % d'IDPs[90] et que 40 % des

protéines possèdent des IDRs significatives longues de plus de 50 résidus.[91] Les IDPs apparaissent donc comme une nouvelle classe protéique gagnant de plus en plus d'intérêt de par leur prédominance, autrefois négligée ou insoupçonnée, dans les voies et mécanismes cellulaires.

De par la liberté conformationnelle des IDPs, leur étude par des méthodes de biologie structurale traditionnelles, à l'image de la DRX, se trouve vite limitée. Nonobstant cette contrainte, des signatures propres aux IDPs, de par cette même diversité de propriétés structurales, peuvent être mises en évidence par une combinaison de techniques de caractérisation biophysiques. L'approche combinatoire des analyses expérimentales est essentielle à la description la plus complète possible d'une IDP. Parmi les méthodes les plus populaires sont rencontrés : i) la résonance magnétique nucléaire (abrégée RMN) pour la détection de contacts à longue portée dans un état protéique déplié ; ii) la diffusion de rayons X aux petits angles (abrégée SAXS pour *small-angle X-ray scattering*), la diffusion dynamique de la lumière (abrégée DLS pour *dynamic light scattering*) et la chromatographie d'exclusion stérique (abrégée SEC pour *steric exclusion chromatography*) pour l'évaluation des dimensions hydrodynamiques, de la compacité et de la forme des IDPs ; iii) un ensemble de techniques spectroscopiques dont le dichroïsme circulaire (abrégé CD pour *circular dichroism*) pour l'absence de structures secondaires résiduelles et les fluorescences intrinsèque ou extrinsèque pour décrire le caractère dynamique des IDPs et leur compacité.[92,93]

C. Fonctions biologiques cellulaires

Comme cela a été évoqué précédemment, tous les protéomes des organismes vivants sont incomplets en l'absence des IDPs ou de protéines hybrides renfermant des IDRs. Les protéines désordonnées, d'un point de vue fonctionnel, sont complémentaires aux protéines globulaires, repliées en une structure 3D stable, étant donné qu'elles sont capables d'opérer à des transitions conformationnelles dynamiques essentielles à certaines activités biologiques. Par exemple, leur flexibilité de chaîne leur autorise à interagir de manière spécifique (mais avec une faible affinité) avec une importante variété de partenaires (protéines, acides nucléiques, métaux, peptides, petites molécules). Les IDPs sont par ailleurs considérées comme des *hubs* au sein des réseaux de PPIs.[94,95]

Les fonctions cellulaires couvertes par les IDPs comprennent la régulation des mécanismes de transcription et de traduction, du cycle cellulaire, d'assemblage de complexes macromoléculaires, la transduction du signal, le remodelage de la

chromatine, l'assemblage des MTs ou encore le transport moléculaire. De par leur plasticité conformationnelle, octroyant une grande accessibilité aux régions désordonnées, les IDPs sont particulièrement sensibles aux PTMs (phosphorylation, méthylation, acétylation, glycosylation). Ces dernières font office de modulateurs des interfaces d'interaction des IDPs en modifiant leur état d'organisation.[96,97]

D. Pathologies associées et stratégies thérapeutiques

Il est connu que les IDPs sont activement impliquées dans l'apparition et le développement de maladies sévères. Il est prédit que 52 à 67 % des protéines humaines associées aux cancers, aux diabètes, aux maladies cardiovasculaires et neurodégénératives possèdent des IDRs ou tombent sous la définition d'IDP.[98] Plus précisément, plus de 79 % des protéines identifiées dans divers types de cancer chez l'Homme contiendraient des IDRs au sein de leur séquence.[99] Quant aux maladies cardiovasculaires, ce chiffre s'élèverait à environ 57 %.[100] Les IDPs se présentent conséquemment comme de potentielles nouvelles cibles dans le traitement des maladies citées, ce qui implique la conception de nouveaux inhibiteurs spécifiques à cette classe de protéines. Néanmoins, une telle rationalisation médicamenteuse repose actuellement sur la connaissance de la structure 3D des protéines cibles et s'avère dès lors complexe voire non pertinente vis-à-vis des IDPs étant donné leur nature intrinsèquement désordonnée. Autrement dit, la pauvreté en motifs de structure secondaire et l'absence d'une structure tertiaire ordonnée réduisent considérablement les possibilités d'inhibition.

Plusieurs stratégies ont été proposées pour contourner cette difficulté. La première consiste à développer des molécules capables de se lier spécifiquement au(x) domaine(s) ou région(s) ordonné(s) de l'IDP afin de provoquer une transition désordre-ordre des IDRs. Il s'agit d'une inhibition allostérique. La seconde vise à élaborer des molécules capables de se lier spécifiquement à l'un des partenaires de structure ordonnée interagissant avec l'IDP cible. Il s'agit de bloquer les PPIs entre l'IDP et son partenaire, ce qui pourrait être appliqué à l'inhibition de la formation du complexe DPF3-PLK1. Ces deux approches sont compatibles avec la conception médicamenteuse traditionnelle.[101] Toutefois, leur champ d'application se retrouve vite restreint. La plupart des IDPs ne possèdent pas de domaines suffisamment structurellement ordonnant et les complexes dans lesquels elles sont impliquées se limitent rarement à un partenaire d'interaction dont la nature peut également être désordonnée.

C'est pourquoi, la stratégie la plus spécifique consisterait en l'élaboration de molécules pouvant interagir ou se lier aux régions désordonnées des IDPs afin de bloquer la protéine-cible dans une structure 3D stable par transition désordre-ordre. Toutefois, pour ce faire, la molécule doit être capable de stabiliser l'ensemble conformationnel de l'IDP. Il s'agit de l'approche la plus exigeante mais certainement la plus prometteuse aussi bien d'un point de vue fondamental que de ses applications dans le domaine pharmaco-médical. Par exemple, cette méthode a déjà été exploitée dans l'étude d'inhibiteurs potentiels de l'oncoprotéine c-Myc, utilisée comme IDP modèle, par criblage virtuel d'une bibliothèque de composés capables de lier un ensemble conformationnel.[101]

Enfin, une quatrième méthode d'inhibition existe et se concentre sur la propension des IDPs à former des agrégats sous la forme d'oligomères ou de fibrilles amyloïdes impliquées dans de nombreuses maladies neurodégénératives. Il s'agit d'élaborer des anti-agrégants capables d'empêcher la formation de macrostructures amyloïdes.[102]

Partie n°2

—

Objectifs et stratégies

I. Objectifs de recherche

Ce mémoire a pour objectif général la caractérisation des interactions régissant la formation du possible complexe DPF3-PLK1. En effet, il s'agit de définir la nature de ces interactions, les domaines et acides aminés essentiels à la reconnaissance de la DPF3 par la PLK1, la structure et dynamique du complexe ainsi que l'affinité entre les deux partenaires. L'élucidation de cette interaction préférentielle conduirait non seulement à une meilleure compréhension du mécanisme général de la division cellulaire mais également à des nouvelles pistes pour la conception de nouveaux médicaments dans la lutte contre le cancer.

En outre, la nature désordonnée des isoformes de la DPF3 est investiguée par la combinaison de techniques biophysiques et prédictives. Il s'agit de différencier les propriétés propres à chaque isoforme afin de comprendre leur spécificité, autrement dit pourquoi la cellule opère à l'épissage alternatif de la DPF3, d'un point de vue structural et fonctionnel ainsi que leur potentielle multifonctionnalité. Par suite des raisons détaillées au point IV de la Partie n°1, cette étude se concentre avant tout sur l'isoforme DPF3a dans le contexte du complexe DPF3-PLK1.

Les résultats attendus de ce projet de recherche fourniraient la première esquisse d'une étude complète des caractéristiques biophysiques et biochimiques des protéines DPF3a et DPF3b. Ainsi, la mise en évidence de la nature et du comportement *in vitro*, en regard du comportement *in vivo*, de chacun de ces deux isoformes permettrait leur identification comme de nouvelles cibles thérapeutiques dans le traitement de maladies impliquant un réseau d'IDPs. Le développement d'inhibiteurs spécifiques des PPIs au sein du complexe DPF3-PLK1 et des isoformes de la DPF3 quant à la conception respective de nouveaux antimitotiques et principes anti-IDPs fait donc office de finalité à long terme.

II. Stratégies de production et de purification

Il sera nécessaire de passer par une série d'étapes expérimentales préliminaires dans le but de produire le matériel protéique en quantité suffisante. Avant de procéder aux étapes de caractérisation, les protéines complètes PLK1, DPF3a et DPF3b (abrégées FL pour *full length*) doivent être surexprimées dans des systèmes bactériens compétents. Il en est de même pour les domaines respectifs de la PLK1, à savoir KIN et PBD, et de la DPF3b, c'est-à-dire N-ter+C_2H_2 (commun aux deux isoformes de la DPF3), PHD-1, PHD-2 et PHD-1+2. Pour rappel, l'organisation des séquences de ces partenaires protéiques est détaillée au point III de la Partie n°1. Dans le cadre de ce

projet de recherche, les constructions FL de la PLK1 et des isoformes de la DPF3 ainsi que le domaine PHD-1+2 seront purifiés. Les conditions de purification seront également optimisées sur ces mêmes constructions.

Afin de faciliter le processus de purification et d'augmenter la stabilité ainsi que la solubilité des protéines recombinantes, ces dernières sont fusionnées à une étiquette ou *tag* protéique de glutathion S-transférase (abrégée GST) à leur extrémité N-terminale. Ce *tag* GST, dont la séquence provient à l'origine du ver parasite *Schistosoma japonicum*, est inclus dans les constructions plasmidiques des protéines PLK1 et DPF3.[103]

Les protéines de fusion d'intérêt seront surexprimées dans des souches de *Escherichia coli* préalablement transformées avec les plasmides de production. Les bactéries transformées seront sélectivement cultivées dans un milieu contenant un antibiotique spécifique au vecteur de clonage employé. La surexpression sera amorcée par induction à l'isopropyl β-D-1-thiogalactopyranoside (abrégé IPTG). Les protéines de fusion seront récupérées par lyse mécanique et purifiées par chromatographie d'affinité. Avant quelque procédure de caractérisation, les protéines purifiées pourront être subséquemment clivées de leur *tag* GST. Le clivage sera opéré à l'aide d'une protéase spécifique au vecteur de clonage employé capable de reconnaître un motif d'acides aminés entre la GST et la protéine qui y est fusionnée.

Trois formats de purification seront considérés. Dans un premier temps, les lysats seront purifiés par écoulement par gravité sur une colonne d'agarose coulée manuellement. Dans un second temps, à des fins comparatives, la purification suivra un format *batch* pour lequel le lysat et la matrice sont directement mélangés et les fractions sont récoltées par centrifugations successives. Dès lors que la sélectivité du processus de purification par l'étiquette GST sera vérifiée par ces deux méthodes et que les paramètres opératoires seront optimisés, le traitement des lysats sera uniformisé et automatisé par chromatographie en phase liquide de protéine rapide (abrégée FPLC pour *fast protein liquid chromatography*) sur des colonnes pré-packées.

III. Stratégies de caractérisation

Concernant les stratégies de caractérisation envisagées, les échantillons obtenus après purification seront de prime abord analysés par électrophorèse sur gel d'acrylamide (abrégée SDS-PAGE pour *sodium dodecyl sulfate-polyacrylamide gel electrophoresis*). L'identité protéique des échantillons sera confirmée ou déterminée plus avant par spectrométrie de masse en tandem par ionisation en électrospray couplée à

une chromatographie en phase liquide (abrégée LC-ESI-MS/MS pour *liquid chromatography-electrospray ionisation-tandem mass spectrometry*).

Le comportement spectroscopique des protéines fusionnées et clivées sera étudié par spectroscopie d'absorption dans le domaine UV-visible et par fluorimétrie. Les spectres d'absorbance en UV-visible rendront également compte de la présence de la protéine et en permettront le dosage à des fins quantitatives. L'analyse fluorimétrique consistera en la mise en évidence de signatures propres aux protéines et leurs structures : la détection de l'activité fluorescente des tryptophanes par leur fluorescence intrinsèque (abrégée ITF pour *intrinsic tryptophan fluorescence*).

Les partenaires protéiques constituant le complexe étudié seront également caractérisés quant à leur taille par rapport à une protéine globulaire type ou non complexée. À cet effet, leur dimension sera évaluée par collation de mesures du diamètre hydrodynamique par DLS et du volume d'élution par SEC. Les informations apportées par ces deux méthodes devraient permettre de dégager des indices sur la nature désordonnée de la DPF3.

D'un point de vue structural, les modifications conformationnelles des partenaires protéiques ainsi que leurs proportions en structures secondaires seront déterminées par CD.

Parallèlement, l'étude du caractère IDP de la DPF3 sera aussi développée selon une approche prédictive par la détermination et la comparaison d'une série de paramètres physicochimiques sur la seule connaissance de la structure primaire. Ces paramètres sont sélectionnés selon leur disponibilité, leur pertinence et leur récurrence dans la littérature dédiée à la prédiction des propriétés des IDPs. Dans les cas où cela est possible, les prédictions théoriques seront confrontées aux preuves expérimentales sur des critères identiques. Une représentation schématique de l'ensemble des stratégies expérimentales et de rationalisation mises en place est proposée à la Fig. 22.

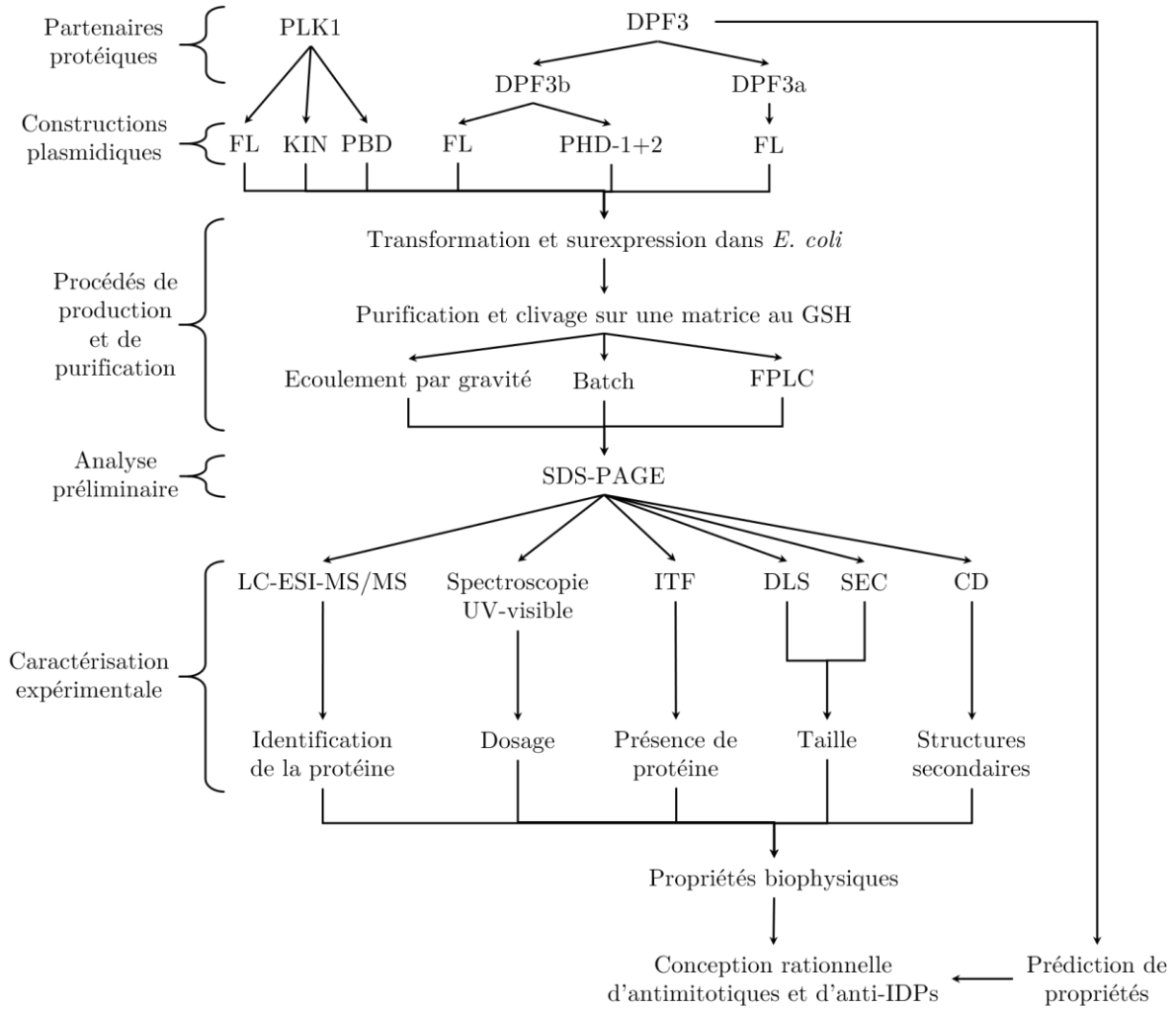

Figure 22 – Représentation schématique des stratégies expérimentales et prédictives visant la caractérisation du complexe DPF3-PLK1 et de ses partenaires protéiques constitutifs.

Partie n°3

—

Résultats et discussions

Section a

—

Purification et caractérisation des protéines de fusion

I. Transformations et surexpressions

Les transformations des constructions de la PLK1 et ses domaines, à savoir FL, KIN et PBD, ont été exécutées avec succès. Pour chaque plasmide, les souches efficacement transformées et cultivées sur un milieu LB-agar sont présentées avant (cf. Fig. 23a) et après reconcentration (cf. Fig. 23b). Chaque point blanc représente une colonie bactérienne composée d'individus identiques et ayant intégré le plasmide codant pour la protéine recombinante. La différence notable en termes de densité de population est due à la différence de concentration (d'un facteur dix).

Considérant les surexpressions, le paramètre observé le plus variable concerne les temps de culture nécessaires aux souches de *E. coli* pour atteindre leur phase exponentielle de croissance. Alors que la densité optique (abrégée DO) mesurée à 600 nm n'atteint sa gamme idéale comprise entre 0.5 et 0.7 qu'en 2h30-3h00 pour les constructions des protéines DPF3a et DPF3b, les souches transformées à l'aide du plasmide PLK1 FL entrent dans leur phase stationnaire de croissance caractérisée par une DO supérieure à 1.0 après seulement 2h30 d'incubation.

Aussi, il peut être noté une certaine divergence quant à l'efficience de production des constructions de DPF3. Les niveaux de surexpression sont équivalents pour DPF3a FL, N-ter+C_2H_2, PHD-1 et PHD-1+2, particulièrement élevés pour PHD-2 et faibles pour DPF3b FL. La différence la plus remarquable est celle existant entre les deux isoformes de la DPF3. Bien que leur vecteur de clonage (pGEX-3X), leur souche (*E.coli* BL21 Rosetta 2^{TM}) et leurs paramètres de culture soient identiques, l'expression hétérologue de la DPF3b semble bien plus limitée que celle de la DPF3a. Les conditions de transformation, de culture et d'induction à l'IPTG pour les protéines d'intérêt fusionnées à la GST avaient déjà été optimisées par l'équipe du Dr. Denis Mottet du GIGA. D'un point de vue global, elles ont été ici efficacement reproduites.

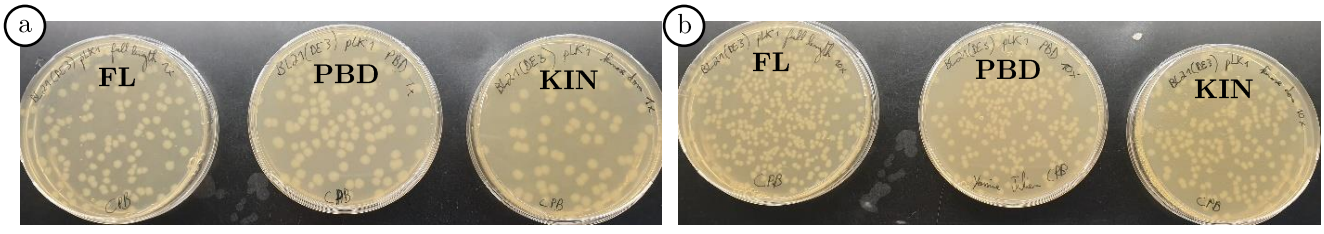

Figure 23 – (a) Colonies bactériennes de E. coli BL21(DE3) transformées avec les plasmides associés à PLK1 et ses domaines une fois concentrés et (b) dix fois reconcentrés. Pour (a) et (b), les plasmides sont de gauche à droite PLK1 FL, PBD et KIN. Les photographies ont été réalisées après environ 16 heures d'incubation à 37 °C sur gélose.

II. Purifications par chromatographie d'affinité sur colonne d'écoulement par gravité

A. Vérification de la spécificité du processus de purification

Concernant le procédé de purification (cf. Fig. 24), il repose sur une matrice composée de billes d'agarose (Sepharose^{TM}) auxquelles est couplé le substrat de la GST, à savoir le glutathion réduit (abrégé GSH) ; sous sa forme oxydée le glutathion est abrégé GSSG de par sa dimérisation via un pont disulfure. En tant qu'enzyme, la GST possède une affinité particulière pour le GSH. Le caractère sélectif du procédé de purification est assuré par cette reconnaissance enzyme-substrat et permet de se débarrasser des protéines non recombinantes dans le *flow-through* (abrégé FT).

La liaison GST-GSH étant réversible, il est alors possible d'éluer les protéines de fusion marquées à la GST par passage d'un tampon concentré en GSH. Le GSH présent dans le tampon d'élution entre en compétition avec le glutathion greffé au support d'agarose et finit par saturer les sites actifs des GST, ces dernières fusionnées à la protéine d'intérêt, ce qui en provoque le décrochage.[103–105] Le clivage est réalisable soit en solution après élution de la protéine-GST soit sur colonne autorisant directement la récupération séparée de la protéine d'une part et de l'étiquette GST d'autre part à condition d'utiliser des tampons d'élution distincts.[106]

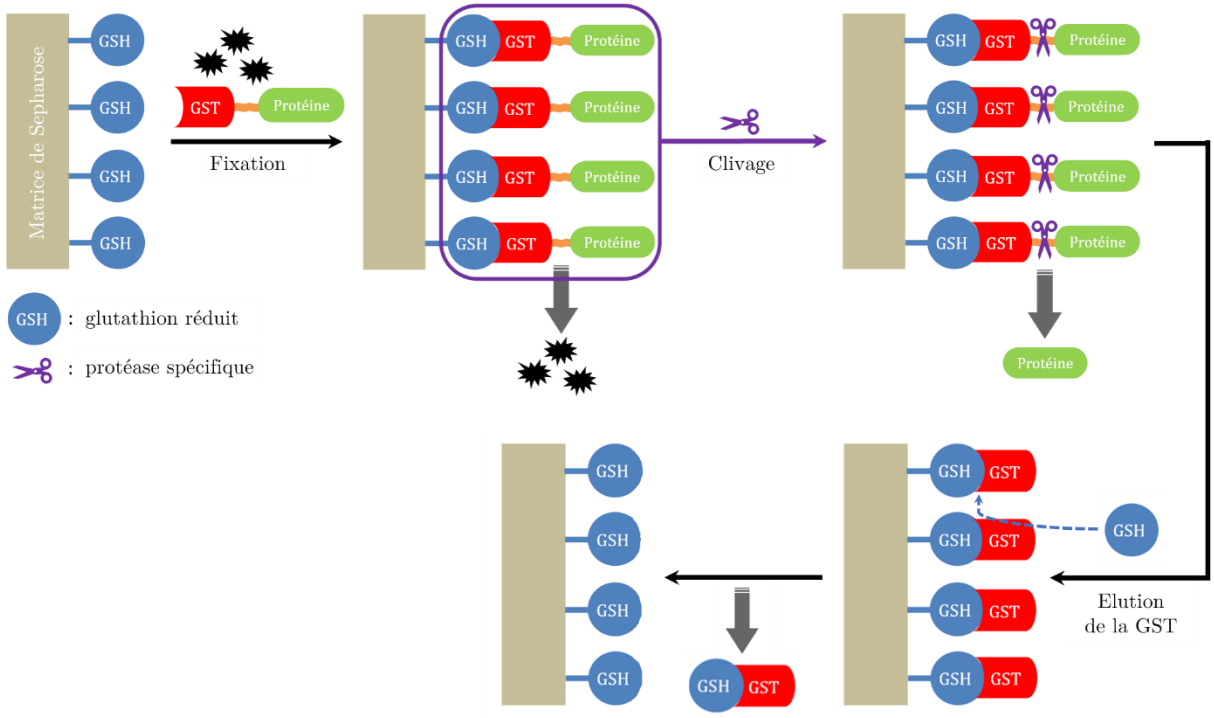

Figure 24 – Représentation schématique générale des procédés de purification et de clivage des protéines de fusion marquées à la GST.

L'efficacité de fixation de la matrice d'agarose couplée au GSH a été dans un tout premier temps testée sur un format plus facilement optimisable et modulable que sont les colonnes d'écoulement par gravité. La première purification a été effectuée sur la construction du domaine PHD-1+2 de la DPF3b dont le gel SDS-PAGE résultant est présenté à la Fig. 25a. Il vient qu'une unique bande est visible dans les échantillons d'élution E2 à E5 (avec une intensité de bande plus marquée pour E2) située entre 25 et 37 kDa selon l'échelle du marqueur de masse moléculaire (abrégé MWSM pour *molecular-weight size marker*). Cette masse moléculaire équivaut à celle de la protéine de fusion PHD-1+2-GST ($M_{PHD\text{-}1+2\text{-}GST} \sim 38$ kDa). Aucun autre matériel protéique n'est détecté sur gel par coloration au bleu de Coomassie. Le procédé de purification via un *tag* GST apparaît donc efficace dans l'obtention de fractions protéiques de pureté satisfaisante en regard de la construction PHD-1+2.

Etant donné le succès de ce premier test, il est répété sur trois nouvelles constructions : PLK1 FL, DPF3a FL et DPF3b FL. Contrairement au domaine PHD-1+2, le gel de la PLK1 (cf. Fig. 25b) présente des bandes inattendues dans les fractions E2 à E4. Alors qu'une première bande aux alentours de 100 kDa est associée à la protéine de fusion PLK1-GST ($M_{PLK1\text{-}GST} \sim 94$ kDa), deux autres bandes se manifestent à des tailles inférieures : l'une localisée entre 50 et 75 kDa, l'autre au niveau de 25 kDa en regard du MWSM. La bande du milieu concorde en termes de masse moléculaire avec celle de la PLK1 FL dénuée de son *tag* ($M_{PLK1} = 68.25$ kDa) tandis que celle du bas correspond à la GST seule ($M_{GST} = 25.50$ kDa). Bien que la présence de ces deux dernières bandes puisse être due à des protéines contaminantes, il est très improbable d'en rencontrer possédant la même masse moléculaire que la GST ainsi qu'une affinité spécifique au GSH. Il est plus plausible que ces bandes soient le résultat d'un clivage post-lyse de la PLK1-GST par des protéases endogènes de *E. coli* ; d'autant plus que le lysat a été conservé quatre jours à 4 °C avant purification.

Cette hypothèse est appuyée par le motif de bandes similaire observé sur les gels de la DPF3a-GST (cf. Fig. 25c) et de la DPF3b-GST (cf. Fig. 25d). Concernant la DPF3a-GST, dans les échantillons E3 à E5, il y a de nouveau occurrence de trois bandes distinctes : la première (celle du haut) vers 75 kDa est sûrement relative à la DPF3a-GST ($M_{DPF3a\text{-}GST} \sim 66$ kDa), celle du milieu située entre 37 et 50 kDa à la DPF3a sans la GST ($M_{DPF3a} = 40.24$ kDa) et la dernière légèrement au-dessus de 25 kDa à la GST identiquement à la Fig. 25b. Quant à la DPF3b-GST, la bande supérieure proche de 75 kDa, observée dans les échantillons E5 à E7, correspondrait à la DPF3b-GST ($M_{DPF3b\text{-}GST} \sim 69$ kDa), celle intermédiaire entre 37 et 50 kDa, de E5 à E7, à la DPF3b

clivée de son *tag* (M_{DPF3b} = 43.08 kDa ; la bande est particulièrement peu intense en regard des autres identifiées sur le gel) et celle inférieure, de E3 à E7, vers 25 kDa à la GST. En conclusion, ce clivage partiel semble apparaître uniquement pour les constructions FL des protéines de fusion.

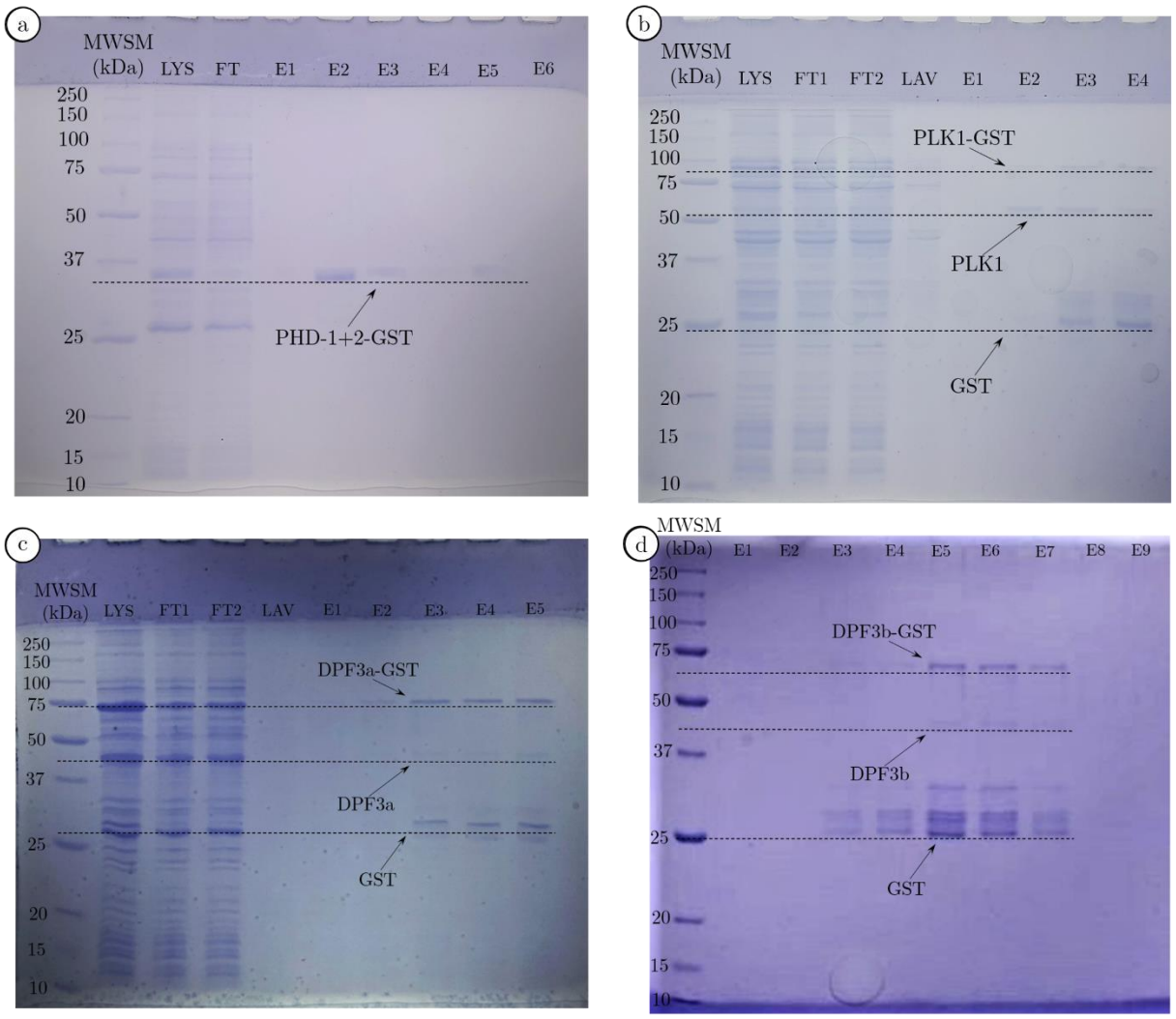

Figure 25 – (a) Gels SDS-PAGE 12 % des protéines de fusion PHD-1+2-GST, (b) PLK1-GST, (c) DPF3a-GST et (d) DPF3b-GST purifiées par chromatographie d'affinité sur colonne d'écoulement par gravité. Pour chaque gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through (FT1 le premier, FT2 le second), LAV est le lavage au tampon de fixation avant élution et E correspond aux fractions d'élution. Les lysats ont été stockés quatre jours avant purification.

B. Sensibilité à la protéolyse

Afin de vérifier l'hypothèse de protéolyse des protéines de fusion, la nature de chacune des trois bandes de PLK1 FL et DPF3a FL a été déterminée par LC-ESI-MS/MS. D'un point de vue qualitatif, la PLK1 et la DPF3a sont détectées dans

l'ensemble des trois bandes de leur gel respectif avec une meilleure couverture de séquence pour celle du haut. L'identification de peptides uniques aux protéines FL peut s'expliquer par la migration de fragments protéiques de plus petite masse moléculaire vers la région de 25 kDa. Ces fragments seraient également le produit d'une protéolyse non spécifique des protéines de fusion. En outre, un motif de migration continue est observé en particulier sur le gel de PLK1 FL (cf. Fig. 25b) et de DPF3b FL (cf. Fig. 25d) entre 37 et 25 kDa. Concernant la GST, elle est uniquement retrouvée dans les bandes du haut et du bas avec la meilleure couverture de séquence pour la dernière, ce qui corrobore l'hypothèse du clivage. Quelque nuance est à apporter à cette analyse. En effet, au regard des bandes obtenues sur gel et de l'identification qui en a été faite par spectrométrie de masse, la présence de la PLK1 ou de la DPF3a sans l'étiquette GST est troublante étant donné qu'elles ne possèdent aucune affinité particulière pour le substrat greffé à la colonne. Jusqu'à présent aucune explication cohérente ne permet de justifier un tel comportement.

Au-delà des protéines d'intérêt, une chaperonine de *Escherichia fergusonii* est également identifiée dans la bande intermédiaire de PLK1 FL. Toutes les bandes du gel de DPF3a FL contiennent, en faible concentration, quelques protéines chaperones et ribosomiales de *E. coli*. Toutefois, très peu de crédit est apporté à ces contaminations provenant probablement de la colonne chromatographique en phase liquide utilisée pour les analyses par spectrométrie de masse.

