



HAL
open science

Analyse structurale des régions prédites comme dépliées de l'enveloppe nucléaire : exemple de l'émerine et de la lamine A.

Florian Celli

► **To cite this version:**

Florian Celli. Analyse structurale des régions prédites comme dépliées de l'enveloppe nucléaire : exemple de l'émerine et de la lamine A.. Biophysique [physics.bio-ph]. Université Paris-Saclay, 2018. Français. NNT : 2018SACLS405 . tel-02060756

HAL Id: tel-02060756

<https://theses.hal.science/tel-02060756>

Submitted on 7 Mar 2019

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

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

Analyse structurale des régions prédites comme dépliées de l'enveloppe nucléaire : exemples de l'émerine et de la lamine A.

Thèse de doctorat de l'Université Paris-Saclay

préparée à l'Université Paris-Sud

École doctorale n°569 :

Innovation thérapeutique : du fondamental à l'appliqué (ITFA)

Spécialité de doctorat : Biochimie et biologie structurale

Thèse présentée et soutenue au CEA de Saclay, le 23/11/2018, par

Florian CELLI

Composition du Jury :

Herman Van Tilbeurgh , Professeur, I2BC	Président
Bruno Kieffer , Professeur, IGBMC	Rapporteur
Karima Djabali , Professeur, TUM	Rapporteur
Nathalie Sibille , Docteur, CBS	Examineur
Sophie Zinn-Justin , Docteur, I2BC	Directeur de thèse
Javier Pérez , Docteur, Synchrotron-SOLEIL	Co-Directeur de thèse

« Dans la vie rien n'est à craindre, tout est à comprendre ». Marie Curie.

A mes grands-pères Rivaldo et André, qui sont pour moi une source constante d'inspiration.

Remerciements

Le travail présenté dans cette thèse fut réalisé au sein de deux laboratoires : le Laboratoire de Biologie Structurale et Radiobiologie (LBSR) du CEA de Saclay ainsi qu'au Synchrotron SOLEIL.

Je souhaiterais tout d'abord adresser mes sincères remerciements au Dr. Sophie Zinn-Justin, ma directrice de thèse. Merci de m'avoir permis de réaliser cette thèse à vos côtés. Merci également pour vos constants encouragements, votre enthousiasme et votre motivation tout au long de ce projet. Enfin, merci pour tous vos conseils tant sur un plan professionnel que personnel. Ce fut un réel plaisir de travailler et de discuter avec vous.

Un immense merci au Dr. Javier Perez, pour avoir accepté de co-encadrer cette thèse et m'avoir permis de réaliser une partie de mon travail sur la ligne SWING dans un environnement aussi stimulant. Merci beaucoup pour ton temps, tes conseils et ta motivation. Merci également de m'avoir permis de travailler en collaboration avec des personnes aussi intéressantes et enrichissantes que Mattia ou Fabio.

Un grand merci au Dr. Philipp Selenko, grâce à qui j'ai eu le privilège de passer 9 mois à cheval sur mon stage de M2 et le début de ma thèse au Leibniz Institute für Molekulare Pharmakologie de Berlin. Cette expérience fut très enrichissante et m'a permis d'en apprendre beaucoup sur les derniers développements en matière de RMN.

Vorrei ringraziare il Dr. Mattia Rocco. Mattia grazie mille per tutto, mi ha fatto molto piacere lavorare con te anche fino a tarda notte. Mi piacerebbe rimanere in contatto con te in futuro. Grazie anche per il tuo aiuto nel trovare il mio nuovo lavoro.

Merci également à Youssef Liatimi. Merci Youssef pour ton aide précieuse, ton temps, ta bonne humeur et tes compétences en électronique lors de mes venues à SOLEIL.

Grazie Mille anche al Dr. Fabio Ferri per avermi dato il benvenuto nel suo laboratorio di Como. Grazie Fabio per il tuo aiuto.

Je souhaiterais également remercier le Dr. François-Xavier Theillet pour son aide et ses précieux conseils notamment concernant la RMN.

Thanks also to Dr. Ana-Andreaa Arteni for her help in analyzing my electron microscopy experiments. This was a pleasure to work with you.

Je souhaiterais adresser un remerciement spécial au Prof. Bruno Kieffer, pour avoir accepté de faire partie de mon comité de suivi de thèse, pour ses précieux conseils au cours

des différentes formations que j'ai pu faire et pour les discussions que nous avons pu avoir concernant mon projet de thèse.

J'adresse mes sincères remerciements aux Prof. Karima Djabali et Prof. Bruno Kieffer mais également au Dr. Nathalie Sibille pour avoir accepté de faire partie de mon jury de thèse.

Aus Berlin, möchte ich Maximillian Zinke bedanken. Danke Max, für deine gute Laune und für diese lustigen Momente.

Enfin, je souhaiterais adresser des remerciements plus particuliers aux personnes avec qui j'ai travaillé au quotidien au CEA de Saclay et qui m'ont soutenu tout au long de ma thèse malgré les moments parfois difficiles. Je souhaite remercier Virginie, Jingqi, Arun, Seydou, Ambre, Simona, Dani, Manon, Rania, Chafiaa et tous les autres... Merci pour votre bonne humeur, les soirées que nous avons pu faire et tous ces bons moments passés en votre compagnie.

Merci à Fouad de m'avoir aidé à augmenter le niveau de testostérone du laboratoire quand je me sentais seul parmi toutes ces femmes. On a bien rigolé et c'était très agréable d'avoir quelqu'un comme toi au laboratoire. Bon courage pour la suite de ta thèse.

Amandine, la ch'tite nordiste, je sais que je me suis moqué de nombreuses fois de ton accent, mais tu me connais, je viens du nord et je n'ai pu qu'apprécier ta compagnie. Merci pour ces éclats de rire, ta bonne humeur et ces bons moments.

Un grand merci à Ania pour m'avoir soutenu et fait rire surtout au court de cette dernière année. Nos fous rire ont été une arme précieuse.

Un immense merci à Camille Samson qui m'a formé et m'a donné de nombreux conseils concernant la biochimie des protéines et qui m'a supporté (et soutenu) durant neuf mois à Berlin, merci à toi, je garderai d'excellents souvenirs de ces années en ta compagnie.

ندي، حبيبتي، البست فريند، اختي الروحية. شكراً جداً علي وقفنك جمبي في الثلاثة سنين دول. شكراً علي الاوقات دي، علي سعدتك و ضحكك التجنن. مكننش اتمني صحبة احسن منك.

Merci à Simon ! Tu as toujours été là pour me soutenir, pour me booster lorsque j'en avais besoin. Seize ans d'amitié, qui n'a fait que se renforcer au cours de ces années de thèse. Merci encore à toi. PS : nos rendez-vous hebdomadaires à la Villa Borghèse vont me manquer.

Un grand merci à mon beau-frère, Jérémy Peultier, qui s'est assuré que la force me guide tout au long de ces trois années de thèse. Merci à toi and may the force be with you!

Un énorme merci à ma belle-famille, Maryse, Jean-Marie, Sébastien, Vanessa, Emilie et Damien. Merci pour vos encouragements, votre soutien et vos visites régulières à Cherbourg ou à Paris qui m'ont beaucoup aidé à tenir le coup.

Merci à mes grands-parents qui ont toujours été là, en toutes circonstances, qui se sont intéressés à mon projet de recherche et qui m'ont soutenu dans tout ce que j'ai pu entreprendre dans ma vie. Merci. Une pensée particulière pour Rivaldo, qui n'est plus là mais qui j'espère sera fier de ce travail.

Je souhaite adresser de chaleureux remerciements à mes sœurs Jeanne-Laure et Charline sans qui tout cela n'aurait pas été possible. Merci à vous les filles, des sœurs comme vous, ça n'a pas de prix.

Je n'ai que peu de mots pour exprimer la gratitude que j'ai envers mes parents qui, tout au long de ma vie ont fait de moi ce que je suis aujourd'hui. Merci pour votre soutien, vos conseils, votre humour à toute épreuve. Merci de m'avoir, chacun à votre manière, aidé tout au long de ces années de thèse.

Enfin, je voudrais remercier la personne qui partage ma vie depuis neuf ans et avec qui j'ai eu le plaisir de m'unir au cours de cette thèse, la femme de ma vie, Laetitia. Merci à toi de m'avoir guidé, soutenu et supporté. Nancy, Paris, Berlin, Cherbourg... nous en avons vu du pays. Avec nos thèses, nous sommes passés par tous les états, des hauts et des bas, mais nous en sortons forts et unis, maintenant, à nous la belle vie.

Table des matières

I.	Introduction.....	16
I.1.	Protéines : Structure, dynamique et fonction.....	17
I.1.1	Le dogme fondamental structure-fonction des protéines.....	17
I.1.2	Les protéines intrinsèquement désordonnées	18
I.1.2.1.	Spécificités et structure.....	19
I.1.2.2.	Modifications post-traductionnelles.....	23
I.1.2.3.	Auto-assemblage et agrégation	25
I.1.2.4.	Rôles	27
I.1.2.4.1	Les protéines désordonnées impliquées dans la régulation de la transcription des gènes	28
I.1.2.4.2	Les protéines désordonnées jouant un rôle dans le remodelage de la chromatine	28
I.1.2.4.3	Protéines désordonnées impliquées dans la signalisation cellulaire.....	29
I.2.	Enveloppe nucléaire.....	30
I.2.1	Présentation générale de l'enveloppe nucléaire.....	30
I.2.2	Les protéines ancrées à l'enveloppe nucléaire.....	32
I.2.2.1.	L'émerine.....	32
I.2.2.1.1	Auto-assemblage de l'émerine	33
I.2.2.1.2	Interaction émerine-BAF.....	35
I.2.2.2.	SUN1 et le complexe LINC.....	36
I.2.2.2.1	Auto-assemblage des protéines à domaine SUN et interaction avec les protéines à domaine KASH.....	36
I.2.2.2.2	Interaction entre SUN1 et la lamine A	37
I.3.	La Lamina nucléaire.....	38
I.3.1	La lamine A.....	39
I.3.1.1.	Structure.....	40
I.3.1.2.	Fonctions	41
I.3.1.2.1	Rôles de la lamine A dans la signalisation cellulaire	41
I.3.1.2.2	Implications de la lamine A dans la régulation de la transcription.....	41
I.3.1.2.3	Rôle de la lamine A dans le remodelage de la chromatine et le maintien de la structure du noyau.....	42

I.3.1.3.	Modifications post-traductionnelles	43
I.3.1.3.1	Régulation de la lamine A par phosphorylation	43
I.3.1.3.2	La farnésylation de la lamine A	44
I.3.2	Les pathologies associées à l'émerine, SUN1/2 et la lamine A/C.....	47
I.3.3	Une pathologie sévère associée à la région spécifique de la lamine A : le syndrome d'Hutchinson-Gilford ou progéria	49
I.3.4	Les syndromes progéroïdes récessifs associés aux lamines A et C	51
I.4.	Les partenaires de l'enveloppe nucléaire associés à la chromatine	53
I.4.1	BAF.....	53
I.4.1.1.	Les protéines à domaine LEM de la membrane nucléaire interne.	54
I.4.1.2.	Les lamines.	55
I.4.1.3.	Histones et protéines de remodelage de la chromatine.....	56
I.4.2	RBBP4	56
I.5.	Hypothèses et objectifs	57
II.	Résultats.....	59
II.1.	Analyse structurale de la région nucléoplasmique de l'émerine	61
II.1.1	Purification de la région nucléoplasmique de l'émerine	61
II.1.2	Analyse structurale de la région nucléoplasmique de l'émerine monomérique.....	62
II.1.3	Etude de l'auto-assemblage de l'émerine.....	68
II.1.3.1.	Structure	68
II.1.3.2.	Fonction.....	72
II.2.	Détermination d'un facteur de forme des fibres au cours du temps Etude de filaments protéiques grâce au système SF-WA-MALS-SAXS.....	73
II.2.1	Adaptation du modèle émerine à ce type d'expérience.....	74
II.2.2	Développement du système SF-WA-MALS-SAXS à l'aide du Fibrinogène	75
II.2.2.1.	Automatisation du système de SF-WA-MALS-SAXS	75
II.2.2.2.	Insertion d'une bulle d'air	79
II.2.2.3.	Etude de polymérisation du fibrinogène.....	82
II.3.	Etude de la région C-terminale intrinsèquement désordonnée de la lamine A	84
II.3.1	Production des échantillons nécessaires à l'étude structurale des régions désordonnées de la prélamine A et de la progérine.....	86
II.3.1.1.	Production des régions C-terminales de la prélamine A et de la progérine	86
II.3.1.2.	Production des partenaires	87
II.3.1.2.1	BAF.....	88
II.3.1.2.2	SUN1.....	89

II.3.1.2.3	RBBP4	92
II.3.2	Etude par RMN des régions C-terminales désordonnées de la prélamine A et de la progérine93	
II.3.2.1.	Les deux peptides possèdent une région conservée adoptant transitoirement une conformation en hélice α	93
II.3.2.2.	Le peptide C-terminal de la lamine A reste difficile à obtenir sous une forme structuralement homogène	96
II.3.3	Etude de la farnésylation de la prélamine A et de son mutant progérine	101
II.3.4	Interactions médiées par les régions désordonnées de la prélamine A et de la progérine102	
II.3.4.1.	Rôle de la région C-terminale désordonnée de la lamine A dans la régulation de l'interaction BAF-Lamine A	102
II.3.4.2.	Interaction entre la région C-terminale désordonnée de la lamine A et SUN1 ..	107
II.3.4.3.	Etude de l'interaction entre la région C-terminale désordonnée de la lamine A avec RBBP4	107
III.	Discussion	110
IV.	Matériel et méthode	121
IV.1.	Production des échantillons protéiques.....	122
IV.1.1	Tampons utilisés	122
IV.1.2	Production des protéines à partir des corps d'inclusion.....	122
IV.1.3	Production des échantillons solubles.....	122
IV.2.	Microscopie électronique.....	123
IV.3.	Cinétiques de fluorescence à la thioflavine T.....	124
IV.4.	RMN en phase liquide	124
IV.5.	RMN du solide	124
IV.6.	Farnésylation <i>in vitro</i>	125
IV.7.	Microcalorimétrie.....	125
V.	Références.....	126
VI.	Annexes	151
	Publications	152
	Attribution de l'émerine 67-170 (0M urée)	190
	Attribution de l'émerine 67-170 (8M urée)	201
	Attribution du peptide ProgC _{CtoA}	209
	Attribution du peptide ProgC	214
	Attribution du peptide PreLamC _{CtoA}	219
	Détermination d'un facteur de forme des fibres au cours du temps	228

Utilisation d'un système de mélange rapide.....	228
Analyse des données de WA-MALS.....	228
Modélisation des données	230

Tableaux

Tableau 1 : Calcul des rendements obtenus lors de la production des différents fragments d'émerine	62
Tableau 2 : Rendements de purification obtenus pour les différentes constructions des peptides C-terminaux de prélamine et progérine utilisées.....	87

Figures

Figure 1 : Impact des acides aminés sur la propension de désordre d'une protéine ¹⁸	19
Figure 2 : Comparaison de la charge nette et de l'hydrophobicité des protéines intrinsèquement désordonnées et des protéines possédant une structure tertiaire stable.	19
Figure 3 : Des degrés variables de désordre au sein des protéines ¹³	20
Figure 4 : Structuration de domaine de régulation de la calcineurine lors de son interaction avec la calmoduline ²¹	21
Figure 5 : Illustration des différents groupes de fuzzy complexes par ordre croissant de désordre ²⁷ .	23
Figure 6 : Influence des conditions physico-chimiques sur les transitions de phase observées avec le domaine « low complexity » (LC) de la protéine FUS avec ou sans coupure de son étiquette de purification ⁶⁹	26
Figure 7 : Modèle de l'enveloppe nucléaire d'une cellule somatique de mammifères, construit à partir de données de cryoélectrotomographie ⁹⁴	30
Figure 8 : Architecture de l'enveloppe nucléaire ⁹⁴	31
Figure 9 : Structure du domaine LEM de l'émerine résolue par RMN ^{103,104}	32
Figure 10 : Modulation du rôle du domaine LEM grâce à l'auto-assemblage de l'émerine ¹¹¹	34
Figure 11 : L'émerine 1-187 (notée EmN) s'auto-assemble pour former des filaments ¹¹³	35
Figure 12 : Structure du domaine LEM de l'émerine lié à un dimère de BAF (code PDB : 2ODG).....	36
Figure 13 : Modèle de l'activation de SUN2 pour la liaison de KASH dans l'espace péri-nucléaire.....	37
Figure 14 : SUN1 interagit via son domaine N-terminal à la lamine A spécifiquement ¹²⁴	38
Figure 15 : Etapes de maturation des lamines (d'après Dechat et al., 2008).	39
Figure 16 : Représentation schématique du dimère de prélamine A ¹³⁰ ..	40
Figure 17 : Illustration de l'assemblage des dimères de lamines en filaments ¹³⁴	40
Figure 18 : Régulation de l'activité d'AP-1 via l'interaction entre la lamine A et ERK1/2 ¹⁴⁴	42
Figure 19 : Représentation topologique de la molécule de farnésylpyrophosphate.....	44
Figure 20 : Localisation cellulaire de la progérine et la progérine non farnésylée (CS) ¹⁶⁸	45
Figure 21 : SUN1 lie la lamine A farnésylée ainsi que la progérine farnésylée ¹⁶⁹	46
Figure 22 : Mutations des lamines et laminopathies ¹⁸¹	48
Figure 23 : Comparaison par microscopie à fluorescence d'un noyau d'une cellule saine (gauche) avec le noyau d'une cellule HGPS (droite) ¹⁹⁷	50
Figure 24 : Impact des mutations causant des syndromes progéroïdes récessifs sur l'interaction avec BAF (Samson et al., 2018).....	52
Figure 25 : Structure du dimère de BAF en interaction avec l'ADN ²³⁸ (code PDB : 2BZF).....	54
Figure 26 : L'accumulation de lamine A entraîne la translocation de BAF vers le noyau ²⁴²	55
Figure 27 : Superposition des 18 structures de Rbbp4 disponibles dans la PDB.	56
Figure 28 : La lamine A interagit avec le complexe NuRD (adapté de Pegoraro et al., 2009).	57
Figure 29 : Régions prédites comme déstructurées des protéines de l'enveloppe nucléaire interne étudiées lors de ma thèse.	58
Figure 30 : Gels SDS-PAGE des différentes constructions purifiées de l'émerine.....	62

Figure 31 : Prédiction de la structure secondaire et de la propension au désordre de l'émerine entière, obtenue sur le site http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1	63
Figure 32 : Superposition des spectres ¹ H- ¹⁵ N HSQC des différents fragments de l'émerine.....	64
Figure 33 : Suivi des signaux RMN de quelques résidus du fragment 67-170 de l'émerine à 8 M (brun), 7 M (rouge), 6 M (orange), 4 M (vert), 2 M (bleu clair), 1 M (bleu), 0 M urée (bleu foncé).....	65
Figure 34 : Attribution de l'émerine 67-170.....	66
Figure 35 : Différences entre les déplacements chimiques secondaires des C α et C β en fonction de la séquence de l'émerine mesurés en présence de 0 M et 8 M d'urée auxquels ont été soustraits les déplacements chimiques secondaires prédits en l'absence de structure secondaire par le logiciel nclDP (http://nmr.chem.rug.nl/nclDP/)(Mulder et al.,2010).....	67
Figure 36 : Attribution de la région 1-170 de l'émerine sur le spectre ¹ H- ¹⁵ N HSQC de la construction 1-187 enregistré à 750 MHz et 303K et à 500 μ M dans un tampon 20 mM phosphate pH 6.5, 30 mM NaCl.	68
Figure 37 : Filaments obtenus avec le fragment d'émerine 67-221.	69
Figure 38 : Cinétique de formation des filaments de l'émerine 67-221 (notée EmC221 ; points violet), comparée aux cinétiques obtenues sur l'émerine 1-187 (notée EmN WT ; points rouge) et sur l'émerine 67-187 (notée Em67C ; points bleu).	69
Figure 39 : Constructions utilisées lors de l'analyse du rôle du domaine LEM dans l'oligomérisation de l'émerine 1-187.	70
Figure 40 : Formation des filaments par le mutant delK37 ²⁵¹	70
Figure 41 : Le domaine LEM de l'émerine est destructuré lors de la formation des filaments.	71
Figure 42 : Interaction entre le domaine globulaire IgFold de la lamine A et les oligomères d'émerine ²³⁰	73
Figure 43 : Profil de diffusion de lumière dynamique, illustrant le caractère oligomérique de l'émerine à 500 μ M à T = 0 min.....	74
Figure 44 : Système SF-WA-MALS-SAXS complet, installé sur la ligne SWING.....	76
Figure 45 : Lavage du circuit par le vide apporté grâce au robot d'injection.....	77
Figure 46 : Insertion de tampon dans le circuit afin d'obtenir une ligne de base.....	78
Figure 47 : Injection de mélange réactionnel vers le MALS et le capillaire de SAXS.....	79
Figure 48 : Profil de données enregistrées avec le logiciel ASTRA lors d'une injection de BSA illustrant la diminution de la densité optique par dilution avec le tampon.....	80
Figure 49 : Profil de données MALS et UV.....	80
Figure 50 : Configuration du robot d'injection lors de l'injection d'un mélange réactionnel.....	81
Figure 51 : Configuration du robot d'injection permettant l'injection d'une bulle d'air après l'échantillon protéique.	81
Figure 52 : profil de données MALS illustrant une expérience de polymérisation du fibrinogène à l'aide d'une solution de thrombine.....	82
Figure 53 : Exemple de modélisation des données de polymérisation du fibrinogène obtenues avec le MALS et analysées avec la méthode de Fisher-Burford.....	83
Figure 54 : Analyse bioinformatique de la séquence de la région C-terminale de la prélamine A (567-664) (http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1)	86
Figure 55 : Alignement de séquence de la région C-terminale de la prélamine A en fonction des espèces.....	86
Figure 56 : Gels SDS-PAGE des différentes constructions de lamine A.....	87
Figure 57 : Filtration sur gel du dimère de BAF.....	88

Figure 58 : Spectre $^1\text{H}^{15}\text{N}$ HSQC de BAF enregistré à 100 μM , 700 MHz, 293 K dans un tampon contenant 20 mM phosphate pH 6.5, 150 mM NaCl et 2 mM DTT.	89
Figure 59 : Prédiction de structure secondaire et de propension de désordre de la région 1-139 de SUN1 murine et humaine, comparaison avec SUN2.	90
Figure 60 : Purification et analyse par RMN de la protéine SUN1. (....	91
Figure 61 : Gels SDS-PAGE de la purification de RBBP4.	92
Figure 62 : Profil de filtration sur gel de la protéine RBBP4 réalisée dans un tampon contenant 50 mM Tris-HCl pH8 ; 150 mM NaCl ; 2 mM DTT avec une colonne Superdex 75 10/300 GL.	93
Figure 63 : Spectres RMN $^1\text{H}^{15}\text{N}$ HSQC des peptides C-terminaux désordonnés de la prélamine A et de la progérine avec et sans cystéines (Celli et al., 2018).	95
Figure 64 : Les régions C-terminales désordonnées de la prélamine A et de la progérine possèdent une région ayant une tendance à former une hélice α transitoire. (....	96
Figure 65 : Gels SDS-PAGE montrant l'hétérogénéité d'un échantillon PreLamC.	97
Figure 66 : Spectres RMN $^1\text{H}^{15}\text{N}$ HSQC illustrant les difficultés à obtenir un échantillon homogène de PreLamC.	98
Figure 67 : Caractérisation de la construction prélamine A-GB1. (A) Schéma illustrant la construction PreLamC-GB1 (saumon)	99
Figure 68 : Variations de déplacements chimiques combinés entre le peptide PreLamC et PreLamC-GB1 en fonction de la séquence en acides aminés.	100
Figure 69 : Intensités mesurées sur les spectres $^1\text{H}^{15}\text{N}$ HSQC des peptides ProgC, PreLamC-2 et PreLamC-GB1.	100
Figure 70 : Farnésylation de la région C-terminale de la progérine.	101
Figure 71 : La farnésylation de la région C-terminale dépliée du peptide progérine provoque l'apparition d'échanges conformationnels autour de la position farnésylée.	102
Figure 72 : Etude des interactions entre les régions C-terminales dépliées des prélamine et progérine et l'IgFold.	103
Figure 73 : Etude de l'interaction du peptide ProgFarn seul ou en présence de BAF d'une part ou de l'IgFold d'autre part.	106
Figure 74 : Etude de l'interaction entre le peptide PreLamC-GB1 et RBBP4.	107
Figure 75 : Interaction entre RBBP4 et BAF.	108
Figure 76 : Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide PreLamC-GB1 seul (saumon) ou en présence de BAF et RBBP4 (cyan).	108
Figure 77 : Analyse par RMN des phosphorylations de l'émerine 1-187 par différentes kinases.	114
Figure 78 : Structure cristallographique du domaine IgFold de la lamine A/C (gris, gauche) en interaction avec le dimère de BAF (bleu et bleu-clair, centre) lui-même en interaction avec le domaine LEM de l'émerine (bleu foncé, droite) (Code PDB : 6GHD).	115
Figure 79 : Courbe de titration suivie par ITC, illustrant l'interaction entre le fragment IgFProg (domaine Igfold et région C-terminale de la progérine) et BAF.	116
Figure 80 : Représentation de Zimm. Les points représentent différentes valeurs de concentration (indiquées en bleu) mais également des angles de diffusion différents (indiqués en rouge).	230
Figure 81 : Comparaison des graphiques obtenus par la méthode de Zimm après avoir choisi manuellement les degrés de polynômes ainsi que les angles à inclure, avec les résultats obtenus avec l'équation de Fisher-Burford.	232
Figure 82 : Modes de polymérisation du fibrinogène.	233

Abréviations

ADN : Acide Désoxyribonucléique

BAF : BArrier-to-autointegration Factor 1

CBP : CREB Binding Protein

CEA : Commissariat à l'Energie Atomique

C.elegans : *Caenorhabditis elegans*

EDMD : dystrophie musculaire d'Emery-Dreifuss

DTT Dithiothréitol

EGF : Epidermal Growth Factor

ERK : Extracellular signal-regulated Kinase

E.coli : *Escherichia coli*

Fus3p : Fusion 3 protein

HES1 : Hairy and Enhancer of Split

HGPS : Hutchinson Gilford Progeria Syndrom

HSQC : Heteronuclear Single Quantum Correlation

Icmt : isoprénylcystéine carboxyméthyl-transférase

JNK : c-Jun amino terminal Kinase

IDPs : protéines intrinsèquement désordonnées

IDR : régions intrinsèquement dépliées

MAP : Mitogen Activated Protein

Mok2 : Mouse Krüppel-like 2

MSCs : cellules souches mésenchymateuses

ME : Microscope Electronique

NETs : domaines transmembranaires

NICD : Notch intracellulaire (NICD)

PCNA : Proliferating Cell Nuclear Antigen

PKC : Protéine Kinase C

pKID : phosphorylated Kinase Inducible Domain

PTMs : modifications post-traductionnelles

RHIM : RIP homotypic interaction motif

RIP : Receptor Interacting Protein Kinase

RMN : Résonance Magnétique Nucléaire

SREBP1 : Sterol Regulatory Element-Binding Protein 1

SOFAST-HMQC : band-Selective Optimized-Flip-Angle Short-Transient-Heteronuclear Multiple Quantum Correlation

I. Introduction

1.1. Protéines : Structure, dynamique et fonction

Les protéines sont des macromolécules biologiques présentes au sein de toutes les cellules. Elles sont constituées de chaînes polypeptidiques, elles-mêmes composées d'acides aminés reliés entre eux par une liaison peptidique. Ces acides aminés possèdent une chaîne latérale variable leur conférant diverses propriétés physico-chimiques.

1.1.1 Le dogme fondamental structure-fonction des protéines

Les protéines assurent un ensemble de fonctions au sein des cellules mais également à l'échelle des tissus et des organismes. Elles peuvent avoir un rôle (i) enzymatique en catalysant les réactions chimiques, (ii) structural comme notamment les protéines impliquées dans l'architecture du nucléosquelette, (iii) de signalisation cellulaire comme les protéines impliquées dans la réponse à un dommage de l'ADN, (iv) de compaction de l'Acide Désoxyribonucléique (ADN) comme les histones, ou encore (v) de régulation de l'expression des gènes comme les facteurs de transcription.

Les protéines peuvent être définies par leur séquence en acides aminés ou structure primaire. Les éléments de type hélice α , feuillet β ou coudes formés par ces séquences polypeptidiques constituent la structure secondaire. Enfin, les protéines adoptent une structure d'ordre supérieur appelée structure tertiaire, soit un repliement résultant de différents types d'interaction (hydrophobe, électrostatique, ponts-disulfure). Le repliement des protéines est un processus thermodynamique permettant de diminuer l'énergie du système en optimisant les contacts intramoléculaires ainsi que les contacts avec le milieu extérieur. Enfin, les protéines peuvent s'auto-assembler pour former des oligomères, on parle alors de structure quaternaire.

La biologie structurale des protéines est fondée sur un dogme fondamental structure-fonction selon lequel la fonction d'une protéine est définie par ses structures tertiaire et quaternaire. Ainsi, les méthodes de résolution de la structure tridimensionnelle des protéines, à savoir la diffraction des rayons X ainsi que la Résonance Magnétique Nucléaire (RMN) ou encore, plus récemment la Microscopie Electronique (ME), ont permis d'apporter une explication moléculaire à de nombreux mécanismes biologiques. Cependant, au cours des dix dernières décennies, il a été découvert que les mouvements des protéines avaient aussi une importance fonctionnelle (Huber *et al.*, 1979). De plus, de nombreuses protéines possèdent des régions absentes des cartes de densités électroniques après diffraction aux rayons X^{1 2}; ³). Ces régions sont flexibles et leur structure primaire est caractérisée par une distribution spécifique en acides aminés.^{4 5 6}). Leur analyse a permis le développement de serveurs de prédiction de désordre comme par exemple DISOPRED3 (<http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1>⁷), ou même de métaserveurs permettant

d'utiliser simultanément des méthodes de prédiction de désordre différentes comme par exemple Metadisorder (<http://genesilico.pl/metadisorder/>⁸).

Il a été observé que plus de 30% du protéome eucaryote était composé de protéines comportant des régions désordonnées de plus de 30 acides aminés⁹, appelées «protéines intrinsèquement désordonnées» (IDPs)¹⁰. A l'heure actuelle, la communauté scientifique parle de trinité protéique, un principe selon lequel une protéine native pourrait adopter **une structure bien définie, ou partiellement désordonnée, ou encore entièrement déstructurée**¹¹.

1.1.2 Les protéines intrinsèquement désordonnées

Les IDPs sont des protéines ne possédant pas de structure tridimensionnelle fixe. En effet, ces protéines échantillonnent un grand ensemble de conformations et leurs structures ne peuvent pas être représentées par un jeu de coordonnées cartésiennes. Par ailleurs, si les protéines globulaires ont des angles de Ramachandran moyens très comparables au cours du temps, il n'en est rien pour les IDPs qui ont, elles, des angles de Ramachandran variant au cours du temps¹². Contrairement aux régions charnières flexibles ayant essentiellement pour but de relier les différents modules structurés et fonctionnels des protéines, les IDPs sont fonctionnelles par elles-mêmes¹³. Les premières protéines intrinsèquement désordonnées et fonctionnelles décrites sont les hormones polypeptidiques telles que le glucagon monomérique, désordonné en solution¹⁴.

L'intérêt pour les IDPs a crû de manière exponentielle ces dernières années car de nombreuses protéines possédant des régions désordonnées ont été identifiées dans des mécanismes de régulation cruciaux tels que la signalisation cellulaire, le remodelage de la chromatine ainsi que dans la transcription des gènes. Les avancées des méthodes de biologie structurale, notamment la RMN et la diffusion des rayons X aux petits angles (SAXS), permettent maintenant la caractérisation structurale des IDPs. Il est possible **de mesurer la propension d'une protéine à adopter une structure secondaire ou un repliement globulaire mais également de décrire sa dynamique sur une grande gamme d'échelles de temps**. Enfin, un lien intime existe entre la conformation des IDPs et les pathologies du repliement, notamment les pathologies neurodégénératives telles que les maladies d'Alzheimer ou de Parkinson. En effet, dans ce type de pathologies, les IDPs adoptent de manière transitoire des conformations, dans lesquelles leurs régions hydrophobes sont accessibles au solvant, entraînant leur agrégation. La bonne compréhension des mécanismes de régulation auxquels les IDPs contribuent permettra de mieux prédire la survenue des pathologies qui leur sont associées.

I.1.2.1. Spécificités et structure

La structure primaire des IDPs est caractérisée par une faible complexité en acides aminés avec notamment un faible pourcentage en acides aminés hydrophobes (Val, Leu, Ile, Met, Phe, Trp and Tyr) et un pourcentage élevé en acides aminés polaires et chargés (Gln, Ser, Gly, Ala, Lys, Pro and Glu)^{15 16 17}; Figure 1). Cette faible complexité empêche les IDPs d'adopter une structure tertiaire stable dans le temps. L'observation des valeurs de charge nette et d'hydrophobicité d'une protéine permet ainsi de prédire son aptitude à se comporter comme un IDP (Figure 2).

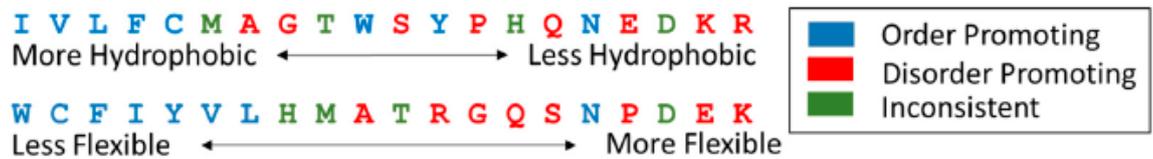


Figure 1 : Impact des acides aminés sur la propension de désordre d'une protéine¹⁸.

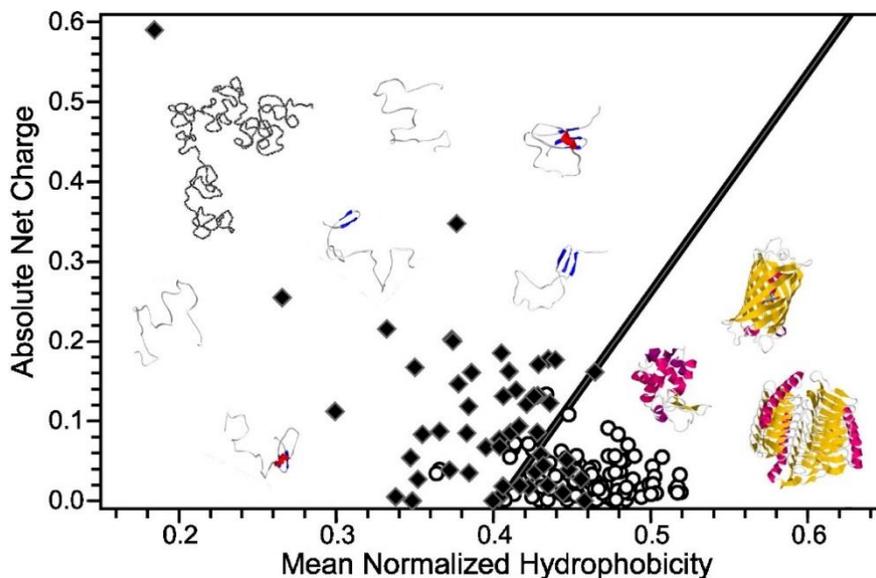


Figure 2 : Comparaison de la charge nette et de l'hydrophobicité des protéines intrinsèquement désordonnées et des protéines possédant une structure tertiaire stable. Les protéines structurées (ronds blancs) sont caractérisées par une hydrophobicité élevée et une charge nette faible. Les IDPs (losanges noirs), possèdent une charge nette importante mais une faible hydrophobicité moyenne¹⁹

D'autre part, les IDPs ne sont pas toujours entièrement désordonnées. Si elles n'ont pas la possibilité d'adopter une structure tertiaire bien définie, elles ont tout de même parfois tendance, en échantillonnant diverses conformations, à former des structures transitoires tels que des structures secondaires résiduelles, ainsi que des contacts tertiaires transitoires dus à

des rapprochements entre résidus hydrophobes²⁰. Ces structures résiduelles aussi sont retrouvées dans des régions intrinsèquement dépliées (IDR) de protéines en partie structurées (Figure 3).

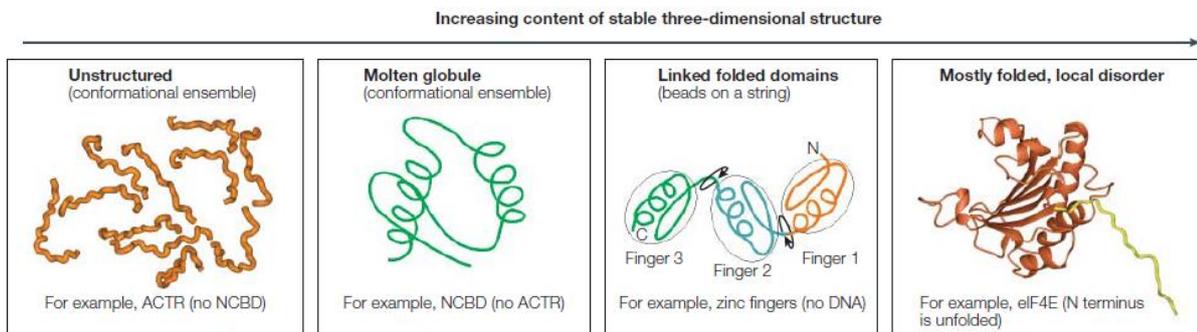


Figure 3 : Des degrés variables de désordre au sein des protéines¹³. La proportion de structure tridimensionnelle stable augmente de gauche à droite à partir d'une protéine totalement désordonnée, puis un globule fondu pouvant être vu comme un intermédiaire de repliement, puis une protéine modulaire, qui contient des régions de liaison désordonnées entre ses modules structurés, et enfin une protéine structurée pouvant néanmoins contenir des régions localement désordonnées.

Les IDPs et IDRs sont impliquées dans divers mécanismes de régulation très importants. Afin d'avoir un rôle régulateur, elles se lient à d'autres partenaires protéiques, promouvant ou empêchant des interactions protéine-protéine. Il arrive, lors d'un tel mécanisme, qu'une IDP/IDR se structure en se liant à son partenaire. Ce partenaire peut être une autre protéine, comme dans le cas de la calcineurine dont le domaine de régulation subit une transition d'un état désordonné à un état structuré lors de son interaction avec la calmoduline²¹ (Figure 4) ou un brin d'ADN comme pour la protéine de levure GCN4 (homologue de l'oncogène humain c-jun). En effet, ce facteur de transcription comporte une région basique désordonnée en l'absence d'ADN. Lorsque GCN4 se lie à l'ADN, le domaine basique adopte une conformation en hélice α ²². C'est un exemple de structuration par liaison à l'ADN. Cette liaison est facilitée par le fait que GCN4 seule n'est pas entièrement désordonnée mais possède une structure transitoire en hélice α ²³. Les protéines membranaires sont également souvent dépliées en solution et se replient en adoptant des structures en hélice au contact de la membrane^{24,25}. Une même IDP/IDR peut se lier à plusieurs partenaires en adoptant chaque fois une structure liée différente, ce qui lui permet d'influer dans un sens ou dans un autre sur une régulation²⁶. Ceci est un autre avantage de ne pas posséder de structure définie. Lors du repliement d'une IDP/IDR, il y a un coût entropique qui est compensé par l'enthalpie de liaison au partenaire, thermodynamiquement favorable. Les complexes formés de cette manière donnent généralement des systèmes à haute spécificité mais ayant une faible affinité. Ceci permet une grande réversibilité car la dissociation de ces complexes est rapide. Une haute spécificité, avec une réversibilité rapide, sont deux caractéristiques parfaitement adaptées à la transduction d'un signal cellulaire par exemple⁹.

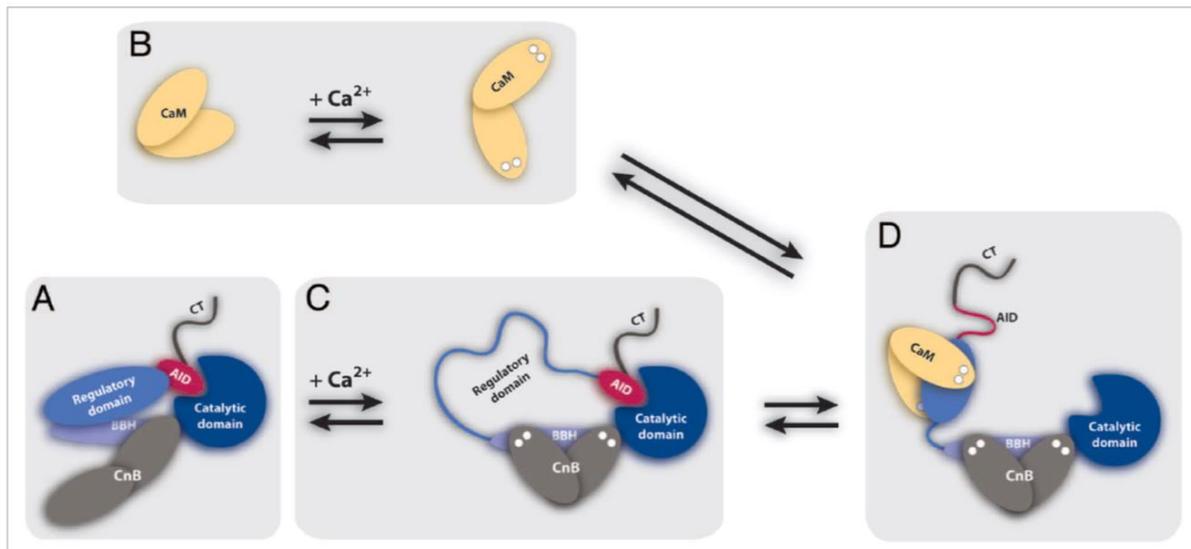


Figure 4 : Structuration de domaine de régulation de la calcineurine lors de son interaction avec la calmoduline²¹. (A) En absence de calcium, la calcineurine est en conformation inactive. (B) La calmoduline adopte une conformation active après liaison au calcium. (C) La calcineurine lie aussi le calcium et son domaine de régulation devient temporairement désordonné. (D) La calmoduline active se fixe sur le domaine de régulation de la calcineurine, entraînant sa structuration.

Les IDPs/IDR peuvent également être impliqués dans des complexes multi-protéiques montrant une quantité non négligeable de désordre après liaison des différents partenaires. Il s'agit alors de « fuzzy complexes ». Les fuzzy complexes ont été classés en 4 catégories selon leur degré de désordre une fois la liaison effectuée²⁷ (Figure 5).

- Le modèle polymorphique :

Une protéine désordonnée peut adopter plusieurs conformations différentes lors de l'interaction avec son partenaire. Il s'agit alors d'un modèle polymorphique. Par exemple, pour permettre l'adhésion médiée par les cadhérines, la β -caténine est transloquée vers le noyau où elle jouera un rôle de coactivateur de facteurs de transcription comme le facteur de transcription de lymphocytes T : Tcf4. Le domaine désordonné de Tcf4 lie la β -caténine en adoptant une conformation étendue (Figure 5a). Cependant, son segment intermédiaire acide établit des ponts salins alternatifs qui sont cruciaux pour l'interaction²⁶. De nombreuses conformations alternatives sont également connues dans le cas des prions sous forme de fibres amyloïdes²⁸. Certaines conformations possèdent un rôle physiologique bien établi²⁹.

- Le modèle « Clamp » :

Dans ce modèle, une protéine possédant deux domaines structurés reliés par une région charnière désordonnée peut interagir via ses domaines structurés sans pour autant entraîner une structuration de sa région charnière. Le plus souvent, la région charnière n'établit aucun contact avec le partenaire lors de l'interaction. C'est le cas de la protéine Scaffolding Protein Sterile 5 (Step5) qui se lie à la protéine Fusion 3 protein (Fus3p) durant la reproduction des levures haploïdes. Step5 interagit via deux segments protéiques séparés par une région charnière désordonnée de 8 résidus qui reste dépliée lors de l'interaction³⁰ (Figure 5b). De

manière intéressante, cette interaction n'a pas lieu si l'on étudie l'interaction des deux domaines structurés sans la présence de la région charnière, ce qui montre son importance dans la régulation de cette interaction.

- Le modèle « flanquant » :

Les IDPs se lient souvent à leurs partenaires via de courts éléments de reconnaissance enfouis au sein d'un environnement désordonné^{31,32}. Cependant, il arrive que les régions désordonnées entourant les éléments de reconnaissance améliorent l'affinité du complexe formé. Le domaine KID désordonné^{33,34} de la protéine CREB par exemple, se lie au domaine KIX de la protéine CBP via un fragment de 29 acides aminés adoptant une structure hélice-boucle-hélice lors de l'interaction³³ (Figure 5c). Bien que la portion restante de KID reste désordonnée dans le complexe, lorsque ces régions désordonnées sont délétées, elles entraînent une diminution du Kd d'un facteur 5³⁵. Le facteur d'épissage SF1 est un second exemple de régions flanquantes désordonnées impliquées dans une interaction. Lorsque SF1 interagit avec U2AF, l'affinité de l'interaction entre le domaine de liaison de SF1 isolé et U2AF est bien moins élevée que l'affinité de l'interaction effectuée avec SF1 entière, montrant la nécessité de ces régions flanquantes³⁶. Les régions flanquantes désordonnées peuvent également être le siège d'autres modes de régulation. En effet, elles peuvent subir des PTMs permettant de moduler l'interaction³⁷. Elles peuvent également initier la dégradation par le protéasome³⁸ ou encore empêcher l'agrégation des protéines³⁹.

- Le modèle « random »

Ce modèle représente la dynamique du désordre extrême. Il apparaît lorsqu'une protéine est majoritairement désordonnée lors d'une interaction. Un exemple de ce phénomène est la structure du récepteur de lymphocytes T en interaction avec un antigène (Figure 5d). L'élastine, qui forme des fibres par auto-assemblage des monomères de tropoélastine, illustre également ce modèle. Une étude de ces fibres par RMN du solide a montré un fort degré de désordre dynamique à l'intérieur de ces fibres avec notamment des valeurs de déplacement chimiques secondaires des carbones de la chaîne principale caractéristiques d'une protéine désordonnée⁴⁰. Dans ce cas, l'auto-assemblage sous forme de fibre est médié par des interactions hydrophobes ainsi que par des liaisons hydrogènes médiées par l'eau⁴¹.

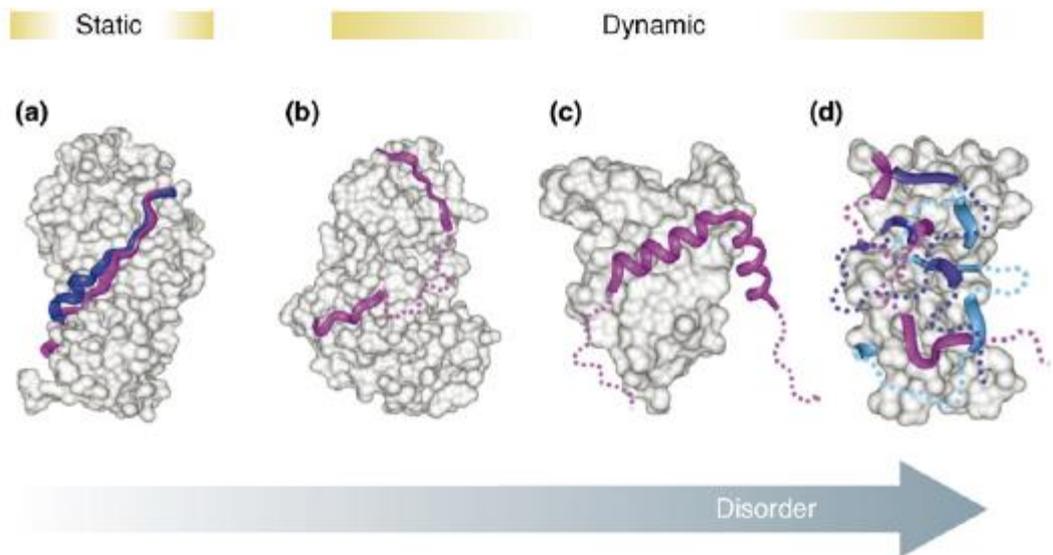


Figure 5 : Illustration des différents groupes de fuzzy complexes par ordre croissant de désordre²⁷. (a) Modèle polymorphique illustré par l'interaction entre Tcf4 et la β -caténine. (b) Modèle "clamp" utilisé par Step5p en interaction avec la MAP kinase Fus3p. (c) Modèle flanquant montré grâce à l'interaction entre CREB KID et CBP-KIX. (d) Modèle «random» illustré par une structure hypothétique de récepteur de lymphocyte T en interaction avec un antigène.

Enfin, l'état désordonné des IDPs/IDRs leur permet d'être facilement accessibles aux protéases. Cette dégradation rapide permettrait d'accroître leur versatilité et il est possible que la dégradation par des protéases soit un mécanisme de régulation physiologique de la fonction des IDPs/IDRs⁴².

1.1.2.2. Modifications post-traductionnelles

Grâce à leur grande accessibilité, les IDPs sont une cible privilégiée des PTMs. Ces PTMs peuvent également être rapidement retirées⁴³, permettant ainsi une fine régulation des processus biologiques. Différents types de PTMs sont observés sur les IDPs : soit l'ajout de groupements chimiques (glycosylation, acétylation, phosphorylation, etc), soit l'ajout d'une petite protéine (sumoylation, ubiquitinylation).

Les PTMs influent sur la protéine cible de différentes manières. Premièrement, l'ajout d'un groupement chimique modifie la structure primaire de la protéine. Ceci entraîne la modification des propriétés chimiques de l'acide aminé, comme l'ajout de charges lors d'une phosphorylation, ou au contraire la neutralisation de charges lors d'une acétylation. Les IDPs peuvent comporter une, plusieurs, ou une combinaison de différentes PTMs. Ceci permet de déclencher des mécanismes d'amplification ou de compétition pour une régulation fine du signal porté par l'IDP. Une seconde manière de moduler l'activité d'une IDP via une PTM est de permettre la stabilisation ou non de structures secondaires résiduelles⁴⁴. Enfin, une PTM peut permettre le rapprochement ou l'éloignement de résidus éloignés dans la séquence en induisant un changement de la structure tridimensionnelle de l'IDP, comme dans le cas de la protéine p27 qui se lie au complexe Cdk4/cycline D dans une conformation ouverte et qui,

grâce à la phosphorylation de sa tyrosine 88, adopte une conformation fermée qui entraîne la levée de l'inhibition du complexe Cdk4/cycline D⁴⁵.

Enfin, l'ajout d'un groupement chimique peut entraîner une stabilisation ou une déstabilisation locale de la structure de la protéine, lui permettant de passer d'un état désordonné à un état structuré⁴⁶⁻⁴⁹, ou de s'auto-associer pour former des structures complexes de type fibrilles⁴⁹ ou des transitions de phase⁵⁰, ou encore d'interagir avec un partenaire²².

La PTM la plus fréquente est la phosphorylation. Il est estimé qu'environ un tiers des protéines eucaryotes pourraient être phosphorylées, et la plupart des phosphorylations se situent dans des IDRs^{37,51}. Cette modification consiste en l'ajout d'un groupement phosphoryl tétraédrique portant deux charges négatives à la place d'une fonction hydroxyle. Le groupement phosphoryl permet de nouvelles interactions électrostatiques grâce aux deux charges négatives présentes, mais également des liaisons hydrogènes supplémentaires grâce aux atomes d'oxygènes ajoutés. Par ailleurs les sérines phosphorylées ont un rôle stabilisateur d'hélice lorsqu'elles sont situées en N-terminal d'une hélice α tandis qu'elles peuvent déstabiliser une telle structure si la phosphorylation est positionnée au cœur ou en C-terminal de l'hélice α ⁵². De plus, les IDPs possédant des sérines et des thréonines phosphorylées ont été montrées comme interagissant avec différents domaines fréquemment retrouvés dans les protéines, tels que les domaines SH2, WW, WD40 et BRCT⁵³. Ainsi, la phosphorylation des IDPs permet d'accroître drastiquement les possibilités d'interaction avec d'autres partenaires protéiques. De la même manière, l'acétylation des lysines des IDPs ainsi que la méthylation des arginines leur permet d'interagir avec les protéines à domaines Bromo⁵⁴ et Tudor⁵⁵, respectivement.

Les histones sont des protéines impliquées dans la compaction de l'ADN, elles forment le nucléosome autour duquel s'enroule l'ADN. Les histones font partie des protéines ayant des IDRs et dont les PTMs entraînent des changements flagrants de comportement vis-à-vis de leurs partenaires. En effet, les queues N-terminale et C-terminale des histones sont dépliées et sont le siège de nombreuses PTMs. Ces PTMs ont un rôle central dans la régulation de l'accessibilité de l'ADN au sein de la chromatine⁵⁶. Par exemple l'acétylation des lysines des queues des histones neutralise la charge positive de la lysine acétylée et modifie l'interaction de l'histone à la molécule d'ADN chargée négativement⁵⁷. Comme dit précédemment, les acétylations et méthylation de certains résidus des histones permettent la liaison de protéines possédant des domaines Bromo et Tudor. Il se trouve que de nombreuses protéines impliquées dans les complexes de remodelage de la chromatine possèdent ce type de domaine. Les PTMs peuvent agir sur le recrutement de ces complexes de remodelage de la chromatine et entraîner localement la compaction ou non de l'ADN, régulant ainsi l'expression des gènes.

Certaines PTMs favorisent également l'ancrage de protéines non membranaires ayant des régions désordonnées à des membranes phospholipidiques. En effet, l'ajout de

groupements myristate, palmitate, isoprénoïdes, glycosylphosphatidylinositol⁵⁸ ou encore farnésylpyrophosphate, contribuent à la relocalisation des IDPs et à leur ancrage à la membrane.

1.1.2.3. Auto-assemblage et agrégation

Une caractéristique très intéressante des IDPs est leur capacité à s'auto-associer. Cette propriété leur vient de la faible complexité de leur séquence en acides aminés. De plus en plus d'IDPs étant capables de s'assembler en des structures complexes sont découvertes. Ces auto-assemblages sont parfois fonctionnels⁵⁹. Ils peuvent également entraîner la survenue de pathologies.

Par exemple, l' α -synucléine, une protéine cytosolique désordonnée, exprimée dans les cellules neuronales humaines, est capable de s'associer de manière réversible à des lipides pour stabiliser une conformation en hélice⁶⁰. Malheureusement, il arrive que cette association avec des lipides entraîne un défaut de structuration et que la protéine s'agrège en formant des fibres riches en feuillets β ⁶¹. Ces agrégats sont les principaux composants des corps de Lewy, associés à des maladies diverses appelées synucléopathies, telle que la maladie de Parkinson⁶². La fibrillation de type amyloïde de l' α -synucléine est donc un comportement pathologique de la protéine. Le même phénomène survient avec le peptide A β -40 dans le cadre de la maladie d'Alzheimer.

Pour autant, toutes les formes de fibres amyloïdes ne sont pas pathogènes, il a d'ailleurs été montré que de nombreuses protéines sont capables, si elles se trouvent dans les conditions adéquates, de former des structures de type amyloïde⁶³. Par exemple, la protéine Pmel17 s'associe en fibres amyloïdes de manière à accélérer la formation de mélanine dans les cellules de mammifères⁶⁴.

Très récemment, différentes études ont montré que les IDPs étaient capables de s'auto-assembler pour former des organelles sans membrane⁶⁵. Ces transitions de phase jouent de nombreux rôles et de plus en plus d'équipes démontrent leur importance dans des mécanismes de régulation divers et variés. Elles sont observées lorsque deux espèces chimiques distinctes se repoussent en étant diluées dans un milieu aqueux⁶⁶. De cette manière, les protéines impliquées dans ces transitions de phase se trouvent dans un état d'énergie libre inférieure à celle de leur état monomérique en solution⁶⁷. Les premières transitions de phase résultant de l'association de protéines entre elles ont été mises en évidence en conditions de stress. En effet, ces conditions de stress entraînent un mauvais repliement des protéines et ainsi un auto-assemblage qui aboutit à la formation d'une particule protéique de haut poids moléculaire⁶⁸.

Une caractéristique commune aux protéines formant des transitions de phase est la présence de régions riches en répétitions d'acides aminés polaires (glycines, glutamines,

asparagines et sérines), apolaires (proline), positivement chargés (arginine, lysine), négativement chargés (aspartate, glutamate), aromatiques (tryptophane, phénylalanine, tyrosine). Par contre, les acides aminés aliphatiques et hydrophobes sont très peu représentés. Il semble que les protéines impliquées dans les transitions de phase restent désordonnées au sein des organelles^{69,70}.

Les transitions de phase sont des phénomènes très sensibles aux conditions physico-chimiques du milieu extérieur. Comme détaillé précédemment, les modifications post-traductionnelles peuvent avoir une influence sur la capacité d'une protéine à former de telles structures. Cependant, une variation de la température, de la concentration en sel ou de la concentration locale de la protéine impliquée dans les transitions de phase permet également de moduler la survenue de ce phénomène. Par exemple, il a été montré que les granules formés par le domaine « low-complexity » de la protéine FUS (RNA-Binding Protein FUS, également nommée « Fused in Sarcoma » ; Figure 6) se dissipent avec l'augmentation de la température, mais qu'ils ont une plus forte tendance à se former lorsqu'on augmente la concentration en sel et également la concentration locale en FUS⁶⁹. Récemment, il a également été montré que la protéine FUS n'est plus capable de réaliser ce type de structure lorsque le résidu arginine du troisième motif « RGG » est méthylé⁷¹.

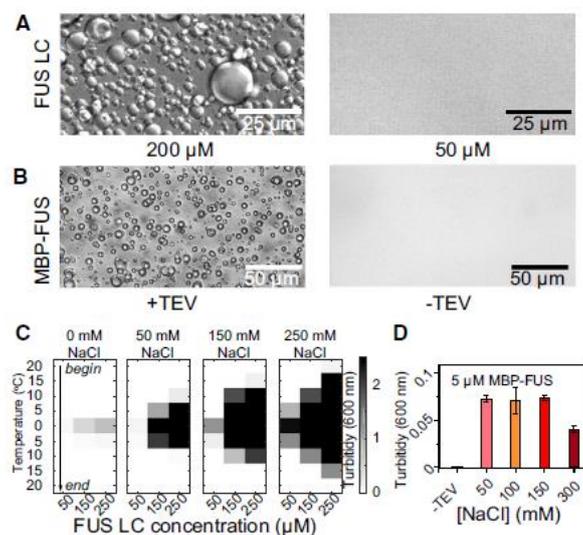


Figure 6 : Influence des conditions physico-chimiques sur les transitions de phase observées avec le domaine « low complexity » (LC) de la protéine FUS avec ou sans coupure de son étiquette de purification⁶⁹. (A) : Image de microscopie par interférences différentielles montrant la formation de transitions de phase par le domaine LC de FUS à haute concentration (gauche) mais pas à faible concentration (droite). (B) : Images de microscopie par interférence différentielle montrant la formation de transitions de phase par la protéine FUS entière après coupure de l'étiquette MBP grâce à la protéase TEV (gauche) mais pas lorsque l'étiquette est présente (droite). (C) : Diagramme montrant par densité optique la capacité du domaine LC de FUS à former des transitions de phase en fonction de la température et de la concentration en protéine. Ces tests ont été réalisés à différentes concentrations de NaCl. (D) : Après coupure de l'étiquette MBP avec la protéase TEV, la concentration en NaCl ne modifie pas la capacité de la protéine FUS entière à former des transitions de phases. Une concentration en NaCl supérieure à 150 mM diminue tout de même ce phénomène.

Les transitions de phase ont souvent été associées à des pathologies, par exemple la cataracte est une pathologie entraînant un changement d'état des protéines du cristallin d'un état clair à une phase opaque, empêchant le passage de la lumière. Par ailleurs, les transitions de phase peuvent évoluer et former des hydrogels⁷² dérivant vers une polymérisation de type amyloïde⁷³. Enfin, récemment, une équipe a suggéré que la formation de transitions de phase par la protéine Tau est une des prémices de l'agrégation de la protéine observée dans la maladie d'Alzheimer⁵⁰.

Cependant, de plus en plus d'études suggèrent que ces transitions de phase liquide-liquide peuvent être fonctionnelles et nécessaires⁷⁴. Notamment, une équipe a étudié la dynamique des granules P au sein des corps riches en ARN. Les granules P montraient une capacité à fusionner entre eux en adoptant un état de transition de phase liquide-liquide⁷⁵. Différents rôles ont été mis en évidence pour de telles structures. On leur attribue en effet la capacité de se lier à des acides nucléiques⁷⁵, à des protéines associées à l'ARN⁷⁶, de servir de cargos protéiques pour permettre une traduction localisée d'un ARNm⁷⁷ ou encore d'augmenter la concentration locale des protéines impliquées⁶⁹.

Enfin, les protéines désordonnées sont également source de nombreux challenges lors de leur étude. Lorsque leur concentration est trop importante, les contacts défavorables s'amplifient et une agrégation aspécifique de la protéine est alors observée. L'encombrement cellulaire peut également être une source de contacts défavorables et le repliement d'une protéine peut être vu comme une stratégie de la cellule visant à maintenir une protéine soluble⁷⁸. Par ailleurs, les IDPs étant particulièrement accessibles au solvant, elles encourent un risque accru d'oxydation, entraînant la formation de ponts disulfures non natifs intra- ou intermoléculaires pouvant mener à de l'agrégation. De la même manière, leur accessibilité au solvant accroît le risque de dégradation par des protéases. Une IDP dégradée peut alors se comporter différemment de la protéine native et ce phénomène peut promouvoir l'agrégation anarchique de la protéine. D'extrêmes précautions sont donc nécessaires lorsqu'une étude structurale sur des protéines dépliées doit être menée.

1.1.2.4. Rôles

Les IDPs sont impliquées dans de nombreuses fonctions cellulaires. Dans cette partie je vais m'intéresser plus particulièrement aux IDPs impliquées dans la régulation de la transcription des gènes, dans le remodelage de la chromatine ainsi que dans la signalisation cellulaire. Je vais détailler plusieurs exemples de protéines en montrant leur impact dans les rôles biologiques précités.

I.1.2.4.1 Les protéines désordonnées impliquées dans la régulation de la transcription des gènes

L'interaction du facteur de transcription HIF1- α (Hypoxia Inducible Factor 1- α) avec CBP/p300 est une bonne illustration de l'importance des régions désordonnées dans la régulation de la transcription des gènes. En effet, la protéine CREB Binding Protein (CBP) ainsi que son paralogue p300 sont des co-activateurs transcriptionnels. Ils possèdent une fonction histone-déacétylase leur conférant un rôle potentiel dans le remodelage de la chromatine mais également dans la modification de facteurs de transcription. Par ailleurs, ils possèdent la capacité d'agir comme chaperonne en recrutant un ensemble de protéines impliquées dans la machinerie de transcription. Le facteur de transcription HIF1- α se lie au domaine TAZ1 de CBP/p300 et régule ainsi la réponse à l'hypoxie⁷⁹. Le facteur HIF1- α est désordonné lorsqu'il est sous forme libre. Lors de son interaction avec CBP, il adopte une structuration en hélice α et s'enroule autour de CBP, créant ainsi une surface d'interaction de très grande taille, menant à une affinité très importante de 7 nM^{80,81}, une affinité qui ne pourrait pas être atteinte si cette interaction était effectuée avec des protéines structurées de tailles similaires. Ce mode de liaison illustre parfaitement la compensation enthalpie-entropie caractéristique des protéines désordonnées.

I.1.2.4.2 Les protéines désordonnées jouant un rôle dans le remodelage de la chromatine

Les histones H1 sont impliquées dans la compaction de la chromatine. Elles se fixent à la surface externe du nucléosome, près du point d'entrée/sortie de l'ADN⁸². Les histones H1 ont trois domaines distincts : un domaine N-terminal de 20 à 35 acides aminés, un domaine globulaire central d'environ 80 acides aminés ainsi qu'un long domaine C-terminal d'environ 100 acides aminés⁸³. Les domaines terminaux sont intrinsèquement désordonnés et se structurent lors de la liaison à l'ADN⁸⁴⁻⁸⁶. Le domaine C-terminal déplié de H1 est impliqué dans la condensation de la chromatine via son association à l'ADN et la neutralisation des charges situées sur l'ADN⁸⁷. H1 est phosphorylée au cours du cycle cellulaire par des kinases dépendantes des cyclines (CDKs) sur des séquences consensus (S/T)-P-X-(K/R). Le niveau de phosphorylation est au plus bas en phase G1, puis augmente au cours des phases S et G2 pour être maximal en métaphase, ce niveau décroît ensuite rapidement^{88,89}. La phosphorylation partielle du domaine C-terminal de H1 réduit considérablement la capacité de la chromatine à s'agréger et empêche ainsi la formation d'assemblages d'ordres supérieurs. La phosphorylation complète permet de restaurer cette capacité. Ainsi H1 est peu phosphorylée en interphase avec une chromatine relâchée et est hyperphosphorylée durant la métaphase avec une chromatine condensée. Le phénomène menant à un relâchement de la chromatine peut s'expliquer par le fait qu'une phosphorylation partielle du domaine C-terminal déplié de H1 entraîne une augmentation de structure β ainsi qu'une diminution des hélices α et des boucles (données de spectroscopie infra-rouge⁹⁰). La diminution du potentiel électrostatique

du domaine C-terminal déplié de H1 proviendrait de sa structuration en feuillet β . En effet, 70% des lysines du domaine C-terminal de H1 sont en doublet qui seraient projetées dans des directions opposées dans le feuillet β . Ceci entraînerait une diminution d'affinité lors de l'interaction avec l'ADN⁹¹, et donc une chromatine moins condensée.

1.1.2.4.3 Protéines désordonnées impliquées dans la signalisation cellulaire

La protéine p120 est un exemple de protéine jouant un rôle dans la signalisation cellulaire en intervenant dans l'adhésion cellule-cellule. En effet, p120 se lie à la queue désordonnée cytoplasmique des cadhérines⁹². La région cœur de la queue des cadhérines apporte la spécificité à cette interaction et se lie très fortement à p120 grâce à une interface structurée statique. Par contre, une région N-terminale flanquante interagit faiblement et de manière dynamique avec p120 en créant des fluctuations entre les états libre et lié. Si cette interaction dynamique contribue peu à l'affinité de la liaison, elle joue un rôle essentiel dans la détermination de l'avenir cellulaire des cadhérines en masquant un motif leucine-leucine qui est nécessaire pour démarrer l'internalisation par endocytose médiée par les clathrines.

Un second exemple de protéines intrinsèquement désordonnées ayant une fonction de signalisation est celui des kinases RIP1 et RIP3 (pour *Receptor Interacting Protein Kinase*). Ces deux dernières s'assemblent pour former un complexe de signalisation requis pour la nécrose programmée qui est une fibrille amyloïde hétérodimérique⁹³. La formation de cette fibrille est médiée par de courtes séquences amyloïdogéniques RHIM (RIP homotypic interaction motif), situées dans des régions intrinsèquement désordonnées de RIP1 et RIP3. Il a été suggéré que les séquences RHIM sont cachées dans l'état inactif et deviennent exposées lors de la formation du nécrosome (sous forme de fibrille amyloïde), ce qui permet l'activation de ces kinases par phosphorylation⁹³.

1.2. Enveloppe nucléaire

Les cellules eucaryotes se différencient des cellules procaryotes par la présence d'un noyau. Le noyau a pour but de stocker l'information génétique et de la protéger. Il contient un ensemble de protéines formant les machineries de réplication des chromosomes, de compaction, de protection et de réparation de l'ADN ainsi que les complexes régulant la transcription des gènes. Habituellement, les cellules eucaryotes ne comportent qu'un seul noyau, cependant, les érythrocytes ne contiennent pas de noyau et les cellules musculaires en possèdent plusieurs. La séparation entre le noyau et le cytoplasme est réalisée par une double membrane appelée enveloppe nucléaire. Cette enveloppe (Figure 7) contient des complexes protéiques appelés pores nucléaires, permettant les échanges entre les milieux intra- et extranucléaires.

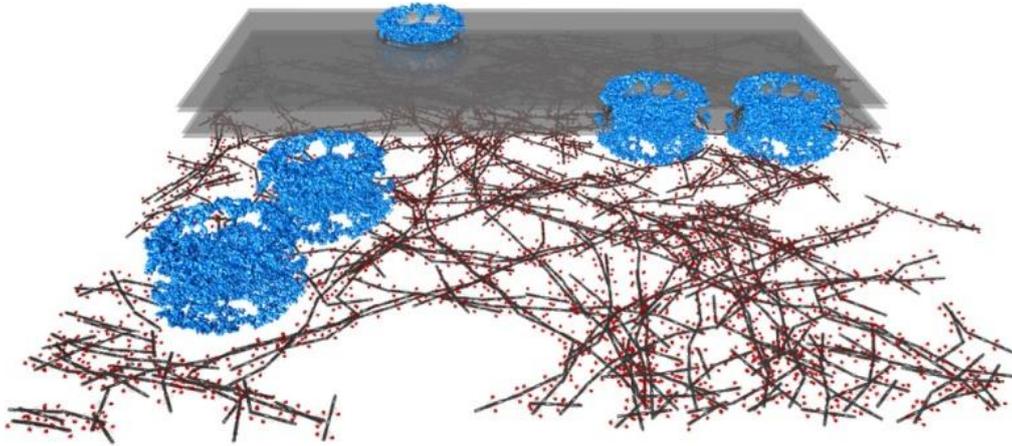


Figure 7 : Modèle de l'enveloppe nucléaire d'une cellule somatique de mammifères, construit à partir de données de cryoélectrotomographie⁹⁴. Les membranes nucléaires externes et internes sont visibles en gris. Les pores nucléaires traversant l'enveloppe nucléaire sont représentés en bleu. Le réseau de lamines est représenté par des filaments noirs. Les domaines globulaires des lamines (IgFold) sont représentés par des points rouges.

1.2.1 Présentation générale de l'enveloppe nucléaire

L'enveloppe nucléaire est composée de deux bicouches phospholipidiques appelées membranes nucléaires interne et externe ainsi que de nombreuses protéines associées à ces bicouches. L'enveloppe nucléaire interne qui va nous intéresser comporte différents types de protéines :

- D'une part, les protéines composant les pores nucléaires qui traversent l'enveloppe nucléaire (Figures 7 et 8). Ces complexes facilitent les échanges de molécules entre le noyau et le cytoplasme.
- D'autre part, les protéines ancrées à la membrane interne par un domaine transmembranaire (Figure 8). Parmi ces protéines, il existe les protéines à domaine LEM telles que LAP2 (pour Lamina-Associated Protein 2), émerine et MAN1.

- Enfin, un réseau de protéines qui tapisse l'enveloppe nucléaire interne (Figures 7 et 8). Il est composé de protéines appelées lamines. Ces lamines peuvent être de 2 types : type A et type B. Les lamines s'associent en dimères pour ensuite s'assembler en filaments protéiques d'environ 3 nm de diamètre (Figure 7).

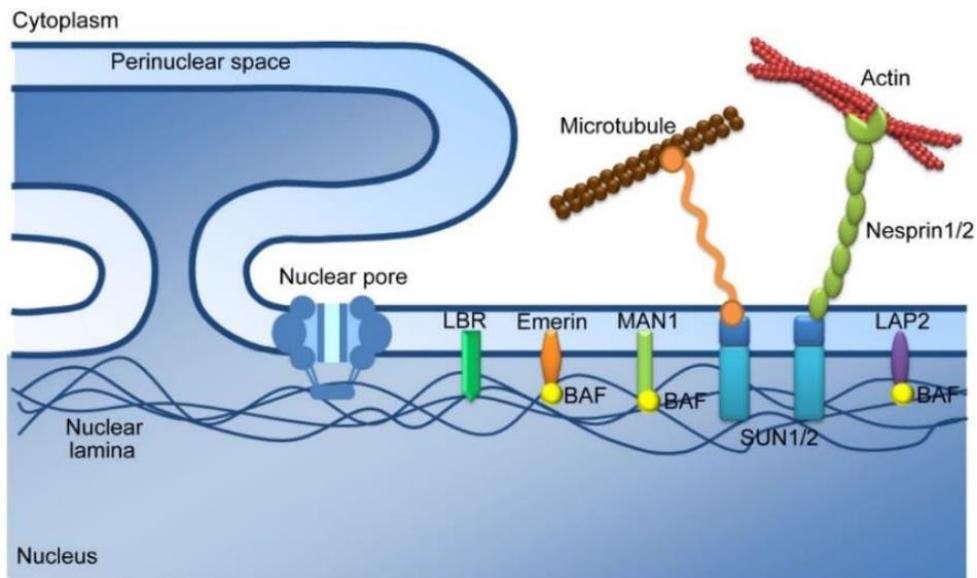


Figure 8 : Architecture de l'enveloppe nucléaire⁹⁴. Les protéines SUN1/2 et nesprines s'associent dans l'espace périnucléaire pour créer un lien physique entre le cytosquelette (ici représenté par le microtubule et filament d'actine) et le nucléosquelette (formé par les lamines et aussi nommé lamina nucléaire). Les lamines tapissent la membrane interne du noyau et interagissent avec de nombreuses protéines ancrées à cette membrane, comme par exemple les protéines à domaine LEM (LAP2, Emerine, MAN1). Ces interactions sont stabilisées par la protéine BAF associée à l'ADN.

La composition protéique de l'enveloppe nucléaire interne diffère selon le type cellulaire étudié. Si une partie des protéines contenant des domaines transmembranaires (NETs pour Nuclear Envelope Transmembrane) est ubiquitaire, une autre partie est spécifique du tissu considéré. A ce jour, 200 NETs ont été découvertes suite à l'étude de trois types cellulaires différents via des méthodes de protéomique. 67 ont été mises en évidence chez des cellules de foie de rats⁹⁵, 29 ont été trouvées chez des cellules musculaires squelettiques de rats⁹⁶ et enfin 87 étaient présentes chez des leucocytes humains⁹⁷. Au sein de ces NETs, certaines protéines possèdent d'importantes régions désordonnées. Je vais, dans la partie suivante, détailler les protéines NETs émerine (que j'ai étudiée lors de ma thèse) et SUN1 (qui est un partenaire de la région C-terminale de la lamine A, aussi étudiée au cours de ma thèse).

1.2.2 Les protéines ancrées à l'enveloppe nucléaire

1.2.2.1 L'émerine

L'émerine est codée par le gène *EMD* se trouvant sur le chromosome X. C'est une protéine de 254 acides aminés prédite comme étant majoritairement dépliée. Elle est ubiquitaire mais montre de forts niveaux d'expression dans les tissus musculaires cardiaques et squelettiques, plus particulièrement au niveau de la membrane nucléaire interne⁹⁸. Elle contribue à l'interface entre le nucléosquelette et la chromatine, cependant, les fonctions de l'émerine sont difficilement identifiables car elle agit en collaboration avec de nombreuses autres protéines de l'enveloppe nucléaire interne. Notamment, dans les myoblastes, les fonctions de l'émerine se superposent à celles des autres protéines à domaine LEM comme MAN1⁹⁹. En revanche, l'absence d'émerine entraîne une pathologie musculaire et cardiaque : la dystrophie musculaire d'Emery-Dreifuss (EDMD¹⁰⁰).

L'émerine comporte en N-terminal un domaine LEM (pour LAP2, Emerine, MAN1) allant des acides aminés 1 à 49¹⁰¹. La structure du domaine LEM de l'émerine a été résolue par Résonance Magnétique Nucléaire¹⁰², de plus, la structure de ce domaine en complexe avec la protéine BAF est également connue¹⁰³ (Figure 9).

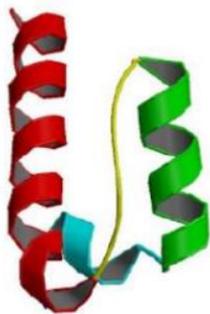


Figure 9 : Structure du domaine LEM de l'émerine résolue par RMN^{103,104}. Ce domaine est composé d'un tour d'hélice de 3 résidus (bleu) suivi de deux hélices parallèles (verte et rouge) connectées par une longue région désordonnée (jaune).

Ce domaine est composé majoritairement de deux hélices α parallèles séparées par un fragment étendu (Figure 9). Après ce domaine, l'émerine comporte une grande région prédite comme intrinsèquement désordonnée ainsi qu'un domaine transmembranaire (résidus 223 à 243) permettant son ancrage à la membrane nucléaire interne.

Il a été montré que la protéine émerine intervient dans la réponse au stress mécanique via des événements de phosphorylation. En effet, une étude a été réalisée sur des noyaux isolés qui subissaient un stress mécanique via des pinces magnétiques poussant sur des billes recouvertes d'anticorps anti-nesprine 1. Après application d'impulsions successives sur ces noyaux, une diminution du déplacement engendré par ces impulsions a été observée. Ceci signifie que la résistance du noyau au stress mécanique augmente¹⁰⁴. Il a été observé que plusieurs protéines nucléaires sont phosphorylées après application de ce stress mécanique. L'émerine est très fortement phosphorylée, et la mutation de 2 tyrosines phosphorylées (Y74

et Y95) perturbe l'adaptation du noyau au stress mécanique. Enfin, il a été montré que la kinase Src est responsable de la phosphorylation de ces tyrosines dans ces conditions.

Il a aussi été montré que l'émerine régule la voie de signalisation Wnt/ β -caténine. Cette voie de signalisation est impliquée dans le développement embryonnaire ainsi que dans le renouvellement des tissus adultes¹⁰⁵. Elle passe par un cofacteur transcriptionnel appelé β -caténine. En l'absence de Wnt, la β -caténine cytoplasmique subit une dégradation via le protéasome. En présence de Wnt, un complexe protéique contenant les protéines frizzled (une famille de récepteurs à protéines G¹⁰⁶, mais également les protéines liées aux récepteurs de lipoprotéines LRP5 et 6^{107,108} est formé, qui entraîne l'accumulation de la β -caténine, qui est alors transloquée dans le noyau pour se lier au facteur de transcription TCF (pour T Cell Factor) et induire l'expression de gènes cibles. Bien que la formation d'un complexe trimérique entre Wnt et les récepteurs de la famille frizzled et LRP5/6 n'a pas été démontrée, leur expression en surface est nécessaire au déclenchement du signal Wnt. Des études récentes montrent l'implication de l'émerine dans la régulation du signal Wnt/ β -caténine. Premièrement, la β -caténine se lie à un domaine APC-like (Adenomatous Polyposis Coli) de l'émerine (résidus 169-180), et la surexpression d'émerine fusionnée à la GFP dans des cellules HEK293 entraîne une accumulation cytoplasmique de β -caténine et inhibe son activité. De la même manière, l'expression d'un mutant de l'émerine ne comportant pas le domaine APC-like provoque une augmentation de la β -caténine nucléaire ainsi que de son activité transcriptionnelle¹⁰⁹. Deuxièmement, des fibroblastes n'exprimant pas d'émerine montrent une croissance augmentée ainsi qu'une accumulation nucléaire de la β -caténine et une activité transcriptionnelle accrue¹⁰⁹. Ainsi, l'interaction entre la β -caténine et l'émerine pourrait inhiber la voie Wnt- β -caténine en restreignant l'accès de la β -caténine au noyau.

1.2.2.1.1 Auto-assemblage de l'émerine

Il semble que les fonctions de l'émerine soient en partie médiées par son auto-assemblage. Les détails moléculaires de cet auto-assemblage ont été analysés par l'équipe de K. Wilson. Un fragment d'émerine 1-221 a été immobilisé sur une colonne de nickel grâce à une étiquette histidine, et différents fragments d'émerine portant une étiquette GST ont été déposés sur cette colonne¹¹⁰. L'étiquette GST de l'émerine partenaire permettait de détecter l'interaction grâce à un western blot à l'aide d'un anticorps anti-GST. Deux modes d'assemblage ont ainsi été démontrés. L'un, médié par l'association des résidus 170-220 d'une molécule d'émerine avec cette même région d'une seconde molécule d'émerine, et l'autre, mettant en jeu les résidus 1-132 et 170-220 permettant ainsi une interaction intramoléculaire. Ces interactions émerine-émerine seraient potentiellement un moyen de moduler les contacts entre l'émerine et la protéine BAF associée à l'ADN, mais également son interaction avec les lamines. En effet, il est proposé que le domaine LEM puisse être en configuration « up » permettant à l'émerine d'interagir avec elle-même et/ou des partenaires

à l'enveloppe nucléaire tandis que la configuration « down » permettrait l'interaction du domaine LEM avec la protéine BAF¹¹⁰ (Figure 10).

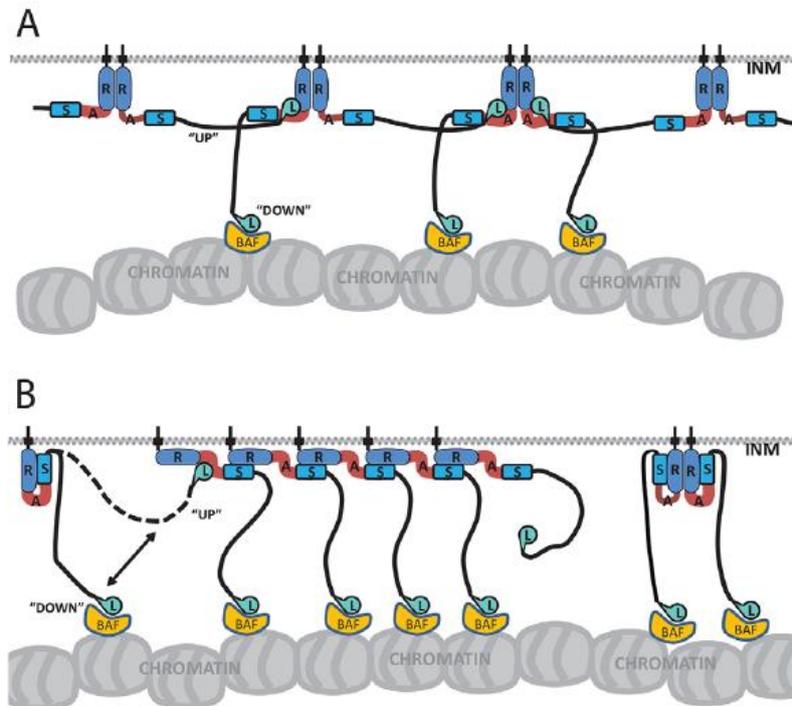


Figure 10 : Modulation du rôle du domaine LEM grâce à l'auto-assemblage de l'émerine¹¹¹. (A) Modèle illustrant l'association intermoléculaire de la région 168-220 (région A R) ainsi que l'association de la région N-terminale comprenant le domaine LEM (noté L) avec les régions 168-186 (notée A) et 187-220 (notée R), en accord avec les résultats des figures 1 et 2 de Berk et al (2014)¹¹¹. (B) Modèle illustrant l'association de la région AR d'une molécule d'émerine avec la région L, la région 147-174 (notée S) et la région R d'une autre molécule d'émerine, en accord avec les résultats de la figure 7 de Berk et al (2014)¹¹¹. La configuration « up » permet à l'émerine de s'auto-assembler et d'interagir avec des partenaires à l'enveloppe nucléaire. La configuration « down » permet à l'émerine d'interagir avec BAF via son domaine LEM.

Par ailleurs, au laboratoire, notre équipe a montré que le fragment d'émerine 1-187 était capable de s'auto-assembler pour former des oligomères de grandes tailles assimilables à des filaments curvilinéaires d'un diamètre d'environ 10 nm¹¹¹. La cinétique de formation de ces filaments a été suivie en utilisant la fluoresscence de la thioflavine-T, une molécule capable de se lier à des structures en feuillets β et d'émettre un signal de fluoresscence lorsqu'elle est excitée à une longueur d'onde spécifique. Cet assemblage sous forme de filaments a par la suite été validé par microscopie électronique à transmission (Figure 11).

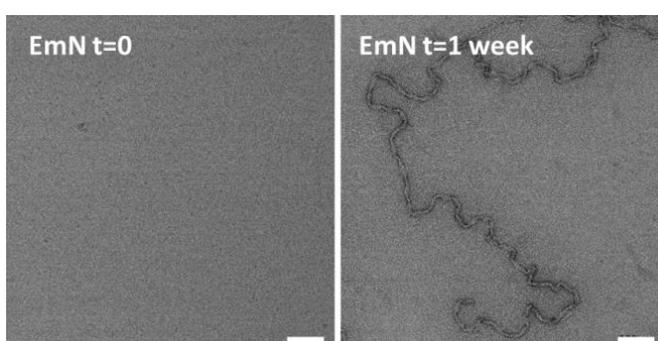


Figure 11 : L'émerine 1-187 (notée EmN) s'auto-assemble pour former des filaments¹¹³. Ces images de microscopie électronique à transmission ont été obtenues à $T = 0$ et $T = 1$ semaine en utilisant comme agent de contraste l'acétate d'uranyl. La protéine était incubée à 293 K et 600 μ M. Les barres blanches représentent 100 nm.

I.2.2.1.2 Interaction émerine-BAF

L'émerine réalise de nombreuses interactions ayant un rôle primordial dans l'assemblage de l'enveloppe nucléaire, sa structure et son interface avec la chromatine. Les régions nécessaires à son auto-assemblage sont essentielles à sa localisation à l'enveloppe nucléaire, ce qui suggère que l'auto-association de l'émerine régule sa localisation permettant l'interaction avec ses partenaires à l'enveloppe nucléaire¹¹⁰. L'émerine interagit avec les protéines SUN1, SUN2 et les nesprines, qui sont responsables de la connexion physique entre l'enveloppe nucléaire et le cytosquelette¹¹². Il a été montré en particulier que le fragment 1-221 de l'émerine interagit avec la région nucléoplasmique de SUN1 comprise entre les résidus 223 et 302 (numérotation de SUN1 murin¹¹²).

De plus, l'émerine contribue à l'interface entre l'enveloppe nucléaire et la chromatine par l'intermédiaire de son interaction avec la protéine BAF, qui lie l'ADN. En configuration « down », c'est-à-dire avec un domaine LEM libre, l'émerine est en lien avec la chromatine par l'établissement de contacts avec BAF¹¹⁰ (Figure 10). La structure du complexe du domaine LEM en interaction avec un dimère de BAF a été résolue par RMN¹⁰³ (Figure 12). Cette interaction possède une constante de dissociation $K_d = 0.6 \mu\text{M}$ déterminée par ITC.

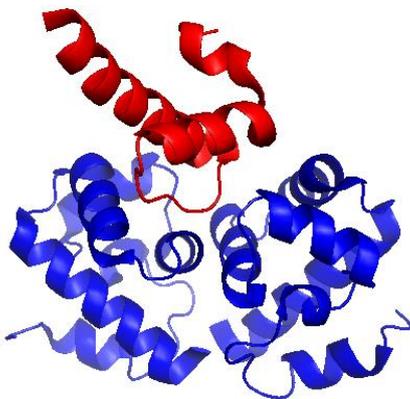


Figure 12 : Structure du domaine LEM de l'émerine lié à un dimère de BAF (code PDB : 2ODG). Le domaine LEM de l'émerine est représenté en rouge, le dimère de BAF, en bleu¹⁰⁴. Le fragment étendu du domaine LEM situé entre ses 2 grandes hélices α constitue le site majeur d'interaction avec le dimère de BAF. L'interaction se fait au niveau de l'interface entre les 2 monomères de BAF.

L'émerine est également capable de moduler les propriétés mécaniques du noyau en interagissant avec le réseau de lamines qui en est le principal élément régulateur. En effet, une étude a montré dès les années 2000 à partir d'expériences de double hybride en levure, que l'émerine se liait directement aux lamines A et C avec une préférence notable pour la lamine A¹¹³. Cette étude a révélé que les acides aminés 384-566 des lamines A/C étaient nécessaires à la liaison avec le fragment de l'émerine 1-221. De plus, la préférence de l'émerine pour la lamine A suggérait une influence de la région 567-646, spécifique de la lamine A, sur l'interaction. Enfin, il a été montré que la protéine BAF était capable de favoriser l'interaction entre l'émerine et la lamine A¹¹⁴.

1.2.2.2. SUN1 et le complexe LINC

La protéine SUN1 ainsi que son paralogue SUN2 font partie des protéines à domaine SUN (Sad1/UNC-84). SUN1 participe à un complexe faisant le lien physique entre le nucléosquelette et le cytosquelette, il s'agit du complexe LINC (Linkers for Nucleoskeleton and Cytoskeleton). Ce complexe est chargé de transmettre les informations mécaniques du cytoplasme au noyau¹¹⁵. La réponse aux stress mécaniques est primordiale pour de nombreuses fonctions cellulaires telles que la migration ou encore la différenciation cellulaire. Le complexe LINC joue un rôle important dans le positionnement du noyau dans la cellule en partie grâce à son interaction avec le réseau de filaments d'actine¹¹⁶. Les protéines à domaine SUN composent la partie interne du complexe LINC et sont ancrées à la membrane nucléaire interne. Leur région N-terminale nucléoplasmique est prédite comme déstructurée et contient une région riche en sérines. Leur région C-terminale se situe entre les deux membranes nucléaires, et contient deux régions coiled-coil entourées de régions désordonnées, ainsi qu'un domaine SUN.

1.2.2.2.1 Auto-assemblage des protéines à domaine SUN et interaction avec les protéines à domaine KASH

Les protéines à domaine SUN s'auto-assemblent par leur domaine SUN sous forme de trimères (Figure 13), afin de pouvoir interagir avec le domaine KASH (Klarsicht/ANC-1/SYNE homology) des nesprines dans l'espace entre les deux membranes nucléaires. La structure du trimère de SUN2 en interaction avec un peptide KASH a été résolue par cristallographie^{117,118}. Dans cette structure, le trimère de SUN2 est capable d'interagir avec trois peptides KASH et chaque peptide KASH se fixe dans un sillon formé par deux domaines SUN. L'interaction SUN2/KASH est stabilisée par un pont disulfure reliant le domaine SUN au fragment KASH. Il est à noter que l'interaction entre la protéine SUN2 et le domaine KASH requiert à la fois le domaine coiled-coil de SUN2 (donc la région 335-717) mais également le domaine SUN car l'absence de l'une de ces deux régions ne permet pas d'observer l'interaction¹¹⁷.

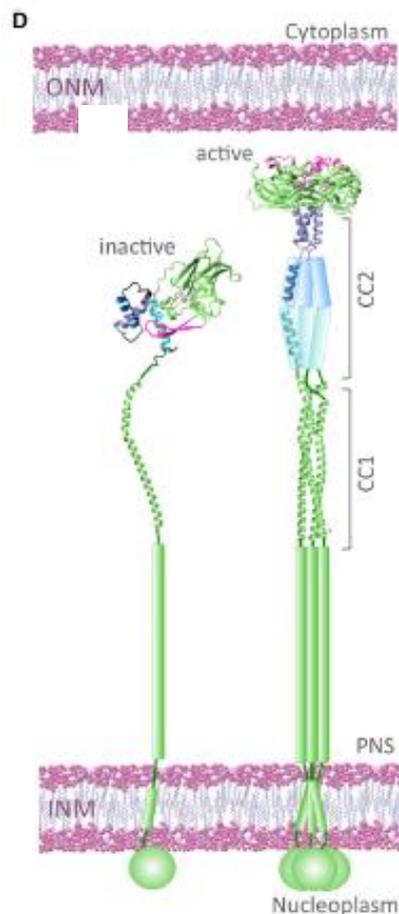


Figure 13 : Modèle de l'activation de SUN2 pour la liaison de KASH dans l'espace péri-nucléaire. La trimérisation de SUN2 est régulée par les domaines coiled-coil 1 et 2 (CC1 et CC2 respectivement)¹²⁰. La structure de SUN2 inactive, avec une région CC2 globulaire, est basée sur la structure de SUN2 murin (5ED8); la structure de SUN2 active, avec une région CC2 allongée, est basée sur la structure de SUN2 homotrimérique humain (4DXT).

Une seconde structure a été résolue chez la souris avec un fragment plus long de SUN2 contenant deux hélices supplémentaires. Dans cette structure, le motif d'interaction avec KASH s'insère dans un fagot de trois hélices maintenant ainsi le domaine SUN monomérique et ne permettant pas son interaction avec le peptide KASH. SUN2 peut donc prendre une forme monomérique inactive mais également une forme trimérique active (Figure 13).

Au contraire de SUN2, SUN1 a peu été étudiée par des méthodes de biologie structurale. L'analyse du mode d'oligomérisation de SUN1 par des méthodes de gel natif ou par « crosslinking » suivi de gel SDS-PAGE a montré que l'auto-assemblage de SUN1 ne semble pas être limité à un trimère. En effet, il semble que l'assemblage du domaine coiled-coil mais également les ponts disulfures inter-chaines soient capables d'entraîner la formation d'oligomères de taille plus conséquente¹¹⁹. Une étude combinant des approches bioinformatiques de simulation en dynamique moléculaire ainsi que de docking réalisés à partir d'un modèle par homologie basé sur la structure de SUN2, soutient l'hypothèse selon laquelle la région C-terminale de SUN1 est capable de trimériser¹²⁰. Cette région de SUN1 serait aussi capable de former des interactions latérales trimère-trimère¹²⁰⁻¹²². Cette association de trimères de SUN1 permettrait la formation de clusters de quatre trimères au sein de l'enveloppe nucléaire. Récemment, l'état oligomérique des régions C-terminales des protéines SUN1 et SUN2 a été mesuré par spectroscopie de fluctuation de fluorescence dans des cellules U2OS¹²³. Les protéines SUN2 ont été observées en équilibre entre monomère et trimère, alors que les protéines SUN1 ont été observées soit en monomère/dimère/tétramère, soit en monomère/trimère/hexamère.