Complémentairement, la résistance à la protéolyse de PLK1 FL est confrontée à celle du domaine PHD-1+2 de la DPF3b. L'expérience de purification a été réitérée dans les mêmes conditions à partir d'un lysat de chacune de ces constructions stocké deux semaines à 4 °C ; au lieu de quatre jours comme cela était le cas pour les premières purifications. Alors que la protéine de fusion PHD-1+2-GST demeure majoritairement intacte (cf. Fig. 26a), seule la bande de la GST apparaît sur le gel de PLK1 FL (cf. Fig. 26b). Cette résistance exceptionnelle à la protéolyse au cours du temps du domaine PHD-1+2 peut s'expliquer par sa nature plus ordonnée et, par conséquent, plus stable. Pour rappel, ce domaine est l'unique structure cristallographique obtenue jusqu'à ce jour de la DPF3b. Quant à la PLK1, elle a été complètement clivée de son *tag* GST avant purification, ce qui explique son absence sur le gel ; elle est perdue dans le FT avec les autres protéines contaminantes.

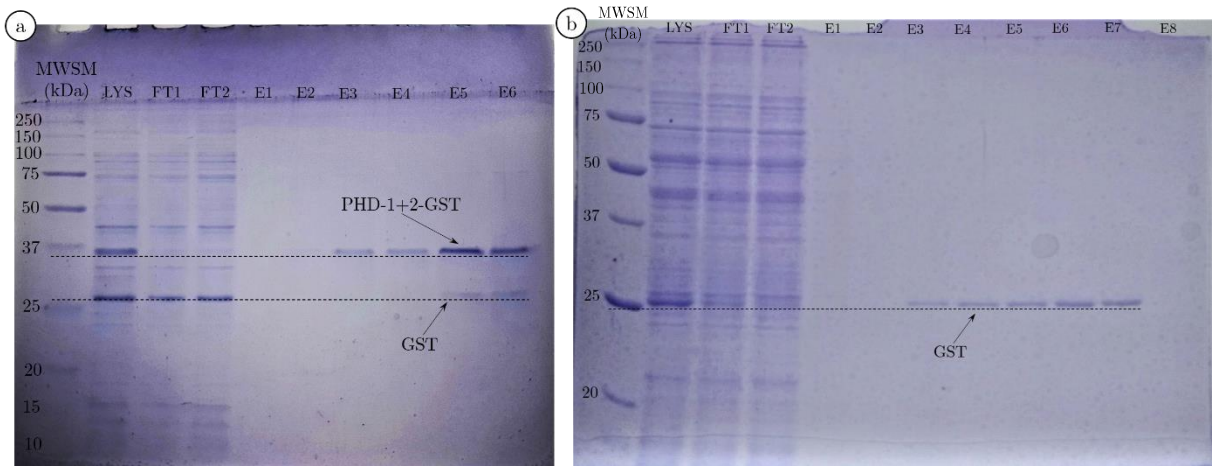

Figure 26 – (a) Gels SDS-PAGE 12 % des protéines de fusion PHD-1+2-GST et (b) PLK1-GST purifiées par chromatographie d'affinité sur colonne d'écoulement par gravité après stockage du lysat respectif deux semaines à 4 °C. Pour chaque gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through (FT1 le premier, FT2 le second) et E correspond aux fractions d'élution.

En outre, l'une des particularités des IDPs est leur sensibilité accrue à la protéolyse. Elle s'explique par la plus grande accessibilité des régions désordonnées aux protéases en regard des structures globulaires au sein desquelles les chaînes latérales sont plus enfouies.[81,107] Les fragments observés sur gel de la DPF3a (cf. Fig. 25c) s'expliquent par cette caractéristique pour autant que la DPF3 soit considérée comme une IDP, ce qui est notamment supportée par l'absence de structure 3D résolue de sa séquence complète. Nonobstant l'absence d'analyse LC-ESI-MS/MS sur la DPF3b, cette explication peut y être étendue. Il s'agit du premier indice expérimental ayant orienté ce travail de recherche sur l'investigation de la nature désordonnée de la DPF3a. Bien que la PLK1 ne soit pas reconnue comme une protéine désordonnée, sa sensibilité protéolytique est ici comparable. Il peut être aussi noté que les protéines exprimées et purifiées sont artificielles de par leur lien de fusion entre la protéine d'intérêt et la GST, ce qui peut également rendre compte de cette plus haute sensibilité vis-à-vis des protéases de *E. coli*. En effet, la fragilité de ce lien est exacerbée vis-à-vis des protéines qu'il fusionne, ces dernières ayant par elles-mêmes leur propre stabilité.

Afin d'éviter la récurrence de ce clivage non-quantitatif dans les prochaines purifications, la période d'activité de l'inhibiteur de protéase, le fluorure de phénylméthylsulfonyle (abrégé PMSF *pour phenylmethylsulfonyl fluoride*), a été étendue par son addition dans tous les tampons de purification à raison de 125 à 200 µM. Le PMSF est un inhibiteur de protéase à sérine habituellement additionné au cours

de la lyse afin de prévenir la protéolyse non-sélective opérée par les protéases endogènes de l'organisme de production.

L'efficacité du procédé de purification au format *batch*, avec en parallèle celle du traitement au PMSF, est vérifiée sur la construction DPF3a FL. Par observation du gel de la DPF3a-GST en présence de PMSF (cf. Fig. 27), il vient que le problème de protéolyse intempestive a été résolu. En effet, seule la bande située entre 50 et 75 kDa correspondant à la protéine de fusion DPF3a-GST est visible dans les deux derniers échantillons d'élution. Aucune procédure de purification au format *batch* n'a été réalisée sur d'autres constructions de protéine de fusion en dehors des essais de clivage présentés à la Partie n°3, Section b.

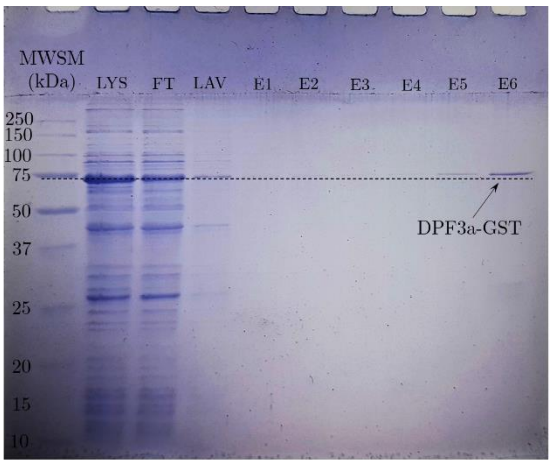

Figure 27 – Gel SDS-PAGE 12 % de la protéine de fusion DPF3a-GST purifiée par chromatographie d'affinité au format batch. MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through, LAV est le lavage au tampon de fixation avant élution et E correspond aux fractions d'élution.

III. Purifications par FPLC

Les paramètres de purification ayant été optimisés précédemment aux formats d'écoulement par gravité et *batch*, l'automatisation du processus est exécutée sur un système FPLC pour les constructions PLK1 FL et DPF3a FL. Concernant la protéine de fusion PLK1-GST, le chromatogramme du pic d'élution (cf. Fig. 28a), suivi par l'absorbance à 280 nm, montre la sortie de la protéine d'intérêt, ce qui est confirmé par le gel SDS-PAGE résultant (cf. Fig. 28b). Sur ledit gel, une démultiplication de bandes est à nouveau observable comme cela était le cas par purification d'écoulement par

gravité (cf. Fig. 25b). Conformément à l'analyse en LC-ESI-MS/MS, une bande supérieure proche de 100 kDa équivalant à la PLK1-GST, une seconde plus basse entre 50 et 75 kDa dont la position correspond à la PLK1 sans son *tag* et une bande inférieure vers 25 kDa identifiée comme étant la GST et précédée d'un motif de migration continue, probablement dû à des fragments de la protéine de fusion initiale. Malgré l'addition de PMSF à raison de 125 à 200 µM au cours de l'étape de lyse, il semble que la protéolyse n'est finalement pas complètement entravée. Le PMSF ne peut être ajouté aux tampons de fixation et d'élution à cause des contraintes imposées par le dégazage de ces solutions en FPLC. Il peut être alors envisagé que la protéolyse se déroule dès la surexpression de la protéine de fusion durant l'induction des souches de *E. coli* transformées. Les protéases endogènes cliveraient une partie du matériel protéique produit, d'où l'apparition marquée de la GST sur gel. Les échantillons récoltés s'apparentent donc à un mélange de PLK1-GST, de PLK1 et de GST clivées.

Le chromatogramme de la DPF3a-GST (cf. Fig. 28c) donne à voir un pic d'élution bien défini à l'instar de la PLK1-GST. Contrairement à la construction PLK1 FL et à ce qui avait été constaté précédemment au format d'écoulement par gravité (cf. Fig. 25c), le nombre de populations de bandes distinguables s'élève à deux sur le gel de la DPF3a-GST (cf. Fig. 28d). Celle du haut proche de 75 kDa s'apparente à la protéine de fusion complète tandis que celle du bas située vers 25 kDa équivaut au *tag* GST. À nouveau, si protéolyse il y a eu, elle s'opère en amont de la lyse et de la purification au sein des cellules bactériennes, autrement dit au cours de l'étape d'induction à l'IPTG. Par ailleurs, la bande intermédiaire, entre 37 et 50 kDa, autrefois présente ne l'est plus et semble même apparaître de manière plus contrastée dans le FT. Cette bande correspondrait à la DPF3a clivée de son étiquette GST. Il est par conséquent plus cohérent de la retrouver au sein du FT de par l'absence d'affinité de la DPF3a pour le GSH.

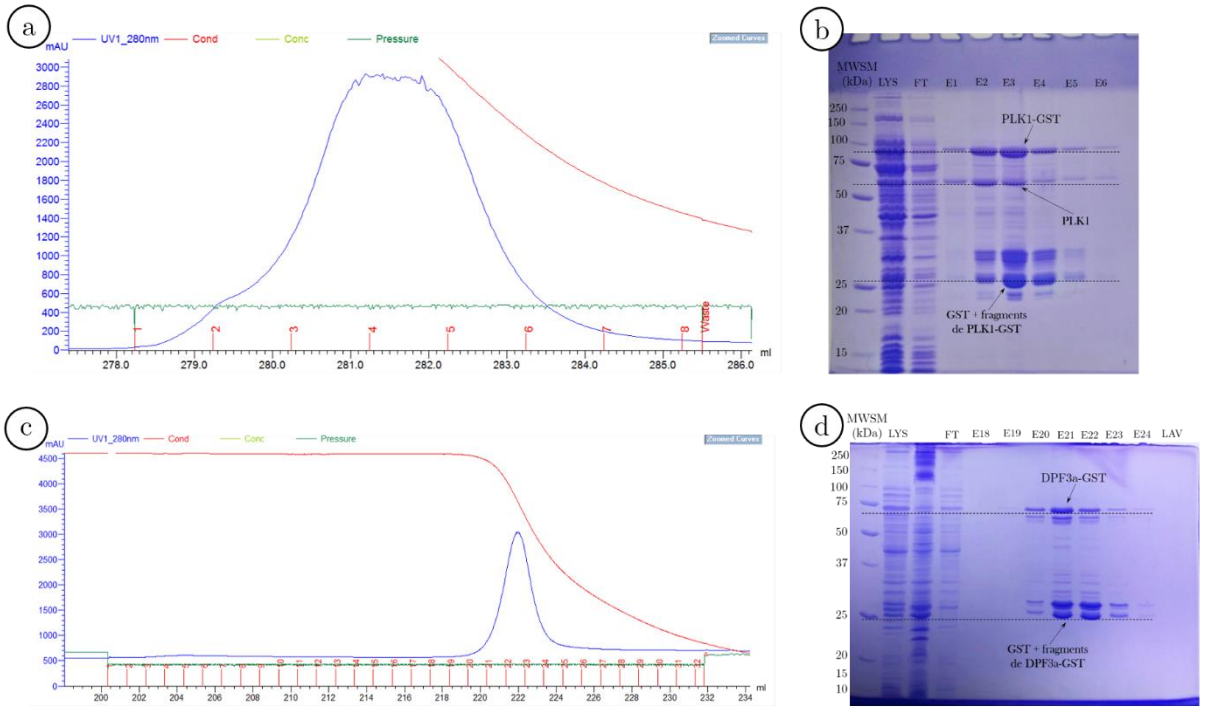

Figure 28 – (a) Chromatogrammes des protéines de fusion PLK1-GST et (c) DPF3a-GST purifiées par FPLC et leur gel SDS-PAGE 12 % associé : (b) gel de la PLK1-GST et (d) gel de la DPF3a-GST. Pour chaque gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through, LAV est le lavage de la colonne pré-packée en fin de purification et E correspond aux fractions d'élution.

IV. Caractérisation biophysique de la protéine de fusion DPF3a-GST

A. Spectroscopie d'absorption UV-visible

Pour la suite des analyses développées dans le présent point IV, la caractérisation se focalise exclusivement sur la protéine de fusion DPF3a-GST. L'analyse de l'absorption dans le domaine UV-visible (la gamme est ici restreinte de 200 à 400 nm) a été réalisée sur la DPF3a-GST purifiée par FPLC. L'absorbance est mesurée pour les fractions d'élution rassemblées et dessalées vers un tampon phosphate sur une colonne PD-10. Les spectres d'absorption (cf. Fig. 29) rendent compte d'un comportement d'absorption classique d'une protéine avec une absorbance maximale à une longueur d'onde de 280 nm. L'absorbance également marquée du tampon d'élution est due à la présence du GSH dont la forme oxydée GSSG commence à absorber aux alentours de 280 nm par formation d'un pont disulfure S-S entre les résidus Cys de deux molécules de GSH. Par ailleurs, comme cela est montré à la Fig. 29, cette absorption du GSSG peut souvent gêner la visualisation de la bande d'absorbance de la protéine d'intérêt qu'il convient préalablement de changer de tampon par dialyse ou sur une colonne PD-10.

La longueur d'onde caractéristique à 280 nm a pour origine l'absorption des résidus aromatiques et des résidus Cys impliqués dans des ponts disulfure.[108] Le tryptophane, dont le coefficient d'extinction molaire à 280 nm (de symbole ε) est le plus élevé (ε_{Trp} = 5690 $M^{-1}.cm^{-1}$), absorbe majoritairement, suivi par la tyrosine (ε_{Tyr} = 1280 $M^{-1}.cm^{-1}$) et enfin les cystéines sous leur forme oxydée au niveau du thiol de leur chaîne latérale (ε_{Cys} = 120 $M^{-1}.cm^{-1}$).[109] Il est à noter qu'il existe des espèces protéiques dont la structure primaire ne renferme aucun acide aminé aromatique et pour lesquelles la longueur d'onde d'absorbance maximale est déplacée à des valeurs plus basses associées à un phénomène d'absorption plus énergétique. Toutefois, ce n'est ni le cas de la DPF3a composée de 2 Trp, 10 Tyr ainsi que 12 Cys, ni celui de la GST dont la séquence renferme 4 Trp, 14 Tyr et 8 Cys. La contribution du *tag* GST aux valeurs d'absorbance enregistrées doit être prise en compte étant donné que son coefficient d'extinction molaire à 280 nm (ε_{GST} = 41640 $M^{-1}.cm^{-1}$) est 1.6 fois supérieur à celui de la DPF3a (ε_{DPF3a} = 25620 $M^{-1}.cm^{-1}$) ; lesdits coefficients ont été calculés à partir des valeurs renseignées pour les résidus Trp, Tyr et Cys déterminées empiriquement par Gill et von Hippel.[109]

Même si les résultats apparaissent reproductibles sur un ensemble d'échantillons, il convient d'en comparer le comportement d'absorption à celui des protéines clivées, autrement dit en l'absence de l'étiquette GST, et de la GST seule.

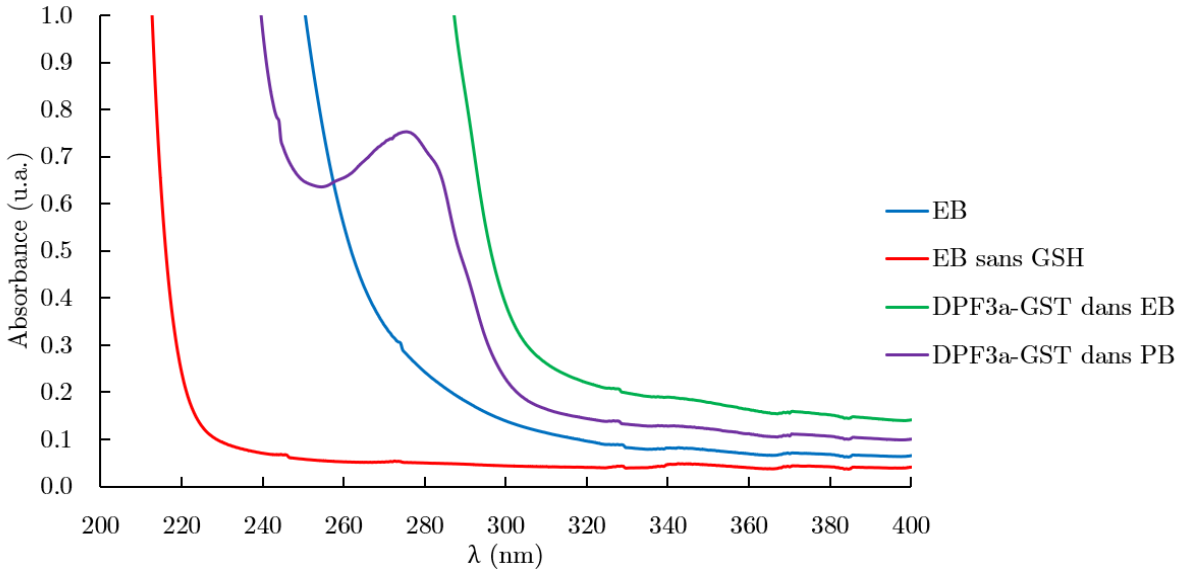

Figure 29 – Spectres d'absorption UV-visible de la protéine de fusion DPF3a-GST. L'évolution de l'absorbance (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). EB correspond au tampon d'élution (Tris-HCl 50 mM pH 8.0, GSH 10 mM) et PB au tampon phosphate 20 mM pH 8.0. Les mesures sont effectuées dans une cuvette en quartz.

B. Fluorescence intrinsèque du tryptophane (ITF)

L'analyse des spectres d'émission en ITF, dont la gamme est ici restreinte de 295 à 600 nm, a été réalisée uniquement sur la protéine de fusion DPF3a-GST purifiée par FPLC ; l'échantillon analysé correspond à la fraction dessalée. Le spectre d'émission de la DPF3a-GST (cf. Fig. 30) présente une signature typique de l'émission du tryptophane. L'excitation peut être établie à 280 nm, équivalant à la longueur d'onde d'absorption maximale (de symbole $\lambda_{abs\ max}$) de l'ensemble des résidus aromatiques et des cystéines. Toutefois, afin de limiter la contribution de la tyrosine et de la phénylalanine à la signature de fluorescence, la longueur d'onde d'excitation est fixée à 295 nm. Les résidus Trp chromophores y sont sélectivement excités étant donné qu'à cette longueur d'onde le coefficient d'extinction molaire du Trp est environ trente fois supérieur à celui de la Tyr.[110] Cela permet d'étudier le comportement des résidus Trp au sein de la protéine et vis-à-vis de son environnement. En effet, l'émission du Trp est dite solvatochromique, à savoir qu'elle dépend des conditions du solvant dans lequel se trouve la protéine, autrement dit de l'exposition des Trp au solvant. Cet effet a pour conséquence l'émission du Trp dans une gamme plus large de longueurs d'onde comprise entre 308 et 355 nm, reflétant l'influence de l'environnement et la dynamique des Trp au sein de la structure 3D. La longueur d'onde d'émission maximale (de symbole $\lambda_{ém\ max}$) est sujette à des déplacements bathochromes (vers le rouge), par interactions électrostatiques au sein d'un milieu polaire, ou hypsochromes (vers le bleu) au sein d'un environnement apolaire.[111,112] Ainsi, de la propriété d'émission du Trp excité à 295 nm, découlent des informations sur l'état de repliement de la protéine étudiée : pour des résidus Trp complètement exposés au solvant ($\lambda_{ém\ max} \sim 350$ nm), partiellement exposés au solvant ($\lambda_{ém\ max} \sim 340$ nm), enfouis mais en interaction avec des résidus polaires et/ou chargés ($\lambda_{ém\ max} \sim 315$-330 nm) et pour des Trp compris dans un noyau apolaire ($\lambda_{ém\ max} \sim 308$ nm). Ces déplacements caractéristiques s'accompagnent également de variations du rendement quantique et de la largeur de la bande d'émission.[112]

La DPF3a-GST possède une $\lambda_{ém\ max}$ à 335 nm. Il peut dès lors être considéré que les résidus Trp sont partiellement exposés à la surface de la protéine ou sont partiellement enfouis tout en interagissant avec des groupements polaires voisins. Etant donné que la GST contient deux fois plus de Trp dans sa séquence que la DPF3a, sa contribution au signal d'émission n'est pas à écarter. Il n'est pas impossible que la bande résultante rende exclusivement compte de l'état d'exposition des Trp de la GST. À des fins comparatives, il convient d'en confronter le comportement d'émission à celui des protéines clivées, autrement dit sans leur *tag* GST, et de la GST seule. L'étude de

la fluorescence relative aux résidus aromatiques (λ_{exc} = 280 nm), de la tyrosine (λ_{exc} = 274-275 nm) et de la phénylalanine (λ_{exc} = 257 nm) peut être aussi envisagée.

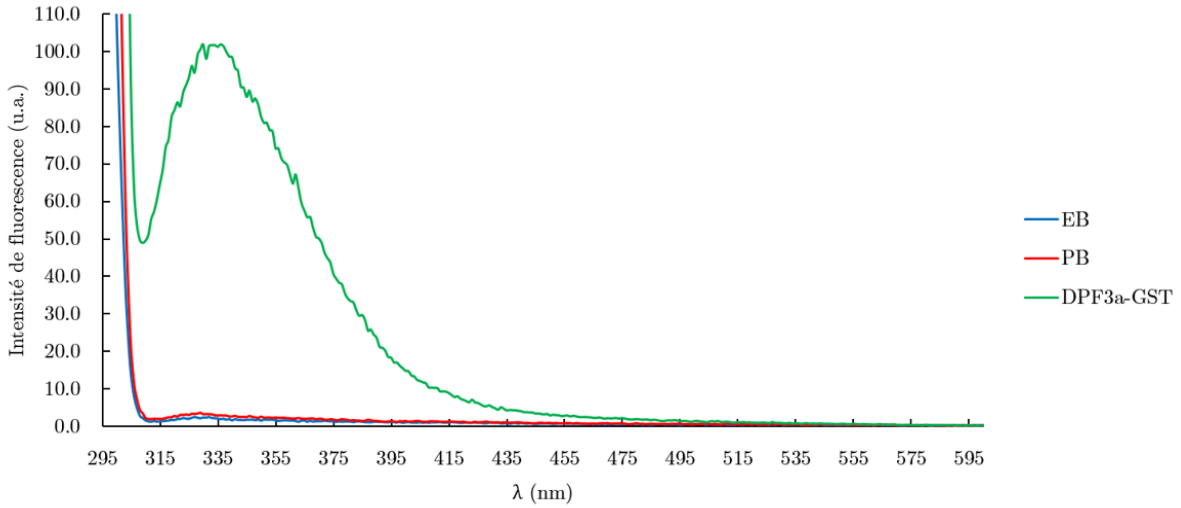

Figure 30 – Spectre d'émission en fluorescence intrinsèque du tryptophane de la protéine de fusion DPF3a-GST. L'évolution de l'intensité de fluorescence (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). EB correspond au tampon d'élution (Tris-HCl 50 mM pH 8.0, GSH 10 mM) et PB au tampon phosphate 20 mM pH 8.0. Les mesures sont effectuées à $\lambda_{excitation}$ = 295 nm et sw = 10 nm.

C. Dichroïsme circulaire (CD)

L'analyse en CD a été réalisée uniquement sur la protéine de fusion DPF3a-GST, purifiée par FPLC et dessalée, de manière qualitative. La technique de CD des protéines se base soit sur l'absorbance dans l'UV lointain (entre 180 nm et 260 nm) soit dans l'UV proche (entre 250 nm et 330 nm) des acides aminés. Étant donné que le faisceau incident se compose de deux types de lumière polarisée circulairement, l'une vers la gauche et l'autre vers la droite, et qu'une protéine est une substance asymétrique (chirale), la préférence d'absorption est mesurable en établissant la différence d'absorbance entre ces deux types de lumière. Concernant l'étude spectroscopique dans l'UV lointain, elle offre des informations quantitatives et qualitatives sur les structures secondaires (cf. Fig. 31).[113,114] Le spectre résultant (cf. Fig. 32) montre une signature typique d'une structure ordonnée pour laquelle il est néanmoins difficile de distinguer le signal des hélices α, se manifestant par deux bandes d'ellipticité (de symbole θ) négative à des longueurs d'onde de 222 et 208 nm ainsi qu'une positive à 193 nm, et celui des feuillets β antiparallèles, caractérisé par une bande négative à 218 nm et une positive à 195 nm.[113,115]

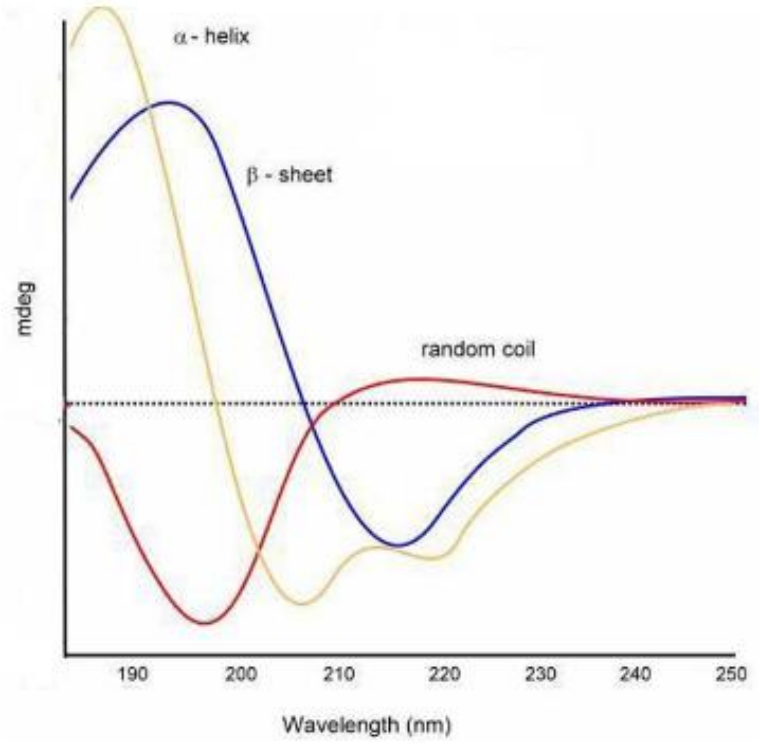

Figure 31 – Signatures spectrales typiques des structures secondaires des protéines en CD.[115]

À la vue de l'allure générale de la courbe de tendance et de ses extrema, le motif d'hélice α semble tout de même prépondérant avec une première bande négative vers 223 nm, une seconde plus marquée vers 206 nm ainsi qu'une forte augmentation de l'ellipticité en se rapprochant de 195 nm. Il n'est pas impossible qu'il y ait une contribution non négligeable de la GST à cette signature étant donné qu'il s'agit d'une protéine dont la structure 3D est complètement résolue et majoritairement composée d'hélices α.[116] En outre, il est remarquable de noter qu'aucun signal propre au désordre n'est détecté pour lequel une ellipticité quasiment nulle est observée au-dessus de 210 nm ainsi qu'une bande négative aux alentours de 200-195 nm.[113] Ce résultat implique que la GST aurait un rôle ordonnant sur la structure de la DPF3a pour peu que cette dernière soit considérée comme désordonnée.[117] Sous sa forme fusionnée, la GST entraînerait une modification conformationnelle de la DPF3a vers un repliement se rapprochant d'un globule. Toutefois, cette hypothèse reste à confronter aux spectres de CD obtenus pour la DPF3a clivée du *tag* GST afin d'en comparer les signatures. L'expérience doit être également reproduite sur d'autres échantillons de la DPF3a-GST.

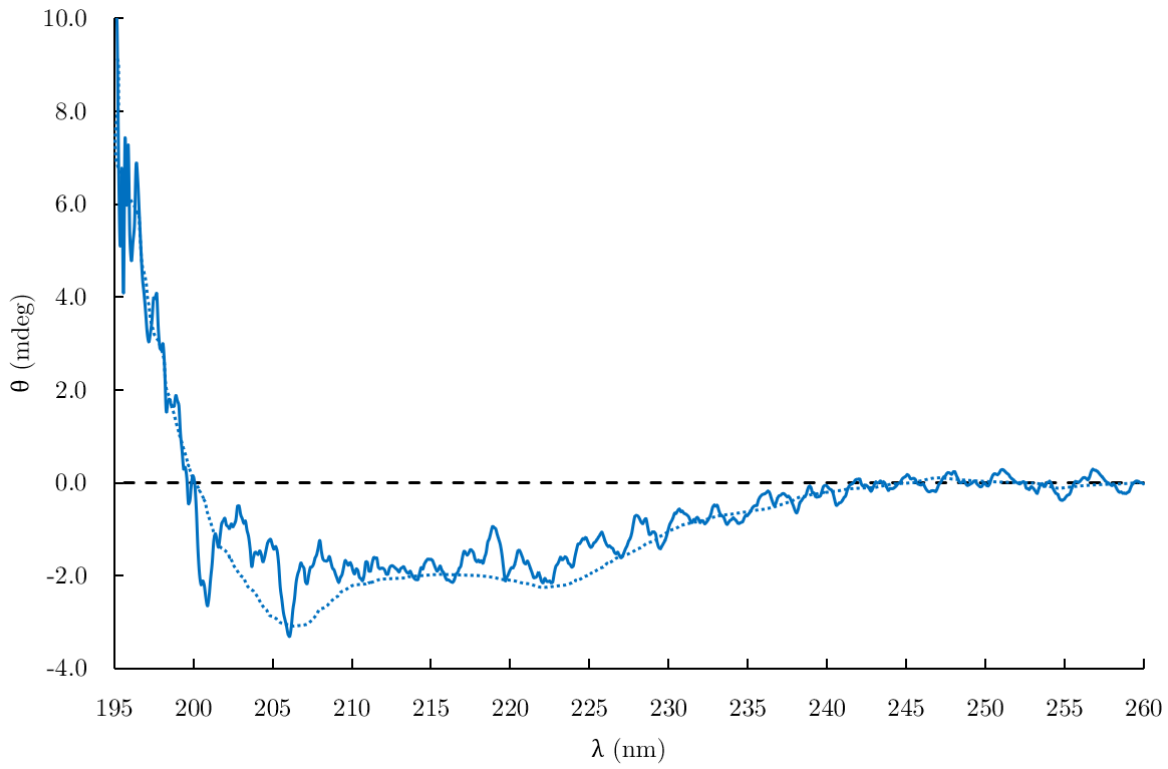

Figure 32 – Spectre de dichroïsme circulaire de la protéine de fusion DPF3a-GST. L'évolution de l'ellipticité θ (exprimée en mdeg) est représentée en fonction de la longueur d'onde λ (exprimée en nm). L'allure générale de la tendance du spectre est représentée par un trait en pointillé.

D. Diffusion dynamique de la lumière (DLS)

L'analyse du diamètre hydrodynamique en DLS a été réalisée uniquement sur la protéine de fusion DPF3a-GST purifiée par FPLC et dessalée. Au sein d'une solution aqueuse, les molécules de soluté se déplacent selon un mouvement brownien. La vitesse de déplacement dépend de la taille desdites molécules : plus elle est volumineuse, plus elle est lente. Par irradiation avec un faisceau de lumière monochromatique, une fraction de ce faisceau est diffusée dont l'intensité de diffusion varie en fonction du temps de par la modification continue de la distance entre les particules. L'analyse DLS permet donc de lier ces variations d'intensité de diffusion au mouvement des particules et, par conséquent, à leur taille hydrodynamique notamment par l'équation de Stokes-Einstein.[118]

Les résultats sont présentés par le diamètre hydrodynamique moyen (de symbole D_h), mesuré à température ambiante, et la déviation standard associée. Il ressort un D_h = 77.2 ± 8.4 nm dont la valeur est d'un ordre de grandeur supérieur à ce qui est attendu pour une protéine globulaire type (D_h compris entre 3 et 10 nm) ou une protéine dénaturée ou une IDP (D_h compris entre 5 et 22 nm) ; les gammes de D_h

varient aussi en fonction de la masse moléculaire de la protéine. À titre comparatif, l'albumine de sérum bovin (abrégée BSA pour *bovine serum albumin*), possédant une masse moléculaire comparable à la DPF3a-GST (M_{BSA} = 66.3 kDa), est caractérisée par un D_h = 16.4 nm sous sa forme dénaturée (guanidinium-HCl 6 M, 25 °C).[119]

Étant en conditions non dénaturantes, il semble qu'il y ait présence d'agrégats protéiques malgré la potentielle stabilisation conformationnelle induite par la GST. Cependant, l'échantillon de la DPF3a-GST a connu un cycle de congélation-décongélation, ce qui en réduit la qualité, et présentait après décongélation des particules en suspension éliminées par centrifugation et filtration avant analyse. Les IDPs sont connues dans la littérature pour avoir une propension accrue au phénomène d'agrégation, aussi bien *in vivo* que *in vitro*, notamment sous une organisation en structures amyloïdes. Même dans le cas d'une protéine bien repliée, ces particules de grande dimension hydrodynamique peuvent s'expliquer par la formation en solution d'agrégats de taille supérieure due à l'âge de l'échantillon protéique ou aux conditions *in vitro*, autrement dit la composition du tampon de stockage.[120] Bien que cette expérience tende à montrer que la DPF3a-GST s'agrège facilement, elle reste encore à reproduire sur d'autres fractions de la DPF3a-GST et à confronter aux valeurs de D_h obtenues pour d'autres constructions de fusion ainsi qu'aux protéines clivées de leur *tag* GST.

E. Chromatographie d'exclusion stérique (SEC)

Dans cette même optique de caractérisation des dimensions hydrodynamiques des protéines, l'analyse du volume d'élution en SEC (sur une colonne HiLoad™ 16/600 Superdex™ 200) a été réalisée sur la protéine de fusion DPF3a-GST purifiée par FPLC et éluée dans le tampon FB dont la composition est la suivante : Tris-HCl 50 mM pH 7.5, NaCl 150 mM et $CaCl_2$ 1 mM. Le tampon FB a été retenu afin de pouvoir comparer les résultats obtenus pour la DPF3a-GST et la GST clivée à ceux de la DPF3a clivée dont les échantillons sont initialement récoltés dans ledit tampon. Les volumes d'élution sont rassemblés à la Table 1, préalablement corrigés par la valeur du volume mort (V_{mort} = 46.7 mL), et mis en relation avec la masse moléculaire estimée à partir de la droite d'étalonnage (cf. Fig. 33). La droite d'étalonnage a été établie expérimentalement à partir de standards protéiques globulaires pour filtration sur gel.

Les valeurs de volume d'élution sont indiquées pour deux échantillons distincts de la DPF3a-GST dont l'un (référencé « vieux ») a été stocké douze jours à température ambiante après l'analyse par DLS et l'autre (référencé « frais ») a été

décongelé avant son dépôt sur colonne ; ce dernier n'a pas été analysé en DLS. Ils proviennent de la même purification. Les deux échantillons présentent des populations d'élution communes. La première sort systématiquement dans le volume mort et excède la limite de séparation de la colonne (> 2 MDa) tandis que la seconde est associée à une masse moléculaire estimée d'environ 45-49 kDa. Cette seconde bande correspond vraisemblablement à la GST ou à un fragment de la DPF3a-GST dont la présence avait déjà été identifiée sur gel (cf. 28d).

Protéine/échantillon	Bande	$V_{élution\ corrigé}$ (mL)	$M_{estimée}$ (Da)	Hypothèse
DPF3a-GST (vieux)	1ère	-6.9	> 2000000	Agrégats
	2ème	33.8	45627	GST
	3ème	70.3	1383	?
DPF3a-GST (frais)	1ère	-6.9	> 2000000	Agrégats
	2ème	22.9	129616	DPF3a-GST
	3ème	32.9	49735	GST
	4ème	62.0	3063	?
	5ème	75.4	849	?

Table 1 – Valeurs de volume d'élution corrigé (exprimé en mL) par le volume mort (V_{mort} = 46.7 mL) et de masse moléculaire estimée (exprimée en Da) en fonction des populations de bande de la protéine de fusion DPF3a-GST stockée douze jours à température ambiante (vieux) et fraîchement décongelée (frais). Les hypothèses associées à la nature de chacune des bandes sont également indiquées.

Pour la première bande, les IDPs sont réputées pour éluer à des volumes bien plus bas que ceux attendus pour un homologue protéique globulaire. Cette propriété s'explique par les dimensions hydrodynamiques plus grandes des IDPs de par leur état désordonné et leur flexibilité de chaîne.[121] Par conséquent, la pertinence de la comparaison de protéines supposées intrinsèquement désordonnées par rapport à une échelle globulaire trouve ses limites, en particulier quant à l'estimation des masses moléculaires. Toutefois, le déplacement de bande vers un volume d'élution très bas est ici d'une trop grande amplitude pour être expliqué par cette propriété. Il apparaît plus cohérent de penser qu'il s'agit d'agrégats de la DPF3a-GST. Cette proposition est notamment soutenue par l'absence de la bande probablement caractéristique de la DPF3a-GST, dont $M_{estimée} \sim$ 129 kDa, dans l'échantillon « vieux » pour lequel le phénomène d'agrégation a eu le temps de s'établir pendant la période de stockage. Par ailleurs, cet échantillon, contrairement au « frais », présentait un caractère légèrement turbide avant injection. Enfin des bandes de plus faible intensité associées à des volumes

d'élution élevés et de très faibles masse moléculaires (~ 800-3000 Da) sont également détectées dans les deux échantillons. Bien qu'il puisse s'agir de fragments peptidiques de la DPF3a-GST ou de la GST, leur identité demeure jusqu'à présent inconnue.

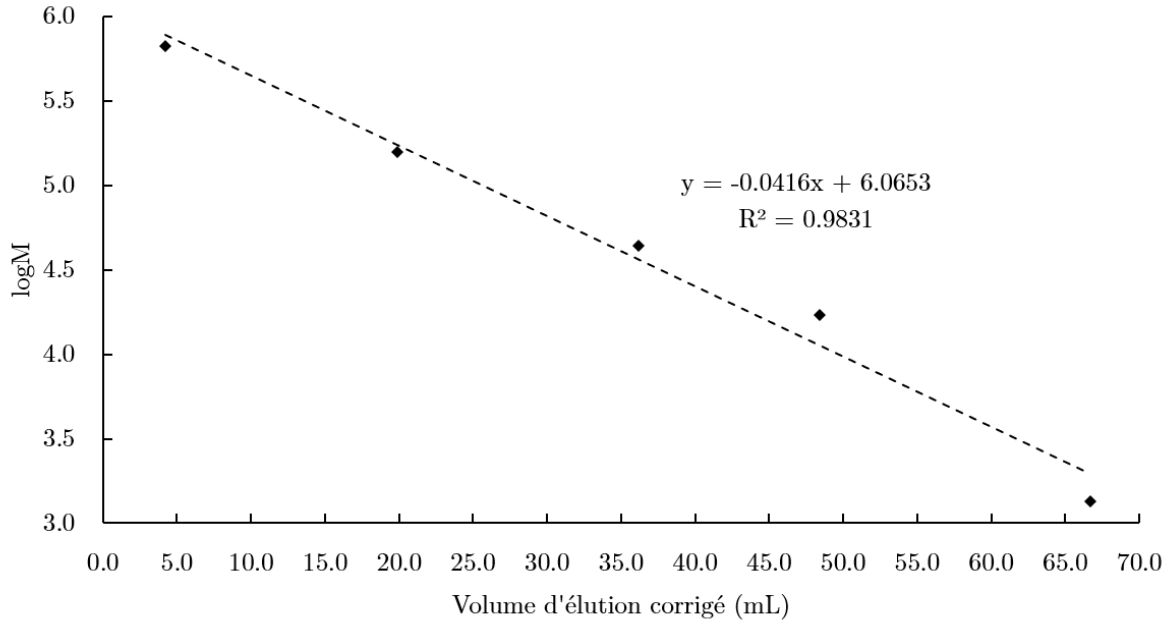

Figure 33 – Droite d'étalonnage (en pointillés) de la chromatographie d'exclusion stérique (sur une colonne HiLoad™ 16/600 Superdex™ 200) à partir de standards protéiques globulaires : thyroglobuline (M = 670 kDa), γ-globuline (M = 158 kDa), ovalbumine (M = 44 kDa), myoglobine (M = 17 kDa) et vitamine B12 (M = 1350 Da). L'évolution du logarithme de la masse moléculaire est représentée en fonction du volume d'élution (exprimé en mL) corrigé par le volume mort (Vmort = 46.7 mL).

Section b

—

Purification et caractérisation de la DPF3a clivée

I. Clivages par chromatographie d'affinité au format *batch*

L'efficacité et la spécificité du clivage des constructions de la DPF3 ont été testées dans un premier temps sur la protéine de fusion PHD-1+2-GST avant l'extension du protocole à la DPF3a-GST. Le clivage de l'étiquette GST est exécuté à l'aide de la protéase facteur Xa. Le gel SDS-PAGE résultant (cf. Fig. 34a) révèle que la protéine de fusion a été clivée spécifiquement avec succès. La bande la plus apparente est retrouvée aux alentours de 15 kDa dans l'ensemble des fractions d'élution récoltées et correspondrait au domaine PHD-1+2 sans GST ($M_{PHD\text{-}1+2}$ = 12.6 kDa). Entre 50 et 75 kDa, des reliquats protéiques sont également observés avec une intensité de bande semblable à celle de PHD-1+2. Ces vestiges peuvent être dus à des contaminations de purifications précédentes au format *batch* étant donné que la matrice est récupérée et régénérée entre chaque procédure.

Par application des mêmes conditions opératoires au clivage de la DPF3a-GST, une unique bande est observée sur le gel SDS-PAGE associé (cf. Fig. 34b). Sa localisation vers 15 kDa ne concorde ni avec la DPF3a-GST ni avec la DPF3a clivée et ni au *tag* GST. De par la similitude de motif de migration avec le gel de la construction PHD-1+2, il n'est pas impossible qu'il s'agisse du domaine PHD-1+2, à trop haute concentration, clivé à la purification précédente et ayant saturé la matrice. Cependant, en dehors de cette contamination, l'absence de toute trace de la DPF3a, qu'elle soit sous forme fusionnée ou clivée, demeure curieux. À la vue des contaminations observées au format *batch*, ce dernier semble à exclure pour la procédure de purification et de clivage pour laquelle la FPLC apparaît plus adaptée.

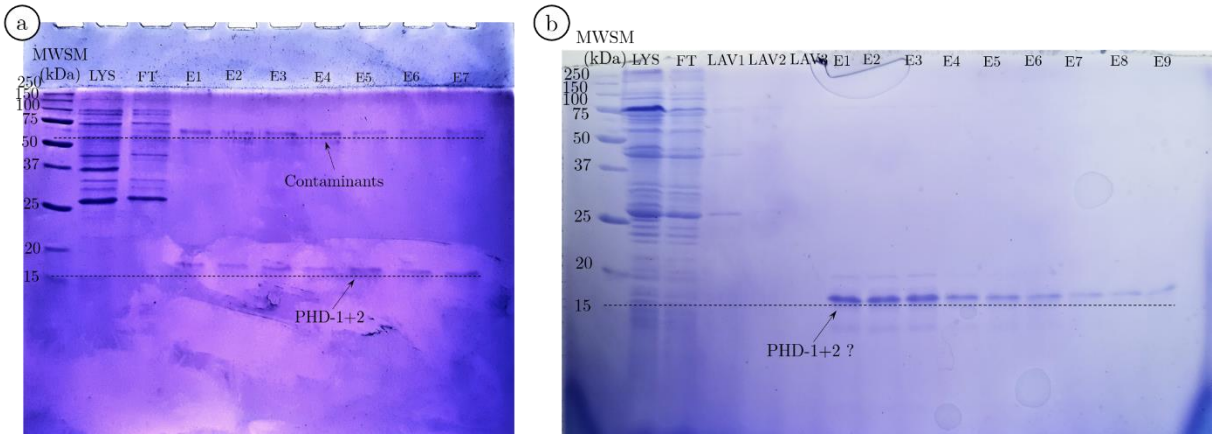

Figure 34 – (a) Gels SDS-PAGE 15 % des protéines de fusion PHD-1+2-GST et (b) DPF3a-GST clivées au facteur Xa et purifiées par chromatographie d'affinité au format batch. Pour chaque gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through, LAV correspond aux différents lavages de matrice avant et après élution et E correspond aux fractions d'élution.

II. Clivages par FPLC et comportement singulier par SDS-PAGE

Bien que les résultats obtenus au format *batch* se soient avérés peu concluants vis-à-vis du clivage de la DPF3a-GST, la procédure est adaptée sur un système FPLC afin d'en faciliter la reproductibilité. Les deux chromatogrammes résultants et les gels SDS-PAGE associés (cf. Fig. 35) ont été obtenus selon les mêmes conditions de purification et de migration ainsi qu'à partir de différents culots de surexpression de DPF3a FL. Par observation des chromatogrammes, la première bande d'élution (la plus à gauche) correspond à la sortie de la colonne de la DPF3a clivée par le facteur Xa tandis que la seconde (la plus à droite) est relative à l'étiquette GST. Il est à noter que, pour l'élution de la DPF3a clivée, le tampon de clivage au facteur Xa est utilisé. L'existence de deux bandes distinctes dont l'une n'élue qu'après incubation avec la protéase spécifique au vecteur de clonage de la protéine rend compte de l'efficacité de la procédure. Aussi, il se dégage une tendance commune quant à l'élution de la DPF3a clivée, à savoir qu'elle est systématiquement accompagnée d'une augmentation de la conductivité. La diminution progressive du signal relatif à la conductivité durant l'élution de la GST clivée est identique à celle observée pour la DPF3a-GST. Cela n'est en aucun point étonnant étant donné que le même tampon d'élution au GSH est employé dans les deux cas. Les différences d'intensité des bandes d'absorbance à 280 nm entre les chromatogrammes sont dues à des variations du niveau de surexpression de la protéine de fusion.

Néanmoins, l'analyse SDS-PAGE se révèle singulière. Alors qu'au sein des échantillons du pic d'élution de la GST cette dernière est nettement identifiée par l'apparition marquée d'une bande aux alentours de 25 kDa vis-à-vis du MWSM, la DPF3a ne se manifeste par aucune bande dans aucune des fractions d'élution migrées. Cette absence de la DPF3a clivée sur gel SDS-PAGE est significative de par sa reproductibilité sur l'ensemble des purifications et migrations effectuées. Ce comportement est d'autant plus curieux à la lueur des chromatogrammes pour lesquels une bande d'élution étant, jusqu'à ce stade, identifiée comme la DPF3a sans son *tag* est détectée. Par ailleurs, les fractions de GST ne dévoilent aucune trace de la DPF3a-GST, montrant une fois de plus que clivage il y a bel et bien eu.

Pour expliquer ce phénomène, plusieurs hypothèses sont avancées. Premièrement, l'hypothèse d'absence de liaison ou d'une interaction instable dans le temps entre le bleu de Coomassie (le colorant du gel SDS-PAGE) et la DPF3a est formulée. Des comportements d'interaction protéine-bleu de Coomassie dynamiques ont, en effet, déjà été répertoriés dans la littérature et sont le corollaire d'une structure

flexible et désordonnée.[122] Cependant, cette hypothèse semble d'ores et déjà peu probable étant donné que la DPF3a contient des résidus chargés (36 Arg, 23 Lys, 23 Asp et 41 Glu), a priori capables de lier le bleu de Coomassie, dans sa séquence et que son absence sur gel n'est pas limitée à ce type de coloration. En effet, à titre comparatif, un gel a été réalisé selon les mêmes paramètres d'analyse à l'exception de la méthode de coloration : une révélation à l'imidazole-zinc (cf. Fig. 36). Le gel coloré par immersion consécutive dans l'imidazole et une solution de chlorure de zinc (II), à l'instar du bleu de Coomassie, n'a révélé aucune trace de la DPF3a clivée. Le *tag* GST est quant à lui toujours bien identifiable vers 25 kDa.

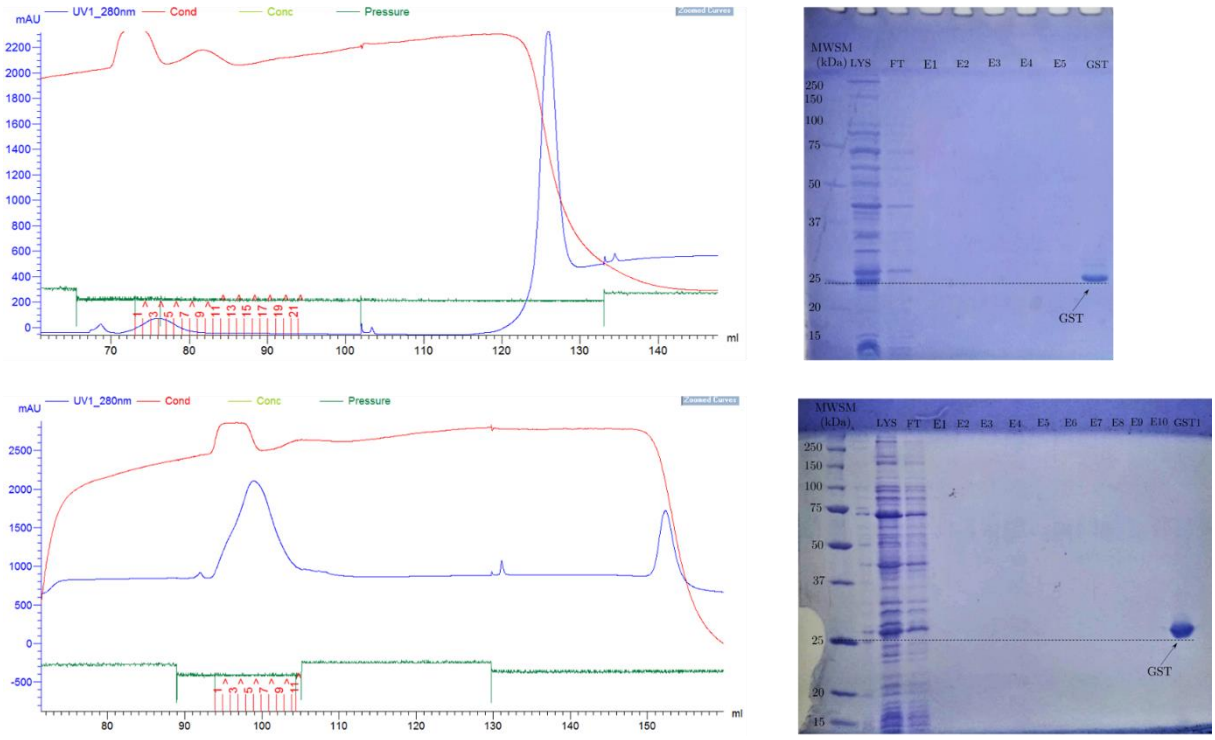

Figure 35 – Chromatogrammes de la protéine de fusion DPF3a-GST clivée au facteur Xa et purifiée par FPLC ainsi que leur gel SDS-PAGE 12 % associé. Pour chaque gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through, E correspond aux fractions d'élution de la DPF3a clivée et GST correspond aux fractions d'élution de la GST clivée.

Secondement, il est possible que, de par sa sensibilité à la protéolyse, la DPF3a ait été lors du processus de clivage fragmentée en peptides de plus faibles masses moléculaires (en-deçà de 10 kDa) ne permettant dès lors pas leur visualisation sur un gel de polyacrylamide. Néanmoins, cette hypothèse est également écartée étant donné que les protéases endogènes de *E. coli* ont été éliminées, selon toute vraisemblance, dans le FT avant incubation au facteur Xa et que ce dernier est spécifique au site de clivage entre la DPF3a et le *tag* GST. La DPF3a ne contient pas dans sa structure

primaire la séquence reconnue par le facteur Xa. En outre, une analyse des échantillons d'élution de la DPF3a par électrophorèse sur gel de Tris-Tricine n'a révélé aucune trace de fragments peptidiques entre 2 et 15 kDa. L'utilisation de la tricine permet d'obtenir une meilleure résolution de séparation des peptides ou protéines dans la gamme de 5 à 20 kDa à de plus faibles concentrations d'acrylamide pour la réticulation du gel.[123]

Troisièmement, l'hypothèse de l'agrégation protéique est à nouveau proposée. Au cours du clivage, la DPF3a perdrait son étiquette GST ayant des propriétés de stabilisation sur la structure de la protéine de fusion ; la GST est conformationnellement stable à l'état clivé.[117] De par son hypothétique nature désordonnée, la DPF3a clivée s'associerait en oligomères ou agrégats récoltés durant la première étape d'élution. Ainsi, bien qu'en condition dénaturante, les agrégats ne seraient pas capables de migrer à travers le réseau réticulé de polyacrylamide, expliquant l'absence de bande relative à la DPF3a. Cela signifie également que les agrégats sont supposément d'une masse moléculaire totale supérieure à 250 kDa ; il s'agit de la limite supérieure de séparation du gel. Cette hypothèse reste à vérifier et à confronter à d'autres résultats d'analyse biophysique.

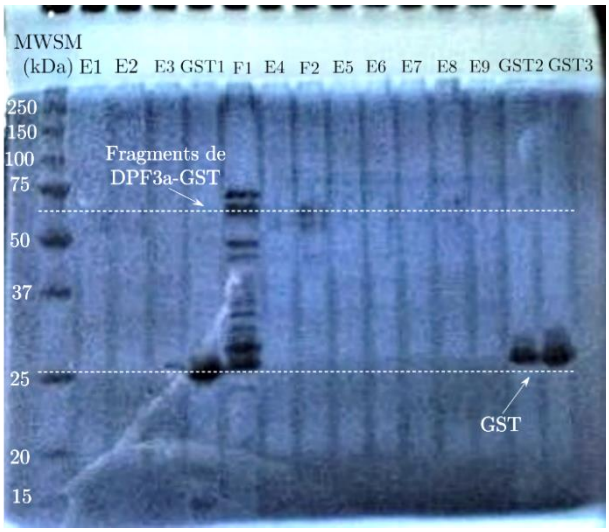

Figure 36 – Gel SDS-PAGE 12 % de la protéine de fusion DPF3a-GST clivée au facteur Xa et purifiée par FPLC. MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), E correspond aux fractions d'élution de la DPF3a clivée, F aux fractions d'élution de la DPF3a-GST et GST correspond aux fractions d'élution de la GST clivée.

III. Identification protéique par spectrométrie de masse (LC-ESI-MS/MS)

À la vue de l'insuffisance des analyses sur gel SDS-PAGE ainsi que par spectroscopie UV-visible et afin de confirmer de manière univoque la présence ou l'absence de la protéine d'intérêt dans les échantillons d'élution récoltés, ces derniers sont digérés en phase liquide par la trypsine (procédure FASP pour *filter-aided sample preparation*) et analysés par spectrométrie de masse. La DPF3a est bel et bien identifiée en tant que protéine majoritaire dans chacune des fractions soumises ; aucune trace de la GST n'est détectée. Néanmoins, les analyses révèlent systématiquement un faible pourcentage de couverture de séquence, compris entre 18 et 25 %, pour lequel les peptides uniques générés proviennent exclusivement du début de la séquence (à l'extrémité N-terminale) et sont limités à la première centaine d'acides aminés. Il apparaît remarquable que le dernier peptide exclusif identifié est systématiquement V^{131}DAREEESIQEIQR^{144} ou E^{135}EESIQEIQR^{144} ; l'ultime acide aminé détecté correspond toujours à Arg^{144}. Le résultat type de l'analyse par LC-ESI-MS/MS de la DPF3a clivée est illustré à la Fig. 37.

(1) M**ATVIHNPLK** ↓ **ALGDQFYK**↓EA IEHCRSYNSR LCAERSVRLP
(41) FLDSQTGVAQ NNCYIWMEKR ↓ **HRGPGLAPGQ LYTYPAR**↓CWR
(81) KKR↓**RLHPPED PK**↓LR↓**LLEIKP EVELPLKK**↓DG FTSESTTLEA
(121) LLRGEGVEKK ↓ **VDAREEESIQ EIQR**↓VLENDE NVEEGNEEED
(161) LEEDIPKRKN RTRGRARGSA GGRRRHDAAS QEDHDKPYVC
(201) DICGKRYKNR PGLSYHYAHT HLASEEGDEA QDQETRSPPN
(241) HRNENHRPQK GPDGTVIPNN YCDFCLGGSN MNKKSGRPEE
(281) LVSCADCGRS AHLGGEGRKE KEAAAAARTT EDLFGSTSES
(321) DTSTFHGFDE DDLEEPRSCR GRRSGRGSPT ADKKGSC

Figure 37 – Représentation schématique des peptides uniques types identifiés dans la séquence de la DPF3a clivée par spectrométrie de masse LC-ESI-MS/MS. Lesdits peptides sont mis en évidence en rouge. Entre chaque peptide, les sites de clivage à la trypsine sont délimités par des flèches bleues.

Une nouvelle fois, la question de l'agrégation protéique est posée. Elle pourrait s'effectuer au niveau du domaine C-terminal, ce qui en restreindrait l'accessibilité à la trypsine pour la digestion peptidique. La trypsine lyse les liaisons peptidiques après les résidus Lys ou Arg à moins que l'acide aminé adjacent (vers la position C-terminale) soit une proline.[124] Par soumission de la séquence de la DPF3a à l'outil *PeptideCutter*

du serveur ExPASy[125], 52 sites de clivage spécifiques à la trypsine sont prédits dont 28 après Arg^{144} ; les peptides générés expérimentalement sont également retrouvés. Cette prédiction conforte l'idée que, dans le cas de la DPF3a clivée, la trypsine pourrait être entravée dans son rôle de peptidase au-delà du résidu Arg^{144} en direction de l'extrémité C-terminale.

IV. Comportement singulier en spectroscopie d'absorption

Étant donné l'absence de la DPF3a clivée sur les gels SDS-PAGE et l'existence d'une bande d'élution sur les chromatogrammes FPLC, l'analyse de l'absorption dans le domaine UV-visible a été réalisée sur les mêmes fractions d'élution. Il s'agit également de vérifier la présence de la protéine d'intérêt par détection d'une signature spectroscopique typique d'une protéine (contenant des résidus aromatiques). À nouveau, similairement à l'analyse SDS-PAGE, les spectres (cf. Fig. 38) ne sont pas ceux attendus. Tandis que la GST clivée absorbe classiquement à $\lambda_{\text{abs max}} \sim 280$ nm (à l'instar de la DPF3a-GST), l'ensemble des fractions choisies de la DPF3a présentent une absorbance maximale comprise entre 254 et 260 nm. Certains échantillons, issus des différentes purifications, présentent deux populations d'absorption, ce qui n'est pas non plus commun chez les protéines : une première dans une gamme de longueurs d'onde comprise entre 250 et 255 nm et une seconde vers 270-275 nm. Il est probable que la bande observée en majorité vers 256-260 nm soit due au recouvrement de ces deux populations ; d'ailleurs certains signaux ont un profil patatoïdal entre 255 et 270 nm. L'apparition de la seconde population est plus marquée en fin d'élution dans les dernières fractions récupérées.

Complémentairement, le suivi de l'absorbance UV à d'autres longueurs d'onde que celle à 280 nm sur les chromatogrammes précédemment présentés donne aussi à voir cette tendance. Les courbes d'absorbance enregistrées à 280, 265 et 254 nm pour une purification type avec clivage de la DPF3a (cf. Fig. 39a) montrent que les bandes d'élution à 265 et 254 nm ont une absorbance beaucoup plus élevée qu'à 280 nm. L'absorbance à 254 nm est supérieure à celle à 265 nm. Comparativement, la tendance inverse est observée pour la GST clivée (cf. Fig. 39b). La bande d'élution à 280 nm est caractérisée par la plus forte absorbance, suivie de celle à 265 nm et la plus faible à 254 nm. Concernant la DPF3a-GST (cf. Fig. 39c), son profil d'absorbance semble être le fruit de la combinaison de celui de la DPF3a et de la GST. L'absorbance des trois bandes d'élution est équivalente. Ce comportement spectroscopique singulier semble caractéristique de la DPF3a à l'état clivé, en regard de la DPF3a-GST, étant donné la

reproductibilité vis-à-vis des signatures observées. Aucune fraction n'absorbe significativement à 280 nm malgré les résidus Trp et Tyr contenus dans la séquence de la DPF3a.

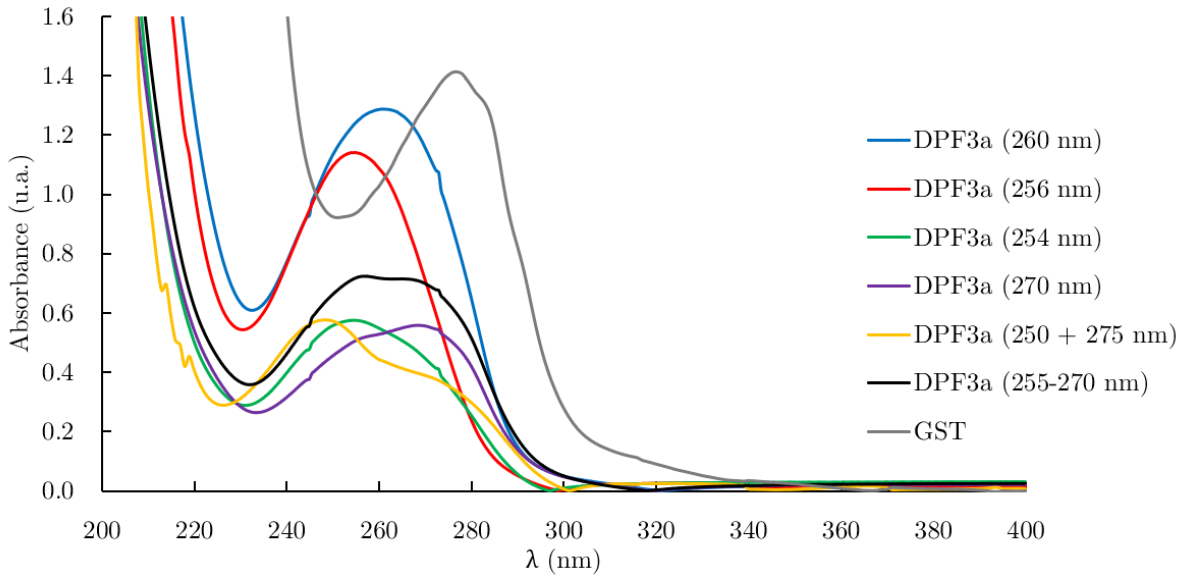

Figure 38 – Spectres d'absorption UV-visible des protéines DPF3a et GST clivées. L'évolution de l'absorbance (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). À chaque spectre de la DPF3a est associée la longueur d'onde ou la gamme de longueurs d'onde caractéristique. Les mesures sont effectuées dans une cuvette en quartz pour les courbes relatives à la DPF3a et dans une UV-cuvette pour la GST.

Pour expliquer ces signatures spectroscopiques d'absorption, plusieurs hypothèses sont développées. Premièrement, il est connu dans la littérature relative à la biochimie des protéines que l'apparition d'une bande de forte absorbance vers 260 nm est synonyme d'une contamination par l'ADN ; il peut aussi s'agir d'ARN dans certains cas moins couramment rencontrés. En effet, les acides nucléiques sont caractérisés par une longueur d'onde maximale d'absorption à 260 nm et un coefficient d'extinction molaire élevé.[126] Étant donné l'implication de la DPF3 dans le complexe de remodelage de la chromatine BAF, il n'est pas impossible qu'elle puisse interagir favorablement avec l'ADN. Une complexation avec l'ADN pourrait également expliquer l'absence de bande sur les gels SDS-PAGE.

Secondement, le phénomène d'agrégation protéique peut provoquer l'apparition d'artefacts sur les mesures d'absorbance, ce qui en modifie l'allure des spectres générés. Cela est dû à la différence de dimension entre une protéine et un agrégat de protéines. Cela concorderait avec la formation d'agrégats de la DPF3a trop volumineux, incapables de traverser la maille de polyacrylamide lors des analyses par SDS-PAGE.

Troisièmement, la protéine clivée pourrait interagir avec des ions métalliques présents en solution, provoquant un déplacement de la $\lambda_{abs\ max}$. Des déplacements ou effets batho-, hypso-, hyper- ou hypochromes en absorption UV-visible ont déjà été observés pour des complexes métal-protéine par transfert de charge du ligand au métal (abrégé LMCT pour *ligand to metal charge transfer*), notamment par l'intermédiaire des cystéines.[127–129] Cette hypothèse reste cohérente avec l'organisation de séquence de la DPF3a contenant deux ZNF, C_2H_2 et PHD-1/2, capables de lier les cations Zn^{2+} grâce aux résidus Cys et His. En outre, le tampon d'élution dans lequel sont contenus les échantillons de la DPF3a clivée contient du Ca^{2+} en concentration 1 mM. Une potentielle interaction avec ces cations calciques divalents n'est pas à écarter.

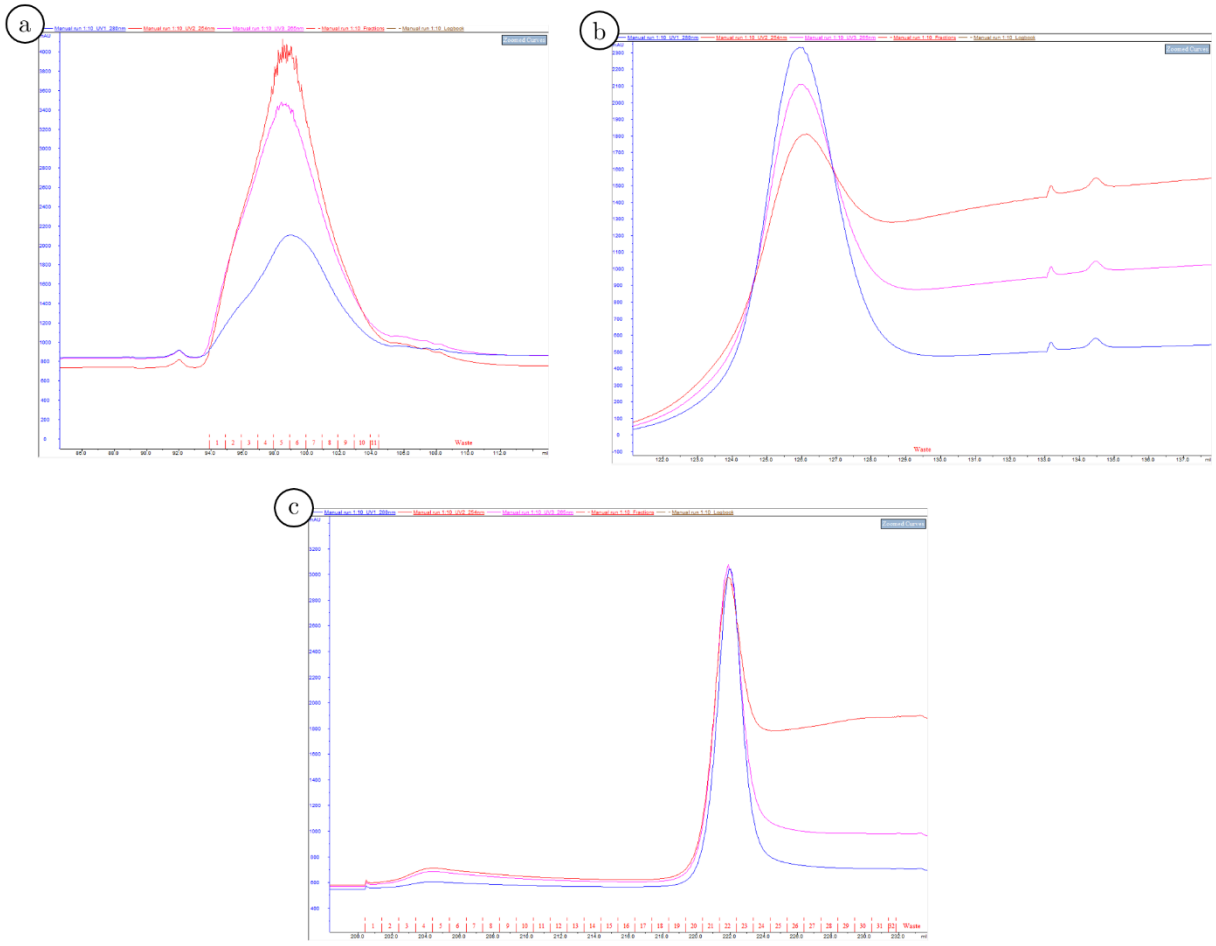

Figure 39 – (a) Chromatogrammes et courbes d'absorption associées de la protéine DPF3a clivée, (b) de la GST clivée et (c) de la protéine de fusion DPF3a-GST purifiées par FPLC. L'évolution de l'absorbance (exprimée en milli-unités arbitraires mu.a.) est représentée en fonction du volume d'élution (exprimé en mL). Pour chaque chromatogramme, la courbe bleue suit l'absorbance à 280 nm, la courbe rose à 265 nm et la courbe rouge à 254 nm.