1.2.2.2 Interaction entre SUN1 et la lamine A

Des expériences de double hybride en levure ont été réalisées en utilisant le domaine C-terminal de la lamine A comme proie et en faisant une recherche systématique de partenaires. Ainsi, la protéine SUN1 a été identifiée comme un partenaire de la lamine A¹²⁴. Pour confirmer cette capacité de SUN1 à lier la lamine A, des immunoprécipitations ont été effectuées. Des cellules U2OS ont été co-transfectées avec un plasmide codant pour la protéine SUN1 liée à une étiquette HA et la GFP d'une part ou la GFP-lamine A d'autre part. Des anticorps anti-HA ont ensuite été utilisés et ont été capable de co-immoprécipiter GFP-lamine A mais pas la GFP seule. Par la suite, des expériences de pulldown réalisées à l'aide de deux constructions de SUN1 possédant une étiquette MBP (résidus 1-355 ou 450-913) avec laquelle la lamine A marquée au ³⁵S (transcrite-traduite *in vitro*) a été incubée, ont montré que la région N-terminale de SUN1 était responsable de l'interaction avec la lamine A (Figure 14). Une expérience inverse mettant en jeu le domaine C-terminal de la lamine A (résidus 389-664) fusionné à la GST et la région N-terminale de SUN1 n'a montré qu'une faible liaison des deux partenaires, suggérant la nécessité d'une région contenue dans le domaine tête ou coiled-coil de la lamine A pour une interaction optimale (ou un effet des étiquettes qui

différent entre les 2 expériences). Par ailleurs, des essais ont également été effectués sur la lamine C ainsi que sur la lamine B et n'ont montré aucune interaction entre SUN1 et ces dernières. L'interaction SUN1-lamine semble donc être spécifique à la lamine A, ce qui suggère un rôle prépondérant de la région C-terminale de la lamine A dans l'interaction avec SUN1.

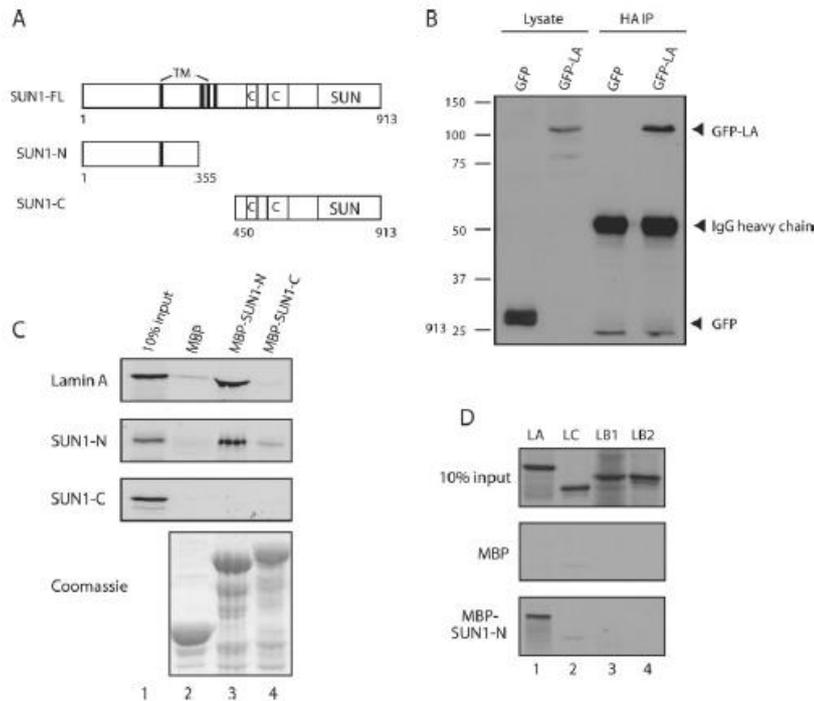


Figure 14 : SUN1 interagit via son domaine N-terminal à la lamine A spécifiquement¹²⁴. (A) Constructions de SUN1 utilisées pour les pulldowns. (B) Coimmunoprecipitation de HA-SUN1 avec GFP-lamine A ou GFP. (C) Pulldowns entre la région N-terminale ou C-terminale de SUN1 étiquetée avec la MBP et la lamine A marquée au ³⁵S. (D) Pulldowns entre MBP-SUN1 et les différents types de lamines produites par traduction in vitro. Les protéines liées sont révélées par autoradiographie.

1.3. La Lamina nucléaire

La lamina nucléaire est un réseau de filaments protéiques situé dans le noyau et soulignant la membrane nucléaire interne, elle est présente chez tous les métazoaires. Elle se compose de différentes protéines appelées lamines. Les lamines peuvent être classées en deux catégories : les lamines de type A, comportant la lamine A et la lamine C, ainsi que les lamines de type B comportant les lamines B1 et B2. La plupart des métazoaires ne comportent que des lamines de type B tandis que les vertébrés comportent des lamines de type A et de type B¹²⁵. Les lamines font partie des filaments intermédiaires de type V, elles s'associent en formant des filaments tétramériques d'une épaisseur de 3.5 nm comme le montre une récente étude réalisée par cryo-électrotomographie¹²⁶. Les lamines de type B sont farnésylées et ancrées à la membrane nucléaire¹²⁷ tandis que les lamines de type A matures, qui ne sont plus farnésylées, interagissent avec les lamines B, et se trouvent le long de l'enveloppe nucléaire mais également dans le nucléoplasme. Bien que les lamines A et B puissent interagir entre elles, il apparaît que les lamines préfèrent former des réseaux filamentaires indépendants qui

restent en contact entre eux¹²⁸. Enfin, ce réseau filamentaire se désassemble de manière réversible durant la mitose¹²⁹.

1.3.1 La lamine A

La lamine A est codée par le gène *LMNA* qui par épissage alternatif peut coder pour la lamine A, la lamine A Δ 10, la lamine C ou encore la lamine C2. La lamine A est produite sous forme de prélamine A et a besoin de plusieurs étapes pour donner la lamine A mature (Figure 15). En effet, la lamine A possède un motif CAAX, c'est-à-dire une cystéine suivie de deux résidus aliphatiques et enfin d'un résidu aléatoire. Le motif CAAX subit une étape de farnésylation sur la cystéine. Il s'agit d'une réaction catalysant l'ajout d'un groupement farnésyl à partir d'un précurseur présent naturellement dans la cellule : le farnésylpyrophosphate. Cette réaction est effectuée par une farnésyl-transférase. Le motif CAAX est ensuite clivé pour ne laisser que la cystéine. Ensuite, une étape de carboxyméthylation est réalisée au niveau de la cystéine C-terminale grâce à la protéine Icmt pour isoprénylcystéine carboxyméthyl-transférase. Enfin, la protéine Zmpste24 clive les 15 derniers acides aminés comprenant les groupements carboxyméthyl et farnésyl pour former la lamine A mature. La structure de l'enzyme Zmpste24 en complexe avec le domaine CSIM C-terminal de la lamine A a d'ailleurs été résolue par cristallographie¹³⁰.

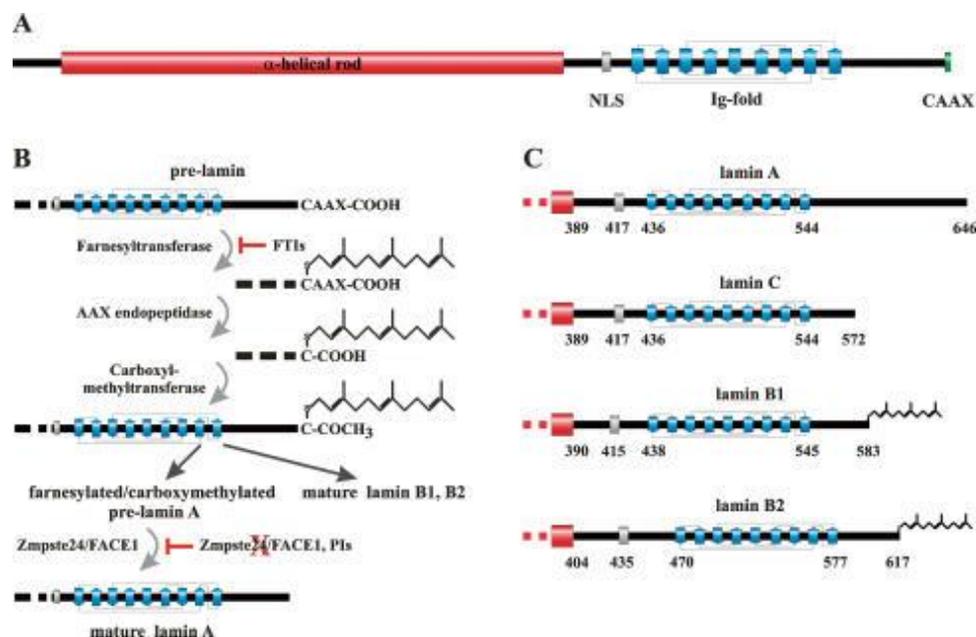


Figure 15 : Etapes de maturation des lamines (d'après Dechat et al., 2008). (A) Schéma représentant la structure d'une lamine. Le domaine coiled-coil est indiqué en rouge et le domaine IgFold est représenté en bleu. Le signal de translocation nucléaire est indiqué en gris (NLS). Le motif de farnésylation est en vert. (B) La prélamine A subit plusieurs modifications chimiques. Elle est farnésylée, ses 3 derniers résidus sont coupés, puis elle est carboxyméthylée. Enfin, l'enzyme Zmpste24 ôte ses 15 derniers résidus. Cette dernière coupure est essentielle car elle retire les modifications post-traductionnelles pour donner la lamine A mature. (C) Représentation schématique des différents produits de maturation des lamines. La lamine C ne possède pas le site de farnésylation. Les lamines B1 et B2 restent constitutivement farnésylées.

1.3.1.1. Structure

Comme toutes les lamines, la lamine A possède une région N-terminale dépliée de 24 résidus, aussi appelée « région tête », suivie d'un domaine de type coiled-coil (ou superhélice α) composé des fragments Coil 1 et Coil 2 et enfin, une région C-terminale appelée « région queue » composée à la fois d'un domaine globulaire de type immunoglobuline (IgFold) et d'un fragment C-terminal déplié. Elle peut s'associer en dimères parallèles grâce à ses domaines Coil 1 et Coil 2¹³¹ (Figure 16). Dans le cas des lamines A et C, les structures 3D des domaines Coil 2 et IgFold sont connues¹³²⁻¹³⁴.

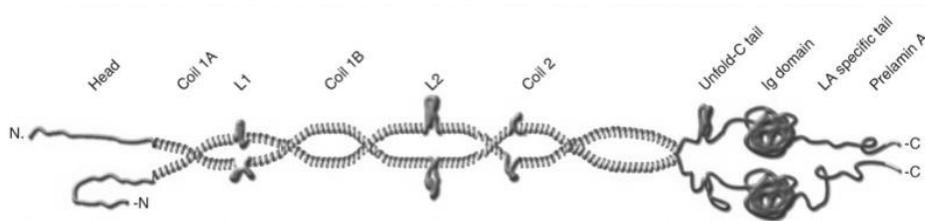


Figure 16 : Représentation schématique du dimère de prélamine A¹³⁰. Le dimère de prélamine A s'assemble en formant des structures de type hélice super enroulée ou super hélice α ou encore coiled-coil (notées ici Coil 1A, Coil 1B et Coil 2). Les extrémités N-terminales et C-terminales sont essentiellement désordonnées. Cependant, dans la région C-terminale se trouve le domaine globulaire au repliement de type immunoglobuline (noté Ig domain ici ou IgFold dans la thèse).

Les dimères de lamine A peuvent alors s'associer de manière parallèle à travers des interactions queue-tête. Puis ces oligomères s'assemblent latéralement de manière antiparallèle pour former un filament de lamines. Ces filaments ont été récemment caractérisés par cryo-tomographie au sein de noyaux purifiés dont la chromatine avait été préalablement digérée¹²⁶ (Figure 17). Leur diamètre est d'environ 3.5 nm. Les domaines Igfold se trouvent distribués tous les 20 nm sur les filaments. Ces filaments forment le réseau de lamines A, qui interagit avec le réseau de lamines B.

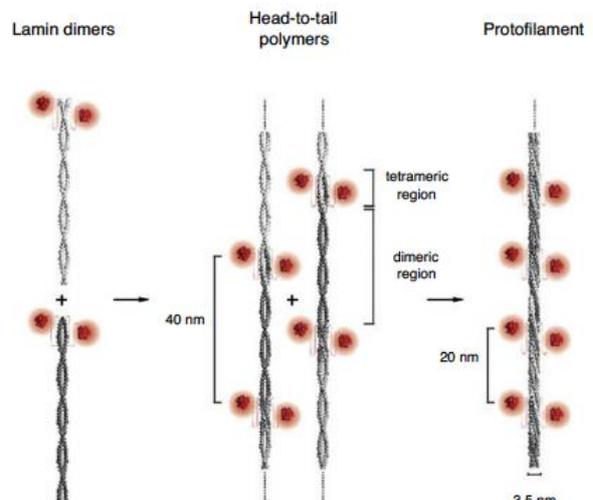
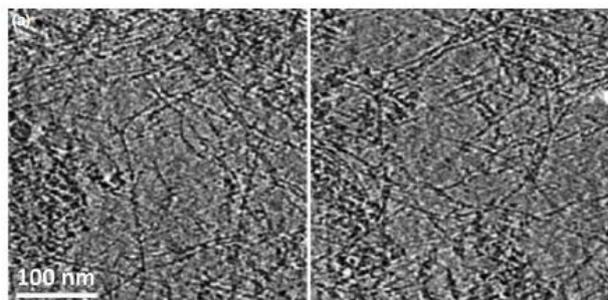


Figure 17 : Illustration de l'assemblage des dimères de lamines en filaments¹³⁴. (Haut) Filaments observés par cryo-tomographie électronique. (Bas) Voie d'assemblage proposée. Les régions en coiled coil sont en grises, les domaines Igfold sont en rouge.

1.3.1.2. Fonctions

1.3.1.2.1 Rôles de la lamine A dans la signalisation cellulaire

La lamine A et ses protéines associées sont impliquées dans un grand nombre d'interactions influant sur la transduction de divers signaux cellulaires. En particulier, il semble que les protéines de l'enveloppe nucléaire soient des régulateurs de l'activité d'acteurs de la voie des MAP (Mitogen Activated Protein) kinases.

En effet, des cellules exprimant un mutant de la lamine A H222P, utilisé comme modèle pour décrire la cardiomyopathie dilatée due à un défaut de lamine A, montrent une activité anormalement élevée de ERK et JNK dans ces cellules¹³⁵. De la même manière, une hyperactivité de ERK1/2 est également observée dans les cellules cardiaques de souris déficientes en émerine¹³⁶. Ces résultats concordent avec une observation faite à partir de fibroblastes de patients EDMD ou atteints de cardiomyopathie dilatée possédant des mutations faux-sens de la lamine A¹³⁷. Ces cellules montrent des fonctions altérées de ERK1/2 entraînant un retard de l'activation d'ERK1/2 suivi d'une hyperactivation de la kinase. Ces effets entraînent une augmentation de la prolifération, une migration élevée ainsi que des défauts de polarisation. Il a été montré récemment que ERK1/2 interagit avec la lamine A et colocalise avec des facteurs de transcription à l'enveloppe nucléaire¹³⁸ (Figure 18). Ceci suggère que l'enveloppe nucléaire module directement l'activité d'ERK1/2 et que des altérations de la lamine A pourraient perturber l'enveloppe nucléaire pour affecter directement ce processus.

1.3.1.2.2 Implications de la lamine A dans la régulation de la transcription

La lamine A interagit directement avec plusieurs régulateurs de la transcription comme c-Fos, Rb, SREBP1¹³⁹ et influence ainsi la transcription de gènes. La lamine A se lie par exemple à son partenaire LAP2 α et ce complexe s'associe avec des séquences promotrices dépendantes de E2F par l'intermédiaire de Rb. Ceci conduit à la répression de l'expression des gènes cibles de E2F. Les cellules qui n'expriment pas de lamine A prolifèrent de manière accrue et montrent un défaut d'arrêt de cycle dépendant de TGF- β du fait d'une perte de stabilité et/ou un défaut de phosphorylation de Rb^{140,141}. Ainsi le complexe lamine A / LAP2 α pourrait contrôler l'équilibre entre prolifération et différenciation dans les cellules souches adultes, essentiel pour assurer la régénération des tissus. La lamine A réprime aussi l'activité de AP-1, qui est impliqué dans l'arrêt du cycle et la différenciation, en séquestrant c-Fos à l'enveloppe nucléaire et en interférant de ce fait avec l'hétérodimérisation c-Fos / c-Jun et la liaison d'AP-1 à l'ADN¹⁴² (Figure 18). De la même façon, la liaison de SREBP1 à la lamina contribue à retenir SREBP1 à l'enveloppe nucléaire, ce qui inhibe l'expression de ses gènes cibles comme PPAR- γ , nécessaire à la différenciation des adipocytes¹⁴³. Ces interactions entre lamine A et facteurs

de transcription participent ainsi également au contrôle de l'équilibre entre prolifération et différenciation.

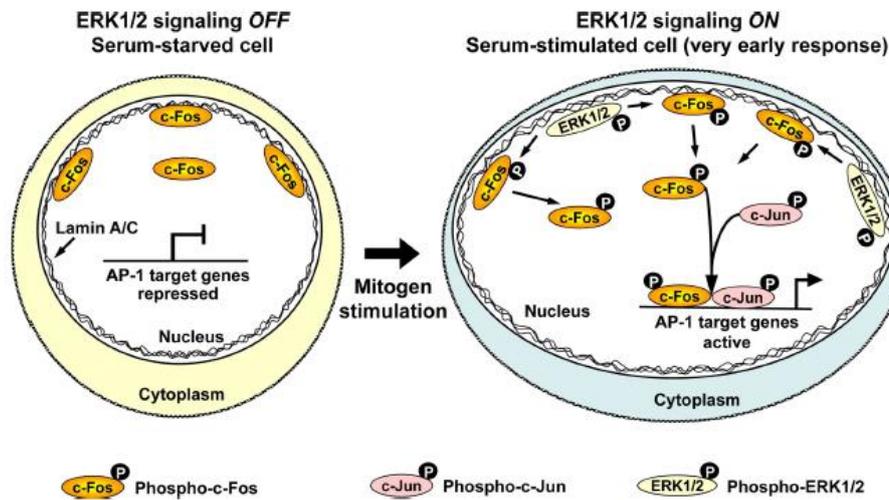


Figure 18 : Régulation de l'activité d'AP-1 via l'interaction entre la lamine A et ERK1/2¹⁴⁴.

I.3.1.2.3 Rôle de la lamine A dans le remodelage de la chromatine et le maintien de la structure du noyau

La lamine A joue un rôle important dans le contrôle de l'expression des gènes de par son implication dans l'organisation et la régulation de la structure de la chromatine. Cette hypothèse est renforcée par des études montrant que la lamine A lie l'ADN, les histones, la protéine BAF associée à l'ADN et aux histones, ou encore les protéines à domaine LEM qui elles-mêmes lient BAF associée à l'ADN et aux histones^{139,144,145}.

Il a été montré que des protéines de l'enveloppe nucléaire relient mécaniquement le nucléosquelette et le cytosquelette via la formation du complexe LINC¹⁴⁶. Ce complexe interagit avec les lamines et l'émerine¹⁴⁷, ce qui contribue au positionnement du noyau ainsi qu'à la réponse du noyau aux stress mécaniques.

Une étude récente a montré que la lamine A/C pouvait induire la formation de câbles apicaux d'actine périnucléaires, responsables de la protection du noyau contre les stress mécaniques exogènes¹⁴⁸. En effet, lorsque les cellules ne possèdent pas de lamine A/C, elles ne comportent pas de câbles apicaux d'actine. Ces câbles apicaux d'actine sont connectés à l'enveloppe nucléaire par le complexe LINC¹⁴⁹. Une telle connexion physique permet la transduction d'un signal entre le cytoplasme et le noyau. La lamine A/C contribue à la formation rapide de ces câbles apicaux d'actine, permettant une résistance accrue aux stress mécaniques et empêchant une déformation importante du noyau¹⁵⁰.

Enfin, des expériences de double-hybride en levure ont montré que la titine, une protéine sarcomérique donnant son élasticité au muscle strié¹⁵¹ et participant à la dynamique des chromosomes, l'expression des gènes et la prolifération cellulaire¹⁵²⁻¹⁵⁴, était en interaction avec les lamines de type A et B¹⁵³. La surexpression du domaine de la titine liant les lamines (M-is7) semble délétère et entraîne l'apparition de hernies sur les noyaux de cellules HeLa et donc une structure nucléaire aberrante.

1.3.1.3. Modifications post-traductionnelles

1.3.1.3.1 Régulation de la lamine A par phosphorylation

Les lamines A forment des homodimères, qui s'assemblent ensuite pour former des filaments. Durant la mitose, ce réseau de lamines A se désassemble. Les régions tête et queue des lamines contiennent deux motifs de phosphorylation très conservés qui régulent le désassemblage des lamines chez la plupart des métazoés¹⁵⁵. Le domaine tête des lamines, très conservé au cours de l'évolution, est phosphorylé par la kinase CDK1 sur la sérine 22 chez la lamine A humaine¹⁵⁶. Cette phosphorylation n'empêche pas l'association de la lamine A en dimères mais bloque l'assemblage des dimères de lamines A en complexes « tête à queue ». La kinase CDK1 cible aussi, à l'extrémité opposée du domaine coiled-coil, la sérine 392 de la lamine A humaine¹⁵⁶. Cette phosphorylation permet de la même façon la dépolymérisation du réseau de lamines A¹⁵⁷. Enfin, la phosphorylation de la sérine 525 dans le domaine Igfold pourrait réguler soit l'assemblage de la lamina, soit son interaction avec la chromatine^{157,158}.

Une autre étude a criblé, par spectrométrie de masse, les sites de phosphorylation retrouvés de manière majoritaire au sein de la lamine A lors de l'interphase. Ainsi, 20 sites de phosphorylation ont pu être identifiés, tous contenus dans 3 régions spécifiques : les régions tête 3-22, début de queue 390-426, puis fin de queue avec la sérine 458 dans l'Igfold et les sérines 628, 636 et 652 dans la région spécifique à la lamine A¹⁵⁹. Les sérines cibles de Cdk1 et régulant l'assemblage des lamines sont retrouvées dans cette étude. Les sérines 403 et 404, situées à côté du signal de localisation nucléaire, sont également détectées : elles régulent le transport au noyau des lamines¹⁵⁷. Pour aller plus loin, les auteurs ont produit des mutants possédant soit des mutations phospho-mimétiques, soit des mutations ne permettant plus à la phosphorylation de se faire et les ont exprimés dans des cellules HeLa. Ils ont mis en évidence que les sérines 22 et 392 étaient également phosphorylées en interphase, ce qui est surprenant compte tenu du fait que ces modifications post-traductionnelles avaient été identifiées durant la mitose. Cependant, ces phosphorylations pourraient réguler la quantité de lamine A nucléoplasmique en interphase également. En effet, le mutant S22D entraîne une localisation nucléoplasmique de la lamine A, et le mutant S392A présente le même effet mais à un degré inférieur.

Enfin, il est intéressant de noter que cette étude met en évidence des sites de phosphorylation situés dans des régions qui permettent l'interaction entre la lamine et des

facteurs associés à la chromatine tels que SREBP1 (Sterol Regulatory Element-Binding Protein 1), Mok2 (Mouse Krüppel-like 2) ou encore PCNA¹⁶⁰⁻¹⁶². Enfin, il est connu que l'interaction entre la lamine A et LAP2 α permet de réguler l'activité du facteur pRB, très important dans le cycle cellulaire^{163,164}. Il semble que la phosphorylation de la lamine soit un bon candidat pour la régulation de l'interaction de la lamine A avec LAP2 α et donc un modulateur potentiel de l'activité de pRB.

1.3.1.3.2 La farnésylation de la lamine A

Au cours de sa maturation, la lamine A, qui est synthétisée sous forme de prélamine A, va subir une modification post-traductionnelle particulière : la farnésylation. Cette étape consiste en l'ajout d'un groupement farnésylpyrophosphate (Figure 19), grâce à une farnésyltransférase.

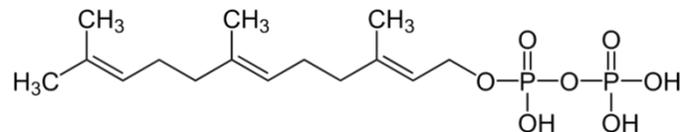


Figure 19 : Représentation topologique de la molécule de farnésylpyrophosphate

Cette farnésylation confère à la lamine A des propriétés chimiques spécifiques. En particulier, elle permet son ancrage à la membrane¹⁶⁵. Une fois le dernier clivage effectué, la lamine A mature est libérée dans le nucléoplasme. Le site de farnésylation est nécessaire pour la fonction de la lamine A. En effet, l'expression d'une lamine A tronquée de son motif CAAX cause des agrégats nucléaires de lamine A tronquée et perturbe la distribution de la lamine A/C endogène¹⁶⁶. Une étude récente a proposé un mécanisme selon lequel la farnésylation de la lamine permettrait d'empêcher la phosphorylation de la sérine 22, modification nécessaire pour la translocation de la lamine A vers le nucléoplasme¹⁶⁷. Une mutation de la lamine A G608G entraîne la production d'une lamine A possédant une délétion de 50 acides aminés et restant constitutivement farnésylée, la progérine. L'accumulation de ce mutant entraîne la survenue d'un syndrome de vieillissement accéléré qui sera décrit dans une partie ultérieure. Si ce mutant reste farnésylé, le fait qu'il ne puisse pas être phosphorylé le maintiendrait à la membrane. Ainsi, le mutant progérine aurait un défaut de localisation et empêcherait la réalisation des fonctions de la protéine au sein du nucléoplasme. De manière intéressante, la phosphorylation sur la sérine 22 peut être réalisée sur la progérine en présence d'inhibiteurs de farnésyltransférase ou pour un mutant de la progérine non farnésylable¹⁶⁷. La progérine ayant ainsi subi une phosphorylation sur la sérine 22 forme des gouttelettes intranucléaires qui retiennent aussi la lamine A/C endogène¹⁶⁷ (Figure 20). Ainsi, un nouveau modèle est proposé, selon lequel la défarnésylation de la prélamine A / progérine facilite la phosphorylation de la sérine 22 et la translocation de la lamine A / progérine dans le nucléoplasme. Un défaut de défarnésylation entraîne un défaut de phosphorylation de la

sérine 22 et de translocation dans le nucléoplasme. Cependant, empêcher la farnésylation de la prélamine A / progérine cause aussi des dysfonctionnements cellulaires. Cela favorise l'accumulation nucléoplasmique de la prélamine A / progérine. La lamine mature reste à l'enveloppe nucléaire après défarnésylation à cause des contacts tête-queue entre les dimères de lamine A (et probablement aussi en partie à cause d'interactions avec d'autres protéines de l'enveloppe nucléaire). Si l'on déstabilise ces contacts, par exemple avec la mutation R50P, alors une absence de farnésylation a un impact encore plus grand. Il a été montré que la progérine forme dans ce cas des agrégats ressemblants à des gouttelettes lipidiques, qui rappellent les transitions de phase observées avec des protéines possédant des régions de faible complexité¹⁶⁷. Cette progérine est alors fortement phosphorylée sur sa sérine 22, ce qui augmente sa vitesse de dégradation.

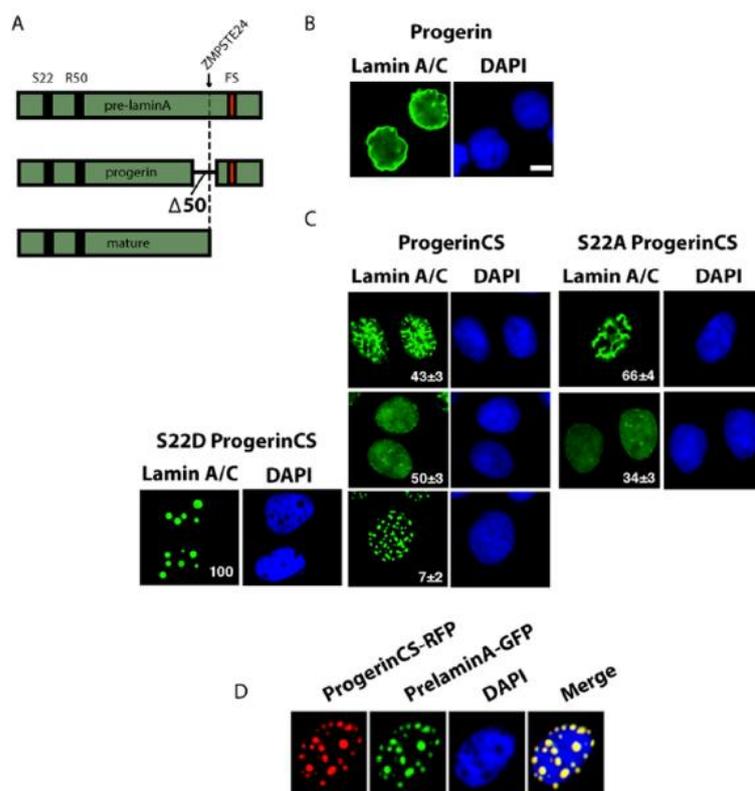


Figure 20 : Localisation cellulaire de la progérine et la progérine non farnésylée (CS)¹⁶⁸. (A) Constructions transfectées dans les cellules U-2 OS. (B-C) Immunofluorescence après transfection avec les différentes constructions de progérine. (D) Immunofluorescence après transfection avec les constructions ProgerinCS et PrelaminA, révélant que la progérine non farnésylée entraîne la lamine A/C dans les gouttelettes nucléoplasmiques.

Les prélamine A et progérine farnésylées ont aussi des partenaires différents de la lamine A/C mature dans la cellule. Une recherche systématique par double hybride en levure menée par l'équipe du Dr. Tom Misteli a permis de cartographier, parmi 41 fragments de lamine A, ceux qui interagissaient avec le plus grand nombre de partenaires¹⁶⁸. Cette équipe a ainsi identifié 337 partenaires de la lamine A. Parmi ces 337 partenaires, 61 sont spécifiques de la lamine A mature (1-646), 225 sont spécifiques de la progérine et 51 interagissent avec les deux formes. De manière intéressante, le mutant de la lamine A C661S qui empêche la farnésylation empêche toute interaction avec les partenaires de la progérine. Ceci permet de poser l'hypothèse de l'importance de la farnésylation dans ces interactions. Cette hypothèse est renforcée par le fait que le mutant K647C, qui est une prélamine A constitutivement

farnésylée, permet à cette prélamine d'interagir avec 75% des partenaires de la progérine. La farnésylation semble donc être la clé des interactions entre les prélamine A et progérine farnésylées et leurs partenaires.

Enfin, il a été montré que la farnésylation régule spécifiquement l'interaction entre la lamine A et la protéine SUN1. Une première étude a étudié par des méthodes de pull-downs cette interaction. Elle a montré que la lamine A est capable de se lier à SUN1 (1-355) mais pas la progérine¹¹². Ceci indique que les 50 résidus absents chez la progérine sont importants pour l'interaction. Cette étude n'a cependant pas pris en compte l'effet de la farnésylation puisque les lamines A utilisées pour les pull-downs étaient produites par transcription-traduction *in vitro*.

Une seconde étude s'est intéressée à l'impact de la farnésylation sur cette interaction¹⁶⁹. Dans cette étude, des co-immunoprécipitations entre différents mutants de la lamine A et SUN1 ont été réalisées.

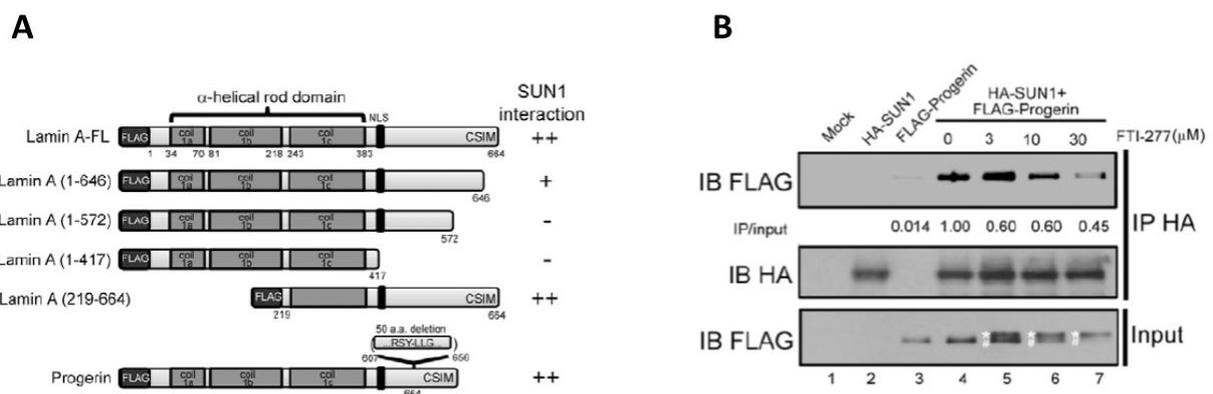


Figure 21 : SUN1 lie la lamine A farnésylée ainsi que la progérine farnésylée¹⁶⁹. (A) Représentation schématique des constructions utilisées pour les études de co-immunoprécipitation avec SUN1. Le résultat de ces co-immunoprécipitations est représenté sur la droite. (B) Co-immunoprécipitations entre la progérine et SUN1 réalisées en présence de concentrations croissantes d'inhibiteurs de farnésyl-transférase, montrant l'importance de la farnésylation dans l'interaction.

Il s'avère que la lamine A farnésylée ainsi que la progérine sont capables de se lier à SUN1. De plus, il apparaît que la lamine A farnésylée lie plus fortement SUN1 que la lamine mature, c'est-à-dire sans son groupement farnésyl (Figure 21A). Par ailleurs, l'incubation de cellules avec des concentrations croissantes d'inhibiteurs de farnésyl-transférase, permettant d'empêcher la farnésylation de la progérine, réduit fortement l'interaction de la progérine avec SUN1 (Figure 21B). La farnésylation semble donc jouer un rôle primordial dans l'interaction de la lamine A (ou la progérine) avec SUN1.

1.3.2 Les pathologies associées à l'émerine, SUN1/2 et la lamine A/C

Comme déjà rapidement évoqué dans les paragraphes précédents, les gènes codant pour l'émerine, SUN1/2 et la lamine A/C sont mutés chez un grand nombre de patients atteints de diverses maladies génétiques. Ce lien a même été à l'origine de la découverte du gène de l'émerine (*STA* ou *EMD*), qui a été identifié par cartographie génétique¹⁰⁰ de la dystrophie récessive d'Emery-Dreifuss liée à l' X^{170} (X-EDMD). Cette pathologie est caractérisée par des contractures des tendons, une perte progressive lente, ainsi qu'un affaiblissement des muscles squelettiques, une cardiomyopathie avec des défauts potentiellement létaux du système de conduction ventriculaire qui peuvent entraîner des arrêts cardiaques soudains¹⁷¹. L'émerine a par la suite été catégorisée dans les protéines de l'enveloppe nucléaire⁹⁸ et classée parmi les protéines entraînant des laminopathies, des pathologies dues à des mutations des lamines ou des protéines en interaction avec les lamines, ou encore des envelopopathies, des pathologies causées par des mutations des protéines de l'enveloppe nucléaire^{172,173}. L'émerine étant présente dans tous les tissus, le fait que seul les muscles striés soient principalement touchés par la maladie suggère une superposition de fonction avec une autre protéine, sans doute à domaine LEM également comme Lem2, qui soit capable de compléter les fonctions de l'émerine dans la plupart des tissus⁹⁹.

Certaines laminopathies sont caractérisées par des altérations du tissu adipeux. Dans la dystrophie musculaire d'Emery-Dreifuss, les myocytes squelettiques et les cardiomyocytes sont progressivement remplacés par du tissu adipeux fibreux. Ces altérations pourraient être dues à une collaboration entre l'émerine, la voie Wnt- β -caténine, et le récepteur γ activé par les proliférateurs des peroxysomes (PPAR- γ). L'adipogénèse est réprimée via l'inhibition PPAR- γ dépendante de la voie Wnt- β -caténine¹⁷⁴. Cette répression passe par la dégradation de la β -caténine par le protéasome¹⁷⁵. L'émerine et la β -caténine influencent mutuellement leur niveau d'expression lors de l'adipogénèse. De plus, en présence de β -caténine activée, des fibroblastes n'exprimant pas d'émerine montrent un signal dépendant de PPAR- γ anormal et une adipogénèse augmentée, ce qui est rétabli lorsque les cellules ont été transfectées avec des siRNA spécifiques de la β -caténine¹⁷⁶.

La protéine SUN1 a été impliquée dans de nombreuses pathologies. SUN1, ainsi que SUN2, sont mutées chez des patients atteints d'EDMD ou autres myopathies similaires. Ces mutations (SUN1 G68D, G338S, W377C ; SUN2 R620C et E438D) affectent le positionnement du noyau dans des fibroblastes, confirmant le défaut de fonction du complexe LINC qui connecte le cytosquelette au noyau¹⁷⁷. De plus, des myotubes portant des mutations hétérozygotes de SUN1 montrent un défaut d'organisation, accompagné d'un défaut de recrutement des marqueurs centrosomaux à l'enveloppe nucléaire ainsi qu'un défaut lors de la nucléation des microtubules à l'enveloppe nucléaire¹⁷⁷. Une étude de 2015 a aussi montré la participation des protéines SUN1 et SUN2 dans le développement de l'ataxie cérébrale post-

natale chez des souris¹⁷⁸. Les souris ne possédant pas SUN1 montraient une remarquable diminution du volume cérébral, phénotype qui était exacerbé lorsque les souris n'exprimaient pas non plus SUN2. Enfin, l'accumulation de SUN1 a été détectée dans des syndromes de type progéria. Des souris exprimant le gène *Lmna* Δ 9 (entraînant une progéria), mais n'exprimant pas la protéine SUN1 montraient un phénotype très réduit de la pathologie avec des tissus moins affectés et une longévité accrue. L'accumulation de SUN1 est donc corrélée aux syndromes de vieillissement accélérés de type progéria¹⁷⁹.

Le gène *LMNA* est le siège de nombreuses mutations entraînant des pathologies, et est très étudié par la communauté scientifique. A ce jour, plus de 400 mutations ont été recensées (Figure 22). Certaines de ces mutations se situent dans la région commune à la lamine A et la lamine C tandis que d'autres sont spécifiques à la lamine A. Les pathologies engendrées par ces mutations sont regroupées sous le terme de laminopathies. Parmi ces laminopathies, on retrouve la dystrophie musculaire d'Emery-Dreifuss (EDMD), la dystrophie musculaire des ceinture de type 1B, la cardiomyopathie dilatée (DCM), la maladie autosomale récessive de Charcot-Marie-Tooth (CMT), la lipodystrophie partielle familiale de type Dunnigan (FPLD), des dysplasies mandibuloacrale avec lipodystrophie de type A (MADA), des syndromes de vieillissement prématuré et notamment le syndrome de Hutchison-Gilford progeria (qui sera détaillé dans la partie suivante), et enfin des dermopathies restrictives¹⁸⁰.

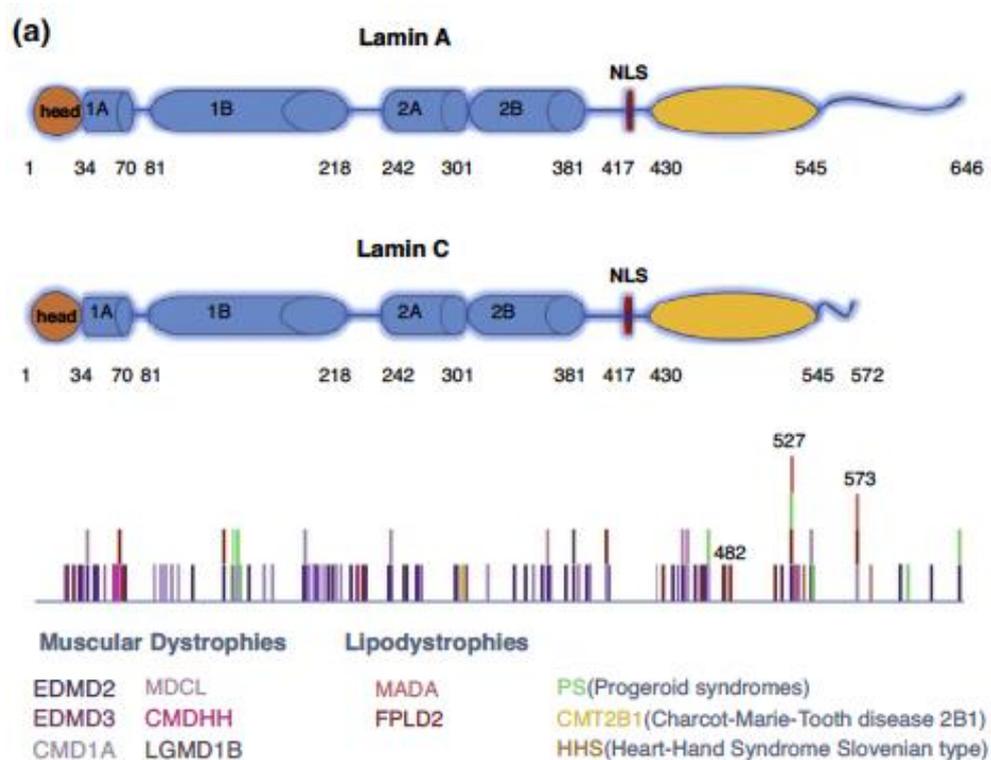


Figure 22 : Mutations des lamines et laminopathies¹⁸¹.

Les mutations causant ces pathologies peuvent affecter des tissus multiples ou être tissu-spécifiques. Bien que ces mutations puissent engendrer un grand nombre de pathologies, de nombreuses similarités peuvent être mises en évidence et permettent de regrouper les laminopathies en quatre groupes¹⁸¹ :

- Les atteintes des muscles striés
- Les lipodystrophies
- Les neuropathies périphériques
- Les syndromes de vieillissement accélérés

Ayant travaillé lors de ma thèse sur la région C-terminale de la lamine A, je vais ici me focaliser sur les syndromes de vieillissement accéléré dont une majorité sont associés avec cette région de la lamine A (Figure 22). Je présenterai tout d'abord le syndrome le mieux décrit, la progéria autosomale dominante d'Hutchinson-Gilford, puis je détaillerai un ensemble de syndromes progéroïdes dominants et récessifs aussi nommés progérias atypiques.

1.3.3 Une pathologie sévère associée à la région spécifique de la lamine A : le syndrome d'Hutchinson-Gilford ou progéria

Une des pathologies les plus dévastatrices liées à la région spécifique de la lamine A est le syndrome de Hutchinson-Gilford, aussi appelé progéria ou HGPS (pour Hutchinson-Gilford Progeria Syndrome) entraînant un vieillissement accéléré¹⁸². Il s'agit d'une maladie rare qui touche un enfant sur 4 millions¹⁸³. La maladie est diagnostiquée peu de temps après la naissance et est caractérisée par un grand nombre de troubles incluant notamment un retard de croissance, une dentition anormale, de l'ostéoporose, une hypoplasie osseuse, des défauts des gonades, une lipodystrophie, une athérosclérose, et une atrophie des muscles squelettiques¹⁸⁴. Habituellement, les enfants atteints de HGPS meurent d'insuffisance cardiaque entre 13 et 16 ans. Au moins 20 mutations distinctes ont été liées à des syndromes HGPS, cependant la mutation la plus fréquente est située en position 1824 et permet l'activation du site cryptique d'épissage alternatif menant à une prélamine A comportant une délétion de 50 acides aminés^{185,186}. Ces 50 acides aminés manquant retirent le site de clivage terminal par l'enzyme ZMPSTE24 sans affecter le motif CAAX. Ainsi, le mutant prélamine A HGPS, appelé « progérine », reste constitutivement farnésylé et carboxyméthylé¹⁸⁷. La progérine est produite de manière native dans les cellules, mais son accumulation est uniquement observée chez les cellules vieillissantes. Ceci renforce le lien entre l'accumulation de progérine et vieillissement cellulaire, la progérine serait même un biomarqueur de vieillissement^{188,189}.

La production de progérine mène à l'apparition progressive de nombreuses altérations cellulaires telles que des défauts de croissance, ainsi qu'une morphologie nucléaire altérée (Figure 23). L'expression ectopique de progérine chez des fibroblastes diploïdes sains

provoque la plupart des symptômes observés chez les cellules HGPS, indiquant que la progérine agit d'une manière dominante négative. De plus, des mutations dans le gène *ZMPSTE24* codant l'enzyme permettant le clivage terminal de la maturation de la prélamine, entraînent une dysplasie mandibuloacrale et/ou une dermatopathie restrictive, deux pathologies proches de la progéria caractérisées par un retard de croissance, des anomalies des muscles squelettiques ainsi qu'une atrophie cutanée¹⁹⁰⁻¹⁹². Il semble également que le vieillissement cellulaire entraîne une diminution du niveau d'expression de *ZMPSTE24*, menant à l'accumulation de progérine¹⁹³. Si l'on surexprime des variants de la prélamine A dans des fibroblastes humains normaux, il apparaît que les variants qui sont constitutivement farnésylés ont un fort effet inhibiteur sur la croissance cellulaire, ont une morphologie de la membrane nucléaire très altérée et induisent la formation d'agrégats de lamine A à la membrane nucléaire^{194,195}. En revanche, l'expression de la prélamine A dont la cystéine C661 (cible de la farnésylation) est mutée en méthionine n'influence que faiblement la croissance cellulaire, même si elle conduit à la formation de petits foyers nucléoplasmiques de prélamine A et à l'apparition de quelques noyaux dysmorphiques. L'expression de ce variant chez la souris conduit à un phénotype beaucoup moins sévère que la progéria caractérisé par une cardiomyopathie¹⁹⁶.

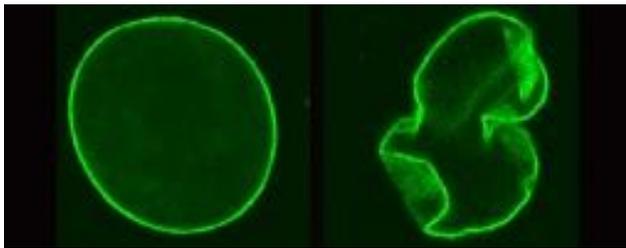


Figure 23 : Comparaison par microscopie à fluorescence d'un noyau d'une cellule saine (gauche) avec le noyau d'une cellule HGPS (droite)¹⁹⁷. Chez la cellule HGPS, le noyau possède une forme aberrante comportant des torsions et des invaginations.

La progérine semble également jouer un rôle dans la solidité du noyau et dans sa réponse à un stress mécanique. En effet, des cellules exprimant la progérine montrent une fragilité accrue du noyau et des signes de stress mécanique^{197,198}. Il est également démontré que l'accumulation de progérine altère l'organisation fonctionnelle des chromosomes en entraînant des changements dans la distribution et la quantité d'hétérochromatine, des défauts dans la ségrégation des chromosomes et un retard dans la formation de l'enveloppe nucléaire pendant la mitose, ainsi qu'une régulation anormale de la longueur des télomères^{188,199-203}. Ces différences sont susceptibles de modifier le niveau d'expression des gènes²⁰⁴⁻²⁰⁶. Elles peuvent être atténuées lors d'un traitement combinant des inhibiteurs de farnésyltransférase avec des inhibiteurs d'histone déacétylases²⁰⁷⁻²⁰⁹. Rien ne montre pour le moment que la progérine influe directement (à travers une interaction par exemple) sur les histone acétyltransférases ou déacétylases. Cependant, on observe un niveau d'expression diminué, ainsi qu'une localisation altérée de régulateurs de la structure de la chromatine tels que *RBBP4/7*, deux membres du complexe NURD du remodelage de la chromatine, chez les cellules progéria²¹⁰. De cette manière, il est possible que l'accumulation de progérine modifie l'organisation de la chromatine ainsi que sa régulation épigénétique. Le défaut observé lors de la mitose pourrait, lui, venir d'une compétition entre la protéine progérine farnésylée et la

protéine centromérique CENP-F, elle aussi farnésylée, au niveau du kinétochore des chromosomes en métaphase²⁰³. Le déplacement de CENP-F pourrait expliquer le retard au niveau de la chromatine (alignement et séparation des chromosomes), des cellules binucléées et une instabilité génomique. Enfin, il a été montré que la lamine A interagit avec TRF2 (Telomere-Repeat Binding Factor 2), ce qui participe à la structuration de l'extrémité des chromosomes^{211,212}. En particulier, le complexe entre la lamine A et TRF2 stabilise l'association entre l'ADN télomérique et des séquences ITS (Interstitial Telomeric Sequences) se trouvant dans des régions non télomériques. Dans les cellules exprimant la progérine, TRF2 ne s'associe plus à ces structures, les boucles d'ADN résultant de ces associations ne sont plus formées, et ceci induit une perte de stabilité des télomères, un défaut de condensation et une fragilité des chromosomes.

1.3.4 Les syndromes progéroïdes récessifs associés aux lamines A et C

Outre le syndrome HGPS, plusieurs syndromes de progéria atypiques (SPP) causés par des mutations du gène *LMNA* ont été décrits, tels que la dermatopathie restrictive atypique (la plus grave), le syndrome de Werner et la dysplasie mandibuloacrée. Certaines mutations causant ces syndromes progéroïdes atypiques ont été identifiées. Elles sont soit hétérozygotes, soit homozygotes.

- Mutations dominantes de la lamine A/C causant des syndromes progéroïdes :

Des mutations ont été identifiées dans les domaines en hélice des lamines A/C, et plus précisément les régions coil 1A (A57P²¹³, E55K, E55G²¹⁴) et 1B (R133L, L140R, S143F et E145K^{213,215-217}), qui sont nécessaires pour la polymérisation des lamines nucléaires en structures d'ordre supérieur. D'autres mutations ont été observées dans la région C-terminale spécifique à la lamine A, qui provoquent des défauts d'épissage. Ces mutations conduisent à la production de différentes prélamines tronquées telles que la prélamine A Δ 35 (del622-656), Δ 50 (progérine, del607-656) et Δ 90 (del567-656)²¹⁸.

- Mutations récessives de la lamine A/C causant des syndromes progéroïdes :

Ces mutations ont été trouvées dans le domaine globulaire de la lamine A/C, l'Igfold (R435C²¹⁹, R471C²²⁰, R527C²²¹, R527H²²²⁻²²⁵, A529T²²⁶, A529V²²⁷ et K542N²²⁸). De manière intéressante, lors de ma thèse, Camille Samson dans l'équipe a montré que l'ensemble des résidus mutés se trouvaient au niveau de la surface de liaison de la lamine A/C à BAF²²⁹. Elle a mesuré l'impact de 5 de ces mutations sur l'affinité de l'Igfold pour BAF : ces mutations diminuent significativement l'affinité de l'Igfold de la lamine A/C pour BAF (Figure 24).

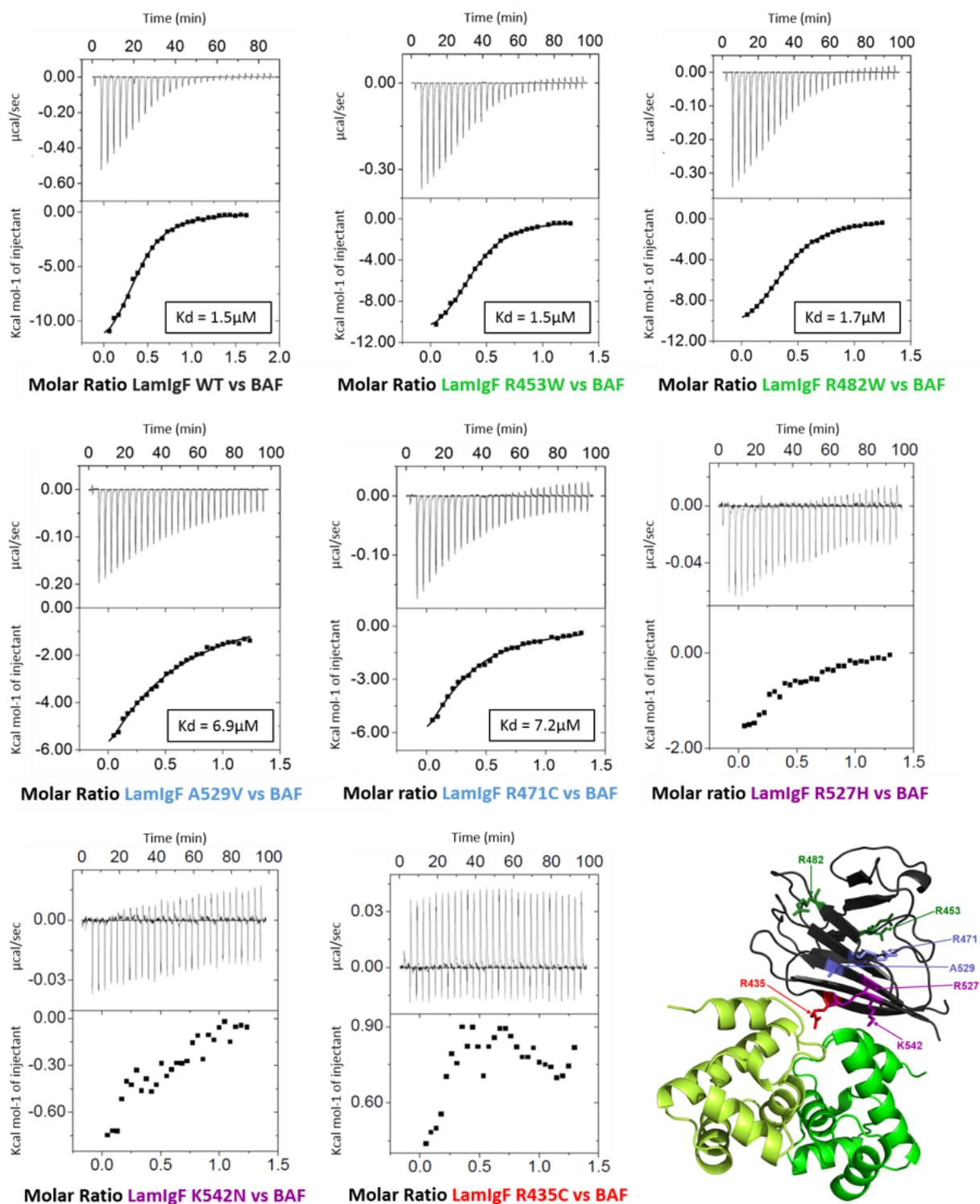


Figure 24 : Impact des mutations causant des syndromes progéroïdes récessifs sur l'interaction avec BAF²³⁰. L'affinité des différents domaines Igfold mutés pour le dimère de BAF a été mesurée par calorimétrie. Deux mutants contrôle ont aussi été étudiés : R453W qui provoque l'EDMD et R482W qui cause une lipodystrophie. Alors que l'affinité pour BAF de ces 2 mutants contrôle est la même que l'affinité de l'Igfold natif, l'affinité des 5 mutants causant des syndromes progéroïdes est significativement plus faible. Les résidus mutés sont positionnés sur la structure 3D de l'Igfold de la lamine A/C en interaction avec BAF (en bas à droite).

Etant donné que la délétion de 50 acides aminés responsable de la production de progérine entraîne la survenue d'un syndrome HGPS, et que la mutation des résidus à l'interface entre l'IgFold de la lamine A/C et BAF entraînent l'apparition de syndromes progéroïdes, il est possible que la présence de la région C-terminale farnésylée de la progérine ait un impact négatif sur l'interaction lamine A/C – BAF.

Dans la partie suivante, pour finir, je vais vous présenter BAF et un autre partenaire des lamines associé à la chromatine nommé Rbbp4, ces deux partenaires étant décrits comme des acteurs des mécanismes pathologiques déclenchés dans les syndromes progéroïdes associés aux lamines.

1.4. Les partenaires de l'enveloppe nucléaire associés à la chromatine

1.4.1 BAF

A l'intérieur du noyau, la chromatine est composée de l'ADN ainsi que des protéines qui permettent sa compaction : les histones. La chromatine se trouve sous deux formes. Une forme euchromatine, transcriptionnellement active, décondensée et principalement située au centre du noyau. Une seconde forme de chromatine est l'hétérochromatine qui est très condensée, très peu transcrite et plutôt située à la périphérie nucléaire^{230,231}. Les histones jouent un rôle essentiel dans ce phénomène de compaction de l'ADN et les modifications chimiques (PTMs) qu'elles portent influencent cette compaction.

BAF (Barrier to Auto-integration Factor 1) est une protéine ubiquitaire localisée à la fois dans le cytoplasme et dans le nucléoplasme, dont la surexpression induit une forte compaction de la chromatine²³². Elle est composée de 89 acides aminés et est très conservée au sein des métazoaires. Elle existe sous forme de dimère en solution, et forme des oligomères de grande taille en présence d'ADN²³³. BAF interagit avec l'ADN double brin sous forme d'homodimère et sans spécificité de séquence²³⁴. La structure du dimère de BAF a été résolue par RMN²³⁵ ainsi que par cristallographie²³⁶. Par la suite, une structure du dimère de BAF en complexe avec l'ADN a été obtenue par cristallographie²³⁷ (Figure 25). Cette structure montre que le dimère de BAF est capable de faire un pont entre deux ADN double brin. De plus, la protéine BAF se lie aux protéines à domaine LEM, la structure du dimère de BAF en complexe avec le domaine LEM de l'émerine a d'ailleurs été résolue par RMN¹⁰³.

BAF joue un rôle dans l'assemblage du noyau après la mitose grâce à sa faculté à lier l'ADN²³⁸. BAF forme un réseau rigide de chromatine permettant de donner sa forme au noyau. BAF se localise sur la couche externe de la chromatine formant les chromosomes et permet, par pontage de l'ADN, d'épaissir la surface de la chromatine. Pendant la mitose, la phosphorylation de BAF par la kinase VRK1 sur la sérine 4 empêche l'interaction de BAF avec

l'ADN. En fin de mitose, cette phosphorylation est retirée et l'interaction entre BAF et l'ADN peut de nouveau se faire, redonnant sa forme au noyau.

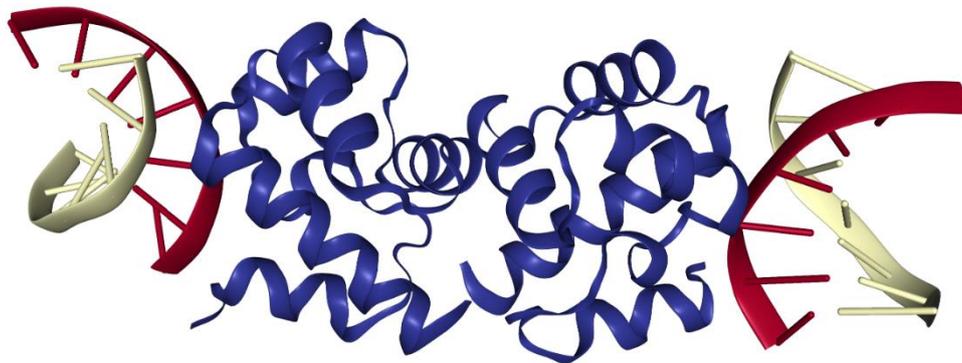


Figure 25 : Structure du dimère de BAF en interaction avec l'ADN²³⁸ (code PDB : 2BZF). Chaque monomère de BAF (bleu) lie une molécule d'ADN (rouge et blanc).

La protéine BAF interagit avec de nombreux partenaires²³⁹. On compte parmi ces protéines : les protéines à domaine LEM, les lamines, les histones, et plusieurs protéines et enzymes qui régulent la chromatine.

1.4.1.1. Les protéines à domaine LEM de la membrane nucléaire interne.

Durant l'interphase, BAF est enrichie à l'enveloppe nucléaire et dans le noyau, mais est aussi observée dans le cytoplasme^{240,241}. BAF interagit avec l'émerine à la membrane nucléaire interne, mais reste beaucoup plus mobile que l'émerine. Lors de la télophase, la protéine est concentrée dans une sous-région des chromosomes, où se trouvent aussi les protéines à domaine LEM et la lamine A. BAF et l'émerine sont alors très peu mobiles. En fin de mitose, BAF joue un rôle essentiel dans la relocalisation des protéines de la membrane nucléaire interne et est un acteur clé de la reformation de l'enveloppe nucléaire²⁴⁰.

Les interactions médiées par la protéine BAF sont modulées des événements de phosphorylation induisant l'ajout de charges négatives. En effet, la kinase VRK1 peut phosphoryler la sérine 4 (site majeur) ainsi que les thréonines 2 et 3 (sites mineurs) de BAF, empêchant ainsi l'interaction de BAF avec l'ADN double brin et réduisant fortement sa liaison avec les protéines de la membrane nucléaire interne à domaine LEM mais également avec la lamine A. La phosphorylation sur la sérine 4 semble aussi réduire l'interaction entre l'émerine et la lamine A *in vitro*, et entraîner une délocalisation de l'émerine vers des agrégats cytoplasmiques (probablement associés au réticulum endoplasmique) en cellules¹¹⁴.

1.4.1.2. Les lamines.

La lamine A subit plusieurs étapes de maturation notamment une étape de farnésylation, suivi d'un clivage, d'une carboxyméthylation et enfin d'un second clivage. Les différentes formes de la lamine A interagissent avec la protéine BAF et modifient sa localisation entre le cytoplasme et le nucléoplasme²⁴². Dans des cellules en interphase, BAF se trouve à la fois dans le cytoplasme et le nucléoplasme. Dans des cellules surexprimant une FLAG-lamine A sauvage, BAF est complètement nucléaire et co-localise avec les lamines A exprimées (Figure 26). Le même résultat est obtenu avec la prélamine A non farnésylée ainsi qu'avec le mutant farnésylé-carboxyméthylé mimant le comportement du mutant progérine responsable du syndrome HGPS (Hutchinson-Gilford Progeria Syndrome), montrant ainsi que la lamine A, ainsi que ses précurseurs, sont essentiels pour le recrutement de BAF au noyau. Cependant la colocalisation entre BAF et lamine A farnésylée-carboxyméthylée est seulement partielle dans le nucléoplasme (Figure 26).

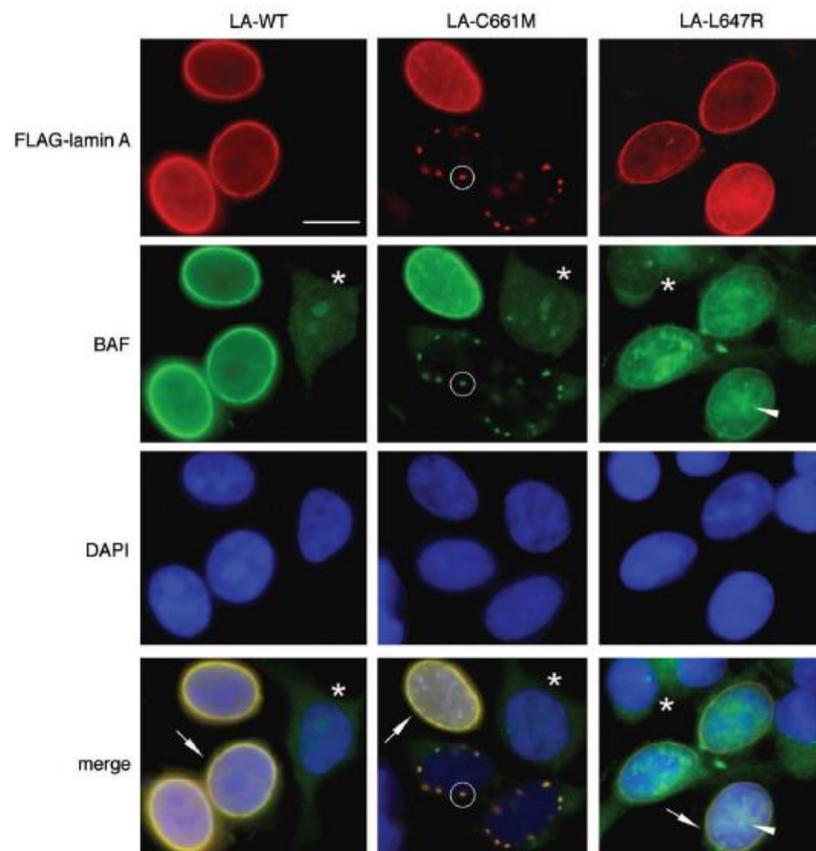


Figure 26 : L'accumulation de lamine A entraîne la translocation de BAF vers le noyau²⁴². L'accumulation de différentes constructions de lamine A possédant une étiquette FLAG dans des cellules HEK293 a été détectée par des anticorps monoclonaux anti-FLAG et visualisée à l'aide d'un anticorps secondaire conjugué avec du FITC (vert). La localisation de BAF a été observée grâce à la fluorescence de la GFP. L'ADN du noyau des cellules a été coloré à l'aide du DAPI. Les cercles indiquent les zones de co-marquage de la prélamine non farnésylée et de BAF au niveau des agrégats de prélamine à l'enveloppe nucléaire interne. Les triangles indiquent les zones d'accumulation de prélamine A farnésylée et de BAF dans le nucléoplasme. Les flèches montrent la colocalisation des différentes constructions de lamine A et de BAF à l'enveloppe nucléaire interne. Les cellules non transfectées sont indiquées par des astérisques.

1.4.1.3. Histones et protéines de remodelage de la chromatine

L'équipe du Dr K. Wilson a montré que BAF interagit directement avec les histones H3 ainsi que H1.1 avec des affinités de l'ordre du micromolaire²⁴³. Elle a montré *in vitro* que la région N-terminale de l'histone H3 et la région C-terminale de l'histone H1.1 sont suffisantes pour la liaison à BAF. Elle a ensuite révélé que l'histone H4 était aussi capable d'interagir avec BAF²⁴⁴. Les études de cette équipe suggèrent que BAF est capable d'influencer les modifications des histones en cellules.

Enfin, une analyse protéomique des partenaires de BAF, aussi réalisée par l'équipe du Dr K. Wilson, propose que plusieurs protéines impliquées dans le remodelage de la chromatine interagissent avec BAF²³⁹. L'une de ces protéines est RBBP4, aussi identifiée comme un partenaire spécifique de la région C-terminale de la prélamine A²¹⁰.

1.4.2 RBBP4

RBBP4 (Retinoblastoma Binding Protein 4, aussi connue sous le nom de Chromatin Assembly Factor-1 p48 subunit ou encore RbAp48) est une chaperone d'histone impliquée dans différents événements cellulaires tels que le remodelage de la chromatine, la transcription des gènes ou encore l'organisation du cytosquelette. Cette protéine possède une taille de 52 kDa et comporte des domaines répétés WD40. Elle fait partie de différents complexes protéiques tels que les complexes NuRD intervenant dans le remodelage de la chromatine²⁴⁵, ou encore CAF-1 jouant un rôle dans la réparation des cassures de l'ADN ou dans l'assemblage de la chromatine lors de la réplication²⁴⁶. Sa structure tridimensionnelle a été résolue de nombreuses fois : 18 structures de Rbbp4 sont disponibles dans la Protein Data Bank, dont la moitié a été résolue en 2017-2018, ce qui témoigne de l'intérêt grandissant des équipes de recherche pour cette protéine (Figure 27).

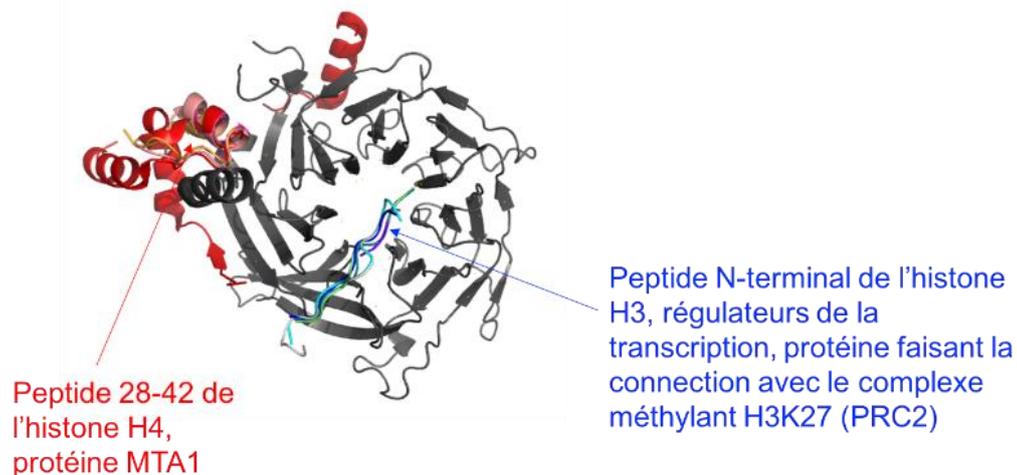


Figure 27 : Superposition des 18 structures de Rbbp4 disponibles dans la PDB. Rbbp4 est en gris et ses ligands en nuances de rouge et bleu. Deux sites d'amarrage ont été identifiés sur Rbbp4, l'un responsable de l'interaction avec l'histone H4 (ici en rouge) et l'autre avec l'histone H3 (ici en bleu).

RBBP4 a été identifiée comme un partenaire de BAF à partir d'expériences de chromatographie d'affinité et spectrométrie de masse, ainsi que décrit plus haut²³⁹. Elle a aussi été identifiée par double hybride en levure comme un partenaire spécifique de la région 562 à 664 de la lamine A²¹⁰ (Figure 28A). Cette interaction a été validée par GST-pulldown (Figure 28B). Il a été montré par immunoprécipitation que Rbbp4 interagit avec la lamine A mais n'interagit pas avec son variant progérine causant la progéria d'Hutchinson-Gilford (Figure 28C). Par ailleurs, il a été montré que dans les cellules HGPS, montrant une accumulation de progérine, le niveau de Rbbp4 ainsi que de son homologue Rbbp7 diminue. Cette observation pose la question de l'impact du défaut d'interaction Rbbp4-progérine dans le développement des syndromes de vieillissement accélérés liés à l'accumulation de progérine.

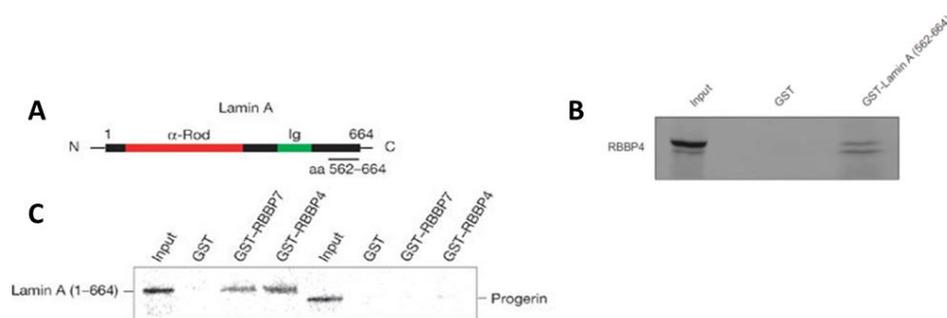


Figure 28 : La lamine A interagit avec le complexe NuRD (adapté de Pegoraro et al., 2009). (A) Fragment 562-664 utilisé dans les expériences de double hybride et de pulldown avec RBBP4. (B) GST-pulldown réalisé avec GST-lamine A (résidus 562-664) et Rbbp4 transcrit-traduit *in vitro* et marqué au ³⁵S. (C) GST-pulldowns réalisés avec un fragment recombinant GST-Rbbp4 (ou Rbbp7) et la lamine A (ou progérine) transcrit-traduite *in vitro* et marquée au ³⁵S.

1.5. Hypothèses et objectifs

Notre laboratoire, le LBSR (Laboratoire de Biologie Structurale et Radiobiologie), travaille depuis de nombreuses années sur la caractérisation structurale des protéines de l'enveloppe nucléaire interne, leurs interactions au sein de l'enveloppe ainsi qu'avec la chromatine. Ces protéines comportent de grandes régions prédites comme désordonnées, dont la structure et la fonction sont mal comprises (Figure 29).

Avant le commencement de ma thèse, une étudiante en thèse, Isaline Herrada, a démontré qu'une protéine ancrée à la membrane nucléaire interne, l'**émerine**, était capable de s'auto-assembler sous forme de filaments. Ce processus d'auto-association nécessitait la présence d'une grande région désordonnée¹¹¹ (notée **1** sur la Figure 29). Dans une première partie de ma thèse, j'ai entrepris d'obtenir des informations structurales sur cette région désordonnée de l'émerine, et d'identifier les changements conformationnels associés à la formation des filaments d'émerine. Une autre étudiante en thèse, Camille Samson, a montré que, si l'émerine monomérique ne semblait pas interagir *in vitro* avec la lamine A, en revanche

l'émerine auto-assemblée était capable d'interagir directement avec le domaine Igfold de la lamine A²²⁹.

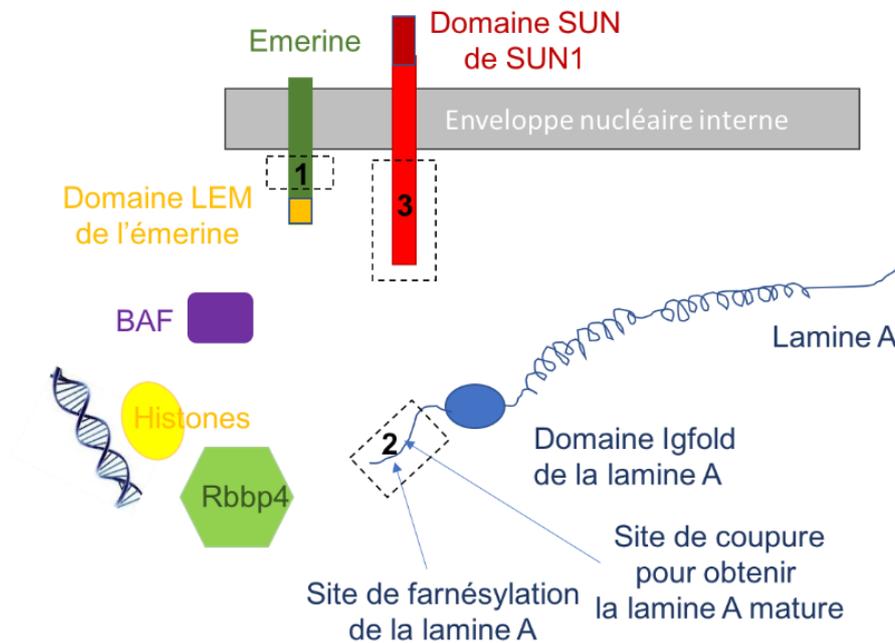


Figure 29 : Régions prédites comme déstructurées des protéines de l'enveloppe nucléaire interne étudiées lors de ma thèse.

Par la suite, je me suis intéressé à la région C-terminale désordonnée de la **lamine A** (notée **2** sur la Figure 29) et de son mutant progérine (responsable de la progeria d'Hutchinson-Gilford), pour tenter d'élucider les différences structurales et fonctionnelles entre ces deux formes protéiques. Un élément crucial de la toxicité de la progérine est sa farnésylation. J'ai donc tenté d'analyser l'impact de cette farnésylation sur sa structure et la liaison à ses partenaires. Enfin, l'un des partenaires de la progérine qui contribue à sa toxicité est **SUN1**. J'ai tenté de caractériser la région N-terminale de SUN1, responsable de sa liaison à la progérine farnésylée, et elle aussi prédite comme désordonnée (notée **3** sur la Figure 29).

Au cours de ce manuscrit, je détaillerai donc :

- 1) Une analyse structurale de la région désordonnée de l'émerine
- 2) La caractérisation des filaments formés par l'émerine
- 3) Une analyse structurale de la région C-terminale désordonnée de la lamine A et de son mutant progérine ainsi que de leurs interactions avec plusieurs partenaires
- 4) L'impact de la farnésylation sur les propriétés structurales et fonctionnelles de la progérine
- 5) La mise en place d'une expérience couplant MALS et SAXS pour l'étude de l'auto-assemblage d'une protéine sur la ligne de SAXS du Synchrotron Soleil.

II. Résultats

De nombreuses protéines possèdent des régions désordonnées. Ces régions semblent être essentielles à la régulation des interactions intermoléculaires. Par ailleurs, elles sont aussi responsables de processus d'auto-assemblage.

Dans une première partie de ma thèse, je me suis intéressé à la région intrinsèquement désordonnée de l'émerine, une protéine ancrée à la membrane nucléaire interne. Grâce à une étude structurale utilisant la RMN, j'ai pu obtenir des informations structurales sur cette région. Il a été récemment montré que l'émerine possédait la capacité de former des oligomères par auto-association de plusieurs molécules d'émerine. Je me suis donc également intéressé aux oligomères formés par l'émerine, à leur structure et au rôle de la région désordonnée de l'émerine dans l'assemblage de ces oligomères.

Par la suite, j'ai travaillé sur la région C-terminale intrinsèquement désordonnée de la lamine A, un partenaire de l'émerine appartenant au nucléosquelette. En parallèle, j'ai réalisé la même analyse sur un mutant de la lamine A, nommé progérine, responsable de la survenue d'un syndrome de vieillissement accéléré appelé HGPS (pour Hutchinson Guilford Progeria Syndrome). La différence entre la protéine sauvage et le mutant progérine réside dans la délétion de 50 acides aminés dans la région C-terminale de la protéine, prédite comme intrinsèquement désordonnée. Je me suis donc intéressé à l'étude structurale de ces deux régions prédites comme désordonnées afin de comprendre les différences fonctionnelles liées à cette mutation. Une seconde particularité du mutant progérine est l'absence du site de coupure responsable de la disparition du groupement farnésyl en fin de maturation de la protéine. En effet ce site de coupure se trouve dans la région de 50 acides aminés absente dans le mutant. Ainsi, la progérine reste constitutivement farnésylée et cela représente une piste d'étude pouvant expliquer la survenue de la maladie. J'ai donc mis au point l'étude de la farnésylation de ces régions C-terminales désordonnées *in vitro* afin d'étudier l'influence de cette modification chimique sur le comportement de la molécule dans le cadre des interactions avec ses partenaires, notamment la lamine A ainsi que la protéine BAF associée à la chromatine.

Enfin, j'ai participé au développement d'un système combinant une injection rapide (Stopped-Flow) avec la mesure de diffusion de rayons X aux grands angles (WAXS) ainsi qu'aux petits angles (SAXS) sur la ligne SWING du Synchrotron SOLEIL. Ce système ayant pour objectif de permettre l'étude structurale en temps réel de la polymérisation des protéines formant des filaments. Ainsi, la mise au point de ce système permettra à la ligne SWING de proposer ce type d'expérience aux utilisateurs de la ligne.

II.1. Analyse structurale de la région nucléoplasmique de l'émerine

L'émerine est une protéine de 254 acides aminés, ancrée à la membrane nucléaire interne via une hélice α C-terminale. Son extrémité N-terminale est nucléoplasmique. Cette protéine est mutée dans la dystrophie musculaire d'Emery-Dreifuss liée à l' X^{100} . L'émerine contribue à la connexion physique entre le cytosquelette et le nucléosquelette en interagissant avec les protéines de la membrane nucléaire interne SUN1 et SUN2¹²⁴ ainsi qu'avec les filaments de lamine²⁴⁷. L'émerine est également impliquée dans le remodelage de la chromatine en interagissant, grâce à son domaine LEM, avec la protéine BAF dont le dimère est capable de lier l'ADN¹⁰³. En dehors du domaine LEM, l'émerine possède une région allant des résidus 44 à 221, prédite comme intrinsèquement désordonnée. Des mutations dans cette région entraînent des défauts d'interaction avec de nombreux partenaires tels que la lamine A, MAN1, HDAC3 ou encore l'actine²⁴⁸. A l'heure actuelle, les détails moléculaires régissant ces interactions ne sont pas connus. Afin d'étudier ces interactions et le rôle de la région intrinsèquement désordonnée de l'émerine dans la régulation de ses fonctions, je me suis intéressé à l'étude structurale de la région désordonnée de l'émerine.

II.1.1 Purification de la région nucléoplasmique de l'émerine

Afin d'obtenir des données structurales sur cette protéine, j'ai produit plusieurs constructions de l'émerine avec ou sans son domaine LEM, notamment les fragments 67-170, 67-221, 1-187 et 1-132. Aucune des constructions étudiées ne comportait le domaine transmembranaire de l'émerine (région 223-243).

J'ai exprimé les différents fragments de l'émerine dans des bactéries *E. coli* possédant un système d'induction à l'IPTG. Ces fragments étant produits dans les corps d'inclusion, il est nécessaire, pour les purifier, de les solubiliser dans un tampon contenant 8 M d'urée. Toutes les constructions de l'émerine étudiées possèdent une étiquette de 8 histidines ainsi qu'un site de coupure à la protéase TEV en N-terminal du fragment d'intérêt. Après une chromatographie par affinité sur colonne de nickel, trois étapes de dialyse successives sont nécessaires pour renaturer la protéine et retirer l'urée et l'imidazole présents. Ensuite, l'étiquette est coupée par la protéase TEV, et les différents fragments sont concentrés en vue des études biophysiques. Les gels SDS-PAGE des purifications des différents fragments de l'émerine peuvent être visualisés en Figure 30. Le Tableau 1 indique les rendements obtenus lors de la purification des fragments d'émerine.

Tableau 1 : Calcul des rendements obtenus lors de la production des différents fragments d'émerine

Constructions	67-221	1-132	67-170	1-187
Rendement (mg/L)	50	100	85	100

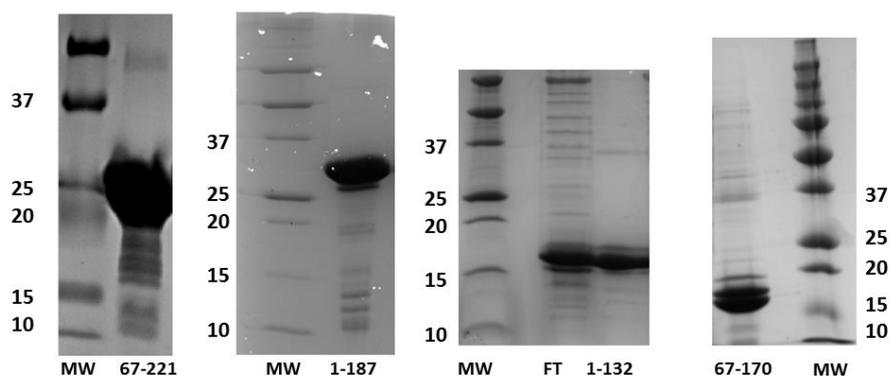


Figure 30 : Gels SDS-PAGE des différentes constructions purifiées de l'émerine. Les marqueurs de taille se trouvent dans la colonne « MW » (en kDa), la fraction non retenue sur colonne est indiquée par la colonne « FT ».

II.1.2 Analyse structurale de la région nucléoplasmique de l'émerine monomérique

Afin d'obtenir une première information structurale sur les fragments d'émerine avec lesquels je comptais travailler, j'ai effectué une prédiction de structure secondaire grâce au serveur Psipred²⁴⁹ (Figure 31A) et de désordre en utilisant le serveur Disopred³⁷. (Figure 31B) (serveur utilisé : <http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1>). Il semble ainsi que seuls les domaines LEM (acides aminés 1 à 49) et transmembranaire (acides aminés 221 à 254) soient structurés (Figure 27 A et B). Le reste de la protéine est prédit comme désordonné. La région reliant ces deux domaines ne semble pas adopter de structure secondaire.

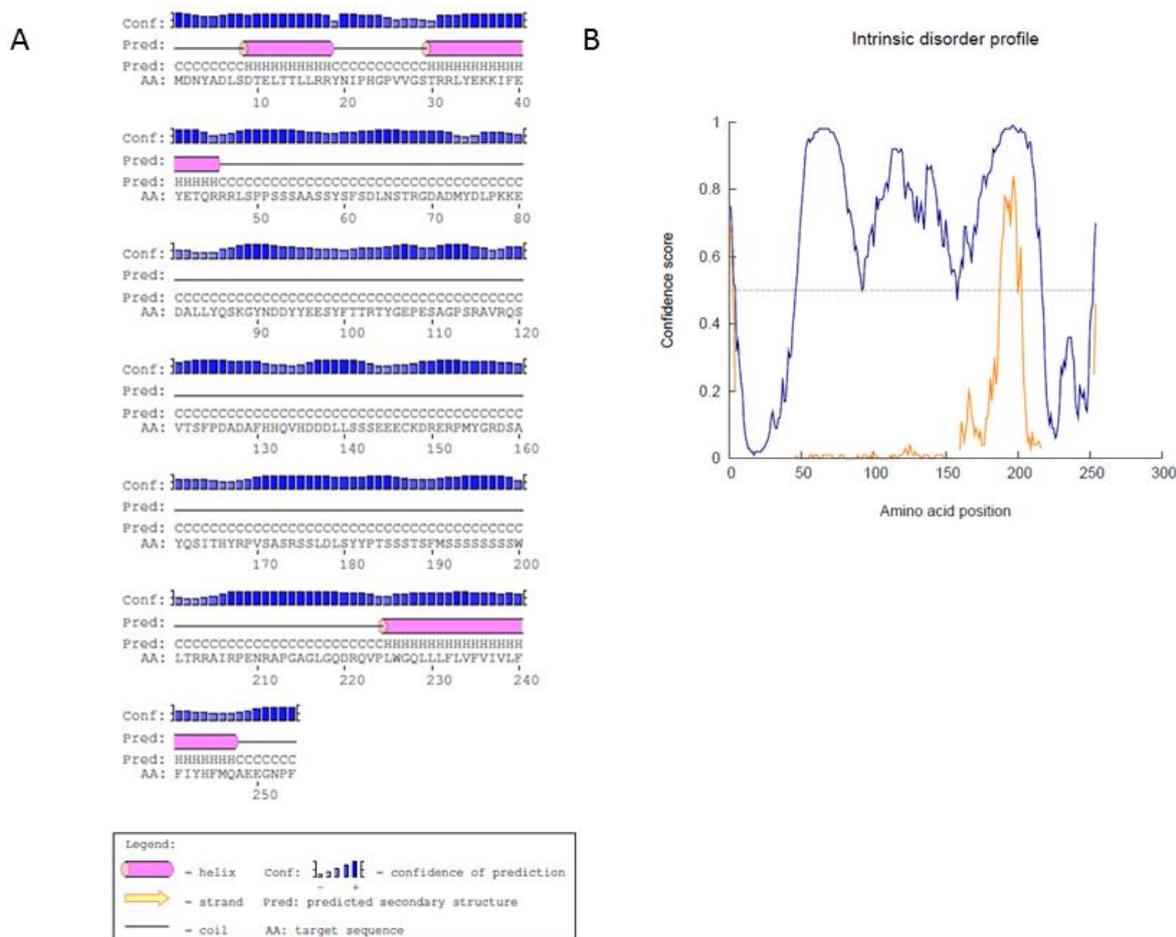


Figure 31 : Prédiction de la structure secondaire et de la propension au désordre de l'émerine entière, obtenue sur le site <http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1>. (A) Prédiction de structure secondaire à partir de la séquence en acides aminés de l'émerine réalisée grâce au serveur Psipred. Le diagramme annoté "Conf" en bleu représente l'indice de confiance de la prédiction pour chaque acide aminé. Les acides aminés montrant une tendance à former une hélice sont représentés par la lettre « H » dans la séquence ainsi qu'avec un cylindre rose. L'absence de structure est indiquée par la lettre « C » ainsi que par un trait noir continu. (B) Prédiction de la propension de désordre en fonction de la séquence représentée par la courbe bleue. Les régions prédites comme étant un site potentiel d'interaction avec des partenaires sont visualisés via la courbe jaune.

Afin de confirmer expérimentalement ces prédictions, les différents fragments de l'émerine ont été produits en milieu minimum enrichi en isotopes stables (^{15}N et ^{13}C), permettant leur étude par RMN. Ainsi, des spectres ^1H - ^{15}N HSQC ont été enregistrés afin d'obtenir l'« empreinte RMN » de chaque peptide (Figure 32A), les différentes constructions utilisées dans cette étude sont représentées en Figure 32B.

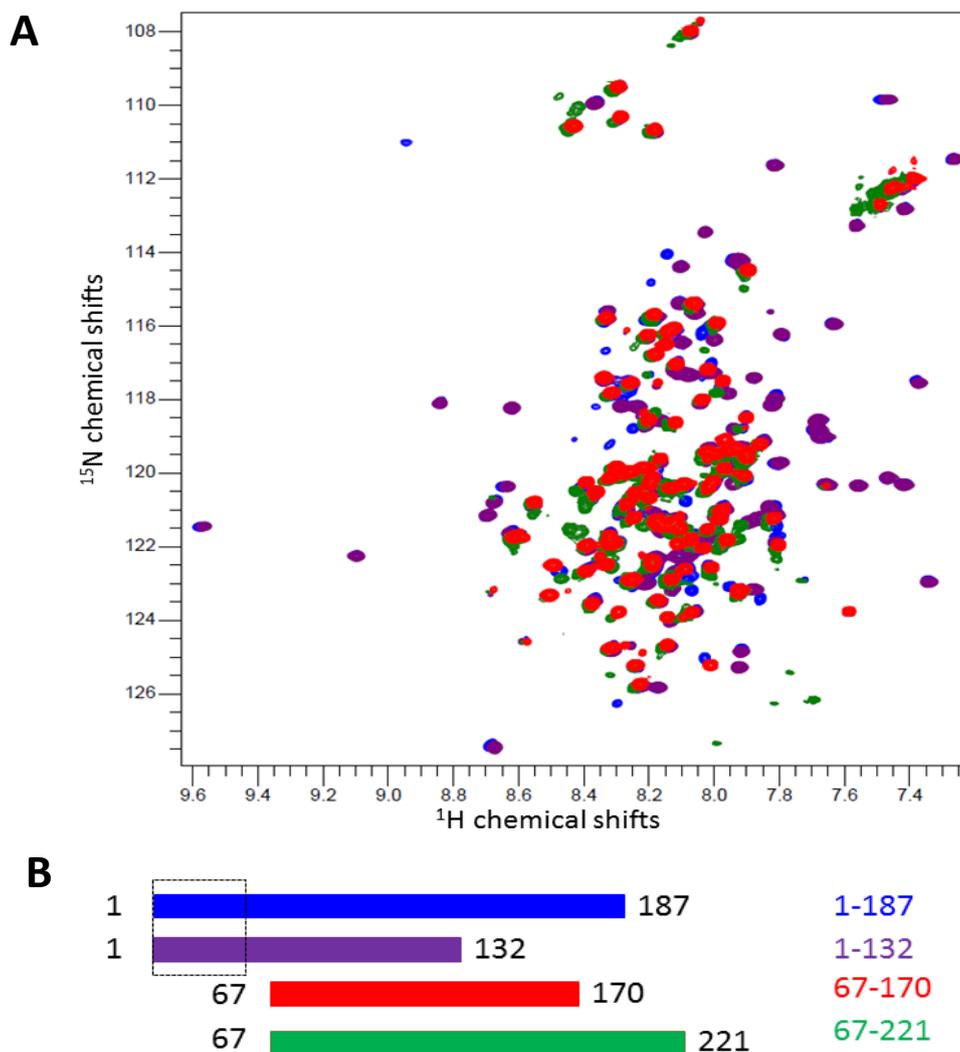


Figure 32 : Superposition des spectres ^1H - ^{15}N HSQC des différents fragments de l'émerine. (A) Spectres ^1H - ^{15}N HSQC des fragments 1-187 (bleu foncé) ; 67-221 (vert) ; 1-132 (violet) ; 67-170 (rouge). Les spectres ont été enregistrés à une concentration de 150 μM dans un tampon contenant 20 mM phosphate pH 6.5 ; 30 mM NaCl ; 5 mM DTT à 30°C. Les spectres de l'émerine 67-170 et 1-187 ont été enregistrés à 750 MHz (FMP Berlin) tandis que les autres spectres ont été enregistrés à 700 MHz (CEA Saclay). (B) Représentation schématique des différentes constructions de l'émerine utilisées. Le rectangle en pointillés indique la position du domaine LEM.

D'après les spectres RMN enregistrés, les régions 67-170 et 67-221 semblent effectivement intrinsèquement désordonnées puisque les signaux RMN correspondants se situent dans la zone caractéristique des acides aminés n'adoptant pas de structure secondaire stable (entre 7.5 et 8.5 ppm en dimension proton). Par ailleurs, les spectres des différentes constructions se superposent bien, indiquant que la structure ne varie pas d'une construction à l'autre. Les spectres des constructions 1-132 et 1-187 montrent des signaux bien dispersés de 7.4 à 0.6 ppm indiquant la présence d'un domaine structuré. Il s'agit de la région 1-49 de l'émerine qui compose le domaine LEM de l'émerine.