V. Investigation de l'hypothèse de contamination par l'ADN

A. Traitement à la DNase I

Dans un premier temps, afin de confirmer ou d'infirmer la validité de l'hypothèse de la contamination par l'ADN, la purification par FPLC est réitérée en procédant à un traitement préalable du lysat par la DNase I. La DNase I (pour désoxyribonucléase I) est une endonucléase responsable de la fragmentation de l'ADN en plus petits polynucléotides par hydrolyse des liaisons phosphodiester.[130] À des fins de reproductibilité, le traitement à la DNase I est répété lors d'une seconde purification. Les chromatogrammes et gels SDS-PAGE associés de ces deux purifications (cf. Fig. 40) donnent à voir un comportement en aucun point distinct vis-à-vis de la DPF3a clivée sans traitement à la DNase I. Le pic d'élution de la DPF3a, qui est uniquement présenté sur les chromatogrammes, s'accompagne d'une augmentation de la conductivité et seule la GST clivée apparaît sur gel par coloration au bleu de Coomassie ; il n'y a toujours aucune trace de la DPF3a. De plus, le clivage apparaît complet de par l'absence de tout excès de DPF3a-GST. Il ne semble pas y avoir formation d'un complexe protéine-ADN suffisamment volumineux pour en empêcher la visualisation sur gel de par l'état fragmenté de l'ADN par action de la DNase I.

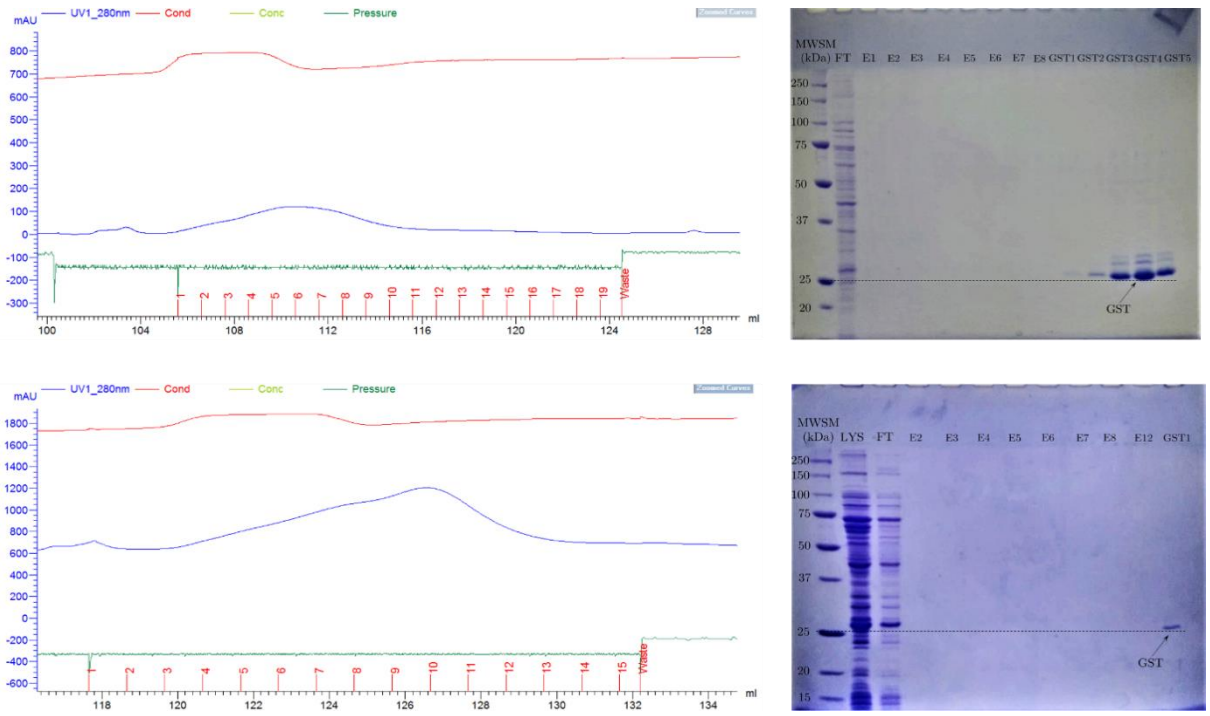

Figure 40 – Chromatogrammes de la protéine de fusion DPF3a-GST traitée à la DNase I, clivée au facteur Xa et purifiée par FPLC ainsi que leur gel SDS-PAGE 12 % associé. Pour chaque gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through, E correspond aux fractions d'élution de la DPF3a clivée et GST correspond aux fractions d'élution de la GST clivée. Seul le pic d'élution de la DPF3a est représenté sur le chromatogramme.

Dans un second temps, ces résultats de purification sont confrontés à l'analyse par spectrophotométrie UV-visible. Les spectres (cf. Fig. 41) révèlent des comportements spectroscopiques similaires à ce qui avait déjà été observé en l'absence du traitement à la DNase I. Certaines des populations de bande précédemment mises en évidence sont ici retrouvées. En regard de l'absorbance maximale, une bande majoritaire centrée sur 258-260 nm, une autre vers 254 nm et une dernière présentant un étalement entre 255 et 270 nm. L'absorption de la GST clivée demeure inchangée ($\lambda_{max} \sim 280$ nm). En outre, la tendance des courbes d'absorbance à 280, 265 et 254 nm sur les chromatogrammes est également identique. L'absorbance du pic d'élution à 254 et 265 nm de la DPF3a clivée demeure supérieure à celle à 280 nm. Il semble de plus en plus évident que ces signatures spectrales d'absorption sont caractéristiques de la DPF3a dans son état clivé. De par leur diversité et leur récurrence entre les purifications, elles ne peuvent être assimilées à une contamination par l'ADN. D'un point de vue global, il ressort que le traitement du lysat à la DNase I n'a aucun impact sur les propriétés sur gel et d'absorption de la DPF3a clivée.

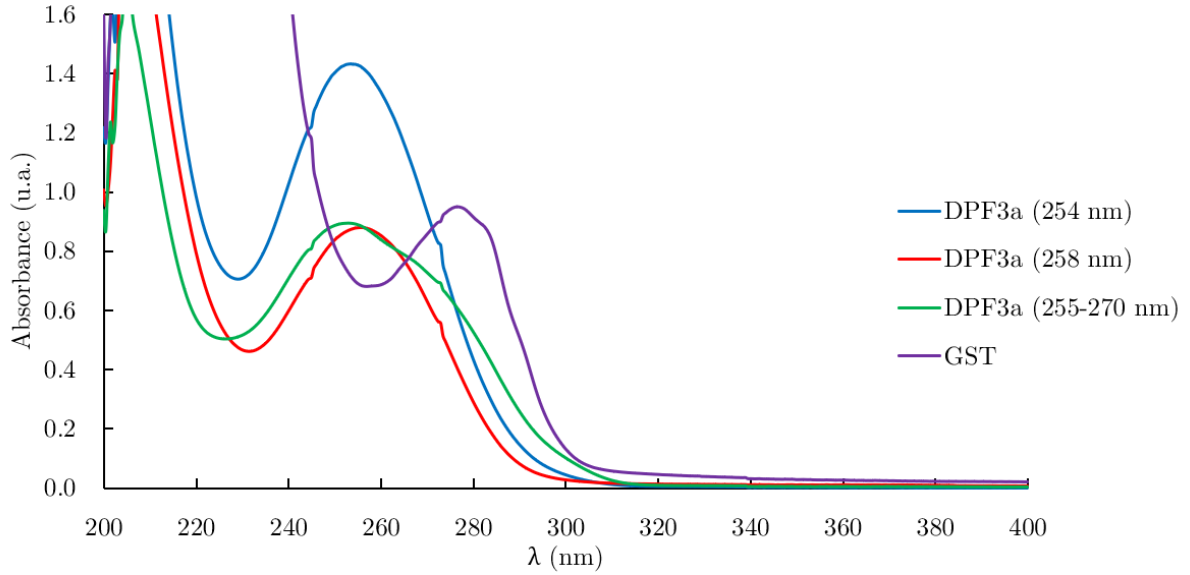

Figure 41 – Spectres d'absorption UV-visible des protéines DPF3a et GST clivées et traitées à la DNase I. L'évolution de l'absorbance (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). À chaque spectre de la DPF3a est associée la longueur d'onde ou la gamme de longueurs d'onde caractéristique. Les mesures sont effectuées dans une cuvette en quartz pour l'ensemble des courbes.

B. Électrophorèse sur gel d'agarose (AGE)

En tant que vérification ultime, la DPF3a, dans ses états fusionné et clivé, est migrée par voie électrophorétique sur gel d'agarose (abrégée AGE pour *agarose gel electrophoresis*) révélé au bromure d'éthidium (abrégé EtBr). Cette technique permet de révéler la présence d'ADN monocaténaire (ou simple brin) à l'aide du EtBr, un agent intercalant de l'ADN hautement toxique. Afin de désapparier les doubles brins d'ADN, les échantillons sont préalablement chauffés à 97 °C. Le gel d'agarose résultant (cf. Fig. 42) montre, aussi bien avant qu'après clivage, que les fractions de DPF3a ne contiennent pas d'ADN même sous forme résiduelle. Conséquemment, l'hypothèse de contamination par l'ADN est écartée à la vue de l'ensemble des résultats présentés. Au-delà de la confirmation apportée par AGE, cette décision est également motivée par l'absence significative de signal spectroscopique d'absorption des acides nucléiques sur les spectres de la DPF3a-GST (cf. Fig. 29). En effet, si l'ADN est éliminé avant clivage, il n'y a pas de raison pour qu'il y soit après.

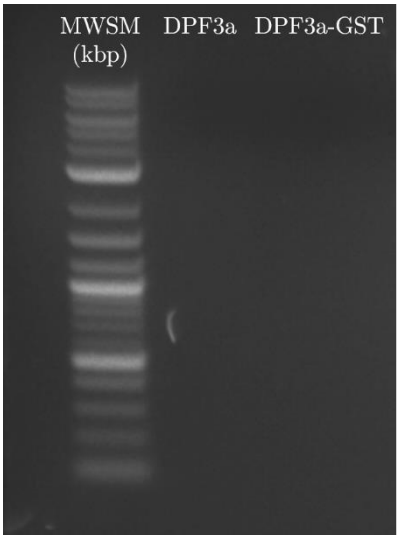

Figure 42 – Gel d'agarose 2 % de la protéine de fusion DPF3a-GST et clivée DPF3a révélé au EtBr. MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kilo-paire de bases kbp).

VI. Investigation de l'hypothèse d'agrégation protéique

A. Stabilisation de protéine en solution par le glycérol

Dès lors qu'il est employé comme cosolvant en solution aqueuse, le glycérol est réputé pour stabiliser les protéines en favorisant le repliement et en prévenant le mécanisme d'agrégation. À des fins stabilisatrices, la concentration de glycérol est généralement comprise entre 0 et 40 % (v/v) ; au-delà, les tampons deviennent trop visqueux, ce qui peut entraîner des complications lors de la purification en particulier par FPLC sur des colonnes pré-packées.[131]

Dans le présent cas, le pouvoir anti-agrégant du glycérol est évalué pour le clivage de la DPF3a à une concentration choisie de 10 % dans l'ensemble des tampons utilisés au cours de la purification. L'influence de la présence de cet agent stabilisant est initialement examinée via le chromatogramme et gel SDS-PAGE résultants (cf. Fig. 43). Contrairement aux purifications précédentes, le pic d'élution de la protéine d'intérêt s'accompagne dans ce cas-ci d'une diminution de la conductivité plutôt que d'une augmentation. Il apparaît également plus juste de parler de plusieurs pics d'élution. En effet, cette dernière semble s'établir en deux voire trois temps comme si le procédé de purification permettait la séparation de trois populations protéiques distinctes (annotées de 1 à 3 sur la Fig. 43). L'étude du gel coloré au bleu de Coomassie ne révèle aucune différence entre ces trois pics, se manifestant tous les trois par l'absence de toute trace de la DPF3a clivée ou d'autres matériels protéiques contaminants.

Cependant, la GST clivée demeure bel et bien identifiée vers 25 kDa à l'instar des analyses précédentes. Bien qu'il soit probable que le glycérol ait participé à l'émergence de cette seconde population avant le pic d'élution principal, sa présence ne contribue pas à la visualisation de la DPF3a sur gel et, par conséquent, à entraver son agrégation ; tout du moins à une concentration de l'ordre de 10 %. Dans le contexte de cette hypothèse, des concentrations plus élevées en glycérol sont à envisager pour de prochaines purifications afin d'en évaluer l'influence. Il n'est pas impossible qu'à partir d'une concentration critique l'agrégation de la DPF3a soit partiellement ou majoritairement bloquée, ce qui permettrait la visualisation de cette dernière sur gel SDS-PAGE.

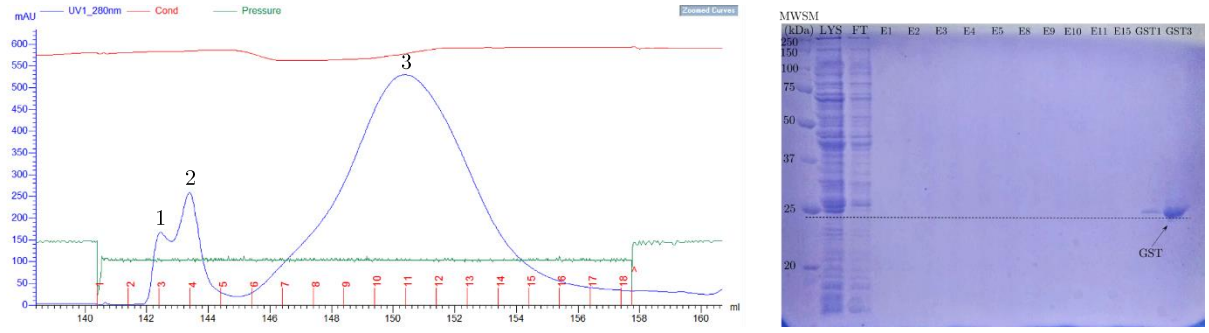

Figure 43 – Chromatogramme de la protéine de fusion DPF3a-GST en présence de glycérol 10 %, clivée au facteur Xa et purifiée par FPLC ainsi que le gel SDS-PAGE 12 % associé. Pour le gel, MWSM est l'échelle du marqueur de masse moléculaire (exprimée en kDa), LYS est le lysat, FT est le flow-through, E correspond aux fractions d'élution de la DPF3a clivée et GST correspond aux fractions d'élution de la GST clivée. Seul le pic d'élution de la DPF3a est représenté sur le chromatogramme.

Malgré l'absence répétée de la DPF3a sur gel même en présence de glycérol, le rôle anti-agrégant de ce dernier peut s'affirmer par la mise en évidence d'autres propriétés. C'est pourquoi, les échantillons de la DPF3a clivée dans le glycérol 10 % sont caractérisés par spectroscopie d'absorption afin de vérifier si les $\lambda_{abs\ max}$ précédemment déterminés se sont vus déplacer. Cette méthode est d'autant plus la bienvenue étant donné que le glycérol n'absorbe pas dans la gamme de longueurs d'onde d'intérêt, à savoir entre 250 et 280 nm.[132]

Les spectres (cf. Fig. 44) rendent à nouveau compte des comportements spectroscopiques détaillés au point IV de la Partie n°3, Section b. Tandis que les populations extrêmes sont détectées (à 250-254 nm ainsi qu'à 270-272 nm), la bande intermédiaire majoritaire, habituellement rencontrée vers 258-260 nm, n'est pas retrouvée. Les profils à deux populations et patatoïdal occurrent également. Bien que cela doive être reproduit pour en confirmer la tendance, le glycérol exacerberait la séparation des bandes d'absorption, conduisant à la disparition de celle centrale à 260 nm. Il est à noter que les deux pics d'élution présentent indistinctement la même diversité de bandes. Par ailleurs, les courbes d'absorbance à 280, 265 et 254 nm sur le chromatogramme sont comparables entre les trois pics d'élution la DPF3a clivée et suivent la même tendance que précédemment établie. L'absorbance à 280 nm est largement excédée par celle à 265 nm et d'avantage par celle à 254 nm. Le comportement d'absorption de la GST clivée reste fidèle et constant aux précédentes observations ($\lambda_{abs\ max} \sim 280$ nm).

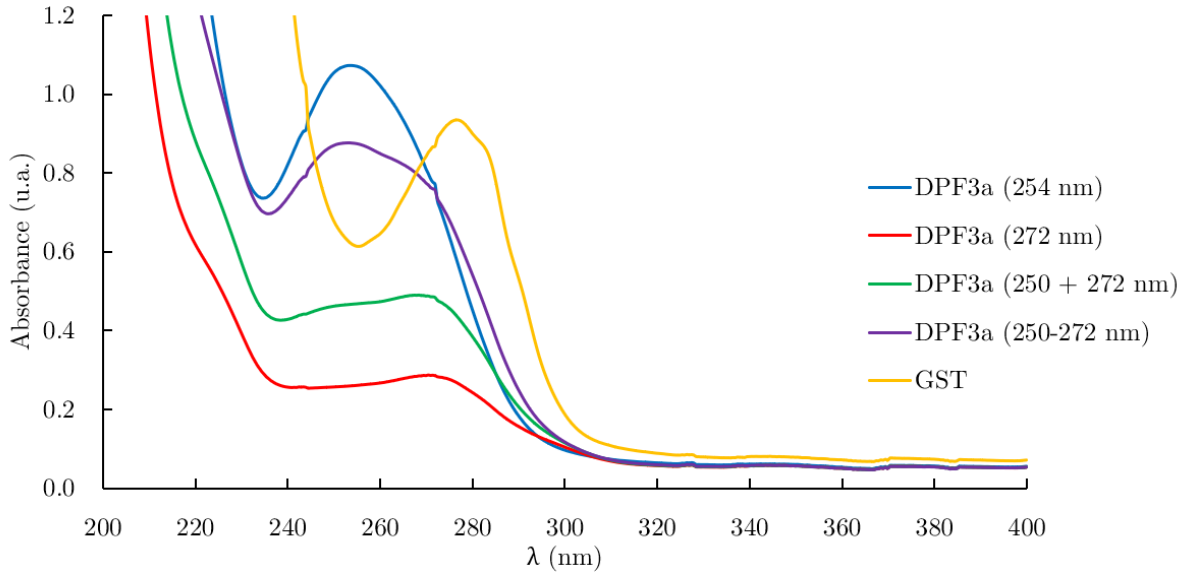

Figure 44 – Spectres d'absorption UV-visible des protéines DPF3a et GST clivées en présence de glycérol 10 %. L'évolution de l'absorbance (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). À chaque spectre de la DPF3a est associée la longueur d'onde ou la gamme de longueurs d'onde caractéristique. Les mesures sont effectuées dans une cuvette en quartz pour l'ensemble des courbes.

B. Comportement en spectroscopie d'émission

1. Fluorescence intrinsèque du tryptophane (ITF)

Étant donné la singularité et la reproductibilité du comportement spectroscopique d'absorption de la DPF3a clivée en regard de son équivalant fusionné, le même travail d'investigation est réalisé vis-à-vis de ses propriétés d'émission par excitation des résidus Trp (λ_{exc} = 295 nm). Des phénomènes singuliers (cf. Fig. 45) émergent des purifications et clivages de la DPF3a. Alors que la DPF3a-GST et la GST (cf. Fig. 30) sont caractérisées par une signature spectroscopique d'émission du tryptophane typique d'une protéine repliée (globule) ou partiellement repliée (globule fondu ; $\lambda_{ém\ max} \sim$ 335 nm), les échantillons de la DPF3a clivée sans traitement particulier présentent soit une bande d'émission déplacée vers 360 nm soit un profil d'émission patatoïdal étendu de 330 à 420 nm. Dans le premier cas, le déplacement bathochrome de $\lambda_{ém\ max}$ au-delà de 350 nm est synonyme d'une plus forte exposition des Trp au solvant, ce qui concorderait avec une nature désordonnée ainsi qu'une plus grande flexibilité de chaîne mettant plus fréquemment les résidus au Trp en contact avec l'eau.[112] Dans le second cas, l'étalement reste curieux.

Néanmoins, le comportement le plus intéressant et inattendu est celui rencontré dans les fractions de la DPF3a traitée avec la DNase I ou le glycérol 10 %. Alors qu'un signal caractéristique des Trp plus enfouis, aux alentours de 330-334 nm, est détecté,

une seconde signature, dont la $\lambda_{\text{ém max}}$ est comprise dans une gamme de 457-475 nm, d'intensité variable est retrouvée. En outre, même si elle est souvent de moindre intensité que la première bande d'émission, elle est majoritaire dans de rares échantillons et le devient systématiquement dès lors que ces derniers sont stockés une certaine durée à température ambiante. À titre exemplatif, après douze jours, tous les échantillons de la DPF3a ne présentant initialement qu'une unique bande vers 360 nm, étalée ou non, sont caractérisés par un déplacement de cette dernière jusqu'à 445 nm. Ainsi, l'apparition de cette deuxième population est indépendante du traitement (avec ou sans DNase I) ou du cosolvant (glycérol 10 %) utilisé. Il peut être aussi noté que l'intensité de fluorescence de cette seconde population connaît une augmentation significative tandis que la première est fortement diminuée à des longueurs d'onde d'excitation plus basses (275 ou 260 nm par exemple), autrement dit à des énergies d'excitation plus élevées.

En l'absence de fluorophore contaminant, ce phénomène d'émission apparaît intrinsèque à la protéine (création d'un chromophore intrinsèque) et sa manifestation concomiterait alors avec la formation d'agrégats amyloïdes. En effet, l'émergence d'une nouvelle fluorescence intrinsèque dans la gamme du visible chez les protéines, en dehors de celle due aux résidus aromatiques, a déjà été observée pour des fibrilles amyloïdes de lysozyme et d'α-synucléine avec une bande d'émission intense à 440 nm après excitation à 350-355 nm. Cette fluorescence dans le bleu-vert ou autofluorescence dans le bleu profond (abrégé dbAF pour *deep-blue autofluorescence*), aussi désigné par fluorescence amyloïde intrinsèque, proviendrait de transitions électroniques de basse énergie par délocalisation des électrons π des liaisons peptidiques le long des liaisons hydrogène entre les feuillets β croisés et est indépendant de la présence de résidus aromatiques. Il s'explique par la formation de structures amyloïdes pour lesquelles ces feuillets β croisés sont superposés les uns sur les autres, de manière ordonnée, de sorte à construire une fibrille.[133,134]

Néanmoins, une étude récente tend à montrer que cette autofluorescence existerait originellement au sein des protéines et des acides aminés avant agrégation. Elle originerait de la double liaison des groupements carbonyles dont la sensibilité au solvant et à l'environnement chimique, à l'instar du tryptophane, en ferait une excellente sonde optique pour l'étude conformationnelle et dynamique des protéines.[135] Cependant, dans le cadre de ce mémoire, l'analyse dbAF est limitée à la recherche d'une signature amyloïde.

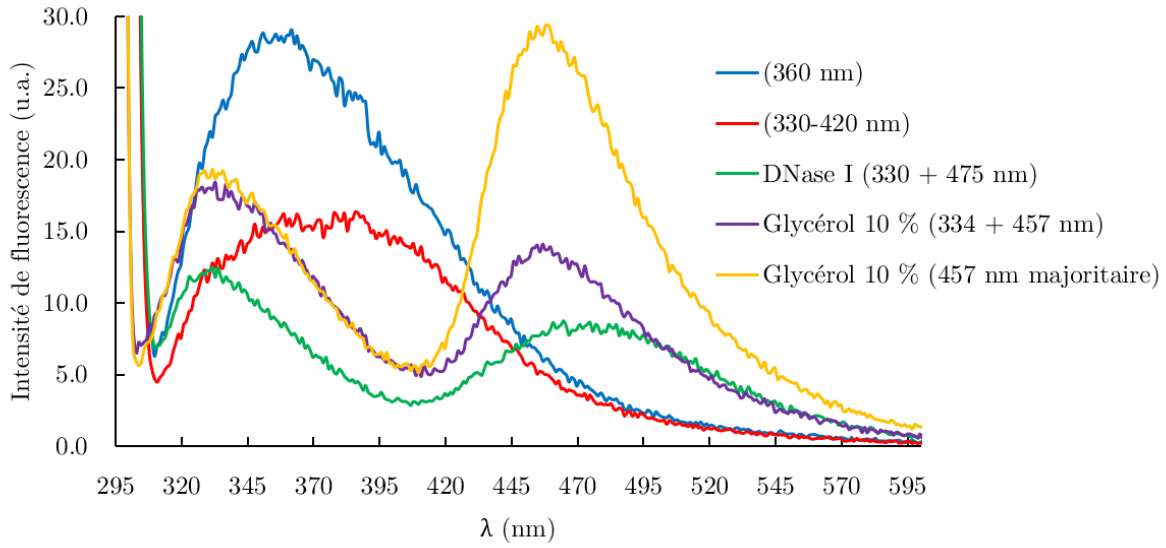

Figure 45 – Spectre d'émission en fluorescence intrinsèque du tryptophane de la protéine clivée DPF3a avec ou sans traitement à la DNase I et purifiée ou non avec une solution contenant du glycérol 10 %. L'évolution de l'intensité de fluorescence (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). À chaque spectre de la DPF3a est associée la longueur d'onde ou la gamme de longueurs d'onde caractéristique. Les mesures sont effectuées à λ_{exc} = 295 nm et sw = 10 nm.

2. Autofluorescence dans le bleu profond (dbAF)

Afin d'approfondir l'analyse du comportement d'émission de la DPF3a et de vérifier si elle donne à voir une signature spectrale caractéristique d'une substance amyloïde, son autofluorescence, évoquée précédemment, est mesurée à une λ_{exc} = 350 nm et confrontée à celle de la DPF3a-GST et d'une protéine globulaire type, le lysozyme C du blanc d'œuf de poule. Les spectres résultants (cf. Fig. 46) rendent compte que ni la DPF3a-GST ni le lysozyme C ne fluoresce après excitation à 350 nm ; elles ne semblent pas exister sous quelque forme agrégée. À l'inverse, la DPF3a clivée, pour laquelle une seconde population de fluorescence était apparue en ITF, est caractérisée par une bande d'émission intense à 440 nm après excitation à 350 nm. Il semble non seulement que l'hypothèse d'agrégation se révèle être la plus exacte mais aussi que ce phénomène prend place selon un schéma propre aux fibrilles amyloïdes. Bien que l'échantillon ayant été traité à la DNase I montre une signature d'intensité moindre, ledit traitement n'est pas impliqué dans cette différence d'intensité ; il est plus probable qu'il s'agisse d'un corollaire d'une différence de concentration entre les échantillons analysés.

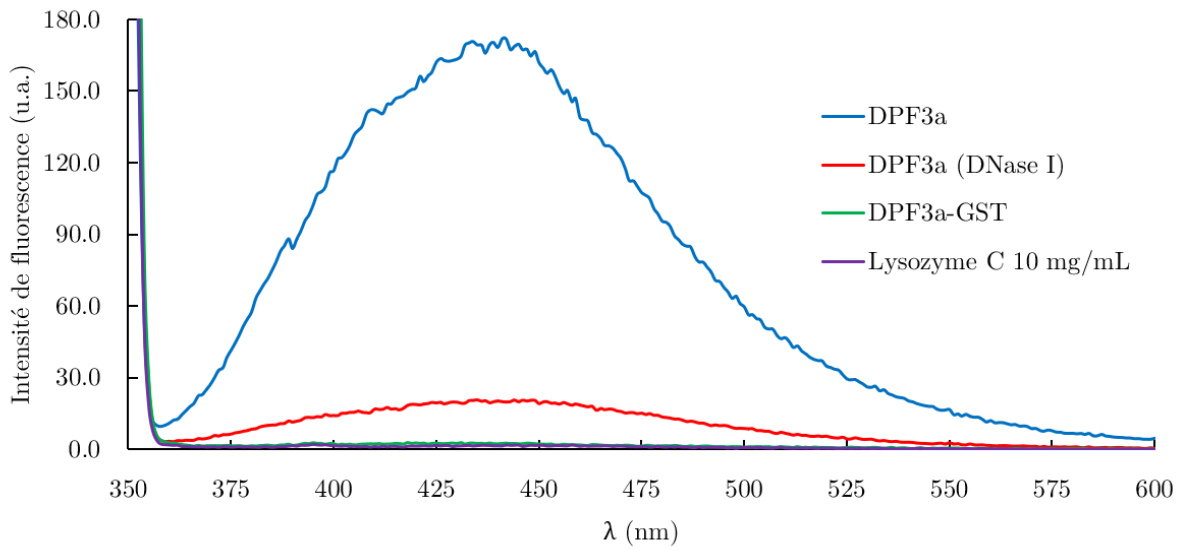

Figure 46 – Spectre d'émission en autofluorescence dans le bleu profond de la protéine de fusion DPF3a-GST, du lysozyme C du blanc d'œuf de poule à une concentration de 10 mg/mL et de la protéine clivée DPF3a avec ou sans traitement à la DNase I. L'évolution de l'intensité de fluorescence (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). Les mesures sont effectuées à λ_{exc} = 350 nm et sw = 5 nm.

C. Dichroïsme circulaire (CD)

L'analyse spectroscopique réalisée en CD (cf. Fig. 47) révèle également une distinction nette de signature entre les états clivé et fusionné de la DPF3a. Alors que la protéine de fusion DPF3a-GST (cf. Fig. 32) montre un signal typique des hélices α, la DPF3a clivée exhibe deux signaux fort distincts n'étant pas associés aux motifs de structure secondaire caractéristiques d'une protéine (cf. Fig. 31). La signature de la GST est fort semblable à celle de la DPF3a-GST. À l'instar de cette dernière, il n'est pas évident de distinguer les contributions respectives des hélices α et des feuillets β au signal. Nonobstant cette difficulté, un minimum (bande négative) est identifiable vers 206 nm, ce qui est caractéristique des hélices α et concordant avec la proportion conséquente en ces motifs de structure secondaire dont est composée la GST.

Quant à la DPF3a clivée, son comportement spectroscopique est une nouvelle fois étonnant de par sa non-continuité entre les échantillons. L'échantillon n'ayant pas été traité à la DNase I a une signature caractéristique d'une structure désordonnée avec une unique bande négative vers 200 nm. Il s'agit de la différence la plus notable vis-à-vis de la DPF3a fusionnée et de la GST clivée et qui rapproche la DPF3a d'une nature intrinsèquement désordonnée. Cependant, la protéine d'intérêt possède un autre signal qui est d'autant plus singulier.

La signature de la DPF3a, dont l'échantillon analysé a été tiré d'une purification ayant intégré un traitement à la DNase I, ne manifeste aucune des composantes typiques d'un spectre CD de protéine. En effet, un minimum global localisé à 227-230 nm ne correspond ni aux hélices α ni aux feuillets β ou encore à une structure désordonnée. Ce signal trouverait son origine dans l'existence, au sein de la structure de la DPF3a, de feuillets β «twistés». Cela se manifeste par déplacement bathochrome de la bande du feuillet β à 218-220 nm provenant de la transition électronique $n \rightarrow \pi^*$ du groupe amide. Ces feuillets β déformés sont notamment impliqués dans la formation de fibrilles amyloïdes comme cela a déjà été mis en évidence pour l'agrégation de l'IDP α-synucléine.[136,137] La signature résultante est également synonyme de l'empilement des feuillets β croisés dans les structures fibrillaires de type amyloïde.[138] Cette bande de feuillets β «twistés» d'un échantillon de la DPF3a n'est pas incompatible avec l'existence d'un signal désordonné pour un autre échantillon étant donné que les IDPs sont connues pour leur propension à s'agréger en substance amyloïde.

À la lueur des résultats obtenus par FPLC et en spectroscopie d'absorption, le traitement à la DNase I ne participe en lui-même pas à cette différence de signature. Il apparaît néanmoins davantage probable qu'elle provient d'une différence de la nature de l'échantillon dont l'état d'agrégation n'est fortuitement pas similaire. Cela conforte l'hypothèse de l'agrégation pour laquelle une signature spectrale caractéristique en CD a été détectée dans une fraction particulière de la DPF3a. Dès lors, il peut être de plus en plus envisagé que la DPF3a subisse un mécanisme d'agrégation précis aboutissant à la construction quasi-systématique de fibrilles amyloïdes. Bien que cette analyse démontre à nouveau la différenciation significative des propriétés biophysiques entre la DPF3a-GST et son homologue clivé, elle doit être opérée sur un panel plus large d'échantillons afin de déterminer si le phénomène est de caractère reproductible et si la signature désordonnée est également retrouvée.

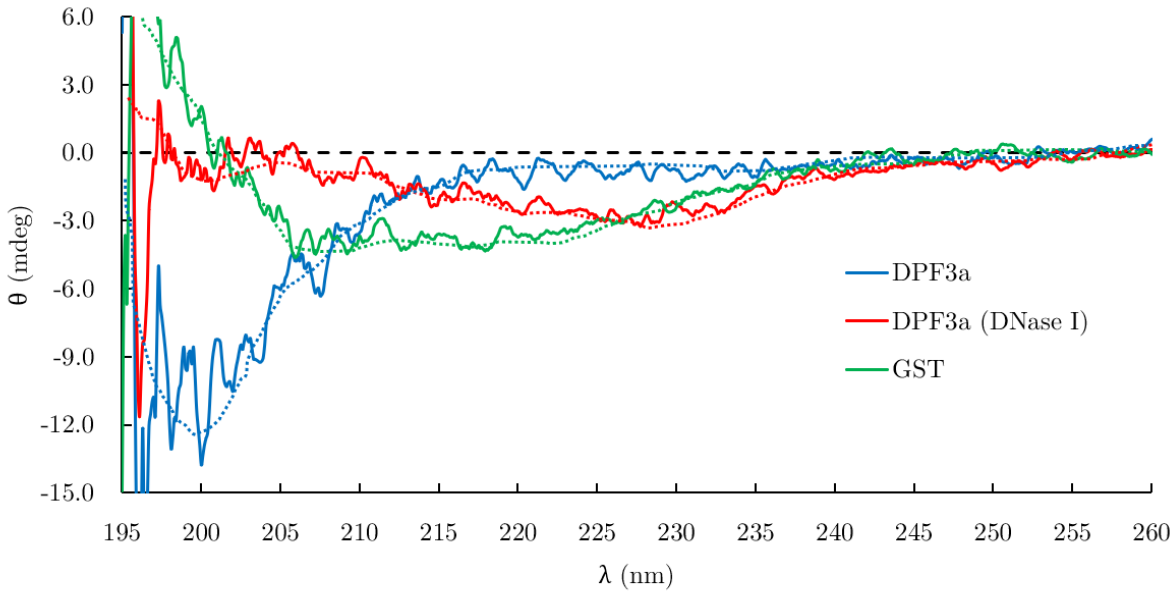

Figure 47 – Spectre de dichroïsme circulaire des protéines DPF3a et GST clivées. L'évolution de l'ellipticité θ (exprimée en mdeg) est représentée en fonction de la longueur d'onde λ (exprimée en nm). L'allure générale de la tendance de chaque spectre est représentée par un trait en pointillé dont le code couleur respecte celui imposé par la légende.

D. Diffusion dynamique de la lumière (DLS)

L'analyse du diamètre hydrodynamique par DLS a été réalisée sur des échantillons de la DPF3a clivée traités ou non à la DNase I et contenus dans le tampon FB. L'influence de la température et du pH a également été testée. Les valeurs de D_h sont renseignées à la Table 2. En regard du D_h mesuré pour la DPF3a-GST, la dimension hydrodynamique de la DPF3a clivée est approximativement deux à sept fois plus grande alors qu'elle est séparée de son *tag* GST. Si un phénomène d'agrégation était déjà soupçonné pour la protéine de fusion, son ampleur est d'autant plus conséquente dans le présent cas. D'un point de vue global, les analyses DLS décrivent pour la DPF3a clivée des objets dont la taille moyenne est d'un ordre de grandeur supérieur à ce qui est attendu pour une protéine globulaire ou même une IDP. La gamme de D_h est comprise entre 180 et 550 nm. Bien que la présence de NaCl 150 mM dans le tampon FB entraîne une augmentation apparente de la taille hydrodynamique des particules mesurée par gonflement des sphères de solvatation, sa contribution est négligeable aux D_h obtenus.[139]

L'autre phénomène le plus remarquable concerne l'influence de la température sur la taille hydrodynamique des agrégats de la DPF3a. L'augmentation de la température s'accompagne d'une augmentation continue du D_h de la DPF3a clivée. Le passage de 25 à 40 °C cause une augmentation de diamètre hydrodynamique de 27 %

tandis qu'elle s'élève à 47 % entre 25 et 70 °C. La chauffe a été opérée sur un autre échantillon de la DPF3a clivée et a montré une tendance similaire. Cette variation de D_h peut s'expliquer soit par une exacerbation de l'agrégation soit par une modification de la structure des agrégats, provoquant leur gonflement. Elle apparaît d'autant plus significative qu'elle dévie de la tendance imposée par l'équation de Stokes-Einstein. En outre, cette augmentation due à la température a déjà été observée dans la littérature sur des fibrilles amyloïdes présentant des dimensions hydrodynamiques du même ordre de grandeur que celles de la DPF3a clivée.[140] Aussi, ce phénomène de gonflement ou d'agrégation apparaît réversible étant donné que lorsque l'échantillon est refroidi de 70 à 15 °C, il retrouve sa taille initiale. À l'inverse, le pH ne semble pas impacter outre mesure la dimension des agrégats protéiques.

Protéine/échantillon	Température (°C)	pH	D_h moyen (nm)
DPF3a-GST	25	8.0	77.2 ± 8.4
DPF3a	25	7.5	550 ± 22
DPF3a (centrifugée)	25	7.5	183.9 ± 15.4
DPF3a (DNase I)	25	7.5	209 ± 18
DPF3a (diluée 5 fois)	25	7.5	226 ± 21
	40	7.5	287 ± 18
	70	7.5	332 ± 63
	15	7.5	228 ± 17
DPF3a (diluée 5 fois)	25	7.0	240 ± 26
	25	2.0	261 ± 27
	25	10.0	258 ± 25

Table 2 – Valeurs de diamètre hydrodynamique moyen D_h et déviation standard associée (exprimés en nm) en fonction de la température (exprimée en °C) et du pH pour la protéine de fusion DPF3a-GST dans le tampon PB et les protéines clivées DPF3a dans le tampon FB avec ou sans traitement à la DNase I.

Le caractère dynamique de l'agrégation est soutenu par l'échantillon présentant un $D_h = 550$ nm ayant été préalablement filtré (0.2 µm). En effet, des agrégats de taille supérieure à la limite autorisée par le filtre se sont formés après l'étape de filtration. Par ailleurs, l'échantillon récupéré après analyse DLS présentait une solution de coloration jaunâtre et turbide et dans laquelle un précipité était nettement visible, ce qui n'était pas le cas après filtration. La centrifugation de cet échantillon a résulté en la formation d'un culot solide et dont la mesure DLS du surnageant a révélé un D_h plus

petit (= 183.9 nm) et comparable à ce qui a été obtenu pour les autres échantillons de la DPF3a clivée sans matériel précipité. L'existence d'un tel précipité appuie l'hypothèse d'agrégation de la DPF3a. Une nouvelle fois, la DNase I n'a aucune influence sur les mesures enregistrées et les D_h sont reproductibles d'une purification à l'autre.

E. Chromatographie d'exclusion stérique (SEC)

Dans le prolongement des analyses effectuées en DLS, les volumes d'élution caractéristiques de la DPF3a et de la GST purifiées et clivées par FPLC sont étudiés par SEC. Comme cela a été expliqué au point E de la Partie n° 3, Section a, l'élution est procédée dans le tampon FB. Les volumes d'élution mesurés sont réunis à la Table 3, préalablement corrigés par la valeur du volume mort (V_{mort} = 46.7 mL), et mis en relation avec la masse moléculaire estimée à partir de la droite d'étalonnage (cf. Fig. 33) de standards protéiques globulaires. Pour les échantillons de la DPF3a clivée soumis, la comparaison est à nouveau établie vis-à-vis du traitement à la DNase I. Quant à la GST clivée, les fractions sont choisies pour deux purifications différentes mais opérées dans des conditions identiques.

Les deux populations observées pour la GST clivée sont consistantes et se retrouvent chacune dans une même gamme de volume d'élution ou de masse moléculaire estimée : ~ 39-42 kDa et ~ 2800-3100 Da. La première correspond à la GST, ce qui est cohérent avec les constations vis-à-vis de la DPF3a-GST, tandis que la seconde reste non-identifiée. La DPF3a clivée, quant à elle, montre, à l'instar de la DPF3a-GST, une bande majoritaire sortant dans le volume mort et associée à un objet de très grande taille. Cela renforce l'idée que la DPF3a existe sous une forme fortement agrégée potentiellement de type amyloïde. Quel que soit l'échantillon, aucune bande d'élution pouvant s'apparenter à la DPF3a seule dans un état désagrégé n'est observée. Des populations à de plus grands volumes d'élution et d'intensité variable sont détectées mais elles n'apparaissent pas spécifiques ou plus particulièrement significatives à la DPF3a étant donné qu'elles sont pour la plupart retrouvées dans les échantillons de la DPF3a-GST et de la GST.

Ces résultats obtenus par SEC sont cohérents avec ceux obtenus par DLS. Il est probable que, de par leurs dimensions hydrodynamiques, les agrégats amyloïdes ne peuvent être séparés par la colonne d'exclusion stérique en regard de sa résolution de séparation et se retrouvent systématiquement dans le volume mort. Étant donné que la DPF3a a une masse moléculaire d'environ 40 kDa, un assemblage de 25 unités

protéiques suffit à dépasser la taille limite de 1 MDa permise par la colonne employée. Il est aussi à noter qu'alors que les bandes relatives à la GST clivée présentent toutes une absorbance maximale à 280 nm, celles de la DPF3a absorbent toutes plus à 265 et 254 nm qu'à 280 nm, ce qui concorde avec les comportements spectroscopiques précédemment constatés. L'analyse par SEC confirme une fois de plus que le traitement à la DNase I n'a aucune influence sur les propriétés mesurées de la DPF3a.

Protéine/échantillon	Bande	$V_{élution\ corrigé}$ (mL)	$M_{estimée}$ (Da)	Hypothèse
DPF3a	1ère	-5.7	> 2000000	Agrégats
	2ème	63.1	2756	?
	3ème	68.1	1707	?
	4ème	72.7	1099	?
DPF3a (DNase I)	1ère	-6.1	> 2000000	Agrégats
	2ème	61.9	3092	?
	3ème	70.7	1331	?
	4ème	75.3	857	?
GST (1)	1ère	35.5	38770	GST
	2ème	62.9	2810	?
GST (2)	1ère	34.7	41858	GST
	2ème	61.9	3092	?

Table 3 – Valeurs de volume d'élution corrigé (exprimé en mL) par le volume mort (V_{mort} = 46.7 mL) et de masse moléculaire estimée (exprimée en Da) en fonction des populations de bande des protéines clivées DPF3a (avec ou sans traitement à la DNase I) et GST (deux échantillons de deux purifications distinctes).

VII. Investigation de l'hypothèse de complexation métal-protéine

A. Influence des ions zinciques sur les propriétés d'absorption

Afin d'apporter des premiers éléments de réponse à l'hypothèse de complexation entre la DPF3a et des cations divalents, pouvant modifier ses propriétés d'absorption dans le domaine UV-visible, l'influence de la présence des ions zinciques Zn^{2+} à différentes concentrations est mise en exergue par la variation des signatures spectroscopiques d'absorption et d'émission. Pour ces essais, le choix du zinc est motivé par l'existence des domaines C_2H_2 et PHD chélatant préférentiellement les ions Zn^{2+} par l'intermédiaire de résidus Cys et His. Les concentrations choisies sont de 0.5 mM, 1.0 mM et 2.0 mM. L'ensemble des échantillons sont établis à partir d'un même éluat protéique de la DPF3a clivée contenue dans le tampon FB et présentant une bande

unique d'absorbance maximale centrée à 258 nm. Ainsi, il est à noter que, de par la composition du tampon FB, les échantillons préparés renferment initialement des ions Ca^{2+} à une concentration de 1.0 mM. Les cations Zn^{2+} introduits s'additionnent par conséquent aux ions calciques dont les interactions avec le domaine de type Krüppel ou PHD-1/2 de la DPF3a en l'absence de zinc sont hypothétiques mais pas inenvisageables.

Les tests préliminaires exécutés par spectroscopie d'absorption entre 200 et 400 nm ont révélé que la valeur de $\lambda_{abs\ max}$ (= 258 nm) ne suit aucun déplacement vers le bleu ou le rouge par addition du zinc, quelle qu'en soit la concentration, mais se caractérise par un effet hypochrome remarquable (à la même $\lambda_{abs\ max}$) et s'accentuant dans le temps. Cet effet se manifeste par une diminution de l'absorbance de l'échantillon, se traduisant par une diminution de son coefficient d'extinction molaire. C'est pourquoi, la cinétique de cette diminution progressive de l'intensité de bande à 258 nm a été suivie et mesurée à intervalle régulier pour chacune des concentrations en Zn^{2+}. Les tendances cinétiques (cf. Fig. 48) donnent à voir que la concentration en zinc n'impacte pas la vitesse de perte d'absorbance étant donné que les profils ont une décroissance similaire et constante durant les 60 premières minutes, ce qui est suivi d'une stagnation du signal (la diminution est nettement moins marquée voire nulle) dans l'intervalle 60-120 minutes. Entre chacune des concentrations, l'effet hypochrome est particulièrement concordant et suit une tendance linéaire si bien qu'il peut être estimé à une diminution de la valeur d'absorbance à $\lambda_{abs\ max}$ d'environ 8-10 mu.a./min sur une période de 60 minutes.

Il est possible qu'il existe sur cette durée une compétition entre les ions Ca^{2+} et Zn^{2+} ou que ces derniers ont une certaine cinétique de liaison aux ZNF. L'origine de cet effet hypochrome reste néanmoins difficile à déterminer dans le cadre de ces seuls essais et méritent d'être approfondis. Il a déjà été constaté sur des polypeptides (inférieurs à cent acides aminés) et protéines en solution qu'un effet hypochrome sur l'absorption est synonyme d'une transition d'une structure de type *random coil* vers une hélice α. Cette diminution du coefficient d'extinction molaire disparaît aussitôt que le polypeptide est placé en conditions dénaturantes, notamment par variation du pH.[141] Ainsi, le zinc pourrait avoir un rôle structurant, en induisant un changement de conformation, par coordination avec les ZNF qui composent les domaines fonctionnels de la DPF3a. Il n'est pas impossible que cet effet hypochrome soit associé au phénomène d'agrégation ou à une modification de l'état agrégé de la protéine, perturbant l'environnement des résidus aromatiques. En effet, certains métaux, dont le zinc, sous

forme cationique promeuvent ou préviennent les processus d'oligomérisation ou d'agrégation de protéines, notamment sous forme de fibrilles amyloïdes.[142–144] Aussi, il est connu que l'absorption du tryptophane est *quenchée* (effet hypochrome) par interaction avec des groupements carboxyles.[145]

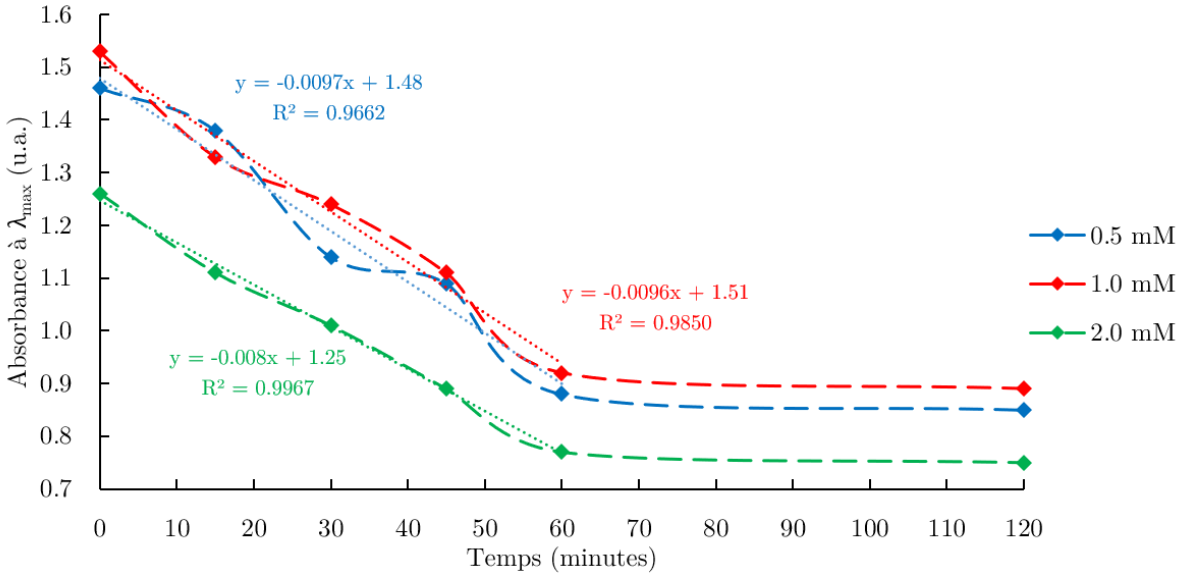

Figure 48 – Tendances cinétiques de l'influence du Zn^{2+} aux concentrations choisies (0.5, 1.0 et 2.0 mM) sur le comportement d'absorption de la protéine DPF3a clivée. L'évolution de l'absorbance (exprimée en unités arbitraires u.a.) mesurée à $\lambda_{abs\ max}$ = 258 nm est représentée en fonction du temps (exprimé en minutes). L'allure de la tendance pour chaque concentration est représentée par un trait discontinu. La courbe de tendance linéaire (de 0 à 60 minutes) pour chaque concentration est représentée par un trait en pointillé. Les mesures sont réalisées dans une cuvette en quartz pour l'ensemble des points.

B. Influence des ions zinciques sur les propriétés d'émission (ITF)

Concernant le comportement spectroscopique d'émission étudié par ITF (cf. Fig. 49), c'est l'effet parfaitement inverse qui est observé. En présence de zinc, l'intensité de fluorescence due aux résidus Trp (λ_{exc} = 295 nm) est augmentée pour les deux populations dont l'existence est décrite au point VI.B de la Partie n°3, Section b. L'augmentation du rendement quantique montre une variation dépendant de la concentration en ions zinciques. Alors que l'effet est particulièrement marqué à 0.5 mM en ions Zn^{2+}, l'augmentation de leur concentration provoque une diminution du signal de fluorescence à 1.0 et 2.0 mM. De plus, la signature d'émission pour la seconde population est caractérisée par un déplacement bathochrome de 460 à 485 nm provoqué par l'addition de 0.5 mM en Zn^{2+}.

Les interactions entre des espèces cationiques et les résidus Trp ont un impact direct sur le comportement de fluorescence de ces derniers.[146] Toutefois, les résidus Trp (W^{56} et W^{79}) ne se situent pas au sein des domaines contenant les ZNF chez la DPF3a. Un changement de conformation s'accompagnant d'une exposition plus conséquente des régions hydrophobes de la DPF3a au solvant peut être alors envisagé. Ce genre de phénomène a déjà été décrit dans la littérature quant aux modifications de conformation de la calréticuline induites par liaison aux ions Ca^{2+} et Zn^{2+}.[147] L'hypothèse de transition conformationnelle apparaît cohérente avec les résultats obtenus en spectroscopie d'absorption. Ainsi, en-deçà d'une certaine concentration (en fonction des rapports stœchiométriques entre la DPF3a et le zinc), les cations Zn^{2+} se lieraient dans un premier temps aux ZNF entraînant un changement conformationnel de la DPF3a, rendant ses résidus Trp plus accessibles au solvant. Au-delà d'une certaine concentration, les sites de liaisons au zinc seraient saturés et l'interaction s'opèrerait directement entre les ions Zn^{2+} excédentaires et les Trp, par l'intermédiaire d'interactions π-cation par exemple, *quenchant* le signal de fluorescence.[146,148]

Il serait pertinent de confronter l'ensemble desdits résultats à des mesures de CD de la DPF3a clivée afin de déterminer la nature de cette transition. Autrement dit, vérifier s'il survient une modification du signal désordonné ou agrégé vers une signature typique d'une structure plus ordonnée ou, au contraire, d'autant plus agrégée.

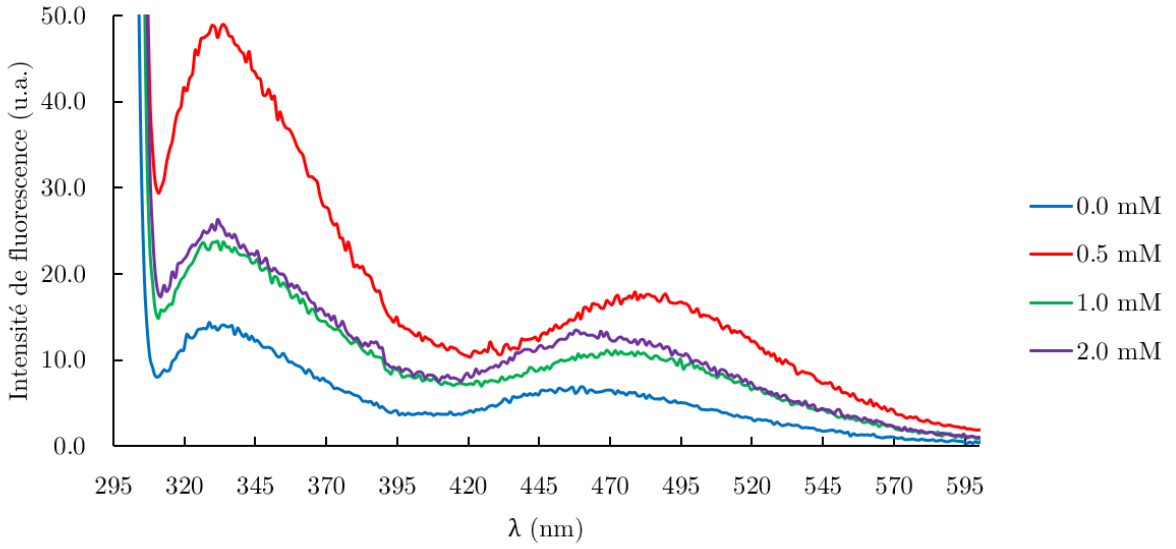

Figure 49 – Spectres d'émission en fluorescence intrinsèque du tryptophane de la protéine DPF3a en présence de Zn^{2+} aux concentrations choisies (0.5, 1.0 et 2.0 mM) après 24 h d'incubation. L'évolution de l'intensité de fluorescence (exprimée en unités arbitraires u.a.) est représentée en fonction de la longueur d'onde λ (exprimée en nm). Les mesures sont effectuées à λ_{exc} = 295 nm et sw = 10 nm.

Section c

—

Prédiction de propriétés associées au désordre

I. Analyses de composition de séquence

La prépondérance de certaines familles (définies par des critères physicochimiques) ou certains types d'acides aminés au sein de la séquence est un premier élément révélateur de la propension au repliement des protéines. L'existence d'un noyau hydrophobe enfoui se manifeste par une plus grande proportion en résidus aromatiques et apolaires tandis que les chaînes latérales polaires ou chargées autorisent une plus grande flexibilité.[149] Dans cette première approche, la distinction des acides aminés se base sur leurs propriétés physicochimiques fondamentales, à savoir leur polarité, état de charge, hydrophilie, hydrophobie (hydrophobicité) ou encore leur caractère aromatique. Ils sont rassemblés comme suit dans les catégories décrites : les résidus considérés polaires sont Arg, Asn, Asp, Cys, Glu, Gln, His, Lys, Ser, Thr et Tyr ; les apolaires sont Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp et Val ; les chargés sont Arg, Asp, Glu et Lys ; les hydrophiles sont Arg, Asn, Asp, Gln, Glu et Lys ; les hydrophobes sont Ala, Cys, Ile, Leu, Met, Phe, Trp et Val ; les neutres (sans affinité préférentielle) sont Gly, His, Pro, Ser, Thr et Tyr ; les aromatiques sont Phe, Tyr et Trp. Les proportions en résidus sont exprimées en pourcentage par rapport à la longueur totale de la séquence et désignées par le taux de composition de séquence (abrégé TCS).

Les catégories explicitées sont appliquées, à des fins comparatives, à une base de données protéiques réduite dans le cadre de ce projet de recherche. Les protéines choisies visent à couvrir une grande variété structurelle, aussi bien globulaire que désordonnée, rencontrée dans la littérature tout en y incluant les partenaires étudiés afin de dégager au mieux une appréciation globale sur le rôle joué par la structure primaire sur l'état de repliement. Lesdites protéines sont les isoformes DPF3b et DPF3a ainsi que la *polo-like kinase* PLK1 pour les partenaires étudiés. Concernant les protéines repliées dites globulaires ou ordonnées (ORDPs), il s'agit des suivantes : la transférase GST de *S. japonicum*[116] ; la transférase en tonneau β PagP de *E. coli*[150] ; l'albumine de sérum bovin BSA[151] ; le lysozyme C du blanc d'œuf de poule (abrégé Lys C)[152] ; la ribonucléase humaine RPP40 de masse moléculaire similaire à la DPF3 ($M_{RPP40} = 41.83$ kDa) et exprimée dans le noyau[153] ; la sérine/thréonine kinase humaine MAPK1 de masse moléculaire similaire à la DPF3 ($M_{MAPK1} = 41.39$ kDa) et également exprimée dans le noyau[154] ; le canal anionique humain à tonneau β VDAC1[155] ; l'hémoglobine glyquée humaine HBA1.[156] Quant aux IDPs, les membres suivants sont considérés (ils sont tous exprimés chez l'humain) : l'alpha-synucléine (abrégée α-syn), l'IDP type, impliquée la régulation de l'activité synaptique[157] ; la protéine oncogène c-Myc[101] ; la

glycoprotéine (cytokine) ostéopontine (abrégée OPN)[158] ; le partenaire d'interaction de la tuftéline TFIP11 (sa nature désordonnée n'est encore qu'hypothétique) ; la protéine membranaire intégrale émerine (abrégée EMD) capable d'interagir avec le complexe BAF[159] ; le facteur de transcription NUPR1 impliqué dans le remodelage de la chromatine[160] ; le facteur de transcription p53 aussi connu comme gène suppresseur de tumeur[161] ; la protéine membranaire présynaptique BASP1.[162] L'ensemble des séquences sont reprises de la base de données UniProtKB.[39]

L'analyse du caractère préférentiellement hydrophile ou hydrophobe (cf. Fig. 50a) des séquences soumises révèle, de manière univoque, que l'ensemble des protéines globulaires ordonnées présentent une plus grande proportion en acides aminés hydrophobes bien que la différence soit moins significative pour la VDAC1 ; la HBA1 est la plus enrichie en acides aminés hydrophobes. La kinase PLK1 est également caractérisée par ce même enrichissement en résidus hydrophobes. Quant aux IDPs, les profils sont plus divers. L'α-syn est la seule IDP dont le TCS en résidus hydrophobes est supérieur à celui en résidus hydrophiles. Néanmoins, leur différence est minime et peut s'expliquer par l'existence d'une hélice α couvrant la moitié de la séquence de l'α-syn. La TFIP11, l'émerine et la protéine p53 montrent des niveaux quasiment équivalents avec une légère préférence pour les acides aminés hydrophiles. L'enrichissement en résidus à caractère hydrophile est nettement observable chez les autres IDPs, ce qui est cohérent avec leur plus grande flexibilité de chaîne en solution aqueuse. Les protéines DPF3a et DPF3b suivent la tendance des IDPs, avec un enrichissement d'autant plus marqué pour la DPF3a. En outre, cette dernière possède le TCS en résidus hydrophiles le plus élevé de toute la base de données. Les variations en résidus neutres sont moins significatives au sein de chacun des groupes considérés.

La différenciation des acides aminés selon leur polarité (cf. Fig. 50b) donne à voir qu'en réalité peu de séquences sont enrichies en résidus apolaires. En effet, seules la transférase PagP, la HBA1 et l'α-syn présentent un tel comportement mais avec une différence relativement faible. Pour les autres ORDPs et la PLK1, les TCS semblent équilibrés et les différences sont également minimes à l'exception de la BSA riche en résidus polaires. Un enrichissement similaire ou d'autant plus exacerbé est observé pour les IDPs c-Myc, OPN, EMD et NUPR1 ainsi que les deux isoformes de la DPF3 ; la différence de TCS en résidus polaires entre la DPF3a et la DPF3b est très peu marquée et de l'ordre de 1.0 %. À nouveau, en regard des propriétés des IDPs, cela apparaît cohérent. Malgré cette tendance, les protéines TFIP11, p53 et BASP1 conservent une

composition plus importante en acides aminés polaires dans l'ordre de grandeur de celle des ORDPs.

Concernant l'analyse de séquence sur base du critère de charge et d'aromaticité (cf. Fig. 50c), l'ensemble des protéines étudiées renferment une proportion moindre en résidus aromatiques à l'exception de PagP particulièrement riche en Trp, Tyr et Phe. En termes de différence de TCS, la majorité des ORDPs contiennent finalement peu de résidus chargés en regard des aromatiques pour lesquels la proportion est proche de deux pour un. Cela ne se vérifie pas chez la PLK1 et la BSA dont le rapport entre les résidus chargés et aromatiques est plutôt de trois pour un. La différence est encore plus conséquente dans le groupe des IDPs et pour la DPF3 avec un rapport moyen de six pour un. La différence la plus flagrante est celle observée chez la BASP1 dont la composition en résidus aromatiques est quasiment nulle ; sa séquence ne renferme aucun Trp ni de Phe. À l'inverse, la composition de séquence de l'EMD se rapproche de celle d'une ORDP.

L'ensemble des critères ont rendu compte que la DPF3, quel que soit son isoforme, tend à exhiber une composition de séquence typique d'une IDP. Toutefois, ces analyses statistiques brutes établies sur des familles d'acides aminés peu discriminantes apparaissent insuffisantes pour dégager une tendance générale au désordre au sein de la famille des IDPs. Par conséquent, il convient d'affiner les critères de comparaison de composition de séquence. Nonobstant cette constatation, une étude de Weathers *et al.* démontre à ce sujet que la connaissance seule de la composition de la séquence d'une protéine est suffisante à la détermination précise de son caractère désordonné. Par ailleurs, les propriétés physicochimiques des acides aminés plutôt que leur occurrence spécifique participe grandement au désordre intrinsèque. Néanmoins, pour conserver la précision de la prédiction, il est conseillé d'appliquer ces critères sur des domaines ou des portions définies de la séquence étant donné que les IDPs peuvent contenir à la fois des IDRs et des régions ordonnées.[163] Le pouvoir de discrimination s'en trouve ainsi limité dès lors que la composition en résidus est comparée sur base de la longueur complète des séquences. Il n'en demeure pas moins intéressant d'envisager une telle analyse restreinte sur les domaines des isoformes de la DPF3.

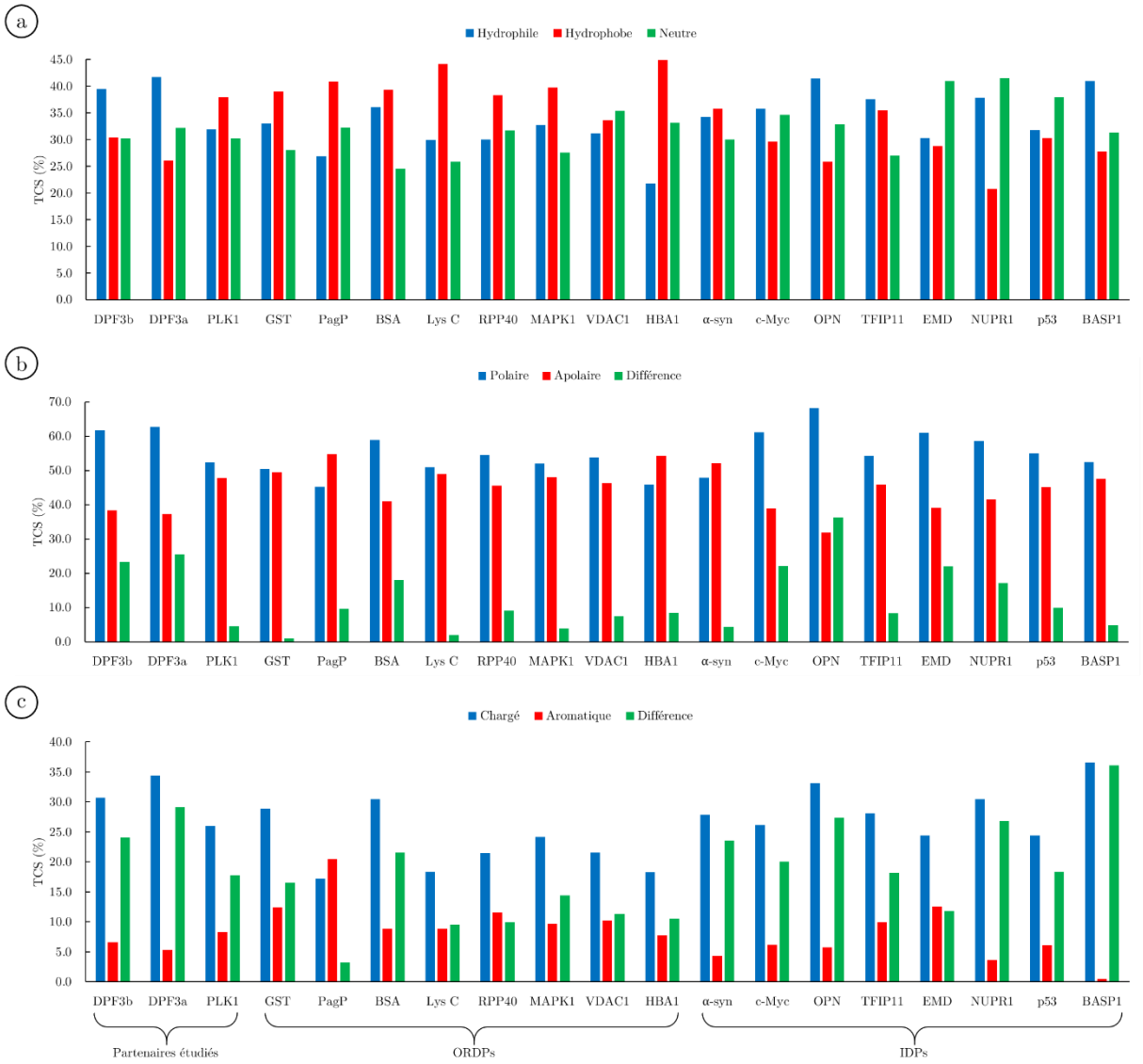

Figure 50 – (a) Analyses histogrammiques de composition de séquence d'une sélection choisie de protéines ordonnées (ORDPs), désordonnées (IDPs) et des partenaires étudiés se basant sur un critère d'affinité aqueuse (hydrophile-hydrophobe-neutre), (b) un critère de polarité (polaire-apolaire) et (c) un critère d'état de charge et d'aromaticité. Pour chaque histogramme, le taux de composition de séquence (TCS) est exprimé en pourcentage par rapport à la longueur de séquence totale. Pour les histogrammes (b) et (c), la différence est présentée en valeur absolue.

II. Acides aminés promoteurs de désordre

Afin de rationaliser d'avantage l'analyse de la base de données de structures primaires, un nouveau critère de classification des résidus est introduit. Il s'agit de les distinguer selon leur propension à promouvoir de l'ordre ou du désordre dans la structure tridimensionnelle de la protéine. L'ensemble des acides aminés naturels sont répartis en trois groupes : les résidus promoteurs d'ordre (abrégés OPRs pour *order-promoting residues*), promoteurs de désordre (abrégés DPRs pour *disorder-promoting*

residues) et non-promoteurs (abrégés NPR pour *non-promoting residues*). Le groupe des OPRs rassemble les acides aminés aromatiques (Phe, Tyr et Trp), apolaires à chaîne longue (Leu, Ile et Val) ainsi que la cystéine de par ses possibilités de ponts disulfure (Cys) et l'asparagine (Asn). Le groupe des DPRs comprend les acides aminés apolaires à chaîne courte (Gly et Ala), trois chargés (Arg, Lys et Glu), deux polaires (Gln et Ser) ainsi que la proline (Pro) de par ses possibilités de rupture de structures secondaires. Le dernier groupe des NPRs inclus un acide aminé chargé (Asp), deux polaires (His et Thr) et un apolaire (Met).[85] Les pourcentages associés à chaque catégorie sont renseignés par le TCS. L'enrichissement en DPRs ainsi que leur récurrence dans les IDRs sont caractéristiques des IDPs ; pour rappel, elles présentent une faible complexité de séquence et ont une grande propension au désordre. Ce critère discriminant est soumis à l'examen de la même base de données protéiques que celle détaillée précédemment au point I de la Partie n°3, Section c.

L'histogramme associé aux TCS en OPRs, DPRs, NPRs et la différence entre les DPRs et OPRs (cf. Fig. 51) montre une certaine diversité chez les protéines globulaires. Tandis que les protéines PagP, Lys C, RPP40 et MAPK1 présentent des niveaux en OPRs et DPRs similaires, la PLK1, GST, BSA, VDAC1 et HBA1 sont enrichies en DPRs sans une différence significative. De manière consistante, chaque membre du groupe des IDPs est caractérisé par un enrichissement net en DPRs vis-à-vis des résidus OPRs. L'enrichissement en DPRs le plus significatif est celui de la BASP1 dont le TCS_{DPR} = 87.6 %. La DPF3a (TCS_{DPR} = 58.8 %) arrive en quatrième place en termes d'enrichissement en DPRs derrière les IDPs BASP1, NUPR1 et α-syn. L'isoforme DPF3b est retrouvé en sixième position, après la protéine p53. La distribution de la proportion en NPRs semble être homogène (entre 12 à 17 %) dans l'ensemble de la base de données à l'exception de l'ostéopontine (TCS_{NPR} = 26.7 %) particulièrement riche en histidines (TCS_{His} = 15.7 %) et de BASP1 (TCS_{NPR} = 8.4 %) ne contenant aucun résidu His.