Afin d'aller plus loin dans la caractérisation structurale de l'émerine, j'ai contribué à attribuer les signaux RMN de la région désordonnée de l'émerine²⁵⁰. Ainsi, j'ai produit l'émerine 67-170 marquée aux ^{15}N et ^{13}C dans le but d'enregistrer des expériences RMN 3D

hétéronucléaires. La protéine montrant des signes d'agrégation à haute concentration, une attribution dans des conditions classiques semblait compliquée. J'ai donc décidé d'attribuer les signaux RMN de l'émerine 67-170 dans un tampon contenant 8M d'urée. J'ai enregistré un jeu d'expériences d'attribution classique (HNCA(N)NH, HNCA, HNCOC, HNCACB, CBCACONH). Ainsi, j'ai pu obtenir l'attribution finale des signaux ^1H , ^{15}N et ^{13}C de la région désordonnée de l'émerine dans 8 M urée. Par la suite, j'ai entamé le suivi de chaque signal en réalisant une titration avec un gradient décroissant d'urée. Ainsi, pour chaque concentration d'urée, un spectre $^1\text{H}^{15}\text{N}$ HSQC a été enregistré et le déplacement de chaque signal a été suivi (Figure 33) pour obtenir une attribution finale dans un tampon ne contenant pas d'urée (Figure 34A et 34B). Un second échantillon d'émerine doublement marqué a été produit sans urée et, à partir de l'HSQC attribuée, les déplacements chimiques carbone associés ont pu être obtenus.

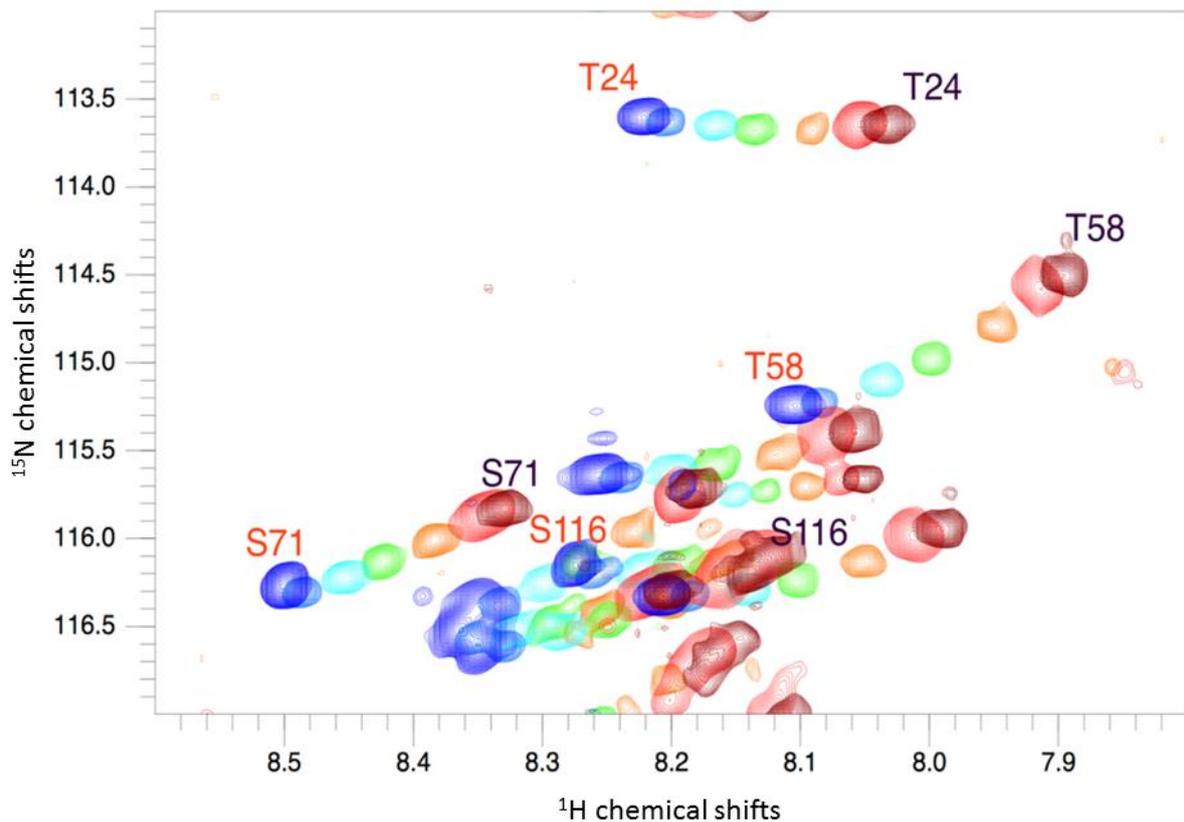


Figure 33 : Suivi des signaux RMN de quelques résidus du fragment 67-170 de l'émerine à 8 M (brun), 7 M (rouge), 6 M (orange), 4 M (vert), 2 M (bleu clair), 1 M (bleu), 0 M urée (bleu foncé). Les spectres ont été enregistrés à 750 MHz (FMP Berlin), dans un tampon contenant 20 mM phosphate pH6.5 ; 30 mM NaCl ; 5 mM DTT ; et les concentrations indiquées en urée.

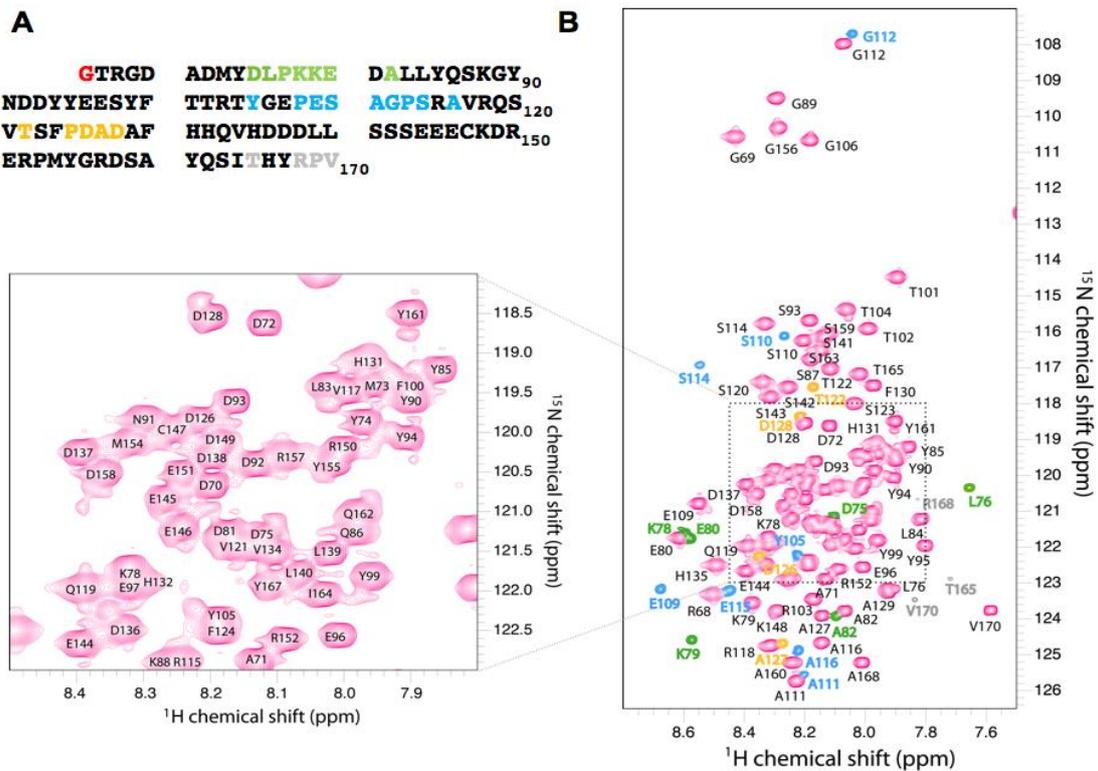


Figure 34 : Attribution de l'émerine 67-170. (A) Séquence de la région désordonnée de l'émerine (67-170)²⁵⁰. (B) Spectre ¹H¹⁵N HSQC de l'émerine 67-170 attribué. Les signaux en vert sont dus à une conformation *cis* de la proline P77 ; en bleu, aux prolines P108 et P113 ; en jaune, à la proline P125 et en gris clair, à la proline P169. Le spectre a été enregistré à 303K, 750 MHz et 600 μ M, dans un tampon contenant 20 mM phosphate pH6.5 ; 150 mM NaCl ; 2 mM DTT.

La différence des déplacements chimiques secondaires des C α et C β a été calculée à 0 et 8 M urée. Le graphique obtenu pour la condition 8M urée confirme que le peptide est complètement déplié. Le graphique obtenu en absence d'urée montre clairement que nous avons toujours affaire à une protéine majoritairement désordonnée, confirmant ce que nous avons vu sur le spectre ¹H¹⁵N HSQC. Seule la région 98-101 se replie transitoirement en hélice α (Figure 35). Même si les prolines sont majoritairement en conformation *trans* (dédduit des valeurs de déplacements chimiques des C β comprises entre 31.9 et 32.2 ppm), certaines prolines (P77, P108, P113, P123 et P169) montrent la présence de signaux correspondant à des conformations *cis* minoritaires avec des valeurs de déplacements chimiques des C β comprises entre 32.9 et 34.8 ppm (Figure 34B).

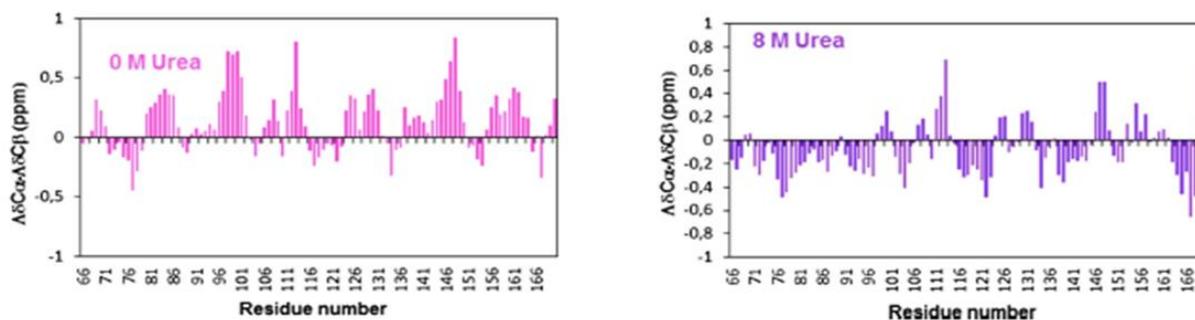


Figure 35 : Différences entre les déplacements chimiques secondaires des $C\alpha$ et $C\beta$ en fonction de la séquence de l'émerine mesurés en présence de 0 M et 8 M d'urée auxquels ont été soustraits les déplacements chimiques secondaires prédits en l'absence de structure secondaire par le logiciel ncIDP (<http://nmr.chem.rug.nl/ncIDP/>)(Mulder et al.,2010). La moyenne des déplacements chimiques secondaires, située autour de 0, confirme la structure globalement désordonnée. La présence d'une hélice α transitoire est révélée par des différences de déplacements chimiques secondaires supérieures à 0.5 sur une séquence d'au moins 3 acides aminés.

L'attribution des déplacements chimiques de ce fragment d'émerine, couplée à l'attribution du domaine LEM¹⁰² et de la région 1-132 (Camille Samson, communication personnelle), précédemment réalisée au laboratoire, a permis d'obtenir l'attribution de la région 1-170 de l'émerine (Figure 36).

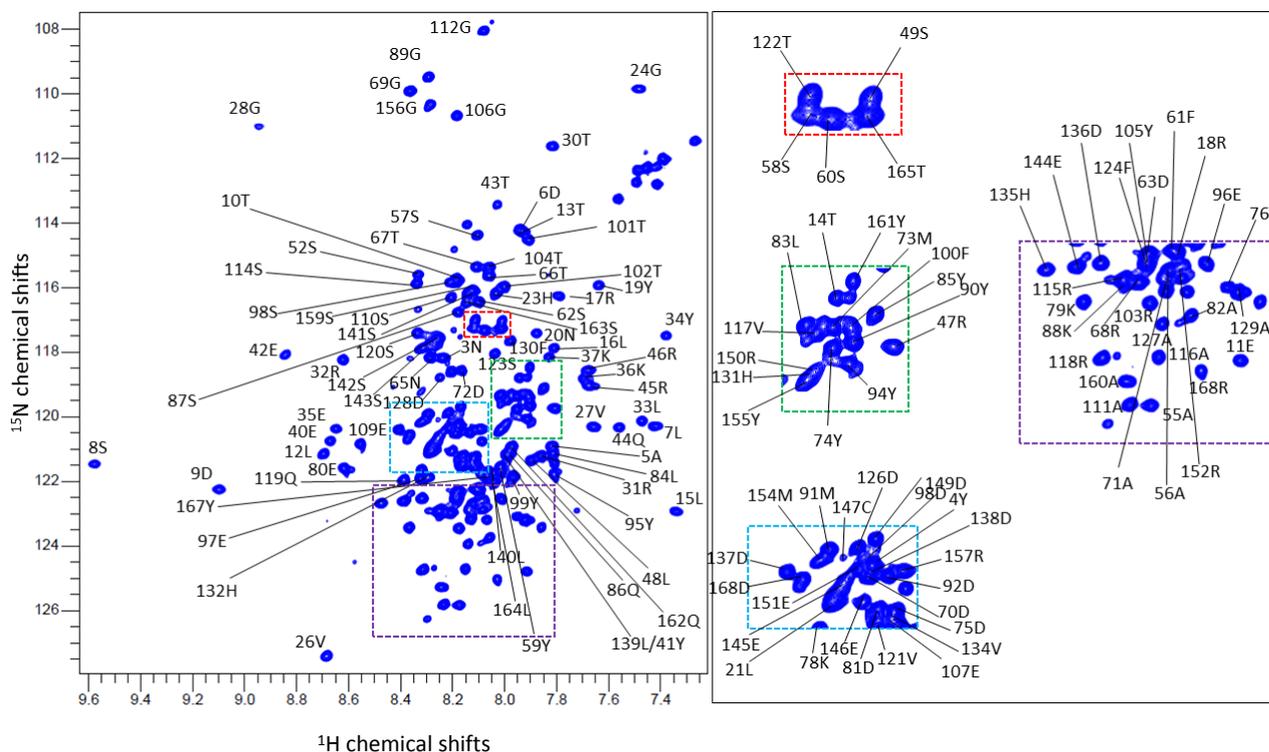


Figure 36 : Attribution de la région 1-170 de l'émerine sur le spectre 1H - ^{15}N HSQC de la construction 1-187 enregistré à 750 MHz et 303K et à 500 μ M dans un tampon 20 mM phosphate pH 6.5, 30 mM NaCl. La vue de gauche montre l'ensemble du spectre. Des zooms sont présentés sur la vue de droite pour plus de clarté.

II.1.3 Etude de l'auto-assemblage de l'émerine

II.1.3.1. Structure

Le groupe du Prof. K. Wilson a été le premier à montrer à partir d'expériences biochimiques réalisées *in vitro* et en cellules que l'émerine pouvait s'auto-associer et former des oligomères ancrés à la membrane nucléaire interne¹¹⁰. Cette étude a révélé la présence d'interactions intermoléculaires émerine-émerine et a suggéré qu'un changement dans l'état oligomérique de l'émerine pourrait réguler ses propriétés d'interaction à l'enveloppe nucléaire. Plus particulièrement, il a été montré que le domaine LEM pouvait se trouver dans un état « down » dans lequel il interagit avec la protéine BAF à la chromatine mais également dans un état « up » au cours duquel il participe à l'oligomérisation de l'émerine et à la reconnaissance potentielle de partenaires à l'enveloppe nucléaire.

Avant mon arrivée en thèse, Isaline Herrada a montré que l'émerine 1-187 était capable de s'auto-assembler pour former des oligomères organisés de type filaments¹¹¹. Cet auto-assemblage a été mis en évidence en suivant la cinétique de formation des filaments de l'émerine par fluorescence de la thioflavine T, une molécule possédant la particularité d'émettre un signal à 480 nm après avoir été excitée à 440 nm uniquement si elle est liée à des structures β . Les filaments ont été observés par microscopie électronique à transmission en coloration négative. Par la suite, j'ai identifié une autre construction capable de former des filaments : l'émerine 67-221²²⁹. A partir des images obtenues en microscopie électronique à transmission (Figure 37) ainsi que des cinétiques de formation des filaments réalisées par fluorescence à la thioflavine T (Figure 38), j'ai montré que le fragment 67-221 est non seulement capable de former des filaments, mais il forme ce type d'oligomères plus rapidement que l'émerine 1-187. En effet, le fragment 67-221 est sous forme oligomérique dès la fin de la purification : des filaments courts sont déjà visibles à $T = 0$ sur les images de microscopie électronique. En revanche j'ai montré que la partie commune à ces 2 constructions, la région 67-187, reste monomérique dans les conditions testées (Figure 38). En l'absence du domaine LEM, la région 187-221 apparaît donc primordiale pour la formation de ce type d'oligomères. Il semble donc que le domaine LEM ainsi que la région 187-221 contiennent des éléments favorisant la formation des filaments. Et si on exprime un fragment combinant 1-187 et 67-221, c'est-à-dire le fragment 1-221, alors on obtient un fragment qui s'agrège systématiquement lors de la renaturation de la protéine. Cette agrégation est peut-être due à un mauvais repliement lors de l'auto-assemblage en absence d'ancrage à la membrane.

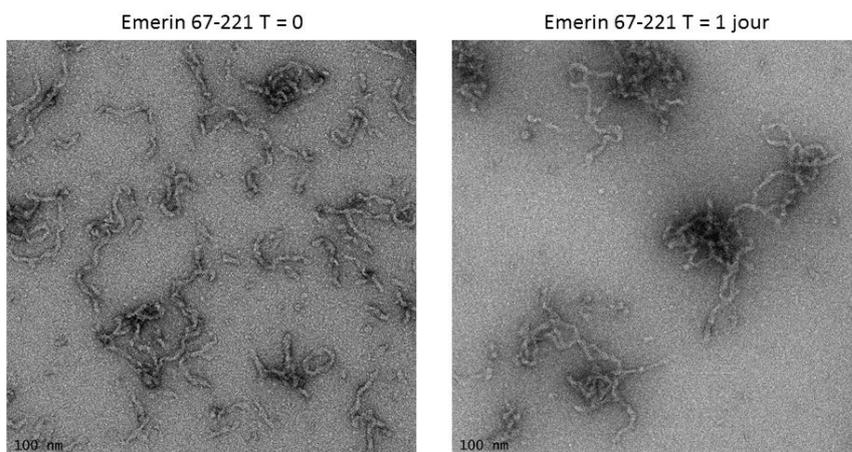


Figure 37 : Filaments obtenus avec le fragment d'émerine 67-221. Images de microscopie électronique à transmission en coloration négative, enregistrées directement après purification de l'émerine 67-221 (gauche) et après 1 jour d'incubation à 20°C (droite), en présence de 2% d'acétate d'uranyle.

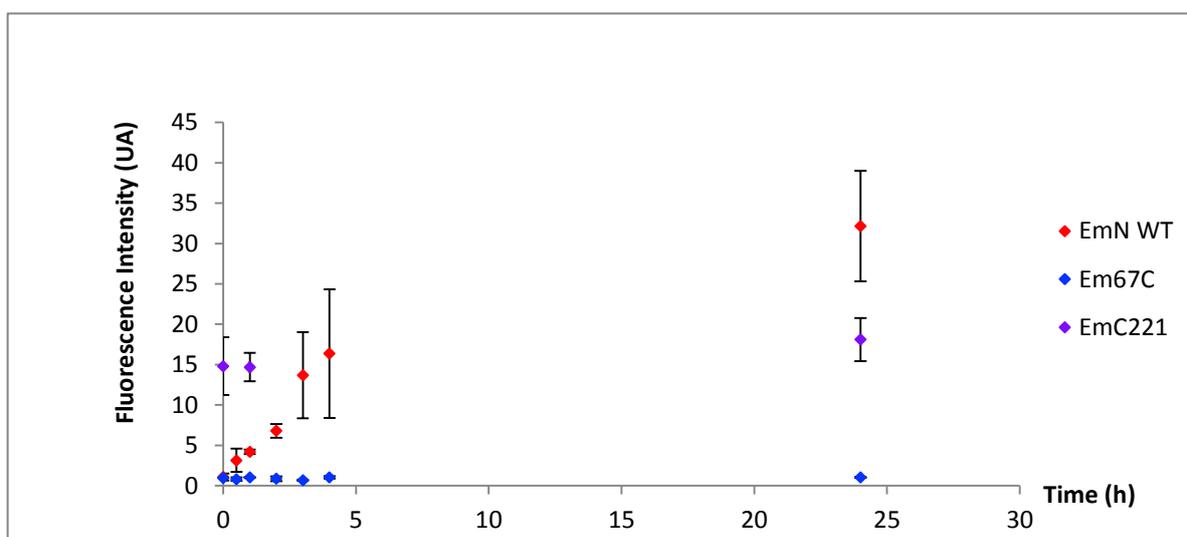


Figure 38 : Cinétique de formation des filaments de l'émerine 67-221 (notée EmC221 ; points violet), comparée aux cinétiques obtenues sur l'émerine 1-187 (notée EmN WT ; points rouge) et sur l'émerine 67-187 (notée Em67C ; points bleu). La fluorescence émise par la thioflavine T à 480 nm est suivie en fonction du temps à 37°C. Les fragments d'émerine sont concentrés à 300 μ M. Les barres d'erreur correspondent à 2 expériences.

Pendant ma thèse, Camille Samson a montré que le domaine LEM de l'émerine seul (1-49) n'est pas capable de former des filaments²⁵¹. En revanche, le fragment 1-132 est capable de s'assembler en filaments. De plus, Isaline Herrada a montré que le mutant de l'émerine 1-187 délété des acides aminés 95-99 (mutation associée à une myopathie) reste monomérique dans nos conditions¹¹¹. Ainsi, le domaine LEM mais également la région désordonnée de l'émerine sont nécessaires pour que les fragments 1-132 et 1-187 de l'émerine puissent s'auto-assembler²⁵¹. J'ai cherché à mieux définir le rôle du domaine LEM dans l'auto-assemblage du fragment 1-132 de l'émerine. Un schéma des fragments d'émerine utilisés dans cette étude est représenté en Figure 39.

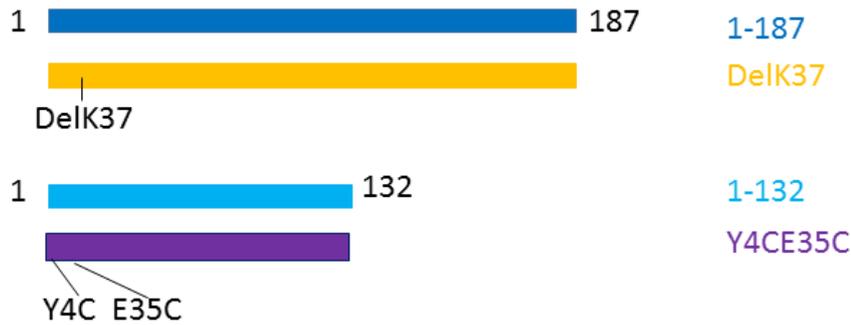


Figure 39 : Constructions utilisées lors de l'analyse du rôle du domaine LEM dans l'oligomérisation de l'émerine 1-187.

Avec Nada Essawy, nous avons montré que le mutant de l'émerine comportant une délétion de la lysine 37 (delK37 ; mutation associée à une atteinte cardiaque) a un domaine LEM désordonné²⁵¹. L'absence de la lysine 37 semble donc déstabiliser la structure de ce domaine. Etant donné l'importance du domaine LEM dans la formation des filaments de l'émerine, nous avons voulu savoir si ce mutant était capable de s'auto-assembler. La confirmation de sa capacité à former des filaments a été obtenue par microscopie électronique à transmission en coloration négative (Figure 40A). Après réalisation d'expériences de cinétiques de fluorescence à la thioflavine T, et de manière surprenante, ce mutant forme des filaments plus rapidement que l'émerine sauvage (Figure 40B). Ceci semble suggérer un changement de conformation du domaine LEM lors de la formation des filaments.

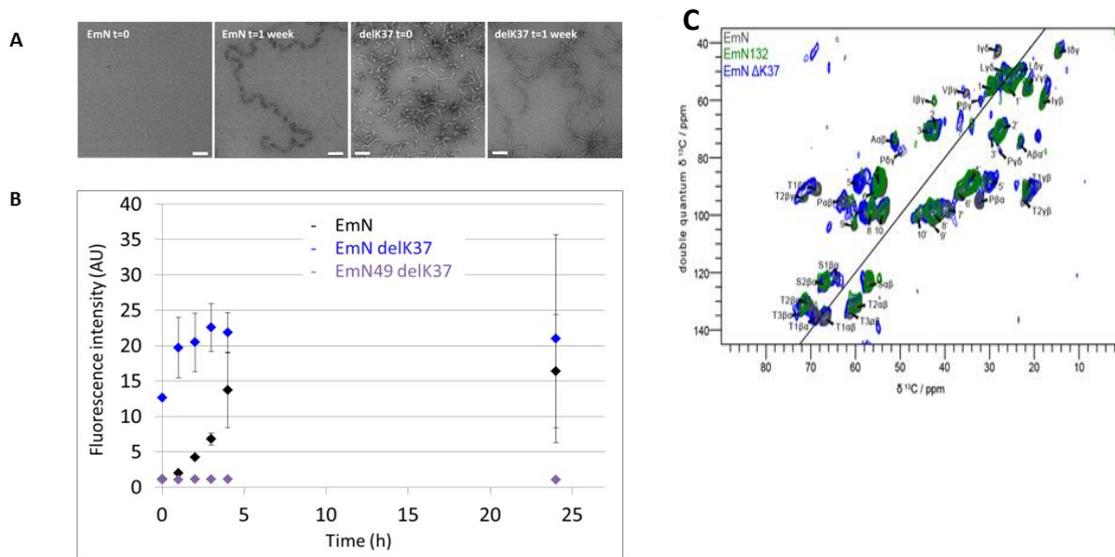


Figure 40 : Formation des filaments par le mutant delK37²⁵¹. (A) Images de microscopie électronique à transmission montrant à T=0 et T=1 semaine, l'état oligomérique du mutant de l'émerine delK37 en comparaison avec l'émerine sauvage. (B) Comparaison de la cinétique de formation des filaments de l'émerine par fluorescence à la thioflavine T. L'émerine delK37 forme des filaments plus rapidement que l'émerine sauvage. (C) Superposition des spectres de RMN du solide révélant les corrélations carbone-carbone double quanta/simple quanta correspondant à l'émerine 1-187 (noir), l'émerine 1-132 (vert) et au mutant delK37 (bleu) (Collab. Prof A. Lange, FMP Berlin).

Nous avons donc posé l'hypothèse d'un dépliement du domaine LEM pour permettre la polymérisation de l'émerine. Il est à noter que la comparaison par RMN du solide des filaments formés par le mutant delK37 et ceux formés par l'émerine sauvage montre qu'ils possèdent la même structure puisque les signaux se superposent parfaitement (Figure 40C).

Pour confirmer notre hypothèse selon laquelle il est nécessaire que le domaine LEM de l'émerine soit déplié pour permettre l'auto-assemblage de l'émerine sous forme de filaments, j'ai produit un mutant artificiel de l'émerine 1-132 (Y4CE35C) dans lequel j'ai ajouté un pont disulfure grâce à la mutation des résidus 4 et 35 en cystéines (Figure 41A). De cette façon, à l'état oxydé, le domaine LEM de la protéine devrait rester bloqué sous forme replié. J'ai ensuite essayé de caractériser la capacité de ce mutant à former des filaments. Pour cela, j'ai réalisé des expériences de cinétiques de formation de filaments par fluorescence à la thioflavine T (Figure 41B). D'après mes résultats, lorsque la protéine est oxydée, c'est-à-dire en conformation fermée, elle n'est pas capable de former des filaments. Ceci confirme bien que lorsque le domaine LEM n'a pas la capacité de changer de conformation, l'auto-assemblage n'est plus possible. Par contre, après réduction par ajout de 5 mM DTT, la cinétique de formation des filaments du mutant Y4CE35C est similaire à celle de l'émerine 1-132 sauvage (Figure 41B). J'ai enregistré des spectres $^1\text{H}^{15}\text{N}$ HSQC avec et sans agent réducteur du mutant. Le spectre du mutant oxydé montre que le domaine LEM est structuré (Figure 41C). Après ajout de 5 mM DTT, j'observe que les signaux structurés du domaine LEM disparaissent pour migrer vers le centre du spectre, confirmant l'aspect désordonné de la protéine résultante (Figure 41C). Cette expérience nous permet donc d'affirmer qu'il est nécessaire pour le domaine LEM de subir un changement de conformation pour permettre la formation des filaments de l'émerine.

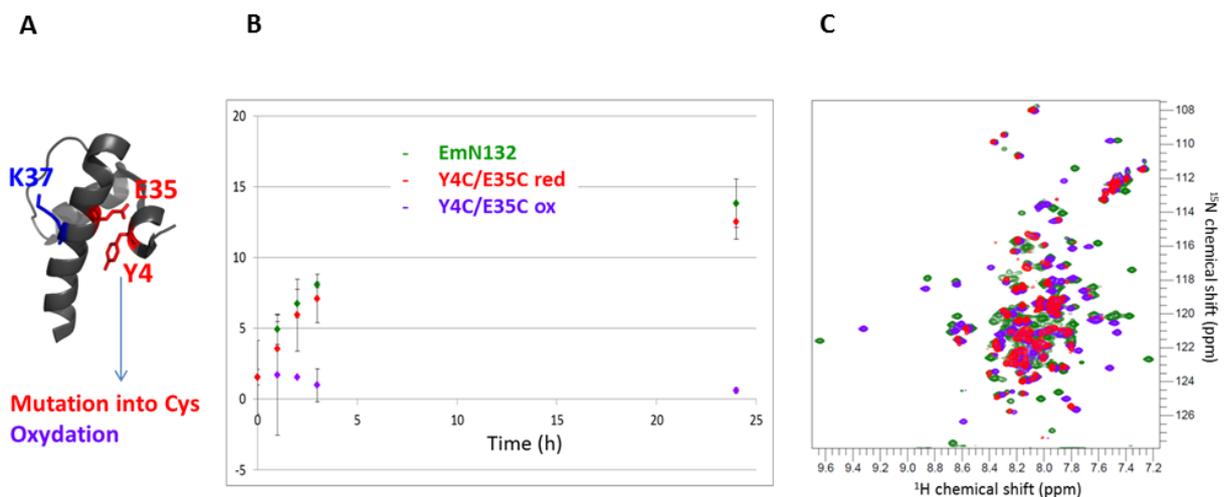


Figure 41 : Le domaine LEM de l'émerine est destructuré lors de la formation des filaments. (A) Structure tridimensionnelle du domaine LEM avec en bleu, la lysine 37 absente chez le mutant delK37, et en rouge, les deux résidus mutés en cystéines chez le mutant Y4CE35C. (B) Cinétique d'assemblage obtenue en mesurant la fluorescence émise par la thioflavine T au cours du temps, montrant l'incapacité du domaine LEM à former des filaments lorsqu'il est maintenu sous forme repliée (Violet). Un dépliement du domaine LEM par réduction du pont disulfure (rouge) permet de retrouver une cinétique semblable à l'émerine sauvage (vert). (C) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de l'émerine 1-132 sauvage (vert), du mutant Y4CE35C sous forme oxydée (violet) et réduite (rouge).

En conclusion, j'ai montré que les émerines 1-132 et 1-187 subissaient un dépliement de leur domaine LEM pour permettre leur auto-assemblage sous forme de filaments, et que la région désordonnée de l'émerine 187-221 ne contenant pas le domaine LEM était aussi capable de s'auto-assembler. Cependant, la construction 67-187 sans domaine LEM mais allant jusqu'au résidu 187 n'est pas capable de former des filaments. En l'absence du domaine LEM, la région 187-221 apparaît donc primordiale pour la formation de ce type d'oligomères.

II.1.3.2. Fonction

Les régions suffisantes de l'émerine permettant son interaction avec la lamine A ont été mises en évidence grâce à des GST-pulldowns réalisés sur des fragments d'émerine possédant une étiquette GST, auxquels la queue C-terminale de la lamine A mature (385-646) a été ajoutée¹¹⁰. Ainsi, les fragments de l'émerine 1-132 et 1-149 semblent permettre la liaison à la lamine A. Une seconde région a été rapportée dans cette étude comme permettant l'association avec la lamine A : il s'agit de la région 159-220. Par ailleurs, il est intéressant de noter que le fragment 170-220 de l'émerine, testé lui aussi, ne permet pas de liaison à la lamine A, indiquant un rôle important des acides aminés 159-170 dans cette interaction.

Afin de comprendre le rôle des régions déstructurées de l'émerine dans cette interaction, en collaboration avec Camille Samson au laboratoire, nous avons étudié l'interaction entre différents fragments de l'émerine et la lamine A. Ainsi, nous avons réalisé par RMN des études d'interaction entre les différents fragments d'émerine monomérique et la région queue de la lamine A. Après plusieurs tentatives, aucune interaction directe n'a été observée avec les monomères d'émerine et la région queue 389-646 de la lamine A²²⁹, quel que soit le fragment d'émerine considéré. Nous avons donc souhaité savoir si l'oligomérisation de l'émerine pouvait avoir une influence sur l'interaction avec la lamine A. Nous avons donc formé des filaments de l'émerine (1-187 et 67-221), et étudié l'interaction des oligomères formés par ces fragments avec le domaine IgFold de la lamine A par RMN (Figure 42A et B). Pour chaque interaction étudiée, l'état filamentaire de l'émerine a été vérifié en parallèle par microscopie électronique à transmission. Les résultats obtenus par RMN montrent une interaction avec les oligomères de l'émerine 1-187 (Figure 42A), cependant, aucune interaction n'est observée avec les filaments formés par le fragment 67-221 (Figure 42B). Ceci nous permet de conclure quant à l'importance du domaine LEM dans l'interaction de l'émerine avec la lamine A. De plus, la formation d'oligomères d'émerine est nécessaire pour obtenir une interaction directe avec le domaine IgFold de la lamine A²²⁹.

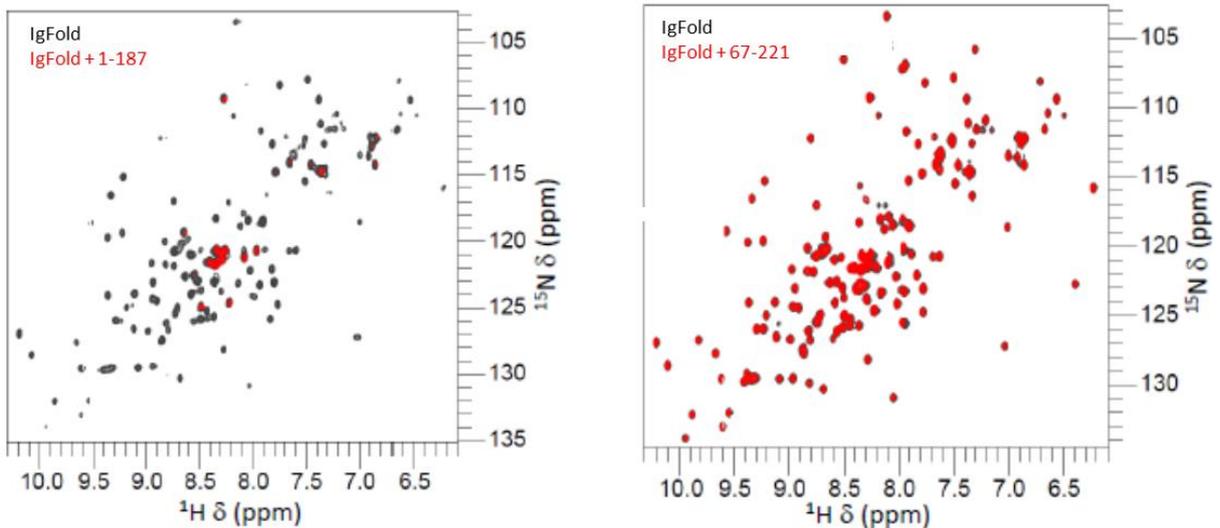


Figure 42 : Interaction entre le domaine globulaire IgFold de la lamine A et les oligomères d'émerine²³⁰. Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de l'IgFold seul (noir) ou en présence des oligomères d'émerine formés par le fragment 1-187 (gauche ; rouge) ou 67-221 (droite ; rouge). Les spectres ont été enregistrés à 700 MHz et 303 K dans un tampon phosphate 20 mM, pH 6.5 ; NaCl 150 mM ; 2 mM DTT ; 10% D_2O et 150 μM de protéine finale.

II.2. Détermination d'un facteur de forme des fibres au cours du temps Etude de filaments protéiques grâce au système SF-WA-MALS-SAXS

Il a été montré au laboratoire que l'émerine était capable de former des filaments¹¹¹. Afin d'obtenir des détails moléculaires permettant de suivre l'assemblage des filaments de l'émerine au cours du temps, une collaboration avec le Dr. Javier Pérez de la ligne SWING du Synchrotron SOLEIL a été initiée. En effet, un système couplant un appareil de mélange rapide, ainsi que l'analyse de diffusion des rayons X aux petits angles et de diffusion de la lumière aux grands angles a été développé par les Drs Javier Pérez, Mattia Rocco et Patrice Vachette²⁵². Ce système sera appelé SF-WA-MALS-SAXS dans la suite de ce travail. Ce système permet d'effectuer des expériences répétées avec détection simultanée par WA-MALS et SAXS tout en évitant une trop longue exposition du même volume d'échantillon aux rayons X très intenses qui pourraient dégrader l'échantillon durant l'acquisition. La description du formalisme permettant l'analyse des données recueillies ainsi que leur modélisation est visible en annexe.

II.2.1 Adaptation du modèle émerine à ce type d'expérience

L'une des principales limitations du système de SF-WA-MALS-SAXS réside dans l'obligation pour le modèle étudié d'être parfaitement monomérique au départ. En effet, comme les espèces de hauts poids moléculaires auront une pondération plus importante dans l'équation finale, afin de détecter les premières étapes de la cinétique, il est crucial d'avoir un objet monomérique à T_0 .

Le protocole de formation des filaments de l'émerine nécessite de concentrer la protéine à environ $500 \mu\text{M}$ afin de lancer la polymérisation. A cette concentration, l'émerine se situe sous plusieurs formes oligomériques comme le montre la Figure 43. En effet, on peut voir que des oligomères de grande taille sont déjà visibles à $T = 0 \text{ min}$ avec un diamètre apparent de 50 nm .

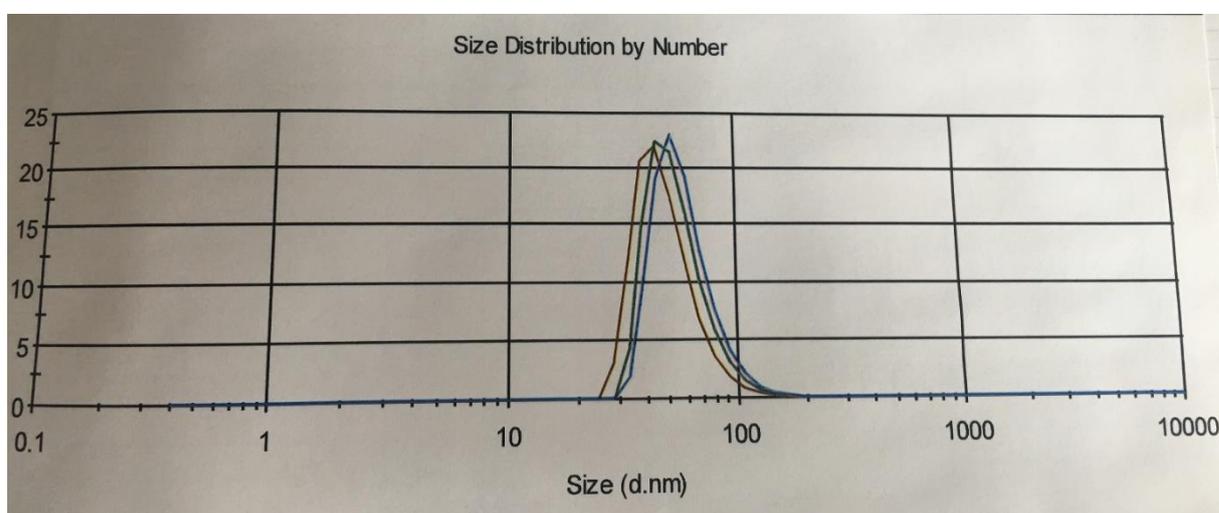


Figure 43 : Profil de diffusion de lumière dynamique, illustrant le caractère oligomérique de l'émerine à $500 \mu\text{M}$ à $T = 0 \text{ min}$.

De plus, à une telle concentration, la formation des filaments de l'émerine nécessite une semaine à 20°C . La cinétique de formation de ces filaments peut être réduite à 1h si la température est augmentée à 37°C . Malheureusement, une température élevée engendrerait des effets délétères pour le système de fluide utilisé. L'utilisation du faisceau de rayons X sur la ligne de SAXS ne doit être que temporaire et une expérience d'une semaine n'est pas envisageable. Aussi, afin d'utiliser l'émerine comme modèle pour ce type d'expérience, il nous a fallu essayer d'obtenir une émerine monomérique à très haute concentration. Cette concentration doit être suffisamment élevée pour entraîner une cinétique rapide de formation des filaments, compatible avec l'utilisation de la ligne de SAXS.

Lors de l'étude des filaments de l'émerine, j'ai pu montrer qu'un mutant artificiel de l'émerine (Y4CE35C) dans lequel un pont disulfure avait été introduit entre les résidus 4 et 35, mutés tous les deux en cystéines, n'était plus capable de former de filaments sous forme

oxydée (Figure 41B) et restait monomérique même à haute concentration, comme le montrent les expériences de cinétique de fluorescence à la thioflavine T. La réduction de ce mutant Y4CE35C permet à la protéine de retrouver une cinétique de formation des filaments semblable à celle de l'émerine WT 1-132 (Figure 41B). Les expériences $^1\text{H}^{15}\text{N}$ HSQC par RMN ont confirmé que la protéine était sous forme repliée à l'état oxydé et se déplaçait lors de sa réduction, puisque les signaux sont visibles dans la région des acides aminés n'ayant pas de structure secondaire stable entre 7.5 et 8.5 ppm (Figure 41C).

Le mutant Y4CE35C semble donc adapté à l'expérience de SF-WA-MALS-SAXS. En injectant une solution réductrice à T0, il devrait être possible de déclencher la formation rapide des filaments de l'émerine, analysable par SAXS dans un court délai de temps. Malheureusement, les essais de réduction avec des concentrations variables de DTT (0.5 ; 1 ; 2 ; 5 mM) réalisés sur cette protéine ont montré qu'à haute concentration, une forte agrégation était visible. Cette agrégation est incompatible avec notre système d'analyse puisque des objets d'aussi grande taille rendraient invisibles toute autre forme oligomérique intéressante. En conclusion, il n'a pas été possible d'adapter l'émerine à notre système. J'ai alors décidé de poursuivre le développement du système SF-WA-MALS-SAXS en utilisant des échantillons de fibrinogène (FG) produit par l'équipe du Dr. Mattia Rocco.

II.2.2 Développement du système SF-WA-MALS-SAXS à l'aide du Fibrinogène

II.2.2.1. Automatisation du système de SF-WA-MALS-SAXS

Pour réaliser des expériences de polymérisation et enregistrer des données de SAXS sur l'échantillon, il est nécessaire de pouvoir réaliser toutes les opérations à distance, de façon à ne jamais être exposé aux rayons X de la ligne de lumière. Ainsi, il a été convenu d'utiliser un système de robot permettant l'injection des échantillons dans le système ainsi que l'enregistrement de divers types d'expériences combinant SAXS et WA-MALS ou permettant d'isoler l'un ou l'autre des appareils. Ce robot d'injection permet également le lavage du système complet avec du tampon. Un second robot, spécifiquement pour le lavage était déjà présent sur le système à mon arrivée au laboratoire et permettait d'injecter grâce à un système de vide entraînant une aspiration de différentes solutions : Hellmanex, acide acétique, tampon, eau distillée.

J'ai participé à la conception du robot d'injection et du schéma du système complet permettant son intégration. J'ai également participé à la programmation de ce robot d'injection comportant une carte Arduino® capable de piloter les 5 vannes et 2 sélecteurs qu'il comprend. Par ailleurs, pour réaliser les expériences de polymérisation, plusieurs

configurations du système SF-WA-MALS-SAXS ont été nécessaires. Ainsi, j'ai programmé des séquences permettant d'obtenir les différentes configurations grâce au logiciel Bossanova. Le système complet sera utilisable facilement par un expérimentateur venant au Synchrotron SOLEIL réaliser ce type d'expériences. Les différentes configurations programmées ainsi que le code du robot d'injection sont visibles en annexe. Le système complet, connecté sur la ligne de SAXS, est visible en Figure 44.

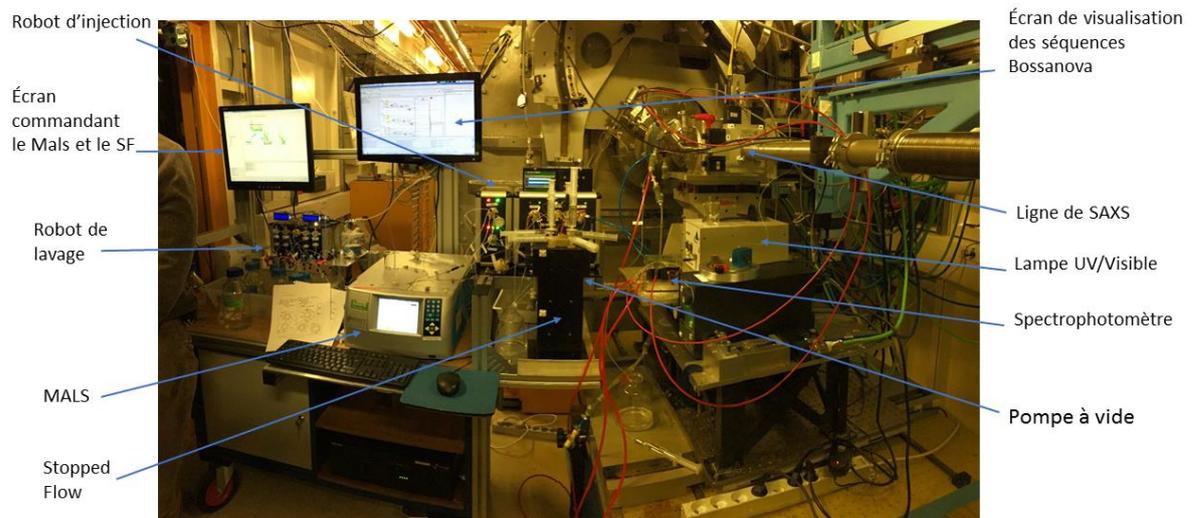


Figure 44 : Système SF-WA-MALS-SAXS complet, installé sur la ligne SWING.

La Figure 45 montre un schéma du circuit dans la configuration de lavage. Sur ce schéma, deux sélecteurs sont présents : les sélecteurs 2 et 3 notés S2 et S3 respectivement. La position dans laquelle se situe un sélecteur est donnée par le numéro suivant le « P ». Par exemple, si le sélecteur 2 est en position 1, il sera indiqué S2P1. Le circuit comporte également 5 vannes notées de V1 à V5 avec chacune, ayant la possibilité de se trouver en position 1 ou 2. La même nomenclature que pour les sélecteurs a été utilisée pour indiquer cette position. Après isolation du MALS, le sélecteur 3 est placé en position 4, permettant ainsi de faire le vide dans le circuit. Le sélecteur 2 nous permet alors de choisir le solvant à insérer dans le circuit pour le lavage. Ainsi, nous pouvons laver le circuit avec de l'eau distillée, du Hellmanex, du tampon ou de l'acide acétique. Il faut cependant ne pas oublier qu'un lavage avec de l'eau distillée doit toujours être fait entre l'Hellmanex et l'acide acétique pour empêcher tout risque de précipitation à l'intérieur du circuit.

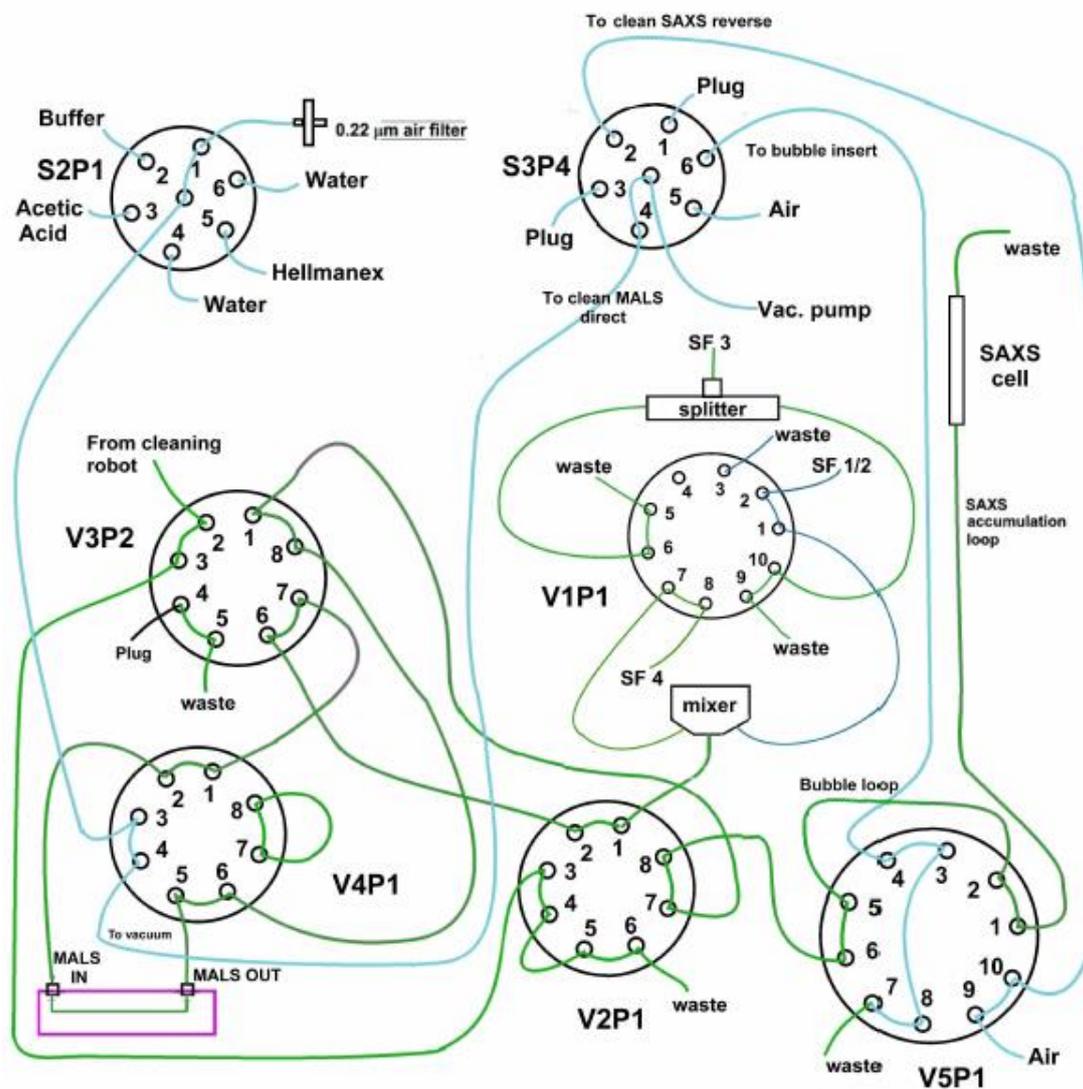


Figure 45 : Lavage du circuit par le vide apporté grâce au robot d'injection. Les connecteurs verts contiennent du tampon, les violets, du mélange réactionnel et les bleus, du vide.

Après lavage du système, il est nécessaire de remplir les circuits de tampon afin de réaliser un « blanc » et obtenir une ligne de base avant analyse de données issues d'un mélange réactionnel. La configuration décrite en Figure 46, permet d'injecter du tampon successivement dans le MALS, puis dans le SAXS à partir de chacune des seringues 1, 2 ou 4 seules ou en combinaison. Le sélecteur 3 est placé en position 1, sur un bouchon, empêchant l'entrée d'air dans le circuit. Les seringues 1, 2 ou 4 injectent le tampon vers la vanne 1 placée en position 1, permettant l'arrivée du tampon au mixeur, à la vanne 2 en position 1, dirigeant le tampon vers la vanne 3 en position 2 afin d'arriver au MALS. Une fois le MALS parcouru, le tampon passera par les vannes 4, 2 et 5 pour finir dans le capillaire de SAXS. D'autres configurations ont été réalisées pour insérer du tampon dans le capillaire de SAXS uniquement ou dans le MALS uniquement. Mais seule la configuration mixte est présentée ici.

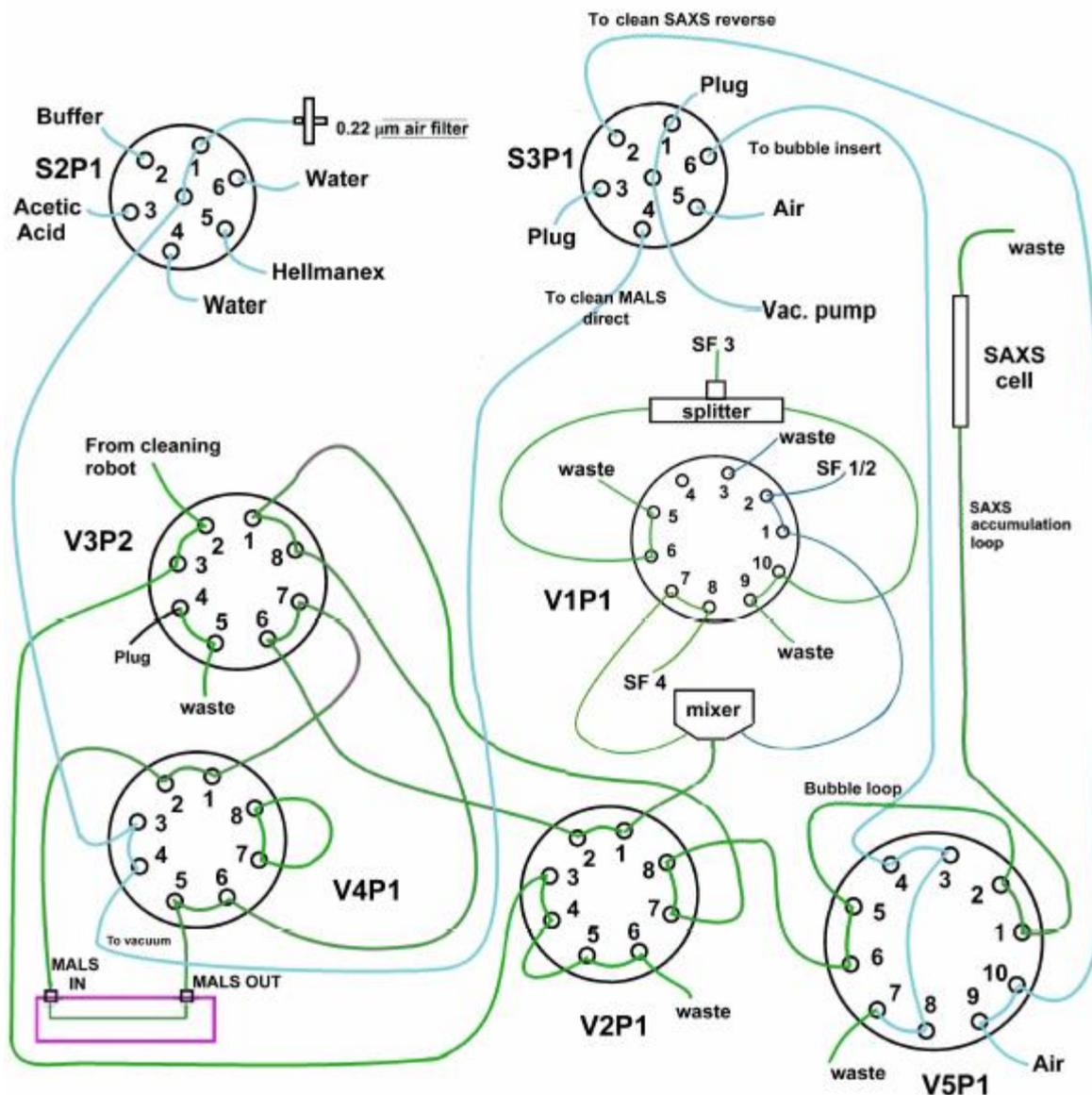


Figure 46 : Insertion de tampon dans le circuit afin d'obtenir une ligne de base. Les connecteurs verts contiennent du tampon, les violets, du mélange réactionnel et les bleus, du vide.

Lors de l'injection d'un mélange réactionnel à partir des seringues 1, 2 et 4 du SF, la configuration du circuit ne change pas, il nous suffit de remplir les seringues avec les échantillons réactionnels afin de déclencher la polymérisation. La Figure 47 représente la configuration d'injection avec le parcours de l'échantillon réactionnel au sein du circuit en violet.

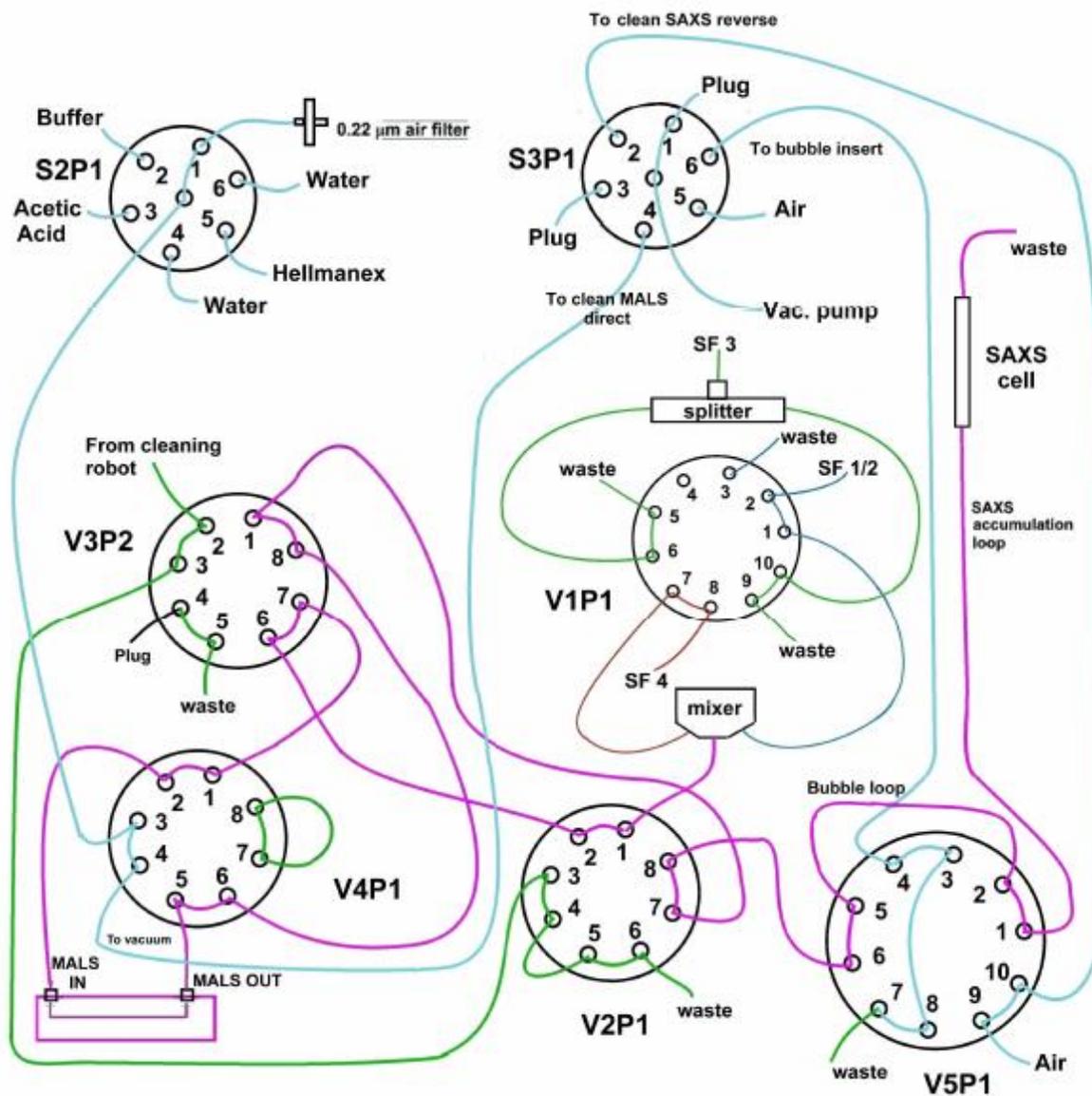


Figure 47 : Injection de mélange réactionnel vers le MALS et le capillaire de SAXS. Les connecteurs verts contiennent du tampon, les violets, du mélange réactionnel et les bleus, du vide.

II.2.2.2. Insertion d'une bulle d'air

Lors d'une injection de solution protéique, nous avons remarqué qu'après un certain laps de temps, la densité optique mesurée au niveau du capillaire SAXS diminuait (en bleu sur Figure 48). Ceci est dû au fait que le tampon qui « pousse » l'échantillon pour éviter la dégradation sous faisceau X dilue progressivement l'échantillon. Afin d'éviter cela et garder une concentration constante au cours du temps, et pouvoir analyser sans biais l'échantillon présent en solution et ses différentes espèces oligomériques, nous avons décidé d'insérer une bulle d'air après l'injection du mélange réactionnel afin de minimiser le phénomène de dilution au sein du circuit.

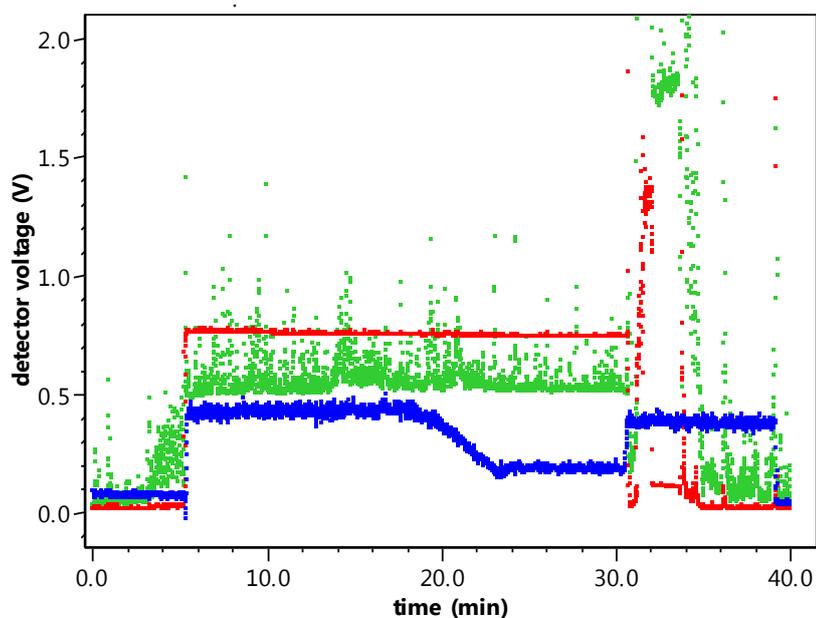


Figure 48 : Profil de données enregistrées avec le logiciel ASTRA lors d'une injection de BSA illustrant la diminution de la densité optique par dilution avec le tampon. 0' : lavage avec du tampon. 2.5' : injection de 1.5 ml de BSA en 15 s. Entre 20' et 23', la densité optique (courbe bleue) diminue à cause de la dilution de l'échantillon par le tampon. 30' : injection de 1.5 ml de BSA en 15 s, , 34' lavage avec du tampon. La courbe verte représente les valeurs enregistrées par le détecteur 4, la rouge, celles enregistrées par le détecteur 11.

La configuration d'injection du mélange réactionnel dans le système nécessite que le sélecteur 3 soit en position 1 (Figure 50). La bulle d'air est insérée grâce à une pompe à vide permettant d'aspirer de l'air. Pour ce faire, il est nécessaire de positionner le sélecteur 3 en position 6 (Figure 51). La détection de la bulle d'air dans la boucle d'accumulation par le détecteur UV est visible en Figure 49.

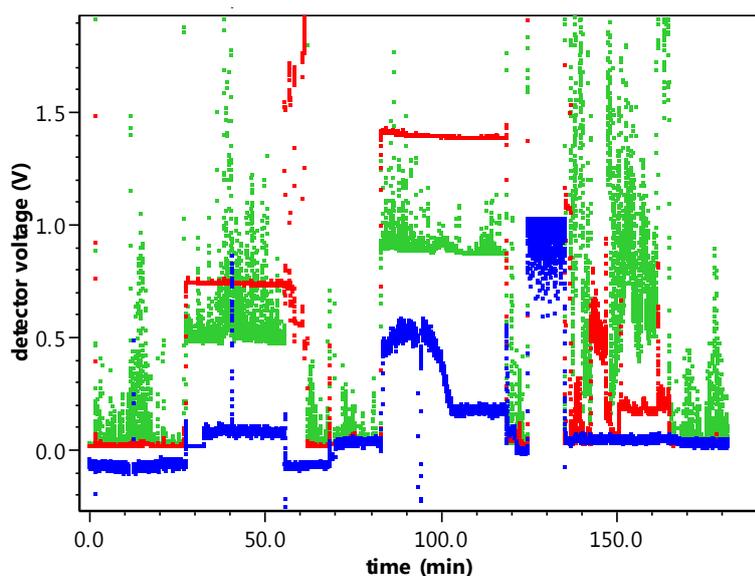


Figure 49 : Profil de données MALS et UV. L'UV montre la présence de la bulle d'air dans la boucle d'accumulation de SAXS. 0' : lavage avec du tampon. 30' injection d'1 ml de BSA en 15 s. 80' passage de la bulle injectée après l'échantillon dans la boucle d'accumulation. Plusieurs lavages ont ensuite été effectués avec du tampon. La courbe bleue représente la valeur de densité optique mesurée, la courbe verte correspond aux valeurs mesurées par le détecteur 4, la rouge, à celles du détecteur 11.

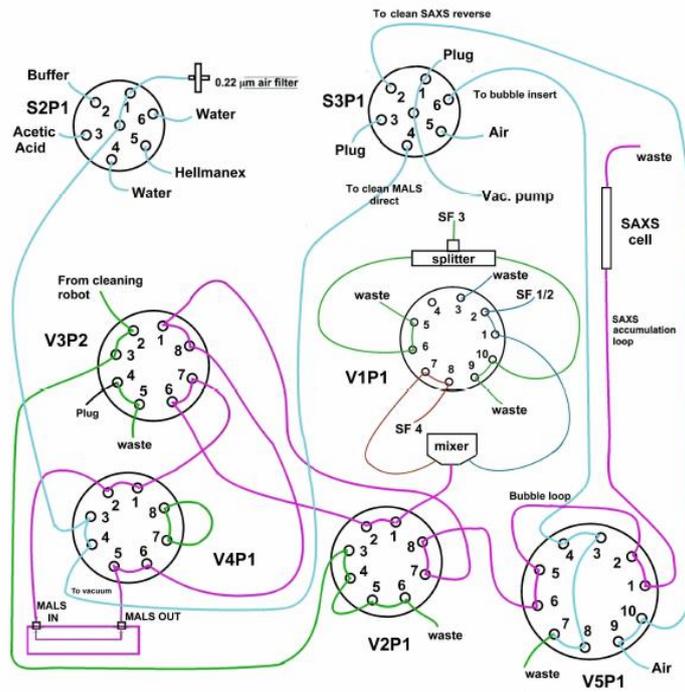


Figure 50 : Configuration du robot d'injection lors de l'injection d'un mélange réactionnel. Cette séquence permet l'injection du mélange réactionnel successivement dans le MALS puis dans le capillaire de SAXS. Cette séquence sera immédiatement suivie de la séquence d'injection d'une bulle d'air. Les connecteurs verts contiennent du tampon, les violets, du mélange réactionnel et les bleus, du vide.

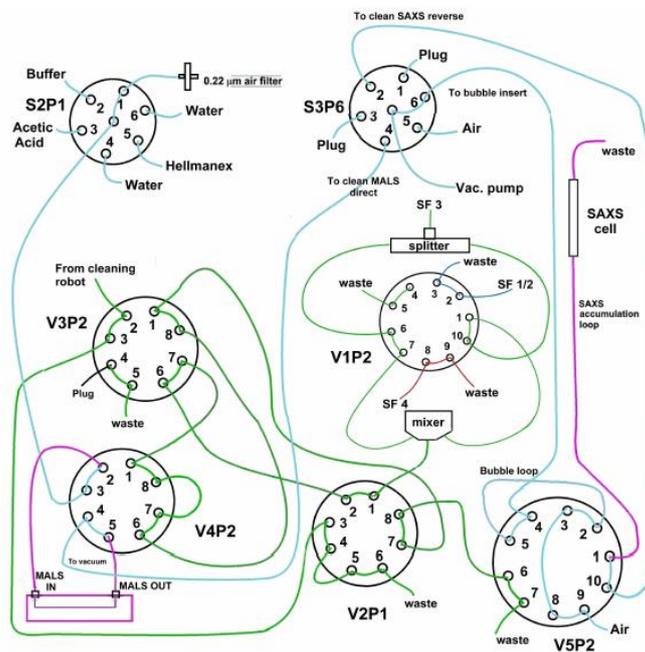


Figure 51 : Configuration du robot d'injection permettant l'injection d'une bulle d'air après l'échantillon protéique. Le sélecteur 3 est placé en position 6 permettant l'aspiration d'air. Ceci doit impérativement être réalisé avec la valve 5 en position 1 pour éviter le reflux de l'échantillon lorsque le sélecteur 3 traverse toutes les positions pour arriver en position 6. La valve 5 est ensuite placée en position 2 pour permettre l'aspiration de la bulle d'air. Les connecteurs verts contiennent du tampon, les violets, du mélange réactionnel et les bleus, du vide.

II.2.2.3. Etude de polymérisation du fibrinogène

Afin de tester le système lors d'une polymérisation, nous avons injecté du fibrinogène en présence de thrombine à l'aide des seringues 1 et 2 du SF. Après injection d'une bulle d'air pour éviter le phénomène de dilution, nous pouvons visualiser une augmentation de la tension détectée par les détecteurs 4 (vert) et 11 (rouge), signifiant l'augmentation de la taille des oligomères présents en solution (Figure 52). Après injection de tampon, nous pouvons voir un retour à la ligne de base du signal. Il nous est donc possible de réaliser plusieurs expériences de polymérisation successivement.

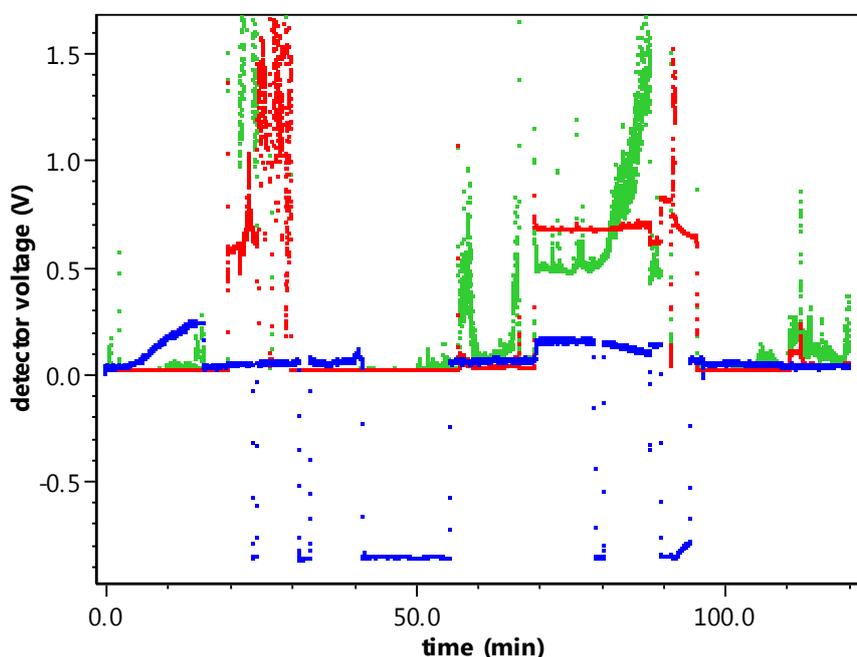


Figure 52 : profil de données MALS illustrant une expérience de polymérisation du fibrinogène à l'aide d'une solution de thrombine. $T = 0$ min : injection de 750 ml de FG ainsi que de 750 ml de thrombine. $T = 28$ min, lavage avec du tampon. $T = 54$ min, injection de 500 ml de FG avec 500 ml de tampon. $T = 90$ min : lavage avec du tampon. La courbe bleu représente la valeur de densité optique mesurée, la courbe verte correspond aux valeurs mesurées par le détecteur 4, la rouge, à celles du détecteur 11.

Ces données obtenues durant la polymérisation du fibrinogène en présence de thrombine peuvent alors être analysées en utilisant la fonction de Fisher-Burford décrite en annexe et visible en Figure 53. Cette fonction nous permettra d'obtenir à chaque instant, le carré du rayon de giration moyen $\langle R_g \rangle^2$ ainsi que la masse moléculaire moyenne $\langle M \rangle_w$. Ce qui nous donnera un facteur de forme en temps réel des oligomères formés. Nous pouvons visualiser que le $\langle R_g \rangle^2$ augmente au même titre que la $\langle M \rangle_w$, montrant bien que la taille des oligomères augmente au cours de la polymérisation.

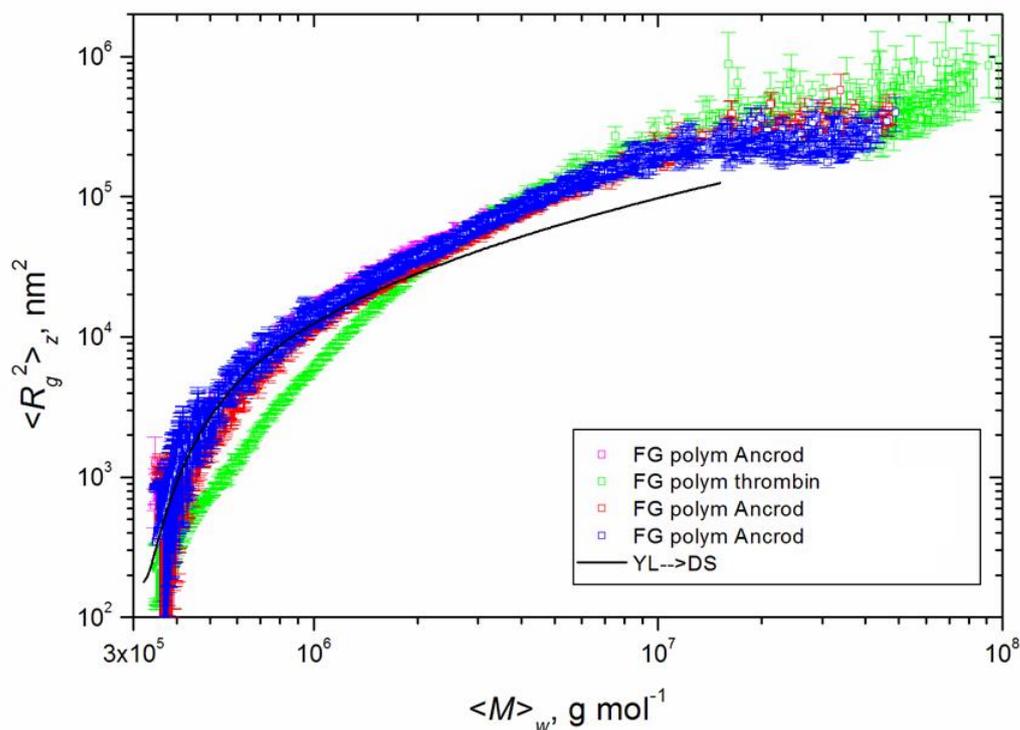


Figure 53 : Exemple de modélisation des données de polymérisation du fibrinogène obtenues avec le MALS et analysées avec la méthode de Fisher-Burford. Des expériences de polymérisation du fibrinogène ont été réalisées par ajout de thrombine ou d'ancrod. Les valeurs de $\langle R_g^2 \rangle$ et de $\langle M \rangle_w$ peuvent alors être obtenues.

En conclusion, le système est maintenant automatisé et pilotable à distance. Il est également prêt à accueillir des utilisateurs sur la ligne SWING pour réaliser des expériences de polymérisation de protéines. Enfin, les données recueillies pourront être analysées grâce aux méthodes de Fisher-Burford pour les données de MALS ainsi que de Guinier pour les données de SAXS.

Les perspectives concernant ce travail seront de veiller à la reproductibilité des expériences de polymérisation, mais également la génération d'un algorithme permettant la modélisation à la fois des données issues du MALS et du SAXS en même temps, sans passer par deux fonctions séparées (Fisher-Burford et Guinier respectivement).

II.3. Etude de la région C-terminale intrinsèquement désordonnée de la lamine A

La région C-terminale de la lamine A est le siège de nombreux événements de modifications post-traductionnelles. En particulier, lors de la maturation de cette protéine, le précurseur de la lamine A, la prélamine A, est tour à tour farnésylé (ajout de 15 carbones), coupé (les 3 acides aminés C-terminaux sont ôtés), carboxyméthylé et enfin encore coupé (les 15 acides aminés C-terminaux sont retirés) ce qui provoque la perte du résidu farnésylé. Par ailleurs, cette région C-terminale interagit avec plusieurs partenaires protéiques de la lamine A. Enfin, la progérine est un mutant de la lamine A ayant subi une mutation ponctuelle entraînant la délétion de 50 acides aminés. Parmi ces acides aminés se trouve le site de clivage qui permet, chez la protéine sauvage, de retirer le groupement farnésyl. La progérine est par conséquent constitutivement farnésylée. Ceci modifie de manière significative ses propriétés de liaison.

Les régions C-terminales de la prélamine A et de son mutant progérine sont prédites comme intrinsèquement désordonnées. Ceci peut être montré par exemple en utilisant le serveur de prédiction de désordre Disopred3 (Figure 54A). Même si le graphique obtenu prédit une absence de structure tertiaire, certaines régions semblent avoir une tendance à adopter une structure locale à l'instar des régions 576-585 et 641-650 (Figure 54A). La région 641-650 est le lieu du deuxième site de clivage (entre les positions 646 et 647), permettant de retirer le groupement farnésyl en fin de maturation.

Concernant les résultats obtenus grâce au serveur Pspred, le graphique obtenu montre que certaines zones ont une tendance à former une structure secondaire, notamment la région 586-589 (prédite comme formant un brin β ; Figure 54B) ainsi que la région 646-649 située au niveau du deuxième site de coupure, permettant le retrait final du groupement farnésyl en fin de maturation (aussi prédite comme formant un brin β par Pspred ; Figure 54B). La prédiction de structure secondaire obtenue avec le serveur Pspred met de plus en évidence une autre région susceptible de former des structures β , entre les acides aminés 620 et 628 (Figure 54B). Cette région, comme la région du deuxième site de coupure, est absente dans la progérine. Les deux régions comprennent un même motif Val-Thr-Arg-Ser-Tyr, de fonction inconnue.

D'après la Figure 55, il est intéressant de noter que la région entre les résidus 579 et 585, présente à la fois dans la prélamine A et la progérine, est particulièrement conservée. Par ailleurs, sur les quatre cystéines de ces régions désordonnées, 3 sont conservées. L'une, en position 661 est le lieu de la farnésylation de la protéine, les deux autres cystéines conservées se situent en positions 588 et 591, donc proches de la séquence conservée. Enfin, l'un des brins β prédits, entre les résidus 585 et 588, se situe aussi proche de la séquence conservée.

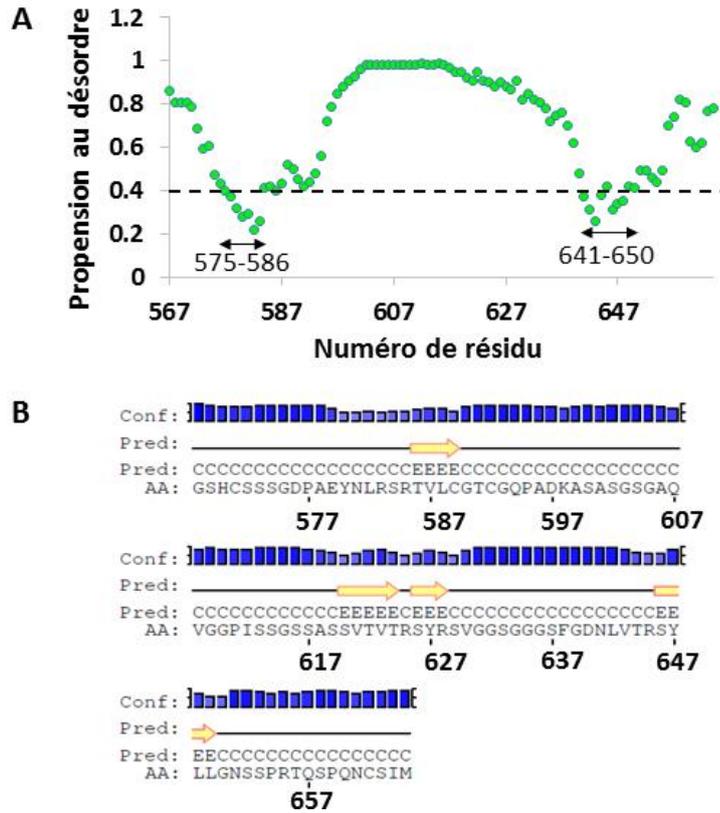


Figure 54 : Analyse bioinformatique de la séquence de la région C-terminale de la prélamine A (567-664) (<http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1>). (A) : Prédiction de propension au désordre grâce au serveur Disopred3. La propension au désordre est représentée en fonction de la séquence en acides aminés (courbe verte). (B) Prédiction de structure secondaire grâce au serveur Psipred. La structure prédite est affichée en fonction de la séquence en acides aminés. L'indice de confiance de la prédiction est représenté par un gradient de couleur bleue.

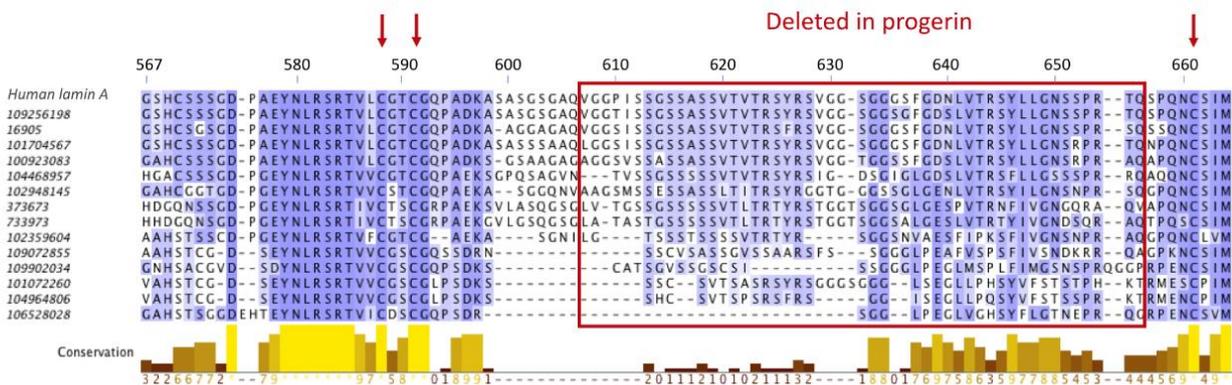


Figure 55 : Alignement de séquence de la région C-terminale de la prélamine A en fonction des espèces. Le taux de conservation peut être visualisé par l'intensité de la couleur bleue sur la séquence ainsi que par les histogrammes du diagramme de conservation calculés par Jalview (Whaterhouse et al., 2009). La région absente chez le mutant progérine est encadrée en rouge. Les flèches rouges indiquent la position des cystéines conservées. La première séquence correspond à la prélamine A humaine. Les autres séquences sont référencées grâce à leur numéro d'identification GenBank. 109256198 : Panthera pardus ; 16905 : Mus musculus ; 101704567 : Heterocephalus glaber ; 100923083 : Sarcophilus harrisii ; 104468957 : Pterocles guttularis ; 102948145 : Chelonia mydas ; 373673 : Xenopus laevis ; 733973 : Xenopus tropicalis ; 102359604 : Latimeria chalumnae ; 109072855 : Cyprinus carpio ; 109902034 : Oncorhynchus kisutch ; 101072260 : Takifugu rubripes ; 104964806 : Notothenia coriiceps ; 106528028 : Austrofundulus limnaeus.

Afin d'obtenir plus d'informations sur la structure des régions C-terminales de la progérine et de la prélamine A, j'ai produit et étudié par RMN les fragments C-terminaux de la progérine et de la prélamine A humaines.

II.3.1 Production des échantillons nécessaires à l'étude structurale des régions désordonnées de la prélamine A et de la progérine

II.3.1.1. Production des régions C-terminales de la prélamine A et de la progérine

Afin de démarrer l'étude des régions C-terminales désordonnées de la prélamine A et de la progérine, je me suis intéressé à la région 567-664 pour la prélamine A ainsi qu'à la région 567-614 pour la progérine. Les premières constructions comportaient une étiquette de 6 histidines en N-terminal ainsi qu'un site de clivage à la protéase TEV. Malheureusement, ces constructions ne donnaient pas de bons niveaux d'expression et ne permettaient donc pas l'analyse structurale des peptides. L'étiquette de 6 histidines a rapidement été remplacée par une étiquette GST, toujours suivie d'un site de clivage à la protéase TEV. Dans ces dernières constructions, pour permettre la quantification des peptides par UV, nous avons rajouté un résidu tryptophane entre le site de clivage à la protéase TEV et le gène codant la protéine. D'autres constructions codant pour des peptides dont les 4 cystéines ont été mutées en alanine ont également été utilisées pour faciliter l'optimisation de la purification et l'analyse des spectres RMN. Les protéines de fusion ont été produites chez *E. coli*, et purifiées en plusieurs étapes. Tout d'abord, après lyse des bactéries, les extraits ont été centrifugés et le surnageant a été placé sur colonne d'affinité pour les protéines avec étiquette GST. Après incubation et lavage, la protéase TEV a été ajoutée directement sur colonne. Le flow-through a été récupéré et concentré. Une étape de précipitation à la chaleur a été réalisée pendant laquelle les protéines repliées ont tendance à s'agréger. Enfin, une étape de filtration sur gel permet de récupérer le peptide final. La Figure 56A montre les gels SDS-PAGE obtenus après purification des différentes constructions utilisées dans ce travail de thèse, la Figure 56B montre les profils de filtration sur gel obtenus pour les peptides PreLamC et Prog. Enfin, le Tableau 2 récapitule les rendements obtenus.

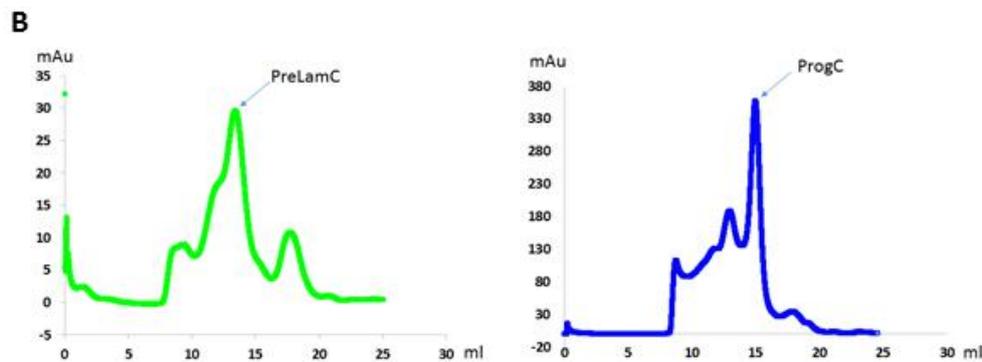
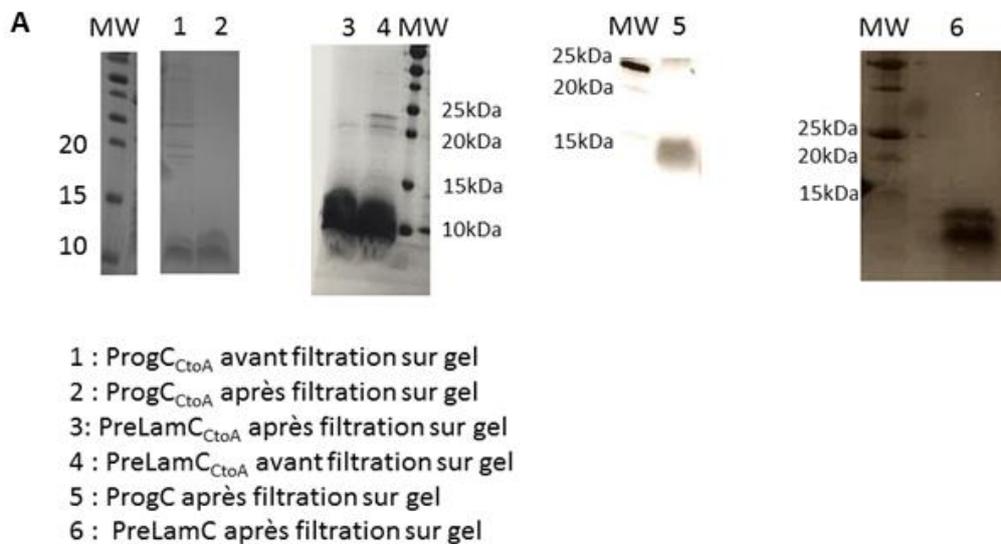


Figure 56 : (A) Gels SDS-PAGE des constructions ProgC_{CtoA}, PreLamC_{CtoA}, ProgC et PreLamC. Le marqueur de taille est indiqué par la mention « MW ». L'indication « FT » représente la fraction non fixée sur colonne d'affinité. (B) Profils de filtrations sur gel des peptides PreLamC et ProgC réalisées dans un tampon 20 mM phosphate pH 6.5, 150 mM NaCl, 2 mM DTT avec une colonne Superdex 70 10/300 GL.

Tableau 2 : Rendements de purification obtenus pour les différentes constructions des peptides C-terminaux de prélamine et progérine utilisées

Construction	ProgC _{CtoA}	PreLamC _{CtoA}	ProgC	PreLamC
Rendement (mg/L)	6	5	6	4

II.3.1.2. Production des partenaires

Afin de caractériser l'interaction des régions C-terminales désordonnées de la prélamine A et de la progérine avec leurs partenaires, j'ai produit plusieurs partenaires protéiques montrés dans la littérature comme interagissant avec la queue C-terminale de la lamine A. Je me suis intéressé à la protéine BAF, à la région N-terminale de SUN1 ainsi qu'à la protéine RBBP4.

II.3.1.2.1 BAF

La protéine BAF a été produite en cellules bactériennes à partir d'un gène possédant une étiquette histidine suivie d'un site de coupure à la protéase TEV. La protéine étant produite en corps d'inclusion, après lyse des cellules et centrifugation, le culot est repris dans un tampon contenant 8 M d'urée. Après une seconde étape de centrifugation, le surnageant est purifié à l'aide d'une colonne de nickel et l'élution se fait à l'aide d'un tampon contenant 8M d'urée et 1 M d'imidazole. Enfin, après plusieurs étapes de dialyse dans un tampon contenant au moins 150 mM NaCl afin d'empêcher la formation d'oligomères et de n'obtenir que le dimère de BAF, la protéine peut être concentrée, et est passée sur colonne de filtration pour s'affranchir des autres formes oligomériques présentes (Figure 57). Un spectre RMN $^1\text{H}^{15}\text{N}$ HSQC de la protéine BAF est visible en Figure 58.

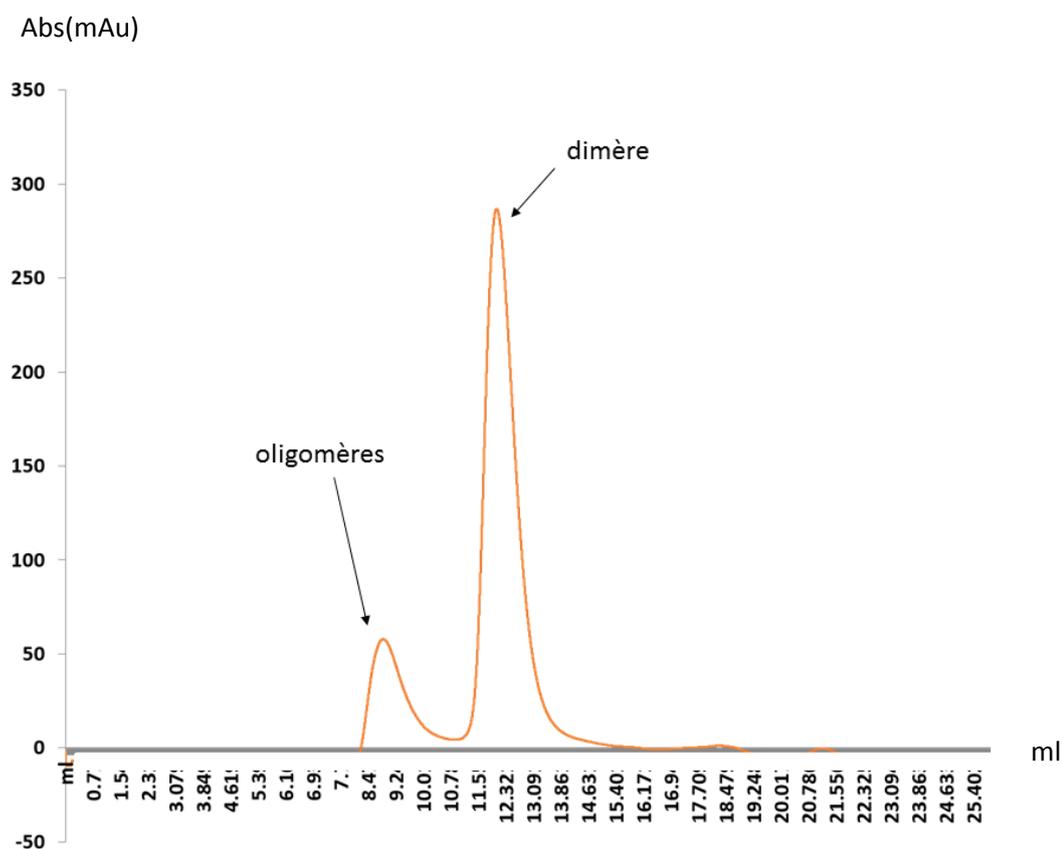


Figure 57 : Profil de filtration sur gel de BAF réalisée dans un tampon contenant 50mM Tris-HCl pH 8 et 150 mM NaCl à l'aide d'une colonne Superdex 76 10/300 GL et illustrant l'état oligomérique de la protéine.