En regard des proportions en DPRs et OPRs au sein de cette base de données restreinte, deux limites arbitraires sont déterminées pour définir l'appartenance à la classe des IDPs. La première est déterminée à partir d'une valeur seuil en DPRs de 50 % au-delà de laquelle la protéine est considérée comme une IDP. L'ensemble des ORDPs sélectionnées sont retrouvées en-deçà de cette limite ainsi que l'ostéopontine dont le plus TCS_{DPR} peut s'expliquer par un enrichissement en NPRs, en particulier en résdisus His. La seconde limite, reposant sur une valeur seuil de 15 % de la différence entre la quantité en DPRs et celle en OPRs, semble être plus consistante étant donné

qu'elle intègre directement l'enrichissement en DPRs dans sa construction. Toutes les ORDPs se situent en-deçà de ce score alors que l'ensemble des IDPs et les isoformes de la DPF3 le dépassent largement (la différence est comprise entre 21.8 et 83.6 %) à l'exception de la TFIP11 qui l'atteint à peine.

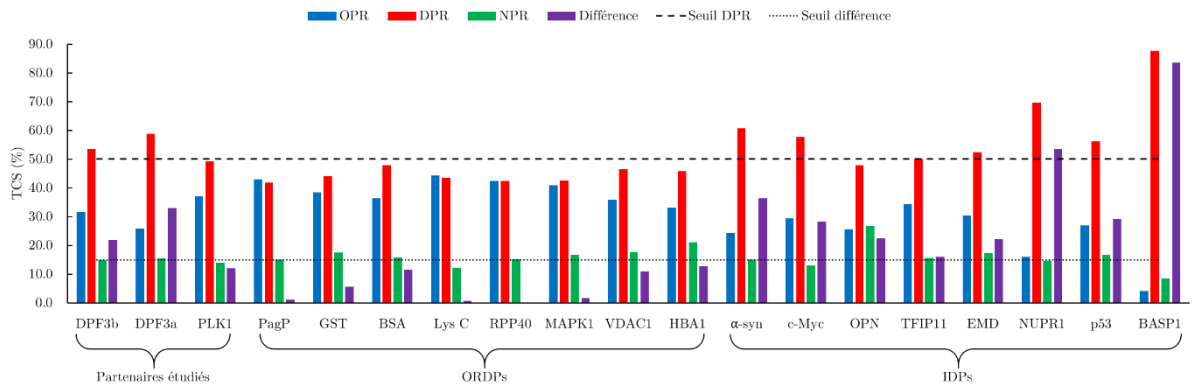

Figure 51 – Analyses histogrammiques de composition de séquence d'une sélection choisie de protéines ordonnées (ORDPs), désordonnées (IDPs) et des partenaires étudiés se basant sur un critère de promotion d'ordre (OPR), de désordre (DPR) et de non-promotion (NPR). Le taux de composition de séquence (TCS) est exprimé en pourcentage par rapport à la longueur de séquence totale. La différence est présentée en valeur absolue. La valeur seuil en DPR des IDPs (50 %) est représentée par un trait discontinu tandis que la valeur seuil de la différence des IDPs (15 %) est représentée par un trait en pointillé.

III. Hydropathie et charge nette moyennes

L'hydropathie (également orthographiée hydropathicité et de symbole H) est une mesure du caractère hydrophobe ou hydrophile d'une protéine sur la base de sa composition en résidus. Conceptuellement, elle peut être considérée comme synonyme d'hydrophilie, hydrophilicité, hydrophobie ou encore d'hydrophobicité. L'échelle d'hydropathie, développée par Kyte et Doolittle en 1982, consiste à attribuer un indice ou une valeur numérique à chacun des vingt acides aminés standards en fonction des propriétés hydrophobes et hydrophiles de leurs chaînes latérales. Il s'agit d'une échelle d'acides aminés comme il peut en exister beaucoup d'autres pour l'hydrophobicité ; plus de 98 échelles ont été élaborées depuis leur première formulation en 1962.[164] Un acide aminé possédant un indice d'hydropathie élevé a un caractère hydrophobe tandis qu'un faible indice (généralement négatif) est révélateur d'un acide aminé hydrophile. Parmi les acides aminés les plus hydrophobes, sont rencontrées la valine (H = 4.2) et l'isoleucine (H = 4.5). La lysine (H = -3.9) et l'arginine (H = -4.5) sont les deux acides aminés les plus hydrophiles de l'échelle. Pour connaître le score hydropathique d'une protéine, les valeurs individuelles des acides aminés composant cette dernière sont

additionnées à partir de l'extrémité N-terminale jusqu'à l'extrémité carboxyle. Le score hydropathique moyen (de symbole <H>) est une grandeur dérivant du score hydropathique total et permettant de mieux rendre compte du comportement de la protéine en milieu aqueux puisqu'elle correspond audit score divisé par la longueur de la séquence protéique (ou nombre de résidus). L'échelle proposée par Kyte J. et Doolittle R. F. peut être normalisée de zéro à un pour simplifier l'expression du score hydropathique total ou moyen.[165]

La charge nette (de symbole R) est définie comme la valeur absolue de la différence entre la proportion en résidus chargés positivement et chargés négativement dans la structure primaire de la protéine. L'état chargé des acides aminés est établi à pH neutre (= 7.0). Les résidus considérés positivement chargés sont la lysine et l'arginine tandis que l'aspartate et le glutamate sont négativement chargés. Bien que la charge de l'histidine à pH physiologique soit quelque peu sujet à controverse, il est généralement admis que les résidus His sont majoritairement déprotonés à pH 7.0. C'est pourquoi, l'histidine n'est pas incluse dans le dénombrement des résidus chargés positivement. Similairement au score hydropathique moyen, la charge nette moyenne (de symbole <R>) est calculée à partir de la charge nette divisée par la longueur de la séquence protéique. La charge nette moyenne peut également être exprimée par rapport au nombre total de résidus chargés plutôt que la totalité de la séquence.[83]

Les études d'Uversky *et al.* ont montré qu'il existe un lien étroit entre l'hydropathie et la charge nette d'une protéine pour déterminer si elle appartient à la classe des IDPs. Elles présentent généralement une plus faible hydropathie (ou hydrophobicité) et une charge nette moyenne plus élevée de par leur plus large composition en résidus polaires et chargés. Ces paramètres peuvent être facilement calculés en connaissant uniquement la séquence de la protéine. La charge nette moyenne <R> (par rapport au nombre total de résidus chargés) est exprimée en fonction du score hydropathique moyen <H> normalisé à un et calculé à partir du serveur ExPASy.[125] La limite entre les IDPs et les ORDPs est exprimée par l'équation suivante $<R> = 2.785<H> - 1.151$.[83,85]

La distribution charge-hydropathie pour la base de données protéiques (cf. Fig. 52) apparaît concordante avec les critères choisis. L'ensemble des ORDPs sont caractérisées par un <H> élevé et une faible <R> à l'exception du Lys C dont la charge nette moyenne est particulièrement élevée. D'ailleurs, à cause de ce score <R>, le Lys C est retrouvé dans la zone des IDPs, ce qui n'est pas le cas des autres ORDPs. La PLK1 est localisée dans la zone des ORDPs en restant proche de la limite à l'instar

de la BSA. Concernant le groupe des IDPs, elles sont toutes comprises dans la zone correspondante avec des configurations de charge-hydropathie diverses. Les isoformes de la DPF3 y sont également retrouvées. La protéine p53 est extrêmement proche de la limite avec les protéines globulaires. L'OPN présente la plus haute charge nette moyenne, IDPs et ORDPs confondues. L'α-syn possède la distribution charge-hydropathie la plus élevée de toutes les IDPs tandis que celle de la DPF3a est la plus basse (faible score hydropathique couplé à une très faible charge nette moyenne). Bien que l'analyse demeure prédictive, la DPF3 est une nouvelle fois associée à la classe des IDPs.

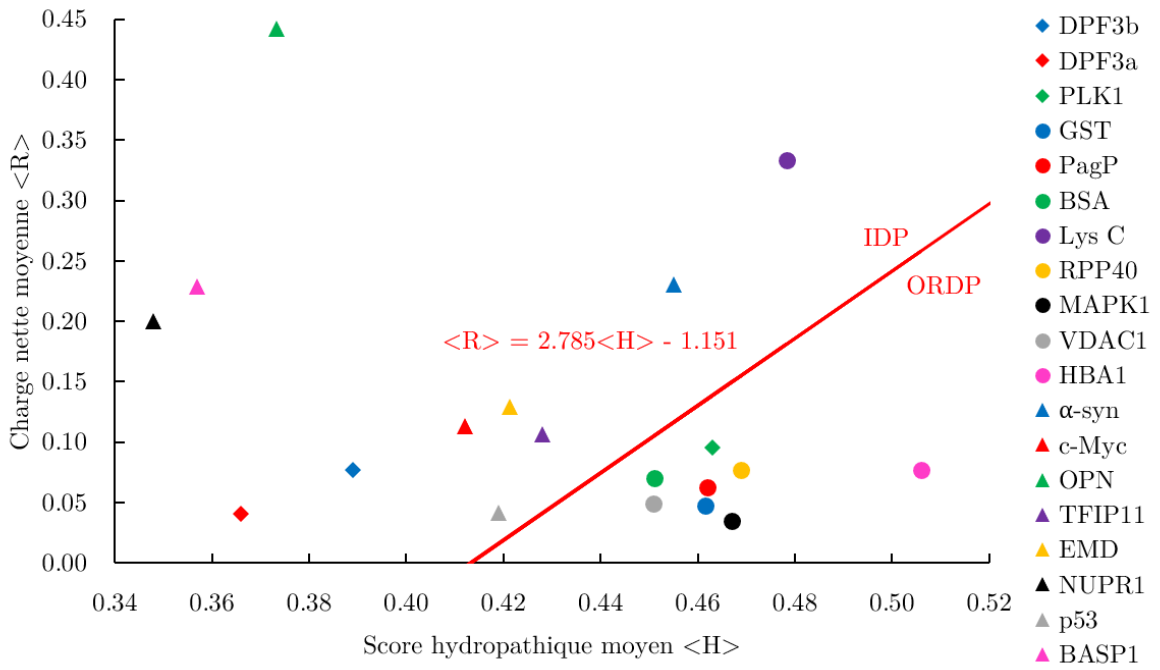

Figure 52 – Distribution charge-hydropathie d'une sélection choisie de protéines ordonnées (ORDPs), désordonnées (IDPs) et des partenaires étudiés. L'évolution de la charge nette moyenne <R> est exprimée en fonction du score hydropathique moyen <H>. La limite ordre-désordre est représentée par un trait continu rouge d'équation <R> = 2.785<H> - 1.151.

IV. Diagramme de phase de Das-Pappu

Le diagramme de phase (ou d'état) de Das-Pappu est subdivisé en cinq régions distinctes (respectivement abrégées R1 à R5) en fonction des rapports en fraction de résidus positivement et négativement chargés (cf. Fig. 53). [166,167] Chacune de ces régions correspond à un ensemble conformationnel protéique type. La région R1 comprend les protéines globulaires ou de type « tétard » (*tadpole-like*) possédant une « tête » globulaire ordonnée et une « queue » désordonnée. De par l'existence d'un cœur hydrophobe stable et la dominance des interactions intra-chaîne, il s'agit de

polyélectrolytes ou polyampholytes faiblement chargés. Une certaine population d'IDPs, à raison de 25 %, est identifiée dans R1. La région R2 est une région limite faisant la transition entre R1 et R3. Elle regroupe les conformations de type enroulé (*coil*) ou les chimères de globules. Les protéines constituant ce groupe sont également désignées par la dénomination de séquences Janus (à l'image de la divinité romaine à double visage) de par la dépendance qu'entretient leur structure avec les conditions dans lesquelles elles se trouvent et que cette structure est définie selon deux états : soit effondrée par effet hydrophobe soit étendue par une solvatation de chaîne favorable. La majorité des IDPs, environ 40 %, sont retrouvées dans la région R2. Les polyampholytes forts à conformation enroulée ou en forme d'épingle à cheveux définissent la région R3 ; environ 30 % des IDPs occupent ce groupe. Enfin, les régions R4 et R5 (parfois simplifiées en R4) représentent respectivement des polyélectrolytes forts chargés négativement ou positivement. D'un point de vue conformationnel, elles s'apparentent à des structures enroulées (*coil-like*) ou semi-flexibles (*rod*) ; seuls environ 5 % des IDPs composent les régions R4 et R5. Les diagrammes sont générés séparément pour les ORDPs et PLK1 (cf. Fig. 54a) et pour les IDPs et les deux isoformes de la DPF3 (cf. Fig. 54b) à l'aide de l'outil en ligne et libre d'accès CIDER.[168]

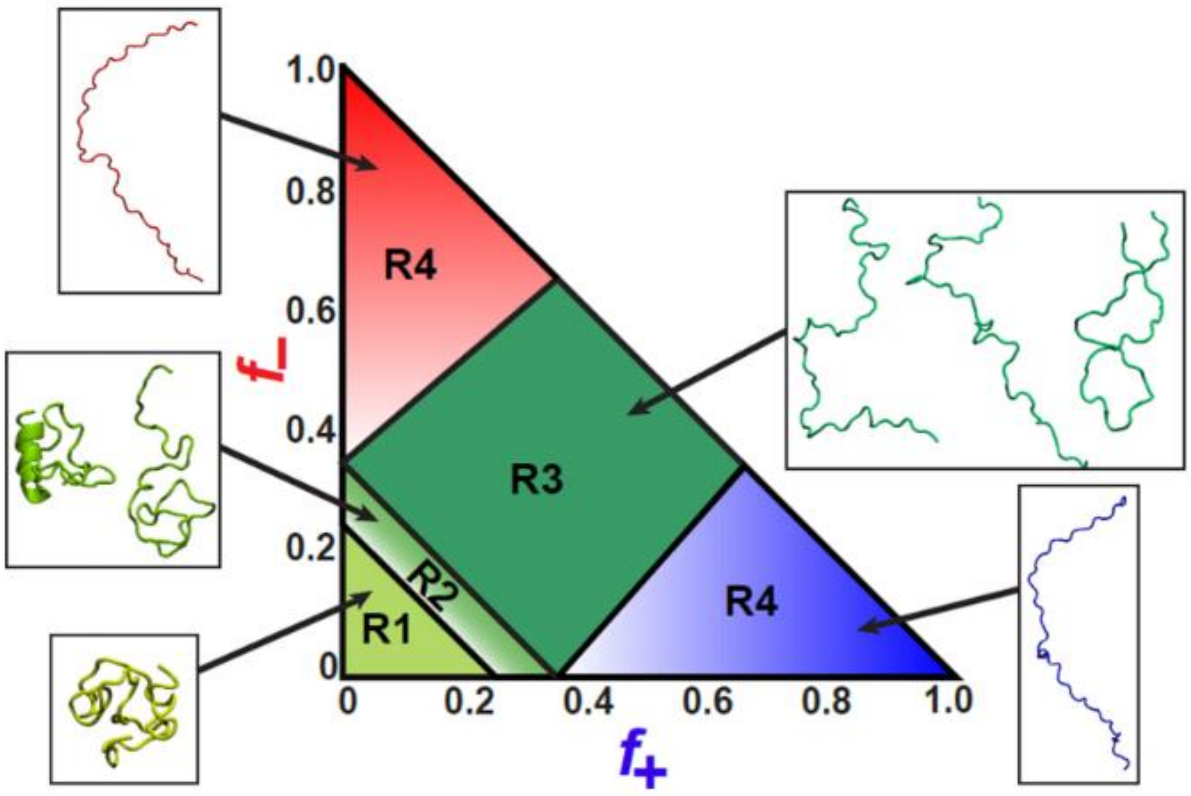

Figure 53 – Diagramme de phase de Das-Pappu pour la classification des classes conformationnelles des séquences des IDPs. Dans cette représentation, les régions R4 et R5 sont indifférenciées.[167]

Au sein des ORDPs sélectionnées, la grande majorité appartient à la région R1, ce qui est cohérent avec leur nature conformationnelle puisqu'il s'agit de protéines globulaires de structure 3D bien définie. Néanmoins, la BSA, la GST ainsi que la PLK1 sont toutes les trois associées à la région de transition R2. Cette répartition n'est pas étonnante étant donné que ces trois protéines présentent les plus grandes proportions en résidus chargés vis-à-vis des autres ORDPs (cf. Fig. 50c), ce qui en fait des polyélectrolytes plus chargés. Concernant les IDPs, la plupart sont réparties entre les régions R2 (α-syn, c-Myc, OPN, TFIP11 et NUPR1) et R3 (BASP1), ce qui est attendu pour des structures désordonnées. Par ailleurs, si la BASP1 est retrouvée dans la région R3, c'est parce qu'il s'agit du polyampholyte le plus fort ($TCS_{\text{Chargé BASP1}} = 36.56$ %) de toutes les protéines de la base de données choisie. Cependant, deux d'entre elles apparaissent dans la région R1 : l'émerine et la protéine p53. Cette différence s'explique par la déplétion plus marquée en résidus chargés chez ces deux protéines vis-à-vis des autres IDPs considérées, ce qui en fait des polyélectrolytes plus faibles. Aussi, elle peut signifier une plus grande stabilité conformationnelle ou la présence d'un cœur hydrophobe structurant, ce qui est le cas pour la p53. Un rapprochement similaire ne peut être effectué pour l'EMD étant donné qu'il n'en existe aucune structure complète déterminée expérimentalement. Les deux isoformes de la DPF3 sont retrouvés dans la région R2. Une nouvelle fois, l'analyse prédictive de la séquence de la DPF3 corrèle les observations précédentes et rend compte d'un profil conformationnel propre à une IDP. Concernant la PLK1, les résultats de l'ensemble des analyses prédictives ne permettent pas de la classer aussi clairement. Alors que sa composition de séquence ainsi que sa distribution charge-hydropathie reflètent les caractéristiques d'une ORDP, le diagramme de Das-Pappu la renvoie à la région R2, propre aux protéines désordonnées ou partiellement désordonnées.

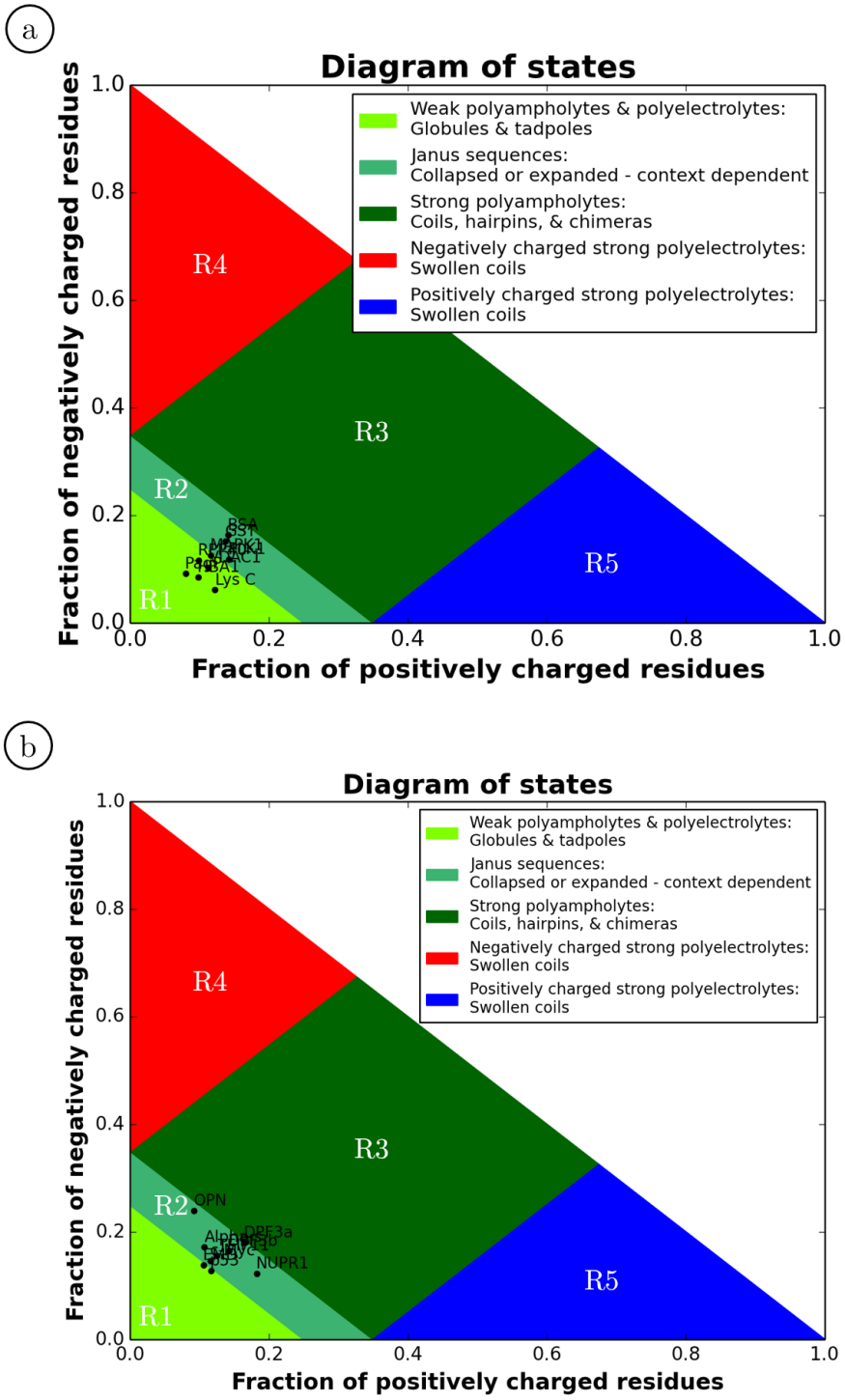

Figure 54 – (a) Diagrammes de phase de Das-Pappu des ORDPs et de PLK1 ainsi que (b) des IDPs et des isoformes de la DPF3. L'évolution de la fraction en résidus chargés négativement est exprimée en fonction de la fraction en résidus positivement chargés.

Partie n°4

—

Conclusions et perspectives

I. Conclusions

Ce travail de recherche s'était donné pour objectifs la purification ainsi que la caractérisation par une combinaison de méthodes biophysiques et prédictives des partenaires protéiques constituant le complexe DPF3-PLK1. Cette démarche était motivée par une meilleure appréhension de la formation dudit complexe afin d'en comprendre l'implication dans la mitose et d'en concevoir des inhibiteurs sélectifs.

D'un point de vue global, l'ensemble des procédés de purification, quel que soit leur format, se sont révélés efficients moyennant la résolution du clivage par des protéases endogènes des souches bactériennes de production. Bien qu'il ait été limité par addition de PMSF dans les tampons de purification, il n'est pas complètement empêché étant donné que le clivage s'opère sur les protéines de fusion au cours de leur surexpression. En outre, cette sensibilité à la protéolyse des protéines complètes (*full length*) vis-à-vis de la résistance du tandem de doigts de zinc PHD-1+2 a fourni le premier indice expérimental quant à la nature IDP de la DPF3 et possiblement de la PLK1.

La protéine de fusion DPF3a-GST s'est avérée suivre les protéiques globulaires typiques en termes de caractérisation biophysique. Elle est détectable sur gel par coloration au bleu de Coomassie, présente une absorbance maximale vers 278-280 nm en spectroscopie UV-visible, une bande d'émission à 335 nm en ITF ainsi qu'un profil propre aux hélices α en CD. Toutefois, il est probable que la GST ait une contribution non négligeable à ces différentes signatures. Les analyses en DLS et en SEC ont mis en évidence un phénomène d'agrégation ou d'oligomérisation de la DPF3a-GST, qui s'accentue au cours du temps.

Contrairement à son homologue fusionné, la DPF3a sous sa forme clivée a dévoilé des propriétés biophysiques particulièrement peu communes dans le domaine des protéines (cf. Table 4). Ces propriétés s'avèrent d'autant plus significatives et propres à la DPF3a de par leur reproductibilité. Les analyses préliminaires par migration SDS-PAGE ont été marquées par l'absence systématique de la protéine d'intérêt sur gel quel que soit le procédé de coloration. En spectroscopie d'absorption UV-visible, la DPF3a absorbe dans une gamme de longueurs d'onde inférieure à 280 nm : entre 250 et 272 nm avec une certaine diversité quant aux profils d'absorbance et une bande principale à 258-260 nm. Pour expliquer ces deux phénomènes distincts en SDS-PAGE et en spectroscopie d'absorption, trois hypothèses ont été formulées. La première décrit une contamination par l'ADN, présentant une forte absorbance à 260 nm. Toutefois, des traitements à la DNase I n'ont montré aucune différence quant au

comportement de la DPF3a clivée sur gel ou en absorption. En outre, aucune trace d'ADN n'a été détectée par AGE. L'hypothèse de contamination par l'ADN a, par conséquent, été écartée.

	DPF3a-GST	DPF3a
SDS-PAGE	Visible sur gel (bleu de Coomassie et imidazole-zinc)	Non-visible sur gel (bleu de Coomassie et imidazole-zinc)
Absorption UV-visible	Une bande d'absorption : $\lambda_{abs\ max}$ = 278-280 nm	Une bande d'absorption : $\lambda_{abs\ max}$ = 250-272 nm
ITF (λ_{exc} = 295 nm)	Une bande d'émission : $\lambda_{ém\ max}$ = 335 nm	Deux bandes d'émission : $\lambda_{ém\ max}$ = 330-334 nm et $\lambda_{ém\ max}$ = 457-475 nm
dbAF (λ_{exc} = 350 nm)	Aucune bande d'émission	Une bande d'émission (signature amyloïde) : $\lambda_{ém\ max}$ = 440 nm
CD	Une signature (bandes négatives) : hélices α (206 et 223 nm)	Deux signatures (bandes négatives) : *random coil* (200 nm) et feuillets β «twistés» (227-230 nm)
DLS	Agrégation : D_h = 77.2 ± 8.4 nm	Forte agrégation : D_h = 180-550 nm
SEC	Trois populations : GST, DPF3a-GST et agrégats (élution dans le volume mort)	Une population : agrégats (élution dans le volume mort)

Table 4 – Récapitulatif comparatif des propriétés biophysiques de la protéine de fusion DPF3a-GST en regard de son homologue clivé DPF3a.

La seconde hypothèse s'appuie sur un phénomène d'agrégation de la protéine, rendant compte de son impossibilité à traverser le réseau réticulé de polyacrylamide du gel et d'artefacts sur les mesures en spectroscopie d'absorption, modifiant l'allure des spectres. L'ensemble des résultats de caractérisation concordent et s'accordent à décrire une grande propension à l'agrégation. Elle est supportée par l'apparition d'une seconde bande d'émission dans le visible en ITF, des objets de très grande taille en SEC, éluant systématiquement dans le volume mort, ainsi que par l'ordre de grandeur (entre 180 et 550 nm) des diamètres hydrodynamiques mesurés en DLS. La caractérisation en DLS a, par ailleurs, montré que les agrégats de la DPF3a présentaient une sensibilité à

caractère réversible à la température, déviant par rapport à la loi de Stokes-Einstein, mais pas au pH. Les agrégats semblent gonfler en chauffant et retrouver leur taille d'origine en refroidissant. Des analyses complémentaires tendent à mieux définir ce phénomène en la formation de structures agrégées spécialisées de type amyloïde. En effet, le spectre CD de certains échantillons de la DPF3a a révélé une signature caractéristique de feuillets β «twistés», à 227-230 nm, qui sont des motifs de structure secondaire récurrents au sein des fibrilles amyloïdes. De plus, l'émergence d'une fluorescence intrinsèque dans le bleu-vert, dont l'occurrence n'est pas due aux résidus aromatiques, a été spécifiquement découverte dans différentes fractions de la DPF3a clivée. Ce phénomène d'émission, connu sous le nom d'autofluorescence dans le bleu profond, est une particularité des agrégats amyloïdes.

Quant à la troisième hypothèse, il s'agit de la complexation de la DPF3a avec des cations métalliques par l'intermédiaire de ses deux doigts de zinc C_2H_2 et PHD-1/2. Les premiers tests ont rendu compte de l'influence des ions Zn^{2+} sur les comportements spectroscopiques de la protéine. En absorption UV-visible, l'addition de chlorure de zinc provoque un effet hypochrome constant et linéaire dans le temps (sur un intervalle de 60 minutes) à la longueur d'onde d'absorbance maximale. Cette diminution du coefficient d'extinction molaire peut être induite par une transition conformationnelle désordre-ordre ou par une perturbation du phénomène d'agrégation étant donné que certains métaux sont structurants et peuvent promouvoir ou entraver l'agrégation protéique. En ITF, la présence de zinc augmente l'intensité de fluorescence et déplace la seconde population de 460 à 485 nm. Ce phénomène s'accompagne d'un *quenching* du signal à trop haute concentration en Zn^{2+}. Le zinc serait à l'origine d'une modification de la conformation de la DPF3a, augmentant l'exposition des résidus Trp au solvant, conduisant aux effets hyper- et bathochromes observés. Le *quenching* surviendrait d'une interaction, par exemple de type π-cation, entre les Trp et les ions Zn^{2+}. Ainsi, les deux hypothèses d'agrégation et de complexation protéine-métal apparaissent compatibles. Leur combinaison est par ailleurs essentielle à la compréhension du comportement *in vitro* de la DPF3a de sorte à obtenir l'aperçu le plus exhaustif de ses propriétés biophysiques.

En regard des analyses prédictives, chacune d'entre elles a montré que les isoformes de la DPF3, DPF3b et DPF3a, ont des propriétés associées au désordre, dont l'expression est plus exacerbée chez la DPF3a. Leur séquence est non seulement enrichie en résidus hydrophiles, polaires et chargés, mais également en DPRs au-dessus des deux valeurs seuils déterminées arbitrairement pour différencier les IDPs des ORDPs. Cette

nature désordonnée est soutenue par leur faible distribution charge-hydropathie les localisant dans la zone propre aux IDPs ainsi que par leur appartenance à la région R2 du diagramme de phase de Das-Pappu, caractéristique des séquences Janus et des conformations du type *random coil*. Toutefois, bien qu'elles tendent à confirmer la DPF3 comme une nouvelle IDP, elles ne peuvent se substituer aux preuves expérimentales. Des résultats préliminaires obtenus en CD concordent avec les analyses prédictives. En effet, une signature propre aux protéines désordonnées ou en *random coil* a été observée pour la DPF3a clivée avec une bande négative vers 200 nm. Le caractère IDP est également cohérent avec l'association en fibrilles amyloïdes étant donné que cette classe protéique est impliquée dans ce genre de structure, à l'image de l'α-synucléine.

II. Perspectives

Jusqu'à présent, ce travail de recherche s'est orienté en direction de la caractérisation biophysique de la DPF3a plutôt que sur les interactions qu'elle, ou son isoforme DPF3b, pourrait entretenir avec la kinase PLK1. C'est pourquoi, la purification et le clivage de la DPF3b ainsi que de la PLK1 doivent être élaborés afin de pouvoir procéder aux premiers essais d'interaction. Ceux-ci prendront la forme de *GST pull-down* entre la DPF3 clivée et la PLK1-GST et feront office de validation préliminaire à l'hypothèse de formation du complexe DPF3-PLK1. Ils seront poursuivis par titrage calorimétrique isotherme (abrégé ITC pour *isothermal titration calorimetry*) afin d'en dégager les constantes d'affinité et la stœchiométrie de liaison. La formation du complexe pourrait également être suivie par fluorescence d'anisotropie à l'aide de l'un des partenaires protéiques fusionné à un fluorophore.[169,170] La taille et dynamique du complexe pourront être évaluées par une analyse combinée en DLS et SEC. La présence de modifications post-traductionnelles sur la séquence de la DPF3, à l'image de motifs phosphorylés par intervention de la PLK1, sera mise en évidence par spectrométrie de masse. Ces techniques de caractérisation des PPIs seront ensuite étendues à l'interaction entre des domaines spécifiques de la PLK1 et des isoformes de la DPF3.

Une fois les propriétés de la DPF3a élucidées, ou en tous les cas précisées, il est intéressant de voir si elles sont transférables à l'isoforme DPF3b. Ainsi, en regard du profil biophysique de la DPF3a, il s'agira d'opérer à un travail de caractérisation analogue sur l'isoforme DPF3b par une combinaison de techniques expérimentales : SDS-PAGE, spectroscopies d'absorption et d'émission, CD, DLS et SEC. Cependant,

le travail est loin d'être terminé pour la DPF3a. En effet, bien que l'investigation de la nature désordonnée de la DPF3a par des méthodes prédictives se soit révélée prometteuse, elle doit être étudiée plus avant sur base d'un plus large panel de preuves expérimentales. À cet effet, certains résultats préliminaires, notamment en dichroïsme circulaire, tendent d'ores et déjà à montrer des signatures typiques d'une IDP. À ce type d'analyse pourront s'additionner des approches structurales en RMN et en SAXS pour lesquelles les IDPs présentent des signatures typiques par rapport à une protéine globulaire ou repliée.[92,171–173] Ces caractérisations devront également être étendues aux protéines DPF3b et PLK1 à des fins comparatives.

Pour y parvenir, l'agrégation de la DPF3a doit être mieux appréhendée afin d'être contrôlée. À cet égard, l'influence du glycérol à différentes concentrations et d'autres agents anti-agrégants sera mesurée par DLS. Des conditions dénaturantes seront également testées pour en évaluer l'impact sur la taille des particules : s'il y a prévention ou exacerbation du phénomène d'agrégation. Aussi, la minimisation du phénomène d'agrégation pourrait se manifester par des modifications des comportements spectroscopiques observés pour la DPF3a à l'état agrégé.

La formation préférentielle de fibrilles amyloïdes doit être confirmée pour la DPF3a. Pour ce faire, leur visualisation peut être envisagée par microscopie électronique à balayage (abrégée SEM pour *scanning electron microscopy*) et investiguée par fluorescence de la thioflavine T (abrégée ThT), un colorant benzothiazolé montrant une affinité pour l'empilement de feuillets β.[174,175] Cette validation serait une preuve supplémentaire concernant la nature désordonnée la DPF3a.

Le mode d'action du zinc sur l'état conformationnel de la DPF3a doit être approfondi par d'autres méthodes biophysiques. À cet égard, l'affinité de liaison ainsi que la stœchiométrie relative pourront être déterminées par ITC. Cette analyse permettrait entre autres d'investiguer l'interaction spécifique entre la DPF3 et les cations Zn^{2+}. Si la présence des ions zinciques provoque une transition d'une structure désordonnée à ordonnée, des études complémentaires en CD seront envisagées pour identifier des modifications au niveau des signatures spectrales. La contribution de ces cations à l'agrégation sera évaluée par des mesures DLS.