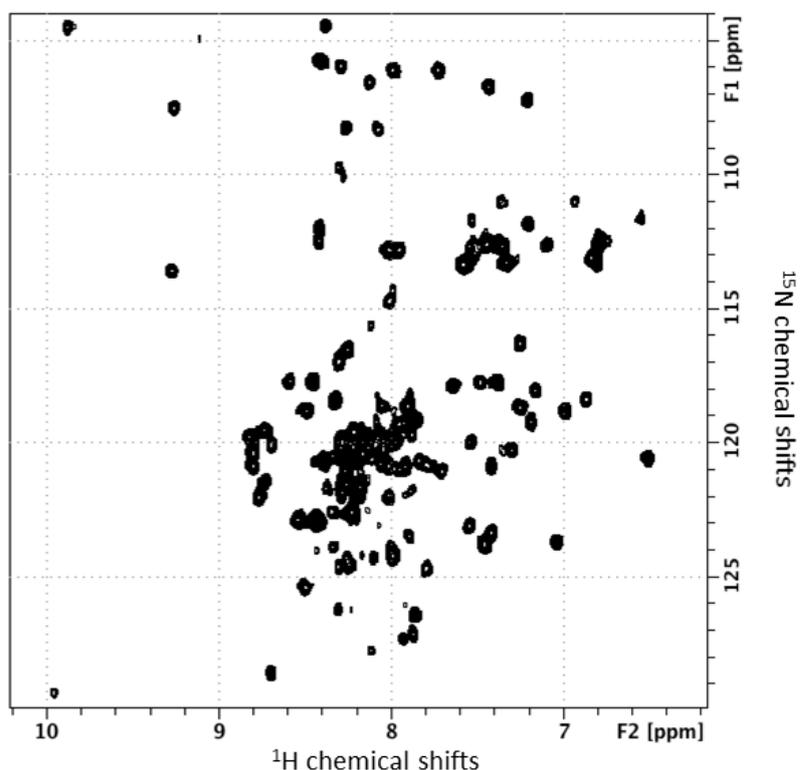


Figure 58 : Spectre $^1\text{H}^{15}\text{N}$ HSQC de BAF enregistré à 100 μM , 700 MHz, 293 K dans un tampon contenant 20 mM phosphate pH 6.5, 150 mM NaCl et 2 mM DTT. La dispersion des pics confirme que BAF est bien structuré. L'attribution des déplacements chimiques de BAF n'étant pas publiée, elle a été entreprise dans l'équipe, mais n'a pas été terminée du fait des temps de relaxation T2 défavorables mesurés sur BAF.

II.3.1.2.2 SUN1

Concernant la protéine SUN1, je me suis intéressé à la région 1-139 de l'isoforme 1 murine, montrée comme interagissant avec la lamine A humaine¹¹² (Figures 59 A et B). Cette région est identique à 68 % avec la région 1-138 de l'isoforme a humaine (Figures 59 C et D). Des prédictions de structure secondaire et de propension au désordre ont été réalisées grâce aux serveurs Psipred et Disopred3 respectivement⁷ (<http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1> ; Figure 59). Il semble que plusieurs régions de ce peptide aient une tendance à former une structure secondaire : notamment la région 22-25 semble adopter une structure en brin β tandis que les régions 35-39 et 89-94 sont prédites comme adoptant une structure en hélice α (Figure 59A). Ces régions prédites comme adoptant des structures secondaires correspondent à des régions où la propension de désordre est moins forte sur le graphique obtenu à l'aide du serveur Disopred3 (Figure 59B). Il est intéressant de noter que les structures prédites pour ces régions semblent conservées entre les protéines SUN1 murine et humaine, mais pas avec SUN2 (Figures 59 E et F).

Afin de vérifier ces prédictions, j'ai voulu produire la région 1-139 de SUN1. Le vecteur utilisé code pour une étiquette histidine en N-terminal, suivie d'un site de coupure à la protéase TEV et de la région N-terminale de SUN1.

La protéine n'étant pas soluble en bactérie, j'ai également réalisé cette purification dans un tampon contenant 8 M d'urée sur une colonne de nickel (Figure 60A). Par la suite, afin de retirer l'urée, j'ai réalisé différents tests de dialyse à différentes concentrations de protéine (100 μ M, 50 μ M, 20 μ M) mais également différentes concentrations de NaCl (0 mM, 150 mM, 300 mM, 500 mM). Malheureusement, aucun de ces tests n'a permis d'obtenir la protéine à plus de 20 μ M de concentration finale. Ceci fut tout juste suffisant pour confirmer le caractère désordonné de la protéine par RMN comme le montre le spectre $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ présenté sur la Figure 60B, spectre rouge, sur laquelle les déplacements chimiques sont situés entre 7.5 et 8.5 ppm. Cependant, une telle concentration n'est pas suffisante pour réaliser une étude d'interaction avec les peptides PreLamC et ProgC.

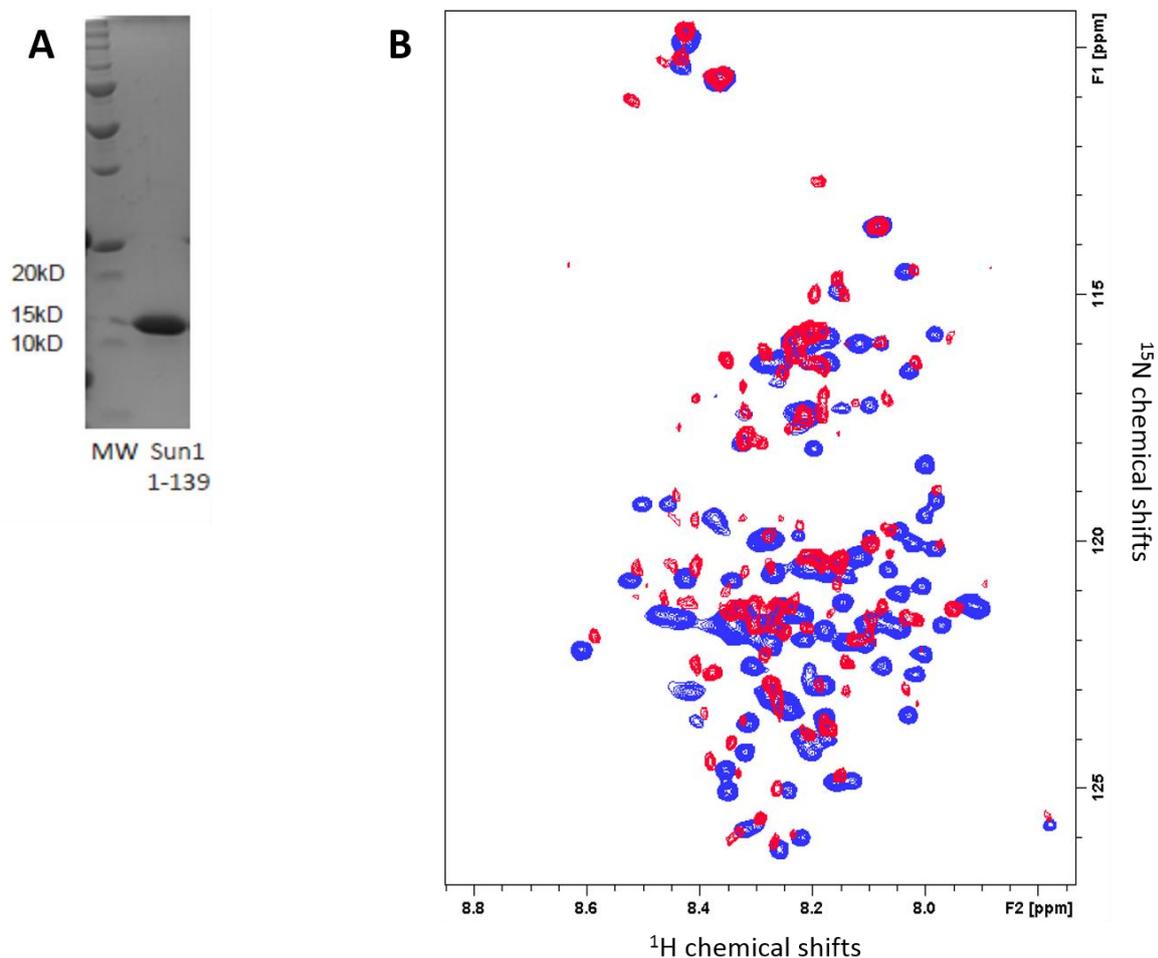


Figure 60 : Purification et analyse par RMN de la protéine SUN1. (A) Gel SDS-PAGE réalisé après purification sur colonne d'affinité de nickel dans un tampon contenant 8M d'urée. (B) Superposition d'un spectre $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ enregistré sur un échantillon à 20 μ M de la protéine SUN1 après renaturation de la protéine (rouge) et d'un spectre $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ enregistré sur la protéine SUN1 dans 2M d'urée (bleu). Les spectres ont été enregistrés dans un tampon contenant 20 mM phosphate pH 6.5 ; 150 mM NaCl ; 2mM DTT, 10% D₂O et à 283 K et 700 MHz.

Il a été possible d'obtenir un spectre RMN $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ de SUN1 à 50 μM , mais uniquement en 2M urée (Figure 60B, spectre bleu). Ce spectre est similaire à celui obtenu sans urée. Malheureusement, une étude d'interaction dans un tampon contenant 2M urée est impossible.

II.3.1.2.3 RBBP4

La protéine RBBP4 est une chaperone d'histone qui est connue d'après la littérature, pour interagir avec la prélamine A ainsi qu'avec la protéine BAF. Cependant, ces interactions ont été montrées soit par double-hybride en levure, soit par pull-down ou encore par co-immunoprécipitation, mais aucun détail moléculaire concernant ces interactions n'est connu. Grâce à l'aide du Dr. Pascal Drevet, nous avons produit la protéine RBBP4 en cellules d'insectes.

Après lyse des cellules, et centrifugation, la protéine RBBP4 contenant une étiquette histidinea été purifiée sur colonne d'affinité de nickel. La protéine éluée a ensuite été coupée à la thrombine pour retirer l'étiquette histidine, puis placée sur une colonne échangeuse d'anions et enfin séparée par filtration sur gel. Les gels de purifications sont visibles en Figure 61, un profil de filtration sur gel est montré en Figure 62.

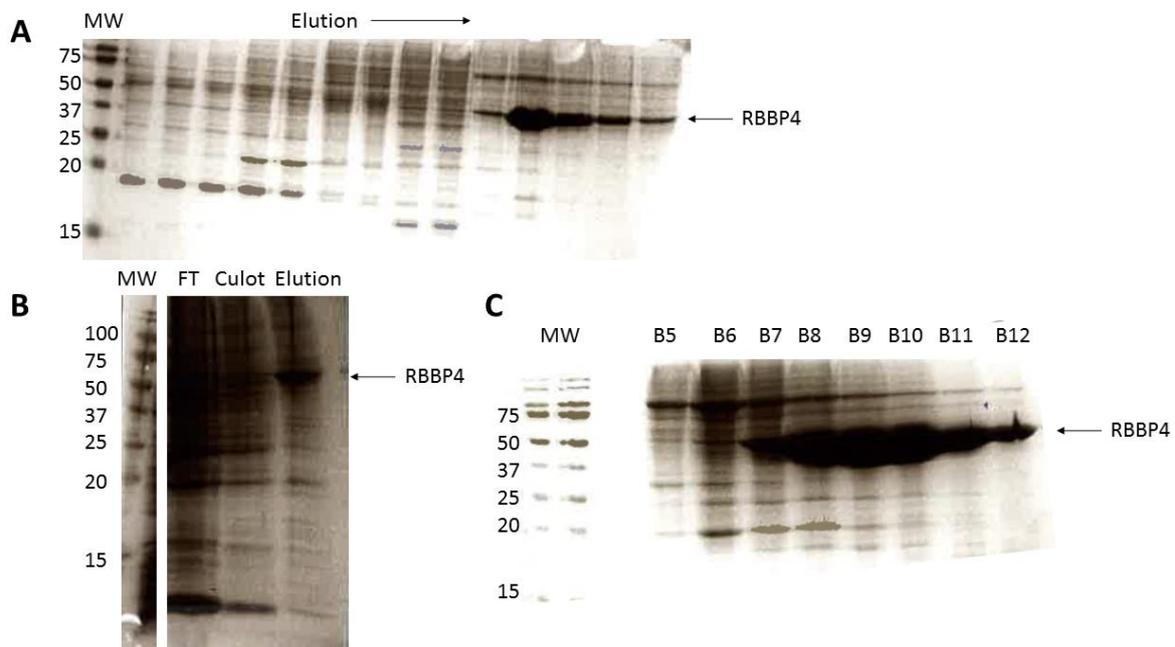


Figure 61 : Gels SDS-PAGE de la purification de RBBP4. (A) Purification sur colonne d'affinité de nickel. (B) Purification sur colonne échangeuse d'anions. L'éluée a été réalisée par un gradient de NaCl croissant orienté dans le sens de la flèche. (C) Purification par filtration sur gel réalisée dans un tampon contenant 50 mM Tris-HCl pH8, 150 mM NaCl, 2 mM DTT à l'aide d'une colonne Superdex75 10/300 GL. Sur tous ces gels, les marqueurs de poids moléculaire (en kDa) se trouvent dans la colonne « MW ».

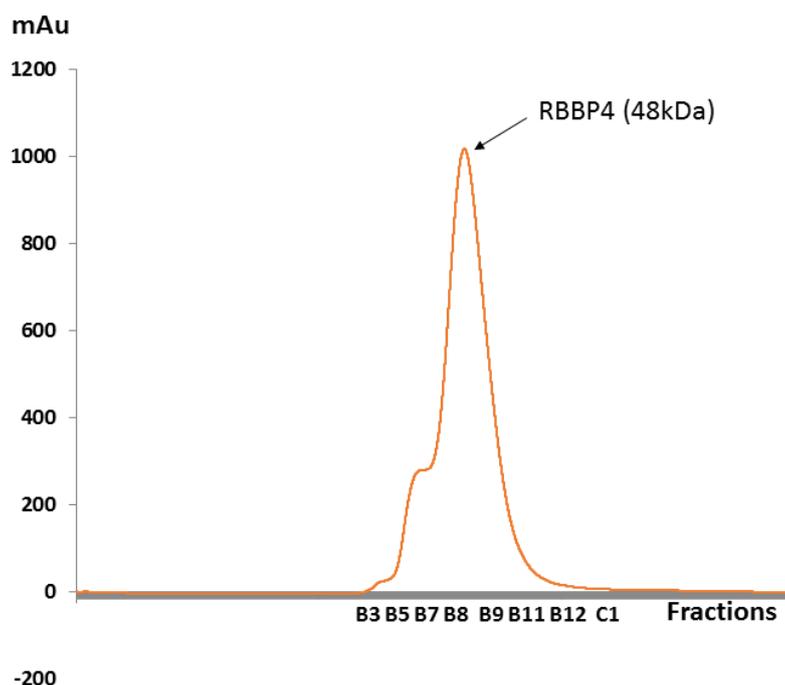


Figure 62 : Profil de filtration sur gel de la protéine RBBP4 réalisée dans un tampon contenant 50 mM Tris-HCl pH8 ; 150 mM NaCl ; 2 mM DTT avec une colonne Superdex 75 10/300 GL.

II.3.2 Etude par RMN des régions C-terminales désordonnées de la prélamine A et de la progérine

II.3.2.1. Les deux peptides possèdent une région conservée adoptant transitoirement une conformation en hélice α

Après production des peptides enrichis aux isotopes stables (^{15}N et ^{13}C) et purification, j'ai enregistré des premiers spectres RMN $^1\text{H}^{15}\text{N}$ HSQC. Dans ces spectres, chaque pic correspond à un acide aminé du peptide. Les signaux RMN observés ont tous des déplacements chimiques ^1H compris entre 7.5 et 8.5 ppm, indiquant le caractère intrinsèquement désordonné des peptides (Figure 63). La superposition des spectres RMN $^1\text{H}^{15}\text{N}$ HSQC des peptides PreLamC et ProgC montre une bonne superposition que ce soit pour les mutants ne comportant pas de cystéines ou pour les spectres des protéines sauvages, confirmant que la délétion des 50 acides aminés n'entraîne pas de différence significative de conformation entre la PreLamC et ProgC (Figure 63).

Une série d'expériences RMN 3D a été enregistrée (HNCACB, HNCA, CBCACONH, HNCO, HNCA(N)NH) et l'attribution des signaux RMN ^1H , ^{15}N et ^{13}C des peptides a pu être réalisée. Tout d'abord, j'ai attribué les spectres RMN des peptides dont les cystéines avaient été mutées en alanine²⁵⁴ (Figure 63A). L'attribution a révélé que les signaux RMN

correspondant à la partie commune des deux peptides (568-606 et 657-664), se superposent parfaitement. Seuls les pics correspondant aux acides aminés proches de la délétion ont une position différente dans les deux spectres (Ala605, Gln606 et Ser657). Le peptide ProgC_{CtoA} a été totalement attribué (100% $^1\text{H}^{15}\text{N}$, 100% $^{13}\text{C}\alpha$, 100% $^{13}\text{C}\beta$, 100% ^{13}CO). Par contre, concernant le peptide PreLamC_{CtoA}, une attribution complète ne fut pas possible. Ainsi, 91% des paires $^1\text{H}^{15}\text{N}$ (84/92), 93% des $^{13}\text{C}\alpha$ (90/97), 93% des $^{13}\text{C}\beta$ (76/82) et 91% des ^{13}CO (88/97) ont été attribués. Le seul fragment de plus de 3 acides aminés non attribué correspond à la région 645-648, correspondant au site de clivage de la dernière étape de maturation de la protéine. L'attribution du peptide ProgC_{CtoA} a ensuite pu être facilement transférée au peptide ProgC (Figure 63B). Des expériences d'attribution similaires à celles précédemment enregistrées ont néanmoins été réalisées afin de confirmer cette attribution. Ainsi, toutes les résonances $^1\text{H}^{15}\text{N}$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ e ^{13}CO ont pu être attribuées. Enfin, les attributions combinées des peptides PreLamC_{CtoA} et ProgC, ont permis de transférer l'attribution au peptide PreLamC (Figure 63C), permettant l'attribution de 82% des paires $^1\text{H}^{15}\text{N}$ (75/92).

L'analyse des déplacements chimiques secondaires des $^{13}\text{C}\alpha$ et $^{13}\text{C}\beta$ a permis de mettre en évidence la présence d'une hélice α partiellement peuplée (ou transitoire) entre les résidus 576-585²⁵⁴ (Figure 64A). La présence de cette hélice α est observée à la fois dans la progérine et la prélamine A, avec et sans cystéines (Figure 64A et B). Par ailleurs, la mesure des nOes hétéronucléaires $^1\text{H}^{15}\text{N}$ enregistrés sur l'échantillon ProgC a permis de confirmer la présence d'une structure secondaire résiduelle correspondant à cette hélice transitoire au sein de la progérine (Figure 64A). Il est intéressant de noter que cette région en hélice α est très conservée en séquence au cours de l'évolution (Figure 55). Il est très probable que cette région constitue un site de liaison pour un partenaire de la lamine A.

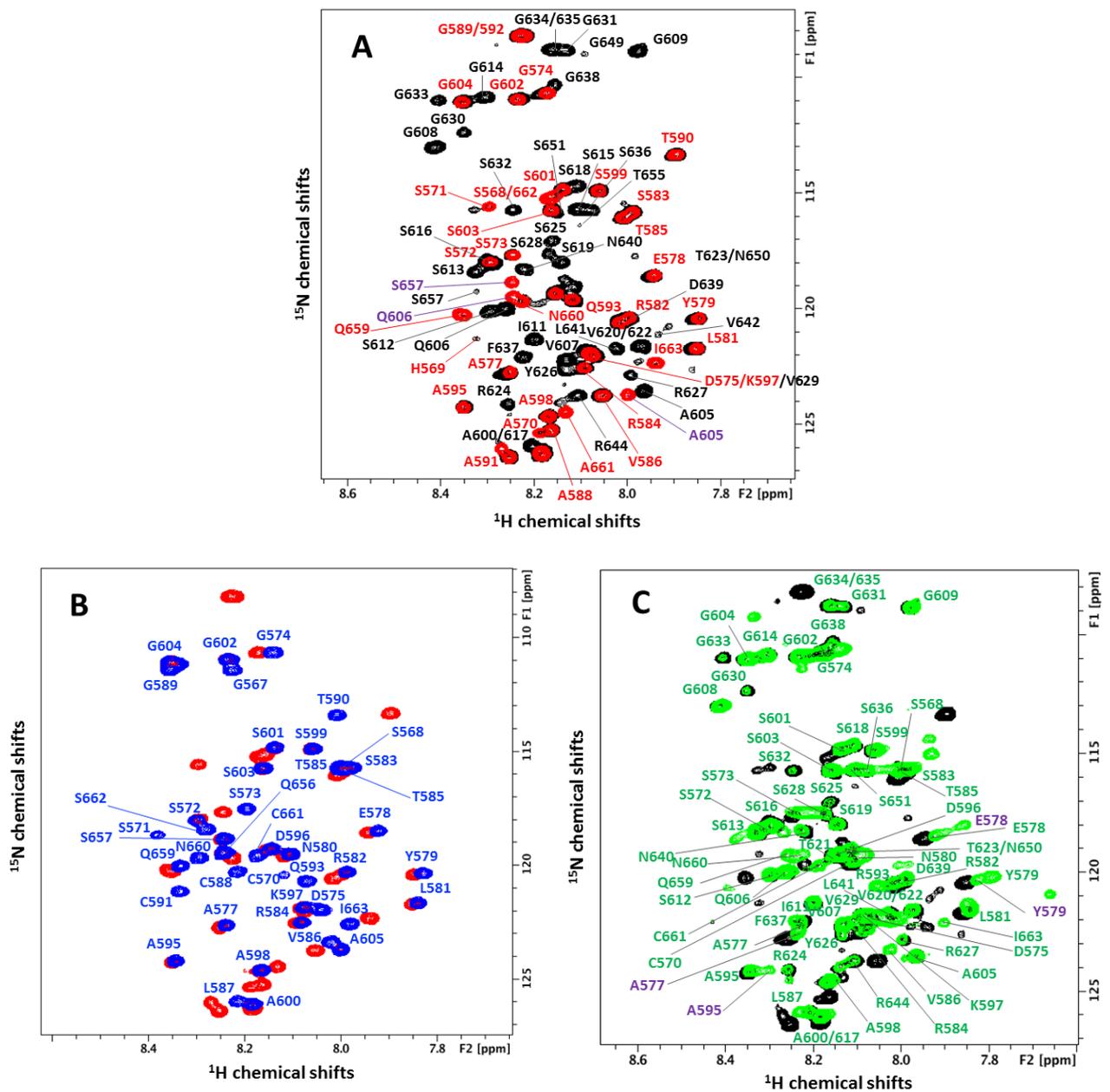


Figure 63 : Spectres RMN $^1\text{H}^{15}\text{N}$ HSQC des peptides C-terminaux désordonnés de la prélamine A et de la progérine avec et sans cystéines (Celli et al., 2018). Chaque pic est associé à un résidu du peptide. Les noms de résidus sont notés dans la couleur correspondant au peptide : rouge pour ProgC_{C10A}, bleu pour ProgC, noir pour PreLamC_{C10A} et vert pour PreLamC. (A) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de ProgC_{C10A} (rouge) sur PreLamC_{C10A} (noir). Les pics de PreLamC_{C10A} ne se superposant pas à ceux de ProgC_{C10A}, sont indiqués en violet. (B) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de ProgC (bleu) sur ProgC_{C10A} (rouge). (C) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de PreLamC (vert) sur PreLamC_{C10A} (noir). Les résidus montrant des conformations minoritaires dues à l'isomérisation des prolines sont indiqués en violet. Les spectres ont été enregistrés à 600 MHz, 283 K et à 600 μM dans un tampon 20 mM phosphate pH 6.5 ; 150 mM NaCl ; 2 mM DTT, 10% D₂O.

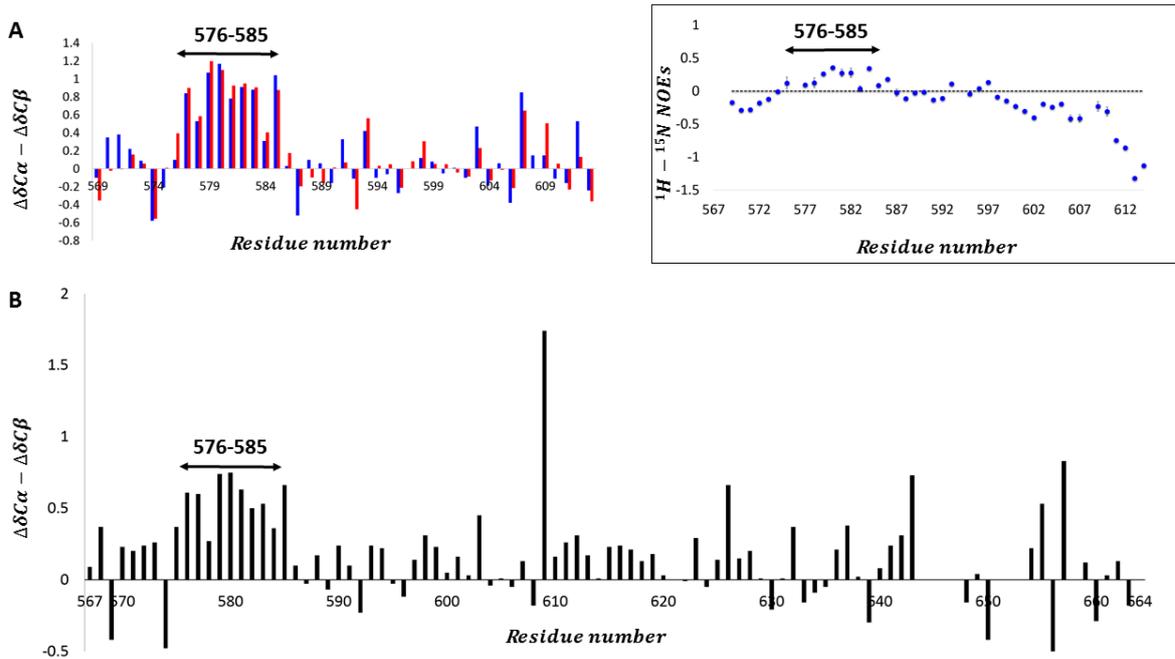


Figure 64 : Les régions C-terminales désordonnées de la prélamine A et de la progérine possèdent une région ayant une tendance à former une hélice α transitoire. (A) La différence entre les déplacements chimiques secondaires des $C\alpha$ et des $C\beta$ est représentée en fonction de la séquence du peptide ProgC WT (bleu) et du mutant ProgC_{CtoA} (rouge). Les nOes hétéronucléaires 1H - ^{15}N sont représentés dans l'encadré de droite en fonction de la séquence du peptide ProgC. (B) La différence entre les déplacements chimiques secondaires des $C\alpha$ et des $C\beta$ est représentée en fonction de la séquence du peptide PreLamC WT (noir). La région caractérisée par plus de 3 résidus avec une valeur supérieure à 0.5 est annotée ; elle correspond à une région formant une hélice transitoire en solution. Les expériences RMN ont été enregistrées à 600MHz, 283 K et à 450 μ M dans un tampon 20 mM Phosphate pH6.5 ; 150 mM NaCl ; 2 mM DTT, 10% D₂O.

11.3.2.2. Le peptide C-terminal de la lamine A reste difficile à obtenir sous une forme structurellement homogène

Plusieurs échantillons de peptides ProgC et PreLamC ont alors été produits. Si les spectres RMN enregistrés sur les échantillons de ProgC ne varient pas, une certaine hétérogénéité a été observée entre les différents spectres de PreLamC. Cette hétérogénéité a pu être identifiée aussi sur gel SDS-PAGE (Figure 65).

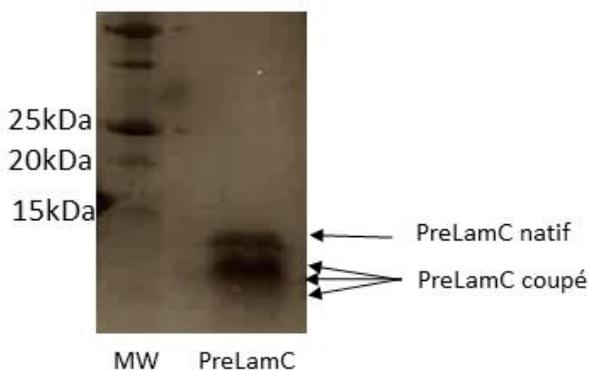


Figure 65 : Gels SDS-PAGE montrant l'hétérogénéité d'un échantillon PreLamC. La colonne MW désigne les marqueurs de taille. Les bandes correspondant au peptide natif ou clivé sont indiquées par des flèches.

La Figure 66 montre une superposition des spectres HSQC de deux échantillons du domaine C-terminal de la prélamine, PreLamC et PreLamC-2. Le spectre de l'échantillon PreLamC a été enregistré à 600 μM tandis que celui de PreLamC-2 a été enregistré à 50 μM . La comparaison des deux spectres montre clairement une disparition des signaux de la région 624-644 sur le spectre de PreLamC-2 qui est faiblement concentré. Cette région semble être en échange intermédiaire. De plus, l'encadré sur la figure montre que l'extrémité C-terminale native de la prélamine A n'apparaît que dans le cas de PreLamC-2. Le caractère natif du C-terminal a été identifié par superposition du spectre de PreLamC-2 avec celui de ProgC (Figure 66). L'encadré montrant des niveaux de contours plus élevés révèle également que sur le spectre de PreLamC, plusieurs C-termini sont présents, indiquant que des molécules de tailles différentes sont produites, comme suspecté à partir du gel SDS-PAGE (Figure 65).

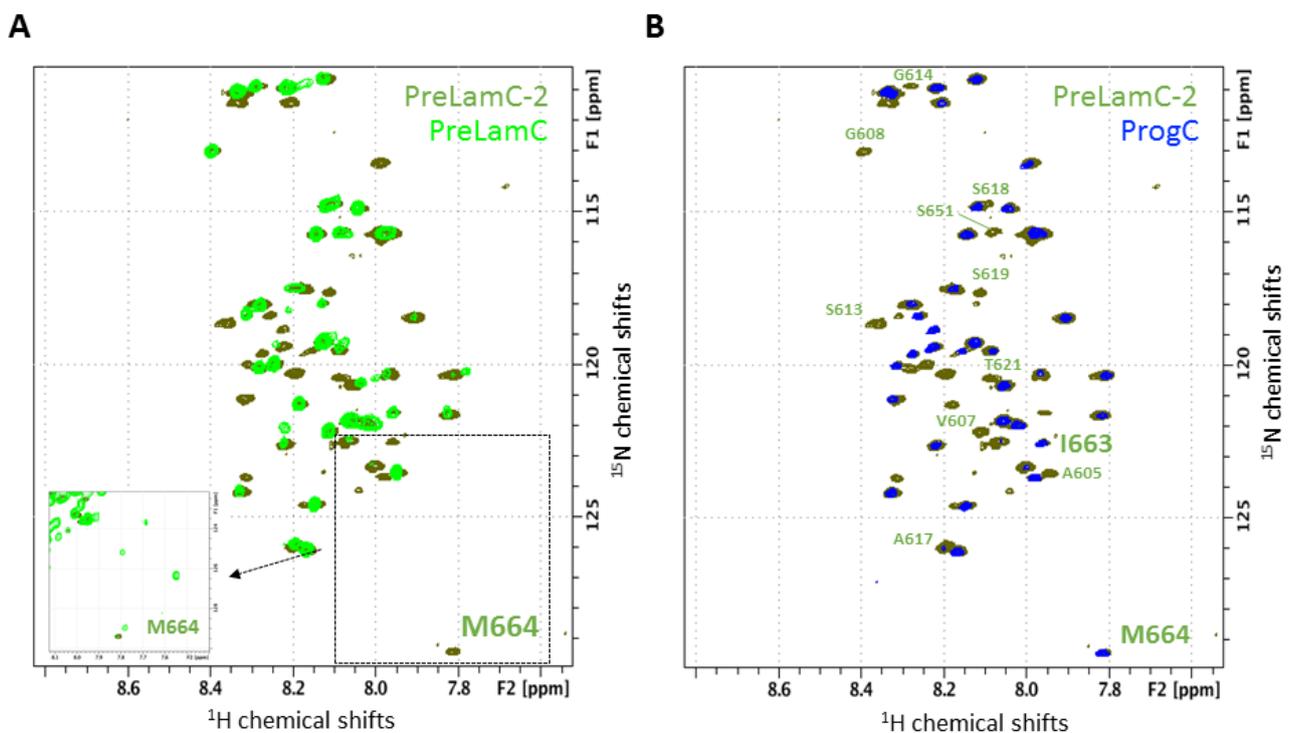


Figure 66 : Spectres RMN ^1H - ^{15}N HSQC illustrant les difficultés à obtenir un échantillon homogène de PreLamC. (A) Superposition des spectres correspondants à deux échantillons PreLamC et PreLamC-2. Le spectre en vert clair correspond à l'échantillon PreLamC de Celli et al. (2018), le spectre en vert foncé correspond à un échantillon purifié précédemment dit PreLamC-2. La région encadrée est de nouveau représentée en bas à gauche du spectre, mais avec un niveau de contour plus bas pour PreLamC. Cet encadré montre que le signal du résidu C-terminal M664 n'apparaissait pas sur la superposition principale dans le cas de PreLamC, car cet échantillon comprend des molécules avec plusieurs C-termini différents. (B) Superposition des spectres correspondants à PreLamC-2 (vert) et ProgC (bleu). Cette superposition montre que les résidus M664 (dans le cas de PreLamC-2) et M614 (dans le cas de ProgC) possèdent le même signal HSQC et donc le même environnement structural. En revanche, sur le spectre de PreLamC-2 les résidus de la région 624-644 ne sont pas visibles contrairement au spectre de PreLamC (Celli et al., 2018). Cette région semble être en échange conformationnel et donc visible uniquement à forte concentration.

Afin d'obtenir la protéine PreLamC en entier, nous avons commandé une nouvelle construction codant pour le peptide PreLamC, suivi d'un site de coupure à la protéase TEV, d'une étiquette GB1 et d'une étiquette 6 histidines pour la purification (Figure 67 A). Cette

construction sera appelée PreLamC-GB1 dans le reste de cette thèse. Une superposition du spectre $^1\text{H}^{15}\text{N}$ HSQC de PreLamC-2 (vert-foncé) et du spectre de PreLamC-GB1 (saumon) est visible en Figure 67B. Les signaux correspondant au peptide PreLamC se superposent globalement bien à ceux de PreLamC-GB1, indiquant que la structure du peptide est maintenue dans la nouvelle construction. Ceci a été confirmé par le calcul des variations de déplacements chimiques combinées (Figure 68) entre le peptide PreLamC et PreLamC-GB1. Très peu de variations de déplacements chimiques sont observables entre les deux constructions à l'exception de l'extrémité C-terminale, ce qui est dû à l'ajout de l'étiquette GB1 qui modifie l'environnement chimique de cette région. Un zoom montre que les glycines de la région 624-648, précédemment identifiées comme étant en échange intermédiaire dans la construction PreLamC-2 (Figure 66), sont de nouveau visibles sur l'échantillon PreLamC-GB1, montrant un changement du régime d'échange (Figure 69B).

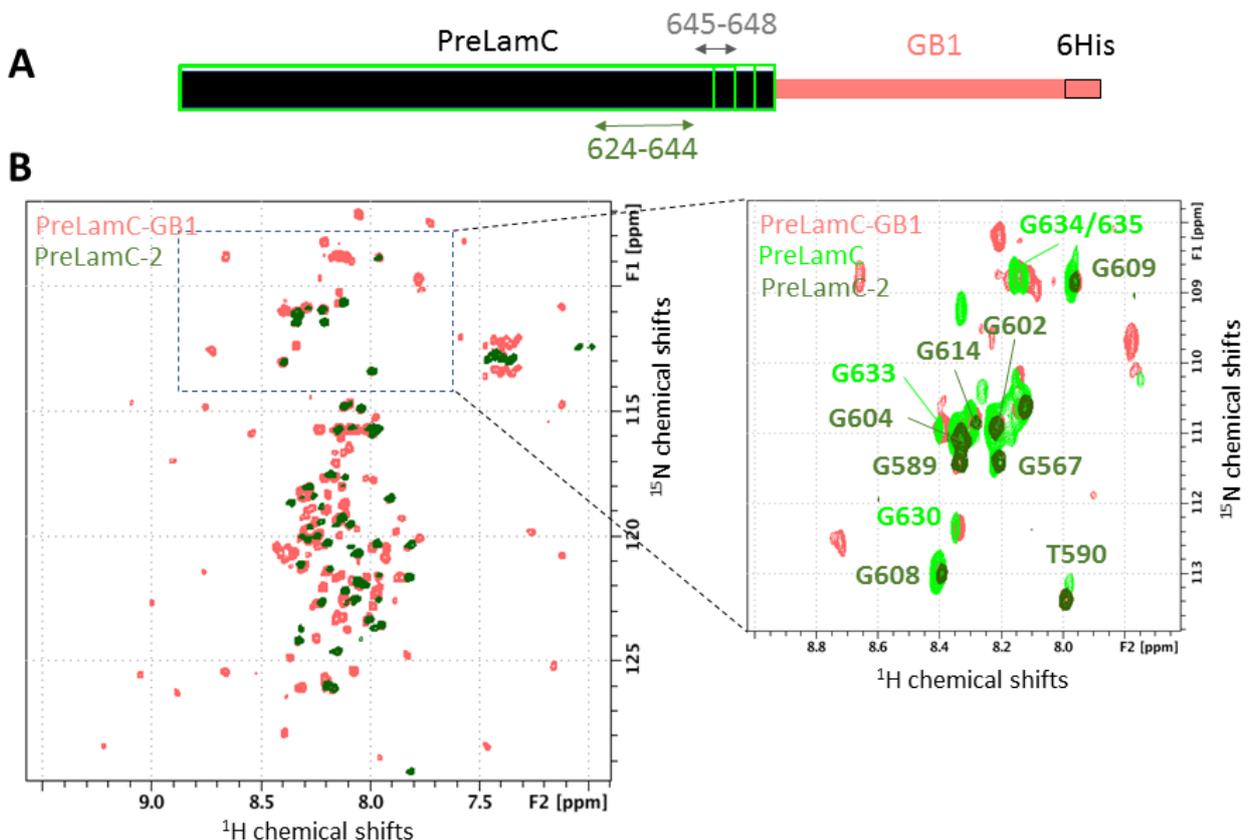


Figure 67 : (A) Schéma illustrant la construction PreLamC-GB1 (saumon). La zone 624-644, en échange intermédiaire dans la construction PreLamC, est représentée à l'aide d'une flèche vert-foncé. Les différentes molécules produites dans le spectre de PreLamC (Celli et al., 2018), sont indiquées par des encadrés vert-clair. (B) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de l'échantillon PreLamC-2 (vert-foncé) et de PreLamC-GB1 (saumon). Un zoom montre la présence des glycines situées dans la région 624-644, en échange intermédiaire dans l'échantillon PreLamC-2, visibles uniquement à haute concentration comme dans l'échantillon PreLamC, qui sont également visibles à faible concentration sur l'échantillon PreLamC-GB1, suggérant que cette région ne subit pas d'échanges intermédiaires dans cette construction.

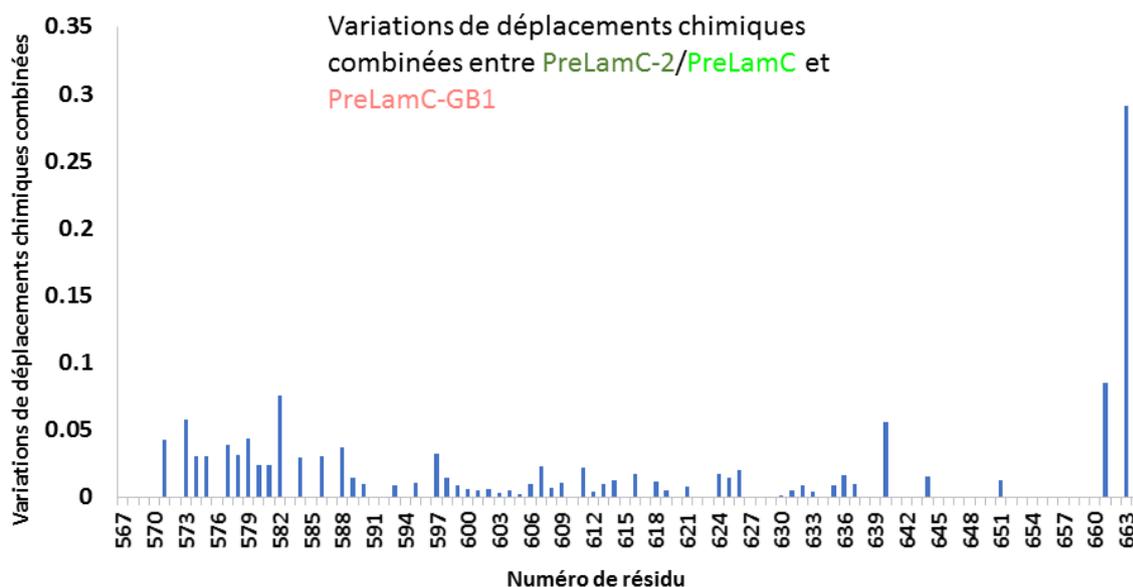
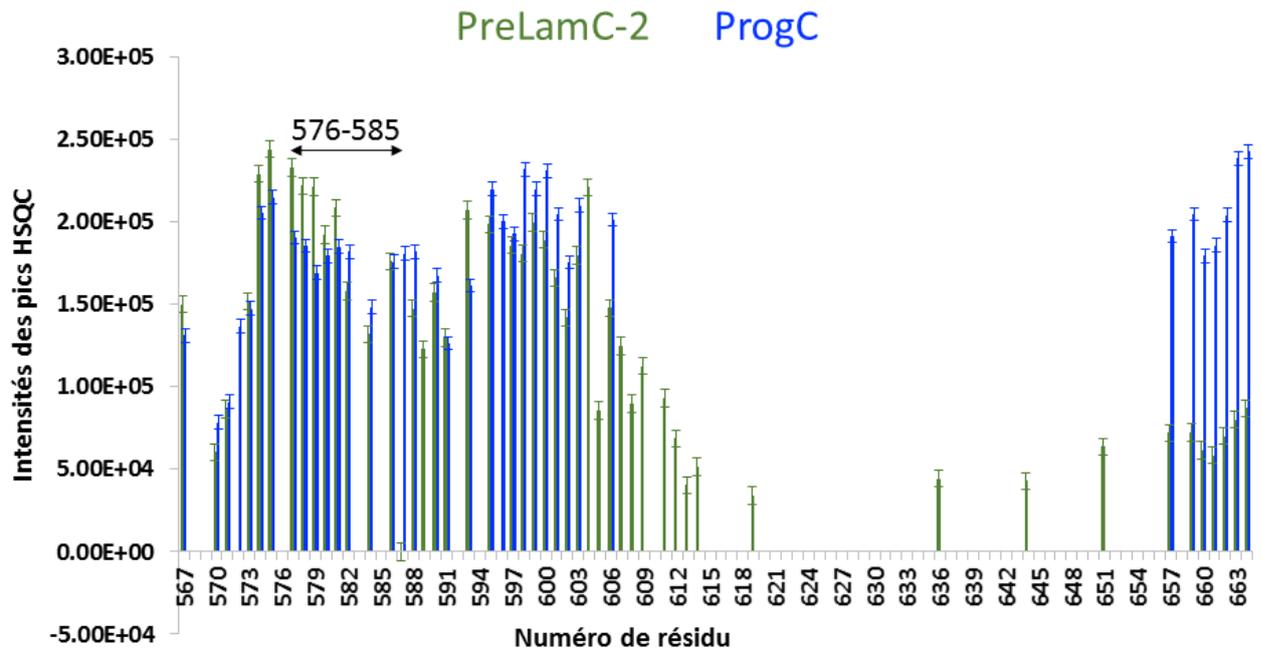
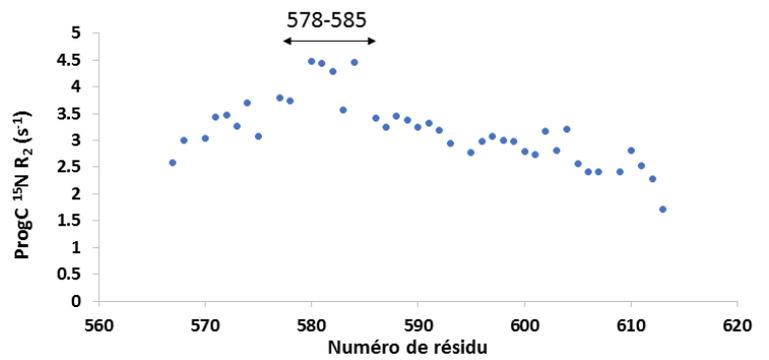


Figure 68 : Variations de déplacements chimiques combinés entre le peptide PreLamC et PreLamC-GB1 en fonction de la séquence en acides aminés.

L'analyse des intensités des pics $^1\text{H}^{15}\text{N}$ HSQC des échantillons ProgC et PreLamC-2, assimilables à la relaxation transversale T_2 des peptides et reflétant leur mobilité interne, montre une bonne corrélation entre les régions communes des peptides (Figure 69B). De manière intéressante, l'intensité diminue chez les deux peptides dans la région 576-585, précédemment identifiée comme ayant une tendance à former une hélice α transitoire (Figures 64 et 69A). Cette tendance à adopter une structure transitoire est confirmée par des expériences de relaxation transversale (Figure 69A) enregistrées sur le peptide ProgC. Ces expériences montrent que la zone 576-585 possède des valeurs de R_2 élevées, c'est-à-dire des T_2 courts, suggérant des mouvements moléculaires plus lents localement qui peuvent être expliqués par une structure transitoire en hélice α .

Une seconde observation a été faite en observant l'intensité des pics HSQC des peptides C-terminaux de la prélamine A et de la progérine. En effet, l'intensité des pics diminue drastiquement dans la région délétée chez la progérine (Figure 69A). Ceci confirme que cette région de la prélamine est en échange intermédiaire. La comparaison de ces profils d'intensités avec celui de la PreLamC-GB1 montre une différence flagrante dans la région délétée chez la progérine. En effet, les intensités qui étaient précédemment très faibles dans l'échantillon PreLamC-2 sont très importantes dans la construction PreLamC-GB1, confirmant le changement du régime d'échange (Figure 69B). Une explication possible serait que cette région est impliquée dans des interactions intermoléculaires caractérisées par des constantes d'échange de l'ordre de la milliseconde dans le cas du peptide PreLamC. En revanche, la présence de GB1 augmente la solubilité de la protéine de fusion, empêche ces interactions, et dans ce cas, la région impliquée devient mobile sur une échelle de temps rapide (ps-ns).

A



B

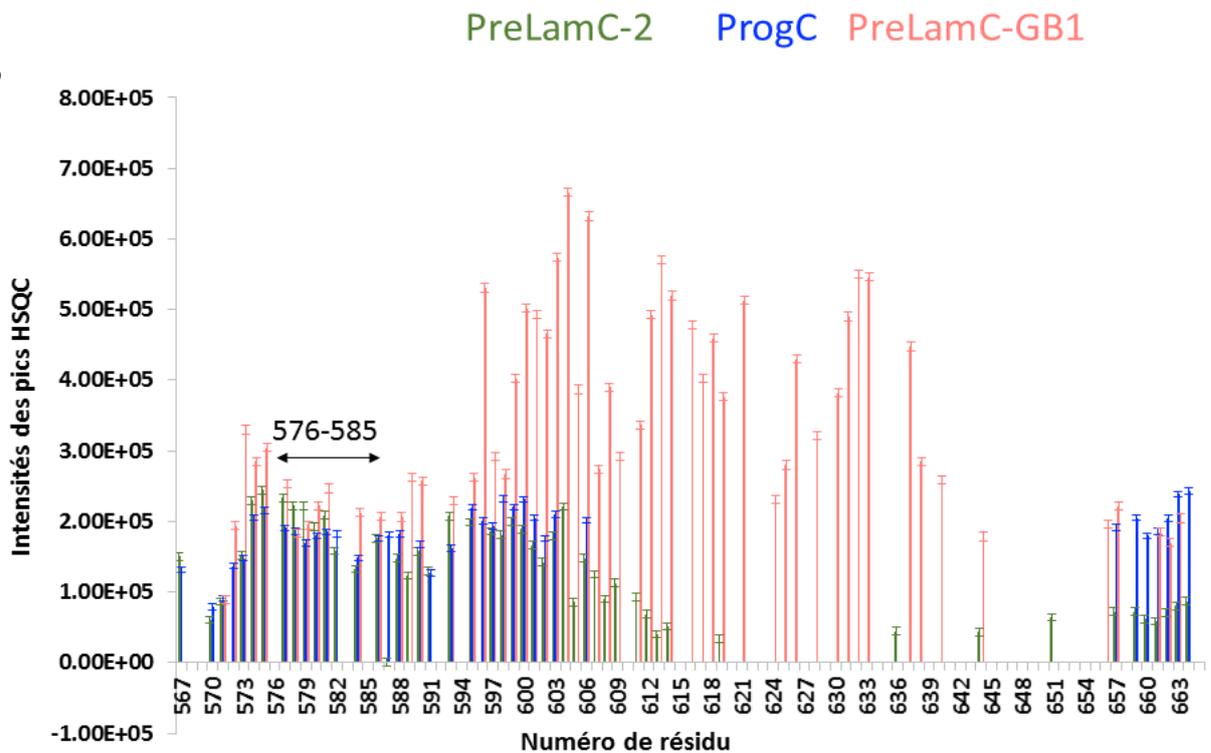


Figure 69 : Intensités mesurées sur les spectres $^1\text{H}^{15}\text{N}$ HSQC des peptides ProgC, PreLamC-2 et PreLamC-GB1. (A) Superposition des intensités en fonction de la séquence, mesurées sur les spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide ProgC (bleu) et PreLamC-2 (vert). (B) Superposition des intensités en fonction de la séquence, mesurées sur les spectres $^1\text{H}^{15}\text{N}$ HSQC des peptides ProgC (bleu), PreLamC-2 (vert) et PreLamC-GB1 (rouge). La région 576-585 ayant une tendance à former une hélice α transitoire est indiquée par une flèche.

II.3.3 Etude de la farnésylation de la prélamine A et de son mutant progérine

Lors de sa maturation, la région C-terminale désordonnée de la prélamine A subit une modification post-traductionnelle particulière, consistant en l'ajout d'un groupement farnésyl, un hydrocarbure naturel. Ce groupement est retiré lors de l'étape finale de la maturation mais reste présent chez le mutant progérine. Pour étudier l'impact de cette modification chimique sur la structure de cette région désordonnée, j'ai mis au point une méthode permettant de farnésyler *in vitro* les peptides précédemment purifiés.

Pour ce faire, j'ai dû mettre au point la purification de l'enzyme responsable de la farnésylation, c'est à dire la farnésyl-transférase (Figure 70A). Cette enzyme est constituée de deux sous unités, visibles sur le gel SDS-PAGE de purification. Une fois l'enzyme purifiée, j'ai incubé la protéine pure à farnésyler, précédemment placée dans un tampon de farnésylation, avec l'enzyme ainsi que du farnésyl-pyrophosphate (co-facteur de la réaction) et une grande quantité d'agent réducteur pour être certain que la cystéine cible de la modification soit réduite.

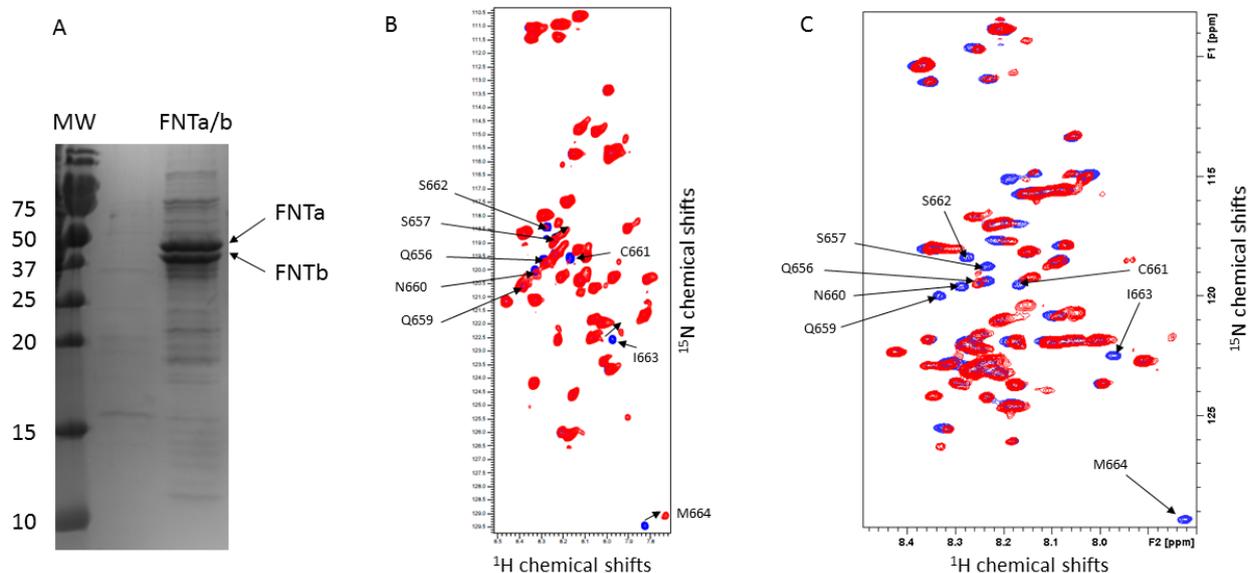


Figure 70 : Farnésylation de la région C-terminale de la progérine. (A) Gel SDS-PAGE de la farnésyl-transférase purifiée. (B) Superposition des spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC 253 du peptide progérine avant (bleu) et après farnésylation (rouge). Les pics du spectre bleu montrant une perte d'intensité sont annotés. Des flèches indiquent les déplacements des pics correspondants aux résidus I663 et M664 dus à la farnésylation. (C) Zoom montrant la superposition des spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC 253 de l'IgFoldProgérine avant (bleu) et après farnésylation (rouge). (A,B) Les spectres RMN ont été enregistrés à 700 MHz, 283K dans un tampon phosphate 20mM pH6.5, 150 mM NaCl, 2 mM DTT, 10% D $_2$ O.

Après farnésylation des peptides, des spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ ont été enregistrés pour les peptides ProgC, et l'IgFProg. La superposition des spectres du peptide ProgC avec et sans farnésylation montre clairement l'apparition de nouveaux signaux et la disparition des signaux proches de la cystéine C661 dans la séquence, confirmant que la modification a bien eu lieu (Figure 70B). Les mêmes disparitions de signaux sont visibles sur les spectres de la ProgC et de l'IgFProg farnésylés (Figure 70B et C). Par contre, les signaux correspondant à l'IgFProg farnésylé ne sont pas visibles, ceci est dû à la faible concentration de l'échantillon. Cependant, une grande partie des signaux se superpose très bien, montrant que la farnésylation de la région C-terminale de ProgC n'entraîne pas de modifications structurales profondes. Par ailleurs, le ratio des intensités entre les spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC du peptide ProgC avec et sans farnésylation, montre une diminution significative et locale de l'intensité des signaux RMN du peptide farnésylé, suggérant un échange intermédiaire et donc une interaction transitoire (Figure 71) et expliquant pourquoi les signaux farnésylés ne sont pas visibles sur le spectre de l'IgFProg peu concentré (Figure 70C).

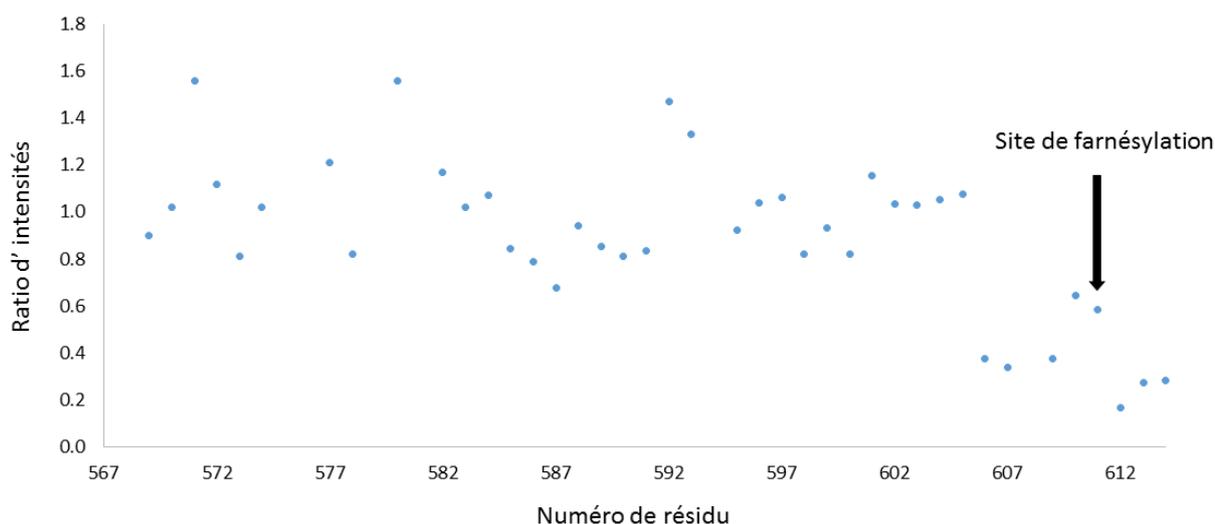


Figure 71 : La farnésylation de la région C-terminale dépliée du peptide progérine provoque l'apparition d'échanges conformationnels autour de la position farnésylée. Ratio des intensités mesurées sur les spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ de la progérine farnésylée versus non farnésylée en fonction de la séquence. Le site de farnésylation est indiqué par une flèche.

II.3.4 Interactions médiées par les régions désordonnées de la prélamine A et de la progérine

II.3.4.1. Rôle de la région C-terminale désordonnée de la lamine A dans la régulation de l'interaction BAF-Lamine A

Je me suis posé la question du rôle de la région C-terminale désordonnée dans la progéria. Une hypothèse était que cette région est capable d'interférer avec les interactions

normales entre lamines A/C, déstructurant ainsi le nucléosquelette. L'autre hypothèse était que cette région est capable d'inhiber l'interaction entre l'IgFold de la lamine A et le dimère de BAF. J'ai donc produit et purifié l'IgFold de la lamines A et la protéine BAF. La superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC des peptides PreLamC et ProgC en présence ou non de l'IgFold (Figures 72 A et B) ne permet pas d'observer de modification des signaux RMN après ajout de l'Igfold, ce qui démontre que les peptides n'interagissent pas avec l'Igfold dans ces conditions.

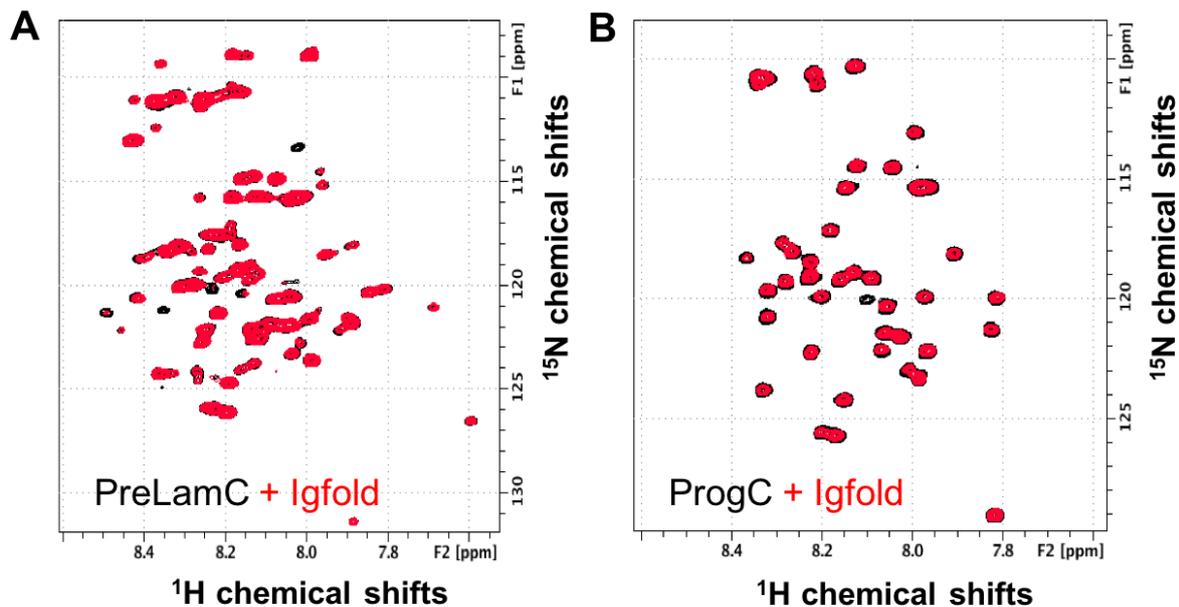


Figure 72 : Etude des interactions entre les régions C-terminales dépliées des prélamine et progérine et l'IgFold. Superpositions des spectres $^1\text{H}^{15}\text{N}$ HSQC (A) du peptide PreLamC seul (noir) ou avec BAF (rouge) ; (B) du peptide ProgC seul (noir) ou avec IgFold (rouge). Ces spectres ont été enregistrés avec 150 μM de peptide marqué à ^{15}N , à 600 MHz et 283 K dans un tampon contenant 20 mM Phosphate pH 6.5 ; 150 mM NaCl et 2 mM DTT, 10% D_2O .

Cependant, au sein de la prélamine et de la progérine, l'IgFold est relié covalamment au peptide C-terminal désordonné, ce qui augmente l'affinité apparente entre les 2 fragments. Pour évaluer la possibilité qu'il y ait une interaction transitoire entre l'IgFold et ProgC au sein de la progérine, j'ai superposé le spectre $^1\text{H}^{15}\text{N}$ HSQC de l'IgFold avec le spectre $^1\text{H}^{15}\text{N}$ HSQC du peptide ProgC. Le calcul des variations de déplacements chimiques combinées entre le peptide IgFProg et ProgC (Figure 73A) montre clairement une zone perturbée. Ainsi, les résidus 571, 572, 573, 577, 580, 582 et 584 semblent être affectés. Ces variations de déplacements chimiques ont été mesurées en comparant les déplacements chimiques de deux spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide ProgC à ceux de deux spectres de l'IgFProg. Les variations de déplacements chimiques entre les deux échantillons d'IgFProg d'une part et les deux échantillons de ProgC d'autre part utilisés pour cette étude sont représentées en Figure 73B, pour comparaison avec la figure 73A. Il est fort probable que les variations de déplacements chimiques observées entre IgFProg et ProgC pour la région 571-573 soient dues à la proximité du domaine globulaire dans la construction IgFProg par rapport au peptide ProgC. En revanche, les variations de déplacements chimiques observées pour la région 577-584 (aussi

illustrées sur la Figure 73C) ne peuvent être expliquées par la proximité en séquence du domaine globulaire. Je propose qu'elles soient dues à une interaction transitoire entre le domaine globulaire IgFold et la région désordonnée C-terminale.

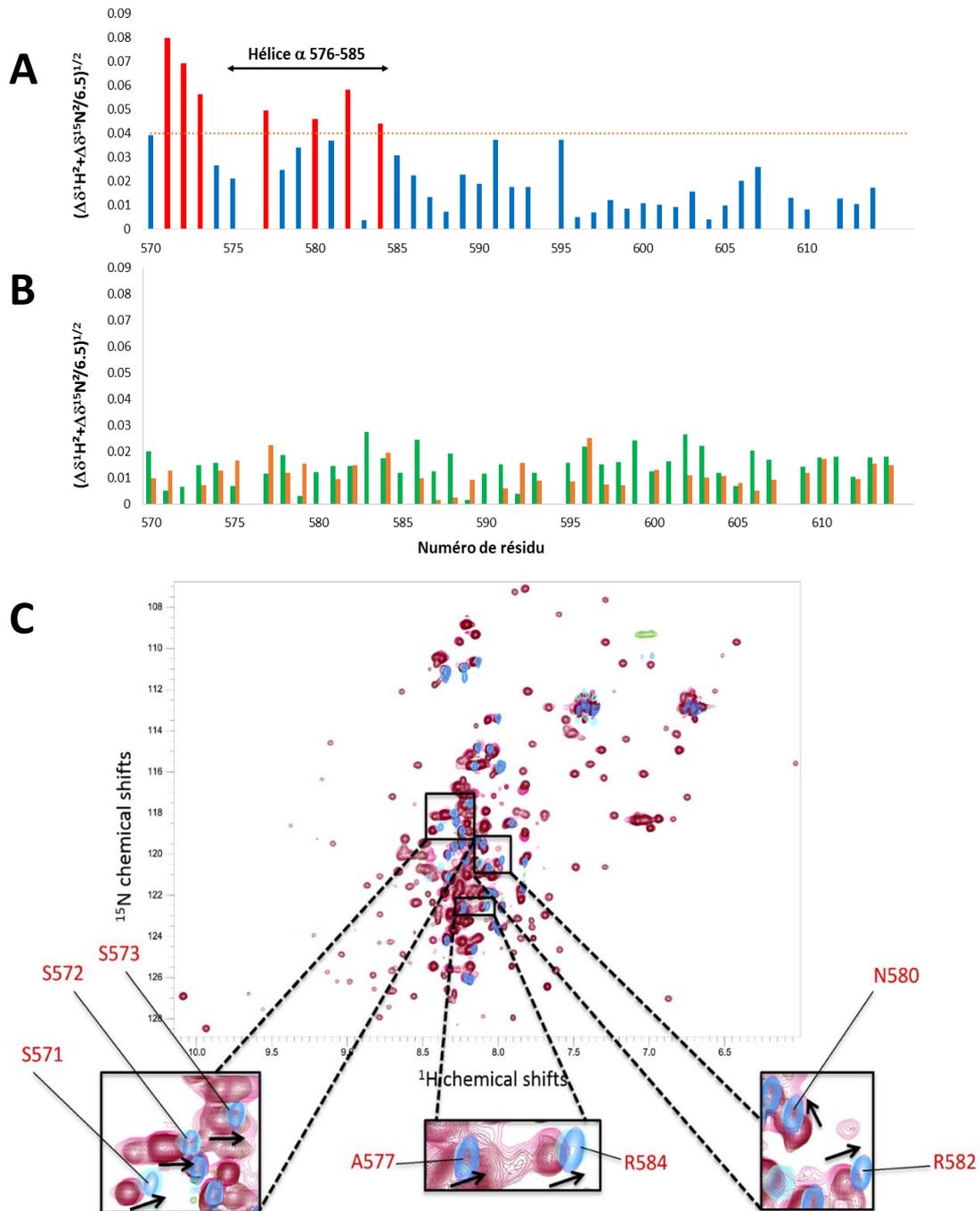


Figure 73 : Interaction transitoire entre la région intrinsèquement désordonnée de la progérine et le domaine globulaire IgFold. (A) Variations de déplacements chimiques combinés entre les spectres $^1\text{H}^{15}\text{N}$ HSQC de la région désordonnée de la progérine seule ou liée à l'IgFold. Les résidus montrant une forte variation sont représentés en rouge. (B) Variations de déplacements chimiques combinés observées entre les spectres $^1\text{H}^{15}\text{N}$ HSQC de deux échantillons d'IgFoldProgérine (orange) et entre les spectres $^1\text{H}^{15}\text{N}$ HSQC de deux échantillons de peptide progérine (vert). (C) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC de l'IgFoldProgérine (rose et brun), et du peptide progérine (bleu clair et bleu foncé). Les zones correspondant aux acides aminés subissant des variations importantes (en rouge dans (A)) sont encadrées et un zoom a été effectué pour visualiser les variations de déplacements chimiques ainsi que les résidus concernés. (A,B,C) Les spectres $^1\text{H}^{15}\text{N}$ HSQC ont été enregistrés à 283K et 600MHz, dans un tampon phosphate 20 mM pH 6.5 ; 150 mM NaCl ; 2 mM DTT ; 10% D_2O .

J'ai aussi analysé la superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC des peptides PreLamC et ProgC en présence ou non de BAF (Figure 74). Lorsque j'ai ajouté BAF à mon échantillon PreLamC-2, j'ai observé des faibles différences d'intensités mais l'échantillon était trop peu concentré pour pouvoir conclure (Figure 74A). Lorsque j'ai voulu répéter l'expérience avec une plus grande concentration de PreLamC, je n'ai obtenu que des échantillons hétérogènes en C-terminus, et je n'ai pas observé d'interaction avec BAF (Figure 74B). J'ai alors décidé de tester l'interaction de PreLamC^{CtoA} avec BAF mais je n'ai toujours pas vu d'interaction (Figure 74C). Enfin, j'ai testé l'interaction de ProgC avec BAF et je n'ai pas vu d'interaction (Figure 74D). Ainsi, si PreLamC interagit avec BAF, alors cette interaction implique le C-terminus du peptide et la cystéine du motif CAAX est essentielle. Pour conclure, il faudrait marquer BAF à l' ^{15}N et enregistrer un spectre $^1\text{H}^{15}\text{N}$ HSQC en absence et en présence d'un peptide synthétique correspondant au C-terminus de PreLamC.

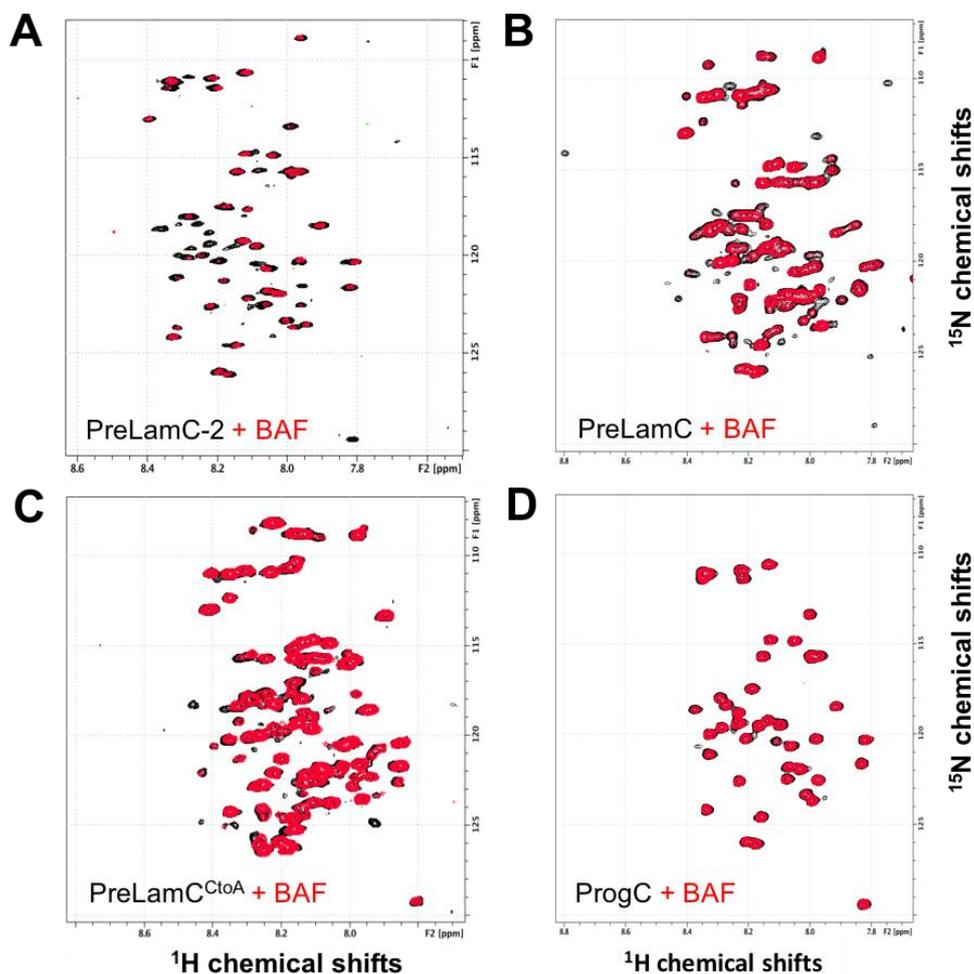


Figure 74 : Etude des interactions entre les régions C-terminales dépliées des prélamine et progérine et BAF. Superpositions des spectres $^1\text{H}^{15}\text{N}$ HSQC (A) du peptide PreLamC-2 seul (noir) et avec BAF (rouge) ; (B) du peptide PreLamC seul (noir) et avec BAF (rouge) ; (C) du peptide PreLamC^{CtoA} seul (noir) et avec BAF (rouge) ; (D) du peptide ProgC seul (noir) et avec BAF (rouge). Ces spectres ont été enregistrés à 600 MHz et 283 K dans un tampon contenant 20 mM Phosphate pH 6.5 ; 150 mM NaCl et 2 mM DTT, 10% D_2O .

Enfin, puisque l'accumulation de progérine restant constitutivement farnésylée entraîne l'apparition de la progéria, le groupement farnésyl pourrait permettre de médier une interaction avec le domaine IgFold de la lamine A ou le dimère de BAF, empêchant ainsi l'interaction de s'établir. J'ai donc enregistré par RMN des spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ de la région C-terminale désordonnée de la progérine farnésylée (ProgFarn) et marquée ^{15}N , en présence ou non de l'IgFold de la lamine A d'une part, et du dimère de BAF, d'autre part. La superposition des spectres du peptide ProgFarn seul et avec l'IgFold (Figure 75B) ou avec le dimère de BAF (Figure 75A) ne montre aucune modification du signal RMN du peptide après ajout du partenaire. En conclusion, la farnésylation du peptide ProgFarn n'augmente pas l'affinité du peptide ni pour l'IgFold, ni pour BAF. Le mécanisme régissant la survenue du syndrome HGPS pourrait donc ne pas être dû à un empêchement de l'interaction entre l'IgFold et BAF.

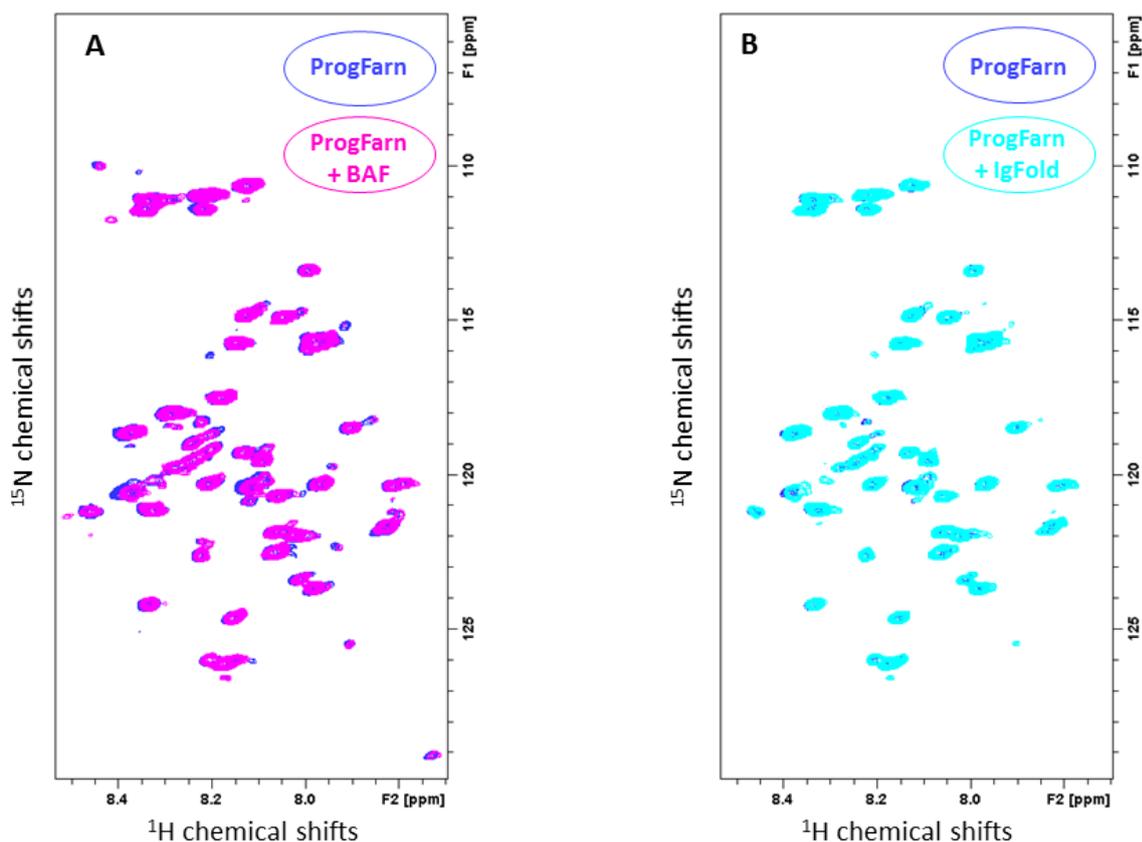


Figure 75 : Etude de l'interaction du peptide ProgFarn seul ou en présence de BAF d'une part ou de l'IgFold d'autre part. (A) Superposition des spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ du peptide ProgFarn seul (bleu) ou en présence de BAF (violet). (B) Superposition des spectres $^1\text{H}^{15}\text{N}$ SOFAST-HMQC²⁵³ du peptide ProgFarn seul (bleu) ou en présence de l'IgFold (cyan). Les spectres ont été enregistrés à 700 MHz, 283 K, et 150 μM de protéines dans un tampon contenant 20 mM phosphate pH 6.5 ; 150 mM NaCl ; 2 mM DTT, 10% D_2O .

II.3.4.2. Interaction entre la région C-terminale désordonnée de la lamine A et SUN1

Lorsque j'ai voulu tester l'interaction entre les domaines C-terminaux désordonnés de la prélamine A et de la progérine et la région N-terminale de SUN1, j'ai été confronté au problème de solubilité de cette dernière. En effet, SUN1 étant produit dans les corps d'inclusion bactériens, j'ai dû purifier cette protéine en urée. Malheureusement, il m'a été impossible de renaturer cette protéine. Cependant, une possibilité pour étudier cette interaction aurait été de purifier le complexe formé par SUN1 et la prélamine A. J'ai donc purifié le peptide PreLamC *in vitro* et l'ai mis en contact avec SUN1 1-139 purifié mais toujours dans un tampon contenant 8M d'urée. L'objectif était donc de renaturer les deux protéines en même temps, l'interaction entre les deux partenaires avait pour finalité de stabiliser SUN1 1-139 et de permettre sa co-purification. Malheureusement, la protéine a tout de même précipité dans ces conditions. Afin de poursuivre dans cette voie, il nous faudrait essayer cette même méthode en co-purifiant SUN1 1-139 en présence d'un échantillon de PreLamC farnésylé ou de ProgFarn.

II.3.4.3. Etude de l'interaction entre la région C-terminale désordonnée de la lamine A avec RBBP4

Afin d'obtenir des informations moléculaires sur l'interaction entre le peptide PrelamC et RBBP4, j'ai produit par RMN un échantillon de PrelamC-GB1 marqué à l'azote ^{15}N , et grâce au Dr. Pascal Drevet et à Ambre Petitalot, j'ai obtenu RBBP4 purifié. J'ai donc enregistré des spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide PrelamC-GB1 seul ou en présence de RBBP4 (Figure 76). Malheureusement, aucun changement ne fut visible sur le spectre, suggérant l'absence d'interaction dans nos conditions.

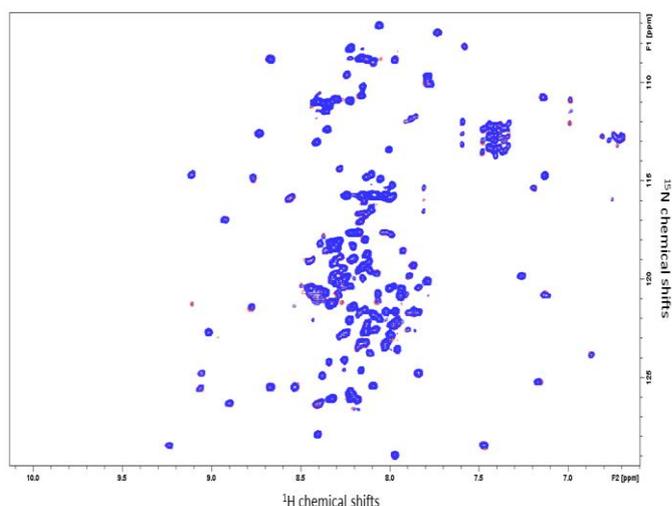


Figure 76 : Etude de l'interaction entre le peptide PrelamC-GB1 et RBBP4. Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide PrelamC-GB1 seul (saumon) ou en présence de RBBP4 (bleu). Les spectres ont été enregistrés avec 50 μM de protéine, à 283 K et 700 MHz dans un tampon contenant 20 mM phosphate pH 6.5 ; 150 mM NaCl ; 2 mM DTT et 10% D_2O .

En 2009, une étude a montré que BAF était capable de se lier à RBBP4²¹⁰. J'ai fait l'hypothèse que cette interaction est directe. Afin de confirmer cette hypothèse, j'ai réalisé en collaboration avec Ambre Petitalot, des expériences de filtration sur gel avec BAF seul ou en présence de RBBP4 (Figure 77). Un décalage de volume d'élution est bien visible sur les gels réalisés entre BAF seul et en présence de RBBP4, indiquant une interaction entre les deux protéines.

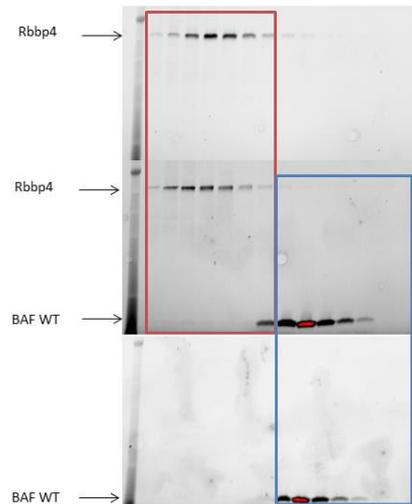


Figure 77 : Interaction entre RBBP4 et BAF. Filtrations sur gels réalisées à 100 μ M de protéines dans un tampon contenant 50 mM Tris-HCl, pH 8, 150 mM NaCl, 2 mM DTT grâce à une colonne Superdex 75 10/300 GL.

J'ai donc voulu savoir si l'interaction entre la lamine A et RBBP4 ne passait pas par la formation d'un complexe avec BAF, permettant aux trois protéines d'interagir ensemble. J'ai donc enregistré un spectre $^1\text{H}^{15}\text{N}$ HSQC du peptide PreLamC-GB1 en présence de BAF et de RBBP4 (Figure 78). Malheureusement, aucune interaction n'est visible par RMN dans nos conditions.

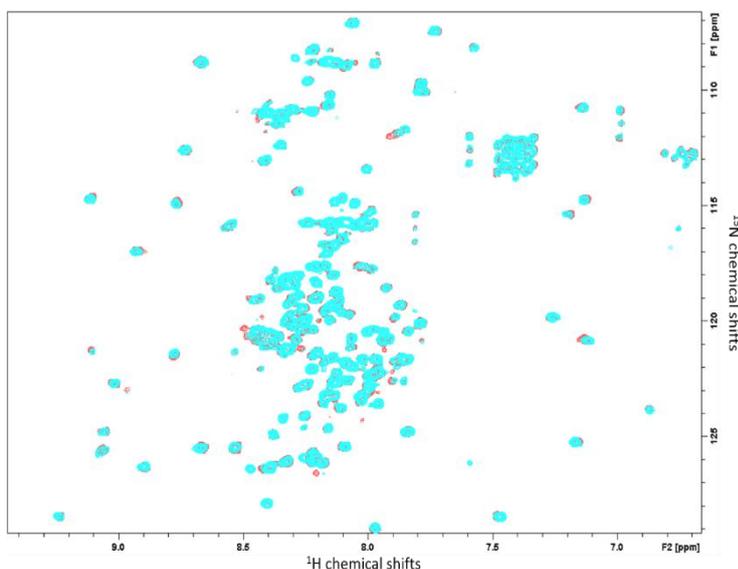


Figure 78 : Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide PreLamC-GB1 seul (saumon) ou en présence de BAF et RBBP4 (cyan). Les spectres ont été enregistrés avec 50 μ M de protéines, à 283 K et 700 MHz dans un tampon contenant 20 mM phosphate pH 6.5 ; 150 mM NaCl ; 2 mM DTT ; 10% D_2O .

Ainsi, dans nos conditions, la protéine RBBP4 interagit faiblement avec BAF, et n'interagit pas avec la région 567-664 de la lamine A. Et le complexe entre RBBP4 et BAF n'interagit pas non plus avec cette région C-terminale de la lamine A.

III. Discussion

Les protéines intrinsèquement désordonnées jouent, grâce à leurs caractéristiques uniques, des rôles essentiels dans les mécanismes biologiques tels que la signalisation cellulaire, l'expression des gènes ou encore le remodelage de la chromatine. Au cours de ma thèse, je me suis intéressé aux IDR de protéines impliquées dans l'architecture de l'enveloppe nucléaire : l'émerine et la prélamine A ainsi qu'à leurs partenaires.

L'émerine est impliquée dans la résistance du noyau aux stress mécaniques au cours desquels elle est phosphorylée pour augmenter la résistance du noyau à ces stress¹⁰⁴. Dans la littérature, il est décrit que l'émerine est capable de s'auto-assembler. L'émerine pourrait ainsi adopter une conformation « up » qui lui permettrait de s'auto-associer et d'interagir avec ses partenaires à l'enveloppe nucléaire, tandis que sa conformation « down » permettrait son interaction avec la protéine BAF associée à l'ADN¹¹⁰. La conformation « up » résulterait de son auto-assemblage, qui mettrait en jeu des interactions inter-moléculaires impliquant les résidus 1-132 et 170-220. Il a été montré dans notre équipe que l'émerine pouvait s'associer sous forme d'oligomères *in vitro*¹¹¹. Sous cette forme, l'émerine est capable d'interagir directement avec la lamine A²²⁹. Ces résultats suggèrent que les oligomères observés *in vitro* ne sont peut-être pas ceux qui se trouvent à l'enveloppe nucléaire, mais qu'ils contiennent une sous-structure fonctionnelle qui elle doit être présente dans les cellules. Je me suis alors intéressé au rôle de la région désordonnée de l'émerine dans la formation de ces oligomères.

La région désordonnée de l'émerine participe à l'assemblage de deux types de filaments

L'émerine possède un domaine LEM structuré (résidus 1 à 49), suivi d'une grande région prédite comme désordonnée (résidus 50 à 221) et d'un domaine transmembranaire (résidus 222 à 254). Isaline Herrada a montré que l'émerine 1-187 est capable de s'auto-assembler¹¹¹. J'ai montré que l'émerine 67-221 est aussi capable de s'auto-assembler, alors que le fragment 67-187 reste toujours monomérique dans nos conditions²²⁹. Ainsi, je peux conclure que la région prédite comme désordonnée est impliquée dans l'auto-assemblage de l'émerine et que deux mécanismes d'assemblage sont possibles, l'un nécessitant le fragment 1-66 et l'autre le fragment 188-221.

Afin de comprendre comment ces filaments s'assemblent, j'ai tout d'abord caractérisé la structure 3D des fragments 1-187 et 67-221 monomériques. Pour cela, j'ai utilisé la Résonance Magnétique Nucléaire qui permet de décrire la conformation d'une protéine en solution, qu'elle soit complètement dépliée, partiellement repliée ou complètement repliée. En collaboration avec Camille Samson, j'ai montré que le fragment 1-187 comprend bien un domaine LEM replié en N-terminal, puis une grande région déstructurée incluant une petite hélice α partiellement peuplée entre les acides aminés 98 et 101. Le fragment 67-221 est globalement déstructuré²²⁹. J'ai attribué les signaux RMN de sa région 67-170, qui ne comporte pas de structure secondaire détectable à part l'hélice α transitoire entre les acides aminés 98 et 101²⁵⁰. Il reste à attribuer et caractériser plus précisément au niveau structural la région 171-221. Cependant, cette région contient une séquence poly-sérine entre les acides

aminés 185 et 199 (12 sérines parmi 16 résidus) qu'il sera certainement difficile d'attribuer de manière non ambiguë.

Puis, comme Camille Samson avait alors montré que les oligomères d'émerine 1-187 étaient fonctionnels (ils sont capables d'interagir directement avec le domaine IgFold de la lamine A/C), je me suis concentré sur le mode d'auto-assemblage de l'émerine 1-187. Camille Samson avait montré que les fragments 1-187 mais aussi 1-132 étaient capables de s'auto-assembler, et Isaline Herrada avait montré que le fragment 1-187 ayant perdu ses acides aminés 95 à 99 (une délétion causant la dystrophie musculaire d'Emery-Dreifuss) restait monomérique dans nos conditions^{111,251}. L'auto-assemblage nécessite donc à la fois la région 1-66, contenant le domaine LEM, et la région 95-99, contenant la petite hélice α transitoire. De manière intéressante, Nada Essawy avait identifié un autre mutant de l'émerine, delK37 (une délétion causant des atteintes cardiaques), dont le domaine LEM est désordonné et qui s'auto-assemble avec une cinétique plus rapide que celle de l'émerine WT²⁵¹. C'est ainsi que j'ai émis l'hypothèse d'un changement de structure du domaine LEM durant la formation des filaments de l'émerine. Afin de valider cette hypothèse, j'ai produit un mutant de l'émerine 1-132 (Y4CE35C) dans lequel j'ai créé un pont disulfure artificiel en mutant les résidus 4 et 35 en cystéine. Ce mutant est produit sous forme oxydée afin de maintenir le domaine LEM sous forme repliée. La structure repliée du domaine LEM de ce mutant a pu être validée par RMN²⁵¹. Les expériences de cinétiques de fluorescence à la thioflavine T ont montré que ce mutant n'était pas capable de former de filaments lorsqu'il est oxydé. Cependant, après ajout d'un agent réducteur, le domaine LEM se déplie (les signaux RMN se déplacent vers le centre du spectre, correspondant aux déplacements chimiques de protéines désordonnées). Les expériences de cinétique par fluorescence à la thioflavine T ont montré que le mutant, une fois réduit, pouvait former des filaments avec la même cinétique que l'émerine WT. En conclusion, un changement de conformation du domaine LEM de l'émerine est nécessaire pour obtenir la formation des filaments.

Seule l'équipe du Dr. K. Wilson a aussi cherché à identifier quels sont les mécanismes responsables de l'auto-assemblage de l'émerine¹¹⁰. Leurs résultats prouvent que les régions 1-132 et 170-220 sont capables de s'associer avec l'émerine 1-220 mais aussi avec son fragment 170-220. Ceci est compatible avec nos observations que 3 régions contiennent des motifs essentiels à l'auto-assemblage : les régions 1-66, 95-99 et 188-221. De plus l'équipe de K. Wilson montre que les fragments d'émerine 1-132 mais aussi 159-220 sont capables de se lier directement à la région queue de la lamine A/C¹¹⁰. Suivant nos observations qui montrent que l'émerine doit s'auto-associer pour interagir avec la région queue de la lamine A/C, ces résultats impliqueraient que les régions 1-132 et 159-220 sont capables de s'auto-associer, ce qui est encore compatible avec nos résultats. Enfin, l'équipe de K. Wilson montre par surexpression d'émerines étiquetées dans des cellules exprimant l'émerine WT et coimmunoprécipitation que la région la plus critique pour l'auto-assemblage de l'émerine est la région 187-220¹¹⁰. J'ai montré de manière cohérente que l'émerine 67-221, mais pas 67-187, était capable de s'auto-assembler plus rapidement que l'émerine 1-187. De manière

intéressante, cette équipe montre aussi de la même façon que la délétion de la région 168-186 facilite l'auto-assemblage : est-ce parce que les différents motifs nécessaires à l'auto-assemblage sont alors plus proches les uns des autres ? Cependant ce mutant s'agrège au sein de foyers proches de la membrane nucléaire. Alors qu'en l'absence des 3 régions que nous avons identifiées comme nécessaires à l'auto-assemblage (notées L, M et R dans leur étude), l'émerine a tendance à rester dans le réticulum endoplasmique / cytoplasme. L'auto-assemblage de l'émerine semble donc nécessaire à sa localisation nucléaire.