Afin de compléter les analyses déjà existantes, les différentes bandes obtenues en SEC seront caractérisées par LC-ESI-MS/MS afin d'en identifier la distribution protéique. Complémentairement et en tant qu'alternative aux migrations non

concluantes en SDS-PAGE, des *Western blot* avec un anticorps de la DPF3a pourront être effectués afin d'en confirmer la présence dans les échantillons purifiés et clivés.

À l'aide d'une série d'algorithmes, les analyses prédictives seront poursuivies et approfondies, notamment par la prédiction de désordre le long de la chaîne.

La conception et le développement d'inhibiteurs spécifiques à l'interaction entre la DPF3 et la PLK1 ou encore capables de stabiliser l'ensemble conformationnel des isoformes de la DPF3 (de par leur nature IDP) font toujours office d'objectifs à long terme. Ils nécessitent une compréhension globale aussi bien d'un point de vue structural que fonctionnel des protéines ciblées. C'est pourquoi, les caractérisations biophysiques de chacun des partenaires protéiques sont essentielles.

Partie n°5

—

Matériels et méthodes

I. Référence et composition des tampons

- Tampon de lyse ou LB (*lysis buffer*) : PBS pH 7.3, TritonX™-100 0.5 %, 200 mM KCl.
- Tampon de lyse ou LB (*lysis buffer*) pour le traitement à la DNase I : 10 mM Tris-HCl pH 7.6, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, TritonX™-100 0.5 %.
- Tampon de fixation ou BB (*binding buffer*) : PBS pH 7.3, 200 mM KCl.
- Tampon d'élution ou EB (*elution buffer*) : 50 mM Tris-HCl pH 8.0, 10 mM ou 20 mM GSH.
- Tampon du facteur Xa ou FB (*factor Xa buffer*) : 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$.
- Tampon d'élution du facteur Xa ou FEB (factor Xa elution buffer) : 50 mM glycine-HCl pH 3.0.
- Tampon acétate de stockage de la benzamidine ou BSB (*benzamidine storage buffer*) : 50 mM $NaCOOCH_3$ pH 4.0 dans l'éthanol 20 %.

II. Transformation de souches compétentes de *E. coli* pour l'expression d'une protéine de fusion marquée à la GST

A. DPF3, isoformes et domaines associés

Les plasmides de DPF3a FL, DPF3b et PHD-1+2 ont été construits avec le vecteur pGEX-3X de GE Healthcare, préalablement transformés et donnés par l'équipe du Pr. Denis Mottet de la Grappe Interdisciplinaire de Génoprotéomique Appliquée (GIGA) de l'Université de Liège. Les transformations reçues pour chaque construction ont été réalisées dans des souches de *E. coli* BL21 Rosetta™ 2 et stockées dans du glycérol 60 %. Le vecteur pGEX-3X code pour la protéine de fusion recombinante marquée à la GST à l'extrémité N-terminale et inclut un site de clivage spécifique au facteur Xa inséré entre le *tag* et la protéine d'intérêt. Le facteur Xa, la forme activée du facteur X ou de Stuart-Prower, est une sérine endopeptidase intervenant dans la cascade de la coagulation. Le motif de reconnaissance de cette protéase est une séquence de quatre à cinq acides aminés au sein de laquelle le site de clivage est marqué par une flèche rouge : I-E-G-R↓(G). Les deux isoformes de DPF3 ne contiennent pas dans leur structure primaire ledit motif de reconnaissance du facteur Xa. Le vecteur pGEX-3X est pourvu d'un gène de résistance à l'ampicilline, un antibiotique de la famille des bêtalactamines.

B. PLK1 et domaines associés

Pour le procédé de transformation de PLK1, les plasmides de PLK1 FL, KIN et PBD ont été construits avec le vecteur pGEX-6P-1 de chez GE Healthcare et donnés par l'équipe du Pr. Izabela Sumara de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch en France. Le vecteur pGEX-6P-1 code pour la protéine de fusion recombinante marquée à la GST à l'extrémité N-terminale et inclut un site de clivage spécifique à la PreScission™ protéase inséré entre le *tag* et la protéine d'intérêt. La PreScission™ protéase est une protéine de fusion du rhinovirus humain 3C et de la GST. Le motif de reconnaissance de cette protéase est une séquence de huit acides aminés au sein de laquelle le site de clivage est marqué par une flèche rouge : L-E-V-L-F-Q↓G-P. PLK1 ne contient pas dans sa structure primaire ledit motif de reconnaissance de la PreScission™ protéase. Le vecteur pGEX-6P-1 est pourvu également d'un gène de résistance à l'ampicilline.

L'ensemble des étapes de transformation sont réalisées en conditions stériles. Les solutions plasmidiques sont centrifugées à l'aide d'une microcentrifugeuse Eppendorf™ 5420 GLP. Pour chaque construction, 50 µL de la souche BL21(DE3) décongelée ainsi que 3 µL de solution plasmidique sont mélangés dans un tube Eppendorf™ stérilisé, qui est incubé sur glace entre 30 et 45 minutes. Six boîtes de Petri (à savoir deux boîtes par construction pour PLK1 FL, KIN et PBD) sont coulées avec du LB Lennox™ agar (gélose) stérilisé contenant 0.24 mM d'ampicilline. Les souches BL21(DE3) subissent un choc thermique dans un bain thermostaté à 42 °C pendant 45 secondes. Elles sont replacées sur glace pendant 3 minutes et 450 µL de LB Lennox™ stérilisé sont additionnés. Les transformations sont incubées dans un thermomixeur Eppendorf™ Comfort AP-895 à 37 °C sous agitation constante réglée à 800 rpm pendant une heure. Après incubation, 100 µL de transformation sont déposés dans une première série de trois boîtes de Petri annotées 1X (pour une fois concentrées) et du code associé à la construction (PLK1 FL, KIN ou PBD). La transformation est étalée sur toute la surface de la boîte à l'aide d'une dizaine de billes de verre stériles. Le restant de la transformation est centrifugé à 8000 rpm pendant une minute à l'aide d'une microcentrifugeuse Eppendorf™ 5420 GLP et 350 µL de surnageant sont écartés afin de reconcentrer environ dix fois les bactéries transformées ; le culot bactérien est resuspendu dans les 50 µL de milieu encore présents. Ces 50 µL sont transférés complètement dans la seconde série de trois boîtes de Petri annotées 10X (pour dix fois concentrées) et du code associé à la construction (PLK1 FL, KIN ou PBD). La transformation est également étalée à l'aide de billes de verre stériles. Les boîtes sont

refermées et incubées à 37 °C o/n. La présence de colonies efficacement transformées est vérifiée la journée du lendemain.

C. Stockage à long terme de souches de *E. coli* efficacement transformées

Après incubation des boîtes de Petri, une préculture est lancée pour chaque construction à partir d'une colonie de BL21(DE3) transformée. À l'aide d'une boucle d'inoculation, une colonie bactérienne est récupérée et déposée dans un erlenmeyer stérilisé de 500 mL avec chicanes contenant 100 mL de LB Lennox™ avec 0.125 mg/mL d'ampicilline. Les précultures sont incubées à 37 °C o/n sous agitation. Après incubation, pour chaque construction 500 µL de préculture sont prélevés et mélangés à 500 µL de glycérol 60 % préalablement stérilisé. Chaque aliquot est vortexé, annoté et stocké à -80 °C.

III. Surexpression hétérologue dans *E. coli* d'une protéine de fusion marquée à la GST

L'expression des protéines de fusion marquées à la GST est contrôlée par induction à l'IPTG. L'ensemble des vecteurs pGEX comprennent dans leur construction un promoteur tac ainsi qu'un gène $lacI^q$. Le produit du gène $lacI^q$ est un répresseur assurant le contrôle de l'expression de l'insert et de la protéine par liaison à l'opérateur du promoteur tac. En présence d'IPTG, qui est un analogue de lactose, le mécanisme de répression est bloqué permettant l'expression de la protéine d'intérêt.[103]

Le procédé de surexpression se déroule en quatre étapes : préculture, culture, induction à l'IPTG, récupération et stockage des culots bactériens. L'ensemble des étapes de surexpression sont réalisées en conditions stériles. Concernant l'étape de préculture, 100 mL de LB Lennox™ préalablement stérilisé sont transférés dans un erlenmeyer stérile avec chicanes (ou déflecteurs) de 500 mL. De l'ampicilline y est additionnée afin d'en obtenir une concentration de 0.125 mg/mL. Une fraction du stock glycérol de la construction souhaitée est déposée dans le milieu LB à l'aide d'une boucle d'inoculation. La préculture est incubée à 37 °C o/n sous agitation.

Concernant l'étape de culture, quatre erlenmeyers stériles avec chicanes sont préparés identiquement par addition de 250 mL de LB Lennox™ avec 0.125 mg/mL d'ampicilline. Dans chacun, 5.0 mL de préculture sont transférés et les cultures sont incubées à 37 °C sous agitation. La croissance bactérienne est suivie par mesure de la densité optique (abrégée DO) à 600 nm à l'aide d'un spectrophotomètre Thermo

Spectronic Genesys 8 ; la mesure d'absorbance à cette longueur d'onde rend compte de la turbidité du milieu. Les cultures ne sont induites que lorsqu'elles présentent une DO suffisante, à savoir comprise entre 0.5 et 0.7 après soustraction de l'absorbance du blanc (LB Lennox™). Cette gamme de DO est caractéristique d'une population bactérienne ayant atteint sa phase exponentielle de croissance.[176] Chaque souche bactérienne transformée possède un temps de croissance optimal lui étant propre à déterminer par des mesures de DO à intervalles réguliers. À titre indicatif, PLK1 FL entre en phase exponentielle de croissance après environ 2h00-2h10 d'incubation, DPF3a FL après 2h30-2h40 et DPF3b FL après seulement 2h50-3h00.

Concernant l'étape d'induction, une solution d'IPTG est additionnée dans chaque culture afin d'en obtenir une concentration de 0.5 mM. L'induction est lancée pendant quatre heures à 37 °C sous agitation.

Concernant l'étape de récupération des culots bactériens induits, le contenu de chaque erlenmeyer est transféré après induction dans un flacon à centrifuger stérile de 250 mL. Les flacons sont refermés et centrifugés pendant 30 minutes à 4000 rpm à 4 °C à l'aide d'une centrifugeuse Eppendorf™ 5804 R. Après centrifugation, le culot bactérien est conservé et le surnageant est éliminé dans une tourie d'eau de Javel. Les culots sont resuspendus dans 15 à 20 mL de LB Lennox™ et les suspensions sont transférées dans des tubes Falcon™ de 50 mL. Elles sont à nouveau centrifugées pendant 15 minutes à 4000 rpm à 4 °C, les surnageants sont écartés et les culots sont finalement annotés et stockés à -20 °C.

IV. Lyse par sonication de souches de *E. coli* induites à l'IPTG

Les culots bactériens sont décongelés à température ambiante et suspendus dans quelques millilitres de tampon de lyse, abrégé LB (pour *lysis buffer*), de composition suivante : tampon PBS (pour *phosphate-buffered saline*) pH 7.3, TritonX™-100 0.5 %, KCl 200 mM. La solution stock de PBS pH 7.3 est de composition suivante : KH_2PO_4 1.76 mM, NaCl 137 mM, KCl 2.7 mM, Na_2HPO_4 10 mM. Pour les manipulations ayant nécessité un traitement du lysat à la DNase I, le LB est de composition suivante : 10 mM Tris-HCl pH 7.6, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, TritonX™-100 0.5 % ; les cations divalents Mg^{2+} et Ca^{2+} sont nécessaires à l'activité de la DNase I.

Le tube Falcon™ contenant la suspension est porté à environ 15 mL à l'aide du LB adéquat. Le PMSF, un inhibiteur non spécifique de sérine protéases, est ajouté à raison de 120 µL d'une solution stock de 25 mM dans l'isopropanol afin d'en obtenir une concentration finale de 200 µM. Le sonicateur est calibré sur une puissance 6 et

chaque suspension est soniquée sur glace six fois 30 secondes avec 30 secondes de repos entre chaque cycle pour éviter toute surchauffe. Après sonication, la suspension est centrifugée pendant 30 minutes à 12000 x g à 4 °C à l'aide d'une centrifugeuse VWR™ Mega Star 600. Le surnageant, autrement dit le lysat, est récupéré et le culot contenant les débris cellulaires est jeté.

Si un traitement à la DNase I est nécessaire, le lysat est additionné d'environ 500 µg de DNase I lyophilisée pour une concentration finale de 33 µg/mL et incubé deux heures à 4 °C sous agitation légère. Du PMSF est ajouté à raison de 120 µL d'une solution stock 25 mM dans l'isopropanol après une heure d'incubation et à la fin de cette dernière. En présence d'une solution aqueuse, le PMSF est particulièrement instable et se dégrade endéans une certaine période de temps au-delà de laquelle il perd son activité inhibitrice. Ce temps de demi-vie est fonction du pH et de la température : à 25 °C, le temps de demi-vie du PMSF dans l'eau est de 110 minutes à pH 7.0, 55 minutes à pH 7.5 et 35 minutes à pH 8.0.[177]

V. Méthodes de purification et de clivage en phase liquide d'une protéine de fusion marquée à la GST

A. Purification sur colonne d'écoulement par gravité

Le procédé de purification par écoulement par gravité est réalisé dans une colonne chromatographique de 12 mL de chez GoldBio dans laquelle est coulée manuellement la matrice de Glutathione Sepharose™ 4B (abrégée GS4B). Chaque tampon employé est additionné de PMSF 200 µM en respectant le temps de demi-vie de ce dernier en regard du pH du tampon. Un total de 4 mL de suspension GS4B dans l'éthanol 20 % est transféré dans la colonne chromatographique afin d'obtenir environ 3 mL de matrice (1.33 mL de suspension ≅ 1 mL de matrice). La colonne est drainée de sa solution d'éthanol 20 % de stockage et tassée.

La matrice est équilibrée par l'équivalant de cinq volumes de matrice (environ 15 mL) au tampon de fixation, abrégé BB (pour *binding buffer*), de composition suivante : tampon PBS pH 7.3, KCl 200 mM. Après équilibration, le lysat centrifugé est déposé sur la matrice et la colonne fermée est incubée quelques minutes avant récupération d'un premier *flow-through*, abrégé FT1. FT1 est redéposé sur la matrice et un second *flow-through* (FT2) est récupéré afin de maximiser la fixation de la protéine de fusion à la matrice. La matrice est lavée au BB avec l'équivalant de cinq volumes de colonne (environ 15 mL). Cette opération est répétée pour un total de deux à trois lavages ; ils sont également récupérés pour analyse. Concernant l'élution de la

protéine marquée à la GST, 1.0 mL de tampon d'élution, abrégé EB (pour *elution buffer*), est déposé et incubé sur la colonne fermée pendant dix minutes avant récupération de l'éluat. Le tampon EB est de composition suivante : 50 mM Tris-HCl pH 8.0, 10 mM ou 20 mM GSH. Il est à noter que le tampon EB employé doit être fraîchement préparé le jour même de la purification étant donné que le GSH s'oxyde vite en GSSG en solution ; le temps de demi-vie du GSH dans l'eau est d'environ neuf heures à pH 7.5 et 20 °C.[178] L'opération d'élution est répétée pour un certain total d'échantillons dont le nombre est laissé au choix de l'expérimentateur.

Afin de régénérer la matrice, cette dernière est lavée par l'équivalant de trois volumes de colonne (environ 9 mL) au BB, à l'eau distillée et par l'équivalant de deux volumes de colonne (environ 6 mL) à l'éthanol 20 %. La colonne est stockée fermée à 4 °C avec 2 à 3 mL d'une solution d'éthanol 20 % en surface. Quand la matrice semble perdre de sa capacité de fixation, elle est lavée abondement à l'éthanol 70 %.

B. Purification au format *batch*

Le procédé de purification au format *batch* est réalisé dans un tube Falcon™ de 50 mL dans lequel sont versés 4 mL de suspension de GS4B dans l'éthanol 20 % afin d'obtenir environ 3 mL de matrice. Chaque tampon employé est additionné de PMSF 200 µM en respectant le temps de demi-vie de ce dernier en regard du pH du tampon. Le commentaire établi au point V.A de la Partie n°5 quant à la stabilité du GSH en solution aqueuse est également d'application.

Concernant la procédure générale de lavage au format *batch*, le tube Falcon™ est fermé, agité manuellement ou vortexé et centrifugé à 600 rpm pendant 3 minutes à température ambiante à l'aide d'une centrifugeuse VWR™ Mega Star 600 afin de tasser la matrice et de faciliter l'élimination du surnageant. Afin d'optimiser le tassement, la vitesse de rotation peut être augmentée jusqu'à 2000 rpm, le temps de centrifugation allongé jusqu'à 10 minutes et la température diminuée jusqu'à 4 °C. La matrice est équilibrée par deux fois 10 à 12 mL de tampon de lyse LB. Le lysat y est additionné et incubé pendant deux heures à 4 °C sous agitation. Le tube Falcon™ est récupéré et centrifugé. Le surnageant, autrement dit le FT, est conservé. La matrice est lavée par trois fois 25 mL de tampon BB et transférée dans une colonne chromatographique en de 12 mL de chez GoldBio. La protéine de fusion marquée à la GST est éluée sur colonne par fraction de 0.5 mL de tampon EB pour un certain total d'échantillons dont le nombre est laissé au choix de l'expérimentateur.

Afin de régénérer la matrice, cette dernière est lavée par trois volumes de matrice (environ 9 mL) d'une solution d'éthanol 20 %. La colonne est stockée fermée à 4°C avec 2 à 3 mL d'une solution d'éthanol 20 % en surface.

Le procédé de clivage au format *batch* décrit ci-après ne concerne que le clivage au facteur Xa des constructions de DPF3 choisies, à savoir DPF3a FL et PHD-1+2. Il suit les mêmes étapes d'équilibration et de fixation de la protéine de fusion décrites précédemment. La procédure de lavage reste aussi identique à l'exception qu'une centrifugation à 4000 rpm est préférée afin de minimiser d'autant plus les pertes de matrice lors du pipetage du surnageant par meilleur tassement. Chaque tampon employé est additionné de PMSF 200 µM en respectant le temps de demi-vie de ce dernier en regard du pH du tampon.

Après les deux heures d'incubation du lysat à 4 °C et récupération du FT, la matrice est lavée par deux fois 15 mL de tampon du facteur Xa, abrégé FB (pour *factor Xa buffer*), de composition suivante : 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$; les cations Ca^{2+} sont nécessaires à l'activité du facteur Xa. La solution de facteur Xa est préparée par addition de 20 µL d'une solution stock de facteur Xa 1 mg/mL dans 2980 µL de tampon FB et transférée sur la matrice délicatement resuspendue et incubée o/n à 22-25 °C sous agitation légère ; les 20 µg de facteur Xa ici employés sont déterminés pour cliver approximativement 1 mg de protéine étiquetée à la GST. Après clivage o/n, la matrice est récupérée et centrifugée à 2000 rpm pendant 5 minutes à température ambiante. Le surnageant, à savoir l'éluat contenant la protéine d'intérêt clivée, est collecté par fraction de 1.0 mL pour un premier total de quatre échantillons. La matrice est lavée par 4 mL de tampon FB et à nouveau centrifugée aux conditions sus-citées. Le surnageant est collecté par fraction de 1.0 mL pour quatre échantillons supplémentaires. Cette dernière étape est répétée une dernière fois pour un total de douze échantillons.

Afin de régénérer la matrice, cette dernière est lavée sur colonne par 10 mL d'eau distillée, 10 mL d'éthanol 70 % et 10 mL d'une solution d'éthanol 20 %. La colonne est stockée fermée à 4 °C avec 2 à 3 mL d'une solution d'éthanol 20 % en surface afin de conserver la matrice pour une prochaine purification au format *batch*.

C. Purification par FPLC sur colonnes pré-packées

Le procédé de purification par FPLC sur colonnes pré-packées est réalisé sur le système FPLC Äkta Purifier et une colonne de glutathion pré-packée GSTrap™ FF de 5 mL de chez GE Healthcare. Les chromatogrammes sont générés via le programme

Unicorn. À cause des contraintes temporelles dues au dégazage et passage des solutions tampons, le PMSF n'y est pas additionné à l'exception du tampon LB dans lequel se trouve le lysat. Il est considéré que toutes les protéases sont éliminées dans le FT. L'ensemble des tampons employés et solutions de lavage sont préalablement dégazés avant toute introduction dans le système. Le commentaire établi au point V.A de la Partie n°5 quant à la stabilité du GSH en solution aqueuse est également d'application.

Le système FPLC est couplé à une boucle d'injection de 5 mL et le tout est rincé à l'eau distillée à un débit de 3.0 mL/min jusqu'à stabilisation des signaux de conductivité et d'absorbance à 280 nm. La colonne GSTrap™ est couplée au système et drainée de sa solution d'éthanol 20 % de stockage par passage d'eau distillée à un débit de 3.0 mL/min. La colonne ainsi que le système sont équilibrés par passage de tampon BB à un débit de 1.0 mL/min jusqu'à observer un plateau de conductivité dû à la présence de KCl 200 mM dans le tampon BB.

Pendant l'étape d'équilibration, le lysat centrifugé est transféré dans un berlin et 120 µL d'une solution stock de PMSF 25 mM dans l'isopropanol sont additionnés afin d'en étendre le temps d'action d'approximativement une heure, ce qui correspond au temps de passage de la totalité du lysat sur la colonne. À l'aide d'une seringue, le lysat est chargé en trois fois 5 mL dans la boucle d'injection de par la limitation volumique imposée par ladite boucle. Le lysat est injecté sur la colonne à un débit de 0.5 à 1.0 mL/min jusqu'à l'apparition d'un FT se manifestant par une augmentation abrupte de l'absorbance à 280 nm et d'une variation de la conductivité. Les trois FT passés et rassemblés, la colonne est lavée par passage de tampon BB à un débit de 3.0 à 5.0 mL/min jusqu'à retrouver des signaux similaires à ceux de l'étape d'équilibration.

Concernant l'étape d'élution de la protéine de fusion marquée à la GST, cette dernière est décrochée de la colonne GSTrap™ par passage de tampon EB à un débit de 0.5 mL/min et récoltée par fraction de 1.0 mL. Après l'apparition du pic d'élution, cette dernière est passée à un débit de 3.0 mL/min pendant 15 à 30 mL jusqu'à ne plus observer quelque variation de signal.

La colonne, le système ainsi que la boucle sont nettoyés une première fois par passage d'eau distillée et une seconde fois par passage d'une solution d'éthanol 20 % à un débit de 3.0 mL/min jusqu'à stabilisation des signaux de conductivité et d'absorbance à 280 nm. Le système est laissé dans l'éthanol 20 %. La colonne GSTrap™ est stockée à 4 °C.

Le procédé de clivage par FPLC sur colonnes pré-packées décrit ci-après ne concerne que le clivage au facteur Xa des constructions de DPF3, à savoir DPF3a FL,

DPF3b FL, N-ter+C_2H_2, PHD-1, PHD-2 et PHD-1+2. Aussi, il suit les mêmes étapes d'équilibration et de fixation de la protéine de fusion décrites précédemment. Le clivage est réalisé sur un système de deux colonnes couplées en série : la colonne pré-packée GSTrap™ FF de 5 mL connectée à une colonne HiTrap™ Benzamidine FF de 5 mL de chez Ge Healthcare pour l'élimination des traces de facteur Xa.

Après passage des FT et lavage au tampon BB, la colonne GSTrap™ est lavée par passage de tampon FB à un débit de 3.0 mL/min jusqu'à observer une stabilisation des signaux relatifs à la conductivité et à l'absorbance à 280 nm. La solution de facteur Xa est préparée par addition de 20 µL d'une solution stock de facteur Xa 1 mg/mL dans 4980 µL de tampon FB ; les 20 µg de facteur Xa ici employés sont déterminés pour cliver approximativement 1 mg de protéine étiquetée à la GST. Cette solution est chargée en une seule fois dans la boucle d'injection et 6 mL sont injectés sur la colonne à un débit de 1.0 mL/min afin d'y assurer la présence de la protéase. La colonne GSTrap™ est retirée du système FPLC, fermée et incubée o/n à 22-25 °C. Avant la poursuite de la procédure de clivage, le système ainsi que la boucle sont entre-temps nettoyés à l'eau distillée ainsi qu'à l'éthanol 20 %.

Après clivage o/n, la colonne HiTrap™ Benzamidine est drainée de sa solution de stockage et rincée par passage d'eau distillée à un débit de 3.0 mL/min. Elle est équilibrée par passage de tampon FB à un débit de 3.0 mL/min jusqu'à observer un plateau de conductivité dû à la présence de NaCl 150 mM dans le tampon FB. Les colonnes sont connectées en série au système FPLC, à savoir la GSTrap™ au-dessus de la HiTrap™ Benzamidine. La protéine d'intérêt clivée est éluée par passage de tampon FB à un débit de 0.5 à 1.0 mL/min et récoltée par fraction de 1.0 mL. Après l'apparition du pic d'élution, cette dernière est passée à un débit de 2.0 à 3.0 mL/min pendant 15 à 30 mL jusqu'à ne plus observer quelque variation de signal.

La colonne HiTrap™ Benzamidine est déconnectée du système. Séparément, le *tag* GST clivé est élué de la colonne GSTrap™ par passage de tampon EB à un débit de 1.0 mL/min et récolté manuellement. Après l'apparition du pic d'élution, cette dernière est passée à un débit de 2.0 à 3.0 mL/min pendant 15 à 30 mL jusqu'à ne plus observer quelque variation de signal.

La colonne GSTrap™ est déconnectée du système et la colonne HiTrap™ Benzamidine y est à nouveau couplée. Séparément, le facteur Xa sous forme de traces est élué par passage de tampon d'élution du facteur Xa, abrégé FEB (pour *factor Xa elution buffer*), à un débit de 1.0 mL/min pour environ 30 à 50 mL ; le facteur Xa est souvent trop peu concentré pour observer quelque variation de signal et aucune fraction

n'est récoltée. Le tampon FEB est de composition suivante : glycine-HCl 50 mM pH 3.0. La colonne HiTrap™ est nettoyée par passage de tampon acétate de stockage, abrégé BSB (pour *benzamidine storage buffer*), à un débit de 1.0 mL/min jusqu'à stabilisation des signaux relatifs à la conductivité et à l'absorbance à 280 nm. Le tampon BSB est de composition suivante : $NaCOOCH_3$ 50 mM pH 4.0 dans l'éthanol 20 %. La colonne HiTrap™ Benzamidine est stockée fermée à 4 °C.

Le système FPLC est nettoyé par passage d'eau distillée à un débit de 3.0 mL/min jusqu'à l'élimination complète du tampon BSB résiduel dans les pompes. La colonne GSTrap™ est à nouveau couplée au système, nettoyée par passage d'eau distillée et d'une solution d'éthanol 20 % à un débit de 3.0 mL/min jusqu'à stabilisation des signaux. Le système est laissé dans l'éthanol 20 %. La colonne GSTrap™ est stockée fermée à 4 °C.

VI. Dessalage et échange de tampon sur colonne PD-10 d'une protéine purifiée

A partir d'une colonne de dessalage pré-packée PD-10 contenant 8.3 mL de matrice Sephadex™ G-25 M de chez GE Healthcare, la matrice est drainée de sa solution de stockage et équilibrée avec le tampon d'échange par une série de lavages successifs pour un total de 20 à 30 mL (correspondant à plus de deux volumes de colonne). Un maximum de 2.5 mL d'échantillon protéique purifié est déposé sur la colonne fermée. Si le volume d'échantillon disponible est inférieur à 2.5 mL, il est complété à 2.5 mL à l'aide du tampon d'échange. Le FT est récupéré. Sur la colonne, 3.5 mL de tampon d'échange sont déposés, expliquant la dilution de la protéine dans le nouveau tampon. L'éluat protéique dessalé est récupéré. En fin de procédure, la colonne est lavée avec environ 20 à 30 mL d'eau distillée et stockée fermée à 4 °C dans de l'eau distillée.

VII. Analyse par SDS-PAGE d'échantillons protéiques

Les gels de migration sont préparés dans des cassettes Novex Life Technologies™ de 1.0 mm en deux temps. Dans un premier temps, le *running* (ou *resolving*) *gel*, dont la solution est préparée par addition dans l'ordre suivant d'eau déminéralisée, d'une solution d'acrylamide-bisacrylamide, de Tris-HCl 1.5 M à pH 8.8, de dodécylsulfate de sodium (abrégé SDS pour *sodium dodecyl sulfate*) 10 %, de persulfate d'ammonium (abrégé APS pour *ammonium persulfate*) 10 % et de tétraméthyléthylènediamine (abrégée TEMED), est coulé et laissé polymériser pendant 30 minutes. Les proportions

en chacun des composants du gel sont déterminées par le degré de réticulation souhaitée. Dans un second temps, le *stacking gel*, dont la solution est préparée par addition dans l'ordre suivant d'eau déminéralisée, d'une solution d'acrylamide-bisacrylamide, de Tris-HCl 0.5 M à pH 6.8, de SDS 10 %, d'APS 10 % et de TEMED, est coulé et additionné d'un peigne pour le nombre de puits souhaité. Le gel est laissé polymériser pendant 30 minutes. Les gels sont stockés à 4 °C emballés dans du papier Tork™ humidifié.

La migration sur gel est réalisée dans une cuve d'électrophorèse Novex Life Technologies™ XCell SureLock™. Après avoir réalisé le montage de la cuve et placé la cassette du gel, la cuve est remplie en son centre à l'aide du tampon de migration de SDS concentré 1X. Les échantillons protéiques sont préparés par addition de 10 µL de tampon de charge à 10 µL d'échantillon. Le tampon de charge contient du SDS 10 % pour la dénaturation, du bleu de bromophénol 0.5 % pour la visualisation de la migration et de β-mercaptoéthanol pour la réduction des ponts disulfure. Le peigne est retiré de la cassette. Dans un puits, 8 µL de marqueur de masse moléculaire sont déposés. Dans les autres puits, 10 µL du mélange échantillon protéique-tampon de charge sont déposés. La cuve est connectée au générateur réglé sur 200 V et la migration est lancée. Elle est surveillée et dure généralement entre 50 minutes à 1h30.

Après migration, la cassette est récupérée, rincée à l'eau distillée et ouverte pour récupérer le gel. Le gel est rincé à l'eau distillée et coloré pendant 30 à 45 minutes dans une solution de bleu de Coomassie R-250 sur un agitateur orbital. Après coloration, le gel est transféré dans une solution de décoloration de composition suivante : méthanol 30 % et acide acétique glacial 10 % dans l'eau distillée. La décoloration prend place pendant une heure à o/n. Le gel est rincé abondamment à l'eau distillée et photographié sur une plaque blanche en PMMA rétroéclairée. Concernant la méthode de révélation à l'imidazole-zinc, le gel est immergé après migration et rinçage dans une solution d'imidazole 0.2 M pendant 5 à 10 minutes sur l'agitateur orbital. Le gel est transféré dans une solution de chlorure de zinc 0.3 M jusqu'au développement des bandes ; cette étape dure généralement entre 30 à 45 secondes. Le gel est rincé abondamment à l'eau distillée et photographié contre une plaque en PMMA noire.

VIII. Méthodes de caractérisation biophysiques

A. Spectrophotométrie UV-visible

Les mesures sont réalisées sur un spectrophotomètre UV-visible VWR™ UV-6300PC, collectées et enregistrées par l'intermédiaire du logiciel UV-Vis Analyst. Les

spectres d'absorption résultants sont obtenus en mode balayage de longueur d'onde dans une gamme choisie entre 200 et 800 nm par pas de 0.5 nm. Les cellules de mesure sont soit des UV-cuvettes Brand™ ultra-micro de chemin optique 10 mm soit une cuvette Hellma™ en quartz QS ultra-micro à haute précision de chemin optique 10 mm. La capacité minimale d'échantillonnage s'élève à 700 µL pour les UV-cuvettes et à 150 µL pour la cellule en quartz. La nature de la cuvette employée est précisée dans la légende de chaque spectre. L'ensemble des mesures d'absorption sont réalisées en simple faisceau. Les valeurs d'absorbance des blancs sont préalablement déterminées dans la gamme de longueurs d'onde d'étude avant analyse et sont soustraites soit automatiquement par auto-zéro de la machine soit manuellement a posteriori.

B. Fluorimétrie (ITF et dbAF)

Les mesures sont réalisées sur un spectrophotomètre de fluorescence Agilent Cary Eclipse, collectées et enregistrées par l'intermédiaire du logiciel Cary Eclipse. Les spectres d'émission résultants sont obtenus en mode balayage de longueur d'onde à partir de la longueur d'onde d'excitation jusque 600 nm par pas de 1.0 nm. Les filtres d'excitation et d'émission sont réglés en mode *Auto* et la largeur de fente (abrégée sw pour *slit width*) d'excitation et d'émission à 5 ou 10 nm en fonction de l'intensité des signaux d'émission. La valeur de sw est référencée dans la légende de chaque spectre. La cellule de mesure est une cuvette Hellma™ en quartz QS ultra-micro à haute précision pour fluorescence de chemin optique 10 mm. Les valeurs d'intensité de fluorescence des blancs sont également déterminées dans la gamme de longueurs d'onde d'étude à des fins comparatives.

C. Spectrométrie de masse (LC-ESI-MS/MS)

Concernant l'analyse par spectrométrie de masse sur gel, les bandes d'intérêt sont excisées et lavées une première fois à l'eau déminéralisée puis à l'acétonitrile. La digestion est opérée o/n à 37 °C par injection de 8 µL de trypsine pour chaque bande suspendue dans un tampon NH_4HCO_3 50 mM. La digestion complétée, les échantillons sont stockés à -20 °C avant analyse. L'analyse est réalisée par injection de 15 µL d'échantillon digéré sur un spectromètre de masse Bruker™ nano-LC-ESI-MS/MS maXis Impact UHR-TOF couplé à un Thermo™ 2D-LC Dionex Ultimate 3000. Les échantillons sont séparés par chromatographie liquide en phase inverse sur une colonne Thermo™ Acclaim PepMap 100 C18 75 µm x 250 mm dans un système chromatographique en phase liquide Ultimate 3000. La phase mobile A est constituée

d'un mélange eau 95 %, acétonitrile 5 % et acide formique 0.1 %. La phase mobile B est constituée d'un mélange eau 20 %, acétonitrile 80 % et acide formique 0.1 %. Un gradient d'élution linéaire est appliqué de sorte à augmenter la proportion en phase mobile B de 5 à 40 % en 25 minutes et de 40 à 100 % en 5 minutes. Les éluats sont connectés à un ionisateur Bruker™ Captive Spray. Les diagrammes sont générés à l'aide du logiciel Bruker™ DataAnalysis 4.2 et confrontés à la base de données UniProt à l'aide du logiciel Mascot™ 2.4. Le logiciel Scaffold 4.8 est employé pour valider les peptides et identifier les protéines d'intérêt.