Comment la phosphorylation de l'émerine régule son auto-assemblage ?

Notre analyse structurale de l'émerine est limitée par au moins 2 choix que nous avons fait pour simplifier notre étude: (1) nous avons travaillé avec la région nucléoplasmique de l'émerine, en absence de la région transmembranaire ; or la région nucléoplasmique seule (résidus 1 à 221) n'est pas soluble à une concentration supérieure au micromolaire, il a donc fallu l'étudier en la fragmentant en deux segments qui se chevauchent : les segments 1-187 et 67-221 ; (2) nous avons travaillé avec une émerine non modifiée ; or les études de spectrométrie de masse montrent que l'émerine est très fortement phosphorylée à la fois en interphase et pendant la mitose²⁴⁸; ces phosphorylations influencent très probablement sa structure locale et sa capacité à s'auto-assembler. En particulier, il a été montré qu'après un stress mécanique, l'émerine est phosphorylée sur ses tyrosines 74 et 95, et ceci est essentiel pour la réponse du noyau à ce stress¹⁰⁴. Or la tyrosine 95 appartient à un motif important pour l'auto-assemblage de l'émerine 1-187. Cette phosphorylation pourrait empêcher ou faciliter l'auto-association de l'émerine.

Comme j'ai attribué les signaux RMN du fragment 1-170 de l'émerine, il est maintenant possible d'utiliser cette attribution pour identifier *in vitro* quels résidus de l'émerine sont phosphorylés par les kinases régulant sa fonction. A titre d'exemple, je vous présente des expériences préliminaires que j'ai réalisées en collaboration avec Nada Essawy dans l'équipe. Nous avons incubé le fragment d'émerine 1-187 marqué à l'azote ¹⁵N avec les kinases PKA, p38 et PLK1, prédites comme phosphorylant l'émerine (et dans le cas de PLK1 par exemple, la voie de signalisation dépendante de cette kinase est déficiente dans les cellules n'exprimant pas d'émerine²⁵⁵). La Figure 79 montre les expériences RMN 2D ¹H-¹⁵N HSQC enregistrées après phosphorylation avec les différentes kinases. Je peux maintenant identifier les pics de sérine et thréonine ayant été fortement déplacés sur le spectre du fait de la phosphorylation, ainsi que les pics des résidus voisins dont l'environnement chimique est modifié lors de la phosphorylation du peptide. De cette manière, il apparaît par exemple que PKA phosphoryle tout d'abord la sérine S120 alors que PLK1 phosphoryle préférentiellement la thréonine T67. Une fois ces événements de phosphorylation bien caractérisés, il sera possible d'analyser leur impact sur la structure 3D et l'auto-assemblage de l'émerine.

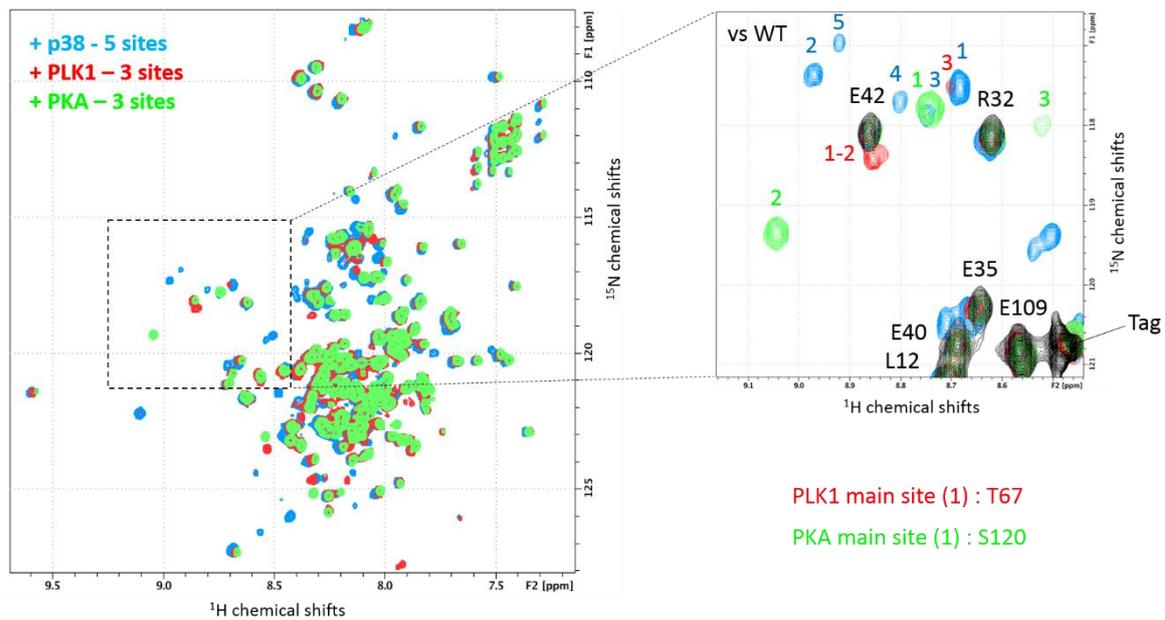


Figure 79 : Analyse par RMN des phosphorylations de l'émerine 1-187 par différentes kinases. Les spectres de l'émerine phosphorylée sont affichés en différentes couleurs selon la kinase utilisée : bleu pour p38, rouge pour PLK1 et vert pour PKA. Ils résultent d'expériences $^1\text{H}^{15}\text{N}$ SOFAST-HSQC enregistrées avec $100\ \mu\text{M}$ d'émerine 1-187, à 700 MHz, 303 K, dans un tampon contenant 20 mM phosphate pH 6.8 ; 30 mM NaCl ; 2 mM DTT. Les pics correspondants aux résidus phosphorylés sont montrés dans le zoom de droite : ce sont les pics qui ne sont pas présents dans le spectre de référence, en noir, enregistré sur un échantillon d'émerine avant phosphorylation. Le nombre de ces pics correspond au nombre de résidus phosphorylés. L'identification des pics du spectre de référence ayant disparu lors de la phosphorylation permet de connaître quels sont les résidus phosphorylés par la kinase.

Les régions C-terminales de la lamine A et de la progérine forment une hélice α transitoire conservée en séquence

Après avoir étudié la structure des filaments formés par l'émerine et permettant son interaction avec la lamine A, je me suis intéressé à la région C-terminale désordonnée qui est spécifique de la prélamine A (résidus 567-664). Cette région subit une série d'événements de maturation et notamment l'ajout d'un groupement farnésyl sur la cystéine 661, qui sera par la suite retiré par l'enzyme Zmpste24 (coupure entre les résidus 646 et 647) pour former la lamine A mature. Un mutant de la lamine A, G608G, ne comporte plus le site de clivage nécessaire pour retirer le groupement farnésyl en fin de maturation. Par conséquent, ce mutant reste constitutivement farnésylé. La présence de ce mutant, appelé progérine, est un phénomène naturel lié au vieillissement cellulaire, cependant, son accumulation entraîne la survenue d'un syndrome de vieillissement accéléré appelé progéria.

Afin d'étudier les différences structurales entre les régions C-terminales désordonnées de la prélamine A et de la progérine, j'ai produit, purifié et étudié par RMN les régions 567-664 pour la prélamine A (PreLamC), et 567-614 pour la progérine (ProgC). J'ai confirmé le caractère déplié de ces peptides. De plus, l'étude des déplacements chimiques secondaires $\text{C}\alpha$ et $\text{C}\beta$ montre la présence d'une structuration transitoire en hélice α dans la région 576-585, commune à la prélamine A et à la progérine. La présence de cette hélice α transitoire et

conservée en séquence suggère un site de liaison à un partenaire. La séquence 581-584 est reconnue par le serveur ELM (pour Eukaryotic Linear Motif : <https://elm.eu.org>) comme un motif di-arginine de rétention au réticulum endoplasmique. Elle contient la mutation R582H, trouvée chez une famille avec un FPLD atypique²⁵⁶.

Enfin, j'ai montré que l'hélice α transitoire était capable d'interagir aussi de manière transitoire avec le domaine Igfold au sein de la région queue de la progérine. En effet, la comparaison des spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide ProgC et de la queue de la progérine IgFProg met en évidence des variations faibles mais significatives des signaux RMN de la région 577-584 entre ces deux fragments de progérine, variations qui s'expliquent par un environnement chimique différent de cette région dans les deux fragments. Une telle interaction intramoléculaire pourrait permettre de réguler l'affinité de la progérine pour ses partenaires, qui doivent déplacer cette interaction pour pouvoir interagir avec l'Igfold ou la région 577-584. Cependant cette interaction reste faible : si l'Igfold n'est pas lié covalamment au peptide ProgC, je n'ai pas observé d'interaction entre ces deux régions de la progérine, que le peptide ProgC soit non modifié ou même farnésylé.

La région C-terminale de la progérine n'empêche pas l'interaction entre le domaine Igfold et le dimère de BAF

Récemment, Camille Samson au laboratoire a résolu la structure par cristallographie aux rayons X du domaine IgFold de la lamine A en complexe avec le domaine LEM de l'émerine et le dimère de BAF²²⁹ (Figure 80). Camille Samson a également montré que plusieurs mutants de la lamine A associés à des syndromes progéroides autosomaux récessifs (A529V, R471C, R527H, K542N et R435C) possèdent un défaut d'interaction avec le dimère de BAF²²⁹. Ainsi chez les patients atteints de ces pathologies, la lamine A, qui est mutée sur les deux allèles, lie de manière moins efficace la protéine BAF. Ceci s'explique par le fait que les mutations sont toutes situées à l'interface entre l'IgFold et le dimère de BAF.

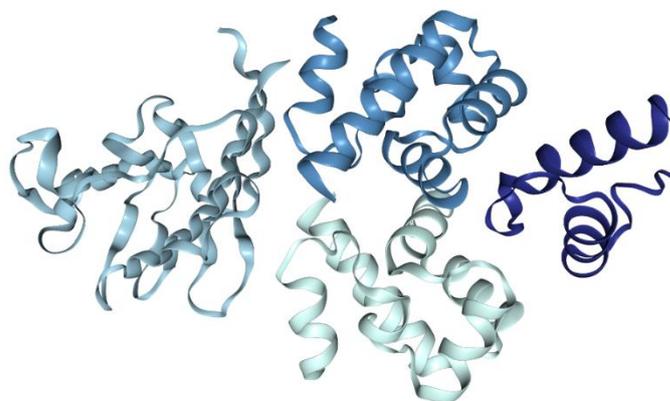


Figure 80 : Structure cristallographique du domaine IgFold de la lamine A/C (gris, gauche) en interaction avec le dimère de BAF (bleu et bleu-clair, centre) lui-même en interaction avec le domaine LEM de l'émerine (bleu foncé, droite) (Code PDB : 6GHD).

La progéria HGPS est une maladie dominante. Je me suis quand même demandé si un défaut d'interaction entre la lamine A et BAF pouvait aussi contribuer à ce syndrome de vieillissement prématuré. J'ai donc supposé que la présence de la progérine pouvait empêcher toute interaction entre la lamine A normale et BAF. Plus précisément, il est possible que la région C-terminale désordonnée de la progérine agisse comme élément de régulation en se glissant à l'intérieur de l'interface entre les deux partenaires et, par interaction avec l'un des deux partenaires, empêche la formation de ce complexe. J'ai donc étudié l'interaction de la queue de la progérine IgProg ou du peptide ProgC avec l'Igfold et BAF. Le peptide ProgC ne lie pas l'Igfold, comme déjà dit plus haut. Et l'IgProg lie très bien BAF, ainsi que montré par ITC par Camille Samson (Figure 81).

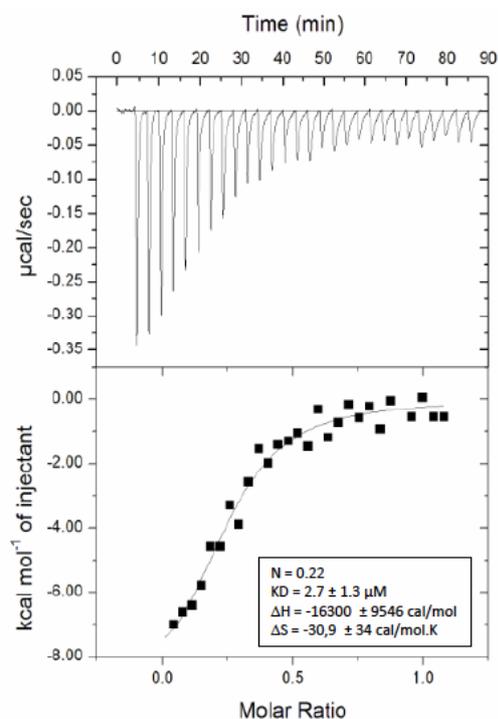


Figure 81 : Courbe de titration suivie par ITC, illustrant l'interaction entre le fragment IgFProg (domaine Igfold et région C-terminale de la progérine) et BAF. Injections successives de 10 μ l d'IgFProg concentré à 120 μ M dans une solution de BAF concentré à 24 μ M. Toutes les protéines ont été dialysées contre le même tampon contenant 20 mM Tris-HCl pH8, 150 mM NaCl, 10 mM β -mercaptoethanol et des inhibiteurs de protéase Roche®. Les expériences ont été faites à 288 K.

Enfin, le peptide ProgC ne lie pas BAF, comme je l'ai vérifié par RMN. J'ai également testé l'interaction entre le peptide ProgFarn et le dimère de BAF mais sans pouvoir observer d'interaction. Enfin, des résultats préliminaires que j'ai obtenus par RMN montrent que l'IgFProg farnésylé est toujours capable d'interagir avec BAF. Ceci suggère donc que le mécanisme par lequel la progérine est pathologique ne passe pas par une absence d'interaction entre l'IgFold et le dimère de BAF.

Cependant l'idée que la progérine farnésylée capture la lamine A endogène et empêche ses interactions avec d'autres partenaires est présente dans la littérature. En particulier, Lee et al. 2016²⁵⁷ propose d'utiliser des inhibiteurs de l'interaction entre progérine et lamine A/C pour améliorer le phénotype des patients HGPS.

Quels partenaires pour la région C-terminale spécifique à la lamine A ?

La région C-terminale désordonnée de la prélamine A, et plus particulièrement le fragment 562-664, interagit avec la protéine RBBP4, membre du complexe NuRD. Cette interaction n'est pas retrouvée dans le cas de la progérine non farnésylée²¹⁰. Par ailleurs, la protéine SUN1 a également été caractérisée comme liant la lamine A mais pas la progérine¹¹². Cependant, une seconde étude a montré que la farnésylation de la lamine A augmentait l'affinité de l'interaction et permettait à la progérine d'interagir avec SUN1¹⁶⁹.

Afin d'étudier l'interaction de la région C-terminale de la lamine A avec RBBP4, j'ai produit les peptides PreLamC et PreLamC-GB1 marqués à l'azote ¹⁵N. Il m'a été très difficile d'obtenir un échantillon pur et concentré du peptide PreLamC. J'ai donc choisi de produire PreLamC-GB1. Dans ce cas, les rendements sont très faibles, mais la protéine obtenue est pure. En parallèle, RBBP4 a été produit en cellules d'insectes sur la plateforme de production de protéines du laboratoire (Resp. : Pascal Drevet). La purification a montré une tendance de la protéine à dimériser avant la coupure de l'étiquette 6His. Ambre Petitalot a donc mis au point un protocole de purification en 3 étapes, qui comprend une étape de coupure de l'étiquette à la thrombine. J'ai alors testé l'interaction entre le peptide PreLamC-GB1 marqué à l'azote ¹⁵N et le monomère de RBBP4 par RMN. Aucune interaction n'a été observée. RBBP4 étant connue pour interagir avec BAF (dans un extrait cellulaire) et la lamine étant également un partenaire de BAF, j'ai envisagé la possibilité de former un complexe à trois entre PreLamC-GB1, BAF et RBBP4. Ainsi, j'ai testé par RMN l'interaction entre le peptide PreLamC-GB1 marqué ¹⁵N et les deux autres partenaires. A nouveau, aucune interaction n'a été observée.

La première hypothèse que l'on peut émettre pour expliquer ces observations est que notre peptide contient les acides aminés 567 à 664 de la prélamine A. Or l'interaction avec RBBP4 a été vue par double hybride en levure en utilisant le fragment 562-664 de la prélamine A comme appât²¹⁰. La séquence 562-566 est : LHHHH. Cette séquence est peut-être essentielle à l'interaction avec RBBP4. Cette hypothèse est renforcée par l'observation que RBBP4 avec une étiquette 6His a tendance à dimériser. Je voudrais donc maintenant tester l'interaction des fragments 389-646 de la lamine A (comprenant l'Igfold et la région C-terminale mature), 389-664 (comprenant l'Igfold et la C-terminale de la prélamine A) et IgFProg avec RBBP4. La deuxième hypothèse est que la présence de la protéine GB1 gêne la reconnaissance des derniers résidus de PreLamC, essentiels à l'interaction avec RBBP4. J'ai donc commandé des peptides synthétisés chimiquement correspondants aux 15 derniers résidus de PreLamC et ProgC. Je voudrais tester l'interaction de RBBP4 avec ces peptides. Enfin, j'ai montré que RBBP4 interagit faiblement mais spécifiquement avec BAF. Si une interaction directe entre RBBP4 et la lamine A est observée, il faudra alors tester si BAF est capable de stabiliser cette interaction.

Concernant la protéine SUN1, j'ai tenté de produire et de purifier le fragment N-terminal 1-139 de SUN1 murin. Cependant, la purification s'effectuant dans l'urée, je n'ai jamais réussi à renaturer la protéine sans qu'elle précipite. J'ai donc essayé de mettre SUN1

1-139 en contact avec le peptide PreLamC avant la renaturation. Ceci, dans le but de stabiliser le complexe au cours de la dialyse. Malheureusement, la protéine a tout de même précipité dans ces conditions. Comme, dans la littérature, une interaction plus forte est décrite lorsque la prélamine est farnésylée¹⁶⁹, et une interaction est observée dans le cas de la progérine farnésylée¹⁶⁹, peut-être qu'une renaturation du fragment 1-139 de SUN1 en présence du peptide PreLamC farnésylé ou ProgC farnésylé permettrait d'obtenir le complexe sous forme soluble.

La région C-terminale spécifique à la lamine A contribue-t-elle à la formation de foyers nucléoplasmiques de lamine A ?

Malgré les difficultés rencontrées lors de l'expression et la purification de PreLamC, j'ai pu caractériser la structure et la dynamique de ce peptide ainsi que de son mutant ProgC. L'analyse des intensités mesurées sur les spectres ¹H¹⁵N HSQC enregistrés sur les peptides PreLamC et ProgC nous donne une information sur la dynamique interne du peptide, à l'échelle du résidu, qui est assimilable à des mesures de relaxation transversale (T₂). Ces mesures m'ont permis de mettre en évidence que les régions situées avant la délétion de 50 acides aminés chez la progérine, possédait la même mobilité dans les deux constructions. Cependant, la région spécifique à la lamine A subit une forte diminution d'intensité dans la construction PreLamC, ce qui indique une différence de mobilité par rapport au reste de la molécule. Cette diminution d'intensité peut être expliquée par la présence d'interactions intermoléculaires transitoires possédant des constantes d'échange de l'ordre de la milliseconde. De manière intéressante, le construit de la région C-terminale dépliée de la prélamine A possédant une étiquette GB1 en C-terminal (PreLamC-GB1) ne possède plus ce comportement. J'ai donc fait la supposition que la présence de l'étiquette GB1 empêche l'apparition d'interactions transitoires entre molécules de PreLamC. Il est alors possible d'observer la région 600-635 du peptide, caractérisée par une hausse d'intensité du signal RMN par rapport au reste du peptide PreLamC-GB1. Cette hausse d'intensité reflète des mouvements d'amplitude particulièrement importante sur une échelle de temps rapide (de la picoseconde à la nanoseconde). Il est important de noter que cette région 600-635 contient 28% de glycines et 31% de sérines, ce qui explique sa flexibilité.

J'ai obtenu récemment des résultats préliminaires par RMN et microscopie électronique à partir desquels je pense que le peptide PreLamC est capable de former des objets de grande taille par transition de phase (Figure 82). En effet, le peptide PreLamC concentré à 450 μM pour la RMN peut évoluer en une nuit pour former des structures au sein desquelles la majorité des glycines et sérines sont en échange intermédiaire du point de vue de la RMN (Figure 82A). Ces structures sont de haut poids moléculaire ainsi que révélé par chromatographie (Figure 82B) et montrent une organisation de type gouttelette lipidique par microscopie électronique (Figure 82C). Les régions responsables des interactions intermoléculaires dans ces structures sont des zones riches en glycines et sérines : résidus 599-606, 612-619 et 626-636, présents uniquement chez la prélamine A. Il me faudrait

maintenant confirmer ces résultats. De plus, il reste à montrer que la région PreLamC est capable d'induire des interactions transitoires entre lamines et/ou des transition de phase lorsqu'elle se trouve au sein des prélamine et/ou lamine A entières.

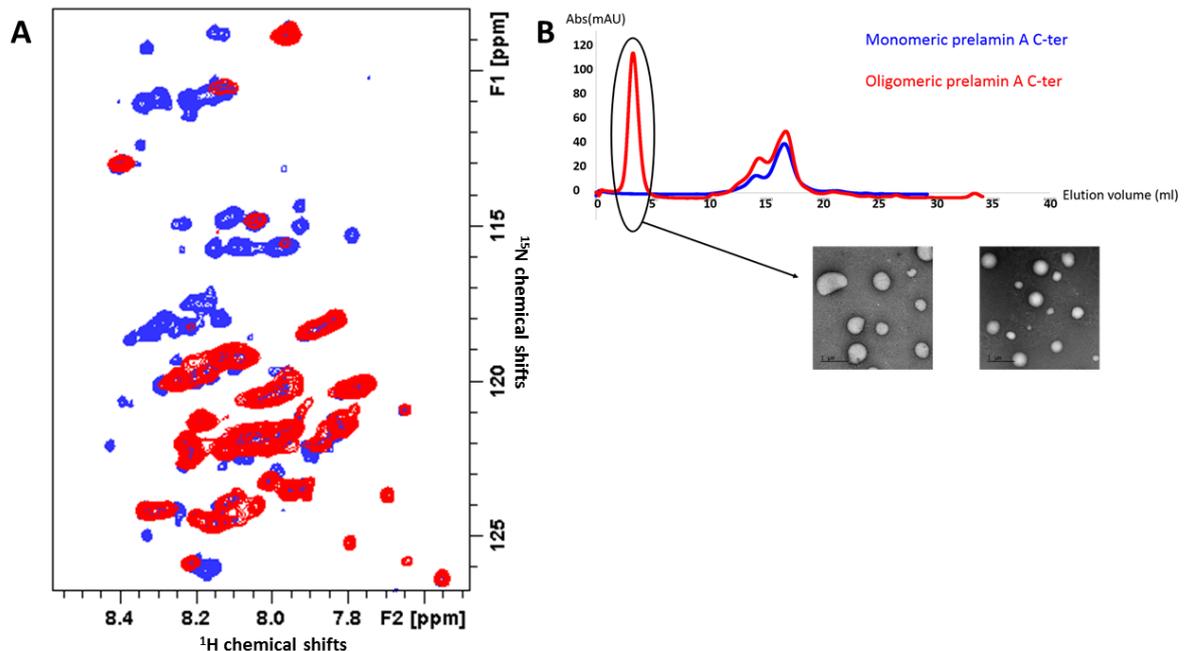


Figure 82 : Le peptide PreLamC est capable de former des structures de haut poids moléculaire. (A) Superposition des spectres $^1\text{H}^{15}\text{N}$ HSQC du peptide PreLamC monomérique (noir) et oligomérique (rouge), montrant qu'une majorité de glycines et sérines sont en échange intermédiaire dans le spectre rouge (leurs pics ne sont plus visibles). (B) Profil de filtration sur gel illustrant la présence d'objets de grande taille dans l'échantillon donnant le spectre RMN affiché en rouge (courbe rouge) en comparaison du profil de filtration sur gel du peptide PreLamC donnant le spectre RMN affiché en noir (courbe bleue). (C) Images de microscopie électronique du pic purifié d'objets de grande taille, révélant la présence de structures de type gouttelettes lipidiques. L'agent de contraste est l'acétate d'uranyl.

Cependant, ces résultats rappellent certaines observations publiées sur l'organisation des lamines en cellules. En effet, plusieurs études ont montré que la progérine non farnésylée pouvait former des agrégats nucléoplasmiques ressemblants à des gouttelettes lipidiques et que la lamine A/C endogène colocalisait alors avec la progérine dans ses foyers^{258,259}. Il a aussi été suggéré que la phosphorylation de la sérine 22 par CDK4/6 permettrait à la lamine A de quitter la membrane nucléaire et d'aller réaliser ses fonctions nucléoplasmiques par le biais de la formation de transitions de phase¹⁶⁵. La surexpression de mutants de lamine A causant des myopathies induit aussi la formation de foyers nucléoplasmiques contenant prélamine A et lamine A²⁶⁰. D'autres publications ont mentionné que la lamine A/C était présente dans des foyers nucléoplasmiques dans certaines conditions, comme après choc thermique en colocalisation avec des chaperones²⁶¹. Enfin, la lamine A a été identifiée comme présente dans les plaques amyloïdes et dans les enchevêtrements neurofibrillaires observés dans la maladie d'Alzheimer^{256,262}.

Conclusion-Perspective

Au cours de ma thèse, j'ai donc caractérisé la région désordonnée de l'émerine et montré son implication dans l'auto-association de la protéine. Cette auto-association est primordiale pour permettre à l'émerine d'interagir avec la lamine A. La région désordonnée de l'émerine étant la cible de modifications post-traductionnelles, il serait intéressant d'étudier l'impact de ces modifications chimiques sur la capacité de l'émerine à s'auto-assembler, ainsi que sur la structure des oligomères formés.

Concernant la région C-terminale désordonnée de la lamine A, j'ai montré la présence d'une hélice α transitoire dans la région commune aux régions C-terminales désordonnées de la prélamine A et de son mutant progérine, impliqué dans la progéria. Je propose de commander des peptides codant pour la région formant cette hélice et de réaliser des expériences de pull-downs en présence d'extraits nucléaires afin d'identifier les partenaires potentiels de cette région. Par ailleurs, j'ai montré que la prélamine A a une tendance à s'auto-assembler grâce à la région absente dans la progérine. Cette différence de comportement, ajouté au fait que la progérine est farnésylée en permanence, pourrait être responsable des défauts de localisation de la progérine et de la lamine A en présence de progérine, ainsi que des défauts d'interaction de ces protéines. Ainsi, il serait intéressant de poursuivre cette étude et d'affiner les conditions permettant à la lamine A de s'agréger et de former les transitions de phase que je suspecte.

Enfin, l'interaction entre le nucléosquelette et la chromatine / les protéines associées à la chromatine est défectueuse chez les patients atteints de progéria, mais aucun détail moléculaire n'est disponible sur ces interactions. Je propose de pousser plus loin l'analyse des interactions entre la région queue de la lamine A, RBBP4 et BAF, sur la base de mes résultats préliminaires, afin de comprendre comment des mutations dans la lamine A et BAF peuvent impacter la formation de ce complexe chez les patients atteints de syndromes progéroides.

IV. Matériel et méthodes

IV.1. Production des échantillons protéiques

IV.1.1 Tampons utilisés

Tampon C : 50 mM Tris-HCl, pH 8, 30 mM NaCl, 30 mM imidazole

Tampon C8 : Tampon C contenant 8M d'urée

Tampon RMN : 20 mM phosphate de sodium, 150 mM NaCl, 5 mM DTT

Tampon de farnésylation : 50 mM Tris-HCl, pH 8, 20 mM KCl, 5 mM MgCl₂, 10 μM ZnCl₂

Tampon de lyse : 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 40 mM imidazole, 5% glycerol, 1% Triton X-100, and 1 mM phenylmethanesulfonylfluoride

IV.1.2 Production des protéines à partir des corps d'inclusion

Les échantillons d'émerine, de SUN1 et de BAF ont été produits en cellules de *E.coli* BL21 (DE3) à 37°C en présence de kanamycine avec une induction sur la nuit à 20°C après ajout d'1 mM IPTG. Les cultures ont été centrifugées 30 min à 3500 rpm (rotor Beckman JA 8100, centrifugeuse Beckman Coulter) afin de récupérer les cellules. Les culots ont été stockés dans du tampon de lyse. Après lyse du culot bactérien, les extraits cellulaires ont été centrifugés durant 30 min à 20 000 rpm (rotor Beckman JA 25.50, centrifugeuse Beckman Coulter). Le culot récupéré a ensuite été repris dans 20 ml de tampon C8 et incubé durant 1h à température ambiante sous agitation. Une seconde étape de centrifugation de 30 min à 20 000 rpm a été effectuée pour récupérer le surnageant et l'incuber sur colonne de nickel (toutes ces protéines possédant une étiquette histidine). Après lavage avec du tampon C8, l'élution a été faite grâce à l'ajout de tampon E8. Les protéines ont ensuite été dialysées par bains successifs dans du tampon C contenant 30 mM NaCl pour l'émerine et 150 mM NaCl pour SUN1 et BAF (1 bain sur la nuit, puis deux fois 2h).

IV.1.3 Production des échantillons solubles

Pour les échantillons des différentes constructions de lamine A, de progérine ainsi que pour la farnésyl-transférase, les cellules bactériennes ont été transformées et mises en culture à 37°C en présence de kanamycine à l'exception de l'IgFold et de l'IgFoldprogérine qui

possèdent un gène de résistance à l'ampicilline. L'induction a été réalisée à l'aide d'1 mM IPTG et durant 3h à 37°C pour les constructions des peptides prélamine A et progérine. Les IgFold et IgFoldprogérine ainsi que la farnésyl-transférase ont été induits sur la nuit à 20°C. Pour la farnésyl-transférase, l'induction a été réalisée par ajout de 0,1 mM IPTG.

Pour les différentes constructions de lamine, à l'exception de la prélamine-GB1, les cellules ont été stockées dans du tampon de lyse, lysées puis centrifugées 30 min à 3500 rpm (rotor Beckman JA 8100, centrifugeuse Beckman Coulter). Le vecteur utilisé est Le surnageant a ensuite été incubé sur billes de GST équilibrées dans du tampon 1 durant 2h à 4°C après ajout de 5 mM DTT. Après lavage à dans du tampon de lavage, les peptides prélamine A et progérine ont été clivés sur colonne directement par ajout de protéase TEV et de 5 mM DTT durant 3h à 4°C. Le flow-through est ensuite récupéré. Après concentration, une précipitation à la chaleur (15 min à 95°C) est réalisée pour retirer les contaminants structurés par centrifugation 20 min à 20 000 rpm. Le surnageant est récupéré, concentré et passé sur colonne de filtration sur gel, équilibrée dans le tampon de RMN pour permettre une étude structurale. Les constructions d'IgFold et d'IgFoldprogérine ont été éluées après lavage et dialysées contre du tampon C, après concentration, une coupure à la protéase TEV a été faite durant 3h à 4°C. Une étape de filtration sur gel a ensuite été effectuée après équilibration dans du tampon RMN pour permettre une étude structurale.

Concernant la farnésyl-transférase et la prélamine-GB1, le surnageant a été placé sur colonne de nickel (les protéines possédant une étiquette histidine). Après incubation durant 2h à 4°C, un lavage a été fait avec du de lavage, puis les protéines ont été éluées sans coupure dans du tampon C.

Pour la protéine RBBP4, le culot (Uniprot : [Q09028](#)) nous a été donné par le Dr. Pascal Drevet après clonage à partir d'un vecteur pFBDM codant pour une étiquette 6-histidines ainsi que pour un site de coupure à la thrombine (fourni par le Dr. E.D. Laue, département de Biochimie, Université de Cambridge), en cellules d'insectes. Une fois la lyse réalisée par sonication, le lysat a été centrifugé 30 min à 20 000 rpm. Le surnageant a ensuite été incubé sur colonne de nickel 2h à 4°C dans un tampon contenant 50 mM Tris-HCl, pH8 ; 150 mM NaCl et 2 mM DTT. L'éluion a été faite dans le tampon précédant, supplémenté avec 1M imidazole. Après dialyse dans un tampon sans imidazol, la protéine a subi une coupure à la thrombine sur la nuit à 4°C. Des étapes de purification par colonne échangeuse d'anions ainsi que par filtration sur gel ont ensuite été réalisées pour obtenir la protéine pure.

IV.2. Microscopie électronique

Les échantillons d'émerine ayant permis l'obtention ont été produits selon le protocole décrit plus haut. Une fois purifiée, l'émerine a été concentrée à une concentration de 600 µM en présence de 5 mM DTT, et incubée à température ambiante durant 1 semaine afin de

permettre la formation des filaments. Les échantillons ont ensuite été dilués à une concentration de 50 μM . 10 μl ont alors été déposés sur grille de microscopie. Les images de microscopie électronique ont été collectées à l'I2BC de Gif-sur-Yvette avec un microscope électronique Tecnai Spirit équipé d'un filament LaB_6 à 100 kV ainsi qu'une caméra K2 (Gatan, 4Kx4K) après ajout de 2% d'acetate d'uranyl comme élément contrastant sur des grilles de cuivre recouvertes de carbone.

IV.3. Cinétiques de fluorescence à la thioflavine T

Après production des échantillons d'émerine, la cinétique de formation des filaments a été suivie en incubant les échantillons à une concentration de 300 μM et 310 K. Des aliquots ont été prélevés à 0 h; 1h ; 2h ; 3h ; 4h et 24h. Chaque aliquot a été dilué dans un tampon contenant 20 mM Tris-HCl pH 8, 30 mM NaCl, 5 mM DTT, de manière à obtenir une concentration finale de 20 μM de protéine et 2.5 mM de Thioflavine T. Les mesures de fluorescences ont été faites dans une cuve de 60 μl , à 293 K grâce à un fluorimètre JASCO ADP-303T en utilisant une longueur d'onde d'excitation de 440 nm et en enregistrant l'émission à une longueur d'onde de 480 nm.

IV.4. RMN en phase liquide

Les échantillons de protéines purifiées ont été dialysés contre un tampon contenant 20 mM de phosphate de sodium à pH 6.5, 150 mM NaCl (30 mM pour les constructions d'émerine), 5 mM DTT et 10% D_2O . Les expériences RMN ont été réalisées sur des spectromètres Bruker Avance II 600, 700 ou 750 MHz équipés de cryo-sondes. Les spectres RMN 2D $^1\text{H}^{15}\text{N}$ HSQC ont été enregistrés dans des tubes RMN 3 mm contenant des échantillons concentrés à 200 μM marqués uniformément au ^{15}N à 283 K (303 K pour les constructions d'émerine). Les expériences nOes hétéronucléaires $^1\text{H}^{15}\text{N}$ ont été enregistrées grâce à des séquences NOESY avec des tubes RMN 3mm contenant des échantillons marqués uniformément ^{15}N et concentrés à 400 μM . Les expériences RMN 3D HN(CA)CO, HNCO, HNCACB, CBCA(CO)NH, HN(CO)(CA)NH ont été enregistrées à 283 K (303 K pour les échantillons d'émerine) avec des tubes RMN 3mm contenant des échantillons marqués uniformément $^{13}\text{C}^{15}\text{N}$, concentrés à 500 μM . Les données RMN ont été transformées grâce au logiciel Topspin 4.0 et analysées en utilisant le logiciel CCPNMR.

IV.5. RMN du solide

Après purification et coupure de l'étiquette histidine, les protéines ont été dialysées contre un tampon contenant 20 mM Tris-HCl, pH8, 30 mM NaCl et 5 mM DTT, pour former les filaments de l'émerine. Les expériences de RMN du solide ont été pour l'émerine 1-187 et 1-132 ont été enregistrées sur un spectromètre Bruker Avance III HD 800 MHz avec un champ magnétique externe de 18.8 T ou sur un spectromètre Bruker Avance HD 700 MHz avec champ magnétique externe de 16.4 T. La température de l'échantillon a été utilisée en utilisant le signal de référence du proton de l'eau référencé grâce à du DSS, et maintenue à 283 K. Les expériences double quanta-simple quanta ont été enregistrées avec rotation à l'angle magique à 8 KHz et les spectres de corrélation C-C avec une fréquence de 11 KHz. Un découplage proton à haute puissance a été fait durant les périodes d'évolution et de détection grâce à SPINAL-64 à une puissance radio-fréquence de 83 KHz. Les spectres ont été processés avec Topspin3.2 et analysés avec CCPNMR. Les spectres double quanta-simple quanta ont été enregistrés en utilisant l'excitation SPC5 et des périodes de reconversion de 0.5 ms. La durée maximum d'acquisition en dimension directe était de 10 ms et de 1.2 ms pour la dimension indirecte. Pour les spectres C-C PDSO, le temps de mélange était de 50 ms. La période maximale d'acquisition était de 15 ms en dimension directe et de 10 ms en dimension indirecte.

IV.6. Farnésylation *in vitro*

Les farnésylations *in vitro* de la région C-terminale désordonnée de la progérine et de l'IgFoldProerine ont été réalisées en incubant la protéine purifiée à 50 μ M dans un tampon contenant du tampon de farnésylation ainsi que 5 équivalents de farnésyl-pyrophosphate par rapport à la concentration de protéine à farnésyler et 10 mM DTT avec 50 nM de farnésyl-transférase pure dialysée dans le même tampon. L'incubation est réalisée durant 3h à 303 K. Une étape de purification sur colonne de nickel est ensuite effectuée pour retirer la farnésyl-transférase dont l'étiquette histidine n'est pas coupée. Les protéines farnésylées ont ensuite été placées dans un tampon permettant leur étude par RMN.

IV.7. Microcalorimétrie

Les expériences d'ITC ont été enregistrées sur un appareil VP-ITC. Les protéines ont été dialysées dans un même tampon contenant 20 mM Tris-HCl, pH8 ; 150 mM NaCl et 10 mM β -mercaptoethanol et des inhibiteurs de protéase Roche®. Pour l'interaction IgFPro-BAF, des injections successives de 10 μ l d'IgFProg concentré à 120 μ M, ont été faites dans une solution de BAF concentrée à 24 μ M. Pour l'interaction IgFold-BAF, des injections successives de 10 μ l d'IgFold concentré à 270 μ M ont été faites dans une solution de BAF concentrée à 22.5 μ M. Les expériences ont été faites à 288 K

V. Références

1. Arnone, A. *et al.* A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. I. Experimental procedures and chain tracing. *J. Biol. Chem.* **246**, 2302–2316 (1971).
2. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature* **276**, 362–368 (1978).
3. Bode, W. & Huber, R. Crystal structure analysis and refinement of two variants of trigonal trypsinogen: trigonal trypsin and PEG (polyethylene glycol) trypsinogen and their comparison with orthorhombic trypsin and trigonal trypsinogen. *FEBS Lett.* **90**, 265–269 (1978).
4. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are ‘natively unfolded’ proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).
5. Dunker, A. K. *et al.* Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 473–484 (1998).
6. Uversky, V. N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci. Publ. Protein Soc.* **11**, 739–756 (2002).
7. Jones, D. T. & Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinforma. Oxf. Engl.* **31**, 857–863 (2015).
8. Kozlowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, 111 (2012).
9. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
10. Dunker, A. K. & Gough, J. Sequences and topology: intrinsic disorder in the evolving universe of protein structure. *Curr. Opin. Struct. Biol.* **21**, 379–381 (2011).
11. Dunker, A. K. & Obradovic, Z. The protein trinity—linking function and disorder. *Nat. Biotechnol.* **19**, 805–806 (2001).

12. Dunker, A. K., Brown, C. J. & Obradovic, Z. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* **62**, 25–49 (2002).
13. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
14. Boesch, C., Bundi, A., Oppliger, M. & Wüthrich, K. ¹H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur. J. Biochem.* **91**, 209–214 (1978).
15. Dunker, A. K. *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).
16. Williams, R. M. *et al.* The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 89–100 (2001).
17. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).
18. DeForte, S. & Uversky, V. N. Order, Disorder, and Everything in Between. *Mol. Basel Switz.* **21**, (2016).
19. Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* **43**, 1090–1103 (2011).
20. Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60 (2002).
21. Creamer, T. P. Transient disorder: Calcineurin as an example. *Intrinsically Disord. Proteins* **1**, e26412 (2013).
22. Weiss, M. A. *et al.* Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575–578 (1990).
23. Bracken, C., Carr, P. A., Cavanagh, J. & Palmer, A. G. Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.* **285**, 2133–2146 (1999).
24. Coleman, D. E. *et al.* Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis. *Science* **265**, 1405–1412 (1994).

25. Coleman, C. M. *et al.* A novel mutation in the helix termination motif of keratin K12 in a US family with Meesmann corneal dystrophy. *Am. J. Ophthalmol.* **128**, 687–691 (1999).
26. Graham, T. A., Ferkey, D. M., Mao, F., Kimelman, D. & Xu, W. Tcf4 can specifically recognize beta-catenin using alternative conformations. *Nat. Struct. Biol.* **8**, 1048–1052 (2001).
27. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).
28. Tanaka, M., Collins, S. R., Toyama, B. H. & Weissman, J. S. The physical basis of how prion conformations determine strain phenotypes. *Nature* **442**, 585–589 (2006).
29. Fowler, D. M., Koulov, A. V., Balch, W. E. & Kelly, J. W. Functional amyloid--from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224 (2007).
30. Bhattacharyya, R. P. *et al.* The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science* **311**, 822–826 (2006).
31. Fuxreiter, M., Tompa, P. & Simon, I. Local structural disorder imparts plasticity on linear motifs. *Bioinforma. Oxf. Engl.* **23**, 950–956 (2007).
32. Mohan, A. *et al.* Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* **362**, 1043–1059 (2006).
33. Radhakrishnan, I. *et al.* Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91**, 741–752 (1997).
34. Richards, J. P., Bächinger, H. P., Goodman, R. H. & Brennan, R. G. Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. *J. Biol. Chem.* **271**, 13716–13723 (1996).
35. Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R. & Wright, P. E. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **277**, 42241–42248 (2002).
36. Selenko, P. *et al.* Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell* **11**, 965–976 (2003).

37. Iakoucheva, L. M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).
38. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat. Struct. Mol. Biol.* **11**, 830–837 (2004).
39. Lindner, R. A. *et al.* Mouse Hsp25, a small shock protein. The role of its C-terminal extension in oligomerization and chaperone action. *Eur. J. Biochem.* **267**, 1923–1932 (2000).
40. Pometun, M. S., Chekmenev, E. Y. & Wittebort, R. J. Quantitative observation of backbone disorder in native elastin. *J. Biol. Chem.* **279**, 7982–7987 (2004).
41. Rauscher, S., Baud, S., Miao, M., Keeley, F. W. & Pomès, R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Struct. Lond. Engl. 1993* **14**, 1667–1676 (2006).
42. Huber, A. H. & Weis, W. I. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **105**, 391–402 (2001).
43. Babu, M. M., Kriwacki, R. W. & Pappu, R. V. Structural biology. Versatility from protein disorder. *Science* **337**, 1460–1461 (2012).
44. Baker, J. M. R. *et al.* CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat. Struct. Mol. Biol.* **14**, 738–745 (2007).
45. Ou, L. *et al.* Incomplete folding upon binding mediates Cdk4/cyclin D complex activation by tyrosine phosphorylation of inhibitor p27 protein. *J. Biol. Chem.* **286**, 30142–30151 (2011).
46. Bah, A. *et al.* Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109 (2015).
47. Metskas, L. A. & Rhoades, E. Folding upon phosphorylation: translational regulation by a disorder-to-order transition. *Trends Biochem. Sci.* **40**, 243–244 (2015).
48. Bah, A. & Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* **291**, 6696–6705 (2016).

49. Wang, Y.-T. *et al.* The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* **288**, 9049–9057 (2013).
50. Wegmann, S. *et al.* Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* **37**, (2018).
51. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1**, (2011).
52. Andrew, C. D., Warwicker, J., Jones, G. R. & Doig, A. J. Effect of phosphorylation on alpha-helix stability as a function of position. *Biochemistry (Mosc.)* **41**, 1897–1905 (2002).
53. Jin, J. & Pawson, T. Modular evolution of phosphorylation-based signalling systems. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 2540–2555 (2012).
54. Tamkun, J. W. *et al.* brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561–572 (1992).
55. Callebaut, I. & Mornon, J. P. The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in Drosophila melanogaster development. *Biochem. J.* **321** (Pt 1), 125–132 (1997).
56. Kowalski, A. Abundance of intrinsic structural disorder in the histone H1 subtypes. *Comput. Biol. Chem.* **59 Pt A**, 16–27 (2015).
57. Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends Genet. TIG* **27**, 389–396 (2011).
58. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed Engl.* **44**, 7342–7372 (2005).
59. McGlinchey, R. P. & Lee, J. C. Why Study Functional Amyloids? Lessons from the Repeat Domain of Pmel17. *J. Mol. Biol.* (2018). doi:10.1016/j.jmb.2018.06.011
60. Yang, M.-L., Hasadsri, L., Woods, W. S. & George, J. M. Dynamic transport and localization of alpha-synuclein in primary hippocampal neurons. *Mol. Neurodegener.* **5**, 9 (2010).

61. Dikiy, I. & Eliezer, D. Folding and misfolding of alpha-synuclein on membranes. *Biochim. Biophys. Acta* **1818**, 1013–1018 (2012).
62. Araki, K. *et al.* Synchrotron FTIR micro-spectroscopy for structural analysis of Lewy bodies in the brain of Parkinson's disease patients. *Sci. Rep.* **5**, 17625 (2015).
63. Chiti, F. *et al.* Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3590–3594 (1999).
64. Fowler, D. M. *et al.* Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4**, e6 (2006).
65. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
66. Zimmerman, R. J., Chao, H., Fullerton, G. D. & Cameron, I. L. Solute/solvent interaction corrections account for non-ideal freezing point depression. *J. Biochem. Biophys. Methods* **26**, 61–70 (1993).
67. St George-Hyslop, P. *et al.* The physiological and pathological biophysics of phase separation and gelation of RNA binding proteins in amyotrophic lateral sclerosis and fronto-temporal lobar degeneration. *Brain Res.* **1693**, 11–23 (2018).
68. Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* **11**, 777–788 (2010).
69. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell* **60**, 231–241 (2015).
70. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).
71. Hofweber, M. *et al.* Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **173**, 706–719.e13 (2018).

72. Lin, Z. *et al.* A comparative investigation between paclitaxel nanoparticle- and nanocrystal-loaded thermosensitive PECT hydrogels for peri-tumoural administration. *Nanoscale* **8**, 18782–18791 (2016).
73. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
74. Hyman, A. A. & Brangwynne, C. P. Beyond stereospecificity: liquids and mesoscale organization of cytoplasm. *Dev. Cell* **21**, 14–16 (2011).
75. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
76. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* **26**, 668–679 (2016).
77. Holt, C. E. & Schuman, E. M. The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron* **80**, 648–657 (2013).
78. Malinowska, L., Kroschwald, S. & Alberti, S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim. Biophys. Acta* **1834**, 918–931 (2013).
79. Arany, Z. *et al.* An essential role for p300/CBP in the cellular response to hypoxia. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12969–12973 (1996).
80. Freedman, S. J. *et al.* Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5367–5372 (2002).
81. Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5271–5276 (2002).
82. Travers, A. Chromatin modification by DNA tracking. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13634–13637 (1999).

83. Hartman, P. G., Chapman, G. E., Moss, T. & Bradbury, E. M. Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule. *Eur. J. Biochem.* **77**, 45–51 (1977).
84. Vila, R. *et al.* DNA-induced alpha-helical structure in the NH₂-terminal domain of histone H1. *J. Biol. Chem.* **276**, 46429–46435 (2001).
85. Vila, R., Ponte, I., Collado, M., Arrondo, J. L. & Suau, P. Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study. *J. Biol. Chem.* **276**, 30898–30903 (2001).
86. Roque, A., Iloro, I., Ponte, I., Arrondo, J. L. R. & Suau, P. DNA-induced secondary structure of the carboxyl-terminal domain of histone H1. *J. Biol. Chem.* **280**, 32141–32147 (2005).
87. Lu, X. & Hansen, J. C. Identification of specific functional subdomains within the linker histone H10 C-terminal domain. *J. Biol. Chem.* **279**, 8701–8707 (2004).
88. Talasz, H., Helliger, W., Puschendorf, B. & Lindner, H. In vivo phosphorylation of histone H1 variants during the cell cycle. *Biochemistry (Mosc.)* **35**, 1761–1767 (1996).
89. Bradbury, E. M. Reversible histone modifications and the chromosome cell cycle. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **14**, 9–16 (1992).
90. Lopez, R. *et al.* Linker histone partial phosphorylation: effects on secondary structure and chromatin condensation. *Nucleic Acids Res.* **43**, 4463–4476 (2015).
91. Roque, A., Ponte, I. & Suau, P. Role of charge neutralization in the folding of the carboxy-terminal domain of histone H1. *J. Phys. Chem. B* **113**, 12061–12066 (2009).
92. Ishiyama, N. *et al.* Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. *Cell* **141**, 117–128 (2010).
93. Li, J. *et al.* The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell* **150**, 339–350 (2012).
94. Chi, Y.-H., Chen, Z.-J. & Jeang, K.-T. The nuclear envelopathies and human diseases. *J. Biomed. Sci.* **16**, 96 (2009).

95. Schirmer, E. C., Florens, L., Guan, T., Yates, J. R. & Gerace, L. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301**, 1380–1382 (2003).
96. Wilkie, G. S. *et al.* Several novel nuclear envelope transmembrane proteins identified in skeletal muscle have cytoskeletal associations. *Mol. Cell. Proteomics MCP* **10**, M110.003129 (2011).
97. Korfali, N. *et al.* The leukocyte nuclear envelope proteome varies with cell activation and contains novel transmembrane proteins that affect genome architecture. *Mol. Cell. Proteomics MCP* **9**, 2571–2585 (2010).
98. Manilal, S., Nguyen, T. M., Sewry, C. A. & Morris, G. E. The Emery-Dreifuss muscular dystrophy protein, emerin, is a nuclear membrane protein. *Hum. Mol. Genet.* **5**, 801–808 (1996).
99. Huber, M. D., Guan, T. & Gerace, L. Overlapping functions of nuclear envelope proteins NET25 (Lem2) and emerin in regulation of extracellular signal-regulated kinase signaling in myoblast differentiation. *Mol. Cell. Biol.* **29**, 5718–5728 (2009).
100. Bione, S. *et al.* Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat. Genet.* **8**, 323–327 (1994).
101. Lin, F. *et al.* MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerin. *J. Biol. Chem.* **275**, 4840–4847 (2000).
102. Wolff, N. *et al.* Structural analysis of emerin, an inner nuclear membrane protein mutated in X-linked Emery-Dreifuss muscular dystrophy. *FEBS Lett.* **501**, 171–176 (2001).
103. Cai, M. *et al.* Solution NMR structure of the barrier-to-autointegration factor-Emerin complex. *J. Biol. Chem.* **282**, 14525–14535 (2007).
104. Guilluy, C. *et al.* Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nat. Cell Biol.* **16**, 376–381 (2014).
105. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469–480 (2006).
106. Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).

107. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
108. Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).
109. Markiewicz, E. *et al.* The inner nuclear membrane protein emerin regulates beta-catenin activity by restricting its accumulation in the nucleus. *EMBO J.* **25**, 3275–3285 (2006).
110. Berk, J. M. *et al.* The molecular basis of emerin-emerin and emerin-BAF interactions. *J. Cell Sci.* **127**, 3956–3969 (2014).
111. Herrada, I. *et al.* Muscular Dystrophy Mutations Impair the Nuclear Envelope Emerin Self-assembly Properties. *ACS Chem. Biol.* **10**, 2733–2742 (2015).
112. Haque, F. *et al.* Mammalian SUN protein interaction networks at the inner nuclear membrane and their role in laminopathy disease processes. *J. Biol. Chem.* **285**, 3487–3498 (2010).
113. Sakaki, M. *et al.* Interaction between emerin and nuclear lamins. *J. Biochem. (Tokyo)* **129**, 321–327 (2001).
114. Bengtsson, L. & Wilson, K. L. Barrier-to-autointegration factor phosphorylation on Ser-4 regulates emerin binding to lamin A in vitro and emerin localization in vivo. *Mol. Biol. Cell* **17**, 1154–1163 (2006).
115. Lombardi, M. L. *et al.* The interaction between nesprins and sun proteins at the nuclear envelope is critical for force transmission between the nucleus and cytoskeleton. *J. Biol. Chem.* **286**, 26743–26753 (2011).
116. Tapley, E. C. & Starr, D. A. Connecting the nucleus to the cytoskeleton by SUN-KASH bridges across the nuclear envelope. *Curr. Opin. Cell Biol.* **25**, 57–62 (2013).
117. Sosa, B. A., Rothballer, A., Kutay, U. & Schwartz, T. U. LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins. *Cell* **149**, 1035–1047 (2012).
118. Zhou, Z. *et al.* Structure of Sad1-UNC84 homology (SUN) domain defines features of molecular bridge in nuclear envelope. *J. Biol. Chem.* **287**, 5317–5326 (2012).

119. Lu, W. *et al.* Sun1 forms immobile macromolecular assemblies at the nuclear envelope. *Biochim. Biophys. Acta* **1783**, 2415–2426 (2008).
120. Jahed, Z. *et al.* Molecular Insights into the Mechanisms of SUN1 Oligomerization in the Nuclear Envelope. *Biophys. J.* **114**, 1190–1203 (2018).
121. Zhou, Z. *et al.* Structure of Sad1-UNC84 homology (SUN) domain defines features of molecular bridge in nuclear envelope. *J. Biol. Chem.* **287**, 5317–5326 (2012).
122. Wang, W. *et al.* Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Res.* **22**, 1440–1452 (2012).
123. Hennen, J., Saunders, C. A., Mueller, J. D. & Luxton, G. W. G. Fluorescence fluctuation spectroscopy reveals differential SUN protein oligomerization in living cells. *Mol. Biol. Cell* **29**, 1003–1011 (2018).
124. Haque, F. *et al.* SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide a physical connection between the nuclear lamina and the cytoskeleton. *Mol. Cell. Biol.* **26**, 3738–3751 (2006).
125. Melcer, S., Gruenbaum, Y. & Krohne, G. Invertebrate lamins. *Exp. Cell Res.* **313**, 2157–2166 (2007).
126. Turgay, Y. *et al.* The molecular architecture of lamins in somatic cells. *Nature* **543**, 261–264 (2017).
127. Stick, R., Angres, B., Lehner, C. F. & Nigg, E. A. The fates of chicken nuclear lamin proteins during mitosis: evidence for a reversible redistribution of lamin B2 between inner nuclear membrane and elements of the endoplasmic reticulum. *J. Cell Biol.* **107**, 397–406 (1988).
128. Shimi, T. *et al.* The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription. *Genes Dev.* **22**, 3409–3421 (2008).
129. Gerace, L. & Blobel, G. The nuclear envelope lamina is reversibly depolymerized during mitosis. *Cell* **19**, 277–287 (1980).

130. Quigley, A. *et al.* The structural basis of ZMPSTE24-dependent laminopathies. *Science* **339**, 1604–1607 (2013).
131. Kapinos, L. E. *et al.* Characterization of the head-to-tail overlap complexes formed by human lamin A, B1 and B2 ‘half-minilamin’ dimers. *J. Mol. Biol.* **396**, 719–731 (2010).
132. Strelkov, S. V., Kreplak, L., Herrmann, H. & Aebi, U. Intermediate filament protein structure determination. *Methods Cell Biol.* **78**, 25–43 (2004).
133. Krimm, I. *et al.* The Ig-like structure of the C-terminal domain of lamin A/C, mutated in muscular dystrophies, cardiomyopathy, and partial lipodystrophy. *Struct. Lond. Engl.* **1993** **10**, 811–823 (2002).
134. Dhe-Paganon, S., Werner, E. D., Chi, Y.-I. & Shoelson, S. E. Structure of the globular tail of nuclear lamin. *J. Biol. Chem.* **277**, 17381–17384 (2002).
135. Muchir, A. *et al.* Activation of MAPK pathways links LMNA mutations to cardiomyopathy in Emery-Dreifuss muscular dystrophy. *J. Clin. Invest.* **117**, 1282–1293 (2007).
136. Muchir, A., Pavlidis, P., Bonne, G., Hayashi, Y. K. & Worman, H. J. Activation of MAPK in hearts of EMD null mice: similarities between mouse models of X-linked and autosomal dominant Emery Dreifuss muscular dystrophy. *Hum. Mol. Genet.* **16**, 1884–1895 (2007).
137. Emerson, L. J. *et al.* Defects in cell spreading and ERK1/2 activation in fibroblasts with lamin A/C mutations. *Biochim. Biophys. Acta* **1792**, 810–821 (2009).
138. González-Granado, J. M. *et al.* Nuclear envelope lamin-A couples actin dynamics with immunological synapse architecture and T cell activation. *Sci. Signal.* **7**, ra37 (2014).
139. Vlcek, S. & Foisner, R. A-type lamin networks in light of laminopathic diseases. *Biochim. Biophys. Acta* **1773**, 661–674 (2007).
140. Johnson, B. R. *et al.* A-type lamins regulate retinoblastoma protein function by promoting subnuclear localization and preventing proteasomal degradation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9677–9682 (2004).

141. Van Berlo, J. H. *et al.* A-type lamins are essential for TGF-beta1 induced PP2A to dephosphorylate transcription factors. *Hum. Mol. Genet.* **14**, 2839–2849 (2005).
142. Ivorra, C. *et al.* A mechanism of AP-1 suppression through interaction of c-Fos with lamin A/C. *Genes Dev.* **20**, 307–320 (2006).
143. Boguslavsky, R. L., Stewart, C. L. & Worman, H. J. Nuclear lamin A inhibits adipocyte differentiation: implications for Dunnigan-type familial partial lipodystrophy. *Hum. Mol. Genet.* **15**, 653–663 (2006).
144. Mattout, A., Goldberg, M., Tzur, Y., Margalit, A. & Gruenbaum, Y. Specific and conserved sequences in *D. melanogaster* and *C. elegans* lamins and histone H2A mediate the attachment of lamins to chromosomes. *J. Cell Sci.* **120**, 77–85 (2007).
145. Dechat, T., Adam, S. A. & Goldman, R. D. Nuclear lamins and chromatin: when structure meets function. *Adv. Enzyme Regul.* **49**, 157–166 (2009).
146. Crisp, M. *et al.* Coupling of the nucleus and cytoplasm: role of the LINC complex. *J. Cell Biol.* **172**, 41–53 (2006).
147. Mislow, J. M. K. *et al.* Nesprin-1alpha self-associates and binds directly to emerin and lamin A in vitro. *FEBS Lett.* **525**, 135–140 (2002).
148. Khatau, S. B. *et al.* A perinuclear actin cap regulates nuclear shape. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19017–19022 (2009).
149. Kim, D.-H., Chambliss, A. B. & Wirtz, D. The multi-faceted role of the actin cap in cellular mechanosensation and mechanotransduction. *Soft Matter* **9**, 5516–5523 (2013).
150. Kim, J.-K. *et al.* Nuclear lamin A/C harnesses the perinuclear apical actin cables to protect nuclear morphology. *Nat. Commun.* **8**, 2123 (2017).
151. Granzier, H. L. & Labeit, S. The giant protein titin: a major player in myocardial mechanics, signaling, and disease. *Circ. Res.* **94**, 284–295 (2004).
152. Machado, C. & Andrew, D. J. D-Titin: a giant protein with dual roles in chromosomes and muscles. *J. Cell Biol.* **151**, 639–652 (2000).

153. Zastrow, M. S., Flaherty, D. B., Benian, G. M. & Wilson, K. L. Nuclear titin interacts with A- and B-type lamins in vitro and in vivo. *J. Cell Sci.* **119**, 239–249 (2006).
154. Qi, J., Chi, L., Labeit, S. & Banes, A. J. Nuclear localization of the titin Z1Z2Zr domain and role in regulating cell proliferation. *Am. J. Physiol. Cell Physiol.* **295**, C975-985 (2008).
155. Torvaldson, E., Kochin, V. & Eriksson, J. E. Phosphorylation of lamins determine their structural properties and signaling functions. *Nucl. Austin Tex* **6**, 166–171 (2015).
156. Machowska, M., Piekarowicz, K. & Rzepecki, R. Regulation of lamin properties and functions: does phosphorylation do it all? *Open Biol.* **5**, (2015).
157. Haas, M. & Jost, E. Functional analysis of phosphorylation sites in human lamin A controlling lamin disassembly, nuclear transport and assembly. *Eur. J. Cell Biol.* **62**, 237–247 (1993).
158. Leukel, M. & Jost, E. Two conserved serines in the nuclear localization signal flanking region are involved in the nuclear targeting of human lamin A. *Eur. J. Cell Biol.* **68**, 133–142 (1995).
159. Kochin, V. *et al.* Interphase phosphorylation of lamin A. *J. Cell Sci.* **127**, 2683–2696 (2014).
160. Lloyd, D. J., Trembath, R. C. & Shackleton, S. A novel interaction between lamin A and SREBP1: implications for partial lipodystrophy and other laminopathies. *Hum. Mol. Genet.* **11**, 769–777 (2002).
161. Dreuillet, C., Tillit, J., Kress, M. & Ernoult-Lange, M. In vivo and in vitro interaction between human transcription factor MOK2 and nuclear lamin A/C. *Nucleic Acids Res.* **30**, 4634–4642 (2002).
162. Shumaker, D. K. *et al.* The highly conserved nuclear lamin Ig-fold binds to PCNA: its role in DNA replication. *J. Cell Biol.* **181**, 269–280 (2008).
163. Dorner, D. *et al.* Lamina-associated polypeptide 2alpha regulates cell cycle progression and differentiation via the retinoblastoma-E2F pathway. *J. Cell Biol.* **173**, 83–93 (2006).
164. Pekovic, V. *et al.* Nucleoplasmic LAP2alpha-lamin A complexes are required to maintain a proliferative state in human fibroblasts. *J. Cell Biol.* **176**, 163–172 (2007).

165. Hennekes, H. & Nigg, E. A. The role of isoprenylation in membrane attachment of nuclear lamins. A single point mutation prevents proteolytic cleavage of the lamin A precursor and confers membrane binding properties. *J. Cell Sci.* **107 (Pt 4)**, 1019–1029 (1994).
166. Izumi, M., Vaughan, O. A., Hutchison, C. J. & Gilbert, D. M. Head and/or CaaX domain deletions of lamin proteins disrupt preformed lamin A and C but not lamin B structure in mammalian cells. *Mol. Biol. Cell* **11**, 4323–4337 (2000).
167. Moiseeva, O., Lopes-Paciencia, S., Huot, G., Lessard, F. & Ferbeyre, G. Permanent farnesylation of lamin A mutants linked to progeria impairs its phosphorylation at serine 22 during interphase. *Aging* **8**, 366–381 (2016).
168. Dittmer, T. A. *et al.* Systematic identification of pathological lamin A interactors. *Mol. Biol. Cell* **25**, 1493–1510 (2014).
169. Chen, Z.-J. *et al.* Dysregulated interactions between lamin A and SUN1 induce abnormalities in the nuclear envelope and endoplasmic reticulum in progeric laminopathies. *J. Cell Sci.* **127**, 1792–1804 (2014).
170. Emery, A. E. & Dreifuss, F. E. Unusual type of benign x-linked muscular dystrophy. *J. Neurol. Neurosurg. Psychiatry* **29**, 338–342 (1966).
171. Bonne, G., Leturcq, F. & Ben Yaou, R. Emery-Dreifuss Muscular Dystrophy. in *GeneReviews*[®] (eds. Adam, M. P. *et al.*) (University of Washington, Seattle, 1993).
172. Méndez-López, I. & Worman, H. J. Inner nuclear membrane proteins: impact on human disease. *Chromosoma* **121**, 153–167 (2012).
173. Burke, B. & Stewart, C. L. The laminopathies: the functional architecture of the nucleus and its contribution to disease. *Annu. Rev. Genomics Hum. Genet.* **7**, 369–405 (2006).
174. Ross, S. E. *et al.* Inhibition of adipogenesis by Wnt signaling. *Science* **289**, 950–953 (2000).
175. Moldes, M. *et al.* Peroxisome-proliferator-activated receptor gamma suppresses Wnt/beta-catenin signalling during adipogenesis. *Biochem. J.* **376**, 607–613 (2003).

176. Tilgner, K., Wojciechowicz, K., Jahoda, C., Hutchison, C. & Markiewicz, E. Dynamic complexes of A-type lamins and emerin influence adipogenic capacity of the cell via nucleocytoplasmic distribution of beta-catenin. *J. Cell Sci.* **122**, 401–413 (2009).
177. Meinke, P. *et al.* Muscular dystrophy-associated SUN1 and SUN2 variants disrupt nuclear-cytoskeletal connections and myonuclear organization. *PLoS Genet.* **10**, e1004605 (2014).
178. Wang, J.-Y. *et al.* Sun1 deficiency leads to cerebellar ataxia in mice. *Dis. Model. Mech.* **8**, 957–967 (2015).
179. Chen, C.-Y. *et al.* Accumulation of the inner nuclear envelope protein Sun1 is pathogenic in progeric and dystrophic laminopathies. *Cell* **149**, 565–577 (2012).
180. Broers, J. L. V., Ramaekers, F. C. S., Bonne, G., Yaou, R. B. & Hutchison, C. J. Nuclear lamins: laminopathies and their role in premature ageing. *Physiol. Rev.* **86**, 967–1008 (2006).
181. Worman, H. J. & Bonne, G. ‘Laminopathies’: a wide spectrum of human diseases. *Exp. Cell Res.* **313**, 2121–2133 (2007).
182. Ackerman, J. & Gilbert-Barness, E. Hutchinson-Gilford progeria syndrome: a pathologic study. *Pediatr. Pathol. Mol. Med.* **21**, 1–13 (2002).
183. Pollex, R. L. & Hegele, R. A. Hutchinson-Gilford progeria syndrome. *Clin. Genet.* **66**, 375–381 (2004).
184. Hennekam, R. C. M. Hutchinson-Gilford progeria syndrome: review of the phenotype. *Am. J. Med. Genet. A.* **140**, 2603–2624 (2006).
185. Eriksson, M. *et al.* Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* **423**, 293–298 (2003).
186. De Sandre-Giovannoli, A. *et al.* Lamin a truncation in Hutchinson-Gilford progeria. *Science* **300**, 2055 (2003).
187. Dechat, T. *et al.* Alterations in mitosis and cell cycle progression caused by a mutant lamin A known to accelerate human aging. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4955–4960 (2007).

188. Scaffidi, P. & Misteli, T. Lamin A-dependent nuclear defects in human aging. *Science* **312**, 1059–1063 (2006).
189. McClintock, D. *et al.* The mutant form of lamin A that causes Hutchinson-Gilford progeria is a biomarker of cellular aging in human skin. *PLoS One* **2**, e1269 (2007).
190. Agarwal, A. K., Fryns, J.-P., Auchus, R. J. & Garg, A. Zinc metalloproteinase, ZMPSTE24, is mutated in mandibuloacral dysplasia. *Hum. Mol. Genet.* **12**, 1995–2001 (2003).
191. Navarro, C. L. *et al.* Loss of ZMPSTE24 (FACE-1) causes autosomal recessive restrictive dermopathy and accumulation of Lamin A precursors. *Hum. Mol. Genet.* **14**, 1503–1513 (2005).
192. Moulson, C. L. *et al.* Homozygous and compound heterozygous mutations in ZMPSTE24 cause the laminopathy restrictive dermopathy. *J. Invest. Dermatol.* **125**, 913–919 (2005).
193. Ragnauth, C. D. *et al.* Prelamin A acts to accelerate smooth muscle cell senescence and is a novel biomarker of human vascular aging. *Circulation* **121**, 2200–2210 (2010).
194. Kudlow, B. A., Stanfel, M. N., Burtner, C. R., Johnston, E. D. & Kennedy, B. K. Suppression of proliferative defects associated with processing-defective lamin A mutants by hTERT or inactivation of p53. *Mol. Biol. Cell* **19**, 5238–5248 (2008).
195. Candelario, J., Borrego, S., Reddy, S. & Comai, L. Accumulation of distinct prelamin A variants in human diploid fibroblasts differentially affects cell homeostasis. *Exp. Cell Res.* **317**, 319–329 (2011).
196. Davies, B. S. J. *et al.* An accumulation of non-farnesylated prelamin A causes cardiomyopathy but not progeria. *Hum. Mol. Genet.* **19**, 2682–2694 (2010).
197. Dahl, K. N. *et al.* Distinct structural and mechanical properties of the nuclear lamina in Hutchinson-Gilford progeria syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10271–10276 (2006).
198. Verstraeten, V. L. R. M., Ji, J. Y., Cummings, K. S., Lee, R. T. & Lammerding, J. Increased mechanosensitivity and nuclear stiffness in Hutchinson-Gilford progeria cells: effects of farnesyltransferase inhibitors. *Aging Cell* **7**, 383–393 (2008).

199. Goldman, R. D. *et al.* Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8963–8968 (2004).
200. Shumaker, D. K. *et al.* Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8703–8708 (2006).
201. Decker, M. L., Chavez, E., Vulto, I. & Lansdorp, P. M. Telomere length in Hutchinson-Gilford progeria syndrome. *Mech. Ageing Dev.* **130**, 377–383 (2009).
202. Huang, S., Risques, R. A., Martin, G. M., Rabinovitch, P. S. & Oshima, J. Accelerated telomere shortening and replicative senescence in human fibroblasts overexpressing mutant and wild-type lamin A. *Exp. Cell Res.* **314**, 82–91 (2008).
203. Eisch, V., Lu, X., Gabriel, D. & Djabali, K. Progerin impairs chromosome maintenance by depleting CENP-F from metaphase kinetochores in Hutchinson-Gilford progeria fibroblasts. *Oncotarget* **7**, 24700–24718 (2016).
204. Csoka, A. B. *et al.* Genome-scale expression profiling of Hutchinson-Gilford progeria syndrome reveals widespread transcriptional misregulation leading to mesodermal/mesenchymal defects and accelerated atherosclerosis. *Aging Cell* **3**, 235–243 (2004).
205. Park, W. Y. *et al.* Gene profile of replicative senescence is different from progeria or elderly donor. *Biochem. Biophys. Res. Commun.* **282**, 934–939 (2001).
206. Scaffidi, P. & Misteli, T. Lamin A-dependent misregulation of adult stem cells associated with accelerated ageing. *Nat. Cell Biol.* **10**, 452–459 (2008).
207. Mattioli, E. *et al.* Drugs affecting prelamin A processing: effects on heterochromatin organization. *Exp. Cell Res.* **314**, 453–462 (2008).
208. Columbaro, M. *et al.* Rescue of heterochromatin organization in Hutchinson-Gilford progeria by drug treatment. *Cell. Mol. Life Sci. CMLS* **62**, 2669–2678 (2005).

209. Krishnan, V. *et al.* Histone H4 lysine 16 hypoacetylation is associated with defective DNA repair and premature senescence in Zmpste24-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12325–12330 (2011).
210. Pegoraro, G. *et al.* Ageing-related chromatin defects through loss of the NURD complex. *Nat. Cell Biol.* **11**, 1261–1267 (2009).
211. Wood, A. M. *et al.* TRF2 and lamin A/C interact to facilitate the functional organization of chromosome ends. *Nat. Commun.* **5**, 5467 (2014).
212. Smith, E. D., Garza-Gongora, A. G., MacQuarrie, K. L. & Kosak, S. T. Interstitial telomeric loops and implications of the interaction between TRF2 and lamin A/C. *Differ. Res. Biol. Divers.* **102**, 19–26 (2018).
213. Chen, L. *et al.* LMNA mutations in atypical Werner's syndrome. *Lancet Lond. Engl.* **362**, 440–445 (2003).
214. Soria-Valles, C. *et al.* Novel LMNA mutations cause an aggressive atypical neonatal progeria without progerin accumulation. *J. Med. Genet.* **53**, 776–785 (2016).
215. Caux, F. *et al.* A new clinical condition linked to a novel mutation in lamins A and C with generalized lipoatrophy, insulin-resistant diabetes, disseminated leukomelanodermic papules, liver steatosis, and cardiomyopathy. *J. Clin. Endocrinol. Metab.* **88**, 1006–1013 (2003).
216. Capanni, C. *et al.* Altered pre-lamin A processing is a common mechanism leading to lipodystrophy. *Hum. Mol. Genet.* **14**, 1489–1502 (2005).
217. Eriksson, M. *et al.* Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* **423**, 293–298 (2003).
218. Barthélémy, F. *et al.* Truncated prelamin A expression in HGPS-like patients: a transcriptional study. *Eur. J. Hum. Genet. EJHG* **23**, 1051–1061 (2015).
219. Madej-Pilarczyk, A. *et al.* Progeroid syndrome with scleroderma-like skin changes associated with homozygous R435C LMNA mutation. *Am. J. Med. Genet. A.* **149A**, 2387–2392 (2009).

220. Zirn, B. *et al.* Association of homozygous LMNA mutation R471C with new phenotype: mandibuloacral dysplasia, progeria, and rigid spine muscular dystrophy. *Am. J. Med. Genet. A.* **146A**, 1049–1054 (2008).
221. Agarwal, A. K., Kazachkova, I., Ten, S. & Garg, A. Severe mandibuloacral dysplasia-associated lipodystrophy and progeria in a young girl with a novel homozygous Arg527Cys LMNA mutation. *J. Clin. Endocrinol. Metab.* **93**, 4617–4623 (2008).
222. Novelli, G. *et al.* Mandibuloacral dysplasia is caused by a mutation in LMNA-encoding lamin A/C. *Am. J. Hum. Genet.* **71**, 426–431 (2002).
223. Simha, V., Agarwal, A. K., Oral, E. A., Fryns, J.-P. & Garg, A. Genetic and phenotypic heterogeneity in patients with mandibuloacral dysplasia-associated lipodystrophy. *J. Clin. Endocrinol. Metab.* **88**, 2821–2824 (2003).
224. Shen, J. J., Brown, C. A., Lupski, J. R. & Potocki, L. Mandibuloacral dysplasia caused by homozygosity for the R527H mutation in lamin A/C. *J. Med. Genet.* **40**, 854–857 (2003).
225. Garavelli, L. *et al.* Mandibuloacral dysplasia type A in childhood. *Am. J. Med. Genet. A.* **149A**, 2258–2264 (2009).
226. Kosho, T. *et al.* Mandibuloacral dysplasia and a novel LMNA mutation in a woman with severe progressive skeletal changes. *Am. J. Med. Genet. A.* **143A**, 2598–2603 (2007).
227. Garg, A., Cogulu, O., Ozkinay, F., Onay, H. & Agarwal, A. K. A novel homozygous Ala529Val LMNA mutation in Turkish patients with mandibuloacral dysplasia. *J. Clin. Endocrinol. Metab.* **90**, 5259–5264 (2005).
228. Plasilova, M. *et al.* Homozygous missense mutation in the lamin A/C gene causes autosomal recessive Hutchinson-Gilford progeria syndrome. *J. Med. Genet.* **41**, 609–614 (2004).
229. Samson, C. *et al.* Structural analysis of the ternary complex between lamin A/C, BAF and emerin identifies an interface disrupted in autosomal recessive progeroid diseases. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky736

230. Finlan, L. E. & Bickmore, W. A. Porin new light onto chromatin and nuclear organization. *Genome Biol.* **9**, 222 (2008).
231. Mekhail, K. & Moazed, D. The nuclear envelope in genome organization, expression and stability. *Nat. Rev. Mol. Cell Biol.* **11**, 317–328 (2010).
232. Segura-Totten, M., Kowalski, A. K., Craigie, R. & Wilson, K. L. Barrier-to-autointegration factor: major roles in chromatin decondensation and nuclear assembly. *J. Cell Biol.* **158**, 475–485 (2002).
233. Lee, M. S. & Craigie, R. A previously unidentified host protein protects retroviral DNA from autointegration. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1528–1533 (1998).
234. Zheng, R. *et al.* Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8997–9002 (2000).
235. Cai, M. *et al.* Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. *Nat. Struct. Biol.* **5**, 903–909 (1998).
236. Umland, T. C., Wei, S. Q., Craigie, R. & Davies, D. R. Structural basis of DNA bridging by barrier-to-autointegration factor. *Biochemistry (Mosc.)* **39**, 9130–9138 (2000).
237. Bradley, C. M., Ronning, D. R., Ghirlando, R., Craigie, R. & Dyda, F. Structural basis for DNA bridging by barrier-to-autointegration factor. *Nat. Struct. Mol. Biol.* **12**, 935–936 (2005).
238. Samwer, M. *et al.* DNA Cross-Bridging Shapes a Single Nucleus from a Set of Mitotic Chromosomes. *Cell* **170**, 956–972.e23 (2017).
239. Montes de Oca, R., Shoemaker, C. J., Gucek, M., Cole, R. N. & Wilson, K. L. Barrier-to-autointegration factor proteome reveals chromatin-regulatory partners. *PLoS One* **4**, e7050 (2009).
240. Haraguchi, T. *et al.* BAF is required for emerin assembly into the reforming nuclear envelope. *J. Cell Sci.* **114**, 4575–4585 (2001).
241. Shimi, T. *et al.* Dynamic interaction between BAF and emerin revealed by FRAP, FLIP, and FRET analyses in living HeLa cells. *J. Struct. Biol.* **147**, 31–41 (2004).

242. Capanni, C. *et al.* Lamin A precursor induces barrier-to-autointegration factor nuclear localization. *Cell Cycle Georget. Tex* **9**, 2600–2610 (2010).
243. Montes de Oca, R., Lee, K. K. & Wilson, K. L. Binding of barrier to autointegration factor (BAF) to histone H3 and selected linker histones including H1.1. *J. Biol. Chem.* **280**, 42252–42262 (2005).
244. Montes de Oca, R., Andreassen, P. R. & Wilson, K. L. Barrier-to-Autointegration Factor influences specific histone modifications. *Nucl. Austin Tex* **2**, 580–590 (2011).
245. Zhang, Y. *et al.* Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13**, 1924–1935 (1999).
246. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).
247. Clements, L., Manilal, S., Love, D. R. & Morris, G. E. Direct interaction between emerin and lamin A. *Biochem. Biophys. Res. Commun.* **267**, 709–714 (2000).
248. Berk, J. M., Tifft, K. E. & Wilson, K. L. The nuclear envelope LEM-domain protein emerin. *Nucl. Austin Tex* **4**, 298–314 (2013).
249. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
250. Samson, C., Herrada, I., Celli, F., Theillet, F.-X. & Zinn-Justin, S. 1H, 13C and 15N backbone resonance assignment of the intrinsically disordered region of the nuclear envelope protein emerin. *Biomol. NMR Assign.* **10**, 179–182 (2016).
251. Samson, C. *et al.* Emerin self-assembly mechanism: role of the LEM domain. *FEBS J.* **284**, 338–352 (2017).
252. Rocco, M. *et al.* A comprehensive mechanism of fibrin network formation involving early branching and delayed single- to double-strand transition from coupled time-resolved X-ray/light-scattering detection. *J. Am. Chem. Soc.* **136**, 5376–5384 (2014).

253. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* **33**, 199–211 (2005).
254. Celli, F., Petitalot, A., Samson, C., Theillet, F.-X. & Zinn-Justin, S. ¹H, ¹³C and ¹⁵N backbone resonance assignment of the lamin C-terminal region specific to prelamin A. *Biomol. NMR Assign.* **12**, 225–229 (2018).
255. Iyer, A., Koch, A. J. & Holaska, J. M. Expression Profiling of Differentiating Emerin-Null Myogenic Progenitor Identifies Molecular Pathways Implicated in Their Impaired Differentiation. *Cells* **6**, (2017).
256. Speckman, R. A. *et al.* Mutational and haplotype analyses of families with familial partial lipodystrophy (Dunnigan variety) reveal recurrent missense mutations in the globular C-terminal domain of lamin A/C. *Am. J. Hum. Genet.* **66**, 1192–1198 (2000).
257. Lee, S.-J. *et al.* Interruption of progerin-lamin A/C binding ameliorates Hutchinson-Gilford progeria syndrome phenotype. *J. Clin. Invest.* **126**, 3879–3893 (2016).
258. Yang, S. H. *et al.* Blocking protein farnesyltransferase improves nuclear blebbing in mouse fibroblasts with a targeted Hutchinson-Gilford progeria syndrome mutation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10291–10296 (2005).
259. Wang, Y. *et al.* Blocking farnesylation of the prelamin A variant in Hutchinson-Gilford progeria syndrome alters the distribution of A-type lamins. *Nucl. Austin Tex* **3**, 452–462 (2012).
260. Hübner, S., Eam, J. E., Hübner, A. & Jans, D. A. Laminopathy-inducing lamin A mutants can induce redistribution of lamin binding proteins into nuclear aggregates. *Exp. Cell Res.* **312**, 171–183 (2006).
261. Adhikari, A. S., Sridhar Rao, K., Rangaraj, N., Parnaik, V. K. & Mohan Rao, C. Heat stress-induced localization of small heat shock proteins in mouse myoblasts: intranuclear lamin A/C speckles as target for alphaB-crystallin and Hsp25. *Exp. Cell Res.* **299**, 393–403 (2004).

262. Ayyadevara, S. *et al.* Proteins that mediate protein aggregation and cytotoxicity distinguish Alzheimer's hippocampus from normal controls. *Aging Cell* **15**, 924–939 (2016).

VI. Annexes

¹H, ¹³C and ¹⁵N backbone resonance assignment of the intrinsically disordered region of the nuclear envelope protein emerin

Camille Samson¹ · Isaline Herrada¹ · Florian Celli¹ · Francois-Xavier Theillet^{1,2} · Sophie Zinn-Justin¹ 

Received: 13 September 2015 / Accepted: 28 December 2015
© Springer Science+Business Media Dordrecht 2016

Abstract Human emerin is an inner nuclear membrane protein involved in the response of the nucleus to mechanical stress. It contributes to the physical connection between the cytoskeleton and the nucleoskeleton. It is also involved in chromatin organization. Its N-terminal region is nucleoplasmic and comprises a globular LEM domain from residue 1 to residue 43. The three-dimensional structure of this LEM domain in complex with the chromatin BAF protein was solved from NMR data. Apart from the LEM domain, the nucleoplasmic region of emerin, from residue 44 to residue 221, is predicted to be intrinsically disordered. Mutations in this region impair binding to several emerin partners as lamin A, actin or HDAC3. However the molecular details of these recognition defects are unknown. Here we report ¹H, ¹⁵N, ¹³CO, ¹³C α and ¹³C β NMR chemical shift assignments of the emerin fragment from residue 67 to residue 170, which is sufficient for nuclear localization and involved in lamin A binding. Chemical shift analysis confirms that this fragment is intrinsically disordered in 0 and 8 M urea.

Keywords Nuclear envelope · Emerin · Intrinsically disordered protein · Urea · NMR spectroscopy · Muscular dystrophy

Biological context

Inner nuclear envelope proteins play an essential role in cell response to mechanical demands. They regulate cell cytoskeletal tension as a function of tissue stiffness and activate mechano-regulated signaling pathways (Ho et al. 2013; Bertrand et al. 2014; Swift and Discher 2014). In particular emerin, SUN1/2 and lamin A contribute to nuclear envelope assembly as well as to the nuclear response to a mechanical stress. However these proteins contain large regions predicted to be intrinsically disordered regions (IDRs), which hinders their study by classical structural biology tools (Herrada et al. 2015, in press). There is a need today to describe the molecular events involved in nuclear envelope assembly and function, and to understand how IDRs contribute to these events.

Human emerin is an inner nuclear membrane protein of 254 amino acids. It is anchored to the membrane through its C-terminal α -helix. Its N-terminal region is nucleoplasmic. It is mutated in patients with X-linked Emery–Dreifuss Muscular dystrophy (Bione et al. 1994). Disease-causing mutations generally lead to an absence of emerin. A small number of missense mutations found in the emerin predicted IDR lead to loss-of-function variants (Ellis et al. 1998). Emerin contributes to the physical connection between the cytoskeleton and the nucleoskeleton. In particular it binds to the inner nuclear membrane proteins SUN1/2 (Haque et al. 2006) and the nuclear filaments lamins (Clements et al. 2000). It is also involved in chromatin organization. It is present at the interface between

Camille Samson, Isaline Herrada have contributed equally to this work.

✉ Sophie Zinn-Justin
sophie.zinn@cea.fr

¹ Laboratoire de Biologie Structurale et Radiobiologie, Institute for Integrative Biology of the Cell (I2BC), CNRS, Univ. Paris South and IBITECS CEA, CEA Saclay Bât. 144, 91191 Gif-sur-Yvette Cedex, France

² Department of NMR-assisted Structural Biology, Leibniz-Institut für Molekular Pharmakologie (FMP), 13125 Berlin, Germany

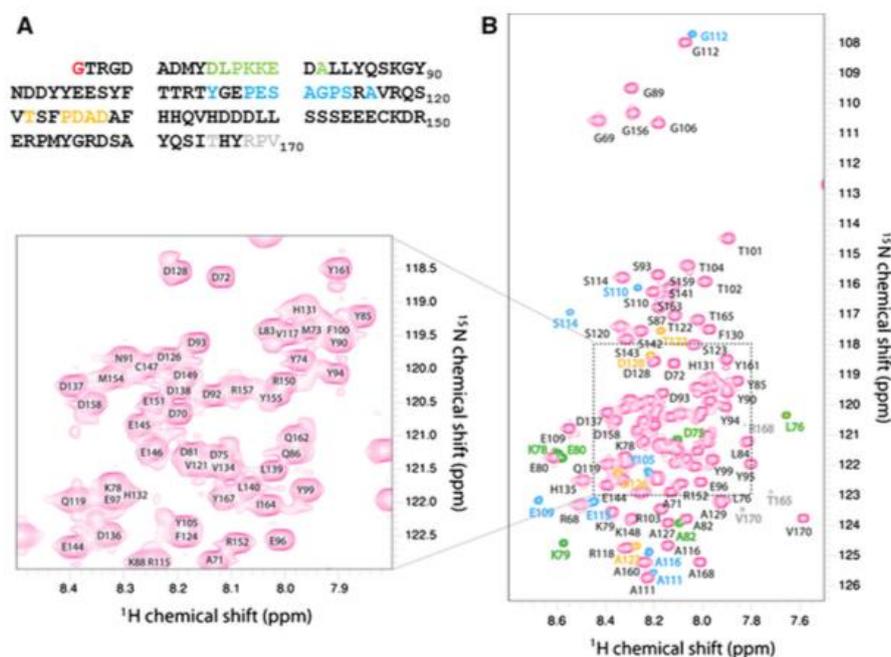


Fig. 1 **a** Amino acid sequence of the emerlin fragment produced for this study, comprising an N-terminal glycine residue (shown in red) that results from the TEV-cleavage of the purification tag, and the emerlin residues between 67 and 170. The colored residues are observed as two signals on the ^1H - ^{15}N HSQC spectrum, the less intense signal corresponding to a proline *cis*-conformation. They are represented in *green* (residues around P77), *blue* (P108 and P113),

yellow (P125) and *light grey* (P169). **b** 2D ^1H - ^{15}N HSQC spectrum of the emerlin fragment in 0 M urea at pH 6.5 and 303 K. Assigned backbone amide resonances are labelled. The spectrum was acquired on a 750 MHz spectrometer equipped with a TCI cryoprobe. An enlarged view of the most crowded region of the spectrum is shown in the *down-left corner*

the nuclear envelope and chromatin, and it directly interacts with the DNA-binding protein BAF. The three-dimensional structures of emerlin N-terminal LEM domain alone (residues 1–43; Wolff et al. 2001) and in complex with BAF (Cai et al. 2007) were solved from NMR data. Analysis of the LEM–BAF structure revealed an asymmetric mode of binding in which one LEM domain binds to two BAF molecules. Apart from this LEM domain, the nucleoplasmic region of emerlin from residue 44 to residue 221 is predicted as intrinsically disordered. Mutations in this region hinder binding to a large number of emerlin partners: lamin A, actin, MAN1, HDAC3, GCL, Btf, YT521-B and Lmo7 (Berk et al. 2013). However the molecular details of these recognition events are unknown. In order to understand the molecular basis of emerlin binding properties and disease-causing defects, we undertook the NMR study of emerlin nucleoplasmic region from residue 67 to residue 170, which is sufficient for nuclear localization and involved in lamin A binding, in both 0 and 8 M urea. We report here the NMR resonance assignments in these two conditions.

Methods and experiments

Protein expression and purification

The human emerlin fragment from residue 67 to residue 170 was expressed using a pETM13 vector as a N-terminal octahistidine fusion protein with a TEV site between the histidine tag and the emerlin fragment in *Escherichia coli* BL21 DE3 Star (Novagen). The emerlin fragment cDNA was optimized for expression in *Escherichia coli* (Genscript). Bacteria were grown in ^{15}N and ^{13}C labelled M9 minimum medium and induced at an optical density of 1 with 0.5 mM isopropyl β -D-1-thiogalactopyranoside at 20 °C for 18 h. Cells were lysed in 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 40 mM imidazole, 5 % glycerol, 1 % Triton X-100 and 1 mM phenylmethanesulfonyl fluoride. As the emerlin fragment is produced in inclusion bodies, the pellet was resuspended in buffer A (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole, 10 mM β -mercaptoethanol and 8 M urea) and centrifuged at 20,000g for 20 min. The protein was then purified on a Ni-NTA column (GE-Healthcare) equilibrated

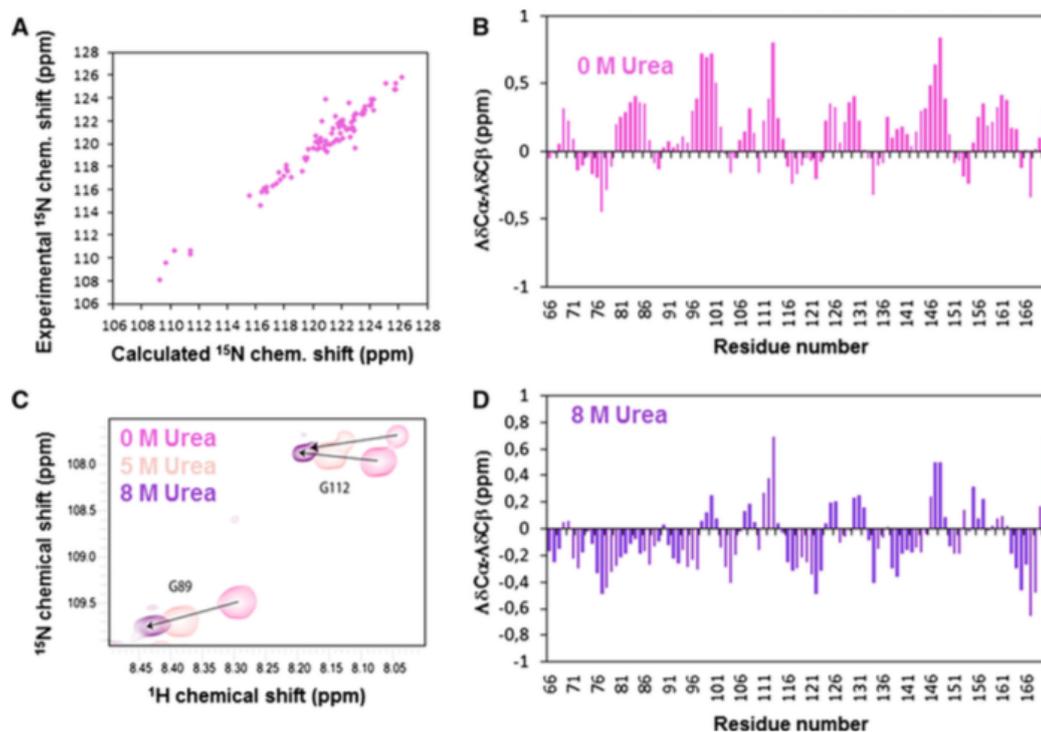


Fig. 2 **a** Experimental ^{15}N chemical shift values measured in 0 M urea plotted against the calculated ^{15}N chemical shift values predicted for a random coil conformation by ncIDP (<http://nmr.chem.rug.nl/ncIDP/>). **b** Plot of the difference between the secondary ^{13}C chemical shifts of the $\text{C}\alpha$ (noted as $\Delta\delta\text{C}\alpha$) and the $\text{C}\beta$ (noted as $\Delta\delta\text{C}\beta$) nuclei. Secondary chemical shifts $\Delta\delta\text{C}\alpha_i$ and $\Delta\delta\text{C}\beta_i$ were obtained by subtracting the random coil chemical shifts (predicted by ncIDP) from the observed chemical shifts. Differences between the secondary chemical shift deviations

$\Delta\delta\text{C}\alpha - \Delta\delta\text{C}\beta$ were plotted against the amino acid residue numbers, taking into account next neighbor effects. Thus $\Delta\delta\text{C}\alpha - \Delta\delta\text{C}\beta$ was calculated as: $\Delta\delta\text{C}\alpha - \Delta\delta\text{C}\beta = (\Delta\delta\text{C}\alpha_{i-1} + \Delta\delta\text{C}\alpha_i + \Delta\delta\text{C}\alpha_{i+1}) - (\Delta\delta\text{C}\beta_{i-1} + \Delta\delta\text{C}\beta_i + \Delta\delta\text{C}\beta_{i+1})$ **c** Superposition of the 2D ^1H - ^{15}N HSQC spectra of the emerlin fragment recorded at pH 6.5 and 303 K in buffers containing 0 (magenta), 5 (beige) and 8 M urea (purple). Zoom on 2 glycine residues. **d** Same as (b) but deduced from the NMR analysis of the emerlin fragment in 8 M urea

in buffer A and eluted directly with 100 % of buffer B (buffer A + 1 M imidazole). After dialyzing the sample to remove imidazole and urea, the octa-histidine tag was cleaved using site-specific TEV proteolysis and the purity of the sample was checked on SDS-PAGE. The sample was finally dialyzed against either buffer D (20 mM sodium phosphate buffer (pH 6.5), 30 mM NaCl and 8 M urea, "Denatured sample" or DS), buffer DN (20 mM sodium phosphate buffer (pH 6.5), 30 mM NaCl and 5 M urea) or buffer N (20 mM sodium phosphate buffer (pH 6.5), 30 mM NaCl and 0 M urea, "Native sample" or NS). All samples were concentrated up to 500–600 μM for the NMR assignment experiments.

NMR spectroscopy

All NMR experiments were performed on a 5 mm diameter NMR sample tube containing 500–600 μM uniformly ^{15}N and ^{13}C labelled emerlin fragment in 20 mM sodium phosphate buffer (pH 6.5), 30 mM NaCl, 10 mM β -

mercaptoethanol or 5 mM dithiothreitol, 90 %:10 %, $\text{H}_2\text{O}:\text{D}_2\text{O}$ and either 0, 5 or 8 M urea. 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was added for chemical shift referencing. Experiments were recorded at 303 K on a 750 MHz Bruker Avance spectrometer equipped with a triple resonance cryogenic probe. The data were processed using Topspin3.1 (Bruker) and analyzed with CCPNMR (Vranken et al. 2005). ^1H , ^{13}C and ^{15}N resonance frequencies were assigned using 3D HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO and HN(CO)(CA)NH experiments.

Extent of assignments and data deposition

The sequence of the studied protein fragment and its ^1H - ^{15}N 2D HSQC spectrum in 0 M urea are shown in Fig. 1. Using a combination of 3D experiments recorded on the NS sample, 96 % of ^1H - ^{15}N pairs (95 out of 99), 99 % of $^{13}\text{C}\alpha$ (104 out of 105), 99 % of $^{13}\text{C}\beta$ (98 out of 99) and 98 % of ^{13}CO (103 out of 105) resonances were assigned for the emerlin fragment in 0 M urea. Almost all the backbone ^1H , ^{13}C , and ^{15}N

resonances were completely assigned except for residues G66, T67, Q133 and H166, probably because of broadening of their spectral signatures. The expressed fragment (sequence shown in Fig. 1a) contains 6 glycine residues: the N-terminal G66 remaining after TEV cleavage (that is not present in the original emerin sequence), G69, G89, G106, G112 and G156. The ^1H - ^{15}N HSQC signals of the 5 emerin glycines are clearly visible (Fig. 1b). The expressed fragment also contains 6 proline residues: P77, P108, P113, P125, P153 and P169. The main conformations of these prolines are *trans*, as deduced from their C β chemical shift values, which are comprised between 31.9 and 32.2 ppm. However the 5 prolines P77, P108, P113, P123 and P169 give rise to supplementary sets of weak signals corresponding to minor *cis* conformations characterized by proline C β chemical shift values comprised between 32.9 and 34.8 ppm (Shen and Bax 2010). In the 2D ^1H - ^{15}N HSQC spectrum, the signals due to proline *cis* conformations and the resulting different chemical environments in their vicinity are colored as a function of the proline in *cis* (Fig. 1b). The limited chemical shift dispersion of the ^1H - ^{15}N HSQC signals, particularly in the ^1H dimension (between 7.6 and 8.7 ppm), indicates that the emerin fragment is not folded in the NMR experimental conditions (20 mM phosphate, 30 mM NaCl, pH 6.5). Comparison of experimental and calculated backbone ^{15}N chemical shifts confirms that the emerin chemical shifts correspond to random coil (Fig. 2a; Tamiola et al. 2010). Consistently, the difference between the secondary chemical shifts of C α and C β yields small values, generally under 1 ppm (Fig. 2b). The largest difference amplitudes are observed for residues 98–101, 113 and 147–148 (differences between the secondary chemical shifts of C α and C β in the 0.5–1 ppm range). Residue 113 is a proline and 147 a cysteine, their chemical shifts could be influenced by their *cis/trans* conformation and oxidation states, respectively. However region 98–101 corresponds to a conserved sequence rich in hydrophobic residues and deletion of the overlapping fragment 95–99 causes Emery-Dreifuss Muscular dystrophy. This small region might fold into a partially populated α -helix in our NS conditions. In the case of the 8 M urea-denatured emerin fragment (sample DS), chemical shifts of 97 % ^1H - ^{15}N pairs (96 out of 99 non-proline residues, the 3 missing residues being D70, Q133 and H166), 100 % $^{13}\text{C}\alpha$ (105 out of 105) and 98 % $^{13}\text{C}\beta$ (97 out of 99) were assigned (Fig. 2c, d). Here again, analysis of the differences between the C α and C β secondary chemical shifts confirms that the emerin fragment is disordered in 8 M urea (Fig. 2d). In this condition no difference value larger than 0.5 ppm is observed, except at residue 167, next to histidine 166; a small variation in pH could explain this difference.

The chemical shift data have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>) under accession

numbers 26654 for the emerin fragment in 0 M urea and 26655 for the same fragment in 8 M urea.

Acknowledgments We thank the French Association against Myopathies (AFM) (research Grant No. 17243 to S.Z.J. and fellowship No. 18159 to C.S.) and the Foundation for Medical Research (FRM) (Grant. FDT20140931008 to I.H.) for providing financial support to this project. We also thank the Leibniz-Institut für Molekular Pharmakologie (FMP) for access to the NMR facility.

References

- Berk JM, Tiftt KE, Wilson KL (2013) The nuclear envelope LEM-domain protein emerin. *Nucleus* 4:298–314. doi:10.4161/nucl.25751
- Bertrand AT et al (2014) Cellular microenvironments reveal defective mechanosensing responses and elevated YAP signaling in LMNA-mutated muscle precursors. *J Cell Sci* 127:2873–2884. doi:10.1242/jcs.144907
- Bione S, Maestrini E, Rivella S, Mancini M, Regis S, Romeo G, Toniolo D (1994) Identification of a novel X-linked gene responsible for Emery–Dreifuss muscular dystrophy. *Nat Genet* 8:323–327. doi:10.1038/ng1294-323
- Cai M, Huang Y, Suh JY, Louis JM, Ghirlando R, Craigie R, Clow GM (2007) Solution NMR structure of the barrier-to-autointegration factor–Emerin complex. *J Biol Chem* 282:14525–14535. doi:10.1074/jbc.M700576200
- Clements L, Manilal S, Love DR, Morris GE (2000) Direct interaction between emerin and lamin A. *Biochem Biophys Res Commun* 267:709–714. doi:10.1006/bbrc.1999.2023
- Ellis JA, Craxton M, Yates JR, Kendrick-Jones J (1998) Aberrant intracellular targeting and cell cycle-dependent phosphorylation of emerin contribute to the Emery–Dreifuss muscular dystrophy phenotype. *J Cell Sci* 111(Pt 6):781–792
- Haque F et al (2006) SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide a physical connection between the nuclear lamina and the cytoskeleton. *Mol Cell Biol* 26:3738–3751. doi:10.1128/MCB.26.10.3738-3751.2006
- Herrada I, Bourgeois B, Samson C, Buendia B, Worman HJ, Zinn-Justin S (2015) Purification and structural analysis of LEM-domain proteins. *Methods Enzymol*
- Ho CY, Jaalouk DE, Vartiainen MK, Lammerding J (2013) Lamin A/C and emerin regulate MKL1-SRF activity by modulating actin dynamics. *Nature* 497:507–511. doi:10.1038/nature12105
- Shen Y, Bax A (2010) Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR* 46:199–204. doi:10.1007/s10858-009-9395-y
- Swift J, Discher DE (2014) The nuclear lamina is mechanoresponsive to ECM elasticity in mature tissue. *J Cell Sci* 127:3005–3015. doi:10.1242/jcs.149203
- Tamiola K, Acar B, Mulder FA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* 132:18000–18003. doi:10.1021/ja105656t
- Vranken WF et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. doi:10.1002/prot.20449
- Wolff N, Gilquin B, Courchay K, Callebaut I, Worman HJ, Zinn-Justin S (2001) Structural analysis of emerin, an inner nuclear membrane protein mutated in X-linked Emery–Dreifuss muscular dystrophy. *FEBS Lett* 501:171–176



^1H , ^{13}C and ^{15}N backbone resonance assignment of the lamin C-terminal region specific to prelamin A

Florian Celli¹ · Ambre Petitalot¹ · Camille Samson¹ · François-Xavier Theillet¹ · Sophie Zinn-Justin¹

Received: 24 November 2017 / Accepted: 20 March 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Lamins are the main components of the nucleoskeleton. They form a protein meshwork that underlies the inner nuclear membrane. Mutations in the *LMNA* gene coding for A-type lamins (lamins A and C) cause a large panel of human diseases, referred to as laminopathies. These diseases include muscular dystrophies, lipodystrophies and premature aging diseases. Lamin A exhibits a C-terminal region that is different from lamin C and is post-translationally modified. It is produced as prelamin A and it is then farnesylated, cleaved, carboxymethylated and cleaved again in order to become mature lamin A. In patients with the severe Hutchinson–Gilford progeria syndrome, a specific single point mutation in *LMNA* leads to an aberrant splicing of the *LMNA* gene preventing the post-translational processing of prelamin A. This leads to the accumulation of a permanently farnesylated lamin A mutant lacking 50 amino acids named progerin. We here report the NMR ^1H , ^{15}N , ^{13}CO , $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shift assignment of the C-terminal region that is specific to prelamin A, from amino acid 567 to amino acid 664. We also report the NMR ^1H , ^{15}N , ^{13}CO , $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shift assignment of the C-terminal region of the progerin variant, from amino acid 567 to amino acid 614. Analysis of these chemical shift data confirms that both prelamin A and progerin C-terminal domains are largely disordered and identifies a common partially populated α -helix from amino acid 576 to amino acid 585. This helix is well conserved from fishes to mammals.

Keywords Nuclear envelope · Nucleoskeleton · Intrinsically disordered protein · NMR spectroscopy

Biological context

The nuclear lamina is a universal feature of metazoan nuclear envelopes (Gerace and Huber 2012; Burke and Stewart 2013). It forms a filamentous layer at the nuclear face of the inner nuclear membrane (Mahamid et al. 2016; Turgay et al. 2017) and is composed primarily of A- and B-type lamins (Xie et al. 2016). Lamins provide structural integrity to the nuclear envelope and contribute to genome organization and function (Gruenbaum and Foissner 2015). Lamin levels correlate with tissue rigidity and with mechanical stability of the nucleus (Swift et al. 2013). The two A-type lamin isoforms lamins A and C encoded by the gene *LMNA* are mainly expressed in fully differentiated cells. They are largely identical, diverging only in

their C-terminal regions. The lamin A C-terminal region undergoes extensive processing. It is produced as prelamin A and is then farnesylated, cleaved, carboxymethylated and cleaved again to lead to mature lamin A. Yet a specific single point mutation in *LMNA* leads to the internal deletion of 50 amino acids preventing the post-translational processing of prelamin A (Fig. 1). The deletion mutant, called progerin, remains farnesylated and accumulates at the nuclear periphery, causing nuclear architectural defects such as blebs and herniations of the nuclear envelope and thickening of the nuclear lamina (Maraldi and Lattanzi 2007). This mutant causes an extremely rare genetic disorder characterized by premature, rapid aging shortly after birth called Hutchinson–Gilford progeria syndrome (HGPS (De Sandre-Giovannoli et al. 2003; Eriksson et al. 2003)). Progerin is also present to a lower extent in non-pathological cells and is a biomarker of cellular aging (McClintock et al. 2007). The prelamin A-specific C-terminal region is highly conserved from fishes to mammals (Fig. 1). It is essential for binding to other nuclear envelope proteins such as SUN1 (Haque et al. 2010), to histone modifying enzymes such as SIRT6,

✉ Sophie Zinn-Justin
sophie.zinn@cea.fr

¹ Institut de Biologie Intégrative de la Cellule (I2BC),
CEA, CNRS, Univ. Paris Sud, Université Paris-Saclay,
Gif-sur-Yvette, France

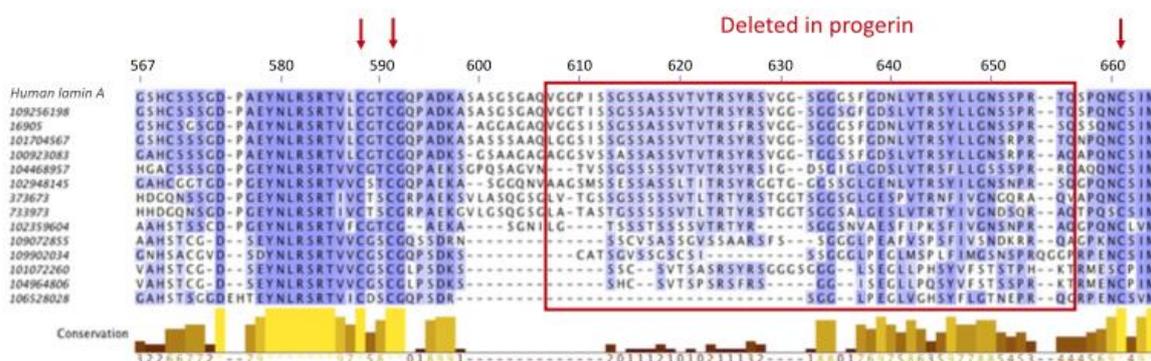


Fig. 1 Sequence alignment of human C-terminal region specific to prelamin A, including a representative set of 14 homologous lamins from mammals, frogs to fishes. The upper sequence is that of human prelamin A. The other sequences are marked using their GenBank Identifier number: 109256198, *Panthera pardus*; 16905, *Mus musculus*; 101704567, *Heterocephalus glaber*; 100923083, *Sarcophilus harrisii*; 104468957, *Pterocles guttularis*; 102948145, *Chelonia mydas*; 373673, *Xenopus laevis*; 733973, *Xenopus tropicalis*;

102359604, *Latimeria chalumnae*; 109072855, *Cyprinus carpio*; 109902034, *Oncorhynchus kisutch*; 101072260, *Takifugu rubripes*; 104964806, *Notothenia coriiceps*; 106528028, *Austrofundulus limnaeus*. The conserved cysteines are indicated by red arrows. The fragment deleted in progerin is highlighted by a red box. Under the sequence alignment, a graph highlighting the amino acid conservation as calculated by Jalview is displayed (Waterhouse et al. 2009)

which is known to deacetylate histone H3 at lysine 9 and 56 (Ghosh et al. 2015), and to the shelterin complex at telomeres (Wood et al. 2014). To further investigate the binding properties of prelamin A and progerin C-terminal regions, before and after post-translational modifications, we have produced and purified these regions and we have described their conformations based on the analysis of their NMR ^1H , ^{15}N , ^{13}CO , $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts.

Methods and experiments

Protein expression and purification

The human prelamin A peptide PreLamC, from amino acid 567 to amino acid 664, and the human progerin peptide ProgC, from amino acid 567 to amino acid 614, were expressed in *Escherichia coli* BL21 DE3 Star(Novagen), using a pETM13 vector coding for a GST-tagged protein with a TEV protease site and an additional tryptophan between the cleavage site and the protein. Their cDNAs were optimized for expression in *Escherichia coli* (Genscript). Moreover, mutated versions of both peptides PreLamC and ProgC in which cysteines are replaced by alanines were also expressed as GST fusion proteins, with a TEV site followed by an additional methionine and an alanine instead of Gly567. These will be further called PreLamC^{CtoA} and ProgC^{CtoA}, respectively. All these peptides were produced using the following protocol. Bacteria were grown in ^{15}N and ^{13}C labelled M9 minimum medium at 37 °C and induced at an optical density of 0.8 with 1 mM Isopropyl

β -D-1-thiogalactopyranoside at 37 °C during 3 h. Cells were lysed in 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 40 mM imidazole, 5% glycerol, 1% Triton X-100 and 1 mM phenylmethanesulfonyl fluoride. Then bacteria were centrifuged at 20,000 $\times g$ for 20 min. Proteins were purified using Glutathione-Sepharose 4B beads (GE Healthcare) equilibrated with a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl (and 10 mM dithiothreitol in the case of cysteine-containing peptides). The tag cleavage was performed by adding TEV protease on the beads before washing with the buffer. Peptides collected during washing were concentrated and aliquoted in 1 ml tubes to be heated 10 min at 95 °C. After centrifugation at 14,000 $\times g$ for 15 min, the supernatants were loaded on a gel filtration column Superdex 75 10/300GL (GE Healthcare) equilibrated in 20 mM Phosphate pH 6.5, 150 mM NaCl, 2 mM dithiothreitol. Finally, all the samples were concentrated up to 600 μM for NMR experiments.

NMR spectroscopy

Most NMR experiments were performed on uniformly ^{15}N and ^{13}C labelled ProgC^{CtoA}, PreLamC^{CtoA} and ProgC peptides in 20 mM phosphate buffer pH 6.5, 150 mM NaCl, 2 mM dithiothreitol, 90%: 10% $\text{H}_2\text{O}:\text{D}_2\text{O}$. Only the $^1\text{H} \rightarrow ^{15}\text{N}$ nOe experiments were recorded on a ^{15}N labelled sample of ProgC. In the case of PreLamC, a ^{15}N labelled sample was produced and an $^1\text{H} \rightarrow ^{15}\text{N}$ HSQC spectrum was recorded. All the NMR experiments were recorded at 283 K on a 600 MHz Bruker Advance II spectrometer equipped with a triple resonance cryogenic probe. ^1H , ^{13}C and ^{15}N resonance frequencies were assigned using 3D HNCACB, CBCA(CO)

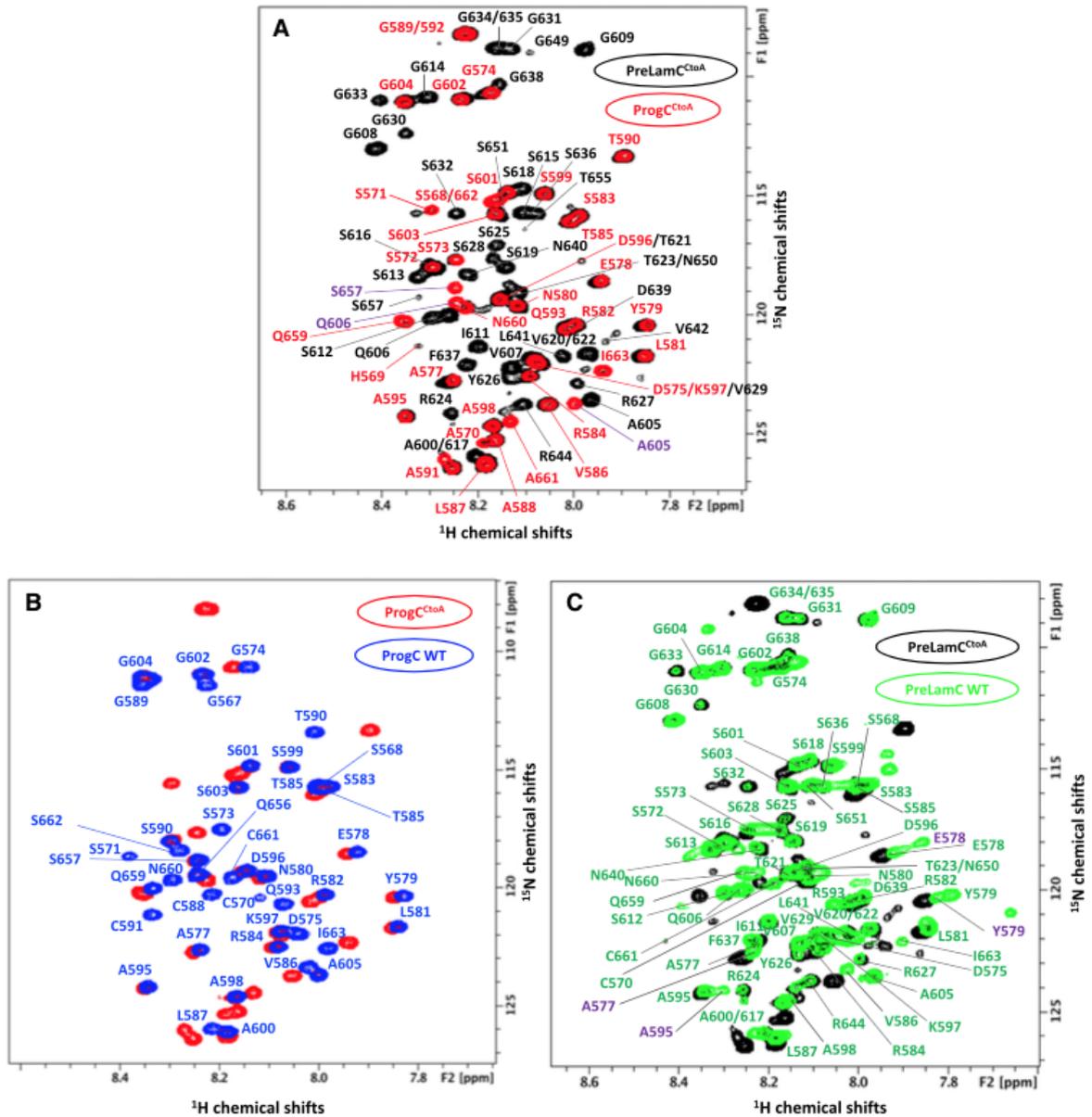


Fig. 2 Superimposition of the ^1H - ^{15}N HSQC spectra of the different lamin fragments analyzed in this study. All these spectra were recorded at 283 K and 600 MHz. **a** The ProgC^{CtoA} spectrum (in red) is displayed onto the PreLamC^{CtoA} spectrum (in black). The peaks from ProgC^{CtoA} that do not overlap with peaks from PreLamC^{CtoA} are labeled in purple. **b** The ProgC spectrum (in blue) is displayed onto the ProgC^{CtoA} spectrum (in red). **c** The PreLamC spectrum (in green)

is displayed onto the PreLamC^{CtoA} spectrum (in black). Peaks corresponding to PreLamC and PreLamC^{CtoA} minor conformations due to proline isomerization are labeled in purple. Indeed, based on their C β chemical shifts, proline residues exhibit a *trans* conformation in the most populated PreLamC^{CtoA} conformation. However, proline residues P576 and P594 clearly show a minor *cis* conformation that give rise to peaks corresponding to their neighboring residues

NH, HNC(O), HN(CA)CO and HN(CO)(CA)NH experiments. The data were processed using Topspin3.1 (Bruker) and analyzed with CCPNMR (Vranken et al. 2005).

Extent of assignments and data deposition

First, NMR assignment was carried out by analyzing spectra recorded on ProgC^{CtoA} and PreLamC^{CtoA}, which did not

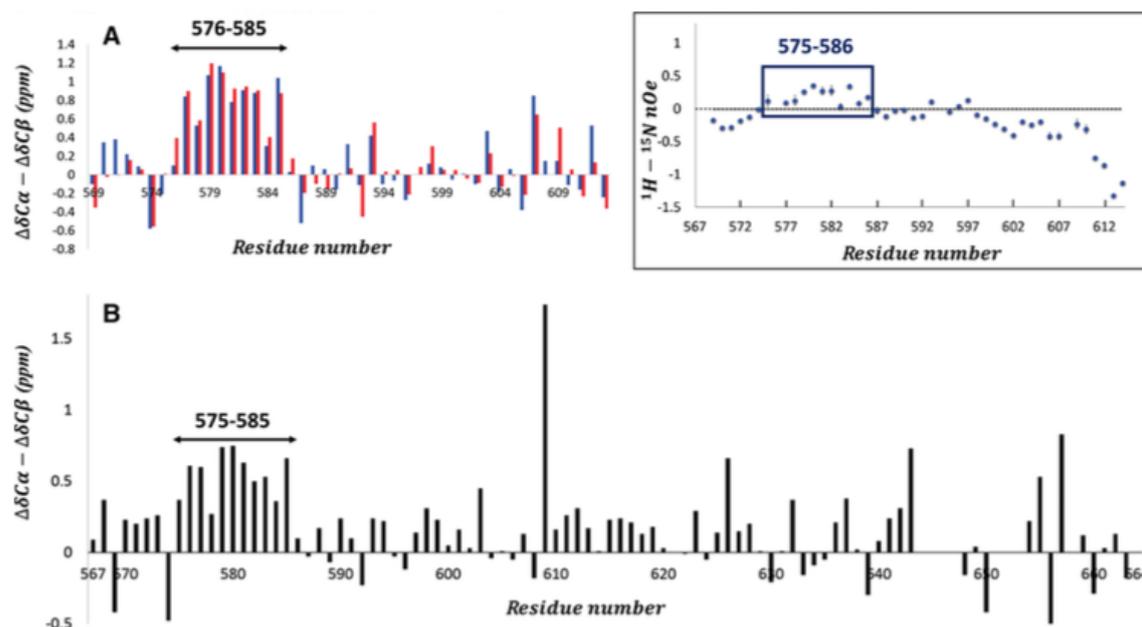


Fig. 3 Analysis of the ^{13}C chemical shifts of ProgC^{CtoA}, ProgC and PreLamC^{CtoA}. **a** Superimposition of the differences between the C α and C β secondary chemical shifts of both ProgC^{CtoA} (in red) and

ProgC (in blue), and in the boxed panel $^1\text{H}\rightarrow^{15}\text{N}$ heteronuclear nOe recorded on a ^{15}N labelled ProgC sample. **b** Differences between the C α and C β secondary chemical shifts of PreLamC^{CtoA}

oxidize with time. Superimposition of the $^1\text{H}\text{-}^{15}\text{N}$ HSQC NMR spectra of ProgC^{CtoA} (red) and PreLamC^{CtoA} (black) is displayed in Fig. 2a. It revealed that the peaks corresponding to the common lamin fragment, i.e. aa 568 to aa 606 and aa 657 to aa 664, nicely overlap. Only peaks corresponding to residues close to the deletion (Ala605, Gln606 and Ser 657) have different positions in the two spectra. In the case of peptides ProgC^{CtoA} and PreLamC^{CtoA}, 100 and 91% (84 out of 92) of $^1\text{H}\text{-}^{15}\text{N}$ pairs, 100 and 93% (90 out of 97) of $^{13}\text{C}\alpha$, 100 and 93% (76 out of 82) of $^{13}\text{C}\beta$ and 100 and 91% (88 out of 97) of ^{13}CO resonances were assigned, respectively. The only unassigned PreLamC^{CtoA} fragment of more than 3 residues corresponds to Ser645-Tyr646-Leu647-Leu648, i.e. the cleavage site of the last maturation step. Further superimposition of the $^1\text{H}\text{-}^{15}\text{N}$ HSQC NMR spectra of ProgC^{CtoA} (red) and ProgC (blue) is displayed in Fig. 2b. Assignment of most ProgC^{CtoA} NMR signals could be easily transferred to ProgC signals. Analysis of the ^1H , ^{13}C and ^{15}N NMR 3D experiments recorded on ProgC was performed in order to confirm the assignment. Thus, the chemical shifts of all $^1\text{H}\text{-}^{15}\text{N}$ pairs, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and ^{13}CO resonances of ProgC could be unambiguously assigned. Finally, superimposition of the $^1\text{H}\text{-}^{15}\text{N}$ HSQC NMR spectra of PreLamC^{CtoA} (black) and PreLamC (green) is displayed in Fig. 2c. Here again most $^1\text{H}\text{-}^{15}\text{N}$ HSQC peaks (82%, 75 out of 92) of PreLamC could be assigned from the comparison with PreLamC^{CtoA}.

Judging by the narrow range of backbone amide ^1H chemical shifts (between 7.8 and 8.6 ppm) and the distribution of the $\Delta\delta\text{C}\alpha\text{-}\Delta\delta\text{C}\beta$ values in ProgC^{CtoA}, ProgC and PreLamC^{CtoA} (Fig. 3), all the peptides are intrinsically disordered. However, a partially populated α -helix is present going from aa 576 to aa 585 of ProgC^{CtoA}, ProgC and PreLamC^{CtoA}. Consistently, positive nOe values were measured at 600 MHz for aa 575 to aa 586 of ProgC (Fig. 3, boxed panel). Interestingly, this partially populated α -helix is located in the most conserved region of ProgC and PreLamC (Fig. 1). We propose that lamin A region from aa 576 to aa 585 is a conserved binding site for a yet unidentified partner. This region contains Arg582, a residue mutated into His in patients with a Dunnigan type familial partial lipodystrophy (Speckman et al. 2000). Mutation could impair binding to a lamin partner. More work is now needed in order to identify how the C-terminal region specific to lamin A recognizes its partners at the nuclear envelope. The chemical shift data have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu/>) under accession numbers 27375 for ProgC^{CtoA}, 27374 for ProgC and 27376 for PreLamC^{CtoA}.

Acknowledgements This work was supported by CEA, CNRS and University Paris South, by the French Infrastructure for Integrated Structural Biology (<https://www.structuralbiology.eu/networks/frisbi>), Grant Number ANR-10-INSB-05-01, Acronym FRISBI) and by the

French Association against Myopathies (AFM; research Grants Nos. 17243 and 20018 to S.Z.-J. and PhD fellowship No. 18159 to C.S.).

References

- Burke B, Stewart CL (2013) The nuclear lamins: flexibility in function. *Nat Rev Mol Cell Biol* 14:13–24. <https://doi.org/10.1038/nrm3488>
- De Sandre-Giovannoli A, Bernard R, Cau P, Navarro C, Amiel J, Boccardo I, Lyonnet S, Stewart CL, Munnich A, Le Merrer M, Lévy N, 2003. Lamin a truncation in Hutchinson-Gilford progeria. *Science* 300:2055. <https://doi.org/10.1126/science.1084125>
- Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, Dutra A, Pak E, Durkin S, Csoka AB, Boehnke M, Glover TW, Collins FS (2003) Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature* 423:293–298. <https://doi.org/10.1038/nature01629>
- Gerace L, Huber MD (2012) Nuclear lamina at the crossroads of the cytoplasm and nucleus. *J Struct Biol* 177:24–31. <https://doi.org/10.1016/j.jsb.2011.11.007>
- Ghosh S, Liu B, Wang Y, Hao Q, Zhou Z (2015) Lamin A is an endogenous SIRT6 activator and promotes SIRT6-mediated DNA repair. *Cell Rep* 13:1396–1406. <https://doi.org/10.1016/j.celrep.2015.10.006>
- Gruenbaum Y, Foisner R (2015) Lamins: nuclear intermediate filament proteins with fundamental functions in nuclear mechanics and genome regulation. *Annu Rev Biochem* 84:131–164. <https://doi.org/10.1146/annurev-biochem-060614-034115>
- Haque F, Mazzeo D, Patel JT, Smallwood DT, Ellis JA, Shanahan CM, Shackleton S (2010) Mammalian SUN protein interaction networks at the inner nuclear membrane and their role in laminopathy disease processes. *J Biol Chem* 285:3487–3498. <https://doi.org/10.1074/jbc.M109.071910>
- Mahamid J, Pfeffer S, Schaffer M, Villa E, Danev R, Cuellar LK, Förster F, Hyman AA, Plitzko JM, Baumeister W (2016) Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* 351:969–972. <https://doi.org/10.1126/science.aad8857>
- Maraldi NM, Lattanzi G (2007) Involvement of prelamin A in laminopathies. *Crit Rev Eukaryot Gene Expr* 17:317–334
- McClintock D, Ratner D, Lokuge M, Owens DM, Gordon LB, Collins FS, Djabali K (2007) The mutant form of lamin A that causes Hutchinson-Gilford progeria is a biomarker of cellular aging in human skin. *PLoS ONE* 2:e1269. <https://doi.org/10.1371/journal.pone.0001269>
- Speckman RA, Garg A, Du F, Bennett L, Veile R, Arioglu E, Taylor SI, Lovett M, Bowcock AM (2000) Mutational and haplotype analyses of families with familial partial lipodystrophy (Dunnigan variety) reveal recurrent missense mutations in the globular C-terminal domain of lamin A/C. *Am J Hum Genet* 66:1192–1198. <https://doi.org/10.1086/302836>
- Swift J, Ivanovska IL, Buxboim A, Harada T, Dingal PCDP, Pinter J, Pajeroski JD, Spinler KR, Shin J-W, Tewari M, Rehfeldt F, Speicher DW, Discher DE (2013) Nuclear lamin-A scales with tissue stiffness and enhances matrix-directed differentiation. *Science* 341:1240104. <https://doi.org/10.1126/science.1240104>
- Turgay Y, Eibauer M, Goldman AE, Shimi T, Khayat M, Ben-Harush K, Dubrovsky-Gaupp A, Sapra KT, Goldman RD, Medalia O (2017) The molecular architecture of lamins in somatic cells. *Nature* 543:261–264. <https://doi.org/10.1038/nature21382>
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. <https://doi.org/10.1002/prot.20449>
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wood AM, Rendtlew Danielsen JM, Lucas CA, Rice EL, Scalzo D, Shimi T, Goldman RD, Smith ED, Le Beau MM, Kosak ST (2014) TRF2 and lamin A/C interact to facilitate the functional organization of chromosome ends. *Nat Commun* 5:5467. <https://doi.org/10.1038/ncomms6467>
- Xie W, Chojnowski A, Boudier T, Lim JSY, Ahmed S, Ser Z, Stewart C, Burke B (2016) A-type lamins form distinct filamentous networks with differential nuclear pore complex associations. *Curr Biol* 26:2651–2658. <https://doi.org/10.1016/j.cub.2016.07.049>

Structural analysis of the ternary complex between lamin A/C, BAF and emerin identifies an interface disrupted in autosomal recessive progeroid diseases

Camille Samson¹, Ambre Petitalot¹, Florian Celli¹, Isaline Herrada¹, Virginie Ropars¹, Marie-Hélène Le Du¹, Naïma Nhiri², Eric Jacquet², Ana-Andrea Arteni¹, Brigitte Buendia³ and Sophie Zinn-Justin^{1,*}

¹Institut de Biologie Intégrative de la Cellule (I2BC), CEA, CNRS, Université Paris Sud, Université Paris-Saclay, Gif-sur-Yvette, France, ²Institut de Chimie des Substances Naturelles, Université Paris Sud, Université Paris-Saclay, CNRS UPR 2301, Gif-sur-Yvette, France and ³Unité de Biologie Fonctionnelle et Adaptative (BFA), CNRS UMR 8251, Université Paris Diderot, Sorbonne Paris Cité, Paris, France

Received December 26, 2017; Revised July 23, 2018; Editorial Decision July 31, 2018; Accepted August 02, 2018

ABSTRACT

Lamins are the main components of the nucleoskeleton. Whereas their 3D organization was recently described using cryoelectron tomography, no structural data highlights how they interact with their partners at the interface between the inner nuclear envelope and chromatin. A large number of mutations causing rare genetic disorders called laminopathies were identified in the C-terminal globular Igfold domain of lamins A and C. We here present a first structural description of the interaction between the lamin A/C immunoglobulin-like domain and emerin, a nuclear envelope protein. We reveal that this lamin A/C domain both directly binds self-assembled emerin and interacts with monomeric emerin LEM domain through the dimeric chromatin-associated Barrier-to-Autointegration Factor (BAF) protein. Mutations causing autosomal recessive progeroid syndromes specifically impair proper binding of lamin A/C domain to BAF, thus destabilizing the link between lamin A/C and BAF in cells. Recent data revealed that, during nuclear assembly, BAF's ability to bridge distant DNA sites is essential for guiding membranes to form a single nucleus around the mitotic chromosome ensemble. Our results suggest that BAF interaction with lamin A/C also plays an essential role, and that mutations associated with progeroid syndromes leads to a dysregulation of BAF-mediated chromatin organization and gene expression.

INTRODUCTION

Laminopathies are rare genetic disorders caused by mutations in genes encoding lamins or by abnormalities in the processing of lamin A (1). They display a large variety of clinical symptoms including cardiomyopathy, muscular dystrophy, lipodystrophy, mandibuloacral dysplasia, restrictive dermopathy, peripheral neuropathy, and premature ageing. A major fraction of the disease-causing mutations comprises single point mutations with poorly understood functional implications. Most of them are distributed all along the *LMNA* gene coding for both prelamin A and lamin C. These two proteins are, together with lamins B1 and B2, essential components of the nucleoskeleton. They share an N-terminal region of 566 residues, and whereas lamin C only possesses six additional specific C-terminal residues, prelamin A presents 98 additional residues, is far-nesyated, carboxymethylated and then cleaved to become mature lamin A. Disease-causing mutations are observed in the large region common to lamins A and C (lamin A/C) as well as in the lamin A specific C-terminal region (2). At the cellular level, lamins A and C form filaments that are mainly present at the nuclear periphery but are also observed in the nucleoplasm (3). Laminopathies are characterized by nuclear morphological abnormalities and an altered pattern of heterochromatin distribution that is more severe in progeroid syndromes, including Hutchinson-Gilford progeria syndrome (HGPS), mandibuloacral dysplasia (MAD), atypical-Werner syndrome (WS) and restrictive dermopathy (RD) (4,5). The lamin filament network interacts with chromatin at different stages of the cell cycle: in ana-telophase, lamin A and C dimers are recruited at the core regions of sister chromosomes; in interphase cells, they relocate both at the nuclear envelope and within the nucleoplasm, where lamins A and C provide anchor points

*To whom correspondence should be addressed. Tel: +33 169083026; Email: sophie.zinn@cea.fr

for chromatin (6–9). The complex molecular organization of the nuclear envelope, nucleoskeleton and their interface with chromatin was recently described using cryo-electron tomography (10,11). Within the densely packed environment observed close to the nuclear envelope, only nuclear pores, lamin filaments and chromatin could be recognized. Lack of high resolution 3D structures for proteins anchored at the inner nuclear membrane as well as complexes between these proteins, the lamina and chromatin still precludes any further description of the nuclear periphery architecture and of the structural defects caused by mutations in the *LMNA* gene coding for lamins A and C.

We set out to understand how the C-terminal immunoglobulin-like (Igfold) domain of lamin A/C interacts with the inner nuclear membrane protein emerin and the chromatin-associated protein Barrier-to-Autointegration Factor (BAF). Indeed, a large number of disease-causing mutations were identified in this lamin Igfold domain (12,13). Mutations causing muscle diseases affect residues of the hydrophobic core, suggesting that destabilization of the lamin A/C Igfold domain is responsible for these diseases (13). Mutations causing either lipodystrophies or progeroid syndromes involve residues localized on two different solvent-exposed sites of the lamin A/C Igfold domain, suggesting that these accessible regions are binding sites for unknown lamin partners and that disruption of these binding events cause the diseases (12–14). Lamin A/C binds to the inner nuclear membrane protein emerin, as shown by blot overlay, coimmunoprecipitation and yeast 2-hybrid assays (15–17). Moreover, the tail region of lamin A/C (residues 384–566, comprising the Igfold domain) is responsible for emerin binding (17,18). The central region of emerin (residues 70–178) is essential for lamin binding (15,18), and its N-terminal domain (residues 1–45, the LEM domain) interacts with BAF, as shown by a panel of techniques going from blot overlay and coimmunoprecipitation assays to X-ray crystallography analyses (15,19).

The structural inter-dependence of lamins, emerin and BAF was revealed by downregulating either lamins A and C or BAF or the LEM domain proteins emerin and LEM2 in *Caenorhabditis elegans* embryos. If any one component was missing, the other two failed to co-assemble, with severe consequences for mitotic spindle assembly and positioning (20), mitotic chromosome segregation and postmitotic nuclear assembly (21,22). *In vitro* experiments revealed that human lamin A binds weakly (K_d of 1 μ M) but directly to BAF and that lamin A and BAF simultaneously bind to emerin (23). Moreover, BAF enhances binding of prelamin A tails to emerin (24). Finally, during nuclear assembly, BAF enriches around the mitotic chromosome ensemble, and its ability to bridge distant DNA sites is essential for guiding membranes to form a single nucleus (25).

We now demonstrate that emerin oligomerization regulates direct lamin A/C binding, and that emerin monomer also interacts with lamin A/C through BAF. We present the 3D structure of a complex involving lamin A/C, emerin and BAF. We propose a model in which BAF dimer bound to two DNA molecules interacts on one side with emerin and on the other side with lamin A/C, and we analyse the

impact of mutations causing autosomal recessive progeroid syndromes of the assembly of this complex.

MATERIALS AND METHODS

Protein expression and purification

Human WT emerin fragments from amino acid 1 to amino acid 187 (EmN), from amino acid 67 to 170 (EmC170), from amino acid 67 to 187 (EmC187) and from amino acid 67 to amino acid 221 (EmC), human BAF with all cysteines mutated in alanine and human lamin B1 tail from amino acid 409 to amino acid 586 were expressed using a pETM13 vector as N-terminal octa-histidine fusions in *Escherichia coli* BL21 DE3 Star (Novagen). The emerin and BAF cDNAs were optimized for expression in *Escherichia coli* (Genscript). Human wild-type lamin A/C fragment from amino acid 411 to amino acid 566 (LamIgF) was expressed using a pGEX vector as an N-terminal GST fusion in *E. coli* BL21 DE3 Star (Novagen). All LamIgF mutant expression vectors (R435C, R453W, R471C, R482W, R527H, A529V and K542N) were obtained by mutagenesis using the Quikchange (Agilent) kit from the LamIgF expression vector. Bacteria were cultured in rich medium (*lysogeny broth*, LB) or 15 N-labeled minimum medium, induced at an optimal density of 1 with 0.5 mM isopropyl β -D-1-thiogalactopyranoside, grown overnight at 293 K, and lysed in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 40 mM imidazole, 5% glycerol, 1% Triton X-100 and 1 mM phenylmethanesulfonylfluoride.

For EmN, EmC170, EmC187, EmC and BAF, after sonication at room temperature and centrifugation at 20 000 g, for 20 min, at 277 K, the pellet was resuspended in buffer C8 (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 8 M urea). A second centrifugation step was performed at 20 000 g, for 20 min, at 293 K. The soluble extract was then loaded onto Ni-NTA beads (GE-Healthcare) equilibrated with buffer C8. Proteins were eluted directly with buffer E8 (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 M imidazole, 8 M urea). Then, proteins were refolded by dialysis in buffer D1 (20 mM Tris-HCl pH 8.0, 30 mM NaCl, 2 mM DTT) for EmN, EmC170, EmC187 and EmC and buffer D2 (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM DTT) for BAF. After refolding, they were incubated with the TEV protease during 3 h at room temperature and finally dialyzed into the selected buffer. The final yield was typically 10 mg of purified protein per liter of bacterial culture (26). Self-assembly was initiated using proteins that were dialyzed in a buffer containing 20 mM Tris-HCl pH 8.0 and 30 mM NaCl. After dialysis, proteins were concentrated until 600 μ M, reduced with 5 mM DTT and stored at room temperature during one week.

For lamin B1 tail, after sonication at 277 K, benzonase (SigmaAldrich) addition and centrifugation at 20 000 g, for 30 min, at 277 K, the soluble extract was loaded onto Ni-NTA beads (GE-Healthcare) equilibrated with buffer C (50 mM Tris-HCl pH 8.0, 150 mM NaCl and 20 mM imidazole). The protein was then eluted directly with buffer E (50 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 M imidazole) and dialyzed in buffer C. It was incubated with the TEV protease during 3 h at room temperature, loaded again onto

Ni-NTA beads and finally, after flow-through recovery, dialyzed into the selected buffer. The final yield was typically 20 mg of purified protein per liter of bacterial culture.

For LamIgF (WT and mutants), after sonication at 277 K, benzonase (SigmaAldrich) addition and centrifugation at 20 000 g, for 30 min, at 277 K, the soluble extract was loaded onto GST-beads (GE-Healthcare) equilibrated with buffer T2 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT). After 2 h of incubation at 277 K, GST-beads were washed with buffer T2 and 1 M NaCl, and then thrombin was added and incubated with the beads overnight. The protein was recovered in the flow-through and finally dialyzed into the selected buffer. The final yield was typically 2 mg of purified protein per liter of bacterial culture.

Liquid-state NMR spectroscopy

NMR experiments were performed on a 700 MHz Bruker Avance spectrometer equipped with a triple resonance cryogenic probe. To study interaction between the lamin fragments and BAF, two-dimensional ^1H - ^{15}N HSQC spectra were acquired at 303 K, on a 3-mm-diameter NMR sample tube containing 100–200 μM uniformly ^{15}N -labeled lamin peptides and non-labeled BAF, in 20 mM sodium phosphate pH 7.0, 100 mM NaCl and 80:20% $\text{H}_2\text{O}/\text{D}_2\text{O}$. To study the interaction between lamin fragments and EmN or EmC, two-dimensional ^1H - ^{15}N HSQC spectra were acquired at 303 K, in a 3-mm-diameter NMR sample tube containing 150 μM uniformly ^{15}N -labeled lamin peptides and non-labeled EmN or EmC, in 20 mM Tris-HCl pH 8.0, 30 mM NaCl and 80:20% $\text{H}_2\text{O}/\text{D}_2\text{O}$. The ^1H , ^{15}N , ^{13}C NMR signals of lamin A/C domain from aa 411 to aa 553 were already assigned previously in a phosphate buffer at pH 6.3 and 303 K (27). In order to assign the NMR signals of lamin A/C region from aa 411 to aa 566 (LamIgF) in our conditions (20 mM sodium phosphate pH 7.0, 100 mM NaCl, 303 K), we took advantage of these previous assignments, but also recorded additional 3D ^1H - ^{15}N - ^{13}C correlation spectra using 3D HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO and HN(CO)(CA)NH pulse sequences on a 3-mm-diameter NMR sample tube containing 500 μM uniformly $^{15}\text{N}/^{13}\text{C}$ -labeled LamIgF. The data were processed using Topspin 3.1 (Bruker) and analyzed using CCPNMR (28).

Size-exclusion chromatography

Size-exclusion chromatography experiments aiming at identifying interactions between emerin, BAF and lamin fragments (LamIgF WT and mutants, lamin B1 tail) were performed using a Superdex-75 10/300 GL column (GE Healthcare) pre-equilibrated with buffer G (20 mM Tris-HCl pH 8.0, 30 mM NaCl, 2 mM DTT). 500 μl of proteins concentrated at 150 μM were injected for each experiment at a flow rate of 0.5 ml/min at 277 K.

Isothermal titration calorimetry

ITC was performed using a high-precision VP-ITC calorimetry system. To characterize interactions between the BAF dimer and LamIgF (WT and mutants), all proteins

were dialyzed against 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM β -mercaptoethanol and protease inhibitors (Roche). BAF (20 μM) in the calorimetric cell at 288 K was titrated with LamIgF WT (at a concentration of 100 μM in the injection syringe) or BAF in the calorimetric cell at 283 K was titrated with LamIgF mutants (at a concentration of 100 μM in the injection syringe). To characterize interactions between the BAF dimer and EmN, all proteins were dialyzed against 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM β -mercaptoethanol and protease inhibitors (Roche). BAF (30 μM) in the calorimetric cell at 288 K was titrated with EmN (at a concentration of 150 μM in the injection syringe). Analyses of the data were performed using the Origin software provided with the instrument.

Negative-staining electron microscopy

To observe self-assembled EmN and EmC proteins, with or without LamIgF WT or lamin B1 tail, protein samples containing 2% uranyl acetate were deposited on glow-discharged carbon-coated copper grids. Data collection was performed using a Tecnai Spirit transmission electron microscope (FEI) equipped with a LaB6 filament, operating at 100 keV. Images were recorded using a K2 Base camera (Gatan, 4k \times 4k) at 15 000 magnification (at the level of the microscope) with a pixel size of 0.25 nm at the specimen level.

X-ray crystallography

Crystallization and data collection. The EmN-Baf-LamIgF complex was purified by gel filtration (Superdex 75, 10/300 GL) in 50 mM Tris-HCl pH 8.0, 100 mM NaCl and concentrated to 3 mg/ml. Crystallization was initiated one week after the gel filtration and in these conditions, EmN was proteolyzed and the final complex was composed of LEM-BAF-LamIgF. For crystallization, 1 μl of the complex was mixed with 1 μl of the reservoir solution and equilibrated against a 500- μl reservoir by hanging drop at 277 K. Crystals were grown in 25% (w/v) polyethylene glycol (PEG) 3.350, 20 mM Bis-Tris pH 5.5, 30 mM NaCl and 0.2 M NH_4SO_4 . They were flash-cooled in liquid nitrogen, using a cryo-protection solution of 25% PEG 3.350, 20 mM Bis-Tris pH 5.5 and 0.2 M NH_4SO_4 , supplemented with 27% (V/V) ethylene glycol.

Structure determination and refinement. The 3D structure of the complex was solved by molecular replacement using Molrep in CCP4 (29,30). The coordinates of the BAF dimer with DNA (PDB entry: 2BZF), the coordinates of lamin A/C globular domain (PDB entry 1IFR) and the coordinates of the emerin LEM domain (PDB entry: 2ODC) were used as templates. The resulting model was rebuilt using PHENIX suite (31), manual correction was performed with Coot (32) according to $|F_o| - |F_c|$ and $2|F_o| - |F_c|$ electron density maps, and further refinement was carried out with phenix.refine. All structure figures were prepared using PyMOL (Schrödinger, LLC). Structure coordinates were deposited to the PDB, with entry 6GHD.

Fluorescence-based thermal shift assay

The thermal stability of LamIgF proteins was monitored by a fluorescence-based thermal shift assay performed with a QuantStudio 12K Flex instrument (Life Technologies). 10 μ g of purified protein was mixed with the SYPRO Orange dye (diluted 800-fold from a 5000-fold stock solution, Invitrogen) in 50 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol. Reactions were carried out in duplicate in a 96-well fast PCR plate at a final volume of 20 μ l (33 μ M) and each experiment was repeated at least twice independently. The samples were submitted to a denaturation kinetic from 283 to 368 K at a rate of 3 K/min and fluorescence of Sypro Orange dye was recorded in real time. The protein denaturation temperature ($T_d \pm$ s.e.m.) was calculated using the Protein Thermal Shift software v1.3 (Life Technologies) as the maximum of the derivative of the resulting fluorescence curves.

Proximity ligation assays

Transfection. HeLa cells were obtained from American Type Culture Collection and cultured in Minimum Essential Medium containing Glutamax (Gibco), 1% non-essential amino acids and 10% fetal bovine serum. HeLa cells were transfected using XtremeGene 9 (Roche) with the plasmids pEGFC3-BAF encoding full length BAF and pcmvTAG2A-FLAG-Lamin A encoding mature lamin A (aa 1-646). After 24 h, cells were processed for either immunoblotting, immunofluorescence or Proximity ligation assay.

Immunoblotting. Whole cell protein extracts were suspended in Laemmli sample buffer, separated by SDS-PAGE and transferred to nitrocellulose membranes. Membranes were blocked for 1.5 h in TBST (10 mM Tris pH 8.0, 150 mM NaCl, 0.05% Tween-20) containing 5% dry milk, incubated with mouse anti FLAG antibody (Sigma; 1:600) and rabbit anti-BAF (SantaCruz, 1:200) for 1 h in TBST/1% milk, washed 4 times and incubated with HRP-conjugated secondary antibodies. After four washes in TBST, proteins were detected by enhanced chemiluminescence.

Immunofluorescence microscopy. Cells were fixed with 3% paraformaldehyde for 12 min at R.T., permeabilized with phosphate buffered saline containing 0.5% Triton for 5 min at R.T., and quenched with 2% bovine serum albumin diluted in phosphate buffered saline containing 0.1% Triton. Primary antibodies were rabbit anti-FLAG (Sigma, 1:300) and mouse anti-GFP (Roche, 1:200). Fluorescent labeled secondary antibodies (donkey anti-mouse Cy2 1:60 and donkey anti-rabbit 594 1:200) were from Jackson ImmunoResearch. DNA was stained with Hoechst 33258 (1 μ g/ml).

Proximity ligation assay. PLA was used to detect interactions between GFP-BAF and FLAG-lamin A, either WT or mutant (R435C, R527H, A529V, K542N, R453W, R482W), based on proximity (<40 nm) of two secondary antibodies directed against these tagged proteins. After cell fixation, cell permeabilisation and quenching (as above), pairs of primary antibodies, rabbit anti-FLAG 1:300 and mouse

anti-GFP 1:200, were added on HeLa cells expressing GFP-BAF together with FLAG-lamin A, for 40 min at R.T. Next, Duolink PLA probe anti-rabbit plus, Duolink PLA probe anti-mouse minus and Duolink detection reagents orange (detected with a Cy3 filter) were used according to manufacturer's instructions (Olink, Bioscience). At last, cells were mounted in Duolink mounting medium with Dapi (Olink, Bioscience). Confocal microscopy image acquisition was performed using a LSM 700 Laser scanning microscope (Zeiss) at the imaging facility of the BFA institute. Quantitative analysis of PLA signals was done on images using ImageJ, as reported in (33). Data were then analyzed by comparing median values for the total number of pixels showing a signal (Cy3) per nucleus. Data of the three independent experiments were normalized to 1 (100%) for cells expressing FLAG-lamin A WT. Statistical analysis were performed using Kruskal–Wallis tests.

RESULTS

The lamin A/C Igfold domain interacts with the self-assembled emerlin fragment EmN

To determine how the lamin A/C tail interacts with emerlin, we generated two fragments of the emerlin nucleoplasmic region (Figure 1A). We tested binding between the tail region common to lamins A and C (LamIgF comprising aa 411 to aa 566) and these emerlin fragments. As previously published, the whole emerlin nucleoplasmic region from aa 1 to aa 221 is poorly soluble (34). We thus produced two overlapping emerlin peptides, which we called EmN (26) and EmC (Figure 1A). EmN (aa 1 to aa 187) comprises a LEM domain and a region that is intrinsically disordered (34,35). We previously showed that, *in vitro*, it can be observed either as a monomer or as an oligomer (26) (Figure 1B, left panel). EmC (aa 67 to aa 221) is entirely unstructured as observed by NMR (Supplementary Figure S1A), and after purification immediately oligomerizes as observed by fluorescence (Supplementary Figure S1B) and negative-staining electron microscopy (EM; Figure 1B, right panel). In order to identify an interaction between lamin A/C and emerlin, we produced a 15 N labeled LamIgF peptide, and mixed the labeled peptide with either monomeric EmN, oligomeric EmN or oligomeric EmC (Figure 1C). The oligomeric state of emerlin fragments in the NMR samples was systematically checked by EM (Supplementary Figure S2). Nuclear Magnetic Resonance (NMR) analysis of these samples revealed that LamIgF only interacts with oligomeric EmN in these conditions. This interaction is specific to lamin A/C as NMR analysis of a sample containing 15 N labeled lamin B1 tail and oligomeric EmN did not demonstrate any binding (Figure 1C, lower and right panel; the lamin A/C and B1 Igfold domains share 52% of sequence identity).

The Igfold domain of lamin A/C forms a ternary complex with the LEM domain of emerlin and BAF

The chromatin-associated protein BAF binds to both lamin A/C and emerlin (36,37). We examined the role of BAF in mediating an interaction between LamIgF, EmN and BAF by Size-Exclusion Chromatography (SEC). Consistent with our previous NMR results, we observed that LamIgF and

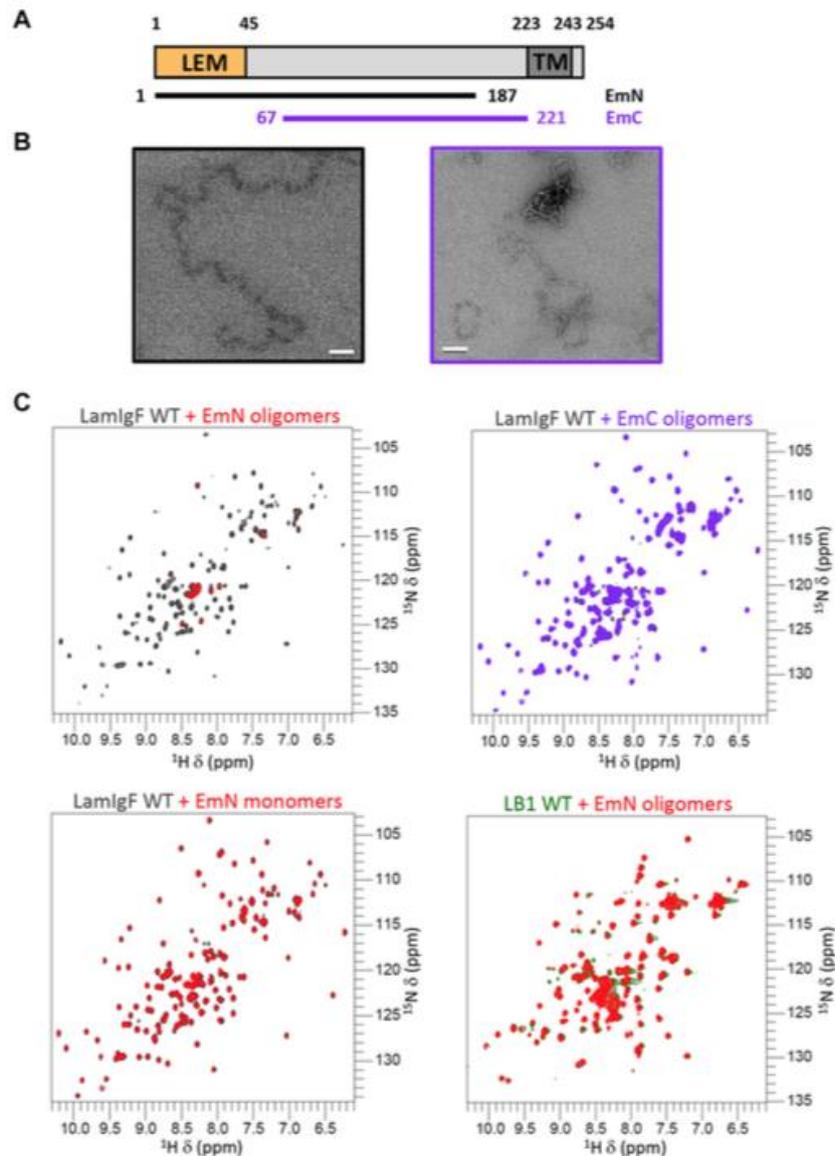


Figure 1. The lamin A/C Igfold domain directly binds to emerin self-assembled nucleoplasmic region. (A) Architecture of the emerin protein, highlighting its LEM domain (in orange) and its transmembrane domain (in dark grey), as well as the 2 fragments of the nucleoplasmic region that could be obtained as soluble peptides (EmN and EmC). (B) Negative staining EM images of EmN and EmC fragments obtained after purification, concentration and incubation at 293 K (white bar: 100 nm). EmN filaments were observed at 600 μM after 1 week (26), whereas EmC filaments were already visible at 300 μM after 1 day. The fragment corresponding to the region common to EmN and EmC did not self-assemble in these conditions (Supplementary Figure S1B). (C) Superimposition of the solution-state NMR ^1H - ^{15}N HSQC spectra of the lamin Igfold domain (lamin A/C in gray or lamin B1 in green) recorded in the absence or presence of emerin fragments EmN or EmC (EmN in red, EmC in purple; 1:1 ratio). The emerin fragments were EmN filaments (upper left and lower right views), EmN monomers (lower left view) and EmC filaments (upper right view). All these spectra were recorded at 303 K on a 700 MHz spectrometer.

monomeric EmN did not co-elute under the conditions used (Figure 2A). However, they co-eluted in the presence of BAF. Similarly, we tested the binding between the lamin B1 tail fragment, EmN and BAF. In this case, no binding could be observed between lamin B1 and BAF, precluding the formation of a ternary complex with EmN. NMR further confirmed that the lamin B1 tail fragment did not interact with BAF (Supplementary Figure S3). Isothermal Titr-

ation Calorimetry (ITC) experiments showed that LamIgF and EmN bind to BAF with an affinity of $3.2 \pm 1.2 \mu\text{M}$ and $0.7 \pm 0.2 \mu\text{M}$, respectively (Figure 2B). The LamIgF/BAF interaction is enthalpy driven, whereas the EmN / BAF interaction is both enthalpy and entropy driven (Supplementary Table S1). Crystal structure of the ternary complex was solved at 1.9 Å resolution (Figure 3A; Supplementary Table S2). One BAF dimer interacts on one side with the LEM do-

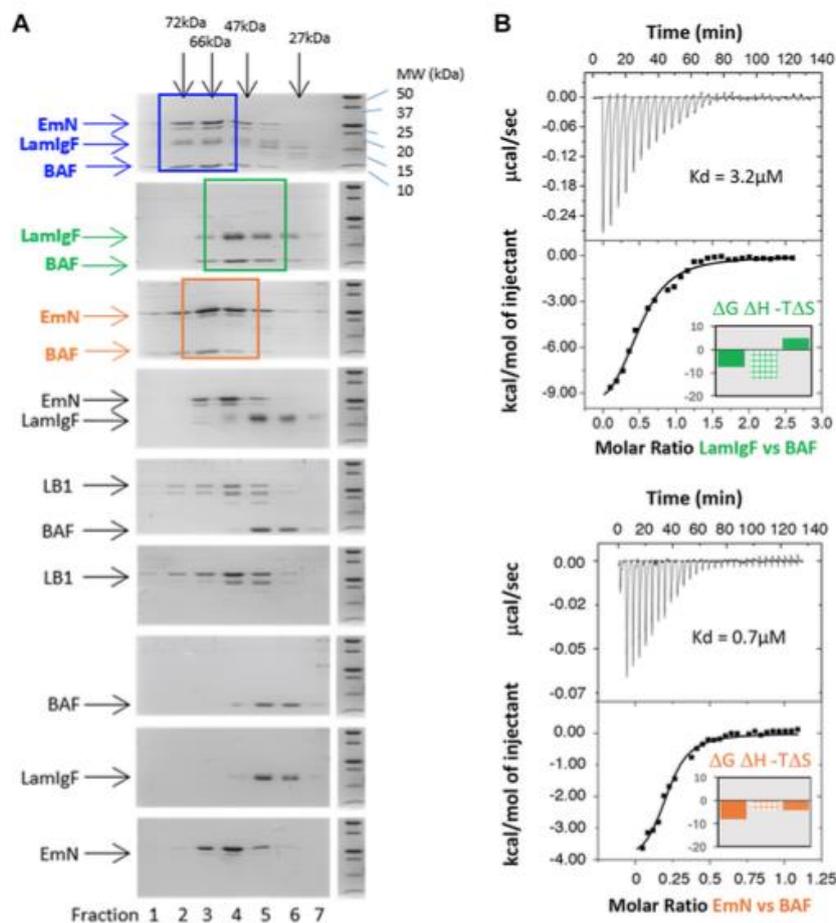


Figure 2. The lamin A/C Igfold domain binds to emerin monomeric LEM domain through BAF. (A) Detection of a ternary complex including LamIgF, EmN and BAF by size exclusion chromatography using a Superdex 75 10/300GL column. Each gel corresponds to the elution of a different set of proteins, using the same column, the same protein concentrations and the same buffer: from the bottom EmN, LamIgF, BAF, lamin B1 tail (LB1; the lower band corresponds to a degradation product), BAF and lamin B1 tail, EmN and LamIgF, EmN and BAF, LamIgF and BAF and finally EmN, LamIgF and BAF. The bands corresponding to the different complexes are boxed: EmN and BAF in orange, LamIgF and BAF in green and the ternary complex in blue. (B) ITC titration of BAF onto LamIgF (upper panel) and BAF onto EmN (lower panel). 200 μM LamIgF or EmN were injected in a cell containing 40 μM of BAF at a temperature of 288 K. The affinities deduced from the fitting are $3.2 \pm 1.2 \mu\text{M}$ and $0.7 \pm 0.2 \mu\text{M}$, respectively. The corresponding thermodynamics parameters (in kcal/mol) are plotted in the low right corner of the ITC figures.

main of emerin and on the other side with the Igfold domain of lamin A/C. The emerin fragment outside of the LEM domain is not visible in the electron density map. SDS-PAGE revealed that it was proteolyzed between protein purification and crystallization.

Dimerization of BAF is essential for both lamin and emerin binding

The 3D structure of the ternary complex highlights that BAF dimerization is essential for emerin and lamin A/C binding. First, BAF monomers 1 and 2 both interact with the LEM domain of emerin (interfaces of 395 and 211 \AA^2 , respectively) (Figure 3B). BAF monomer 1 loop $\alpha 2\alpha 3$ (Phe39), helix $\alpha 3$ (Gln48, Val51, Leu52) and helix $\alpha 4$ (Leu58, Glu61, Trp62) contact a LEM surface comprising the β -strand (Gly24, Pro25), loop $\beta 1\alpha 2$ (Ser29, Thr30) and helix $\alpha 2$ (Leu33, Tyr34, Lys36, Lys37). BAF monomer

2 loop $\alpha 2\alpha 3$ (Arg37, Phe39) and helix $\alpha 3$ (Glu48) contact an overlapping LEM surface comprising helix $\alpha 1$ (Thr13), the β -strand (Pro25) and loop $\beta 1\alpha 2$ (Val27). A similar 3D structure between BAF dimer and emerin LEM domain was previously modeled based on NMR data (Figure 3C; PDB reference 2ODG (19)). When we superimposed the $C\alpha$ atoms of the 2 structures, we obtained root mean square deviation values of 1.3, 1.4 and 0.9 \AA for the LEM domain, BAF monomer 1 and BAF monomer 2, respectively, demonstrating that the NMR and X-ray structures are highly similar. If we superimposed only the $C\alpha$ atoms of dimeric BAF, we measured a root mean square deviation of 2.8 \AA between the LEM domains of the 2 structures. Consistently, the surface buried at the LEM/BAF interface yielded 510 \AA^2 in the NMR structure compared to 606 \AA^2 in the X-ray structure. Altogether, we observed a slightly different positioning of the LEM domain onto the BAF dimer

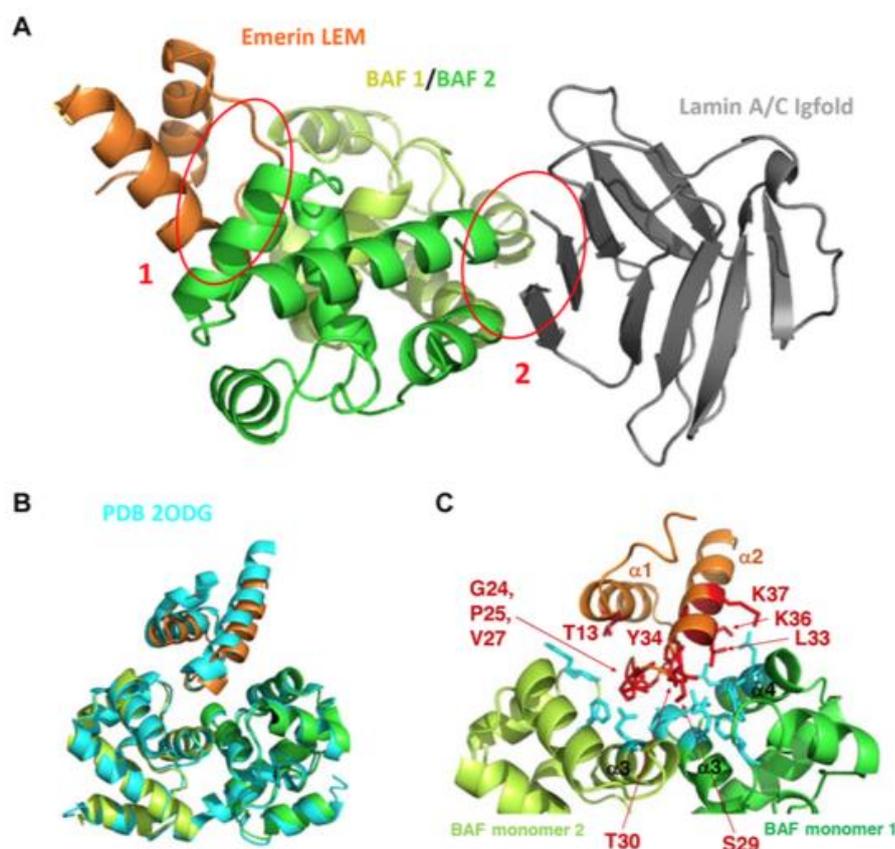


Figure 3. Three-dimensional structure of the complex between LamIgF, emerlin LEM domain and BAF. (A) Cartoon representation of the complex, with emerlin LEM domain (residues 3–44) in orange, BAF dimer in yellow-green and green (residues 4–89), and lamin A/C Igfold domain (residues 432–544) in gray. The interfaces corresponding to the emerlin / BAF and BAF / lamin interactions are indicated by red circles and numbered as 1 and 2, respectively. (B) Superimposition of the 3D structure of the BAF dimer bound to the emerlin LEM domain, as determined in this work (same colors as in (A)), and as revealed using NMR by Clore *et al.* (PDB reference 2ODG (19); in cyan). (C) Zoom on the EmN/BAF interface, with residues >30% buried within the interface displayed in sticks. On the emerlin side, the interface is mainly formed by residues Gly24 to Lys37 (labeled residues on loop $\alpha 1\alpha 2$ and helix $\alpha 2$). On the BAF side, it is mainly formed by helices $\alpha 3$ and $\alpha 4$ of one monomer and loop $\alpha 2\alpha 3$ and helix $\alpha 3$ of the other monomer.

and a consistent larger surface complementarity in the X-ray structure.

Both BAF monomers also contribute to lamin A/C binding, through interfaces of 270 and 192 Å², respectively (Figure 4A). BAF monomer 1 helix $\alpha 1$ (Val11, Ala12), loop $\alpha 1\alpha 2$ (Pro14) and helix $\alpha 5$ (Glu83, Ala87, Phe88) contact lamin A/C strand $\beta 1$ (His433, Ala434, Arg435) and strand $\beta 9$ (Met540, Lys542), whereas BAF monomer 2 helix $\alpha 1$ (Ala12), loop $\alpha 1\alpha 2$ (Pro14) and helix $\alpha 5$ (Ala87, Phe88) contact lamin A/C strand $\beta 1$ (Arg435) and loop $\beta 8\beta 9$ (Glu537). It was reported that BAF mutant G47E lost its ability to bind to lamin A as well as prelamin A in cells (5). Gly47 is located at the center of the dimerization interface (in magenta on Figure 4A). Mutation G47E might destabilize the BAF dimer, therefore hindering BAF interaction with both emerlin LEM domain and lamin A/C Igfold.

The interface between lamin A/C and BAF involves residues mutated in progeroid syndromes

By solving the 3D structure of a ternary complex formed by BAF, EmN and LamIgF, we revealed a yet undescribed

interface between the BAF dimer and the lamin A/C Igfold domain. On the BAF side, Ala12, Pro14 and Phe88 of both monomers are more than 50% buried in the complex (Figure 4A). An Ala12 to Thr amino acid substitution occurs in patients with a Nestor-Guillermo progeria syndrome (38). Our 3D structure predicts that this substitution causes a defect in BAF/lamin binding. Consistently, in cells, it significantly decreases BAF binding to lamin A and prelamin A (5).

On the lamin side, Arg435, Met540 and K542 are more than 50% buried at the lamin/BAF monomer 2 interface (Figure 4A). Arg435 side-chain is hydrogen-bonded to Phe88 backbone oxygens of both BAF monomers, while Lys542 side-chain is bonded to BAF monomer 2 Asp86 backbone oxygen. These lamin residues are mutated in patients with severe autosomal recessive progeroid syndromes (RD and HGPS (14,39–43)). They belong to a hot spot for mutations causing progeroid diseases (14). In order to verify that in solution, this hot spot surface is contacting BAF, we followed the NMR ¹H and ¹⁵N signals of LamIgF while adding BAF to the NMR sample. We observed that the intensities of the NMR signals corresponding to several

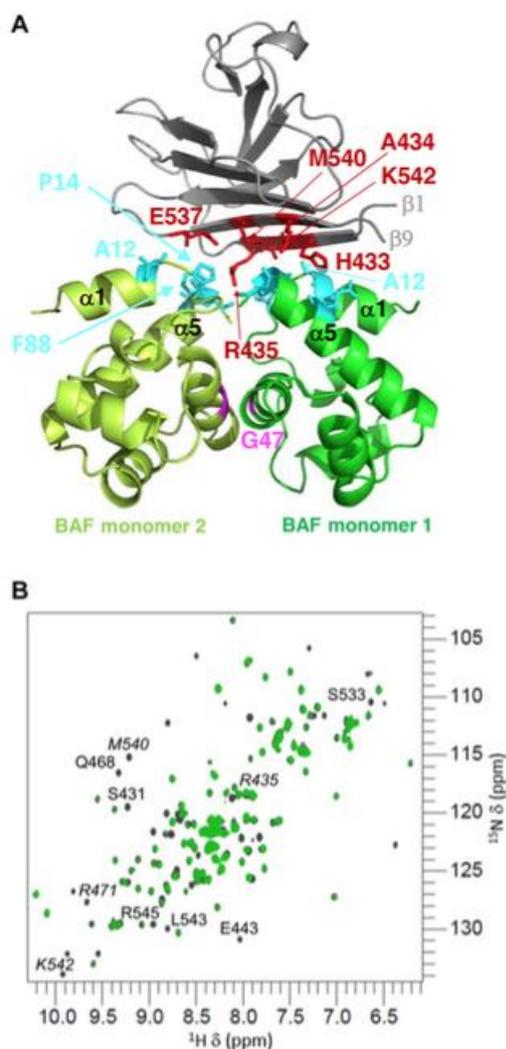


Figure 4. Lamin A/C Igfold domain binds BAF through a β -sheet including residues mutated in recessive progeroid syndromes. (A) Cartoon representation of the complex between the lamin A/C Igfold domain and the BAF dimer (main chain colors as in Figure 3). On the lamin side, the interface is mainly polar: it is formed by residues His433, Ala434, Arg435, Glu537, Met540 and Lys542 (in red sticks). On the BAF side, it is formed by residues from helix $\alpha 1$, loop $\alpha 1\alpha 2$ and helix $\alpha 5$ from both monomers. At the BAF monomer-monomer interface, Gly47 is displayed in magenta. (B) ^1H - ^{15}N HSQC spectrum of LamIgF (150 μM) in the absence (grey) and presence (green) of BAF (150 μM) (ratio: 1 LamIgF for 0.5 BAF dimer). The labelled peaks exhibit a significant intensity decrease after addition of BAF. The peaks labelled in italics correspond to residues mutated in patients with progeroid diseases (R435(39,40),(42), R471C (44), M540 (43),(14), K542 (41)).

residues of strand $\beta 1$ (as Arg435), loop $\beta 1\beta 2$, strand $\beta 4$ (as Arg471), loop $\beta 8\beta 9$ and strand $\beta 9$ (as Met540 and Lys542) were significantly decreased (Figure 4B). This demonstrated that the β -sheet formed by strands $\beta 1$ and $\beta 9$ is indeed involved in BAF binding in solution.

Lamin A/C Igfold mutations causing progeroid syndromes decrease binding affinity for BAF

To identify the consequences of the LamIgF mutations causing progeroid diseases on the formation of the lamin-BAF complex, we produced a set of LamIgF peptides mutated on sites responsible for severe diseases. We focused our interest on five homozygous LamIgF mutations, R435C (strand $\beta 1$) (39,40,42), R471C (strand $\beta 4$) (44), R527H (45–48), A529V (49) (strand $\beta 8$) and K542N (strand $\beta 9$) (41), which were identified in patients with atypical progeroid syndromes. Two of these mutations cause syndromes generally appearing during early childhood as RD (R435C) and HGPS (K542N), characterized by severe premature aging features, whereas mutations R471C, R527H and A529V cause a disease generally appearing later, called MAD, characterized by growth retardation and skeletal abnormalities. In the case of R471C, the disease was particularly severe and the phenotype combined MAD, progeria, and rigid spine muscular dystrophy (44). We also produced as control variants the mutants R453W and R482W that cause Emery-Dreifuss Muscular Dystrophy (EDMD) and Dunnigan-type Familial Partial Lipodystrophy (FPLD), respectively. We first verified that the LamIgF mutated peptides were well folded and had a thermal stability close to that of WT LamIgF (i.e. ± 5 K relatively to WT; Supplementary Figure S4; see also the X-ray structure of mutant R482W in (50)). We then measured their affinity for BAF by ITC (Figure 5A; Supplementary Table S1). We observed that the 2 control mutants R453W and R482W bind BAF as LamIgF WT, whereas mutant R435C does not bind BAF in our conditions. Mutants R471C and A529V exhibit a 5-fold decreased affinity for BAF. In the case of K542N and R527H, the heat release due to BAF binding was so weak that no affinity could be measured. We also tested the binding of LamIgF mutants to BAF by SEC (Figure 5B). We observed that mutants R453W and R482W coeluted together with BAF as WT LamIgF. However, mutant R435C did not induce any elution volume shift of BAF, and mutants R471C, A529V, R527H and K542N only weakly shifted BAF elution volume. We concluded that the five tested mutations causing atypical progeroid syndromes decrease the affinity of lamin A/C for BAF. Moreover, the binding defect seems to globally correlate with the disease severity (Figures 5A and B).

Finally, we transfected HeLa cells with plasmids coding for GFP-BAF and FLAG-mature lamin A (aa 1 to aa 646) WT or variants and detected using an *in situ* Proximity Ligation Assay the proximities between BAF and lamin A in these cells. We thus tested the impact of mutations related to RD (R435C), HGPS (K542N) and MAD (R527H, A529V), as well as EDMD (R453W) and FPLD (R482W), on these proximities. FLAG-lamin variants were expressed at similar protein levels as FLAG-lamin WT; they localized both at the nuclear periphery and in the nucleoplasm (Figure 6A and B). We observed that all 4 lamin A mutations associated with atypical progeroid syndromes similarly reduced the frequency of the proximity events between BAF and lamin A (Figures 6C and 7). Lamin A mutations associated with muscle and adipose tissue diseases had an intermediate impact: they also reduced the number of proximity

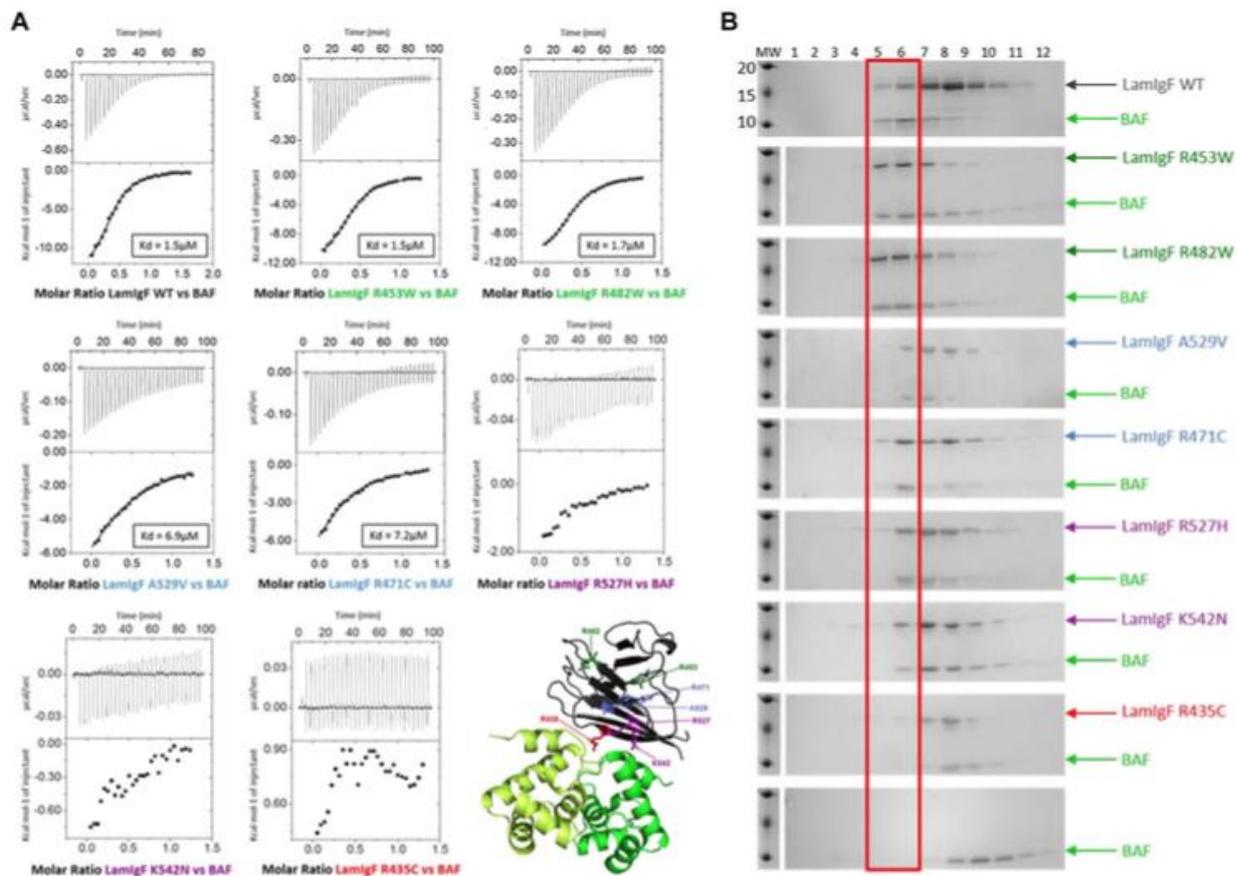


Figure 5. Impact of mutants causing recessive progeroid syndromes on BAF binding depends on the disease severity. (A) ITC titration of BAF onto LamIgF, as measured at 283K. Whereas variants R453W and R482W show WT binding, variants A529V and R471C exhibit a significantly lower binding enthalpy and affinity as compared to LamIgF WT. No affinity could be measured for R527H and K542N and no binding could be detected for R435C. In the lower right panel, the spatial distribution of residues mutated in (A) is displayed on the lamin A/C Igfold structure (PDB reference 1IFR (12)). The two residues mutated in control variants, causing muscle and adipose tissue diseases respectively, are colored in green, whereas the 5 residues mutated in variants with a lower affinity for BAF are colored in cyan, purple and red (for weak binding, no measurable affinity and no detected binding, respectively). (B) Size exclusion chromatography of LamIgF mutants mixed to BAF (Superdex 75 10/300GL). Each gel corresponds to the elution of a different set of proteins, using the same column, the same concentrations and the same buffer: from bottom BAF, BAF and R435C, BAF and K542N, BAF and R527H, BAF and R471C, BAF and A529V, BAF and R482W, BAF and R453W and BAF and WT LamIgF. The bands corresponding to the BAF-LamIgF complex are boxed in red.

events but their impact was significantly lower than the impact of mutations associated with progeroid syndromes in cells (Figures 6C and 7).

DISCUSSION

The mechanical roles of lamins and their functions in gene regulation are often viewed as independent activities, but recent findings suggest a highly cross-linked and interdependent regulation of these different activities (51–53). Lamins interact with inner nuclear membrane proteins to mediate mechanical signaling but also contact chromatin-associated proteins that contribute to gene transcription regulation. Most of these interactions depend on the lamin ability to assemble into higher-order structures (54,55). Therefore, lamin assembly has been studied intensively *in vitro* and the resulting filaments have been observed in cells (11). In contrast, little is known about the contacts of lamins with their partners *in vivo*. Lamins are extensively modified and inter-

act with a large number of proteins (56,57), the inner nuclear membrane (58) and chromatin (59). Their tail region is responsible for association with partners. It contains a globular Igfold domain that is mutated in several laminopathies (13). Until this study, it was not known how this domain interacted with partners anchored at the inner nuclear membrane and established contacts with chromatin.

The lamin A/C Igfold domain directly binds to self-assembled emerin but also indirectly interacts with monomeric emerin through BAF

We showed that the lamin A/C Igfold domain interacts with the inner nuclear membrane protein emerin through two mechanisms: it either directly binds to self-assembled EmN or interacts indirectly with monomeric EmN through the chromatin-associated protein BAF. It was previously published that direct lamin A tail interaction with emerin is disrupted not by mutations in the LEM domain, but by

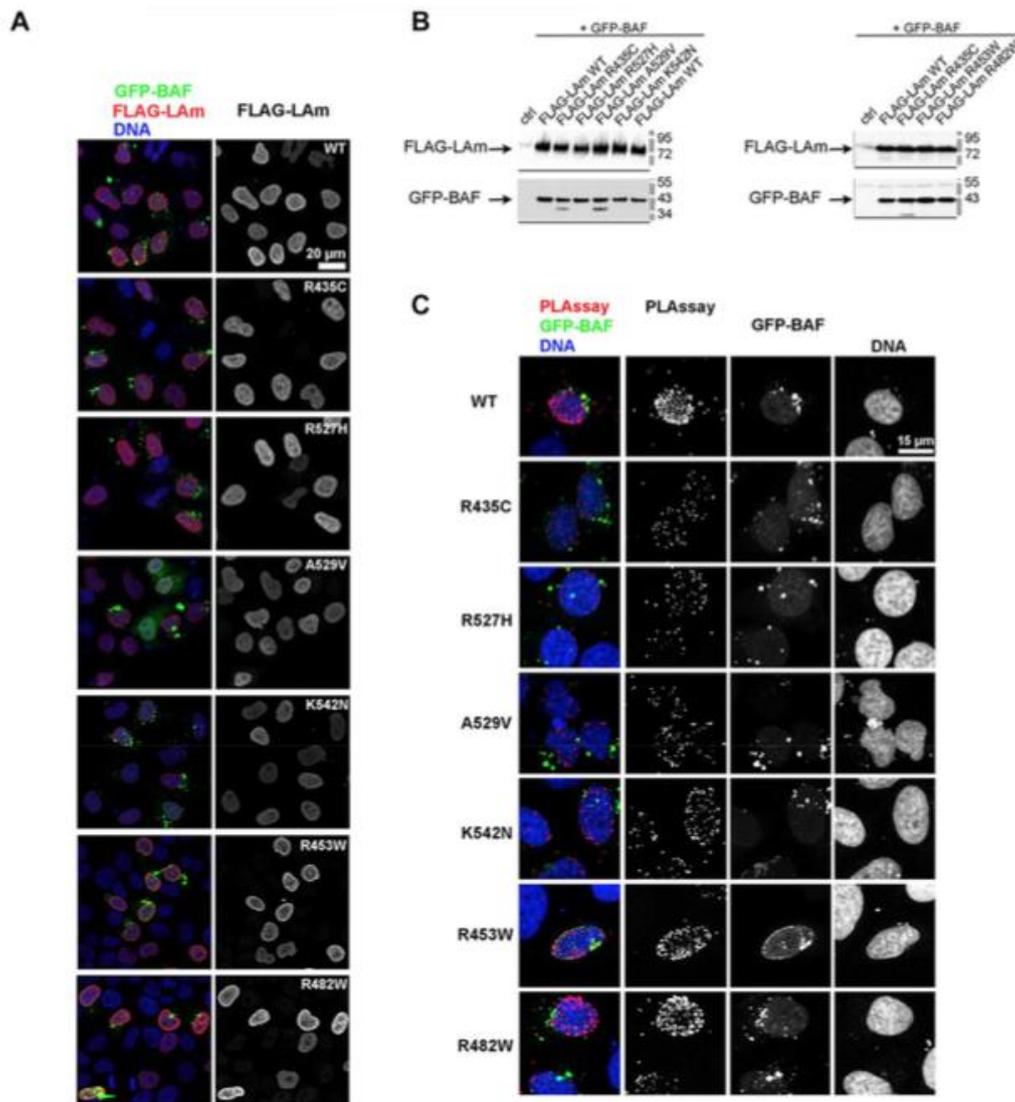


Figure 6. Mature lamin mutants with mutations causing recessive progeroid syndromes impair BAF binding in cells. **(A)** HeLa cells expressing GFP-BAF (WT) together with either WT, R435C, R527H, A529V, K542N, R453W or R482W FLAG-LAm were fixed, and processed for immunofluorescence using as primary antibodies, a mix of anti GFP and anti-FLAG antibodies. Observation was done at the confocal microscope. Are shown in the left panels, merges for GFP-BAF (green), FLAG-LAm (red) and DNA (blue), and in the right panels, the FLAG-LAm alone. Scale bar: 20 μ m. **(B)** Whole cell protein extracts prepared from HeLa cells control (ctrl) or expressing GFP-BAF (WT) together with either WT, R435C, R527H, A529V, K542N, R453W or R482W FLAG-LAm were analyzed by western blot using anti FLAG and anti BAF antibodies. **(C)** HeLa cells expressing GFP-BAF (WT) together with either WT, R435C, R527H, A529V, K542N, R453W or R482W FLAG-LAm were fixed, labeled with a mix of anti GFP and anti-FLAG antibodies and further processed for PLA before observation at the confocal microscope. The PLA signals (red) are shown merged with DNA (blue) and GFP-BAF (green) or alone (PLAssay). Signals for GFP-BAF and DNA are shown alone in the two right panels, respectively. Scale bar: 15 μ m.

mutations in emerin region from aa 70 to aa 164: mutations around positions 70, 76, 95, 112, as well as 141 and 164, decreased binding of lamin A to emerin (15,37). These results are consistent with our analysis. Indeed, the disordered region of emerin between aa 50 and aa 132 is necessary for emerin self-assembly (26) and LamIgF only binds directly to self-assembled emerin. Similarly, GST fusions of an emerin fragment from aa 1 to aa 132 is sufficient to bind mature lamin A tails (18), and we showed that emerin aa 1 to aa 132 is sufficient for self-assembly (26). Finally, the EDMD-associated emerin deletion mutation del195-99,

which impairs EmN self-assembly *in vitro*, causes a decrease of lamin A/C binding in cells (26). Altogether these results strongly suggest that emerin self-assembly is necessary for direct binding to lamin A/C in cells. On the other hand, we showed that the monomeric LEM domain (emerin aa 1 to aa 45) indirectly interacts with the lamin A/C Igfold domain through a dimer of BAF, consistent with previous data showing that BAF enhanced binding of prelamin A tails to emerin (24).

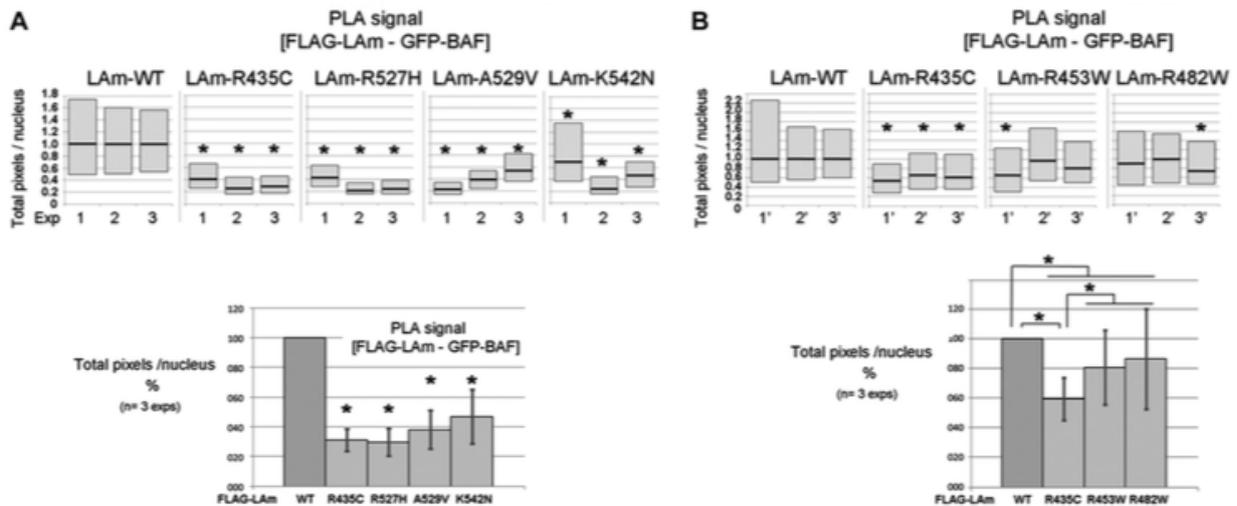


Figure 7. Quantification of the decrease in proximity events between mature lamin A and BAF due to mutations causing either recessive progeroid syndromes, EDMD or FPLD. Quantification of PLA signals per nucleus was performed as reported in (33). (A) Proximity events were compared between either WT lamin A and BAF, or lamin A variants associated with progeroid syndromes and BAF. (B) The impact of lamin mutations causing either a progeroid syndrome (R435C), EDMD (R453W) or FPLD (R482W) on the proximity between mature lamin A and BAF was measured. The upper graphs show the results of three independent experiments carried out with each mutant. The lower graphs represent the mean values \pm stdevpa ($n = 3$ experiments; $*P < 0.05$ with Kruskal and Wallis test).