Concernant l'analyse par spectrométrie de masse en phase liquide, les échantillons sont digérés par la procédure FASP (pour *filter-aided sample preparation*). Les étapes d'analyse après digestion sont identiques à celles détaillées précédemment. L'ensemble des analyses ont été réalisées par l'équipe de Marc Dieu de la Plateforme Technologique de Spectrométrie de Masse de l'Université de Namur (UNamur) MaSUN.

D. Diffusion dynamique de la lumière (DLS)

Les mesures sont réalisées sur un analyseur de nanoparticules Horiba Zetasizer SZ-100. Avant analyse, les échantillons sont filtrés sur un filtre 0.2 µm soit en cellulose recomposée (abrégée RC pour *recomposed cellulose*) soit en polyéthersulfone (abrégé PES) de par leur faible taux de rétention protéique. Si les échantillons présentent par ailleurs des particules visibles en suspension, ils sont préalablement centrifugés pendant 10 minutes à 12000 x g à 4 °C à l'aide d'une centrifugeuse VWR™ Mega Star 600 avant filtration. L'ensemble des analyses ont été réalisées par le Dr. Éric Perpète du Laboratoire de Chimie Physique des Biomolécules (CPB).

E. Dichroïsme circulaire (CD)

Les mesures sont réalisées sur un spectropolarimètre Jasco J-810 aux paramètres suivants : température de 20 °C, D.I.T. de 1 s, largeur de bande de 1.00 nm, vitesse de balayage à 50 nm/min. Les spectres résultants sont obtenus en mode balayage (répété quatre fois) dans la gamme de longueurs d'onde 260-195 nm par pas de 0.1 nm, moyennés et corrigés par rapport au spectre du blanc. Il n'y a pas eu de lissage de courbe. La cellule de mesure est une cuvette Hellma™ en quartz Suprasil de chemin optique 1 mm. Les données sont présentées selon la variation de l'ellipticité en mdeg par rapport à la longueur d'onde en nm. L'ensemble des analyses ont été réalisées au

laboratoire d'Enzymologie et Repliement des Protéines du Pr. André Matagne de l'ULiège.

F. Chromatographie d'exclusion stérique (SEC)

Le procédé de séparation de populations protéiques par filtration sur gel est réalisé sur le système FPLC Äkta Purifier et une colonne pré-packée d'agarose réticulé couplé au dextrane HiLoad™ 16/600 Superdex™ 200 prep grade de 120 mL de chez GE Healthcare. Les chromatogrammes sont générés via le programme Unicorn. L'ensemble des tampons employés et solutions de lavage sont préalablement dégazés avant toute introduction dans le système.

Concernant la procédure d'équilibration de la colonne de séparation, le système FPLC est couplé à une boucle d'injection de 2 mL et le tout est rincé à l'eau distillée à un débit de 10.0 mL/min jusqu'à stabilisation des signaux de conductivité et d'absorbance à 280 nm. La colonne Superdex™ est connectée au système et est drainée de sa solution d'éthanol 20 % de stockage par passage d'eau distillée à un débit de maximum 1.0 mL/min pour l'équivalant d'un volume de colonne, c'est-à-dire environ 120 mL. Il est crucial de ne pas excéder le débit maximal (1.0 mL/min) supporté par la colonne ; il en résulte une pression maximale de 0.30 MPa. La colonne est équilibrée avec le tampon de clivage FB qui, pour rappel, est de composition suivante : 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$. Le choix de ce tampon d'équilibration est motivé par la conservation d'une composition de solution identique à celle du tampon d'élution dans lequel l'ensemble des échantillons de la DPF3 clivée sont récoltés et stockés. L'équilibration est exécutée soit à un débit de 1 mL/min en quelques heures soit à un débit de 0.1 à 0.2 mL/min o/n pour l'équivalant d'un à deux volumes de colonne jusqu'à stabilisation des signaux de conductivité et d'absorbance à 280 nm. Dans le présent cas, l'équilibration est achevée dès lors qu'un palier de conductivité est atteint étant donné la présence de NaCl 150 mM dans le tampon FB.

Concernant l'établissement de la droite de calibration, le volume mort de la colonne est déterminé par injection d'une solution de bleu de dextrane de haute masse moléculaire (M = 1 MDa) 1 mg/mL à raison de 1.2 mL (correspondant à 1 % du volume de colonne) à un débit de 1.0 mL/min. La sortie du bleu de dextrane est suivie par variation de l'absorbance à 280 nm. Après le passage du bleu de dextrane, l'élution est continuée jusqu'à atteindre l'équivalant d'un volume de colonne pour stabiliser les signaux d'absorbance et de conductivité. La solution d'étalons protéiques est préparée par dissolution dans 0.5 mL d'eau distillée de 18 mg d'un mélange lyophilisé de

marqueurs de masse moléculaire Gel Filtration Standard de chez BioRad. Le mélange est contenu dans une fiole scellée, qui est agitée doucement et laissée reposer sur glace pendant 2 à 3 minutes. La solution est à nouveau mélangée doucement avant prélèvement. Les standards sont la thyroglobuline (bovin ; M = 670 kDa ; m = 5 mg), la γ-globuline (bovin ; M = 158 kDa ; m = 5 mg), l'ovalbumine (poule ; M = 44 kDa ; m = 5 mg), la myoglobine (cheval ; M = 17 kDa ; m = 2.5 mg) et la vitamine B12 (M = 1350 Da ; m = 0.5 mg). L'étalonnage est réalisé par injection de la solution d'étalons à raison de 0.5 mL à un débit de 1.0 mL/min. La sortie des standards est suivie par variation de l'absorbance à 280 nm. La droite de calibration est construite à partir du logM en fonction du volume d'élution de chaque étalon en tenant compte du volume mort de colonne.

Les échantillons purifiés sont injectés sur la colonne Superdex™ à raison d'environ 1.0 mL à un débit de 1.0 mL/min. La sortie des populations protéiques est suivie parallèlement par variation de l'absorbance à 254, 265 et 280 nm. L'élution est toujours poursuivie jusqu'à atteindre au minimum l'équivalant d'un volume de colonne. Si les échantillons présentent des particules en suspension, ils sont préalablement filtrés sur un filtre 0.2 µm en PES avant injection. En fin d'analyse, le système, la boucle ainsi que la colonne sont lavés par passage dans un premier temps d'eau distillée et dans un second temps d'une solution d'éthanol 20 % à un débit de 1.0 mL/min pour l'équivalant d'un à deux volumes de colonne. Le système est laissé dans l'éthanol 20 %. La colonne Superdex™ est stockée fermée dans l'éthanol 20 % à température ambiante.

Partie n°6

—

Bibliographie

1. Cooper, G. M. & Hausman, R. E. *The Cell: A Molecular Approach. Sinauer Associates, Inc.* (ASM Press, 2007).
2. Lew, D. J. *Cell Cycle. Brenner's Encyclopedia of Genetics: Second Edition* vol. 1 (Elsevier Inc., 2013).
3. Tan, E. P., Duncan, F. E. & Slawson, C. The sweet side of the cell cycle. *Biochem. Soc. Trans.* 45, 313–322 (2017).
4. Ohkura, H. Meiosis: An overview of key differences from mitosis. *Cold Spring Harb. Perspect. Biol.* 7, 1–15 (2015).
5. Ong, J. Y. & Torres, J. Z. Dissecting the mechanisms of cell division. *J. Biol. Chem.* 294, 11382–11390 (2019).
6. Rhind, N. & Russell, P. Signaling pathways that regulate cell division. *Cold Spring Harb. Perspect. Biol.* 4, 1–15 (2012).
7. Fisher, D., Krasinska, L., Coudreuse, D. & Novák, B. Phosphorylation network dynamics in the control of cell cycle transitions. *J. Cell Sci.* 125, 4703–4711 (2012).
8. Chen, R., Kang, R., Fan, X. G. & Tang, D. Release and activity of histone in diseases. *Cell Death Dis.* 5, 1–9 (2014).
9. Mcintosh, J. R. Mitosis. *Cold Spring Harb. Perspect. Biol.* 1–16 (2016).
10. Rossetto, D., Avvakumov, N. & Côté, J. Histone phosphorylation. *Epigenetics* 7, 1098–1108 (2012).
11. Zhiteneva, A. *et al.* Mitotic post-translational modifications of histones promote chromatin compaction in vitro. *Open Biol.* 7, (2017).
12. Gräff, J. & Tsai, L. H. Histone acetylation: Molecular mnemonics on the chromatin. *Nat. Rev. Neurosci.* 14, 97–111 (2013).
13. Annunziato, A. T. & Hansen, J. C. Role of histone acetylation in the assembly and modulation of chromatin structures. *Gene Expr.* 9, 37–61 (2000).
14. Nelissen, H., Boccardi, T. M., Himanen, K. & Van Lijsebettens, M. Impact of core histone modifications on transcriptional regulation and plant growth. *CRC. Crit. Rev. Plant Sci.* 26, 243–263 (2007).
15. Walczak, C. E., Cai, S. & Khodjakov, A. Mechanisms of chromosome behaviour during mitosis. *Nat. Rev. Mol. Cell Biol.* 11, 91–102 (2010).
16. Makrantoni, V. & Marston, A. L. Cohesin and chromosome segregation. *Curr. Biol.* 28, R688–R693 (2018).
17. Giménez-Abián, J. F., Díaz-Martínez, L. A., Wirth, K. G., De La Torre, C. & Clarke, D. J. Proteasome activity is required for centromere separation independently of securin degradation in human cells. *Cell Cycle* 4, 1558–1560 (2005).
18. Dominguez-Brauer, C. *et al.* Targeting Mitosis in Cancer: Emerging Strategies. *Mol. Cell* 60, 524–536 (2015).
19. Barnum, K. J. & O'Connell, M. J. Cell cycle regulation by checkpoints. *Methods Mol. Biol.* 1170, 29–40 (2014).

20. Vermeulen, K., Berneman, Z. N. & Van Bockstaele, D. R. Cell cycle and apoptosis. *Cell Prolif.* 36, 165–175 (2003).
21. Sesso, A. *et al.* Structural elements common to mitosis and apoptosis. *Tissue Cell* 31, 357–371 (1999).
22. Kroemer, G. & Galluzzi, L. Classification of cell death. *M Piac.* 16, 3–11 (2009).
23. Castedo, M. *et al.* Cell death by mitotic catastrophe: A molecular definition. *Oncogene* 23, 2825–2837 (2004).
24. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* 144, 646–674 (2011).
25. Levine, M. S. & Holland, A. J. The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev.* 32, 620–638 (2018).
26. Gascoigne, K. E. & Taylor, S. S. How do anti-mitotic drugs kill cancer cells? *J. Cell Sci.* 122, 2579–2585 (2009).
27. Van Vuuren, R. J., Visagie, M. H., Theron, A. E. & Joubert, A. M. Antimitotic drugs in the treatment of cancer. *Cancer Chemother. Pharmacol.* 76, 1101–1112 (2015).
28. L. Garner, A. & D. Janda, K. Protein-Protein Interactions and Cancer: Targeting the Central Dogma. *Curr. Top. Med. Chem.* 11, 258–280 (2010).
29. Kar, G., Gursoy, A. & Keskin, O. Human cancer protein-protein interaction network: A structural perspective. *PLoS Comput. Biol.* 5, (2009).
30. de Las Rivas, J. & Fontanillo, C. Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* 6, 1–8 (2010).
31. Lu, H. *et al.* Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduct. Target. Ther.* 5, (2020).
32. Garland, W., Benezra, R. & Chaudhary, J. *Targeting protein-protein interactions to treat cancer-recent progress and future directions. Annual Reports in Medicinal Chemistry* vol. 48 (Elsevier Inc., 2013).
33. Llamazares, S. *et al.* polo encodes a protein kinase homolog required for mitosis in Drosophila. *Genes Dev.* 5, 2153–2165 (1991).
34. Lowery, D. M., Lim, D. & Yaffe, M. B. Structure and function of Polo-like kinases. *Oncogene* 24, 248–259 (2005).
35. Okamura, E., Sakamoto, T., Sasaki, T. & Matsunaga, S. A plant ancestral polo-like kinase sheds light on the mystery of the evolutionary disappearance of polo-like kinases in the plant kingdom. *Cytologia (Tokyo).* 82, 261–266 (2017).
36. Lee, S.-Y., Jang, C. & Lee, K.-A. Polo-Like Kinases (Plks), a Key Regulator of Cell Cycle and New Potential Target for Cancer Therapy. *Dev. Reprod.* 18, 65–71 (2014).
37. Nakojima, H., Toyoshima-Morimoto, F., Taniguchi, E. & Nishida, E. Identification of a consensus motif for PlK (Polo-like kinase) phosphorylation

reveals Myt1 as a Plk1 substrate. *J. Biol. Chem.* 278, 25277–25280 (2003).
38. Liu, Z., Sun, Q. & Wang, X. PLK1, A potential target for cancer therapy. *Transl. Oncol.* 10, 22–32 (2017).
39. UniProt. UniProtKB, Protein knowlegdebase. https://www.uniprot.org/ (2020).
40. Schmucker, S. & Sumara, I. Molecular dynamics of PLK1 during mitosis. *Mol. Cell. Oncol.* 1, 1–9 (2014).
41. García-Álvarez, B., De Cárcer, G., Ibañez, S., Bragado-Nilsson, E. & Montoya, G. Molecular and structural basis of polo-like kinase 1 substrate recognition: Implications in centrosomal localization. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3107–3112 (2007).
42. Weiß, L. & Efferth, T. Polo-like kinase 1 as target for cancer therapy. *Exp. Hematol. Oncol.* 1, 1–6 (2012).
43. Qin, T. *et al.* Discovery of Novel Polo-Like Kinase 1 Polo-Box Domain Inhibitors to Induce Mitotic Arrest in Tumor Cells. *J. Med. Chem.* 59, 7089–7096 (2016).
44. Yun, T., Qin, T., Liu, Y. & Lai, L. Identification of acylthiourea derivatives as potent Plk1 PBD inhibitors. *Eur. J. Med. Chem.* 124, 229–236 (2016).
45. Ninkina, N. N. *et al.* Cerd4, third member of the d4 gene family: Expression and organization of genomic locus. *Mamm. Genome* 12, 862–866 (2001).
46. Chestkov, A. V., Baka, I. D., Kost, M. V., Georgiev, G. P. & Buchman, V. L. The d4 gene family in the human genome. *Genomics* 36, 174–177 (1996).
47. Buchman, V. L. *et al.* Differential splicing creates a diversity of transcripts from a neurospecific developmentally regulated gene encoding a protein with new zinc-finger motifs. *Nucleic Acids Res.* 20, 5579–5585 (1992).
48. Aasland, R., Gibson, T. J. & Stewart, A. F. The PHD finger: Implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* 20, 56–59 (1995).
49. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* 31, 35–40 (2006).
50. Boamah, D. *et al.* Characteristics of a PHD Finger Subtype. *Biochemistry* 57, 525–539 (2018).
51. Sanchez, R. & Zhou, M. M. The PHD finger: A versatile epigenome reader. *Trends Biochem. Sci.* 36, 364–372 (2011).
52. Xiong, X. *et al.* Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. *Nat. Chem. Biol.* 12, 1111–1118 (2016).
53. Schindler, U., Beckmann, H. & Cashmore, A. R. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J.* 4, 137–150 (1993).
54. Gabig, T. G. *et al.* Expression and chromosomal localization of the Requiem gene. *Mamm. Genome* 9, 660–665 (1998).
55. Mertsalov, I. B. *et al.* Generation of mutant mice with targeted disruption of two members of the d4 gene family: Neuro-d4 and cer-d4. *Dokl. Biochem. Biophys.*

419, 65–68 (2008).
56. Kulikova, D. A., Mertsalov, I. B. & Simonova, O. B. d4 family genes: Genomic organization and expression. *Russ. J. Dev. Biol.* 44, 1–6 (2013).
57. Lange, M. *et al.* Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev.* 22, 2370–2384 (2008).
58. Li, W., Zhao, A., Tempel, W., Loppnau, P. & Liu, Y. Crystal structure of DPF3b in complex with an acetylated histone peptide. *J. Struct. Biol.* 195, 365–372 (2016).
59. Son, E. Y. & Crabtree, G. R. The role of BAF (mSWI/SNF) complexes in mammalian neural development. *Am. J. Med. Genet. Part C Semin. Med. Genet.* 166, 333–349 (2014).
60. Panamarova, M. *et al.* The BAF chromatin remodelling complex is an epigenetic regulator of lineage specification in the early mouse embryo. *Dev.* 143, 1271–1283 (2016).
61. McNally, J. G. Transcription, chromatin condensation, and gene migration. *J. Cell Biol.* 185, 7–9 (2009).
62. Hohmann, A. F. & Vakoc, C. R. A rationale to target the SWI/SNF complex for cancer therapy. *Trends Genet.* 30, 356–363 (2014).
63. Ishizaka, A. *et al.* Double plant homeodomain (PHD) finger proteins DPF3a and -3b are required as transcriptional co-activators in SWI/SNF complex-dependent activation of NF-κB RelA/p50 heterodimer. *J. Biol. Chem.* 287, 11924–11933 (2012).
64. Zeng, L. *et al.* Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466, 258–262 (2010).
65. Zhu, X. *et al.* HRP2-DPF3a-BAF complex coordinates histone modification and chromatin remodeling to regulate myogenic gene transcription. *Nucleic Acids Res.* 48, 6563–6582 (2020).
66. Cui, H. *et al.* Phosphorylation of the chromatin remodeling factor DPF3a induces cardiac hypertrophy through releasing HEY repressors from DNA. *Nucleic Acids Res.* 44, 2538–2553 (2015).
67. Fu, K. *et al.* A temporal transcriptome and methylome in human embryonic stem cell-derived cardiomyocytes identifies novel regulators of early cardiac development. *Epigenetics* 13, 1013–1026 (2018).
68. Kosova, G. *et al.* Variants in DPF3 and DSCAML1 are associated with sperm morphology. *J. Assist. Reprod. Genet.* 31, 131–137 (2014).
69. Pandey, A. *et al.* The dynamics of gene expression during and post meiosis sets the sperm agenda. *Mol. Reprod. Dev.* 86, 1921–1939 (2019).
70. Liu, S. Y. *et al.* Strong association of SLC1A1 and DPF3 gene variants with idiopathic male infertility in Han Chinese. *Asian J. Androl.* 18, 486–492 (2016).
71. Sato, Y. *et al.* Association of TUSC1 and DPF3 gene polymorphisms with male

infertility. *J. Assist. Reprod. Genet.* 35, 257–263 (2018).

72. Liu, H. *et al.* Expression profiles of HA117 and its neighboring gene DPF3 in different colon segments of Hirschsprung's disease. *Int. J. Clin. Exp. Pathol.* 7, 3966–3974 (2014).
73. Vieira, A. R., McHenry, T. G., Daack-Hirsch, S., Murray, J. C. & Marazita, M. L. Candidate gene/loci studies in cleft lip/palate and dental anomalies finds novel susceptibility genes for clefts. *Genet. Med.* 10, 668–674 (2008).
74. Theodorou, M. *et al.* Identification of a STAT5 Target Gene, Dpf3, Provides Novel Insights in Chronic Lymphocytic Leukemia. *PLoS One* 8, 2–11 (2013).
75. Hoyal, C. R. *et al.* Genetic polymorphisms in DPF3 associated with risk of breast cancer and lymph node metastases. *J. Carcinog.* 4, 1–9 (2005).
76. Lin, W. hao *et al.* Downregulation of DPF3 promotes the proliferation and motility of breast cancer cells through activating JAK2/STAT3 signaling. *Biochem. Biophys. Res. Commun.* 514, 639–644 (2019).
77. Hiramatsu, H. *et al.* The role of the SWI/SNF chromatin remodeling complex in maintaining the stemness of glioma initiating cells. *Sci. Rep.* 7, 1–13 (2017).
78. Wu, Y. *et al.* Novel Pheretima guillelmi-derived antithrombotic protein DPf3: Identification, characterization, in vitro evaluation and antithrombotic mechanisms investigation. *Int. J. Biol. Macromol.* 154, 545–556 (2020).
79. Deiana, A., Forcelloni, S., Porrello, A. & Giansanti, A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One* 14, 1–16 (2019).
80. Zhao, B., Katuwawala, A., Uversky, V. N. & Kurgan, L. IDPology of the living cell: intrinsic disorder in the subcellular compartments of the human cell. *Cell. Mol. Life Sci.* (2020) doi:10.1007/s00018-020-03654-0.
81. Tompa, P. *Structure and Function of Intrinsically Disordered Proteins*. (2009).
82. Lieutaud, P. *et al.* 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, 1–33 (2016).
83. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.* 41, 415–427 (2000).
84. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331 (1999).
85. Uversky, V. N. Intrinsically disordered proteins and their 'Mysterious' (meta)physics. *Front. Phys.* 7, 8–23 (2019).
86. Dunker, A. K. & Obradovic, Z. The protein trinity - Linking function and disorder. *Nat. Biotechnol.* 19, 805–806 (2001).
87. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59 (2001).
88. Uversky, V. N. Natively unfolded proteins: A point where biology waits for

physics. *Protein Sci.* 11, 739–756 (2002).
89. Deforte, S. & Uversky, V. N. Intrinsically disordered proteins in PubMed: What can the tip of the iceberg tell us about what lies below? *RSC Adv.* 6, 11513–11521 (2016).
90. Wang, J., Cao, Z., Zhao, L. & Li, S. Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). *Int. J. Mol. Sci.* 12, 3205–3219 (2011).
91. Liu, Z. & Huang, Y. Advantages of proteins being disordered. *Protein Sci.* 23, 539–550 (2014).
92. Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19, 23–30 (2009).
93. Uversky, V. N. Chapter 7 - Biophysical Methods to Investigate Intrinsically Disordered Proteins: Avoiding an "Elephant and Blind Men" Situation. in *Intrinsically Disordered Proteins Studied by NMR Spectroscopy* 215–260 (Springer, 2015). doi:10.1007/978-3-319-20164-1.
94. Uversky, V. N. Dancing protein clouds: The strange biology and chaotic physics of intrinsically disordered proteins. *J. Biol. Chem.* 291, 6681–6688 (2016).
95. Uversky, V. N. Unreported intrinsic disorder in proteins: Disorder emergency room. *Intrinsically Disord. Proteins* 3, (2015).
96. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208 (2005).
97. Wright, P. E. & Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29 (2015).
98. Metallo, S. J. Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14, 481–488 (2010).
99. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z. & Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323, 573–584 (2002).
100. Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K. & Uversky, V. N. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* 45, 10448–10460 (2006).
101. Yu, C. *et al.* Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc. *Sci. Rep.* 6, 1–11 (2016).
102. Uversky, V. N. Intrinsically disordered proteins and novel strategies for drug discovery. *Expert Opin. Drug Discov.* 7, 475–488 (2012).
103. GE Healthcare. *GST Gene Fusion System - Handbook*. (2014).
104. Volkman, H. Affinity purification of GST fusion proteins (Protocol). *Trends Biochem. Sci.* 21, 400–401 (1996).
105. Harper, S. & Speicher, D. W. Purification of proteins fused to glutathione S-tranferase. *Methods Mol. Biol.* 681, 151–175 (2011).
106. Amersham Biotech. *Efficient, rapid protein purification and on-column cleavage*

using GSTrap FF columns.

107. Suskiewicz, M. J., Sussman, J. L., Silman, I. & Shaul, Y. Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Sci.* 20, 1285–1297 (2011).
108. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4, 2411–2423 (1995).
109. Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182, 319–326 (1989).
110. Ghisaidoobe, A. B. T. & Chung, S. J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on förster resonance energy transfer techniques. *Int. J. Mol. Sci.* 15, 22518–22538 (2014).
111. Vivian, J. T. & Callis, P. R. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.* 80, 2093–2109 (2001).
112. Eftink, M. R. Intrinsic Fluorescence of Proteins. *Top. Fluoresc. Spectrosc. Protein Fluoresc.* 6, 1–15 (2000).
113. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1, 2876–2890 (2007).
114. Miles, A. J. & Wallace, B. A. Circular Dichroism Spectroscopy for Protein Characterization: Biopharmaceutical Applications. *Biophys. Charact. Proteins Dev. Biopharm.* 109–137 (2015) doi:10.1016/B978-0-444-59573-7.00006-3.
115. Veronica Isabel Dodero, Zulma Beatriz Quirolo, M. A. S. Biomolecular studies by circular dichroism. *Front. Biosci.* 16, 61–73 (2011).
116. Rufer, A. C., Thiebach, L., Baer, K., Klein, H. W. & Hennig, M. X-ray structure of glutathione S-transferase from Schistosoma japonicum in a new crystal form reveals flexibility of the substrate-binding site. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 61, 263–265 (2005).
117. Kachel, W. B. Applications of the GST- Affinity Tag in the Purification and Characterization of Proteins. *Theses Diss.* 1625, (2016).
118. Lorber, B., Fischer, F., Bailly, M., Roy, H. & Kern, D. Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochem. Mol. Biol. Educ.* 40, 372–382 (2012).
119. Tcherkasskaya, O. & Uversky, V. N. Denatured collapsed states in protein folding: Example of apomyoglobin. *Proteins Struct. Funct. Genet.* 44, 244–254 (2001).
120. Song, J. Why do proteins aggregate? 'Intrinsically insoluble proteins'? And 'dark mediators'? Revealed by studies on 'insoluble proteins'? Solubilized in pure water. *F1000Research* 2, (2013).
121. Uversky, V. N. Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. *Intrinsically Disord. Protein Anal. Vol. 2, Methods Exp. Tools* 896, 179–194 (2012).

122. Juhasova, A., Baliova, M. & Jursky, F. A Dynamic Interaction of Coomassie Dye with the Glycine Transporters N-termini. *Protein J.* 35, 371–378 (2016).
123. Schägger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368–379 (1987).
124. Rodriguez, J., Gupta, N., Smith, R. D. & Pevzner, P. A. Does Trypsin Cut Before Proline? *J. Proteome Res.* 7, 300–305 (2008).
125. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–608 (2005) doi:10.1385/1592598900.
126. Gallagher, S. Quantitation of Nucleic Acids with Absorption Spectroscopy. *Curr. Protoc. Protein Sci.* A.4K.1-A.4K.3 (1998) doi:10.1002/0471140864.psa04ks13.
127. Pountney, D. L., Schauwecker, I., Zarn, J. & Vašák, M. Formation of Mammalian Cu8-Metallothionein in Vitro: Evidence for the Existence of Two Cu(I)4-Thiolate Clusters. *Biochemistry* 33, 9699–9705 (1994).
128. Grzyska, P. K., Hausinger, R. P. & Proshlyakov, D. A. Metal and substrate binding to an Fe(II) dioxygenase resolved by UV spectroscopy with global regression analysis. *Anal. Biochem.* 399, 64–71 (2010).
129. Tarasava, K., Loebus, J. & Freisinger, E. Localization and spectroscopic analysis of the Cu(I) binding site in wheat metallothionein Ec-1. *Int. J. Mol. Sci.* 17, 1–15 (2016).
130. Kishi, K., Yasuda, T. & Takeshita, H. DNase I: Structure, function, and use in medicine and forensic science. *Leg. Med.* 3, 69–83 (2001).
131. Vagenende, V., Yap, M. G. S. & Trout, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry* 48, 11084–11096 (2009).
132. Xu, H., Zhu, T. & Yu, R.-P. Study on the Absorption and Fluorescence Spectra of Ethylene Glycol and Glycerol. *Spectrosc. Spectr. Anal.* 27, 1381–1384 (2007).
133. Chan, F. T. S. *et al.* Protein amyloids develop an intrinsic fluorescence signature during aggregation. *Analyst* 138, 2156–2162 (2013).
134. Sirangelo, I., Borriello, M., Irace, G. & Ianuzzi, C. Intrinsic blue-green fluorescence in amyloyd fibrils. *AIMS Biophys.* 5, 155–165 (2018).
135. Niyangoda, C., Miti, T., Breydo, L., Uversky, V. & Muschol, M. Carbonyl-based blue autofluorescence of proteins and amino acids. *PLoS One* 12, 1–15 (2017).
136. Manning, M. C., Illangasekare, M. & W. Woody, R. Circular dichroism studies of distorted α-helices, twisted β-sheets, and β-turns. *Biophys. Chem.* 31, 77–86 (1988).
137. Iyer, A. *et al.* C-Terminal Truncated α-Synuclein Fibrils Contain Strongly Twisted β-Sheets. *J. Am. Chem. Soc.* 139, 15392–15400 (2017).
138. Lara, C. *et al.* ILQINS hexapeptide, identified in lysozyme left-handed helical ribbons and nanotubes, forms right-handed helical ribbons and crystals. *J. Am. Chem. Soc.* 136, 4732–4739 (2014).

139. Calzolai, L. Measuring Batch Mode DLS. 1–12 (2016).
140. Saiki, M., Shiba, K. & Okumura, M. Structural stability of amyloid fibrils depends on the existence of the peripheral sequence near the core cross-β region. *FEBS Lett.* 589, 3541–3547 (2015).
141. Rosenheck, K. & Doty, P. The Far Ultraviolet Absorption Spectra of Polypeptide And Protein Solutions And Thier Dependence On Conformation. *Proc. Natl. Acad. Sci. U. S. A.* 47, 1775–1785 (1961).
142. Pedersen, J. T. *et al.* Rapid formation of a preoligomeric peptide-metal-peptide complex following copper(II) binding to amyloid β peptides. *Angew. Chemie - Int. Ed.* 50, 2532–2535 (2011).
143. Borghesani, V., Alies, B. & Hureau, C. CuII Binding to Various Forms of Amyloid-β Peptides: Are They Friends or Foes? *Eur. J. Inorg. Chem.* 2 (2018) doi:10.1002/ejic.201701374.
144. Cragnell, C., Staby, L., Lenton, S., Kragelund, B. B. & Skepö, M. Dynamical oligomerisation of histidine rich intrinsically disordered proteins is regulated through zinc-histidine interactions. *Biomolecules* 9, (2019).
145. Fasman, G. D., Bodenheimer, E. & Pesce, A. Fluorescence studies on poly-alpha-amino acids: models of protein conformation. *J. Biol. Chem.* 241, 916–923 (1966).
146. Callis, P. R. & Liu, T. Quantitative Prediction of Fluorescence Quantum Yields for Tryptophan in Proteins. *J. Phys. Chem. B* 108, 4248–4259 (2004).
147. Khanna, N. C., Tokuda, M. & Waisman, D. M. Conformational changes induced by binding of divalent cations to calregulin. *J. Biol. Chem.* 261, 8883–8887 (1986).
148. Chen, J., Callis, P. R. & King, J. Mechanism of the very efficient quenching of tryptophan fluorescence in human γD- and γS-crystallins: The γ-crystallin fold may have evolved to protect tryptophan residues from ultraviolet photodamage. *Biochemistry* vol. 48 3708–3716 (2009).
149. Uversky, V. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord. proteins* 1, e24684 (2013).
150. Hwang, P. M. *et al.* Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13560–13565 (2002).
151. Huang, B. X., Kim, H. Y. & Dass, C. Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* 15, 1237–1247 (2004).
152. Knubovets, T., Osterhout, J. J., Connolly, P. J. & Klibanov, A. M. Structure, thermostability, and conformational flexibility of hen egg-white lysozyme dissolved in glycerol. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1262–1267 (1999).
153. Wu, J. *et al.* Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell* 175, 1393-1404.e11 (2018).
154. Canagarajah, B. J., Khokhlatchev, A., Cobb, M. H. & Goldsmith, E. J.

Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* 90, 859–869 (1997).

155. Bayrhuber, M. *et al.* Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15370–15375 (2008).
156. Feng, L. *et al.* Structure of oxidized α-haemoglobin bound to AHSP reveals a protective mechanism for haem. *Nature* vol. 435 697–701 (2005).
157. Lucas, H. & Fernández, R. Navigating the dynamic landscape of alpha-synuclein morphology: A review of the physiologically relevant tetrameric conformation. *Neural Regen. Res.* 15, 407–415 (2020).
158. Kurzbach, D. *et al.* Cooperative unfolding of compact conformations of the intrinsically disordered protein osteopontin. *Biochemistry* 52, 5167–5175 (2013).
159. Berk, J. M. *et al.* The molecular basis of emerin-emerin and emerin-BAF interactions. *J. Cell Sci.* 127, 3956–3969 (2014).
160. Santofimia-Castaño, P. *et al.* Intrinsically disordered chromatin protein NUPR1 binds to the C-terminal region of polycomb RING1B. *Proc. Natl. Acad. Sci. U. S. A.* 114, E6332–E6341 (2017).
161. Uversky, V. N. p53 proteoforms and intrinsic disorder: An illustration of the protein structure–function continuum concept. *Int. J. Mol. Sci.* 17, (2016).
162. Forsova, O. S. & Zakharov, V. V. High-order oligomers of intrinsically disordered brain proteins BASP1 and GAP-43 preserve the structural disorder. *FEBS J.* 283, 1550–1569 (2016).
163. Weathers, E. A., Paulaitis, M. E., Woolf, T. B. & Hoh, J. H. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett.* 576, 348–352 (2004).
164. Simm, S., Einloft, J., Mirus, O. & Schleiff, E. 50 years of amino acid hydrophobicity scales: Revisiting the capacity for peptide classification. *Biol. Res.* 49, 1–19 (2016).
165. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132 (1982).
166. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* 110, 13392–13397 (2013).
167. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32, 102–112 (2015).
168. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* 112, 16–21 (2017).
169. Gijsbers, A., Nishigaki, T. & Sánchez-Puig, N. Fluorescence anisotropy as a tool to study protein-protein interactions. *J. Vis. Exp.* 2016, 1–9 (2016).
170. Gell, D. A., Grant, R. P. & MacKay, J. P. The detection and quantitation of

protein oligomerization. *Adv. Exp. Med. Biol.* 747, 19–41 (2012).
171. Sibille, N. & Bernadó, P. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.* 40, 956–962 (2012).
172. Bernadó, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* 8, 151–167 (2012).
173. CurcÓ, D. *et al.* Stochastic simulation of structural properties of natively unfolded and denatured proteins. *J. Mol. Model.* 18, 4503–4516 (2012).
174. Takai, E. *et al.* Scanning electron microscope imaging of amyloid fibrils. *Am. J. Biochem. Biotechnol.* 10, 31–39 (2014).
175. Randrianjatovo-Gbalou, I., Marcato-Romain, C. E. & Girbal-Neuhauser, E. Quantification of amyloid fibrils using size exclusion chromatography coupled with online fluorescence and ultraviolet detection. *Anal. Biochem.* 488, 19–21 (2015).
176. Parija, S. C. *Textbook of Microbiology & Immunology.* (Elsevier Ltd, 2012).
177. James, G. T. Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers. *Anal. Biochem.* 86, 574–579 (1978).
178. Stevens, R., Stevens, L. & Price, N. C. The Stabilities of Various Thiol Compounds used in Protein Purifications. *Biochem. Educ.* 11, (1983).