A defect at the interface between lamin A/C and the chromatin-associated BAF protein is observed in atypical progeroid syndromes

By analyzing the 3D structure of our ternary complex comprising domains of lamin A/C, BAF and emerin, we noticed that the interface between lamin A/C and BAF coincides with a hot spot of lamin A/C residues mutated in progeroid syndromes (14). Proximities between lamin A/C and BAF were detected *in situ* using a proximity assay; they were particularly enriched at the nuclear periphery in HeLa cells (Supplementary Figure S5). The four tested mutations associated with progeroid syndromes reduced the frequency of proximity events between lamin A and BAF. This decrease was significantly larger than the decrease observed in the case of control mutations R453W and R482W associated with muscle and adipose tissue diseases, respectively: mutations associated with progeroid syndromes reduced the number of PLA signals to $\sim 50\%$ of the signals measured for WT lamin A and BAF. Consistently, *in vitro*, SEC experiments revealed that LamIgF mutants R471C, A529V, R527H and K542N weakly interact with BAF, and mutant R435C does not bind BAF. ITC experiments provided only estimations for affinity decreases, due to the weak affinities of the mutants. However, an affinity decrease of ~ 5 -fold could be calculated for mutants R471C and A529T, whereas essentially no binding could be detected for R527H, K542N and R435C. Interestingly, this affinity scale matches with the severity of the associated diseases: R471C and A529V cause MAD that generally appears during late childhood (44,49), whereas K542N causes a HGPS with an onset between 18 and 24 months (41) and R435C causes the most severe disease, RD, which appears during the first months after birth (39,40,42); only mutant R527H, which causes a MAD (45–48), also shows a strong BAF binding defect un-

der the conditions we used. We conclude that a destabilization of the interface between lamin A/C and BAF might mediate the disease-mechanism in atypical progeroid syndromes due to homozygous mutations in LamIgF.

Another progeroid syndrome called the Nestor-Guillermo Progeria Syndrome occurs in patients older than the average lifespan of progeroid patients. This syndrome is due to the homozygous BAF mutation A12T (38). The mutated protein is expressed and stable (60). In cells, it still co-immunoprecipitates with endogenous emerin, lamin A and histone H3. Our preliminary data show that, *in vitro*, mutation A12T does not compromise BAF 3D structure (Supplementary Figure S6A), and BAF mutant A12T interacts with LamIgF, but with an affinity lower than BAF WT and BAF S4E (Supplementary Figure S6B). Immunoprecipitation experiments performed in co-transfected cells also showed a reduced interaction between BAF mutant A12T and lamin A (5). Finally, BAF mutant A12T exhibited a marked defect in double-stranded DNA binding compared to WT BAF (60). Altogether, these data strongly suggest that the complex between BAF, lamin A/C, DNA and emerin is destabilized by the mutation A12T, highlighting the similarity between the mechanisms of autosomal recessive progeroid syndromes due to mutations in LamIgF and BAF.

Complexes including lamin A/C and emerin are assembled at different stages of the cell cycle

In cells, formation and localization of complexes including lamin A/C, BAF and emerin is highly regulated and depend on the cell cycle, in accord with the disassembly/reformation of the nuclear envelope that delimits the interphase nuclear compartment. We have identified two types of complexes including lamin A/C and

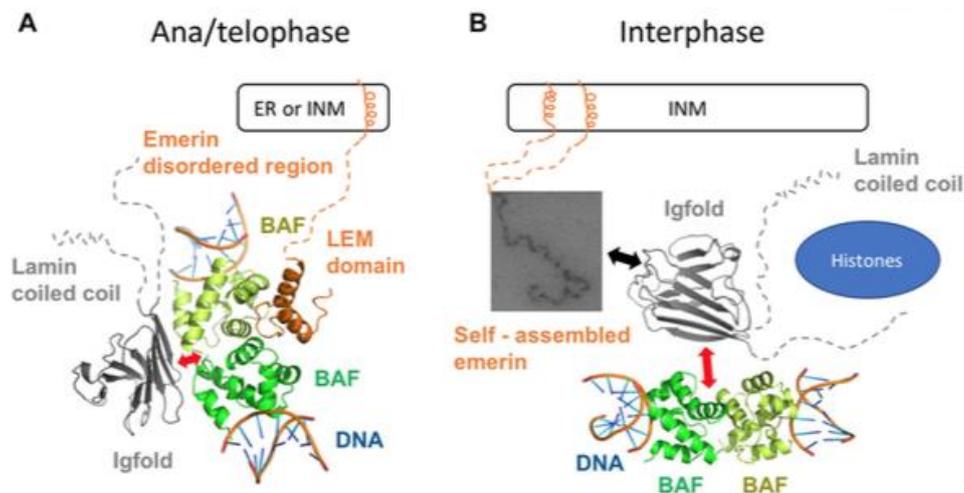


Figure 8. A first model of the interface between the LEM-domain proteins, the nucleoskeleton and the chromatin-associated protein BAF. (A) During the ana/telophase, BAF mediates an interaction between the lamin A/C Igfold domain and emerin anchored at the endoplasmic reticulum (ER) or at the inner nuclear membrane (INM). Superimposition of the X-ray structure of the LamIgF-BAF-LEM complex onto the X-ray structure of the BAF protein complexed to DNA (PDB reference 2BZF (62)) suggests that BAF can simultaneously interact with lamin A/C, emerin and DNA. The interface between lamin A/C and BAF is disrupted by mutations causing autosomal recessive progeroid syndromes, as symbolized by the red arrow. (B) In interphase cells, the lamin A/C Igfold domain binds self-assembled emerin anchored at the inner nuclear membrane (INM), as symbolized by the black arrow. It is also able to interact with BAF and histones, thus creating an interface between the lamina and chromatin. This interface is disrupted by mutations causing autosomal recessive progeroid syndromes, as symbolized by the red arrow.

emerin, however the biological relevance of these two modes of interaction is still unclear. It is known that at the end of mitosis, in ana-telophase, first BAF, and then emerin and lamin A/C accumulate at the core region of chromosomes (61). On the BAF side, mutation G25E, at the interface with DNA, impairs localization of BAF at the core region of chromosomes, emerin localization during telophase and further emerin and lamin A accumulation at reforming nuclear envelopes. On the emerin side, the LEM domain is necessary and sufficient for the core localization of emerin. Superimposition of our 3D structure with that of a complex between BAF dimer and DNA (PDB reference 2BZF (62)) revealed that the BAF dimer can simultaneously bind to emerin, lamin and DNA (Figure 8, left panel). We propose that BAF mediates the interaction between emerin and lamin A at the core region.

In interphase cells, lamin A/C is assembled into polymers that bind to emerin (37). This interaction might not require a well-folded LEM domain. Indeed, an emerin protein mutated in its LEM domain (residues 24–27 being changed into alanines), which does not interact with BAF (15), localizes normally at the nuclear envelope (63), consistent with its ability to directly bind lamin A/C *in vitro* (15). We suggest that lamin A/C directly interacts with emerin that is self-assembled at the nuclear envelope (18,64) (Figure 8, right panel). During *in vitro* self-assembly, emerin exhibits a large LEM domain conformational change, strongly suggesting that self-assembled emerin does not bind BAF (34). However, we still detect proximities between lamin A/C and BAF at the nuclear periphery of interphase cells (Supplementary Figure S5).

In conclusion, the Igfold domain of lamin A/C can both directly interact with self-assembled emerin and bind to

monomeric emerin through the chromatin-associated protein BAF. A defect in the interaction between lamin A/C and BAF might lead to a less efficient targeting of emerin at the reforming nuclear envelope in telophase. However, diseases associated with mutations disrupting the lamin A/C–BAF interaction are much more severe than diseases associated with emerin loss of function. This suggests that the impact of these lamin A/C and BAF mutations goes beyond LEM-domain protein mislocalization and loss of function. We propose that the autosomal recessive progeroid syndrome-causing mutations perturb the BAF-mediated interaction between lamin A/C and chromatin, thus leading to a dysregulation of chromatin organization and gene expression. Consistently, an abnormal nuclear envelope–chromatin interface was previously described using electron microscopy in the case of progeroid syndromes caused by lamin A or BAF mutations (5,65–67). More work is needed to understand how the lamin–BAF complex interacts with DNA and histones, and how a defect in this complex compromises the interactions between lamina and chromatin at the nuclear periphery in autosomal recessive progeroid syndromes.

DATA AVAILABILITY

Structure coordinates were deposited to the PDB, with entry 6GHD.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. H.J. Worman for providing the plasmid coding for GST-LamIgf and reviewing the manuscript and Prof. G. Lattanzi for providing the plasmid coding for GFP-BAF. We also thank N. Vadrot for her help with cell cultures.

FUNDING

Commissariat à l'énergie atomique et aux énergies alternatives, Centre National de la Recherche Scientifique, University Paris South and University Paris Diderot; French Infrastructure for Integrated Structural Biology (<https://www.structuralbiology.eu/networks/frisbi>) [ANR-10-INSB-05-01, Acronym FRISBI]; French Association against Myopathies (AFM) [17243 and 20018 to S.Z.-J. and PhD fellowship number 18159 to C.S.]. Funding for open access charge: Grant 20018 (French Association against Myopathies).

Conflict of interest statement. None declared.

REFERENCES

- Worman, H.J. (2012) Nuclear lamins and laminopathies. *J. Pathol.*, **226**, 316–325.
- Worman, H.J. and Bonne, G. (2007) "Laminopathies": a wide spectrum of human diseases. *Exp. Cell Res.*, **313**, 2121–2133.
- Gruenbaum, Y. and Medalia, O. (2015) Lamins: the structure and protein complexes. *Curr. Opin. Cell Biol.*, **32**, 7–12.
- Robin, J.D. and Magdiner, F. (2016) Physiological and pathological aging affects chromatin dynamics, structure and function at the nuclear edge. *Front. Genet.*, **7**, 153.
- Loi, M., Cenni, V., Duchi, S., Squarzone, S., Lopez-Otin, C., Foisner, R., Lattanzi, G. and Capanni, C. (2016) Barrier-to-autointegration factor (BAF) involvement in prelamin A-related chromatin organization changes. *Oncotarget*, **7**, 15662–15677.
- Burke, B. and Gerace, L. (1986) A cell free system to study reassembly of the nuclear envelope at the end of mitosis. *Cell*, **44**, 639–652.
- Glass, J.R. and Gerace, L. (1990) Lamins A and C bind and assemble at the surface of mitotic chromosomes. *J. Cell Biol.*, **111**, 1047–1057.
- Taniura, H., Glass, C. and Gerace, L. (1995) A chromatin binding site in the tail domain of nuclear lamins that interacts with core histones. *J. Cell Biol.*, **131**, 33–44.
- Lopez-Soler, R.I., Moir, R.D., Spann, T.P., Stick, R. and Goldman, R.D. (2001) A role for nuclear lamins in nuclear envelope assembly. *J. Cell Biol.*, **154**, 61–70.
- Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L.K., Forster, F., Hyman, A.A., Plitzko, J.M. and Baumeister, W. (2016) Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science*, **351**, 969–972.
- Turgay, Y., Eibauer, M., Goldman, A.E., Shimi, T., Khayat, M., Ben-Harush, K., Dubrovsky-Gaupp, A., Sapra, K.T., Goldman, R.D. and Medalia, O. (2017) The molecular architecture of lamins in somatic cells. *Nature*, **543**, 261–264.
- Dhe-Paganon, S., Werner, E.D., Chi, Y.I. and Shoelson, S.E. (2002) Structure of the globular tail of nuclear lamin. *J. Biol. Chem.*, **277**, 17381–17384.
- Krimm, I., Ostlund, C., Gilquin, B., Couprie, J., Hossenlopp, P., Mornon, J.P., Bonne, G., Courvalin, J.C., Worman, H.J. and Zinn-Justin, S. (2002) The Ig-like structure of the C-terminal domain of lamin A/C, mutated in muscular dystrophies, cardiomyopathy, and partial lipodystrophy. *Structure*, **10**, 811–823.
- Verstraeten, V.L., Broers, J.L., van Steensel, M.A., Zinn-Justin, S., Ramaekers, F.C., Steijlen, P.M., Kamps, M., Kuijpers, H.J., Merckx, D., Smets, H.J. et al. (2006) Compound heterozygosity for mutations in LMNA causes a progeria syndrome without prelamin A accumulation. *Hum. Mol. Genet.*, **15**, 2509–2522.
- Lee, K.K., Haraguchi, T., Lee, R.S., Koujin, T., Hiraoka, Y. and Wilson, K.L. (2001) Distinct functional domains in emerin bind lamin A and DNA-bridging protein BAF. *J. Cell Sci.*, **114**, 4567–4573.
- Vaughan, A., Alvarez-Reyes, M., Bridger, J.M., Broers, J.L., Ramaekers, F.C., Wehnert, M., Morris, G.E., Whitfield, W.G.F. and Hutchison, C.J. (2001) Both emerin and lamin C depend on lamin A for localization at the nuclear envelope. *J. Cell Sci.*, **114**, 2577–2590.
- Sakaki, M., Koike, H., Takahashi, N., Sasagawa, N., Tomioka, S., Arahata, K. and Ishiura, S. (2001) Interaction between emerin and nuclear lamins. *J. Biochem. (Tokyo)*, **129**, 321–327.
- Berk, J.M., Simon, D.N., Jenkins-Houk, C.R., Westerbeck, J.W., Gronning-Wang, L.M., Carlson, C.R. and Wilson, K.L. (2014) The molecular basis of emerin-emerin and emerin-BAF interactions. *J. Cell Sci.*, **127**, 3956–3969.
- Cai, M., Huang, Y., Suh, J.Y., Louis, J.M., Ghirlando, R., Craigie, R. and Clore, G.M. (2007) Solution NMR structure of the barrier-to-autointegration factor-Emerin complex. *J. Biol. Chem.*, **282**, 14525–14535.
- Qi, R., Xu, N., Wang, G., Ren, H., Li, S., Lei, J., Lin, Q., Wang, L., Gu, X., Zhang, H. et al. (2015) The lamin-A/C-LAP2alpha-BAF1 protein complex regulates mitotic spindle assembly and positioning. *J. Cell Sci.*, **128**, 2830–2841.
- Liu, J., Lee, K.K., Segura-Totten, M., Neufeld, E., Wilson, K.L. and Gruenbaum, Y. (2003) MAN1 and emerin have overlapping function(s) essential for chromosome segregation and cell division in *Caenorhabditis elegans*. *PNAS*, **100**, 4598–4603.
- Margalit, A., Segura-Totten, M., Gruenbaum, Y. and Wilson, K.L. (2005) Barrier-to-autointegration factor is required to segregate and enclose chromosomes within the nuclear envelope and assemble the nuclear lamina. *PNAS*, **102**, 3290–3295.
- Holaska, J.M., Lee, K.K., Kowalski, A.K. and Wilson, K.L. (2003) Transcriptional repressor germ cell-less (GCL) and barrier to autointegration factor (BAF) compete for binding to emerin in vitro. *J. Biol. Chem.*, **278**, 6969–6975.
- Bengtsson, L. and Wilson, K.L. (2006) Barrier-to-autointegration factor phosphorylation on Ser-4 regulates emerin binding to lamin A in vitro and emerin localization in vivo. *Mol. Biol. Cell*, **17**, 1154–1163.
- Samwer, M., Schneider, M.W.G., Hoefler, R., Schmalhorst, P.S., Jude, J.G., Zuber, J. and Gerlich, D.W. (2017) DNA Cross-Bridging shapes a single nucleus from a set of mitotic chromosomes. *Cell*, **170**, 956–972.
- Herrada, I., Samson, C., Velours, C., Renault, L., Ostlund, C., Chervy, P., Puchkov, D., Worman, H.J., Buendia, B. and Zinn-Justin, S. (2015) Muscular dystrophy mutations impair the nuclear envelope emerin self-assembly properties. *ACS Chem. Biol.*, **10**, 2733–2742.
- Krimm, I., Couprie, J., Ostlund, C., Worman, H.J. and Zinn-Justin, S. (2002) 1H, 13C and 15N resonance assignments of the C-terminal domain of human lamin A/C. *J. Biomol. NMR*, **22**, 371–372.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D, Biol. Crystallogr.*, **67**, 235–242.
- Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.*, **30**, 1022–1025.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 213–221.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 486–501.
- Vadrot, N., Duband-Goulet, I., Cabet, E., Attanda, W., Barateau, A., Vicart, P., Gerbal, F., Briand, N., Vigouroux, C., Oldenburg, A.R. et al. (2015) The p.R482W substitution in A-type lamins deregulates SREBP1 activity in Dunnigan-type familial partial lipodystrophy. *Hum. Mol. Genet.*, **24**, 2096–2109.
- Samson, C., Celli, F., Hendriks, K., Zinke, M., Essawy, N., Herrada, I., Arteni, A.A., Theillet, F.X., Alpha-Bazin, B., Armengaud, J. et al.

- (2017) Emerin self-assembly mechanism: role of the LEM domain. *FEBS J.* **284**, 338–352.
35. Samson, C., Herrada, I., Celli, F., Theillet, F.X. and Zinn-Justin, S. (2016) ¹H, ¹³C and ¹⁵N backbone resonance assignment of the intrinsically disordered region of the nuclear envelope protein emerin. *Biomol. NMR Assign.* **10**, 179–182.
 36. Margalit, A., Brachner, A., Gotzmann, J., Foisner, R. and Gruenbaum, Y. (2007) Barrier-to-autointegration factor—a BAFfling little protein. *Trends Cell Biol.* **17**, 202–208.
 37. Berk, J.M., Tift, K.E. and Wilson, K.L. (2013) The nuclear envelope LEM-domain protein emerin. *Nucleus*, **4**, 298–314.
 38. Puente, X.S., Quesada, V., Osorio, F.G., Cabanillas, R., Cadinanos, J., Fraile, J.M., Ordóñez, G.R., Puente, D.A., Gutierrez-Fernandez, A., Fanjul-Fernandez, M. *et al.* (2011) Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome. *Am. J. Hum. Genet.* **88**, 650–656.
 39. Madej-Pilarczyk, A., Rosinska-Borkowska, D., Rekawek, J., Marchel, M., Szalus, E., Jablonska, S. and Hausmanowa-Petruśiewicz, I. (2009) Progeroid syndrome with scleroderma-like skin changes associated with homozygous R435C LMNA mutation. *Am. J. Med. Genet. A*, **149A**, 2387–2392.
 40. Youn, G.J., Uzunyan, M., Vachon, L., Johnson, J., Winder, T.L. and Yano, S. (2010) Autosomal recessive LMNA mutation causing restrictive dermopathy. *Clin. Genet.* **78**, 199–200.
 41. Plasilova, M., Chattopadhyay, C., Pal, P., Schaub, N.A., Buechner, S.A., Mueller, H., Miny, P., Ghosh, A. and Heinemann, K. (2004) Homozygous missense mutation in the lamin A/C gene causes autosomal recessive Hutchinson-Gilford progeria syndrome. *J. Med. Genet.* **41**, 609–614.
 42. Starke, S., Meinke, P., Camozzi, D., Mattioli, E., Pfaeffle, R., Siekmeyer, M., Hirsch, W., Horn, L.C., Paasch, U., Mitter, D. *et al.* (2013) Progeroid laminopathy with restrictive dermopathy-like features caused by an isodisomic LMNA mutation p.R435C. *Aging (Albany NY)*, **5**, 445–459.
 43. Bai, S., Lozada, A., Jones, M.C., Dietz, H.C., Dempsey, M. and Das, S. (2014) Mandibuloacral dysplasia caused by LMNA mutations and uniparental disomy. *Case Rep. Genet.* **2014**, 508231.
 44. Zirn, B., Kress, W., Grimm, T., Berthold, L.D., Neubauer, B., Kuchelmeister, K., Müller, U. and Hahn, A. (2008) Association of homozygous LMNA mutation R471C with new phenotype: mandibuloacral dysplasia, progeria, and rigid spine muscular dystrophy. *Am. J. Med. Genet. A*, **146A**, 1049–1054.
 45. Novelli, G., Muchir, A., Sangiuolo, F., Helbling-Leclerc, A., D'Apice, M.R., Massart, C., Capon, F., Sbraccia, P., Federici, M., Lauro, R. *et al.* (2002) Mandibuloacral dysplasia is caused by a mutation in LMNA-encoding lamin A/C. *Am. J. Hum. Genet.* **71**, 426–431.
 46. Simha, V., Agarwal, A.K., Oral, E.A., Fryns, J.P. and Garg, A. (2003) Genetic and phenotypic heterogeneity in patients with mandibuloacral dysplasia-associated lipodystrophy. *J. Clin. Endocrinol. Metab.* **88**, 2821–2824.
 47. Shen, J.J., Brown, C.A., Lupski, J.R. and Potocki, L. (2003) Mandibuloacral dysplasia caused by homozygosity for the R527H mutation in lamin A/C. *J. Med. Genet.* **40**, 854–857.
 48. Garavelli, L., D'Apice, M.R., Rivieri, F., Bertoli, M., Wischmeijer, A., Gelmini, C., De Nigris, V., Albertini, E., Rosato, S., Virdis, R. *et al.* (2009) Mandibuloacral dysplasia type A in childhood. *Am. J. Med. Genet. A*, **149A**, 2258–2264.
 49. Garg, A., Cogulu, O., Ozkinay, F., Onay, H. and Agarwal, A.K. (2005) A novel homozygous Ala529Val LMNA mutation in Turkish patients with mandibuloacral dysplasia. *J. Clin. Endocrinol. Metab.* **90**, 5259–5264.
 50. Magracheva, E., Kozlov, S., Stewart, C.L., Wlodawer, A. and Zdanov, A. (2009) Structure of the lamin A/C R482W mutant responsible for dominant familial partial lipodystrophy (FPLD). *Acta Crystallogr. F, Struct. Biol. Cryst. Commun.* **65**, 665–670.
 51. Ho, C.Y., Jaalouk, D.E., Vartiainen, M.K. and Lammerding, J. (2013) Lamin A/C and emerin regulate MKL1-SRF activity by modulating actin dynamics. *Nature*, **497**, 507–511.
 52. Cho, S., Irianto, J. and Discher, D.E. (2017) Mechanosensing by the nucleus: From pathways to scaling relationships. *J. Cell Biol.* **216**, 305–315.
 53. Osmanagic-Myers, S., Dechat, T. and Foisner, R. (2015) Lamins at the crossroads of mechanosignaling. *Genes Dev.* **29**, 225–237.
 54. Strelkov, S.V., Herrmann, H., Geisler, N., Lustig, A., Ivaninskii, S., Zimbelmann, R., Burkhard, P. and Aebi, U. (2001) Divide-and-conquer crystallographic approach towards an atomic structure of intermediate filaments. *J. Mol. Biol.* **306**, 773–781.
 55. Herrmann, H., Strelkov, S.V., Burkhard, P. and Aebi, U. (2009) Intermediate filaments: primary determinants of cell architecture and plasticity. *J. Clin. Invest.* **119**, 1772–1783.
 56. Dörner, D., Gotzmann, J. and Foisner, R. (2007) Nucleoplasmic lamins and their interaction partners, LAP2alpha, Rb, and BAF, in transcriptional regulation. *FEBS J.* **274**, 1362–1373.
 57. Simon, D.N. and Wilson, K.L. (2013) Partners and post-translational modifications of nuclear lamins. *Chromosoma*, **122**, 13–31.
 58. King, M.C. and Lusk, C.P. (2016) A model for coordinating nuclear mechanics and membrane remodeling to support nuclear integrity. *Curr. Opin. Cell Biol.* **41**, 9–17.
 59. Camozzi, D., Capanni, C., Cenni, V., Mattioli, E., Columbaro, M., Squarzone, S. and Lattanzi, G. (2014) Diverse lamin-dependent mechanisms interact to control chromatin dynamics. Focus on laminopathies. *Nucleus*, **5**, 427–440.
 60. Paquet, N., Box, J.K., Ashton, N.W., Suraweera, A., Croft, L.V., Urquhart, A.J., Bolderson, E., Zhang, S.D., O'Byrne, K.J. and Richard, D.J. (2014) Nestor-Guillermo Progeria Syndrome: a biochemical insight into Barrier-to-Autointegration Factor 1, alanine 12 threonine mutation. *BMC Mol. Biol.* **15**, 27.
 61. Haraguchi, T., Kojidani, T., Koujin, T., Shimi, T., Osakada, H., Mori, C., Yamamoto, A. and Hiraoka, Y. (2008) Live cell imaging and electron microscopy reveal dynamic processes of BAF-directed nuclear envelope assembly. *J. Cell Sci.* **121**, 2540–2554.
 62. Bradley, C.M., Ronning, D.R., Ghirlando, R., Craigie, R. and Dyda, F. (2005) Structural basis for DNA bridging by barrier-to-autointegration factor. *Nat. Struct. Mol. Biol.* **12**, 935–936.
 63. Haraguchi, T., Koujin, T., Segura-Totten, M., Lee, K.K., Matsuoka, Y., Yoneda, Y., Wilson, K.L. and Hiraoka, Y. (2001) BAF is required for emerin assembly into the reforming nuclear envelope. *J. Cell Sci.* **114**, 4575–4585.
 64. Barton, L.J., Soshnev, A.A. and Geyer, P.K. (2015) Networking in the nucleus: a spotlight on LEM-domain proteins. *Curr. Opin. Cell Biol.* **34**, 1–8.
 65. Columbaro, M., Capanni, C., Mattioli, E., Novelli, G., Parnaik, V.K., Squarzone, S., Maraldi, N.M. and Lattanzi, G. (2005) Rescue of heterochromatin organization in Hutchinson-Gilford progeria by drug treatment. *Cell. Mol. Life Sci.* **62**, 2669–2678.
 66. Shumaker, D.K., Dechat, T., Kohlmaier, A., Adam, S.A., Bozovsky, M.R., Erdos, M.R., Eriksson, M., Goldman, A.E., Khuron, S., Collins, F.S. *et al.* (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *PNAS*, **103**, 8703–8708.
 67. Scaffidi, P. and Misteli, T. (2005) Reversal of the cellular phenotype in the premature aging disease Hutchinson-Gilford progeria syndrome. *Nat. Med.* **11**, 440–445.

Emerin self-assembly mechanism: role of the LEM domain

Camille Samson¹, Florian Celli¹, Kitty Hendriks², Maximilian Zinke², Nada Essawy³,
Isaline Herrada¹, Ana-Andreea Arteni⁴, François-Xavier Theillet¹, Béatrice Alpha-Bazin⁵,
Jean Armengaud⁵, Catherine Coirault³, Adam Lange^{2,6} and Sophie Zinn-Justin¹

¹ Laboratory of Structural Biology and Radiobiology, Institute for Integrative Biology of the Cell (CEA, CNRS, University Paris South), University Paris-Saclay, Gif-sur-Yvette, France

² Department of Molecular Biophysics, Leibniz-Institut für Molekulare Pharmakologie, Berlin, Germany

³ Center for Research in Myology (INSERM, CNRS), Université Pierre et Marie Curie Paris 06, Sorbonne Universités, France

⁴ Department of Structural Virology, Institute for Integrative Biology of the Cell (CEA, CNRS, University Paris South), University Paris-Saclay, Gif-sur-Yvette, France

⁵ Laboratory 'Innovative technologies for Detection and Diagnostics', Institute of Biology and Technology Saclay, CEA, Bagnols-sur-Côze, France

⁶ Institut für Biologie, Humboldt-Universität zu Berlin, Germany

Keywords

folding; intrinsically disordered region; lamin; nuclear envelope; nucleoskeleton; oligomerization

Correspondence

S. Zinn-Justin, Laboratory of Structural Biology and Radiobiology, Institute for Integrative Biology of the Cell (CEA, CNRS, University Paris South), University Paris-Saclay, Bât 144, 91191 Gif-sur-Yvette Cedex, France
Fax: +33 0 169084712
Tel: +33 0 169083026
E-mail: Sophie.zinn@cea.fr

(Received 22 September 2016, revised 18 November 2016, accepted 5 December 2016)

doi:10.1111/febs.13983

At the nuclear envelope, the inner nuclear membrane protein emerin contributes to the interface between the nucleoskeleton and the chromatin. Emerin is an essential actor of the nuclear response to a mechanical signal. Genetic defects in emerin cause Emery–Dreifuss muscular dystrophy. It was proposed that emerin oligomerization regulates nucleoskeleton binding, and impaired oligomerization contributes to the loss of function of emerin disease-causing mutants. We here report the first structural characterization of emerin oligomers. We identified an N-terminal emerin region from amino acid 1 to amino acid 132 that is necessary and sufficient for formation of long curvilinear filaments. In emerin monomer, this region contains a globular LEM domain and a fragment that is intrinsically disordered. Solid-state nuclear magnetic resonance analysis identifies the LEM β -fragment as part of the oligomeric structural core. However, the LEM domain alone does not self-assemble into filaments. Additional residues forming a β -structure are observed within the filaments that could correspond to the unstructured region in emerin monomer. We show that the delK37 mutation causing muscular dystrophy triggers LEM domain unfolding and increases emerin self-assembly rate. Similarly, inserting a disulfide bridge that stabilizes the LEM folded state impairs emerin N-terminal region self-assembly, whereas reducing this disulfide bridge triggers self-assembly. We conclude that the LEM domain, responsible for binding to the chromatin protein BAF, undergoes a conformational change during self-assembly of emerin N-terminal region. The consequences of these structural rearrangement and self-assembly events on emerin binding properties are discussed.

Abbreviations

BAF, barrier-to-autointegration factor; DTT, dithiothreitol; EDMD, Emery–Dreifuss muscular dystrophy; LEM, Lap2-emerin-Man1; LINC, linker of nucleoskeleton and cytoskeleton; NMR, nuclear magnetic resonance; SDS/PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; ThT, thioflavin T.

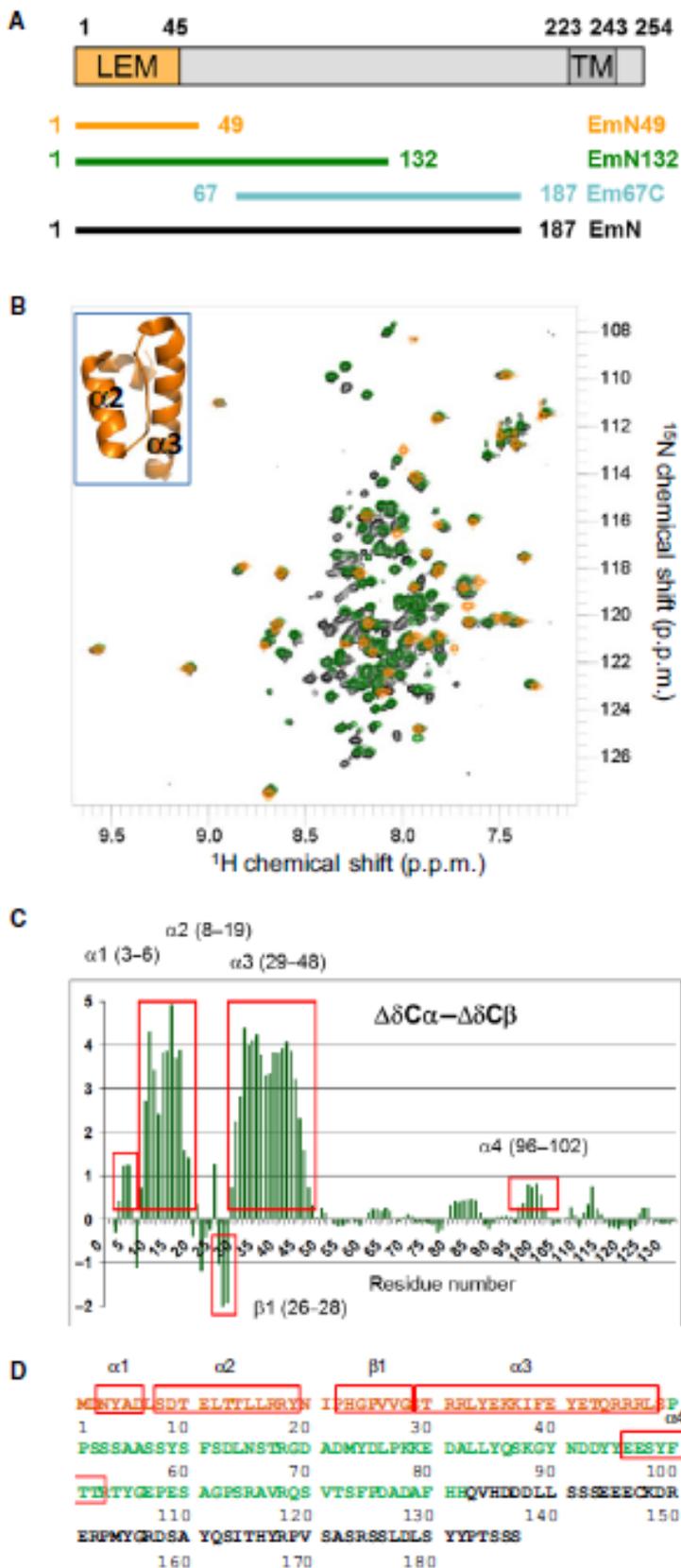


Fig. 1. A LEM domain with a $\alpha\alpha\beta\alpha$ motif is present in the monomeric emerin fragments EmN49, EmN132, and EmN. (A) Emerin architecture and fragments studied by solution NMR. (B) Superimposition of the solution-state NMR ^1H - ^{15}N HSQC spectra of EmN49 [the LEM domain, in orange; its 3D structure is boxed in the upper left corner of the figure (PDB references 1JE1 and 2ODC)], EmN132 (in green), and EmN (in black), recorded at pH 6.5, 303 K and 750 MHz. Peak overlap reveals that the LEM domain adopts the same 3D structure in the three fragments. (C) Analysis of the solution-state NMR ^{13}C signals of EmN132 (pH 6.5, 303 K, 750 MHz), confirming that its N-terminal domain exhibits three helices and one extended strand, and further showing that region 50–132 is essentially unstructured. Analysis of the NMR ^{13}C signals of emerin fragment 67–170 consistently showed that this fragment is intrinsically disordered (Samson *et al.* [29]). (D) Sequence of the emerin fragment EmN. The secondary structure elements deduced from (C) are boxed in red.

comprising residues 1–45 of emerin, plays a critical role in EmN oligomerization, and that the small in-frame deletion delK37 significantly favors emerin

self-assembly. On these bases, oligomerization mechanisms and their functional consequences are discussed.

Results

The N-terminal LEM domain is the only globular domain observed in the different monomeric emerlin fragments

We have previously proposed that the monomeric emerlin fragment from residue 1 to residue 187 (named EmN; see Fig. 1A) is composed of a globular LEM domain, from residue 1 to residue 45 [28], and a large intrinsically disordered region, from residue 46 to residue 187 [25,29]. EmN is also capable of oligomerizing into filaments rich in β -structure. In order to identify the minimal region of EmN that still self-associates into filaments, we produced three different recombinant fragments of EmN called EmN49, EmN132, and Em67C, corresponding to the LEM domain [28], the smallest N-terminal EmN fragment-binding lamins [24], and the smallest C-terminal EmN fragment containing all residues whose mutation impairs lamin A binding [30], respectively (Fig. 1A). Superimposition of the solution-state NMR ^1H - ^{15}N HSQC spectra of fragments EmN49, EmN132, and EmN confirmed that the LEM domain fold of EmN49 [28,31] is also present in monomeric EmN132 and EmN (Fig. 1B). Significant differences are found between the chemical shifts of the last four residues of helix $\alpha 3$ in EmN132 or EmN as compared to EmN49. This is most probably due to edge effects related to the C-terminal extension in EmN132 and EmN, which is absent in EmN49. It may highlight a moderate difference in the stability of the helix $\alpha 3$ C terminus. Analysis of the EmN132 NMR carbon signals consistently established that this fragment contains N-terminal helix $\alpha 1$ (3–6), helix $\alpha 2$ (8–19), strand $\beta 1$ (22–28, with a bulge at Pro25), and helix $\alpha 3$ (29–48) (Fig. 1C,D). It also contains an additional helix $\alpha 4$ (96–102) that is, as helix $\alpha 1$, only transiently observed: the corresponding region exists both as an α -helix and a disordered peptide. Our recent NMR analysis of the intrinsically disordered emerlin fragment 67–170 also demonstrated α -helical propensities between residue 98 and residue 101 [29].

The LEM domain and the intrinsically disordered region are both needed to form filaments

We tested the self-assembly capacities of EmN49, EmN132, and Em67C. Therefore, we incubated the different fragments at a fixed concentration of 600 μM during 1 week at 293 K. We then observed the samples using negative staining electron microscopy. Figure 2A shows that long curvilinear filaments were detected for EmN, as reported earlier [25], and also for EmN132. No

filaments could be identified for EmN49 and Em67C. Consistently, only EmN and EmN132 were capable of forming self-assembled particles interacting with thioflavin T, as detected by fluorescence on samples incubated at a concentration of 600 μM at 310 K (Fig. 2B). Thus, both region 1–66, mostly restricted to the LEM domain, and region 50–132, which is intrinsically disordered, are necessary for filament assembly. This result is consistent with our previous study showing that residues from 95 to 99, in the intrinsically disordered region, are necessary for EmN self-assembly [25].

EDMD causing small in-frame deletion delK37 impairs folding of the LEM domain

The emerlin mutation delK37 was identified in a patient with an EDMD phenotype characterized by a strong cardiac dysfunction [18]. Although the previously reported missense mutations or small in-frame deletions all corresponded to defects in emerlin intrinsically disordered region, this mutation targets a LEM domain residue. We used this mutation in order to probe the role of the LEM domain in emerlin self-assembly. Solution-state NMR 1D ^1H spectrum analysis of EmN mutant delK37 showed that this protein is completely unfolded (Fig. 3A). Superimposition of the solution-state NMR 2D ^1H - ^{15}N HSQC spectra of fragments of EmN, wild-type and mutated, revealed that the peaks corresponding to folded LEM residues (in orange on Fig. 1B) are absent in the case of the mutated EmN (Fig. 3B). The NMR signals corresponding to protons of side chains buried in the LEM hydrophobic core, which are probes for domain folding, are absent in the 1D ^1H spectrum of the EmN mutant delK37 (Fig. 3A,C). Thus, deletion of residue K37 in helix $\alpha 3$ impairs folding of the whole LEM domain.

LEM domain unfolding is associated with faster self-assembly in the case of the delK37 mutant

In order to probe for the role of the LEM domain in EmN self-assembly, we incubated in parallel EmN wild-type and delK37 at a concentration of 600 μM during 1 week at 293 K. Negative staining electron microscopy revealed that EmN mutant delK37 has also formed long filaments after 1 week (Fig. 4A), and SDS/PAGE analysis of chymotrypsin digested proteins validated that both wild-type and mutated EmN are partially protected from degradation when they are self-assembled (Fig. 4B). However, EmN mutant delK37 has already formed short filaments at $t = 0$ (Fig. 4A). Consistently, following protein self-assembly at 310 K and 300 μM using thioflavin T as a probe

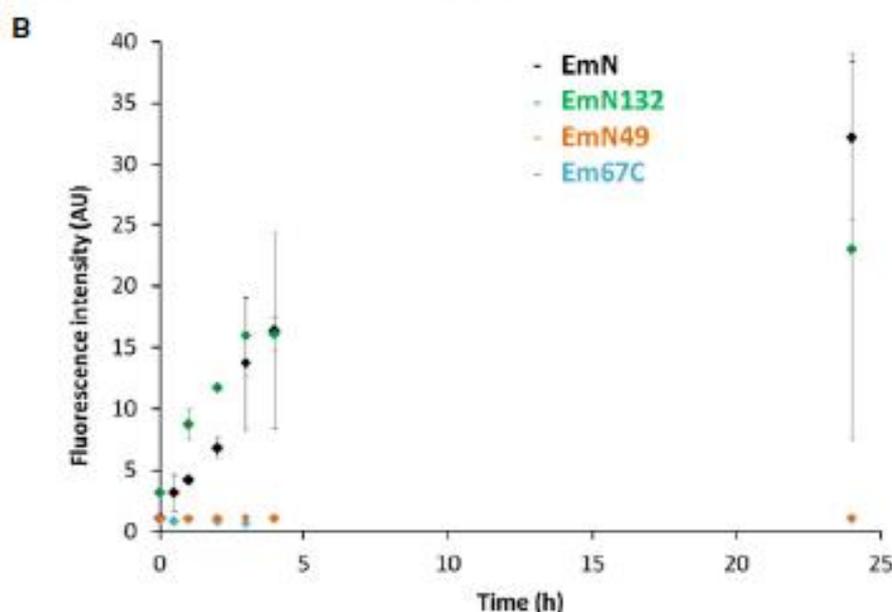
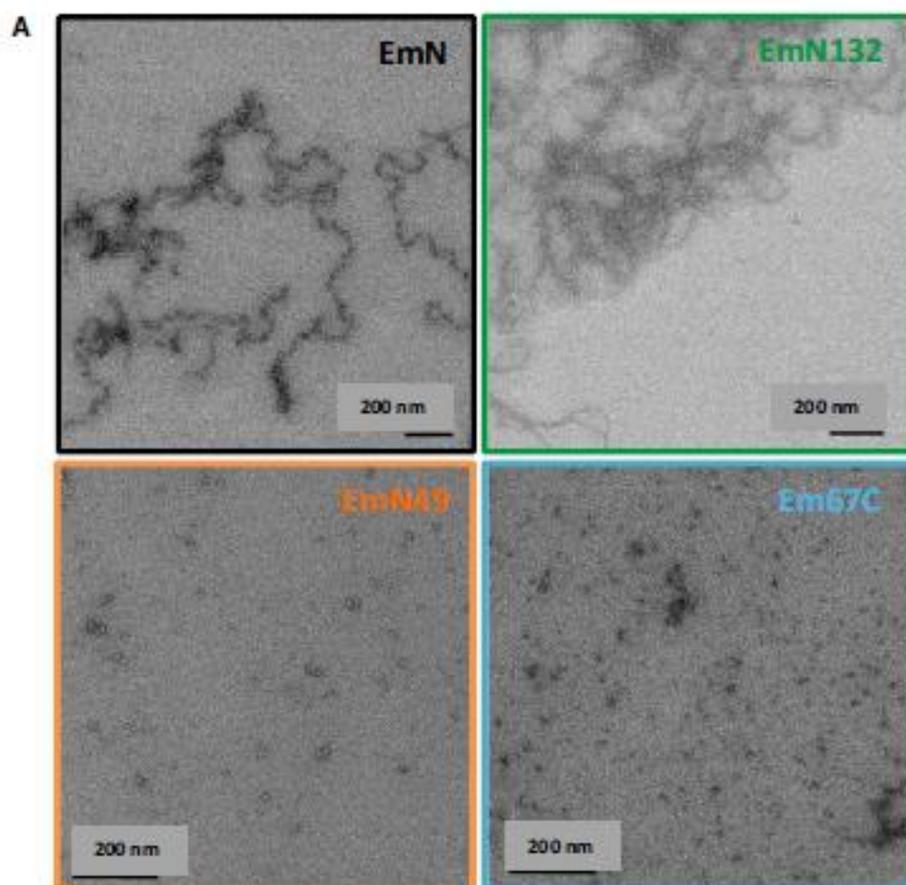


Fig. 2 The LEM sequence and part of the unstructured region are necessary for emerlin self-assembly. (A) Negative staining EM images of emerlin fragments EmN, EmN132, EmN49, and Em67C concentrated at 600 μM and incubated at 293 K during 1 week. (B) ThT fluorescence as a function of the incubation time at 310 K, measured on emerlin fragments concentrated at 300 μM ($n = 2$; error bars correspond to SD). Only EmN132 and EmN form filaments. They self-assemble with similar kinetics. EmN49 and Em67C do not form filaments, as also observed by EM.

demonstrated that the EmN mutant delK37 triggers significantly faster fluorescence increase as compared to wild-type EmN (Fig. 4C). As a control we observed that the LEM domain delK37 alone does not provoke any thioflavin fluorescence increase with time, showing that a destabilized LEM domain does not *per se* bind thioflavin T through its exposed hydrophobic core in

the absence of filament formation. We concluded that the intrinsically disordered EmN mutant delK37 forms oligomers faster than wild-type EmN.

To further probe the relationship between LEM domain folding and oligomerization kinetics, we introduced a disulfide bridge into the LEM domain of EmN132 (which contains no cysteine) by mutating Y4

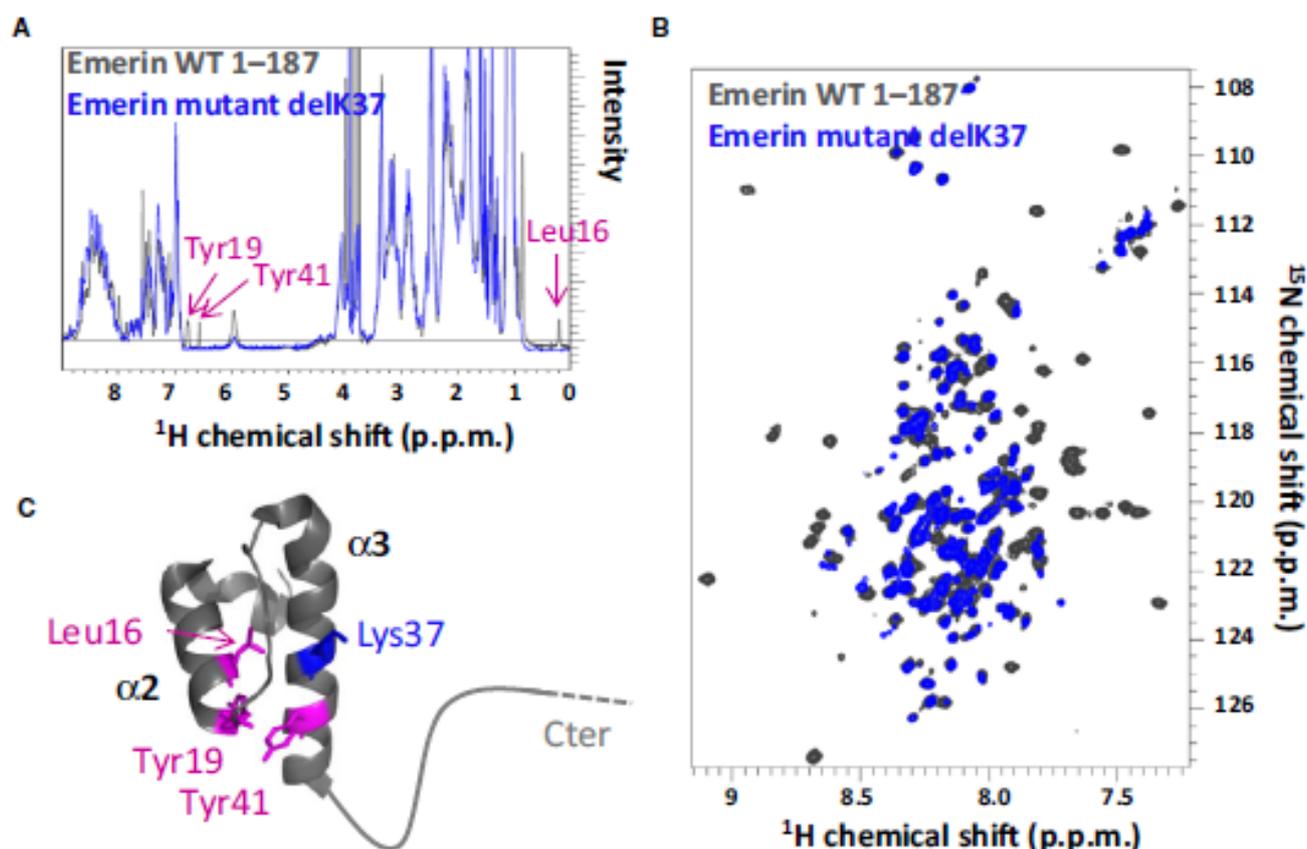


Fig. 3. The LEM domain of disease-causing mutant delK37 is unstructured. (A) Superimposition of the solution-state NMR 1D spectra of EmN (in black) and delK37 (in blue), recorded at pH 6.5, 303 K, 750 MHz. Three signals corresponding to side-chain protons of residues Leu16, Tyr19, and Tyr41 buried in the LEM domain of EmN are identified with magenta arrows. These signals are strongly shifted in the spectrum of delK37. (B) Superimposition of the solution-state NMR 2D ¹H-¹⁵N HSQC spectra of EmN (in black) and delK37 (in blue), recorded at pH 6.5, 303 K, 750 MHz. Signals corresponding to the LEM domain (in orange on Fig. 1B) are moved toward regions corresponding to random coil in the spectrum of delK37. (C) Representation of the 3D structure of EmN, with residue K37, lacking in the mutant, colored in blue, and residues Leu16, Tyr19, and Tyr41, exhibiting different chemical environments in the mutant, colored in magenta, as in (A).

and E35 into cysteines (Fig. 4D). The presence of this disulfide bridge impairs EmN132 self-assembly, whereas reducing this disulfide bond triggers filament formation, as shown by fluorescence (Fig. 4E). Solution-state ¹H-¹⁵N HSQC spectra of EmN132 wild-type (green on Fig. 4F) and Y4C/E35C (mauve) were recorded in the absence of DTT to verify that the oxidized mutant adopts a well-defined LEM-like 3D structure (well-dispersed cross-peaks of homogeneous linewidths on Fig. 4F). Analysis of the spectrum of the Y4C/E35C mutant in the presence of 5 mM DTT showed that in these conditions the mutant exhibits conformational exchange (heterogeneous linewidths on Fig. 4F).

The LEM domain region contributes to the structural core of the EmN filaments

In order to identify residues forming the structural core of EmN132 and EmN filaments, we searched for

residues protected from proteolysis by mass spectrometry. Therefore, we incubated filaments of these emerin fragments with either endoproteinase GluC or chymotrypsin during 4 h at 310 K for a limited proteolysis. Degradation products were then resolved by SDS/PAGE and analyzed by means of high-resolution tandem mass spectrometry (Fig. 5). First, two main bands with apparent molecular weights of 6 and 10 kDa are observed after proteolysis by endoproteinase GluC in the cases of both EmN132 and EmN. They all contain peptides with a LEM domain (boxed in yellow on Fig. 5). The band at 6 kDa probably corresponds to the LEM region itself. The band at 10 kDa corresponds to: (a) the LEM region and part of the intrinsically disordered region from residue 48 to residue 103, in the case of EmN132; (b) a mixed of peptides containing either the N-terminal LEM domain or the C-terminal peptide 158–187 in the case of EmN. Limited proteolysis by chymotrypsin led to a main degradation

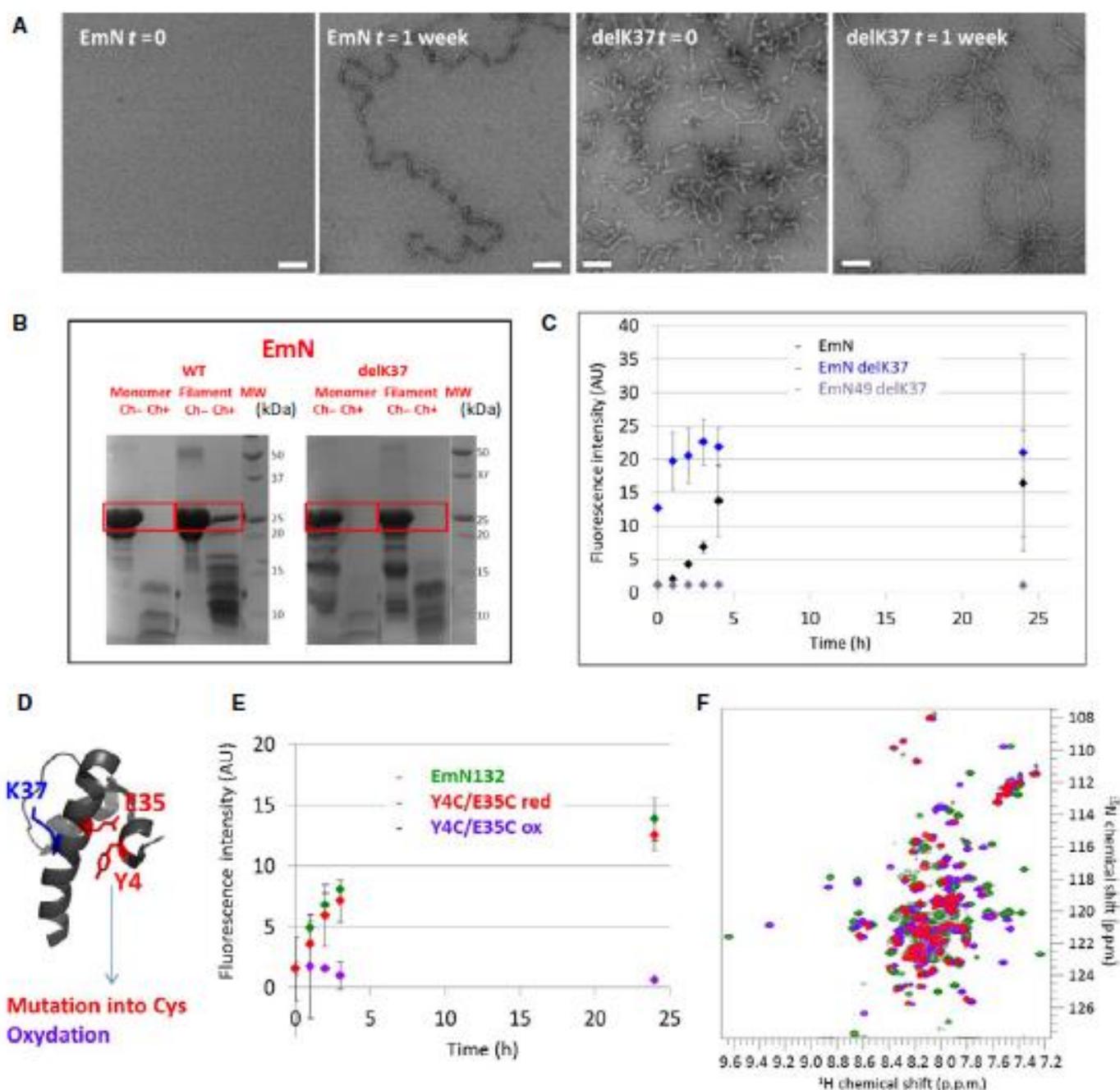


Fig. 4. The EmN mutant delK37 self-assembles faster than wild-type EmN. (A) Negative staining EM images of emerlin fragments EmN wild-type and delK37, before ($t = 0$) and after ($t = 1$ week) incubation during 1 week at 293 K and 600 μm . White bars represent 100 nm. (B) SDS/PAGE gels of EmN wild-type and delK37, corresponding to the samples observed in (A). The samples were deposited before and after incubation during 1 h at 310 K with chymotrypsin 1 : 1000. Self-assembly protects both wild-type and mutated EmN from proteolysis. Degradation products migrate similarly in both cases, suggesting that the 3D structures of wild-type and mutated filaments are closely related. (C) ThT fluorescence as a function of the incubation time at 310 K, measured on EmN wild-type and delK37 concentrated at 300 μm ($n = 2$; error bars correspond to SD). The mutant delK37 forms filaments faster than WT EmN. (D) 3D structure of the LEM domain with K37 in blue and the two residues Y4 and E35 mutated into cysteine in mutant EmN132 Y4C/E35C in red. (E) ThT fluorescence intensity as a function of time measured on EmN132 wild-type (green), Y4C/E35C in the presence of 5 mM DTT (red), and Y4C/E35C in the absence of DTT (mauve), measured at 310 K and 300 μm ($n = 2$; error bars correspond to SD). The mutant forms filaments only in reducing conditions. (F) Superimposition of the solution-state NMR 2D ^1H - ^{15}N HSQC spectra of EmN132 wild-type (green, Tris pH 7.5 100 mM NaCl), Y4C/E35C in the presence of 5 mM DTT (red, Tris pH 7.5 30 mM NaCl), and Y4C/E35C in the absence of DTT (mauve, Tris pH 7.5 30 mM NaCl), recorded at 303 K, 700 MHz. Although in the absence of DTT, the HSQC peaks are thin and well-dispersed, as for EmN132 wild-type, in the presence of DTT, the HSQC peaks are only located in the central region of the spectrum and show heterogeneous linewidths.

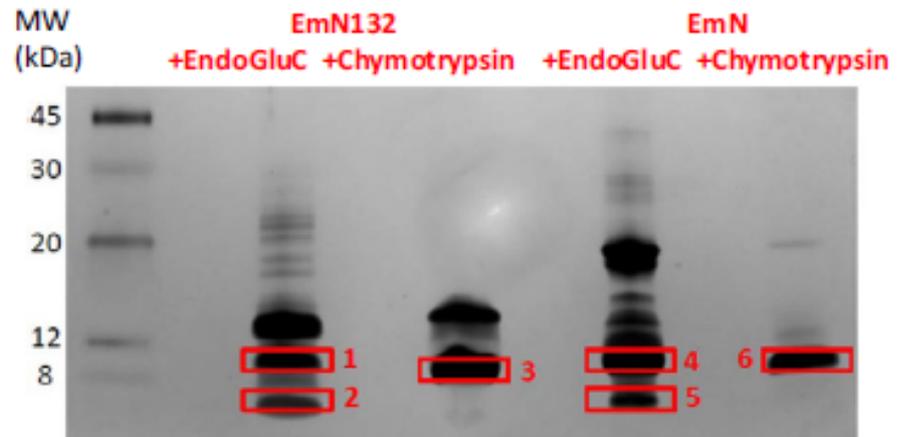


Fig. 5. The filament structural core contains the LEM domain. SDS/PAGE gel of emerin fragments EmN132 and EmN after 4 h of incubation at 310 K with either endoproteinase Glu-C or chymotrypsin (1 : 1000). The bands squared in red were characterized by mass spectrometry and their peptide composition is summarized in the table below. As judged by spectral count significant amounts of peptides are indicated by +, whereas low and no signal are indicated by -. Clearly, all samples contain at least part of the LEM region. Sample 4 contains a mixed of peptides comprising either the N terminus or the C terminus of EmN, but further proteolysis (sample 5) shows that the LEM region is the most protected fragment of EmN. This experiment was performed in duplicate.

Peptide sequence	Start	Stop	EmN132			EmN		
			Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
LTLLR	12	17	+	+	-	+	-	-
RYNIPHGPPVVGSTR	18	31	+	+	+	+	+	+
RYNIPHGPPVVGSTRR	18	32	+	+	-	+	+	-
YNIPHGPPVVGSTR	19	31	+	+	+	+	+	-
NIPHGPPVVGSTR	20	31	-	-	+	-	-	+
KIFEYETQR	37	45	+	-	+	+	-	+
IFEYETQR	38	45	+	-	+	+	-	+
RLSPSSSAASSY	47	59	-	-	+	-	-	+
LSPSSSAASSY	48	59	-	-	+	-	-	+
LSPSSSAASSYSFSDLN								
STR	48	68	+	-	-	+	-	-
LSPSSSAASSYSFSDLN								
STRGDADMY	48	74	-	-	+	-	-	-
GDADMYDLPK	69	78	+	-	-	+	-	-
GDADMYDLPKKE	69	80	+	-	-	+	-	-
GYNDYYEESYFTTR	89	103	+	-	-	-	-	-
DSAYQSITHYRPVSASR	158	174	-	-	-	+	-	-
SSLDSYPTSSS	175	187	-	-	-	+	-	-

band with an apparent molecular weight of 9 kDa in the case of EmN132 and 10 kDa in the case of EmN. These bands again correspond to degradation products comprising the LEM region and an additional fragment: (a) fragment 47–74 in the case of EmN132; (b) probably a slightly larger fragment in the case of EmN, even if only region 47–59 is detected by mass spectrometry. Thus an N-terminal region with an apparent molecular mass of about 10 kDa is protected from proteolysis in self-assembled EmN132 and EmN with the LEM region being the most protected from cleavage by either endoproteinase Glu-C or chymotrypsin.

The EmN filaments, either wild-type or delK37, share a common structural core rich in β -structure comprising the LEM domain β 1-strand

We finally characterized the filament structure of EmN, EmN132, and EmN delK37 using solid-state NMR. Superimposing the ^{13}C - ^{13}C proton-driven spin diffusion spectra of EmN wild-type and delK37 showed that the correlation peaks overlap nicely (Fig. 6A). However, the spectra are dominated by random coil signals which are relatively broad. In a double quantum–single quantum experiment (Fig. 6B), double quantum coherence is generated from and later

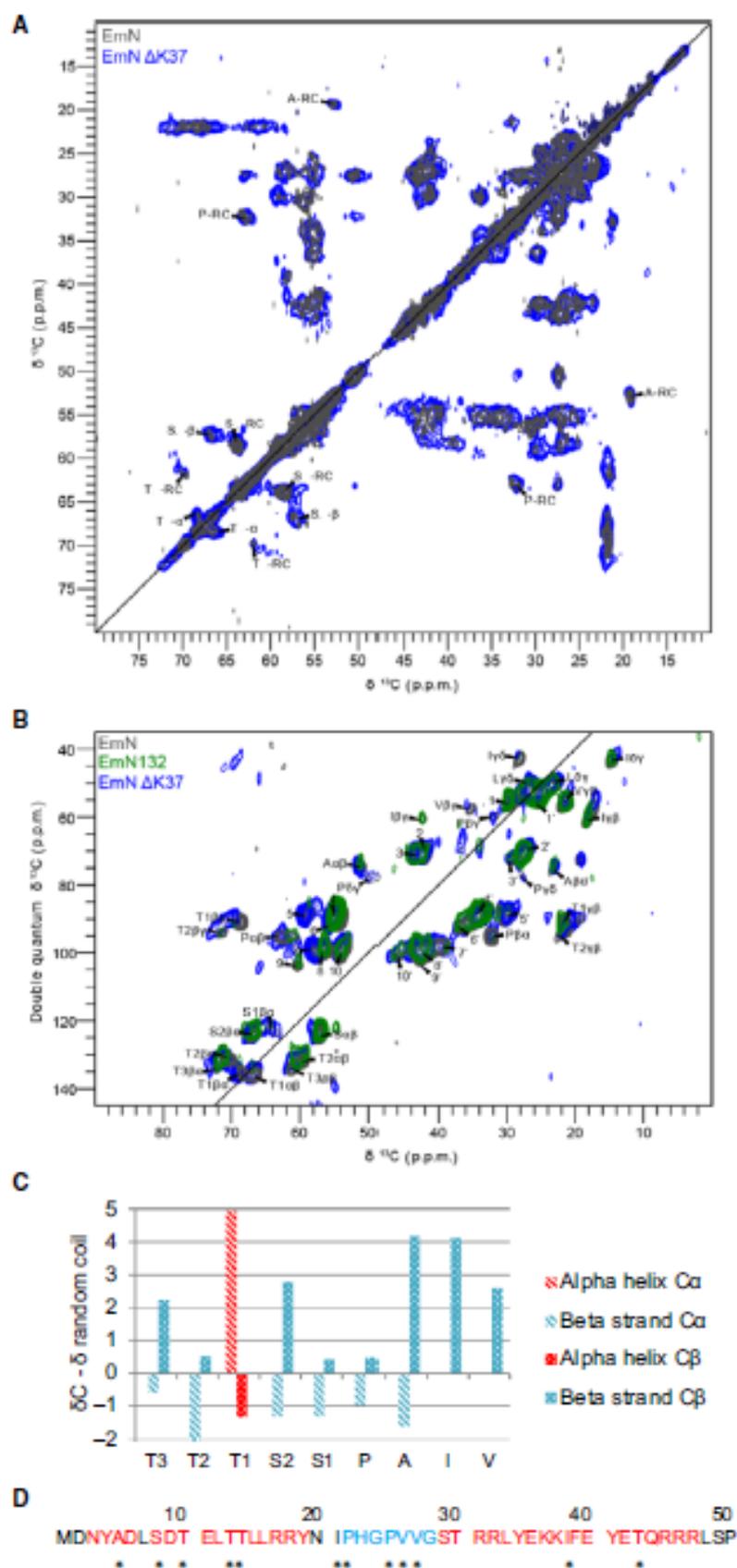


Fig. 6. The filament structural core is similar in EmN wild-type and delK37 and contains β -structure. (A) Solid-state NMR 2D PDSD spectra recorded with a mixing time of 50 ms for EmN wild-type (black) and delK37 (blue); secondary structure classifications are indicated (α = alpha helical, β = beta sheet, RC = random coil). (B) Solid-state NMR double quantum-single quantum C-C correlation spectra for EmN wild-type (black) and delK37 (blue) and EmN132 (green), with the identified residue types indicated. (C) Secondary structure analysis for the assigned residues from the double quantum-single quantum C-C correlation spectrum of EmN as shown in (B). Depicted is the difference between ^{13}C observed chemical shift and random coil reference [39]. Negative differences for $\text{C}\alpha$ and positive differences for $\text{C}\beta$ indicate β -sheet secondary structure; corresponding bars are colored in cyan. Bars indicating α -helix secondary structure are colored in red. (D) Sequence of the LEM domain colored as a function of the secondary structure observed in solution (see Fig. 1C,D), with α -helices displayed in red and β -strands in cyan. Asterisks indicate residues identified in (C) and discussed in the text.

reconverted to single quantum coherence by means of SPC-5 [32] excitation and reconversion periods. During these recoupling periods, an effective dipolar double-

quantum Hamiltonian is active, which will be scaled down in the presence of molecular motions. Thus, only very rigid parts of the protein are visible in the

resulting spectrum. Approximately 20 amino acids contribute to the double quantum–single quantum ^{13}C – ^{13}C correlation spectrum of EmN filaments (Fig. 6B): these residues form the filament core structure. They comprise three threonines, two serines, one proline, one alanine, one valine, one leucine, and one isoleucine. EmN132 shows a very similar double quantum–single quantum ^{13}C – ^{13}C correlation spectrum, with a clear isoleucine signal, and possesses only two isoleucines, both in the LEM domain region, precisely at positions 21 and 38. This first observation confirms that the LEM domain region is part of the filament structural core. Valine residues are found at positions 26 and 27 and proline residues at positions 22 and 25 in this same LEM domain region. This suggests that the structural core comprises at least fragment 21–26, which corresponds to the strand βI of the LEM domain (Fig. 1D). Analysis of the secondary chemical shifts of the isoleucine and valine $^{13}\text{C}_\beta$ measured in the double quantum–single quantum spectrum predicts that these residues also adopt a β -conformation in the filament (Fig. 6B,C). Additionally, secondary structure predictions for an alanine, two serines and two threonines identify another β -structure element in the filament. Only one alanine is present in the LEM domain region, at position 5 (Fig. 6D). Two serines are present at positions 8 and 29, and several threonines are present at positions 10, 13, 14, 30, and 43. These residues are all in α -helices in monomeric EmN. So either a fragment of the LEM domain region switches from an α - to a β -conformation (as fragment 5–14 for example; see Fig. 6D) or a fragment of the intrinsically disordered region folds into a β -conformation during EmN self-assembly. Both processes could also take place simultaneously. A more detailed description of the filament core structure would necessitate further assignment of the solid-state NMR signals. This is prevented by structural heterogeneity/resolution issues in our conditions.

Discussion

Very few structural data are yet available on the architecture of the nuclear envelope. Inner nuclear membrane proteins exhibit large regions predicted to be intrinsically disordered that interact with the nuclear intermediate filaments called the lamins lining the inner nuclear membrane. However, the details of these intermolecular interactions and their mode of regulation through oligomerization and post-translational modifications are still unclear. At the nuclear envelope, LEM domain proteins share the ability to bind lamins and tether repressive chromatin at the nuclear

periphery (for review, see [33]). Their LEM domain binds BAF, a metazoan DNA-binding protein [34,35]. Here we focus on the inner nuclear envelope emerin, one of the most studied LEM domain proteins. The group of Prof K. Wilson was the first to reveal, from biochemical analyses performed *in vitro* and in cells, that emerin could indeed be assembled as oligomers anchored at the inner nuclear membrane [24]. They showed that several intra- and/or intermolecular interactions could be observed within emerin molecules, and suggested that changes in emerin oligomeric state were responsible for regulating emerin binding properties at the inner nuclear envelope. In particular, the LEM domain could be found in a ‘DOWN’ state in which it interacts with BAF at the chromatin, and also in a ‘UP’ state in which it participates in emerin self-assembly and potentially partner recognition at the nuclear envelope. In this study, we demonstrate that the LEM domain can indeed be found in two different quaternary structures, either as a globular domain in a monomeric EmN fragment, or as part of the structural core of long curvilinear filaments formed by EmN. The molecular details of these LEM structures and their potential functional role are discussed in the following.

The LEM domain plays a central role in EmN oligomerization

Our results reveal for the first time that emerin N-terminal region from residue 1 to residue 132 can self-assemble through interactions involving the LEM domain and part of the intrinsically disordered region 48–132. Region 21–26 might be involved in emerin self-assembly, as strongly suggested by solid-state NMR spectroscopy. This region corresponds to the LEM domain strand βI . Region 95–99 is also critical for emerin oligomerization: EDMD causing small in-frame deletion del95-99 in EmN impaired filament formation in our conditions [25]. It exhibits a serine residue (Ser98) and is close to a threonine-rich patch (Thr101, Thr102, Thr104), which correspond to amino acid types observed by solid-state NMR in the rigid part of the filaments.

LEM domain unfolding favors EmN oligomerization

Analysis of the solution structure of EmN mutant delK37 revealed that this single amino acid deletion in helix α3 destroys the LEM domain fold. However, delK37 self-assembles faster than wild-type EmN and is capable of forming oligomers with a similar

structural organization, as shown by negative staining electron microscopy, limited proteolysis, and solid-state NMR. Figure 4B clearly shows that the first steps of self-assembly are significantly facilitated by deletion of K37, suggesting that structural rearrangement of the LEM domain is a rate-limiting step in filament assembly. Interactions between strand $\beta 1$ and the rest of the LEM domain might have to be disrupted in order to enable intermolecular interactions with another emerlin molecule. His23 is located in strand $\beta 1$ and its side chain is directed toward helix $\alpha 2$. Protonation of this histidine under pH 6.5 might favor filament assembly. In our conditions, lowering the pH under 6.5 triggered EmN aggregation, thus preventing any EmN study under this pH [25]. However, in the case of mutant Q133H, which self-assembles slower than wild-type EmN, going from pH 8.0 to pH 6.8 clearly favored self-assembly (Suppl. Figure 4 in [25]).

Functional role of the self-assembled form of emerlin

The best described LEM domain function is to interact with BAF in order to create a physical link between the nuclear envelope and the chromatin. All studied LEM domain proteins bind BAF. The 3D structure of a complex between emerlin LEM domain and a dimer of BAF was determined from solution-state NMR data [31]. The two main α -helices ($\alpha 2$ and $\alpha 3$) as well as the β -strand ($\beta 1$) all directly contact BAF dimer, suggesting that the LEM fold is critical for BAF recognition. Thus the emerlin mutant delK37 and the filaments of EmN, either wild-type or delK37, might also have lost part of their BAF-binding capacity. On the opposite, mutant del95-99, which is defective in EmN self-assembly [25], might be more often in a conformation efficient for BAF binding.

Emerlin also binds lamin A/C, and this interaction is critical for emerlin localization at the nuclear envelope. It was reported that region EmN132 is necessary and sufficient for binding to the tails of lamin A/C and lamin B1 [24]. We showed in this study that EmN132 is necessary and sufficient for filament formation. It was further reported that emerlin mutant del95-99 does not bind lamin A/C as observed by blot overlay assays [30], and that in cells, proximities between this emerlin mutant and lamin A/C are less frequently observed by proximity ligation assays [25]. We showed that EmN mutant del95-99 does not self-associate in our conditions [25]. Finally, several LEM domain mutants, and in particular a mutant with alanines at positions 24, 25, 26, and 27 (in strand $\beta 1$) as well as a mutant with alanines at positions 34, 35, 36, and 37 (in helix $\alpha 3$),

bind lamin A/C reproducibly better than wild-type emerlin as observed by blot overlay assays [30]. Such large number of alanine substitutions in secondary structure elements probably destabilizes the LEM domain fold. As our results suggest that destabilization of the LEM fold favors EmN self-assembly, we propose that the alanine mutants have a less stable LEM domain, which favors emerlin self-assembly and thus lamin A/C binding.

It was recently proposed that LEM domain proteins self-associate in order to form a 3D complex network at the nuclear envelope. We propose a first analysis of the emerlin self-association mechanisms, which reveals that the LEM domain (amino acid 1 to amino acid 49) interacts with residues from the intrinsically disordered region (amino acid 50 to amino acid 132) in order to oligomerize. Our analysis strongly suggests that the LEM domain strand $\beta 1$ undergoes a structural rearrangement in order to contact fragments of other emerlin molecules, at the N terminus of the LEM fold or/and in the intrinsically disordered region, thus triggering emerlin self-association into filaments. Further studies will be needed in order to understand if this self-assembly pathway also occurs in cells, and how it is influenced by emerlin membrane anchoring, post-translational modifications, and partner binding. However, as emerlin is highly phosphorylated in cells, and in particular at positions 4, 8, 10, and 28 at the N termini of helices $\alpha 1$, $\alpha 2$, and $\alpha 3$, we propose that phosphorylation impacts the stability of the LEM domain and thus regulate emerlin self-assembly. Moreover our results, as well as previous analyses [24,25], suggest that emerlin self-assembly regulates BAF and lamin binding. Modification and oligomerization were identified as essential events for the regulation of the LEM domain protein network at the nuclear envelope [33]. Deciphering these molecular mechanisms provides first elements to explain the impact of LEM domain protein mutations on nuclear envelope function, in the frame of human envelopopathies.

Materials and methods

Protein expression and purification

Human wild-type emerlin fragments from amino acid 1 to amino acid 187 (EmN) and from amino acid 67 to amino acid 187 (EmN67C) were expressed in *Escherichia coli* BL21 DE3 Star (Novagen, EMD Biosciences, Madison, WI, USA) using a pETM13 vector as N-terminal octahistidine fusions. The emerlin cDNAs were optimized in terms of codon usage for expression in *E. coli* (Genscript, Hong Kong, China). To encode emerlin variants from

structural organization, as shown by negative staining electron microscopy, limited proteolysis, and solid-state NMR. Figure 4B clearly shows that the first steps of self-assembly are significantly facilitated by deletion of K37, suggesting that structural rearrangement of the LEM domain is a rate-limiting step in filament assembly. Interactions between strand $\beta 1$ and the rest of the LEM domain might have to be disrupted in order to enable intermolecular interactions with another emerlin molecule. His23 is located in strand $\beta 1$ and its side chain is directed toward helix $\alpha 2$. Protonation of this histidine under pH 6.5 might favor filament assembly. In our conditions, lowering the pH under 6.5 triggered EmN aggregation, thus preventing any EmN study under this pH [25]. However, in the case of mutant Q133H, which self-assembles slower than wild-type EmN, going from pH 8.0 to pH 6.8 clearly favored self-assembly (Suppl. Figure 4 in [25]).

Functional role of the self-assembled form of emerlin

The best described LEM domain function is to interact with BAF in order to create a physical link between the nuclear envelope and the chromatin. All studied LEM domain proteins bind BAF. The 3D structure of a complex between emerlin LEM domain and a dimer of BAF was determined from solution-state NMR data [31]. The two main α -helices ($\alpha 2$ and $\alpha 3$) as well as the β -strand ($\beta 1$) all directly contact BAF dimer, suggesting that the LEM fold is critical for BAF recognition. Thus the emerlin mutant delK37 and the filaments of EmN, either wild-type or delK37, might also have lost part of their BAF-binding capacity. On the opposite, mutant del95-99, which is defective in EmN self-assembly [25], might be more often in a conformation efficient for BAF binding.

Emerlin also binds lamin A/C, and this interaction is critical for emerlin localization at the nuclear envelope. It was reported that region EmN132 is necessary and sufficient for binding to the tails of lamin A/C and lamin B1 [24]. We showed in this study that EmN132 is necessary and sufficient for filament formation. It was further reported that emerlin mutant del95-99 does not bind lamin A/C as observed by blot overlay assays [30], and that in cells, proximities between this emerlin mutant and lamin A/C are less frequently observed by proximity ligation assays [25]. We showed that EmN mutant del95-99 does not self-associate in our conditions [25]. Finally, several LEM domain mutants, and in particular a mutant with alanines at positions 24, 25, 26, and 27 (in strand $\beta 1$) as well as a mutant with alanines at positions 34, 35, 36, and 37 (in helix $\alpha 3$),

bind lamin A/C reproducibly better than wild-type emerlin as observed by blot overlay assays [30]. Such large number of alanine substitutions in secondary structure elements probably destabilizes the LEM domain fold. As our results suggest that destabilization of the LEM fold favors EmN self-assembly, we propose that the alanine mutants have a less stable LEM domain, which favors emerlin self-assembly and thus lamin A/C binding.

It was recently proposed that LEM domain proteins self-associate in order to form a 3D complex network at the nuclear envelope. We propose a first analysis of the emerlin self-association mechanisms, which reveals that the LEM domain (amino acid 1 to amino acid 49) interacts with residues from the intrinsically disordered region (amino acid 50 to amino acid 132) in order to oligomerize. Our analysis strongly suggests that the LEM domain strand $\beta 1$ undergoes a structural rearrangement in order to contact fragments of other emerlin molecules, at the N terminus of the LEM fold or/and in the intrinsically disordered region, thus triggering emerlin self-association into filaments. Further studies will be needed in order to understand if this self-assembly pathway also occurs in cells, and how it is influenced by emerlin membrane anchoring, post-translational modifications, and partner binding. However, as emerlin is highly phosphorylated in cells, and in particular at positions 4, 8, 10, and 28 at the N termini of helices $\alpha 1$, $\alpha 2$, and $\alpha 3$, we propose that phosphorylation impacts the stability of the LEM domain and thus regulate emerlin self-assembly. Moreover our results, as well as previous analyses [24,25], suggest that emerlin self-assembly regulates BAF and lamin binding. Modification and oligomerization were identified as essential events for the regulation of the LEM domain protein network at the nuclear envelope [33]. Deciphering these molecular mechanisms provides first elements to explain the impact of LEM domain protein mutations on nuclear envelope function, in the frame of human envelopopathies.

Materials and methods

Protein expression and purification

Human wild-type emerlin fragments from amino acid 1 to amino acid 187 (EmN) and from amino acid 67 to amino acid 187 (EmN67C) were expressed in *Escherichia coli* BL21 DE3 Star (Novagen, EMD Biosciences, Madison, WI, USA) using a pETM13 vector as N-terminal octahistidine fusions. The emerlin cDNAs were optimized in terms of codon usage for expression in *E. coli* (Genscript, Hong Kong, China). To encode emerlin variants from

amino acid 1 to amino acid 49 (EmN49) or amino acid 132 (EmN132), stop codons were inserted into the EmN cDNA using a standard QuikChange Site-Directed Mutagenesis kit (Stratagene, San Diego, CA, USA). Mutations delK37 and Y4C/E38C were inserted into the EmN, EmN49 cDNAs, and the EmN132 cDNA, respectively, using the same Mutagenesis kit. Bacteria were cultured in rich medium or ^{15}N -labeled minimum medium, induced at an optimal density of 1.0 with 0.5 mM isopropyl β -D-1-thiogalactopyranoside, grown overnight at 293 K, and lysed in 50 mM Tris/HCl (pH 8.0), 300 mM NaCl, 40 mM imidazole, 5% glycerol, 1% Triton X-100, and 1 mM phenylmethanesulfonylfluoride. After centrifugation at 20 000 *g* for 20 min at 277 K, the pellet was resuspended in buffer C8 (50 mM Tris/HCl pH8.0, 150 mM NaCl, 20 mM imidazole, 8 M urea). A second centrifugation step was performed at 20 000 *g* for 20 min at 293 K. The soluble extract was then loaded onto Ni-NTA beads (GE Healthcare, Piscataway, NJ, USA) equilibrated with buffer C8. Proteins were eluted directly with buffer E8 (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 M imidazole, 8 M urea). Proteins were refolded by dialysis in buffer D (20 mM Tris/HCl pH8.0, 30 mM NaCl, 2 mM DTT). After refolding, they were incubated with the TEV protease during 3 h at room temperature and finally dialyzed into the desired buffer. The final yield was typically 10 mg of purified protein per liter of bacterial culture [25].

Liquid-state NMR spectroscopy

After purification and His-Tag cleavage, proteins were dialyzed against 20 mM sodium phosphate pH 6.5, 30 mM NaCl, 5 mM DTT (20 mM Tris/HCl pH 7.5, 30–100 mM NaCl, 5 mM DTT in Fig. 4F). NMR experiments were performed on 700 or 750 MHz Bruker Avance II spectrometers equipped with a cryogenic probe. 2D NMR ^1H ^{15}N correlation spectra were acquired using a HSQC pulse sequence at 303 K, on a 3-mm-diameter NMR sample tube containing 100–200 μM uniformly ^{15}N -labeled EmN, EmN67C, EmN49, EmN132, EmN delK37, or EmN132 Y4C/E35C. 3D ^1H ^{15}N ^{13}C correlation spectra were acquired using 3D HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, and HN(CO)(CA)NH pulse sequences at 303 K, on a 3-mm-diameter NMR sample tube containing 500–600 μM uniformly ^{15}N ^{13}C -labeled EmN132. The data were processed using TOPSPIN3.1 (Bruker Biospin, Wissembourg, France) and analyzed with CCPNMR [36].

Solid-state NMR spectroscopy

After purification and His-Tag cleavage, proteins were dialyzed against 20 mM Tris/HCl pH 8, 30 mM NaCl, and 5 mM DTT. Solid-state NMR experiments for the EmN and EmN132 constructs were recorded on a Bruker Avance III HD 800 MHz spectrometer with 18.8 T external

magnetic field or on a Bruker Avance III HD 700 MHz spectrometer with 16.4 T external magnetic field. All solid-state NMR experiments for the EmN delK37 mutant were recorded on a Bruker Avance III HD 700 MHz spectrometer with 16.4 T external magnetic field. The sample temperature was determined using the water-proton signal referenced to DSS (3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt) and kept at 283 K. Double quantum–single quantum spectra were recorded with magic-angle spinning at 8 kHz and C–C correlation spectra with 11 kHz MAS rate. High-power ^1H decoupling was performed during evolution and detection periods using SPINAL-64 [37] at a radio frequency strength of 83 kHz. All spectra were processed using Bruker TOPSPIN 3.2 and analyzed with CCPNMR [36].

Double quantum–single quantum C–C correlation spectra were recorded using SPC5 [32] excitation and reconversion periods of 0.5 ms. Maximum acquisition times for the direct dimension were 10 ms and for the indirect dimension 1.2 ms. For the EmN and EmN132 samples, the number of scans was set to 336 and it was set to 192 for the mutant delK37. C–C correlation spectra using proton-driven spin diffusion (PDS) were recorded using 50 ms mixing time [38]. Maximum acquisition times were 15 ms for the direct and 10 ms for the indirect dimension. For the EmN and EmN132 samples, the number of scans was set to 112 and it was set to 184 for the mutant delK37. Secondary structure classifications were based on the carbon chemical shifts as described by Wang *et al.* [39].

Electron microscopy

To observe self-assembled proteins, EmN, EmN49, EmN132, EmN67C, and EmN delK37 were dialyzed against 20 mM Tris/HCl (pH 8.0), 30 mM NaCl, and 5 mM DTT. They were concentrated up to 600 μM and incubated at 293 K during 1 week. Specimens were prepared by negative staining with 2% uranyl acetate on glow-discharged carbon-coated copper grids. Data collection was performed using a Tecnai Spirit transmission electron microscope (FEI) equipped with a LaB₆ filament, operating at 100 kV. Images were recorded on a K2 Base camera (Gatan, Pleasanton, CA, USA, 4k \times 4k) at 15 000 or 4400 magnification (pixel size at specimen level – 0.25 and 0.83 nm, respectively).

Thioflavin kinetics

After purification and His-Tag cleavage, proteins were dialyzed against 20 mM Tris/HCl pH 8.0, 30 mM NaCl, and 5 mM DTT. Oligomer assembly was followed with time by incubating proteins at 300 μM and 310 K and by regularly taking protein aliquots to be analyzed by fluorescence. These aliquots were diluted in a thioflavin T (ThT) containing buffer so as to obtain 20 μM protein and 2.5 μM ThT in

20 mM Tris (pH 8.0), 30 mM NaCl, and 5 mM DTT. The fluorescence measurements were carried out in a 60 μ L cuvette at 293 K using a fluorimeter JASCO ADP-303T. The fluorescence was monitored using an excitation wavelength of 440 nm and fluorescence emission was read at 480 nm.

Limited proteolysis

After purification and His-Tag cleavage, EmN, EmN132, and mutant delK37 were dialyzed against 20 mM Tris/HCl pH 8.0, 30 mM NaCl, and 5 mM DTT. For experiments on monomers, proteins were concentrated up to 125 μ M. Filament formation was performed by concentrating proteins until 600 μ M and heating the samples during 1 h at 338 K (65 °C) [25]. Before proteolysis, filaments were diluted at 125 μ M and both monomers and filaments were incubated with Chymotrypsin or EndoproteinaseGluC (1 : 1000) during 1 h at 310 K. Protease reactions were stopped by the addition of 10 μ L of Laemmli blue and sample heating during 5 min at 368 K (95 °C). Degradation products were then resolved by SDS/PAGE and analyzed by means of high-resolution tandem mass spectrometry.

Tandem mass spectrometry

For mapping fragments generated by limited proteolysis, excised polyacrylamide bands were first rinsed with MilliQ water, destained, and then subjected to in-gel protein reduction with DTT followed by protein alkylation with iodoacetamide, as described previously [40]. They were digested with either sequencing-grade Trypsin Gold and ProteaseMax surfactant (Promega, Madison, WI, USA) for 1 h at 323 K or sequencing-grade chymotrypsin for 1 h at 298 K as previously reported [41]. Trifluoroacetic acid was added to the samples at 0.5% final concentration to stop the enzymatic reaction. The resulting peptides were identified by tandem mass spectrometry with an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Rockford, IL, USA) coupled to an UltiMate 3000 nRSLC system (Dionex; ThermoFisher), under conditions similar to those previously described [42,43]. Briefly, peptide mixtures (10 μ L injection) were first desalted online in a reverse-phase Acclaim PepMap100 C18 μ -precolumn (5 μ m, 100 Å, 300 μ m inner diameter \times 5 mm, LC Packings, Amsterdam, the Netherlands), and then were separated on a nanoscale Acclaim PepMap100 C18 capillary column (3 μ m, 100 Å, 75 μ m i.d. \times 50 cm, LC Packings) using a 90-min gradient from 4% to 40% solvent B (CH₃CN, 0.01% formic acid) with solvent A being H₂O, 0.01% formic acid. The full-scan mass spectra of high mass accuracy were measured in the Orbitrap analyzer from *m/z* 300 to 1800 in data-dependent mode. The MS/MS scans in the linear ion trap were acquired on the seven most abundant precursor ions. MS/MS data processing was done as

described previously [44] against the Swissprot_Human database comprising 20 192 protein sequences using the following parameters: semitryptic peptides with a maximum of two miss-cleavages; mass error tolerances set at 5 ppm for parent ions and at 0.5 Da for fragments; fixed modification for carbamidomethylated cysteine and variable modification for methionine oxidation. Peptides were filtered with a *P*-value threshold below 0.05 and a rank set at 1.

Conflicts of interest

The authors declare no competing interest.

Acknowledgements

The authors acknowledge Pascal Fricke for his help in setting up and analyzing the solid-state experiments and the FMP NMR facility (Leibniz-Institut für Molekulare Pharmakologie, Berlin, Germany) for access to the NMR spectrometers.

Funding

This work was supported by CEA, CNRS and University Paris South, by the French Infrastructure for Integrated Structural Biology (<https://www.structuralbiology.eu/networks/frisbi>, grant number ANR-10-INSB-05-01, Acronym FRISBI) and by the French Association against Myopathies (AFM; research grants no. 17243 and 20018 to SZJ and PhD fellowship no. 18159 to CS).

Author contributions

CS, FC, NE, IH, FXT performed the biochemistry, fluorescence, and liquid-state NMR experiments. KH and MZ carried out the solid-state NMR experiments. CS and AAA were involved in the EM analyses. BAB and JA performed and analyzed the mass spectrometry experiments. CC, AL, and SZJ supervised the project. CS, AL, SZJ wrote the paper.

References

- 1 Petrie RJ, sKoo H & Yamada KM (2014) Generation of compartmentalized pressure by a nuclear piston governs cell motility in a 3D matrix. *Science* **345**, 1062–1065.
- 2 Osmanagic-Myers S, Dechat T & Foisner R (2015) Lamins at the crossroads of mechanosignaling. *Genes Dev* **29**, 225–237.
- 3 Bertrand AT, Ziaei S, Ehret C, Duchemin H, Mamchaoui K, Bigot A, Mayer M, Quijano-Roy S,

- Desguerre I, Laine J *et al.* (2014) Cellular microenvironments reveal defective mechanosensing responses and elevated YAP signaling in LMNA-mutated muscle precursors. *J Cell Sci* **127**, 2873–2884.
- 4 Swift J & Discher DE (2014) The nuclear lamina is mechano-responsive to ECM elasticity in mature tissue. *J Cell Sci* **127**, 3005–3015.
- 5 Guilluy C, Osborne LD, Van Landeghem L, Sharek L, Superfine R, Garcia-Mata R & Burrridge K (2014) Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nat Cell Biol* **16**, 376–381.
- 6 Ho CY, Jaalouk DE, Vartiainen MK & Lammerding J (2013) Lamin A/C and emerin regulate MKL1-SRF activity by modulating actin dynamics. *Nature* **497**, 507–511.
- 7 Harr JC, Gonzalez-Sandoval A & Gasser SM (2016) Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Rep* **17**, 139–155.
- 8 Bertrand AT, Renou L, Papadopoulos A, Beuvin M, Lacene E, Massart C, Ottolenghi C, Decostre V, Maron S, Schlossarek S *et al.* (2012) DeIK32-lamin A/C has abnormal location and induces incomplete tissue maturation and severe metabolic defects leading to premature death. *Hum Mol Genet* **21**, 1037–1048.
- 9 Azibani F, Muchir A, Vignier N, Bonne G & Bertrand AT (2014) Striated muscle laminopathies. *Semin Cell Dev Biol* **29**, 107–115.
- 10 Lammerding J, Schulze PC, Takahashi T, Kozlov S, Sullivan T, Kamm RD, Stewart CL & Lee RT (2004) Lamin A/C deficiency causes defective nuclear mechanics and mechanotransduction. *J Clin Invest* **113**, 370–378.
- 11 Lammerding J, Hsiao J, Schulze PC, Kozlov S, Stewart CL & Lee RT (2005) Abnormal nuclear shape and impaired mechanotransduction in emerin-deficient cells. *J Cell Biol* **170**, 781–791.
- 12 Hale CM, Shrestha AL, Khatau SB, Stewart-Hutchinson PJ, Hernandez L, Stewart CL, Hodzic D & Wirtz D (2008) Dysfunctional connections between the nucleus and the actin and microtubule networks in laminopathic models. *Biophys J* **95**, 5462–5475.
- 13 Emerson LJ, Holt MR, Wheeler MA, Wehnert M, Parsons M & Ellis JA (2009) Defects in cell spreading and ERK1/2 activation in fibroblasts with lamin A/C mutations. *Biochim Biophys Acta* **1792**, 810–821.
- 14 Worman HJ & Bonne G (2007) "Laminopathies": a wide spectrum of human diseases. *Exp Cell Res* **313**, 2121–2133.
- 15 Krimm I, Ostlund C, Gilquin B, Couprie J, Hossenlopp P, Momon JP, Bonne G, Courvalin JC, Worman HJ & Zinn-Justin S (2002) The Ig-like structure of the C-terminal domain of lamin A/C, mutated in muscular dystrophies, cardiomyopathy, and partial lipodystrophy. *Structure* **10**, 811–823.
- 16 Bione S, Maestrini E, Rivella S, Mancini M, Regis S, Romeo G & Toniolo D (1994) Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat Genet* **8**, 323–327.
- 17 Ellis JA, Craxton M, Yates JR & Kendrick-Jones J (1998) Aberrant intracellular targeting and cell cycle-dependent phosphorylation of emerin contribute to the Emery-Dreifuss muscular dystrophy phenotype. *J Cell Sci* **111** (Pt 6), 781–792.
- 18 Ben Yaou R, Toutain A, Arimura T, Demay L, Massart C, Peccate C, Muchir A, Lense S, Debrugre N, Leturcq F *et al.* (2007) Multitissular involvement in a family with LMNA and EMD mutations: role of digenic mechanism? *Neurology* **68**, 1883–1894.
- 19 Berk JM, Tift KE & Wilson KL (2013) The nuclear envelope LEM-domain protein emerin. *Nucleus* **4**, 298–314.
- 20 Sosa BA, Rothbaler A, Kutay U & Schwartz TU (2012) LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins. *Cell* **149**, 1035–1047.
- 21 Wang W, Shi Z, Jiao S, Chen C, Wang H, Liu G, Wang Q, Zhao Y, Greene MI & Zhou Z (2012) Structural insights into SUN-KASH complexes across the nuclear envelope. *Cell Res* **22**, 1440–1452.
- 22 Zhou Z, Du X, Cai Z, Song X, Zhang H, Mizuno T, Suzuki E, Yee MR, Berezov A, Murali R *et al.* (2012) Structure of Sad1-UNC84 homology (SUN) domain defines features of molecular bridge in nuclear envelope. *J Biol Chem* **287**, 5317–5326.
- 23 Nie S, Ke H, Gao F, Ren J, Wang M, Huo L, Gong W & Feng W (2016) Coiled-coil domains of SUN proteins as intrinsic dynamic regulators. *Structure* **24**, 80–91.
- 24 Berk JM, Simon DN, Jenkins-Houk CR, Westerbeck JW, Gronning-Wang LM, Carlson CR & Wilson KL (2014) The molecular basis of emerin-emerin and emerin-BAF interactions. *J Cell Sci* **127**, 3956–3969.
- 25 Herrada I, Samson C, Velours C, Renault L, Ostlund C, Chervy P, Puchkov D, Worman HJ, Buendia B & Zinn-Justin S (2015) Muscular dystrophy mutations impair the nuclear envelope emerin self-assembly properties. *ACS Chem Biol* **10**, 2733–2742.
- 26 Manilal S, Nguyen TM, Sewry CA & Morris GE (1996) The Emery-Dreifuss muscular dystrophy protein, emerin, is a nuclear membrane protein. *Hum Mol Genet* **5**, 801–808.
- 27 Yorifuji H, Tadano Y, Tsuchiya Y, Ogawa M, Goto K, Umetani A, Asaka Y & Arahata K (1997) Emerin, deficiency of which causes Emery-Dreifuss muscular dystrophy, is localized at the inner nuclear membrane. *Neurogenetics* **1**, 135–140.

- 28 Wolff N, Gilquin B, Courchay K, Callebaut I, Worman HJ & Zinn-Justin S (2001) Structural analysis of emerin, an inner nuclear membrane protein mutated in X-linked Emery-Dreifuss muscular dystrophy. *FEBS Lett* **501**, 171–176.
- 29 Samson C, Herrada I, Celli F, Theillet FX & Zinn-Justin S (2016) ¹H, ¹³C and ¹⁵N backbone resonance assignment of the intrinsically disordered region of the nuclear envelope protein emerin. *Biomol NMR Assign* **10**, 179–182.
- 30 Lee KK, Haraguchi T, Lee RS, Koujin T, Hiraoka Y & Wilson KL (2001) Distinct functional domains in emerin bind lamin A and DNA-bridging protein BAF. *J Cell Sci* **114**, 4567–4573.
- 31 Cai M, Huang Y, Suh JY, Louis JM, Ghirlando R, Craigie R & Clore GM (2007) Solution NMR structure of the barrier-to-autointegration factor-Emerin complex. *J Biol Chem* **282**, 14525–14535.
- 32 Hohwy M, Rienstra CM, Jaroniec CP & Griffin RG (1999) Fivefold symmetric homonuclear dipolar recoupling in rotating solids: application to double quantum spectroscopy. *J Chem Phys* **110**, 7983–7992.
- 33 Barton LJ, Soshnev AA & Geyer PK (2015) Networking in the nucleus: a spotlight on LEM-domain proteins. *Curr Opin Cell Biol* **34**, 1–8.
- 34 Shumaker DK, Lee KK, Tanhehco YC, Craigie R & Wilson KL (2001) LAP2 binds to BAF-DNA complexes: requirement for the LEM domain and modulation by variable regions. *EMBO J* **20**, 1754–1764.
- 35 Cai M, Huang Y, Ghirlando R, Wilson KL, Craigie R & Clore GM (2001) Solution structure of the constant region of nuclear envelope protein LAP2 reveals two LEM-domain structures: one binds BAF and the other binds DNA. *EMBO J* **20**, 4399–4407.
- 36 Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J & Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696.
- 37 Fung BM, Khitritin AK & Ermolaev K (2000) An improved broadband decoupling sequence for liquid crystals and solids. *J Magn Reson* **142**, 97–101.
- 38 Bloembergen N (1949) On the interaction of nuclear spins in a crystalline lattice. *Physica* **15**, 386–426.
- 39 Wang Y & Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **11**, 852–861.
- 40 Hartmann EM & Armengaud J (2014) Shotgun proteomics suggests involvement of additional enzymes in dioxin degradation by *Sphingomonas wittichii* RW1. *Environ Microbiol* **16**, 162–176.
- 41 Kish A, Gaillard JC, Armengaud J & Elie C (2016) Post-translational methylations of the archaeal Mre11: Rad50 complex throughout the DNA damage response. *Mol Microbiol* **100**, 362–378.
- 42 Clair G, Armengaud J & Duport C (2012) Restricting fermentative potential by proteome remodeling: an adaptive strategy evidenced in *Bacillus cereus*. *Mol Cell Proteomics* **11**, M111.013102.
- 43 Rubiano-Labrador C, Bland C, Miotello G, Guerin P, Pible O, Baena S & Armengaud J (2014) Proteogenomic insights into salt tolerance by a halotolerant alpha-proteobacterium isolated from an Andean saline spring. *J Proteomics* **97**, 36–47.
- 44 Christie-Oleza JA, Pina-Villalonga JM, Bosch R, Nogales B & Armengaud J (2012) Comparative proteogenomics of twelve *Roseobacter* exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol Cell Proteomics* **11**, M111.013110.

Attribution de l'émerine 67-170 (0M urée)

n° résidu	type résidu	atome	type	isotope	frequence (ppm)
2	THR	CA	C	13	62.02
2	THR	CB	C	13	70.02
3	ARG	H	H	1	8.55
3	ARG	C	C	13	176.83
3	ARG	CA	C	13	56.49
3	ARG	CB	C	13	30.79
3	ARG	N	N	15	123.36
4	GLY	H	H	1	8.48
4	GLY	C	C	13	174.12
4	GLY	CA	C	13	45.36
4	GLY	N	N	15	110.66
5	ASP	H	H	1	8.247
5	ASP	C	C	13	176.42
5	ASP	CA	C	13	54.72
5	ASP	CB	C	13	41.02
5	ASP	N	N	15	120.73
6	ALA	H	H	1	8.18
6	ALA	C	C	13	177.66
6	ALA	CA	C	13	52.88
6	ALA	CB	C	13	19.28
6	ALA	N	N	15	122.96
7	ASP	H	H	1	8.17
7	ASP	C	C	13	176.43
7	ASP	CA	C	13	54.6
7	ASP	CB	C	13	41.13
7	ASP	N	N	15	118.76
8	MET	H	H	1	8
8	MET	C	C	13	175.94
8	MET	CA	C	13	55.95
8	MET	CB	C	13	32.77
8	MET	N	N	15	119.49
9	TYR	H	H	1	8.016
9	TYR	C	C	13	175.33
9	TYR	CA	C	13	57.73
9	TYR	CB	C	13	38.79
9	TYR	N	N	15	119.92
10	ASP	H	H	1	8.163
10	ASP	C	C	13	175.49
10	ASP	CA	C	13	54.1
10	ASP	CB	C	13	41.28
10	ASP	N	N	15	121.36
11	LEU	H	H	1	7.97
11	LEU	C	C	13	175.2

11	LEU	CA	C	13	53.15
11	LEU	CB	C	13	42.03
11	LEU	N	N	15	123.31
12	PRO	C	C	13	176.87
12	PRO	CA	C	13	63.18
12	PRO	CB	C	13	32.16
13	LYS	H	H	1	8.37
13	LYS	C	C	13	176.92
13	LYS	CA	C	13	55.94
13	LYS	CB	C	13	33.23
13	LYS	N	N	15	121.77
14	LYS	H	H	1	8.42
14	LYS	C	C	13	176.88
14	LYS	CA	C	13	56.77
14	LYS	CB	C	13	33.04
14	LYS	N	N	15	123.6
15	GLU	H	H	1	8.66
15	GLU	C	C	13	176.45
15	GLU	CA	C	13	56.97
15	GLU	CB	C	13	30.03
15	GLU	N	N	15	121.79
16	ASP	H	H	1	8.22
16	ASP	C	C	13	176.42
16	ASP	CA	C	13	54.72
16	ASP	CB	C	13	41.27
16	ASP	N	N	15	121.33
17	ALA	H	H	1	8.115
17	ALA	C	C	13	178.22
17	ALA	CA	C	13	53.22
17	ALA	CB	C	13	19.15
17	ALA	N	N	15	123.85
18	LEU	H	H	1	8.07
18	LEU	C	C	13	177.82
18	LEU	CA	C	13	55.67
18	LEU	CB	C	13	41.94
18	LEU	N	N	15	119.48
19	LEU	H	H	1	7.86
19	LEU	C	C	13	177.4
19	LEU	CA	C	13	55.62
19	LEU	CB	C	13	42.19
19	LEU	N	N	15	121.22
20	TYR	H	H	1	7.91
20	TYR	C	C	13	175.88
20	TYR	CA	C	13	58.02
20	TYR	CB	C	13	38.52

20	TYR	N	N	15	119.27
21	GLN	H	H	1	8.03
21	GLN	C	C	13	175.82
21	GLN	CA	C	13	55.82
21	GLN	CB	C	13	29.6
21	GLN	N	N	15	121.26
22	SER	H	H	1	8.23
22	SER	C	C	13	174.67
22	SER	CA	C	13	58.65
22	SER	CB	C	13	63.74
22	SER	N	N	15	116.86
23	LYS	H	H	1	8.3
23	LYS	C	C	13	176.85
23	LYS	CA	C	13	56.42
23	LYS	CB	C	13	33.07
23	LYS	N	N	15	122.96
24	GLY	H	H	1	8.34
24	GLY	C	C	13	173.77
24	GLY	CA	C	13	45.17
24	GLY	N	N	15	109.55
25	TYR	H	H	1	7.941
25	TYR	C	C	13	175.5
25	TYR	CA	C	13	58.01
25	TYR	CB	C	13	38.88
25	TYR	N	N	15	119.63
26	ASN	H	H	1	8.35
26	ASN	C	C	13	174.7
26	ASN	CA	C	13	53.16
26	ASN	CB	C	13	39.16
26	ASN	N	N	15	119.93
27	ASP	H	H	1	8.18
27	ASP	C	C	13	175.94
27	ASP	CA	C	13	54.82
27	ASP	CB	C	13	41.22
27	ASP	N	N	15	120.46
28	ASP	H	H	1	8.21
28	ASP	C	C	13	175.85
28	ASP	CA	C	13	54.61
28	ASP	CB	C	13	41.01
28	ASP	N	N	15	119.66
29	TYR	H	H	1	7.95
29	TYR	C	C	13	175.29
29	TYR	CA	C	13	58.27
29	TYR	CB	C	13	38.88
29	TYR	N	N	15	120.14

30	TYR	H	H	1	7.85
30	TYR	C	C	13	175.21
30	TYR	CA	C	13	57.98
30	TYR	CB	C	13	39.17
30	TYR	N	N	15	122.06
31	GLU	H	H	1	8.05
31	GLU	C	C	13	176.26
31	GLU	CA	C	13	56.43
31	GLU	CB	C	13	30.49
31	GLU	N	N	15	122.62
32	GLU	H	H	1	8.37
32	GLU	C	C	13	176.89
32	GLU	CA	C	13	57.37
32	GLU	CB	C	13	30.12
32	GLU	N	N	15	121.95
33	SER	H	H	1	8.23
33	SER	C	C	13	174.6
33	SER	CA	C	13	58.85
33	SER	CB	C	13	63.67
33	SER	N	N	15	115.76
34	TYR	H	H	1	8.01
34	TYR	C	C	13	175.76
34	TYR	CA	C	13	58.62
34	TYR	CB	C	13	38.65
34	TYR	N	N	15	121.87
35	PHE	H	H	1	7.948
35	PHE	C	C	13	175.99
35	PHE	CA	C	13	58.04
35	PHE	CB	C	13	39.55
35	PHE	N	N	15	119.58
36	THR	H	H	1	7.94
36	THR	C	C	13	174.58
36	THR	CA	C	13	62.12
36	THR	CB	C	13	69.86
36	THR	N	N	15	114.6
37	THR	H	H	1	8.04
37	THR	C	C	13	174.45
37	THR	CA	C	13	62.12
37	THR	CB	C	13	69.81
37	THR	N	N	15	115.96
38	ARG	H	H	1	8.22
38	ARG	C	C	13	176.1
38	ARG	CA	C	13	56.06
38	ARG	CB	C	13	30.93
38	ARG	N	N	15	123.5

39	THR	H	H	1	8.11
39	THR	C	C	13	174.11
39	THR	CA	C	13	61.82
39	THR	CB	C	13	69.84
39	THR	N	N	15	115.49
40	TYR	H	H	1	8.24
40	TYR	C	C	13	176.1
40	TYR	CA	C	13	58.07
40	TYR	CB	C	13	38.99
40	TYR	N	N	15	122.46
41	GLY	H	H	1	8.23
41	GLY	C	C	13	173.6
41	GLY	CA	C	13	44.96
41	GLY	N	N	15	110.69
42	GLU	H	H	1	8.15
42	GLU	C	C	13	174.81
42	GLU	CA	C	13	54.46
42	GLU	CB	C	13	29.68
42	GLU	N	N	15	121.61
43	PRO	C	C	13	177.19
43	PRO	CA	C	13	63.4
43	PRO	CB	C	13	32.1
44	GLU	H	H	1	8.6
44	GLU	C	C	13	176.79
44	GLU	CA	C	13	57.03
44	GLU	CB	C	13	30.12
44	GLU	N	N	15	120.87
45	SER	H	H	1	8.25
45	SER	C	C	13	174.23
45	SER	CA	C	13	58.35
45	SER	CB	C	13	63.85
45	SER	N	N	15	116.35
46	ALA	H	H	1	8.28
46	ALA	C	C	13	177.75
46	ALA	CA	C	13	52.5
46	ALA	CB	C	13	19.54
46	ALA	N	N	15	125.78
47	GLY	H	H	1	8.12
47	GLY	C	C	13	178.02
47	GLY	CA	C	13	44.66
47	GLY	N	N	15	108.06
48	PRO	C	C	13	177.36
48	PRO	CA	C	13	63.36
48	PRO	CB	C	13	32.18
49	SER	H	H	1	8.38

49	SER	C	C	13	174.8
49	SER	CA	C	13	58.62
49	SER	CB	C	13	63.77
49	SER	N	N	15	115.83
50	ARG	H	H	1	8.299
50	ARG	C	C	13	175.9
50	ARG	CA	C	13	56.32
50	ARG	CB	C	13	30.8
50	ARG	N	N	15	122.93
51	ALA	H	H	1	8.2
51	ALA	C	C	13	177.65
51	ALA	CA	C	13	52.44
51	ALA	CB	C	13	19.3
51	ALA	N	N	15	124.76
52	VAL	H	H	1	8.045
52	VAL	C	C	13	176.2
52	VAL	CA	C	13	62.31
52	VAL	CB	C	13	32.83
52	VAL	N	N	15	119.57
53	ARG	H	H	1	8.36
53	ARG	C	C	13	176.1
53	ARG	CA	C	13	56.06
53	ARG	CB	C	13	30.91
53	ARG	N	N	15	124.75
54	GLN	H	H	1	8.44
54	GLN	C	C	13	175.86
54	GLN	CA	C	13	55.86
54	GLN	CB	C	13	29.64
54	GLN	N	N	15	122.04
55	SER	H	H	1	8.38
55	SER	C	C	13	174.47
55	SER	CA	C	13	58.33
55	SER	CB	C	13	63.8
55	SER	N	N	15	117.47
56	VAL	H	H	1	8.2
56	VAL	C	C	13	176.29
56	VAL	CA	C	13	62.4
56	VAL	CB	C	13	32.86
56	VAL	N	N	15	121.49
57	THR	H	H	1	8.17
57	THR	C	C	13	174.2
57	THR	CA	C	13	61.79
57	THR	CB	C	13	69.65
57	THR	N	N	15	117.06
58	SER	H	H	1	8.08

58	SER	C	C	13	173.35
58	SER	CA	C	13	58.03
58	SER	CB	C	13	63.9
58	SER	N	N	15	118.13
59	PHE	H	H	1	8.24
59	PHE	C	C	13	174.01
59	PHE	CA	C	13	55.57
59	PHE	CB	C	13	38.82
59	PHE	N	N	15	122.59
60	PRO	C	C	13	176.63
60	PRO	CA	C	13	63.56
60	PRO	CB	C	13	31.94
61	ASP	H	H	1	8.26
61	ASP	C	C	13	176.35
61	ASP	CA	C	13	54.37
61	ASP	CB	C	13	41.17
61	ASP	N	N	15	119.9
62	ALA	H	H	1	8.19
62	ALA	C	C	13	177.93
62	ALA	CA	C	13	53.15
62	ALA	CB	C	13	19.29
62	ALA	N	N	15	123.92
63	ASP	H	H	1	8.25
63	ASP	C	C	13	176.47
63	ASP	CA	C	13	54.63
63	ASP	CB	C	13	41.17
63	ASP	N	N	15	118.75
64	ALA	H	H	1	7.97
64	ALA	C	C	13	177.85
64	ALA	CA	C	13	53.16
64	ALA	CB	C	13	19.01
64	ALA	N	N	15	123.33
65	PHE	H	H	1	8.02
65	PHE	C	C	13	175.81
65	PHE	CA	C	13	58.04
65	PHE	CB	C	13	39.25
65	PHE	N	N	15	117.56
66	HIS	H	H	1	8.01
66	HIS	C	C	13	174.62
66	HIS	CA	C	13	55.89
66	HIS	CB	C	13	29.91
66	HIS	N	N	15	119.3
67	HIS	H	H	1	8.36
67	HIS	CA	C	13	55.85
67	HIS	N	N	15	121.96

68	GLN	C	C	13	175.61
68	GLN	CA	C	13	55.82
68	GLN	CB	C	13	29.72
69	VAL	H	H	1	8.18
69	VAL	C	C	13	175.49
69	VAL	CA	C	13	62.42
69	VAL	CB	C	13	32.77
69	VAL	N	N	15	121.46
70	HIS	H	H	1	8.54
70	HIS	C	C	13	174.56
70	HIS	CA	C	13	55.46
70	HIS	CB	C	13	30.05
70	HIS	N	N	15	122.66
71	ASP	H	H	1	8.38
71	ASP	C	C	13	176.15
71	ASP	CA	C	13	54.95
71	ASP	CB	C	13	41.2
71	ASP	N	N	15	122.46
72	ASP	H	H	1	8.44
72	ASP	C	C	13	176.58
72	ASP	CA	C	13	54.82
72	ASP	CB	C	13	41.15
72	ASP	N	N	15	120.31
73	ASP	H	H	1	8.25
73	ASP	C	C	13	176.56
73	ASP	CA	C	13	54.82
73	ASP	CB	C	13	41.12
73	ASP	N	N	15	120.36
74	LEU	H	H	1	8.07
74	LEU	C	C	13	177.72
74	LEU	CA	C	13	55.62
74	LEU	CB	C	13	42.18
74	LEU	N	N	15	121.6
75	LEU	H	H	1	8.11
75	LEU	C	C	13	177.72
75	LEU	CA	C	13	55.47
75	LEU	CB	C	13	42.12
75	LEU	N	N	15	121.86
76	SER	H	H	1	8.18
76	SER	C	C	13	174.75
76	SER	CA	C	13	58.53
76	SER	CB	C	13	63.79
76	SER	N	N	15	116.25
77	SER	H	H	1	8.3
77	SER	C	C	13	174.78

77	SER	CA	C	13	58.47
77	SER	CB	C	13	63.92
77	SER	N	N	15	117.56
78	SER	H	H	1	8.36
78	SER	C	C	13	174.9
78	SER	CA	C	13	58.67
78	SER	CB	C	13	63.87
78	SER	N	N	15	117.86
79	GLU	H	H	1	8.44
79	GLU	C	C	13	176.92
79	GLU	CA	C	13	57.22
79	GLU	CB	C	13	30.12
79	GLU	N	N	15	122.66
80	GLU	H	H	1	8.32
80	GLU	C	C	13	176.92
80	GLU	CA	C	13	57.17
80	GLU	CB	C	13	30.24
80	GLU	N	N	15	120.96
81	GLU	H	H	1	8.29
81	GLU	C	C	13	176.72
81	GLU	CA	C	13	57.22
81	GLU	CB	C	13	30.22
81	GLU	N	N	15	121.26
82	CYS	H	H	1	8.31
82	CYS	C	C	13	175.12
82	CYS	CA	C	13	59.02
82	CYS	CB	C	13	27.62
82	CYS	N	N	15	120.06
83	LYS	H	H	1	8.34
83	LYS	C	C	13	176.62
83	LYS	CA	C	13	56.92
83	LYS	CB	C	13	32.82
83	LYS	N	N	15	123.86
84	ASP	H	H	1	8.24
84	ASP	C	C	13	176.32
84	ASP	CA	C	13	54.82
84	ASP	CB	C	13	40.92
84	ASP	N	N	15	120.21
85	ARG	H	H	1	8.05
85	ARG	C	C	13	176.32
85	ARG	CA	C	13	56.33
85	ARG	CB	C	13	30.89
85	ARG	N	N	15	120.26
86	GLU	H	H	1	8.29
86	GLU	C	C	13	176.22

86	GLU	CA	C	13	56.52
86	GLU	CB	C	13	30.32
86	GLU	N	N	15	120.56
87	ARG	H	H	1	8.14
87	ARG	C	C	13	174.05
87	ARG	CA	C	13	54.01
87	ARG	CB	C	13	30.29
87	ARG	N	N	15	122.66
88	PRO	C	C	13	176.84
88	PRO	CA	C	13	63.14
88	PRO	CB	C	13	32.05
89	MET	H	H	1	8.37
89	MET	C	C	13	175.94
89	MET	CA	C	13	55.52
89	MET	CB	C	13	32.96
89	MET	N	N	15	120.22
90	TYR	H	H	1	8.07
90	TYR	C	C	13	176.36
90	TYR	CA	C	13	57.86
90	TYR	CB	C	13	38.81
90	TYR	N	N	15	120.43
91	GLY	H	H	1	8.34
91	GLY	C	C	13	174.28
91	GLY	CA	C	13	45.38
91	GLY	N	N	15	110.37
92	ARG	H	H	1	8.14
92	ARG	C	C	13	176.21
92	ARG	CA	C	13	56.33
92	ARG	CB	C	13	30.88
92	ARG	N	N	15	120.32
93	ASP	H	H	1	8.41
93	ASP	C	C	13	176.55
93	ASP	CA	C	13	54.53
93	ASP	CB	C	13	41.29
93	ASP	N	N	15	120.54
94	SER	H	H	1	8.17
94	SER	C	C	13	174.73
94	SER	CA	C	13	58.79
94	SER	CB	C	13	63.78
94	SER	N	N	15	116.1
95	ALA	H	H	1	8.29
95	ALA	C	C	13	177.77
95	ALA	CA	C	13	52.93
95	ALA	CB	C	13	18.98
95	ALA	N	N	15	125.33

96	TYR	H	H	1	7.95
96	TYR	C	C	13	175.88
96	TYR	CA	C	13	58.28
96	TYR	CB	C	13	38.57
96	TYR	N	N	15	118.56
97	GLN	H	H	1	8.03
97	GLN	C	C	13	175.6
97	GLN	CA	C	13	55.84
97	GLN	CB	C	13	29.6
97	GLN	N	N	15	121.15
98	SER	H	H	1	8.2
98	SER	C	C	13	174.76
98	SER	CA	C	13	58.52
98	SER	CB	C	13	63.77
98	SER	N	N	15	116.56
99	ILE	H	H	1	8.08
99	ILE	C	C	13	176.35
99	ILE	CA	C	13	61.46
99	ILE	CB	C	13	38.78
99	ILE	N	N	15	122.07
100	THR	H	H	1	8.07
100	THR	C	C	13	174.15
100	THR	CA	C	13	61.9
100	THR	CB	C	13	69.66
100	THR	N	N	15	117.21
101	HIS	C	C	13	174.45
101	HIS	CA	C	13	55.86
101	HIS	CB	C	13	30.2
102	TYR	H	H	1	8.16
102	TYR	C	C	13	174.98
102	TYR	CA	C	13	57.97
102	TYR	CB	C	13	38.82
102	TYR	N	N	15	121.96
103	ARG	H	H	1	8.06
103	ARG	C	C	13	173.29
103	ARG	CA	C	13	53.33
103	ARG	CB	C	13	30.85
103	ARG	N	N	15	125.31
104	PRO	C	C	13	176.06
104	PRO	CA	C	13	63.25
104	PRO	CB	C	13	31.89
105	VAL	H	H	1	7.64
105	VAL	C	C	13	175.12
105	VAL	CA	C	13	63.46
105	VAL	CB	C	13	33.33

105	VAL	N	N	15	123.86
-----	-----	---	---	----	--------

Attribution de l'émerine 67-170 (8M urée)

n° résidu	type résidu	atome	type	isotope	frequence (ppm)
1	GLY	H	H	1	7.99
1	GLY	CA	C	13	45.38
1	GLY	N	N	15	109.47
2	THR	H	H	1	8.27
2	THR	CA	C	13	61.94
2	THR	CB	C	13	70.31
2	THR	N	N	15	113.66
3	ARG	H	H	1	8.61
3	ARG	CA	C	13	56.51
3	ARG	CB	C	13	31.13
3	ARG	N	N	15	123.42
4	GLY	H	H	1	8.54
4	GLY	CA	C	13	45.47
4	GLY	N	N	15	110.15
5	ASP	CA	C	13	54.58
5	ASP	CB	C	13	41.49
6	ALA	H	H	1	8.32
6	ALA	CA	C	13	53
6	ALA	CB	C	13	19.39
6	ALA	N	N	15	123.56
7	ASP	H	H	1	8.37
7	ASP	CA	C	13	54.67
7	ASP	CB	C	13	41.52
7	ASP	N	N	15	118.97
8	MET	H	H	1	8.19
8	MET	CA	C	13	56.03
8	MET	CB	C	13	33.01
8	MET	N	N	15	119.9
9	TYR	H	H	1	8.2
9	TYR	CA	C	13	58.25
9	TYR	CB	C	13	39.05
9	TYR	N	N	15	120.18
10	ASP	H	H	1	8.36
10	ASP	CA	C	13	54.45
10	ASP	CB	C	13	41.66
10	ASP	N	N	15	121.6
11	LEU	H	H	1	8.14
11	LEU	CA	C	13	53.02
11	LEU	CB	C	13	41.96

11	LEU	N	N	15	123.28
12	PRO	CA	C	13	63.21
12	PRO	CB	C	13	32.52
13	LYS	H	H	1	8.57
13	LYS	CA	C	13	56.35
13	LYS	CB	C	13	33.37
13	LYS	N	N	15	122.4
14	LYS	H	H	1	8.64
14	LYS	CA	C	13	56.64
14	LYS	CB	C	13	33.33
14	LYS	N	N	15	124
15	GLU	H	H	1	8.78
15	GLU	CA	C	13	56.91
15	GLU	CB	C	13	30.45
15	GLU	N	N	15	122.85
16	ASP	H	H	1	8.46
16	ASP	CA	C	13	54.58
16	ASP	CB	C	13	41.62
16	ASP	N	N	15	121.96
17	ALA	H	H	1	8.32
17	ALA	CA	C	13	53.02
17	ALA	CB	C	13	19.36
17	ALA	N	N	15	124.2
18	LEU	H	H	1	8.26
18	LEU	CA	C	13	55.36
18	LEU	CB	C	13	42.16
18	LEU	N	N	15	120.64
19	LEU	H	H	1	8.16
19	LEU	CA	C	13	55.45
19	LEU	CB	C	13	42.49
19	LEU	N	N	15	122.36
20	TYR	H	H	1	8.21
20	TYR	CA	C	13	57.96
20	TYR	CB	C	13	38.88
20	TYR	N	N	15	120.44
21	GLN	H	H	1	8.3
21	GLN	CA	C	13	55.42
21	GLN	CB	C	13	29.95
21	GLN	N	N	15	122.06
22	SER	H	H	1	8.4
22	SER	CA	C	13	58.54
22	SER	CB	C	13	64.02
22	SER	N	N	15	117.39
23	LYS	H	H	1	8.53
23	LYS	CA	C	13	56.78

23	LYS	CB	C	13	33.33
23	LYS	N	N	15	123.32
24	GLY	H	H	1	8.48
24	GLY	CA	C	13	45.32
24	GLY	N	N	15	109.76
25	TYR	H	H	1	8.16
25	TYR	CA	C	13	58.13
25	TYR	CB	C	13	39.14
25	TYR	N	N	15	119.99
26	ASN	H	H	1	8.55
26	ASN	CA	C	13	53.27
26	ASN	CB	C	13	39.27
26	ASN	N	N	15	120.36
27	ASP	H	H	1	8.34
27	ASP	CA	C	13	54.65
27	27 ASP	CB	C		13
27	27 ASP	N	N		15
28	28 ASP	H	H		1
28	28 ASP	CA	C		13
28	28 ASP	CB	C		13
28	28 ASP	N	N		15
29	29 TYR	H	H		1
29	29 TYR	CA	C		13
29	29 TYR	CB	C		13
29	29 TYR	N	N		15
30	30 TYR	H	H		1
30	30 TYR	CA	C		13
30	30 TYR	CB	C		13
30	30 TYR	N	N		15
31	31 GLU	H	H		1
31	31 GLU	CA	C		13
31	31 GLU	CB	C		13
31	31 GLU	N	N		15
32	32 GLU	H	H		1
32	32 GLU	CA	C		13
32	32 GLU	CB	C		13
32	32 GLU	N	N		15
33	33 SER	H	H		1
33	33 SER	CA	C		13
33	33 SER	CB	C		13
33	33 SER	N	N		15
34	34 TYR	H	H		1
34	34 TYR	CA	C		13
34	34 TYR	CB	C		13
34	34 TYR	N	N		15

35	35 PHE	H	H	1
35	35 PHE	CA	C	13
35	35 PHE	CB	C	13
35	35 PHE	N	N	15
36	36 THR	H	H	1
36	36 THR	CA	C	13
36	36 THR	CB	C	13
36	36 THR	N	N	15
37	37 THR	H	H	1
37	37 THR	CA	C	13
37	37 THR	CB	C	13
37	37 THR	N	N	15
38	38 ARG	H	H	1
38	38 ARG	CA	C	13
38	38 ARG	CB	C	13
38	38 ARG	N	N	15
39	39 THR	H	H	1
39	39 THR	CA	C	13
39	39 THR	CB	C	13
39	39 THR	N	N	15
40	40 TYR	H	H	1
40	40 TYR	CA	C	13
40	40 TYR	CB	C	13
40	40 TYR	N	N	15
41	41 GLY	H	H	1
41	41 GLY	CA	C	13
41	41 GLY	N	N	15
42	42 GLU	H	H	1
42	42 GLU	CA	C	13
42	42 GLU	CB	C	13
42	42 GLU	N	N	15
43	43 PRO	CA	C	13
43	43 PRO	CB	C	13
44	44 GLU	H	H	1
44	44 GLU	CA	C	13
44	44 GLU	CB	C	13
44	44 GLU	N	N	15
45	45 SER	H	H	1
45	45 SER	CA	C	13
45	45 SER	CB	C	13
45	45 SER	N	N	15
46	46 ALA	H	H	1
46	46 ALA	CA	C	13
46	46 ALA	CB	C	13
46	46 ALA	N	N	15

47	47	GLY	H	H	1
47	47	GLY	CA	C	13
47	47	GLY	N	N	15
48	48	PRO	CA	C	13
48	48	PRO	CB	C	13
49	49	SER	H	H	1
49	49	SER	CA	C	13
49	49	SER	CB	C	13
49	49	SER	N	N	15
50	50	ARG	H	H	1
50	50	ARG	CA	C	13
50	50	ARG	CB	C	13
50	50	ARG	N	N	15
51	51	ALA	H	H	1
51	51	ALA	CA	C	13
51	51	ALA	CB	C	13
51	51	ALA	N	N	15
52	52	VAL	H	H	1
52	52	VAL	CA	C	13
52	52	VAL	CB	C	13
52	52	VAL	N	N	15
53	53	ARG	H	H	1
53	53	ARG	CA	C	13
53	53	ARG	CB	C	13
53	53	ARG	N	N	15
54	54	GLN	H	H	1
54	54	GLN	CA	C	13
54	54	GLN	CB	C	13
54	54	GLN	N	N	15
55	55	SER	H	H	1
55	55	SER	CA	C	13
55	55	SER	CB	C	13
55	55	SER	N	N	15
56	56	VAL	H	H	1
56	56	VAL	CA	C	13
56	56	VAL	CB	C	13
56	56	VAL	N	N	15
57	57	THR	H	H	1
57	57	THR	CA	C	13
57	57	THR	CB	C	13
57	57	THR	N	N	15
58	58	SER	H	H	1
58	58	SER	CA	C	13
58	58	SER	CB	C	13
58	58	SER	N	N	15

59	59 PHE	H	H	1
59	59 PHE	CA	C	13
59	59 PHE	CB	C	13
59	59 PHE	N	N	15
60	60 PRO	CA	C	13
60	60 PRO	CB	C	13
61	61 ASP	H	H	1
61	61 ASP	CA	C	13
61	61 ASP	CB	C	13
61	61 ASP	N	N	15
62	62 ALA	H	H	1
62	62 ALA	CA	C	13
62	62 ALA	CB	C	13
62	62 ALA	N	N	15
63	63 ASP	H	H	1
63	63 ASP	CA	C	13
63	63 ASP	CB	C	13
63	63 ASP	N	N	15
64	64 ALA	H	H	1
64	64 ALA	CA	C	13
64	64 ALA	CB	C	13
64	64 ALA	N	N	15
65	65 PHE	H	H	1
65	65 PHE	CA	C	13
65	65 PHE	CB	C	13
65	65 PHE	N	N	15
66	66 HIS	H	H	1
66	66 HIS	CA	C	13
66	66 HIS	CB	C	13
66	66 HIS	N	N	15
67	67 HIS	H	H	1
67	67 HIS	CA	C	13
67	67 HIS	N	N	15
68	68 GLN	CA	C	13
68	68 GLN	CB	C	13
69	69 VAL	H	H	1
69	69 VAL	CA	C	13
69	69 VAL	CB	C	13
69	69 VAL	N	N	15
70	70 HIS	H	H	1
70	70 HIS	CA	C	13
70	70 HIS	CB	C	13
70	70 HIS	N	N	15
71	71 ASP	H	H	1
71	71 ASP	CA	C	13

71	71 ASP	N	N	15
72	72 ASP	H	H	1
72	72 ASP	CA	C	13
72	72 ASP	CB	C	13
72	72 ASP	N	N	15
73	73 ASP	H	H	1
73	73 ASP	CA	C	13
73	73 ASP	CB	C	13
73	73 ASP	N	N	15
74	74 LEU	H	H	1
74	74 LEU	CA	C	13
74	74 LEU	CB	C	13
74	74 LEU	N	N	15
75	75 LEU	H	H	1
75	75 LEU	CA	C	13
75	75 LEU	CB	C	13
75	75 LEU	N	N	15
76	76 SER	H	H	1
76	76 SER	CA	C	13
76	76 SER	CB	C	13
76	76 SER	N	N	15
77	77 SER	H	H	1
77	77 SER	CA	C	13
77	77 SER	CB	C	13
77	77 SER	N	N	15
78	78 SER	H	H	1
78	78 SER	CA	C	13
78	78 SER	CB	C	13
78	78 SER	N	N	15
79	79 GLU	H	H	1
79	79 GLU	CA	C	13
79	79 GLU	CB	C	13
79	79 GLU	N	N	15
80	80 GLU	H	H	1
80	80 GLU	CA	C	13
80	80 GLU	CB	C	13
80	80 GLU	N	N	15
81	81 GLU	H	H	1
81	81 GLU	CA	C	13
81	81 GLU	CB	C	13
81	81 GLU	N	N	15
82	82 CYS	H	H	1
82	82 CYS	CA	C	13
82	82 CYS	CB	C	13
82	82 CYS	N	N	15

83	83	LYS	H	H	1
83	83	LYS	CA	C	13
83	83	LYS	CB	C	13
83	83	LYS	N	N	15
84	84	ASP	H	H	1
84	84	ASP	CA	C	13
84	84	ASP	CB	C	13
84	84	ASP	N	N	15
85	85	ARG	H	H	1
85	85	ARG	CA	C	13
85	85	ARG	CB	C	13
85	85	ARG	N	N	15
86	86	GLU	H	H	1
86	86	GLU	CA	C	13
86	86	GLU	CB	C	13
86	86	GLU	N	N	15
87	87	ARG	H	H	1
87	87	ARG	CA	C	13
87	87	ARG	CB	C	13
87	87	ARG	N	N	15
88	88	PRO	CA	C	13
88	88	PRO	CB	C	13
89	89	MET	H	H	1
89	89	MET	CA	C	13
89	89	MET	CB	C	13
89	89	MET	N	N	15
90	90	TYR	H	H	1
90	90	TYR	CA	C	13
90	90	TYR	CB	C	13
90	90	TYR	N	N	15
91	91	GLY	H	H	1
91	91	GLY	CA	C	13
91	91	GLY	N	N	15
92	92	ARG	H	H	1
92	92	ARG	CA	C	13
92	92	ARG	CB	C	13
92	92	ARG	N	N	15
93	93	ASP	H	H	1
93	93	ASP	CA	C	13
93	93	ASP	CB	C	13
93	93	ASP	N	N	15
94	94	SER	H	H	1
94	94	SER	CA	C	13
94	94	SER	CB	C	13
94	94	SER	N	N	15

95	95 ALA	H	H	1
95	95 ALA	CA	C	13
95	95 ALA	CB	C	13
95	95 ALA	N	N	15
96	96 TYR	H	H	1
96	96 TYR	CA	C	13
96	96 TYR	CB	C	13
96	96 TYR	N	N	15
97	97 GLN	H	H	1
97	97 GLN	CA	C	13
97	97 GLN	CB	C	13
97	97 GLN	N	N	15
98	98 SER	H	H	1
98	98 SER	CA	C	13
98	98 SER	CB	C	13
98	98 SER	N	N	15
99	99 ILE	H	H	1
99	99 ILE	CA	C	13
99	99 ILE	CB	C	13
99	99 ILE	N	N	15
100	100 THR	H	H	1
100	100 THR	CA	C	13
100	100 THR	CB	C	13
100	100 THR	N	N	15
101	101 HIS	CA	C	13
101	101 HIS	CB	C	13
102	102 TYR	H	H	1
102	102 TYR	CA	C	13
102	102 TYR	CB	C	13
102	102 TYR	N	N	15
103	103 ARG	H	H	1
103	103 ARG	CA	C	13
103	103 ARG	CB	C	13
103	103 ARG	N	N	15
104	104 PRO	CA	C	13
104	104 PRO	CB	C	13
105	105 VAL	H	H	1
105	105 VAL	CA	C	13
105	105 VAL	CB	C	13
105	105 VAL	N	N	15

Attribution du peptide ProgC_{ctoA}

n° résidu	type résidu	atome	type	isotope	frequence (ppm)
2	SER	H	H	1	8.17

2	SER	N	N	15	115.185
3	HIS	H	H	1	8.288
3	HIS	C	C	13	174.287
3	HIS	CA	C	13	55.455
3	HIS	CB	C	13	29.459
3	HIS	N	N	15	120.73
4	ALA	H	H	1	8.195
4	ALA	C	C	13	177.639
4	ALA	CA	C	13	52.44
4	ALA	CB	C	13	19.043
4	ALA	N	N	15	125.285
5	SER	H	H	1	8.299
5	SER	C	C	13	174.624
5	SER	CA	C	13	58.275
5	SER	CB	C	13	63.64
5	SER	N	N	15	115.499
6	SER	H	H	1	8.296
6	SER	C	C	13	174.541
6	SER	CA	C	13	58.536
6	SER	CB	C	13	63.716
6	SER	N	N	15	117.89
7	SER	H	H	1	8.244
7	SER	C	C	13	174.704
7	SER	CA	C	13	58.554
7	SER	CB	C	13	63.596
7	SER	N	N	15	117.566
8	GLY	H	H	1	8.16
8	GLY	C	C	13	173.277
8	GLY	CA	C	13	44.745
8	GLY	N	N	15	110.609
9	ASP	H	H	1	8.076
9	ASP	C	C	13	174.873
9	ASP	CA	C	13	52.152
9	ASP	CB	C	13	40.89
9	ASP	N	N	15	121.917
10	PRO	C	C	13	177.027
10	PRO	CA	C	13	63.553
10	PRO	CB	C	13	31.79
11	ALA	H	H	1	8.243
11	ALA	C	C	13	178.326
11	ALA	CA	C	13	53.037
11	ALA	CB	C	13	18.607
11	ALA	N	N	15	122.642
12	GLU	H	H	1	7.934
12	GLU	C	C	13	176.624

12	GLU	CA	C	13	56.845
12	GLU	CB	C	13	29.709
12	GLU	N	N	15	118.468
13	TYR	H	H	1	7.84
13	TYR	C	C	13	176.013
13	TYR	CA	C	13	58.549
13	TYR	CB	C	13	38.1
13	TYR	N	N	15	120.33
14	ASN	H	H	1	8.111
14	ASN	C	C	13	175.274
14	ASN	CA	C	13	53.535
14	ASN	CB	C	13	38.386
14	ASN	N	N	15	119.539
15	LEU	H	H	1	7.845
15	LEU	C	C	13	177.761
15	LEU	CA	C	13	55.923
15	LEU	CB	C	13	41.757
15	LEU	N	N	15	121.622
16	ARG	H	H	1	7.991
16	ARG	C	C	13	176.582
16	ARG	CA	C	13	56.517
16	ARG	CB	C	13	30.177
16	ARG	N	N	15	120.317
17	SER	H	H	1	7.989
17	SER	C	C	13	174.531
17	SER	CA	C	13	58.616
17	SER	CB	C	13	63.361
17	SER	N	N	15	115.749
18	ARG	H	H	1	8.09
18	ARG	C	C	13	176.387
18	ARG	CA	C	13	56.297
18	ARG	CB	C	13	30.394
18	ARG	N	N	15	122.501
19	THR	H	H	1	8.006
19	THR	C	C	13	174.309
19	THR	CA	C	13	62.354
19	THR	CB	C	13	69.536
19	THR	N	N	15	115.976
20	VAL	H	H	1	8.052
20	VAL	C	C	13	175.968
20	VAL	CA	C	13	62.318
20	VAL	CB	C	13	32.382
20	VAL	N	N	15	123.643
21	LEU	H	H	1	8.19
21	LEU	C	C	13	176.92

21	LEU	CA	C	13	54.824
21	LEU	CB	C	13	41.95
21	LEU	N	N	15	126.265
22	ALA	H	H	1	8.167
22	ALA	C	C	13	178.099
22	ALA	CA	C	13	52.577
22	ALA	CB	C	13	18.953
22	ALA	N	N	15	125.227
23	GLY	H	H	1	8.226
23	GLY	C	C	13	174.375
23	GLY	CA	C	13	45.113
23	GLY	N	N	15	108.117
24	THR	H	H	1	7.896
24	THR	C	C	13	174.444
24	THR	CA	C	13	61.682
24	THR	CB	C	13	69.738
24	THR	N	N	15	113.275
25	ALA	H	H	1	8.256
25	ALA	C	C	13	178.046
25	ALA	CA	C	13	52.651
25	ALA	CB	C	13	18.872
25	ALA	N	N	15	126.388
26	GLY	H	H	1	8.232
26	GLY	C	C	13	173.646
26	GLY	CA	C	13	44.849
26	GLY	N	N	15	108.1
27	GLN	H	H	1	8.014
27	GLN	C	C	13	173.889
27	GLN	CA	C	13	53.394
27	GLN	CB	C	13	28.491
27	GLN	N	N	15	120.549
28	PRO	CA	C	13	62.98
28	PRO	CB	C	13	31.856
29	ALA	H	H	1	8.349
29	ALA	C	C	13	177.515
29	ALA	CA	C	13	52.422
29	ALA	CB	C	13	18.881
29	ALA	N	N	15	124.192
30	ASP	H	H	1	8.15
30	ASP	C	C	13	176.285
30	ASP	CA	C	13	54.101
30	ASP	CB	C	13	40.944
30	ASP	N	N	15	119.25
31	LYS	H	H	1	8.077
31	LYS	C	C	13	176.49

31	LYS	CA	C	13	56.464
31	LYS	CB	C	13	32.62
31	LYS	N	N	15	121.84
32	ALA	H	H	1	8.169
32	ALA	C	C	13	177.957
32	ALA	CA	C	13	52.675
32	ALA	CB	C	13	18.819
32	ALA	N	N	15	124.579
33	SER	H	H	1	8.06
33	SER	C	C	13	174.457
33	SER	CA	C	13	58.34
33	SER	CB	C	13	63.605
33	SER	N	N	15	114.871
34	ALA	H	H	1	8.186
34	ALA	C	C	13	177.793
34	ALA	CA	C	13	52.515
34	ALA	CB	C	13	18.905
34	ALA	N	N	15	126.062
35	SER	H	H	1	8.137
35	SER	CA	C	13	58.451
35	SER	CB	C	13	63.612
35	SER	N	N	15	114.769
36	GLY	H	H	1	8.233
36	GLY	C	C	13	174.276
36	GLY	CA	C	13	45.153
36	GLY	N	N	15	110.911
37	SER	H	H	1	8.162
37	SER	C	C	13	175.054
37	SER	CA	C	13	58.471
37	SER	CB	C	13	63.662
37	SER	N	N	15	115.678
38	GLY	H	H	1	8.354
38	GLY	C	C	13	173.842
38	GLY	CA	C	13	45.071
38	GLY	N	N	15	111.015
39	ALA	H	H	1	7.997
39	ALA	C	C	13	177.642
39	ALA	CA	C	13	52.362
39	ALA	CB	C	13	18.985
39	ALA	N	N	15	123.592
40	GLN	H	H	1	8.245
40	GLN	C	C	13	175.707
40	GLN	CA	C	13	55.329
40	GLN	CB	C	13	29.156
40	GLN	N	N	15	119.389

41	SER	H	H	1	8.246
41	SER	C	C	13	172.608
41	SER	CA	C	13	56.411
41	SER	CB	C	13	62.993
41	SER	N	N	15	118.842
42	PRO	C	C	13	176.924
42	PRO	CA	C	13	63.229
42	PRO	CB	C	13	31.912
43	GLN	H	H	1	8.348
43	GLN	C	C	13	175.707
43	GLN	CA	C	13	55.967
43	GLN	CB	C	13	29.01
43	GLN	N	N	15	120.092
44	ASN	H	H	1	8.228
44	ASN	C	C	13	174.708
44	ASN	CA	C	13	53.082
44	ASN	CB	C	13	38.595
44	ASN	N	N	15	119.658
45	ALA	H	H	1	8.134
45	ALA	CA	C	13	52.477
45	ALA	CB	C	13	18.95
45	ALA	N	N	15	124.398
46	SER	H	H	1	8.169
46	SER	CA	C	13	58.342
46	SER	CB	C	13	63.551
46	SER	N	N	15	115.12
47	ILE	H	H	1	7.949
47	ILE	CA	C	13	61.192
47	ILE	CB	C	13	38.426
47	ILE	N	N	15	122.336
48	MET	H	H	1	7.817
48	MET	CA	C	13	56.804
48	MET	CB	C	13	33.371
48	MET	N	N	15	129.302

Attribution du peptide ProgC

n° résidu	type résidu	atome	type	isotope	frequence (ppm)
1	TRP	H	H	1	8.46
1	TRP	N	N	15	121.06
2	GLY	H	H	1	8.219

2	GLY	N	N	15	111.33
3	SER	H	H	1	7.991
3	SER	N	N	15	115.52
4	HIS	C	C	13	174.431
4	HIS	CA	C	13	55.729
4	HIS	CB	C	13	29.06
5	CYS	H	H	1	8.126
5	CYS	C	C	13	174.338
5	CYS	CA	C	13	58.404
5	CYS	CB	C	13	27.895
5	CYS	N	N	15	120.343
6	SER	H	H	1	8.382
6	SER	C	C	13	174.496
6	SER	CA	C	13	58.285
6	SER	CB	C	13	63.575
6	SER	N	N	15	118.567
7	SER	H	H	1	8.294
7	SER	C	C	13	174.475
7	SER	CA	C	13	58.356
7	SER	CB	C	13	63.472
7	SER	N	N	15	117.905
8	SER	H	H	1	8.189
8	SER	C	C	13	174.67
8	SER	CA	C	13	58.472
8	SER	CB	C	13	63.486
8	SER	N	N	15	117.441
9	GLY	H	H	1	8.135
9	GLY	C	C	13	173.219
9	GLY	CA	C	13	44.722
9	GLY	N	N	15	110.579
10	ASP	H	H	1	8.032
10	ASP	C	C	13	174.882
10	ASP	CA	C	13	51.994
10	ASP	CB	C	13	40.953
10	ASP	N	N	15	121.867
11	PRO	C	C	13	177.042
11	PRO	CA	C	13	63.557
11	PRO	CB	C	13	32.084
12	ALA	H	H	1	8.229
12	ALA	C	C	13	178.345
12	ALA	CA	C	13	53.044
12	ALA	CB	C	13	18.675
12	ALA	N	N	15	122.531
13	GLU	H	H	1	7.913
13	GLU	C	C	13	176.621

13	GLU	CA	C	13	56.882
13	GLU	CB	C	13	29.801
13	GLU	N	N	15	118.393
14	TYR	H	H	1	7.821
14	TYR	C	C	13	176.013
14	TYR	CA	C	13	58.452
14	TYR	CB	C	13	38.137
14	TYR	N	N	15	120.26
15	ASN	H	H	1	8.097
15	ASN	C	C	13	175.278
15	ASN	CA	C	13	53.588
15	ASN	CB	C	13	38.365
15	ASN	N	N	15	119.419
16	LEU	H	H	1	7.833
16	LEU	C	C	13	177.751
16	LEU	CA	C	13	55.819
16	LEU	CB	C	13	41.801
16	LEU	N	N	15	121.567
17	ARG	H	H	1	7.977
17	ARG	C	C	13	176.56
17	ARG	CA	C	13	56.479
17	ARG	CB	C	13	30.175
17	ARG	N	N	15	120.203
18	SER	H	H	1	7.975
18	SER	C	C	13	174.502
18	SER	CA	C	13	58.459
18	SER	CB	C	13	63.231
18	SER	N	N	15	115.635
19	ARG	H	H	1	8.073
19	ARG	C	C	13	176.314
19	ARG	CA	C	13	56.229
19	ARG	CB	C	13	30.416
19	ARG	N	N	15	122.443
20	THR	H	H	1	7.986
20	THR	C	C	13	174.284
20	THR	CA	C	13	62.257
20	THR	CB	C	13	69.28
20	THR	N	N	15	115.69
21	VAL	H	H	1	8.018
21	VAL	C	C	13	175.87
21	VAL	CA	C	13	62.214
21	VAL	CB	C	13	32.425
21	VAL	N	N	15	123.315
22	LEU	H	H	1	8.209
22	LEU	C	C	13	177.072

22	LEU	CA	C	13	54.874
22	LEU	CB	C	13	41.913
22	LEU	N	N	15	125.925
23	CYS	H	H	1	8.214
23	CYS	C	C	13	174.825
23	CYS	CA	C	13	58.512
23	CYS	CB	C	13	27.943
23	CYS	N	N	15	120.216
24	GLY	H	H	1	8.345
24	GLY	C	C	13	174.158
24	GLY	CA	C	13	45.169
24	GLY	N	N	15	111.347
25	THR	H	H	1	8
25	THR	C	C	13	174.617
25	THR	CA	C	13	61.683
25	THR	CB	C	13	69.497
25	THR	N	N	15	113.309
26	CYS	H	H	1	8.329
26	CYS	C	C	13	174.84
26	CYS	CA	C	13	58.539
26	CYS	CB	C	13	27.745
26	CYS	N	N	15	121.068
27	GLY	H	H	1	8.337
27	GLY	CA	C	13	45.02
27	GLY	N	N	15	111.085
28	GLN	H	H	1	8.064
28	GLN	C	C	13	173.854
28	GLN	CA	C	13	53.41
28	GLN	CB	C	13	28.645
28	GLN	N	N	15	120.602
29	PRO	CA	C	13	62.919
29	PRO	CB	C	13	31.931
30	ALA	H	H	1	8.339
30	ALA	C	C	13	177.493
30	ALA	CA	C	13	52.433
30	ALA	CB	C	13	19.007
30	ALA	N	N	15	124.111
31	ASP	H	H	1	8.139
31	ASP	C	C	13	176.277
31	ASP	CA	C	13	53.985
31	ASP	CB	C	13	40.88
31	ASP	N	N	15	119.191
32	LYS	H	H	1	8.07
32	LYS	C	C	13	176.479
32	LYS	CA	C	13	56.355

32	LYS	CB	C	13	32.598
32	LYS	N	N	15	121.766
33	ALA	H	H	1	8.16
33	ALA	C	C	13	177.953
33	ALA	CA	C	13	52.627
33	ALA	CB	C	13	18.961
33	ALA	N	N	15	124.531
34	SER	H	H	1	8.053
34	SER	C	C	13	174.443
34	SER	CA	C	13	58.226
34	SER	CB	C	13	63.462
34	SER	N	N	15	114.804
35	ALA	H	H	1	8.18
35	ALA	C	C	13	177.786
35	ALA	CA	C	13	52.532
35	ALA	CB	C	13	19.018
35	ALA	N	N	15	126.028
36	SER	H	H	1	8.131
36	SER	C	C	13	173.97
36	SER	CA	C	13	58.381
36	SER	CB	C	13	63.489
36	SER	N	N	15	114.729
37	GLY	H	H	1	8.225
37	GLY	C	C	13	174.247
37	GLY	CA	C	13	45.143
37	GLY	N	N	15	110.88
38	SER	H	H	1	8.153
38	SER	C	C	13	175.045
38	SER	CA	C	13	58.566
38	SER	CB	C	13	63.521
38	SER	N	N	15	115.643
39	GLY	H	H	1	8.336
39	GLY	C	C	13	173.816
39	GLY	CA	C	13	45.01
39	GLY	N	N	15	111.012
40	ALA	H	H	1	7.994
40	ALA	C	C	13	177.621
40	ALA	CA	C	13	52.442
40	ALA	CB	C	13	18.992
40	ALA	N	N	15	123.619
41	GLN	H	H	1	8.236
41	GLN	C	C	13	175.686
41	GLN	CA	C	13	55.369
41	GLN	CB	C	13	29.36
41	GLN	N	N	15	119.336

42	SER	H	H	1	8.238
42	SER	C	C	13	172.599
42	SER	CA	C	13	56.515
42	SER	CB	C	13	62.897
42	SER	N	N	15	118.783
43	PRO	C	C	13	176.875
43	PRO	CA	C	13	63.324
43	PRO	CB	C	13	31.854
44	GLN	H	H	1	8.329
44	GLN	C	C	13	175.688
44	GLN	CA	C	13	55.806
44	GLN	CB	C	13	29.207
44	GLN	N	N	15	119.953
45	ASN	H	H	1	8.289
45	ASN	C	C	13	174.915
45	ASN	CA	C	13	53.256
45	ASN	CB	C	13	38.529
45	ASN	N	N	15	119.568
46	CYS	H	H	1	8.168
46	CYS	C	C	13	174.285
46	CYS	CA	C	13	58.315
46	CYS	CB	C	13	27.978
46	CYS	N	N	15	119.462
47	SER	H	H	1	8.275
47	SER	C	C	13	173.958
47	SER	CA	C	13	58.334
47	SER	CB	C	13	63.451
47	SER	N	N	15	118.337
48	ILE	H	H	1	7.975
48	ILE	C	C	13	175.208
48	ILE	CA	C	13	61.263
48	ILE	CB	C	13	38.372
48	ILE	N	N	15	122.491
49	MET	H	H	1	7.825
49	MET	N	N	15	129.37

Attribution du peptide PreLamC_{CtoA}

n° résidu	type résidu	atome	type	isotope	frequence (ppm)
2	MET	CA	C	13	55.53
2	MET	CB	C	13	32.55
2	MET	N	N	15	120.1
3	ALA	H	H	1	8.27
3	ALA	CA	C	13	52.77

3	ALA	CB	C	13	18.68
3	ALA	N	N	15	125.87
4	SER	H	H	1	8.18
4	SER	CA	C	13	55.452
4	SER	CB	C	13	61.041
4	SER	N	N	15	115.09
5	HIS	H	H	1	8.31
5	HIS	CA	C	13	55.51
5	HIS	CB	C	13	29.28
5	HIS	N	N	15	120.59
6	ALA	H	H	1	8.206
6	ALA	CA	C	13	52.75
6	ALA	CB	C	13	18.81
6	ALA	N	N	15	125.239
7	SER	H	H	1	8.32
7	SER	CA	C	13	58.56
7	SER	CB	C	13	63.92
7	SER	N	N	15	115.53
8	SER	H	H	1	8.32
8	SER	CA	C	13	58.6
8	SER	CB	C	13	63.92
8	SER	N	N	15	117.96
9	SER	H	H	1	8.26
9	SER	CA	C	13	58.77
9	SER	CB	C	13	63.97
9	SER	N	N	15	117.65
10	GLY	H	H	1	8.2
10	GLY	CA	C	13	44.82
10	GLY	N	N	15	110.59
11	ASP	H	H	1	8.1
11	ASP	CA	C	13	52.28
11	ASP	CB	C	13	40.96
11	ASP	N	N	15	121.91
12	PRO	CA	C	13	63.89
12	PRO	CB	C	13	31.68
13	ALA	H	H	1	8.28
13	ALA	CA	C	13	53.15
13	ALA	CB	C	13	18.47
13	ALA	N	N	15	122.78
14	GLU	H	H	1	7.97
14	GLU	CA	C	13	56.93
14	GLU	CB	C	13	29.72
14	GLU	N	N	15	118.54
15	TYR	H	H	1	7.866
15	TYR	CA	C	13	58.61

15	TYR	CB	C	13	38.25
15	TYR	N	N	15	120.299
16	ASN	H	H	1	8.154
16	ASN	CA	C	13	53.55
16	ASN	CB	C	13	38.41
16	ASN	N	N	15	119.718
17	LEU	H	H	1	7.889
17	LEU	CA	C	13	55.92
17	LEU	CB	C	13	41.87
17	LEU	N	N	15	121.72
18	ARG	H	H	1	8.05
18	ARG	CA	C	13	56.52
18	ARG	CB	C	13	30.24
18	ARG	N	N	15	120.5
19	SER	H	H	1	8.035
19	SER	CA	C	13	58.65
19	SER	CB	C	13	63.83
19	SER	N	N	15	116.05
20	ARG	H	H	1	8.13
20	ARG	CA	C	13	56.48
20	ARG	CB	C	13	30.32
20	ARG	N	N	15	122.63
21	THR	H	H	1	8.045
21	THR	CA	C	13	62.4
21	THR	CB	C	13	69.93
21	THR	N	N	15	116.08
22	VAL	H	H	1	8.078
22	VAL	CA	C	13	62.41
22	VAL	CB	C	13	32.27
22	VAL	N	N	15	123.64
23	LEU	H	H	1	8.203
23	LEU	CA	C	13	55.12
23	LEU	CB	C	13	41.94
23	LEU	N	N	15	126.1
24	ALA	H	H	1	8.177
24	ALA	CA	C	13	52.85
24	ALA	CB	C	13	18.7
24	ALA	N	N	15	125.125
25	GLY	H	H	1	8.245
25	GLY	CA	C	13	45.2
25	GLY	N	N	15	108.172
26	THR	H	H	1	7.911
26	THR	CA	C	13	61.98
26	THR	CB	C	13	70.14
26	THR	N	N	15	113.23

27	ALA	H	H	1	8.273
27	ALA	CA	C	13	52.84
27	ALA	CB	C	13	18.6
27	ALA	N	N	15	126.31
28	GLY	H	H	1	8.25
28	GLY	CA	C	13	45.07
28	GLY	N	N	15	108.14
29	GLN	H	H	1	8.036
29	GLN	CA	C	13	53.66
29	GLN	CB	C	13	28.47
29	GLN	N	N	15	120.49
30	PRO	CA	C	13	63.34
30	PRO	CB	C	13	31.79
31	ALA	H	H	1	8.369
31	ALA	CA	C	13	52.49
31	ALA	CB	C	13	18.74
31	ALA	N	N	15	124.24
32	ASP	H	H	1	8.176
32	ASP	CA	C	13	54.23
32	ASP	CB	C	13	40.92
32	ASP	N	N	15	119.249
33	LYS	H	H	1	8.096
33	LYS	CA	C	13	56.66
33	LYS	CB	C	13	32.66
33	LYS	N	N	15	121.89
34	ALA	H	H	1	8.192
34	ALA	CA	C	13	52.95
34	ALA	CB	C	13	18.57
34	ALA	N	N	15	124.642
35	SER	H	H	1	8.079
35	SER	CA	C	13	58.56
35	SER	CB	C	13	63.99
35	SER	N	N	15	114.888
36	ALA	H	H	1	8.23
36	ALA	CA	C	13	52.65
36	ALA	CB	C	13	18.72
36	ALA	N	N	15	125.94
37	SER	H	H	1	8.18
37	SER	CA	C	13	58.67
37	SER	CB	C	13	63.96
37	SER	N	N	15	114.8
38	GLY	H	H	1	8.26
38	GLY	CA	C	13	45.27
38	GLY	N	N	15	110.96
39	SER	H	H	1	8.186

39	SER	CA	C	13	58.8
39	SER	CB	C	13	64
39	SER	N	N	15	115.71
40	GLY	H	H	1	8.38
40	GLY	CA	C	13	45.16
40	GLY	N	N	15	111.056
41	ALA	H	H	1	7.987
41	ALA	CA	C	13	52.51
41	ALA	CB	C	13	18.76
41	ALA	N	N	15	123.54
42	GLN	H	H	1	8.283
42	GLN	CA	C	13	55.74
42	GLN	CB	C	13	29.02
42	GLN	N	N	15	119.913
43	VAL	H	H	1	8.141
43	VAL	CA	C	13	62.69
43	VAL	CB	C	13	32.36
43	VAL	N	N	15	122.088
44	GLY	H	H	1	8.43
44	GLY	CA	C	13	45
44	GLY	N	N	15	112.959
45	GLY	H	H	1	7.994
45	GLY	CA	C	13	44.49
45	GLY	N	N	15	108.869
46	PRO	CA	C	13	63.21
46	PRO	CB	C	13	31.84
47	ILE	H	H	1	8.219
47	ILE	CA	C	13	61.38
47	ILE	CB	C	13	38.46
47	ILE	N	N	15	121.246
48	SER	H	H	1	8.315
48	SER	CA	C	13	58.42
48	SER	CB	C	13	63.95
48	SER	N	N	15	120.017
49	SER	H	H	1	8.332
49	SER	CA	C	13	58.68
49	SER	CB	C	13	63.91
49	SER	N	N	15	118.365
50	GLY	H	H	1	8.327
50	GLY	CA	C	13	45.25
50	GLY	N	N	15	110.83
51	SER	H	H	1	8.128
51	SER	CA	C	13	58.43
51	SER	CB	C	13	63.97
51	SER	N	N	15	115.659

52	SER	H	H	1	8.317
52	SER	CA	C	13	58.56
52	SER	CB	C	13	63.99
52	SER	N	N	15	117.97
53	ALA	H	H	1	8.228
53	ALA	CA	C	13	52.81
53	ALA	CB	C	13	18.73
53	ALA	N	N	15	125.879
54	SER	H	H	1	8.132
54	SER	CA	C	13	58.49
54	SER	CB	C	13	63.87
54	SER	N	N	15	114.634
55	SER	H	H	1	8.162
55	SER	CA	C	13	58.47
55	SER	CB	C	13	64.02
55	SER	N	N	15	117.96
56	VAL	H	H	1	7.992
56	VAL	CA	C	13	62.41
56	VAL	CB	C	13	32.41
56	VAL	N	N	15	121.56
57	THR	H	H	1	8.169
57	THR	CA	C	13	63.34
57	THR	CB	C	13	69.97
57	THR	N	N	15	119.17
58	VAL	H	H	1	7.99
58	VAL	CA	C	13	62.36
58	VAL	CB	C	13	32.34
58	VAL	N	N	15	121.55
59	THR	H	H	1	8.14
59	THR	CA	C	13	62.03
59	THR	CB	C	13	69.96
59	THR	N	N	15	118.97
60	ARG	H	H	1	8.27
60	ARG	CA	C	13	56.03
60	ARG	CB	C	13	30.57
60	ARG	N	N	15	124.03
61	SER	H	H	1	8.19
61	SER	CA	C	13	58.37
61	SER	CB	C	13	63.91
61	SER	N	N	15	116.9
62	TYR	H	H	1	8.13
62	TYR	CA	C	13	58.32
62	TYR	CB	C	13	38.58
62	TYR	N	N	15	122.53
63	ARG	H	H	1	8.02

63	ARG	CA	C	13	55.84
63	ARG	CB	C	13	30.69
63	ARG	N	N	15	122.68
64	SER	H	H	1	8.19
64	SER	CA	C	13	58.37
64	SER	CB	C	13	63.87
64	SER	N	N	15	117.49
65	VAL	H	H	1	8.11
65	VAL	CA	C	13	62.51
65	VAL	CB	C	13	32.39
65	VAL	N	N	15	121.81
66	GLY	H	H	1	8.37
66	GLY	CA	C	13	45.1
66	GLY	N	N	15	112.34
67	GLY	H	H	1	8.115
67	GLY	CA	C	13	45.09
67	GLY	N	N	15	108.83
68	SER	H	H	1	8.27
68	SER	CA	C	13	58.72
68	SER	CB	C	13	64
68	SER	N	N	15	115.74
69	GLY	H	H	1	8.42
69	GLY	CA	C	13	45.23
69	GLY	N	N	15	111.002
70	GLY	H	H	1	8.182
70	GLY	CA	C	13	45.15
70	GLY	N	N	15	108.78
71	GLY	H	H	1	8.184
71	GLY	CA	C	13	45.03
71	GLY	N	N	15	108.79
72	SER	H	H	1	8.1
72	SER	CA	C	13	58.47
72	SER	CB	C	13	63.98
72	SER	N	N	15	115.704
73	PHE	H	H	1	8.243
73	PHE	CA	C	13	58.47
73	PHE	CB	C	13	39.21
73	PHE	N	N	15	122.061
74	GLY	H	H	1	8.189
74	GLY	CA	C	13	45
74	GLY	N	N	15	110.365
75	ASP	H	H	1	8.04
75	ASP	CA	C	13	53.96
75	ASP	CB	C	13	40.85
75	ASP	N	N	15	120.44

76	ASN	H	H	1	8.24
76	ASN	CA	C	13	53.42
76	ASN	CB	C	13	38.41
76	ASN	N	N	15	118.3
77	LEU	H	H	1	8.04
77	LEU	CA	C	13	55.58
77	LEU	CB	C	13	41.96
77	LEU	N	N	15	121.76
78	VAL	H	H	1	7.97
78	VAL	CA	C	13	62.62
78	VAL	CB	C	13	32.34
78	VAL	N	N	15	121.13
79	THR	H	H	1	8.14
79	THR	CA	C	13	62.47
79	THR	CB	C	13	70.01
79	THR	N	N	15	118.96
80	ARG	H	H	1	8.13
80	ARG	CB	C	13	32.45
80	ARG	N	N	15	123.54
84	LEU	CA	C	13	55.18
84	LEU	CB	C	13	41.96
85	GLY	H	H	1	8.11
85	GLY	CA	C	13	45.29
85	GLY	N	N	15	109
86	ASN	H	H	1	8.15
86	ASN	CA	C	13	52.66
86	ASN	CB	C	13	38.44
86	ASN	N	N	15	118.84
87	SER	H	H	1	8.16
87	SER	N	N	15	116.14
90	ARG	H	H	1	8.13
90	ARG	CA	C	13	56.21
90	ARG	CB	C	13	30.5
90	ARG	N	N	15	122.83
91	THR	H	H	1	8.1
91	THR	CA	C	13	62.39
91	THR	CB	C	13	69.99
91	THR	N	N	15	116.43
92	GLN	CA	C	13	55.33
92	GLN	CB	C	13	29.45
92	GLN	N	N	15	122.57
93	SER	H	H	1	8.34
93	SER	CA	C	13	56.88
93	SER	CB	C	13	63.04
93	SER	N	N	15	119.11

94	PRO	CA	C	13	63.24
94	PRO	CB	C	13	32.1
95	GLN	H	H	1	8.38
95	GLN	CA	C	13	55.91
95	GLN	CB	C	13	28.96
95	GLN	N	N	15	120.11
96	ASN	H	H	1	8.23
96	ASN	CA	C	13	52.88
96	ASN	CB	C	13	38.67
96	ASN	N	N	15	119.5
97	ALA	H	H	1	8.14
97	ALA	CA	C	13	52.77
97	ALA	CB	C	13	18.68
97	ALA	N	N	15	124.36
98	SER	H	H	1	8.18
98	SER	CA	C	13	58.51
98	SER	CB	C	13	63.67
98	SER	N	N	15	115.08
99	ILE	H	H	1	7.94
99	ILE	CA	C	13	61.16
99	ILE	CB	C	13	38.37
99	ILE	N	N	15	122.19
100	MET	H	H	1	7.83
100	MET	CA	C	13	57.08
100	MET	CB	C	13	33.46
100	MET	N	N	15	129.15

Détermination d'un facteur de forme des fibres au cours du temps

Utilisation d'un système de mélange rapide

Pour suivre la formation de fibres au cours du temps, il est primordial d'obtenir des informations précises sur les étapes d'initiation de cette polymérisation à partir des monomères protéiques. Malheureusement ces étapes sont difficiles à obtenir. Ainsi, un système de mélange rapide fut nécessaire pour les étudier. L'équipe du Dr. Javier Perez a donc pour cela choisi d'utiliser un appareil de Stopped-Flow développé par la société Bio-Logic. Ce système permet également d'injecter les réactifs permettant la polymérisation simultanément dans le circuit de WA-MALS ainsi que dans le capillaire de SAXS afin d'obtenir un temps initial T_0 simultané pour les deux analyses.

Analyse des données de WA-MALS

Les données de WA-MALS ont pour but d'extraire la masse molaire ainsi que le rayon de giration des polymères observés au cours du temps. Les données ont été analysées grâce à la méthode de Zimm (Zimm B., 1948) dans laquelle $K^*c/R(\Theta)$ est représenté en fonction de $\sin^2(\Theta/2)$.

Avec :

- Θ : Angle de diffusion
- K^* : Constante optique du système ($\text{cm}^2 \cdot \text{g}^{-2}$) (Equation 1)
- c : Concentration de l'échantillon
- $R(\Theta)$: Ratio de Rayleigh. Le ratio de Rayleigh représente l'excédent de diffusion par rapport à la diffusion du solvant seul pour un détecteur centré à l'angle Θ .

K^* est reliée à l'indice de réfraction de l'échantillon par la relation suivante :

$$\text{Equation 1: } K^* = \frac{4\pi^2 n_0^2 \left(\frac{dn}{dc}\right)^2}{N_A \lambda^4}$$

Avec :

- n_0 : indice de réfraction du solvant

- $\frac{dn}{dc}$ ($\text{cm}^3 \cdot \text{g}^{-1}$) : indice de réfraction de l'échantillon
- N_A : Nombre d'Avogadro
- λ : Longueur d'onde de la lumière incidente dans le vide ($\text{g} \cdot \text{cm}^{-3}$)

Il est alors possible d'extraire la masse molaire pondérée $\langle M_w \rangle$ (Equation 2) ainsi que le carré du rayon de giration $\langle R_g^2 \rangle$ (Equation 3) à partir de la représentation de Zimm.

$$\text{Equation 2 : } \langle M_w \rangle = \sum M_i \cdot w_i$$

$$\text{Equation 3 : } \langle R_g^2 \rangle = \frac{\sum M_i \cdot w_i \cdot (R_g^2)_i}{\sum M_i \cdot w_i}$$

avec :

- M_i : masse molaire
- R_g^2 : carré du rayon de giration
- w_i : fraction pondérée de l'espèce i

L'équation de Zimm est alors :

$$\text{Equation 4 : } \frac{K^*c}{R(\theta)} = \frac{1}{\langle M \rangle_w} \left(1 + \frac{16\pi^2 \cdot n_0^2 \cdot \langle R_g^2 \rangle_z}{3\lambda^2} \cdot \sin^2 \frac{\theta}{2} \right)$$

Lorsque $\theta = 0$, on obtient :

$$\text{Equation 5 : } \frac{K^*c}{R(0)} = \frac{1}{\langle M \rangle_w}$$

On trace alors la courbe $\frac{K^*c}{R(\theta)}$ en fonction de $\sin^2 \frac{\theta}{2} + kc$ où k est une constante arbitraire qui sert à espacer les courbes pour les différents c , sur un même diagramme : le diagramme de Zimm (Figure 1). On peut alors construire les courbes pour $\theta = 0$ et $c = 0$, en extrapolant les courbes pour c constante et θ constant respectivement. Les courbes pour $\theta = 0$ et $c = 0$ devraient alors intercepter l'axe $\frac{K^*c}{R(\theta)}$ en un même point correspondant à $\frac{1}{\langle M \rangle_w}$.

L'intersection de la courbe donnera alors $\frac{1}{\langle M \rangle_w}$.

Pour trouver $\langle R_g^2 \rangle_z$, nous utilisons le fait que la courbe à $c = 0$ a la forme :

Equation 6 : $\frac{K^*c}{R(\theta)} = a + b \cdot \sin^2 \frac{\theta}{2}$ où $b = 1 + \frac{16\pi^2 \cdot n_0^2 \cdot \langle R_g^2 \rangle_z}{3\lambda^2}$ on peut alors mesurer la pente de cette courbe pour $c = 0$ et, connaissant la valeur de λ , il nous est facile de trouver $\langle R_g^2 \rangle_z$.

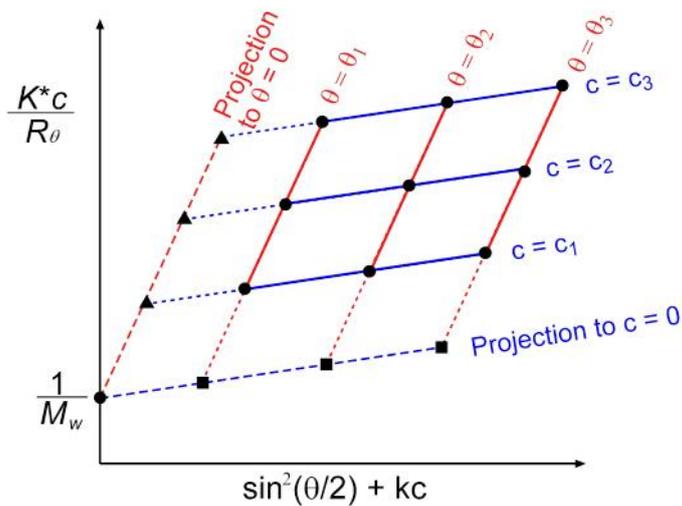


Figure 82 : Représentation de Zimm. Les points représentent différentes valeurs de concentration (indiquées en bleu) mais également des angles de diffusion différents (indiqués en rouge).

Ainsi, les données de WA-MALS nous permettent d'extraire la masse moyenne ainsi que le rayon de giration des filaments formés.

Modélisation des données

La méthode de Zimm a d'abord été développée pour des particules monodisperses. Lors de l'étude de solutions de particules monodisperses et sphériques, une méthode de régression linéaire est utilisée et le résultat est une droite. Malheureusement, lorsqu'une solution contient des particules polydisperses et allongées, la méthode de Zimm donne des courbes. L'extrapolation à $\theta = 0$ ne peut alors être faite qu'avec des polynômes. Une limite intrinsèque à la méthode de Zimm est la nécessité de choisir des angles ainsi que les degrés des polynômes lors de sa modélisation. De plus, les résultats obtenus à des θ très petits étant très bruités. Il est donc très difficile de choisir le bon polynôme ainsi que les angles θ à inclure. Une seconde méthode de modélisation, plus robuste a alors dû être employée.

Pour la modélisation de nos données de MALS, nous utiliserons l'équation de Fisher-Burford. La méthode de Zimm utilisait $\frac{K^*c}{R(\theta)}$ permettant d'obtenir $\frac{1}{\langle M \rangle_w}$. Dans la méthode de Fisher-Burford, nous calculons $I^*(q)$ qui est en fait $\frac{R(\theta)}{K^*c}$ donnant alors par intersection directement $\langle M \rangle_w$.

L'équation de Fisher-Burford est alors décrite par :

$$\text{Equation 7 : } I^*(q) = \frac{\langle M \rangle_w}{[1+q^2 \langle \xi^2 \rangle_z] \frac{D_m}{2}}$$

Où :

- $I^*(q)$ est l'amplitude de l'intensité diffusée.
- q est le vecteur de diffusion $q = 4\pi \frac{\sin \frac{\theta}{2}}{\lambda}$

La limite de $I^*(q)$ lorsque q tend vers 0 fournit directement le paramètre $\langle M \rangle_w$.

Le paramètre $\langle R_g^2 \rangle_z$ est obtenu à partir de $\langle \xi^2 \rangle_z$ grâce à la relation

$$\text{Equation 8 : } \langle R_g^2 \rangle_z = \frac{3D_m}{2} \langle \xi^2 \rangle_z$$

La fonction de reconstruction aura donc comme paramètres libres : $\langle \xi^2 \rangle_z$ et D_m .

Une comparaison des valeurs de $\langle R_g \rangle$ et $\langle M \rangle$ obtenues respectivement par les méthodes de Fisher-Burford et de Zimm est visible en Figure 2.

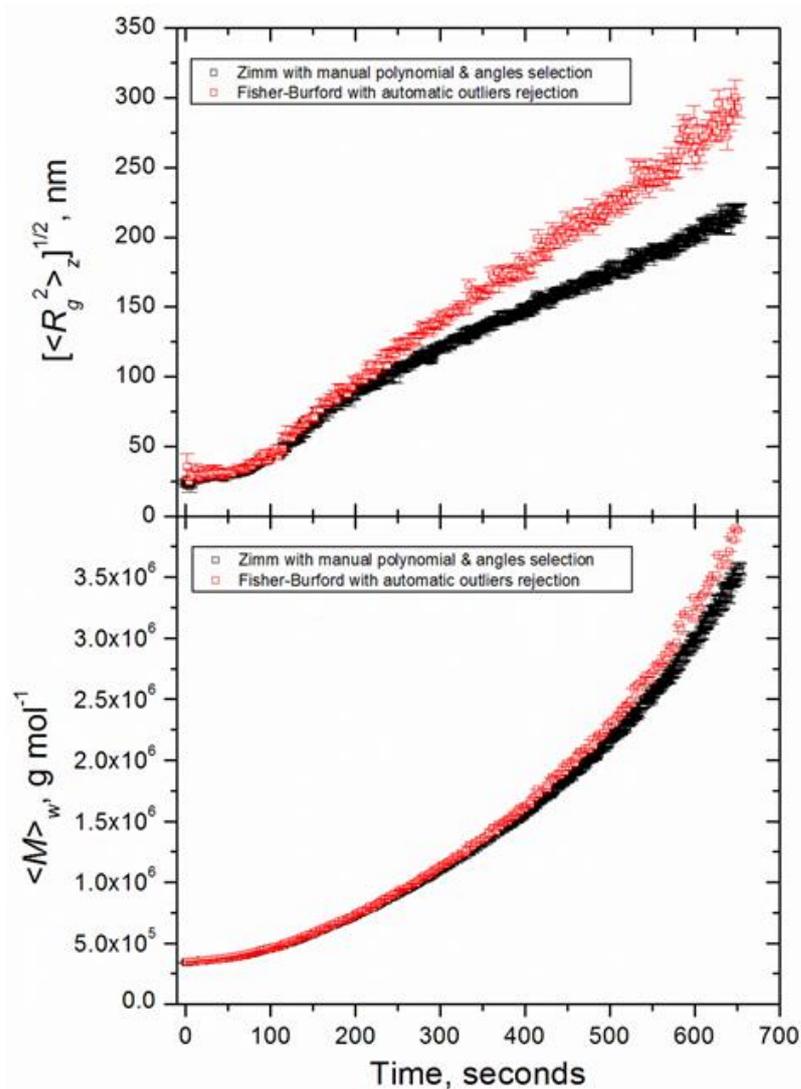


Figure 83 : Comparaison des graphiques obtenus par la méthode de Zimm après avoir choisi manuellement les degrés de polynômes ainsi que les angles à inclure, avec les résultats obtenus avec l'équation de Fisher-Burford.

La fonction de Fisher-Burford semble fonctionner assez bien avec les données de MALS enregistrées sur le fibrinogène. Cependant, afin de tester sa robustesse, il a été nécessaire de confronter cette méthode à un grand nombre de jeux de données. Ainsi, un générateur de polymères a ensuite été utilisé. Chaque monomère est alors représenté par un cylindre. Les cylindres peuvent être associés entre eux de différentes manières (Figure 3):

- Mode simple brin
- Double brin
- Branchés en forme de « Y »

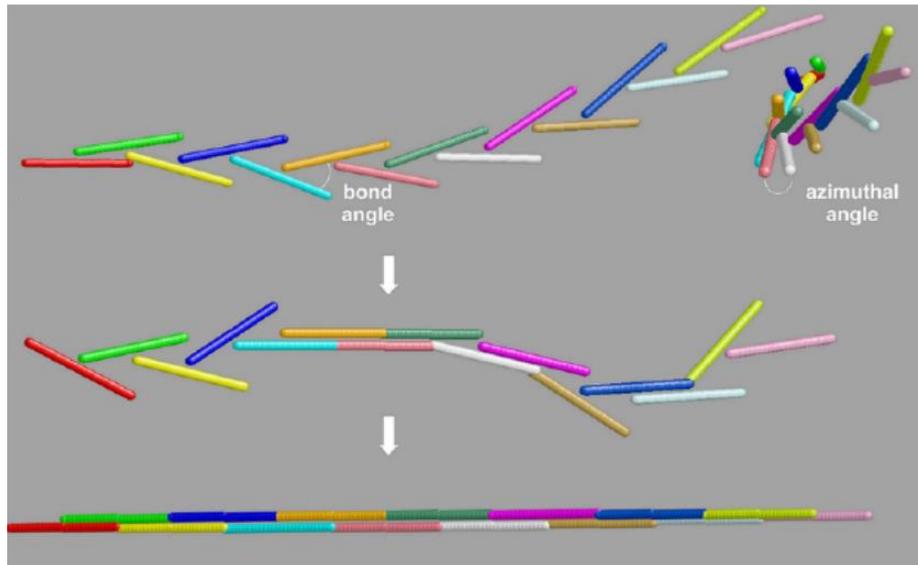


Figure 84 : Modes de polymérisation du fibrinogène. Haut : branchement des cylindres modèles en forme de "Y" avec un angle de liaison de 30°. La torsion est donnée par l'angle azymuthal. Milieu : les cylindres peuvent s'associer en mode « double brin ». Bas : la transition « Y » vers « double brin » est autorisée.

Chaque polymère est décrit par les coordonnées initiales et finales de chaque cylindre qui le compose. Afin de produire une distribution d'espèces de polymères, le programme utilise les statistiques de polycondensation bifonctionnelle de Flory régissant la formation des gels. A partir d'un ensemble défini de polymères, il est possible de générer grâce à des distributions théoriques, les paramètres $\langle R_g^2 \rangle_z$, R_c^2 et $\langle M \rangle_w$.

Une seconde méthode pour générer la courbe de diffusion donnée par l'ensemble des polymères est effectuée à partir des intensités diffusées en fonction de q pour des ensembles de cylindres connectés. Le jeu de données synthétique est généré en prenant en compte, pour chaque cylindre k d'un polymère composé de N monomères, le champ électrique diffusé $E_k(\theta)$ à tous les angles de diffusion, puis en sommant toutes les contributions provenant de tous les cylindres et en recréant le facteur de forme du polymère $P_N(\theta)$ en tant que carré du module normalisé du champ électrique total :

$$\text{Equation 9 : } P_N(\theta) = \frac{|\sum_k^N E_k(\theta)|^2}{|\sum_k^N E_k(0)|^2}$$

Pour chaque cylindre, $E_k(\theta)$ est calculé en prenant en compte son orientation par rapport à la direction de diffusion grâce à l'approximation de Rayleigh-Gans. Les facteurs de forme des polymères ayant même N sont alors moyennés en fonction de leurs orientations et de leurs configurations de manière à obtenir une estimation robuste statistiquement de $P_N(\theta)$. Enfin, le ratio de diffusion de Rayleigh total est calculé en prenant la moyenne sur z des facteurs de forme de tous les polymères en solution tel que :

$$\text{Equation 10 : } R(\theta) = K^* c \langle M \rangle_w \langle P(\theta) \rangle_z$$

Où $\langle P(\theta) \rangle_z = \sum_i^{N_{max}} w_i \cdot M_i \cdot P_i(\theta)$ et N_{max} est le nombre de monomère du plus grand polymère. La validation se fait alors en comparant les valeurs exactes de $\langle R_g^2 \rangle_z$ et

$\langle M \rangle_w$ à chaque instant (dérivé des polymères connus et utilisés pour calculer la courbe de diffusion pour chaque polymère), avec ceux dérivés de $I^*(q)$ obtenus grâce à la fonction de Fisher-Burford. Un bruit statistique est ajouté au préalable aux données de $I^*(q)$ synthétiques pour reproduire au mieux les conditions expérimentales théoriques.

Analyse des données de SAXS

L'analyse des courbes obtenues par SAXS sont faites à l'aide de la représentation de Guinier adaptée aux particules cylindriques. Cette représentation consiste à tracer $\ln(qI^*(q))$ en fonction de q^2 et permet d'évaluer le rayon de section moyen de la particule grâce à la formule :

$$\text{Equation 11 : } \ln \frac{qI^*(q)}{c} = \ln \frac{qI^*(0)}{c} - \left(\frac{R_c^2}{2}\right) \cdot q^2 \text{ pour } \frac{2\pi}{L} < q < \frac{1}{R_c}$$

Avec :

- q : vecteur de diffusion $q = 4\pi \frac{\sin \frac{\theta}{2}}{\lambda}$
- $I^*(q)$: Intensité de diffusion normalisée réduite en fonction du transfert de moment q
- $\langle R_c \rangle$: Section du rayon de giration de particules cylindriques $R_c^2 = \frac{r^2}{2}$ avec r étant le rayon géométrique de la particule étudiée
- c : Concentration de l'échantillon
- L : Longueur d'un monomère de l'espèce étudiée

Ainsi, les données de SAXS nous permettent d'extraire la section des filaments formés R_c^2 . Cependant, il n'est pas possible d'obtenir simultanément $\langle R_g^2 \rangle_z$ et $\langle M \rangle_w$ par SAXS au-delà d'oligomères de petites tailles puisque rapidement, les valeurs deviennent hors d'atteinte des capacités de mesure du système. Ainsi, dans le but de mesurer simultanément ces trois valeurs, il a fallu coupler la mesure par SAXS avec le système WA-MALS.

Titre : Analyse structurale des régions prédites comme dépliées de l'enveloppe nucléaire : exemples de l'émerine et de la lamine A.

Mots clés : Lamines, Résonance Magnétique Nucléaire, Protéines Intrinsèquement Désordonnées, Progéria

Résumé

Les lamines sont le principal composant du nucléosquelette. Elles sont principalement localisées à l'enveloppe nucléaire, où elles interagissent avec la membrane nucléaire interne, les protéines associées à la chromatine ainsi qu'avec des modulateurs de la signalisation cellulaire. Le gène *LMNA* code pour la prélamine A et la lamine C. La région C-terminale de la prélamine A est prédite pour être désordonnée et est la cible de plusieurs événements de maturation. En effet, la protéine est farnésylée, coupée, carboxyméthylée, puis coupée à nouveau ; perdant finalement son groupement farnésyl. Un mutant de cette protéine, dont 50 acides aminés sont manquants, est responsable du syndrome d'Huchinson-Gilford, appelé progéria (Eriksson et al., 2003). Chez ce mutant, appelé progérine, le site de coupure finale est absent et la protéine reste constitutivement farnésylée.

La lamine A est connue pour interagir avec la protéine de la membrane nucléaire interne, l'émerine. L'absence d'émerine est responsable de la dystrophie musculaire d'Emery Dreifuss. L'émerine contient un LEM, suivi d'une région prédite comme désordonnée, essentielle pour l'auto-assemblage de l'émerine (Berk et al., 2014). L'oligomérisation de l'émerine régule ses interactions avec plusieurs partenaires à la membrane nucléaire interne et à la chromatine. Nous avons auparavant démontré que la région nucléoplasmique de l'émerine peut s'auto-associer pour former des filaments *in vitro* (Herrada et al., 2015) et nous avons récemment révélé que ces filaments sont capables d'interagir directement avec la queue de la lamine A (Samson et al., 2018).

Ici, je me suis intéressé à l'analyse structurale des régions prédites comme désordonnées chez (1) l'émerine (2) la prélamine A. Dans le cas de l'émerine, j'ai analysé la conformation de la région nucléoplasmique d'émerine avant et après auto-assemblage, en travaillant avec l'émerine sauvage et plusieurs mutants entraînant des myopathies. J'ai montré que deux fragments de l'émerine 1-187 et 67-221 peuvent polymériser, tandis que leur région commune 67-187, reste toujours monomérique dans nos conditions (Samson et al., 2018). Nous avons aussi montré que le domaine LEM est au moins partiellement déplié au cours de l'assemblage de la région 1-187. J'ai également attribué les signaux RMN de la région désordonnée 67-170, dans le but d'étudier par la suite l'impact des phosphorylations de cette région sur la structure de l'émerine et sur ses propriétés d'auto-assemblage (Samson et al., 2016). Dans le cas de la lamine A, j'ai étudié la région C-terminale de la prélamine A, prédite comme dépliée et qui est le siège de nombreuses modifications post-traductionnelles. J'ai attribué les signaux RMN du peptide prélamine A ainsi que de son mutant progérine (Celli et al., 2018). J'ai montré que ces deux peptides sont en effet dépliés et possèdent une hélice α transitoire très conservée. Je propose cette hélice comme site de liaison pour un partenaire encore non identifié. J'ai également démontré que le peptide prélamine A possède une tendance à s'auto-assembler. Cependant, la prélamine A et le peptide progérine sauvages et farnésylés, n'interagissent pas avec le domaine IgFold de la lamine A ni avec BAF, deux domaines associés avec la progéria. J'ai étudié par la suite les interactions de ces peptides avec deux autres partenaires associés à la progéria : la protéine de la membrane nucléaire interne SUN1 et la protéine associée à la chromatine RBBP4. SUN1 est également intrinsèquement désordonnée et très peu soluble dans nos conditions. Les résultats montrent que le peptide prélamine A ne lie pas RBBP4 mais pourrait avoir besoin de la partie C-terminale qui la précède. Cependant, RBBP4 lie directement le partenaire de la lamine BAF. Sur les bases de ces résultats, je propose une série d'expériences pour identifier les détails moléculaires des interactions entre la queue C-terminale de la lamine A, BAF et RBBP4.

Université Paris-Saclay

Espace technologique / immeuble Discovery

Route de l'Orme aux Merisiers RD 128/91190 Saint-Aubin, France

Title : Structural analysis of nuclear envelope regions predicted as disordered : examples of emerin and lamin A.

Keywords : Lamins, Nuclear Magnetic Resonance, Intrinsically Disordered Proteins, Progeria

Abstract

Lamins are the main components of the nucleoskeleton. They are primarily located at the nuclear envelope, where they interact with inner nuclear membrane proteins, chromatin-associated proteins and cell signaling modulators. The *LMNA* gene codes for prelamin A and lamin C. The C-terminal region of prelamin A is predicted to be unfolded and is the target of several maturation events. Indeed, the protein is farnesylated, cleaved, carboxymethylated and cleaved again; losing eventually its farnesyl group. A mutant of this protein, lacking 50 amino acids, is responsible for the Hutchinson-Gilford Progeria Syndrome (Eriksson et al., *Nature* 2003). In this mutant, called progerin, the final cleavage site is absent and the protein stays constitutively farnesylated.

Lamin A is reported to interact with the inner nuclear membrane protein emerin. Lack of emerin is responsible for Emery Dreifuss Muscular Dystrophy. Emerin contains a folded LEM domain, followed by a region that is predicted to be disordered and is essential for emerin self-assembly (Berk et al., 2014). Emerin oligomerization regulates its interaction with several partners at the inner nuclear membrane and at the chromatin. We previously showed that the nucleoplasmic region of emerin can self-assemble to form curvilinear filaments *in vitro* (Herrada et al., 2015) and we recently revealed that these filaments are able to directly bind to the lamin A tail (Samson et al., 2018).

Here I focused on the structural analysis of regions that are predicted to be unfolded in (1) emerin, (2) prelamin A. In the case of emerin, I analysed the conformation of the nucleoplasmic region of emerin before and after self-assembly, working on wild-type emerin as well as several mutants causing myopathies. I showed that the two fragments of emerin 1-187 and 67-221 were able to self-assemble, whereas their common region, 67-187, is always a monomer in our conditions (Samson et al., 2018). I also revealed that the LEM domain is at least partially unfolded during self-assembly of region 1-187, as a mutant with a destabilized LEM domain self-assembles faster and a mutant with a LEM domain locked in its folded conformation cannot self-assemble (Samson et al., 2017). I also assigned all the NMR signals of the unfolded region 67-170, in order to further study by NMR the impact of phosphorylation of this region on emerin structure and self-assembly properties (Samson et al., 2016). In the case of lamin A, I studied the C-terminal region of prelamin A that is predicted as unfolded and is heavily post-translationally modified. I assigned the NMR signals of this prelamin A peptide as well as its mutant peptide corresponding to the progerin sequence (Celli et al., 2018). I showed that both peptides are indeed unstructured and exhibit a partially populated α helix that has a highly conserved sequence. I propose that this helix is a binding site for a yet unidentified partner. I also revealed that the prelamin A peptide has a tendency to self-assemble. However, the monomeric prelamin A and progerin peptides, wild-type as well as farnesylated, do not interact with the immunoglobulin-like domain of lamin A/C and with BAF, two domains associated with progeria. Then, I investigated the interactions mediated by these peptides and two other important partners associated to progeria: the inner nuclear membrane SUN1 and the chromatin-associated protein RBBP4. However, SUN1 is also intrinsically disordered and poorly soluble in our conditions. First results showed that the prelamin peptide does not bind to RBBP4 but might need the remaining part of the lamin A tail for this interaction. However, RBBP4 directly binds to the lamin partner BAF. Based on my results, I propose a set of experiments to identify the molecular details of the interactions between the lamin A tail, BAF and RBBP4.

Université Paris-Saclay

Espace technologique / immeuble Discovery

Route de l'Orme aux Merisiers RD 128/91190 Saint-Aubin, France

